University College London

UCL

Biocatalysis Using Plant and Metagenomic Enzymes for Organic Synthesis

Sophie Alice Newgas

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

2018

Declaration

I, Sophie Alice Newgas, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Dated:

Abstract

Biocatalysts provide an excellent alternative to traditional organic chemistry strategies, with advantages such as mild reaction conditions and high enantio- and stereoselectivities. The use of metagenomics has enabled new enzymes to be sourced with high sequence diversity. At UCL a metagenomics strategy has been developed for enzyme discovery, in which the library generated is annotated and searched for desired enzyme sequences.

In this PhD, a metagenomic approach was used to retrieve 37 short chain reductase/dehydrogenases (SDRs) from an oral environment metagenome. Eight enzymes displayed activity towards cyclohexanone and their substrate selectivities were investigated. Four of the SDRs displayed activity to the Wieland-Miescher ketone (WMK), a motif found in several pharmaceutically relevant compounds. SDR-17 displayed high conversions and stereoselectivities and was co-expressed with the co-factor recycling enzyme glucose-6-phosphate dehydrogenase. This system was then successfully used to reduce (R)-WMK on a preparative scale reaction in 89% isolated yield and >99% e.e..

In further studies using reductases, the substrate specificities of two ketoreductases known as tropinone reductase I and II (TRI and TRII respectively) from the plant *D. stramonium* and MecgoR from *E. coca* were investigated. These studies expanded on reported substrate activities with these enzymes in the literature. A selection of symmetric and asymmetric tropinone analogues were synthesised, towards which MecgoR and TRI showed high activities, providing a strategy to access novel alcohols. Furthermore, sixteen ketoreductases were selected from a drain metagenome based on their sequence similarity of over 24% to MecgoR. They were annotated as aldo/keto reductases (ARKs) and five were successfully expressed in *E. coli*. Interestingly, the novel enzyme AKR-3 displayed activities toward aromatic ketones and aldehydes such as 2-indanone, phenylacetaldehyde and benzaldehyde.

Transaminases (TAms) from the enzyme library toolbox at UCL were also tested with tropinone analogues and related cyclic compounds, several of which showed good activities.

Impact Statement

This PhD project explores on the use of biocatalysis as an environmentally friendly alternative to the use of traditional chemical catalysts. The use of enzymes to perform chemical reactions is a sustainable alternative because the method results in low energy consumption and waste production. The reactions can be highly efficient with excellent selectivity, and the use of enzymes is both renewable and may be less toxic than alternatives. Because of this, there has been a recent increase in interest in this sustainable topic.

The use of enzymes to perform organic chemical reactions relies on the availability of enzymes which accept the targeted molecule of interest. Searching for novel enzymes is critical both inside academia, as well as in industry, for the sustainable production of pharmaceuticals and fine chemicals.

This thesis addresses this issue by mining a metagenomic database to discover enzymes. From the identified enzymes, several were identified as performing useful reactions with a selection of substrates. Highlighting the industrial applicability of this work, one of these enzymes was scaled up to a preparative scale with a particularly useful molecule, the Wieland Miescher Ketone, which is a motif found in a wide selection of pharmaceutical drugs. A co-factor recycling system was employed to achieve excellent conversions and stereoselectivities.

This research also utilises plant enzymes to efficiently perform biocatalytic reactions, with a focus on the use of the bicyclic natural product tropinone as a substrate. The tropinone structure is a motif that is found in a wide selection of pharmaceutical and recreational drugs and the tropinone alkaloid family is known for its anticholinergic and chemotherapeutic properties, with applications ranging from treatments for Parkinson's disease to intestinal tract problems. Plant enzymes were employed to perform reactions on a selection of embellished tropinone substrates and accomplished high conversions.

With a focus on the interface between biology and chemistry, a broad range of techniques were utilised including organic synthesis, molecular biology, analytical

chemistry and biochemical approaches. This research has been presented at multiple conferences, both nationally and internationally, and will be used by other academics working in the discipline. The research presented in this thesis overcomes challenges associated with reduction and transamination reactions by utilising biocatalysis and contributes to the development of sustainable processes within organic chemistry.

Acknowledgements

I am very grateful to my fantastic supervisor Professor Helen Hailes for all her guidance, support and teaching. She has helped me every step of the way during my PhD, and has been so encouraging. I also would like to thank Professor John Ward for all his help and ideas and also being an excellent supervisor. Both Helen and John have created a wonderful lab group and I couldn't have asked better supervisors.

I also would like to thank other who have helped me with this work, including Jack Jeffries, Maria Bawn, Damien Baud, Dragana Dobrijevic, Fiona Truscott, Nadine Ladkau, Fabiana Subrizi, and Dani Mendez-Sanchez. They have taught me so much day to day in the lab and have provided so much advice.

My fellow PhD students including Leona, Alice, Rachael, and Harriet have been fantastic friends, and have supported me during this PhD. I couldn't have done it without you guys. My whole lab group has been amazing, so I want to express my gratitude to everyone in the group. In addition, all the PhD students and PDRAs in John's group have also been amazing, and have taught me so much.

I also thank all the other friends and family from outside UCL who have been so important to me over that last few years. My parents, John and Celia, and my brothers, Adam and David, have helped me an indescribable amount and I am so grateful. My housemates Inez and Amy have also put up with me for three years and have been there for all the highs and lows. I'm also so grateful to my RSY friends, as well as all my school friends for their support and continuous encouragement. My friends from both Warwick University and those that I have made whilst at UCL have also been made these last three years unforgettable.

Abbreviations

2,5-DMTHF	Dimethoxy tetrahydrofuran
AcOH	Acetic acid
ACP	Acyl-carrier-protein
ADH	Alcohol dehydrogenase
AKR	Aldo/keto reductase
Aq.	Aqueous
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BVMO	Baeyer-Villager Monooxygenases
°C	Degrees centigrade
CBT	Carbomethoxytropinone
CCL	Clarified cell lysate
$CDCl_3$	Deuterated chloroform
CH_2Cl_2	Dichloromethane
$CHCl_3$	Chloroform
Contig	Contiguous sequences
d.e.	Diastereomeric excess
Da	Dalton
DMAP	4-(Dimethylamino)pyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
e.e.	Enantiomeric excess
EC	Enzyme Commission
Et	Ethyl
Et_2O	Diethyl ether
Et_3N	Triethylamine
EtOAC	Ethyl acetate
EtOH	Ethanol
FID	Flame ionisation detector

G6P	Glucose-6-phosphate
G6PDH	Glucose-6-phosphate dehydrogenase
GC	Gas Chromatography
GDH	Glucose dehydrogenase
h	Hour(s)
His-tag	Hexahistidine tag
HPK	Hajos-Parrish ketone
HPLC	High Performance Liquid Chromatography
HSD	Hydroxysteroid dehydrogenase
Hz	Hertz
IDABO	8,8- Dimethyl-3-oxo-8-azonia-bicyclo[3.2.1]octane iodide
$iPrNH_2$	Isopropylamine
iPrOH	Isopropanol or 2-propanol
KRED	Ketoreductase
LB	Lysogeny broth
LC-MS	Liquid chromotography mass spectrometry
LDA	Lithium diisopropylamide
LEH	Limonine-1,2-epoxide hydrolases
L-Orn	L-Ornithine
MADO	(1R,6S)-10-methyl-10-azabicyclo[4.3.1]decan-8-one
MBA	Methylbenzylamine
Me	Methyl
MeOH	Methanol
mins	Minute(s)
mol	Mole(s)
Mp	Melting point
MS	Mass spectroscopy
Mw	Molecular weight
n.d.	Not determined
n/a	Not applicable
$NAD(P)^+$	Nicotinamide adenine dinucleotide (phosphate)

NAD(P)H	Reduced nicotinamide adenine dinucleotide (phosphate)
n-BuLi	n-Butyllithium
NCBI	National Centre of Biotechnology Information
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein Data Bank
Pet. Ether	Petroleum ether
Pi	Phosphate buffer
PLE	Porcine liver esterase
PLP	Pyridoxal 5'-phosphate
PMP	Pyridoxamine 5' phosphate
Ppm	Parts per million
Pr	Propyl
R	Generic alkyl group
Rf	Retention factor
Rt	Retention time
RT	Room temperature
s	Second(s)
Sat.	Saturated
SDR	Short chain dehydrogenase/reductase
TAm	Transaminase
TB	Terrific broth
TBON	8-Thiabicyclo[3.2.1.]-octan-3-one
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TK	Transketolase
TLC	Thin layer chromatography
TMS	Tetramethylsilane
TR	Tropinone reductase
TRI	Tropinone reductase 1

TRII	Tropinone reductase 2
TRL	Tropinone reductase-like SDR
UV	Ultra violet
WMK	Wieland-Miescher ketone

Table of Contents

1 Introduction

1.1 Biocatalysts

1.1.1 What biocatalysts are

The use of enzymes in biocatalytic reactions has been developed over the past century, with an expansion over the past 25 years in the selection of biocatalysts available for organic synthesis. Early biocatalysts were commonly whole-cell yeast strains, which suffered from insufficient activity and enantioselectivity, slow growth, significant levels of unwanted side reactions and by-product or substrate inhibition.[1] Additionally, the use of specialised equipment such as fermenters to grow the organism containing the biocatalysts meant they were not generally accepted as a viable synthetic method in the chemical industry. These obstacles have been lessened by increased usage and publications on successful apparatus, as well as the introduction of techniques such as recombinant DNA technology and the use of cofactor regeneration systems for those classes of enzymes that needed cofactors.

Biocatalysts have been used in reactions in three ways: whole-cell biocatalysts, cell-free extracts and isolated enzymes, as well as in solution or in an immobilised format. Whole-cell biocatalysts are implemented without destruction of the cells in which the enzyme was expressed. This has several advantages, such as the availability of the endogenous co-factors and metabolic pathways, and the enzymes and co-factors are protected in their natural environment, making the reaction more stable.[2] However, many interesting substrates are toxic to the wild-type living organism, solubility may be an issue and only low substrate concentrations are tolerated. Additionally, the high levels of by-products complicates product purification.[2] Alternatively, biocatalyst-containing organisms can be lysed, and the cell-debris removed *via* centrifugation or similar technique leaving the cell-free extract, or clarified cell lysate (CCL). This allows for adaptation of the reaction conditions and higher enzyme concentrations. However, this adds another stage to biocatalyst preparation which ideally would be avoided in an industrial setting. Another disadvantage is the associated issues with other enzymes and endotoxins present in the lysate, which can affect the reaction. The enzymes may also be isolated so the enzyme and reactions can be investigated in more detail as a known concentration of pure enzyme is used.

1.1.2 Advantages and disadvantages of biocatalysis

The benefits of using biocatalysts compared to traditional organic chemistry strategies include mild reaction conditions, excellent enzymatic regio- and enantioselectivities and no requirements for: high-pressure hydrogen apparatus, use and removal of expensive and often toxic transition metals, use of organic solvents, or use of extreme temperatures.[3] Additionally, use of a biocatalyst can reduce the number of necessary steps in a compound's synthesis. These advantages can be exploited in the pharmaceutical sector as well as in novel routes to fine chemicals. Biocatalysts can minimise the use of difficult synthetic procedures which may include complex protecting group strategies, lengthy purification procedures using solvents and expensive catalysts.

Limitations when using biocatalysts include equipment cost of fermenters or other apparatus, low final product concentrations, removal of water and complex downstream processing methods. There is also a limited range of reactions that are covered by biocatalysts, and the enzymes themselves may suffer from low organic solvent tolerances, narrow working pH ranges, and low thermostability.[4] Scaling the biocatalysts to an industrial scale can also prove challenging, for example, compound and catalyst degradation is an issue in batch operations.[5] Techniques that allow for enzyme recycling, such as immobilisation, are therefore important. A further consideration is that biocatalytic processes that take place late in a product synthetic route may require the removal of unwanted biological material such as nucleic acids and endoxin.[5]

Protein engineering has been used to overcome some of these challenges, focusing initially on the rational design of the protein structure in the 1980s. As the primary sequence encodes the protein and affect protein characteristics, directed evolution can be used to modify the characteristics of the protein by introduction mutations (Figure 1). Common strategies for introducing mutations include: error prone PCR (introducing random mutation throughout the gene); saturation mutagenesis (using degenerate primers to substitute some or all codons at chosen sites within the gene); DNA shuffling (using restriction enzymes or DNAse to fragment homologous parent genes and then recombining the genes using PCR and

overlapping homologous regions).[6] In recent years, molecular biology techniques have also advanced - especially in areas of active site modelling and docking, rational design and mutagenesis.

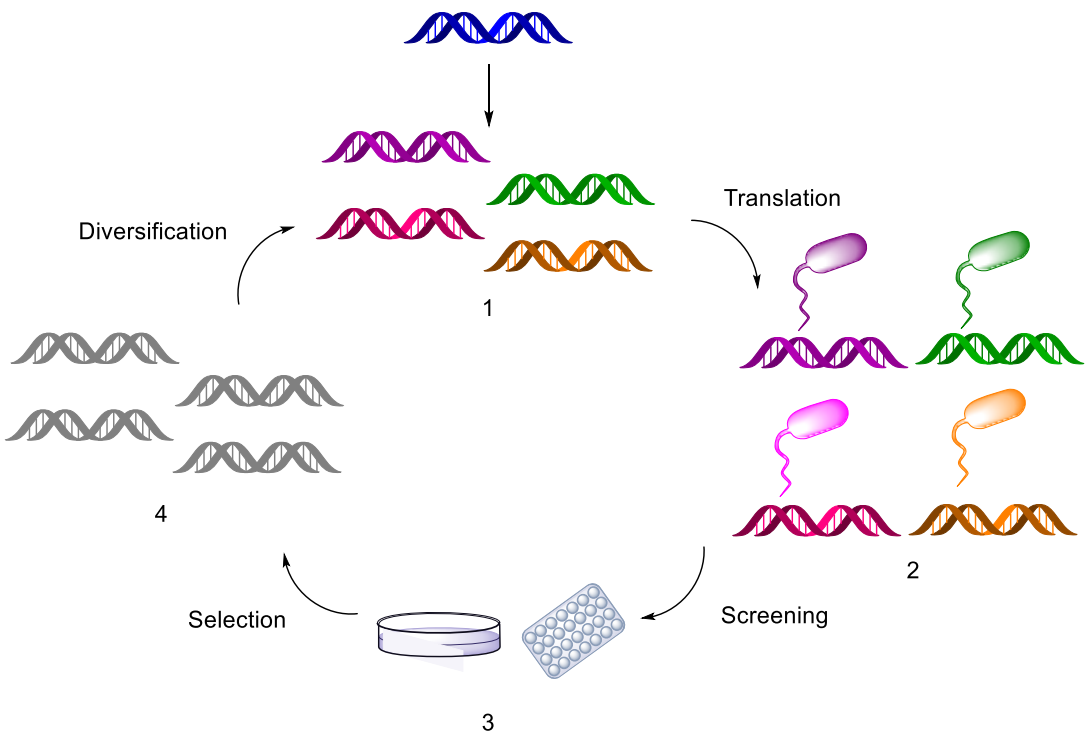


Figure 1: Cycle of directed evolution.[6] 1: Mutation of wild type sequence generates a pool of diverse sequences. 2: Mutated sequences are put back into bacteria where translation of sequences into protein allows screening for desired characteristics. 3: Selected mutants are amplified. 4: Mutant containing bacteria are ready to restart the cycle.

1.1.3 Enzyme discovery

Enzyme engineering relies on sequence diversity and subsequent selection (Figure 1). When developing an enzyme to do a chemical transformation, it is beneficial to have many examples of the enzyme in question (homologues), which vary in the primary sequence. This provides a wider range of starting characteristics to evolve and increasing the chances of generating mutants with the desired characteristic. This can be achieved one of two ways: function- or sequence- based.

Function based discovery relies on the observation of the desired functionality by screening thousands of clones. This method does not rely on prior knowledge and has the potential to reveal completely unrelated sequences and novel enzyme classes. However, having identified activity, the sequence must then be

retrieved, which can be challenging. Expression and purification of the enzyme may be demanding, as the enzyme of interest may not work in isolation.[7] Enrichment cultivation is a strategy by which the growth of a particular microorganism is favoured by application of a specific growth medium, to select for a particular characteristic. For example, using an amine as the sole nitrogen source has been employed to identify new transaminases (TAms), such as the TAm from *Vibrio fluvialis* JS17 (Vf-TAm), as discussed in Section 1.6.3.1.[8,9]

In sequence-based discovery the whole genome of the organism is sequenced, the gene(s) of interest identified and then cloned into a suitable expression vector and assayed. Function can be inferred from the primary sequence, and the development of sequencing techniques means enzymes with related primary sequences can be identified on large databases. For example, *Chromobacterium violaceum* 2025 TAm (Cv-TAm, Section 1.6.3.2) was identified through its 38% sequence identity to Vf-TAm and both portray high activities towards aromatic compounds.[10] Motif based searching can be used for function prediction or screening and is described in Section 1.1.5. The inherent process of sequence-based discovery limits novel hits to known functional classes.

1.1.4 Metagenomics

Metagenomics is the analysis of organisms *via* the genomic DNA from a whole community as opposed to the analysis of organism by culturing and growing organisms in the laboratory.[11] The term was first used in 1998 when describing a method to analyse DNA extracted from soil environments.[12] Less than 1% of microbes can be cultivated in the laboratory due to slow growth rates, unknown optimal conditions or other organisms required to provide nutrients.[13,14] A clonal culture will fail to represent this, the population variety and biological functions.[13]

Sequencing and *in silico* strategies provide opportunities to further expand the selection of known enzymes, using isolated microorganisms and the sequencing of their genomes in the laboratory. The drastic reduction in the cost of sequencing has enabled large growth in this area of research and an increased selection of enzymes could help overcome some of the limitations of biocatalysts mentioned above (Section 1.1.2).[11] Environmental metagenomic sequencing has limitations

because the sequencing techniques produce short fragments which are difficult to assemble into large meaningful contiguous DNA sequences (contigs). This difficulty in the assembly of short fragments increases with the diversity and number of species in a single sample, where the full genome is not available for homologous species. The volume of data is much larger than that for single organism sequencing[13].

A research focus at UCL is the development of metagenomics techniques to expand the range of enzymes that catalyse the same reaction with differing primary sequences, in order to more effectively engineer biocatalytic reactions. Current work has created a library of novel enzymes including ketoreductases (KREDs), transaminases (TAms) and transketolases (TKs).

1.1.5 Metagenomics project at UCL

Recent work at UCL has developed a novel genome mining strategy, which differs from other techniques due to several of the initial stages being performed *in silico* as opposed to *in vitro* (Figure 2). The strategy can be summarised as follows: the complete DNA genome is extracted from an environmental sample and sequenced to yield the corresponding metagenome library of contiguous (contig) reads. This is formatted into a library, which can then be accessed by using query sequences with BLAST, PFam, CATH or Uniprot interrogations.[15–17] BLAST searches hold the advantage that the library may be made into an easily searchable BLAST library, and may be quickly interrogated.[18] CATH and PFam databases classify enzymes differently, and thus potentially some different sequences can be pulled out from the library. There will be a large overlapping set of sequencing that these techniques identify. Primers can be designed for the ORFs and thus the DNA sequences can be amplified by PCR and cloned into an expression system. A function based assay can then be used to assess a specific activity.[19]

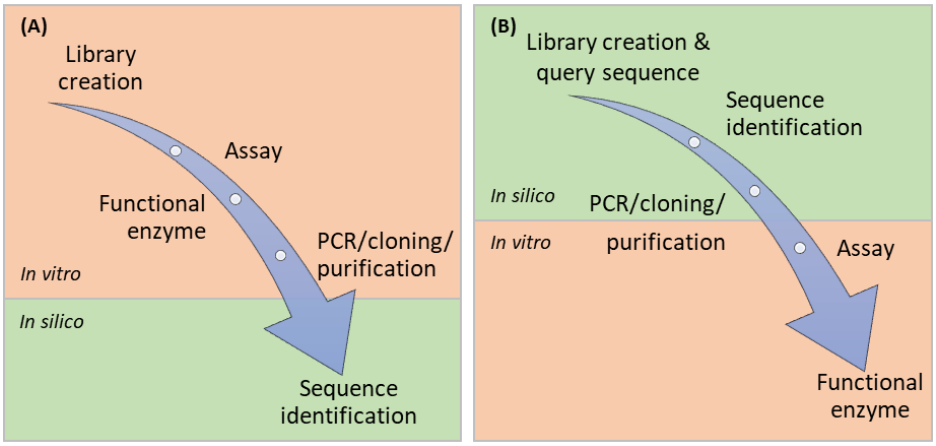


Figure 2: Schematic representation of (a) existing functional metagenomics strategies and (b) the strategy developed at UCL.[19] Traditionally, the majority of work is done *in vitro* and requires more time and resources. The strategy from UCL shifts the focus to *in silico* work, and aims to select enzymes on this basis. Figure adapted from Jeffries *et al*..[19]

1.1.5.1 Tongue metagenome

Work using metagenomics to discover new enzymes from the oral microbiome began in 2009.[20,21] Bacterial DNA from tongue scrapings from 9 volunteers was extracted and sequenced using the Roche 454 Titanium platform, producing 1.1 million individual sequences of length ~400 base pairs (bp). To ensure high quality data, corrupt sequences, reverse complement duplicates and human sequences were removed. The sequences were then assembled to make up contigs when they had at least a 50 base pair overlap and a similarity score of 98%. This generated a total of 39,971 contigs, ranging from 104 bp to 41.9 kb, which were formatted into a searchable BLAST database (http://gene3d.biochem.ucl.ac.uk/TongueBlast/).

The contig library was also scanned with the GeneMark tool to predict open reading frames (ORFs) and predicted 125, 211 ORFs.[22] These ORFs were then scanned with the Pfam standalone tool (Dr Natalie Dawson) to mark them with an ID.[15] This generated 88, 247 annotations, belonging to 3702 protein classes.

The Roche 454 system was the earliest commercially available first-generation sequencing platform. A disadvantage of this system is that the length of homopolymers must be inferred from signal intensity, which can lead to insertion and deletion errors. The system however can generate longer read lengths compared to other platforms.[23]

1.1.5.2 Drain Metagenome

DNA was also extracted from a residential drain and was sequenced using the Illumina Miseq platform to generate 10 million individual reads. This went through a similar process to the tongue metagenome and was assembled into contigs which were also scanned with the Pfam standalone tool (Dr Natalie Dawson & Dr Dragana Dobrijevic).[15]

The Illumina technology differs from the 454 system, by using bridge PCR and reversible termination chemistry. Illumina sequencing is more accurate than the Roche 454 platform, and generates more data. However, the read length is more limited, with early versions of the technology giving lengths of around 36 base pairs and longer reads having a higher error rate. The later version of Illumina technology used here give reads of approximately 140 base pairs and a greater depth of coverage.

1.1.6 Other metagenomics research

Research with metagenomic have been growing in recent years. An example of this is the Hotzyme project which has collected samples from hot springs in 13 environments ranging in temperature from 46 to 92 °C and pH from 1.8 to 10.2.[24] With similar methodology to that used at UCL, the DNA was sequenced using Roche/454 Titanium FLX or Illumina HiSeq plateforms, assembled into contigs and submitted to the programme BLAST. This was then used for *in silico* screening with two known limonine-1,2-epoxide hydrolases (LEHs) from *Rhodococcus erythropolis* and *Mycobacterium tuberculosis* (GenBank Q9ZAG3.3 and CCP45539.1). This identified two novel LEHs. These displayed different substrate and stereoselectivity to the previously reported LEH from *Rhodococcus erythropolis* (Re-LEH), as well as higher optimal and melting temperatures. [24]

Popovic *et al.* also reported the screening of metagenomic libraries from 16 different environments. This generated 714 positive hits, 80 of which displayed validated esterase activity, and four of which were characterised biochemically.[25]

1.2 Ketoreductases

1.2.1 Classification and past work

Ketoreductases (KREDs) refer to a broad class of enzymes which can catalyse the generation of a hydroxy group from a carbonyl moiety by addition of a hydride. These versatile enzymes can also catalyse the opposite reaction yielding the corresponding carbonyl derivatives depending on the reaction conditions.[26] These enzymes have been investigated since the 1970's and are widespread in organisms including mammals, plants, bacteria, yeasts, fish and insects. In these organisms, they have important anabolic and catabolic reactions.[26]

1.2.2 Short chain dehydrogenase/reductases

One subset of KREDs is the superfamily short chain dehydrogenase/reductases (SDRs) which first identified in the 1980s by sequence analysis of insect, yeast and mammalian alcohol dehydrogenases (ADHs).[27] Other NAD(P)(H) dependant oxidoreductases can belong to long chain (600-750 amino acids) or medium chain (~ 350 amino acids) dehydrogenases, and are defined by chain length, sequence motifs, and structural comparisons.[27–29] SDRs are characterised by a conserved Rossmann fold, consisting of a twisted β-sheet flanked by 3 to 4 α-helices on either side.[30] Residue identity is commonly low at approximately 15-30%; sequence homology is lower towards the C-terminal, and it has been suggested that this is involved with substrate specificity.[28] However the folds are commonly similar in three dimensions.[31] SDRs are commonly 250 to 300 residues in length and have wide functional diversity including isomerization, decarboxylation, epimerization, C=N bond reduction, dehydratase activity, dehalogenation, enoyl-CoA reduction, and carbonyl-alcohol oxidoreduction.[32] Substrates for these enzymes include sugars, steroids, alcohols, and aromatic compounds.[30,31]

The SDRs can be divided in seven families (Table 1), two of which are: classical and extended - defined by differing Gly-motifs in the co-enzyme binding region and chain lengths (250 and 350 residues respectively).[31] The SDRs have been regrouped into several types using different systems, such as sequence identities or structural

motifs.[30] SDRs can be annotated using the Pfam ID system as adh_short (PF00106) and adh_short_C2 (PF13561) (also known as enoyl-(acyl carrier protein) reductase).[15]

Human 15-hydroxyprostaglandin dehydrogenase (15-PGDH) is commonly used as an example SDR to number characteristic residues. For example. Tyr-151 is a critical catalytic residue and is often found in a conserved YxxxK motif. The active site is also often constructed of an additional upstream Ser-138 and/or Asn-107 or another Ser in complex SDRs[29]

SDR Type	One-letter title	No. of SDR forms	No. of SDR families	Example activity	Sequence patterns
Classical	C	67762	285	Oxidoreductases	TGxxxGxG for NAD(P)(H)
Extended	E	37604	81	Epimerases & Dehydratases	TGxxGxxG for NAD(P)(H)
Atypical	A	419	2	NmrA-like	Lacks catalytic Tyr
Intermediate	I	445	1	Drosophila ADH	G/AxxGxxG/A
Divergent	D	3415	4	Enoyl reductases	GxxxxxSxA for NAD(P)(H) YxxMxxxK altered active site
Complex	X	3976	15	β-Ketoacyl reductases	GGXGXXG for NAD(P)(H) YxxxN altered active site
Unassigned	U	9319	76	n/a	n/a

Table 1: The seven SDR types and their typical characteristics.[30]

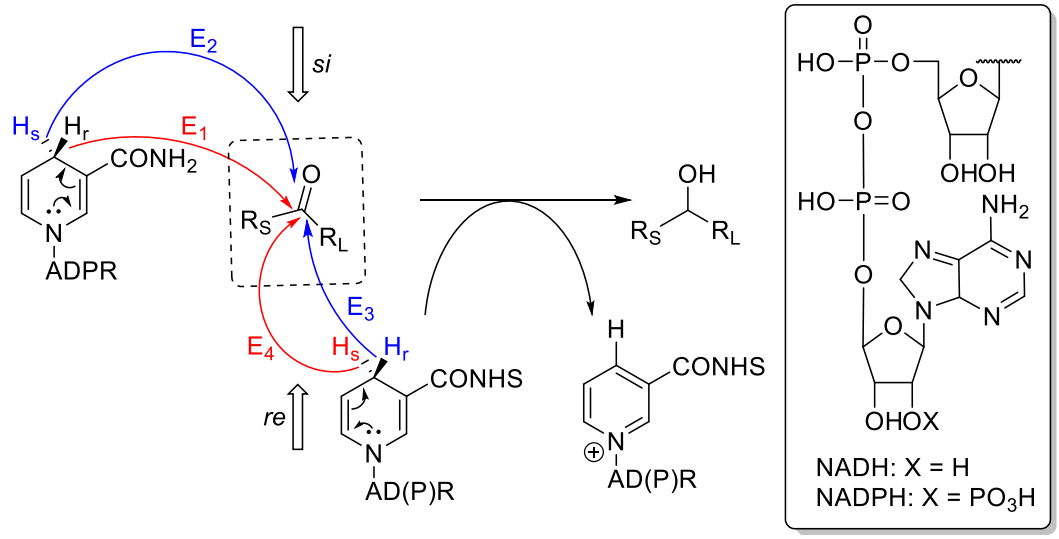
1.2.3 Aldo/keto reductases

AKRs are typically 320 residues in length, and mainly soluble monomeric proteins of 34-37 kDa. There are 16 AKR families, with sequence identities of over 60% defining proteins in each family. In plants, AKRs can be organised into four functional groups and may be annotated as PF00248 in the Pfam ID system.[15] AKRs catalyse a sequentially ordered bi-bi reaction: the cofactor binds first and leaves last. The cofactor is bound in an extended *anti*-conformation, so that the pro-*R*-hydride can be transferred (Scheme 1). AKRs are characterised by a triose-phosphate isomerase TIM barrel or an $(\alpha/\beta)_8$-barrel with additional α-helices inserted.[33] Loops at the rear of the barrel define substrate specificity and there is a conserved catalytic tetrad: Tyr, Lys, His and Asp.[33] Pi-stacking and H-bonding holds the NADP(H) head group in the correct orientation, and electrostatic forces hold the 2'phosphate group.[33] In 2007, there were 12 known KRED enzymes that belonged to both the AKR and SDR superfamilies.

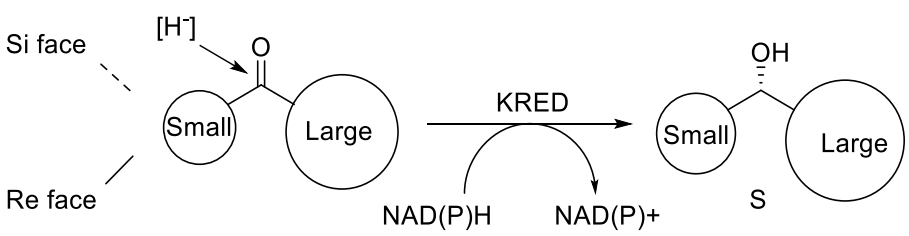
1.2.4 Mechanism

The reduction reactions commonly proceed as follows: the enzyme-reduced coenzyme complex binds the ketone substrate, hydride is transferred from the reduced cofactor NAD(P)H to the ketone, the alcohol product is released from the enzyme-oxidised cofactor complex and the oxidised cofactor then uncouples from the enzyme and may be reduced *via* a regeneration system.[1]

Hydride transfer may follow one of four stereochemical patterns. E1 and E2 adds to the *si* face and E3 and E4 add to the *re* face of the hybridised C=O moiety and generate the (*R*)- or (*S*)-alcohol from the pro-*R* or pro-*S* hydride of NAD(P)H respectively (Scheme 1). There are many examples of these mechanisms in the literature, though E4 is much less common.[1,34] The reduction of the ketones to the (*S*)-enantiomer is the most frequent stereochemical outcome, thus following Prelog's rules, which predicts the enantiomer generated based on the size of the substituents on the carbonyl group (Scheme 2).[2] This is knowns as following "Prelog's rule. Conversely, hydride transfer through the other face generates an alcohol with (R)-configuration, and the ADH is said to be "anti-Prelog".



Scheme 1: Hydride transfer mechanisms from NAD(P)H to ketone substrate. R_S = small sized substitution group, R_L = medium/large sized substitution group.[1] AD(P)R is highlighted in the box.



Scheme 2: Scheme representing asymmetric reduction according to Prelog's rule, assuming that the order of the "large" substituent is greater than that of the "small".[2,35]

1.2.5 Reductases used in biocatalysis

1.2.5.1 Chemical reductions

In traditional organic synthesis, chiral reagents such as chiral aluminium hydrides[36] and chiral boranes[37] may be used to perform reduction reactions. Although these sometimes use mild conditions, they are not atom economical, and can be expensive. Chiral auxiliaries have also been used to improve enantioselectivies in reduction reactions, but add extra steps in a synthetic route and require cleavage conditions that do not damage the product. In terms of chemo-catalysis, there are two methods to perform asymmetric reduction reactions: the use of metals and the use of organo-catalysts. Boranes and borohydrides are common catalysts and developments in this area have generated new homogeneous and immobilised catalysts which achieve high enantioselectivities and/or may be recoverable.[38] However biocatalysts offer an environmentally friendly alternative (Section 1.1.2).

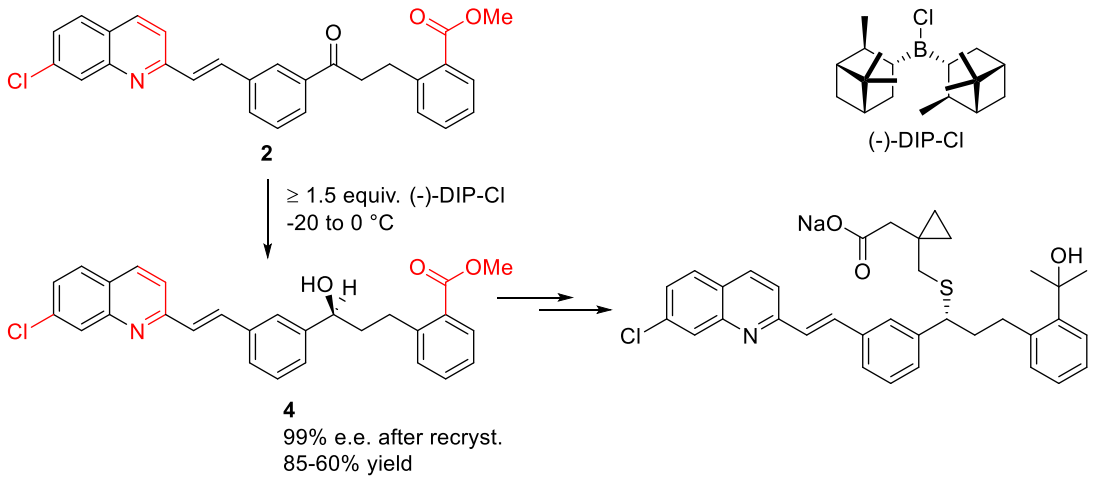
1.2.5.2 Use of biocatalysts

In early biocatalytic work, ketone reductions to optically active chiral alcohols predominantly used Baker's yeast, which contains various KREDs, or microbial enzymes.[39] Although whole cell catalysts have been implemented inexpensively on a large scale, the factors that regulate expression of the ketoreductases are poorly understood.[40] Therefore, expensive fermentation optimisation was required to consistently produce the enzyme.[40] Currently, there are more than 100 recombinant ketoreductases commercially available. Accessibility has been enhanced by the improved technologies in modern enzyme development and screening tools, for example, directed evolution and metagenome screening.[1] The use of metagenomics is therefore an opportunity to expand the selection of off-the-shelf enzyme libraries.

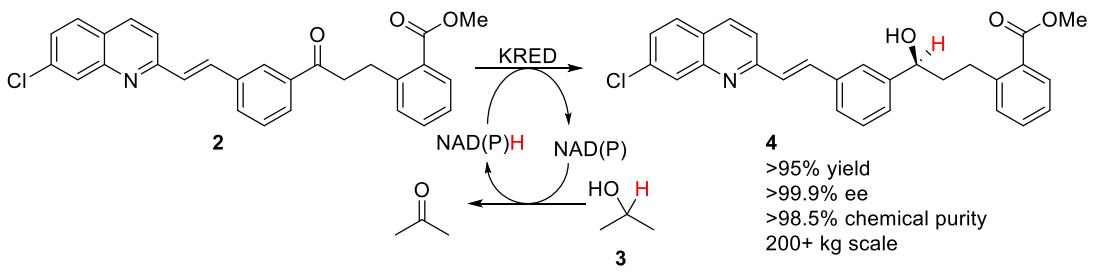
Substrate specificities for the KRED family of enzymes are not completely understood, but they accept a broad range of substrates.[41] A very wide variety of substrates have been tested against a number of KREDs. Bulky substrates tend to be poorly accepted by most known alcohol dehydrogenases, resulting in low activities.[1]

1.2.6 Montelukast case study

An excellent example of the industrial applicability of biocatalysis is in the synthesis of Montelukast sodium salt, marketed as *Sinulair* by Merck-Codexis **1**.[42] The drug is a leukotriene receptor antagonist that controls the symptoms of asthma and allergies with one stereocentre. The mild and selective reagent (-)-B-chlorodiisopinocampheylborane ((-)-DIP-Cl) was used by Merck to reduce the ketone to an alcohol, which was necessary due to the sensitive functional groups found on **2** (labelled red on Scheme 3). Starting with a commercially available KRED, three rounds of directed evolution led to a variant that was able to withstand the high substrate, product and organic solvent concentrations (Scheme 4).[42] As **2** is highly hydrophobic, a suitable solvent system was a ternary system of 1:5:3 toluene/isopropanol (**3**)/buffer was employed. This has many advantages: **3** acts as the hydride donor making an efficient recycling system, (*S*)-**4** readily crystallises, the reaction is "self-driven" to >99% conversion. Implementation of the KRED allowed for a more efficient, environmentally friendly and stereospecific route to the target compound.



Scheme 3: Formation of the alcohol on a montelukast intermediate 2 using (-)-DIP-Cl. Sensitive functional groups have been highlighted in red.[42]



Scheme 4: Formation of the alcohol on the montelukast intermediate 2 using a biocatalytic technique. [42]

1.2.7 Assay techniques

Chiral assay methods for KRED activity include the use of chiral stationary phases on HPLC or GC formats.[43] These methods are accurate to within 1% and can be used to measure both enantiomeric excess (e.e.) and reaction conversions. However, each sample assay typically takes around 5-20 minutes to run and optimal peak separation conditions need to be determined, which may make the retention times longer. The collection of rate data requires sampling at multiple time points, which is time intensive. Therefore there has been much work in the area of rapid assay techniques such as fluorescence reaction microarrays[44], infrared thermography[45], mass spectrometry[46], capillary electrophoresis[47] and pH indicators[48].

1.2.8 Recycling systems for NAD(P)H

A wide variety of cofactor recycling systems (Figure 3) have been used with ketoreductases due to the purchase price of NADH and NADPH.[1] A coupled enzyme approach may be employed for regeneration of NADH as well as NADPH (Figure 3A).[1] Enzymes have included hydrogenases, formate dehydrogenases, aldehyde dehydrogenases, glucose-6-phosphate dehydrogenases (G6PDH), phosphite dehydrogenases and glucose dehydrogenases (GDH). One benefit is that such enzyme recycling systems have been proven to be scalable.[49] Glucose dehydrogenase is very active and can be used for both NADH and NADPH recycling. However, this can cause the pH of the reaction to drop due to the production of gluconic acid, and therefore requires a pH control system.

An alternative option is that the KRED can be used to oxidise another substrate, for example, isopropanol (propan-2-ol or iPrOH, **3**) to acetone, although acetone may have to be removed in order to shift the equilibrium position (Figure 3B).[1] This approach has been successfully used for NAD(P)H recycling, and mutants have been developed when necessary.[49] As discussed in Section 1.1.1, whole cell biocatalysts may be employed as the necessary co-factor recycling enzymes and

compounds *in situ* in the cell (Figure 3C). There are other non-biocatalyst-based approaches summarised in Figure 3 D to F.

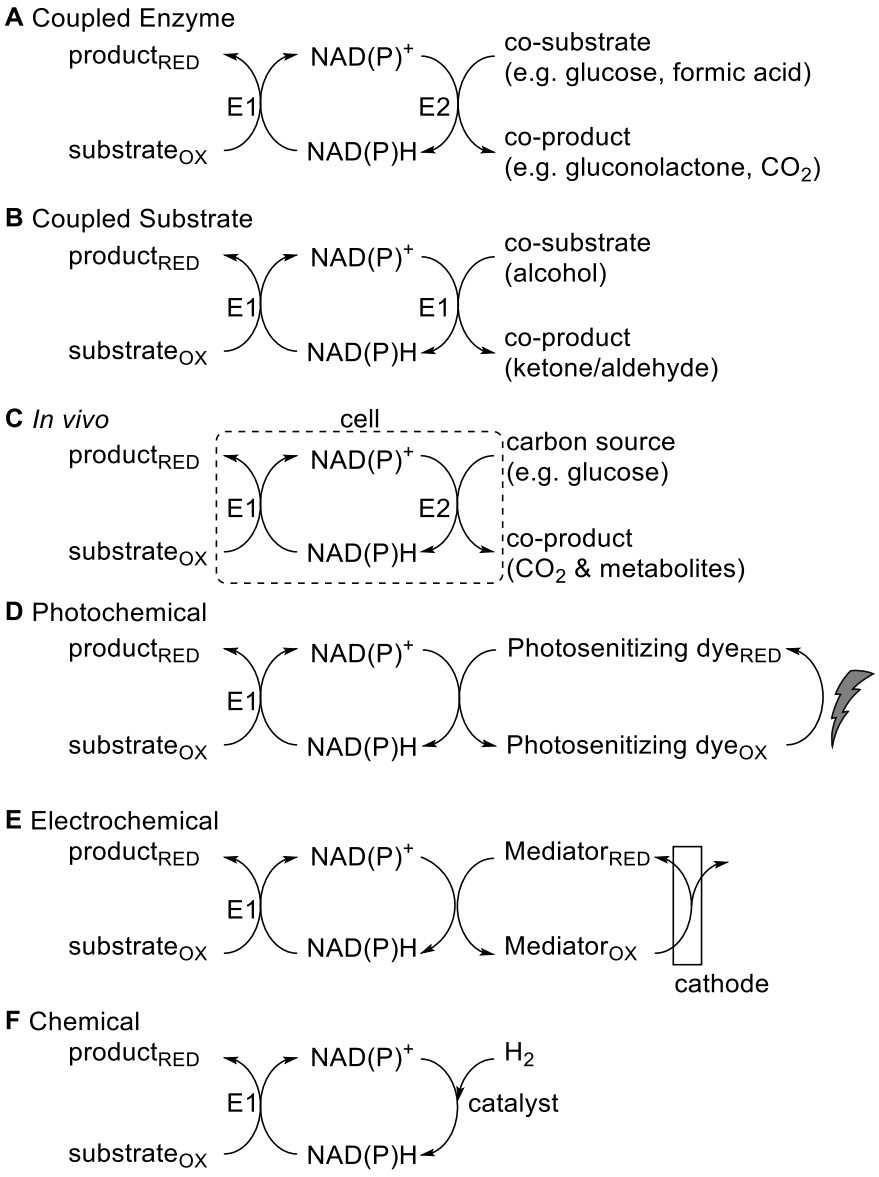
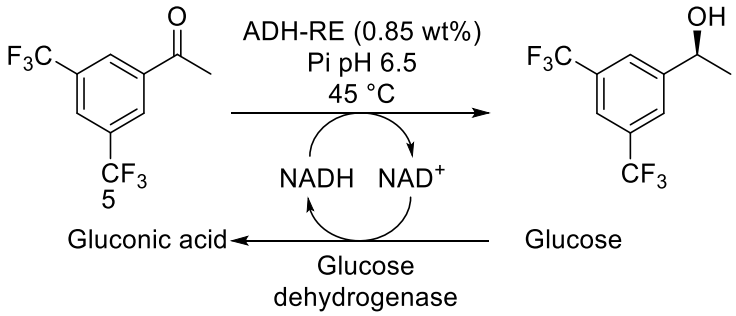


Figure 3: Overview of cofactor regeneration concepts.[1] E denotes enzyme.

1.2.8.1 Substrates including fluorinated compounds

Ketoreductases have been used for substrates across varied structural classes. Fluorinated compounds are important in the pharmaceutical and agrochemical industry as they make up building blocks for biologically active compounds.[50] For example, the quantitative reduction of 3,5-bistrifluoromethyl acetophenone **5** has been reported (Scheme 5).[51] A GDH recycling system was chosen based on a pH optimum that suited both this enzyme and the ketoreductase, ADH-RE (an ADH from

Rhodococcus erythropolis). As both these enzymes are thermostable, the solubility issues relating to the substrate were negated by using an elevated reaction temperature of 45 °C. 2,2,2-Trifluoroacetophenone has also been an important substrate, with 15 articles having reported use of an alcohol dehydrogenase for the reduction of this compound.

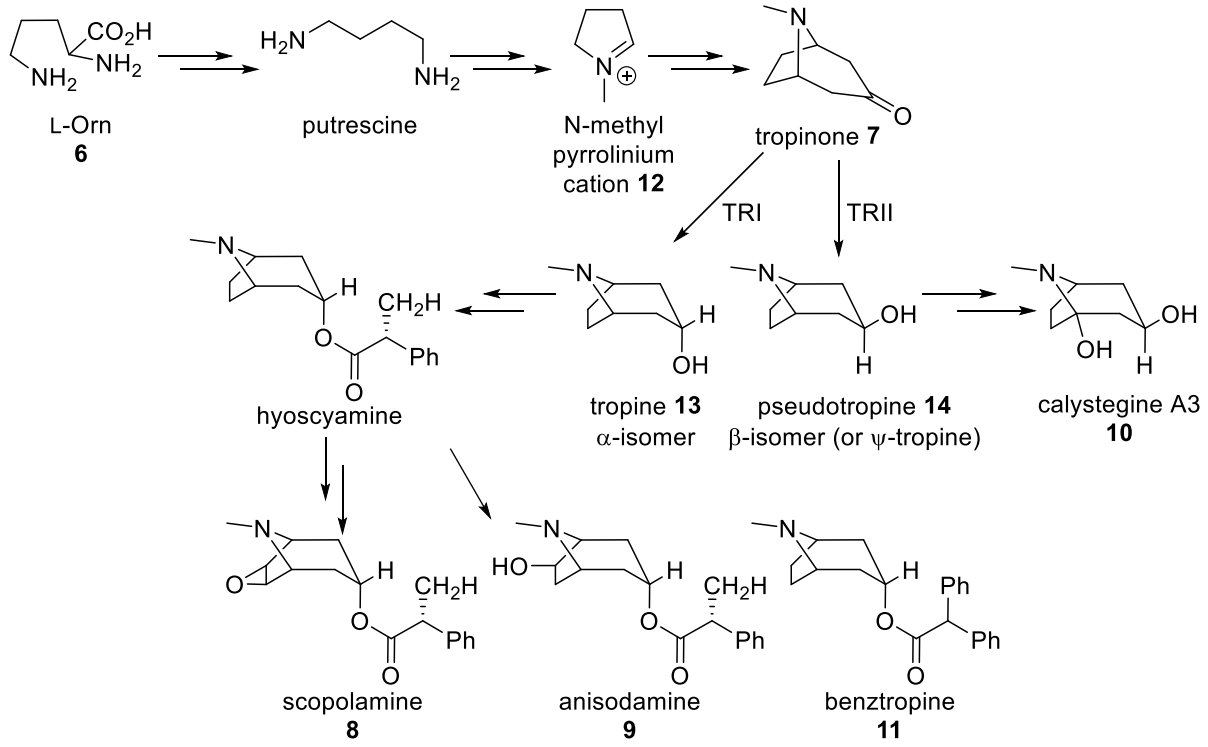


Scheme 5: Reduction of 3,5-bistrifluoromethyl acetophenone 5.[51] ADH-RE is an alcohol dehydrogenases from alcohol dehydrogenase *Rhodococcus erythropolis.*

1.3 Tropane alkaloids

1.3.1 The alkaloid family of natural products

Alkaloids are a diverse group of nitrogen containing secondary metabolites. In general, the nitrogen atoms originate from an amino acid, the skeleton of which is retained in the alkaloid structure.[52] One major class of alkaloids are those derived from L-ornithine (L-Orn **6**), a non-proteinogenic amino acid. It is produced from L-arginine by arginase enzymes in mammals as part of the urea cycle and from L-glutamic acid in plants. L-Orn is converted into the alkaloid tropinone **7** *via* several steps, including the addition of other carbons derived from acetyl-CoA, *via* an intermolecular Mannich-like reaction, a Claisen condensation and an intramolecular Mannich reaction (Scheme 6).[52,53]



Scheme 6: Overview of the biosynthesis of tropinone and its derivatives.[52] TRI and TRII represent the enzymes tropinone reductase I and II respectively. A more detailed scheme is in Appendix Section 8.1.

Tropane alkaloids are commonly found in plants belonging to three families: *Solanaceae*, *Erthroxylaceae* and *Convolvulaceae*, and have an *N*-methyl-8-azabicyclo[3.2.1]-octane core structure. Grantane alkaloids have a methyl-9-azabicyclo[3.3.1]-nonane core scaffold and appear in fewer metabolites. The tropinone alkaloid family include several medically important derivatives (Scheme 6), known for their anticholinergic and chemotherapeutic properties. For example, Scopolalmine butylbromide (from scopolamine or hyoscyamine **8**) is the largest tropane drug produced by volume (original preparation: Buscopan®) and is used for intestinal tract problems.[54] Atropine, the racemic form of **8,** has been used to treat peptic ulcers, Parkinson's disease, diarrhoea and bronchial asthma.[54] Additionally, anisodamine **9** is used in the treatment of acute circulatory shock in China, calystegine A3 **10** is a glucosidase inhibitor (offering an option for treatment of diabetes mellitus type) and benzatropine **11** is used in the treatment of Parkinson's disease and dystonia.[54,55] Investigating the biocatalytic routes of and modifications to this motif is of interest as it is a precursor to many compounds possessing biological activity.

Previously, *in vivo* feeding experiments using radiolabelled precursors have been used to elucidate the general pathways involved in tropane alkaloid biosynthesis in plants.[56–58] Putrescine becomes *N*-methylated and is then oxidised to 4-methyl-1-amino butanal. Under normal physiological conditions, this then spontaneously cyclises into *N*-methyl-Δ-pyrrolinium **12**. The origin of the second ring is still under discussion with two hypotheses: condensation with acetoacetate, oxidation, and a second aldol condensation; or two rounds of polyketide-type extension with malonyl-CoA.[59–61] The reduction of tropinone is a branch point in the biosynthetic pathways and the keto group is reduced to the α or β position **13** or **14**. The α isomer **13** is a precursor to a variety of esterified alkaloids, whereas the β-isomer **14** is a precursor to non-esterified alkaloids calls calystegines.

1.4 Tropinone Reductases I and II

1.4.1 Background

Tropinone reductases I and II (TRI and TRII, EC 1.1.1.206 and 1.1.1236) are at the branch point in the biosynthesis of the tropane alkaloids (Scheme 6) as both reduce the 3-carbonyl group on tropinone *via* transfer of the pro-*S* hydride from the NADPH cofactor to this site.[57,62] As such, both enzymes are NADPH-dependant oxioreductases and members of the short chain dehydrogenase/reductase (SDR) family, previously discussed in Section 1.2.[63,64]

The two TRs generate different products: TRI forms tropine with a 3α-hydroxy group **13**, whereas TRII makes psuedotropine or ψ-tropine with a 3β-hydroxy group **14** (Scheme 6). These are not interconverted *in vivo* and downstream are further metabolised in to a variety of alkaloid products, some of which have been mentioned. Tropinone reductase-like SDRs (TRLs) belong to the SDR group, and have been collected into the SDR65C protein group by the SDR nomenclature initiative, which contains over 350 members.[65] Several genes have been identified as tropinone reductases from Solanaceae such as *Atropa belladonna* or *Datura stramonium*. Recently, a TR isolated from the Brassicaeae *Cochlearia officialis* has been found to have different catalytic specificities, and has been tested with a wide variety of substrates, including terpenes.[66] TRs substrate specificities known to date are summarised in Appendix Section 8.2, which includes other species that have been

cloned including *A. belladonna, S. tuberosum, C. officinalis,* and *S. dulcamara.*[62,64,66–73] As this PhD focused on TRs from *D. stramonium,* Table 2 summarises the substrate selected of the TRs reported in the literature. Activities have been referred to as Vmax or v values in some articles and reviews.[62,69,74]

	TRI		TRII	
Substrate	**Km (µM)**	**Vmax (%)**	**Km (µM)**	**Vmax (%)**
(3)-Quinuclidinone	2200	80	0	nd
2-Fluoroethyl-*N*-nortropinone	3400	nd	670	nd
3-Methylcyclohexanone **38**	nd	38	nd	33
4-Ethylcyclohexanone	45	41	nm	58
4-Methylcyclohexanone **66**	30	39	2800	22
4-Tetrahydro-thiopyrone	30	83	2000	115
6-Hydroxytropinone	5500	20	nm	10
7-Hydroxytropinone	5500	15	nd	6
8-Thiabicyclo[3.2.1.]-octan-3β-one (β-TBON, β-**15**)	nd	35	nd	nd
8-Thiabicyclo[3.2.1.]-octan-3-one (TBON, **15**)	33	30	0	0
N-Ethylnorpseudopelletierine	2500	65	250	45
N-Ethylnortropinone	nd	0	nd	5
N-iso-propylnortropinone	20000	55	20000	nd
N-Methyl-4-piperidone **21**	20000	180	1400	28
Nortropinone **18**	nd	3 (2.5 mM)	nd	23 (2.5 mM)
Norpseudopelletierine	nd	0	nd	8
N-Propyl-4-piperidone	0	78	nm	
Pseudopelletierine **86**	nd	0	nd	7
Tetrahydrothiopyran-4-one	nd	83	nd	nd
Tropinone **7**	1300	100	110	100

Table 2: Summary of the activities toward ketone substrates of TRI and TRII as a percentage of activity as relative to that with tropinone from *Datura stramonium* reported in the literature.[68,69,75,76]

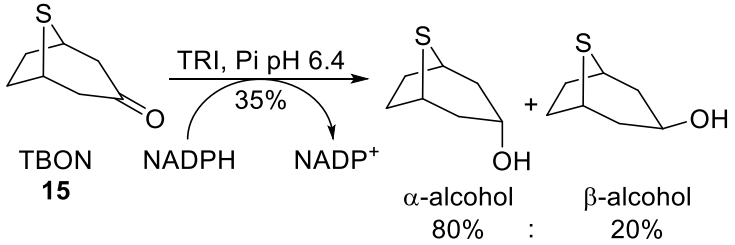
An interesting finding from Brock *et al.* was that tropinone was not the preferred substrate for *Cochlearia officinalis* derived TRII (Co-TR).[64] Although it accepted a broad range of substrates, much like other TRs, many of these substrates were reduced more rapidly than tropinone.[64] This demonstrates that substrate specificity is very much species dependant and thus one of the limitations of current *in silico* annotations, as the retrieved enzyme has differing substrate specificity to that which was expected. Co-TR is not stereospecific, acting as both a TRI and TRII, and so is considered a possible ancestor of specialised TRs.[64,77]

TRs from *D. stramonium* (Ds-TRs) have been analysed by X-ray crystallography, revealing that the two enzymes have very similar folding and share a catalytic Tyr residue and cofactor binding site.[62] The two enzymes share 64% residue identity. Substrates that have been tested with Ds-TRI and Ds-TRII have been

limited to bicyclic systems similar to tropinone, methylcyclohexanones and piperidone based rings. At pH 7.0, Ds-TRI was reported to be 11 times more active than Ds-TRII.[74]

Work has been reported on a selection of TRs to elucidate the mechanism of action and show how TRI and TRII generate the different stereocentres. Nakajima *et al.* used X-ray crystallographic data of Ds-TRs to demonstrate that 5 of the 11 amino acid residues that are associated with substrate binding could also be correlated with making the tropine **13** or ψ-tropine product **14**.[78] The NADPH cofactor is bound in the same way on both TRI and TRII.[78]

The two TRs bind the tropinone substrate in different orientations, causing the α- or β- conformation in the product.[62] The pKa of tropinone **7** is 8.9 the substrate substantially charged.[68,79,80] In TRI the orientation of tropinone is fixed by the repulsion between the positive charge of His112 and the nitrogen atom on tropinone.[74,78,81] The positively charged His112 is replaced by the polar, though not basic, Tyr-100 in TRII and the hydrophobic Val-168 in TRI is replaced by Glu156.[78] In TRII, the attraction of the nitrogen by a negatively charged glutamic acid (Glu156) side chain holds tropinone in place.[74,78,81] Site-directed mutagenesis also confirmed this hypothesis[82]. Yamashita *et al*. noted that positioning of the substrate at an optimal angle for hydride transfer is an important requirement for the TRII to be active.[62,81]Additionally Nakajima *et al*. reported that 8-thiabicyclo[3,2,1]octane-3-one (TBON, **15,** Scheme 7) was reduced to either the α- or β-alcohol, suggesting that the nitrogen is key for the stereoselectivity.



Scheme 7: Reduction of TBON 15 by TRI.[68,74]

Portseffen *et al.* investigated Ds-TRs activity at differing pH environments. Ds-TRI activity is optimum at pH 6.4, is inactive below pH 5.5 and above pH 7.5 tropinone reduction decreases. Several reasons have been proposed for this, such as the concentration of uncharged tropine becoming sufficiently high to compete for the

active site and pH having an effect on the catalytic site itself.[68] It was proposed that acidic pH conditions favoured turnover but uncharged tropinone more readily formed complexes at the active site. Despite TRI having a higher K_{cat}, TRI has a much lower affinity to tropinone than TRII (Table 3). Nakajima *et al*. therefore proposed that the low affinity of tropine may positively impact the rapid turnover of TRI. However, increasing the pH has been shown to decrease the K_m for TRI.[62,82]

Ds-TR	TRI	TRII
K_m for tropinone (mM)	0.117 ± 0.010	0.0476 ± 0.0122
K_m for NADPH (mM)	11.6 ± 4.0	18.7 ± 5.4
K_{cat} (s^{-1})	25.6 ± 0.9	2.73 ± 0.16

Table 3: Kinetic parameters of Ds-TRs measured at pH 7.0 at 28 °C.[62]

Ds-TRII has a similar optimal pH of 6.25, however, keeping 90% of maximal activity between pH 5.2 and 8.2.[68] The pH optimum of TRs varies with species (Table 4).

Species	Enzyme	pH optimum	Purification process
S. tuberosum	TRI	6.0[71]	Enzyme expressed in *E. coli* and purified
S. tuberosum	TRII	6.4[72]	Enzyme expressed in *E. coli* and purified
D. stramonium	TRI	6.4[68]	Enzyme purified from root cultures
D. stramonium	TRII	6.4[68]	Enzyme purified from root cultures
H. niger	TRI	5.9[70]	Enzyme purified from root cultures
H. niger	TRII	5.9[70]	Enzyme purified from root cultures
A. belladonna	TRII	6.25[67]	Enzyme purified from root cultures

Table 4: A comparison of TRs isolated from different species and a comparison of the pH at which they are the most active towards tropinone.

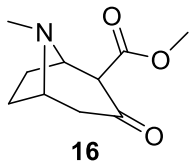
Freydank *et al.* highlighted the effect of appending a hexahistidine-tag (His-tag) to TRII from *S. dulcamara*.[83] The specific activity of TRII with a His-tag on the N-terminal was over four-fold higher and the K_m was approximately tenfold lower compared to TRII with the His-tag on the C-terminal. Though the N-terminal His-tag had no direct effect on the active site, the C-terminal His-tag was found to inhibit the catalytic activity of TRII. They postulated that the His-tag is close to the active site if appended to the C-terminal, which inhibits induced fit movements of the enzyme and hinders substrate entry, slowing down catalysis.[64]

1.5 MecgoR

1.5.1 The tropinone reductase MecgoR

The reduction of tropinone in *Solanaceae* is well described in the literature and summarised above. However, these *Solanaceae* enzymes do not show high activity with methylecgonone or 2-carbomethoxy tropinone (2-CBT, **16**, Figure 4), the precursor involved in cocaine biosynthesis (Section 5.9). *Erythroxylum coca* (*Erythroxylaceae*) is a species that accumulates cocaine and other tropane alkaloids in abundance and is used to harvest this natural product as a drug.

The enzyme methylecgonone reductase (MecgoR) has been characterised by Jirschitzka *et al.*, who purified the enzyme from young *Erythroxylaceae* leaf extracts after detecting enzyme activity.[84] Confirmation of MecgoR's role in tropane biosynthesis came from immunoprecipitation experiments. The stereospecificity of the expressed protein results in the β-configuration of the alcohol from **16**, which is the same as those of the tropane alkaloids in *Erythroxylaceae*. MecgoR is capable of catalysing the oxidation reaction, whereas TRII is not.[68,72] Furthermore, MecgoR can use NADH as a cofactor, which is much more economically viable, but activity is reduced to 14% of the value with NADPH.[61] Additionally, MecgoR can be inhibited by its own substrate which has not been noted for other TRs described in the literature.[61] Optimum pH for the reduction of 2-CBT was 6.8 and 9.8 for the oxidation. Table 5 shows the substrate specify of MecgoR, which displayed some activity to **16**, **17**, **7** and **18**, but not to **15**, **19**, **20** or **21**.

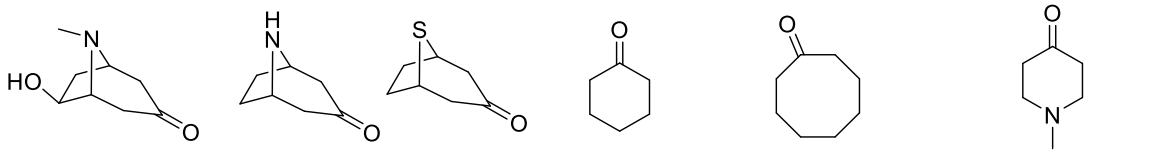


16

Figure 4: Methylecgonone or 2-carbomethoxy tropinone (2-CBT, 16), the natural substrate for MecgoR.

Substrate	Activity relative to that of 2-CBT (%)
2-CBT **16**	100
6-hydroxytropinone **17**	45
Tropinone **7**	36
Nortropinone **18**	6
TBON **15**, cyclohexanone **19**, cyclooctenone **20**, N-methyl-4-piperidone **21.**	0

Table 5: Substrate that have been tested with MecgoR to date.[61] Substrates shown in Figure 5.



6-Hydroxytropinone **17** | Nortropinone **18** | TBON **15** | Cyclohexanone **19** | Cyclooctanone **20** | *N*-methyl-4-piperidone **21**

Figure 5: Reported substrates that have been tested with MecgoR.[61]

MecgoR belongs to the aldo-keto reductase (AKR) family. This large family consists of enzymes from plants, mammals, amphibians, yeast, protozoa, and bacteria[85]. All enzymes in this group are NAD(P)H -dependent and possess a similar cofactor binding site. The AKR catalytic tetrad is also conserved, and consists of Asp-50, Tyr-55, Lys-85 and His-118 in MecgoR from sequence identities. MecgoR, TRI and TRII are therefore from two different protein groups but carry out the same function, and have evolved independently in the two plant families.

MecgoR shows high sequence identity to AKR family proteins found in the genome of *Handroanthus impetious* plant found in South and Mesoamerica with a 63% sequence identity.[86] This protein has been annotated as such in a genome sequenced using the Illumina HiSeq system. Jirschitzka *et al*. also highlighted MecgoR shares 48%, 50%, and 58% identity deoxymugineic acid synthase (*Zea mays*), codeinone reductase (*Papaver somniferum*), and chalcone reductase *(Sesbania rostrata*) respectively, all of which are also AKRs.[61]

1.5.2 The cocaine biosynthetic pathway

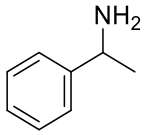
MecgoR is the penultimate step in the cocaine biosynthetic pathway (Appendix Section 8.1). Schmidt *et al.* have recently described the final step, in which a member of the BAHD acyltransferase family (EC 2.3.1) of enzymes produces both cocaine and cinnamoylcocaine *via* activated benzoyl- or cinnamoyl-Coenzyme A thioesters.[87] They

also reported that tropane production occurs in the above-ground parts of the plant in *Erythroxylaceae*, but in the roots in *Solanaceae*.[87]

1.6 Transaminses

1.6.1 Background

Transaminases (TAms, EC 2.6.1) are pyridoxal-5'-phosphate (PLP)-dependant enzymes that catalyse the transfer of an amine group from a donor to an acceptor substrate by a ping-pong bi-bi mechanism (Scheme 8).[88] There are six classes of TAms, I to VI, defined by amino acid sequence alignment (Table 6).[89] Though a wide range of possible small molecules have been recorded in the literature as amine donors, methylbenzylamine (MBA, **22**, Figure 6) has been used the most frequently.[90] TAms have the advantages that they demonstrate high stereoselectivity, have no requirement for redox cofactor recycling and high reaction rates have been reported.[91] The PLP cofactor is recycled during the catalytic cycle. However, problems such as substrate and/or product inhibition, low tolerance to organic solvents and unfavourable reaction equilibria have hindered TAms from increased industrial uptake.



22

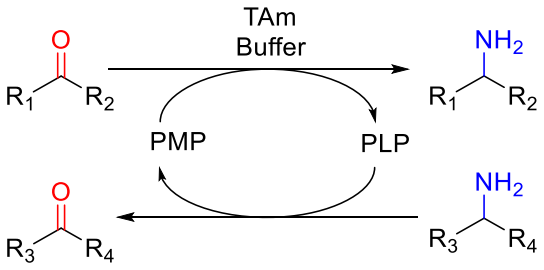
Figure 6: Methylbenzylamine 22.

Subgroup	Enzyme
I/II	Aspartate TAm Aromatic TAm
III	ω-TAm
IV	Branched-chain TAm (*R*)-selective TAm
V	Phosphoserine TAm
VI	Sugar TAm

Table 6: Classification of TAms.[92]

Most TAms naturally occur as dimers, though there are some tetramer examples, such as from *Arthrobacter aurescens.*[93] The active site is located at the dimer interface, so there are 2 active sites per dimer. Currently known wild-type

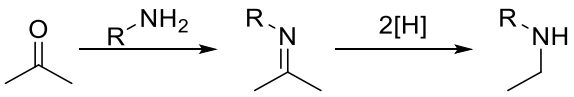
TAms have an active site which is made up of a large and a small binding pocket. For Class III or ω-TAms, it is common that the smaller of these pockets only accepts a methyl group, and so protein engineering has focussed on expanding this.[94]



Scheme 8: Simplified transaminase reaction scheme. The amino donor binds first to the enzyme, and PLP is aminated to pyridoxamine 5'phosphate (PMP), releasing the respective keto product from the amine donor. The amino group is then transferred from the PMP to the acceptor molecule, and liberating PLP.[95]

There are two general strategies used with TAm in biocatalystic syntheses: kinetic resolution of a racemic starting material and asymmetric synthesis starting from a prochiral starting material. Both methods can also be used together in deracemisations involving the kinetic resolution of a racemic amine and the subsequent ketone product can be used as a substrate for asymmetric synthesis using a TAm of opposing enantiopreference.[39]

In traditional approaches to single isomer amines, chiral resolutions of amines can be performed with up to a 50% yield, using a chiral acid such as (*R*)-mandelic acid or (*R,R*) tartaric acid. Another major method of synthesising amines is reductive amination (Scheme 9).[96]



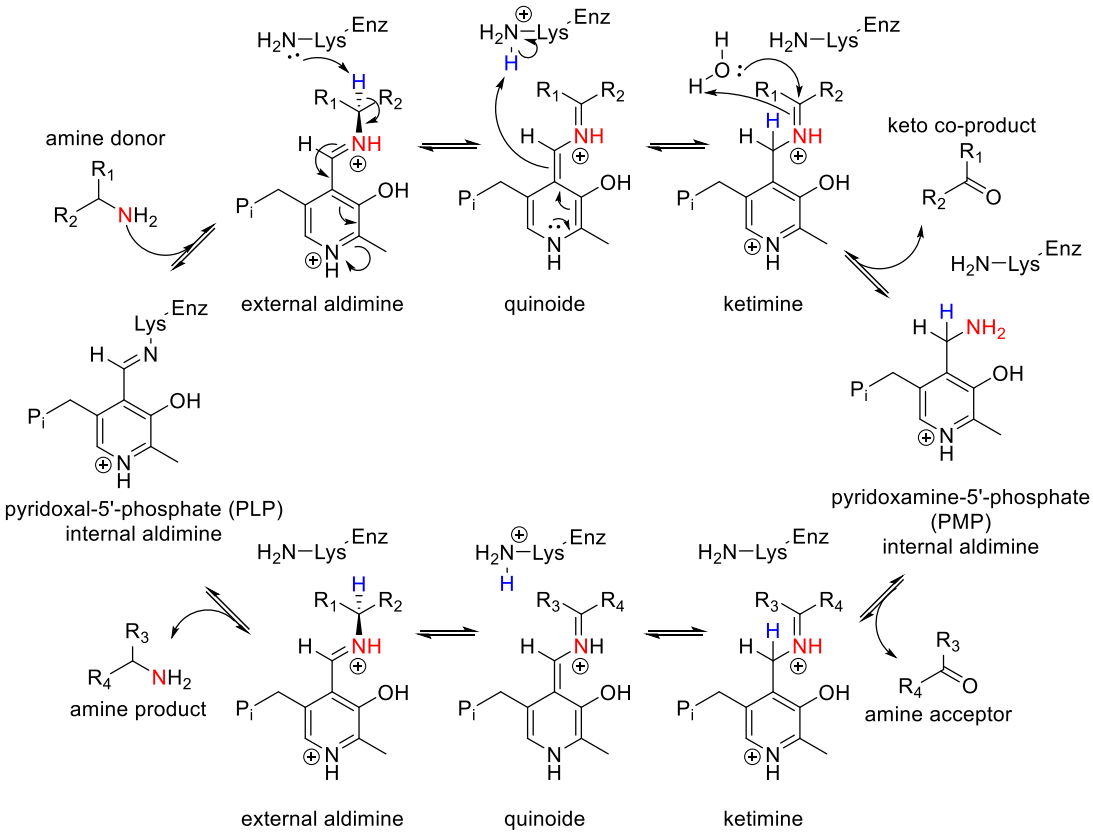
Scheme 9: General scheme of reductive amination.

By comparison, the use of TAm enzymes with prochiral ketones would give up to a 100% yield. The replacement of these methods with TAms would reduce the use of toxic transition metal catalysts and the energy required as these reactions are often carried out at high pressures or temperatures.

1.6.2 TAm catalytic mechanism

As shown in Scheme 10, TAms use PLP and lysine in a catalytic fashion to asymmetrically synthesize amines.[94,97] The reaction is proceeds via a ping-pong bi-bi mechanism and consists of two half reactions. The active sites of TAms have a larger

binding pocket in which R_L is directed and a smaller one for R_S, orientating the substrate for asymmetric transamination.



Scheme 10: Schematic catalytic mechanism of TAms, with a mechanism for the first half reaction. [94,97–101]

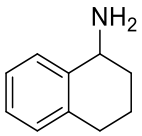
1.6.3 Reported TAms

Transaminases have been used on a wide range of substrates and several examples are highlighted.

1.6.3.1 Vf-TAm

An early example of a TAm is (*S*)-selective *Vibrio fluvialis* JS17 TAm (Vf-TAm), first described in 2003 by Shin *et al.*[9] Vf-Tam was discovered *via* enrichment cultivation using MBA **22** as the sole nitrogen source (Section 1.1.3) and the enzyme has been used as an catalyst with aromatic amine donors, such as MBA or 1,2,3,4-tetrahydro-1-naphthylamine **23** (Figure 7), and as such has been widely applied.[8] The optimal pH for the enzyme was reported to be pH 9.2 and 37 °C was the optimal temperature.[9] The presence of PLP was reported to increase enzyme stability and (*S*)-MBA was found to reversibly destabilise the enzyme.[9] Vf-TAm has been extensively

mutated, guided by homology modelling and substrate docking, to increase its tolerance towards aromatic compounds.[102]



23

Figure 7: 1,2,3,4-tetrahydro-1-naphthylamine 23.

1.6.3.2 Cv-TAm

The bacterial (*S*)-selective ω-TAm from *Chromobacterium violaceum* (DSM30191, CV2025, or Cv-TAm) was first identified *via* sequence homology to Vf-TAm. It has 38% sequence identity, and has been tested on a range with aromatic substrates, including 1-indole.[10] Cv-TAm is a homodimer in which each monomer has an active site at the dimeric interface involving amino acids from both subunits (Figure 8).[103] There are three critical loops that are involved in the structure of the active site. In the presence of PLP, major structural rearrangements take place, such as the ordering of a small helical domain at the N-terminus of each monomer. Crystal structures for both the apo- and holo-enzyme have been resolved, which identified residues such as an active-site loop (residues 81-93) which potentially be significant for substrate specificity.[104]

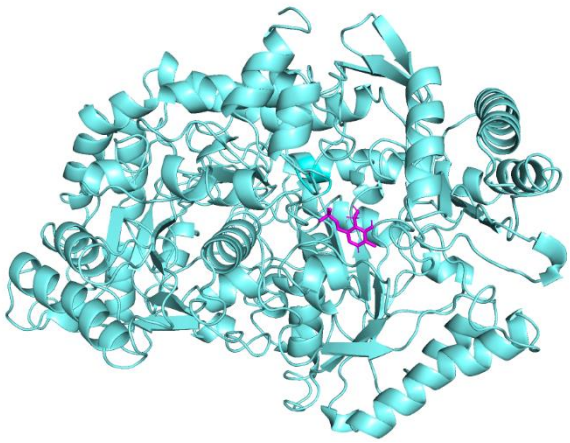


Figure 8: Crystal structure of Cv-TAm dimer (PDB accession number: 4AH3), coloured in cyan. PMP is coloured magenta.

1.6.3.3 Mv-TAm

The transaminase from *Mycobacterium vanbaalenii* (Mv-TAm) is selective towards (*R*)-MBA. The majority of work performed using this enzyme has been with ketone or amines as acyclic moieties, but aromatic cyclic ketones, many of which have been functionalised with fluorine and aliphatic groups, have also been investigated.[105] Mv-TAm operates optimally at a pH of 7-7.5.[106]

1.6.3.4 As-TAm

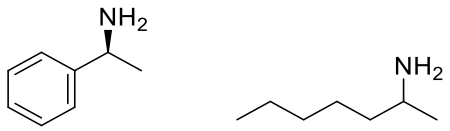
An *R*-selective variant of the ω-TAm from *Arthebrobacter sp.,* known as ArRMut11 (As-TAm), was developed by Savile *et al.* after 11 rounds of mutagenesis.[3] In total, 8% of the enzyme sequence was subjected to protein engineering and was used with 1,3-ketoamides to generate the (*R*)-functionality in sitagliptin.[94] It has been reported to transaminate a number of tetralone and chromone bicyclic compounds, with up to 99% conversion yields.[107]

1.6.3.5 Other TAms

Pp-Tam-2 from *Pseudomonas putida* PP_3718 has been shown to tolerate dopamine as a substrate.[108] BS-TAm-2 and Bs-TAm3 are from *Bacillus subtilis* strains BSU09260_1971 and BSU09260_402 respectively, but the enzymes are the same.

1.6.4 TAm protein engineering

Mutagenesis of TAms to increase tolerance to larger and bulkier molecules has been investigated extensively.[3,93,102,104,109,110] For example, Vf-TAm (Section 1.6.3.1), was engineered to increase activity towards **(*S*)-22** and **24** (Figure 9).[102] Using homology modelling and the substrate-structure relationship, the active site for the TAm was shown to have a large and small substrate-binding site. Among several residues selected to influence binding, the residue W57 in the larger substrate-binding site had bulky side chains, and mutagenesis was used to improve substrate binding. The W57G variant displayed over a 2-fold increase in specific activity towards (*S*)-MBA **(S)-22** compared to the wild type, and a 6.7-fold increase towards **24.**[102]

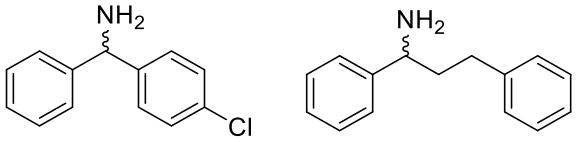


(*S*)-MBA
(*S*)-22

heptan-2-amine
24

Figure 9: Substrates that Vf-TAm was engineered to accept more readily.[102]

One notable study by Pavlidis *et al*. in 2016 reported the engineering of a TAm to accept bulky aromatic substrates (**25** and **26**).[111] One enzyme from Ruegeria sp. TM1040 (3FCR) was selected from a panel of TAms for mutagenesis as it displayed activity towards **25** and **26**. *In silico* analysis and subsequent mutagenesis led to a variant containing only 4 mutations. This variant was employed in a number of preparative scale reactions, giving conversions of over 95% and over 99% e.e.. This illustrates the strength of rational design and Pavlidis *et al*. suggested that the incorporation of these mutations is broadly applicable to other TAms. However, Vf-TAm and Cv-TAm already contained 3 of the suggested mutations as well as other structural motifs that provide the structural flexibility.[111] Both CV-TAm and Vf-TAm did not share activity with the targeted substrates, despite these similarities in the motif.

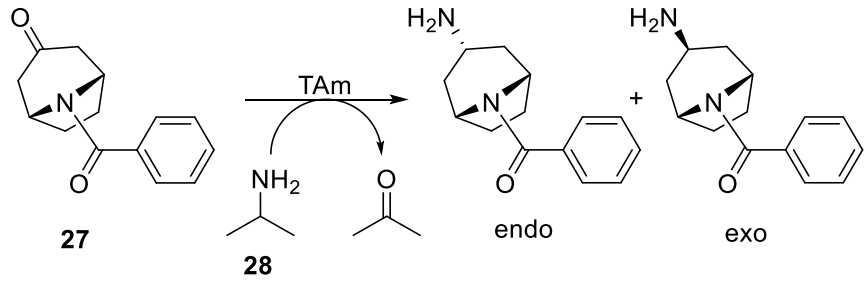


25 **26**

Figure 10: Targeted compounds in the recent work of Pavlidis *et al.* for TAm mutagenesis.[94]

Another publication by Weiß *et al.* in 2016 focussed on protein engineering, to identify a TAm variant that accepts the bicyclic compound **27** (Scheme 11).[112] This bicyclic compound **27** does not follow the traditional TAm target substrates with a large and small side chain, so rational protein engineering was hampered and an error-prone PCR technique in combination with a high-throughput glycine oxidase assay was utilised. This compound is structurally related to tropinone and is a building block for pharmaceutically active ingredients.[112] The group initially screened variants from the previous study, but found only low levels of activities.[94] One variant (3FCR_QM) was then subjected to site directed mutagenesis at the 5 residues.

Introduction of a Y152F mutation lowered selectivity towards the *exo*-isomer (100% to 40% *exo* to *endo* preference), whilst a T59L mutation caused a 3-fold increase in activity: 145 mU mg^{-1} compared to 45 mU mg^{-1}. A non-optimised small-scale preparative scale reaction was completed using isopropylamine **28** as amine donor, a 75% conversion (60% isolated yield) and >99.5% selectivity for the *exo*-amine was achieved.



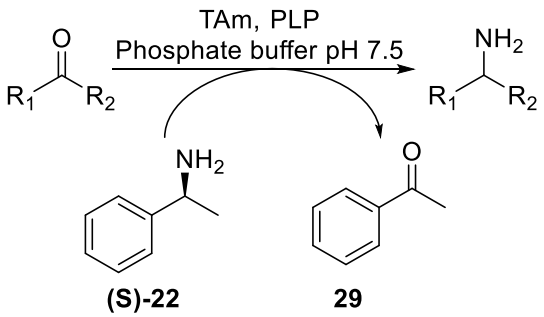
Scheme 11: Targeted transamination reaction investigated by Weiβ *et al.*.[112]

1.6.5 Detection techniques and assays

There are a wide number of methods that have been developed to detect transaminase activity. HPLC can be used to detect product formation or depletion of the starting material, or colourimetric assays can be employed.

1.6.5.1 Acetophenone assay

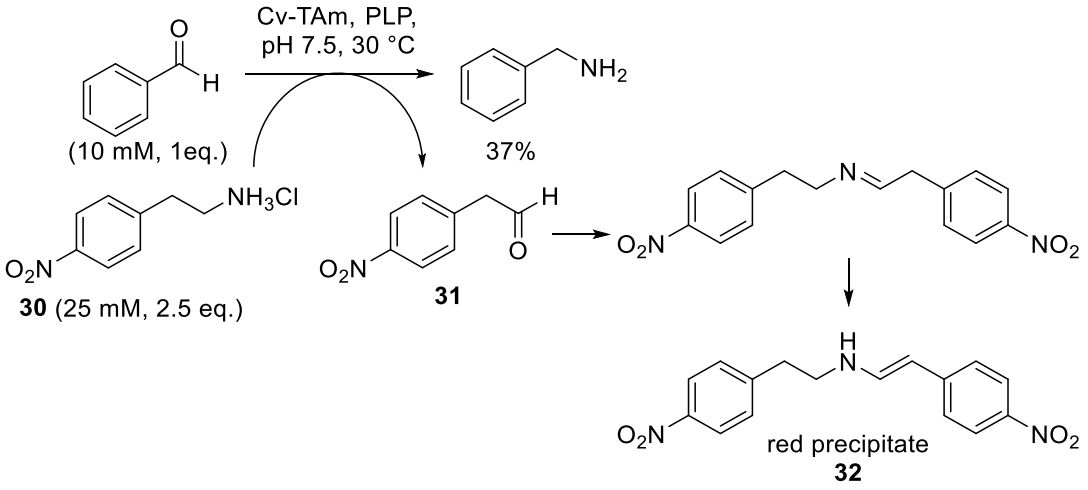
(*S*)-MBA has been widely used as a model amine donor as it is commonly accepted by many ω-TAms. Acetophenone **29** is formed upon transamination, which is detectable by HPLC at 254 nm and can be used to calculate conversion, as well as a method for enzyme characterisation.[113] Schätzle *et al.* illustrated the use of the assay in 2009 to characterise a novel TAm from *Rhodobacter spharroides.*[113]



Scheme 12: Transaminase assay using (*S*)-MBA (*S*)-22 as the amine donor. (*R*)-MBA may also be used with R-selective TAms.

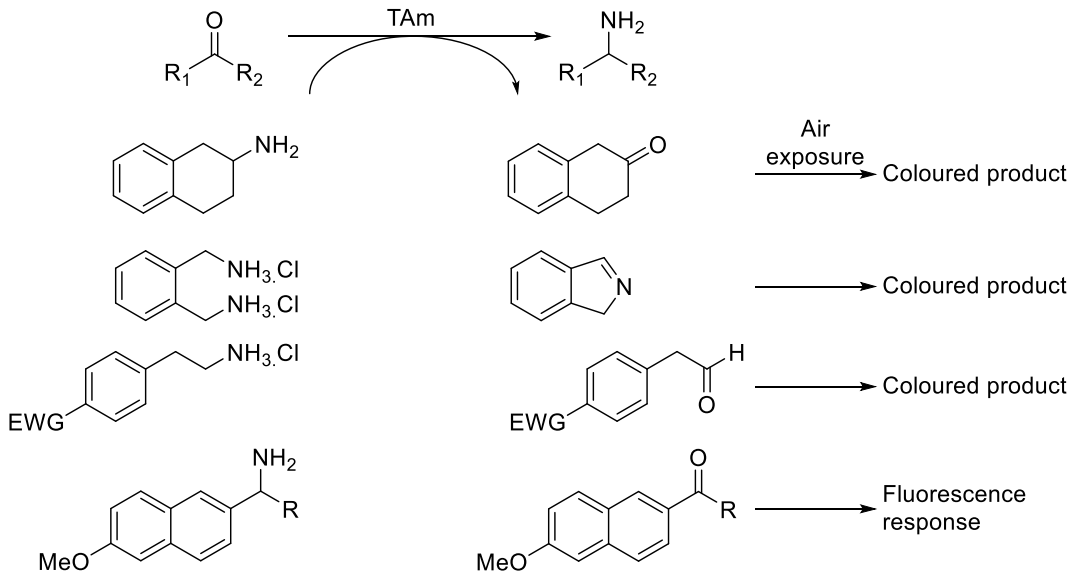
1.6.5.2 Colourimetric assays

A novel assay developed at UCL is a rapid and low-cost method to evaluate the transaminase activity, using 2-(4-nitrophenol)ethan-1-amine **30** as the amine donor (Scheme 13).[89] Conversion of the amine donor **30** in a TAm reaction generates the aldehyde **31**, which upon reaction with **30** and tautomersiation, gives the red precipitate **32**, indicative of activity. The assay may be used as a qualitative and high-throughput method of detecting conversion.



Scheme 13: Use of 2-(4-nitrophenyl)ethan-1-amine 32 as the amine donor in a transamination reaction with Cv-TAm.

There are other examples of colorimetric assays reported that use various amine donors to monitor conversion (Scheme 14). Other examples include: a Phenol red assay using lactate dehydrogenase (LDH) and glucose dehydrogenase (GDH), copper sulphate assay, and glycine oxidase assays.[8,114]



Scheme 14: A summary of selected colourimetric assays for high throughput screening of TAms.[8,89,115–117]

1.7 Aims of this PhD

This project aimed to build on earlier work at UCL which developed a novel strategy for sequence mining from metagenomic data. As discussed earlier, new biocatalysts are needed with enhanced properties or substrate selectivities for use in synthesis. High sequence diversity in the enzymes identified would make a good starting point for future evolution projects. The tropinone structure is found in several pharmaceutically relevant compounds and a focus of this PhD project was the natural product tropinone and tropinone reductases (TRs) found in plants. MecgoR has been reported to be active towards compounds even bulkier than the TRs, and a goal of the project was search for MecgoR activity within the metagenomic libraries at UCL and find novel TR- or MecgoR- like enzymes, as well as building on reported work on TRs. Transaminase activity towards tropinone and tropinone analogues has not been reported in nature. Here, one aspect of the project was to also identify TAms that are active towards tropinone or its analogues. Several analytical techniques will be employed to assess this, such as HPLC, spectrophotometer, NMR and mass spectrometry characterisation.

The overall aims of this PhD project were:

(i) Investigate enzymes, specifically ketoreductases, from metagenomic sources and use them for biocatalytic applications.

(ii) Expand the knowledge of currently reported TRs and using them with a wide selection of substrates and in biocatalytic applications.

(iii) Search for MecgoR-like enzymes in metagenomic libraries and assess the activity of retrieved enzymes.

(iv) Identify any transaminase activity with tropinone or tropinone analogues.

2 Tongue metagenome short-chain dehydrogenases/reductases

2.1 Introduction and previous work: Metagenomic preparation, gene selection and expression

Short-Chain Dehydrogenases/Reductases (SDRs) are a family of KREDs which have recently been identified in some of the metagenomes work at UCL. As discussed in Section 1.1.2, biocatalysis poses an environmentally friendly, cost effective and efficient alternative to traditional chemistry techniques. As of 2009, about 10% of total drug synthesis depends on biocatalysts.[34] SDRs, and ketoreductases (KREDs) in general, are therefore of industrial and environmental significance: oxidation and reduction have accounted for the second largest number of the studies on biocatalysis.[34] KREDs also provide an efficient route to single enantiomers, which is important in the pharmaceutical industry: In 2006, 75% of all dosage form drug sales in the USA were single enantiomers.[118]

In previous work at UCL, 37 potential SDR enzymes were selected from a metagenomic library derived from the microbiome scraped from a human tongue from 9 volunteers at UCL and cloned into expression systems*. These SDRs were then investigated in this PhD for their activity against a range of aldehydes and ketones including tropinone.

In previous work at UCL*, an *in silico* metagenome library was generated from sequencing DNA samples collected from the oral human cavity using the Roche 454 platform (Section 1.1.3). MetaGeneMark software was used to predict open reading frames in this, and these were scanned using the Pfam stand-alone tool to mark them with a Pfam Identification (ID). The sequences of SDRs from *L. brevis* (Uniprot ID Q84EX5), *L. kefiri* (Uniprot ID Q6WVP7), *L. Minor* (Uniprot ID A0A0R1Z9F9) and *W. thailandensis* (Uniprot ID G0UH95) were used as search sequences due to their preference for the (*R*)-alcohol.[6] This retrieved 55 contigs that contained some similarity to the query sequences. These were then interrogated on the Pfam web server, and generated two PFam IDs for short chain alcohol dehydrogenases: adh_short_C2 and adh_short. The tongue metagenome was then used to search the

* Tongue metagenome library creation was carried out by John Ward, Natalie Dawson, Stephanie Easton, Samantha Whiting & interrogation, PCR and cloning procedures were performed by Jack Jeffries at Biochemical Engineering, UCL.[6]

marked contig database and generated 139 sequences which were considered to be full length and non-redundant. Of these open reading frames (ORFs), sequences were judged to be acceptable for retrieval if they had an initiator methionine and stop codon and were between 230 and 250 amino acids in length. The remaining ORFs were then clustered based on similarity and one sequence from each cluster was chosen along with a non-redundant sequence for primer design, totalling 38 sequences for retrieval.[6]

PCR was used to amplify the selected DNA, followed by cloning with pET29a vectors using Gibson assembly, a quick restriction-independent method. This was then transformed into a Top10 cloning strain of *E. coli* to capture the recombinant plasmids, and subsequently transformed into BL21 Star DE3 pLys for expression. This was successful for all selected genes except for enzyme 14, for which PCR was unsuccessful.[6]

The retrieved genes showed high sequence similarity to homologues from fully sequenced individual genomes in the NCBI databases: 27 genes had ≥ 95% identity to registered genes. Enzyme 3 was the lowest with 77% identity, and it was proposed that this was due to variation in the retrieval process, as opposed to a corrupt *in silico* sequence that was ambiguous at certain points in its DNA sequence.[6]

2.2 Metagenomics alignment

With this preliminary work in place, further work was presented as part of this PhD. The sequences were aligned using the EMBL-EBI Clustal Omega web tool, and this was used to compile the results as a matrix heatmap of percent identity (Figure 11).[119] The SDRs were identified *via* homology with SDRs from *Lactobacillus brevis, Lactobacillus kefiri*, and *Lactobacillus minor,* and all belonged to the SDR family but showed low sequence similarity to each other in general (0-40%). Some pairs of enzymes had particularly high similarity, such as SDR-3 & SDR-29; SDR-1 & SDR-21; SDR-18, SDR-37 & SDR-5; SDR-36 & SDR-20; SDR-25 & SDR-7; SDR-27 & SDR-30. SDRS- SDR-25, SDR-27 and SDR-30 showed particularly low sequence homology to the other SDRs except for each other. This may suggest that they have similar

activities. An alignment generated using the M-coffee web tool was also visualised in a tree format, using ITOL, and shows similar pairings (Figure 12).[120,121]

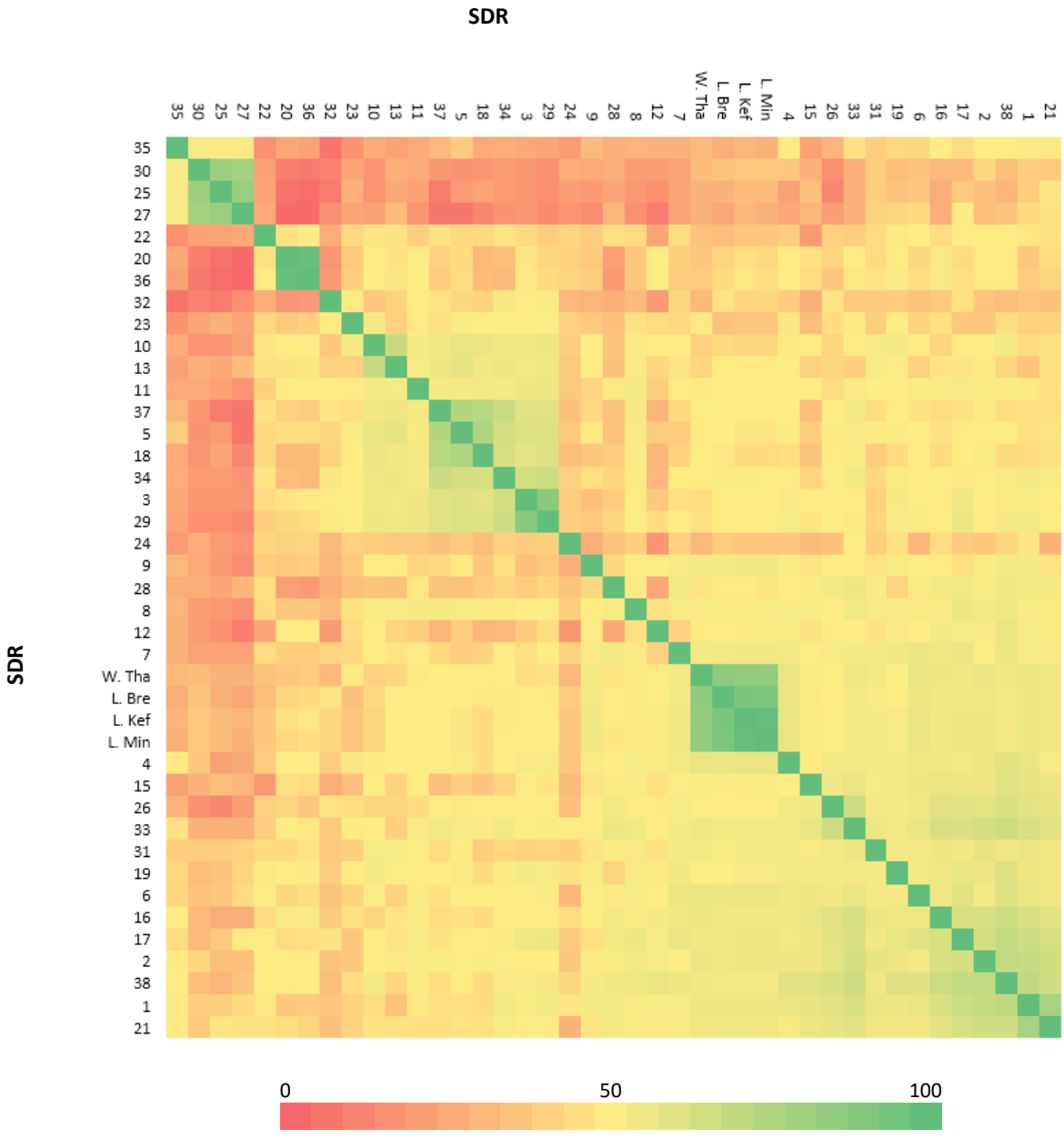


Figure 11: Heatmap of sequence identities (%) of the SDRs from the tongue metagenome as calculated by Clustal Omega online tool.[17,121] The colour key is displayed showing overall percent sequence identity.

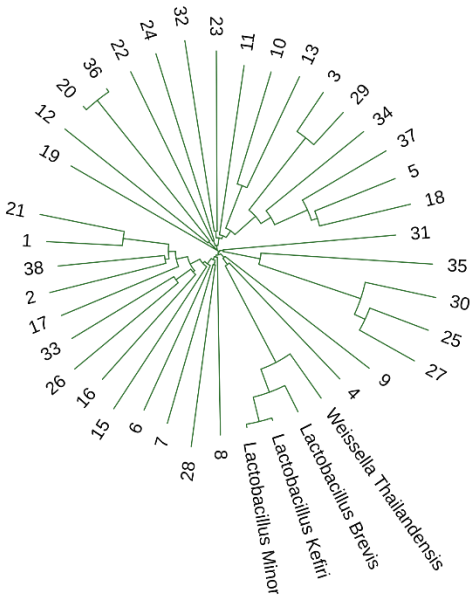


Figure 12: Phylogenetic tree to visualise sequence identities made using the ITOL tool and the sequence identities (Figure 11).[120]

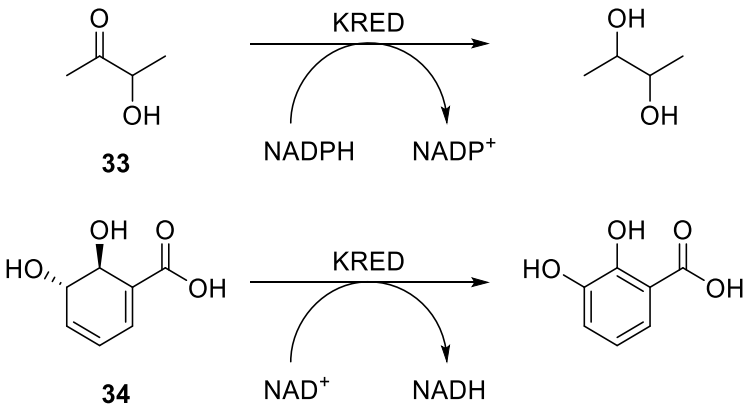
NCBI was also used to predict the enzyme functions: NCBI BLAST was used to query the amino acid sequence, and the assignment with the highest percent identity (Table 7). 3-Oxoacyl-ACP reductases, also knowns as β-ketoacyl-ACP reductases, were a common enzyme assignment and are involved in fatty acid synthesis. Other enzymes are given the broad classification of oxidoreductase, and largely were annotated proteins on genomes, as opposed to isolated and experimentally investigated enzymes. SDR-4 and SDR-9 were predicted to accept acetoin **33** and 2,3-dihydro-2,3-dihydroxybenzoate **34** as substrates respectively (Scheme 15).[122,123] All the sequences often aligned to sequences from annotated genomes as opposed to expressed and experimentally tested enzymes with known functionality.

Enzyme Number	Contig ID	Number of amino acids	Molecular weight (Da)	NCBI annotation	Identity (%)
1	4926	248	26132.1	3-Oxoacyl-[acyl-carrier-protein] reductase [*Porphyromonas somerae*]	100
2	28	260	27587.7	3-Oxoacyl-[acyl-carrier-protein] reductase [*Neisseria sp.* HMSC06F02]	99
3	26973	249	26577.3	KR domain-containing protein [*Veillonella dispar]*	99
4	3701	254	26486.9	Acetoin reductase [*Mycobacterium abscessus subsp. abscessus*]	100

5	7073	249	27324.3	SDR family NAD(P)-dependent oxidoreductase [*Streptococcus sp.* HMSC072G04]	99
6	1819	261	27854.7	3-Oxoacyl-ACP reductase [*Megasphaera micronuciformis]*	99
7	8450	254	27118.8	SDR family NAD(P)-dependent oxidoreductase [*Veillonella sp.* ICM51a]	99
8	35929	231	25107.5	3-Ketoacyl-ACP reductase [*Peptoniphilus lacrimalis* DNF00528]	99
9	25723	249	26169.5	2,3-Dihydro-2,3-dihydroxybenzoate dehydrogenase [*Rothia sp.* HMSC065C12]	99
10	2262	267	28773.5	Short-chain dehydrogenase [*Mycobacterium tuberculosis*]	100
11	2625	277	30369.9	Oxidoreductase [*Rothia sp.* HMSC071C12]	99
12	32846	340	35604.4	KR domain-containing protein [*Isoptericola variabilis*]	99
13	7745	249	26719.4	KR domain-containing protein [*Sanguibacter keddieii*]	99
15	2725	191	20030.6	SDR family NAD(P)-dependent oxidoreductase [*Atopobium sp.* BS2]	99
16	272	238	25500.3	Beta-ketoacyl-ACP reductase [*Rothia mucilaginosa*]	99
17	5881	244	25744.6	3-Oxoacyl-[acyl-carrier-protein] reductase [*Streptococcus parasanguinis*]	99
18	714	274	30434.6	KR domain-containing protein [*Sanguibacter keddieii*]	99
19	39036	266	28512.5	SDR family NAD(P)-dependent oxidoreductase [*Atopobium sp.* BS2]	99
20	17633	247	27193.9	Oxidoreductase, short chain dehydrogenase/reductase family protein [*Prevotella melaninogenica* D18]	98
21	4991	248	26121.9	3-Oxoacyl-[acyl-carrier-protein] reductase [*Prevotella sp.* F0091]	99
22	32000	280	31201.2	Short chain dehydrogenase [Prevotella sp. C561]	99
23	5299	242	26812.1	KR domain-containing protein [*Prevotella sp.* ICM33]	99
24	11819	256	28676.4	KR domain-containing protein [Streptococcus parasanguinis]	93
25	609	286	31472.9	Enoyl-ACP reductase [*Prevotella pallens*]	99
26	440	242	26125	Putative 3-oxoacyl-(acyl-carrier-protein) reductase [*Neisseria flavescens* SK114]	99
27	8107	273	30216.4	enoyl-ACP reductase [*Prevotella salivae*]	99
28	18110	241	26393.5	Oxidoreductase, short chain dehydrogenase/reductase family protein [*Oribacterium sinus* F0268]	94

29	3555	248	26477.21	NAD(P)-dependent oxidoreductase [*Veillonella atypica*]	99
30	11294	276	29831.9	Enoyl-(acyl-carrier-protein) reductase family protein [*Porphyromonas sp.* KLE 1280]	100
31	8366	237	24707	NAD(P)-dependent oxidoreductase [*Actinomyces graevenitzii*]	97
32	66	249	27373.1	KR domain-containing protein [Prevotella sp. ICM33]	99
33	7731	241	26188	3-Oxoacyl-ACP reductase FabG [*Prevotella sp.* ICM33]	99
34	1480	249	26491.3	KR domain-containing protein [*Prevotella histicola*]	99
35	225	261	27931.9	Enoyl-[acyl-carrier-protein] reductase [*Haemophilus sp*. HMSC073C03]	99
36	34673	235	25663.2	KR domain-containing protein [Prevotella melaninogenica]	99
37	3780	249	27377.1	SDR family NAD(P)-dependent oxidoreductase [*Haemophilus parainfluenzae*]	100
38	15028	245	25623.6	3-Oxoacyl-[acyl-carrier-protein] reductase [*Veillonella sp.* ICM51a]	99

Table 7: Functional assignments of the amino acid sequences of the retrieved SDRs from the tongue metagenome using NCBI BLAST. Protparam tool was used to predict enzyme featues.[121,124]



Scheme 15: Predicted substrates for SDR-4 and SDR-9, and the literature reactions associated with the homologous enzymes: acetoin 33 and 2,3-dihydro-2,3-dihydroxybenzoate 34.[122,123]

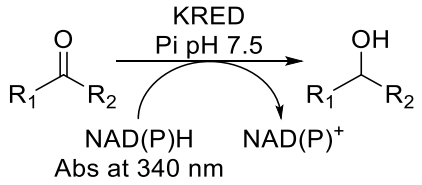
2.3 Initial activity assays, assay optimisation and selection of cofactor

To assess the activity of the 37 SDR against selected ketones and aldehydes, the cells were cultured from glycerol stocks† in parallel. Assays were performed using either the clarified cell lysates or purified proteins (concentrations displayed in

† Glycerol stocks prepared by Jack Jeffries.

Section 8.6 and 8.7) in parallel using 96-well plates and using a plate-reader detection method, which had been performed in the literature. [6,73,125,126]

Plasmids containing genes selected from metagenomics sources were transformed into BL21 *E. coli* expression cells. The assay protocol used the clarified cell lysate (10% v/v), 0.1 M sodium phosphate buffer (pH 7.2), 1mM NAD(P)H and 0 to 15 mM substrate, added in that order. These were incubated for 100 minutes and the NAD(P)H depletion was monitored using UV-visible spectrophotometer at 340 nm. The λmax of NAD(P)H is 340 nm and this decreases when the cofactor is oxidised, and therefore this indicated enzymatic activity as it turns over the substrate (Scheme 16).[127] Table 8 shows change in absorption over time with darker green shades representing larger changes and indicting higher activity.



Scheme 16: Reaction of SDR assay. NAD(P)H is detected at 340 nm, and its depletion indicates reduction of the ketone. Pi denotes phosphate buffer.

This initial screen, though useful as an indication of SDR activity, had some limitations. When growing cell cultures on a small scale in parallel, as in this case, it is not possible to take an OD_{600} reading for all the cultures, and a representative sample was grown using the same procedure. This can cause differences in the number of cells produced and the enzyme concentrations in the assay. NAD(P)H can be depleted during the assay due to background reactions and degradation, so a negative control without the addition of the substrate was performed and the change in absorption deducted.[128] NAD(P)H is also very sensitive to moisture and over time its ability to absorb at 340 nm is depleted. The assays were also completed in triplicate. Additionally, controls using no clarified cell lysate at all were performed, shown in Appendix Section 8.9. This supported the results showing that the addition of the enzymes is associated with the changes. A concentration of 1 mM NADPH was used to give the enzyme more cofactor than if sometimes used in literature spectrophotometric assays.

Initially, all 37 SDR enzymes were screened against tropinone **7** at concentrations of 5 to 15 mM using the SDRs as clarified cell lysate (Column 1, Table 8). No significant change in absorption was recorded when the negative control was deducted. There was a large background activity for some enzymes however: there was a change in absorption over time even without a negative control, as shown in Figure 13 as an example.

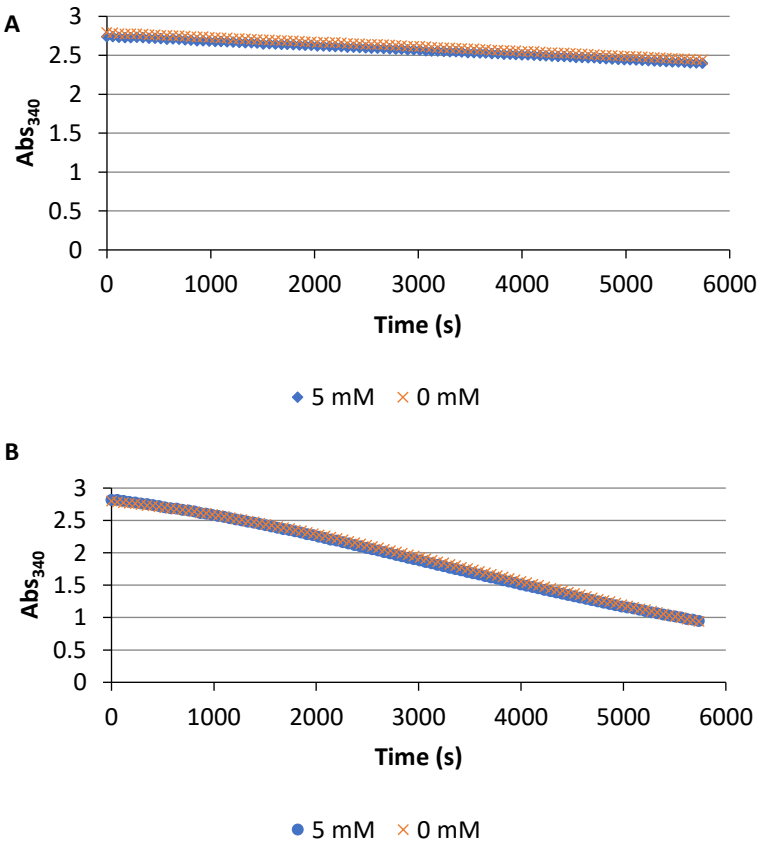
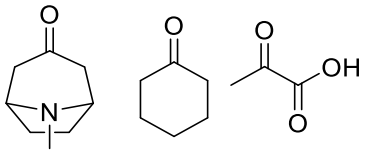


Figure 13: A graph to show change over time of (A) SDR-1 and (B) SDR-3 with and without the tropinone substrate (200 µL): tropinone (0-5 mM), clarified cell lysate (10% v/v, 0.1-0.2 mg/mL), NADH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were shaken for 100 min, at 25 °C and quantified using the spectrophotometer at 340 nm.

	Column 1			Column 2		Column 3					
	Cell lysate			Purified enzyme		Purified and desalted enzyme					
	NADH			NADH		NADH			NADPH		
SDR	5 mM Tropinone	10 mM Tropinone	15 mM Tropinone	5 mM Tropinone	5 mM Cyclohexanone	5 mM Tropinone	5 mM Cyclohexanone	5 Mm Pyruvic acid	5 mM Tropinone	5 mM Cyclohexanone	5 Mm Pyruvic acid
1	-0.04	-0.06	-0.02	0.06	0.04	0.08	0.03	-0.02	0.04	0.01	0
2	0.02	-0.02	-0.08	-0.01	-0.01	0.09	0.03	-0.04	0.06	0	0

3	0.33	0.02	-0.43	0.09	-0.28	-0.01	0.01	0.04	0.01	0.06	2.16
4	0.08	-0.1	-0.47	0.03	-0.01	-0.05	0.54	0.04	-0.02	0	-0.01
5	-0.06	-0.08	-0.07	0	0.07	-0.26	0.03	2.16	0.6	0.31	1.05
6	-0.09	-0.1	-0.1	0	-0.01	-0.07	0.02	0.91	-0.03	0	0
7	-0.09	-0.13	-0.15	-0.04	0	-0.09	0.03	-0.03	-0.01	-0.02	-0.01
8	0.04	-0.06	-0.31	0.07	0.05	-0.06	0.15	0.06	-0.04	-0.04	-0.01
9	0.07	-0.11	-0.56	-0.01	-0.09	-0.04	-0.08	-0.1	0.01	0.02	-0.16
10	0.02	-0.15	-0.39	0.08	0.15	-0.08	0	0.3	0.03	0.08	0.18
11	-0.01	-0.23	-0.43	0	0.01	-0.05	0.02	-0.02	0.05	0.06	-0.12
12	0.45	-0.05	-0.48	0	0.03	0.14	0.02	0.34	0	0.05	0.03
13	-0.14	-0.22	-0.19	-0.11	-0.02	-0.08	0.11	0.02	-0.01	0.02	0.23
15	0.31	0.06	-0.29	-0.03	0.01	0.03	0.02	0.8	-0.01	0.01	-0.02
16	-0.23	-0.1	-0.12	-0.05	0	-0.08	0	-0.02	-0.05	0.22	-0.02
17	0.36	-0.06	-0.26	0.2	-0.1	-0.04	0.03	0.06	-0.02	1.12	0
18	0.27	-0.21	-0.21	-0.02	0.02	-0.05	0.02	0.24	-0.02	0.01	0.04
19	-0.01	-0.08	-0.26	-0.01	-0.01	-0.1	0.02	1.33	-0.01	0	-0.02
20	0	-0.02	-0.28	0.11	0.02	-0.04	0.03	0.02	-0.01	0.01	-0.01
21	0.42	0.1	-0.18	-0.08	0.05	-0.14	0.02	1.04	-0.02	0	-0.02
22	0.07	0.07	-0.06	0	-0.01	-0.12	0.05	1.9	-0.01	-0.01	-0.11
23	0.04	0.07	-0.08	-0.01	-0.01	-0.12	0.13	0.94	0.02	0	-0.02
24	0.17	0.1	-0.07	0.03	0	0.19	-0.01	1.12	-0.05	0	-0.06
25	0.19	-0.46	-0.06	0.07	0.02	0	0.1	0.6	-0.03	-0.04	0
26	-0.24	-0.22	-0.2	0	0	0	0.02	0.34	-0.06	0	-0.04
27	0.6	0.08	-0.12	-0.01	0.01	0	0.03	0.5	-0.01	0.01	-0.03
28	0.07	-0.19	-0.14	0.07	0.02	-0.02	0.01	-0.02	-0.02	0	-0.01
29	-0.16	-0.24	-0.15	0	-0.01	0.08	0.04	-0.03	0.08	0.01	-0.01
30	0.21	-0.18	-0.36	0.22	0.01	-0.01	0.01	-0.01	-0.03	0.02	0
31	-0.13	-0.18	-0.14	0.11	-0.03	0	0.02	0.86	-0.03	0.82	-0.02
32	-0.2	-0.26	-0.24	-0.01	0.01	0	0.03	-0.01	-0.03	0.01	-0.03
33	0	-0.27	-0.43	-0.05	-0.03	0.01	0.01	0	-0.02	0.05	-0.01
34	-0.07	-0.14	-0.07	0.04	0.01	0	0.03	-0.03	-0.01	0.01	0.01
35	0.2	-0.15	-0.25	-0.02	0.01	-0.06	0.08	0.35	-0.03	0.01	-0.02
36	-0.08	-0.48	-0.6	0.05	0	0.3	-0.07	-0.1	0.1	0.02	0.06
37	-0.33	-0.34	-0.39	0	0.01	0	0.12	0.04	-0.02	0.88	2.36
38	0.12	0.26	-0.72	0.02	0.02	-0.01	0.04	1.78	-0.05	-0.05	-0.02

Table 8: Table to show the change in absorption at 340 nm after 100 min between no substrate and at various concentrations of substrate (200 µL): substrate (0-15 mM), enzyme (10% v/v, 0.1-0.7 mg/mL), NADP(H) (1 mM), Pi (100 mM, pH 7.2). The reactions were shaken for 100 min, at 25 °C, and quantified using the spectrophotometer at 340 nm. The larger the change (showing activity) over the control, the darker green the box. The substrates are shown in Figure 14.



7 **19** **35**

Figure 14: Substrates used in Table 8.

To reduce potential background reactions from other enzymes present in the cell lysates, the enzymes were then purified using His-tag Ni-NTA spin columns. To further establish the protocol, cyclohexanone **19**, which has shown to be active with

many other SDRs in the literature, was used as another substrate. All the enzyme activity towards tropinone and cyclohexanone was diminished (Column 2, Table 8).

At this stage in the initial assays it was unclear whether the problem was due to the protocol or the inherent enzyme activity. The enzymes were therefore buffer exchanged into sodium phosphate buffer pH 7.4, as the imidazole in the elution buffer used during purification may have been affecting activity and altering the pH to beyond working conditions. Enzyme expression was demonstrated with SDS-PAGE gels. All the SDRs were expressed and purified successfully, though some in low concentrations. Figure 15 showed three SDRs as an example, and the remainder are shown in Appendix Section 8.7. A Bradford assay was also used to estimate the purified and desalted protein concentration as approximately in the range of 1-2 mg/mL. The assay diluted ten-fold from this stock concentration, making the final concentration used 0.1 – 0.2 mg/mL (Section 8.7).

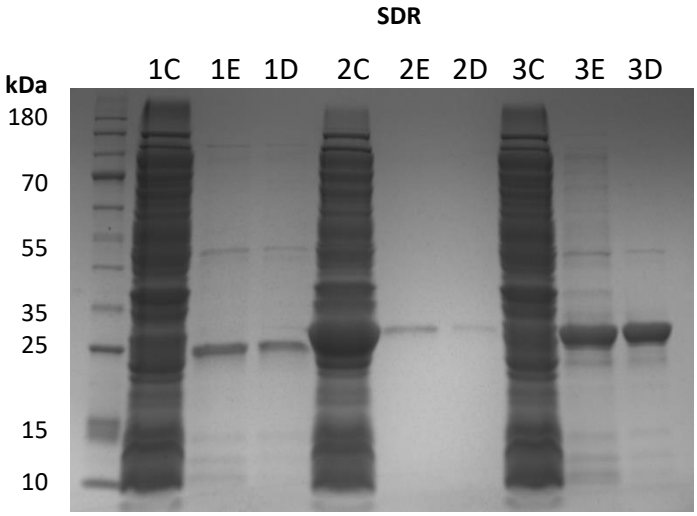


Figure 15: Example SDS-PAGE gels of SDR-1 (26.1 kDa), SDR-2 (28.6 kDa). SDR-3 (26.6 kDa). C = cell lysate; E = elution from purification column; D = desalted enzyme.

These purified and buffer-exchanged enzymes regained their activity and SDR-4 was found to be active toward cyclohexanone using NADH as at cofactor (Column 3, Table 8). Pyruvic acid **35** was also used as a standard substrate, and activity towards this substrate was noted for 13 of the enzymes.

The enzymes may be involved in anabolic functions and likely to use NADPH as the cofactor metabolic functions, as suggested by the NCBI annotation (Table 7).

NADPH was therefore also included in the assay (Column 3, Table 8). When NADPH is applied as the cofactor, four of the SDRs were active towards cyclohexanone (SDR-5, SDR-17, SDR-31 and SDR-37) and three towards pyruvic acid (SDR-3, SDR-5 and SDR-37). The natural substrates of the enzymes were likely dissimilar to those tested, and without further optimisation of reaction conditions and substrate specificity, the activities of the other enzyme have not been revealed.

Overall, significant activity was found using NADH with SDR-4 and with NADPH with SDR-3, SDR-5, SDR-11, SDR-17, SDR-31, and SDR-37. Additionally, previous assays completed by Jack Jefferies (results not shown), highlighted SDR-11 and SDR-23 as interesting. Although many enzymes were active towards pyruvic acid, this was not used as a selection method as the project focused on cyclic substrates. These enzymes (3, 4, 5, 11, 17, 23, 31, and 37) were therefore taken forward for wider substrate specify testing.

SDR-17 and SDR-31 were both assigned as 3-oxoacyl-(acyl-carrier-protein) reductases (EC 1.1.1.100), using Clustal Omega (Table 7).[17] Toomey *et al.* showed that one of these enzymes shows optimal activity between pH 6 and 7.[129] The other selected enzymes aligned to oxidoreductases. When compared to each other, the selected enzymes did not display high sequence homologies to each other, except for SDR-3, SDR-5 and SDR-37 (Figure 16), fulfilling the aim of the project to have high variation in the primary sequence of the enzymes.

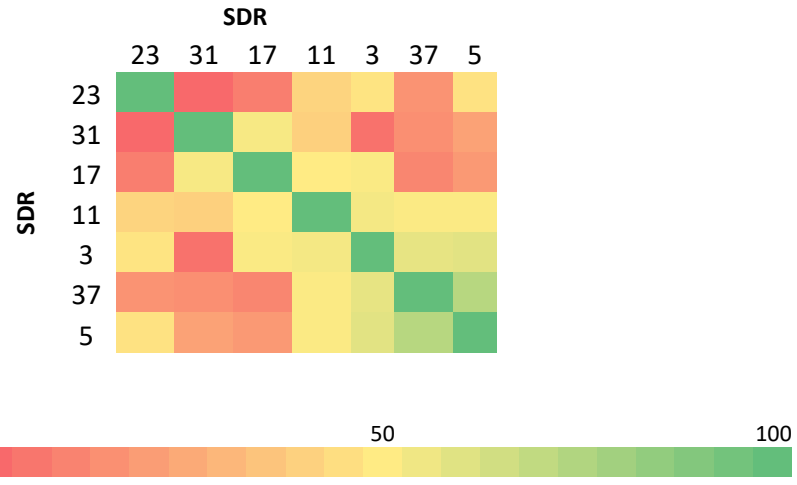
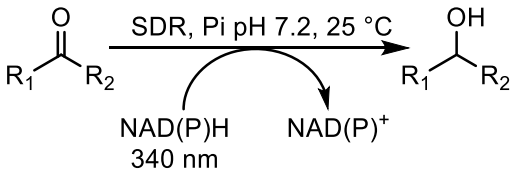


Figure 16: Heatmap of sequence identities (%) of the SDRs from the tongue metagenome that were selected for further investigation as calculated by Clustal Omega online tool.[17121] The colour key is displayed showing overall percent sequence identity.

2.4 Substrate specificities

The spectrophotometric assay was an efficient method to test large numbers of SDRs and substrates, so was further used to test the selected SDRs with a broad range of substrates (Scheme 17, Table 9 and Figure 18). As discussed in Section 2.3 there were problems associated with the spectrophotometric assay, therefore the assay was used as an initial indicator of activity. Likewise, in Section 2.3, activities were calculated from the change over time in absorbance at 340 nm and the negative water/DMSO control was deducted. The percent change in NADPH concentration was calculated from the absorbance at 340 nm ($\xi_{NAD(P)H}$ 622 $M^{-1}mm^{-1}$ and pathlength 4.55 mm). Control assays were also conducted without the presence of an enzyme (Appendix, Section 5.15). DMSO was added as a co-solvent to aid solubilisation of the substrates.



Scheme 17: Ketone reduction using SDRs was monitored by measuring absorbance at 340 nm to determine NAD(P)H depletion.

Clarified cell lysate was employed in this assay as opposed to the purified enzymes. Clarified cell lysate has the advantage of not requiring additional purification steps, so is more industrially relevant and faster to prepare when testing multiple enzymes. The enzymes exhibited high levels of expression, except for SDR-23 (Figure 17). After expression and cell lysis, the clarified cell lysates were aliquoted out in to 96-well plates, and flash frozen in liquid nitrogen for storage. Typical total protein concentrations are displayed in Section 8.6.

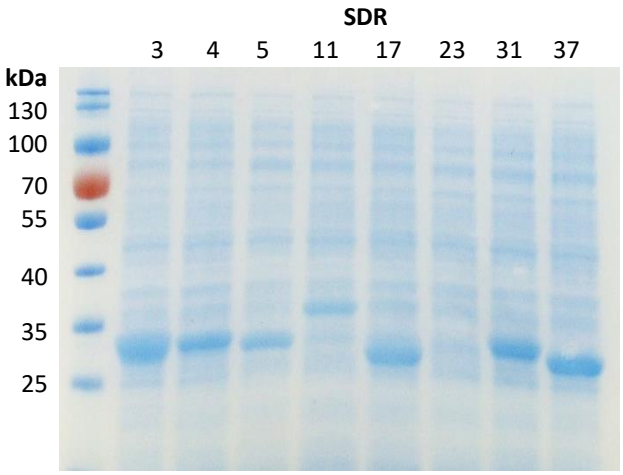


Figure 17: An SDS polyacrylamide gel of clarified cell lysates of a selection of SDRs of interest (labelled at the top of the gel). Enzyme 23 was poorly expressed. Protein sizes in kDa calculated using the online tool Expasy ProtParam as follows: SDR-3 (29.5 kDa), SDR-4 (29.4kDa), SDR-5 (30.2 kDa), SDR-11 (33.2 kDa), SDR-17 (28.5 kDa), SDR-23 (29.7 kDa), SDR-31 (27.4 kDa), SDR-37 (30.2 kDa).[124]

Tropinone was not shown to be tolerated by any of the tested SDRs. Hence, piperidone compounds which mimicked tropinone without the di-carbon bridge (**21**, **36**, and **37**) where used. Cyclic and bicyclic aliphatic and aromatic structures used to probe the substrate tolerance of the SDRs. Some fluorinated compounds were also tested, as these have important applications in pharmaceuticals, agrochemicals, radiotracers and high performance materials.[130]

In general, SDRs 4, 17, 31 and 37 were tolerant to a wide range of aromatic and aliphatic compounds, including those with a fluorine group. Enzyme 3 showed lower activities towards aromatic cyclic compounds and did not accept aliphatic cyclic compounds. Enzyme 23 showed very little activity with any of the substrates, perhaps due to low expression of enzyme 23 (Figure 17). Enzyme 5 also exhibited poor activity with all of the tested substrates.

		Conversion (%)							
		NADPH							NADH
		3	**5**	**11**	**17**	**23**	**31**	**37**	**4**
19	Cyclohexanone	2	1	5	16	2	54	54	4
38	2-Methylcyclohexanone	4	0	3	58	3	60	27	45
39	3-Methylcyclohexanone	8	2	5	58	2	64	55	31
40	2-Cyclohexen-1-one	4	3	9	8	7	11	11	0
51	3-(2-Oxocyclohexyl)acetaldehyde	nd	nd	nd	34	nd	23	52	19
36	1-(2-Methyl propyl)-4-piperidone	0	0	61	0	0	32	51	0
37	Ethyl-4-oxo-1-piperidone carboxylate	66	6	66	64	2	66	58	24

21	*N*-Methyl-4-piperidone	4	0	7	-3	0	11	51	0
29	Acetophenone	0	0	2	0	0	6	10	31
49	1-Phenyl propan-1-one	17	2	0	22	4	43	0	5
50	1-Phenyl propan-2-one	14	0	59	0	0	0	0	2
41	2,2,2-Trifluoroacetophenone	28	1	4	66	2	70	50	35
43	2,2,2,4′-Tetrafluoroacetophenone	22	1	0	53	1	59	49	8
42	2,2,2,3′-Tetrafluoroacetophenone	22	1	2	53	4	59	38	25
44	4'-*N*-Propyl-2,2,2-trifluoroacetophenone	2	0	0	34	0	11	8	7
52	1-Indanone	17	0	0	0	0	0	0	0
53	2-Indanone	0	0	0	0	0	0	0	0
54	Wieland–Miescher ketone	nd	nd	nd	51	nd	59	37	10
45	Benzaldehyde	0	0	68	30	0	40	54	16
46	Phenylacetaldehyde	23	3	0	0	0	0	0	0
47	2-Heptanone	20	2	33	0	13	0	0	26
48	Methyl acetoacetate	17	0	50	0	0	0	23	0

Table 9: Table to show the conversion of NADPH (Abs_{340}) after 100 min (200 µL): substrate (5 mM), enzyme (10% v/v, 0.1-0.2 mg/mL), NAD(P)H (1 mM), Pi (100 mM, pH 7.2), DMSO (10% v/v). The reactions were shaken for 100 min, at 25 °C and quantified using the spectrophotometer at 340 nm. The larger the change (showing activity) over the control, the darker green the box. Performed in triplicate with standard deviation <10%. Italic font denotes work completed by Jack Jeffries and "nd" denotes activities not determined. Figure 18 shows the structures of the compounds tested.

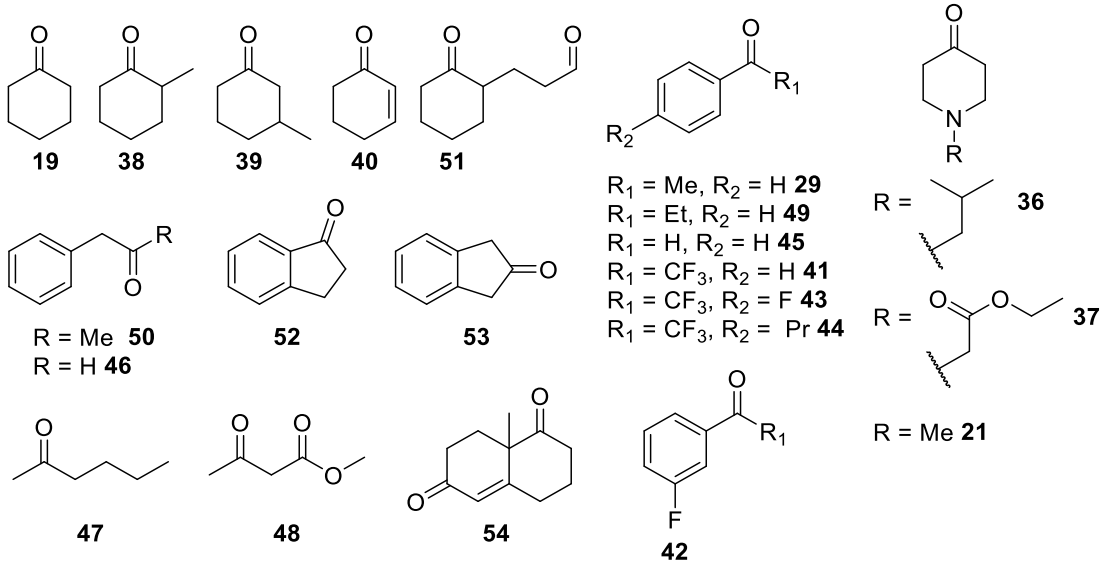


Figure 18: Structures of compounds tested against SDRs and later the TRs and MecgoR.

The 6-membered aliphatic ring systems were readily accepted by SDR-17, SDR-31, SDR-37 and SDR-4. The addition of a methyl group in substrates 2-methylcyclohexanone **38** and 3-methylcyclohexanone **39** increased activity of SDRs-4 and SDR-17 compared to activity towards cyclohexanone **19**. None of the tested enzymes were active towards 2-cyclohexen-1-one **40**, suggesting that the enzymes

cannot accept enones. Piperidone systems **21**, **38** and notably **39** were also generally accepted by the SDRs tested. 1-(2-methyl propyl)-4-piperidone **38**, which has an aliphatic group on the nitrogen, was poorly accepted by many of the enzymes. Ethyl-4-oxo-1-piperidone carboxylate **39** has a more polar *N*-methyl group and was widely accepted with high activities by most of the enzymes. The applications of **39** have not been widely reported on in the literature, but Uruno *et al.* reported in 2015 that a subtype-selective activator of muscarinic acetylcholine receptors can be synthesised from **39** connected with an aromatic ring *via* an amine linker. This could be used as a treatment for schizophrenia.[131]

Though acetophenone **29** was not well accepted by the enzymes, **41**, **42**, and **43**, were readily accepted by SDRs-3, 17, 31, 37 and 4. The compound 4'-*N*-propyl-2,2,2-trifluoroacetophenone **44**, which has an alkyl tail, was only readily accepted by SDR-17, and the activity was reduced compared to the other trifluoroacetophenone substrates for the other enzymes.

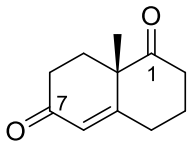
In terms of aldehyde tolerance, many of the enzymes tested accepted benzaldehyde **45**, but not phenylacetaldehyde **46**. The linear substrates tested, **47** and **48**, were accepted by some enzymes including SDRs-3, 11 and 37. Aromatic compounds 1-phenyl propan-1-one **49** and 1-phenyl propan-2-one **50** were accepted by only SDR-3, SDR-17 and SDR-31 and SDR-3 and SDR-11 only. Compound 3-(2-oxocyclohexyl)acetaldehyde **51** was also only accepted by SDR-4-, SDR-17, SD-31 and SDR-37.

Interestingly, SDR-4 portrays a 100% sequence homology to acetoin reductases from many S*treptococcus* species using NCBI BLAST, which have been shown in literature to reduce acetoin, diacetyl and 2,3-pentane-dione.[121,132] Although acetoin was not tested, SDR-4 did not readily accept such as pyruvic acid, which has a similar size but may be too polar, or 2-heptanone, which has an alkyl chain.

SDR-17 and SDR-31 displayed 100% sequence identity to 3-oxoacyl-ACP reductases (EC 1.1.1.100) using NCBI BLAST.[121] 3-Oxoacyl-ACP reductases or β-keto-acyl ACP reductases have been shown to exhibit a wide specificity with respect to the chain length of the β-ketoacyl group.[129] They are involved in the first reductive step in Type II fatty acid synthesis (FAS), are a classical SDR and display a GGxGxxG NAD(P) binding motif and an altered active site motif YxxxN (Section 1.2) (Fungal type

ketoacyl reductases have a TGXXXGX$_{(1-2)}$G NAD(P)-binding motif).[32] When the residue sequence of SDR-17 is searched on NCBI Blast, there are 28 logged proteins of greater than 98% sequence identity, all of which are from the *Streptococcus* genus. Many of the retrieved sequences were annotated with the active site N112, S140, Y153 and K157, and this motif was found in over 160 of the retrieved most similar sequences. SDR-17 and SDR-31 also had wide substrate specificities, though no β-ketoacyl-ACP compounds were tested. Methyl acetoacetate **48**, which is closely related, was tested but neither SDR-17 or SDR-31 demonstrated activity towards this compound.

The screened enzymes did not readily accept 1- and 2-indanone **52** and **53** respectively. Notably an interesting result was that the Wieland–Miescher ketone (WMK) **54** was readily accepted by SDRs-17, 31 and 37 (Figure 19). It was proposed that the non-conjugated ketone (C-1) was more likely to be reduced, as the enzymes displayed low activities towards 2-cyclohexen-1-one **40** possessing an enone.



51

Figure 19: Wieland-Miescher ketone (*S*)-55.

2.5 Determination of enantiomeric purity

The spectrophotometric assay with acetophenone **29** as a substrate showed NADPH depletion was approximately 3:1 ratio for enzymes 4 and 37 respectively (Table 9). To determine the enzyme stereoselectivities when using SDR-4 and SDR-37, the reaction was scaled to 500 µL, the product of the reaction extracted and a Chiracel OD column was used for HPLC analysis using commercially available acetophenone and (*R*)- and (*S*)-phenylethanol **(*R*)-55** and **(*S*)-55** as standards.[133] Both enzymes 4 and 37 produced solely (*S*)-phenylethanol, with conversion to a product concentration of 29% and 12% with respect to the 1 mM NAD(P)H used (Figure 20 and Scheme 18). This obeys Prelog's stereoselectivity rule. The stereoselectivities of the SDRs towards trifluroacetophenone **41** was also attempted, but sufficient peak separation was not found (Figure 21).

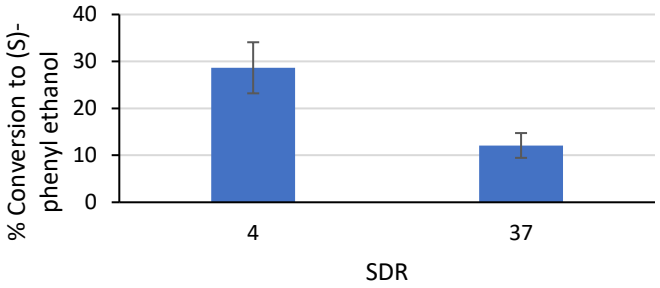
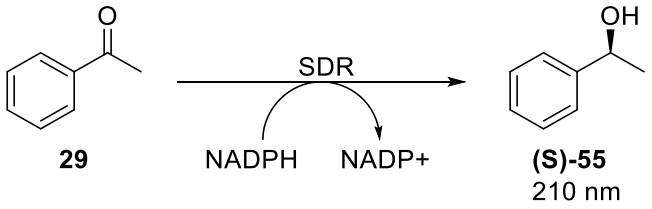


Figure 20: Graph showing conversion to (*S*)-phenylethanol (500 µL): acetophenone (5 mM), enzyme (10% v/v, 0.1-0.2 mg/mL), NAD(P)H (1 mM), Pi (100 mM, pH 7.2), DMSO (10% v/v). The reaction was performed in duplicate and shaken for 100 min, at 25 °C. The reaction was extracted into Et_2O and quantified by HPLC.



Scheme 18: The reduction of acetophenone 29 to (*S*)-phenyl ethanol using SDR-4 or SDR-37.

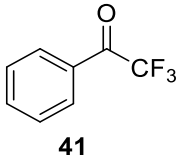


Figure 21: SDR substrate trifluroacetophenone 41.

2.6 Reductions of Wieland-Miescher Ketone

2.6.1 Introduction

The Wieland-Miescher ketone (WMK, **54**) is a bicyclic diketone and a structural motif present in over 50 natural products, including a great number of medically useful compounds (Figure 22).[134] Chemoselective reduction at the C-1 carbonyl to the **(4a*S*,5*S*)-56** and (**4a*R*,5*R*)-56** using $NaBH_4$ or $Zn(BH_4)_2$ has been reported in the literature.[135,136] However, a chemical reduction to **(4a*S*,5*R*)-56** or **(4a*R*,5*S*)-56** has not been described (Figure 23). Though C-7 reduction is less facile, it can be achieved *via* initial protection of the C-1 carbonyl.[134]

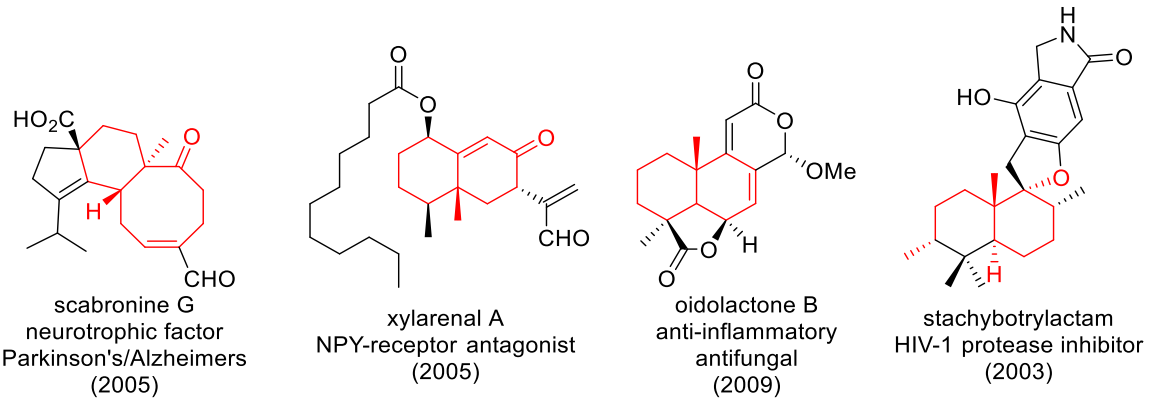


Figure 22: Selected example of application of the WMK motif.[134]

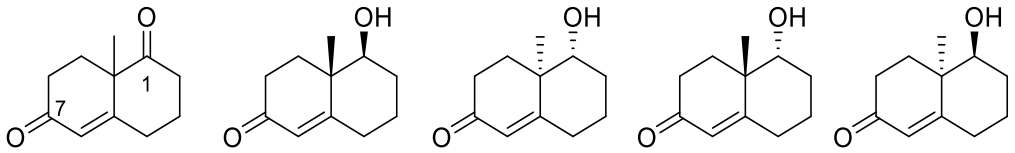


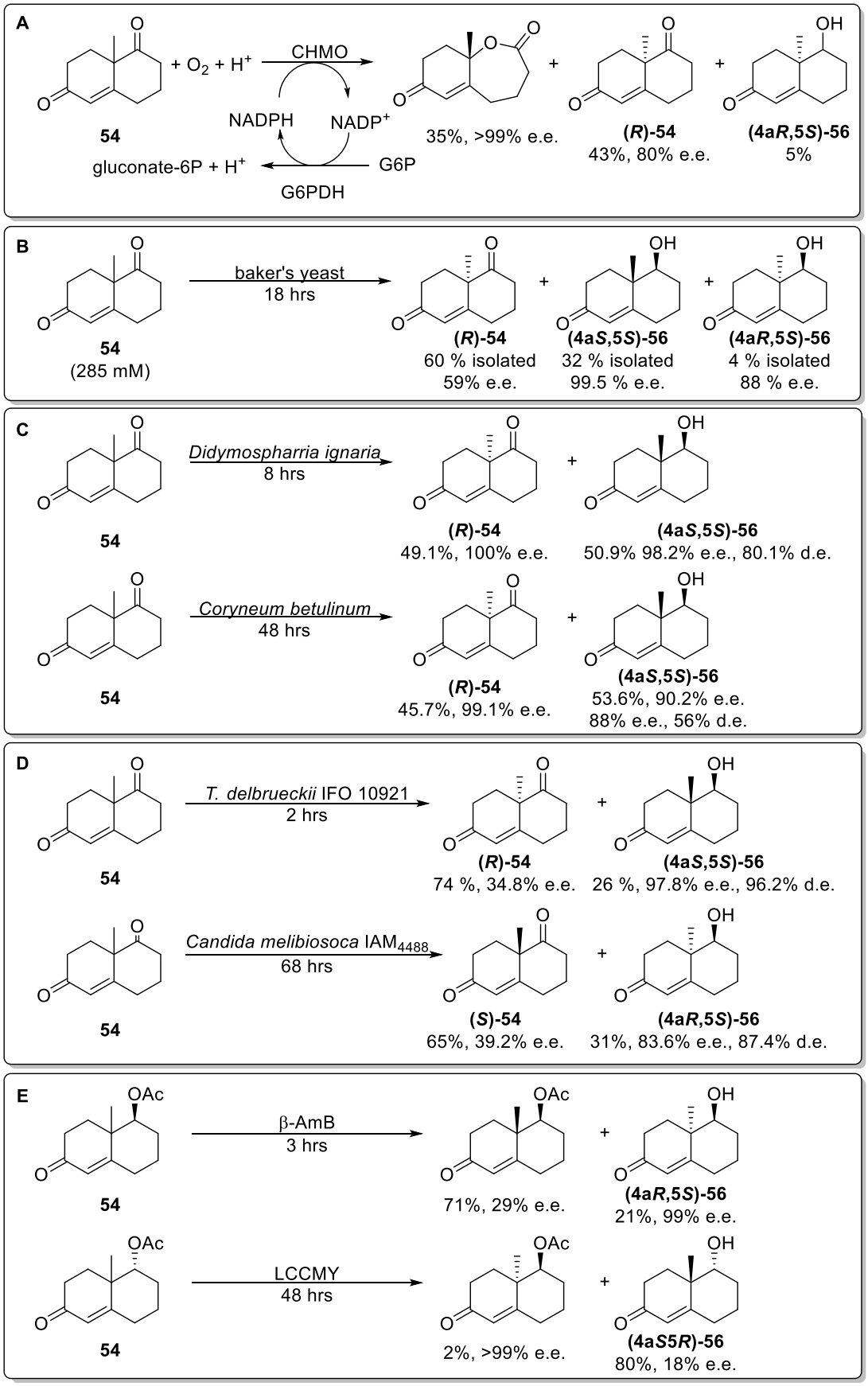
Figure 23: The WMK, and diastereomers of 5-hydroxy-4a-methyl-4,4a,5,6,7,8-hexahydronaphthalen-2(3H)-one.

There has been limited investigation into the use of biocatalysts to perform the reduction of **54**, though the reduction the of analogues and derivatives has been reported, with varying yields and enantioselectivities.

A number of biocatalytic reductions have been described, for example, Ottolina *et al.* reported the use of a mono-oxygenase to generate **(4a*R*,5*S*)-56** and (**4a*R*,5*R*)-56** in a total yield of 5% (Scheme 19A).[137] Yeast has also been used to generate varying products (Scheme 19B,C and D).[138–140] An amylase and a lipase have also been used to generate **(4a*R*,5*S*)-56** and **(4a*S*,5*R*)-56** (Scheme 19E).[141]

Preparative scale reactions have also been employed. Hioki *et al.* used 4500 g/L of baker's yeast (Scheme 19B). Fushuku *et al.* reduced enatiomerically pre-enhanced forms of WMK using *T. delbrueckii IFO* 10921 at a 2 L scale: (*R*)-WMK (70% e.e.) was reduced to **(4a*S*,5*S*)-56** in 18% yield (1.84 g) and 72.0% d.e.. (*S*)-WMK (70% e.e.) was reduced to **(4a*S*,5*S*)-56** in 78% yield (7.91 g) and 94.4% d.e.. Both reactions required the use of flash column chromatography and recrystallization to obtain the enantiomerically pure product. Shimizu *et al.* also reports two lager scale reactions generating **(4a*R*,5*S*)-56** in 21% yield, (99% e.e.) using β-amylase type VIII-A from

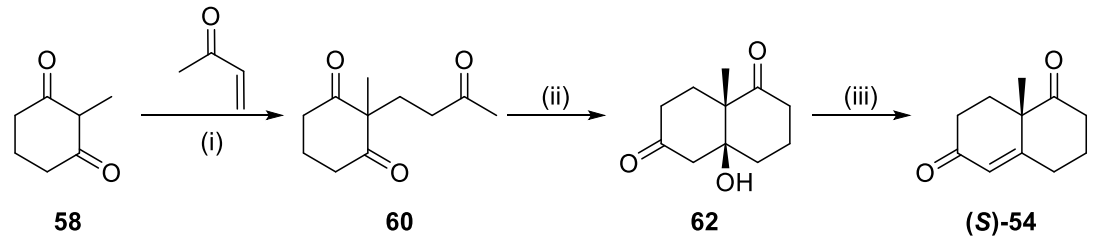
barley, and **(4a*R*,5*S*)-56** in 48% yield (93% e.e.) using a lipase from *Candida cylindacea*, although this required a reaction time of 54 hours.



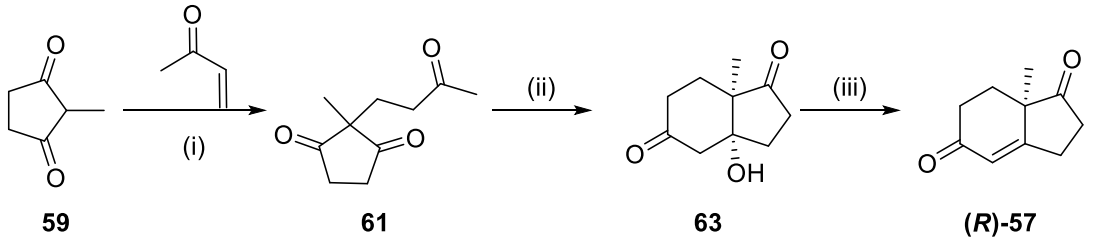
Scheme 19: Literature reductions of WMK: (A) Ottolina *et al.* using flavoprotein cyclohexanone monooxygenase (CHMO) from *Acinetobacter calcoaceticus* NCIMB 9871;[137] (B) Hioki *et al.* using baker's yeast;[138] (C) Janeczko *et al.* using yeast;[139] (D) Fuhshuku *et al.* using yeast;[140] (E) Shimizu *et al.* using β-amylase type VIII-A (β-Am-B) from barley.[141] When a range of conditions was used to generate various yields and enantioselectivies, the yields highlighted in the text or those with the highest e.e. has been highlighted in the scheme.

2.6.2 Synthetic route to WMK and SDR activities

The WMK **54** and the 6,5-membered bicyclic analogue the Hajos-Parish ketone (HPK, **57**) can be synthesised in an enantioenriched fashion using Proline or other asymmetric catalysts in the Hajos-Parrish-Eder-Sauer-Wiechert reaction (Scheme 20 and Scheme 21).[142–144] Starting from commercially available **58** and **59**, triketones **60** and **61** were generated. (*S*)-(-)- and (*R*)-(+)-proline was used to catalyse the asymmetric aldol reaction to make **62** or **63**, followed by loss of water to make **(*S*)-54** or **(*R*)-57**. The selected SDRs were assayed for activity towards the intermediates in the formation of the WMK **54** and HPK **57**. If there was low activity, then a one-pot process for synthesis and subsequent reduction could be considered in future investigations. A preliminary investigation of the feasibility of this was conducted.



Scheme 20: Synthetic routes to the WMK 54: (i) aq. AcOH, 72-74 °C, 1 h, 62%; (ii) (*S*)-(-)-proline, DMSO, 25 °C, 120 h, 64%; (iii) H_2SO_4, DMF, 95 °C, 1 h, 58%, 71% e.e

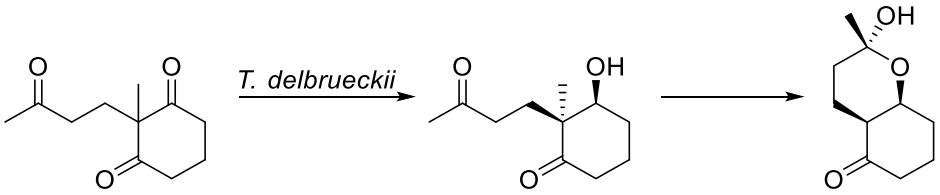


Scheme 21: Synthetic routes to the HPK 57: (i) aq. AcOH, 75 °C, 14 h, 91%; (ii) (*R*)-(+)-proline, DMSO, 25 °C, 5 d, 29%; (iii) H_2SO_4, DMF, 95 °C, 1 h, 44%, >99% e.e.

The activities of the SDRs with the WMK and HPK derivatives are shown in (Table 10 and Table 11), monitoring NADPH consumption by measuring absorption

at 340 nm with substrate (5 mM), NADPH (1 mM) in phosphate buffer (100 mM, pH 7.2) at pH 7.4. As before, the negative control without the addition of a substrate was deducted from change in absorption over time. As NAD(P)H is the limiting reagent, an NAD(P)H concentration curve was used to calculate change in NAD(P)H concentrations and show conversion relative to this from the change in absorbance. The substrates consisted of compounds **60**, **62**, **61**, and **63** which were synthesised, and commercially purchased compounds **58**, **(*S*)-54**, **59**, and **(*R*)-54**.

SDR-3 and SDR-5 did not display activity to any of the tested compounds. SDR-11 accepted **60** and **62** and did not reduce the WMK **54**. Notably, SDR-17 only accepted the (*R*)-WMK **(*R*)-54** out of all 5 compounds tested in Table 10, making this a possible target for future investigations. SDR-31 was more promiscuous, and showed activity across many of the intermediates tested (**60**, **(*S*)-54**, **(*R*)-54,** and **61**). SDR-37 showed similar activity to SDR-17, but with lower conversions. **58** was not tolerated by any of the SDRs tested. Fuhshuku *et al.* has discussed reduction of the triketone **60** using a *T. delbrueckii* strain (Scheme 22), which give **64**.[140,145]



60 **64**

Scheme 22: Reduction of 60 by *T. delbrueckii* IFO 10921. The maximum yield observed of 64 was 60% in 24 hours.[145]

SDR	58	60	62	(*S*)-54	(*R*)-54
3	0	1	0	0	0
5	0	0	0	0	0
11	0	38	32	0	0
17	0	0	0	0	29
31	0	13	0	13	37
37	0	5	2	0	0

Table 10: Conversions (%) of NADPH with SDRs and WMK derivatives (200 µL): substrate (5 mM), enzyme (10% v/v, 0.1-0.2 mg/mL), NAD(P)H (1 mM), Pi (100 mM, pH 7.2), DMSO (10% v/v). Reaction performed in triplicate (standard deviation <10%) and shaken for 100 min, at 25 °C; quantification using the spectrophotometric assay at 340 nm. The larger the change (showing activity) over the control, the darker green the box. Substrates depicted in Scheme 20.

SDR-3 and 5 did not show activity towards most of the tested HPK derivative compounds, except SDR-5 showed a small activity towards **(*S*)-57**. SDR-11 showed

interesting activities, as the triketone **61** was accepted, as was **63** and **(*S*)-57**. SDR-17 only accepted the triketone **61** and both SDRs-31 and 37 accepted **61** and **(*S*)-57**. **(*R*)-57** was not tested.

SDR	59	61	63	(*S*)-57
3	0	3	0	0
5	0	0	0	14
11	0	49	11	5
17	0	8	0	0
31	0	21	0	23
37	0	48	0	32

Table 11: Conversions (%) of NADPH with SDRs and HPK derivatives (200 µL): substrate (5 mM), enzyme (10% v/v, 0.1-0.2 mg/mL), NAD(P)H (1 mM), Pi (100 mM, pH 7.2), DMSO (10% v/v). Reaction performed in triplicate (standard deviation <10%) and shaken for 100 min, at 25 °C; quantification using the spectrophotometric assay at 340 nm. The larger the change (showing activity) over the control, the darker green the box. Substrates depicted in Scheme 21.

2.6.3 Wieland-Miescher ketone selectivity and pH dependence on conversion

SDRs-17 and SDR-31 displayed activity towards the WMK **54** in the spectrophotometric assay, so this was examined in greater detail. As demonstrated in Table 10, SDR-31 accepted both enantiomers, but preferentially accepted (S)-WMK. Conversely, SDR-17 showed a strong preference towards the (*R*)-WMK.

These reactions were then investigated at pH conditions ranging from 5.5 to 8.0 (Figure 24). SDR-17 maintained its preference for the (*R*)-WMK **(R)-54** at all the tested pH conditions. SDR-31 showed a preference towards the (*S*)-WMK **(*S*)-54** at all the tested pH conditions, but the activity towards (*R*)-WMK decreases more above pH 7.0 than with the (*S*)-WMK. The highest conversion was shown at pH 7.0 for SDR-17 with (*R*)-WMK and SDR-31 with (*S*)- and (*R*)-WMK. SDR-17 demonstrates highest conversion towards the (*S*)-WMK at pH 6.5.

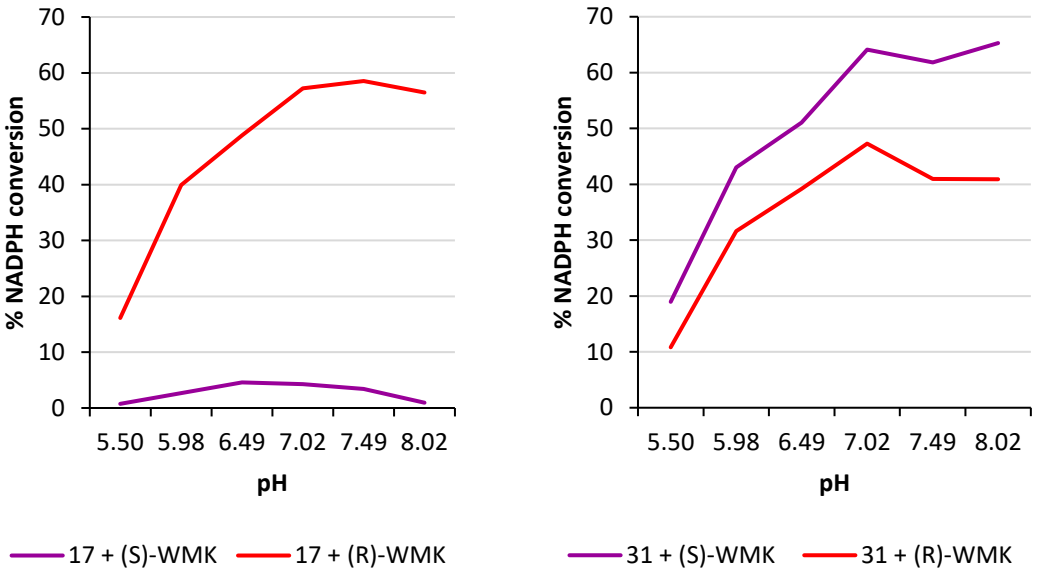
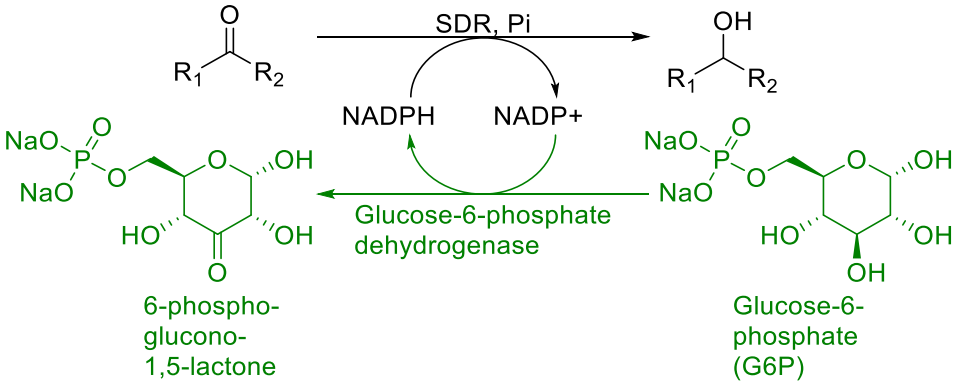


Figure 24: Graphs displaying activity with WMKs (*S*)-54 and (R)-54 with SDR-17 and SDR-31 at a range of pH conditions (200 µL): (±)-Wieland Miescher ketone (5 mM), clarified cell lysate (0.4 mg/mL), NAD(P)H (1 mM), Pi (100 mM), DMSO (10%, v/v). The reactions were shaken for 100 min at 25 °C, performed in duplicate (standard deviation >5%) and quantified by the spectrophotometric assay at 340 nm. Commercial substrates were used.

2.6.4 Co-expression of SDR-17 and SDR-31 with Glucose-6-phosphate dehydrogenase (G6PDH)

Having established interesting activities towards WMK, one potential problem with using this system is the use of NADPH as it is an expensive material. In the above work this was the limiting reagent as the spectrophotometric assay monitored its depletion. As discussed in Section 1.2.8, NADPH can be efficiently recycled using a co-enzyme. Glucose-6-phospate dehydrogenase (G6PDH, EC 1.1.1.49) was selected because it was a suitable method for regeneration of NADPH (Scheme 23). The enzyme has several advantages, including the fact that glucose-6-phosphate (G6P) is stable in solution under the reaction conditions, is unlikely to be reduced by the reductase, is active towards to $NADP^+$ and the reduction of $NADP^+$ is also thermodynamically favoured.[146]



Scheme 23: Reaction scheme showing use of G6PDH as a co-factor recycling system in the biocatalytic reduction of a ketone using an SDR.

Commercial G6PDH may be a costly addition to the reaction, so it was proposed that co-expressing G6PDH and the SDR could provide an efficient means of delivering the recycling system. The SDRs have been employed in these assays as an unpurified clarified cell lysate, so G6PDH would also be in the reaction mixture if co-expressed.

G6PDH from *Saccharomyces cerevisiae* SF838 (Uniprot P11412) on a pACYduet-1 vector backbone‡ was co-transformed into the BL21 expression strain of *E. coli* along with the plasmid containing either SDR-17 or 31. This was then expressed, and the use of 5- and 16-hour induction periods were investigated. More expression was observed after a 16-hour induction period (Figure 25). Expression the SDRs was good and although expression of G6PDH was low, it was produced in sufficient quantities to recycle NADPH (Section 2.6.5).

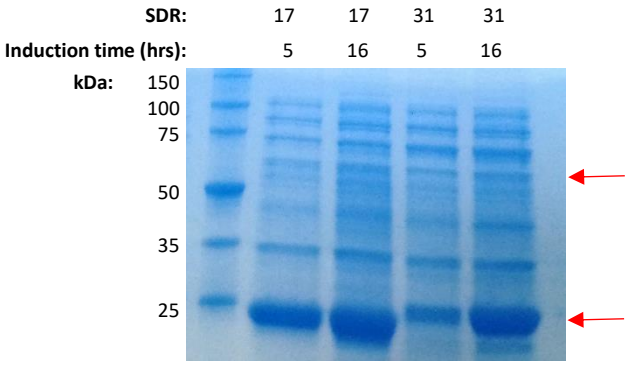


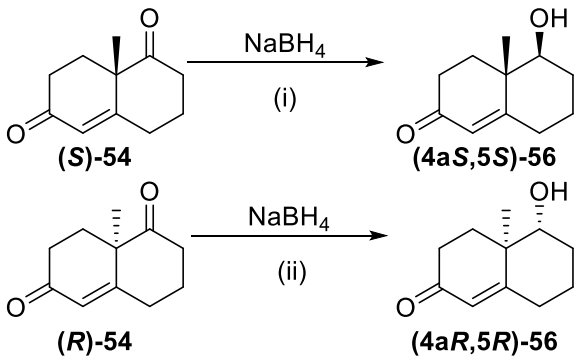
Figure 25: InstaBlu stained SDS-PAGE gel of clarified cell lysates after two induction periods (5 and 16 hours) containing SDR-17 (28.5 kDa) and SDR-31 (27.4 kDa) co-expressed with G6PDH (57.5 kDa) highlighted by red arrows.

‡ G6PDH recombinant vector sourced from Nadine Ladkau.

2.6.5 Application of the co-factor recycling system

The co-factor recycling system was then used for 500 µL scale reactions, initially to establish that the system worked, but additionally to examine more closely at the products of the reaction. The reactions were performed in duplicate with G6P and $NADP^+$ with the SDR alone or SDR and G6PDH (Scheme 23). The reactions were also left for longer than the previous microscale reactions in order to achieve higher conversions. One disadvantage of using the spectrophotometer to monitor NADPH depletion is the potential for false-positive results due to the decomposition of NADPH.[146] This is due to oxidation over time during storage reducing the absorbance at 340 nm, and also due to decomposition in solution during the reaction.[146] The reaction was monitored using HPLC, as this was unaffected by NADPH instability. The substrate concentration was also doubled compared to the spectrophotometric assay to 10 mM and the enzymes were tolerant to DMSO co-solvent (10% v/v), which was employed in all the assays to aid substrate solubilisation.

To monitor the reaction, commercial and synthesised compounds were used as standards. Commercial (*R*)-WMK **(*R*)-54** and (*S*)-WMK **(*S*)-54** were used (Figure 26). (*R*)-WMK and (*S*)-WMK were reduced using $NaBH_4$, which preferentially makes **(4a*R*,5*R*)-56** and **(4a*S*,5*S*)-56** respectively due to steric hindrance caused by the neighbouring methyl group and causing the substrate to react at the least hindered side of the compound (Scheme 24 and Figure 27). The compounds **(4a*S*,5*R*)-56** and **(4a*R*,5*S*)-56** were less easy to make chemically. **(4a*S*,5*R*)-56** was not synthesized and was generated in a later preparative scale reaction (Section 2.6.7). These compounds were then ran on the HPLC using a selection of columns and gradients until satisfactory peak separation was found. The mobile phase was isocratic at 4% isopropanol/hexane, and the flow rate was lowered to 0.5 mL/min to ensure the pressure in the system did not rise too much. This made broad but distinguishable peaks of all the available compounds.



Scheme 24: Reduction of (R)-WMK (*R*)-54 and (S)-WMK (*S*)-54 to make standards for HPLC: (i) NaBH4, EtOH, 1 h, 0 °C, 63%, 96% d.e; (ii) $NaBH_4$, EtOH, AcOH, 1 h, 0 °C, 96%, 98% d.e..

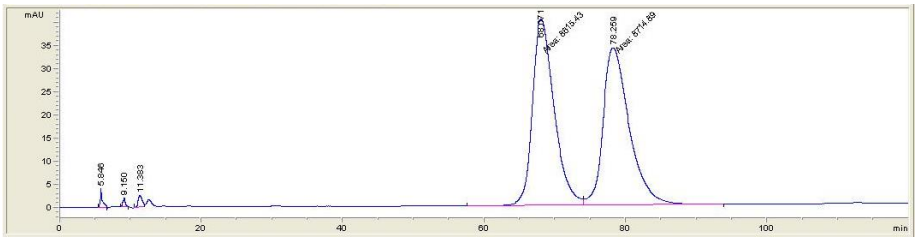


Figure 26: HPLC trace of commercial WMK (1 mM): (S)-WMK (*S*)-54 Rt = 64 min, (R)-WMK (*R*)-54 Rt = 73 min. HPLC conditions were 4% ^{i}PrOH:hexane, 120 min, 0.5 mL/min at 230 nm, on a Chiracel OJ column.

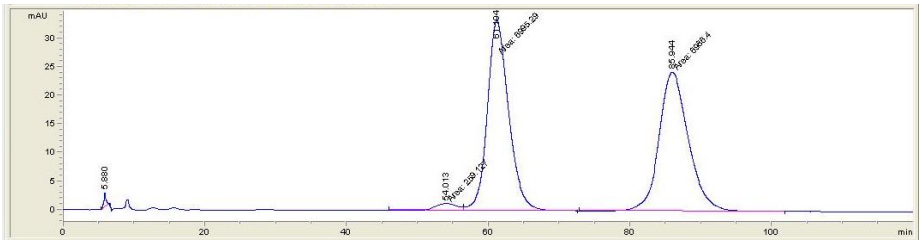


Figure 27: HPLC trace of the products of the reduction using $NaBH_4$ of commercial WMK: (4a*R*,5*S*)-56 Rt = 54 min, (4a*R*,5*R*)-56 Rt = 61 min, (4a*S*,5*S*)-56 Rt = 85 min. HPLC conditions were 4% iPrOH:hexane, 120 min, 0.5 mL/min at 230 nm, on a Chiracel OJ column.

SDR-31 without the recycling system showed a 100% conversion of the ketone starting material (±)-WMK. 100% of the (*S*)-WMK was converted into **(4a*S*,5*S*)-56**, whereas the (*R*)-WMK, there was a small conversion to **(4a*R*,5*R*)-56** with the remainder being **(4a*R*,5*S*)-56** (Figure 28). There is a strong preference to obeying Prelog's rule. With the addition of G6PDH, the final reaction medium consisted of similar ratios of the compounds to without the recycling system present.

The reaction of SDR-17 alone did not proceed to completion: both (*S*)-WMK and (*R*)-WMK remained in the reaction medium after 24 hours (Figure 29). The conversion of (*S*)-WMK was low as indicated from the spectrophotometric assay

(Sections 2.6.2 and 2.6.3). With the introduction of the G6PDH recycling system, SDR-17 converted all of the (±)-WMK in to the (*S*)-alcohols **(4a*S*,5*S*)-56** and **(4a*R*,5*S*)-56**, following Prelog's rule. Both SDRs-17 and SDR-31 showed a preference for the formation of the (*S*)-alcohol.

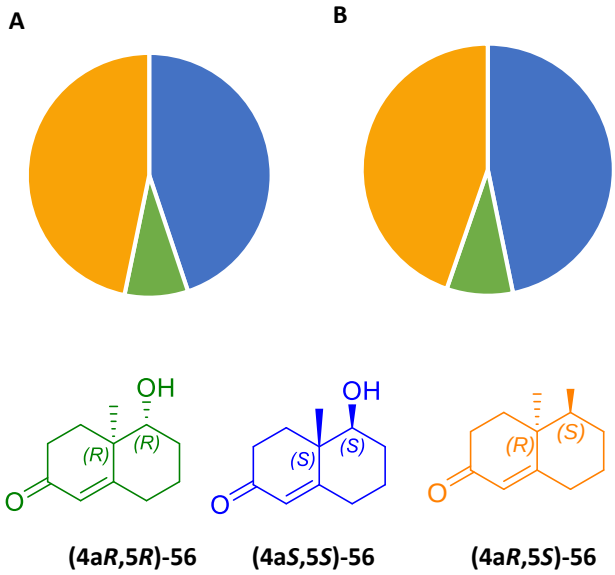


Figure 28: Chart showing (A) SDR-31 (B) SDR-31 co-expressed with G6PDH (500 µL): (±)-WMK (10 mM), clarified cell lysate (0.4 mg/mL), NADP(H) (3 mM), glucose-6-phosphate sodium salt (100 mM), Pi (100 mM, pH 8), DMSO (10%, v/v). Shake 24 h, 37 °C, 250 rpm. Quantification by HPLC at 230 nm.

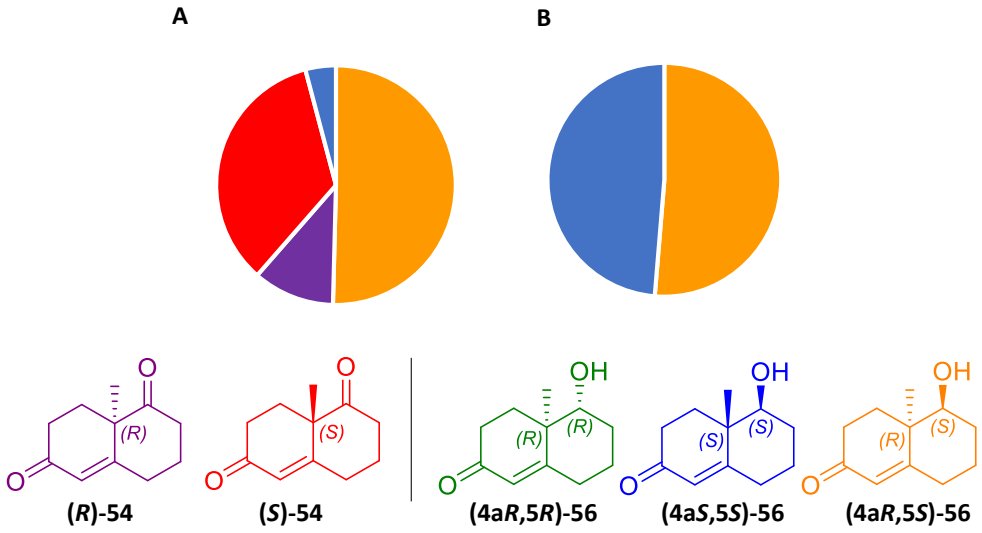


Figure 29: Chart showing (A) SDR-17 (B) SDR-17 co-expressed with G6PDH (500 µL): (±)-WMK (10 mM), clarified cell lysate (0.4 mg/mL), NADP(H) (3 mM), glucose-6-phosphate sodium salt (100 mM), Pi (100 mM, pH 8), DMSO (10%, v/v). Shake 24 h, 37 °C, 250 rpm. Quantification by HPLC at 230 nm.

2.6.6 Product formation over time

SDR-17 co-expressed with G6PDH appeared to make exclusively the (*S*)-alcohols **(4a*R*,5*S*)-56** and **(4a*S*,5*S*)-56** in a stereoselective fashion. The reaction was followed over time to establish the enzyme's preference for enantiomers of the WMK. SDR-17 showed a preference for the (*R*)-WMK and reduced it, but after 24 hours, both enantiomers were reduced to the (*S*)-alcohol.

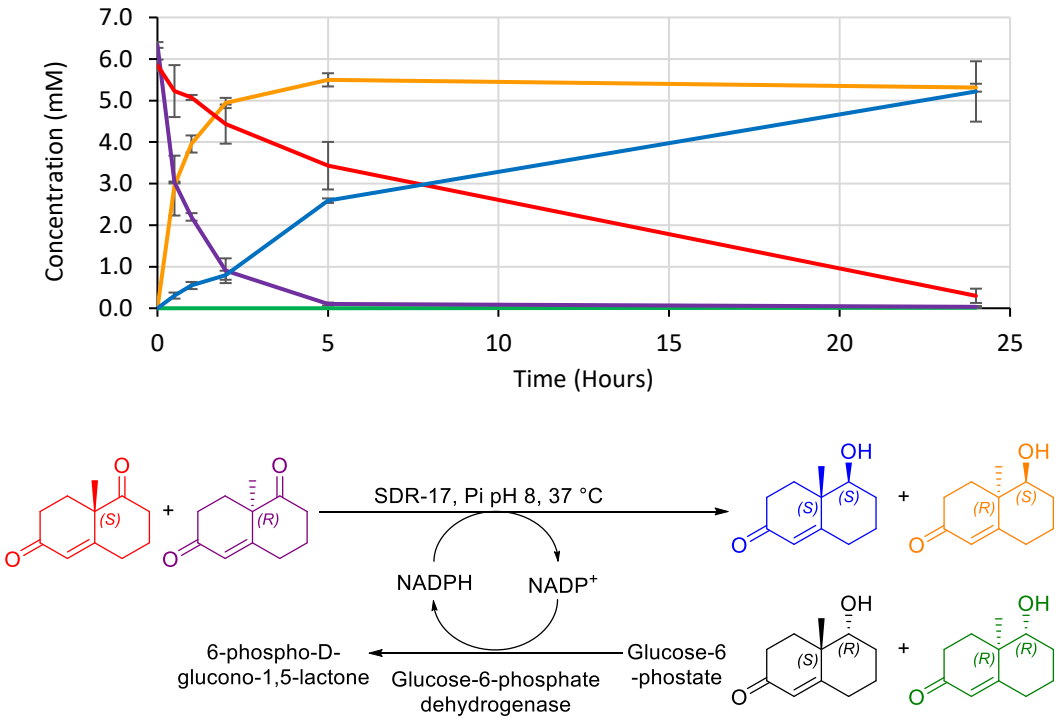
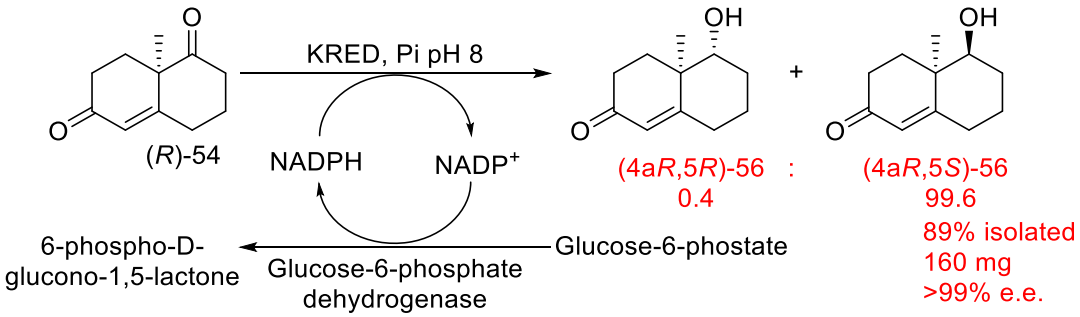


Figure 30: Graph following the reaction of SDR-17 with the WMK (500 µL): (±)-Wieland Miescher ketone (12 mM), clarified cell lysate (0.4 mg/mL), NADP⁺ (3 mM), glucose-6-phosphate sodium salt (100 mM), Pi (100 mM, pH 8), DMSO (10%, v/v). The reactions were performed in duplicate and shaken for 24 h at 37 °C and 500 rpm. Quantification by HPLC at 230 nm.

2.6.7 Scale up to preparative scale of SDR-17 and (*R*)-WMK

SDR-17 (informally referred to as Newgase) was shown to reduce the WMK in a stereoselective manner. Compound **(4a*R*,5*S*)-56** is challenging to make synthetically: no chemical reduction has been reported in the literature. Although enzymatic reactions have used bakers' yeast or mono-oxygenases have been prepared, **(4a*R*,5*S*)-56** had not been generated from a reduction reaction in an entiopure fashion (Section 2.6.1).[137,140,147–149] As such, a preparative scale reaction was conducted at a 50 mL scale to produce **(4a*R*,5*S*)-56** using (*R*)-WMK (Scheme 25).

A higher concentration of (*R*)-WMK was employed (20 mM) and this was tolerated by SDR-17. The product was extracted without the requirement for flash column chromatography and both the HPLC traces and the spectroscopic NMR analysis demonstrated high purity. The reaction time was also only 24 hours, whereas reactions times of up to 68 hours had been reported using yeast.[140]



Scheme 25: Preparative scale reaction of SDR-17 and (R)-WMK (50 mL): *(R)*-WMK (20 mM), clarified cell lysate (0.4 mg/mL), NADP$^+$ (3 mM), glucose-6-phosphate sodium salt (100 mM), Pi (100 mM, pH 8), DMSO (10%, v/v). Shake 24 h, 37 C, 200 rpm. Quantification by HPLC at 230 nm.

An HPLC set up for chiral conditions using a reverse-phase column was used to determine which enantiomer was generated and the yield (Figure 31), as discussed in Section 2.6.5. The peak of the starting material (*R*)-WMK had a retention time (Rt) of 71.6 minutes (Figure 31A). The product **(4a*R*,5*R*)-56** was made chemically using $NaBH_4$ which preferentially generates the alcohol in the (*R*)-configurtation.[150] This is due to steric hindrance caused by the methyl group, causing the substrate to react at the least hindered side and make this product (Figure 31B). The retention time of the product of the reaction was at 56 minutes, suggesting that the (*S*)-alcohol **(4a*R*,5*S*)-56** was generated. This route is more stereo selective, higher yielding and faster than the alternative methods discussed in Section 2.6.1.

The product **(4a*R*,5*S*)-56** was produced with excellent conversions of 100% (determined using HPLC), 89% (160 mg) isolated yield and the e.e. was greater than 99%. This demonstrates an efficient route to **(4a*R*,5*S*)-56**: the reductase and co-factor recycling enzyme were readily expressed together, and so G6P and NADP$^+$ were required as low cost and available reagents.

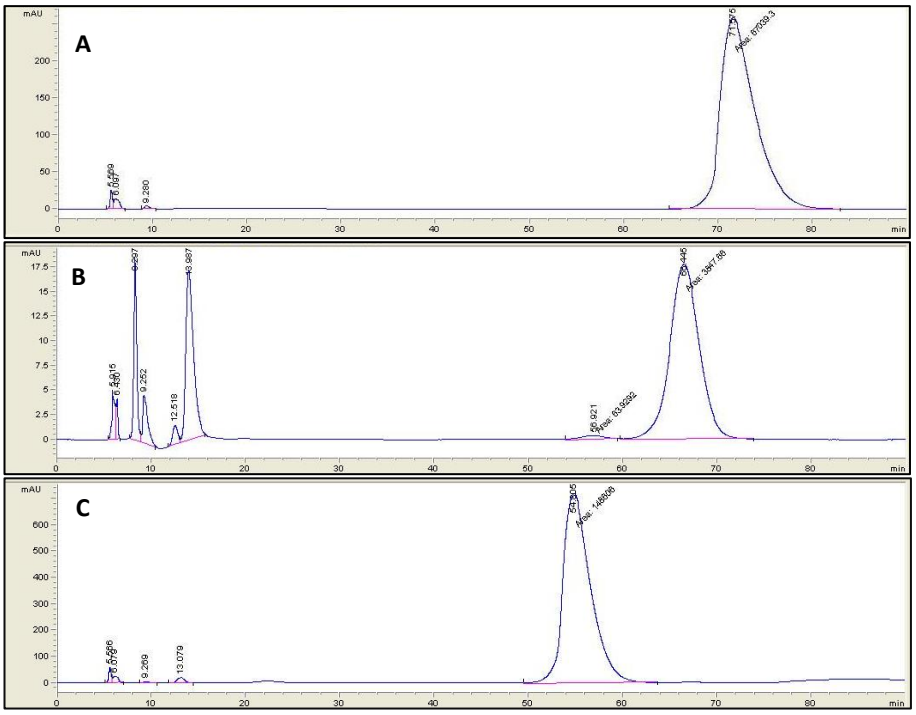


Figure 31: HPCL traces of: (A) (*R*)-WMK (10 mM) Rt = 71.6 min; (B) (4a*R*,5*R*)-56 (1 mM) Rt = 66.4 min; (C) Product of the scale up starting from (*R*)-WMK Rt = 54 min. HPLC used Chiralcel OJ column with 4% isopropanol/hexane mobile phase at f 0.5 mL/min flow rate and detection at 230 nm.

2.7 Conclusions and future work

In summary, this chapter describes the retrieval of 37 SDRs from the tongue metagenome. The activity of the SDRs were established towards pyruvic acid, cyclohexanone and tropinone. Eight active enzymes were then assayed against a broad range of substrates to identify their substrate tolerances. WMKs, HPKs and compounds that are intermediates in their syntheses were tested with six of the SDRs. The optimum pH conditions for SDR-17 and SDR-31 with (*R*)- and (*S*)-WMK were determined to be approximately pH 7. Additionally, SDR-17 was co-expressed with the co-factor recycling system G6PDH and this was used to create an efficient system to reduce (*R*)-WMK at a preparative scale with almost 90% isolated yield. This has therefore expanded the number of useful SDRs, and demonstrated and industrially relevant reaction.

SDR-17 has shown to be a versatile enzyme in terms of its substrate selectivities, its robustness to substrate concentrations (5-20 mM) and tolerance to DMSO as a co-solvent (10% v/v). SDR-17 could be industrially relevant, and further testing would highlight this in future work. For example, reducing the reaction temperature, co-

solvent concentration and reaction time could make this reaction more industrially applicable. An X-Ray crystal structure could be elucidated in the future to more fully understand the active site of the enzyme. In addition, in small scale reactions, SDR-17 did not display activity towards compounds **58**, **60** or **62** (Section 2.6.2). This makes SDR-17 a good candidate for a one-pot reaction in which WMK is both synthesised and then reduced.

Two of the retrieved SDRs were investigated in detail with the WMK, but the other SDRs could also be investigated to fully understand their substrate selectivities. SDR-17, SDR-31 and SDR-37 readily accepted a wide variety of substrates including fluorinated aromatic compounds, and this could be further investigated. Additionally, the tongue metagenome could be searched further for other enzymes.

3 Tropinone Reductases and MecgoR

3.1 Introduction

Tropinone reductases (TRs) have been reported in the literature from a number of plants and have been characterised in terms of their role in plant metabolism, for which their activity towards a number of small molecules has been assessed. Tropinone, the natural substrate for these enzymes, is a structure found in many useful compounds (Section 1.3.1). These enzymes therefore can be used for biocatalytic applications, which will be explored in this chapter.

3.2 Tropinone reductase I and II with/without 6-Histidine tag

Plasmids containing tropinone reductase 1 (TRI) and 2 (TRII) genes from *Datura stramonium* containing plasmids were transformed into BL21 *E. coli* expression cells[§] and successfully expressed (Figure 32). As previous studies have shown that the addition of a hexahistadine (His) tag, as well as its location on the protein, can alter activity, both TRI and TRII enzymes with and without a His-tag were initially assessed.[83] Both in previous studies on these enzymes by Portsteffen et al. and Nakajima et al., the proteins were purified without the use of a His-tag, which were purified by several steps of chromatography.[68,78]

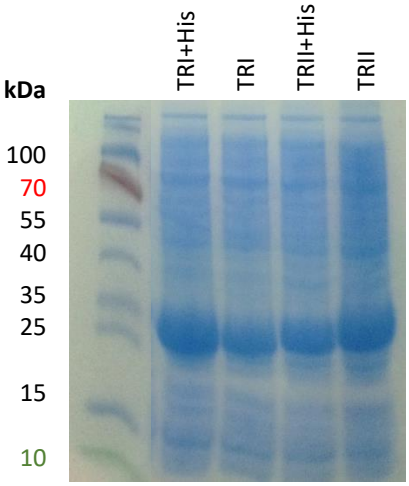
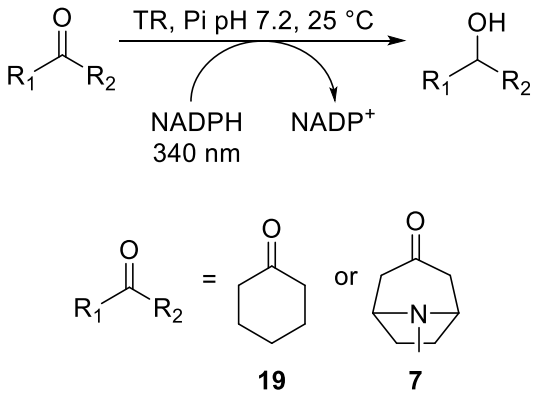


Figure 32: SDS-PAGE gels of TRI+His (30.5 kDa), TRI (29.6 kDa), TRII+His (29.2 kDa), TRII (28.3 kDa) clarified cell lysate.

[§] Procedure carried out by Prof. John Ward at UCL.

A spectrophotometric assay was used in order to assess for activity, which measures the change in absorbance at 340 nm after 100 minutes after the addition of substrate to clarified cell lysate (Scheme 1). The λmax of NADPH is 340 (ξ_{NADPH} 622 $M^{-1}mm^{-1}$), whereas the oxidised form absorbs less at this wavelength and therefore the drop in absorbance indicates enzymatic activity.[127,151] For initial screens, the absorbance at 340 nm was plotted against time and this was used to measure relative initial reaction rates.



Scheme 26: Tropinone reductase assay. NADPH is detected at 340 nm, and its depletion indicates reduction of the ketone.

A reaction time of 100 minutes was employed, as this was adequate time to identify activity, and the reactions completed within this time (Figure 33). The pH of the reaction was 7.2, in order to make these reactions comparable to the SDR reactions (Section 2). In the literature, a variety of pH conditions have been employed. Nakajima *et al.* used pH 7, commenting that this was physiological pH, which is reported as pH 7.6.[68,82] Portsteffen *et al.* used pH 6.4, as this was the reported optimum for TRI, and as such, the activities relative to tropinone were expected to be different.[76]

The substrates tropinone **7** and cyclohexanone **19** were used for initial assays. NADPH depletion in the spectrophotometric assay (Figure 33) indicated that tropinone was reduced by both TRI and TRII, but the reaction with TRII was slower than for TRI. Portsteffen *et al.* have reported that TRI and TRII from *D. stramonium* were both active with tropinone, although only activities relative to tropinone were reported.[68] TRI and TRIII were reported to generate α-alcohol **13** and β-alcohol **14** respectively (Scheme 27).[68]

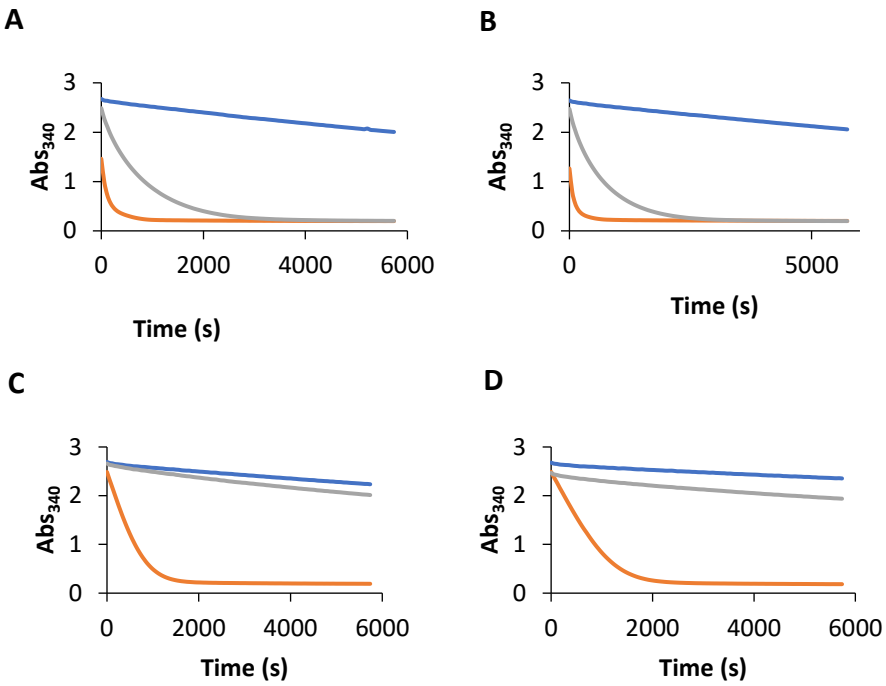
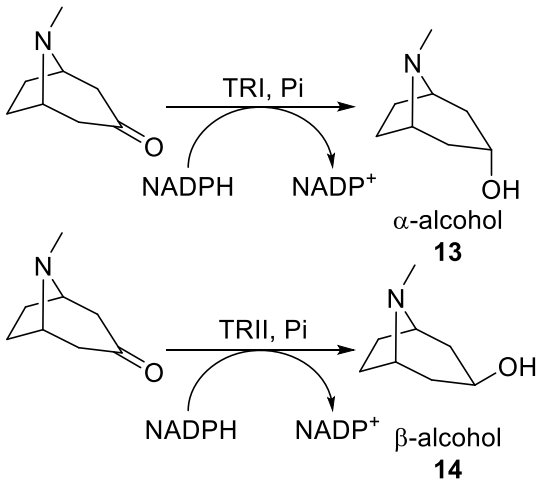


Figure 33: Graphs showing the change in absorbance at 340 nm over time with the following enzymes (200 µL): substrate (5 mM), enzyme (10% v/v, 0.1-0.2 mg/mL), NAD(P)H (1 mM), Pi (100 mM, pH 7.2), DMSO (10% v/v). The reactions were shaken for 100 min, at 25 °C and quantified using the spectrophotometer at 340 nm: (A) TRI with His-tag; (B) TRI without His-tag; (C) TRII with His-tag (D) TRII without His-tag. The substrates were as follows: blue = water control; grey = cyclohexanone; orange = tropinone. Reaction was repeated three times and standard deviation <10%.



Scheme 27: TRI and TRII were reported to reduced tropinone to the α-alcohol and β-alcohol respectively.[68]

Cyclohexanone was also reduced, albeit at a slower rate than tropinone (Figure 33). In the case of TRII, cyclohexanone reduction was only marginally above the negative water control. The background reactions may be caused by endogenous reductases and metabolites, and NADPH decomposition.[128] The background reaction was monitored by using water alone as the substrate (blue line, Figure 33). In

subsequent sections of this PhD, this negative water control was deducted from the percent NADPH conversion. This observation is supported by the literature, in which TRI and TRII from *D. stramonium* also reacted more slowly with 6-membered ring ketones. For example, Porsteffen *et al.* reported that 4-methylcyclohexanone displayed relative activities of 39% and 22% to tropinone with *D. stramonium* TRI and TRII respectively.[68] In comparison, five of the eight selected SDRs showed activity towards cyclohexanone.

The presence of the His-tag on TRI appeared to result in a faster reaction rate with tropinone and reduced rate with cyclohexanone (Figure 33). TRII reacted with tropinone at a similar rate both with and without the His-tag present. Freydank *et al.* previously highlighted that addition of the His-tag to the C terminus of TRI and TRII from *Solanum dulcamara* functionally impaired the enzyme performance and modelling experiments rationalised it could interfere with the active site.[83] To further investigate this, more detailed kinetic work could be carried out in the future.

3.2.1 Tropinone reductases I and II substrate specificities.

A variety of substrates were screened to assess more fully the substrate specificity of the two enzymes, with and without His-tags. Table 12 displays the change in absorbance at 340 nm after 100 minutes with the negative control (lacking the substrate) deducted and as a percentage of the activity towards tropinone.[69,72,74] The endpoint absorbance was used as opposed to the early converions noted in the previous section. This is due to rapid initial reaction rates with well tolerated substrates before UV-visible spectra measurements were started, and so this initial activity was missed. To reduce any errors resulting from the assay and problems with the degradation of NADPH (discussed in Section 2.6.4), tropinone was used as a calibration standard on each 96-well plate and this was used to calculate relative percentage NADPH conversions. Comparing enzyme activity to that towards tropinone has been reported in the literature in several TR studies.[68,69,73,152,153] Using 10% DMSO in the assay did not affect activity markedly, and so was used to aid solubilisation of the substrates.

		Activity relative to tropinone (%)			
	Substrate	**TR1+His**	**TR1**	**TR2+His**	**TR2**
7	Tropinone*	100	100	100	100
7	Tropinone	99	103	96	100
19	Cyclohexanone	100	100	11	19
65	Cycloheptanone	80	84	0	1
38	2-Methylcyclohexanone*	91	90	36	27
39	3-Methylcyclohexanone	91	105	3	20
(*R*)-39	(R)-3-Methylcyclohexanone*	96	99	12	8
66	4-Methylcyclohexanone*	97	132	16	17
40	2-Cyclohexen-1-one	37	24	0	0
51	3-(2-Oxocyclohexyl)propanal	73	73	44	25
52	1-Indanone	1	0	0	0
53	2-Indanone	62	58	0	0
54	Wieland–Miescher ketone	82	79	13	4
21	*N*-methyl-4-piperidone	83	120	74	90
37	Ethyl-4-oxo-1-piperidine carboxylate	93	125	78	91
36	1-(2-Methyl propyl)-4-piperidone	115	147	82	89
43	2,2,2,4′-Tetrafluoroacetophenone	98	104	46	42
42	2,2,2,3′-Tetrafluoroacetophenone	103	105	43	45
41	2,2,2-Trifluoroacetophenone	103	106	87	76
44	4'-N-propyl-2,2,2-Trifluoroacetophenone	15	20	0	0
19	Acetophenone	65	70	0	0
45	Benzaldehyde	91	105	61	38
50	1-Phenyl propan-2-one	90	99	13	9
46	Phenylacetaldehyde	0	0	0	0
47	2-Heptanone	80	84	6	2
48	Methylacetoacetate	90	93	15	2
35	Pyruvic Acid	nd	75	17	10
71	(±)-Camphor	0	0	0	0
67	Testosterone	0	0	0	0
68	Progesterone	0	0	0	0
69	Trans andersterone	0	0	0	0
70	11-Keto progesterone	0	0	0	0
(1*R*)-71	(1R)-(+)-camphor	0	0	0	0
72	*Cis*-jasmone	0	0	0	0
73	(+)-Pulegone	0	0	0	0

Table 12: Table to show the activity of TRI and TRII (+/- His-tag) towards various substrates relative to that towards tropinone (%) (200 µL): Substrate (5 mM), clarified cell lysate (0.4 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were shaken for 100 minutes, 25 °C and performed in triplicate (standard deviation <10%). NADPH depletion was quantified by spectrophotometer at 340 nm. * Denotes that 10% DMSO was not used. The structures of the compounds are shown in Figure 34.

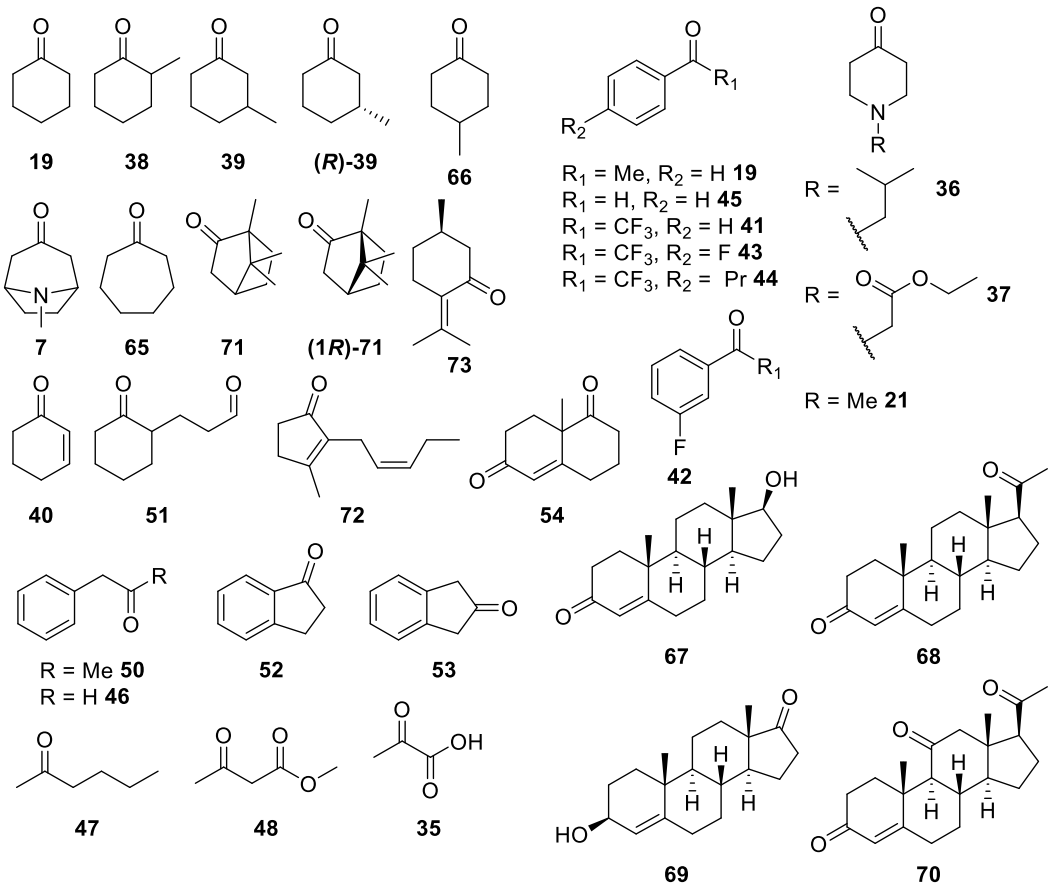


Figure 34: Structures of compounds tested towards TRs.

Cyclohexanone **19** was readily accepted by TRI, though poorly accepted by TRII, regardless of the presence of a His-tag, as discussed above. Similarly, cycloheptanone **65** and the methylcyclohexanones **38**, **39**, **(*R*)-39**, **66** also displayed good activities of over 90% relative to that towards tropinone. 4-Methylcyclohexanone **66** had been tested in the literature and the relative activity demonstrated was much higher than that reported in the literature: Relative Vmax of 39% and 22% to TRI and TRII were reported respectively, whereas in this study, relative percent NADPH conversion was found to be comparable to tropinone in this study.[68] Portsteffen *et al.* however used purified enzyme and different reaction conditions including reaction temperature of 30 °C, NADPH concentrations of 200 µM and the substrate concentrations were varied. Notably a pH of 6.4 as opposed to 7.2 used here. The pH 6.4 is the optimum for TRs with tropinone, and so this biases the results which are measured relative to that of tropinone. SDR-4, SDR-17, SDR-31 an SDR-37 showed high activities towards **19, 38** and **39.**

Interestingly, with introduction of the enone functionality, **40** was still accepted by TRI (with and without the His-tag). Additionally, 3-(2-oxocyclohexyl)acetaldehyde **51** was also accepted by both enzymes, although it is unclear if one or both the carbonyl groups were reduced.

When a nitrogen was introduced into the ring in the 4-piperidone substrates (**21, 37**, and **36**), the activity increased notably for both enzymes and TRII was almost as active as TRI. This may be due to the structure similarity to tropinone. The pKa of **21** is 7.9, so the substrate would be substantially charged.[154] Compounds **37** and **36** were readily accepted by SDR-3, SDR-11, SDR-17, SDR-31, SDR- 37, and SDR-4, though N-methyl-4-piperidone accepted with much lower activities.

The tested steroids and terpenes (**67**, **68**, **69**, **70**, **71**, **(1*R*)-71**, **72**, and **73**) did not show activity with the TRs. This may have been because many of them precipitated out of solution despite using 10% DMSO as a co-solvent. Compounds 2-heptanone **47**, methylacetoacetate **48** and pyruvic acid **35** were more readily accepted by TRI than TRII.

The bicyclic ring systems 2-indanone **53** and the WMK **54**, were accepted by TRI giving good conversions. Neither were accepted by TRII and 1-indanone **52** was inactive with both enzymes. Activities towards these bicyclic systems have not been reported in the literature with plant TRs (Section 8.2). However, it is interesting to note that the SDRs from the tongue metagenome in Section 2.4 showed similar activities with WMK: SDR-4, SDR-17, SDR-31 and SDR-37 accepted WMK, however none of the SDRs showed activity towards 1- or 2-indanone **52** or **53**.

The TRs displayed varied activities towards a range of aromatic compounds. For example, benzaldehyde **45** was accepted by both enzymes although phenylacetaldehye **46** was not. Acetophenone **29** and 1-phenyl propan-2-one **50** were also accepted by TRI and not TRII.

Fluorinated compounds (**41**, **42** and **43**) were readily accepted by TRI and to some degree by TRII. However, 4'-*N*-propyl-2,2,2-trifluoroacetophenone **44** was poorly accepted by TRI and not at all by TRII, potentially due to its larger size. A similar

trend was demonstrated by the SDRs, **41**, **42**, and **43** were readily accepted by SDR-4, SDR-17, SDR-31 and SDR-37, but none of the SDRs showed activity towards **44**.

3.2.2 Modelling to explain experimental results

In general, TRI accepted a wider range of substrates than TRII, and TRI appeared to be more active towards the same substrate. As a guide to rationalise some of the experimental findings, some initial docking experiments were performed. Using ChemBioDraw 3D the tested substrates were modelled, energy minimised, docked into the reported *D. stamonium* model using Autodock Vina. Pymol was used to measure the distance between the carbonyl group of the substrate and the cofactor hydride transferred. The X-ray crystal structures used were downloaded from the PDB with accession codes 1AE1 and 2AE1 for TRI and TRII respectively, as reported by Nakajima *et al.*.[155]

Nakajima *et al.* highlighted that the residues that come in to contact with tropinone are: V110, S158, I159, A160, L165 V168, Y171, V203, L208, V209, I223, F226 for TRI and V98, Y100, S146, V147, S148, V153, E156, Y159, V191, L210, L213 for TRII.[78] These have been highlighted in blue in Figure 35 (the crystal structure of TRII is missing residues that correspond to L208 and V209 on TRI). X-ray crystal structures of the two enzymes showed that TRI has a larger and more open active site and may allow access for larger molecules to the active site (Figure 35). It should be noted that in TR-II, Leu196 and Val197 missing from B because of their disorder in the crystal. The residues would occupy the positions that correspond to Leu208 and Val209 in TRI.[78]

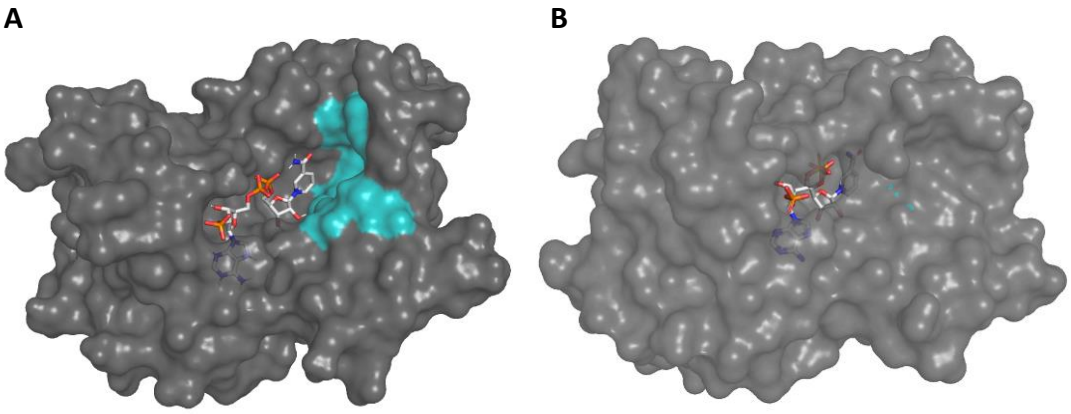


Figure 35: Crystal structures of enzymes (A) TRI and (B) TRII (PDB accession codes 1AE1 and 2AE1 respectively) visualised using Pymol. NADPH has been co-crystallised in both cases with these enzymes. The substrate binding region is highlighted in teal (mostly hidden in TRII).[78] TRI has a more open and large active site.

When 3-(2-oxocyclohexyl)acetaldehyde **51** was docked into the crystal structure of TRI and TRII, both ketone groups appeared to be the same distance (4.9 Å) to the hydride transferred from the cofactor (Section 3.2.2, Figure 36). Therefore was not possible to propose which (or both) of the ketones was reduced, and further product analysis would be necessary to discern this.

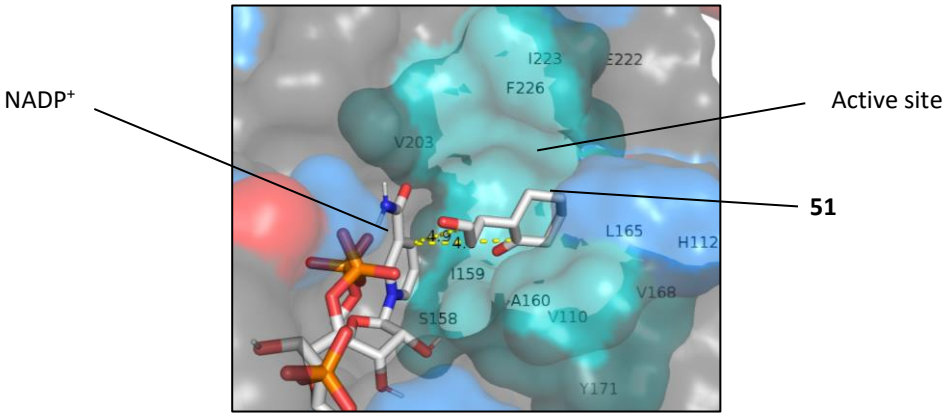


Figure 36: Substrate 3-(2-oxocyclohexyl)acetaldehyde 51 docked into TRI.

As noted previously, steroids and terpenes were poorly tolerated which may have been due to the poor water solubility. Docking of these substrates was attempted, for example, 11-ketoprogesterone **70** in Figure 37. The TRI and TRII appeared unlikely to accommodate these substrates. Even TRI, which had a larger active site, is unlikely to accept steroids: the carbonyl group is over 5 Å from the NADPH hydride transferred, making reduction unlikely.

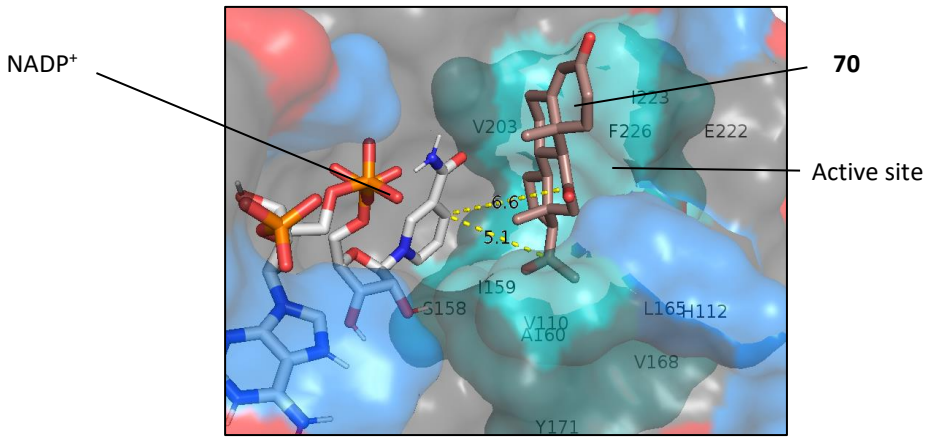


Figure 37: 11-Ketoprogesterone 70 docked into TRI.

Modelling 2-indanone **53** into the active site showed that more "linear" substrates fitted more readily in the active site (Figure 38). TRI has a more open active site than TRII, and this may explain why TRI is more receptive to more variation in substrate molecular architecture.

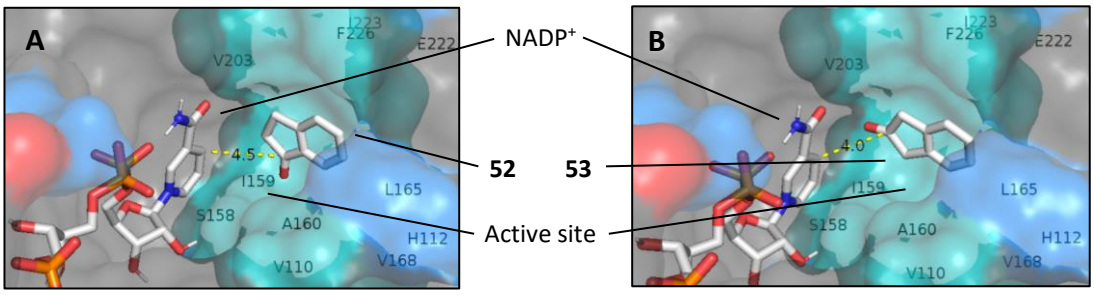


Figure 38: TRI crystal structure with NADPH and (A) 1-indanone 52 and (B) 2-indanone 53 docked. 2-indanone is more linear and fits in to the active site better.

Compounds containing a piperidone motif were also readily accepted by the TRs. The substrate 1-(2-methyl propyl)-4-piperidone **36** was very well accepted by TRI and TRII with relative activities of over 110% and 80% respectively. This compound was docked with TRI, displayed in Figure 39. This may be due the hydrophobic residues (V110, I159, A160, L165 V168, Y171, V203, L208, V209, I223, F226, L208 and V209) and resulting interactions in the central areas of the pocket closest to the cyclohexanone ring and the aliphatic tail on the substrate. In TRII, His112 is replaced by Tyr100, which is a polar, non-basic residue and this may also contribute to differing substrate selectivities.[78]

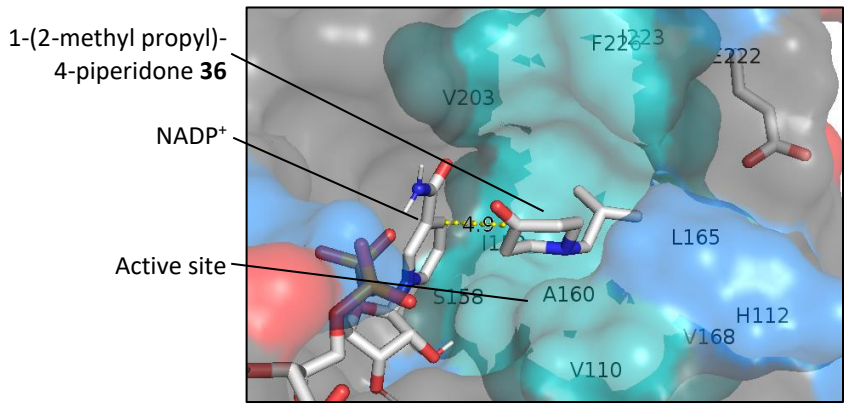
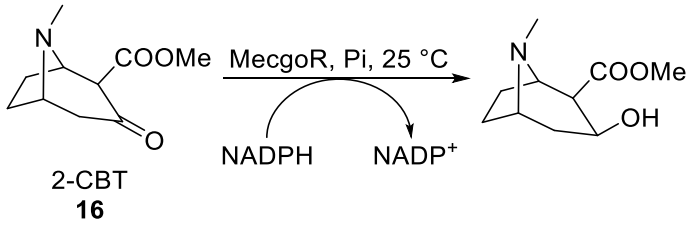


Figure 39: Close-up image of TRI substrate binding pocket, with the NADP cofactor, the substrate 1-(2-methyl propyl)-4-piperidone 36 labelled. Electrostatic charge distribution has been highlighted (red for negative and blue for positive) and the remainder of the active site highlighted in teal.[78]

3.3 MecgoR Substrate specificities

Having investigated preliminary substrate tolerances of TRI and TRII from *D. stramonium*, it was found that bulkier substrates were not readily accepted. MecgoR from *Erythroxylum coca* accepts 2-CBT **16** as its natural substrate, so may be active with larger "non-linear" substrates (Scheme 28), and be used as a biocatalyst.



Scheme 28: Jirschitzka *et al.* reported that MecgoR reduced 2-CBT 16 to the β-alcohol.

To investigate this idea, the MecgoR gene from *E. coca* containing plasmids were transformed into BL21(DE3 Star)pLysS *E. coli* expression cells[**] and expressed (Figure 40) and tested with a selection of substrates (Table 13). A His-tag was appended to the N-terminal. The spectrophotometric assay was used with the same method as discussed earlier using 96-well plates (Scheme 29 and Section 3.1). An NADPH concentration curve was used to infer activity from the change in absorbance at 340 nm after 100 minutes with the negative control (lacking the substrate) deducted. The percent NADPH conversion has been shown as a percentage of that with tropinone, as this common in the literature for TRs.[68,69,153,155,156]

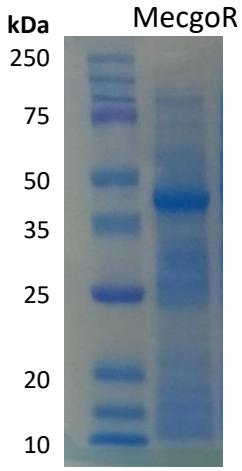
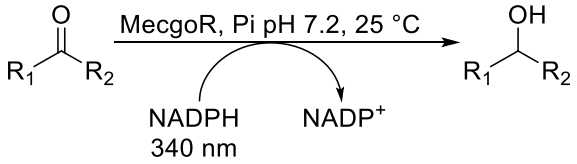


Figure 40: SDS-PAGE gel of MecgoR clarified cell lysate (39.6 kDa).

[**] Procedure carried out by Prof. John Ward at UCL.



Scheme 29: MecgoR spectrophotometric assay

	Substrate	Relative activity (%)		Substrate	Relative activity (%)
7	Tropinone	100	**52**	1-Indanone	0
7	Tropinone*	43	**53**	2-Indanone	0
16	2-CBT	93	**41**	Trifluoroacetopenone	0
19	Cyclohexanone*	46	**42**	2,2,3-Trifluoroacetophenone	86
65	Cycloheptatnone*	52	**43**	2,2,4-Trifuloroacetophenone	62
39	2-Methylcyclohexanone*	64	**44**	4'-*N*-Propyl-2,2,2-trifluroacetophenone	28
39	3-Methylcyclohexanone	44	**29**	Acetophenone	0
66	4-Methylcyclohaxanone*	82	**45**	Benzaldehyde	83
40	2-Cyclohexene-1-one	34	**50**	Phenyl-propan-2-one	13
21	1-Methyl-4-piperidone*	52	**46**	Phenylacetaldehyde	0
37	Ethyl-4-oxo-1-piperidone*	78	**71**	(±)-Camphor*	15
36	1-(2-Methyl)-4-piperidone*	66	**68**	Progesterone	0
47	2-Hepatanone	9	**74**	Andersterone	0
48	Methylacetoacetate	19	**70**	11-Keto-progesterone	0
35	Pyruvic acid*	0	**72**	*Cis*-jasmone*	0

Table 13: Table to show the positive activity of MecgoR towards various substrates relative to that towards tropinone (%) (200 µL): Substrate (5 mM), MecgoR clarified cell lysate (0.4 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were performed in triplicate (standard deviation <10%) and shaken for 100 min, at 25 °C. NADPH depletion was quantified using the spectrophotometer at 340 nm. Blank spaces denote no activity. * Denotes that 10% DMSO co-solvent was not used. The structures of the compounds are shown in Figure 41.

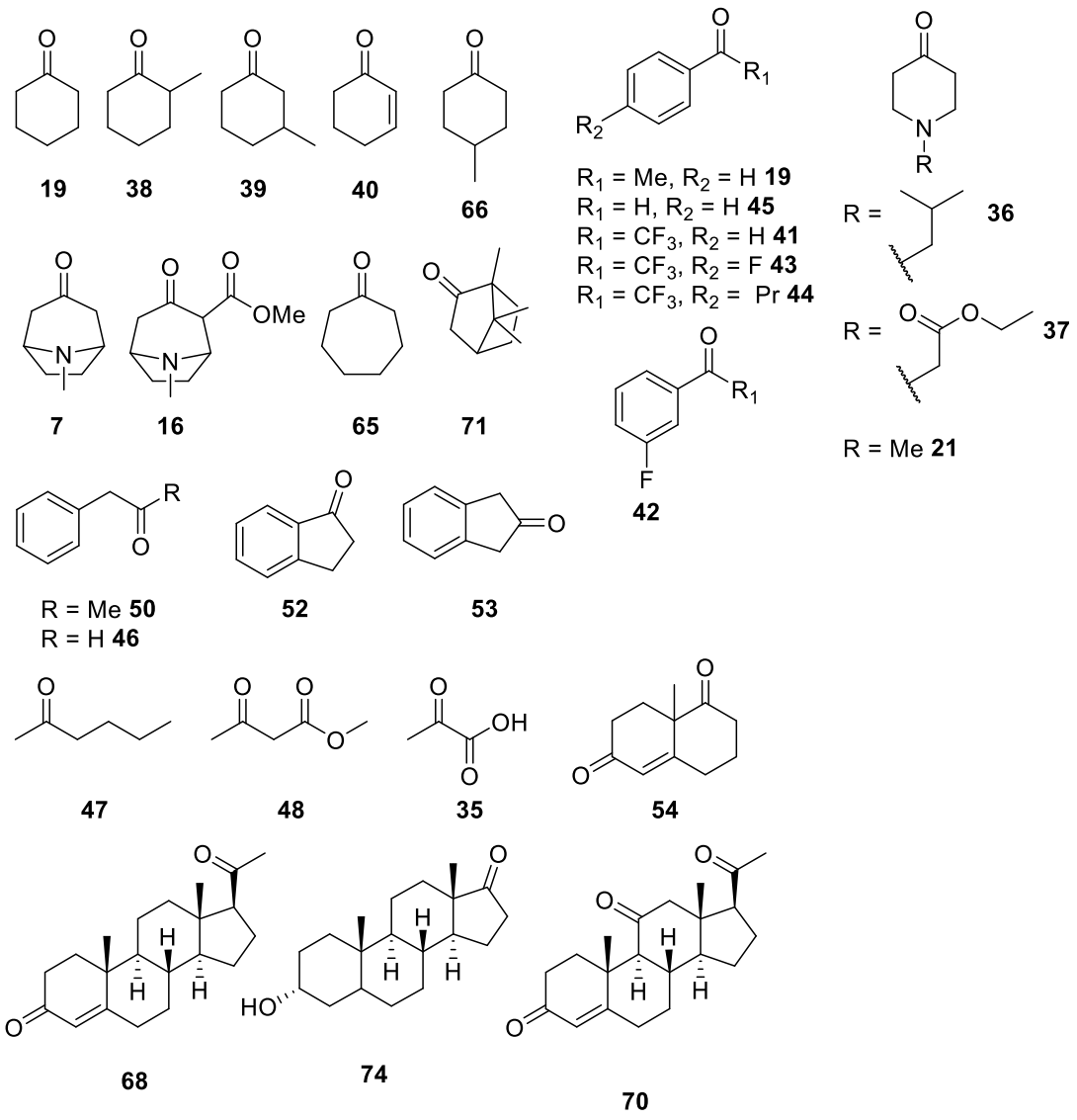


Figure 41: Substrates used in Table 13.

In general, MecgoR was more active towards larger substrates than TRI or TRII was. For example, MecgoR displayed 15% relative conversion towards (±)-camphor **72**, compared to 0% with TRI and TRII. Compound **44** was reduced with 28% relative conversion by MecgoR, but 15% and 0% by TRI and TRII respectively. Activity towards tropinone was affected by the use of DMSO. Addition of 10% DMSO to the reaction mixture more than doubled the activity towards tropinone, and this may well be due to solubility issues. DMSO was initially added to aid solubilisation of the substrates. After MecgoR was demonstrated to be tolerant to 10% DMSO, this was used for all the later reactions as standard to circumvent potential solubility problems.

6-Membered non-aromatic compounds, and piperidone based compounds (**19**, **65, 38**, **39**, **66**, **21, 37**, **36**) were accepted by the enzyme MecgoR with good conversions of 44 to 78% relative to activity with tropinone. Previous work on the MecgoR by Jirschitzka *et al.* showed that cyclohexanone and *N*-methyl-4-piperidone **21** was not active with the enzyme, however here they were readily accepted. With differing reaction conditions, such as pH 6.8, lack of 10% v/v DMSO and a 30 minute reaction time, the results were not directly comparable. The shorter assay conducted by may not have given time for the reaction to progress further to higher conversions.[61] Additionally, at the reported lower pH, the more of *N*-methyl-4-piperidone **21** may be protonated and so this may affect activity (pK_a = 7.9).[154] These cyclic substrates were accepted more readily by TRI than TRII, and were also accepted by SDR-4, SDR-11, SDR-31 and SDR-37.

As with TRI and TRII, MecgoR was not active with any of the steroids used. Several of these precipitated out when added to the reaction buffer, causing problems with the assay as the plate reader was unable to read opaque samples, and the concentration of the substrate was unknown. Small activities were demonstrated with (±)-camphor **71**. (±)-Camphor is a bicyclic structure related to tropinone and 2-CBT, but is much more hydrophobic in nature.

The fluorinated compounds tested were all accepted by MecgoR, even 4'-*N*-propyl-2,2,2-trifluroacetophenone **41** which was not accepted by TRI and TRII, illustrating the capacity of MecgoR to accept more bulky substrates. Interestingly, like TRII, benzaldehyde **45** was accepted by MecgoR but phenylacetaldehyde **46** was not. However, TRI accepted this compound. Additionally, pyruvic acid **35** was not accepted, although this was active with TRI and TRII. MecgoR was expected to accept pyruvic acid, but other factors may be determining activity, such as complexation of the acid group to histidine or arginine residues in the active site. 2-Heptanone **47** and methylacetoacetate **48** were accepted but had low activity.

3.3.1 Enones

Like TRI, MecgoR accepted cyclohexenone **40** and so it was proposed that MecgoR could reduce conjugated ketones. Several enone-containing compounds

(**75**-**81** and **57**) were used in the spectrophotometric assay on a small scale (Figure 42). In general, tolerance to these substrates was low, with furfuraldehyde **81** showing some activity (approx. 50% conversion). Cyclohex-1-en-2-one **40**, penten-2-one **75**, (S)- and (R)-carvone **(*S*)-80** and **(*R*)-80** showed low activities of around 10% NADPH conversion.

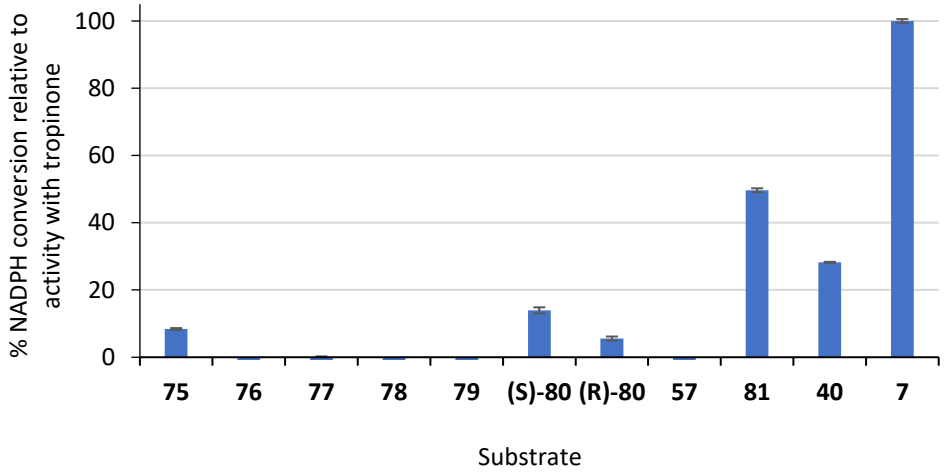


Figure 42: Graph displaying % conversion of NADPH as a percentage of that with tropinone for a selectin of enone-containing substrates with MecgoR (200 µL): substrate (5 mM), clarified cell lysate (0.4 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.6), DMSO (10%, v/v). The reactions were performed in triplicate and shaken for 100 min, at 25 °C and quantified by spectrophotometric assay. Substrate structures are depicted in Figure 43.

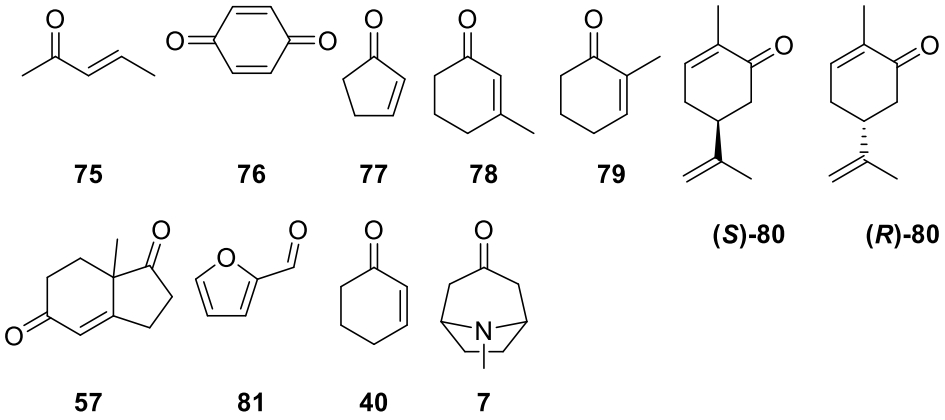


Figure 43: Structures of substrates used in Figure 42.

3.3.2 MecgoR optimum pH conditions

The spectrophotometric assay was also used to test the activity of MecgoR as clarified cell lysate towards tropinone at a range of pH conditions in phosphate buffer (Figure 44 and Scheme 30). Activity increased with pH from 5.8 to 7.2, when activity began to decrease, making the optimum pH between 7 and 7.5. Jirschitzka *et al.*

highlighted an optimum at pH 6.8 for the purified MecgoR towards 2-CBT, though does not specify how much activity changes when pH conditions are altered from this. TRI is more pH dependant: Portsteffen *et al.* reported that activity toward tropinone with TRI peaks at pH 6.4 and drastically reduces when the pH are changed from this, especially at more acidic conditions. TRII however was active at pH 4.5 to 8.5.[157]

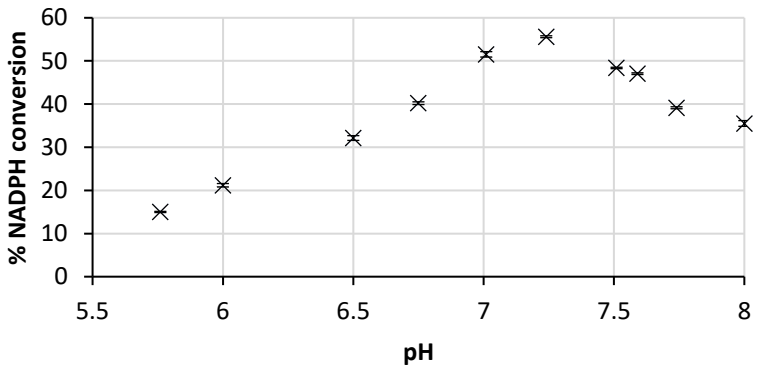
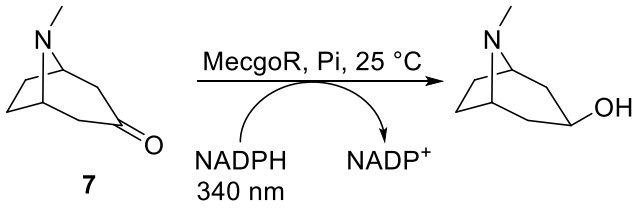


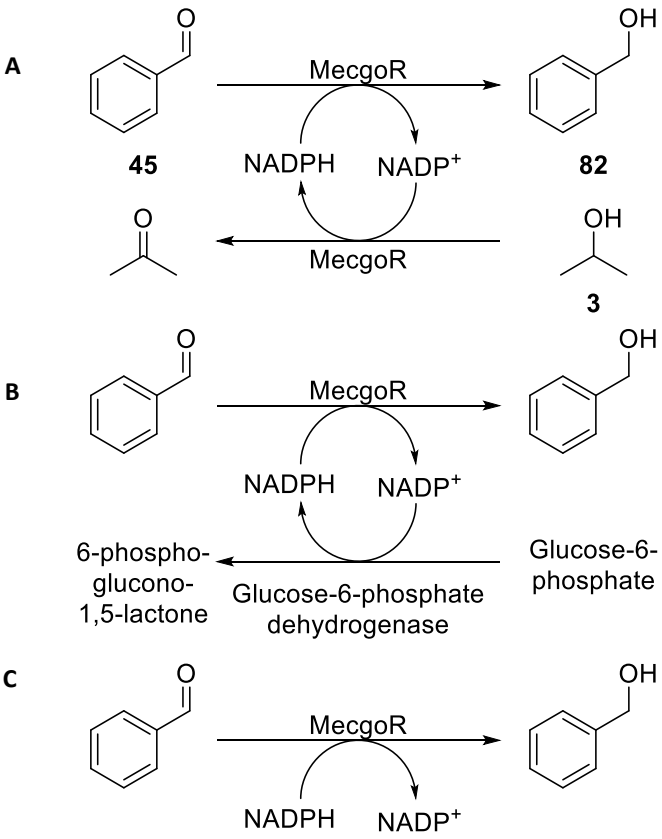
Figure 44: Graph to show % conversion of NADPH with MecgoR and tropinone at various pH conditions (200 µL): tropinone (5 mM), clarified cell lysate (0.4 mg/mL), NADPH (1 mM), Pi (100 mM), DMSO (10%, v/v). The reactions were shaken for 100 min, at 25 °C and performed in triplicate. NADPH depletion was quantified using the spectrophotometer at 340 nm.



Scheme 30: Benzaldehyde was reduced using MecgoR at various pH conditions (Figure 44).

3.3.3 Using a recycling system with MecgoR

As discussed in Section 1.2.8, NADPH is a costly co-factor necessary for these reduction reactions. Use of a co-factor recycling system mitigates these costs. Isopropanol (^{i}PrOH **3**) was initially attempted as a hydride donor (Scheme 31A), as this poses an efficient recycling system with few additional resources required.[127,158,159] For example, this has been effectively used by Mendez-Sanchez *et al.* for the reduction of 3-phenylbutan-2-one using an ADH with up to 98% yield.[159] An alternative recycling system, discussed in Section 2.6.5, is the use of G6PDH (Scheme 31B).



Scheme 31: NADPH recycling systems employed using (A) isopropanol or (B) G6PDH as a hydride donor for the reduction of benzaldehyde with MecgoR. Scheme (C) shows the reaction without a recycling system.

The reduction of benzaldehyde **45** with MecgoR performed using both recycling systems and was monitored by the production of benzyl alcohol **82**, analysed using HPLC (Figure 45). The addition of ^{i}PrOH did not result in a functioning recycling system: the percentage conversion remained the same as the negative control. The amount of ^{i}PrOH employed was reduced as this may have a detrimental effect on enzyme function, but this did not increase the activity. When commercial glucose-6-phosphate dehydrogenase (G6PDH) was used, the conversion was restored, and with 10% higher conversions than without the recycling system. G6PDH was therefore used in further reactions. The substrate concentration of benzaldehyde was increased to 20 mM in this assay, showing that MecgoR can tolerate higher substrate concentration. Reactions using NADPH were also conducted to show the reactions were functioning.

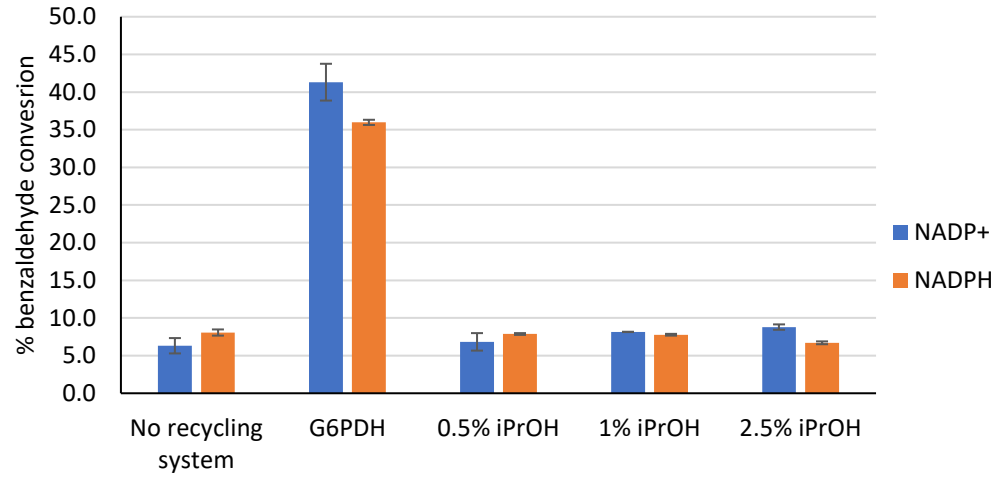


Figure 45: Graph displaying % conversion of benzaldehyde with MecgoR (200 µL): Benzaldehyde (20 mM), clarified cell lysate (0.4 mg/mL), NADP(H) (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v), ^{i}PrOH (0.5-2.5% v/v) or G6PDH (0.36 mg/mL). Reaction performed in triplicate and shook for 22 hours, at 25 °C, 400 rpm. Quantification of benzyl alcohol production by HPLC at 210 nm. MecgoR clarified cell lysates flash frozen and defrosted.

The recycling system assay was also conducted using freeze dried cells that were rehydrated before lysis, clarification and use (Figure 46). This method is preferable to the flash freezing method as a more accessible and long-term storage solution for enzyme usage. There was 15% decrease in activity relative to the used of defrosted enzyme (Figure 45) and it can be concluded that this method may also be used for enzyme storage.

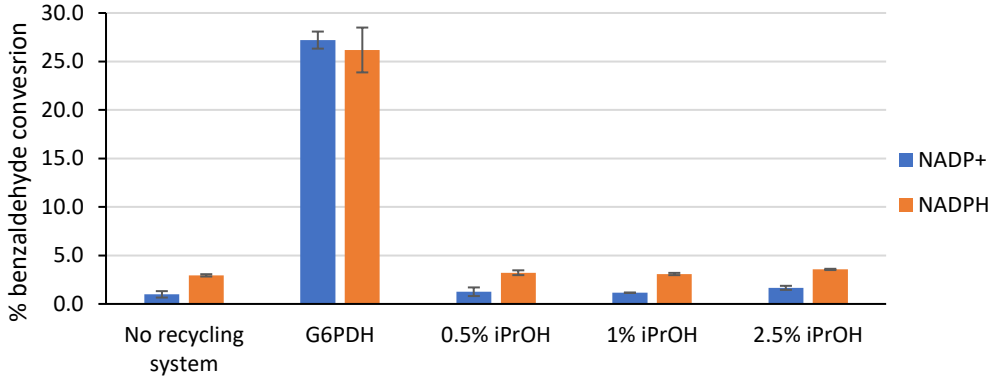


Figure 46: Graph displaying % conversion of benzaldehyde with MecgoR (200 µL): Benzaldehyde (20 mM), clarified cell lysate (0.4 mg/mL), NADP(H) (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v), ^{i}PrOH (0.5-2.5% v/v) or G6PDH (0.36 mg/mL). Reaction performed in triplicate and shook for 22 hours, at 25 °C, 400 rpm. Quantification of benzyl alcohol production by HPLC at 210 nm. MecgoR was employed as clarified cell lysates from freeze dried cells.

The effect of reaction temperature was also investigated (Figure 47). Altering the temperature of the reaction did not affect conversions significantly, although lower yields were observed at 50 °C.

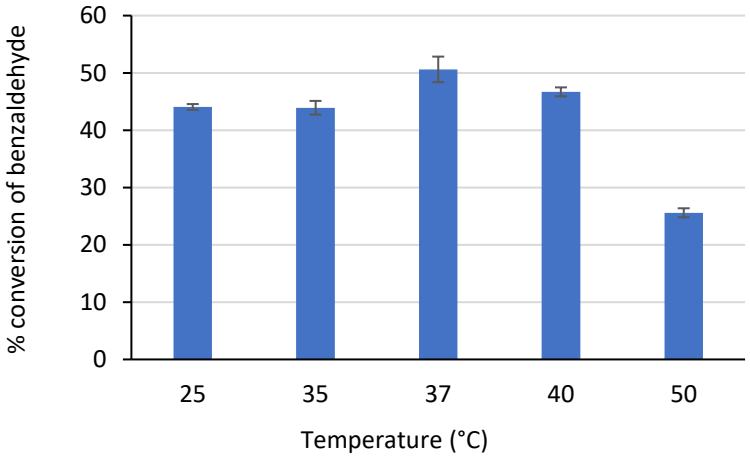
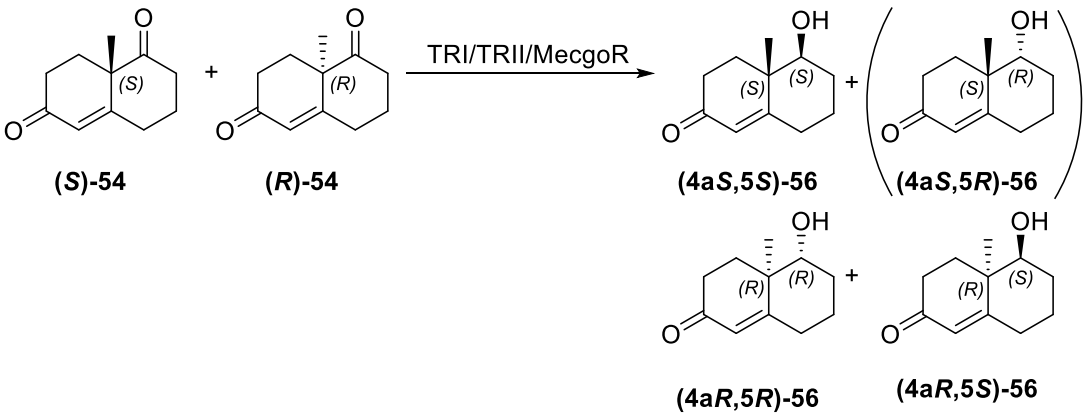


Figure 47: Graph displaying % conversion of benzaldehyde with MecgoR at various temperatures (200 µL): Benzaldehyde (20 mM), clarified cell lysate (0.4 mg/mL), NADP (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v), G6PDH (0.36 mg/mL). Reaction performed in triplicate and shook for 22 hours, at 25 °C, 400 rpm. Benzyl alcohol production was quantified by HPLC at 210 nm. MecgoR was employed as clarified cell lysates from freeze dried cells.

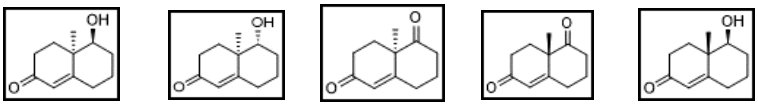
3.4 Wieland-Miescher Ketone reductions

With a motif that is employed in a large number of useful natural products, the Wieland-Miescher ketone (WMK) is of particular interest. The WMK was investigated in detail in Section 2.6 with SDRs. The 200 µL scale reactions were analysed after 100 minutes in more detail by chiral HPLC analysis as described in Section 2.6 (Table 14).

Compounds (*S*)-WMK **(*S*)-54** and (*R*)-WMK **(R)-54** were commercially available and used as starting material standards in the HPLC analysis. The compounds **(4a*S*,5*S*)-56** and **(4a*R*,5*R*)-56** were generated *via* the chemical reduction of (*S*)-WMK and (*R*)-WMK respectively using $NaBH_4$ (Scheme 24 in Section 2.6.5).[150,160] **(4a*R*,5*S*)-56** was generated from (*R*)-WMK *via* a preparative scale reaction employing SDR-17. These compounds were then used as standards for the products (calibration curves in Section 8.14). Compound **(4a*S*,5*R*)-56** was not chemically synthesised as it was more chemically challenging.[161,162]



Scheme 32: Scheme showing the reaction employed in Table 14.



Purified enzyme	Conversion (%)					
	(4a*R*,5*S*)-56	**(4a*R*,5*R*)-56**	**(4aS,5S)-56**	**(*R*)-54**	**(*S*)-54**	**Total**
TRI	0	25	25	25	1	77
TRII	0	10	0	35	53	98
MecgoR	2	0	0	35	51	88

Table 14: Table displaying % conversions in reactions of (±)-WMK with TRI, TRII & MecgoR (200 µL): (±)-WMK (5 mM), purified enzyme (0.4 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were shaken for 100 min, at 25 °C, and quantified by chiral HPLC at 230 nm.

TRI reduced the (*R*)-WMK with a 50% conversion to **(4a*R*,5*R*)-56**. (*S*)-WMK was almost totally consumed by TRI, however only half of this was converted to **(4a*S*,5*S*)-56**, as determined *via* HPLC analysis. This suggested that **(4a*S*,5*R*)-56** may be formed or reduction of the carbonyl occurred on the enone to generate **(*S*)-83**. This issue could not be resolved however because **(4a*S*,5*R*)-56** was not made synthetically and so a standard was not available. Additionally, **83** was poorly detected *via* HPLC due to the loss of conjugation in the molecule. Product analysis *via* NMR spectrometry of the reaction mixture or from an isolated yield could be used in future to determine the products of the reactions.

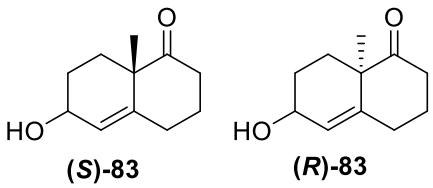


Figure 48: Potential products of the reduction of enone (*S*)- and (*R*)-83.

TRII did not reduce (*S*)-WMK, and consumed 20% of (*R*)-WMK to **(4a*R*,5*R*)-56**, demonstrated by HCPL analysis. MecgoR also did not reduce (*S*)-WMK, but 30% of the of the (*R*)-WMK was consumed. However, as **(4a*R*,5*S*)-56** nor **(4a*R*,5*R*)-** was produced, it could be suggested that the carbonyl on the enone has been reduced generating **(*R*)-83**.

In future work, preparative scale reactions will be used in order to isolate the products of these reactions, and investigate if **(4a*S*,5*R*)-56**, **(*S*)-83** or **(*R*)-83** are produced.

3.5 Quantification of reduction reaction and use of purified enzymes

Spectrophotometric assays have been extensively used in the literature with ketoreductases and also TRs specifically. However, this follows co-factor depletion (Section 2.3), so to confirm the results, HPLC and GC analysis was used to follow the substrate and product concentrations to determine reaction yields. Clarified cell lysate and purified enzymes (Figure 49) were used to reduce (±)-WMK **54** and acetophenone **29** as example substrates (Scheme 16).

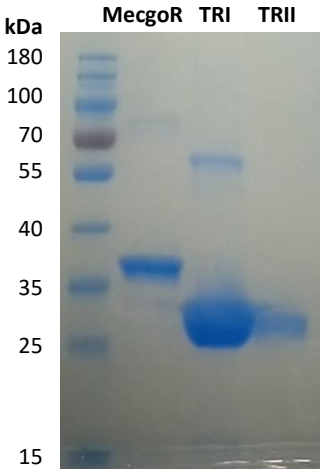
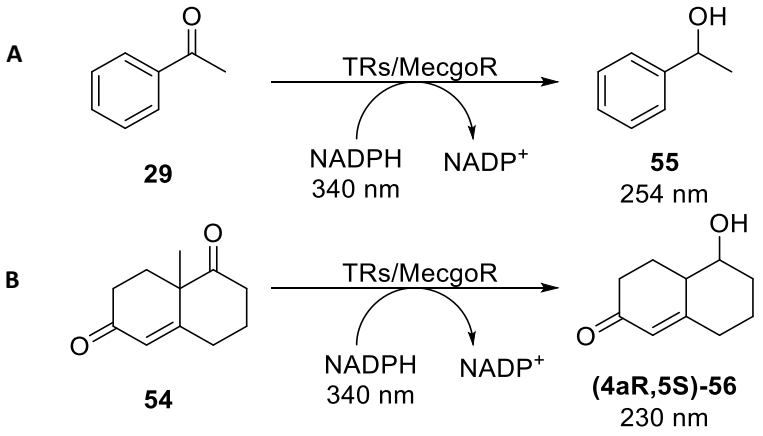


Figure 49: SDS-PAGE gel of purified MecgoR+His (36.9 kDa), TRI+His (29.6 kDa) and TRII+His (28.3 kDa).

The conversion of NADPH was measured using the spectrophotometric assay and a NADPH calibration curve. This was used to show the percent conversion as discussed in previous sections. Commercially available phenyl ethanol **55** and reduced WMK **(4a*R*,5*S*)-56** (generated from the preparative scale reaction in Section

2.6.7) were used to make a calibration curve and analyse product formation by GC and HPLC respectively (Section).



Scheme 33: The reactions for the TRs/MecgoR with (A) Acetophenone 19 and (B) WMK 54.

In general, there was a good correlation between the results from the spectrophotometric assay and the quantification assay (Table 15 and Table 16). Additionally, the purified enzyme and clarified cell lysate reactions were similar, which supported the accuracy of the previous assays which used clarified cell lysates.

WMK	Method	TRI+His	TRI	TRII+His	TRII	MecgoR
CCL	UV	62	66	11	14	3
CCL	HPLC	78	74	20	20	3
Pure	UV	80	n/a	24	n/a	13
Pure	HPLC	74	n/a	12	n/a	14

Table 15: Table displaying % conversions relative to NADPH in reactions of (±)-WMK with TRI, TRII & MecgoR (200 µL): (±)-WMK (5 mM), purified enzyme (0.4 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were performed in triplicate (standard deviation <10%) and shook for 100 min, at 25 °C. Quantification was completed by HPLC or spectrophotometric assay (denoted by "UV"). "n/a" denotes not applicable, as only enzyme with a His-tag were purified.

Acetophenone	Method	TRI+His	TRI	TRII+His	TRII	MecgoR
CCL	UV	34	40	4	5	3
CCL	GC	31	28	0	0	0
Pure	UV	66	n/a	5	n/a	3
Pure	GC	102	n/a	nd	n/a	nd

Table 16: Table displaying % conversions relative to NADPH in reactions of acetophenone with TRI, TRII & MecgoR (200 µL): acetophenone (5 mM), purified enzyme (0.4 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were performed in triplicate (standard deviation <10%) and shaken for 100 min, 25 °C. Quantification by GC or spectrophotometric assay (denoted by "UV"). "na" and "nd" denote not applicable and not determined (due to low activities) respectively.

Additionally, the products of the reactions were identified using chiral HPLC (Table 14 and Table 16). The WMK reaction is discussed in Section 3.4. TRI showed a small stereoselectivity towards the formation of (*S*)-phenylethanol (57%, 14% e.e), however was much less stereoselective than with tropinone as described in the literature. Stereoselectivities were not determined using TRII and MecgoR as these had very low conversions associated with acetophenone.

To support the previous results, the reactions were repeated for those which previously showed positive activity with purified enzymes. This is displayed in Table 17 which illustrates the data as both percent NADPH conversion and percent conversion relative to that towards tropinone. The conversions are similar to that using the clarified cell lysates to perform the biotransformations (Table 12). TR1 showed the largest substrate tolerance, and showed good activities relative to that with tropinone for all the tested substrates except 2-methylcyclohexanone **38,** pyruvic acid **35** and (±)-camphor **71**.

		% NADPH Conversion		% NADPH Conversion relative to tropinone	
	Substrate	**TRI+His**	**TRII+His**	**TRI+His**	**TRII+His**
7	Tropinone	72	75	100	100
29	Acetophenone	66	5	92	7
62	Cycloheptanone*	46	4	65	5
38	2-Methylcyclohexanone	11	7	16	9
39	3-Methylcyclohexanone	54	13	75	17
40	2-Cyclohexen-1-one	42	6	59	8
52	1-Indanone	32	5	45	7
53	2-Indanone	79	10	110	13
54	Wieland–Miescher ketone	80	24	111	32
21	1-Methyl-4-piperidone*	84	61	117	81
37	Ethyl-4-oxo-1-piperidinecarboxylate	62	65	87	87
36	1-(2-Methyl propyl)-4-piperidone	78	69	109	92
43	2,2,2,4'-Tetrafluoroacetophenone	69	57	96	76
42	2,2,2,3'-Tetrafluoroacetophenone	85	67	119	89
41	2,2,2-Trifluoroacetophenone	84	74	118	99
44	4'-*N*-Propyl-2,2,2-trifluoroacetophenone	91	45	128	60
19	Cyclohexanone	84	18	118	24
45	Benzaldehyde	84	7	117	9
50	1-Phenyl propan-2-one	73	8	102	11
47	2-Heptanone	28	10	39	13
48	Methylacetoacetate	78	7	108	9
35	Pyruvic acid*	2	0	3	0
71	Camphor	11	9	15	12

Table 17: Purified TRs with a selection of substrates that had been found to be active in Table 12 and Table 13. Ratio with tropinone's activity Table to show the % NADPH conversion portrayed as a % of tropinone (200 µL): substrate (5 mM), clarified cell lysate (0.9 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were shaken for 100 min, at 25 °C. The reactions were performed in triplicate (standard deviation <10%) and quantification by the spectrophotometric assay. * denotes that DMSO (10% v/v) co-solvent was not used (negative control was adjusted accordingly). Substrates are depicted in Figure 50.

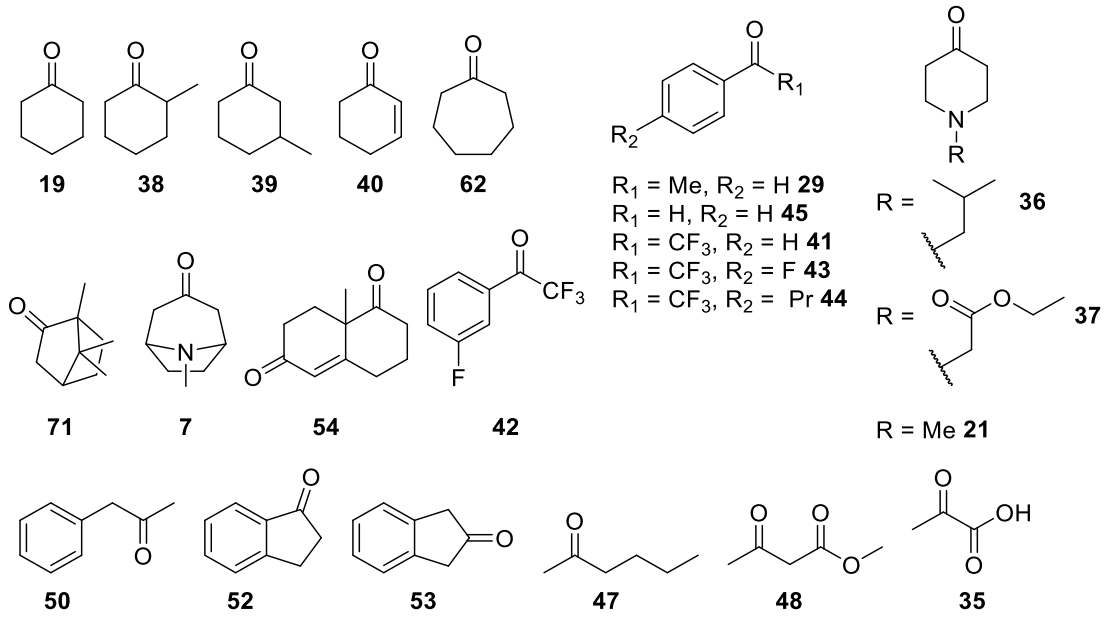


Figure 50: Substrates used in Table 17.

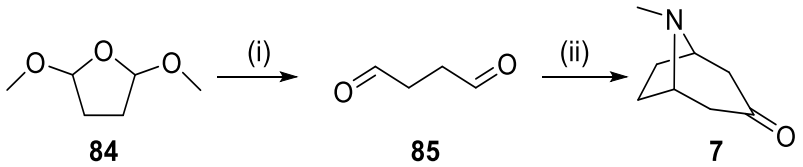
3.6 Testing TRs and MecgoR with tropinone analogues

3.6.1 Synthesis of tropinone analogues

As outlined in Section 1.7, an aim of this project is to generate analogues of tropinone and ultimately used these to investigate enzymatic substrate specificities and generate complex products in high stereoselectivities.

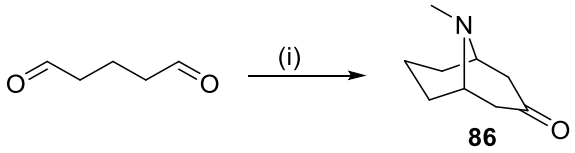
3.6.1.1 Tropinone and expanding the ring size

A synthetic route to tropinone was important so that the protocol could be used to make analogous structures. Commercially available 2,5-dimethoxy THF **84** was used as starting material for the generation of tropinone *via* the formation of **85**, following Chiou's protocol (Scheme 34).[163] Purification of the product **7** was carried out using a Kugelrohr double-distillation method and tropinone **7** was isolated in 22% yield.



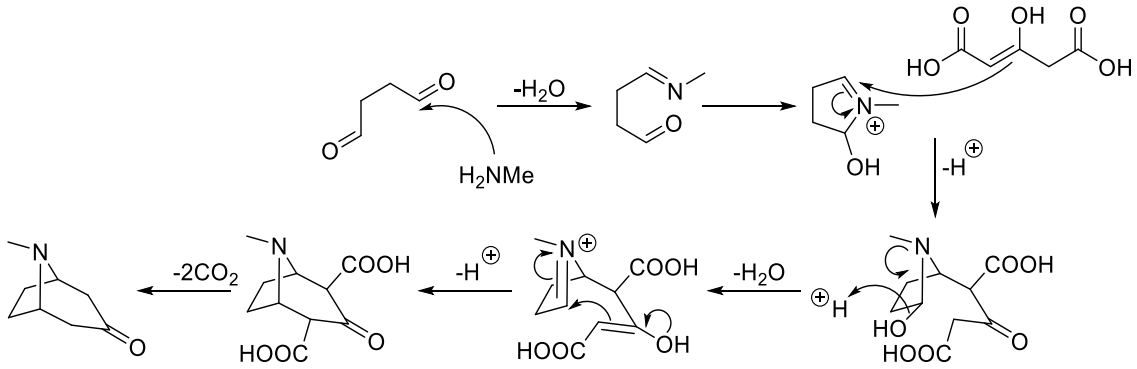
Scheme 34: Synthesis of tropinone 7, using Chiou's protocol.[163] (i) HCl, 70 °C, 30 min; (ii) NaOH (3.9 equiv), AcOH (3.9 equiv), acetonedicarboxylic acid (1.1 equiv), CH_3NH_2·HCl (1.1 equiv), 40 °C, 1 h, pH 5, 70%.

A similar route was used to synthesise pseudopelletierine **86** in 72% yield, following a route by Vernekar *et al.* using cheap and commercially available starting materials (Scheme 35).[164]

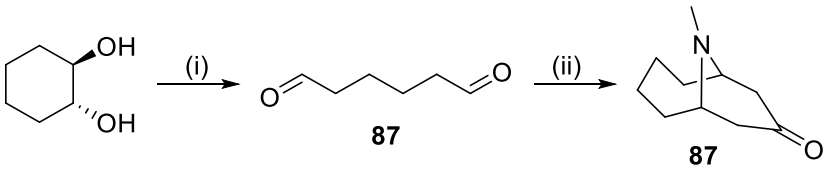


Scheme 35: Pseudopelletierine 86 was synthesised: (i) NaOH (0.36 equiv), acetonedicarboxylic acid (0.17 equiv), CH_3NH_2·HCl (1.48 equiv), 25 °C, 16 h, then 75 °C, 1 h, 72%.

These reaction conditions were modified and were used to make compound 10-methyl-10-azabicyclo[4.3.1]decan-8-one (MADO, **87**) based on the proposed mechanism for the synthesis of tropinone **7** *via* a well-known double mannich reaction in 94% yield (Scheme 36).[163,165] This compound is a ring expanded analogue of pseudopelletierine, and was used to identify the enzyme's tolerance of larger ring systems. 1,6-Hexandial **88** was synthesized using a literature procedure and directly used in the next step (Scheme 37).



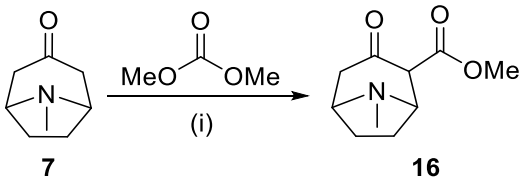
Scheme 36: Proposed mechanism for the synthesis of tropinone *via* Chiou's route.[163]



Scheme 37: Synthesis of MADO 87. (i) $NaIO_4$, SiO_2, H_2O/CH_2Cl_2, 85%. (ii) $NH_2Me.HCl$, 1,3-acetonedicaboxylic acid, phosphate buffer pH 5, 24h, 25 °C, then HCl, 1 h, 75 °C, 94%.

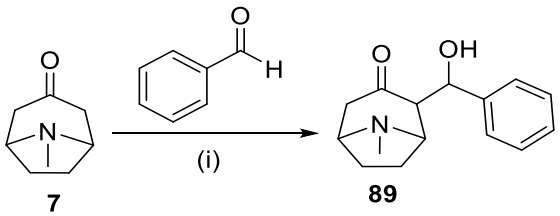
3.6.1.2 Other tropinone analogues

2-Carbomethoxytropinone (2-CBT, **16**) has a chiral centre and an ester side chain. This would probe the capacity of the TRs and MecgoR to accept substrates that are more sterically demanding than tropinone and generate more complex products. Compound **16** was generated *via* the deprotonation of tropinone and the addition of dimethyl carbonate in 22% yield after purification by silica flash column chromatography (10% $^{i}PrNH_2$/30% EtOAc/59% hexane/1% Et_3N) (Scheme 38).[166]

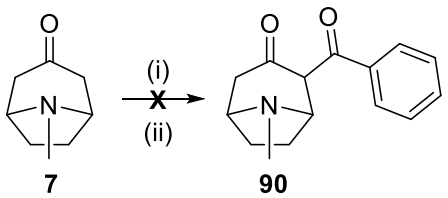


Scheme 38: Synthesis of 2-CBT 16.[166] (i) NaH (60%), cyclohexanone, 22%.

Following a similar mechanism of tropinone deprotonation and subsequent addition, *n*-Butyllithium (*n*-BuLi) solution was used to make **89** in 13% yield (Scheme 39).[167] Both NaH and *n*-BuLi were employed to in the synthesis of **90**, however neither approach yielded the desired product (Scheme 40).

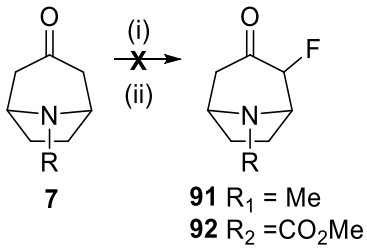


Scheme 39: Synthesis of 89.[167] (i) LDA, nBuLi, 13%.



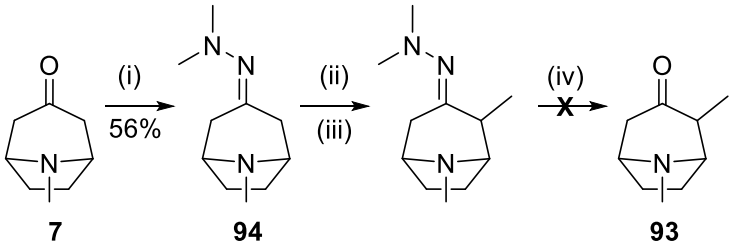
Scheme 40: Attempted synthesis of 90. Method 1: (i) NaH (60%), cyclohexanone, diphenyl carbonate, 80 °C. [168] **Method 2: (i) diisopropylamine, THF, n-BuLi, 0 °C, 25 min; (ii) benzaldehyde, -78 °C, 15 min.**[167]

The synthesis of a fluorinated asymmetrical analogue of tropinone was also attempted following a literature procedure (Scheme 41).[169] Though both **91** and **92** products were identified using mass spectrometry, the products could not be isolated. Increased equivalents of Selectfluor® were used (up to three equivalents), but this did not improve yields.



Scheme 41: Attempted synthesis of fluorinated tropinone analogues: (i) Diisoproylamine, THF, *n*-BuLi, TMSCl, -78 °C, 30 min; (ii) Selectfluor, CH_3CN, 25 °C, 16 h.

The literature procedure recorded by Lazny *et al.* was attempted for the synthesis of **93** (Scheme 42).[170] The first intermediate **94** was successfully isolated in 56% yield. The following steps in Scheme 42 were performed, and the reactions were followed using a combination of mass spectrometry, TLC, and NMR spectroscopy. Although the mass of the final product **93** was noted using LC-MS, the product yields were minimal and the compounds were not isolated.

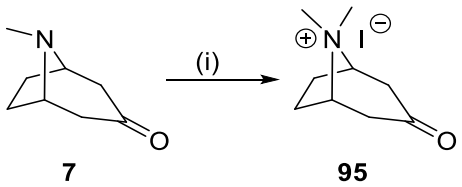


Scheme 42: Attempted synthesis of a methylated tropinone analogue 93: (i) Me_2NNH_2, pTsOH, 75 °C, 16 h; (ii) *n*-BuLi, THF, 0 °C, 4 h; (iii) MeI, 25 °C, 16 h (iv) TFA/H_2O/THF (2:1:7), 25 °C, 4 h. [170]

3.6.1.3 Functional groups on the Nitrogen heteroatom

8,8-Dimethyl-3-oxo-8-azonia-bicyclo[3.2.1]octane iodide (IDABO) **95** was also synthesised in 97% yield in order to investigate whether the enzymes can accept tropinone analogues with a quaternary amine present (Scheme 43).[171] This could also

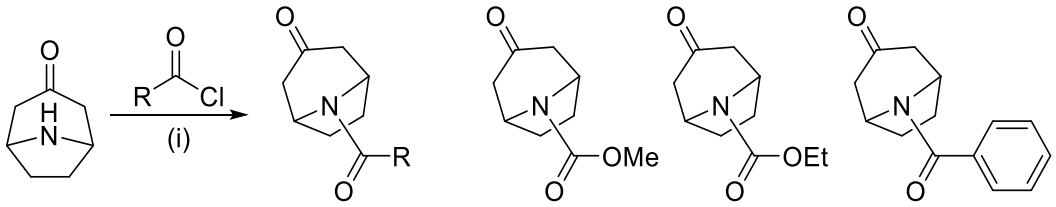
give an indication of the protonation state of tropinone preferred in the TRI and TRII active site.



7 95

Scheme 43: Synthesis of IBADO 95.[171] (i) MeI (1.1 equiv), acetone, 97%.

Other groups on the nitrogen were also incorporated in **96**, **97**, **98** all of which were generated using acid chlorides and purified *via* flash column chromatography in 18%, 92% and quantitative yields respectively (Scheme 44).[172–174]



96 97 98

Scheme 44: Synthesis of N-functionalised tropinone. Compound 96: R = OMe, DMAP, Et_3N, 18%. Compound 97: R = OEt, toluene, K_2CO_3, 92 %. Compound 98: R= Ph, CH_2Cl_2, Et_3N, Quant.[172–174]

3.6.2 Activities of TRs and MecgoR towards tropinone analogues

After synthesis of a selection of compounds related to the tropinone motif, the TRs and MecgoR was assayed against them. Purified enzymes were used in the 200 µL scale spectrophotometric assay as previously described, in which NADPH depletion was monitors at 340 nm, and a NADPH calibration curve used to calculate the percent conversion of NADPH, and infer activity (Table 18).

	NADPH conversion (%)			NADPH conversion relative to tropinone (%)		
Substrate	**TRI**	**TRII**	**MecgoR**	**TRI**	**TRII**	**MecgoR**
Tropinone **7**	71	49	79	100	100	100
Pseudopelletierine **86**	55	9	61	77	18	77
MADO **87**	0	0	17	0	0	21
2-CBT **16**	17	9	82	24	17	103
89	79	14	46	110	29	59
Nortropinone.HCl **18**	72	71	84	102	143	107
IDABO **95**	0	0	81	0	0	103
Benzyl nortropinone.HCl **99**	66	82	82	92	166	103

98	57	0	82	80	0	103
96	69	1	74	96	2	94
97	64	0	76	90	0	97

Table 18: Table to show the percent conversion of NADPH (200 µL): Substrate (5 mM), purified enzyme (0.4 mg/mL), NADPH (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were shaken for 100 min, at 25 °C and performed in duplicate. NADPH depletion was quantified by spectrophotometer. Structures of the substrate used are depicted in Figure 51.

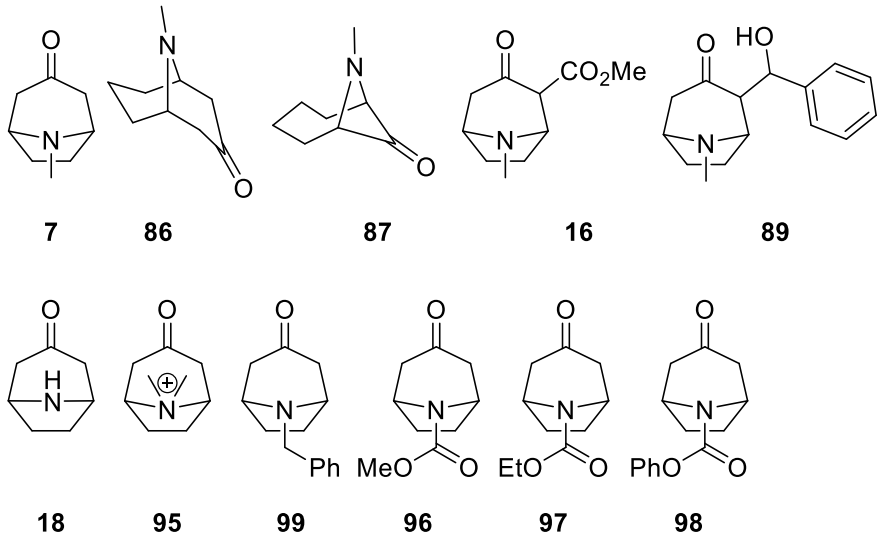


Figure 51: Structures of the compounds using in Table 18.

TRI was active towards all the substrates tested except MADO **87** and IDABO **85**. TRI was tolerant to additional functionally on the nitrogen bridge, and showed activity towards *n*-benzylnortropinone **99**, **97**, **96**, and **98**. TRI could accept pseudopelletierine **86**, but displayed low activities towards esterified tropinone analogue 2-CBT **16**.[68]

Boswell *et al.* have shown that there was little activity towards pseudopelletierine **86** when using TRI from *D. stramonium*, and this differs from the finding here that TRI accepts the compound.[69] Docking this compound in to the active site of TRI also shows that may complex in the same orientation as tropinone. There appears to be 3.9 Å between the carbonyl and the NADPH hydride, which is a reasonable distance noted for a reaction to take place and the same distance when tropinone was docked in to the active site (Figure 52). Boswell *et al.* used different reaction conditions, notably, 2.5 mM concentration of **86**, 0.2mM NADPH, pH 6.4 buffer and a 30 °C reaction temperature, and the reaction duration is not defined, as opposed to 5 mM concentration of **86**, 1 mM NADPH, pH 7.4 and 25 °C used here, so

results cannot be directly compared.[69] Monitoring production of 86 using GC or NMR spectroscopy in the future would determine conversion yields.

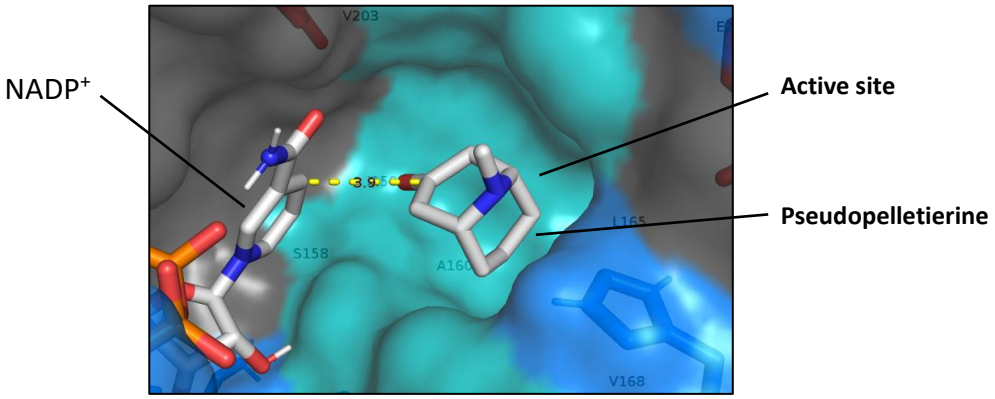


Figure 52: Pseudopelletierine 86 docked into TRI. Figure 53: Pseudopelletierine docked to TRI with the active site residues in teal colour, and His112 in blue.

One very surprising result was that **89** was accepted by TRI. 2-CBT is smaller in size than **89** and is it was unlikely that a larger asymmetrical functional group would be accepted. MADO **87** and IBADO **95** were also not accepted by TRI, suggesting that there was limit to the ring size tolerated and that the quaternary amine inhibited activity.

TRII solely displayed high activities with tropinone **7**, nortropinone **18**, and benzyl nortropinone **99**. Activities towards **18** and **99** were greater than that towards tropinone with 143% and 166% relative activity. TRII however showed lower or no activities towards the other substrates tested and as discussed before, this may be due to its less exposed active site. **98** was not accepted by TRII, but benzyl nortropinone.HCl **99** was, perhaps due to electronic effects.

MecgoR showed similar substrate specificities as TRI, as they both accepted **86**, **96**, **97** and **98**, although TRII did not. MecgoR is stereospecific for the β-configuration of tropine, similar to TRII and unlike TRI, and so MecgoR could be used as a strategy to access these β-alcohol compounds. MecgoR did display similar activity towards 2-CBT **16** and with tropinone **7**. Jirschitzka *et al.* noted that tropinone displayed only 45% of the activity that 2-CBT demonstrated with MecgoR.[61] The previous results in Table 13 showed that DMSO has a clear effect on activity: the absence of DMSO a co-solvent at 10% v/v caused a reduction to 43% relative activity

towards tropinone. As Jirschitzka did not record the use of DMSO and it is proposed that the findings here agree when considering the use of co-solvents.

Compound **89** was readily accepted by MecgoR, suggesting that MecgoR may be capable of processing bulkier compounds. MADO **87** was also accepted but showed low activities. Furthermore, IDABO **95** was accepted by MecgoR, suggesting that the enzyme is more accepting of quaternary amine groups than TRI or TRII. This also may suggest a wide tolerance to pH, as the enzyme may be more able to accept charged tropinone, an idea that is supported the activity of MecgoR at various pH conditions above (Section 3.3.2 and Figure 44).

3.7 Larger scale reactions using MecgoR

3.7.1 Benzaldehyde reductrions

Initially, a 5mL scale reaction was attempted with benzaldehyde **45** was used a standard substrate as it was easy to monitor both substrate consumption and product formation by HPLC (Table 19). This had been used for testing the recycling system earlier (Figure 45). As the concentration of benzaldehyde increased the reaction gave lower conversions, suggesting substrate or product inhibition effects. When using 5 mM of benzaldehyde, all the substrate was consumed in the reaction. The reaction was further scaled up to 50 mL, and yielded 100 % of benzyl alcohol by HPLC analysis.

[Benzaldehyde] (mM)	% Benzaldehyde conversion
5	Quant.
25	50
50	9

Table 19: Table displaying % conversion of benzaldehyde with MecgoR (5 mL): Benzaldehyde (5-50 mM), clarified cell lysate (0.4 mg/mL), $NADP^+$ (3 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v), G6P (100 mM), G6PDH (3000 U). Shake 24 h, 37 °C. Quantification of benzyl alcohol by HPLC at 210 nm. MecgoR clarified cell lysates from freeze dried cells.

3.7.2 2-CBT reduction

2-CBT **16** is the natural substrate for MecgoR and was well accepted in small scale reactions (Section 1.5.1) and in the literature.[61] This however yielded only 62% conversion at 5mL scale after 24 hours, analysis by substrate depletion using HPLC analysis. Jirschitzka *et al.* only reported reactions with MecgoR at a 100 µL scale.

Despite the lower conversions, this is the first use of the enzyme at a larger scale, and should be further optimised in the future.

3.8 Conclusions and future work

In summary when investigating the effect of a His-tag on TRs from *D. stramonium,* the activities towards cyclohexanone and tropinone were similar. TRI was also demonstrated to be tolerant to a wider range of substrates than TRII. This is consistent with reports by Portsteffen *et al.* which noted higher activities of TRI. The substrate scope for these enzymes has been expanded compared the literature, for example, conversions of substrates included fluorinated rings, bicyclic rings and tropinone analogues have not been reported in the literature.

Furthermore, the substrate scope of MecgoR was greatly expanded upon from that recorded in the literature. MecgoR showed a similar substrate tolerance to TRI, but also tolerated some asymmetric compounds with the tropinone motif, and was tolerant to wider range of pH conditions. The optimum pH for MecgoR was 7.4 and a recycling system was employed to scale the reaction using benzaldehyde as a model substrate.

MecgoR and TRI have accepted a wide range of substrates, including tropinone analogues. The reduction of useful substrates will in future work be performed at a preparative scale, as well as with WMK in order to identify the products of the reaction discussed in Section 3.4. Furthermore, product isolation will support the conversions illustrated and activities described in the literature.

Reductions using tropinone analogues **96**, **97**, **98**, **89** and **87** have not been reported in the literature, and isolation of the products of MecgoR will be investigated further for use as a biocatalyst, and the x-ray crystal structure could be resolved for the enzyme to aid this. Sequence homology modelling could alternatively be used to provide additional information about the enzyme.

4 Drain metagenome aldo/keto reductases

4.1 Introduction

Having established that MecgoR was a versatile enzyme (see Section 3.3), it was proposed that similar enzymes may be found using the metagenomic strategies developed at UCL. As discussed in Section 1.1.5.2, DNA has been extracted from a residential drain and sequenced using the Illumina MiseQ sequencing platform. In a similar approach to that used for the identification of SDRs, various quality control applications were applied, and Pfam IDs were assigned (Section 2.1). These measures include ensuring the presence of an initiator methionine and a stop codon, being of adequate sequence length and removal of redundant sequences.

4.2 Bioinformatics and gene selection

MecgoR is classed as an Aldo-Keto Reductase (AKR) and there is potential for finding similar enzymes from metagenomic databases. BLAST sequence alignment was used to select for sequences with 24% or higher sequence identity to MecgoR in the following metagenomic samples: tongue, drain, soil A and soil B (two other annotated metagenome libraries maintained by UCL). All the retrieved sequences were annotated as AKRs, though this was not a prerequisite for selection, and found in the drain metagenome.[††] When the genes were translated and interrogated on the NCBI protein BLAST online tool, all the genes most closely aligned to either AKRs or oxidoreductases (Table 21). As discussed in Section 1.1.5.2, the drain metagenome generated about ten times as much data as the tongue genome. A total of 16 sequences were selected as showing higher sequence identities, and were judged as full length, based on the presence of a stop codon and the length of the gene (Table 20). However, the AKR Pfam ID generated 82 reads in total. As a comparison, the SDR Pfam ID generated 374 sequences in the drain metagenome, whereas the tongue metagenome generated 139. This low number in the tongue metagenome was also due to highly similar sequences that were considered redundant and not counted.

[††] Dragana Dobrijevic prepared the metagenomic database.

AKR No.	Contig name	Gene Length (bp)	Protein Length (aa)	(kDa)	GC content (%)	Restriction sites in primer	Identity to MecgoR (%)
1	contig-120_19664_1	846	289	31.6	63	*BsaI*	31
2	contig-120_6751_1	774	265	28.83	65	*NdeI Xho1*	34
3	contig-120_28243_1	831	284	30.68	58	*NdeI Xho1*	35
4	contig-120_19634_1	981	336	36.37	64	*NdeI Xho1*	25
5	contig-120_8191_2	969	332	35.16	69	*BsaI*	24
6	contig-120_9148_1	969	332	35.16	68	*BsaI*	26
7	contig-120_562_6	828	283	36.17	51	*BsaI*	24
8	contig-120_3698_1	981	336	60.36	63	*NdeI XhoI*	24
9	contig-120_3273_2	1035	354	36.73	69	*BsaI*	24
10	contig-120_16375_1	1242	425	45.52	66	*NdeI XhoI*	25
11	contig-120_3402_3	1035	354	37.85	69	*NdeI XhoI*	28
12	contig-120_982_2	1035	354	37.9	68	*BsaI*	25
13	contig-120_1078_6	1035	354	37.65	68	*NdeI XhoI*	27
14	contig-120_33576_1	690	235	25.87	62	*NdeI XhoI*	25
15	contig-120_49236_1	651	222	23.78	68	*NdeI HindIII*	25
16	contig-120_1656_3	1038	355	38.33	60	*NdeI XhoI*	32

Table 20: AKR genes selected from the drain metagenome and sequence identity to MecgoR. Values were calculated using the online ExPASy ProtParam tool.[124]

The sequence residues were interrogated using the NCBI Blast online tool to find similar known enzymes (Table 21). This showed that many of the enzymes that were selected for retrieval displayed low sequence identities to recorded enzymes, so are considered novel. Additionally, as predicted the functionality assigned to the most similar enzymes were all reductases.

AKR	NCBI Annotation	Identity (%)
1	aldo/keto reductase [Bradyrhizobium valentinum]	88
2	aldo/keto reductase [Mesorhizobium sp. ORS3324]	79
3	oxidoreductase [Pelagibacterium sp. SCN 63-23]	68
4	aldo/keto reductase [Paenibacillus sp. 1_12]	54
5	aldo/keto reductase [Pseudoxanthomonas sp. Root630]	89
6	aldo/keto reductase [Pseudoxanthomonas mexicana]	91
7	40% ? oxidoreductase [Pseudarthrobacter siccitolerans]	40

8	NADP(H)-dependent aldo-keto reductase [Alcanivorax pacificus]	63
9	aldo/keto reductase [Sphingopyxis sp. H115]	95
10	aldo/keto reductase [Pseudoxanthomonas wuyuanensis]	83
11	aldo/keto reductase [Pseudoxanthomonas wuyuanensis]	88
12	aldo/keto reductase [Pseudoxanthomonas sp. CF385]	97
13	aldo/keto reductase [Pseudoxanthomonas mexicana]	97
14	aldo/keto reductase [Bradyrhizobium paxllaeri]	90
15	oxidoreductase [Brevundimonas subvibrioides]	88
16	aldo/keto reductase [Candidatus Moduliflexus flocculans]	76

Table 21: Table to show functional assignment of the retrieved enzymes, using the online NCBI protein BLAST tool.[121]

4.3 Primer design and PCR

Primers were designed consisting of 3 adenosine nucleotides, followed by the restriction enzyme recognition site and the gene itself. The ATG start site was incorporated into the recognition site of *Nde*I, and so this was removed from the gene section of the primer when necessary. The reverse primer was constructed from the reverse complement of the last line of the gene without the stop codon (final 3 nucleotides). To optimise the efficacy of the primer, nucleotides were removed from the gene so that the primer melting temperature (Tm) was greater or equal to 72 °C, the primer was greater or equal to 20 nucleotides in length and also ended in a guanine or a cytosine. The Tm was estimated using Multiple Primer analyser online tool from ThermoFisher Scientific.[175]

PCR was used to amplify the desired gene from the metagenomic DNA. This was attempted several times to retrieve the largest number of genes (Figure 54). Initially attempts used a mixture consisting of dNTPs (40 µM), phusion polymerase (1% v/v), GC buffer (5x, 20% v/v) (or a similar buffer made up freshly) and 3% DMSO was used for genes 9 to 16 to successfully amplify the DNA (Figure 54A). Genes 1 to 8 were not successfully retrieved using this reaction mixture. As 3% v/v of DMSO was used, it was proposed that a higher DMSO concentration may aid the PCR process. Genes for AKR-1, AKR-2, ARK-5 and AKR-6 were amplified successfully with the addition of 5% v/v DMSO (Figure 54B). For the remaining genes, a selection of different conditions were attempted, which had the following additional components added to the above mixture: betaine (2 M); betaine (2 M) and DMSO (10% v/v); ethylene glycol (1 M); ethylene glycol (1 M) and betaine (2 M); Q5 polymerase (0.5%

v/v) and GC enhancer (5x, 10% v/v); DMSO (10% v/v); DMSO (20% v/v). The addition of Q5 polymerase (0.5% v/v) and GC enhancer (5x, 10% v/v) successfully yielded amplified gene 8 (Figure 54C). Amplification of the other genes was not successful (AKR-3, AKR-4, AKR-7).

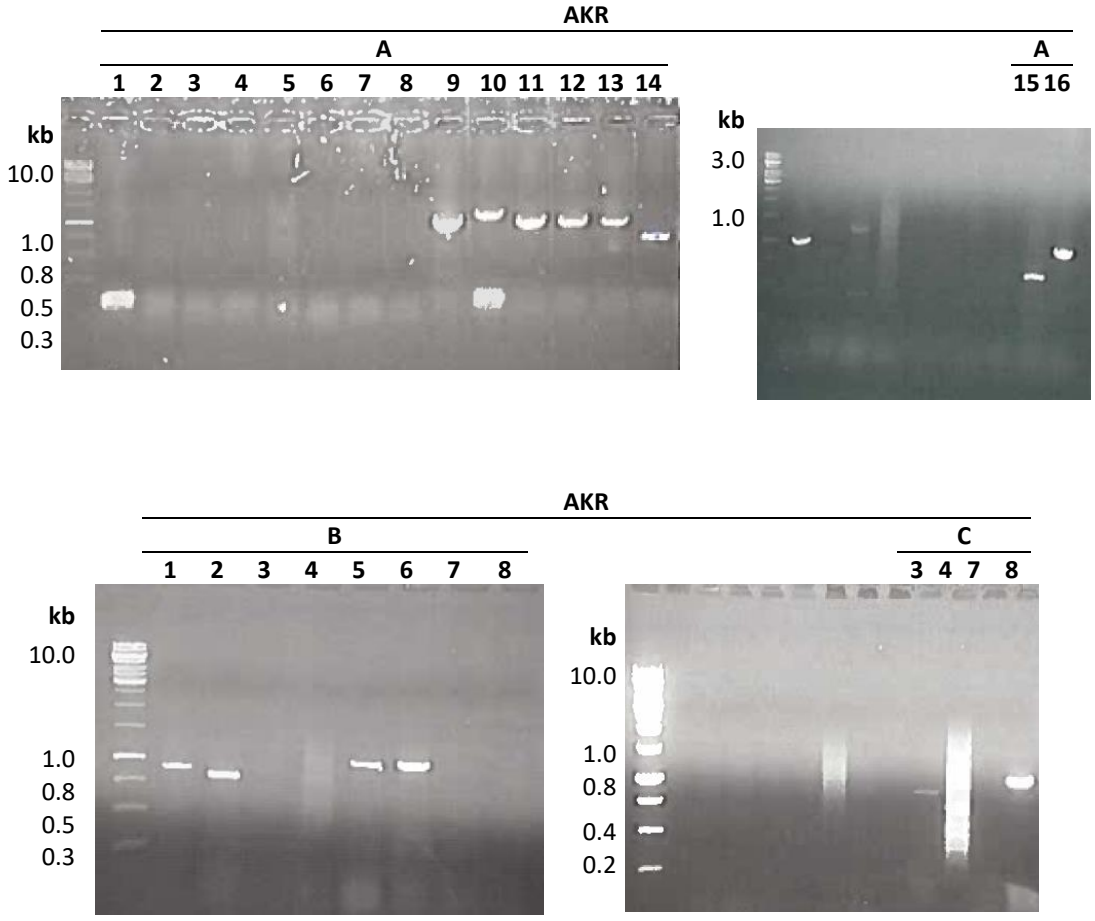


Figure 54: Products of the PCR on agarose gel and visualised using ethidium bromide and *uv* light with the following additives: (A) DMSO (3%); (B) DMSO (5%); (C) Q5 polymerase (0.5% v/v) and GC enhancer (5x, 10% v/v). The following products were retrieved: (A) AKRs-9 to 16; (B) AKRs-1, 2, 5, 6; (C) AKR-8. Unlabelled sections correspond to some unsuccessful reaction (data not shown for all unsuccessful reactions). See Table 20 for sequence lengths.

4.4 Cloning of AKRs from the drain metagenomic DNA

Two strategies were employed for cloning the genes into vectors: a novel method developed at UCL (Dobrijevic, unpublished) and classical cloning.

4.4.1 Novel cloning strategy

The latter strategy was employed for genes which did not have the restriction site for *BsaI* in the sequence as this method is a faster one-pot procedure, using a plasmid constructed by Dragana Dobrijevic (pET29:SacB). This method was attempted for ARK-1, ARK-5, ARK-6, ARK-9, and ARK-14. It used a one-pot ligation process followed by transformation into competent Top10 *E. coli* cells and plating out onto an agar plate supplemented with sucrose (10%) and kanamycin (50 µg/mL). AKR-1, ARK-5, ARK-6, and ARK-9 were successfully cloned using this method although AKR-14 was not so was not pursued further.

The AKR genes were sequenced using Eurofins Sequencing service, and compared to the genes from the metagenome (Table 22). The retrieved enzymes showed high sequence identities and successful incorporation of the gene into the vector. For AKR genes 5 and 9, the identity is lower. This is because the sequencing stopped prior to the end of the gene and reverse sequencing was not completed. Sequencing was completed up to 89% of AKR-5 and 81% of AKR-9. Of the sequenced sections, there was 100% sequence identity.

Gene	Identity (%)
1	99.2
5	92.2
6	99.5
9	80.7

Table 22: Table to show the sequence identity between the genes selected from the drain metagenome and the retrieved gene sequences (calculated using EMBOSS Needle Pairwise sequence alignment online tool).[17]

4.4.2 Classic cloning strategy

A restriction and ligation cloning approach was used for the following AKRs: 2, 3, 8, 10, 11, 12, 13, 15, and 16. Classical cloning involved digesting both the gene insert and vector, and then ligating these together to form a complete plasmid with the required gene, using restriction enzymes and DNA ligase respectively. The vectors were then transformed into chemically competent OneShot Top10 *E. coli* cells, which have a high insert stability. Successful sequence insertion was monitored *via* a restriction digest followed by visualisation on an agarose gel (Figure 55).

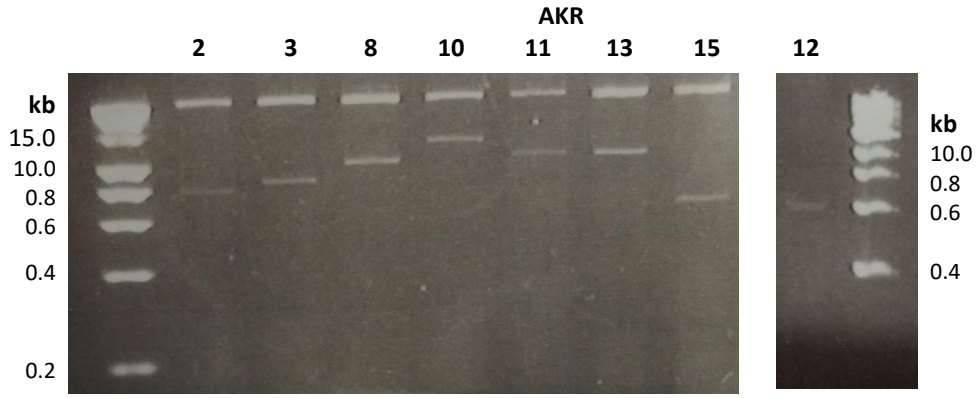


Figure 55: The restriction digest on agarose gel and visualised using ethidium bromide and *uv* light. The heavier band is thepET29a vector (1660 kDa), and the smaller bands show the presence of the smaller insert. See Table 20 for sequence lengths.

A commercial miniprep kit was then used to collect the plasmids for all the AKR and these were then transformed in to BL21 DE3 *E. coli* cells for efficient expression. AKRs-3, 5, 6, 12, 13 all expressed well and could be readily purified using nickel resin packed columns (Figure 56).

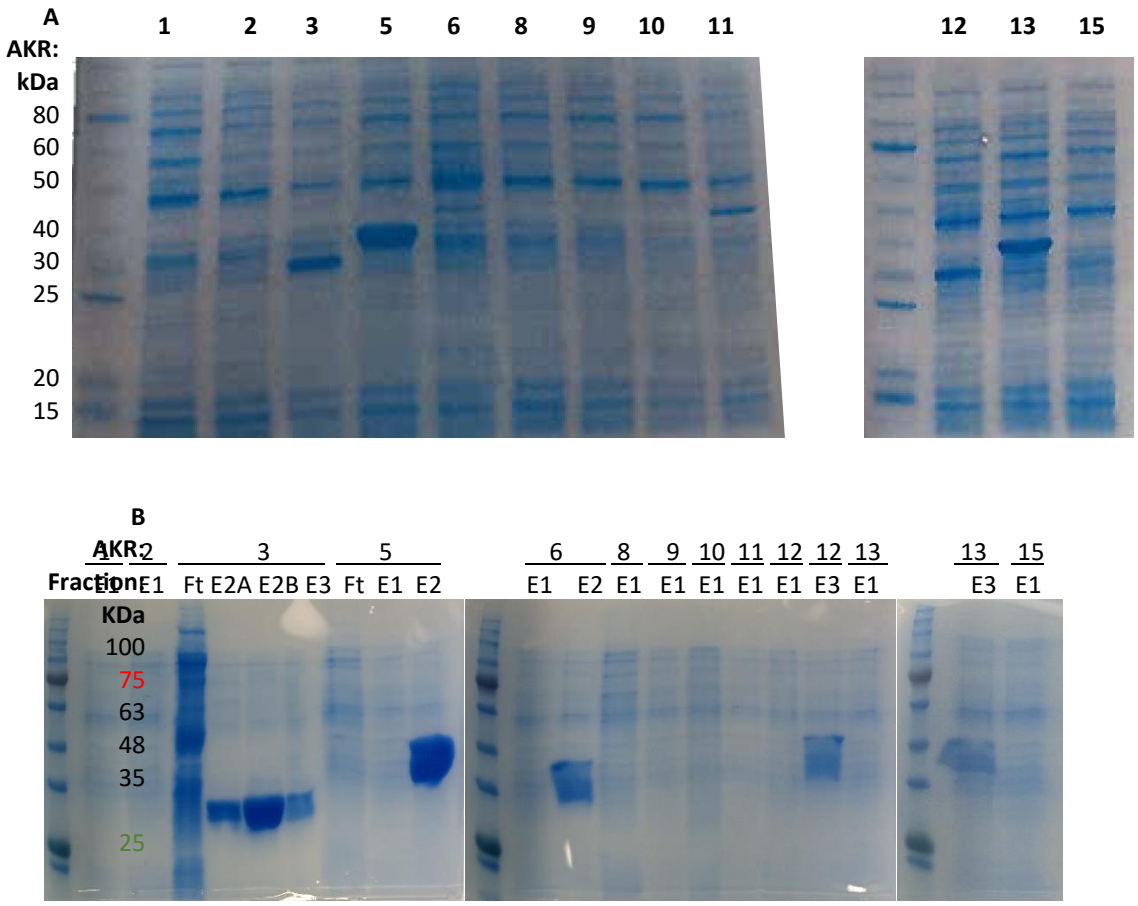
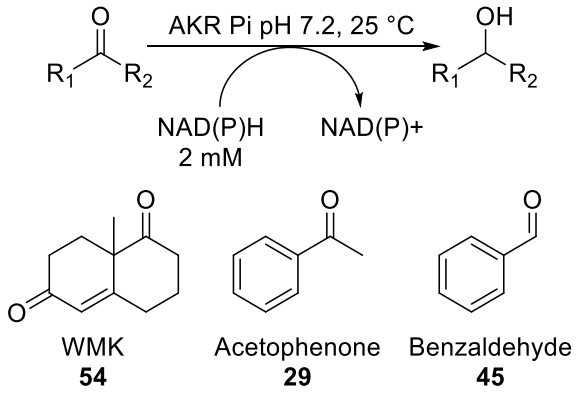


Figure 56: SDS-PAGE gels of (A) clarified cell lysate and (B) protein-containing fractions of purified AKRs from the drain metagenome. Ft = Flow through E= elution fraction.

4.5 Initial activity assays and assay optimisation

Activity was tested with the AKRs which expressed well: AKRs-3, 5, 6, 12, 13 (Figure 56) towards **54**, **29** and **45**. Product formation was quantified using HPLC (Scheme 45 and Figure 57).



Scheme 45: Reaction to quantify AKR activity towards compounds 54, 29, and 45.

When using NADH as the cofactor, AKR-3 and AKR-12 showed good activity with benzaldehyde **45**. No other activity was noted. Greater activity was demonstrated when using NADPH as the cofactor: AKR-3, AKR-12 and AKR-13 all showed activity with benzaldehyde, and AKR-6 to a lesser extent. Small activities towards the WMK **54** and acetophenone **29** with AKR-3 and AKR-12 were also noted.

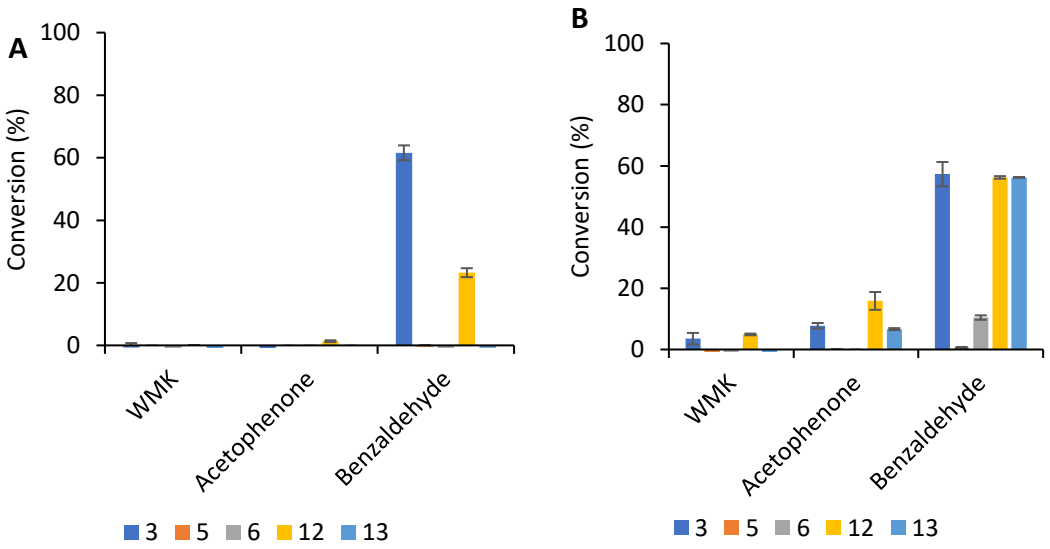


Figure 57: % conversion relative to NAD(P)H limiting reagent with a selection of AKRs from the drain metagenome. (A) NADH and (B) NADPH (200 µL): substrate (5 mM), purified enzyme (0.4 mg/mL), NAD(P)H (2 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). Shake 16 h, 30 °C, 500 rpm. Performed in

triplicate. Quantification of product by HPLC at 254 nm (for compound 56) or 210 nm (for compounds 55 and 82).

The spectrophotometric assay described in Sections 2.3 and 3.1, was then used with the selected five AKRs with a view to expand the substrates screened (Figure 58). Initially it was used with benzaldehyde **45** and reaction conversions were comparable to that using HPLC with NADH or NADPH as the cofactor, validating the use of the assay.

For the other AKRs, negligible conversions were observed with NADH as the cofactor. With NADPH, AKR-3 showed activity towards 2-CBT **16** and benzaldehyde **45**. This is a particularly interesting result, as AKR-3 displayed the highest sequence homology to MecgoR, whose natural substrate was 2-CBT. AKRs-12 and AKR-13 (and to a lesser extent AKR-3) showed activity towards cyclohexanone **19** with NADPH.

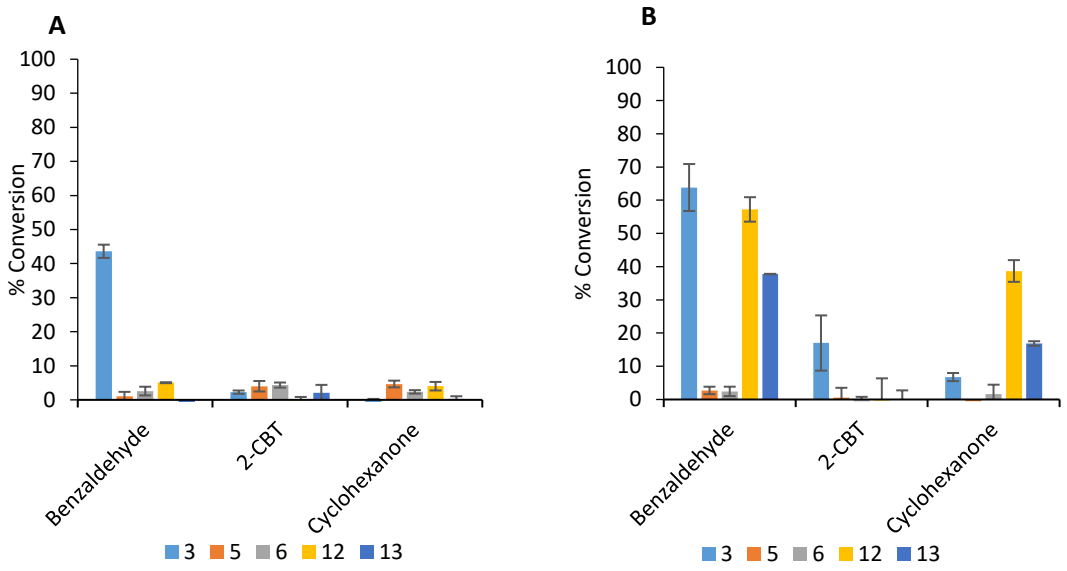


Figure 58: The % NAD(P)H conversion with a selection of AKRs from the drain metagenome (A) NADH and (B) NADPH (200 µL): substrate (5 mM), purified enzyme (0.4 mg/mL), NAD(P)H (1 mM), Pi (100 mM, pH 7.2), DMSO (10%, v/v). The reactions were shaken for 100 min at 25 °C. Reactions were performed in triplicate and quantified by spectrophotometric assay. The substrates used are depicted in Figure 59.

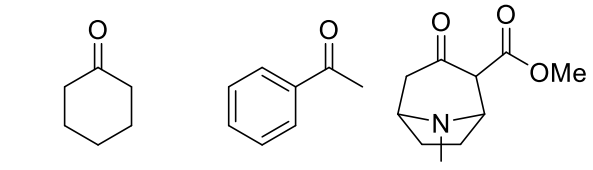


Figure 59: Substrates used in Figure 58.

4.6 Substrate specificities

Since AKR-3 showed some activity towards 2-CBT using NADPH as the cofactor, further substrate specificity assays were carried out in order to determine if the enzyme showed similar substrates tolerances to MecgoR (Table 23). AKR-3 was therefore tested for activity with the same substrates was tested with MecgoR. This covered a range of aliphatic and aromatic cyclic and linear small molecules, as well as a selection of tropinone analogues. Activity was demonstrated towards with 2-indanone **53**, benzaldehyde **45** and phenylacetaldehyde **46**. This is clearly different to the activities displayed by MecgoR, which was active towards a range of monocyclic aliphatic and aromatic rings. Thus, searching the metagenomes using sequence identities did not yield enzymes with similar activities. Reinhardt *et al.* discussed the functional activity of SDRs with >50% sequence identity to proven TRs from Solanaceae. Despite selection based on Hidden Markov models, it was also concluded that they lacked the functionality of TRs and neither reduced tropinone, nor nitrogen-containing tropinone analgoues.[66]

The BLAST NCBI database was queried with the residue sequence of ARK-3 and the closest match is oxidoreductase [*Pelagibacterium* sp. SCN 63-23] (Accession no ODT70425.1) with 68% sequence identity. This was recorded in the metagenomic samples of two bioreactors which were used to study cyanide and thiocyanate contaminated waste water.[176]

Substrate		% NADPH conversion	Substrate		% NADPH conversion
29	Acetophenone	2	**44**	4'-*N*-Propyl-2,2,2-trifluoroacetophenone	5
45	Benzaldehyde	45	**50**	1-Phenyl propan-2-one	4
19	Cyclohexanone	5	**46**	Phenylacetaldehyde	60
65	Cycloheptanone	0	**47**	2-Heptanone	3
38	2-Methylcyclohexanone	0	**48**	Methylacetoacetate	7
39	3-Methylcyclohexanone	2	**35**	Pyruvic acid	0
40	2-Cyclohexen-1-one	2	**71**	Camphor	0
52	1-Indanone	1	**68**	Progesterone	0
53	2-Indanone	20	**7**	Tropinone	0
54	Wieland–Miescher ketone	0	**86**	Psuedopellitierne	0
21	1-Methyl-4-piperidone	0	**18**	Nortropinone	0
37	Ethyl-4-oxo-1-piperidinecarboxylate	3	**99**	Benzyl nortropinone	3
36	1-(2-Methyl propyl)-4-piperidone	2	**95**	IDABO	0

41	2,2,2-Trifluoroacetophenone	3	**96**	Methyl 3-oxo-8-azabicyclo[3.2.1]octane-8-carboxylate	2
42	2,2,2,3'-Tetrafluoroacetophenone	4	**97**	Ethyl 3-oxo-8-azabicyclo[3.2.1]octane-8-carboxylate	1
43	2,2,2,4'-Tetrafluoroacetophenone	3	**98**	Phenyl 3-oxo-8-azabicyclo[3.2.1]octane-8-carboxylate	1

Table 23: The % NADPH conversion with AKR-3 from the drain metagenome (200 µL): substrate (5 mM), purified enzyme (0.4 mg/mL), NADPH (1 mM), Pi (100 mM), DMSO (10%, v/v). Reactions were shaken for 100 min at 25 °C, performed in triplicate (standard deviation <10%), and NADPH depletion was quantified by spectrophotometric assay at 340 nm. Substrates are depicted in Figure 60.

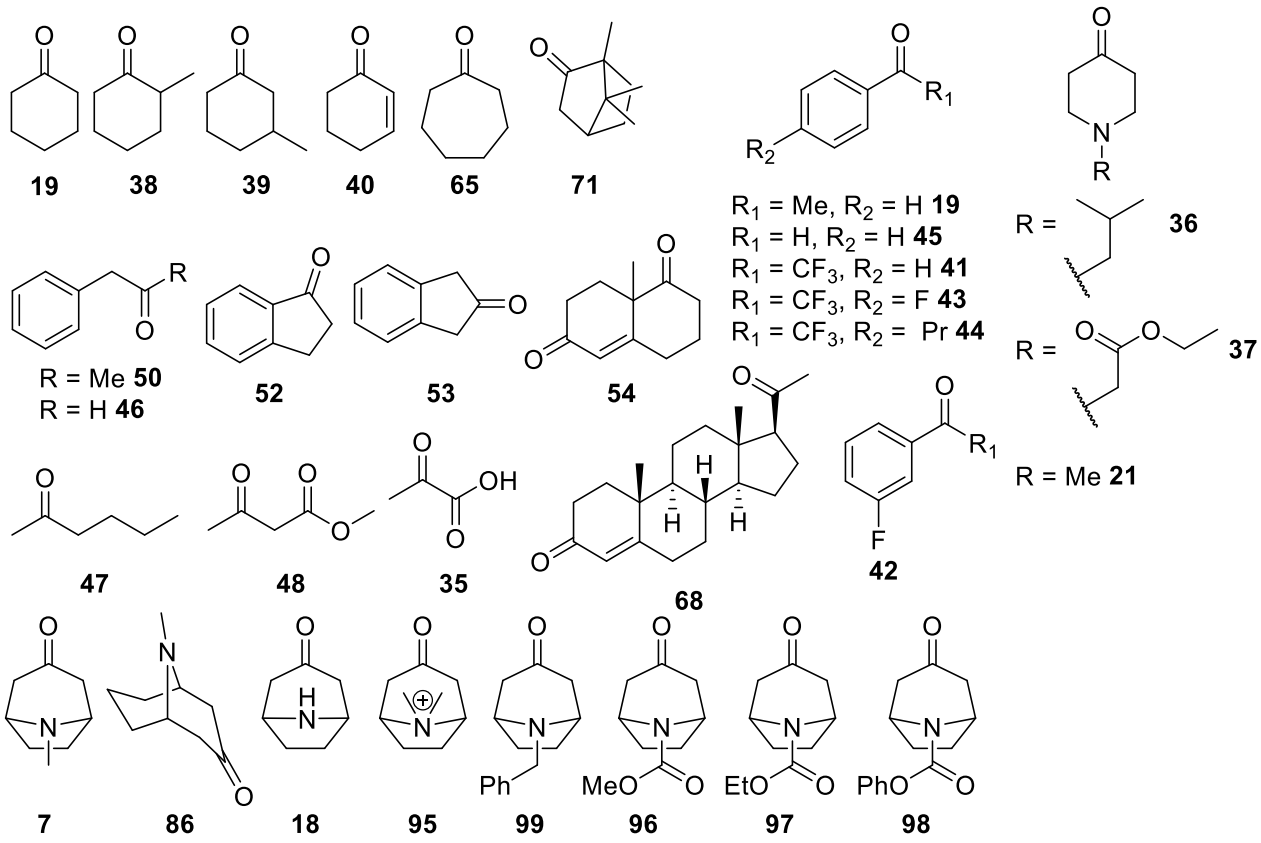


Figure 60: Substrate used in Table 23.

4.7 Conclusions and future work

Several metagenomic libraries have been searched for genes showing sequence identities to MecgoR. A total of 16 genes showing 24% or greater sequence identity were selected, all of which were annotated as AKRs and found in the drain metagenome. Of these, 12 were successfully cloned and 5 expressed well. Initial substrate tolerances were identified for these five AKRs. AKR-3 was assayed against

a panel of substrates, and the enzyme showed activities towards the aromatic substrates benzaldehyde, 2-indanone and phenylacetaldehyde. In comparison with work in Section 2, in which enzymes were selected according to PFam ID, this selection method yielded fewer active enzymes.

In future work, the reaction conditions when using AKR-3 will be optimised, and further work performed will establish trends in the substrate activities. For example, this preliminary work showed that AKR-3 tolerated some aromatic substrates, and this could be further examined. Additionally, the metagenomes could be further searched for more enzymes to identify tropinone or 2-CBT reduction functionality.

5 Transaminases

5.1 Introduction

Though the reduction of tropinone is a key step in biosynthetic pathways, the transamination reaction has not been observed in nature. The transamination of tropinone has been reported by Weiβ *et al.* as discussed in Section 1.6.4 and in parallel we have been working on the identification of transaminases that are able to accept tropinone.[112] UCL's toolbox of transaminases contains a selection of both (*R*)- and (*S*)-selective transaminases, which have been given names for ease of interpretation and are summarised in Table 24.

TAm	pQR Number	Microorganism	Mr
As-TAm	AS	*Arthrobacter sp.*	36.4
Bs-TAm-1	906	*Bacillus subtillis*	40.1
Bs-TAm-2	961	*Bacillus subtillis*	49.9
Bs-TAm-3	960	*Bacillus subtillis*	49.9
Cv-TAm	801	*Chromobacterium violaceum*	51.2
Dg-TAm-1	980	*Deinococcus geothermalis*	45.2
Ec-TAm	907	*Escherichia coli*	40.1
Kp-TAm-1	904	*Klebsiella pneumoniae*	43.4
Kp-TAm-2	1005	*Klebsiella pneumoniae*	45.5
Kp-TAm-3	1006	*Klebsiella pneumoniae*	47.1
Mv-TAm	Mb	*Mycobacterium vanbaalenii*	37.0
Pa-TAm-1	902	*Pseudomonas aeruginosa*	43.3
Pa-TAm-2	426	*Pseudomonas aeruginosa*	48.4
Pp-TAm-1	427	*Pseudomonas putida*	48.5
Pp-TAm-2	958	*Pseudomonas putida*	52.2
Pp-TAm-3	959	*Pseudomonas putida*	49.6
Rr-TAm	1017	*Rhodospirillum rubrum*	50.4
Rs-TAm	1019	*Rhodobacter sphaeroides*	50.1
Sa-TAm-1	806	*Streptomyces avermitilis*	48.2
Sa-TAm-2	807	*Streptomyces avermitilis*	50.8
Se-TAm-1	803	*Saccharopolyspora erythraea*	45.5
Se-TAm-2	804	*Saccharopolyspora erythraea*	50.6
Se-TAm-3	805	*Saccharopolyspora erythraea*	50.4
Vf-TAm	1003	*Vibrio fluvialis*	50.1

Table 24: Summary of the TAms from the UCL library. Molecular weights calculated using ProtParam online tool.[124]

The sequences of the TAms used were compared using the Clustal Omega online tool (Figure 61).[17] Cv-TAm, Pp-TAm-3 and Rr-TAm displayed high sequences homology of greater than 50% to each other, as did Sa-TAm-2, Se-TAm-2 and Se-TAm-3. As-TAm and Mv-TAm were both (*R*)-selective and display high sequence homology.

Several of these TAms are well known in the literature, as described in Section 1.6. Cv-TAm is from *Chromobacterium violaceum* DSM30191 and Vf-TAm is from *Vibrio fluvialis*. Both have been reported to show activities towards a range of cyclic compounds.[9,10,177,178] Pp-TAm from *Pseudomonas putida* PP_3718 has also been shown to tolerate dopamine as a substrate.[108] Two (*R*)-selective TAms were used investigated: As-TAm from *Arthrobacter* sp. has undergone 11 rounds of mutagenesis and showed activity with tetralones[3,107] and Mv-TAm from *Mycobacterium vanbaalenii* has largely been used with acyclic compounds.[179]

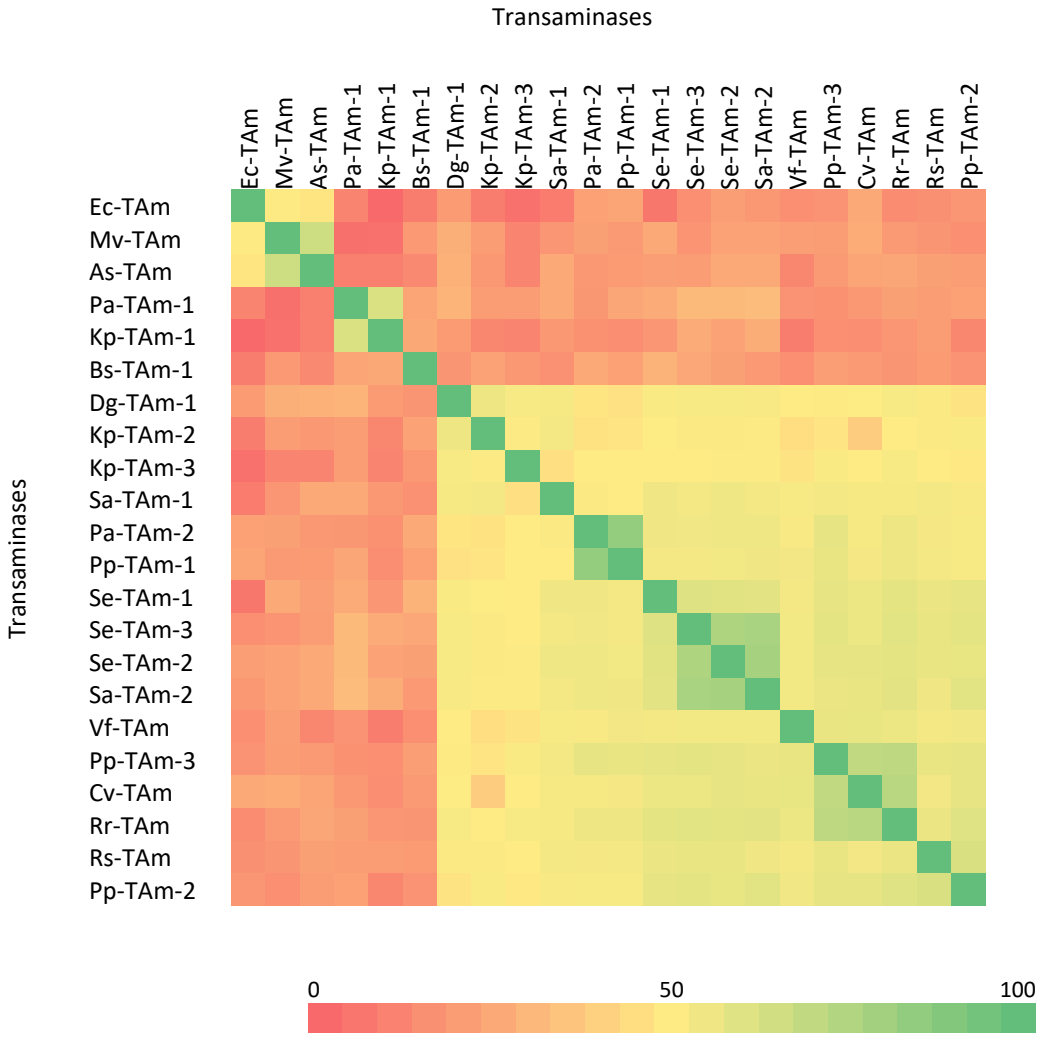


Figure 61: Heatmap of sequence identities (%) of the TAms from the UCL library (Table 24) as calculated by Clustal Omega online tool.[17] The colour key is displayed showing overall sequence identity.

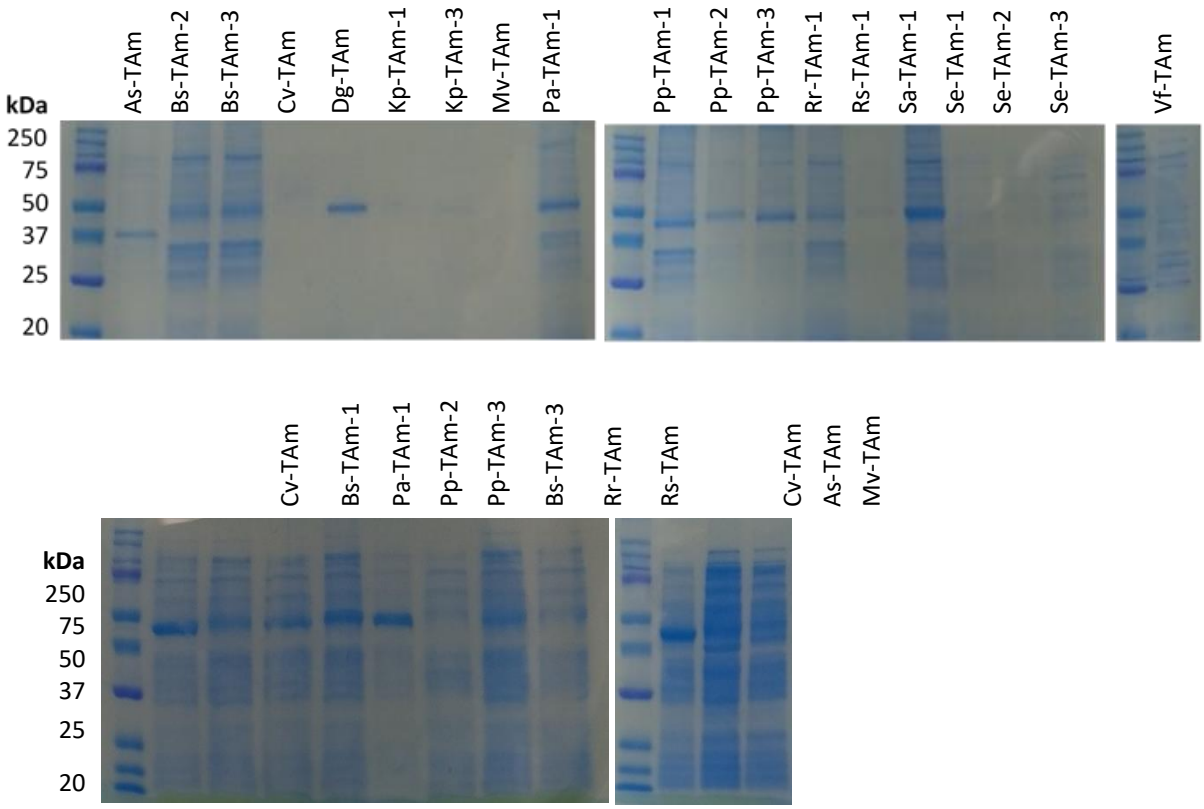
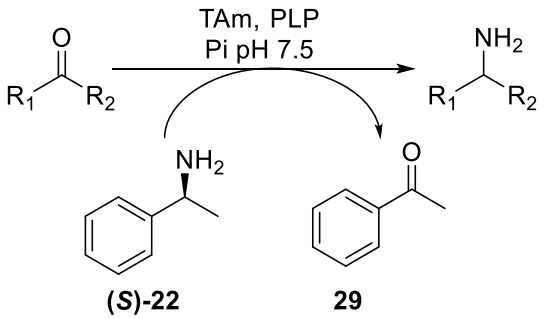


Figure 62: SDS-PAGE gels of the clarified cell lysates of the UCL TAm library. Molecular weights displayed in Table 24.

5.2 Transaminases with tropinone and similar compounds

Eight TAms were screened against the tropinone and analogues (see Section 3.6.1 for synthesis) using (*S*)-MBA **(*S*)-22** as the amine donor and monitoring acetophenone **29** co-product formation by HPLC analysis (Scheme 46 and Table 25). Benzaldehyde was used a standard substrate to confirm that the assay was performed correctly. The assays were completed at a 200 µL scale in parallel using 96-well plates.



Scheme 46: Transaminases assay using (*S*)-MBA (*S*)-22 as the amine donor. The production of acetophenone 29 was monitored using HPLC and used to calculate the conversion (%) of the substrate. (*R*)-MBA was also used for where appropriate with R-selective TAms.

None of the TAms showed high activity towards the compounds tested. Cv-TAm showed a low activity (less than 10%) towards several tropinones such as **96** and **97**. This data is useful information for future mutagenesis studies to enhance TAm activities. For example, Wieβ *et al.* described using mutagenesis to increase a small activity (specific activity of 70 mU mg^{-1}) with **98** to great effect (Section 1.6.4).[112]

	Substrate									
TAm	**45**	**35**	**7**	**18**	**95**	**99**	**96**	**97**	**98**	**16**
H_2O	4	1	0	1	3	2	2	2	3	3
Cv-TAm	47	38	4	0	3	3	7	5	0	5
Pp-TAm-2	27	31	3	0	2	0	0	0	0	0
Pp-TAm-3	23	29	0	0	0	0	0	0	0	0
Bs-TAm-3	14	24	0	0	0	0	0	0	0	0
Rr-TAm	12	21	0	0	0	0	0	0	0	0
Rs-TAm	12	37	0	0	0	0	0	0	0	0
Pa-TAm-1	35	33	0	0	0	0	0	0	0	0
Bs-TAm-1	26	11	4	0	2	2	1	1	0	0

Table 25: MBA assay with selected TAms and tropinones as substrates (200 µL): (*S*)- or (*R*)-MBA (25 mM), substrate (10 mM), PLP (10 mM), phosphate buffer (100 mM, pH 8), clarified cell lysate (0.2-0.5 mg/mL) and DMSO (10% v/v). Shaken for 24 hours, 30 °C, 500 rpm. Quantification was determined by HPLC and reactions were performed in duplicate. Substrates used are depicted in Figure 63.

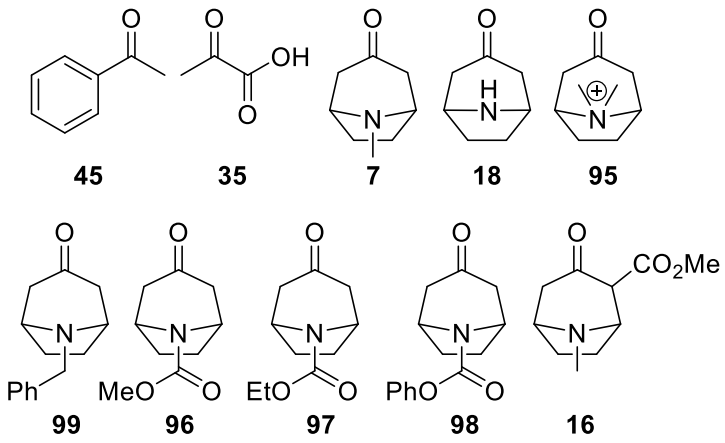


Figure 63: Substrates used in Table 25 and Figure 64.

The enzymes and substrates which demonstrated small activities in the MBA assay were selected and the experiment were repeated using an alternative colourimetric assay using 2-(4-nitrophenol)ethan-1-amine **30** as the amine donor (Table 25, described in Section 1.6.5.2).[89] This is a sensitive assay which has been reported to show conversions as low as 1.5% of benzaldehyde using Cv-TAm, and

could support the earlier findings.[89] Only low level conversions were observed in a couple of cases, such as **98** and **16** with Rs-TAm, Bs-TAm-1 and **97** with Bs-TAm-1. The activity of Pp-TAm-3 towards **7**, **98** and **99** was discounted, as activity was also demonstrated with the negative control H_2O.

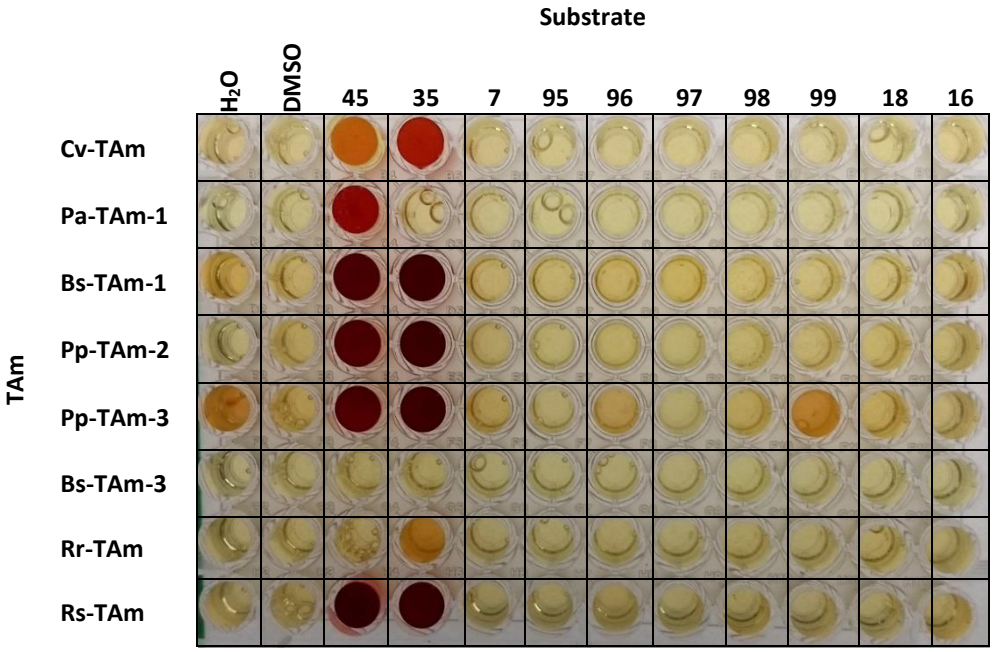


Figure 64: Colourimetric assay of selected TAms with the tropinone based substrates (200 µL): substrate (10 mM), PLP (0.2 mM), phosphate buffer (100 mM, ph 8), (2-(4-nitrophenol)ethan-1-amine (25 mM) and TAm clarified cell lysate (0.3-0.4 mg/mL). Shake 24 h, 30 °C, 500 rpm. Darker red colours show higher activity. The reaction was performed twice (duplicate not shown) and the substrate structures are depicted in Figure 63.

5.3 Transaminases with mono-cyclic compounds

Little activity was found with TAms towards the tropinone analogues, so it was decided to investigate TAm reactivity towards related mono-cyclic structures. Several heterocycles were therefore tested with the TAms (Figure 65).

The colourimetric assay (Section 1.6.5.2) was used with the substrates as this was a quick, qualitative assay to identify activities. 1-benzyl-3-piperidone **100**, which was used later, was not used as this decomposed over time to develop a dark brown colour.

Cv-TAm displayed activity towards all of the compounds. Bs-TAm-1, Pp-TAm-2, Pp-TAm-3, Rr-TAm, and Rs-TAm were active toward several of the compounds. Compounds **100** was readily accepted by these enzymes and **37** was accepted by

these TAms, excluding Rr-TAm. Bs-TAm-1 and Rr-TAm showed activity towards four of the compounds. Pa-TAm-1, Bs-TAm-3 and As-TAm did not tolerate many/any of the compounds.

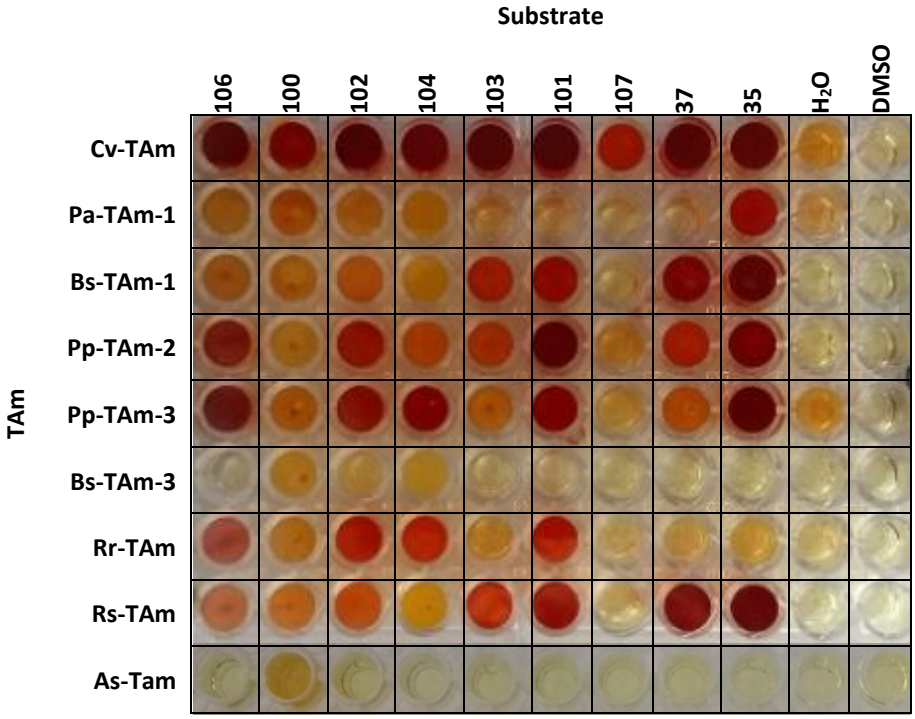


Figure 65: Colourimetric assay of selected TAms with piperidone substrates (200 µL): substrate (10 mM), PLP (0.2 mM), phosphate buffer (100 mM, pH 8), (2-(4-nitrophenol)ethan-1-amine (25 mM) and TAm clarified cell lysate (0.3-0.4 mg/mL). Shake 24 h, 30 °C, 500 rpm. Darker red colours show higher activity. The reaction was performed twice (duplicate not shown). The substrate used are depicted in (Figure 66).

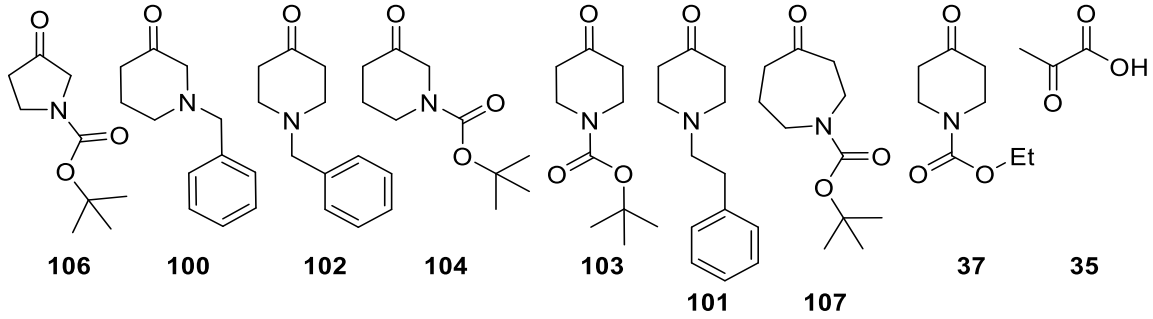


Figure 66: Substrates used in Figure 65 and Table 26.

The MBA assay was then used in order to quantify these results, as many of the TAms displayed activities towards the tested substrates (Table 26). Enzymes that displayed particularly high activity towards piperidones included Cv-TAm, Bs-TAm-1, Pp-TAm-2, Pp-TAm-3, Bs TAm-3, Rr-TAm, Rs-TAm and As-TAm. Activities were generally similar for each of these enzymes with the tested piperidones and ranged between

approximately 20-60% conversion. Other TAms displayed poor conversions. Cv-TAm was the most active of the enzyme with conversions of between 40-70% towards the largest number of substrates tested.

In general, the amine donor 2-(4-nitrophenol)ethan-1-amine **30** was less readily accepted than (*S*)-MBA as an amine donor. For example, Pa-TAm-1 showed low activities towards all the substrates except for pyruvic acid. However, Cv-TAm accepted both the amine donor and all the ketone donors well. Rs-TAm also displayed low activities.

1-Benzyl-4-piperidone **102** was accepted by Bs-TAm-1, Cv-TAm, Pp-TAm-2, Pp-TAm-3, Bs-TAm-3, Rr-TAm, Rs-TAm, and As-TAm, while 1-(-phenylethyl)-4-piperidone **101**, which is a larger structure, was also accepted by the same enzymes, usually with similar conversions. In the case of Pp-TAm-2, yield was greatly increased from 11 to 43%. This may be due to the more flexible n-alkyl chain. Ethyl-4-oxo-1-piperidoncaboxylate **37** was accepted by Bs-TAm-1, Cv-Tam, Pp-TAm-2, Pp-TAm-3, Bs-TAm-3, and As-TAm. 1-Boc-4-piperidone **103** was accepted by these enzymes as well as Rr-TAm. Cv-TAm displayed the highest activity with both these substrates.

With many transaminases showing activities with piperidones, activities with a smaller 5-membered ring (pyrrolidinones) and asymmetric piperidones were also demonstrated (Table 26). Being asymmetric, the product of these reactions would be chiral. 1-Benzyl-3-piperidone **100** was not accepted by any of the TAms. 1-benzyl-4-piperidone **102** was accepted by several TAms, suggesting a preference for more linear shaped compounds. A similar trend is found for 1-boc-3-piperidone **104** compared to 1-boc-4-piperidone **105** with enzymes Bs-TAm-1, As-TAm, Rr-TAm. Cv-Tam, Pp-TAm-2, Pp-TAm-3, and Bs-TAm-3 however had similar activities with both compounds. Though Cv-TAm and Pp-TAm-3 share high sequence identities (Figure 61), they do not show high identities with Pp-TAm-2 or Bs-TAm-3, so a correlation could not be identified. 1-Benzyl-3-pyrrolidone **105** was not accepted by any of the enzymes. The less bulky 1-boc-3-pyrrolidinone **106** was more readily tolerated, again by Bs-TAm-1, Cv-TAm, Pp-TAm-2, Pp-TAm-3 and Kp-TAm-3. Interestingly, *n*-boc-hexahydro-1H-azepin-4-one **107** which contains a 7-membered *n*-heterocycle, was

accepted by the same enzymes but with higher conversions. This suggests a preference towards larger cyclic structures.

	Substrate								
TAm	**102**	**101**	**37**	**105**	**100**	**104**	**106**	**107**	**108**
Pa-TAm-1	3	6	4	2	4	2	0	0	0
Kp-TAm-1	0	1	1	2	0	1	0	0	0
Bs-TAm-1	24	60	62	20	0	5	8	19	8
Ec-TAm	0	0	1	2	0	2	0	0	0
Bs-TAm-2	2	2	1	1	0	1	-2	0	1
Pa-TAm-2	0	1	2	0	2	0	0	0	2
Pp-TAm-1	0	0	1	4	3	2	0	0	0
Cv-TAm	55	54	65	67	2	45	2	41	61
Se-TAm-2	1	2	3	0	0	0	1	0	0
Se-TAm-3	0	1	0	0	0	0	1	0	0
Sa-TAm-1	0	0	1	0	0	0	0	0	1
Ta-TAm	2	3	1	6	2	2	2	1	2
Pp-TAm-2	11	42	26	26	2	19	2	8	37
Pp-TAm-3	15	11	12	25	0	14	1	11	19
Bs-TAm-3	20	25	35	27	0	20	0	13	29
Dg-TAm	0	1	3	0	0	0	0	0	1
Kp-TAm-2	0	0	1	0	1	1	0	0	0
Rr-TAm	10	11	3	2	0	8	0	5	1
Rs-TAm	13	22	7	17	0	7	6	5	8
Se-TAm-1	0	1	3	0	3	0	0	0	0
Sa-TAm-2	1	2	1	0	0	1	1	0	0
Vf-TAm	0	0	2	0	0	0	0	0	0
Kp-TAm-3	0	2	5	8	0	1	0	0	0
As-TAm	21	18	33	20	0	3	3	11	22
Mv-TAm	0	0	1	0	4	6	3	1	2

Table 26: Table of 25 TAms and their activities towards-selected piperidones (200 µL): (*S*)- or (*R*)-MBA (25 mM), substrate (10 mM), PLP (10 mM), phosphate buffer (100 mM, pH 8), clarified cell lysate (0.2-0.5 mg/mL) and DMSO (10% v/v). Shaken for 24 hours, 30 °C, 500 rpm. Quantification was determined by HPLC and repeated in duplicate. Darker green colours denote higher activity. Quantification was determined by monitoring acetophenone production using HPLC analysis (Scheme 11). All reactions were performed in triplicate (standard deviation <10%). The structures of the substrates are depicted in Figure 66.

5.4 Modelling to explain experimental results

In silico analysis was used as an initial guide to rationalise the results in a qualitative way. Docking was performed using AutoDock Vina for energy-minimised substrates.

The piperidone ethyl 4-oxopiperidine-1-carboxylate **37** was accepted readily by Cv-TAm. When docked into CV-TAm, the ketone group is orientated toward the PMP, with the *N*-carboethoxy group fitting in the active-site tunnel (Figure 67). Furthermore, the electronegative nitrogen and oxygen were well aligned with the polar and positively charged residues (Arg416), highlighted in green and blue respectively in the figure.

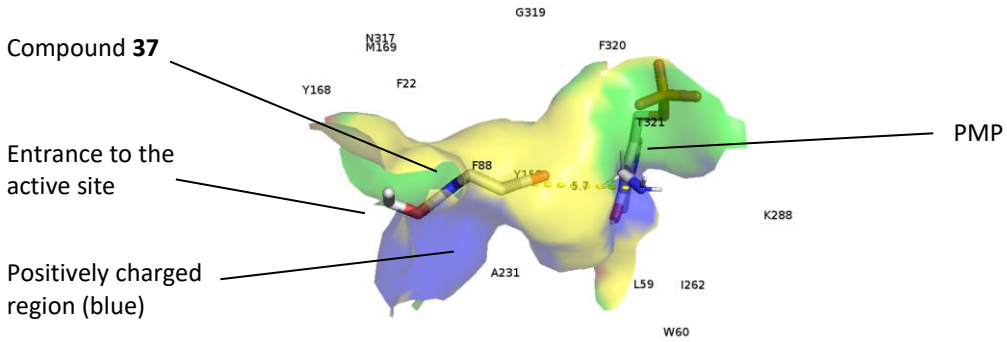


Figure 67: Image of active site tunnel of Cv-TAm with ethyl 4-oxopiperidine-1-carboxylate 37. PMP (labelled) embedded in the enzyme at the end of the tunnel. The residues are coloured depending on their physiochemical properties: blue = positive; red = negative; green = polar; yellow = non-polar.

Tropinone is a larger bicyclic ring system and the docking suggested at it docks at a different location in the active site tunnel in Cv-TAm (Figure 68). The docking study highlighted that Cv-TAm has the potential to accept tropinone, which sits in a productive orientation with the ketone group pointing towards the PMP. This however was not the experimental findings: several 5- and 6- membered heterocycles were accepted and tropinone **7** and pseudopelletierine **86** were not. This may suggest that the larger substrates cannot move easily into the active site entrance, as this study suggested if they were in the active site they would be transaminated. Leaving the active site may also be a potential problem and the polarity of the residues, such as Tyr163 or Tyr168, in the active site may also have an effect.

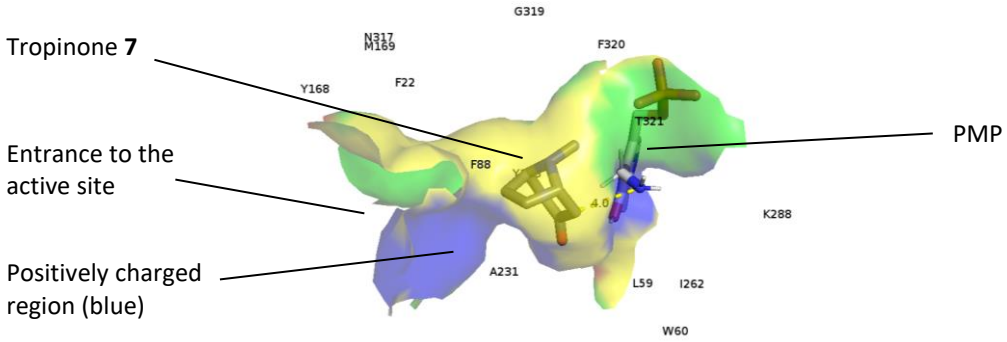
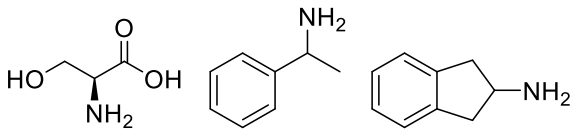


Figure 68: Image of active site tunnel of Cv-TAm tropinone 7. PMP (labelled) embedded in the enzyme at the end of the tunnel. The residues are coloured depending on their physiochemical properties: blue = positive; red = negative; green = polar; yellow = non-polar.

Future work could investigate mutants of Cv-TAm which have a larger entrance to the active site, allowing the larger substrates to enter/exit. Examples of residues include: Phe22, Phe88, Tyr153, Tyr168, Arg416, which are close to the active site entrance and could allow access from larger substrates if they were substituted for less bulky amino acids. Phe22 however is known to cause enzyme precipitation. Deszcz *et al.* performed site-direct mutagenesis for some of these residues (Tyr153X and Phe88X), in order to enhance serine:pyrucate α-transaminase activity in Cv-Tam.[104] It was reported that residues Tyr153, Trp60 and Phe88 were important for substrate specificity, with altered activity to L-serine **108**, (*S*)-MBA **(*S*)-22** and rac-aminoindane **109** reported (Figure 69).



L-108 **(*S*)-22** **109**

Figure 69: Substrates reported by Deszcz *et al.*.[104]

5.5 Conclusions and Future work

This chapter described the substrate selectivities of 25 TAms in the UCL TAm library towards a selection of heterocycles. Eight TAms were tested with compounds containing the tropinone motif, and Cv-TAm was demonstrated to show very low activities towards this structural motif. This could represent a starting point for further engineering to increase activity towards such compounds.

Future work including mutagenesis would improve on the small activities seen with the TAms and tropinone and similar compounds. Locations for mutagenesis have been highlighted to open the active site further and site-directed mutagenesis could be employed to achieve this. Successful candidates could also be used along with other enzymes in cascades in the future.

6 Conclusions and future work

6.1 Overall conclusions

Both plants and metagenomics strategies are valuable sources of enzymes, which may be usefully employed for biocatalytic purposes. From the tongue metagenome, 37 enzymes annotated as short chain reductases/dehydrognases (SDRs) were retrieved from a metagenome. These displayed low sequence similarities, mostly between 0 and 40% fulfilling the aim of generating a wide variation in enzyme primary sequence. Eight of those enzymes, which displayed activities towards cyclohexanone, were selected and tested for activity towards a selection of 22 small molecules. This highlighted enzyme activities and generally a preference for fluorinated aromatic rings as well as piperidone structures was noted. However, two of the enzymes did not display any activity towards the substrates tested. This strategy of *in silico* searching and enzyme selection yielded a higher hit rate of successful reductases than traditional high-throughput screening strategies. Although the sequence of the enzymes selected had been logged on the NCBI database, functional activity had not been identified for them, so this approach also expanded the current knowledge of substrate selectivities of these enzymes.

SDR-17 and SDR-31 were active towards a bicyclic compound, the Wieland-Miescher ketone (WMK), the use of which is widely reported in the literature and whose motif is incorporated into several pharmaceutical products.[134] SDR-17 and SDR-31 readily accepted this substrate. To improve the efficiency of this reaction, these enzymes were successfully co-expressed with glucose-6-phospte dehydrogenase (G6PDH), an NADPH recycling system. SDR-17 and G6PDH converted WMK to the (*S*)-alcohol with 100% conversion in small scale reactions, and this was scaled up to a preparative scale. (*R*)-WMK was employed at a 20 mM concentration as the substrate to make the product: **(4a*R*,5*S*)-56** with excellent yields and selectivity in 89% isolated yield and >99% e.e.. This was a large improvement on literature biocatalytic methods in terms of the enantioselectivity and yield. Additionally, the product was extracted from the reaction mixture without the use of flash column chromatography, recrystallisation or further purification. Overall this approach identified an efficient route to **(4a*R*,5*S*)-56**.

Tropinone reductases I and II (TRI and TRII) from the plant species *D. stramonium* and MecgoR from *E. coca* were also investigated expanding on the substrate selectivities that had been reported in the literature.[61,68,69,157] The steroid and terpene compounds tested were not tolerated by the enzymes, but fluorinated aromatic rings and small cyclic aliphatic compounds were readily accepted. Additionally, it was demonstrated that the addition of a His-tag to the C-terminus of the sequence did not greatly affect activity for TRI and TRII. A selection of tropinone analogues were synthesized chemically, and were tested with the TRs and MecgoR. TRII showed limited conversions, but tolerated nortropinone and benzyl nortropinone. TRI and MecgoR however were tolerant to wide variation in the tropinone structure. This work has expanded on reported substrate specificities and enzymatic reductions using these compounds, many of which have not been reported in the literature to date. This is a novel way to access both isomers of the tropine analogue alcohol products. Larger 5 mL-scale reactions were also performed using MecgoR and 2-carbomethoxytropinone (2-CBT) or benzaldehyde. When scaled to a 50 mL scale reaction with benzaldehyde, a quantitive yield was achieved. Reported reactions using MecgoR have been at a 100 µL scale, and so this is the first use of the enzyme on a larger biocatalytic scale.[61]

The metagenome from the drain environment was searched for enzymes that shared high sequence similarity to MecgoR. Twelve of these enzymes were retrieved *via* PCR and cloned into *E. coli* expression systems. Of the five AKRS that were expressed and purified readily, three of them showed activity towards benzaldehyde, 2-CBT or cyclohexanone. AKR-3, a novel enzyme, was tested with a selection of aliphatic and aromatic substrates, and showed activity towards aromatic substrates 2-indanone, phenylacetaldehyde and benzaldehyde. All the retrieved genes were annotated as AKRs, the same annotation given to MecgoR, showing that sequence similarity may be used as a method to guide enzyme selection in terms of general functionality. However, substrate specificites were dissimilar to the targeted MecgoR.

Lastly, a selection of TAms from the library at UCL were tested for activity towards a selection of tropinone analogues. Small activities with **98** and **16** with Rs-

TAm, Bs-TAm-1 and **97** with Bs-TAm-1 were recorded and by modelling the compounds into the Cv-TAm crystal structed reported highlighted a starting point for future mutatgenesis. Cv-TAm, Bs-TAm-1, Pp-TAm-2, Pp-TAm-3, Rr-Tam and Rs-TAm all readily accepted several piperidone and pyrrolidine based compounds, demonstrated in a colourimetric assay and more quantitively by monitoring the co-product acetophenone produced.

Overall, enzymes from metagenomic sources have been applied to useful substrates that are industrially applicable, for example, with SDR-17 and WMK. This work on the TRs has expanded the current knowledge of these plant derived enzymes. Novel enzymes from the drain metagenome were selected due to their sequence identity with MecgoR, and the functionality of these enzymes has been compared. Initial steps have been made into TAms that can accept tropinone, and this will be used in future work.

6.2 Future work

The enzymes yielded from both the tongue and the drain metagenome provide excellent starting points for reductions with a selection of ketones. With further examination of substrates tolerances using this current work as a guide, these enzymes could in future be used for other applications.

SDR-17 was successfully used in an efficient system for the reduction of (*R*)-WMK. In future work, this reaction could be scaled up further and the conditions optimised. To make the reaction more industrially applicable, the temperature used, DMSO co-solvent concentration and reaction time could be lowered, and the effect on the yield and selectivity of the reaction assessed. The tolerance to higher substrate loading will also be assessed. SDR-17 accepted analogues of WMK such as the Hajos Parish ketone so this is an option for further investigation in the in the future, as well as use in a cascade reaction with other enzymes, such as ene-reductases. A one-pot process combining the synthesis of the WMK and its reduction could also be investigated and optimised, as SDR-17 only showed activity with (*R*)-WMK without reducing its synthetic precursors.

The TRs displayed interesting activities with bulky bicyclic tropinone analogues. The stereochemistry of the products has yet to be established, so this should be investigated. The reduction of many of the compounds used has not yet been reported in the literature, so these will be scaled up to access novel compounds. Additionally, even bulkier substrates which are pharmaceutically relevant will be tested with these enzymes (Figure 70). The structure of MecgoR could be studies *via* sequence homology modelling to guide mutagenesis work, or X-ray crystallography experiments to determine the structures.

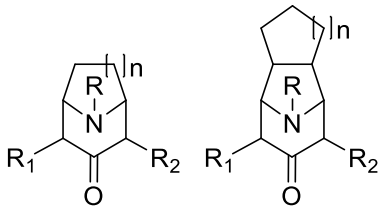


Figure 70:Tropinone analogues which could be synthesised and used in future work.

The drain metagenome should also be further used and searched to find enzymes with similar functionally to MecgoR. For example, structure prediction tools or more complex searches using predicted active sites could be used. The expression of the retrieved AKRs will be optimised and further investigated in terms of the substrate selectivities and optimum reaction conditions in order to find substrate preferences and stereoselectivities.

For the UCL library TAms, the activities towards piperidones, pyrrolidines and tropinone analogues have provided an initial guide for future mutagenesis. For Cv-TAm, modelling suggested that residues such as Phe22, Phe88, Tyr153, Tyr168, Arg416 which lie near the entrance to the active site would be some options for potential mutagenesis targets.

7 Methods and materials

7.1 General experimental detail

Oven (Memmet) dried glassware were used for reactions performed under anhydrous conditions. Flash chromatography was carried out using Silica gel (Geduran, 40-60 µm). Thin layer chromatography was carried out using aluminium backed silica gel plates (Merck Keiselgel) and were visualised using ultra-violet (UV) light or potassium permanganate solution containing $KMnO_4$ (3 g), K_2CO_3 (20 g), 5% NaOH aq. (5 mL) and H_2O (300 mL).

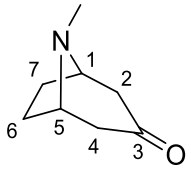
Infrared spectroscopy was carried out using a Spectrum 100 FTIR spectrometer (Perkin-Elmer) and IR peaks were reported in cm^{-1}. ^{1}H and ^{13}C NMR spectroscopy was carried out using AMX300 (Bruker), AMX400 (Bruker), Avance 500 (Bruker), Avance 600 (Bruker) instruments at the field indicated. Chemical shifts were recorded in ppm and referenced to residual pronated solvent. Coupling constants (*J*) were recorded in Hz and were quoted relative to tetramethylsilane (TMS). Mass spectroscopy was carried out using a VG70-SE mass spectrometer Trace 1310 Gas Chromatograph (Thermoscientific) connected to an ISQ single quadrapole MS (Thermoscientific) or a VG70-SE Mass Spectrometer (Waters). Melting points were determined using IA9000 Series melting point apparatus (Electrothermal) and were uncorrected. Optical rotations were measured at 25 °C unless otherwise stated, on a Perkin-Elmer Model 343 Polarimeter. Specific rotations are given in $10^{-1}°cm^2g^{-1}$. HPLC analysis was carried out using Series 1100 (Hewlett Packard) or 1260 (Aligent Technologies) instruments with an Ace 5 C_{18} column (HICHREM) (internal diameter 4.6 mm, length 150 mm) or Chiralcel OJ column (250 x 4.6 mm ID Analytical column, 10 µm beads). GC analysis was carried out with a Series 7820A (Agilent Technologies) instrument fitted with a flame ionisation detector (FID) and a Beta Dex 225 capillary column (30 m x 0.25 mm x 0.25 µm film thickness, Supelco).

Sonication of cell lysates was performed using a probe Sonifier 150 (Branson) for the TAms and Soniprep 150 sonicator (MSE) for the KREDs. The following centrifuges were used: Allegras x-15R centrifuge (Beckman Coulter), Centrifuge 5415R (EppendoRr), Centrifuge 5810R (Eppendorf) and Centrifuge 5430R (Eppendorf). A ShakerX ClimoShaker ISFI-X (Kuhner), Innova 44 (New Brunswick

Scientific) or Mixing Block MB-102 (BIOER) incubating shaker was used. A GENios plate reader spectrophotometer (Tecan) was used. Chemicals, media and apparatus were autoclaved (Priorclave) at 121 °C for 30 min when necessary. All chemicals were from commercial sources, unless otherwise stated.

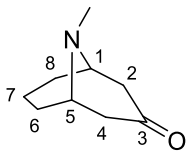
7.2 Compound synthesis

7.2.1 Tropinone **7**



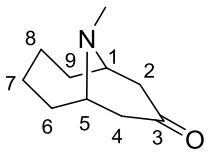
A solution of 2,5-dimethoxy tetrahydrofuran (1.30 mL, 11.6 mmol), concentrated HCl (10.4 M, 150 µL) and water (3.50 mL) was stirred together at 70 °C for 30 min. An acetate buffer solution was prepared at 0 °C, as follows: NaOH (1.80 g, 45.3 mmol), AcOH (2.60 mL, 45.3 mmol), 1,3-acetonedicaboxylic acid (1.87 g, 12.8 mmol), methylamine HCl salt (0.860 g, 12.8 mmol), water (8.50 mL). The acidic solution was added slowly to the buffer solution at 0 °C. The pH of the combined solution was maintained at pH 5, and stirred for 1 h at 40 °C. Once cooled to 25 °C, the mixture was added to a 50% NaOH solution (900 µL) and NaCl (3.00 g, 51.3 mmol) added at 0 °C. The mixture was stirred at 25 °C for 1 h, and then extracted with CH_2Cl_2 (3 x 6 mL). The aqueous layer was then extracted with CH_2Cl_2 (5 x 4 mL). The combined organic layers were dried (Na_2SO_4) and concentrated under reduced pressure to give a brown solid. The crude product was distilled under vacuum using Krugelrohr distillation apparatus (60-65 °C, 4 mmHg) to give **7** as colourless crystals (0.44 g, 28%).[163] Mp 40-42 °C, [lit.: 40-42 °C][163]; Rf = 0.35 (50% EtOAc/Pet. ether 40-60 °C); ν_{max} (neat): 2882 (C-H), 1701 (C=O) cm^{-1}; ^1H NMR (300 MHz, $CDCl_3$): δ 3.51 (1H, br s, 1-H and 5-H), 2.80-2.72 (2H, dd, *J* = 16.1, 2.6, 2-*H*H and 4-*H*H), 2.49 (3H, s, N-C*H*$_3$), 2.27 – 2.17 (2H, d, *J* = 16.1 2-H*H* and 4-H*H*), 2.17 – 2.11 (2H, m, 6-*H*H and 7-*H*H), 1. 67 – 1.60 (2H, m, 6-H*H* and 7-H*H*); ^{13}C NMR (300 MHz, $CDCl_3$): δ 60.9 (C-1 and C-5), 47.7 (C-2 and C-4), 38.5 (N-*C*H_3), 27.8 (C-6 and C-7); *m/z* (ES+) 140 ($[M+H]^+$).

7.2.2 Pseudopelletierine **86**



To a solution of glutaraldehyde (1.78 mL, 10.0 mmol) in H_2O (6.80 mL) a solution of NH_2CH_3.HCl (1.00 g, 14.8 mmol) and acetone-1,3-dicarboxylic acid (1.68 g, 11.5 mmol) in H_2O (16.4 mL) and a solution of Na_2HPO_4.12 H_2O (1.76 g, 4.92 mmol) and NaOH (0.16 g, 3.6 mmol) in H_2O (4.00 mL) were added. Carbon dioxide was evolved and the mixture was stirred for 24 h at 25 °C. Concentrated HCl (0.64 mL) was added and the solution stirred at 75 °C for 1 h to complete the decarboxylation. Once cooled to 25 °C, a solution of NaOH (1.55 g, 38.8 mmol) in H_2O (2 mL) was added and the basic mixture was extracted with EtOAc (4 x 10 mL), dried ($MgSO_4$), concentrated to approximately half volume, filtered through neutral alumina and eluted with EtOAc. This was concentrated under reduced pressure to give **86** as a brown solid (1.11 g, 72%).[164] Mp 54-58 °C, [lit.: 63-67 °C][164]; Rf = 0.29, (20% MeOH/CH_2Cl_2); ν_{max} (neat): 2918 (C-H), 2815 (C-H), 1694 (C=O) cm^{-1}; ^{1}H NMR (300 MHz, $CDCl_3$): δ 3.28 (2H, br s, 1-H and 5-H), 2.81-2.71 (2H, dd, *J* = 17.0, 6.7, 2-*H*H and 4-*H*H), 2.60 (3H, s, N-C*H*$_3$), 2.26 – 2.20 (2H, d, *J* = 17.0, 2-H*H* and 4-H*H*), 1.98 – 1.90 (3H, m, 6-*H*H, 7-*H*H, and 8-*H*H), 1.56 – 1.46 (3H, m, 6-H*H*, 7-H*H*, and 8-H*H*); ^{13}C NMR (300 MHz, $CDCl_3$): δ 206.7 (C-3) 77.1 (C-1 and C-5), 55.8 (C-2 and C-4), 41.5 (N-*C*H_3), 38.2 (C-6 and C-8), 16.0 (C-7); *m/z* (ES+) 154 ([M+H]$^+$); HRMS $C_9H_{16}NO$ calcd. 154.1232, found 154.1229.

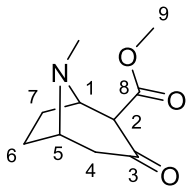
7.2.3 10-Methyl-10-azabicyclo[4.3.1]decan-8-one (MADO) **87**



To a vigorously stirred solution of silica gel (10.5 g) in CH_2Cl_2 (50 mL), $NaIO_4$ (1.47 g, 650 mM, 6.82 mmol) in H_2O (10.5 mL) was added dropwise. *Trans*-cyclohexandial (0.610 g, 5.23 mmol) in CH_2Cl_2 (20 mL) was added and the reaction was stirred for 24 h at 25 °C This was then filtered (sintered funnel) and washed with CH_2Cl_2.

To a solution of 1,6-hexandial (0.479 g, 3.00 mmol) in DMSO (1.5 mL), a solution of NH_2CH_3.HCl (0.305 g, 4.50 mmol) and acetone-1,3-dicarboxylic acid (0.505 g, 3.45 mmol) in H_2O (8 mL) and a solution of Na_2HPO_4.12 H_2O (0.875 g, 2.46 mmol) and NaOH (0.248 g, 6.20 mmol) in H_2O (4.00 mL) were added and the mixture was stirred for 24 h at 25 °C. Concentrated HCl (0.32 mL) was added and the solution stirred at 75 °C for 1 h to complete decarboxylation. Once cooled to 25 °C, the mixture was extracted with EtOAc (4 x 10 mL), dried ($MgSO_4$), concentrated to approximately half volume, filtered through neutral alumina and eluted with EtOAc. This was concentrated under reduced pressure to give **87** as an orange oil (0.471 g, 94%).[164,180] Rf = 0.65 (10:30:60 $^{i}PrNH_2$:EtOAc:Pet. ether); ν_{max} (neat): 2924 (C-H), 1697 (C=O) cm^{-1}; ^{1}H NMR (300 MHz, $CDCl_3$): δ 3.44 (2H, q, *J* = 6.1 Hz, 1-*H*H and 5-*H*H), 2.68-2.74 (2H, m, 2-*H*H and 4-*H*H), 2.64 (3H, s, N-*C*H_3), 2.16 – 2.19 (2H, m, 2-H*H* and 4-H*H*), 1.92-2.00 (2H, m, 7-*H*H and 8-*H*H), 1.63-1.71 (2H, m, 7-H*H* and 8-H*H*), 1.42-1.51 (4H, m, 6-H and 9-H); ^{13}C NMR (300 MHz, $CDCl_3$): δ 210.9 (C-3), 59.7 (C-1 and C-5), 43.6 (C-2 and C-4), 42.5 (N-*C*H_3), 31.2 (C-7 and C-8), 22.6 (C-6 and C-9); ν_{max} (neat): 2925 (C-H), 1700 (C=O) cm^{-1}; *m/z* (EI) 167 ($[M]^+$). HRMS (EI) *m/z*: $[M]^+$ Calcd for $[C_{10}H_{17}NO]^+$ 167.1305; Found 167.1305.

7.2.4 (±)-2-CBT 16

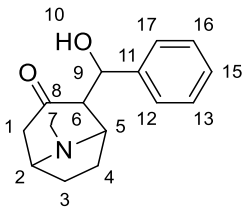


Tropinone (0.251 g, 1.80 mmol) in cyclohexane (1.70 mL) was added dropwise to a mixture of NaH (60% dispersion in mineral oil; 0.143 g, 3.11 mmol), dimethyl carbonate (0.330 mL, 3.68 mmol) and cyclohexane (0.73 mL) and heated at 84 °C. MeOH (625 µL) was then added and the reaction mixture was heated at reflux until effervescence stopped. The mixture was cooled to 25 °C, and water (300 µL) added. The layers were separated and the cyclohexane layer extracted with water (3 x 5 mL). The combined aqueous layers were saturated with NH_4Cl, extracted with CH_2Cl_2 (8 x 5 mL), dried ($MgSO_4$) and concentrated under reduced pressure. The crude product was purified using flash chromatography (10:30:59:1 $^{i}PrNH_2$:EtOAc:hexane:Et_3N) to

16 (0.079 g, 0.40 mmol, 22%)[166]; as a fine yellow powder: Mp 97-100 °C, [lit.: 104-106 °C][164]; Rf = 0.52 (10:30:59:1 ^{i}PrNH$_2$:EtOAc:hexane:Et$_3$N); ν_{max} (neat): 2949 (C-H), 2749 (C-H), 1665 (CO-O-H) cm^{-1}; *m/z* (ES+) 198 ([M+H]$^+$); HRMS $C_{10}H_{16}NO_3$ calcd. 198.1130, found 197.1137.

Compound exists as 3 isomers: enol tautomer (I), (*S*)-(-)-2CBT (II), and (*R*)-(+)-2CBT (III), in the ratio 1:1:1.3 in DMSO-D$_6$. Protons labelled "E" or "A" to indicated either equatorial or axial positions on the ring. NMR assigned by Dr Abil Aliev. **(I)**^{1}H NMR (600 MHz; DMSO-*d*$_6$): δ 11.60 (1H, s, 3-O*H*), 3.70 (3H, s, 9-H), 3.61 (1H, s, 1-H), 3.23 (1H, m, 5-H), 2.64 (1H, M, 4E-H), 2.19 (3H, s, NC*H*$_3$), 2.03 (1H, m, 6E-H), 1.96 (1H, m, 7E-H), 1.85 (1H, d, *J* = 4.4, 4A-H), 1.65 (1H, m, *J* = ,7A-H), 1.45 (1H, m, 6A-H); ^{13}C NMR (600 MHz; DMSO-*d*$_6$): 169.5 (C-3), 101.70 (C-2), 170.9 (C-8), 56.7 (C-5), 56.7 (C-1), 51.5 (C-9), 35.9 (C-10), 34.0 (C-7), 33.8 (C-4), 28.9 (C-6). **(II)**^{1}H NMR (600 MHz; DMSO-*d*$_6$): δ 3.81 (1H, d, J = 1.0, 2-H), 3.61 (3H, s, OC*H*$_3$), 3.55 (1H, m, 1-H), 3.42 (1H, m, 5-H), 2.76 (1H, m, 4E-H), 2.45 (1H, s, NC*H*$_3$), 2.02 (1H, d, *J* = 4.4, 4A-H) , 1.97 (1H, m, 7A-H), 1.91 (1H, m, 7E-H), 1.91 (1H, m, 6E-H), 1.42 (1H, m, 6A-H); ^{13}C NMR (600 MHz; DMSO-*d*$_6$): 203.5 (C-3), 169.2 (C-8), 157.7 (C-2), 62.5 (C-1), 60.8 (C-5), 51.5 (C-9), 45.3 (C-4), 35.7 (C-10), 28.0 (C-6), 25.2 (C-7). **(III)**^{1}H NMR (600 MHz; DMSO-*d*$_6$): δ 3.66 (1H, d, *J* = 6.5, 1-H), 3.64 (3H, s, OC*H*$_3$), 3.35 (1H, m, 5-H), 3.20 (1H, t, *J* = 1.0, 2-H), 2.66 (1H, m, 4E-H), 2.23 (3H, s, NC*H*$_3$), 2.13 (1H, m, 4A-H), 2.11 (1H, m, 7E-H), 2.04 (1H, m, 6E-H), 1.45 (1H, m, 7A-H), 1.45 (1H, m, 6A-H); ^{13}C NMR (600 MHz; DMSO-*d*$_6$): 204.8 (C-3), 169.3 (C-8), 64.5 (C-1), 63.3 (C-2), 61.3 (C-5), 52.1 (C-9), 49.0 (C-4), 40.6 (C-10), 25.9 (C-6), 24.9 (C-7).

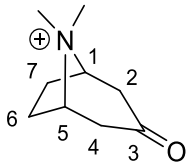
7.2.5 2-(Hydroxy(phenyl)methyl)-8-methyl-8-azabicyclo[3.2.1]octan-3-one 89



Diisopropylamine (310 µL, 2.20 mmol) was added to THF (6.4 mL), cooled to 0 °C and *n*-BuLi (2.3 M solution in hexane; 1.04 mL, 2.20 mmol) was added. The solution was stirred for 25 min at 0 °C, then cooled to -78 °C and tropinone (278 mg, 2.00 mmol) in THF (4 mL) was added dropwise. Phenylethanol (520 µL, 5.20 mmol)

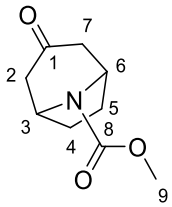
was then added and the solution stirred for 15 min at –78 °C before being quenched with sat. NH_4Cl solution. The mixture was then warmed to RT and extracted with CH_3Cl (3 x 20 mL). The combined organic layers were washed with water, dried ($NaSO_4$) and concentrated *in vacuo*. The crude product was then recrystallised in CH_3Cl and hexane. This yielded **89** as white needles (142 mg, 13%).[167] Mp 125-128 °C [lit. 132.7 °C][167]; Rf = 0.24 (20% EtOAc/Pet. ether 40 – 60 °C); ν_{max} (neat): 3200 (O-H), 2945 (C-H), 1714 (C=O) cm^{-1}; ^1H NMR (600 MHz, CD_3Cl): δ 7.21-7.33 (5H, m, 12-H to 17-H), 5.24 (1H, d, *J* = 3.1, 9-H), 3.61 (1H, d, *J* = 7.0, 5-H), 3.50 (1H, m, 2-H), 2.82-2.91 (1H, m, 1-*H*H), 2.48 (3H, s, 7-H), 2.42-2.45 (1H, m, 6-H), 2.30-2.37 (1H, m, 1-H*H*), 2.12-2.30 (2H, m, 4-H), 1.54 – 1.60 (2H, m, 3-H); ^{13}C NMR (150 MHz, CD_3Cl): 207.94 (C-8), 141.9 (C-11), 128.2 (C-13 & 1 C-6, 2C), 127.5 (C-15), 125.7 (C-12 & C-17), 76.7 (C-9), 67.6 (C-5), 62.0 (C-2), 51.6 (C-1), 40.8 (C-7), 26.6 (C-4), 26.4 (C-3); *m/z* (EI): 245 ([M]$^+$).

7.2.6 8,8- Dimethyl-3-oxo-8-azonia-bicyclo[3.2.1]octane iodide (IDABO) 95



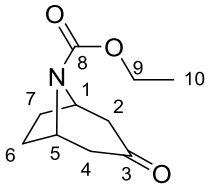
Tropinone (0.104 g, 0.748 mmol) was dissolved in acetone (750 µL) whilst stirring. Methyl iodide (50 µL, 0.82 mmol) was added dropwise and the reaction was stirred for 1 h at 25 °C. The resulting precipitate was filtered and washed with acetone, then a 1:1 pentane:EtOAc solution. This was then concentrated to dryness to give **95** as a white fine powder (0.112 g, 97%).[144,171] Mp 270-275 °C (decomp), [lit.: 273-274 °C (decomp)][181]; Rf = 0.33 (CH_2Cl_2); ν_{max} (neat): 2995 (C-H), 1720 (C=O) cm^{-1}; ^1H NMR (600 MHz; DMSO-d_6): δ 4.15 (2H, br s, 1-H and 5-H), 3.38 (3H, s, N-CH_3), 3.16 (3H, s, N-CH_3), 3.17-3.12 (2H, m, 2-*H*H and 4-*H*H), 2.53 (2H, s, 2-H*H* and 4-H*H*), 2.47-2.44 (2H, m, 6-*H*H and 7-*H*H), 1.95 (2H, d, *J* = 9.2, 6-H*H* and 7-H*H*); ^{13}C NMR (150 MHz; DMSO-d_6): 202.2 (C-3), 68.2 (C-1 and C-5), 50.4 (N-CH_3), 44.1 (C-2 and C-4), 26.0 (C-6 and C-7); *m/z* (EI) 154 ([M]$^+$).

7.2.7 Methyl 3-oxo-8-azabicyclo[3.2.1]octane-8-carboxylate 96



To a solution of nortopinone hydrochloride (0.288 g, 2.07 mmol) in CH_2Cl_2 (12.5 mL), Et_3N (1.5 mL, 10.68 mmol) was added and the mixture was cooled to 0 ˚C. DMAP (0.044 g, 0.356 mmol) was added and methyl chloroformate (0.1 mL, 1.19 mmol) was added dropwise. This was stirred at 25 °C for 16 h. The mixture was then diluted in EtOAc, the solids filtered off and the filtrate was washed with sat. NaCl solution. The aqueous phase was extracted with EtOAc (3 x 20 mL). The combined organic layer was washed with water (3 x 20 mL), dried ($MgSO_4$), and concentrated *in vacuo*. The crude product was filtered through a silica plug with 50% EtOAc/Pet. ether and concentrated *in vacuo* to give **96**, as a white solid (0.0731 g, 18 %)[173]. Mp = 60-61 °C, [Lit. 60-61 °C][182]; Rf = 0.49 (30% EtOAc/Pet. ether 40-60 °C); ν_{max} (neat): 2957 (sat. C-H), 1691 (C=O) cm^{-1}; ^1H NMR (600 MHz, $CDCl_3$, 60 °C): δ 4.55 (3-H & 6-H, br. s., 2H), 3.76 (3H, s, 9-H), 2.66 (2H, dd, *J* = 15.9 Hz, 3.39 2-*H*H & 7-*H*H), 2.30 – 2.38 (2H, m, 2-H*H* & 7-H*H*), 2.00-2.14 (2H, m, 4-*H*H & 5-*H*H), 1.62-1.75 (2H, m, 4-H*H* & 5-H*H*); ^{13}C NMR (150 MHz, $CDCl_3$): δ 208.2 (C-1), 154.3 (C-8), 52.8 (C-3 & C-6, 2C), 49.3 (9-C), 48.8 (C-9), 29.5 (C-4 & C-5, 2C), 28.7 (C-2 & C-7, 2C); *m/z* (ES+) 184 ($[M+H]^+$).

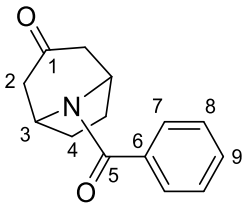
7.2.8 Ethyl 3-oxo-8-azabicyclo[3.2.1]octane-8-carboxylate 97



To tropinone (1.02 g, 7.32 mmol) and K_2CrO_3 (77 mg, 0.43 mmol) was added under argon, anhydrous toluene (9 mL) was added followed by ethylchloroformate (2.10 mL, 17.0 mmol) dropwise. The reaction heated to reflux for 18 h. The mixture was concentrated under vacuum and the brown solid was dissolved in CH_2Cl_2 (10 mL) and washed in H_2O (10 mL), dried ($MgSO_4$) then concentrated under reduced

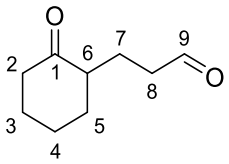
pressure to give **97** as a brown oil (1.12 g, 92%)[174]. The product was sufficiently pure for direct use. Rf = 0.53 (20% EtOAc/Pet. ether 40 – 60 °C); ν_{max} (neat): 2979 (C-H), 1703 (C=O) cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$, 60˚C): δ 4.52 (2H, br s, 1-H and 5-H), 4.17 (2H, q, *J* = 7.1, 9-H), 2.64 (2H, br s, 2-*H*H and 4-*H*H), 2.32 (2H, d, *J* = 15.8, 2-H*H* and 4-H*H*), 2.08 – 2.06 (2H, m**,** 6-*H*H and 7-*H*H), 1. 68 – 1.62 (2H, m, 6-H*H* and 7-H*H*), 1.26 (3H, t, *J* = 7.1, 10-H); ^{13}C NMR (100 MHz, $CDCl_3$, 60˚C): δ 207.5 (C-3), 153.9 (C-8), 61.32 (C-9), 53.2 (C-5 and C-1), 48.8 (C-2 and C-4), 29.0 (C-6 an C-7), 14.6 (C-10); *m/z* (ES+) 198 ($[M+H]^+$).

7.2.9 8-Benzoyl-8-azabicyclo[3.2.1]octan-3-one **98**



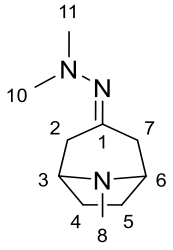
Nortropinone hydrochloride (0.280 g, 1.73 mmol) was dissolved in CH_2Cl_2 (12.5 mL) supplemented with Et_3N (2.5 mL) and was cooled to 0 °C. Benzoyl chloride (200 µL, 1.73 mmol) was added dropwise and the reaction warmed to 25 °C and stirred for 18 h. The solvent was removed *in vacuo* to leave a white powder which was dissolved in CH_2Cl_2 (10 mL) and washed with H_2O (2 x 10 mL). This was dried ($NaSO_4$) and concentrated *in vacuo*. The product was passed through a silica plug and washed with pet. ether, then 40% ethyl acetate/Pet. ether, and dried *in vacuo* to procure **98** as a white solid in quantitative yield. Mp 101.8-102.6 °C, [lit.: 95 °C][172,112]; Rf = 0.36 (40% EtOAc/Pet. ether 40 – 60 °C); ν_{max} (neat): 1715 (C=O), 2931.39 & 2892.57 (sat. C-H) cm^{-1}; ^{1}H NMR (600 MHz; $CDCl_3$): δ 7.52- 7.66 (2H, m, 7-H), 7.43 – 7.51 (3H, m, 8-H & 9-H), 5.08 (1H, br. s, 3-H), 4.41 (1H, br. s, 3-H), 2.96 (1H, br. s, 2-*H*H), 2.24-2.64 (4H, m, 2-H), 2.17 (2H, br. s., 4-*H*H), 1.78 (2H, d, *J* = 8.3, 4-H*H*); ^{13}C NMR (150 MHz; $CDCl_3$): δ 207.6 (C-1), 169.1 (C-5), 135.7 (C-6), 130.7 (C-9), 128.8 (2C, C8), 127.2 (2C, C-7), 56.1 (C-3), 51.6 (C-3), 49.7 (C-2), 49.0 (C-2), 29.7 (C-4), 28.2 (C-4); *m/z* (ESI) 230 ($[M]^+$).

7.2.10 3-(2-Oxocyclohexyl)propanal **51**



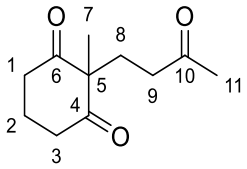
Cyclohexanone (1.10 mL, 10.13 mmol) was added to cyclohexane (30 mL) and then $MgSO_4$ (8 g) was added in one portion under argon at -10 °C. Pyrrolidine (420 µL, 51.1 mmol) was then added dropwise. The reaction was stirred vigorously for 30 min at -10 °C and then at 25 °C for 16 h. The $MgSO_4$ was removed by filtration and washed thoroughly with cyclohexanone. The combine filtrates and washes were concentrated under reduced pressure to yield 1-cyclohexenyl pyrrolidine. A portion of the filtrate oil (0.506 g, 3.31 mmol) was dissolved in dry THF (2.5 mL) under argon and cooled to -12 °C. With vigorous stirring, acrolein (220 µL, 3.32 mmol) in dry THF (0.2 mL) was added slowly dropwise. The temperature was kept between -7 and -2 °C during addition and stirred at 0 °C for 1 h. Water (0.1 mL) was added and stirred for 20 min and 6M HCl (0.53 mL) was added to bring pH 5-6. The organic layer was separated, washed with sat. $NaHCO_3$ solution, dried ($MgSO_4$) and concentrated under reduced pressure. This was purified by distillation (124-125 °C, 10 mmHg) to yield **51**, as a yellow oil (0.0192 g, 4%).[183,184] Rf = 0.28 (1:3 EtOAc:Pet. ether); ν_{max} (neat): 2924 (C-H), 1703 (C=O) cm^{-1}; ^1H NMR (600 MHz; $CDCl_3$): 9.76 (1H, t, *J* = 1.5, 9-H), 2.44-2.57 (2H, m, 8-H), 2.35-2.37 (2H, m, 2-H), 2.09-2.34 (1H, m, 6-H), 2.02-2.06 (H, m, 3-H), 1.99-2.02 (1H, m, 7-H), 1.61-1.67 (2H, m, 4-H), 1.53-1.67 (2H, m, 5-H); ^{13}C NMR (150 MHz; $CDCl_3$): 212.9 (C-1), 202.6 (C-9), 49.8 (C-6), 42.3 (C-8), 41.9 (C-2), 34.5 (C-3), 28.2 (C-7), 25.2 (C-4), 22.2 (C-5); *m/z* (EI) 154 ($[M]^+$).

7.2.11 3-(2,2-Dimethylhydrazineylidene)-8-methyl-8-azabicyclo[3.2.1]octane **94**



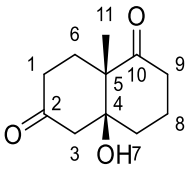
N,N-dimethylhydrazine (2mL) was added to tropinone (1.11 g, 18.5 mmol) and pTsOH (7.5 µL) and stirred under reflux (75 °C) for 16 h under argon. The excess *N,N*-dimethylhydrazine was removed *in vacuo*, and the resulting mixture was purified *via* vacuum distillation using a Kugelrohr apparatus (oven temp 110 °C, 3 mmHg) to yield **94** as a runny yellow oil (802 mg, 56 %).[170] Rf = 0.65 (30% EtOAc/Pet. ether 40-60 °C); ν_{max} (neat): 2946 (C-H), 1628 (C=N-) cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): 3.24-3.27 (1H, br. s., 3-H), 3.18 – 3.22 (1H, m, 6-H), 2.96 (1H, m, 7-*H*H), 2.55 (1H, d, *J* = 15.0 Hz, 2-*H*H), 2.37 (6H, d, *J* = 2.0 Hz, 10-H & 11-H), 2.33 (3H, d, *J* = 2.0 Hz, 8-H), 2.19 (1H, d, *J* = 15.0 Hz, 7-H*H*), 2.09-2.16 (1H, m, 2-H*H*), 1.96 (2H, br. s., 4-*H*H & 5-*H*H), 1.54-1.60 (1H, m, 4-H*H*) 1.39-1.45 (1H, m, 5-H*H*); ^{13}C NMR (150 MHz, $CDCl_3$): 165.35 (C-1), 61.3 (C-3), 60.4 (C-6), 47.3 (C-10), 41.0 (C-2), 39.2 (C-8), 35.1 (C-7), 27.2 (C-5), 26.5 (C-4); *m/z* (EI) 181 ($[M]^+$).

7.2.12 2-Methyl-2-(3-oxobutyl)cyclohexane-1,3-dione 60



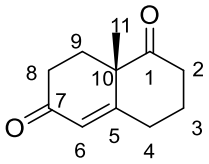
2-methyl-1,3-cyclohexadione (1.26 g, 10.0 mol) was added to H_2O (6 mL), AcOH (6 µL) and methyl vinyl ketone (1.62 mL, 10% hydroquinone, 20 mol). This was stirred at 72-75 °C for 1 h, before NaCl (1 g) was added and extracted with EtOAc (3 x 10 mL), washed with sat. NaCl solution (2 x 10 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The oil was washed using a silica plug with 10% EtOAc/Pet. ether to procure **60** as a brown oil (0.0278 g, 1.4%).[144] Rf = 0.27 (30% EtOAc/Pet. ether); ν_{max} (neat): 2942 (C-H), 1713(C=O); 1690 (C=O) cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): δ 2.53-2.70 (4 H, m, H-1, H-3), 2.25-2.31 (2H, t, H-9), 2.04 (3H, s, H-11), 1.81-2.01 (4H, m, H-2 & H-8), 1.18 (3H, s, H-7); ^{13}C NMR (150 MHz, $CDCl_3$): δ 210.1 (C-4 & C-6) 207.7 (C-10), 64.4 (C-5), 38.4 (C-9), 37.8 (C1 & C-3), 30.1 (C-8), 29.6 (C-11), 20.2 (C-7), 17.7 (C-2) ; *m/z* (ES+) 197 ($[M+H]^+$).

7.2.13 (4a*S*,8a*S*)-4a-Hydroxy-8a-methylhexahydronaphthalene-1,6(2H,5H)-dione **62**



Compound **60** 0.348 g, 1.8 mol) was added to DMF (2 mL), and (*S*)-(-)-proline (0.01 g, 0.09 mol) under argon. This was stirred at 25 °C for 120 h, before the solvent was removed *in vacuo*. The crude product was purified *via* flash column chromatography (40% EtOAc/Pet. ether) to procure **62** as a brown oil (0.197 g, 64%).[143] Rf = 0.34 (40% EtOAc/Pet. ether). $[\alpha]^{25}_{D}$ -4.25 (toluene, c 1.5) [lit. $[\alpha]^{25}_{D}$ -21.97 (CH_3Cl, c 1.1)][185]; ν_{max} (neat): 3437 (O-H), 2933 (C-H), 1706 (C=O) cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): δ 2.60-2.66 (1H, m, 9-*H*H,) 2.49 – 2.59 (2H, 1-*H*H & 3-*H*H), 2.39-2.42 (1H, m, 9-H*H*), 2.26-2.36 (3H, m, 1-H*H* & 3-H*H* & 6-*H*H), 2.09-2.13 (1H, m, 7-*H*H), 2.00-2.07 (1H, m, 8-*H*H), 1.76 -1.83 (1H, m, 6-H*H*), 1.65-1.73 (2H, m, 7-H*H* & 8-H*H*), 1.30 (3H, s, H-11); ^{13}C NMR (150 MHz, $CDCl_3$): δ 212.8 (C-10), 209.8 (C-2), 79.4 (C-4), 53.8 (C-5), 51.3 (C-3), 38.4 (C-1), 36.1 (C-9), 35.0 (C-7), 30.3 (C-6), 20.4 (C-11), 20.3 (C-8); *m/z* (EI) 196 ($[M]^+$).

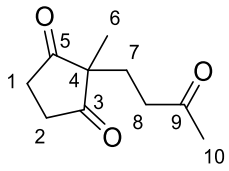
7.2.14 (*S*)-Wieland–Miescher ketone (*S*)-54



Concentrated H_2SO_4 (30 µL) was added to DMF (500 µL) at 0 °C. Compound **62** (0.202 g, 0.102 mol) made up in dry DMF (250 mL) was heated to 70-75 °C. The H_2SO_4 solution (200 µL) was added the reaction mixture was stirred at 95 °C for 1 h and an additional 75 µL. After cooling, the solvent was removed *in vacuo* and residual oil was dissolved in sat. NaCl solution (15 mL) and washed with CH_2Cl_2 (3 x 10 mL). The combine organic layers were washed with sat. NaCl solution (2 x 10 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was then filtered through a

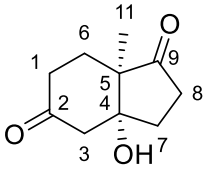
silica plug and washed with EtOAc to procure **(*S*)-54** as a dark brown oil (0.0105 g, 58%).[143,186,187] 71% e.e. by HPLC analysis. Rf = 0.40 (50% EtOAc/Pet. ether); $[\alpha]^{25}_{D}$ 14.31 ($CHCl_3$, c 0.8) [lit. $[\alpha]^{25}_{D}$ 104 (94% e.e., $CHCl_3$, c 0.8)]; ν_{max} (neat): 2951 (C-H), 1709 (C=O), 1663 (C=C) cm^{-1}; ^1H NMR (600 MHz; $CDCl_3$): δ 5.85 (1H, d, 6-H, *J* = 1.7 Hz), 2.68-2.74 (2H, m, 8-H), 2.40-2.49 (4H, m, 2-H and 4-H), 2.01-2.05 (2H, m, 9-H), 1.67-1.74 (2H, m, 3-H), 1.45 (3H, s, 11-H); ^{13}C NMR (600 MHz; $CDCl_3$): δ 211.3 (C-7), 198.5 (C-1), 166.0 (C-5), 126.0 (C-6), 50.7 (C-10), 37.9 (C-2), 33.7 (C-9), 31.9 (C-4), 29.8 (C-8), 23.4 (C-11), 23.0 (C-3). *m/z* (EI) 178 ($[M]^+$).

7.2.15 2-Methyl-2-(3-oxobutyl)cyclopentane-1,3-dione **61**



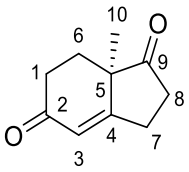
Compound 2-methyl-1,3-cyclopentadione (0.615 g, 5.49 mol) was added to H_2O (0.87 mL), AcOH (20 µL) and methyl vinyl ketone (0.770 mL, 10% hydroquinone, 9.20 mol). This was stirred at 75 °C for 14 h, before extracted with CH_2Cl_2 (3 x 10 mL), washed with sat. NaCl solution (2 x 10 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified *via* flash column chromatography (30% EtOAc/Pet. ether) to procure **61** as a pale orange oil (1.022 g, 91%) [lit. 96%][142]. Rf = 0.50 (40% EtOAc/Pet. ether). ν_{max} (neat): 2926 (C-H), 1760 (C=O), 1716 (C=O) cm^{-1}; ^1H NMR (600 MHz, $CDCl_3$): δ 2.70-2.86 (4H, m, H-1 & H-2), 2.43 (2H, t, *J* = 7.2 Hz, H-8), 2.08 (3H, s, H-10), 1.86 (2H, t, *J* = 7.2 Hz, H-7), 1.08 (3H, s, H-6); 13C NMR (150 MHz, $CDCl_3$): δ 216.0 (C-3 & C-5), 208.1 (C-9), 55.2 (C-4), 37.5 (C-8), 34.8 (C-1 & C-2), 30.2 (C-7), 27.9 (C-10), 19.2 (C-6); *m/z* (ES+) 182 ($[M+H]^+$).

7.2.16 (3a*R*,7a*R*)-3a-Hydroxy-7a-methylhexahydro-1H-indene-1,5(4H)-dione **63**



A foil covered round bottomed flask was charged with (*R*)-(+)-proline (0.0039 g) under argon. Dry DMF (1 mL) was added and stirred at 15-16 °C for 20 min. Compound **61** (0.200 g, 1.1 mol) in DMF (1 mL) was added and stirred at 15-16 °C for 5 days, before the solvent was removed *in vacuo*. The crude product was purified *via* flash column chromatography (50% EtOAc/Pet. ether) to procure **63** as a pale brown solid (0.0558 g, 29%, >99% d.e. by ^{1}H NMR).[143] Mp 97 – 98 °C, [lit.: 108 – 110 °C][188] ; Rf = 0.32 (50% EtOAc/Pet. ether); [α] $^{25}_{D}$ -57.46 ($CHCl_3$, c 0.8) [lit. [α] $^{25}_{D}$ -57.6 ($CHCl_3$, c 0.8)][188]; ν_{max} (neat): 3468 (O-H), 2924 (C-H), 1713 (C=O) cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): δ 2.62 (2H, s, 3-H), 2.48-2.59 (2H, m, 8-H), 2.39-2.47 (1H, m, 1-*H*H), 2.32 (1H, m, 1-H*H*), 1.97-2.03 (2H, m, 7-H), 1.66-1.78 (2H, m, 6-H), 1.26 (3H, s, 11-H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 217.8 (C-9) 207.8 (C-2), 81.6 (C-4), 52.6 (C-5), 50.6 (C-3), 36.7 (C-1), 33.6 (C-8), 33.0 (C-7), 29.9 (C-6), 14.0 (C-11); *m/z* (EI) 182 ([M]$^{+}$).

7.2.17 (4a*S*)-5-Hydroxy-4a-methyl-4,4a,5,6,7,8-hexahydronaphthalen-2(3H)-one **(*R*)-57**

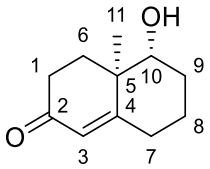


Concentrated H_2SO_4 (30 µL) was added to DMF (500 µL) at 0 °C. Compound **63** (0.0105 g, 0.00579 mmol) in dry DMF (250 mL) was heated to 70-75 °C. The H_2SO_4 solution (200 µL) was added the reaction mixture was stirred at 95 °C for 1 h and an additional 75 µL. After cooling, the solvent was removed *in vacuo* and residual oil was dissolved in sat. NaCl solution (15 mL) and washed with CH_2Cl_2 (3 x 10 mL). The combine organic layers were washed with sat. NaCl solution (2 x 10 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was then filtered through a silica plug and washed with EtOAc to procure ***(R)*-57** as a dark brown oil (0.008 g, 44%)[143]. Rf = 0.36 (50% EtOAc/Pet. ether); [α] $^{25}_{D}$ -70.74 (toluene, c 1) [lit. [α] $^{25}_{D}$ -67 (toluene, c 1)][189]; ν_{max} (neat): 2916 (C-H), 1735 (C=O) 1602 (C=C) cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): δ 5.86 (1H, d, *J* = 1.7 Hz, 3-H), 2.93 – 2.99 (1H, m, 7-*H*H), 2.65-2.78 (2H, m, 7-H*H* & 8-*H*H), 2.42-2.56 (3H, m, 1-H & 8-H*H*), 2.10-2.13 (1H, dd, 6-H), 1.83 -1.87 (1H, td, 6-H), 1.32 (3H, s, H-10); ^{13}C NMR (150 MHz, $CDCl_3$): δ 216.7 (C-9) 198.3 (C-2), 169.8 (C-4), 124.0

(C-3), 48.8 (C-5), 36.0 (C-8), 33.0 (C-1), 29.3 (C-6), 27.0 (C-7), 20.7 (C-10); *m/z* (EI) 164 ([M]$^+$).

7.2.18 (4a*R*,5*R*)-5-Hydroxy-4a-methyl-4,4a,5,6,7,8-hexahydronaphthalen-2(3H)-one

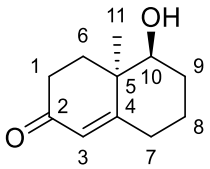
(4a*R*,5*R*)-56



To a stirred solution of (R)-WMK (0.020 g, 0.12 mmol) and EtOH (500 µL) as 0 °C, $NaBH_4$ (0.0055 g, 0.11 mmol) was added portion-wise and stirred for 15 min. AcOH (50 µL) was added and stirred for 15 min at 0 °C. The solvents were evaporated, and the remaining mixture was extracted with EtOAc (3 x 10 mL) and washed with sat. NaCl solution (2 x 10 mL). The organic layer was dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified using a silica plug and washed with EtOAC (10 mL) to procure **(4aR,5R)-56** as a white oil (0.0194 g, 96%), 98% d.e. by HPLC (Ratio 99:1 4a*R*,5*R*:4a*R*,5*S*). Rf = 0.30 (30% EtOAc/Pet. ether); $[\alpha]^{25}_D$ -34.53 ($CHCl_3$, c 1) [lit. $[\alpha]^{25}_D$ -90.4 (benzene, c 1)][190]; ν_{max} (neat): 3407 (O-H), 2932 (C-H), 1719 (C=O), 1614 (C=C) cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): δ 5.80 (1H, d, *J* = 1.9 Hz, 3-H), 3.44 (1H, d, *J* = 11.7 Hz, 10-H), 2.31-2.50 (3H, m, 1-H & 7-*H*H), 2.16-2.26 (2H, m, 6-*H*H & 7-H*H*), 1.81-1.94 (3H, m, 6-H*H* & 8-*H*H & 9-*H*H), 1.66-1.76 (1H, m, 9-H*H*), 1.37-1.48 (1H, m, 8-H*H*),1.21 (3H, s, 11-H); ^{13}C NMR (150 MHz, $CDCl_3$): δ199.6 (C-2), 168.3 (C-4), 125.7 (C-3), 78.5 (C-10), 41.7 (C-5), 34.4 (C-6), 33.8 (C-1), 32.1 (C-7), 30.4 (C-9), 23.3 (C-8), 15.4 (C-11); *m/z* (EI) 180 ([M]$^+$).

7.2.19 (4a*R*,5*S*)-5-Hydroxy-4a-methyl-4,4a,5,6,7,8-hexahydronaphthalen-2(3H)-one

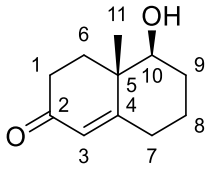
(4a*R*,5*S*)-56



The freeze-dried cells of SDR-17 (25 mg) were resuspended in buffer (NaPi, 100 mM, 1 mL) and were sonicated for ten x 15 s pulses at 0 °C before being

centrifuged (10 min, 3000 g, 4 °C). The enzymatic reaction was performed reaction using a total volume of 50 mL, containing buffer (NaPi, 100 mM, pH 7.2), (*R*)-WMK **(*R*)-54** (20 mM), clarified cell lysate (0.4 mg/mL), NADP (3 mM), DMSO (10%, v/v), G6P (100 mM), Shake 24 h, 37 °C, 500 rpm. The reaction was stopped with the addition of TFA (0.5% v/v), and the product extracted with EtOAc (5 x 15 mL) and washed with sat. NaCl solution (5 x 15 mL) to afford **(4a*R*,5*S*)-56** as a yellow solid without further purification (0.160 g, 89%, >99% e.e. by HPLC). Mp 75-79 °C [lit. 88-90 °C][191]; Rf = 0.61 (33% EtOAc:Pet. ether); $[\alpha]^{25}_{D}$ -111.66° (toluene, c 1.3) [lit. $[\alpha]^{25}_{D}$ -111 (benzene, c 1.3)][161]; ν_{max} (neat): 3427 (O-H), 2963 (C-H), 1638 (C=O), 1600 (C=C) cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): δ 5.87 (1h, d, *J* = 1.9 Hz, 3-H), 3.65 (1H, t, *J* = 2.7 Hz, 10-H), 2.56-2.62 (1H, m, 6-*H*H), 2.44-2.52 (2H, m, 1-H), 2.39-2.44 (1H, m, 7-*H*H), 2.24-2.32 (1H, m, 7-H*H*), 2.01-2.09 (1H, m, 9-*H*H), 1.82-1.93 (1H, m, 8-*H*H), 1.76-1.82 (1H, m, 9-H*H*), 1.67-1.75 (1H, m, 8-H*H*), 1.47-1.53 (1H, m, 6-H*H*), 1.24 (3H, s, 11-H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 199.6 (C-2), 168.0 (C-4), 127.2 (C-3), 75.4 (C-10), 41.0 (C-5), 34.1 (C-1), 31.8 (C-7), 30.9 (C-6), 28.8 (C-9), 21.9 (C-11), 19.9 (C-8); *m/z* (ES+) 181 ($[M+H]^+$).

7.2.20 (4a*S*,5*S*)-5-Hydroxy-4a-methyl-4,4a,5,6,7,8-hexahydronaphthalen-2(3H)-one (4a*S*,5*S*)-56



To a stirred solution of (*S*)-WMK **(*S*)-54** (0.030 g, 0.17 mmol) and EtOH (500 µL) as 0 °C, $NaBH_4$ (0.002 g, 0.057 mmol) was added portion-wise and stirred for 15 min. The solvents were evaporated *in vacuo*, and the remaining mixture was extracted with EtOAc (3 x 10 mL) and washed with sat. NaCl solution (2 x 10 mL). The organic layer was dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified *via* flash column chromatography with 50% EtOAC/Pet. ether to procure **(4a*S*,5*S*)-56** as a colourless oil (0.0189 g, 63%). 96% d.e. by HPLC (Ratio 98:2 (4a*S*,5*S*):(4a*S*,5*R*)). Rf = 0.32 (50% EtOAc/Pet. ether); $[\alpha]^{25}_{D}$ 20.71° ($CHCl_3$, c 1) [lit. $[\alpha]^{20}_{D}$ 78 ($CHCl_3$, c 1)][160]; ν_{max} (neat): 3320 (O-H), 2971 (C-H), 1655 (C=O) cm^{-1}; ^{1}H NMR

(600 MHz, $CDCl_3$): δ 5.79 (1H, d, *J* = 1.8 Hz, 3-H), 3.44 (1H, dd, *J* = 11.7, 4.16 Hz, 10-H), 2.30-2.49 (3H, m, 1-H & 7-*H*H), 2.15-2.26 (2H, m, 6-*H*H & 7-H*H*), 1.37-1.48 (1H, m, 8-*H*H), 1.20 (3H, s, 11-H), 1.63-1.76 (2H, m, 9-*H*H, OH), 1.81-1.94 (3H, m, 6-H*H*, 8-H*H*, 9-H*H*); ^{13}C NMR (150 MHz, $CDCl_3$): δ 199.7 (C-2), 168.4 (C-4), 125.7 (C-3), 78.5 (C-10), 41.7 (C-5), 34.4 (C-6), 33.8 (C-1), 32.1 (C-7), 30.4 (C-9), 23.3 (C-8), 15.4 (C-11); *m/z* (EI) 180 ($[M]^+$).

7.3 Docking studies

Structures were prepared using Chemdraw Professional Version 16.0 (Perkin-Elmer) and were energy minimised using ChemBioDraw3D Version 16.0 (Perkin-Elmer) MM2 energy optimisation. The resulting coordinates were saved in PBDQT format using AutoDock Tools.

7.3.1.1 Cv-TAm

AutoDock Vina was used to dock the compound structures into the crystal structure of the CV2025 homodimer (PDB: 4AH3). The "box" to locate the substrate binding had the following attributes: centre (x,y,z) = (32, 20, 20); size (x, y, z) = 25, 44, 44. Pymol was then used to calculate the distance between the carbonyl carbon group and to the primary amine group on PMP.

7.3.1.2 Ds-TRI and Ds-TRII

AutoDock Vina was used to dock the compound structures into the crystal structure of the TRI complex with NADP (PDB: 1AE1) and TRII complex with NADP (PDB: 1IPE). The "box" to locate the substrate binding had the following attributes: centre (x,y,z) = (43, 105, 47); size (x, y, z) = 24, 24, 24 and centre (x,y,z) = (5, 32, 150); size (x, y, z) = 24, 24, 24 respectively. Pymol was then used to calculate the distance between carbonyl carbon group to the carbon that donates/accepts a hydrogen on NADP(H).

7.4 Molecular biology

7.4.1 PCR to retrieve AKR genes from Metagenome

The following commercial products were used: DMSO (NEB, Catalogue reference #B0515A), 5x Phusion GC reaction buffer (NEB, Catalogue reference #B05195); Phusion high fidelity DNA polymerase (NEB, Catalogue reference

#M05305, 2 U/mL); Phusion high fidelity PCR master mix (NEB, Catalogue reference #M05325); deoxyribonucleotide (dNTP) solution mix (NEB, NO447AA); Q5 high GC enhancer (NEB, Catalogue reference #B9028A), q5 reaction buffer (NEB, Catalogue reference #B9028A); Q5 High fidelity DNA polymerase (NEB, Catalogue reference #M04915) and GeneRuler 1 kb DNA ladder (Thermoscientific, SM031), Monarch gel extraction kit (NEB, Catalogue reference #T1020L).

A sterile PCR Eppendorf was charged with the following mixtures (to a total 20 µL):

Genes 1, 2, 4, 5: Forward primer (10 µM), backwards primer (10 µM), metagenomic DNA (85 ng/µL), DMSO (5% v/v), dNTPs (40 µM), phusion polymerase (1% v/v), GC buffer (5x, 20% v/v), water (62% v/v).

Gene 8: Forward primer (10 µM), backwards primer (10 µM), metagenomic DNA (85 ng/µL), dNTPs (10 µM), phusion polymerase (1% v/v), Q5 polymerase (0.5% /v), 5 x GC enhancer (10% v/v), water (23.5% v/v).

Genes 9-11 (50 µL): Forward primer (10 µM), backwards primer (10 µM), metagenomic DNA (85 ng/µL), DMSO (3% /v/), phusion HF PCR master mix (50% v/v), water (36% v/v).

Genes 12-16 (50 µL): Forward primer (10 µM), backwards primer (10 µM), metagenomic DNA (85 ng/µL), DMSO (3% /v/), dNTPs (40 µM), phusion polymerase (1% v/v), GC buffer (5x, 20% v/v), water (64% v/v).

A TECHNE PCR machine was used following the manufacturer's protocols, with the following conditions: 98 °C for 5 min; 30 cycles of 98 °C for 10 s and 72 °C for 15 s; 72 °C for 7 min; hold at 4 °C. An agarose gel (1%, Section 7.4.4) was prepared and the PCR products were ran on this. A gel extraction was completed with a Monarch kit following the manufacturer's instructions.

The following primers were used (F and R denotes forward and reverse primers respectively): AKR-1F (AAA GGT CTC TTA TGG GCT GCA ATC TCC GGA TGT TCC), AKR-1R (AAA GGT CTC GGG TGA TCC CAT TTC GGC GCG AAG C), AKR-2F (AAA CAT ATG TGG CAG GTG GAT GCC GGC GTC), AKR-2R (AAA CTC GAG GAA CAT GAA AGA

TGC CGT CTC CGG), AKR-3F (AAA CAT ATG AGC GAA CAA CCG ACC AT TCC G), AKR-3R (AAA CTC GAG GAA AGT GGC CGT GAC TGG ATT CG), AKR-4F (AAA CAT ATG GAA TAC CTG ACG CTT TCT CAT ACC GAT TTA CG), AKR-4R (AAA CTC GAG TGG CAT GCG TGC GCC GCG CG), AKR-5F (AAA GGT CTC TTA TGA AGA GCT ATC GCC TGG CCG ATG), AKR-5R (AAA GGT CTC GGG TGA GGC ACC CGC CCA CCC CGC G), AKR-6F (AAA GGT CTC GGG TGA AGA CCT ATA CCC TCG CCG ACG GC), AKR-6R (AAA GGT CTC GGG TGC GGC ACA CGA TCA CCA CGC G), AKR-7F (AAA GGT CTC TTA TGC GAC TCG GAC TTG GCA CGA TG), AKR-7R (AAA GGT CTC GGG TGC GGC GAC GGA TTC GGA TGC TG), AKR-8F (AAA CAT ATG ACC TGG GGC GAG CAG AAC AGC), AKR-8R (AAA CTC GAG CGG CGA CGG ATT CGG ATG CTG), AKR-9F (AAA GGT CTC TTA TGG AAT ATC GGC AAC TCG GAT CGT CG), AKR-9R (AAA GGT CTCG GGTG AAC AAG CGGC GGG TTA AGG CG), AKR-10F (AAA CAT ATG GAA ACC CGT TTTC TCGG CCG C), AKR-10R (AAA CTC GAG GGTT GCG CCT CGA TAT GCG G), AKR-11F (AAA CAT ATG GAA TAC CGC TAT CTC GGC CGC TC), AKR-11R (AAA CTC GAG CAC CGG AGT GGG GTT CCG GTA C), AKR-12F (AAA GGT CTC TTA TGG AAT ACC GTT ACC TCG GCG CGT C), AKR-12R (AAA GGT CTC GGG TGG ACC GGC GTG GGA TTG CGG TAG), AKR-13F (AAA CAT ATG GAA TAC CGC TAC CTC GGC GCA TC), AKR-13R (AAA CTC GAG GAC CGG GGC GGG ATT GCG GTA G), AKR-14F (AAA CAT ATG GCG ACC GACG GCAC CAA ACA AG), AKR-14R (AAA CTC GAG CTT CGT GAT CTG ATC TAT TTC AGC CAT TTC TTC), AKR-15F (AAA CAT ATG GGC GAG GTC ATC GCC GGA CG), AKR-15R (AAA AAG CTT GAT GAT GTC CAG CGG CCG GG), AKR-16F (AAA CAT ATG GAA AAC CAG ACT TTC GCG GTT GCC), AKR-16R (AAA CTC GAG CGC GCC GGC GAT CGT TCC GTC).

7.4.2 Classical cloning method

The following restriction enzymes used: *NdeI* (NEB, Catalogue reference #R0111S), *Xho*1 (NEB, R0146S), *Hind*III (NEB, Catalogue reference #R3104S), *BsaI*-HF (NEB, Catalogue reference #R3535S). The following commercial products were used: CutSmart buffer (NEB, Catalogue reference #B7204S), T4 DNA ligase (Thermo, Catalogue reference #ELD011), One Shot™ TOP10 Chemically Competent E. coli (Thermofisher, Catalogue reference #C404010). Hyper ladder 1kb (Bioline, Catalogue reference #BIO-33053)

The vector was digested in a sterile Eppendorf: pET29a vector (50 ng/µL), CutSmart buffer (1 µL), restriction enzyme (1 µL each) and incubated for 2 h at 37 °C. The genes were also digested in an Eppendorf: (9 µL), CutSmart Buffer (1 µL), restriction enzyme (1 µL each) and incubated for 2 h at 37 °C. Restriction enzymes used were *NdeI*/*Xho*1 for AKRs-2,3,8,10,11,13, and *NdeI*/*Hind*III for AKR-15. The two were ligated together at °C: T4 DNA ligase (1 µL), T4 DNA ligase buffer (10x, 2 µL), vector DNA (50 ng), insert DNA (37.5 ng) and water to make the total volume 20 µL. The reaction was incubated for 16 h and inactivated by incubating or 10 min at 65 °C. This was then transformed into commercial Top10 competent cells using standard protocols. DNA concentrations were determined using the NanoDrop (ng/µL): [AKR-2] = 73.4; [AKR-3] = 69.7; [AKR-6] = 59.1; [AKR-8] = 78.5; [AKR-10] = 71.6; [AKR-11] = 59.6; [AKR-13] = 73.2; [AKR-15] = 74.9.

7.4.3 Novel cloning method

The following commercial product were used: *BsaI*-HF (NEB, Catalogue reference #R3535S), T4 ligase (NEB, Catalogue reference #M0202M), ATP (NEB, Catalogue reference #P0756S), CutSmart buffer (NEB, B7204S), Top10 competent cells (Invitrogen, Catalogue reference #C404003), miniprep (QIAGEN, Catalogue reference #27106), BL21 DE3 (NEB, Catalogue reference #C2530H), S.O.C. media (Thermo Fisher Scientific, Catalogue reference #15544034).

A sterile Eppendorf was charged with: vector (pET29:SacB, 1 µL), PCR product (Section 7.4.1, 100 ng), *BsaI*-HF (1 µL), T4 ligase (1 µL), ATP (1 mM, 1 µL), CutSmart buffer (2 µL) and water to make the total volume 20 µL. This was incubated for up to 35 min at 37 °C.

Top10 cells were incubated at 0 °C for 20 min. To a sterile Eppendorf, Top10 cells (20 µL) and the reaction mixture (5 µL) were added and this was incubated at 0 °C for 30 min. The mixture was heat-shocked at 42 °C for 30 s before being replaced to 0 *°C* for 2 min. S.O.C. media (200 µL) was added and shook for 1 h at 37 °C at 225 rpm. This mixture (20 µL) was plated on LB agar with kanamycin (25-50 µg/mL) and supplemented with 10% sucrose. The plates were incubated for 16 h at 37 °C. LB media (10 mL) with kanamycin (10 µg/ml) was inoculated with a colony from the

plate, and grown for 16 h at 37 °C at 250 rpm. A miniprep (QIAGEN) kit was used according to the manufacturer's instructions to procure the plasmid. DNA concentrations were determined using the NanoDrop (ng/µL): [AKR-1] = 80.2; [AKR-5] = 44.9; [AKR-6] = 59.1; [AKR-9] = 41.9. This DNA (2 µL) was transformed in the same method as before into BL21 DE3 cells, from which glycerol stocks were made.

7.4.4 Agarose gel

The following commercial product were used: Agarose powder (Sigma, Catalogue reference #A9539), ethidium bromide (Sigma, 1239-45-8), loading dye (NEB, Catalogue reference #B70245), DNA ladder (Thermoscientific, Catalogue reference #SM0311, GeneRuler 1 kB, 0.5 g/ µL).

An agarose gel (1%) was prepared: agarose powder (2 g) was dissolved in TAE buffer (200 mL) and heated in the microwave. Upon cooling to approx. 60 °C (may be touched for 5 s by hand), ethidium bromide (5 µL/100 mL v/v) was added and mixed. The agarose gel was poured into the mould and allowed to cool. The gel was loaded on to the electrophoresis unit and submerged in TAE buffer. DNA loading dye (10 µL) was added to PCR reaction product (20 µL) and was loaded on to the gel along with a DNA ladder (consisting of loading dye (5 µL) and DNA ladder (3 µL) and run 50 V for 1 h. The gel was visualized using a UV gel documentation system.

7.4.5 Co-transformation of SDR17 and 31 with G6PDH

Competent *E. coli* BL21 (DE3) cells (NEB, #C2427I) were incubated at 0 °C for 10 min. A pET29a plasmid with SDR gene (3 µL) and glucose-6-phosphate dehydrogenase from *Saccharomyces cerevisiae* SF838 (Uniprot P11412) in a pACYCduet-1 plasmid (3 µL) were added, gently mixed, and incubated at 0 °C for 30 min. The cells were incubated for 45 s at 42 °C before being replaced at 0 °C for 2 min, Pre-warmed LB medium (800 µL, 37 °C) was added and shook at 37 °C, for 45 min at 500 rpm. The mixture was centrifuged at 2000 rpm, at 4 °C for 5 min, and the supernatant discarded, Cells were resuspended in the residual liquid which was transferred to LB agar plates (50 µg/mL kanamycin and 25 µg/mL chloramphenicol). The plates were incubated for 18 h at 37 °C used to make glycerol stocks.

7.5 Enzyme expression

7.5.1 Tropinone Reductase growth

Genes for TRI and TRII, both with and without His-tag, were designed by John Ward. TRI (UniProtKB/Swiss-Prot: P50162.1) and TRII (UniProtKB/Swiss-Prot: P50162.1) protein sequences were retrieved from NCBI online tool[192] and the protein sequences put into the Reverse Translate tool at bioinformatics.org. The "most likely codons" for *E. coli* sequence was used from that site. An *NdeI* site was designed at the N-terminus followed by 6 His codons and an *XhoI* site and two stop codons at the C-terminus. Synthetic genes were made by MWG/Eurofins/Operon and subcloned into pET29a plasmids, and then transformed into *E. coli* BL21(DE3) pLysS and grown on an agar plate containing kanamycin (50 µg/mL). This was used to inoculate LB broth (1 mL) and incubated for 2 h, 37 °C, 300 rpm. This was then added to LB broth (5 mL) containing kanamycin (10 µg/ml) and grown for approx. 4 h until OD_{600} of 0.6-0.8 was achieved. IPTG (200 µg/ml) was added and incubated for 16 h at 25 °C, shaking at 400 rpm and then centrifuged. The pellets were resuspended and combined in buffer (NaPi, 100 mM, pH 7.2, 500 µL). Cells were resuspended in lysis buffer (1 mL) containing BugBuster protein extraction solution (Novagen), DNAse I (20 µL/mL), lysozyme (1 µg/mL) or alternatively sonicated (3 x 10 s on/off at 14 watts). The supernatant after centrifugation (12000 rpm, 4 °C, 5 min) was employed as cell-free extract.

7.5.2 MecgoR growth

MecgoR was designed by John Ward. MecgoR (GenBank accession no. GU562618) was retrieved from NCBI online tool[192] and codon optimised using the reverse translate tool at bioinformatics.org. Synthetic genes were made by Eurofins/MWG and subcloned into pET29a, and then transformed into *E. coli* BL21 (DE3 Star) pLysS and grown on an agar plate containing kanamycin. This was used to inoculate LB broth (1 mL) and incubated for 2 h, 37 °C, 300 rpm. This was then added to LB broth (5 mL) containing kanamycin (10 µg/ml) and grown for approx. 4 h until OD_{600} of 0.6-0.8 was achieved. IPTG (200 µg/ml) was added and incubated for 16 h at 25 °C, shaking at 400 rpm and then centrifuged. The pellets were resuspended and combined in buffer (NaPi, 100 mM, pH 7.2, 500 µL). Cells were resuspended in lysis

buffer (1 mL) containing BugBuster protein extraction solution (Novagen), DNAse I (20 µL/mL), lysozyme (1 µg/mL) or alternatively sonicated (3 x 10 s on/off at 14 watts). The supernatant after centrifugation (12000 rpm, 4 °C, 5 min) was employed as cell-free extract.

7.5.3 SDR growth

E. coli containing cloned SDR genes (see list in Section 0) were prepared by Jack Jeffries: Gibson assembly was used to join the PCR products of the KRED genes directly into pET29a plasmids. This was then transformed into a cloning strain (*E. coli* Top10), which was grown up, a miniprep procedure performed and transformed into the expression strain (BL21*(DE3)pLysS) and stored as a glycerol stock.

These glycerol stocks were used to inoculate Terrific Broth (TB) medium (Merck, 500 µL/well in a 96 deep-square well plate) containing 50 µg/mL kanamycin and 30 µg/mL chloramphenicol and incubated for 16 h, 37 °C, 400 rpm, sealed with a microporous breathable membrane. Then 100 µL of this starter culture was added to TB media (10 mL) containing 50 µg/mL kanamycin, 30 µg/mL chloramphenicol. This was incubated for 6-8 h, until OD_{600} of 0.6-0.8 was achieved. IPTG (1 mM final concentration) was added and incubated for 16 h at 25 °C, shaking at 400 rpm. The cultures were centrifuged for 15 mins at 10,000 rpm. Cells resuspended in lysis buffer (1 mL) containing BugBuster protein extraction solution (Novagen), DNAse I (20 µL/mL), lysozyme (1 µg/mL) or alternatively sonicated (3 x 10 s on/off at 14 watts). The supernatant after centrifugation (12000 rpm, 4 °C, 5 min) was employed as cell-free extract.

The clarified cell lysates were aliquoted out into 20 µL portions in 96-well plates and flash frozen with liquid nitrogen. These were stored at -80 °C until defrosted prior to use.

Larger cultures were also grown as necessary and induced under the same conditions as before. SDR-17 and SDR-31 where co-expressed with G6PDH under the same conditions. Cells were freeze-dried and stored at -20 °C, before use as discussed in Section 7.6.2.

7.5.4 AKR growth

AKR genes in pET29a vectors (see list in Section 8.4) were transformed into *E. coli* BL21 (DE3 Star) pLysS and grown on an agar plate containing kanamycin. This was used to inoculate LB broth (1 mL) and incubated for 2 h, 37 °C, 300 rpm. This was then added to LB broth (5 mL) containing kanamycin (10 µg/ml) and grown for approx. 4 h until OD_{600} of 0.6-0.8 was achieved. IPTG (200 µg/ml) was added and incubated for 16 h at 25 °C, shaking at 400 rpm and then centrifuged. The pellets were resuspended and combined in (NaPi, 100 mM, pH 7.2, 500 µL). Cells were resuspended in lysis buffer (1 mL) containing BugBuster protein extraction solution (Novagen), DNAse I (20 µL/mL), lysozyme (1 µg/mL) or alternatively sonicated (3 x 10 s on/off at 14 watts). The supernatant after centrifugation (12000 rpm, 4 °C, 5 min) was employed as cell-free extract.

7.5.5 TAms growth

7.5.5.1 Small scale growth

Glycerol stocks of recombinant *E. coli* BL21 containing TAm genes (see list in Section 8.12) prepared by Nadine Ladkau were used to inoculate 2 x TY medium (800 µL/ well in a 96 deep-square well plate) containing 50 µg/mL kanamycin, 30 µg/mL chloramphenicol and sealed with a microporous breathable membrane. The cells were cultivated in an orbital shaker at 37 °C, 400 rpm and 85% humidity for 8 h and induced with 1 mM IPTG for 16 h at 30 °C before they were harvested *via* centrifugation (3000 g, 15 min, 4 °C). The supernatant was discarded by inversion of the plate and the cells were resuspended in buffer (NaPi, 100 mM, pH 8, 250 µL). The cell suspension was used either directly for cell lysis or stored at -80 °C until usage for up to 6 months.

Cell lysate were then prepared: 30 µL of the cell suspension were mixed with lysis buffer (225 µL) containing buffer (NaPi, 100 mM, pH 7.2), 0.2 mg mL^{-1} lysozyme, 1 mM PLP and 10% BugBuster protein extraction solution (Novagen) in a standard 96 well plate, and incubated at 21 °C for 10 min and 37 °C for 30 min. After centrifugation (4 °C, 3000 g, 1 h), supernatant was employed as clarified cell lysate and was used either immediately for bioconversion or was stored at -20 °C for up to 1 month.

7.5.5.2 Large growth

Glycerol stocks of recombinant *E. coli* BL21 BL21 containing TAm genes (see list in Section 8.12) prepared by Nadine Ladkau were used to inoculate 2xTY medium (5 mL) containing 50 µg/mL kanamycin, 30 µg/mL chloramphenicol and incubated for 16 h at 37 °C, 400 rpm. 1 mL of this starter culture was then added to 2 x TY medium (100 mL) containing 50 µg/mL kanamycin, 30 µg/mL chloramphenicol. This was incubated for 6-8 h, until OD_{600} of 0.6-0.8 was achieved. IPTG (1 mM final concentration) was added and incubated for 16 h, 30 °C, 400 rpm. Cells were harvested by centrifugation (12000 rpm, 4 °C, 5 min) and resuspended in the minimum quantity of buffer (NaPi, 100 mM, pH 7.2) containing 1 mM PLP. This was then freeze- dried and frozen at -20 °C for storage. When used for the assay, dry cells where resuspended in buffer (Kpi, 100 mM, pH 8.0, 30 g/L) and sonicated (3 x 10 s on/off at 14 watts). The supernatant solution after centrifugation (5 min, 13000 rpm, 4 °C) was employed as cell-free extract.

7.5.6 Purification

NTA-Ni spin column (QAIGEN) columns were used to purify the protein, following the manufacturer's instructions. The protein was buffer exchanged into buffer (NaPi, 100 mM, pH 7.2) using Zebra Spin Desalting plates (Thermoscientific) following the manufacturer's instructions. The protein was aliquotted out into 20 µL volumes and frozen at -80 °C for storage. Alternatively, ammonium sulphate was added to a final concentration of 3.2 M and the precipitate protein was stored at 4 °C for up to 6 months. For use, a portion of the suspension was centrifuged (45 min, 13, 000 rpm, 25 °C) and the supernatant discarded. The pellet was resuspended in water. Absorption was measured at 280 nm, and the following formula was used to calculate the protein concentration. Extinction coefficient was estimated using the ExPASy ProtParam online tool.[124]

$$[protein] = \frac{A_{280}\ x\ dilution\ factor\ x\ Mr}{Extinction\ coefficient}$$

7.5.7 SDS-PAGE

Protein composition was analysed by sodium dodecyl sulphate-12% polyacrylamide gel electrophoresis (SDS-PAGE) using Mini-Protean TGX Gels (Bio-Rad). Protein samples were prepared by heating for 5 minutes at 95 °C in the presence of sample buffer (1:1 dilution of 2 x Laemmli Sample Buffer (Bio-Rad) and 100 mM dithiothreitol:protein). A broad range protein marker (10-25 kDa, New England Biolabs) was used to estimate the molecular mass of the proteins. Instant*Blue* (Thermoscientific) was used to stain the protein following the manufacturer's instructions.

7.5.8 Bradford Assay

Protein concentration was determined using Bradford Reagent (Bio-Rad, 1 mL) mixed with the standard or the cell lysate (50 µL). Bovine serum albumin was used as the standard to calibrate a concentration curve, and a UV spectrophotometer (Eppendorf) was used to determine the absorption at 595 nm of the calibration curves and the cell lystate (diluted appropriately). The calibration curve was used to calculate the protein concentration in the cell lysates.

7.5.9 Recipes for buffers and media

7.5.9.1 Terrific broth (TB) media

Terrific broth (Merck, 47.6 g) and glycerol (Calbiochem, 4 mL) were added to water (1 L) and autoclaved for 15 min, 121 °C.

7.5.9.2 2 x TY broth media

Tryptone (16 g), yeast (10 g) and NaCl (5 g) were added to water (1 L) and autoclaved for 15 min, 121 °C.

7.5.9.3 Sodium phosphate buffer pH 7.2

0.2M monobasic sodium phosphate buffer (48 mL) and 0.2 M basic sodium phosphate buffer (252 mL) were added to water (300 mL) and autoclaved.

7.5.9.4 10% Sucrose plates

Agar (32 g) was added to water (800 mL) and autoclaved. Kanamycin (25-50 µg/mL) and 50% sucrose solution (filter sterile, 200 mL) was added and the plates poured.

7.5.9.5 TAE buffer (50x stock)

Tris-base (242 g), acetic acid (57.1 mL), EDTA (100 mL, 0.5 M) was added to a total volume of 1 L of water.

7.6 Biotransformations

7.6.1 General spectrophotometric ketone reduction assay method (200 µL)

Reduction of the ketone was determined at 25 °C by following the oxidation of NAD(P)H using a GENios microplate reader (Tecan) at 340 nm absorbance during 100 cycles of 57 s with a shake duration of 1 s and a shake settle time between cycles of 1 s. Each reaction mixture (200 µL) contained buffer (NaPi, 100 mM, pH 7.2), 5 mM substrate, and the enzyme (20 µL). The reaction was initiated by the addition of the substrate. Protein concentrations were determined using a Nanodrop 2000c spectrophotometer (Thermoscientific) or Bradford assay (Section 7.5.8). A calibration curve of NAD(P)H (0.062 to 1 mM) was used on each plate to calculate the amount of NADPH consumed. $\xi_{NAD(P)H}$ is 622 $M^{-1}mm^{-1}$ and pathlength on the plate reader is 4.55 mm (calculated from average absorbance of 1 mM NAD(P)H solution and Beer's law).[151]

7.6.2 General ketone reduction method (500 µL – 50 mL)

The freeze-dried cells (25 mg) were resuspended buffer (NaPi, 100 mM, pH 7.2, 1 mL) and were sonicated for ten x 15 s pulses at 0 °C before being centrifuged (10 min, 3000 g, 4 °C). Each assay mixture contained buffer (NaPi, 100 mM, pH 7.2, 500 mL), 5-50 mM substrate, clarified cell lysate (0.4 mg/mL), NADP(H) (3 mM), DMSO (10%, v/v), G6P (100 mM), G6PDH (3000 U). G6P, G6PDH, and NADP(H) was omitted as necessary. The reactions were shaken for 24 h, at 500 rpm, at 25-37 °C. The reaction was stopped with the addition of TFA (0.5% v/v), and centrifuged (4 °C, 3000 g, 20 min). Et_2O (1 mL) was added to the supernatant and vortexed for 30 s. The organic layer was separated, dried ($NaSO_4$), decanted into HPLC vials and evaporated. EtOH (500 uL) was then added.

For the WMK assays: The reactions were analysed by HPLC (Agilent) using a Chiralcel OJ column (250 x 4.6 mm ID Analytical column, 10 µm beads). Concentrations of product were determined using a 4% 2-propanol:hexane mixture,

run-time of 120 min, at 0.5 mL/min flow-rate, and detection at 230 nm. Retention times (min): **(*R*)-54** at 68.2; **(*S*)-54** at 78.3; **(4a*S*,5*S*)-56** at 85.9; **(4a*R*,5*R*)-56** at 61.2; **(4a*R*,5*S*)-56** at 54.0.

For the reduction of acetophenone to phenylethanol, the reaction was quantified by either HPLC or GC. HPLC method: A Chiralcel OD column (250 x 4.6 mm ID Analytical column, 10 µm beads) was used. Product concentrations were determined using a 10% 2-propanol:hexane mixture, run-time of 120 min, at 0.5 mL/min flow-rate and detection at 210 nm. Retention times (min): (*S*)-phenylethanol **(*S*)-55** at 6.6 min; (*R*)-phenylethanol **(*R*)-55** at 5.9 min.

GC method: Beta Dex 225 capillary column (30 m x 0.25 mm x 0.25 µm, Supelco) was used. Injector and FID detector temperatures were kept at 250 and 300 °C respectively. The column was initially set at 90°C and held for 1 min, followed by two heating stages: first from 90 °C to 150 °C with a rate of 5 °C/min; secondly from 150°C to 200 °C with a rate of 20 °C/min and held for 2 min. Retention times (min): (*R*)-phenylethanol at 9.2; (*S*)-phenylethanol at 9.5.

HPLC method: Reactions were also followed using HPLC (Aligent) under non-chiral conditions with the following flow rate (1 mL/min): solvent A (H2O + 0.1 % TFA); solvent B (acetonitrile) and solvent C (water). Elution was monitored over the following gradients: 0 min A: 85%, B: 15%; 1 min A: 85%, B: 15%; 10 min A: 28%, B: 72%; 12 min A: 85%, B: 15%; 13 min A: 85%, B: 15%. The reaction mixute (180 µL) was added to 540 µL of water before HPLC analysis. Detection at 254 nm and retention times (min): acetophenone at 8.8; phenylethanol **55** at 7.5; **56** at 5.9; WMK **54** at 6.9.; Detection at 210 nm: benzaldehyde **45** at 8.8; benzyl alcohol **82** at 6.1; 2-CBT **16** at 3.2.

7.6.3 TAm MBA Assay

The freeze-dried cells (25 mg) were resuspended in buffer (NaPi, 100 mM, pH 7.2, 1 mL) and were sonicated for ten x 15 s pulses at 0 °C before being centrifuged (10 min, 3000 g, 4 °C). Each assay mixture (200 µL) contained buffer (NaPi, 100 mM, pH 7.5), substrate (5 mM), 25 mM (*S*)-MBA, 1 mM PLP and the TAm enzyme (20 µL). The reaction was shaken at 750 rpm at 30 °C for 24 h. The reaction was stopped with

the addition of TFA (1 µL), and centrifuged (4 °C, 3000 g, 20 min). Transamination of the substrate was determined by following acetophenone production by the area of the peak using the HPLC (Aligent) with the following flow rate (1 mL/min): solvent A (H2O + 0.1 % TFA); solvent B (acetonitrile) and solvent C (water). Elution was monitored over the following gradients: 0 min A: 85%, B: 15%; 1 min A: 85%, B: 15%; 10 min A: 28%, B: 72%; 12 min A: 85%, B: 15%; 13 min A: 85%, B: 15%; The acetophenone was eluted at a time of 8.8 min and detected at 254 nm.

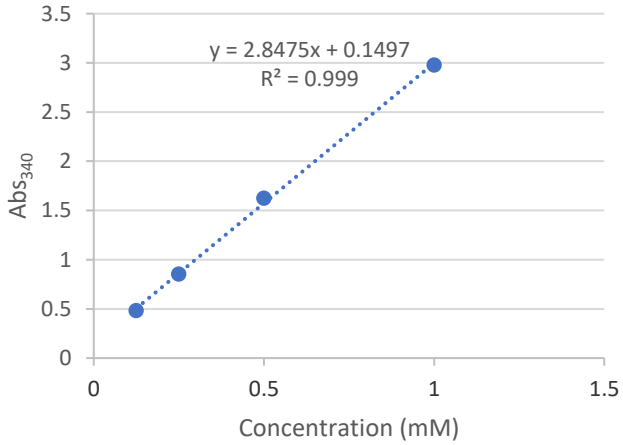
7.6.4 TAm Colourimetric assay

The freeze-dried cells (25 mg) were resuspended in buffer (NaPi, 100 mM, pH 7.2, 1 mL) and were sonicated for ten x 15 s pulses at 0 °C before being centrifuged (10 min, 3000 g, 4 °C). Each assay mixture (200 µL) contained buffer (NaPi, 100 mM, pH 7.5), substrate (10 mM), 2-(4-nitrophenyl)ethan-1-amine hydrochloride (25 mM), PLP (0.2 mM) and the TAm enzyme (4 mg/mL). The reaction was shaken at 500 rpm at 30 °C for 24 h.

7.7 Calibration curves

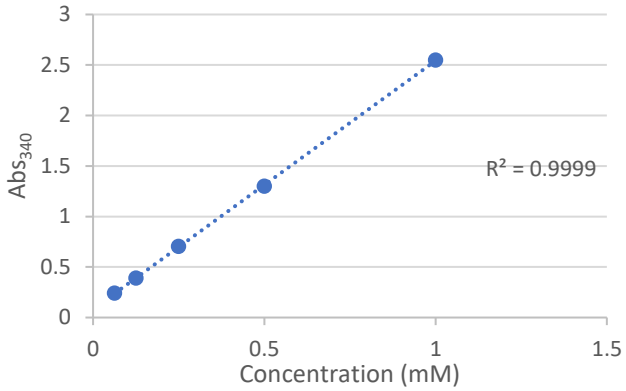
7.7.1 Example NADH for UV vis assays

Spectrophotometer, 340 nm



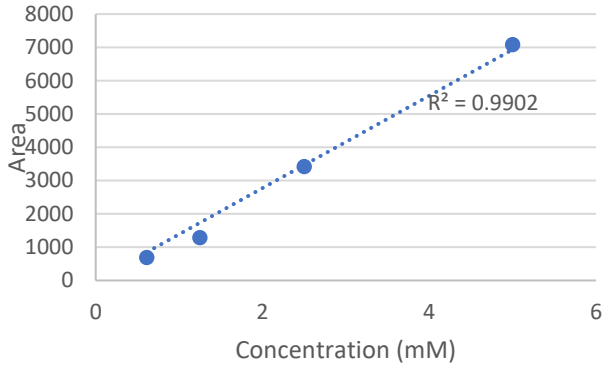
7.7.2 Example NADPH for UV vis assays

Spectrophotometer, 340 nm



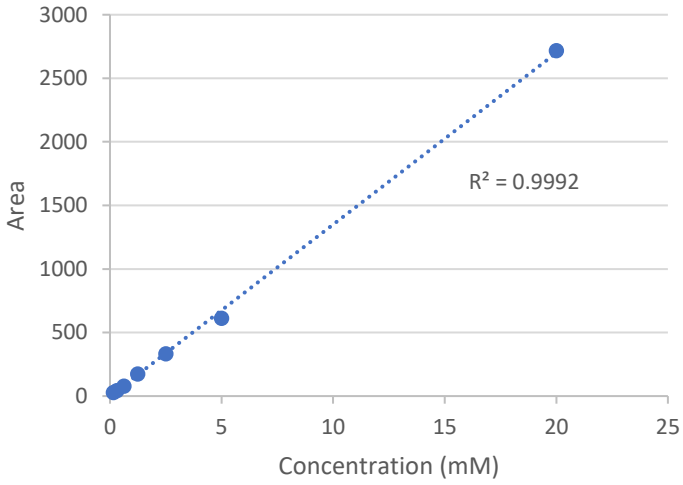
7.7.3 Acetophenone **29**

HPLC, C18 column, 254 nm



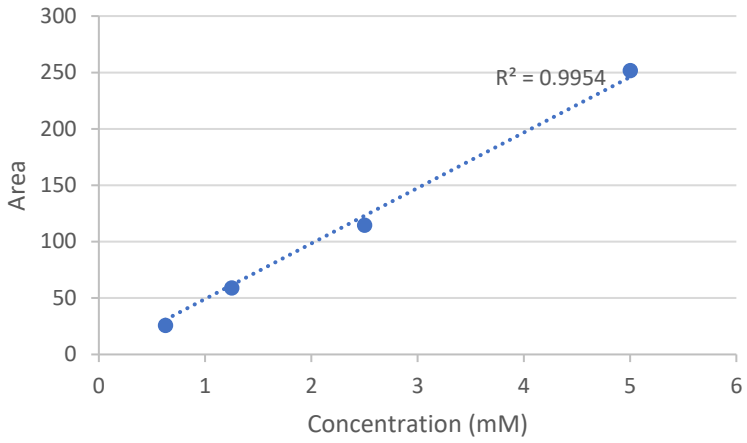
7.7.4 Benzaldehyde **45**

HPLC, C18 column, 254 nm



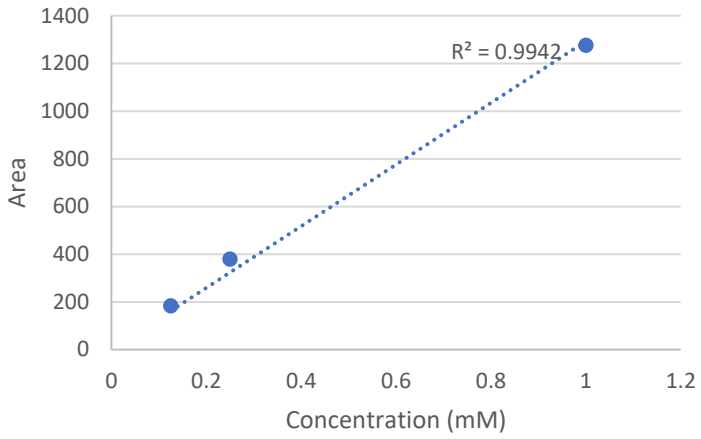
7.7.5 Benzyl Alcohol **82**

HPLC, C18 column, 210 nm



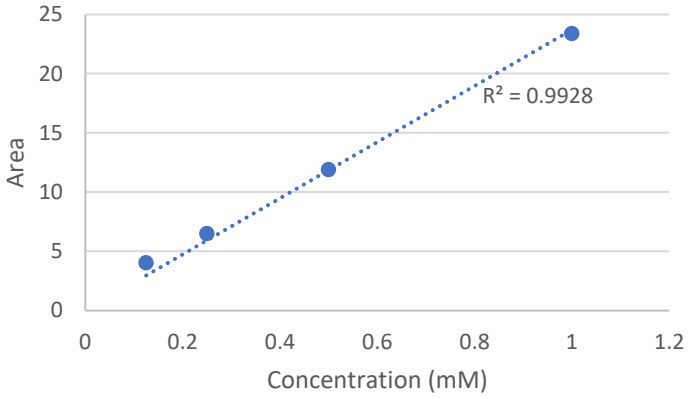
7.7.6 (4a*R*,5*S*)-56

HPLC, C18 column, 210 nm



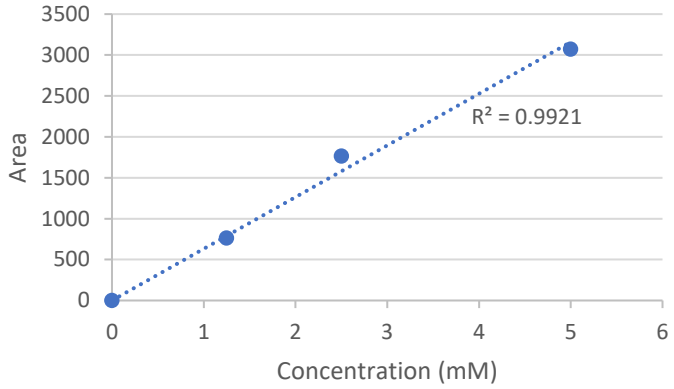
7.7.7 Phenylethanol **55**

HPLC, C18 column, 254 nm



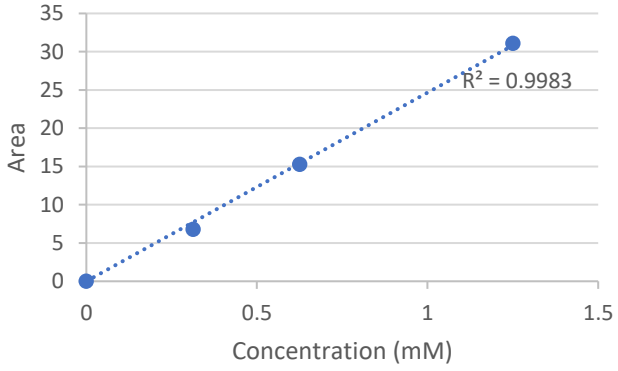
7.7.8 2-CBT **16**

HPLC, C18 column, 254 nm



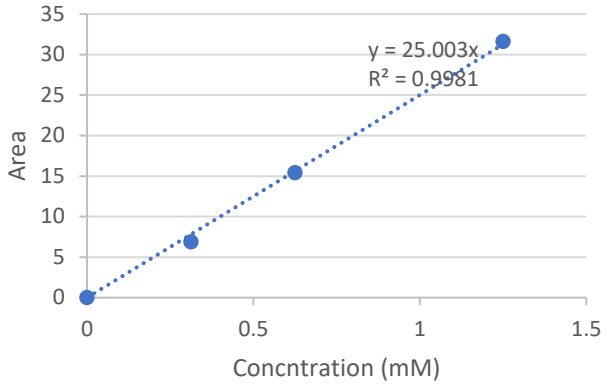
7.7.9 (*R*)-phenylethanol (***R***)-55

GC



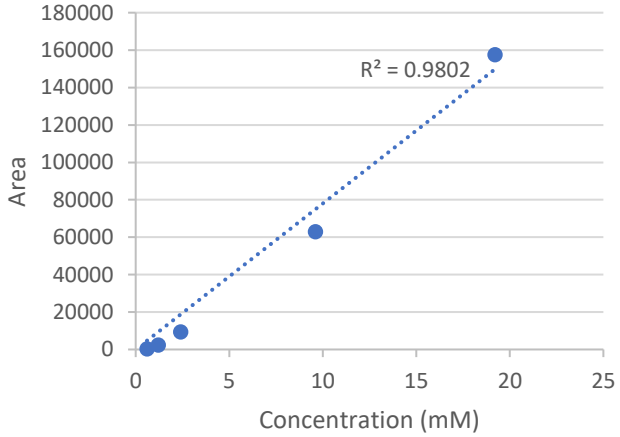
7.7.10 (*S*)-phenylethanol (*S*)-55

GC



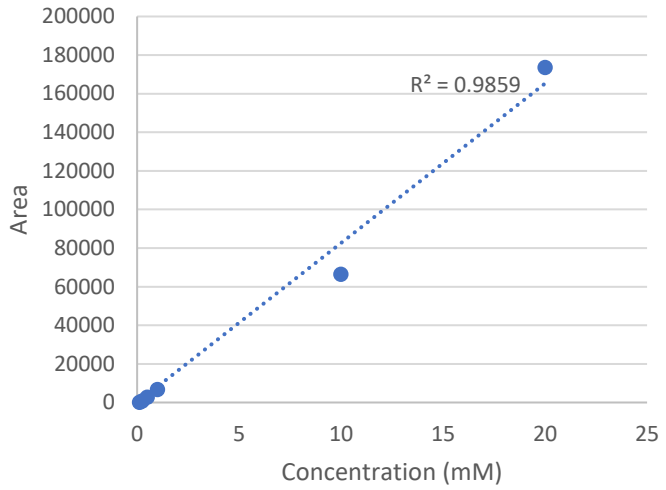
7.7.11 (*S*)-WMK (*S*)-54

HPLC, OJ column, 230 nm



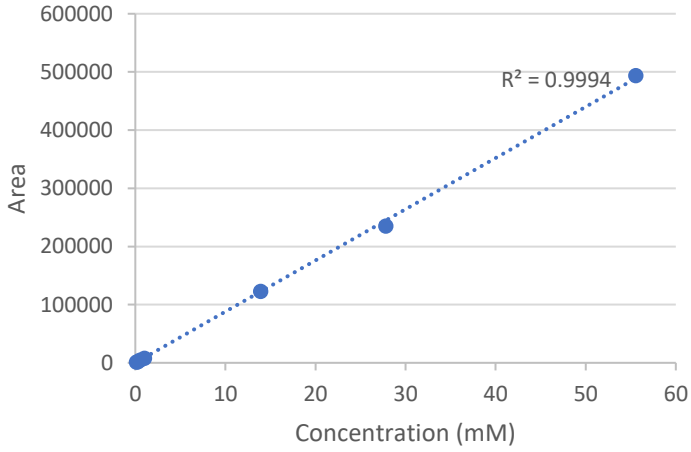
7.7.12 (*R*)-WMK (*R*)-WMK

HPLC, OJ column, 230 nm



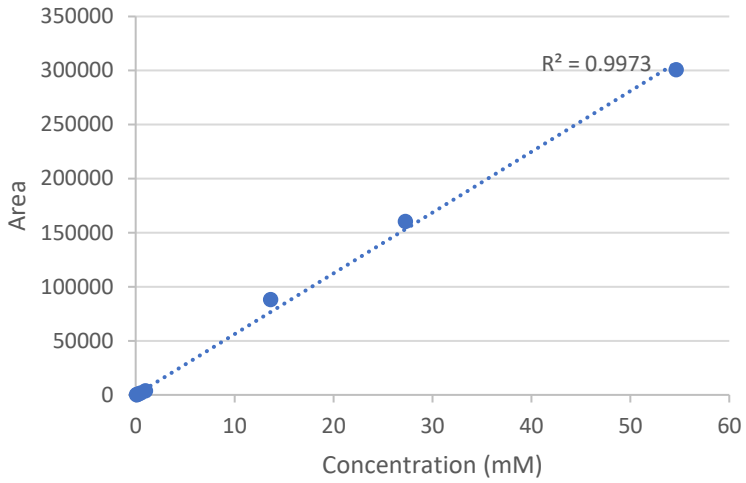
7.7.13 (4aS,5*S*)-56

HPLC, OJ column, 230 nm



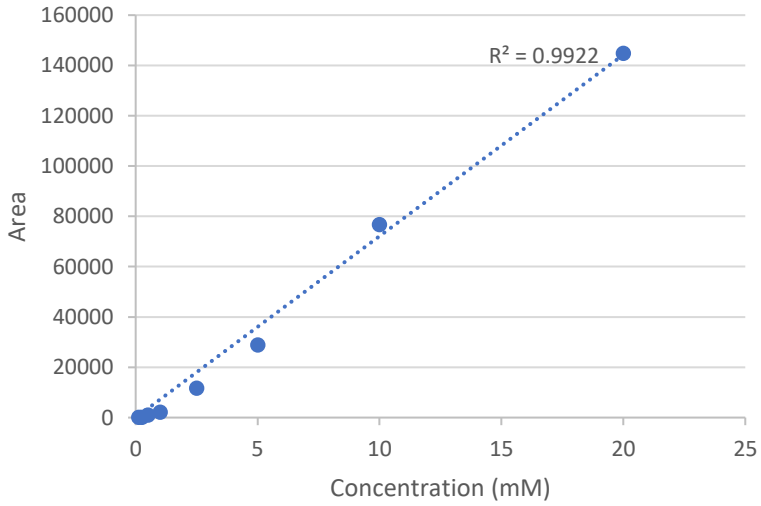
7.7.14 (4a*R*,5*R*)-56

HPLC, OJ column, 230 nm



7.7.15 (4a*R*,5*S*)-56

HPLC, OJ column, 230 nm



8 Appendix

8.1 Biosynthesis of tropane alkaloids

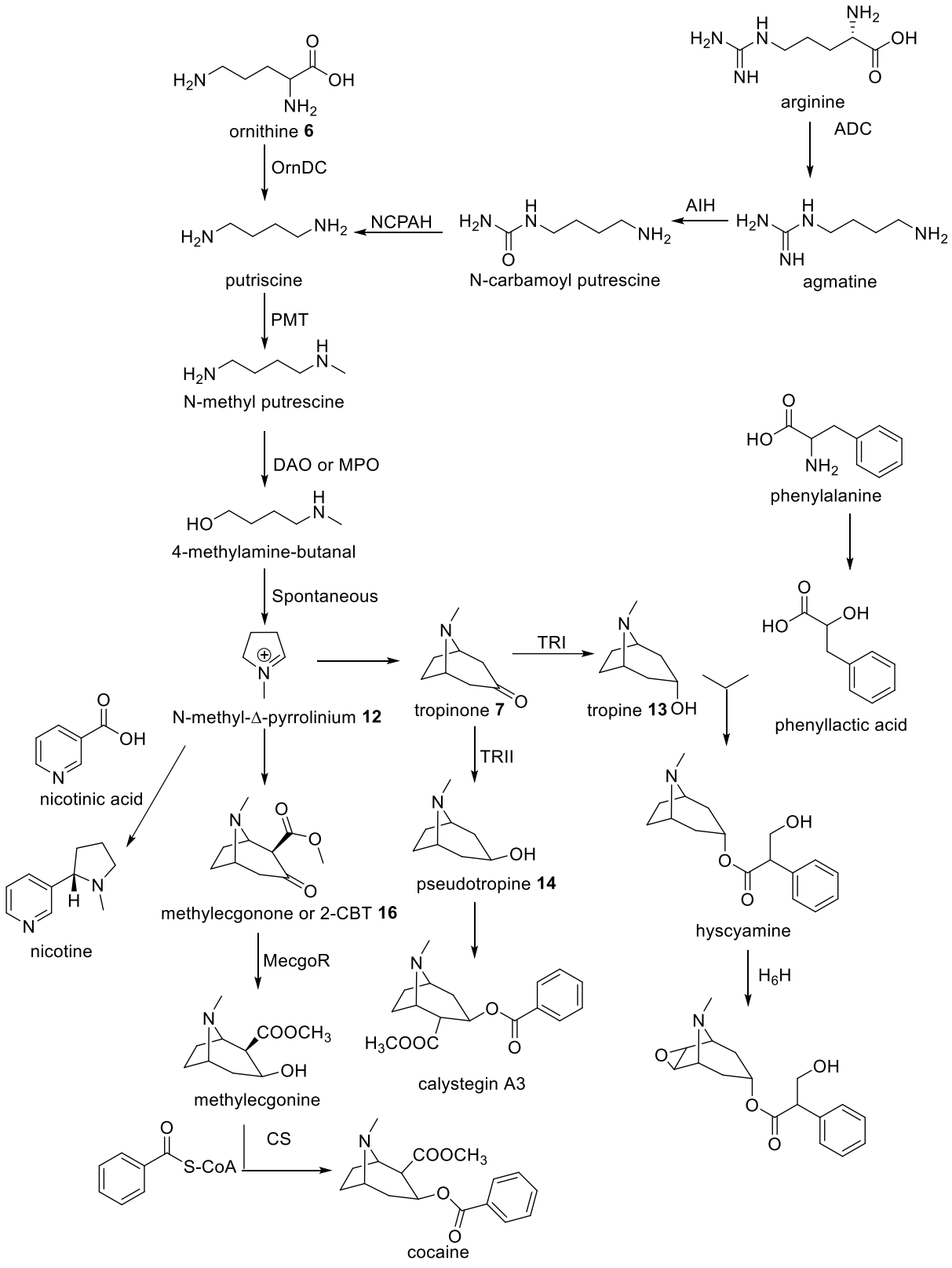


Figure 71: Biosynthetic pathway of selected tropane alkaloids.[61,74] **Abbreviations: ArgDC aginine decarboxylase; DAO diamine oxidase; H6H hyoscyamine 6-hydroxylase; OrnDC ornithine decarboxylase; PMT putrescine N-methyltransferase; TR tropinone reductase; ADC arginine**

decarboxylase; AIH agmatine deiminase; CS cocaine synthase; MPO N-methylputrescine oxidse; NCPAH N-cabamoylputrescine amidolase.

8.2 Reported TR substrate activities

Substrate	*B. candida x aurea* TRI[69]	*B. candida x aurea* TRII[69]
	Activity (%)	Activity (%)
(3)-Quinuclidinone	108	46
2-Fluoroethyl-N-nortropinone	30	56
3-Methylcyclohexanone	63	54
4-Ethylcyclohexanone	106	122
4-Methylcyclohexanone	69	77
7-Hydroxytropinone	3	0
***N*-Ethylnortropinone**	47	60
***N*-iso-propylnortropinone**	nd	22
***N*-Methyl-4-piperidone**	26	156
***N*-Propyl-4-piperidone**	0	133
Tropinone	100	100

Substrate	*W. coagulans* TRI[73]			
	Activity (%)	Km (µM)	Kcat (s-1)	Kcat/Km (sec^{-1} M^{-1})
1-Methyl-2-pyrolidone	0			
1-Methylpyrrolidine	5			
2-Pyrolidone	6			
3-Methylcyclohexanone	88			
4-Chloro-1-methylpiperidine	30			
4-Ethylcyclohexanone	38			
4-Methylcyclohexanone	32			
4-Piperidone	6			
6-Hydroxytropinone	31			
***N*-Methyl-4-piperidone**	25			
***N*-Methylpiperidine**	76			
***N*-Propyl-4-piperidone**	13			
Tetrahydrothiopyran-4-one	50			
Tropine		26.31	1.87	14000
Tropinone	100	1.451	16.74	11530

Substrate	*S. tuberosum* TRI[71]		*S. tuberosum* TRII[72]	
	Km (µM)	pH	Km (µM)	pH
Tropinone	21	6.4	33	6.4

Substrate	*H. Niger* TRI[70]		*H. Niger* TRII[70]	
	Km (µM)	Activity relative v_{max} (%)	Km (µM)	Activity relative v_{max} (%)
(3)-Quinuclidinone	1810	136	>1x10^5	nd
1-Methyl-2-pyrolidone		**nd**		**3**
2-CBT		**67**		**nd**
2-Methylcyclohexanone		**28**		**80**
2-Piperidone		**nd**		**nd**
2-Pyrolidone		**nd**		**nd**
3-Methylcyclohexanone	41	85, **40**	7580	172, **41**

4-Ethylcyclohexanone	30	107, **50**	534	108, **80**
4-Methylcyclohexanone	12	64, **50**	2030	113**, 60**
4-Piperidone	>1x10^4	nd, **7**	21600	1090, **182**
4-Tetrahydro-thiopyrone				
6-Hydroxytropinone		**13**		**27**
8-Thiiabicyclo[3.2.1.]-octan-3-one (TBON)	33	40, **33**	>4000	nd, **3**
Citronellal				
Hygrine		**nd**		**nd**
***N*-Methyl-4-piperidone**	231	13, **20**	770	512, 381
***N*-methyl-piperidone**		**nd**		**nd**
***N*-Propyl-4-piperidone**	0	0, **nd**	265	530, **388**
Pellitierene		**1**		**nd**
Tetrahydrothiopyran-4-one		**80**		**288**
Tropine-N-oxide		**100**		**3**
Tropinone	1010	100	34	100

Substrate	***A. Belladona* TRII**[67]		
	V_{max} (nkat mg^{-1})	V_{max} %	Km (90 uM)
(3)-Quinuclidinone	0	0	0
2-Fluoroethyl-N-nortropinone		46	
4-Tetrahydro-thiopyrone		71	380
6-Hydroxytropinone		25	
8-Thiiabicyclo[3.2.1.]-octan-3-one (TBON)	0	0	0
***N*-Methyl-4-piperidone**		313	650
***N*-Propyl-4-piperidone**		129	
Tropinone	17.5	100	90

Substrate	***A. thaliana* SDR**[64]		
	Vmax (nkact mg^{-1})	Km (uM)	Kcat.KM (sec^{-1} M^{-1})
(3)-Quinuclidinone	6.63	4.16	10.68
4-Methylcyclohexanone	105.63	0.3	0.049

Substrate	***C. officinalis* TR**[64,66]		
	Km (uM)	Kcat (s-1)	Kct/KM (sec^{-1} M^{-1})
(-)-Borneol	126.2	0.09	0.71
(-)-Camphor	0	0	0
(-)-Carvone	62.3	0.12	1.93
(-)-Fenchone	36.2	0.07	1.93
(-)-Methone	282.2	0.31	1.1
(-)-α-Thujone	65	0.46	7.08
(+)-Camphor	1038.1	0.06	0.06
(+)-Neomenthol	20.3	0.02	0.99
(3)-Quinuclidinone	2636		31.8
3-Methylcycloclohexanol	57.2	0.17	2.97
3-Methylcyclohexanone	13 (24.1 at pH 5)	2.34 at pH 5	5171 (97.1 at pH 5)
4-Methylcycloclohexanol	29.2	0.17	5.82
4-Methylcyclohexanone	15 (5.1 at pH 5)	(3.54 at pH 5)	95939 (694.12at pH 5)
8-Thiabicyclo[3.2.1.]-octan-3-one (TBON)	42		1096

Citronellal	app 71.4		app 0.16	2.24
Dimethlpentan-3-one		76.9	0.57	7.41
Fructose				
Gluconolactone				
Glucose				
Nerol		26	0.06	2.31
***N*-Methyl-4-piperidone**		7583		2374
Nortopinone		2002		8.9
Nortpseudopellititerene				
***N*-Propyl-4-piperidone**		1342		510
Pellitierene				
Pentananal	app 74.1		app 0.16	2.16
Tropinone		3290		7.3

8.3 Sequences for selected SDRs from the tongue metagenome

>1
MNLLANKVAIITGAGRGIGRAIALKYAQEGASVVITDLKIDETVEAFVKELEGLGVKAKAYASNAANFEDAHKLVE
AVVADFGRIDVLVNNAGITRDGLMMRMTEEQWDLVINVNLKSAFNLIHAVTPVMVKQRSGSIINMASVVGVS
GNAGQANYSASKAGMIGLAKSIAKELGARGIRANAIAPGFIITDMTGALSEEVRKQWEVQIPLRRGGTPEDVAN
VATFLASDLSSYVSGQTIHVCGGMNM

>2
MSTQDLSGKIALVTGASRGIGAAIADTLAVAGAKVIGTATSESGAAAISERLAQWGGEGRALNSAEPETIENLI
ADIEKEFGKLDILVNNAGITRDNLLMRMKEEEWDDIMQVNLKSVFRASKAVLRGMMKQRAGRIINITSVVGV
MGNAGQTNYAAAKAGLIGFSKSMAREVGSRGITVNCVAPGFIDTDMTRALPEETRKTFEAQTSLGKFGEAQDI
ADAVLFLASDQAKYITGQTLHVNGGMLMPPWLISXPEFQHT

>3
MAHNIFVTGATSGIGLCIAEAYAKHGDNVLISGRRAELLGEVQARLSKEYGVRVETLVLDVRSREDVESKVPAAIE
AFGGVDVLVNNAGLAQGLDPFQDSAVDDAVTMIDTNVKGLLYVTKAVLPFMIDKNEGHIVNMGSTAGIYAYP
NGAVYCATKAAVKTLSDGIRMDTITTDIKVTTIQPGIVETPFSEVRFHGDAERAKSVYAGIDAIQPEDVADVVLYV
TNQPKRLQISDVTIMANQQAAGFMV

>4
MSKVAIVTGAGQGIGFAIAKRLVQDGFKVGVLDYNAETAEKAVAELSADKAFAVVADVSKQAEVAAAFQKVVD
HFGDLNVVVNNAGVAPTTPLDTITEEQFNRTFAINVGGVIWGAQAAQAQFKALGHGGKIINATSQAGVVGNP
NLTVYGGTKFAVRGITQTLARDLADSGITVNAYAPGIVKTPMMYDIAHEVGKNAGKDDEWGMQTFAKDITLKR
LSEPEDVAAAVSFLAGPDSNYITGQTIIVDGGMQFH

>5
MSETILVTGASAGFGQAICRRLVADGYRVIGSARRIDKLQALQEELGEAFYPLQMDVTDLSQVDHALASLLKAW
EKVDVLVNNAGLALGLAPAYEAEVADWLTMIQTNIVGLTYLTRKILPQMVERNDGYIINLGSTAGTVPYPGANV
YGASKAFVKQFSLNLRADLAGKKIRVSNIEPGLCEGTEFSSVRFKGDEKRVEALYRDAHAIQSEDIANTVAWLIQQ
PKHVNVNRIEIMPVSQTFGPQPVYR

>6
MYSELKGKVAVITGGSKGIGTAIAKRFGQEGMKVVINYNSDAAGAELAAEAVRCAGGEAVTVKAHVGTEEGVQ
SLVDAAVENYSGIDVWINNAGMENKVATHEMPLSDWERVINVNLTGVFLGTRAALTYFMEHDVKGSIVNMSS
VHEQIPWPTFAHYCASKGGIKLFTQTVAMEYAKYGIRVNAVGPGAINTPINAKKFSDPVQYEETMSMVPMKRI
GKPEEVAACVAWLASDEASYVTGITLFVDGGMTLYPAFQDGKG

>7
MPHKNDVQVALISGGTSGIGFATAKLLLQEGWCVVINGRDEQAGQRAKMKLRRYSSKVRYVKGDVSSVSDCQ
RIVKETVDFFGSISALVTAAGYYEEELLADVSESAFDEMFGTNVKGTVFLCQAALPYLRQVKGSIVTVSSDAGLQG
NVACSVYGASKGAIVSFTKSLSLEMAPHEVRVNCVCPGDVDTSLVDKQIAQSHQDAEQAKEEMGQHYPLGRIA
KPHEIGEVIAFLISSKASFVTGAAWTIDGGLTS

>8
MTKRVLVTGVSSGIGLAQARLFLENGYQVYGVDQGEKPDLQGNFHFLQRDLTLDLEPIFDWCPQVDILCNTAGI
LDDYKSLLEQSAQEIQEIFEINYVTPVELTRYYLTQMLEYKRGTIINMCSIASSLAGGGGHAYTSSKHALAGFTKQL
ALDYAEAGIQVFGIAPGAVKTGMTAADFEPGGLADWVASETPIKRWIEPEEVAEVSLFLASGKASAMQGQILTI
DGGWSL

>9
MSWVIVTGANGGIGEATVHQLIKNGYSVFAADLAEQPIASFGTYGDAIFRYRAVDVTSEESVTALAEAVAALDE
PITGAVLAAGIAHSQPLLETSFATWKRLHAVNSDGVFLCLREFARIMIDQQESDPTNSRSLVTVASNAARVPRAE
FGAYGASKASAARVSSSFGLQLAAHGIRVNSVCPGTTRTPMVTDAWEGEDRSALPVAGNPQTFRLGIPLGRIAD
PADIAAVNAFLISEAARHITMQEIVA

>10
MTENTHIPASPFLPHTGDGKVAVVTGASSGIGRATVQQLVASGWTVYALARRTDRLYTLYAETGAHPVTCDVT
DEQSVQFVAEQILEEQGTIDALVNIAGGAIGVDKVAEGKPDDYLKMYQMNVLGILHMVRAFAEALRQNGYGTI
LNLTSTAAEHGYEGGAGYNAAKFGARGLTEALRLEEAENNIRVIEICPGMVHTEEFSLNRLGSKEAAERVYAGVE
KPLTAEDVAQTVTFALNVPHHVNLDRITIRPVAQPSQFKVIRKEGP

>11
MPKSLQRPVVLLTGASSGIGYDVAPLLVRYGYIVYGAARRVEKIEELASEGVKALSLDVTDEASMEAAVQQIIDAE
GRIDVLINNAGYGSYGAIEDVPIDEARRQFEVNLFGLARLTQLVLPHMRARGSGRILNISSMAGRITSPLGAWYH

ATKYALEAFSDALRMEVEEFGIDVVIIEPGGIKTPWGLIAADHLEESSRNGVYAAQAQRVAANMRRLYSPSSNLS
EPKVISNAILRALEARRPKTRYLVGFGAKPSVFLHTVLPDRLFDKVARRIF

>12

MSNAPEVRGLFLKALGRPVIVAPSSAEPTVTFDGPLTEVCSCSLKETELPVVVR
AGEETFEVRATSTGERTINGRVALVTGGAQGFGAEIARGLVDAGCFVFIADLNGEGAAAKAAELGGEGVAHPIT
VNVADEESVAAMAAEIERVTGGLDLVVSNAGIVRAGSVLEQDASAFRLSTDINYVAFFLVTKHLGQLLARQHST
APEWLTDIIQINSKSGLVGSNKNAAYAGSKFGGIGLVQSFALEMVAHGVKVNAICPGNFYDGPLWSDPDRGLF
VQYLNSGKVPGAKTVADVKEFYEAKVPMRRGAQGIDVLRAIFYIVEQEYETGQAVPVTGGQVMLS

>13

MNTSRRVVVTGASTGIGQATARLLAKRGWKVVAVARRRERLEALAEQIGCEYWAADLTDEAQVKEMAAHVLE
GGPVDAVVNNAGGAIGVDRVAEGDPARWSAMFERNVLTALHCSRAFLPGMRERGGDLVFLTSTAAHDTYPG
GGGYVAAKHAERIIANTLRQELVGEPVRVIEIAPGMVRTEEFSLNRLGSQEAADRVYEGVAAPLVAEDIAEAIVW
TLERPSHVNIDSMIVRPVAQATNTLVARKTA

>15

MDIADAEGVKKTVAQILEECGDVNVLINCAGRISSVPFTEVDDKEWNNTINTNLTGTYNVTHALWQHFIDRGG
ARIVNVSSVAGKIGGGLLGTVAYASSKAGMNGFTKAIAKEGGKYGISCNAVCPSFTITDMTTALSNDEEKYKKVV
GIIPLGRPAQASEPAQMVLFFASDAASFVNGEVGDCDGGIVME

>16

MSENTPKTVLVTGGNRGIGYEIAKEFQAAGHNVCITYRSGEAPEEFFAVKADVRDADSINEAFKEIEAEFGPVEV
LVANAGITRDMLLMRMKESDFTDVVDTNLTGSFRVVQRAIKGMLKLKRGRIILVSSVVGLYGSPGQVNYSASKA
ALVGMARSITRELGGRNITANVVAPGFINTAMTEVLPEETKKNYLASIPAGRFAXAEEVARVIRWLASDEASYISG
AVIPVDGGLGMGH

>17

MELKNKNVFVTGSTRGIGLAVAHKFASLGANVVLNGRSEISEDLLAQFADYGVTVVGISGDISNGEDAQRMVA
EAIEKLGSVDVLVNNAGITNDKLMLKMTEEDFERVLKINLTGAFNMTQAVLKPMSKARQGAIINMSSVVGLMG
NIGQANYAASKAGLIGFTKSVAREVAARGVRVNAIAPGFIESDMTDAIPEKMKDAMLAQVPMKRIGQAEEVAE
VAAFLAGQEYLTGQTIAIDGGMTMQ

>18

MAILITGASAGFGAAMCRTFVAAGYHVIGAARREDKLQQLAEELGEQFYPLEMDVSRTESIQNALNSLPEHLSEI
DCLINNAGLALGLDTADKADFGDWETMIQTNIIGLTFLTRQILPQMVARKQGYIINLGSIAGSYAYSGSNVYGAT
KAFVRQFSMNLRAELADKNIRITNIEPGLCGDTEFSNVRFKGDDQRAAEVYENVEFIQPQDIADTALWLYQRPA
RMNVNSIEIMPVAQTFAGMKVYRDEPAPAKEETFEKQSMSLFGKIKSWFK

>19

MNDWLNIKGKTVLVTGASSGIGKAIVEELLELGVNVANFDLSDNDLRHPNLLFVKVDVTSRSEVEEGVAKIVERF
GNIDAVVNNAGINIPRLLIDAENPKGPYELDDETFEKVTMINQKGLYLVSQAVGRILVKNGKGVIVNMASEAGLE
GSEGQSAYAATKAAVYSYTRSWAKELGKHGVRVVGIAPGIMEATGLRTLSYEEALAYTRGKTVEDIRAGYASTST
TPLGRSGKLREVADLVAFYISDRSSYITGVTTNIAGXKTRG

>20

MYLCKQIYHTDKHRKTNMAKVLVTGANKGIGYGICKFLGKSGWQIIVGARNSERAEEAMKSLKAEGVDVIGW
QYVNLSDNASLEQTAKEVKEKYHDLELLVNNAGIPGDMEVASYESELKDVIDTVQVNYVGTFCLTKALTPLLSAN
KGRIVNITVPSEVSPYWHPMAYVASKAAQNAMTSIMAMEFEKNNIPVEIFNIHPGATTTDLNNHYTGPGSHSI
DVVSEKIAEVINDGGKHQGEFVELYPI

>21

MKLLEGKTALITGAARGIGKAIALKFAEEGANIAFTDLVIDENGKATEAEIAAKGVKVKGYASNAADFAQSEEVV
KQVKEEFGSIDILVNNAGITKDGLMLRMTEQQWDAVIAVNLKSAFNFIHACVPVMMRQRNGSIINMASVVGV
HGNGGQANYAASKAGMIALAKSVAQEMGPKGIRANAIAPGFIDTAMTQALNDDIRKEWTSKIPLRRGGTVDD
IANTAVYLGSELSSYVSGQVIQVDGGMNM

>22

MMTSLRKKIIAIGGALAYPSISLTPDITREHFSGKWVLVTGASHGIGRALTEKIINAGANVFLIARSEADLRLLC
AKAKQMGSSADYCAIDLRDREKLEQLCQKLRETLPRLDYFFCNAGKSIHRKINDAQDRLHDYDRTMDLNYRSLV
ALSLAILPALKASKGSIIYSSSVSTLYPMAPGWSAYHASKSAANTWCETANSEFAPLGVHVQIAYLPLVHTAMSD
VNEQYKHLPAYTPADAANILLKLAIRKVRTYKPWWAKLSAPIAYLFAPIIHLYYKR

>23

MKKAIVVGASSGIGHEVARLLIAQGWAVGVAARRIDKLTDLQAMAPERVYTAQIDVNNEDAETSLLQLIERMN
GIDLYFHAAGIGWQNPSLNADIELKTMETNAVGFTRMIGCAYRYFANKGGGHIACITSIAGTKGLGPAPAYSAT
KAMQNTYLQALEQLAACKHHNIHFTDIRPGFVDTPLLAGTSHLPMLMTTEKVARSIIKAINSRRHICVIDSRWCV
LTY

>24
MKMEDTIKKGIIIGATGGIGRQLAKELAQRLEHLILVSRDADKLSQVQKELTGSKAQLSILTLDMLDQVALEAFVE
SLDADLLVNCAGLAYFSRESDLDSASEQDLWQVNYHSSVQLIKQVVQKNQKIQLVQLSSLAALFPHPYLAAYSA
SKAALQTFTLALQEELRQSESQVQLGLYILGPVQTAIFPPKLVEALGGSRLQMKSEKVAQQLIRFIERDTSYEIIGLR
YRLLVWLGRLLPQRWIIRLLAIYLKKGT
>25
MSYNLLKGKRGIIFGALNEQSIAWKVAERAVEEGASITLSNTPIAVRMGTVNSLAEKLNCEVIAADATSVEDLENF
KRSMEILGGKIDFVLHSIGMSPNVRKHRTYDDLDYNMLNTTLDISAVSFHKMIQSAKKLDAISDYGSILALSYIAA
QRTFFGYNDMADAKALLESIARSFGYIYGREKNVRINTISQSPTMTTAGQGVKGMDKLYDFANRMAPLGNASA
AECADYCIVMFSDLTKKVTMQNLYHDGGFSNIGMSLRAMTTYEKGIGDEYKDENGKIIYG
>26
MSETILITGSNRGIGKAVALGLAQDGFDIVVHCRSRRDEAEAVAEEIRALGRNARVLQFDVSDREACREILTADI
EANGTYYGVVLNAGLTRDNAFPAFTDDDWDLVLRTNLDGFYNVLHPLTMPMIRRRKAGRIVCMASVSGLTGN
RGQVNYSASKAGLIGAAKALAVELAKRKITVNCVAPGLIDTDIIDENVPVEEILKAVPAARMGLPEEVAHAVRFL
MDEKAAYITRQVIAVNGGLC
>27
MAYNLLKGKRGVIFGALNDMSIAWKVAERCAEEGATLVLSNTEMALRMGSLDELSKKINAPVIAADATNYDDL
ENVFVKAQELLGGKIDFVLHSIGMSPNVRKHRTYDDLDYNWLNKTLDISAISFHKMLQAAKKVDAIAEYGSVVA
LTYVA
SHRTFFGYNDMADAKSLLESIARSFGYIYGREKNVRINTISQSPTETTAGKGIKDIDNMMDFADKMSPLGNATA
DECADYCVTLFSDLTRKVTMQTLFHDGGFSNMGMSLRAMNQYSKTLD
>28
MNIAIITGASRGIGKAAAKRFAREGYSLLLNCEKNWTLLEELKKEIQSDLPENCPEIFLCKDLGTKKGLSRILEGKSL
SKLILIANSGKDAIKLLQDCREEETKALLETNLLQPFILCQKLLPYLLQAEEGRILFSSSVWGNVGASMESLYSLTKG
GISTFAKALGKELAPSHISVNAVAFGAVDTDMNSWLSTEEKQSLEEGIPYGRMATVEEAADFLYLLSQAPLYLTA
QVIPFDGGWI
>29
MNRKIFVTGATSGIGLECARAFAQDGDNVLIAGRRADRLAAIKEDFEQQYGIRVDTLVLDVSKREDVDAKVKPAI
EAFGGIDVLVNNAGLAQGLDPFQDSTVEDAVTMINTNVLGLLYVTKAVLPFMMAQNSGHIVNMGSTAGIYAY
PGGAVYCATKAAVKMLADGIRMDTIATDIKVTTIQPGIVETPFSEVRFHGDAERAASVYAGIEAVQPEDVADVV
LYVTNQPKRLQISDVTIMANQQAAGFM
>30
MSYNLLAGKKGIIFGALNEQSIAWRVAERAVEEGAEIILTNTAVAVRMGQLNELGQKLNAKVVPADATKEEELE
VVFQEAMKSFGKVDFVLHSIGMSPNVRKGRAYDDLDYKMLQTTFDISAVSFHKMLQVAKKLDAIAEGGSVVAL
TYIAAQRTFVGYNDMADAKSLLESIARSFGYIYGRDKGVRINTVSQSPTVTTAGSGVKGMSDLLDFAEDLSPLGN
ADANDCADYCITLFSDLTRKITMQNLFNDGGFSSMGMSAAAIEAFATGRANREK
>31
MTKHLRRAVVTGASSGIGWAITERLVTEGWQVVGISRTGQVPEGALSVSADLAEDGVGEAITQAQNLLGGVD
AYVGAAGSTYEQLAARADLEQVNTQLRLHYLSNYEAISLLLPGMVRARWGRIVLLSSVVAQSGMAGLSAYGAA
KGALEALVKSLALEVGRRAITVNAVAPGYIQTPMTQSLSPRQQERYLQRTGAARPGTPQDVAGPVAFLLSDDAA
YVNGQILHVDGAMGVGNP
>32
MNIFITGGTSGIGLALARFYAAKGHRVGVCGRNTARIDKSDEVNKLLLAYQLDVCDKDALTVAVEVFCADKGLD
MMIVAAGYYRNGVTEEVDFEQTSQMLKVNIAGALNAMEVAREAMNASGGHLVVIASVAGLLHYPCASVYAK
CKRALIQIADAYRRSLADYQITVTTLVPGYIDTPRLREIYRNDLSKCPFCMPLNRAVETMTKAIAQRKEQVVFPPK
MRLSIAILSLLPTCLLSAFMHRKTLWSI
>33
MKYALITGASRGIGRSVALLLAERYSIIINYQSNAEAAQAVKQEIETKGGHVELLPFDVSDPKAIEAAIDTWEASHP
DEFISVLVNNAGIRRDNVMFMMSDEDWHSVLDTNMNGFFYITRRLLKHMMPRKRGGRIINMASLSGLKGLP
GQVNYSAAKAALIGATKALAQEVAARKITVNAVAPGFIQTDMTKELPEDELKKLVPVGRFGTPEEVADVVAFLA
SDAAAYITGEVINVNGGFY
>34
MRKTALITGATSGIGEACARKFAQGGYDVIITGRNKQRLAALKVELETGETKVLALAFDVRNRAAATKAIKSLPAE
WAKIDVLINNAGLALGLEPEYEGDFEDWDTMIDTNIKGLLTMTRLIVPKMVERNSGHIINIGSVAGDAAYAGGN
VYCATKAAVKTITDGLRIDVAHTAVRVTNVKPGLVETHFSNVRFHGDDKRANSVYHGIKPLTGTDIADVAYYAAS
APAHVQIAEVLVLATHQGSGSVIH
>35

MGFLTGKRILVAGLASNRSIAYGIAKAMKEQGAELAFTYLNDKLQPRVEEFAKEFGSDIVLPLDVATDESIQNCFA
ELSKRWEKFDGFVHAIAFAPGDQLDGDYVNAATREGYRIAHDISAFSFVAMAQAARPYLNPNAALLTLSYLGAE
RAIPNYNVMCLAKASLEAATRVMAADLGKEGIRVNAISAGPIRTLAASGIKNFKKMFSAFEKTAALRRTVTIEDV
GNSAAFLCSDLASGITGEIVHVDAGFSITAMGELGE

>36

MAKVMVTGANKGIGYGICKFLGKSGWQVIVGARNSERAEEAMKSLKAEGVDVIGWQYVNLSDNASLEQTAK
EVKEKYHDLELLVNNAGIPGDMKVASYESELKDVIDTVQVNYVGTFCLTKALTPLLSANKGRIVNITVPSEVSPYW
HPMAYVASKAAQNAMTSIMAMEFEKNNIPVEIFNIHPGATTTDLNNHYTGPGSHSIDVVSEKIAEVINDGKKH
QGEFVELYPIVDEGR

>37

MQRTALVTGATAGFGAAICRTLIENGYRVIGTGRRVARLEQLQQELGENFHFLAFDISDRQATEDAFHSLPTNW
QSIDLLVNNAGLGLGLESADKASLDDWMQMIDTNIKGLVTITRLVLPQMVERNSGHIINLGSIAGTYPYPGGNV
YGGTKAFIKQFSLNLRADLAGTQIRVTNVEPGLCGGTEFSNIRFKGDDARAKKLYENVEYVSPQDIANIVLWLNQ
QPEHVNINRIEVMPTAQTFAPLNVAR

>38

MHLEGKVALVTGASRGIGRAVAIQLAQSGADVAVNYSGSEAAAQETVDAILALGRKAIKIKANVANAEEVAAM
VEETHKTFGHIDILVNNAGITRDGLLMRMKDEDFDAVIDINLKGVYLVTKAVSKIMMKQRAGHIINMTSVVGL
MGNAGQANYAASKAGVIGFTKSCAKELASRGITVNAIAPGFINTDMTDVLPEKVKEAMVTQIPLGRMAKAEEV
AAVTTFLASDFASYITGQVINVDGGM

>Lactobacillus_Brevis

MSNRLDGKVAIITGGTLGIGLAIATKFVEEGAKVMITGRHSDVGEKAAKSVGTPDQIQFFQHDSSDEDGWTKLF
DATEKAFGPVSTLVNNAGIAVNKSVEETTTAEWRKLLAVNLDGVFFGTRLGIQRMKNKGLGASIINMSSIEGFV
GDPSLGAYNASKGAVRIMSKSAALDCALKDYDVRVNTVHPGYIKTPLVDDLPGAEEAMSQRTKTPMGHIGEPN
DIAYICVYLASNESKFATGSEFVVDGGYTAQ

>Lactobacillus_Kefiri

MTDRLKGKVAIVTGGTLGIGLAIADKFVEEGAKVVITGRHADVGEKAAKSIGGTDVIRFVQHDASDEAGWTKLF
DTTEEAFGPVTTVVNNAGIAVSKSVEDTTTEEWRKLLSVNLDGVFFGTRLGIQRMKNKGLGASIINMSSIEGFVG
DPTLGAYNASKGAVRIMSKSAALDCALKDYDVRVNTVHPGYIKTPLVDDLEGAEEMMSQRTKTPMGHIGEPN
DIAWICVYLASDESKFATGAEFVVDGGYTAQ

>Lactobacillus_Minor

MTDRLKGKVAIVTGGTLGIGLAIADKFVEEGAKVVITGRHADVGEKAARSIGGTDVIRFVQHDASDETGWTKLF
DTTEEAFGPVTTVVNNAGIAVSKSVEDTTTEEWRKLLSVNLDGVFFGTRLGIQRMKNKGLGASIINMSSIEGFVG
DPALGAYNASKGAVRIMSKSAALDCALKDYDVRVNTVHPGYIKTPLVDDLEGAEEMMSQRTKTPMGHIGEPN
DIAWICVYLASDESKFATGAEFVVDGGYTAQ

>Weissella_Thailandensis

MSERLKGKVAIVTGGSVGIGWSIANRFVQEGAKVVITGRRAQVGEDAAKKIGTPDVIRFIQHDSSDEDGWVDL
FDQTEKAFGPVSTLVNNAGIAVDNSIENTTTEEWRKVMSVNLDGVFFGTRLAIQRMKNKDLGASIINMSSIEGF
VGDSNLGAYNASKGAVRLMSKSAAVDCALKDYDIRINSVHPGYIKTPMVENADGMVEIMSQRSKTPMGHLGD
PDDIAYMCVFLASNESKFATGSEFVVDGGYTAQ

Table 27: SDR proteins selected from the tongue metagenome. 14 was not successfully cloned, and was not included. The final four proteins were used to create the metagenome interrogation.

8.4 Sequences of selected AKRs from the drain metagenome

>contig-120_19664_1
MGCNLRMFLMNVVEANGAKIPGIGLGTWELHGRTCARIVEQALRLGYRHIDTAQIYENEREVGEGVRASGVKR
EIFLTTKVWTTHFAPNDLERSTKESLARLRLTEVDLLLLHWPNPQVPLVETLGALARMKQQGLAKHIGVSNFTVLI
EEAVAVCPEPLACDQVEYHPYLEQTVVREACARHGMALVAYSPVARGRIKNDRALLRIGDRYRKTAAQVCLWL
VQQNVAAIPRTSKLERLSENIEIFDFELSDEDMQEISGMGSATGRLTDYSFAPKWD

>contig-120_6751_1
MWQVDAGVTAKVVGWGIEAGYRSIDTAEGYRNEEGVGEAIRAASVPREEIFITSKLRNSAHGRDDALREFDAT
RKLGLEQLDLFLIHWPLPGQDKYVEAWKTLIELRDAGRIRSIGVSNFEQDHLERIIGETGVTPAVNQIELHPRFQR
DKRAFHDKHGIRIESWSPLGSGRLLDDAGIGAIAKKHGKSIAQTIIRWHLQEGLIVIPKSVNRDRIVANFDVFDEL
DGGDMQRIAAMDKPSGRTGPNPETASFMF

>contig-120_28243_1
MSEQPTIPLNDGKAIPQIGLGVWQTPNDTSVIAVEAALRNGYRHIDTAAIYGNEEGVGEGLRKSGIAREQVYLTT
KLWNDDQGYDAALRACELSLKKLGTDYVDLYLIHWPSPHRNLYLESWKALIKLQQEGKARSIGVSNFGADQLER
ILGETGVKPVLNQVELHPRFQQRELREAHARLGIATQSWSPLGQGQILQDPVVQGIAAKHGKSTAQVLIRWHL
DHKFIVIPKSVTPLRIAENFAVFDFVLDENDRDLLDAMDSAEGRIGPNPVTAT

>contig-120_19634_1
MEYLTLSHTDLRVSRLIAGCMGLGGGWGRGTVIDARIEKQARDFVDAAFELGINFFDHANIYAHGRAEEVFGRI
FKQMPSLRDRMVLQSKVGIRWEGDPAGTPQRFDFSYEHIVEAVDAILKRLNTSHLDILLLHRPDPLMEGEEVAR
AFARLKKEGKVRYFGVSNQNRFTMEYLQHFLPDPLIANQLEMNLLHSGFAEATVSFNQAAPAYPDGWEGVVE
YCRLKGVSLQAWSPLARGLLTGGDLSDVAPDVAKTAALVQELAEAYAVSAEAIVLAWLMRHPAGIMPVLGTSR
ADRLRACAESVKVTLTRDEWYRLFEAARGARMP

>contig-120_8191_2
MKSYRLADDLTVSRIGYGCMHLSRAWDASPITTEERRNAQRLVETALAHGITLFDHADIYARGKSEQVFGEVLR
ASPGLRAKMVLQSKCGIRFADDPPGAPQRYDFSHAHIVASVEGSLARLGVDHLDILLLHRPDPLVEPDEVARAFD
ALHASGKVRHFGVSNHTRGQIDLLRRHVRQPLVAHQVEISLLHLPLIDDGVVANTTGHAYASAAGTLDYCRLHD
IRVQAWSPLASGKLASTSEFSDPAVRDTATLLRQLADAKGVTPEAIQLAWLLRHPAGIQPIVGTTDPVRLVACAA
ADEVPLSREEWYALFTAARGGRVP

>contig-120_9148_1
MKTYTLADGLTASRIAYGCMQLSRAWDATPVTADERRQAQRLMETALAGGITLFDHADIYARGKSERIFGDVL
RASPGLRERMVLQSKCGIRFADDPPGTPGRYDFSHAHIVTSVDGSLARLGVDRLDVLLLHRPDALVEPEEVARAF
DDLHAAGKVRHFGVSNQTAGQIELLRRHVRQPLVANQVEVSLLHHHLIDDGVVANTTGHVYASAAGTLDYCRL
HDIRVQAWSPLAGGKLATTSEFADPVIRTTSTLLRQLAEEKGVTPEAIQLAWLLRHPAGIQPIVGTTDPVRLAACC
AADGISLSREEWYALFTVARGDRVP

>contig-120_562_6
MRLGLGTMSLCGANGWGEPPQFEHMQQFLQLAYEAGIRCFDSAGFYGPDVALRLLKPLFAQDRELIISSKVGL
QREAINTWRVDASAAAIRQQVEHDLRTLGTDCLPLVFLRLGDGKFLPKDPTPLQESLMALVTLQSEGKLRAIGLS
ECTLTDLEQALAITPIAAIQNRYNLRQCDDSVLAWATAQHCPYWAYSPLASGLLLKRVPPQLAHMAQKYQATP
AQIACAWLLARSPMLMPIVGSQNPQHLLEHTKAVSIQMSAVDRAVLEQLLDLDN

>contig-120_3698_1
MTWGEQNSLEEGFAQMDAAIEAGINFFDTAEMYAVPPTAKTYGSTETIIGEWFRARPGRREQVLLATKVTGRS
ETMAYIRGNPRLNRQHIEEAVNASLQRLQTDHIDLYQLHWPDRPTNFFGQLGYRHQEGAETVPIAETQAVLAD
LVKAGKIRHFGLSNETPWGVMSYLHHAETTGLPRPVSIQNPYNLLNRSFEIGLAEIAHREDVGLLAYSPLAFGALT
GKYLDGQWPEGARITLFKRFSRYMNPQAEAAIAAYVKLARDSGIPPAQLALAFINQQPFVTSNIIGATSLAQLQE
NIASVQVKLDDDLLAAIEAIHVQHPNPSP

>contig-120_3273_2
MEYRQLGSSGLRVPALSLGTGTFGGQGPLFSAWGTSDAAEARRLIDISLDAGVTLFDTADVYSNGASEAILGEAI
KGRRDAVLISTKTGLPMGDGPHDWGASRARLIGAVEASLRRLGTDHIDLLQLHAFDASTPVDELMDTLATLIAA
GKLRYAGVSNYPGWQLMKAQAAADRLGVPRFVAHQVYYSLIGRAYEADLMPLAVDQGIGALVWSPLGWGRL
TGKIGRGRPVPAGSRLHETEQFAPPVAEELLYRVIDALEAAAAETGKTVPQVAINWLLRRPTVSSVIIGARNEEQL
RQNLGAVGWELTAEQVAVLDGASDVLPPYPHTPYRQQAGFARLNPPLV

>contig-120_16375_1
METRFLGRSGLKVSALGFGAGTFGGKGPLFSAWGNTDVTEARRMIDLSLDAGLTFFDTADVYSDGASESILGEA
LRGRRDRVILSSKTGLRLGEGANDAGTSRLRLLKAVDASLKRLGTDYLDLLQLHAFDAMTPIEETLSTLDDLVRAG
KVRYLGVSNFSGWQLMKSLGIADRHGWSRYVANQTHYSLIGRDYEWELMPLGIDQGVGAVVWSPLGWGRL

TGKLRRGQPLPATSRLHDTADVGPPVAQERLFRIVDALDIIAEETGKSIPQIAINWLLQRPTVSTVLIGARNEEQLR
QNLGAIGWQLTADQIARLDAASAMEAPYPYYPYWRGQFSERSPLPVQAFPRLPEPPHESRRPDPLPAHRRSAIT
AGCRTRNAYRHRSRHPRARGSRVGEPGRHEGARTEAAYRGAT

>contig-120_3402_3
MEYRYLGRSGFRVPALSFGTGTFGGSGALFSAWGSTDVAQARRLVDICMEAGLNMFDSADVYSKGAAEEILG
QAIKGRPRDSLLVSTKATFRFGDGENQVGSSRYHLINAVDAALRRLGTDYIDLFQLHGFDARPPVEEVLSTLDNL
VRAGKIRYLGVSNFSGWHLMKSLATADRYGWTRYVAHQAYYSLLGRDYEWELMPLAEDQGVGAVVWSPLG
WGRLTGKIRRGQPLPESSRLQSQTAVDSGPPVDVEHVYKVVDALDEIAAQTGRSVPQVALNWLLQRPTVSTVV
IGARNEDQLRQNLGAVGWNLSAEQVAKLDAASAVTKPYPYWHQAGFKYRNPTPV

>contig-120_982_2
MEYRYLGASGFRVPVLSFGTGTFGGQGALFSAWGSTDVAEARRLVDICLDAGLTMFDSADVYSKGAAEEILGEA
IKDRPRDALIISTKATFRFGDGENQVGSSRHHLINAVDAALKRLGTDYIDLFQLHGFDARPPVEEVLSTLDTLVKA
GKIRYLGVSNFSGWHLMKSLAAADRHGWTRYVAHQAYYSLVGRDYEWELMPLAADQGVGAVVWSPLGWG
RLTGKIRRGQPLPENSRLQSQTAHDAGPKVDLEYLYDVVDALDEVAQETGKTIPQVALNWLLQRPTVSTVVIGA
RNEAQLRQNLGAVGWNLTPEQVAKLDKASQRPKPYPYWHQAGFAYRNPTPV

>contig-120_1078_6
MEYRYLGASGFRVPVLSFGTGTFGGQGALFSAWGSTDVAEARRLVDICLDAGLTMFDSADVYSKGAAEEILGEA
IKGKPRDALIISTKATFRFGDGENQVGSSRHHLINAVDAALKRLGTDYIDLFQLHGFDARPPVEEVLSTLDTLVKA
GKIRYLGVSNFSGWHLMKSLAAADRHGWTRYVAHQAYYSLVGRDYEHELMPLAADQGVGAVVWSPLGWGR
LTGKIRRGQPLPETSRLQLQTAQDAGPQVDIDYLYDVVDALDEIAQETGKSIPQVALNWLLQRPTVSTVVIGARN
EDQLKQNLGAVGWNLTAEQVAKLDAASQRPKPYPYWHQAGFAYRNPAPV

>contig-120_33576_1
MATDGTKQGASRRYIMSAVEASLARLKTDYIDLYQQHDYDPLTPMEETLRALDDLVRQGKVRYIGNSNFPAWR
IAEAEMMARQMNVNRFVSCQDEYSLVVRDIEKDLQPAATEYRLGLLPFFPLASGLLTGKYQRGTSAPADTRFGK
MPALKDRYATARNEDIVERLQAFAKARGHTLLELAFSWLAARPQVSSVIAGATRVEQVEQNVKAIDWNLSPEE
MAEIDQITK

>contig-120_49236_1
MGEVIAGRRDEVFLVSKIRPENAGEMTMMLHAEKSLERLGVDRIDLYLLHWESRFPLEEIVAGFEELIDEGMIAR
WGVSNLDLRAMNALTEVEGGEHCAANQLLYNLGSRGIEFDLLPWQQEHLQPVMAYSPLGRGGLLEHPALLAV
ADRHDASPAQIALAAVLRHDGVIAIPKASTVAHVEANAAALEIELDPEDIEVLDRAFPPPTTARPLDII

>contig-120_1656_3
MENQTFAVAADGVDPSRVPYKTLYTGARIPSIGIGTFGSDRFSGEEVADAVLGAISVGYRHVDCASVYSNEHLIG
ESLQKAVAGGIKREDLFITSKVWNNMHGDGDILLSVAKTLKDLRLDYLDLFLVHWPFPNHHATGVSVDSRDPH
AVPYIHENFMKTWRQMERLVDMGLVKAIGTSNVTVPKLKLILRDARIKPAVNEMELHPHFQQPALFQFCLDNN
IVPIGYSPIGSPNRPDRDKTEADTVDVEDPVVVAIGNRLGVHPALVCLKWAVQRGQIPIPFSIHRKNYLANIRSITE
NLLTDEEMKAMAGIDKNCRLIKGQVFLWPSAKDWPDLWDEDGTIAGA

Table 28: AKR proteins selected from the drain metagenome.

8.5 Primers designed for selected AKRs from the drain metagenome

	Gene	Contig name	Primer name	Full primer	Tm	%GC	nt
1	19664	contig-120_19664_1	dAKR19664_F	AAAGGTCTCTTATGGGCTGCAATCTCCGGATGTTCC	72.9	59.1	22
			dAKR19664_R	AAAGGTCTCGGGTGATCCCATTTCGGCGCGAAGC	74.3	60	20
2	6751	contig-120_6751_1	dAKR6751_F	AAACATATGTGGCAGGTGGATGCCGGCGTC	80.4	71.4	21
			dAKR6751_R	AAACTCGAGGAACATGAAAGATGCCGTCTCCGG	73.4	54.2	24
3	28243	contig-120_28243_1	dAKR28243_F	AAACATATGAGCGAACAACCGACCATTCCG	72.4	57.1	21
			dAKR28243_R	AAACTCGAGGAAAGTGGCCGTGACTGGATTCG	72.7	56.2	23
4	19634	contig-120_19634_1	dAKR19634_F	AAACATATGGAATACCTGACGCTTTCTCATACCGATTTACG	72.6	43.8	32
			dAKR19634_R	AAACTCGAGTGGCATGCGTGCGCCGCGCG	87.5	80	20
5	8191	contig-120_8191_2	dAKR8191_F	AAAGGTCTCTTATGAAGAGCTATCGCCTGGCCGATG	72.4	59.1	22
			dAKR8191_R	AAAGGTCTCGGGTGAGGCACCCGCCCACCCCGCG	84.2	85	20
6	9148	contig-120_9148_1	dAKR9148_F	AAAGGTCTCGGGTGAAGACCTATACCCTCGCCGACGGC	74.1	62.5	24
			dAKR9148_R	AAAGGTCTCGGGTGCGGCACACGATCACCACGCG	77.7	70	20
7	562	contig-120_562_6	dAKR562_F	AAAGGTCTCTTATGCGACTCGGACTTGGCACGATG	72.5	61.9	21
			dAKR562_R	AAAGGTCTCGGGTGCGGCGACGGATTCGGATGCTG	77.77	66.7	21

8	3698	contig-120_3698_1	dAKR3698_F	AAACATATGACCTGGGGCGAGCAGAACAGC	74.1	66.7	21
			dAKR3698_R	AAACTCGAGCGGCGACGGATTCGGATGCTG	77.7	66.7	21
9	3273	contig-120_3273_2	dAKR3273_F	AAAGGTCTCTTATGGAATATCGGCAACTCGGATCGTCG	73	54.2	24
			dAKR3273_R	AAAGGTCTCGGGTGAACAAGCGGCGGGTTAAGGCG	74.5	61.9	21
10	16375	contig-120_16375_1	dAKR16375_F	AAACATATGGAAACCCGTTTTCTCGGCCGC	74.4	61.9	21
			dAKR16375_R	AAACTCGAGGGTTGCGCCTCGATATGCGG	73.9	65	20
11	3402	contig-120_3402_3	dAKR3402_F	AAACATATGGAATACCGCTATCTCGGCCGCTC	72.4	60.9	23
			dAKR3402_R	AAACTCGAGCACCGGAGTGGGGTTCCGGTAC	74.4	68.2	22
12	982	contig-120_982_2	dAKR982_F	AAAGGTCTCTTATGGAATACCGTTACCTCGGCGCGTC	73	60.9	23
			dAKR982_R	AAAGGTCTCGGGTGGACCGGCGTGGGATTGCGGTAG	77	68.2	22
13	1078	contig-120_1078_6	dAKR1078_F	AAACATATGGAATACCGCTACCTCGGCGCATC	73.1	60.9	23
			dAKR1078_R	AAACTCGAGGACCGGGGCGGGATTGCGGTAG	79.1	72.7	22
14	33576	contig-120_33576_1	dAKR33576_F	AAACATATGGCGACCGACGGCACCAAACAAG	76.5	63.6	22
			dAKR33576_R	AAACTCGAGCTTCGTGATCTGATCTATTTCAGCCATTTCTTC	72.4	39.4	33
15	49236	contig-120_49236_1	dAKR49236_F	AAACATATGGGCGAGGTCATCGCCGGACG	78.3	75	20
			dAKR49236_R	AAAAAGCTTGATGATGTCCAGCGGCCGGG	76.3	70	20
16	16563	contig-120_1656_3	dAKR16563_F	AAACATATGGAAAACCAGACTTTCGCGGTTGCC	73.9	54.2	24
			dAKR16563_R	AAACTCGAGCGCGCCGGCGATCGTTCCGTC	82.7	76.2	21

Table 29: Primers ordered for PCR amplification of AKR genes from the drain metagenome, and additional information calculated using Thermo Fisher Multiple Primer Analyzer.

8.6 SDR clarified cell lysate total protein concentrations

Enzyme	[Total protein] (mg/mL)	Enzyme	[Total protein] (mg/mL)	Enzyme	[Total protein] (mg/mL)	Enzyme	[Total protein] (mg/mL)
1	0.4	11	0.5	21	0.3	30	0.6
2	0.5	12	0.7	22	0.5	31	0.6
3	0.5	13	0.6	23	0.5	32	0.4
4	0.6	15	0.6	24	0.7	33	0.4
5	0.5	16	0.6	25	0.7	34	0.5
6	0.6	17	0.6	26	0.6	35	0.7
7	0.5	18	0.5	27	0.6	36	0.7
8	0.5	19	0.6	28	0.4	37	0.5
9	0.7	20	0.6	29	0.6	38	0.6
10	0.6						

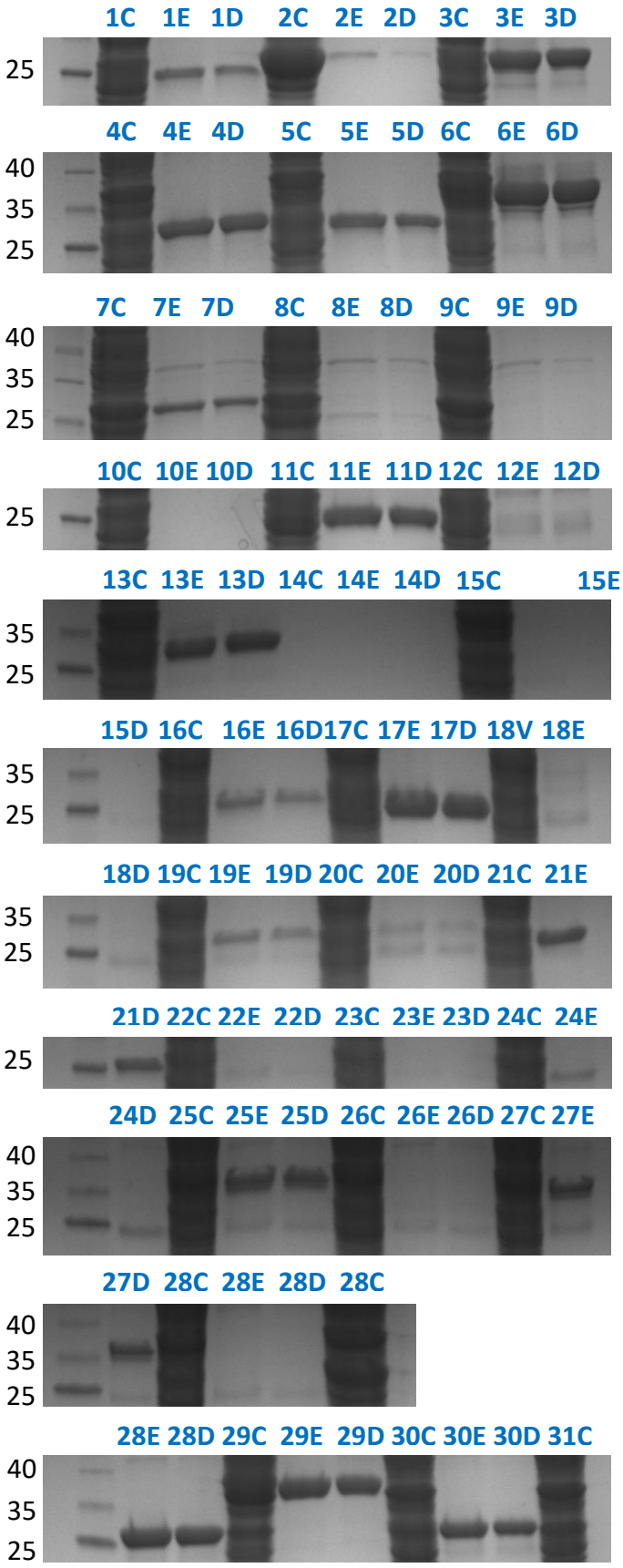
Table 30: Table to show final SDR total protein concentrations used in the spectrophotometric assays. Clarified cell lysates were flash frozen in liquid nitrogen and defrosted prior to use.

8.7 SDR purified and desalted protein concentrations

Enzyme	Concentration (mg/mL)	Enzyme	Concentration (mg/mL)	Enzyme	Concentration (mg/mL)
1	1.310	15	1.081	28	2.465
2	1.597	16	1.696	29	0.963
3	2.164	17	2.079	30	0.142
4	1.471	18	1.597	31	1.076
5	1.112	19	1.280	32	2.315
6	1.543	20	1.766	33	1.591
7	13.47	21	1.662	34	1.959
8	1.426	22	2.025	35	3.013
9	1.763	23	1.167	36	2.014
10	1.699	24	1.122	37	1.201
11	1.368	25	1.631	38	1.216
12	1.040	26	1.369		
13	1.156	27	1.131		

Table 31: Stock concentrations. Enzymes typically diluted ten-fold for assays.

8.8 KRED Gels


1C 1E 1D 2C 2E 2D 3C 3E 3D
25
4C 4E 4D 5C 5E 5D 6C 6E 6D
40
35
25
7C 7E 7D 8C 8E 8D 9C 9E 9D
40
35
25
10C 10E 10D 11C 11E 11D 12C 12E 12D
25
13C 13E 13D 14C 14E 14D 15C 15E
35
25
15D 16C 16E 16D17C 17E 17D 18V 18E
35
25
18D 19C 19E 19D 20C 20E 20D 21C 21E
35
25
21D 22C 22E 22D 23C 23E 23D 24C 24E
25
24D 25C 25E 25D 26C 26E 26D 27C 27E
40
35
25
27D 28C 28E 28D 28C
40
35
25
28E 28D 29C 29E 29D 30C 30E 30D 31C
40
35
25

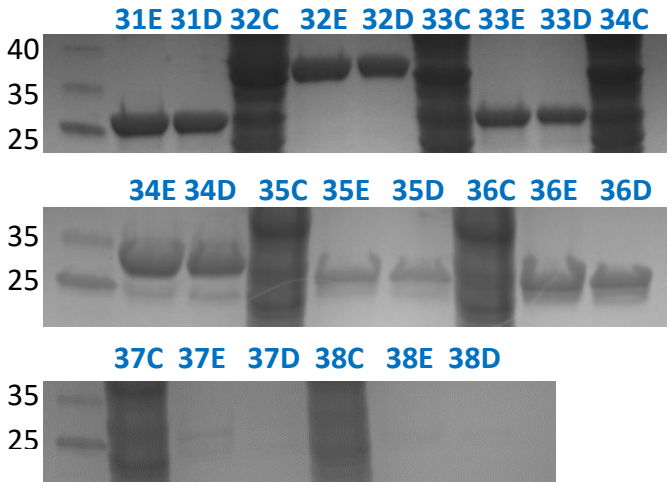


Figure 72: Gels are sorted in to numerical order, with the enzyme number labelled, in the following order: C = cell lysate; E = elution from purification column; D = desalted enzyme. Key on the left referes to molecular weight (kDa). The molecular weights (kDa): 1) 29, 2) 27.5, 3) 29.5, 4) 29.4, 5) 30.2, 6) 30.7, 7) 29.8, 8) 28, 9) 28.9, 10) 31.5, 11) 33.2, 12) 35.6, 13) 29.6, 15) 22.9, 16) 28.4, 17) 28.5, 18) 33.3, 19) 31.2, 20) 30.1, 21) 29, 22) 34, 23) 29.7, 24) 31.5, 25) 34.3, 26) 29, 27) 33.1, 28) 29.1, 29) 29.2, 30) 32.6, 31) 27.4, 32) 30.1, 33) 28.96, 34) 29.2, 35) 30.7, 36) 28.5, 37) 30.2, 38) 28.5.

8.9 Empty vector control

	Substrate	% NADPH conversion
7	Tropinone (DMSO)	8
29	Acetophenone	3
65	Cycloheptanone	-5
38	2-methylcyclohexanone	2
39	3-methylcyclohexanone	5
40	2-cyclohexen-1-one	3
52	1-indanone	3
53	2-indanone	0
54	Wieland–Miescher ketone	5
21	1-methyl-4-piperidone	1
37	Ethyl-4-oxo-1-piperidinecarboxylate	13
36	1-(2-methyl propyl)-4-piperidone	2
43	2,2,2,4'-Tetrafluoroacetophenone	6
42	2,2,2,3'-Tetrafluoroacetophenone	5
41	2,2,2-trifluoroacetophenone	8
44	4'-N-propyl-2,2,2-trifluoroacetophenone	0
40	Cyclohexanone	8
45	Benzaldehyde	12
50	1-phenyl propan-2-one	7
46	Phenylacetaldehyde	0
47	2-heptanone	6
48	Methyl acetoacetate	12
35	Pyruvic acid	0
71	Camphor	1
68	Progesterone	0
86	Pseudopelletierine	1

Table 32: Table displaying % conversion of NADPH without enzyme (200 µL)): substrate (5 mM), NADPH (1 mM), Pi (100 mM), DMSO (10%, v/v) clarified cell lysate of empty pET29a vector containing *E. coli* (20 µL). The reactions were shaken for 100 min at 25 °C and performed in triplicate. Quantification of NADPH depletion was completed by spectrophotometric assay at 340 nm.

8.10 No enzyme control (water instead)

Substrates		NADPH	Standard deviation	NADH	Standard deviation
7	Tropinone	0.00	0.00	0.03	0.01
7	Tropinone*	-0.02	0.01	0.04	0.00
95	IDABO	0.10	0.02	0.11	0.02
86	Pseudopelletierine*	0.17	0.03	0.15	0.01
87	MADO	-0.20	0.00	0.03	0.04
16	2-CBT*	0.01	0.01	0.01	0.02
19	Cyclohexanone*	-0.03	0.01	0.00	0.01
65	Cycloheptatnone*	-0.01	0.00	0.02	0.06
38	2-methylcyclohexanone*	0.00	0.01	-0.03	0.03
(R)-39	(R)-3-methylcyclohexanone*	0.01	0.01	0.02	0.03
39	3-methylcyclohexanone	0.00	0.00	0.02	0.01
66*	4-methylcyclohaxanone*	-0.03	0.01	-0.02	0.01
40	2-cyclohexene-1-one	-0.06	0.00	-0.03	0.01
51	3-(2-oxocyclohexyl)acetaldehyde	0.03	0.01	0.03	0.02

52	1-indanone	0.04	0.01	0.10	0.03
53	2-indanone	-0.01	0.05	0.06	0.02
21	1-methyl-4-piperidone*	-0.09	0.01	-0.06	0.02
37	Ethyl-4-oxo-1-piperidone*	0.02	0.04	-0.01	0.01
36	1-2-methyl-4-piperidone*	-0.09	0.01	-0.07	0.05
43	2,2,4-trifuloroacetophenone	0.02	0.01	0.05	0.03
42	2,2,3-trifluoroacetophenone	0.02	0.03	0.03	0.00
41	2,2,2-trifluoroacetopenone	0.00	0.01	0.02	0.01
44	4'-N-propyl-2,2,2-trifluroacetophenone	-0.18	0.02	-0.11	0.04
45	Benzaldehyde	0.01	0.02	0.02	0.00
50	Phenyl-propan-2-one	0.08	0.00	0.09	0.02
47	2-hepatanone	-0.01	0.00	0.01	0.01
(±)-71	(±)-Camphor	-0.01	0.00	-0.03	0.06
Cis-72*	Cis-jasmone*	0.03	0.00	0.05	0.01
+)-73	(+)-pulegone	-0.05	0.02	-0.01	0.01
H2O*	H2O*	0.00	0.01	0.00	0.01
DMSO	DMSO	0.00	0.01	0.00	0.01
29	Acetophenone	0.01	0.03	0.00	0.00
35	Pyruvic acid	0.00	-0.02	0.00	0.00
35	Pyruvate	0.01	-0.03	0.00	0.00
Trans-69	Transandersterone	0.01	0.04	0.00	0.00
69	Andersterone	0.00	-0.38	0.00	0.00
67	Testosterone	0.00	-0.38	0.00	0.00
68	Progesterone	0.00	-0.38	0.00	0.00
70	11-ketoprogesterone	0.05	-0.26	0.00	0.00
(1R)-(+)-71	(1R)-(+)-Camphor	0.00	0.02	0.00	0.00
46	Phenylacetaldehyde	0.01	-0.14	0.00	0.00
48	Methyl acetoacetate	0.01	0.20	0.02	0.10
29	Acetophenone*	0.01	-0.05	0.00	0.00
60		0.06	0.01		
62		0.09	0.02		
(S)-54		0.02	0.00		
(R)-54		0.00	0.02		
63		0.06	0.01		
58		-0.01	0.01		
59		0.01	0.02		
61		-0.03	0.00		
87		0.00	3.33		
89		0.00	1.52		

Table 33: Table displaying % conversion of NADPH without enzyme (200 µL)): substrate (5 mM), NADPH (1 mM), Pi (100 mM), DMSO (10%, v/v). The reactions were shaken for 100 min at 25 °C and performed in triplicate. Quantification of NADPH depletion was completed by spectrophotometric assay at 340 nm.

8.11 UV-visible spectrum for substrates

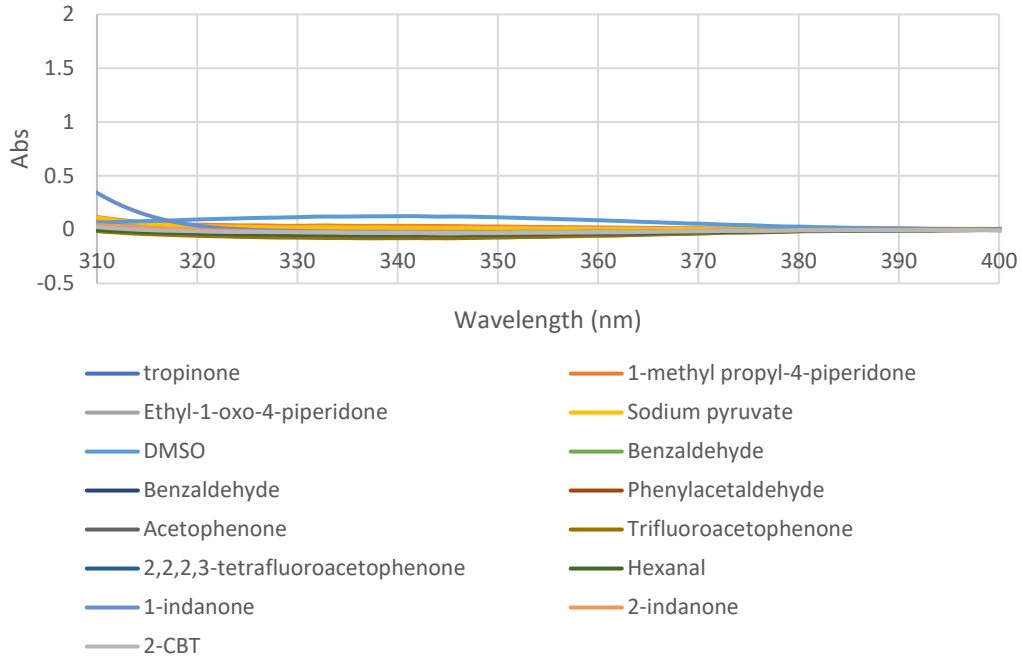


Figure 73: UV-visible spectra of substrates used for spectrophotometric assays in the region near the λmax of NAD(P)H at 340 nm.

8.12 List of Transaminases from UCL Toolbox

Name	pQR Number	Microorganism	Construct Gene ID
***Pa*-TAm-1**	902	*Pseudomonas aeruginosa*	NCBI gi:4382849
***Kp*-TAm-1**	904	*Klebsiella pneumoniae*	NCBI gi:3309659
***Bs*-TAm-1**	906	*Bacillus subtillis*	NCBI:gi16079319
***Ec*-TAm**	907	*Escherichia coli*	NCBI:gi49176403
***Bs*-TAm-2**	961	*Bacillus subtillis*	BSU09260_402
***Pa*-TAm-2**	426	*Pseudomonas aeruginosa*	PAO132
***Pp*-TAm-1**	427	*Pseudomonas putida*	PP0596
***Cv*-TAm**	801	*Chromobacterium violaceum*	CV2025
***Se*-TAm-1**	803	*Saccharopolyspora erythraea*	Sery1824 = SACE3329
***Se*-TAm-2**	804	*Saccharopolyspora erythraea*	Sery1599 = SACE4673
***Se*-TAm-3**	805	*Saccharopolyspora erythraea*	Sery1902 = SACE5670
***Sa*-TAm-1**	806	*Streptomyces avermitilis*	Sav4551
***Sa*-TAm-2**	807	*Streptomyces avermitilis*	Sav2612
***Ta*-TAm**	914	*Thermus aquaticus*	Y51MC23
***Pp*-TAm-2**	958	*Pseudomonas putida*	PP_3718
***Pp*-TAm-3**	959	*Pseudomonas putida*	PP_2180

***Bs*-TAm-3**	960	*Bacillus subtillis*	BSU09260_1971
***Dg*-TAm**	980	*Deinococcus geothermalis*	Dgeo_1416
***Vf*-TAm**	1003	*Vibrio fluvialis*	VF
***Kp*-TAm-2**	1005	*Klebsiella pneumoniae*	KPN_00255
***Kp*-TAm-3**	1006	*Klebsiella pneumoniae*	KPN_00799
***Rr*-TAm**	1017	*Rhodospirillum rubrum*	Rru_A1254
***Rs*-TAm**	1019	*Rhodobacter sphaeroides*	Rsph17025_2835
***As*-TAm**	As	*Arthrobacter sp.*	AS-Merck
***Mb*-TAm**	Mb	*Mycobacterium vanbaalenii*	Mb (Mvan_4516)

Table 34: TAms from the UCL toolbox library

8.13 TAm clarified cell lysate total protein concentrations

TAm	Concentration (mg/mL)
Pa-TAm-1	7.8
Kp-TAm-1	5.6
Bs-TAm-1	7.5
Ec-TAm	8.7
Bs-TAm-2	10.3
Pa-TAm-2	10.6
Pp-TAm-1	9.9
Cv-TAm	8.2
Se-TAm-1	2.5
Se-TAm-2	8.6
Se-TAm-3	7.3
Sa-TAm-1	10.8
Sa-TAm-2	8.3
Ta-TAm	10.4
Pp-TAm-2	9.9
Pp-TAm-3	10.6
Bs-TAm-3	9.1
Dg-TAm	9.4
Vf-TAm	0.0
Kp-TAm-2	10.6
Kp-TAm-3	0.0
Rr-TAm	9.2
Rs-TAm	10.3
As-TAm	5.8
Mb-TAm	7.6
Se-TAm-1	3.9
Sa-TAm-2	2.1
Vf-TAm	2.5
Kp-TAm-3	3.0

Table 35: Table of typical total protein concentrations of TAm clarified cell lystates. These were diluted accordingly for assays.

8.14 HPLC and GC traces of standards

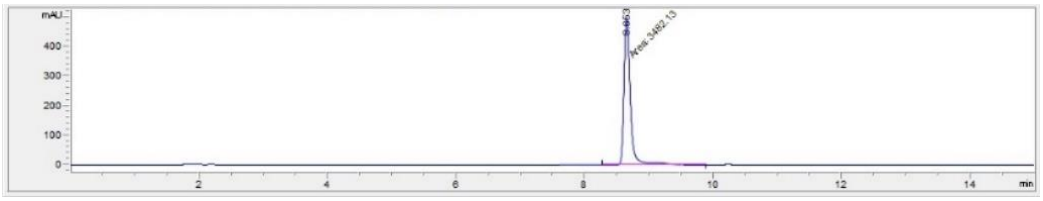


Figure 74: HPLC trace of acetophenone 29 (254 nm)

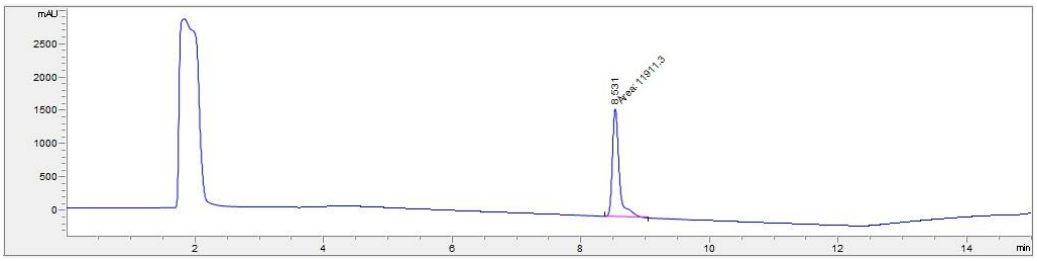


Figure 75: HPLC trace of acetophenone 29 (210 nm)

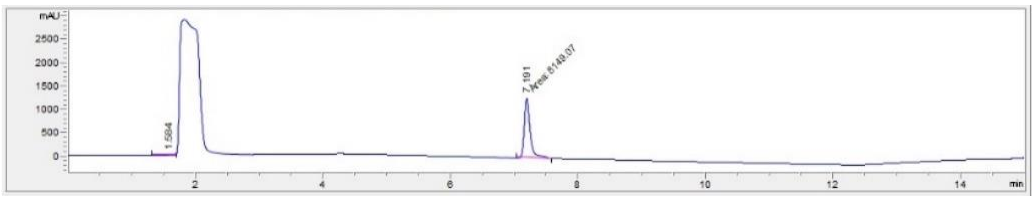


Figure 76: HPLC trace of phenylethanol 55 (210 nm)

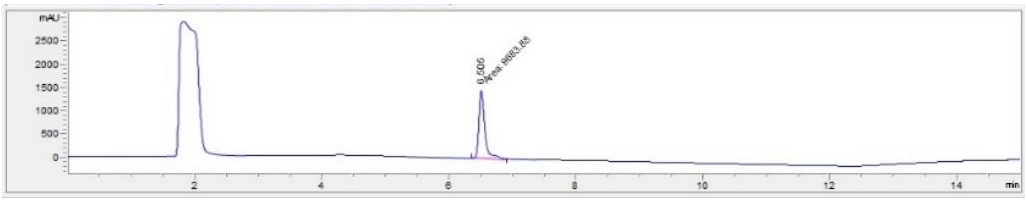


Figure 77: HPLC trace of WMK 54 (210 nm)

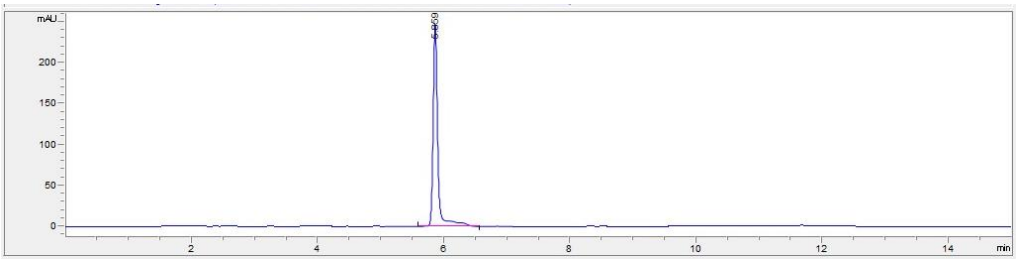


Figure 78: HPLC trace of (4a*R*,5*S*)-56 (254 nm)

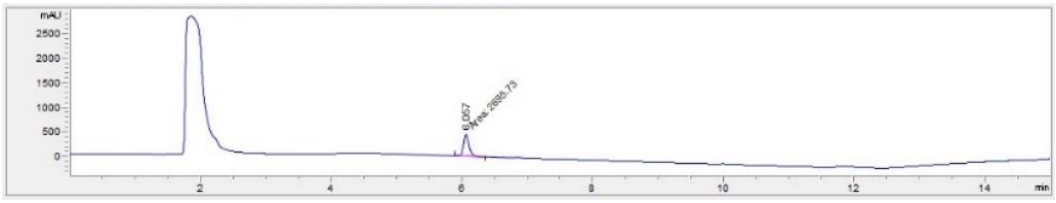


Figure 79: HPLC trace of (4a*R*,5*S*)-56 (210 nm)

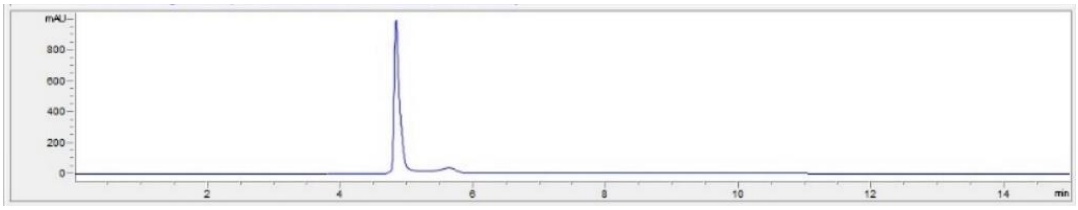


Figure 80: HPLC trace of 2-CBT 16 (254 nm)

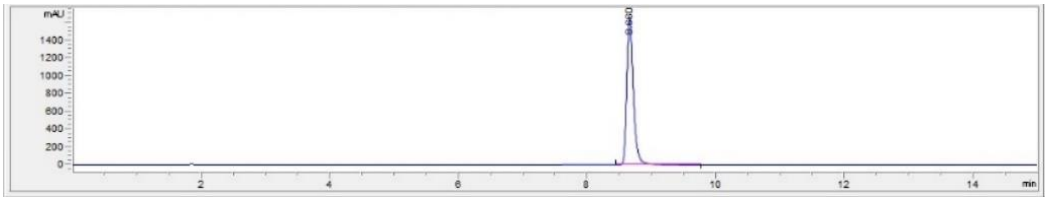


Figure 81: HPLC trace of benzaldehyde 45 (254 nm)

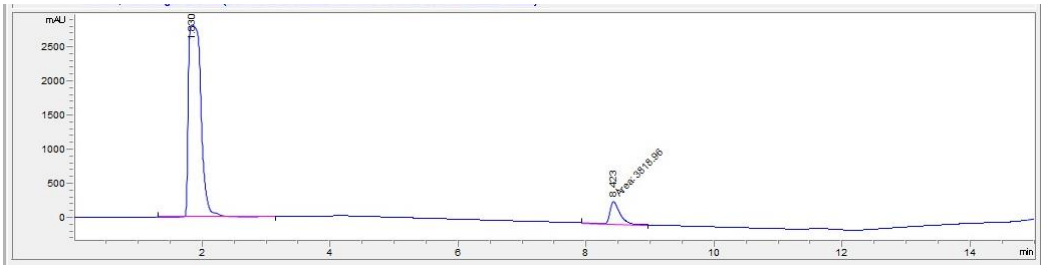


Figure 82: HPLC trace of benzaldehyde 45 (210 nm)

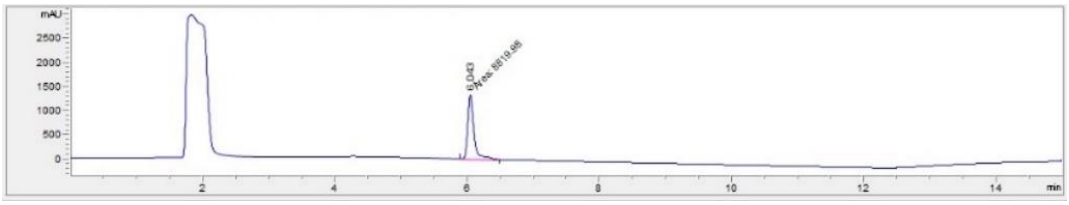


Figure 83: HPLC trace of benzyl alcohol 82 (210 nm)

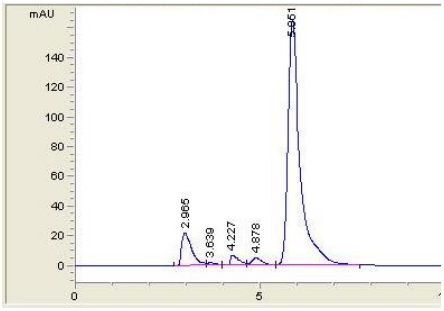


Figure 84: HPLC trace of (*R*)-phenylethanol (*R*)-55 (210nm)

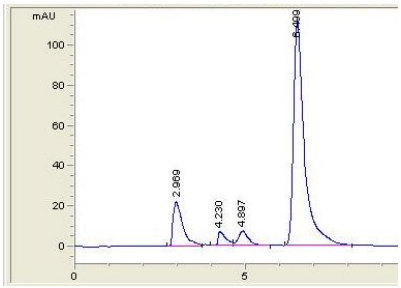


Figure 85: HPLC trace of (*S*)-phenylethanol (*S*)-55 (210nm).

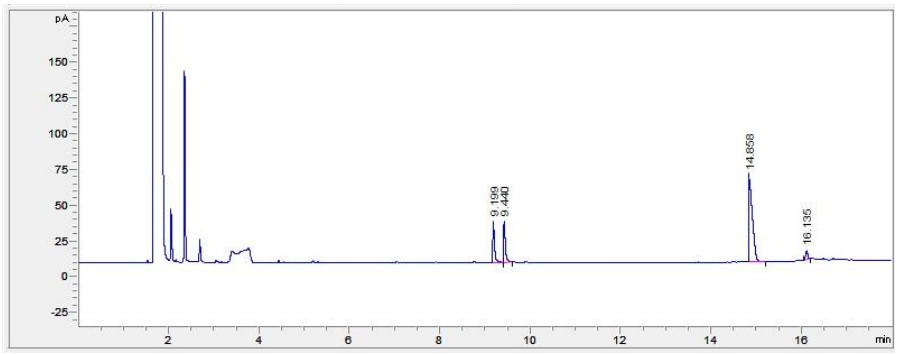


Figure 86: GC trace of (±)-phenylethanol (±)-55 (Rt = 9.2 and 9.4, DMSO Rt = 14.9)

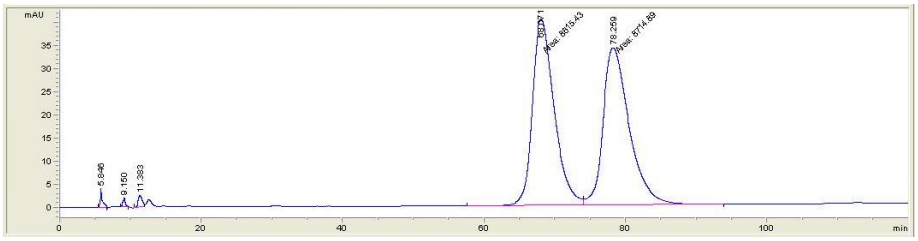


Figure 87: HPLC trace of WMK 54 (230 nm, (S)-WMK (*S*)-54 Rt = 64 min, (R)-WMK (*R*)-54 Rt = 73 min).

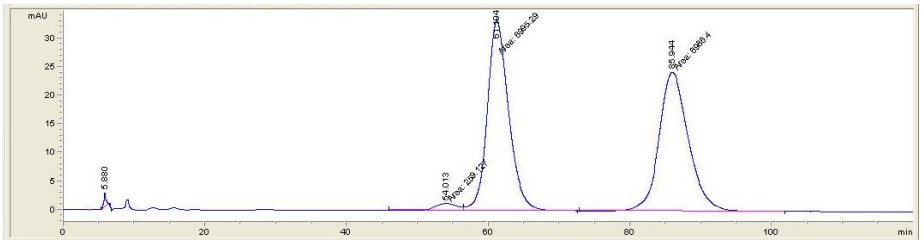


Figure 88: HPLC trace of 56 (230 nm, (4a*R*,5*S*)-56 Rt = 54 min, (4a*R*,5*R*)-56 Rt = 61 min, (4a*S*,5*S*)-56 Rt = 85 min).

9 References

1. De Wildeman, S. M. a, Sonke, T., Schoemaker, H. E. & May, O. Biocatalytic reductions: From lab curiosity to 'first choice'. *Acc. Chem. Res.* **40,** 1260–1266 (2007).

2. Scheidt, B. & Glieder, A. Yeast cell factories for fine chemical and API production. *Microb. Cell Fact.* **7,** 25 (2008).

3. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329,** 305–309 (2010).

4. Jeffries, J. W. E. *et al.* Metagenome Mining: A Sequence Directed Strategy for the Retrieval of Enzymes for Biocatalysis. *ChemistrySelect* **1,** 2217–2220 (2016).

5. Pollard, D. J. & Woodley, J. M. Biocatalysis for pharmaceutical intermediates: the future is now. *Trends Biotechnol.* **25,** 66–73 (2007).

6. Jeffries, J. W. E. Functional Metagenomics: Metagenome Mining for Industrial Biocatalysis. (University College London, 2016).

7. Handelsman, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol. Mol. Biol. Rev.* **69,** 195–195 (2005).

8. Guo, F. & Berglund, P. Transaminase biocatalysis: optimization and application. *Green Chem.* **19,** 333–360 (2017).

9. Shin, J.-S., Yun, H., Jang, J.-W., Park, I. & Kim, B.-G. Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from Vibrio fluvialis JS17. *Appl. Microbiol. Biotechnol.* **61,** 463–471 (2003).

10. Kaulmann, U., Smithies, K., Smith, M. E. B., Hailes, H. C. & Ward, J. M. Substrate spectrum of ω-transaminase from Chromobacterium violaceum DSM30191 and its potential for biocatalysis. *Enzyme Microb. Technol.* **41,** 628–637 (2007).

11. Gilbert, J. a & Dupont, C. L. Microbial metagenomics: beyond the genome. *Ann. Rev. Mar. Sci.* **3,** 347–371 (2011).

12. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5,** R245–R249 (1998).

13. Wooley, J. C., Godzik, A. & Friedberg, I. A primer on metagenomics. *PLoS Comput. Biol.* **6,** 1–13 (2010).

14. Torsvik, V., Øvreås, L. & Thingstad, T. F. Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* **296,** 1064–6 (2002).

15. Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Res.* **42,** 290–301 (2014).

16. Dawson, N. L. *et al.* CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* **45,** 289–295 (2017).

17. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** 580–584 (2015).

18. Pearson, W. R. An Introduction to Sequence Similarity ('Homology') Searching. *Curr. Protoc. Bioinforma.* **4,** 3.1.1–3.1.8 (2013).

19. Jeffries, J. *et al.* Functional Metagenomics : A Strategy for the Sequence directed retrieval of functional enzymes for biocatalysis. *Accepted* 1–17

20. Leis, B. *et al.* Identification of novel esterase-active enzymes from hot environments by use of the host bacterium Thermus thermophilus. *Front. Microbiol.* **6,** 1–12 (2015).

21. Stroobants, A., Portetelle, D. & Vandenbol, M. New carbohydrate-active enzymes identified by screening two metagenomic libraries derived from the soil of a winter wheat field. *J. Appl. Microbiol.* **117,** 1045–1055 (2014).

22. Besemer, J. & Borodovsky, M. Heuristic approach to deriving models for gene finding. *Nucleic Acids Res.* **27,** 3911–3920 (1999).

23. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).

24. Monti, D. *et al.* Cascade Coupling of Ene-Reductases and ω-Transaminases for the Stereoselective Synthesis of Diastereomerically Enriched Amines. *ChemCatChem* **7,** 3106–3109 (2015).

25. Popovic, A. *et al.* Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. *Sci. Rep.* **7,** 1–15 (2017).

26. Hoffmann, F. & Maser, E. Carbonyl reductases and pluripotent hydroxysteroid dehydrogenases of the short-chain dehydrogenase/reductase superfamily. *Drug Metab. Rev.* **39,** 87–144 (2007).

27. Jörnvall, H. *et al.* Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34,** 6003–13 (1995).

28. Bhatia, C. *et al.* Towards a systematic analysis of human short-chain dehydrogenases/reductases (SDR): Ligand identification and structure-activity relationships. *Chem. Biol. Interact.* **234,** 114–125 (2015).

29. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65,** 3895–3906 (2008).

30. Persson, B. & Kallberg, Y. Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chem. Biol. Interact.* **202,** 111–115 (2013).

31. Kallberg, Y., Oppermann, U., Jörnvall, H. & Persson, B. Short-chain dehydrogenases/reductases (SDRs). Coenzyme-based functional assignments in completed genomes. *Eur. J. Biochem.* **269,** 4409–4417 (2002).

32. Persson, B., Kallberg, Y., Oppermann, U. & Jörnvall, H. Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chem. Biol. Interact.* **143–144,** 271–278 (2003).

33. Penning, T. M. The aldo-keto reductases (AKRs): Overview. *Chem. Biol. Interact.* **234,** 236–246 (2015).

34. Matsuda, T., Yamanaka, R. & Nakamura, K. Recent progress in biocatalysis for asymmetric oxidation and reduction. *Tetrahedron Asymmetry* **20,** 513–557 (2009).

35. Zhu, D. & Hua, L. How carbonyl reductases control stereoselectivity: Approaching the goal of rational design. *Pure Appl. Chem.* **82,** 117–128 (2010).

36. Giacomelli, G., Menicagli, R. & Lardicci, L. Alkyl metal asymmetric reduction. III. Stereochemistry of alkyl phenyl ketone reductions by chiral organoaluminum compounds. *J. Org. Chem.* **38,** 2370–2376 (1973).

37. Bravo, P. & Resnati, G. Preparation and properties of chiral fluoroorganic compounds. *Tetrahedron: Asymmetry* **1,** 661–692 (1990).

38. Cho, B. T. Recent development and improvement for boron hydride-based catalytic asymmetric reduction of unsymmetrical ketones. *Chem. Soc. Rev.* **38,** 443–452 (2009).

39. Behrens, G. a., Hummel, A., Padhi, S. K., Schätzle, S. & Bornscheuer, U. T. Discovery and protein engineering of biocatalysts for organic synthesis. *Adv. Synth. Catal.* **353,** 2191–2215 (2011).

40. Moore, J. C., Pollard, D. J., Kosjek, B. & Devine, P. N. Advances in the Enzymatic Reduction of Ketones Background : Whole Cell Bioreductions. *Acc. Chem. Res.* **40,** 1412–1419 (2007).

41. Forrest, G. L. & Gonzalez, B. Carbonyl reductase. *Chem. Biol. Interact.* **129,** 21–40 (2000).

42. Liang, J. *et al.* Development of a biocatalytic process as an alternative to the (-)-DIP-Cl-mediated asymmetric reduction of a key intermediate of montelukast. *Org. Process Res. Dev.* **14,** 193–198 (2010).

43. Truppo, M. D., Escalettes, F. & Turner, N. J. Rapid determination of both the activity and enantioselectivity of ketoreductases. *Angew. Chemie - Int. Ed.* **47,** 2639–2641 (2008).

44. Korbel, G. a., Lalic, G. & Shair, M. D. Reaction microarrays: A method for rapidly determining the enantiomeric excess of thousands of samples. *J. Am. Chem. Soc.* **123,** 361–362 (2001).

45. Millot, N. *et al.* Rapid Determination of Enantiomeric Excess Using Infrared Thermography. *Org. Process Res. Dev.* **6,** 463–470 (2002).

46. Reetz, M. T., Becker, M. H., Klein, H.-W. & Stöckigt, D. A Method for High-Throughput Screening of Enantioselective Catalysts. *Communications* **38,** 1758–1761 (1999).

47. Reetz, M. T., Kühling, K. M., Deege, A., Hinrichs, H. & Belder, D. Super-High-Throughput Screening of Enantioselective Catalysts by Using Capillary Array Electrophoresis. *Angew. Chemie* **39,** 3891–3893 (2000).

48. Morís-Varas, F. *et al.* Visualization of enzyme-catalyzed reactions using pH indicators: rapid screening of hydrolase libraries and estimation of the enantioselectivity. *Bioorg. Med. Chem.* **7,** 2183–2188 (1999).

49. Huisman, G. W., Liang, J. & Krebber, A. Practical chiral alcohol manufacture using ketoreductases. *Curr. Opin. Chem. Biol.* **14,** 122–129 (2010).

50. Smart, B. E. Fluorine substituent effects (on bioactivity). *J. Fluor. Chem.* **109,** 3–11

(2001).

51. Pollard, D., Truppo, M., Pollard, J., Chen, C. Y. & Moore, J. Effective synthesis of (S)-3,5-bistrifluoromethylphenyl ethanol by asymmetric enzymatic reduction. *Tetrahedron Asymmetry* **17,** 554–559 (2006).

52. Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach*. *Medicinal Natural Products: A Biosynthetic Approach* (John Wiley & Sons, 2009).

53. Kim, N., Estrada, O., Chavez, B., Stewart, C. & D'Auria, J. C. Tropane and granatane alkaloid biosynthesis: A systematic analysis. *Molecules* **21,** 1–25 (2016).

54. Grynkiewicz, G. & Gadzikowska, M. Tropane alkaloids as medicinally useful natural products and their synthetic derivatives as new drugs. *Pharmacol. Reports* **60,** 439–463 (2008).

55. Jocković, N., Fischer, W., Brandsch, M., Brandt, W. & Dräger, B. Inhibition of human intestinal α-glucosidases by calystegines. *Journal of Agricultural and Food Chemistry* **61,** 5550–5557 (2013).

56. Ahmad, A. & Leete, E. Biosynthesis of the tropine moiety of hyoscyamine from δ-N-methylornithine. *Phytochemistry* **9,** 2345–2347 (1970).

57. T Hashimoto, and & Yamada, Y. Alkaloid Biogenesis: Molecular Aspects. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45,** 257–285 (1994).

58. Leete, E. Biosynthesis of cocaine and cuscohygrine in Erythroxylon coca. *J. Chem. Soc. Chem. Commun.* 1170–1171 (1980). doi:10.1039/c39800001170

59. Humphrey, A. J. & O'Hagan, D. Tropane alkaloid biosynthesis. A century old problem unresolved. *Nat. Prod. Rep.* **18,** 494–502 (2001).

60. Leete, E. Recent Developments in the Biosynthesis of the Tropane Alkaloids. *Planta Med.* **56,** 339–352 (1990).

61. Jirschitzka, J. *et al.* Plant tropane alkaloid biosynthesis evolved independently in the Solanaceae and Erythroxylaceae. *Proc. Natl. Acad. Sci.* **109,** 10304–10309 (2012).

62. Nakajima, K., Kato, H., Oda, J., Yamada, Y. & Hashimoto, T. Site-directed mutagenesis of putative substrate-binding residues reveals a mechanism controlling the different stereospecificities of two tropinone reductases. *J. Biol. Chem.* **274,** 16563–16568 (1999).

63. Oppermann, U. *et al.* Short-chain dehydrogenases/reductases (SDR): the 2002 update. **144,** 247–253 (2003).

64. Brock, A., Brandt, W. & Dräger, B. The functional divergence of short-chain dehydrogenases involved in tropinone reduction. *Plant J.* **54,** 388–401 (2008).

65. Persson, B. *et al.* The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative. *Chem. Biol. Interact.* **178,** 94–98 (2009).

66. Reinhardt, N. *et al.* Substrate flexibility and reaction specificity of tropinone reductase-like short-chain dehydrogenases. *Bioorg. Chem.* **53,** 37–49 (2014).

67. Dräger, B. & Schaal, A. Tropinone reduction in Atropa belladonna root cultures. *Phytochemistry* **35,** 1441–1447 (1994).

68. Portsteffen, A., Dräger, B. & Nahrstedt, A. The reduction of tropinone in Datura stramonium root cultures by two specific reductases. *Phytochemistry* **37,** 391–400 (1994).

69. Boswell, H. D. *et al.* Specificities of the enzymes of N-alkyltropane biosynthesis in Brugmansia and Datura. *Phytochemistry* **52,** 871–878 (1999).

70. Hashimoto, T., Nakajima, K., Ongena, G. & Yamada, Y. Two Tropinone Reductases with Distinct Stereospecificities from Cultured Roots of Hyoscyamus niger. *Plant Physiol.* **100,** 836–845 (1992).

71. Kaiser, H. *et al.* Immunolocalisation of two tropinone reductases in potato (Solanum tuberosum L.) root, stolon, and tuber sprouts. *Planta* **225,** 127–137 (2006).

72. Keiner, R., Kaiser, H., Nakajima, K., Hashimoto, T. & Dräger, B. Molecular cloning, expression and characterization of tropinone reductase II, an enzyme of the SDR family in Solanum tuberosum (L.). *Plant Mol. Biol.* **48,** 299–308 (2002).

73. Kushwaha, A. K., Sangwan, N. S., Tripathi, S. & Sangwan, R. S. Molecular cloning and catalytic characterization of a recombinant tropine biosynthetic tropinone reductase from Withania coagulans leaf. *Gene* **516,** 238–247 (2013).

74. Dräger, B. Tropinone reductases, enzymes at the branch point of tropane alkaloid metabolism. *Phytochemistry* **67,** 327–37 (2006).

75. Nakajima, K., Oshita, Y., Yamada, Y. & Hashimoto, T. Insight into the molecular evolution of two tropinone reductases. *Bioscience, biotechnology, and biochemistry*

63, 1819–1822 (1999).

76. Nakajima, K., Oshita, Y., Kaya, M., Yamada, Y. & Hashimoto, T. Structures and expression patterns of two tropinone reductase genes from Hyoscyamus niger. *Bioscience, biotechnology, and biochemistry* **63,** 1756–1764 (1999).

77. Qiang, W. *et al.* Functional characterisation of a tropine-forming reductase gene from Brugmansia arborea, a woody plant species producing tropane alkaloids. *Phytochemistry* **127,** 12–22 (2016).

78. Nakajima, K. *et al.* Crystal structures of two tropinone reductases: different reaction stereospecificities in the same protein fold. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 4876–4881 (1998).

79. Dräger, B. Analysis of tropane and related alkaloids. *J. Chromatogr. A* **978,** 1–35 (2002).

80. Polievktov, M. K., Markova, I. G. & Lizgunova, M. V. Investigation of the synthesis of tropinone according to the Robinson-Schöpf reaction by a polarographic method. *Pharm. Chem. J.* **9,** 413–417 (1976).

81. Yamashita, A. *et al.* Capturing enzyme structure prior to reaction initiation: Tropinone reductase-II-substrate complexes. *Biochemistry* **42,** 5566–5573 (2003).

82. Nakajima, K., Hashimoto, T. & Yamada, Y. Opposite stereospecificity of two tropinone reductases is conferred by the substrate-binding sites. *J. Biol. Chem.* **269,** 11695–11698 (1994).

83. Freydank, A. C., Brandt, W. & Dräger, B. Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. *Proteins Struct. Funct. Genet.* **72,** 173–183 (2008).

84. Jirschitzka, J. *et al.* Plant tropane alkaloid biosynthesis evolved independently in the Solanaceae and Erythroxylaceae Supporting Information. **4,** 1–9

85. Jez, J. M., Bennett, M. J., Schlegel, B. P., Lewis, M. & Penning, T. M. Comparative anatomy of the aldo – keto reductase superfamily. **636,** 625–636 (1997).

86. Silva-Junior, O. B., Grattapaglia, D., Novaes, E. & Collevatti, R. G. Genome assembly of the Pink Ipê (Handroanthus impetiginosus, Bignoniaceae), a highly valued, ecologically keystone Neotropical timber forest tree. *Gigascience* **7,** 1–16 (2018).

87. Schmidt, G. W. *et al.* The Last Step in Cocaine Biosynthesis Is Catalyzed by a BAHD Acyltransferase. *Plant Physiol.* **167,** 89–101 (2015).

88. Du, C. J., Rios-Solis, L., Ward, J. M., Dalby, P. A. & Lye, G. J. Evaluation of CV2025 ω-transaminase for the bioconversion of lignin breakdown products into value-added chemicals: synthesis of vanillylamine from vanillin 1. *Biocatal. Biotransformation* **32,** 302–313 (2014).

89. Baud, D., Ladkau, N., Moody, T. S., Ward, J. M. & Hailes, H. C. A rapid, sensitive colorimetric assay for the high-throughput screening of transaminases in liquid or solid-phase. *Chem. Commun.* **51,** 17225–17228 (2015).

90. Schell, U., Wohlgemuth, R. & Ward, J. M. Synthesis of pyridoxamine 5'-phosphate using an MBA:pyruvate transaminase as biocatalyst. *J. Mol. Catal. B Enzym.* **59,** 279–285 (2009).

91. Rudat, J., Brucher, B. R. & Syldatk, C. Transaminases for the synthesis of enantiopure beta-amino acids. *AMB Express* **2,** 11 (2012).

92. Dold, S.-M., Syldatk, C. & Rudat, J. Transaminases and their Applications. in *Green Biocatalysis* (ed. Patel, R. N.) 715–746 (John Wiley & Sons, Inc, 2016). doi:10.1002/9781118828083.ch29

93. Genz, M. *et al.* Alteration of the donor/acceptor spectrum of the (S)-amine transaminase from vibrio fluvialis. *Int. J. Mol. Sci.* **16,** 26953–26963 (2015).

94. Pavlidis, I. V. *et al.* Identification of (S)-selective transaminases for the asymmetric synthesis of bulky chiral amines. *Nat. Chem.* **8,** 1076–1082 (2016).

95. Koszelewski, D., Tauber, K., Faber, K. & Kroutil, W. ω-Transaminases for the synthesis of non-racemic ω-chiral primary amines. *Trends Biotechnol.* **28,** 324–332 (2010).

96. Schätzle, S. Identification , characterization and application of novel (R) -selective amine transaminases. (University of Greifswald, 2011).

97. Cassimjee, K. E., Manta, B. & Himo, F. A quantum chemical study of the ω-transaminase reaction mechanism. *Org. Biomol. Chem.* **13,** 8453–8464 (2015).

98. Hayashi, H. *et al.* Strain and catalysis in aspartate aminotransferase. *Biochim. Biophys. Acta - Proteins Proteomics* **1647,** 103–109 (2003).

99. Höhne, M., Schätzle, S., Jochens, H., Robins, K. & Bornscheuer, U. T. Rational

assignment of key motifs for function guides in silico enzyme identification. *Nat. Chem. Biol.* **6,** 807–813 (2010).

100. Mathew, S., Shin, G., Shon, M. & Yun, H. High throughput screening methods for omega-transaminases. *Biotechnology and Bioprocess Engineering* **18,** 1–7 (2013).

101. Cassimjee, K. E., Manta, B. & Himo, F. A quantum chemical study of the ω-transaminase reaction mechanism. *Org. Biomol. Chem.* **13,** 8453–8464 (2015).

102. Cho, B.-K. *et al.* Redesigning the Substrate Specificity of omega-Aminotransferase for the Kinetic Resolution of Aliphatic Chiral Amines. *Biotechnol. Bioeng.* **99,** 275–284 (2007).

103. Humble, M. S. *et al.* Crystal structures of the Chromobacterium violaceumω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. *FEBS J.* **279,** 779–792 (2012).

104. Deszcz, D. *et al.* Single active-site mutants are sufficient to enhance serine:pyruvate α-transaminase activity in an ω-transaminase. *FEBS J.* **282,** 2512–2526 (2015).

105. Shin, G., Mathew, S. & Yun, H. Kinetic resolution of amines by (R)-selective omega-transaminase from Mycobacterium vanbaalenii. *J. Ind. Eng. Chem.* **23,** 128–133 (2015).

106. Schätzle, S. *et al.* Enzymatic asymmetric synthesis of enantiomerically pure aliphatic, aromatic and arylaliphatic amines with (R)-selective amine transaminases. *Adv. Synth. Catal.* **353,** 2439–2445 (2011).

107. Pressnitz, D. *et al.* Asymmetric amination of tetralone and chromanone derivatives employing ω-transaminases. *ACS Catal.* **3,** 555–559 (2013).

108. Lichman, B. R. *et al.* One-pot triangular chemoenzymatic cascades for the syntheses of chiral alkaloids from dopamine. *Green Chem.* **17,** 852–855 (2015).

109. Höhne, M., Schätzle, S., Jochens, H., Robins, K. & Bornscheuer, U. T. Rational assignment of key motifs for function guides in silico enzyme identification. *Nat. Chem. Biol.* **6,** 807–813 (2010).

110. Martin, A. R. *et al.* Improved activity and thermostability of (S)-aminotransferase by error-prone polymerase chain reaction for the production of a chiral amine. *Biochem. Eng. J.* **37,** 246–255 (2007).

111. Pavlidis, I. V. *et al.* Identification of (S)-selective transaminases for the asymmetric synthesis of bulky chiral amines. *Nat. Chem.* 1–7 (2016). doi:10.1038/NCHEM.2578

112. Weiß, M. S. *et al.* Protein-engineering of an amine transaminase for the stereoselective synthesis of a pharmaceutically relevant bicyclic amine. *Org. Biomol. Chem.* **14,** 10249–10254 (2016).

113. Schätzle, S., Höhne, M., Redestad, E., Robins, K. & Bornscheuer, U. T. Rapid and sensitive kinetic assay for characterization of ω-transaminases. *Analytical Chemistry* **81,** 8244–8248 (2009).

114. Weiß, M. S., Pavlidis, I. V., Vickers, C., Hohne, M. & Bornscheuer, U. T. Glycine oxidase based high-throughput solid-phase assay for substrate profiling and directed evolution of (R)- and (S)-selective amine transaminases. *Anal. Chem.* **86,** 11847–11853 (2014).

115. Martin, A. R. *et al.* Improved activity and thermostability of (S)-aminotransferase by error-prone polymerase chain reaction for the production of a chiral amine. *Biochem. Eng. J.* **37,** 246–255 (2007).

116. Green, A. P., Turner, N. J. & O'Reilly, E. Chiral amine synthesis using ω-transaminases: An amine donor that displaces equilibria and enables high-throughput screening. *Angewandte Chemie - International Edition* **53,** 10714–10717 (2014).

117. Scheidt, T. *et al.* Fluorescence-Based Kinetic Assay for High-Throughput Discovery and Engineering of Stereoselective ω-Transaminases. *Adv. Synth. Catal.* **357,** 1721–1731 (2015).

118. Nguyen, P., West, M., Feske, B. D. & Padgett, C. W. Enantioselectivity and Enzyme-Substrate Docking Studies of a Ketoreductase from Sporobolomyces salmonicolor (SSCR) and Saccharomyces cerevisiae (YOL151w). *Hindawi Publ. Corp.* **2014,** (2014).

119. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).

120. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23,** 127–128 (2006).

121. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).

122. Raedts, J., Siemerink, M. A. J., Levisson, M., van der Oost, J. & Kengen, S. W. M. Molecular characterization of an NADPH-dependent acetoin reductase/2,3-butanediol dehydrogenase from Clostridium beijerinckii NCIMB 8052. *Appl. Environ. Microbiol.* **80,** 2011–2020 (2014).

123. Sakaitani, M. *et al.* Mechanistic Studies on trans-2,3-Dihydro-2,3-dihydroxybenzoate Dehydrogenase (Ent A) in the Biosynthesis of the Iron Chelator Enterobactin. *Biochemistry* **29,** 6789–6798 (1990).

124. Gasteiger E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handbook, Humana Press* 571–607 (2005).

125. Cooper, W. C., Jin, Y. & Penning, T. M. Elucidation of a Complete Kinetic Mechanism for a Mammalian Hydroxysteroid Dehydrogenase (HSD) and Identification of All. **282,** 33484–33493 (2007).

126. Rosen, T. C., Feldmann, R., Dünkelmann, P. & Daußmann, T. Bioreductive synthesis of perfluorinated chiral alcohols. *Tetrahedron Lett.* **47,** 4803–4806 (2006).

127. Bradshaw, C. W., Wong, C. H. & Hummel, W. Lactobacillus kefir Alcohol Dehydrogenase: A Useful Catalyst for Synthesis. *J. Org. Chem.* **57,** 1532–1536 (1992).

128. Wu, T., Wu, L. H., Knight, J. a & Diethylaminoethyl, M. O. Stability of NADPH: Effect of Various Factorson the Kinetics of Degradation. *Clin. Chem.* **32,** 314–319 (1986).

129. Toomey, R. E. & Wakil, S. J. Studies on the mechanism of fatty acid synthesis XV. Preparation and general properties of β-ketoacyl acyl carrier protein reductase from Escherichia coli. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **116,** 189–197 (1966).

130. Borzęcka, W., Lavandera, I. & Gotor, V. Synthesis of Enantiopure Fluorohydrins Using Alcohol Dehydrogenases at High Substrate Concentrations. 1–46 (2013).

131. Uruno, Y. *et al.* Discovery of dihydroquinazolinone derivatives as potent, selective, and CNS-penetrant M1 and M4 muscarinic acetylcholine receptors agonists. *Bioorg. Med. Chem. Lett.* **25,** 5357–5361 (2015).

132. Takusagawa, Y. *et al.* Purification and characterization of L-2,3-butanediol dehydrogenase of Brevibacterium saccharolyticum C-1012 expressed in Escherichia coli. *Bioscience, biotechnology, and biochemistry* **65,** 1876–1878 (2001).

133. Kurbanoglu, E. B., Zilbeyaz, K., Kurbanoglu, N. I. & Kilic, H. Asymmetric reduction of acetophenone analogues by Alternaria alternata using ram horn peptone. *Tetrahedron: Asymmetry* **18,** 2332–2335 (2007).

134. Bradshaw, B. & Bonjoch, J. The Wieland-Miescher Ketone: A Journey from Organocatalysis to Natural Product Synthesis. *Synlett* **23,** 337–356 (2012).

135. Sarkar, D. C., Das, A. R. & Ranu, B. C. Use of zinc borohydride as an efficient and highly selective reducing agent. Selective reduction of ketones and conjugated aldehydes over conjugated enones. *J. Org. Chem.* **55,** 5799–5801 (1990).

136. Ward, D. E., Rhee, C. K. & Zoghaib, W. M. A general method for the selective reduction of ketones in the presence of enones. *Tetrahedron Lett.* **29,** 517–520 (1988).

137. Ottolina, G., De Gonzalo, G., Carrea, G. & Danieli, B. Enzymatic Baeyer-Villiger oxidation of bicyclic diketones. *Adv. Synth. Catal.* **347,** 1035–1040 (2005).

138. Hioki, H., Hashimoto, T. & Kodama, M. Efficient kinetic resolution of (±)-4-methyl-Hajos – Parrish ketone by baker' s yeast reduction. **11,** 829–834 (2000).

139. Janeczko, T., Dmochowska-Gładysz, J. & Kostrzewa-Susłow, E. Chemoenzymatic resolution of racemic Wieland–Miescher and Hajos–Parrish ketones. *World J. Microbiol. Biotechnol.* **26,** 2047–2051 (2010).

140. Fuhshuku, K. I. *et al.* Access to Wieland-Miescher ketone in an enantiomerically pure form by a kinetic resolution with yeast-mediated reduction. *Journal of Organic Chemistry* **65,** 129–135 (2000).

141. Shimizu, N., Akita, H. & Kawamata, T. Enzymatic resolution of cis- and trans-1,2,3,4,6,7,8,8a-octahydro-8a-methyl-6-oxo-naphthyl acetate derivatives. *Tetrahedron Asymmetry* **13,** 2123–2131 (2002).

142. Rance, G. A. & Khlobystov, A. N. The effects of interactions between proline and carbon nanostructures on organocatalysis in the Hajos–Parrish–Eder–Sauer–Wiechert reaction. *Nanoscale* **6,** 11141–11146 (2014).

143. Hajos, Z. G. & Parrish, D. R. (+)-(7aS)-7a-Metyhyl-2,3,7,7a-Tetrahydro-1 H-Indene-1,5-(6H-Dione. *Org. Synth.* **63,** 26 (1985).

144. Buchschacher, P., Fürst, A. & Gutzwiller, J. (S)-8a-Methyl-3,4,8,8a-Tetrahydro-1,6(2H, 7H)-Naphthalenedione. *Org. Synth.* **63,** 37 (1985).

145. Fuhshuku, K., Tomita, M. & Sugai, T. Enantiomerically Pure Octahydronaphthalenone and Octahydroindenone: Elaboration of the Substrate Overcame the Specificity of Yeast-Mediated Reduction. *Adv. Synth. Catal.* **345,** 766–774 (2003).

146. Chenault, H. K. & Whitesides, G. M. Regeneration of Nicotinamide Cofactors for use in Organic Synthesis. *Appl. Biochem. Biotechnol.* **14,** 147–197 (1987).

147. Hioki, H., Hashimoto, T. & Kodama, M. Efficient kinetic resolution of (±) -4-methyl-Hajos – Parrish ketone by baker ' s yeast reduction. *Tetrahedron Asymmetry* **11,** 829–834 (2000).

148. Westen, H. H. Untersuchung von trans-l-Methyl-cyclodecen-(1)-on-(6) auf Atropisomerie mittels einer asymrnetrischen Synthese aus (1 S, 9 R, 10 R)-lO-Hydroxy-l-anethansul€onyloxy-9-methyl-decal. *Helv. Chim. Acta* **47,** 575–590 (1964).

149. Prelog, V. & Acklin, W. 90. Reaktionen mit Mikroorganismen. 1. *Helv. Chim. Acta* **39,** 748–751 (1956).

150. Yeo, S. K., Hatae, N., Seki, M. & Kanematsu, K. Enantioselective synthesis of an oxa-taxane derivative via tandem intramolecular [2+2] cycloaddition and [3,3]-sigmatropic rearrangement of allenyl ether. *Tetrahedron* **51,** 3499–3506 (1995).

151. Li, H., Yang, Y., Zhu, D., Hua, L. & Kantardjieff, K. Highly enantioselective mutant carbonyl reductases created via structure-based site-saturation mutagenesis. *J. Org. Chem.* **75,** 7559–7564 (2010).

152. Drager, B., Hashimoto, T. & Yamada, Y. Pseudotropine Forming Tropinone Reductase from Hyoscyamus. *Agric. Biol. Chem.* **52,** 2663–2667 (1988).

153. Jirschitzka, J. *et al.* Plant tropane alkaloid biosynthesis evolved independently in the Solanaceae and Erythroxylaceae. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 10304–9 (2012).

154. Dekkers, A. W. J. D., Verhoeven, J. W. & Speckamp, W. N. On the nature of sigma-coupled transitions. Through-bond interactions in 1-aza-adamantane derivatives. *Tetrahedron* **29,** 1691–1696 (1973).

155. Nakajima, K. & Hashimoto, T. Two tropinone reductases, that catalyze opposite stereospecific reductions in tropane alkaloid biosynthesis, are localized in plant root with different cell-specific patterns. *Plant Cell Physiol.* **40,** 1099–1107 (1999).

156. Brock, A., Herzfeld, T., Paschke, R., Koch, M. & Dräger, B. Brassicaceae contain

nortropane alkaloids. *Phytochemistry* **67,** 2050–2057 (2006).

157. Portsteffen, A., Draeger, B. & Nahrstedt, A. Two tropinone reducing enzymes from Datura stramonium transformed root cultures. *Phytochemistry* **31,** 1135–1138 (1992).

158. De Bastos, F. M. *et al.* Three different coupled enzymatic systems for in situ regeneration of NADPH. *Biotechnol. Tech.* **13,** 661–664 (1999).

159. Méndez-Sánchez, D., Mangas-Sánchez, J., Busto, E., Gotor, V. & Gotor-Fernández, V. Dynamic Reductive Kinetic Resolution of Benzyl Ketones using Alcohol Dehydrogenases and Anion Exchange Resins. *Adv. Synth. Catal.* **358,** 122–131 (2016).

160. Scheck, M., Koch, M. A. & Waldmann, H. Synthesis of a dysidiolide-inspired compound library and discovery of acetylcholinesterase inhibitors based on protein structure similarity clustering (PSSC). *Tetrahedron* **64,** 4792–4802 (2008).

161. Renouf, P., Poirier, J. & Duhamel, P. Asymmetric hydrolysis of pro-chiral ,3-disubstituted 2,4-diacetoxy- cyclohexa1,4-dienes. *J. Chem. Soc., Perkin Trans. 1* 1739–1745 (1997).

162. Shimizu, T., Hiranuma, A. & Nakat, T. Efficient Method for Inversion of Secondary Alcohols by Reaction of Chloromethanesuifonates with Cesium Acetate Takeshi Shimizu,* Sayoko Hiranuma, and Tadashi Nakata*. *Tetrahedron Lett.* **37,** 6145–6148 (1996).

163. Chiou, W.-H., Chiang, Y.-M. & Chen, G.-T. A facile synthesis of enantiopure 7-benzyloxycarbonyl-7-azabicyclo[2.2.1]heptane-2-carboxylic acid. *Tetrahedron: Asymmetry* **25,** 92–97 (2014).

164. Vernekar, S. K. V *et al.* Toward biophysical probes for the 5-HT3 receptor: Structure-activity relationship study of granisetron derivatives. *J. Med. Chem.* **53,** 2324–2328 (2010).

165. Burks, J. E. *et al.* Development of a Manufacturing Process for Zatosetron Maleate. **1,** 310–319 (1997).

166. Meltzer, P. C., Liang, a Y. & Madras, B. K. The discovery of an unusually selective and novel cocaine analog: difluoropine. Synthesis and inhibition of binding at cocaine recognition sites. *J. Med. Chem.* **37,** 2001–2010 (1994).

167. Majewski, M. & Zheng, G. Z. Stereoselective deprotonation of tropinone and reactions

of tropinone lithium enolate. *Can. J. Chem.* **70,** 2618–2626 (1992).

168. Meltzer, P. C., Liang, A. Y. & Madras, B. K. The Discovery of an Unusually Selective and Novel Cocaine Analog: Difluoropine. Synthesis and Inhibition of Binding at Cocaine Recognition Sites. *J. Med. Chem.* **37,** 2001–2010 (1994).

169. Armstrong, A., Ahmed, G., Dominguez-Fernandez, B., Hayter, B. R. & Wailes, J. S. Enantioselective Epoxidation of Alkenes Catalyzed by 2-Fluoro- N -Carbethoxytropinone and Related Tropinone Derivatives. *J. Org. Chem.* **67,** 8610–8617 (2002).

170. Lazny, R., Wolosewicz, K., Ratkiewicz, A., Pioro, D. & Stocki, M. Synthesis and isomer distribution of 2-alkyltropinones and 2-alkylgranatanones. *Tetrahedron* **70,** 597–607 (2014).

171. Willand, N. *et al.* Efficient, two-step synthesis of N-substituted nortropinone derivatives. *Tetrahedron Lett.* **48,** 5007–5011 (2007).

172. Bollinger, S., Hübner, H., Heinemann, F. W., Meyer, K. & Gmeiner, P. Novel pyridylmethylamines as highly selective 5-HT1A superagonists. *J. Med. Chem.* **53,** 7167–7179 (2010).

173. Kubas, H. *et al.* Scaffold hopping approach towards various AFQ-056 analogs as potent metabotropic glutamate receptor 5 negative allosteric modulators. *Bioorganic Med. Chem. Lett.* **23,** 6370–6376 (2013).

174. Cararas, S. A. *et al.* Further structure-activity relationship studies on 8-substituted-3-[2- (diarylmethoxyethylidenyl)]-8-azabicyclo[3.2.1]octane derivatives at monoamine transporters. *Bioorganic Med. Chem.* **19,** 7551–7558 (2011).

175. Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci.* **83,** 3746–3750 (1986).

176. Kantor, R. S. *et al.* Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. *Environ. Microbiol.* **17,** 4929–4941 (2015).

177. Smithies, K. *et al.* Stereoselectivity of an ω-transaminase-mediated amination of 1,3-dihydroxy-1-phenylpropane-2-one. *Tetrahedron Asymmetry* **20,** 570–574 (2009).

178. Richter, N., Simon, R. C., Kroutil, W., Ward, J. M. & Hailes, H. C. Synthesis of

pharmaceutically relevant 17-α-amino steroids using an ω-transaminase. *Chem. Commun. (Camb).* **50,** 6098–100 (2014).

179. Shin, G., Mathew, S., Shon, M., Kim, B.-G. & Yun, H. One-pot one-step deracemization of amines using ω-transaminases. *Chem. Commun.* **49,** 8629 (2013).

180. López, S., Fernández-Trillo, F., Castedo, L. & Saá, C. Synthesis of callyberynes A and B, polyacetylenic hydrocarbons from marine sponges. *Org. Lett.* **5,** 3725–3728 (2003).

181. Shimizu, M. & Uchimaru, F. Studies om N-Substituted Nortropane Derivatives. I. Synthesis of N-substituted 3-Nortropanones by the Robinson-Schopf Codensation. *Chem. Pharm. Bull. (Tokyo).* **9,** 300–303 (1961).

182. Hayakawa, Y., Baba, Y., Makino, S. & Noyori, R. General Synthesis of Tropane Alkaloids via the Polybromo Ketone-Iron Carbonyl Reaction. *J. Am. Chem. Soc* **100,** 1786–1791 (1978).

183. Zhao, X., Liu, D., Xie, F. & Zhang, W. Enamines: efficient nucleophiles for the palladium-catalyzed asymmetric allylic alkylation. *Tetrahedron* **65,** 512–517 (2009).

184. Cope, A. C., Nealyy, D. L., Peter, S. & Wood, G. Proximity Effects. XLIII. The Solvolysis of 4-Cyclooctene- 1 -methyl Brosylate. *J. Am. Chem. Soc.* **87,** 3130–3135 (1965).

185. Parrish, D. R. & Hajos, Z. Asymmetrische synthese polycyclischer organischer verbindungen. 1–36 (1977).

186. Bui, T. & Barbas, C. F. A proline-catalyzed asymmetric Robinson annylation reaction. *Tetrahedron Lett.* **41,** 6951–6954 (2000).

187. Labadie, J. W. *et al.* Synthetic utility of the palladium-catalyzed coupling reaction of acid chlorides with organotins. *J. Org. Chem.* **48,** 4634–4642 (1983).

188. Sevillano, L. G. *et al.* Inotropic activity of hydroindene amidinohydrazones. *J. Med. Chem.* **45,** 127–136 (2002).

189. Richard, J. A. & Chen, D. Y. K. A chiral-pool-based approach to the core structure of (+)-hyperforin. *European J. Org. Chem.* **6,** 484–487 (2012).

190. Hajos, Z. G., Parrish, D. R. & Oliveto, E. P. Total Synthesis of Optically Active (-)-17β-Hydroxy-Δ9(10)-desA-Androsten-5-one. *Tetrahedron* **24,** 2039–2046 (1968).

191. Hammoumi, A., Revial, G., D'Angelo, J., Girault, J. P. & Azerad, R. Microbial

hydroxylation and functionalization of synthetic bicyclic enones. *Tetrahedron Lett.* **32,** 651–654 (1991).

192. Stothard, P. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28,** 1102–1104 (2000).