



University of  
**Nottingham**

UK | CHINA | MALAYSIA

# **Development of Biocatalytic Approaches and Screening Methodology for the Synthesis of Nitrogen-Containing Small Molecules**

Doctor of Philosophy (PhD)

**Ryan Cairns MSci (Hons)**

Thesis Submitted to the University of Nottingham  
for the Degree of Doctor of Philosophy

2021

## Contents

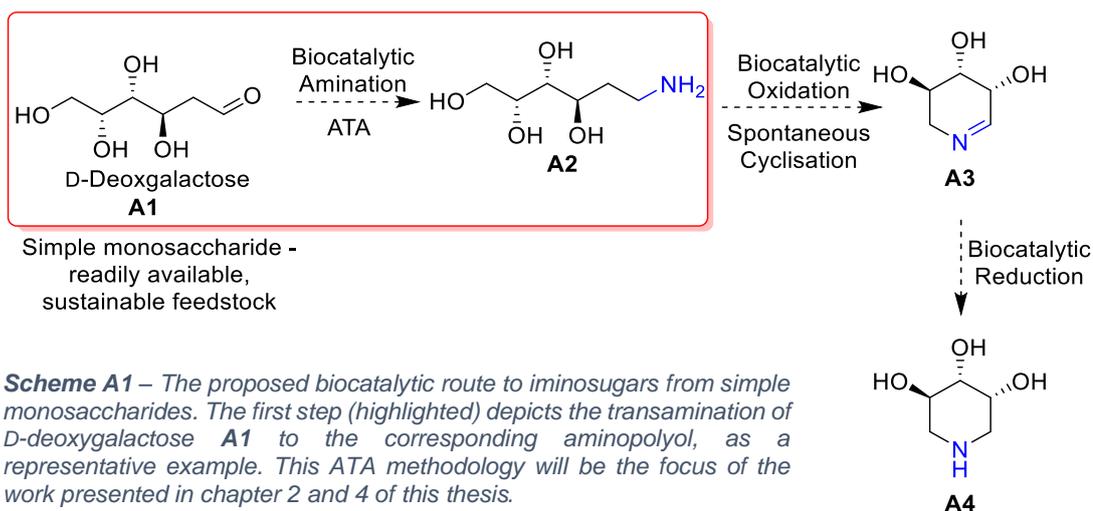
Abstract.....	1
Acknowledgments .....	6
Abbreviations, Acronyms, and Initialisms.....	8
1.0 Introduction.....	10
1.1 Biocatalytic cascades.....	12
1.2 Iminosugars – biological properties, applications, and synthesis .....	17
1.3 Biocatalytic route to iminosugars .....	20
1.4 Chiral amine synthesis.....	22
1.5.0 Transaminases .....	27
1.5.1 Transaminase mechanism.....	28
1.5.2 Mechanistic origin of substrate specificity and stereoselectivity .....	30
1.5.3 Thermodynamic challenges.....	31
1.5.4 Synthetic applications of ATAs.....	35
1.5.5 Exploring novel ATAs .....	39
1.6 Enzyme engineering .....	39
1.7 Screening for ATA activity .....	44
1.8 Project aims and objectives .....	48
2.0 Conversion of aldoses to valuable $\omega$ -aminopolyols using transaminases.....	50
2.1 Introduction .....	50
2.2 Aims and Objectives .....	53
2.3.0 Results and discussion.....	53
2.3.1 Screening for activity.....	53
2.3.2 Analytical Biotransformations .....	54
2.3.3 Preparative scale biotransformations.....	61
2.4 Conclusion .....	62
2.5 Experimental.....	63
3.0 Development of a Comprehensive Quantitative Assay for Amine Transaminases .....	73
3.1 Introduction.....	73
3.2 Aims and objectives .....	76
3.3.0 Results and discussion.....	76
3.3.1 Screening for ketone acceptors using 2-aminoethylaniline (2AEA) .....	76
3.3.2 Quantitative acceptor screen development.....	79
3.3.3 Quantitative amine donor screen development.....	81
3.3.4 Enantiopreference screen.....	85

3.3.5 Development of a solid-phase acceptor screen .....	89
3.4 Conclusion .....	92
3.5 Experimental.....	93
4.0 Molecular modelling of monosaccharides in $\omega$ -ATAs and the semi-rational engineering of 3HMU towards ketose activity.....	104
4.1 Introduction.....	104
4.2 Aims and objectives .....	107
4.3.0 Results and discussion.....	108
4.3.1 Molecular modelling of monosaccharides in $\omega$ -ATAs.....	108
4.3.2 Site directed mutagenesis of 3HMU and HEWT to increase the size of the small pocket .....	116
4.3.3 Optimisation of protein isolation from 48-well plate .....	119
4.3.4 Combinatorial mutagenesis of 3HMU .....	120
4.4 Conclusions and future work .....	122
4.5 Experimental.....	123
5.0 A Transaminase-Mannich Cascade Towards the Generation of Complex Alkaloids .....	130
5.1 Introduction.....	130
5.2 Aims and Objectives .....	134
5.3.0 Results and discussion.....	135
5.3.1 Synthesis of Alkaloid Chemical Standards.....	135
5.3.2 Mannich Reaction Under Biotransformation Conditions .....	136
5.3.3 Time and Temperature Screen .....	138
5.3.4 Amine Donor Concentration Screen.....	140
5.3.5 Whole Cell Biocatalysis Using the Cadaverine Producing <i>C. glutamicum</i> Strain Cada1 .....	142
5.3.6 Preparative Scale Biotransformation towards a Complex Alkaloid Using L- and D-Proline.....	144
5.4 Conclusions and Future Work .....	145
5.5 Experimental.....	146
6.0 Final Conclusions and Future Work.....	156
7.0 Appendix .....	160
8.0 References .....	171

## Abstract

As we move towards a more sustainable global economy, there is a need to develop efficient methods for the synthesis of key chiral molecules that are particularly vital for the pharmaceutical and agrochemical industries. Catalysis has opened up a plethora of synthetic opportunities and biocatalysis, specifically, has now emerged as a sustainable alternative to other catalytic approaches. The application of amine transaminases (ATAs) for the conversion of pro-chiral carbonyls to the corresponding chiral amines in the presence of a suitable amine donor and the coenzyme pyridoxyl-5'-phosphate (PLP) has been well established. While their commonly cited limitations include narrow substrate scope and/or incompatible operational temperatures/pH, there is significant literature precedent for overcoming these challenges through enzyme engineering endeavours, which typically require reliable, high throughput selection assays.

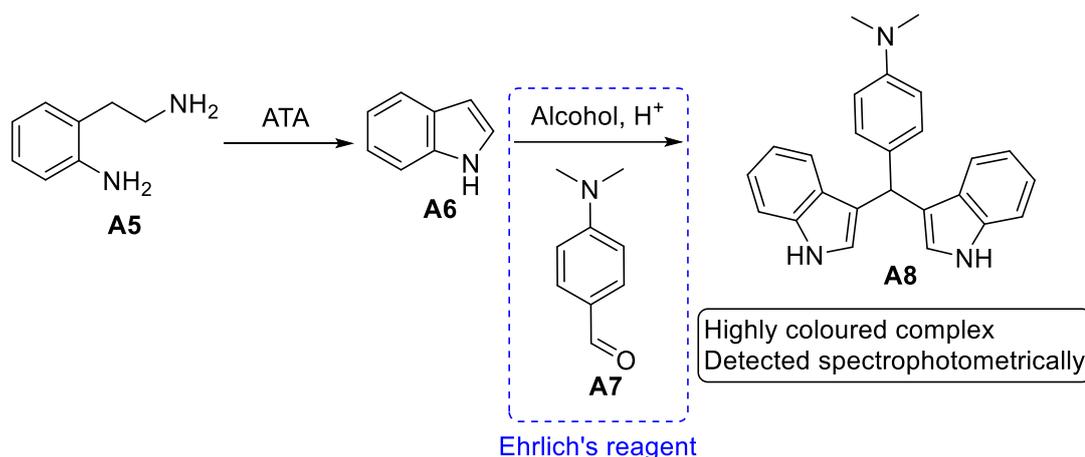
The O'Reilly group is interested in the biocatalytic synthesis of iminosugars from simple monosaccharides. The approach relies on the biocatalytic conversion of a monosaccharide, such as D-deoxygalactose (**A1**), to the corresponding aminopolyol **A2** (**Scheme A1**) using an ATA. Following installation of the amine, regioselective biocatalytic oxidation and spontaneous cyclisation would afford cyclic imine **A3**, which could be selectively reduced to the desired iminosugar **A4**. The work presented in this thesis will predominantly focus on the first step of this process, the enzymatic transamination.



Chapter 1 introduces the field of biocatalysis, with particular attention given to ATAs and their synthetic applications/limitations, and concludes with a summary of the aims and objectives of this doctoral thesis. Chapter 2 describes the previously unreported direct amination of a panel of monosaccharides, which exist predominantly in their cyclic form, using ATA biocatalysts. The use of three commercially available ATAs that are active on a small panel of aldoses is investigated on an analytical scale, allowing optimal conditions with these substrates to be determined. Following this initial optimisation, a selection of amino alcohols were synthesized on a preparative scale using commercially available (*S*)-selective ATA-256. This methodology allowed for facile access to valuable amino alcohol products (up to 475 mg), clearly demonstrating the scalability of this approach and positioning it as a potentially ideal route for the synthesis of such amino alcohols from simple monosaccharides. While this work represented an exciting expansion to the substrate panel available to these enzymes, it is noted that only aldoses were identified as substrates and it is desirable to expand this methodology to

enable the direct conversion of ketoses, where the enzyme mediates the introduction of a chiral amine.

At the commencement of this work, a comprehensive quantitative, high-throughput assay allowing the rapid identification of ATA variants with desired activity from large mutant libraries was not available. As we were interested in expanding our monosaccharide transamination methodology (Chapter 2) to include ketoses, it was necessary to design a robust assay to complement enzyme engineering. The focus of the 3<sup>rd</sup> chapter is on the development of a general and operationally simple end-point assay for ATAs that enables the screening of both amine donors and acceptors in liquid phase. This spectrophotometric screen exploits the novel amine donor 2-aminoethylaniline (2-AEA) (**A5**) and the subsequent reaction of *in situ* generated indole (**A6**) with Ehrlich's reagent, a common reagent to test for the presence of indole in biological systems (**Scheme A2**). A modification of this method also allowed for the screening of amine acceptors, utilising 4-dimethylaminobenzaldehyde (4-DMAB) (**A7**) as an amine donor, whereby indole is added exogenously, and its subsequent depletion measured spectrophotometrically. This assay proved to be the first fully quantitative and qualitative screen that also allowed for the facile determination of ATA enantioference. To increase the throughput, an engineered expression strain (*E. coli* BL21(DE3)  $\Delta tnaA$ ) lacking tryptophanase activity, was generated to enable reliable and direct evaluation of individual colonies on the solid phase.

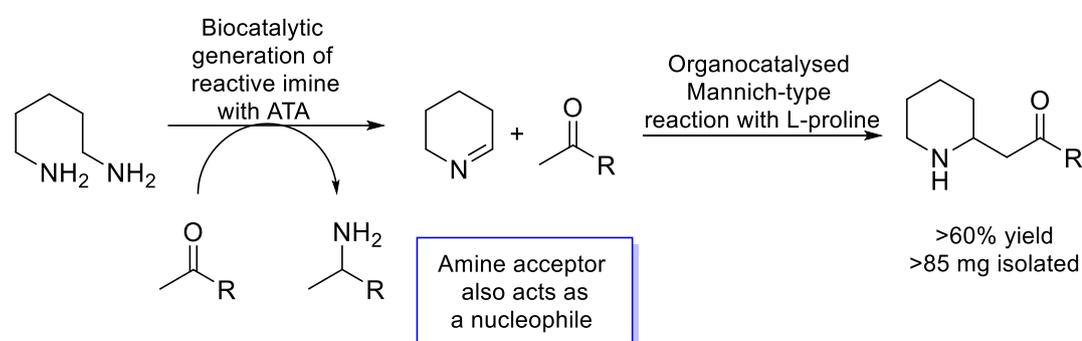


**Scheme A2** – The transamination of 2-aminoethylamine (2-AEA) (**A5**) to generate indole (**A6**), followed by reaction of **A6** with Ehrlich's reagent (including 4-dimethylaminobenzaldehyde (**A7**)) generating a highly coloured species **A8**.

From the preliminary studies, only aldose sugars were identified as readily accepted substrates. However, the production of chiral amino alcohols from ketose sugars would be of great interest to the chemical industry. A recent publication has identified an (*R*)-selective ATA that can perform this transformation on a small panel of ketose sugars. As such, Chapter 4 aims to expand on the work reported in chapter 2 by exploring the structural basis for aldose/ketose selectivity using molecular modelling and subsequent semi-rational engineering. Docking experiments were performed to identify key residues in the active site of 3HMU (from *Rugeria pomeroy*), HEWT (from *Halomonas elongata*) and 3I5T (from *Rhodobacter sphaeroides*) that are ideal for rational design and rounds of iterative saturation mutagenesis, and may expand the substrate scope of these (*S*)-selective enzymes to include ketoses.

There is continued interest in developing cascade processes towards bioactive natural product analogues and chiral building blocks, such as complex alkaloids. The 5<sup>th</sup>, and final chapter strays away from monosaccharides and reports a bio-organocatalytic cascade for the synthesis of a small panel of complex alkaloids, relying on an ATA to generate a key reactive intermediate

for the complexity building organocatalysed Mannich reaction (**Scheme A3**). The successful transamination of cadaverine using commercially available ATA-256 and a variety of ketones as the amine acceptor, led to the production of  $\Delta^1$ -piperideine, which that could then further react with the ketone starting material in the presence of proline, affording 2-substituted N-heterocyclic alkaloids. Conversions of up to 75% were achieved using this methodology and the preparative scale synthesis of the natural product pelletierine (60% yield, 85 mg) is reported.



**Scheme A3** – A bioinspired route for the generation of 2-substituted piperidine alkaloids via a Mannich-type reaction between simple methyl ketones and in situ generated  $\Delta^1$ -piperideine.

## **Acknowledgments**

Firstly, I must thank Dr Elaine O'Reilly for providing me with extensive support and guidance throughout this entire process. Your insightful feedback pushed me to sharpen my thinking and gain a wealth of research experience across an exciting interdisciplinary space.

I would also like to thank Dr James Ryan for not only his continued assistance as our postdoc, but, also as a good friend and housemate. My transition to Dublin would not have been so easy without you. I would also like to thank Dr Michael Sharkey and Dr Andrew Gomm for their advice and stimulating conversation. Much of this work would not have been possible without your help. I extend further thanks to the rest of the O'Reilly group, past and present, for providing such an entertaining and supportive space to work. A special shoutout is reserved for Dr Stylianos Grigoriou and Dr Justyna Kuska for being such patient instructors and good friends.

I would also like to thank the fantastic analytical and support staff across both the University of Nottingham and University College Dublin for their wealth of support and advice. A further thanks must go to Kevin Butler for his assistance regarding quantitative NMR, without this my first publication might still be a distant pipedream.

In addition, I must thank my parents, none of this would be possible without your innumerable sacrifices and encouragement. A special thanks also goes to my grandfather, 'Granda Jimmy', for igniting and fuelling my thirst for knowledge. I also thank the rest of my family for their unconditional support of my pursuits and providing me with a space away from the stresses of PhD life.

Finally, these acknowledgments would not be complete without a special thank you to my closest friends Chantelle Flynn, Jack Ayre, and my partner Seán Keane. You have each provided me with more support and welcomed distractions than you know, go raibh míle maith agat!

## Abbreviations, Acronyms, and Initialisms

<b>2-AEA</b>	2-(Aminoethyl)aniline
<b>4-DMAB</b>	4-Dimethylaminobenzaldehyde
<b>4-DMACA</b>	4-Dimethylaminocinnamaldehyde
<b>AcN</b>	Acetonitrile
<b><i>AspFum</i></b>	<i>Aspergillus fumigatus</i>
<b>ATA</b>	Amine transaminase
<b>BHIS</b>	Brain heart infusion broth
<b><i>Cvi</i></b>	<i>Chromobacterium violaceum</i>
<b>DMSO</b>	Dimethyl sulfoxide
<b>e.e.</b>	Enantiomeric excess
<b>EtOAc</b>	Ethylacetate
<b>FADH</b>	Flavin adenine-dinucleotide
<b>FTIR</b>	Fourier transform infrared spectroscopy
<b>GC-FID</b>	Gas chromatography - flame ionising detection
<b>GDH</b>	Glucose dehydrogenase
<b>HEPES</b>	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
<b>HEWT</b>	<i>Halomonas elongata</i>
<b>HPLC</b>	High performance liquid chromatography
<b>IRED</b>	Imine reductase
<b>IPA</b>	Isopropylamine
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>KPi</b>	Potassium phosphate buffer
<b>LB</b>	Lysogeny broth
<b>LDH</b>	Lactate dehydrogenase

<b>MBA</b>	Methylbenzylamine
<b>MeOH</b>	Methanol
<b><i>MycVan</i></b>	<i>Mycobacterium vanbaalenii</i>
<b>NAD(H)</b>	Nicotinamide adenine dinucleotide
<b>NADP(H)</b>	Nicotinamide adenine dinucleotide phosphate
<b>NMR</b>	Nuclear Magnetic Resonance
<b><i>o</i>-xyl</b>	<i>o</i> -xylylenediamine
<b>PCR</b>	Polymerase chain reaction
<b>PDB</b>	Protein data bank
<b>PLP</b>	Pyridoxal-5'-phosphate
<b>PMP</b>	Pyridoxal-5'-phosphate
<b>SDS-PAGE</b>	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
<b>TB</b>	Terrific broth
<b>WT</b>	Wild-type

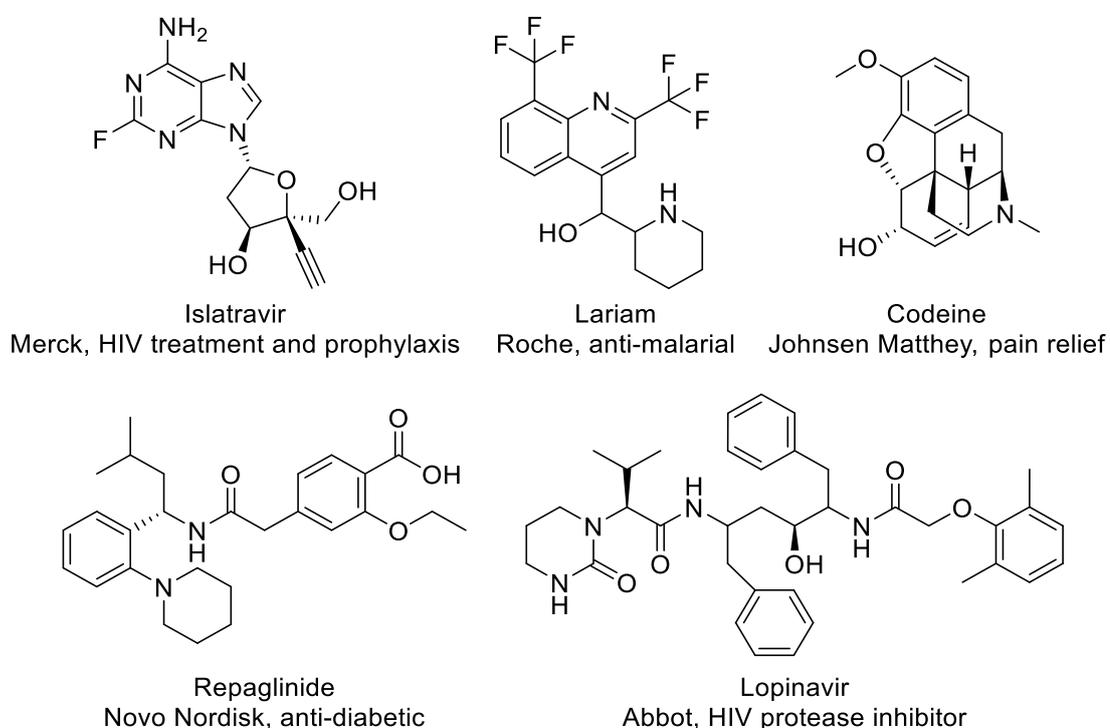
## 1.0 Introduction

The field of biocatalysis has become hugely important to the chemical industries, as it allows for green and sustainable transformations to be performed with excellent selectivity and often replaces syntheses that require the use of organic solvents or metal catalysis.<sup>1-4</sup> Protein engineering enables the optimisation of wild-type enzymes and allows them to carry out reactions with increased substrate scope and productivity. Indeed, this has now been extended to the design of *de novo* enzymes that can perform reactions not seen in nature.<sup>5,6</sup> Enzymatic reactions are no longer restricted to the milligram scale, but play an important role in the industrial preparation of pharmaceuticals<sup>7</sup>, foodstuffs<sup>8</sup>, organics<sup>9</sup>, biofuels<sup>10</sup> and the degradation of plastics<sup>11,12</sup>, to name a few examples. The global market for these specialty enzymes is expected to be worth \$4.044 billion by 2023.<sup>13</sup>

Enzymes are often highly enantioselective and allow for the synthesis of optically pure compounds. More than half of the drugs currently on the market contain at least one chiral moiety, and although enantiomers typically display similar chemical properties, within the chiral environment of the protein, they often exhibit vastly different biological activities such as pharmacology, pharmacokinetics, metabolism and toxicology. One such example is ketamine, an intravenous anaesthetic, where the (*S*)-(+)-isomer is a better NMDA receptor antagonist than the (*R*)-(-)-isomer and hence a much more potent and less toxic drug.<sup>14,15</sup>

Chiral amines consistently prove to be powerful pharmacophores, often explained by their strong predilection for hydrogen bonding (**Figure 1**). Such drugs overwhelmingly contain secondary or tertiary amines that are often

masked as amides. As a result, these motifs are most commonly synthesised via alkylation/amide formation, starting from of the corresponding  $\alpha$ -chiral primary amine and as such,  $\alpha$ -chiral amines are a vital building block in pharmaceutical drug synthesis.<sup>16</sup> This is evidenced by the fact that a roundtable of leading pharmaceutical corporations voted for the sustainable synthesis of chiral amines amongst the key green chemistry research areas of the decade.<sup>17</sup> As we move towards a global green economy, there is significant interest from many of these companies to manufacture such drugs utilising biocatalytic or chemoenzymatic routes.<sup>18–20</sup> An exciting recent example from Merck includes the design of an *in vitro* biocatalytic cascade for the manufacture of islatravir, an investigational HIV drug (**Figure 1**).<sup>21</sup> This involved a three step cascade including 9 enzymes, resulting in a synthesis requiring less than half the steps of previously reported synthetic routes.



**Figure 1** - Small selection of amine-based pharmaceutical drugs currently on the market or in human trials.

## 1.1 Biocatalytic cascades

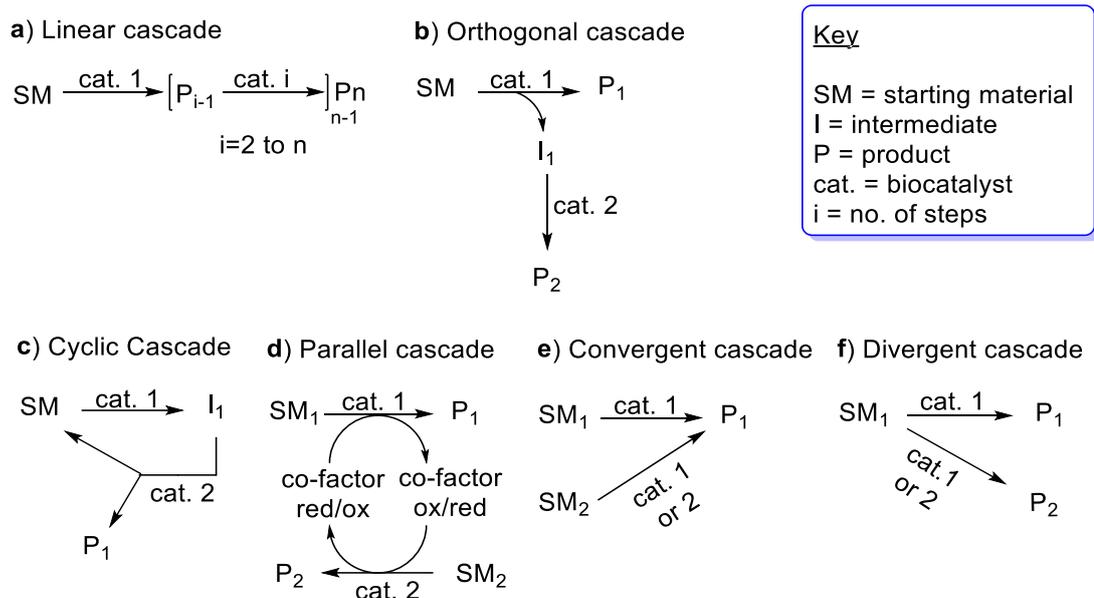
Nature has evolved to convert a limited pool of simple starting materials to a plethora of multi-functionalised natural products. In most cases, these complex biomolecules are built through the actions of multiple enzymes acting in tandem *via* complex biosynthetic pathways in organisms. These cascade reactions take place in the cell surrounded by numerous other bio-available molecules, and it is therefore essential that the enzymes involved in these pathways are highly selective. These enzymes have already been perfectly adapted to work under similar physiological conditions (pH, temperature etc.) in aqueous media and this makes them suitable for the design and application of 'artificial' biocatalytic cascades that can generate valuable natural and non-natural organic molecules.<sup>22-24</sup>

The design of these biocatalytic cascades must consider the requirements for any successful route; i.e. the overall cascade is thermodynamically feasible ( $\Delta G_{\text{cascade}} < 0$ ) and all enzymes involved must be highly selective with high functional group orthogonality to avoid unwanted cross reactivity. This allows for the coupling of thermodynamically unfavourable steps to achieve a more favourable position of equilibrium.<sup>25</sup> Cascades circumvent the need for the purification/isolation of intermediates, substantially reducing the time required and waste generated when compared to traditional chemical synthesis. As the product of one reaction immediately becomes the starting material of the next catalyst, they can also prevent major limitations prevalent in one-step transformations, including unstable product/intermediate(s) and inhibition. Furthermore, as the toolbox of available biocatalysts expands, so too does the complexity of the transformations that can be introduced to these systems.<sup>6,26,27</sup>

These processes have been developed both *in vitro* and *in vivo*. *In vitro* cascades can use purified enzymes, cell free extracts, or freeze-dried whole cells and are advantageous as they allow for the addition of known amount of each enzyme to carefully control the reaction kinetics to maximise the yield of desired product(s). However, the industrial scale-up of such processes can become unfeasible, as many enzymes require expensive co-factors, which must either be introduced to the system stoichiometrically or regenerated by a recycling system.<sup>28,29</sup> This can be overcome in *in vivo* systems whereby the enzymes are simultaneously expressed in host organisms, which are also constantly generating the necessary co-factor(s). However, the use of whole cells can lead to downstream processing issues due to the presence of cell metabolites, although the environmental cost of purification is still often much less than traditional chemical synthesis.<sup>30</sup> Another limitation of such one-pot systems is that all enzymes and chemistry occurring in the process must be fully compatible; i.e. the operating pH, solvent tolerance and temperature must be similar and the product/starting material of one catalyst must not be the inhibitor of another. This is not always possible and has led to the engineering of compartmentalised flow systems whereby non-compatible enzymes are fully separated, and unused starting material can be pumped back into the cascade to maximise product recovery.<sup>31,32</sup>

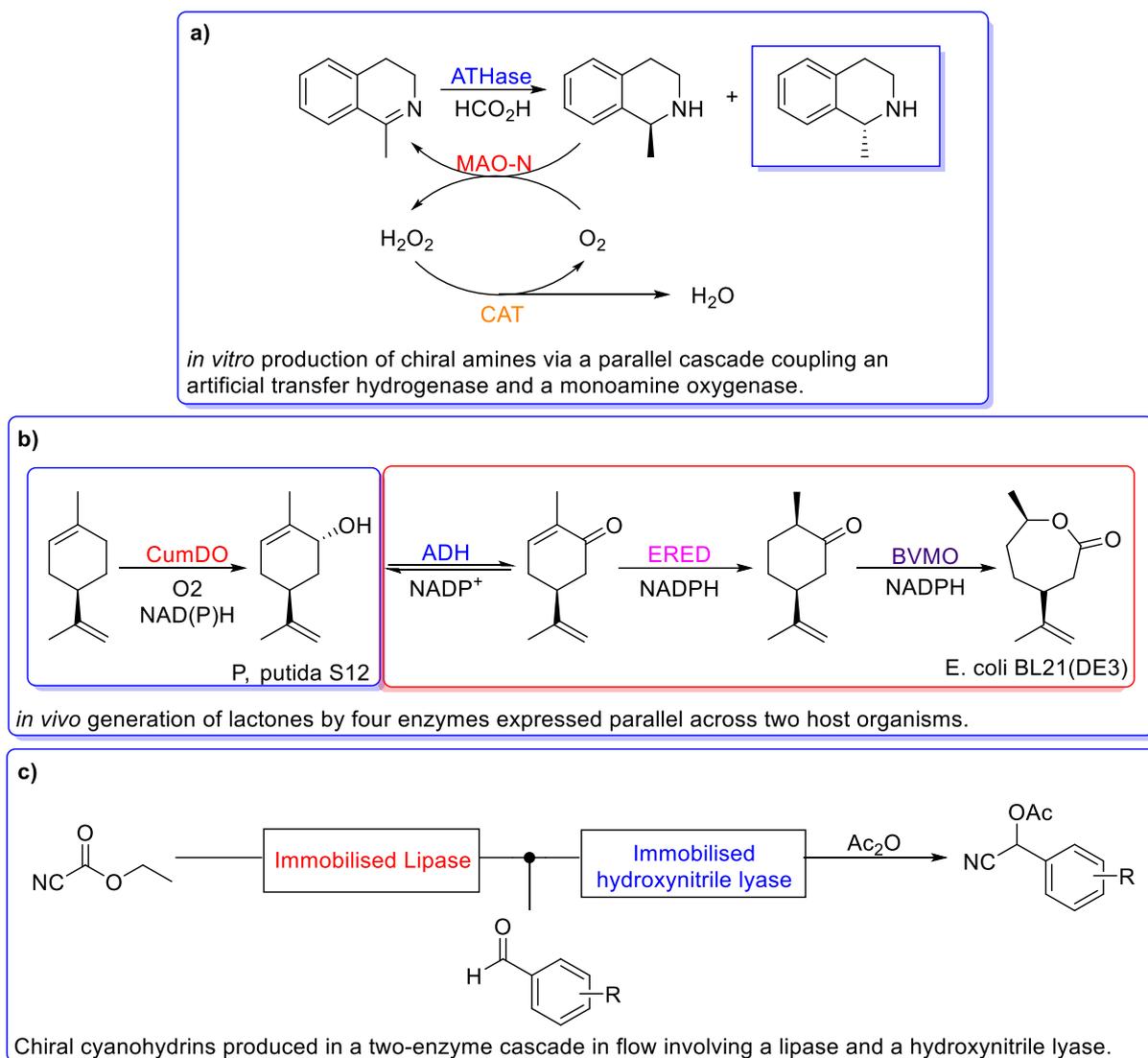
The first reported study of an enzymatic cascade conducted *in vitro* was the transformation of sugars to ethanol and CO<sub>2</sub>, which led researchers to the enzymes and intermediates of glycolysis.<sup>33</sup> This was followed by the first artificial *in vitro* cascade involving two enzymes; the transformation of lactic acid to L-alanine via pyruvate in a flow system in 1984.<sup>34,35</sup> Since then, the

complexity and types of reported cascades has soared (**Scheme 1**). This led to Kroutil *et al.* suggesting a classification system for biocatalytic cascades that considers things such as; number of steps, number of catalysts, types of catalyst, reaction topology, mode of cofactor regeneration, spatial organisation of the biocatalysts and the final product formed.<sup>22</sup>



**Scheme 1** – Simplified illustration of the concepts behind the diverse types of biocatalytic cascades abundant in the literature. a) linear cascade. b) Orthogonal cascade. c) Cyclic cascade. d) Parallel cascade. e) Convergent cascade. f) Divergent cascade.

The concepts described thus far have been applied to generate a wide variety of functionalised organic molecules. These include industrially relevant motifs such as; chiral amines produced *via* a parallel cascade coupling an artificial transfer hydrogenase and a monoamine oxygenase *in vitro* (**Scheme 2a**),<sup>36</sup> lactones generated *in vivo* by four enzymes expressed parallel in two host organisms (**Scheme 2b**),<sup>37</sup> and chiral cyanohydrins produced in a two enzyme cascade in flow involving a lipase and a hydroxynitrile lyase (**Scheme 2c**).<sup>38</sup> This brief snapshot of reported cascades highlights the versatility of functionalisation and complexity available to these systems.



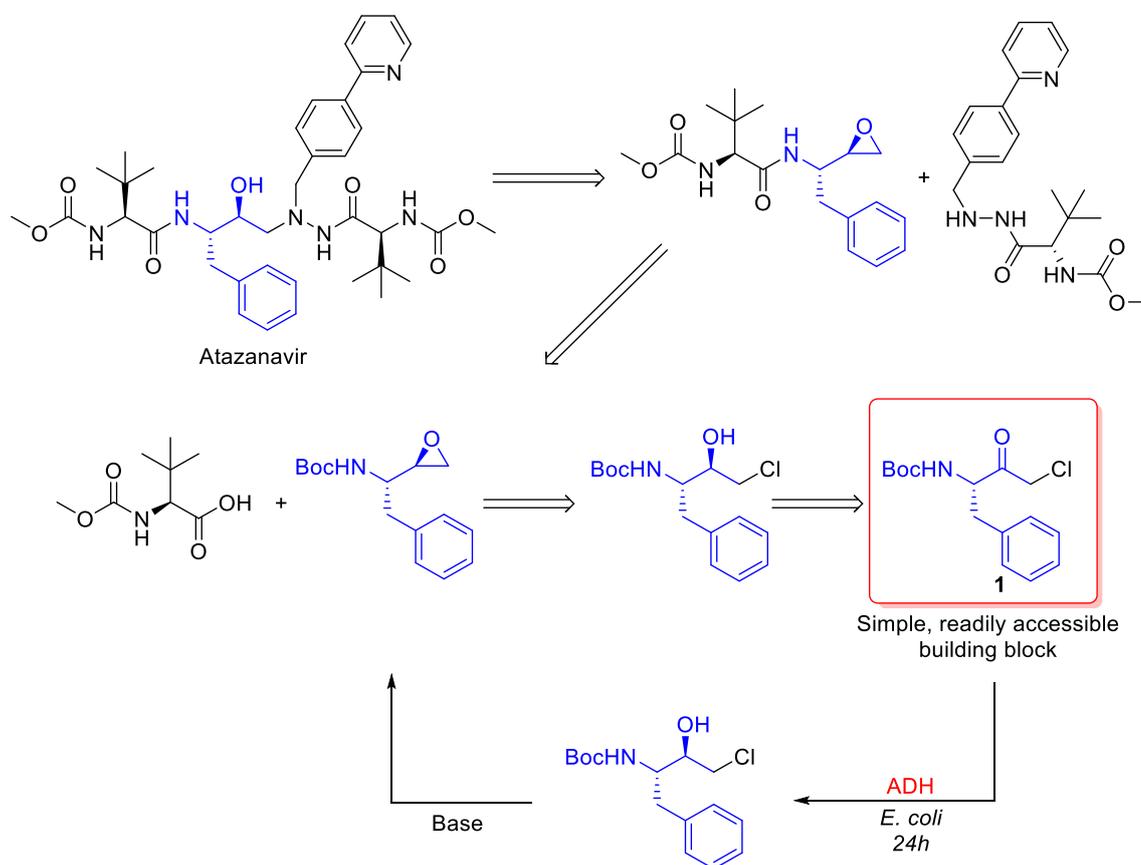
**Scheme 2** – The generation of industrially relevant motifs via enzymatic cascades. a) The *in vitro* production of chiral amines via a parallel cascade coupling an artificial transfer hydrogenase and a monoamine oxygenase. b) The *in vivo* generation of lactones by four enzymes expressed parallel across two host organisms. c) Chiral cyanohydrins produced in a two-enzyme cascade in flow involving a lipase and a hydroxynitrile lyase.

The toolbox of commercially available enzymes is continuously increasing in size and diversity, allowing chemists to avail of a wide selection of catalysts that can perform a varied range of transformations on a broad array of substrates. This has led to the concept of biocatalytic retrosynthesis maturing over the past decade, allowing undergraduate and research chemists alike to avail of its principles to design elegant cascades to target molecules.<sup>39–41</sup> Retrosynthetic analysis was first proposed by E.J. Corey in the 1960's and is

based upon the systematic disconnection of chemical bonds linking major components of synthetic targets until simple building blocks or readily available starting materials are derived.<sup>42</sup> Here, synthons are generated by the cleaving of C-X or C-C bonds in the reverse direction, in order to achieve this, functional group interconversions (FGIs) are commonly used to allow for synthetically reliable disconnections. Importantly, each disconnection must correspond to a known synthetic transformation and this thinking has historically inspired the development of new synthetic chemistry transformations. This approach revolutionised how synthetic chemists design routes to target molecules and is now widely used in organic chemistry groups.

Biocatalytic retrosynthesis is an analogous concept, whereby synthetic disconnections and FGI's are replaced by enzyme catalysed reactions. This has been founded on the basis that biocatalysts are readily available for the key bond forming steps and enzymatic FGIs can be used to prepare the building blocks subsequently derived. This methodology can improve the yield and mass recovery of multi-step syntheses that use traditional synthetic chemistry. One such example is in the synthesis of Atazanavir (**Scheme 3**), an antiretroviral medication for HIV/AIDS. Here, a retrosynthetic analysis leads to an  $\alpha$ -halo ketone intermediate that is chemically reduced using lithium hydride to the desired diastereoisomer in moderate to good yields.<sup>43</sup> However, de Miranda *et al.* report a biocatalytic approach using an alcohol dehydrogenase (ADH) that can achieve 99% conversion and 84% *de* after 24 h (**Scheme 3**).<sup>44</sup> This illustrates how a biocatalytic retrosynthesis approach can significantly increase the efficiency and sustainability of syntheses towards industrially relevant molecules.

The O'Reilly group is particularly interested in applying the principles of biocatalytic retrosynthesis to produce natural products and synthetically challenging chiral molecules.

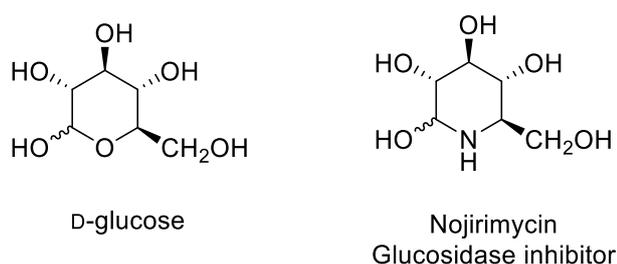


**Scheme 3** - A retrosynthetic disconnection of Atazanavir, an antiretroviral medication for HIV/AIDS, including a forward biocatalytic synthesis to a key intermediate from a readily accessible building block (**1**) using an alcohol dehydrogenase (ADH).

## 1.2 Iminosugars – biological properties, applications, and synthesis

Iminosugars are a class of biologically active compounds that resemble monosaccharides, where the ring oxygen is replaced with a nitrogen, forming polyhydroxylated secondary and tertiary amines (**Figure 2**).<sup>45,46</sup> They were historically identified and studied for their ability to interact with and inhibit glycosidases, which catalyse the hydrolysis of glycosidic bonds in complex sugars. Mainstream pharmaceutical companies had a long-standing reluctance to working with carbohydrate-like drug targets as they were deemed

too polar for delivery and bioavailability. Nevertheless, many of the iminosugars studied so far display excellent oral bioavailability. Glyset® and Zavesca® are two commercially available drugs that were derived from the glucosidase inhibiting natural product l-deoxynojirimycin (DNJ) (**Figure 2**), and are used to treat diabetes mellitus type 2 and type I Gaucher disease (GD1) respectively.<sup>47,48</sup>



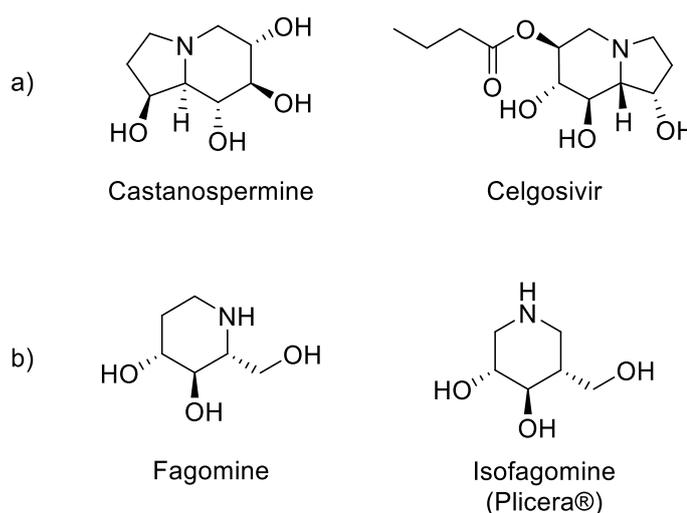
*Figure 2 - Structure of the much-studied iminosugar, nojirimycin, compared with its monosaccharide equivalent, D-glucose.*

It is now understood that this class of compound is involved in a broad range of biological activities that do not involve glycosidase inhibition, and in fact, this inhibition can lead to off-target effects therapeutically. Iminosugars can act as chaperones for misfolded proteins by binding to catalytic sites, leading to better trafficking in the endoplasmic reticulum and improvement in function.<sup>49,50</sup> They can function as immune modulators, antiviral agents (against herpes simplex virus, bovine diarrhoea virus, hepatitis C and HIV-1), display anticancer activity and possess the ability to be non-competitive and competitive inhibitors of carbohydrate manipulation enzymes.<sup>51-57</sup> Such enzymes are involved in the attachment of oligosaccharides to the surface of tumour cells and play a key role in the malignant phenotype of tumour growth.<sup>58</sup>

The specificity of enzyme inhibition is key to utilising such therapies without off-target effects. One such example is with castanospermine, which is an  $\alpha$ -

glucosidase I inhibitor that exhibits antiviral activity towards several viruses (**Figure 3a**). However, the compound also acts as an inhibitor of intestinal sucrases, leading to osmotic diarrhoea.<sup>59,60</sup> On the other hand, the 6-*o*-butanoyl derivative, celgosivir, exhibits 30-fold greater activity, does not act as a sucrase inhibitor and appears to be non-toxic to the gastrointestinal tract (**Figure 3a**).<sup>61</sup>

There is a vast structural space available for small molecules such as iminosugars, but thus far most of these compounds have been isolated from plants as natural products.<sup>62</sup> Several groups have broadened the versatility of motifs available by synthesising iminosugar derivatives using these natural iminosugars as templates. Plicera® is an example of a therapeutic drug that was designed using a natural product template, fagomine (**Figure 3b**). The natural product exhibits a weak ability to bind to glucocerebrosidase, whereby the modified drug has a threefold increase in activity and is used to treat Gaucher's disease (**Figure 3b**).<sup>61</sup>



**Figure 3** - a) Example of an antiviral drug, celgosivir, that is a derivative of the natural product castanospermine currently in clinical trials. The modified natural product exhibits improved activity and higher specificity/less off-target activity. b) Example of a marketed drug, Plicera®, which was designed using the natural product fagomine as a template.

Despite their apparent wide range of possible uses, the high toxicity associated with the 1st generation of iminosugars has hampered their development as pharmaceutical drugs and the difficulties associated with synthesising structurally diverse derivatives has been cited as a key reason for their lack of development.<sup>45</sup> To date iminosugars intended for therapeutic purposes, such as DNJ, have been extracted from higher plants (mulberry leaves/dayflower)<sup>63,64</sup> resulting in a limited availability for study. To overcome this issue it has been suggested that the compounds could instead be extracted from the fermentation of microorganisms, such as *Bacillus* strains, however, the yield from this process is currently underwhelming.<sup>65</sup> Traditional chemical approaches for the synthesis of iminosugars have proved difficult, largely due to the analogous reactivity of the hydroxyl groups found on simple sugars. Furthermore, they often involve many steps that use metal catalysis, high pressure, extreme temperatures and/or expensive ligands with a poor overall yield resulting in large amounts of waste.<sup>66-68</sup> Therefore, a much more sustainable synthesis capable of producing a diverse selection of iminosugars is required to allow for a true exploration of the structural space available.

### **1.3 Biocatalytic route to iminosugars**

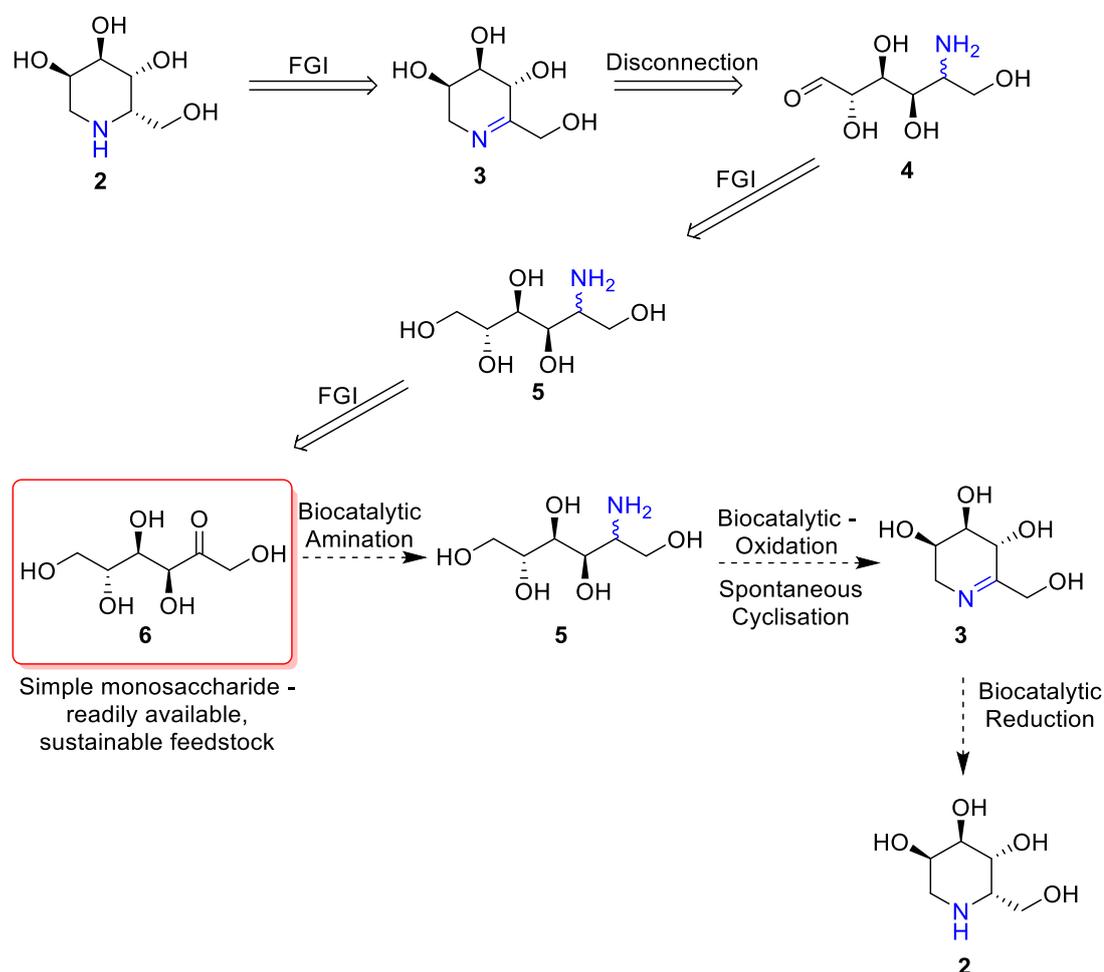
The O'Reilly group proposes using an entirely biocatalytic approach to iminosugar synthesis. The high regioselectivity of enzymes would avoid the challenges often associated with the use of polyhydroxylated reagents with analogous reactivities that are often encountered in traditional chemical synthesis, whilst also providing a much more 'green' and sustainable process. Retrosynthetic analysis of iminosugar **2** results in a simple monosaccharide starting material (**6**) after two FGIs and a disconnection (to a minor equilibrium

state) (**Scheme 4**) and this route is transferrable to numerous other motifs. Such sugars are low-cost, readily available starting materials that can be obtained from sustainable feedstocks. Our group proposes the forward reaction could proceed *via* a three step biocatalytic synthesis. Initially the existing carbonyl is aminated, which is followed by oxidation of a distal primary alcohol, resulting in spontaneous cyclisation to a cyclic imine (**3**). This is finally reduced to the iminosugar, and the approach can be substantially diversified by the choice of monosaccharide.

The initial phase of the overall research programme requires that enzymes are identified that can carry out each individual step of this process, and can subsequently be combined in an *in vitro* or *in vivo* biocatalytic cascade to generate a wide variety of iminosugars. This will involve screening the substrates against large enzyme libraries, likely followed by the engineering of selected 'hits' to maximise conversion. Once this has been optimised *in vitro* the same enzymes can be added to a host organism whereby the cascade takes place *in vivo*. These organisms can produce the sugar starting materials in large quantities *via* fermentation.<sup>69</sup> It is foreseen that this will eventually lead to a fully sustainable process, where all the catalysts, reagents and cofactors are produced *in vivo* and the iminosugar product can be isolated from the resultant broth.

There is already a significant literature precedent for the biocatalytic reduction of cyclic imines using imine reductases (IREDs) or monoamine oxidases (MAO's).<sup>70</sup> Several sugar oxidases have also been described that are capable of oxidising primary alcohols on polyhydroxylated reagents.<sup>71</sup> However, at the commencement of this project, there were no reports of an enzymatic

amination of the carbonyl on sugars that exist predominantly in their cyclic form at equilibrium. For this reason, priority was placed on finding a suitable candidate for this transformation.



*Scheme 4 - A retrosynthetic analysis of an iminosugar to a simple monosaccharide. Followed by a three-step forward biocatalytic synthesis proposed by the O'Reilly group.*

## 1.4 Chiral amine synthesis

Traditionally, chiral amines had been prepared by the resolution of racemic mixtures using chiral chromatography (HPLC) or dynamic kinetic resolution (DKR).<sup>72,73</sup> However, a plethora of research has gone into the development of asymmetric syntheses of chiral amines and many of these synthetic routes require metal catalysis, high temperature/pressure and are performed in organic solvents.<sup>16</sup> Hence, they are often non-sustainable and have prompted

the development of 'green' biocatalytic routes to chiral amines, as the chemical industry strives to achieve a higher level of sustainability.<sup>74-76</sup>

The structural scope of chiral amines is immensely large, and as such, research into their biocatalytic synthesis has identified several enzyme classes that can perform the necessary transformations on an industrial scale. These include;

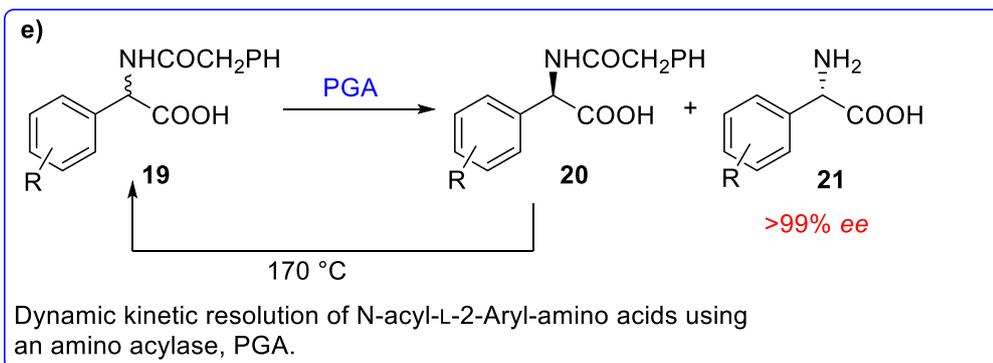
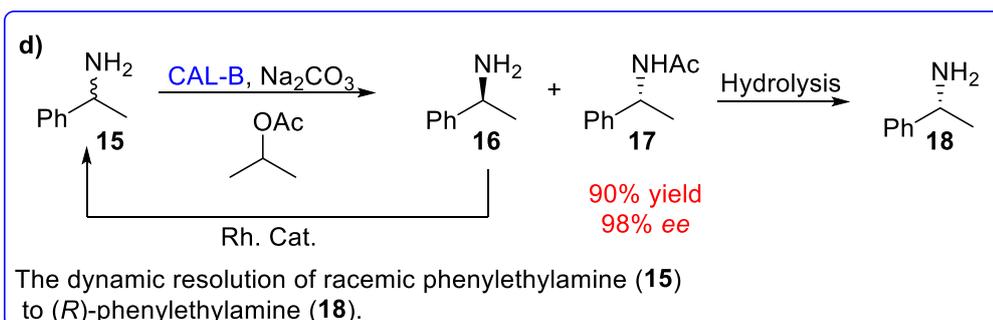
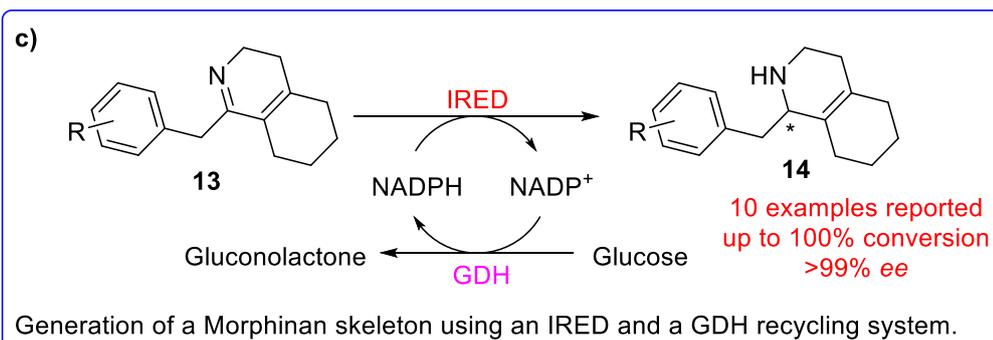
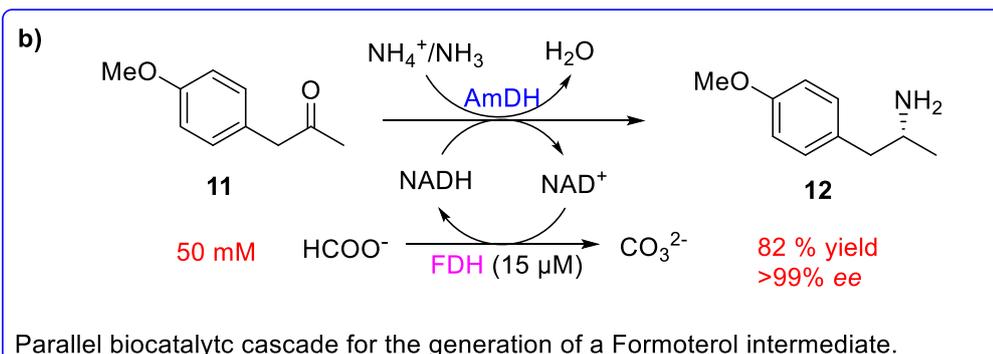
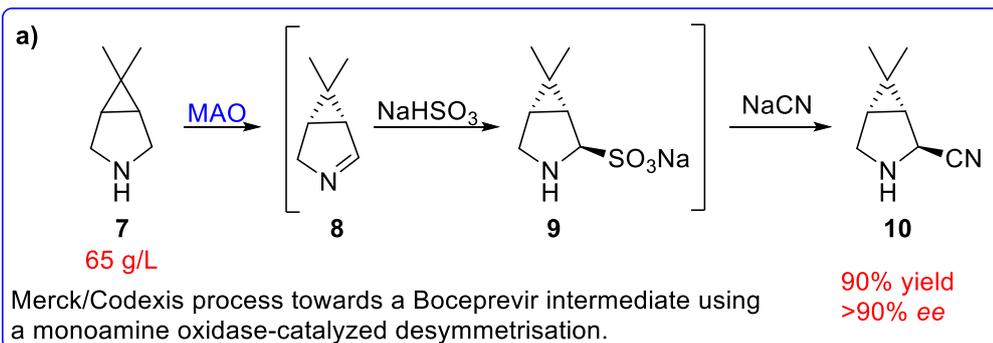
1. Monoamine oxidases' (MAO's), flavin adenine-dinucleotide (FADH) dependent enzymes that can catalyse the reversible oxidation of amines to imines and are therefore utilised in the DKR of racemic amines or in the chemoenzymatic synthesis of chiral amines.<sup>77-81</sup> However, they are largely limited to cyclic amines, presumably due to the increased stability of the imine generated.<sup>82</sup> A collaboration between Merck and Codexis developed a chemoenzymatic process towards an intermediate (**10**) of boceprevir using an engineered MAO. This work resulted in a one-pot net oxidative Strecker reaction involving the enzymatic oxidative desymmetrisation of the prochiral amine substrate (**Scheme 5a**).<sup>83</sup>
2. Amine dehydrogenases (AmDH's), a class of enzymes that can lucratively use ammonia as a nitrogen source to aminate carbonyls using nicotine adenine-dinucleotide (NADH) as a co-factor.<sup>84,85</sup> Reported examples require a 'small' and a 'bulky' substituent on the carbonyl group, which greatly reduces their substrate scope. An exciting example of their application is in the generation of an intermediate (**12**) of Formoterol, a bronchodilator used in the management of asthma and

COPD.<sup>86,87</sup> This involved a parallel biocatalytic cascade to successfully aminate *para*-methoxyphenylacetone (**11**) to **12**, incorporating an FDH for cofactor regeneration, and achieved high yields and excellent enantioselectivity (**Scheme 5b**). Amine **12** has also been reported as a key intermediate in the synthesis of tacrine–selegiline hybrids that are candidates for the treatment of Alzheimer's disease.<sup>88</sup>

3. Imine reductases (IRED's) are nicotinamide adenine dinucleotide phosphate (NADPH) dependent oxidoreductases capable of reducing imines to amines and have a strong preference for cyclic secondary imines.<sup>89,90</sup> Their use is 'limited by their requirement for expensive NADPH cofactor or a cofactor recycling system, however, significant enzyme engineering is being undertaken to modify their preference to the less expensive NADH.<sup>91</sup> The relevance of IREDs to the pharmaceutical industry has been demonstrated by Yao *et al.*, who have succeeded in the generation of diversified morphinan skeletons *via* the enantioselective reduction of bulky  $\alpha,\beta$ -unsaturated imines by an IRED (**Scheme 5c**).<sup>92</sup> This required the use of a GDH recycling system for the regeneration of NADPH. Morphinan is the prototype chemical structure of a large class of psychoactive drugs and this route allows for the facile diversification of this motif to access such therapeutics.<sup>93</sup> More recently, reductive aminases (RedAms) have been identified as a subclass of IREDs that can also catalyse the imine formation from the amine and carbonyl.<sup>94,95</sup>
4. Lipases, a hydrolytic enzyme class, were amongst the first class of enzymes to be used industrially and are capable of hydrolysing both

amide and ester bonds to release an amine, carboxylic acid or alcohol reversibly.<sup>96,97</sup> These enzymes are often used for the kinetic resolution of racemic mixtures of amines (theoretical yield of 50%) or in chemoenzymatic DKRs to produce a single chiral acetylated product, which can be cleaved to the chiral amine with a theoretical yield of 100%.<sup>98</sup> The biotransformations catalysed by lipases can be performed in low/non-aqueous conditions to achieve the acylation, and provides the possibility of using hydrophobic substrates with a high substrate concentrations.<sup>99</sup> However, this approach for chiral amine synthesis relies on the availability of a racemic mixture of the desired amine. BASF have patented an industrial-scale process for the generation of (*R*)-phenylethylamine *via* a DKR using a lipase, CAL-B (Scheme **5d**)<sup>100</sup> and this is currently applied on the multi-ton scale.

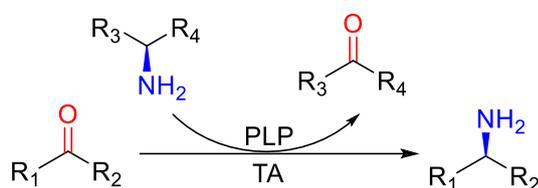
5. Acylases are an alternative class of hydrolytic enzymes. Amino acylases enantioselectively deacetylate *N*-acyl-amines and so are used for the DKR of racemic mixtures to generate chiral amines.<sup>101,102</sup> They have been used for the industrial generation of non-natural L-amino acids since the 1950s.<sup>103</sup> One example includes the chemoenzymatic synthesis of L-2-aryl-amino acids using a penicillin G acylase (PGA) from *Bacillus subtilis* (Scheme **5e**).<sup>104</sup> The *R*-enantiomer of such *N*-acyl-amino acids can be racemised at 170 °C, allowing for high theoretical yield to be obtained in an industrial setting.
6. Transaminases are a well-studied family of enzyme that can be used for the asymmetric production of chiral amines and will be the subject of the rest of this thesis.



**Scheme 5** – The use of different enzyme classes to generate industrially relevant chiral amines: a) Merck/Codexis process towards a Boceprevir intermediate using a monoamine oxidase-catalysed desymmetrisation) Parallel biocatalytic cascade for the generation of a Formoterol intermediate. c) Generation of a Morphinan skeleton using an IRED and a GDH recycling system. d) The dynamic resolution of racemic phenylethylamine (**15**) to (R)-phenylethylamine (**18**.) e) Dynamic kinetic resolution of N-acyl-L-2-Aryl-amino acids using an amino acylase, PGA.

### 1.5.0 Transaminases

Aminotransferases (also commonly referred to as transaminases) (TAs) are pyridoxal-5'-phosphate-dependant (PLP-dependent) enzymes that can convert prochiral carbonyls to their corresponding chiral amine in the presence of a suitable sacrificial amine donor, often with excellent enantioselectivity (**Scheme 6**).



**Scheme 6** - The conversion of a prochiral ketone to a chiral amine by a transaminase using a sacrificial amine donor.

PLP-dependent enzymes have been classified by Percudani *et al.* as belonging to 7 major fold types and are known to perform at least 238 distinct catalytic functions including decarboxylation, racemization, Claisen condensation and transamination.<sup>105–107</sup> TAs belong to fold types I and IV, which are further broken into 6 classes depending upon their structural and sequence similarity.<sup>108</sup> Classes I and II include L-aspartate and L-alanine TAs; class III are ω-TAs (see below); class IV contain branched chain and D-amino acid TAs; class V are L-serine TAs and VI are sugar TAs. It is noteworthy to mention that sugar TAs are only reported to transfer amine groups to exocyclic ketones on sugar derivatives, rather than the endocyclic masked carbonyl group of simple sugars.

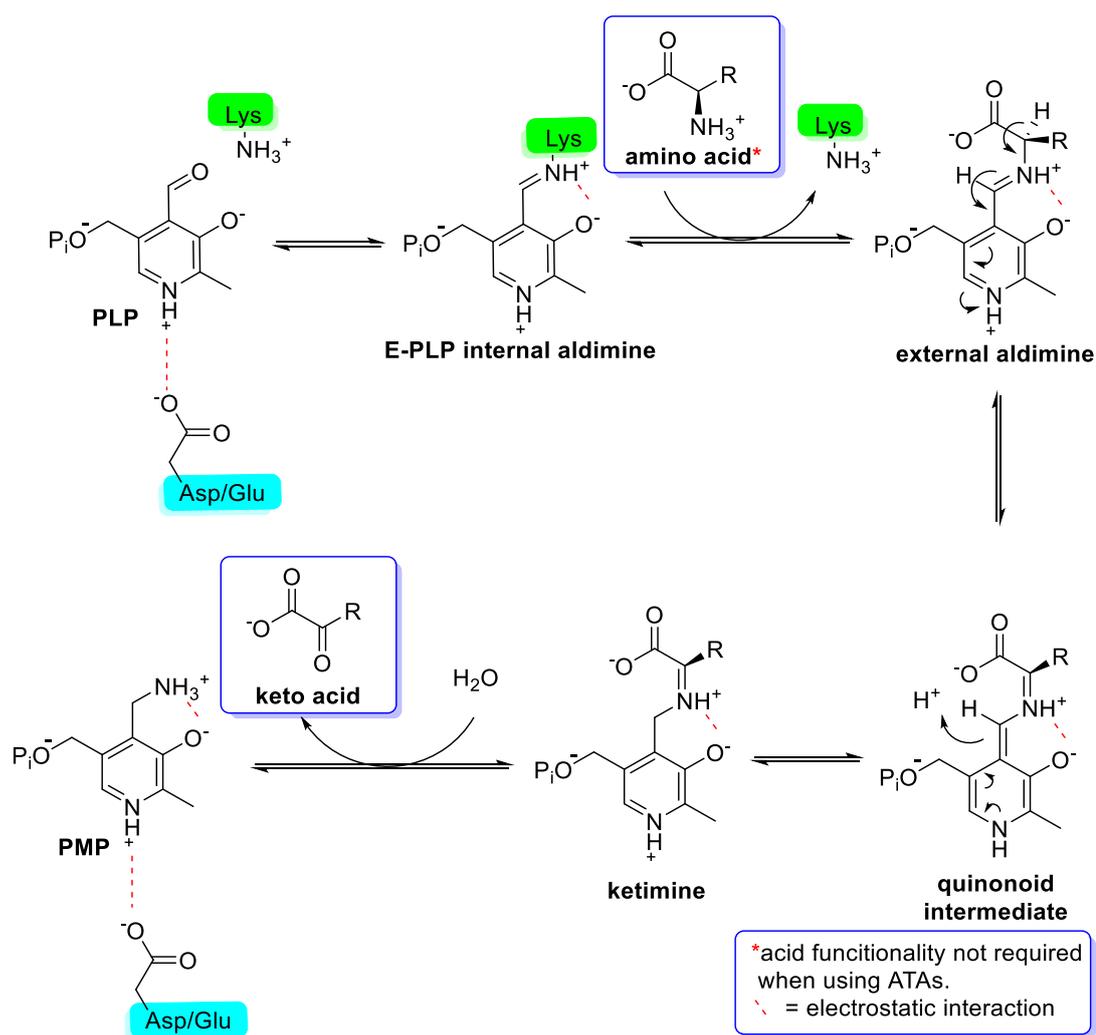
These enzymes can also be more broadly classified as either  $\alpha$ -TAs, which catalyse the transfer of the amino group at the  $\alpha$ -carbon, or  $\omega$ -TAs where the amino group is transferred at a site distal from the carbonyl. Two subgroups of  $\omega$ -TAs studied for synthetic approaches are  $\beta$ -TAs that act on amino groups  $\beta$  to the carboxylate, and amine TAs (ATAs) that can act on a large variety of substrates that lack carboxylic functionality.<sup>109</sup> As we are interested in the asymmetric synthesis of chiral amines without such functionality the remainder of this thesis will focus on ATAs.

### 1.5.1 Transaminase mechanism

The mechanism of ATAs was first described by Oshima *et al.* in 1962 and is an example of a ping-pong bi-bi reaction.<sup>110,111</sup> This means it is a non-sequential mechanism involving two substrates and two products, whereby only one substrate must be bound to the enzyme to release the first product. Firstly, PLP diffuses into the active site where it is stabilised by the negative charge on an aspartic acid or glutamic acid residue. The general mechanism follows the formation of an imine between PLP and a lysine residue in the active site followed by reaction of an amino acid (or an alternative amino acceptor) to form an imine between PLP and the amino acid (**Scheme 7**). This is followed by deprotonation of the  $\alpha$ -C leading to the formation of a quinonoid intermediate, which is resonance stabilised. Spontaneous protonation of the C4' of PLP then generates a ketimine and hydrolysis subsequently occurs to form PMP and release the keto acid (or carbonyl in the case of ATAs), the product of the first half of the mechanism.

At this point PMP is only weakly bound to the enzyme and is readily expelled from the active site to form the unstable apoenzyme leading to insoluble

protein aggregates, therefore, a PLP concentration of 0.1-10 mM is required for operational stability.<sup>112</sup> However, too much PLP will inhibit the enzyme-PMP complex and cause aggregation. To complete the second half of the mechanism, the PMP forms an external aldimine with the amino acceptor substrate and a mirrored sequence of steps takes place to expel PLP and the chiral amine product.

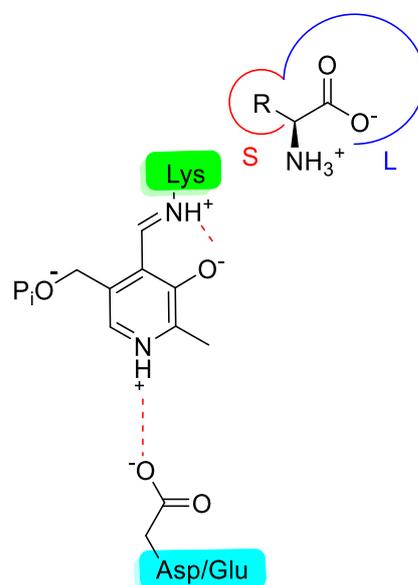


**Scheme 7** - A general mechanism of biocatalysis for the first half of the ping-pong bi-bi reaction of  $\omega$ -ATAs. Starting from the electrostatically bound PLP coenzyme that reacts with the catalytic lysine residue to give the covalently bound internal aldimine, this then reacts with the amine substrate to give the external aldimine, also electrostatically bound in the active site. Followed by deprotonation of the  $\alpha$ -carbon to give the quinonoid

intermediate that is itself protonated to give the ketimine. The addition of H<sub>2</sub>O across the imine results in the keto product and PMP.

### 1.5.2 Mechanistic origin of substrate specificity and stereoselectivity

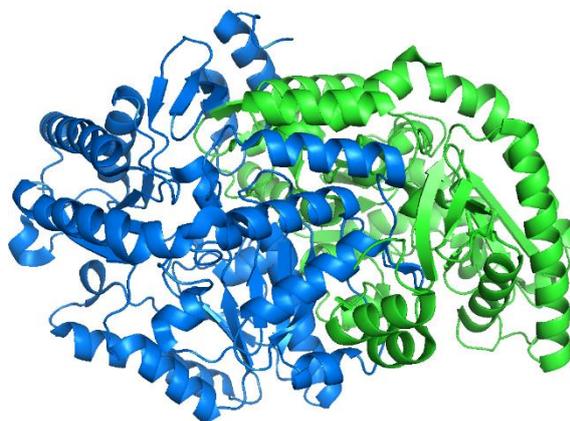
Reported crystal structures and active site models can explain the mechanism of substrate specificity and stereoselectivity displayed by  $\omega$ -TAs.<sup>113,114</sup> Shin *et al.* propose a widely accepted active site model (**Figure 4**).<sup>115</sup> This illustrates the existence of two binding pockets in the active site, a large pocket that binds to the carboxylate group and a small pocket that accommodates the side chain. Due to the strong repulsion experienced by the carboxylate in the small pocket, this binding domain determines the substrate specificity and stereoselectivity. In ATAs, the large pocket can accommodate a wide variety of aliphatic/phenyl side chains, however, the small pocket of wild is often restricted to methyl groups.



*Figure 4 - Model of an ATA active site, illustrating the existence of a small (S) and large (L) pocket for substrate binding. Just note there that L=large and S=small*

It is also worth noting that ATAs exist as homo-(di/tetra/hexa)-mers in aqueous media, with the active sites situated at the interface(s) (**Figure 5**). The interactions between chains determines the quaternary structure and

orientation of the small and large pockets. Hence, changes to the primary structure, specifically to those residues at the interface can alter the substrate specificity and stereoselectivity.<sup>116,117</sup>



*Figure 5 - A cartoon derived from the crystal structure of 3HMU, an  $\omega$ -TA, clearly demonstrating the homodimeric structure. Image generated from the protein databank (PDB: 3HMU).*

### 1.5.3 Thermodynamic challenges

All enzymatic biotransformations, including those incorporating ATAs, are inherently reversible and exist as dynamic equilibria whereby a steady state is reached, and the concentration of starting material/product remains unchanged. Single-step ATA reactions *in vitro* present a particularly unfavourable thermodynamic equilibrium as there is often no driving force for the reaction, meaning they do not meet industrial processes requirements. These challenges are well described, and several creative solutions have been devised to shift the equilibrium towards product formation.<sup>118–120</sup> These strategies improve the conversion and efficiency of such biotransformations by shifting the unfavourable equilibrium following le Chatelier principles. These include; removal of the co-product product using a chemo-enzymatic cascade<sup>121</sup> or other physical means,<sup>122</sup> the use of ‘smart amine donors’<sup>123,124</sup> that

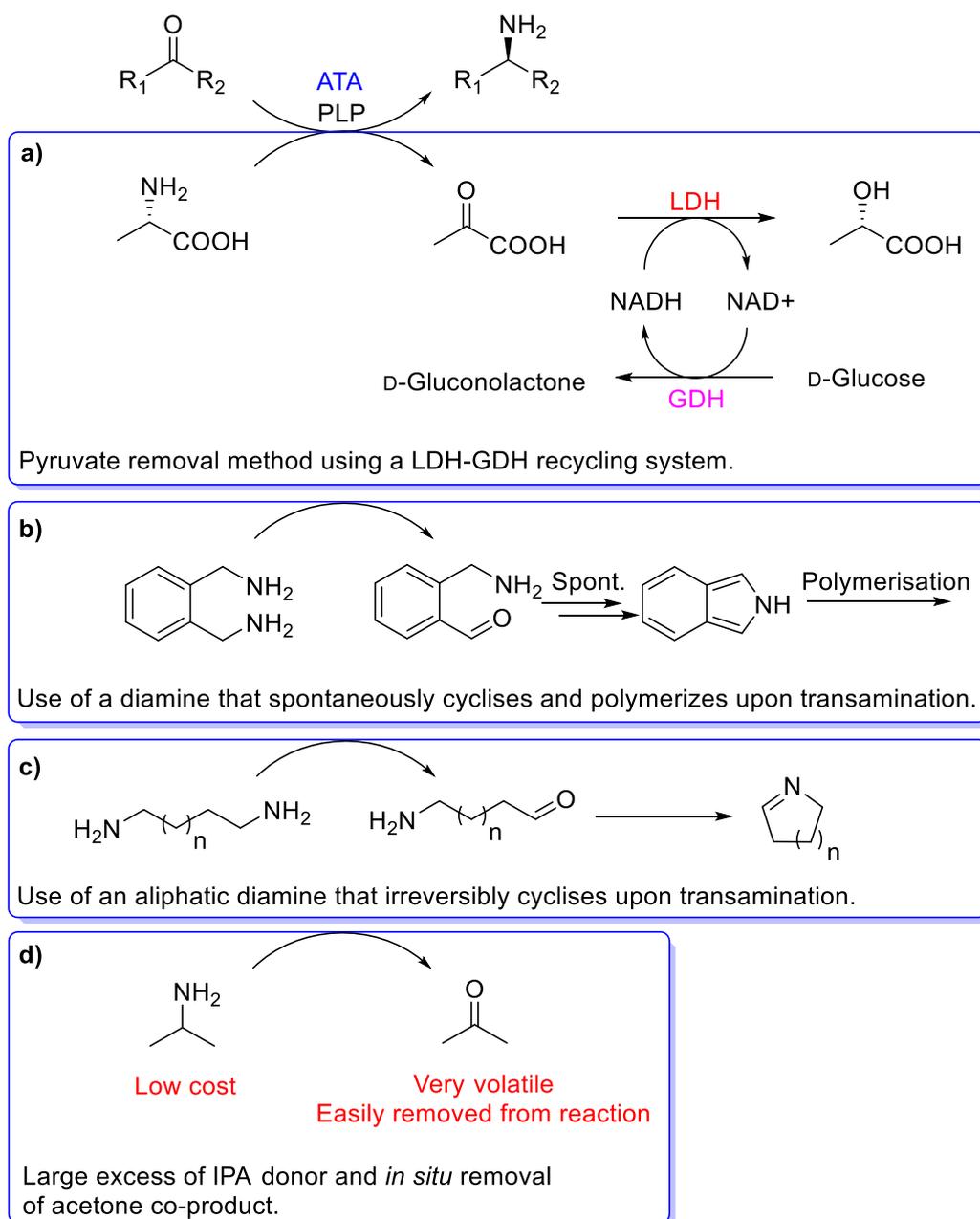
undergo irreversibly intramolecular reactions upon transamination and the use of large excesses of amine donor.

Alanine is the natural substrate for most ATAs and is the amine donor of choice for many transamination reactions. However, the resulting pyruvate coproduct can inhibit the enzyme and is typically removed *in situ*. Shin *et al.* reported a now widely used pyruvate removal system for the whole cell synthesis of (S)-MBA.<sup>121</sup> Here, the pyruvate is removed by converting it to lactate, using a lactate dehydrogenase (LDH), coupled with a glucose dehydrogenase (GDH), for NADH co-factor recycling (**Scheme 8a**).

The use of 'smart amine donors' that irreversibly react upon transamination has been described by several groups in the past decade. Such substrates are spontaneously removed from the reaction and irreversibly drive the equilibrium towards product formation. The O'Reilly group have previously reported the use of *o*-xylylenediamine as an amine donor, upon transamination the coproduct spontaneously cyclises, tautomerizes and subsequently polymerises to form a black precipitate (**Scheme 8b**).<sup>125</sup> This allows for quantitative asymmetric amination of ketone substrates.<sup>126</sup> The use of aliphatic terminal diamines as sacrificial amine donors has also been described, these substrates also cyclise and irreversibly polymerise after transamination, driving the equilibrium (**Scheme 8c**).<sup>123,124</sup> Crucially, it has been demonstrated that they are accepted by a broad range of ATAs that can successfully transaminate a spectrum of prochiral ketones.

Isopropyl amine (IPA) is a very common amine donor in industrial scale biotransformations, largely due to its low-cost and the high volatility of the

acetone coproduct, which can be removed *in situ* (**Scheme 8d**).<sup>122</sup> As IPA is an achiral substrate, there is no issue with acceptance of one enantiomer over the other and this further reduces costs when compared to other racemic amine donors. Its low cost also makes it feasible to use huge excesses (~1 M), which is often required in combination with *in situ* coproduct removal *via* evaporation, to displace the reaction equilibrium towards product formation. However, using IPA on an industrial scale requires substantial levels of enzyme engineering towards substrate acceptance to make the process economically viable.<sup>127,128</sup>



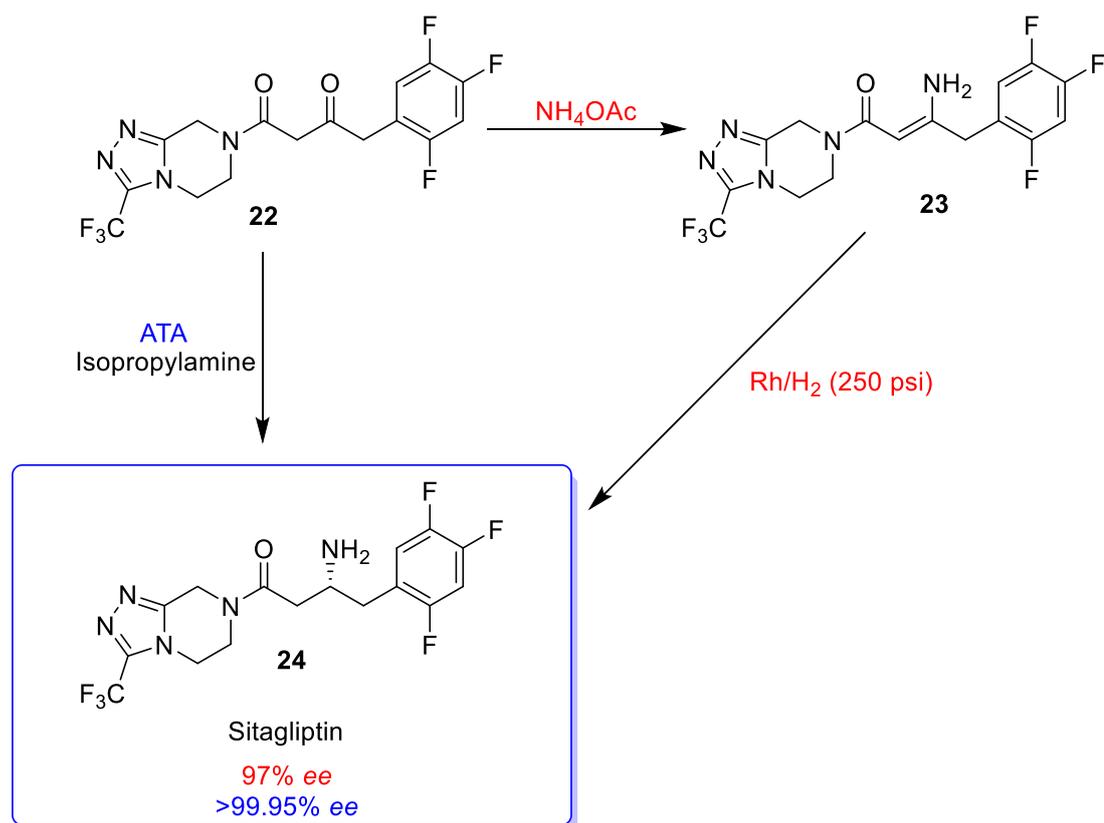
**Scheme 8** – Common methods for displacing the challenging ATA equilibrium in biocatalytic synthesis: a) Pyruvate removal method using a LDH-GDH recycling system. b) Use of a diamine that spontaneously cyclises and polymerizes upon transamination. c) Use of an aliphatic diamine that irreversibly cyclises upon transamination. d) Large excess of IPA donor and *in situ* removal of acetone co-product.

For more challenging equilibria, the use of flow systems incorporating immobilised ATAs have been described.<sup>129,130</sup> This methodology allows products that may be inhibitory to be immediately washed away from the enzyme and can be optimised to allow for repeated use. Furthermore, the use of continuous flow allows for unreacted starting material to be pumped back

into the system, optimising the conversion to product. Contente *et al.* describe a self-sufficient continuous flow system using co-immobilised PLP, which eliminates the need for its exogenous addition to the reaction.<sup>131</sup> This allowed for the synthesis of a variety of biogenic aldehydes with yields greater than 95% in under 15 minutes.

#### 1.5.4 Synthetic applications of ATAs

ATAs are routinely used for the preparative/industrial scale synthesis of chiral amines from prochiral ketones, often as part of biocatalytic or chemoenzymatic cascades to industrially relevant chemicals.<sup>132–135</sup> The first successful report of their development for industrial application involved the engineering of an (*R*)-selective ATA from *Arthrobacter sp.* for the synthesis of the antidiabetic drug sitagliptin (**24**). This tonne-scale process replaced a rhodium-catalysed asymmetric enamine hydrogenation with the biocatalytic amination and afforded a 13% increase in yield, a 19% reduction in total waste and removed the need for high-pressure hydrogenation, whilst achieving >99.95% *ee* (**Scheme 9**).<sup>136</sup> This is often referenced as a landmark moment for the relevance of ATAs in synthetic applications.



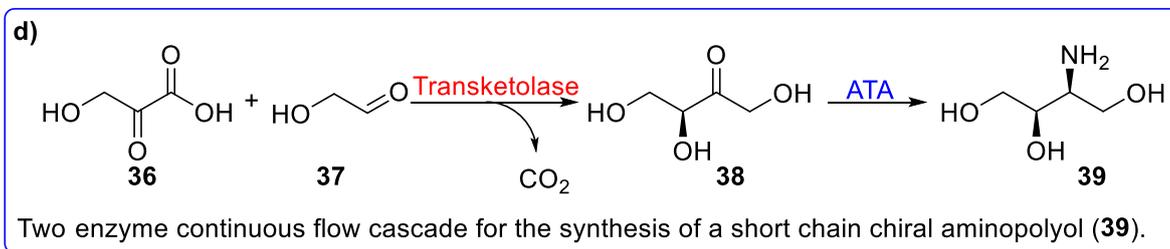
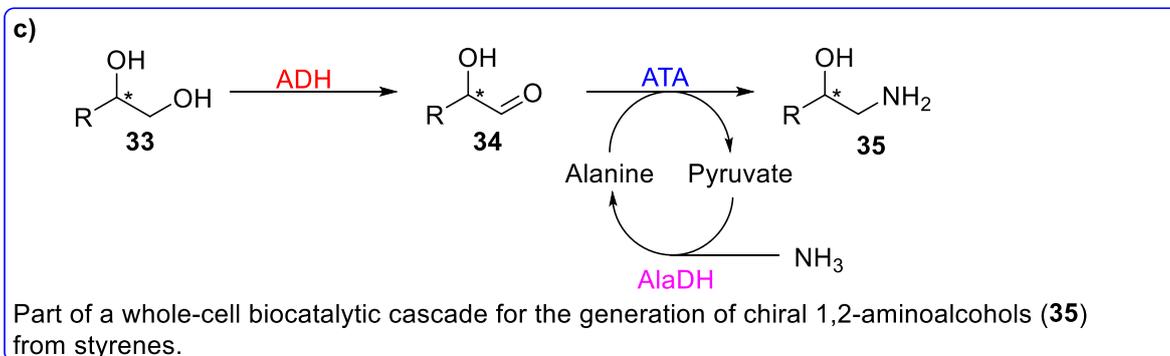
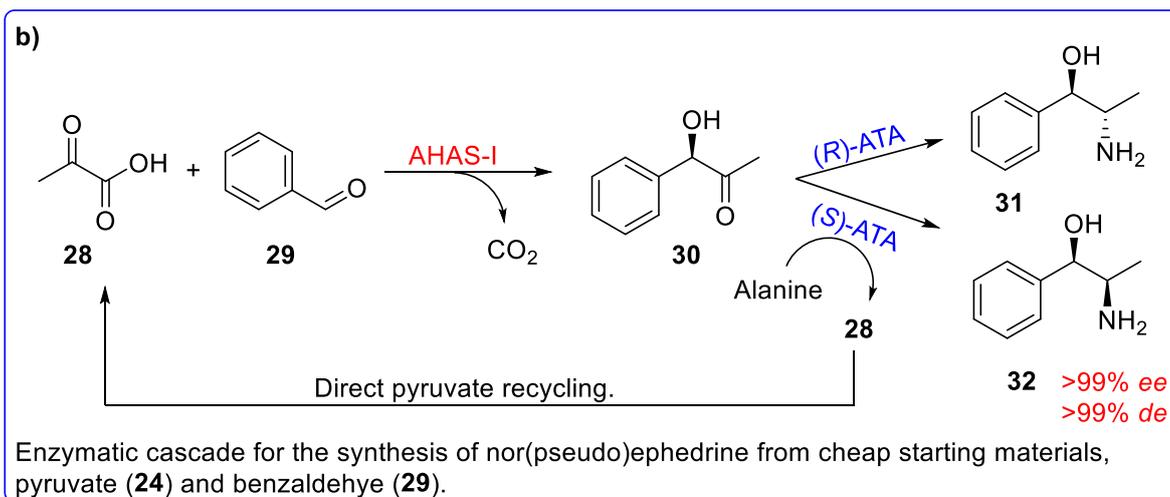
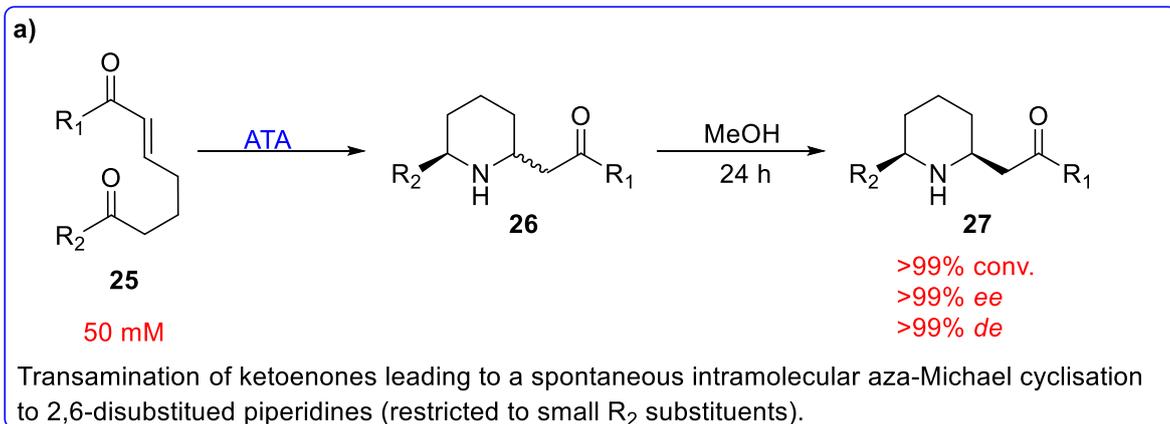
**Scheme 9** – The industrial synthesis of sitagliptin from a key late-stage intermediate. Illustrating the two-step classical synthetic route (in red) using rhodium catalysis to enantioselectively reduce an enamine, compared to the one-step biocatalytic route (in blue) to using an engineered (*R*)-selective ATA.

Ryan *et al.* report the use of commercially available ATAs on ketoenones (**25**) leading to a spontaneous intramolecular aza-Michael cyclisation to 2,6-disubstituted piperidines (**27**), an important scaffold in many biologically active molecules (**Scheme 10a**).<sup>137</sup> The O'Reilly group has recently expanded on this work to transaminate ketoenones, resulting in cyclic  $\beta$ -enaminones, followed by annulation chemistry for a chemoenzymatic synthesis of fused alkaloids.<sup>138</sup> Fully enzymatic cascades using ATAs have also been designed. An elegant example includes the synthesis of nor(pseudo)ephedrine (**31/32**), a psychostimulant, from inexpensive starting materials.<sup>139</sup> This involves the decarboxylation of pyruvate (**28**) by acetohydroxyacid synthase I, followed by its spontaneous ligation to benzaldehyde (**29**). Afterwards either an (*R*)- or (*S*)-selective transaminase can enantioselectively transaminate the ketone to

nor(pseudo)ephedrine (**Scheme 10b**).<sup>139</sup> The use of alanine as an amine donor results in the formation of pyruvate, which is itself the starting material in the cascade, allowing for direct pyruvate recycling that pushes the equilibrium towards product formation.

The complexity of systems involving ATAs is not limited to *in vitro* cascades incorporating a couple of enzymes. Impressive work by Wu *et al.* describes the entirely whole cell biocatalytic synthesis of (S)- $\alpha$ -hydroxyacids, (S)-aminoalcohols (**35**) and (S)- $\alpha$ -aminoacids from styrenes, involving a cascade of 4-8 separate enzymes expressed in *Escherichia coli*.<sup>140</sup> The key transamination step couples an alanine dehydrogenase (AlaDH) and an ATA for the oxidation–transamination of a 1,2-diol (**33**) to a 1,2-amino alcohol (**34**) alongside the recycling of alanine as an amine donor (**Scheme 10c**). The application of ATAs in flow is also well established with Gruber *et al.* reporting a two-step continuous flow system to produce chiral amino alcohols (**Scheme 10d**).<sup>141</sup> This involves the coupling of a transketolase and ATA to convert hydroxypyruvate (**36**) and glycolaldehyde (**37**) to a short chain chiral amino polyol (**39**), an important motif for the generation of complex and optically pure pharmaceuticals. They report full conversion to product in under two hours.

These examples highlight how ATAs have been successfully applied to complex *in vitro*, *in vivo* and flow processes, often with excellent enantioselectivity and yield. Hence, demonstrating their key role in the synthesis of chiral amines and their high versatility to synthetic applications.



**Scheme 10** – The preparative/industrial scale use of ATAs to generate industrially relevant chiral amines: a) Transamination of ketoenones leading to a spontaneous intramolecular aza-Michael cyclisation to 2,6-disubstituted piperidines. b) Enzymatic cascade for the synthesis of nor(pseudo)ephedrine from cheap starting materials, pyruvate (**24**) and benzaldehyde (**29**). c) Part of a whole-cell biocatalytic cascade for the generation of chiral 1,2-aminoalcohols (**35**) from styrenes. d) Two enzyme continuous flow cascade for the synthesis of a short chain chiral aminopolyol (**39**).

### 1.5.5 Exploring novel ATAs

While there are many examples of the synthesis of high-value molecules using enzymes and enzymatic cascades, it is vital to further expand the repertoire of biocatalysts available and increase their substrate scope and selectivity, to allow access to a greater variety of compounds. Several techniques exist to generate novel enzymes with desired activities, these are; activity-based screening, gene mining (sequence-based screening) or protein engineering for tailor made enzymes.<sup>135</sup> Activity based enzyme discovery involves growth of microorganisms on selective media containing the substrate of choice, or screening different organisms for their ability to convert the desired substrate. However, this method is limited by the fact that the majority, up to 99%, of bacteria are unculturable and not all substrates can be used as a carbon/nitrogen source. Gene mining is often preferable due to the rapidly increasing size of the database that can be searched; however, this is a time-consuming process. Combined with the *in-silico* prediction of key active site residues it has become a tantalising technique for the discovery of novel ATAs. This technique was used by Bornscheuer *et al.* to identify 17 novel (*R*)-selective ATAs, which was particularly exciting as previously the vast majority of characterised ATAs had been (*S*)-selective.<sup>142</sup>

### 1.6 Enzyme engineering

Darwinian evolution has given rise to an innumerable selection of diverse and versatile enzymes. However, these biocatalysts have evolved towards a specific natural role and we often want to carry out biotransformations on unnatural substrates under unnatural conditions. Thus, these enzymes are often not readily available to address the current challenges in the industrial

synthesis of valuable molecules. To meet these challenges the enzymes must exhibit; high substrate specificity, high activity/conversion and excellent enantioselectivity. Furthermore, they must remain stable under operational reaction conditions that may include extreme, temperatures, pH, substrate concentrations and the use of organic solvents. Enzymes are routinely optimized by enzyme engineering to meet these requirements for their application in industrial biotechnology. There are two main approaches towards successful enzyme engineering: directed evolution and rational design, these are often combined in a semi-rational design methodology.<sup>143-145</sup>

Directed evolution mimics the process of natural evolution by the careful selection and reproduction of desired phenotypes. This usually involves the initial selection of an enzyme that already has measurable activity towards the desired substrate. Alternatively, previous *de novo* protein design or rational engineering could have introduced the target activity to a protein that requires further optimisation. Regardless of how the enzyme is selected, the encoding gene subsequently undergoes rounds of mutagenesis to generate variant libraries. The protein is then expressed in a suitable host organism and a selection pressure is applied by screening for and selecting desirable mutants. This is routinely achieved by the random recombination of related sequences (e.g. shuffling)<sup>146</sup>, or by the introduction of random mutations in the genetic code (e.g. error-prone PCR)<sup>147</sup>. The key advantage of this approach is that beneficial mutations distal from the active site can be installed without the need for any structural information. However, most mutations cause little to no change in function/activity, therefore, several rounds of mutagenesis are required. Thus, a vast number of variants must be screened to sample a

sufficient percentage of the structural space available. This methodology has been so instrumental in developing biocatalysts for application that the pioneering work in this field was awarded the Nobel Prize in chemistry in 2018.<sup>148</sup>

On the other hand, rational design uses the increasing abundance and reliability of enzyme structures, molecular modelling, and biochemical data to propose potentially advantageous mutations. These are subsequently introduced by site-directed mutagenesis. The main advantage of this method is the increased possibility of introducing desirable mutations, which in turn significantly reduces the number of variants that must be screened. This is particularly important in cases where high-throughput screening protocol is unavailable. Semi-rational protein design combines the principles of directed evolution rational design to gain the advantages of both, this results in smaller 'smart' libraries that have been derived from structural data.<sup>149</sup>

CASTing (combinatorial active site saturation test) is perhaps the most universally used approach to semi-rational protein design.<sup>150-152</sup> This uses information derived from structural/modelling data to identify key residues in interesting regions of the protein, these are commonly mutated by site-directed mutagenesis either individually or as a set. The best variants can be selected, and the process repeated iteratively until an optimal variant is produced, this is known as iterative saturation mutagenesis (ISM). This method is beneficial as unknown synergistic effects that exist between residues/mutations are identified that would otherwise be missed by a purely rational approach.

It is impossible to sample the entire sequence space available to a protein as this would require screening an insurmountable number of variants. Semi-rational design is a method to reduce the library size. However, as there are 20 natural amino acids and 64 possible codons the size of the library required quickly rises with the number of residues selected for mutagenesis. Hence, clever methods are required that further reduce the library to a suitable size whilst also increasing the probability of a successful mutation. A regularly employed technique involves the use of specialised oligonucleotides that reduce codon bias and/or redundancy to give a subset of amino acids with varying properties. For instance, the NDT codon (where N = A/T/G/C, D = A/G/T) codes for just 12 amino acids that differ in size, charge, and polarity.<sup>153</sup> This greatly reduces the library size required to sample a wide variety of chemical environments at any given residue(s).

The techniques discussed thus far for enzyme engineering have been successfully employed to ATAs, greatly increasing their applicability in industrial biotechnology.<sup>118,135,154,155</sup> The previously discussed landmark engineering of an ATA from *Arthrobacter sp.* towards the industrial synthesis of sitagliptin combined computational modelling and ISM to enlarge the small and large binding pockets.<sup>136</sup> The resultant ATA-117 had 27 mutations from the wild type, many of which were distal from the active site. Several of these mutations occur at the dimer interface suggesting they increase stabilisation to allow for a solvent tolerance of 50% DMSO.

The small pocket of ATAs is instrumental in substrate recognition as it often prevents bulky substituents from binding. Often limiting their substrate scope to ketones that contain an  $\alpha$ -functional group no larger than a methyl

substituent. It has been demonstrated that active site residues in this pocket are not significantly involved in the catalytic mechanism.<sup>156</sup> This has allowed for the engineering of binding pocket residues to allow for the accommodation of bulkier substituents. The Bornscheuer group have been particularly effective in this area. They initially reported the rational engineering of a *Vibrio fluvialis* ATA to 30-fold increased activity towards bulky-bulky aryl ketones.<sup>157</sup> They then published the rational design and optimization of an active site motif for the engineering of highly active and stereoselective transaminases for bulky amines. The resultant ATAs exhibited up to 8,900-fold higher activity than the starting scaffold towards the synthesis of industrially relevant bulky chiral amines. They further demonstrate that this motif can be applied to other transaminase scaffolds to the same effect.

The use of directed evolution to generate transaminases with increased substrate activities has also been reported. Martin *et al.* describe the evolution of a transaminase from *Athrobacter citreus* using error-prone PCR.<sup>158</sup> The best mutant after 6 rounds of evolution had a 268-fold increase in activity towards the substrate and a 5-fold increase in product concentration. This also increased the thermostability of the mesophilic wild type to allow extended operation at >50 °C. Telzerow *et al.* also report increased operational thermostability through rational design of a fold IV ATA that naturally converts biaryl ketones.<sup>159</sup> This was achieved by identifying stabilizing interactions between amino acid residues of a well-studied ATA from *Aspergillus terreus*, and subsequently introducing these interactions into their novel enzyme. The best variant had five mutations and exhibited a thermal inactivation half-life  $t_{1/2}$

at 45 °C improvement of 10 min and a thermal inactivation  $T_{1/2}$  increase of 4.4 °C.

It has also been possible to switch the enantioselectivity of ATAs through rational protein design considering substrate binding. The Burgland group used molecular modelling to analyse the structure of an ATA from *Arthrobacter citreus* and identified key residues in the 'phosphate-binding cup' that they proposed controlled stereoselectivity for the synthesis of 1-(4-fluorophenyl)propan-2-amine.<sup>160</sup> The single point mutation V328A reversed the stereoselectivity from 98% ee (*S*) to 58% ee (*R*), this was at a time when (*R*)-selective ATAs remained largely undiscovered. Scalden *et al.* report the investigation into the enantioselectivity of a *Vibrio fluvialis* ATA towards cyclic ketones.<sup>161</sup> They used protein–substrate complex structure analysis to identify L56 as a key residue due to its orientation relative to the substrate. This achieved improved diastereoselectivity with the mutation L56I demonstrating (*S*)-selectivity of 70% *de* from a starting point of 14% *de*. A further mutation, L56V demonstrated a higher (*R*)-selectivity of 66 % *de*. These mutants were subsequently applied in a cascade alongside an enoate reductase to generate optically active cyclic amines.

### **1.7 Screening for ATA activity**

Successful enzyme engineering towards desired activity often relies upon the creation of large variant libraries with high genetic diversity, and such libraries require high throughput screening methods to identify attractive mutants. The previously mentioned molecular biology approaches to engineering ensures high levels of genetic diversity. This has resulted in the identification of desired variants becoming the key limiting factor in such experiments. Therefore, high-

throughput screening protocols are essential for the directed evolution of enzymes and it is essential that these screens allow for facile genotype-phenotype linkage. These screens commonly use pH, absorbance, colour and fluorescence for quick detection. Many of these approaches will use model substrates as they are easier to work with, however, this often results in enzyme optimisation towards the model substrate rather than the desired substrate. This can lead to additional screening efforts to identify suitable mutants. Hence, screening with substrates and conditions as close as possible to that of the expected biotransformation is highly desirable. Several reviews have been published discussing the techniques used to design a successful high throughput screening protocol.<sup>162–164</sup>

In order to rapidly characterise and engineer novel ATAs, there is a need for robust high-throughput enzymatic assays for the facile identification of desirable activity. The reliability and sensitivity of such screens is important, particularly when screening libraries of mutants as a quantitative measurement is often desirable to identify optimal variants. Some of the first approaches towards ATA screening were based on the direct detection of the transamination products by ion exchange chromatography or capillary electrophoresis.<sup>165</sup> However, several methods have now been described that allow for the quick and facile identification of a successful transamination.

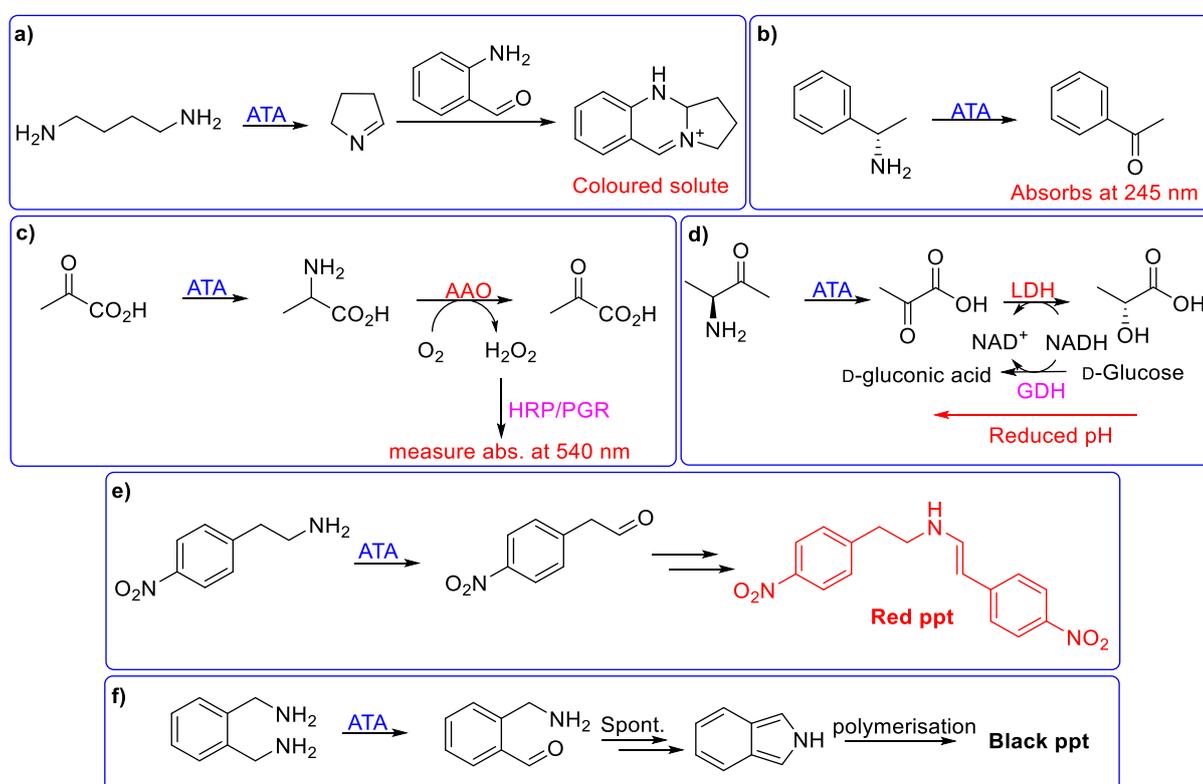
One of the first examples of a colorimetric assay was based on the coloured complex formed between *o*-amino benzaldehyde and either  $\Delta^1$ -piperidine or 1-pyrroline (**Scheme 11a**),<sup>135,166,167</sup> however, this method limited the number of amino substrates that could be screened. Another early high-throughput method was used for the identification of glutamic-pyruvic transaminases,<sup>168</sup>

which catalyse the conversion of alanine to pyruvate. As the Cu(II)-L-alanine complex dissociates the free Cu(II) ions combine with BSA leading to fluorescence quenching (340 nm). Schätzle *et al.* developed a much more facile and rapid kinetic assay using either (*S*)- or (*R*)-MBA and pyruvate as substrates (**Scheme 11b**), which allowed for qualitative and quantitative screening by measuring the production of acetophenone at 300 nm.<sup>169</sup> Due to the high absorbance of acetophenone, this screen is often restricted to ~10 mM by standard laboratory spectrophotometers. The use of UV spectroscopy does not routinely allow for the screening of a broad range of UV active acceptors, furthermore, expensive robotic suites are required for high-throughput application.

The use of several screens using enzymatic cascades has also been described. Hopwood *et al.* report the use of pyruvate as an amine acceptor whereby the resultant amino acid is oxidised by an amino acid oxidase (AAO) generating H<sub>2</sub>O<sub>2</sub> as a by-product (**Scheme 11c**).<sup>170</sup> The addition of horse radish peroxidase and pyrogallol red allows for the colorimetric detection of H<sub>2</sub>O<sub>2</sub> at 540 nm. This screen has also been modified to allow for colorimetric colony-based screening.<sup>171</sup> Using this assay limits the amino acceptor to  $\alpha$ -keto acids (usually pyruvate). Alternatively, physical changes, such as pH, have been measured to identify successful transamination when alanine is used as a substrate and the subsequent addition of a lactate dehydrogenase/glucose dehydrogenase system (**Scheme 11d**).<sup>172</sup> Here, the oxidation of glucose to gluconic acid leads to a measurable reduction in pH.

More recently, a high-throughput screen has been developed using 2-(4-nitrophenyl)ethan-1-amine as the amine donor, which then dimerises to form

a red precipitate, this method can also be used on the solid phase for colony screening (**Scheme 11e**).<sup>173</sup> The O'Reilly group have also reported a now widely used screen for ATAs. Here, *o*-xylylenediamine is used as an amino donor, which subsequently polymerises and forms a black precipitate upon transamination (**Scheme 11f**).<sup>125</sup> This is a highly sensitive screen, with conversions <0.5 mM giving a strong positive result and can be used in both the solid and liquid phase.



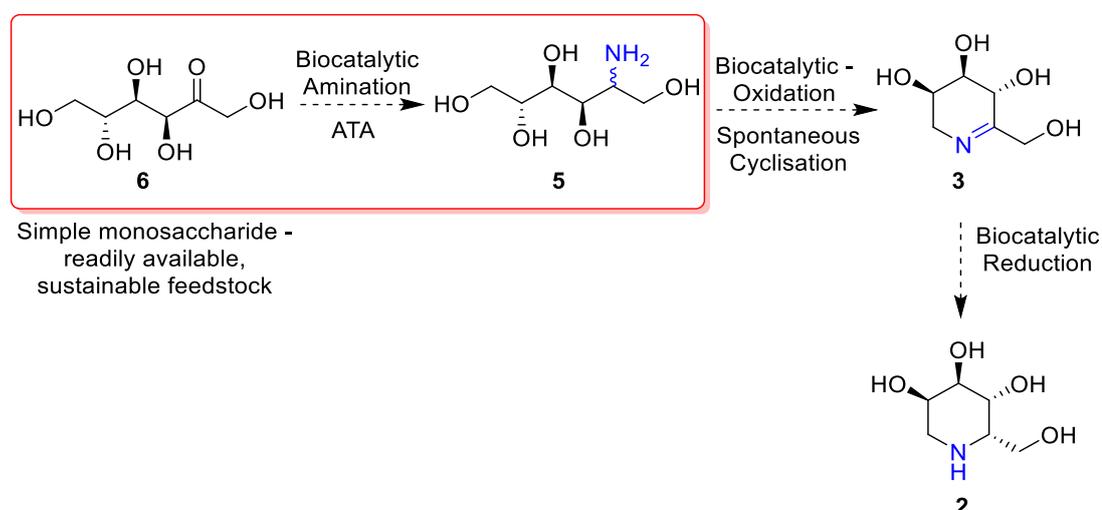
**Scheme 11** - Several colorimetric, UV and pH screening techniques for ATA activity. a) Colorimetric assay based on the coloured complex formed between *o*-amino benzaldehyde and either  $\Delta^1$ -piperidine or 1-pyrroline. b) A rapid kinetic assay using either (*S*)- or (*R*)-MBA to produce acetophenone, which absorbs at 245 nm. c) Colorimetric detection of  $\text{H}_2\text{O}_2$  by-product when using pyruvate as an amino acceptor followed by the enzymatic oxidation of alanine. d) An enzymatic cascade reaction with a lactate dehydrogenase/glucose dehydrogenase system to measure pH change. e) 2-(4-nitrophenyl)ethan-1-amine as a donor that upon transamination dimerises to form a red precipitate. f) *o*-Xylylenediamine as an amino donor, which subsequently polymerises and forms a black precipitate upon transamination.

As these methods often rely on non-natural substrates, they are limited by the enzyme's ability (or lack thereof) to accept the screening amine donor/acceptor. Most TA screens reported thus far only allow for the screening

of amine acceptors and there is a strong need for assays that identify novel activities towards amine donors. Furthermore, these screens are largely purely qualitative and their use for quantitative studies has not been described. Quantitative screens are highly desirable in engineering endeavours as the best variants are ideally identified to be taken forwards. Therefore, there is a clear need for the development of novel quantitative amine donor/acceptor assays that can be applied to the high-throughput and facile screening of ATAs.

### 1.8 Project aims and objectives

As discussed previously, the O'Reilly group is interested in the biocatalytic synthesis of iminosugars from simple monosaccharides. We propose taking a sugar, such as D-fructose (**6**), and using an ATA to generate an aminopolyol **5** (**Scheme 12**). This would then be followed by a regioselective biocatalytic oxidation/cyclisation to the cyclic imine **3** that could then be selectively reduced to the desired iminosugar **2**. The work presented in this thesis will focus on the first step of this process, the transamination.



**Scheme 12** – The proposed biocatalytic route to iminosugars from simple monosaccharides. Highlighting the first step, a transamination, to an aminopolyol that will be the focus of the work presented in this thesis.

At the beginning of this project there was no reported biocatalytic amination of the carbonyl on cyclic sugars to an aminopolyol. Hence, the 2<sup>nd</sup> chapter of this thesis exhibits the identification of ATAs that can perform this transformation on a range of monosaccharides and the work done to apply this on-scale. Furthermore, this enzyme class lacked a comprehensive quantitative screen that we believed to be necessary for the successful engineering of these enzymes to improved activity towards sugars. The focus of the 3<sup>rd</sup> chapter is on the development of such a screen that can be applied to liquid and solid-state screening. Chapter 4 marries the knowledge gained from these chapters to illustrate the molecular modelling of monosaccharides in  $\omega$ -ATAs and the semi-rational engineering of 3HMU towards ketose activity. The final chapter describes a hybrid bio-organocatalytic cascade for the synthesis of complex alkaloids, relying on a transaminase to generate a key reactive intermediate for the complexity building Mannich reaction. There is a continued interest in developing cascade processes for the synthesis of these key chiral building blocks and bioactive natural product analogues.

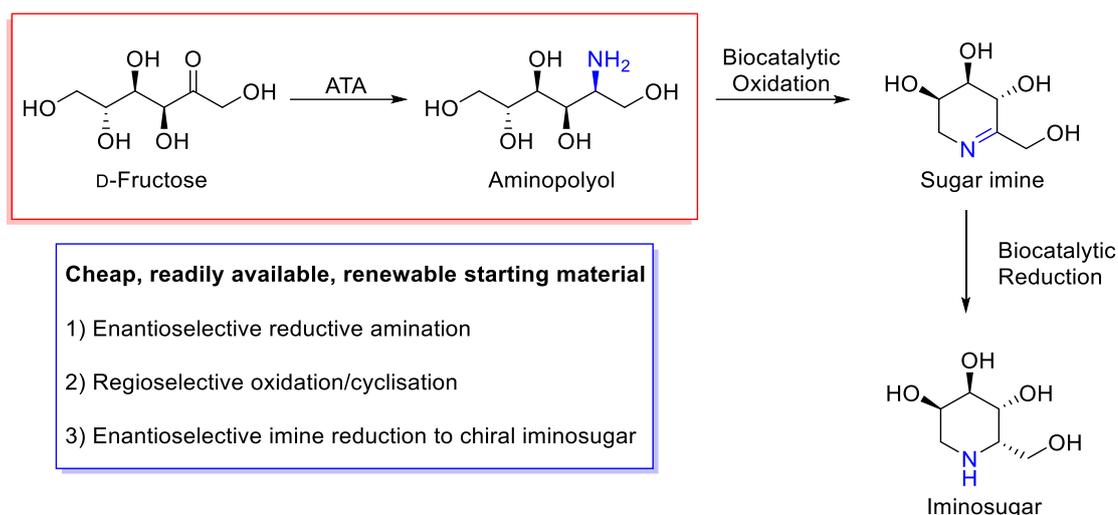
## 2.0 Conversion of aldoses to valuable $\omega$ -aminopolyols using transaminases

The work presented in this chapter has been published: R. Cairns, A. Gomm, J. Ryan, T. Clarke, E. Kulcinskaja, K. Butler and Elaine O'Reilly; *ACS Catal.*; 2019, **9 (2)**, 1220-1223.

Some of the work for this project was carried out in collaboration with other members of the research group, Dr Andrew Gomm and Dr James Ryan. Of the data shown in this chapter, the analytical scale ATA-256 reactions were performed by Dr Andrew Gomm and Dr James Ryan.

### 2.1 Introduction

There is a need to reduce our dependence on fossil fuel-derived chemicals and find alternative, sustainable, and low-cost starting materials to produce pharmaceuticals and agrochemicals that we depend upon in our modern economy. The need for the sustainable synthesis of iminosugars applicable on an industrial scale has already been established (see chapter 1.2). Our group propose an alternative biocatalytic route for the preparation of chiral iminosugars, by exploiting a transaminase to aminate simple keto-sugars, followed by enzymatic oxidation/reduction (**Scheme 13**). This route has the potential to enable the production of large iminosugar libraries for biological screening and thus significantly increase the probability of finding new drug leads. Furthermore, the monosaccharide substrates used would be obtained from readily accessible and sustainable carbohydrate feedstocks. This route requires ketose sugars to produce chiral iminosugars, although non-chiral iminosugars could be produced using aldoses.

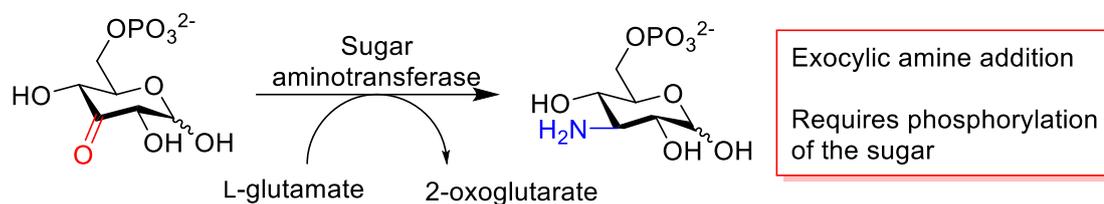


**Scheme 13** - Proposed biocatalytic route to iminosugars, starting from simple ketose sugars. The transamination of the sugar to the corresponding amino alcohol is followed by biocatalytic oxidation and spontaneous cyclisation to the sugar imine and completed by biocatalytic reduction.

The first step of the synthesis generates valuable chiral amino alcohols/aminopolyols *via* the direct amination of monosaccharides by an ATA. These compounds represent valuable synthons for the synthesis of a diverse array of chiral molecules.<sup>174–177</sup> Such amino alcohols are an industrially relevant class of compounds that are commonly used as intermediates (i.e., chiral auxiliaries) in natural product and drug synthesis whilst also exhibiting interesting pharmacological properties.<sup>178,179</sup> This process would allow for the generation of a wide range of high value aminopolyols using a selection of aldose/ketose sugars.

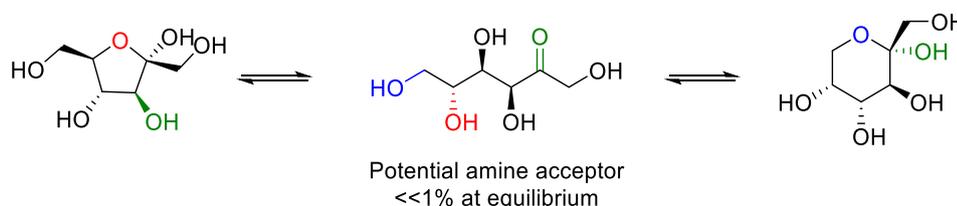
However, thus far the only reported conversion of simple sugars to amino alcohols *via* transamination have been with the tetroses, D-erythrose, and L-erythrulose, which do not readily form cyclic structures, and prior to this work, there was no report of this transformation on monosaccharides that cyclize to form hemiacetals.<sup>180</sup> Sugar aminotransferases have also been identified that play an important biosynthetic role in the amination of oxidised sugar phosphates and sugar nucleotides (**Scheme 14**). while activity on simple

monosaccharides has not been reported.<sup>181–183</sup> Additionally, these enzymes generate the corresponding aminosugar rather than the amino alcohol.



*Scheme 14 – The direct amination of a sugar phosphate to generate an aminosugar using a sugar aminotransferase.*

A traditional chemical synthesis of aminopolyols from simple ketose sugars would involve numerous protection group manipulation steps to mediate a selective reductive amination. We propose that this challenging synthesis could be overcome by using simple sugars as amino acceptors and using an ATA to install the amine functionality. At first glance, cyclic sugars do not resemble typical ATA substrates, however, they are structurally dynamic and exist in several conformations (**Figure 6**), which are all in equilibrium. The linear, open chain conformation of sugars contains a carbonyl group and is structurally similar to classical ATA substrates, albeit more polar and spatially bulky. The open chain conformation is typically present in much less than 1% at equilibrium, so we envisaged that the use of ‘smart amine donors’ will be particularly important to force the already challenging ATA equilibrium.



*Figure 6 - A simple diagram illustrating the structural conformations of fructose.*

Preliminary work undertaken in the O'Reilly group involved the qualitative screening of a small panel of commercial and wild-type ATAs that accepted several aldose sugars. However, the development of an analytical method to quantitatively measure conversion and further investigate substrate scope was required. The three enzymes selected for future study were 3HMU from *Rugeria pomeroy*,<sup>184</sup> HEWT from *Halomonas elongata*<sup>180</sup> and the commercial ATA-256.

## **2.2 Aims and Objectives**

This work aimed to establish the monosaccharide substrate scope of a selection of ATAs. *o*-Xylylenediamine was initially used as an amine donor to screen for sugar substrates under the expectation that spontaneous polymerisation/precipitation of the product would effectively displace the reaction equilibrium towards product formation. This required the development of a quantitative NMR protocol for analysis of monosaccharide conversion, with a variety of amine donors. This allowed for pH, temperature, and amine donor optimisation. This was followed by the preparative-scale biocatalytic synthesis and isolation of  $\omega$ -aminopolyol products.

### **2.3.0 Results and discussion**

#### **2.3.1 Screening for activity**

Following the success of preliminary studies by the O'Reilly group, where 3HMU, HEWT and the commercial ATA-256 were identified as having activity towards a panel of aldoses, a larger panel of ATAs (wild-type and variants) were screened to identify enzymes that display activity towards ketoses, as well as aldoses. To identify ATAs that exhibit activity towards sugars, a robotics system was utilised to allow for high throughput screening of a large library of

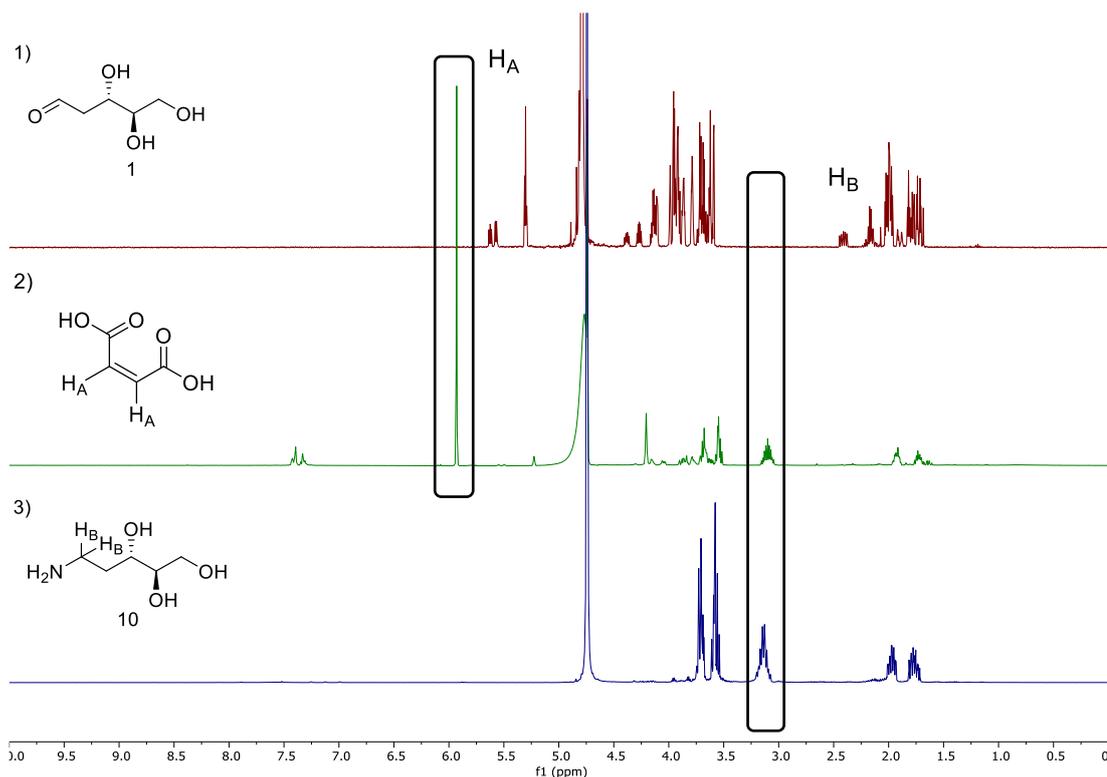
wild type and mutant ATAs. The acceptor panel screened comprised 13 sugars that were commercially available and of relatively low cost and *o*-xylylenediamine was chosen as an amine donor, as it allows for a highly sensitive screen whereby positive results are easily distinguishable from negatives due to the dark black colour of the transamination product (See chapter 1.7). Additionally, as the cyclic structure of the substrates are thermodynamically more stable than the linear form, it was assumed that the position of equilibrium would hinder transamination. For this reason, smart amine donors were employed to help drive the equilibrium towards product formation. A total of ~100 ATAs were screened in this process.

Overall, the vast majority of ATAs screened appeared to have no activity towards monosaccharides and no ATAs were identified that were active towards ketoses. This is not surprising, as monosaccharides are structurally quite distinct from the natural substrates and ketoses would introduce significant steric bulk into the small pocket of the active site, relative to aldoses. The ATAs that were identified were mutants of 3HMU, an enzyme that had already been identified by the O'Reilly group as having activity towards aldoses. Furthermore, this screen was not viable for the ~20 (*R*)-selective enzymes that were screened, as they did not readily accept *o*-xylylenediamine. Therefore, it was decided to focus the remaining studies on the enzymes already identified within the group.

### **2.3.2 Analytical Biotransformations**

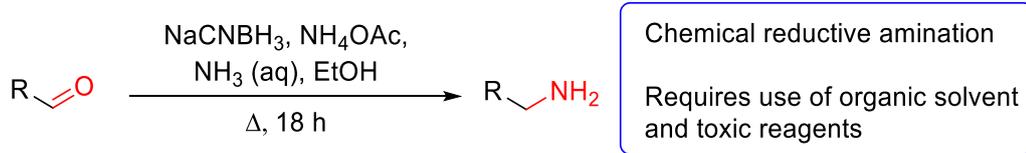
To further investigate the substrate scope of these enzymes towards sugars and allow for the optimal conditions to be established, a reliable quantitative method to calculate conversion was required.

In industry, it is common to measure the purity of compounds using quantitative NMR and in recent years biochemists have started to adapt this method to measure the conversion of a variety of biotransformations.<sup>185</sup> The principal being that the area under the peaks in a proton NMR is directly proportional to the number of protons each peak represents, relative to all other protons in the sample. Therefore, if the sample is doped with an internal standard of known concentration, the quantity of product in the sample can be calculated by integrating the area under a characteristic peak in the product and comparing it to the area of a characteristic peak in the standard. For a reliable method, it is important an NMR experiment is designed in which all the protons in the sample are relaxed between pulses and an internal standard is chosen with peaks present in a region of the spectrum distinct from the rest of the sample. As these experiments were performed in H<sub>2</sub>O, an excellent solvent suppression method was also required. NMR experiments were conducted on a 500 MHz NMR machine with a cryoprobe operating at 20 K to maximise the signal to noise ratio meaning even small conversions of <1% should be detectable. In order to accurately establish the percentage conversion, maleic acid was added to the NMR samples as an internal standard and the formation of the product protons, HB, were monitored (**Figure 7**).



**Figure 7** - Representative example of an NMR spectra displaying the identifiable peak of the amino alcohol product (**49**). 1) 2-Deoxy-D-ribose (**40**); 2) biotransformation of 2-deoxy-D-ribose (**40**) with *o*-xylylenediamine and maleic acid internal standard; 3) (2*R*,3*S*)-5-aminopentane-1,2,3-triol product (**49**).

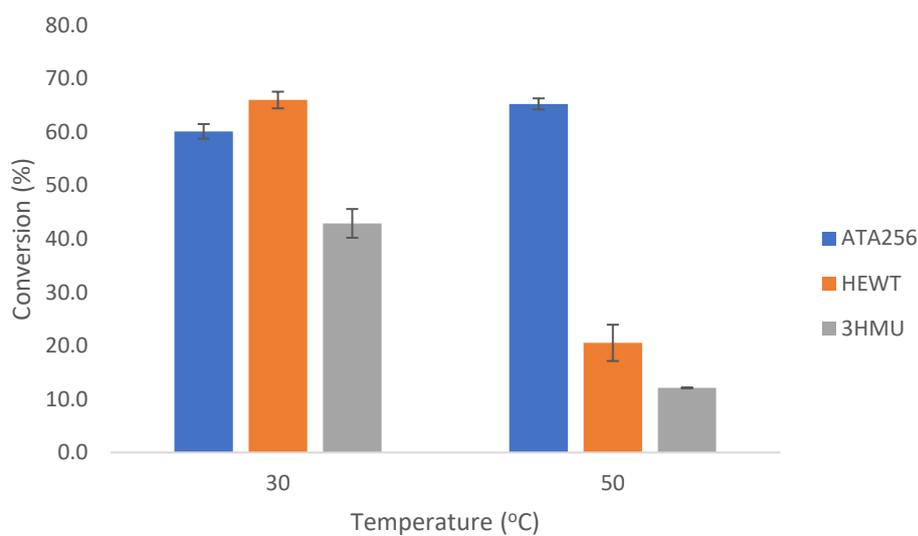
This required the chemical synthesis of the amino alcohol standards (**Scheme 15**), which could be achieved in a single step under mild conditions, as only aldose sugars were under investigation (see experimental for comprehensive data on each standard).<sup>186</sup>



**Scheme 15** - General scheme for the chemical reductive amination of aldose sugars to the amino alcohol standard.<sup>186</sup>

To establish the optimum conditions for performing these reactions, analytical-scale biotransformations were first carried out using the enzymes identified above; 3HMU, HEWT and ATA-256 alongside a single sugar substrate, 2-deoxy-D-ribose (**40**). A broad range of pH values, temperatures, and amine

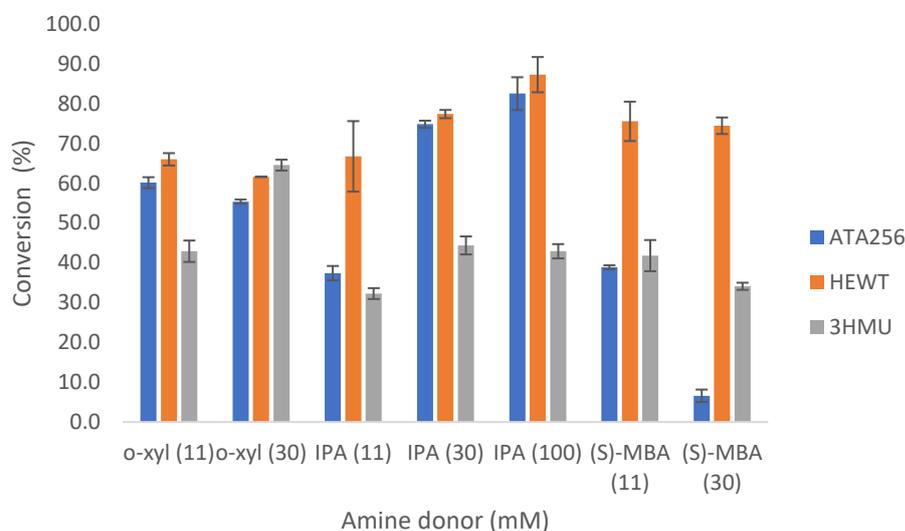
donors/amine donor concentrations were evaluated. The ATAs were screened at 30 °C and 50 °C with commercial ATA-256, which has been engineered to operate at sustained elevated temperatures, performing better as expected at 50 °C (**Figure 8**). The WT enzymes showed a significant drop in conversion at 50 °C, as they have not evolved to function at higher temperatures.



**Figure 8** - Conversion of 2-deoxy-D-ribose (10 mM) to amino alcohol with and o-xylylenediamine (11 mM) in KPi (100 mM, pH 7, 7.5 or 9, 0.1 mM PLP) after 24h at 200 rpm. Exact conditions dependent on enzyme used, see experimental 2.5.

To assess the effectiveness of different amine donors at shifting the equilibrium towards product formation, varying equivalents of o-xylylenediamine, (S)-methylbenzylamine ((S)-MBA) and isopropylamine (IPA) were used (**Figure 9**). Surprisingly, increasing from 1.1 to 3 equivalents of donor had negligible effect on the conversion, with 3 equivalents of (S)-methylbenzylamine causing a significant drop in conversion when used with ATA-256. Using 10 equivalents of IPA resulted in the highest conversion, with up to 87% recorded. This data suggests that although <1% of the linear sugar form exists at dynamic equilibrium, the proposed difficulties this would cause for product formation are insignificant as even the use of donors that have little

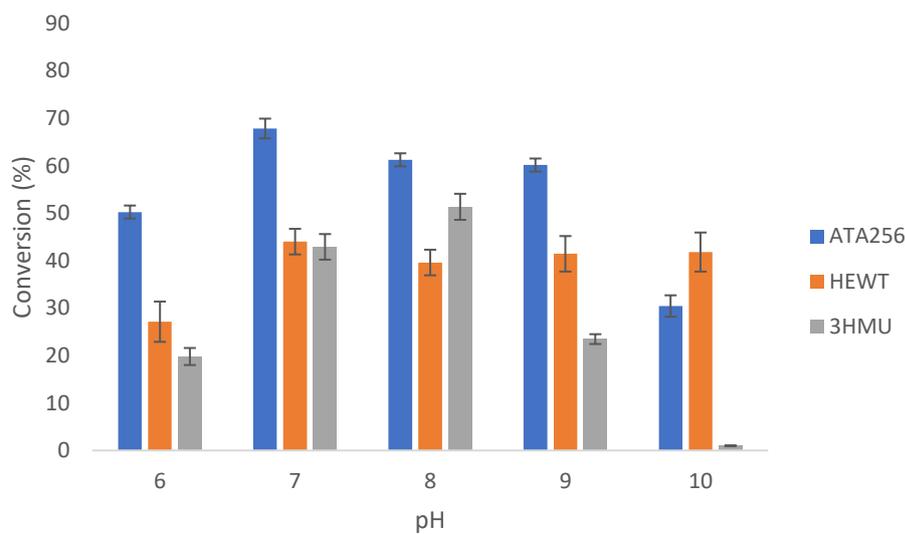
effect on the equilibrium lead to moderate-high conversion. Indeed, IPA and (S)-MBA outperform *o*-xylylenediamine at 1.1 equivalents with HEWT.



**Figure 9** - Conversion of 2-deoxy-D-ribose (10 mM) to amino alcohol with amine donor (11 or 30 mM or 100 mM for IPA) in KPi buffer (100 mM, pH 7.0, 7.5 or 9.0, 0.1 mM PLP) after 24h at 30 °C and 200 rpm. Exact conditions dependent on enzyme used, see experimental 2.5.

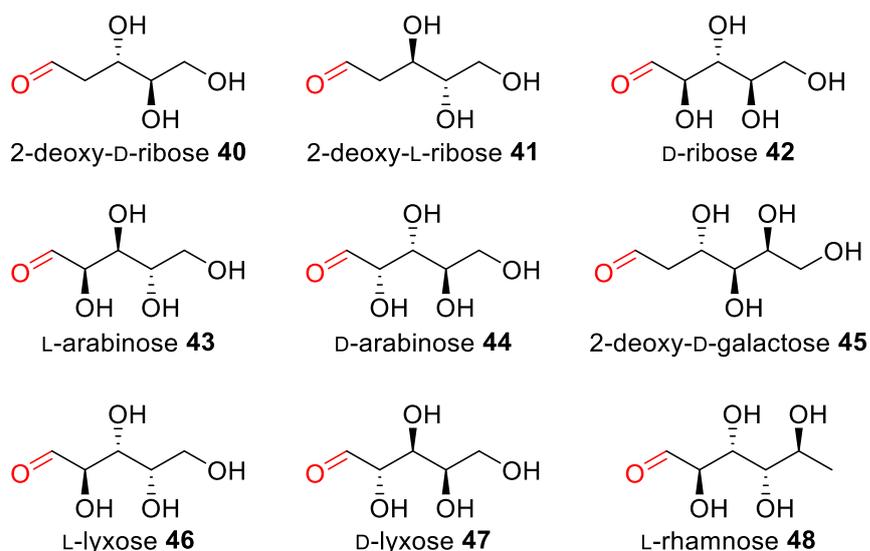
ATA-256 and HEWT were broadly consistent across the full range of pH's screened (**Figure 10**). ATA-256 has been engineered to tolerate a broad range of pH's and its use to generate quantitative yields has been described across a range of pH's, with previous work by our group citing the availability of the amine donor at a given pH the most likely factor determining the optimal pH.<sup>187,188</sup> A full characterisation of HEWT by Cerioli *et al.* revealed that maximum activity was observed pH 10, but the enzyme exhibited increased stability at lower pH.<sup>180</sup> The optimum balance between activity and stability was recorded at pH 8–9, as further evidenced by this work. Characterisation of 3HMU by Steffen-Munsberg *et al.* revealed a broad pH range, however, they identified the optimal pH to be ~9.5.<sup>184</sup> This number was achieved using a photometric activity assay over 10 minutes. Therefore, it is possible the optimal pH for stability differs from this figure as the protein has evolved from *Rugeria*

*pomeroy*, an organism that favours near neutral pH. Hence, the true optimal pH is likely to be lower than that recorded and the use of 3HMU at lower pH has been reported by our group.<sup>189</sup> When using D-deoxyribose, 3HMU showed significant drops in conversion away from the optimum pH range (7-8), with insignificant conversion recorded at pH 10.



**Figure 10** - Conversion of 2-deoxy-D-ribose (10 mM) to aminoalcohol with *o*-xylylenediamine (11 mM) in KPi buffer (100 mM, 0.1 mM PLP) after 24h at 30 °C and 200 rpm. Exact conditions dependent on enzyme used, see experimental 2.5.

Following the data obtained from screening 2-deoxy-D-ribose under a wide variety of conditions, a panel of nine sugars was selected for screening with all three ATAs under the optimal conditions identified (**Figure 11**).



*Figure 11 - Sugar substrates 40–48 used for analytical scale biotransformations.*

All three enzymes performed better with the deoxy sugars **40**, **41**, and **45**, and generally displayed lower activity toward the non-deoxy sugars (**Table 1**). Despite the low concentration of the linear open-chain sugar substrate available at equilibrium, conversions of up to 69% were achieved using just 1.1 equivalent of *o*-xylylenediamine. There appears to be no obvious correlation between the concentration of active carbonyl species in solution and the overall conversion, based on values for the mutarotation rates reported in the literature.<sup>190,191</sup> However, measuring the concentrations of the carbonyl species accurately has historically proved particularly challenging, as the concentration is so low in solution, therefore, this was not explored in detail. Furthermore, no stereochemistry preference pattern emerged from this screen, although this is a relatively small panel of sugars, and a larger panel would be needed to draw any meaningful conclusions on stereoselectivity.

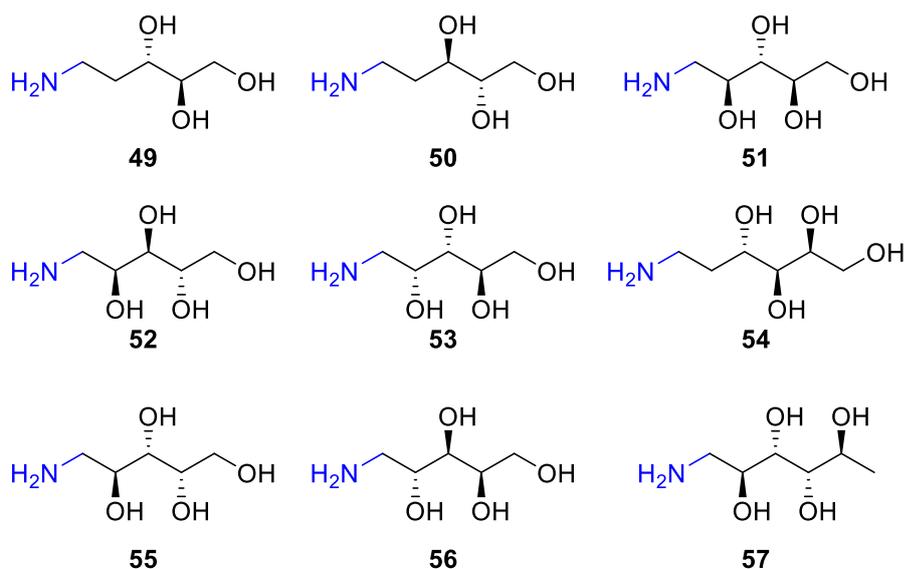
Substrate	Conv. ATA256 (%)	Conv. 3HMU (%)	Conv. HEWT (%)
-----------	------------------------	----------------------	----------------------

2-deoxy-D-ribose <b>40</b>	68	51	66
2-deoxy-L-ribose <b>41</b>	39	35	69
D-ribose <b>42</b>	22	17	41
L-arabinose <b>43</b>	3	6	8
D-arabinose <b>44</b>	7	3	3
2-deoxy-D-galactose <b>45</b>	14	15	35
L-lyxose <b>46</b>	11	13	6
D-lyxose <b>47</b>	7	2	5
L-rhamnose <b>48</b>	10	2	25

**Table 1** - The conversion of sugars **40-48** to the corresponding amino alcohols (10 mM) using ATA256, HEWT and 3HMU with *o*-xylylenediamine (11 mM) in Kpi buffer (100 mM, 0.1 mM PLP). Conversions recorded after 24 h at 30 °C and 200 rpm. See experimental section 2.5 for specific conditions.

### 2.3.3 Preparative scale biotransformations

ATA-256 was selected for preparative scale biocatalytic synthesis of the 9 amino alcohols **49-57** (Figure 12) as it has been engineered to operate at scale with high substrate concentrations, specifically whilst using the cheap IPA as an amine donor. 100 mM or 200 mM of the starting sugar was added as a substrate with 4 equivalents of IPA.



**Figure 12** – Amino alcohols **49-57** synthesised biocatalytically on the preparative-scale using ATA-256 from sugars **40-48** (100 mM or 200 mM), utilising 4 equiv. IPA.

Conversions of up to 94% to the corresponding amino alcohols were achieved for several the substrates after 48 h at 50 °C, with the conversions at this scale following the same trend as seen on the analytical scale, with the exception of the deoxysugar **45**, which achieved a relatively low conversion at this scale. Products were isolated in moderate to good yields (**Table 2**) enabling up to 475 mg of amino alcohol **40** to be isolated. We believe these results demonstrate the applicability of using ATAs for the scalable biosynthesis of valuable  $\omega$ -aminopolyols.

Substrate	Amino alcohol	Conc. mM	Conv. 24 h %	Conv. 48 h %	Yield (%)	Isolated mg
2-deoxy-D-ribose <b>40</b>	<b>49</b>	200	61	94	69	475
2-deoxy-L-ribose <b>41</b>	<b>50</b>	200	69	94	69	474
D-ribose <b>42</b>	<b>51</b>	100	29	62	47	176
L-arabinose <b>43</b>	<b>52</b>	100	19	35	28	104
D-arabinose <b>44</b>	<b>53</b>	100	39	75	38	144
2-deoxy-D-galactose <b>45</b>	<b>54</b>	100	30	30	22	88
L-lyxose <b>46</b>	<b>55</b>	200	15	24	16*	121
D-lyxose <b>47</b>	<b>56</b>	200	12	21	11*	86
L-rhamnose <b>48</b>	<b>57</b>	100	12	48	33	131

*Table 2 - Preparative-scale synthesis of amino alcohols from sugars **40-48** (100 mM or 200 mM) using ATA256 and 4 equiv. of IPA. Conversions were calculated by NMR after 48 hours at 50 °C and isolated yields were recorded after purification on Dowex® resin. \*After purification, compounds were still contaminated with IPA.HCl salt.*

## 2.4 Conclusion

In conclusion, these results demonstrate the first example of the biocatalytic conversion of simple, cyclic monosaccharides to the corresponding  $\omega$ -aminopolyols using ATAs. The screening of a large ATA library for activity towards these sugars using a colorimetric assay employing *o*-xylylenediamine suggests these substrates are not widely accepted by this enzyme class, and as such, no ATAs were identified that accepted ketose sugars. However, three

ATAs were selected for analytical-scale reactions and activity towards a small panel of aldoses has been documented. This allowed for the optimal conditions with these substrates to be determined. Further to this, a selection of amino alcohols were synthesized on a preparative scale using the commercially available ATA-256. This methodology allowed for the facile attainment of up to 475 mg of amino alcohol product, which clearly demonstrates the scalability of this approach giving it the potential to become the primary route for the synthesis of such amino alcohols from simple monosaccharides in industry.

## 2.5 Experimental

### General Methods and Materials

**General:** NMR spectra were recorded on a Bruker AV(III)500 ( $^1\text{H}$  500 MHz), Bruker Avance 400 spectrometer ( $^1\text{H}$  400 MHz,  $^{13}\text{C}$  100 MHz) and are referenced internally according to a residual solvent signal. The chemical shifts ( $\delta$ ) were recorded in ppm and the coupling constants ( $J$ ) were reported in Hz, are corrected with reference to the apparent peak multiplicities. HRMS spectra were recorded on a Bruker MicroTOF (Time of Flight) mass spectrometer using Electron Spray Ionisation (ESI).

**Materials:** Commercially available reagents and solvents were purchased from Acros Chemicals, Fluorochem, Sigma Aldrich and Thermo Fisher Scientific. Commercially available transaminase ATA-256 was purchased from Codexis in the form of lyophilised cell extract.

### Transaminase library screening using *o*-xylylenediamine

Four 96 well plates containing LB and either ampicillin or kanamycin 100  $\mu\text{L}/\text{mL}$  were inoculated with *E. coli* BL21(DE3) glycerol cell stocks, which

transformed with a variety of transaminase wild types and mutants. The plates were shaken at 200 rpm for 6 h at 37°C before being induced with IPTG (0.1 M) or rhamnose (0.2%) and left to express at 20°C overnight, again at 200 rpm. The plates were then centrifuged (4900 rpm, 20 minutes, 4°C) to collect the cell pellets, which were washed with HEPES (50 mM, pH 8), PLP (0.1 mM), before being centrifuged again. The cell pellets were then resuspended in the wash buffer with DNase 1 µg/mL and 1 mg/mL and left for 90 minutes at 30°C to break open the cells. After further centrifugation (4900 rpm, 20 minutes, 4°C) the supernatant containing the cell free lysate was transferred to fresh 96 well plates. 50 µL of the supernatant was then added to 150 µL biotransformations in 96 well plates which contained 7.5 mM of each sugar substrate (either 3 or 4 sugars per well), *o*-xylylenediamine (7.5 mM), HEPES (50 mM), 0.5 mM PLP. The biotransformations were then shaken at 200 rpm for 24 h at 30°C and any colour changes were recorded.

**Enzyme preparation:** The plasmids containing the genes for  $\omega$ -transaminase enzymes HEWT,<sup>180</sup> 3HMU,<sup>184</sup> were used to transform *E. coli* BL21(DE3) competent cells for gene expression.

HEWT was inoculated into auto induction media (containing LB) with ampicillin (100 µL/mL) and grown at 37 °C for 24 h. 3HMU was inoculated into 300 mL LB containing ampicillin (100 µL/mL) and grown at 37 °C until the OD<sub>600</sub> reached ~0.7 and then induced with IPTG (0.1 mM) at 20 °C for 18 h. The cells were harvested by centrifugation at 4000 rpm for 15 min and the cell pellets were resuspended in phosphate buffer (100 mM, pH 8) containing PLP (0.1 mM). Cells were disrupted by ultrasonication for 10 cycles of 30s sonication and 30s of cooling. The suspension was centrifuged at 15000 rpm for 30 min

to yield a clear lysate. Presence of the protein was confirmed using an SDS page and activity was compared to published data using an acetophenone assay on an EPOCH2 microplate reader with (*S/R*)-MBA (2.5 mM), pyruvate (2.5 mM) and enzyme.<sup>169</sup> The cell free lysate was then used for subsequent biotransformations.

**Protein purification:** The lysate was then purified on an ÄKTA start with the resuspension buffer plus imidazole (30 mM) and NaCl (100 mM) used for the wash buffer and imidazole (300 mM) for the elution buffer. The purified protein was then desalted at 4 °C overnight using dialysis tubing in tis buffer (100 mM, pH 9), PLP (0.1 mM).

### **Analytical scale biotransformations of sugars substrates**

#### **General procedure and NMR experiment**

(*S*)-selective ATA 256 (2.5 mg/ml), HEWT (250 µL CFE) or 3HMU (0.1 mg/mL purified enzyme) was rehydrated/diluted in KPi buffer (1 mL, 100 mM , pH 9.0, 7.5 or 8.0) containing PLP (1 mM), and amine donor (11 or 30 mM from a 500 mM stock in KPi buffer, pH 9.0, 7.5 or 8.0). To this was added the sugar substrate (10 mM from a 100 mM stock in KPi buffer, pH 9.0, 7.5 or 8.0). The reaction mixture was incubated at 30 °C, 200 rpm in a shaking incubator. After 24 hours, a 500 µL aliquot of the reaction was added to 50 µL of maleic acid (110 mM in D<sub>2</sub>O) and analysed by NMR. (Water suppressed <sup>1</sup>H NMR spectra were recorded using a zgcprr pulse sequence on a Bruker AV(III)500 instrument fitted with a 5mm autotunable dual <sup>1</sup>H/<sup>13</sup>C (DCH) cryoprobe. Data was collected with 64k points with a sweep width of 20 ppm. Experiments were

run with 64 scans using a relaxation delay of 10 seconds and an acquisition time of 3.2 seconds. Experiments were carried out at 298K.

Post NMR: manual phase correction was performed around the maleic acid peak, when required. Automated baseline correction, Bernstein polynomial fit "order 3". The identifiable peak area of the aminoalcohol was compared to the maleic acid peak area and a conversion factor (x 1.1) was applied to determine the concentration of aminoalcohol (see example in figure 7). The reactions were performed in triplicate and the average conversion and standard deviations were calculated.

#### **Optimisation studies carried out with enzymes and 2-deoxy-D-ribose 40**

Analytical assays were carried out with ATA256, HEWT and 3HMU to establish the best pH, temperature, and amine donor/concentration.

#### **Amine donor screen**

(S)-selective ATA 256 (2.5 mg/ml) was rehydrated in KPi buffer (1 mL, 100 mM, pH 9.0) containing PLP (1 mM), and amine donor (11 mM or 30 mM from a 550 mM stock in KPi buffer, pH 9.0). To this was added 2-deoxy-D-ribose (10 mM from a 100 mM stock in KPi buffer, pH 9.0). The reaction mixture was incubated at 30 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

HEWT (250 µL CFE) or 3HMU (0.1 mg/mL purified enzyme) was diluted in KPi buffer (1 mL, 100 mM, pH 7.5) containing PLP (1 mM), and amine donor (11 mM or 30 mM from a 550 mM stock in KPi buffer, pH 7.5). To this was added 2-deoxy-D-ribose (10 mM from a 100 mM stock in KPi buffer, pH 7.5). The

reaction mixture was incubated at 30 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

### **Temperature screen**

(S)-selective ATA 256 (2.5 mg/ml) was rehydrated in KPi buffer (1 mL, 100 mM, pH 9.0) containing PLP (1 mM), and *o*-xylylenediamine (11 mM from a 550 mM stock in KPi buffer, pH 9.0). To this was added 2-deoxy-D-ribose (10 mM from a 100 mM stock in KPi buffer, pH 9.0). The reaction mixture was incubated at 30 °C or 50 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

HEWT (250 µL CFE) was diluted in KPi buffer (1 mL, 100 mM, pH 7.5) containing PLP (1 mM), and *o*-xylylenediamine (11 mM from a 550 mM stock in KPi buffer, pH 7.5). To this was added 2-deoxy-D-ribose (10 mM from a 100 mM stock in KPi buffer, pH 7.5). The reaction mixture was incubated at 30 °C or 50 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

3HMU (0.1 mg/mL purified enzyme) was diluted in KPi buffer (1 mL, 100 mM, pH 8.0) containing PLP (1 mM), and *o*-xylylenediamine (11 mM from a 550 mM stock in KPi buffer, pH 8.0). To this was added 2-deoxy-D-ribose (10 mM from a 100 mM stock in KPi buffer, pH 8.0). The reaction mixture was incubated at 30 °C or 50 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

### **pH screen**

(S)-selective ATA 256 (2.5 mg/ml), HEWT (250 µL CFE) or 3HMU (0.1 mg/mL purified enzyme) was rehydrated/diluted in KPi buffer (1 mL, 100 mM, pH 7-

10) containing PLP (1 mM), and *o*-xylylenediamine (11 mM from a 550 mM stock in KPi buffer). To this was added 2-deoxy-D-ribose (10 mM from a 100 mM stock in KPi buffer). The reaction mixture was incubated at 30 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

### **Sugar substrate screen**

(S)-selective ATA 256 (2.5 mg/ml) was rehydrated in KPi buffer (1 mL, 100 mM, pH 9.0) containing PLP (1 mM), and *o*-xylylenediamine (11 mM from a 550 mM stock in KPi buffer, pH 9.0). To this was added the sugar substrate (10 mM from a 100 mM stock in KPi buffer, pH 9.0). The reaction mixture was incubated at 30 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

HEWT (250 µL CFE) was diluted in KPi buffer (1 mL, 100 mM, pH 7.5) containing PLP (1 mM), and *o*-xylylenediamine (11 mM from a 550 mM stock in KPi buffer, pH 7.5). To this was added the sugar substrate (10 mM from a 100 mM stock in KPi buffer, pH 7.5). The reaction mixture was incubated at 30 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

3HMU (0.1 mg/mL purified enzyme) was diluted in KPi buffer (1 mL, 100 mM, pH 8.0) containing PLP (1 mM), and *o*-xylylenediamine (11 mM from a 550 mM stock in KPi buffer, pH 8.0). To this was added the sugar substrate (10 mM from a 100 mM stock in KPi buffer, pH 8.0). The reaction mixture was incubated at 30 °C, 200 rpm in a shaking incubator. After 24 hours, the reaction was analysed by NMR following the general procedure.

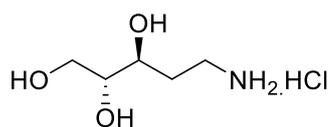
## **General procedure for the chemical synthesis of aminoalcohols<sup>186</sup>**

To the corresponding sugar **40-48** (1 mM) dissolved in a saturated solution of NH<sub>4</sub>OAc in EtOH (20 mL) were added NaCNBH<sub>3</sub> (188 mg) and 30% aq. NH<sub>3</sub> (8 mL). The mixture was stirred at reflux for 18 h, then cooled to room temperature and concentrated under reduced pressure. The crude residue was dissolved in H<sub>2</sub>O and loaded on to Dowex 50WX8 hydrogen form and washed thoroughly with H<sub>2</sub>O. The amine product was then eluted with 30% aq. NH<sub>3</sub> (100 mL). The eluent was concentrated under reduced pressure, 1M HCl (50 mL) was added and concentrated under reduced pressure to provide the corresponding aminoalcohol hydrochloride salts.

## **Preparative scale biotransformations of sugars substrates**

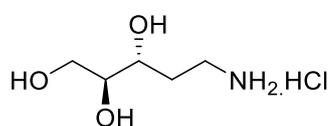
Commercially available (*S*)-selective ATA 256 (50 mg) was rehydrated in ammonia carbonate buffer (20 mL, 100 mM, pH 9.0) containing PLP (1 mM) and amine donor (400 or 800 mM). To this was added the sugar substrate (100 or 200 mM). The reaction mixture was incubated at 50 °C, 200 rpm in a shaking incubator and the reaction monitored by <sup>1</sup>H NMR. After 48 h, the reaction mixture was concentrated under reduced pressure, hot methanol (20 mL) was added, filtered and concentrated under reduced pressure. The crude residue was dissolved in H<sub>2</sub>O and loaded on to Dowex 50WX8 hydrogen form and washed thoroughly with H<sub>2</sub>O. The amine product was then eluted with 30% aq. NH<sub>3</sub> (100 mL). The eluent was concentrated under reduced pressure, 1M HCl (50 mL) was added and concentrated under reduced pressure to provide the amino alcohol hydrochloride salts.

## **(2*R*,3*S*)-5-aminopentane-1,2,3-triol hydrochloride **49****



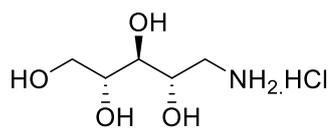
Prepared from 2-deoxy-D-ribose **40** (200 mM). Brown solid (475 mg, 69% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  3.83 – 3.69 (m, 2H), 3.67 – 3.50 (m, 2H), 3.28 – 3.06 (m, 2H), 2.07 – 1.95 (m, 1H), 1.87 – 1.73 (m, 1H).;  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  74.3, 69.6, 62.3, 37.2, 29.3; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_5\text{H}_{14}\text{NO}_3^+$   $[\text{M}+\text{H}]^+$ : 136.0968; found: 136.0971. IR (ATR) 3316, 2962, 1614, 1505, 1464, 1388, 1066, 1029  $\text{cm}^{-1}$ .

### **(2S,3R)-5-aminopentane-1,2,3-triol hydrochloride 50**



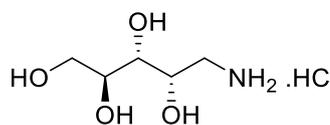
Prepared from 2-deoxy-L-ribose **41** (200 mM). Brown solid (474 mg, 69% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  3.87 – 3.68 (m, 2H), 3.66 – 3.50 (m, 2H), 3.29 – 3.06 (m, 2H), 2.11 – 1.91 (m, 1H), 1.86 – 1.73 (m, 1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  74.3, 69.6, 62.3, 37.2, 29.3; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_5\text{H}_{14}\text{NO}_3^+$   $[\text{M}+\text{H}]^+$ : 136.0968; found: 136.0970. IR (ATR) 3320, 2982, 1628, 1058, 1013  $\text{cm}^{-1}$ .

### **(2R,3S,4S)-5-aminopentane-1,2,3,4-tetraol hydrochloride 51**



Prepared from D-ribose **42** (100 mM). Brown solid (176 mg, 47% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  4.10 – 4.01 (m, 1H), 3.85 – 3.70 (m, 3H), 3.68 – 3.58 (m, 1H), 3.32 – 3.23 (m, 1H), 3.15 – 3.01 (m, 1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  72.8, 71.8, 67.8, 62.5, 41.0; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_5\text{H}_{14}\text{NO}_4^+$   $[\text{M}+\text{H}]^+$ : 152.0917; found: 152.0917. IR (ATR) 3302, 2988, 1631, 1528, 1453, 1394, 1070  $\text{cm}^{-1}$ .

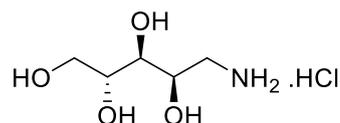
### **(2S,3R,4S)-5-aminopentane-1,2,3,4-tetraol hydrochloride 52**



Prepared from L-arabinose **43** (100 mM). Brown solid (104 mg, 28% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  4.21 – 4.13 (m, 1H), 3.90 – 3.80 (m, 1H), 3.78 – 3.72 (m, 1H), 3.72 – 3.64

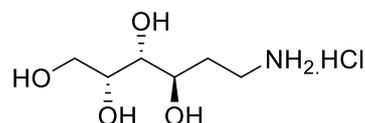
(m, 1H), 3.58 – 3.50 (m, 1H), 3.24 – 3.13 (m, 2H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  71.2, 70.6, 66.4, 62.8, 42.6; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_5\text{H}_{14}\text{NO}_4^+$   $[\text{M}+\text{H}]^+$ : 152.0917; found: 152.0915. IR (ATR) 3348, 2925, 1631, 1514, 1461, 1061  $\text{cm}^{-1}$ .

### **(2R,3S,4R)-5-aminopentane-1,2,3,4-tetraol hydrochloride 53**



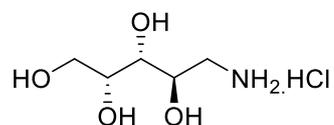
Prepared from D-arabinose **44** (100 mM). Brown solid (144 mg, 38% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  4.24 – 4.11 (m, 1H), 3.88 – 3.80 (m, 1H), 3.78 – 3.72 (m, 1H), 3.71 – 3.64 (m, 1H), 3.58 – 3.50 (m, 1H), 3.22 – 3.14 (m, 2H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  71.2, 70.5, 66.4, 62.8, 42.6; ; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_5\text{H}_{14}\text{NO}_4^+$   $[\text{M}+\text{H}]^+$ : 152.0917; found: 152.0917. IR (ATR) 3294, 2935, 1622, 1513, 1459, 1001  $\text{cm}^{-1}$ .

### **(2R,3R,4R)-6-aminohexane-1,2,3,4-tetraol hydrochloride 54**



Prepared from 2-deoxy-D-galactose **45** (100 mM). Light brown solid (88 mg, 22% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  3.95 – 3.88 (m, 1H), 3.85 – 3.77 (m, 1H), 3.70 – 3.62 (m, 2H), 3.52 – 3.44 (m, 1H), 3.28 – 3.10 (m, 2H), 2.18 – 2.05 (m, 1H), 1.88 – 1.74 (m, 1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  73.1, 70.1, 68.8, 63.0, 37.2, 30.0; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_6\text{H}_{16}\text{NO}_4^+$   $[\text{M}+\text{H}]^+$ : 166.1074; found: 166.1075. IR (ATR) 3359, 2978, 1613, 1506, 1469, 1397, 1380, 1060  $\text{cm}^{-1}$ .

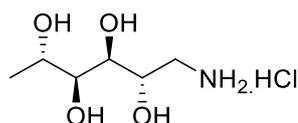
### **(2R,3R,4R)-5-aminopentane-1,2,3,4-tetraol hydrochloride 56**



Prepared from D-lyxose **47**. Brown solid (86 mg, 11% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  4.01 – 3.88 (m, 2H), 3.72 – 3.64 (m, 2H), 3.62 – 3.58 (m, 1H), 3.45 – 3.34 (m, 1H), 3.14 – 2.99 (m,

1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  71.9, 69.8, 67.2, 62.8, 42.3; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_5\text{H}_{14}\text{NO}_4^+$   $[\text{M}+\text{H}]^+$ : 152.0917; found: 152.0915. \*Compound was contaminated with IPA.HCl salt (~15%). This will mean that the isolated yield has been overestimated.

**(2S,3S,4S,5S)-1-aminohexane-2,3,4,5-tetraol hydrochloride 57**



Prepared from L-rhamnose **48** (100 mM). Brown solid (131 mg, 33% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  3.98 – 3.90 (m, 1H), 3.88 – 3.78 (m, 2H), 3.59 – 3.53 (m, 1H), 3.44 – 3.35 (m, 1H), 3.10 – 3.02 (m, 1H), 1.28 (d,  $J$  = 6.3 Hz, 3H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  73.0, 70.9, 67.3, 66.9, 42.4, 18.9; HRMS (ESI)  $m/z$ : Calculated  $\text{C}_6\text{H}_{16}\text{NO}_4^+$   $[\text{M}+\text{H}]^+$ : 166.1074; found: 166.1076. IR (ATR) 3350, 2976, 1613, 1514, 1469, 1396, 1378, 1065  $\text{cm}^{-1}$ .

### **3.0 Development of a Comprehensive Quantitative Assay for Amine Transaminases**

The work shown in this chapter is published: R. Cairns, A. Gomm, C. Peel, M. Sharkey and E. O'Reilly; 2019, *ChemCatChem*, **11**, 4738-4743.

#### **3.1 Introduction**

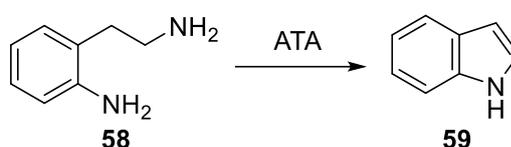
The current limitations surrounding the applications of enzymes in industry are low reaction rates, limited substrate conversion, product inhibition, and poor stability.<sup>192</sup> This is currently overcome by enzyme engineering, where large libraries of variants must be screened for desirable activity.<sup>143,154,193,194</sup> Successful protein engineering requires the development of rapid, facile, and sensitive screening methods whereby a selection bias is applied, and advantageous mutants can be identified for subsequent rounds of evolution.<sup>195</sup> As discussed in chapter 1.7, many screening techniques have been developed for ATAs, however, each comes with some drawbacks. The screens published thus far largely only work in one direction i.e., they are either an amine donor or amine acceptor screen. Furthermore, whilst the novel ATA activity towards sugars described in chapter 2 is exciting, activity on a broader range of monosaccharides, particularly ketoses, is desirable. A fast, reliable high-throughput screening protocol allowing for the facile identification of desirable variants is required to engineer ATAs that can perform this chemistry.

Arguably, the most widely used approach for small-scale screening is the acetophenone assay, which uses either (*R*)- or (*S*)-methylbenzylamine as an amine donor and measures the development of acetophenone using a UV spectrophotometer, allowing for kinetic studies and the detection of very low levels of activity.<sup>169</sup> This assay has several drawbacks, which led us to believe

the development of an alternative screening strategy was necessary. Firstly, the amount of protein used is limited in this assay due to the high levels of protein absorbance in the recorded region. Secondly, it is only practically feasible to screen low absorbing amine acceptors and routine screening of a broad range of UV active acceptors is not possible. Importantly the screening of alternative amine donors is not viable with this method. Also, successful utilisation of this assay on large libraries often relies on expensive robotics platforms and cannot be used for solid-state colony assessment. A screen relying on horse radish peroxidase (Chapter 1.7) is another widely used high-throughput colorimetric assay that requires a costly enzyme cascade and thus has high operational costs on scale.

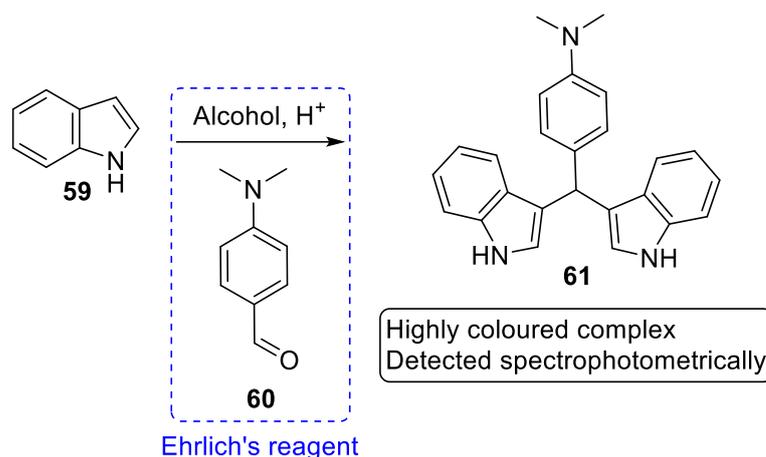
The development of the 'smart amine donor', *o*-xylylenediamine has allowed for a quick, facile and highly sensitive colorimetric screen in both the liquid and solid phase. However, the presence of cellular pyruvate likely results in a strong background after only 20 minutes of performing the colony-based assay.<sup>196</sup> This has recently been improved to over an hour in a recent publication by the introduction of a background depletion step.<sup>197</sup> Another major drawback of this screen is that it is not only purely qualitative, but, highly sensitive, resulting in difficulties distinguishing between low and high conversions. Thus, a qualitative assessment of relative conversion cannot be made, and it is difficult to select the 'best' mutants for subsequent engineering. For these reasons, there is a clear need for a quick, facile, sensitive colorimetric screen that is both qualitative and quantitative and accessible in the liquid and solid phase, allowing for rapid mutant selection in engineering projects.

It is proposed that using 2-aminoethylaniline (2AEA) (**58**) as a novel amine donor would produce the aromatic heterocycle indole (**59**) upon transamination (**Scheme 16**). As indole is significantly more thermodynamically stable than the starting material, it is thought that this transformation will also displace the equilibrium towards product formation. Furthermore, **58** is structurally similar to *o*-xylylenediamine, which many ATAs are already known to be active towards, so it was proposed they would also readily accept **58**.



**Scheme 16** - Simplistic representation of the transamination of 2-aminoethylaniline **58** to the more thermodynamically stable indole **59**.

Colorimetric screens based on the detection of indole in both solid and liquid phase are well established using Ehrlich's reagent, and thought to form the highly coloured complex **61** (**Scheme 17**).<sup>198–201</sup> We hypothesised that this screen could be combined with a subsequent transamination step and allow for the development of a colorimetric ATA screen that can either monitor the formation of indole when **58** is used as a substrate, or, the depletion of **60** when it is used a substrate. This would deliver a qualitative colorimetric assay that could be used to screen in both the forward and reverse directions and enable panels of donors and acceptors to be evaluated in a medium/high-throughput manner. Furthermore, as the concentration of indole is directly proportional to the absorption at a given wavelength, this screen would allow for the spectrophotometric monitoring of *in situ* produced or exogenously added indole.<sup>202</sup> Therefore, allowing for quantitative measurements of conversion.



**Scheme 17** - Reaction of indole (**59**) with Ehrlich's reagent (including 4-dimethylaminobenzaldehyde (**60**)), generating the highly coloured species **61**.

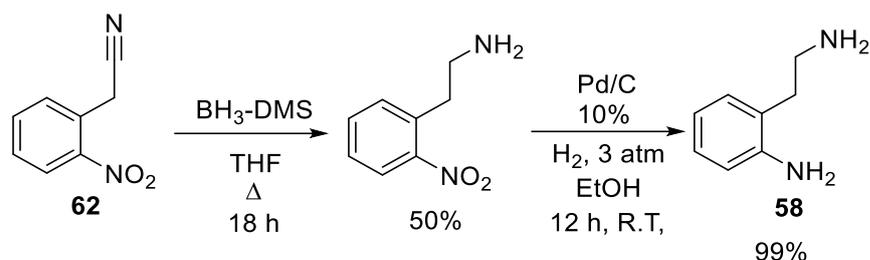
### 3.2 Aims and objectives

This work aimed to develop fully comprehensive, high throughput amine acceptor and donor screens for ATAs that are both quantitative and qualitative. This utilised **58** as an amine donor to screen for carbonyl acceptors and exploit **60** as an amine acceptor to screen for donors, which also allows for the enantiopreference of ATAs to be determined. This screen was further developed to work on solid phase to allow for rapid screening of mutant libraries.

#### 3.3.0 Results and discussion

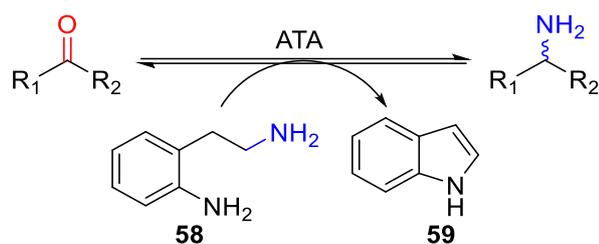
##### 3.3.1 Screening for ketone acceptors using 2-aminoethylaniline (2AEA)

2-Aminoethylaniline (2AEA) (**58**) was synthesised from commercially available (2-nitrophenyl)acetonitrile (**62**) following a previously reported synthetic route (**Scheme 18**).<sup>203</sup>

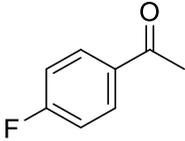
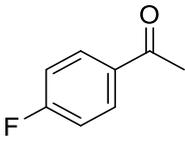
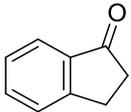


**Scheme 18** - Synthesis of 2AEA novel diamine donor from (2-nitrophenyl)acetonitrile via a borane and subsequent palladium reduction.<sup>203</sup>

As described in the introduction (**Section 3.1**), it was hypothesised **58** could be used as an amine donor generating indole as a coproduct, which could then be used for high throughput screening by adding Ehrlich's reagent (4-DMAB (**60**), conc. HCl and a suitable alcohol) leading to the development of a pink colour in the presence of indole. Initial efforts were concentrated on determining whether **58** could be readily accepted as a substrate with commercially available ATA-113, and four ketones (**63-66**) were selected to determine the applicability of this substrate. Conversions were calculated by GC-FID, with up to 84% achieved with ketone **65** and 3 equivalents of 2AEA (**Table 3**).

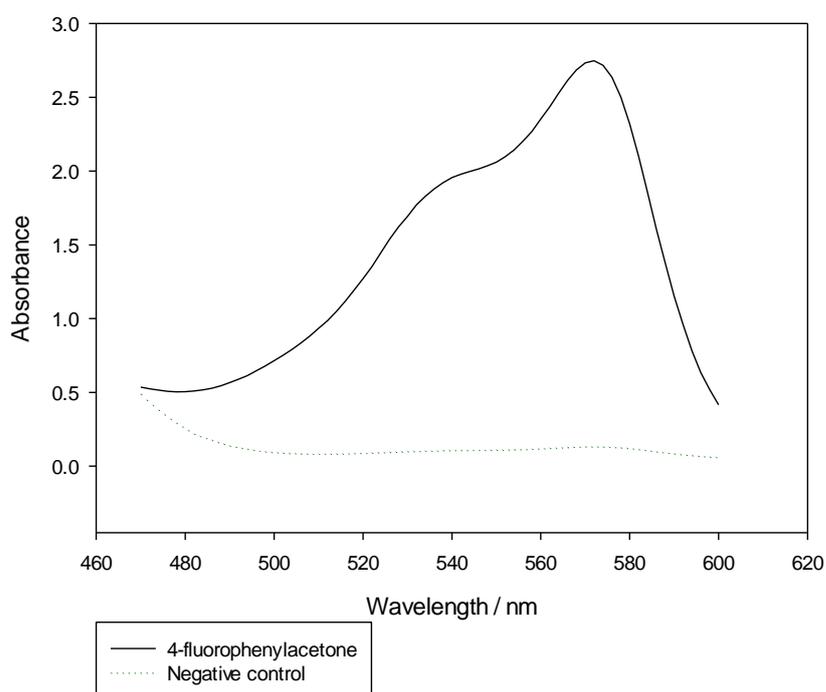


Entry	Ketone Substrate	Equiv.	Conv. %
1		2	8

2		2	62
3		2	69
4		3	84
5		2	<5

**Table 3** - Conversion of ketones (**63-66**) (5 mM) to the corresponding amines in the presence of 2-AEA (2 or 3 eq.), HEPES (100 mM, pH 7.5), PLP (1mM), ATA-113 (2.5 mg/mL), 30 °C, 24 h at 200 rpm. Conversion measured by GC-FID. Results are mean of triplicates.

To demonstrate the applicability of this route for screening purposes, Ehrlich's reagent was subsequently added to the biotransformations, which were then read spectrophotometrically to determine the absorption profile of a successful biotransformation, relative to a negative control (**Figure 13**). The maximum absorbance was recorded at 572 nm, whereby the concentration of indole is reported as being directly proportional to the intensity of absorbance.<sup>202</sup>

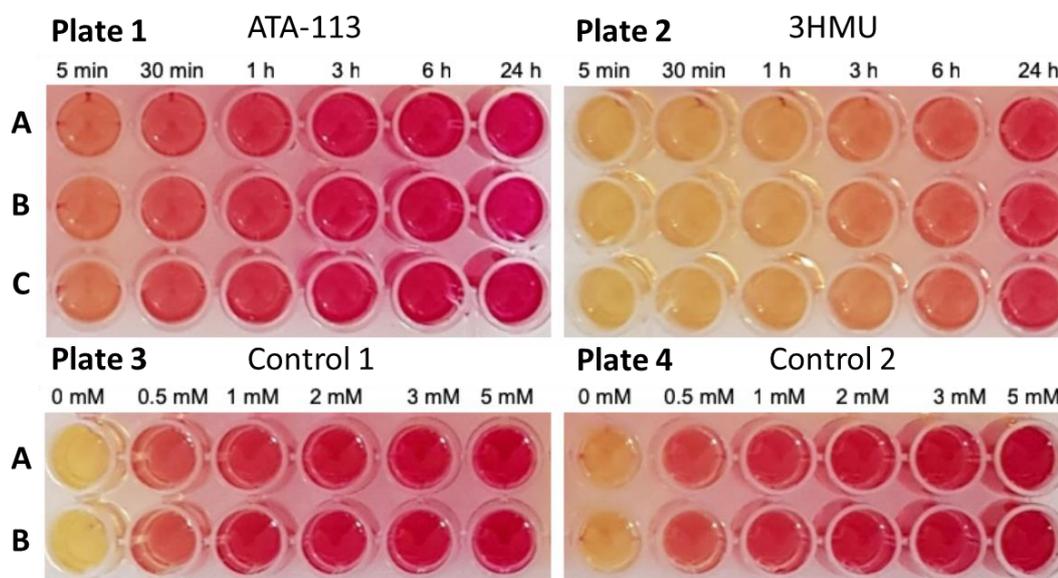


**Figure 13** – UV/vis Wavelength vs absorption from 470-600 nm of ketone 4-fluorophenylacetone **X** and a negative control (no ketone), showing that the maximum absorbance of **X** occurs at 572 nm.

### 3.3.2 Quantitative acceptor screen development

Subsequent efforts were focused on developing a fully quantitative liquid phase assay that could be used to identify ATAs that have activity towards desired ketones, by using **58** as a donor and adding Ehrlich's reagent. To ensure the screen was quantitative over a wide range of conversions, a time-course experiment from 5 mins - 24 h was performed using (4-fluorophenyl)acetone (**65**) as the acceptor and either the commercially

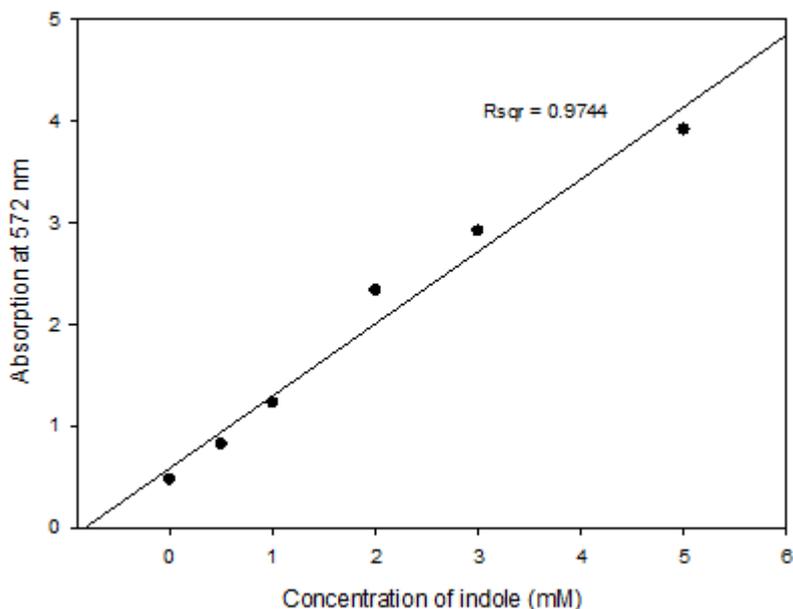
available ATA-113 or wild type 3HMU, whereby conversions by GC-FID from <1-84% using 3 equivalents of 2AEA were recorded (**Figure 14**).



**Figure 14** - Biotransformations with (4-fluorophenyl)acetone (**65**) (5 mM) and 2-AEA (**58**) (15 mM) at various time points, developed using Ehrlich's reagent. Plate 1 & 2: (4-fluorophenyl)acetone (5 mM), 2-AEA (15 mM), ATA113 (2.5 mg/mL) or 3HMU (10% lysate). Plate 3 & 4: same as plate 1 & 2 but enzyme omitted and indole added (0 mM - 5 mM). A-C are triplicates. Control 1 and 2 differ only in reaction setup for the commercial and WT enzyme as the enzymes require different conditions. See experimental for more details.

An indole concentration gradient control was used to produce a standard curve (**Figure 15**) allowing for absorbance to be directly related to the conversion; this is necessary to allow accurate conversions to be calculated from absorption. The intensity of the colour produced increases from left to right as the conversion, and hence, concentration of indole available to react with 4-DMAB increases with time. The intensity of colour on plate 2 with 3HMU is clearly less than that for plate 1 with ATA-113, which is to be expected, as the conversion with the wild type was calculated as being significantly lower (**Table 4**). The spectrophotometric analysis of the wells supported these conclusions and allowed for the calculation of conversion from absorption (**Table 4**), these

values were compared to those obtained *via* GC-FID and the results are closely comparable across the spectrum of conversions recorded.



**Figure 15** - Standard curve relating the absorption values for known concentrations of indole at 572 nm to the % conversion.

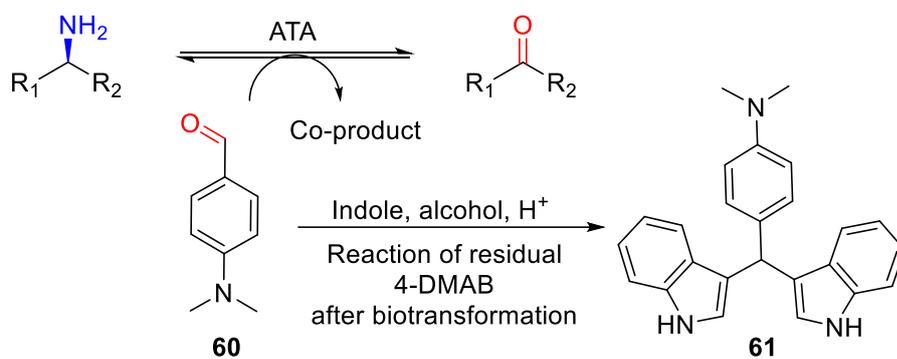
Entry	Reaction time (h)	Abs 572 nm	UV vis. % conv.	GC % conv.
<b>ATA113</b>				
<b>1</b>	1 hour	1.3	19	<b>25</b>
<b>2</b>	3 hours	2.0	38	<b>43</b>
<b>3</b>	6 hours	2.7	60	<b>64</b>
<b>4</b>	24 hours	3.5	81	<b>84</b>
<b>3HMU</b>				
<b>5</b>	1 hour	0.5	<5	<b>&lt;5</b>
<b>6</b>	3 hours	0.5	<5	<b>&lt;5</b>
<b>7</b>	6 hours	0.6	<5	<b>6</b>
<b>8</b>	24 hours	1.5	26	<b>23</b>

**Table 4** -Conversion of (4-fluorophenyl)acetone (**65**) (5 mM) to the corresponding amine with either ATA-113 or 3HMU, in the presence of 2-AEA (**58**) (15 mM), HEPES (100 mM, pH 7.5), PLP (1 mM), 30 °C at 200 rpm. Conversion determined by spectrophotometric reading of the colorimetric assay and GC-FID. Results are mean of triplicates.

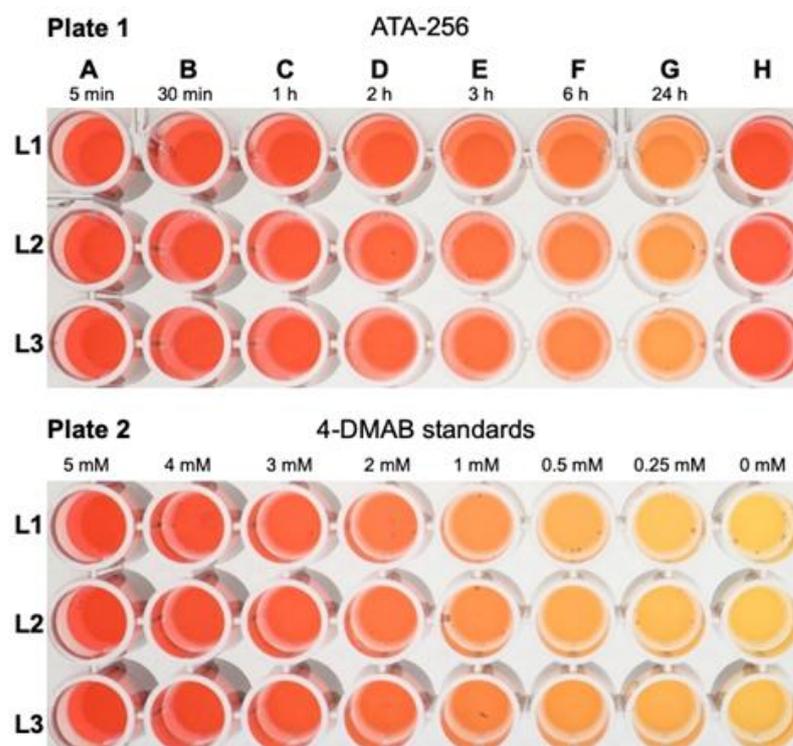
### 3.3.3 Quantitative amine donor screen development

It was then suggested that the Ehrlich test could be used to develop a quantitative amine donor screen, where 4-DMAB could be employed as an

amine acceptor. This is based on the assumption that a given ATA accepts 4-DMAB as a substrate, allowing for a range of amino donor substrates to be evaluated, by testing for the presence of residual 4-DMAB after the biotransformation. This can then react with exogenously added indole to give the bis(indolyl)methane product **61**, as in the acceptor screen (**Scheme 19**). Unlike the previous assay, a lower absorbance value would suggest a higher conversion. This methodology was demonstrated by using commercially available (*S*)-selective ATA-256, (*S*)-MBA and 4-DMAB, and monitored at various time points over a 24 h period (**Figure 16**).

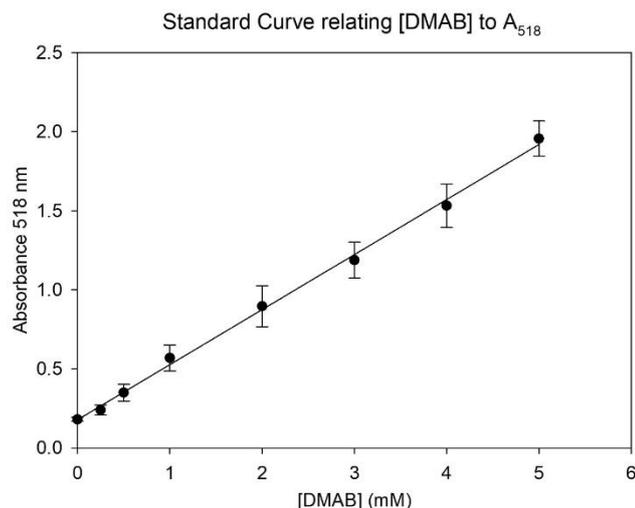


**Scheme 19** - Strategy for screening amine donors, exploiting 4-DMAB as the acceptor and measurement of the unreacted starting material to determine conversion.



**Figure 16** - Donor screen. Plate 1: biotransformations with (S)-MBA (5 mM), 4-DMAB (5 mM) and ATA-256 (1 mg/mL), monitored over 24 hours, and developed using Ehrlich's reagent. H = no enzyme control. Plate 2: 4-DMAB standards with enzyme omitted. L1–3 are replicates.

The reactions were developed using a combination of indole and HCl in EtOH/DMSO. As the reaction time increases (A–G plate 1), there is a substantial decrease in colour intensity, and absorbance, caused by a decrease in the concentration of residual 4-DMAB that reacts with the exogenously added indole. 4-DMAB standards were analysed in parallel to generate a standard curve (**Figure 17**).



**Figure 17** - A standard graph ( $R^2 = 0.998$ ), demonstrating the linear relationship between absorption at 518 nm, and DMAB concentration up to 5 Mm.

UV/VIS spectrophotometrically predicted values correlated well with GC conversions measured independently (**Table 5**) across a broad concentration range. This demonstrates the applicability of this protocol as a fully quantitative, spectrophotometric screen for screening amine donors. Notably, high conversions, of up to 82%, were recorded using 4-DMAB as an acceptor, demonstrating its suitability as a donor for screening.

Entry	Reaction Time	Abs. % conv.	GC % conv.
<b>ATA-256</b>			
<b>1</b>	6 min	<5	<b>&lt;5</b>
<b>2</b>	30 min	14	<b>14</b>
<b>3</b>	1 hour	27	<b>23</b>
<b>4</b>	2 hours	41	<b>40</b>
<b>5</b>	3 hours	51	<b>54</b>
<b>6</b>	6 hours	67	<b>65</b>
<b>7</b>	24 hours	78	<b>82</b>

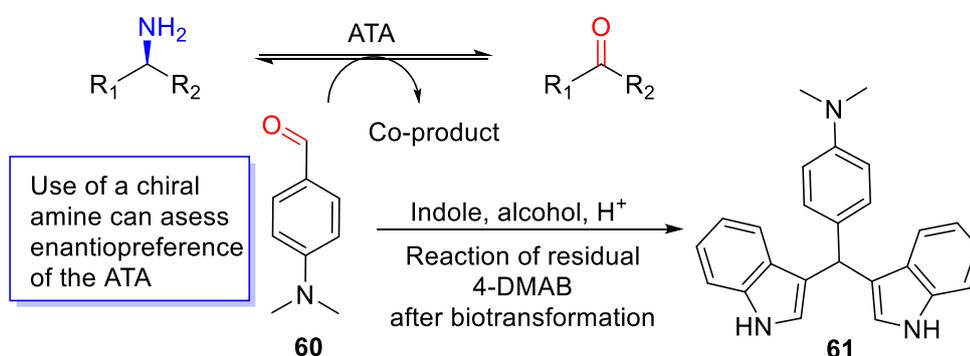
**Table 5** - Donor screen: conversion of (S)-MBA (5 mM) to acetophenone in the presence of 4-DMAB (5 mM) and ATA-256, HEPES (100 mM, pH 7.5), 30 °C at 200 rpm. Results are the mean of triplicates.

Interestingly, exposure to light accelerated colour formation after addition of reagents during the donor screen, whereas it was instantaneous for the acceptor screen. The key difference between both assays is the application of the Ehrlich's test, 4-DMAB is in huge excess for the acceptor screen, whereas it is the amine donor in large excess during the donor screen. This suggests there could be a mechanistic route for the differences in reaction kinetics. There was some debate in the literature as to the exact structure of the coloured compound. Therefore, compound **61** was synthesized chemically and compared to the product isolated subsequent to a biotransformation. The NMR data confirms that compound **61** is formed under these reaction conditions, although, the exact nature of the molecule providing the colour is not fully understood, as it is not clear how light affects the molecule and/or its mechanism of formation. Dr Michael Sharkey, of the O'Reilly group, ran several experiments to determine the effect of light on colour formation, including the wavelengths of light used and the optimisation of conditions (see SI of the published work).<sup>204</sup>

### **3.3.4 Enantiopreference screen**

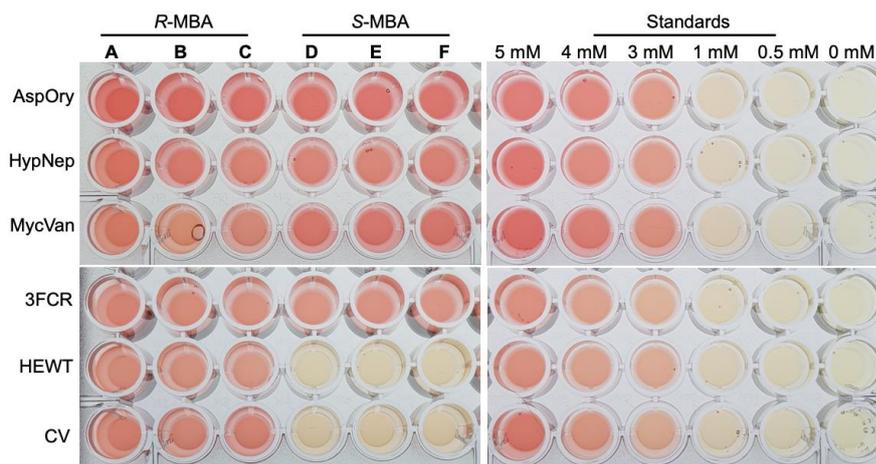
When characterising novel ATAs it is important to accurately determine their enantiopreference and without a suitable screening protocol this can be time consuming, particularly when many enzymes are to be characterised in parallel. Screening methodologies for ATAs published to date most commonly screen for amine acceptors, however, the identification of a novel donor screen presents the opportunity to screen for the enantiopreference of this enzyme class when using chiral amines. Therefore, this concept of using 4-DMAB (**60**) as an acceptor for screening amine donors was then taken and applied to a

panel of (*R*)- and (*S*)-selective ATAs, to demonstrate that this screening protocol can be used for the high-throughput determination of enantiopreference (**Scheme 20**). Here, we can expect the absorption to decrease when a chiral amine with the correct chirality is used, as **60** is consumed. Whereas, the maximum absorption would be observed when the wrong chirality is used due to the amine, and hence, **60** not being transaminated.



**Scheme 20** – Strategy for the high-throughput of ATA enantiopreference, exploiting 4-DMAB as the acceptor and a chiral amine as the donor. Subsequent measurement of the unreacted starting material can determine whether (*S*)- or (*R*)- amines have been accepted.

Three (*R*)-selective enzymes AspOry<sup>142</sup>, HypNep<sup>142</sup>, MycVan<sup>142</sup> and three (*S*)-selective ATAs 3FCR<sup>184</sup>, HEWT<sup>180</sup>, CV<sup>205</sup> were screened using 4-DMAB (**60**) and either (*R*)- or (*S*)-methylbenzylamine (MBA) followed by the subsequent addition of indole/conc. HCl after 24 h incubation (**Figure 18**).



**Figure 18** - Biotransformations with either (R)- or (S)-MBA (5 mM), 4-DMAB (5 mM), cell lysate (25% well volume), KPi buffer (100 mM, pH 8), PLP (1 mM), 30 °C, 24 h. Reactions were performed in triplicate. Standards contain 4-DMAB (0-5 mM) and (S)-MBA (enzyme omitted).

HEWT and CV showed the highest conversions at 81% and 80% respectively, where the decrease in colour is both visually apparent and quantitatively measurable spectrophotometrically. Although significantly lower conversions were recorded for the other ATAs, they were still measurable and the values calculated by UV-vis were comparable to those from GC-FID even at low concentrations, hence, demonstrating the applicability of this methodology as a highly sensitive screen (**Table 6**).

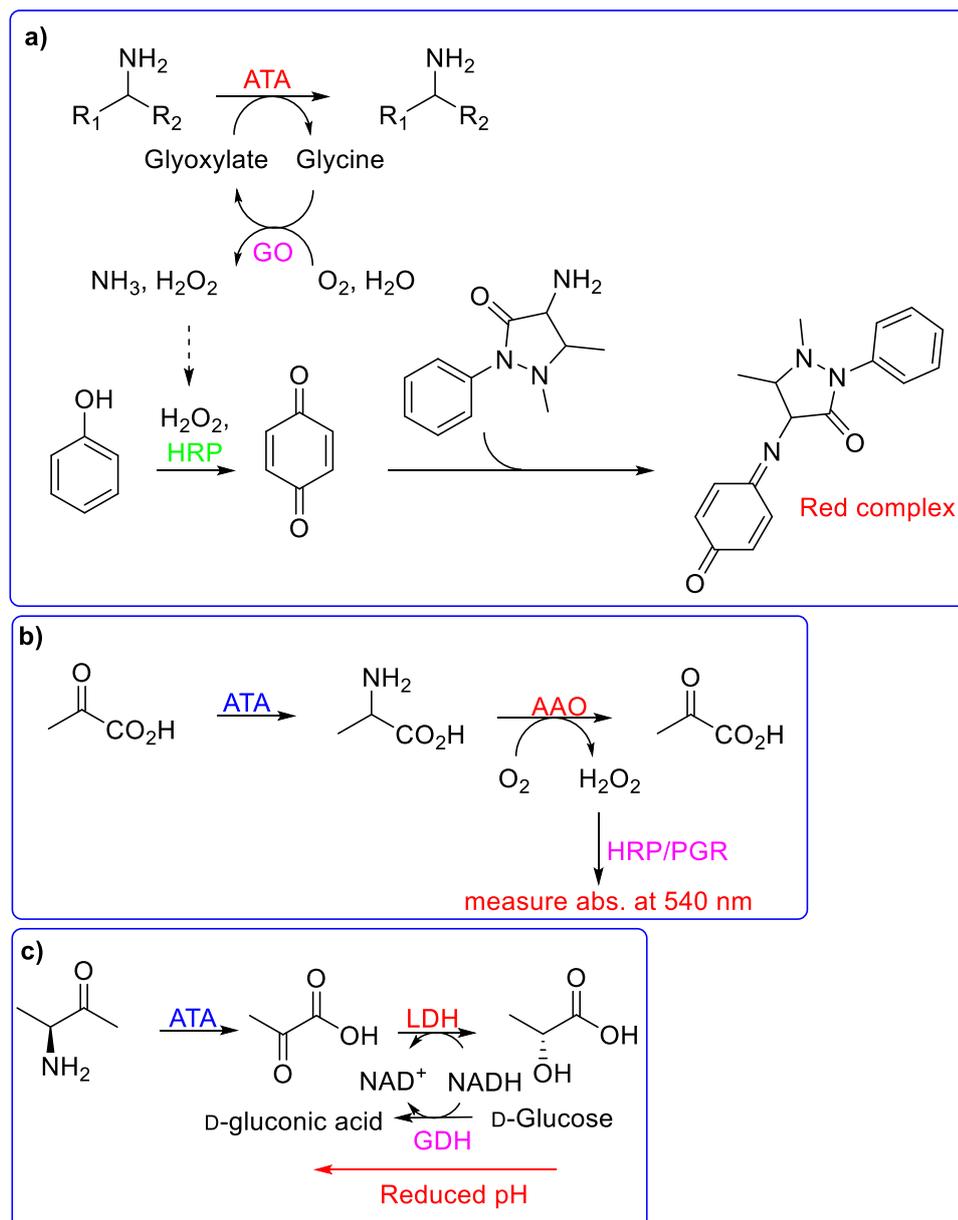
ATA	(R)-MBA		(S)-MBA	
	UV Vis conv. (%)	GC conv. (%)	UV vis conv. (%)	GC conv. (%)
<b>AspOry</b>	<1	<1	<1	<1
<b>HypNep</b>	10	7	6	nd
<b>MycVan</b>	28	27	4	<1
<b>3FCR</b>	nd	<1	3	9
<b>HEWT</b>	<1	<1	73	81
<b>CV</b>	<1	<1	70	80

**Table 6** – Conversion of (R)- and (S)-MBA (5 mM) with 4-DMAB (5 mM), cell lysate (25% well volume), KPi buffer (100 mM, pH 8), PLP (1 mM), 30 °C, 24 h. Conversions calculated and compared from spectrophotometric assay and GC-FID. nd=not detected.

A small number of amine donor screens have been reported that could be easily adapted to determine enantioselectivity. A recent example includes a glycine based high-throughput colorimetric solid-phase screen for substrate

profiling (**Scheme 21a**).<sup>206</sup> Here, transaminase activity is upon the successful amination glyoxylate, generating glycine, which is subsequently oxidised releasing hydrogen peroxide ( $\text{H}_2\text{O}_2$ ). Horseradish peroxidase then uses this hydrogen peroxide to produce benzoquinone, which forms a red complex by a subsequent condensation with 4-aminoantipyrine. Key drawbacks of this screen are the operational complexity and the fact that it is not quantitative. Another example includes the use of pyruvate as an amine acceptor whereby the resultant amino acid is oxidised by an amino acid oxidase (AAO) generating  $\text{H}_2\text{O}_2$  as a by-product (**Scheme 21b**).<sup>170</sup> The addition of horse radish peroxidase and pyrogallol red allows for the colorimetric detection of  $\text{H}_2\text{O}_2$  at 540 nm. This screen has also been modified to allow for colorimetric colony-based screening.<sup>171</sup> Using this assay limits the amino acceptor to  $\alpha$ -keto acids (usually pyruvate), furthermore, it is not quantitative.

Alternatively, physical changes, such as pH, have been measured to identify successful transamination when alanine is used as a substrate and the subsequent addition of a lactate dehydrogenase/glucose dehydrogenase system (**Scheme 21c**).<sup>172</sup> Here, the oxidation of glucose to gluconic acid leads to a measurable reduction in pH. However, this screen is neither colorimetric nor quantitative, reducing its applicability for high-throughput screening. The screen presented herein has significant advantages over these screens, including its operational simplicity, its colorimetric and fully quantitative nature, and its capability for high-throughput applications.



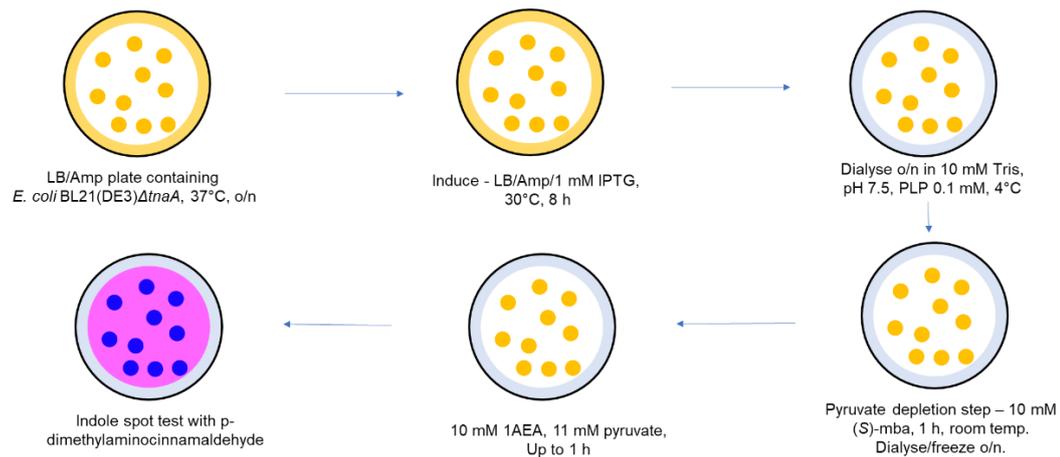
**Scheme 21** – Several amine donor screens that could be adapted to assess enantioselectivity. a) A glycine based high-throughput colorimetric solid-phase screen for substrate profiling. b) Colorimetric detection of  $\text{H}_2\text{O}_2$  by-product when using pyruvate as an amino acceptor followed by the enzymatic oxidation of alanine. c) An enzymatic cascade reaction with a lactate dehydrogenase/glucose dehydrogenase system to measure pH change.

### 3.3.5 Development of a solid-phase acceptor screen

In order to expand the applicability of these assay methodologies to allow for high-throughput screening of mutant libraries, a solid-phase screen was developed. This will allow for a much higher number of variants to be screened in parallel. However, the enzyme tryptophanase (EC4.1.99.1), which hydrolytically cleaves tryptophan to give indole, pyruvate and ammonia, exists

in indole positive bacteria.<sup>207</sup> In *E. coli* BL21(DE3), which is a commonly used laboratory expression strain that is also indole positive, the enzyme is encoded by the gene *tnaA*.<sup>208</sup> Initial studies showed that the concentration of indole generated *via* this pathway was high enough that it was not possible to distinguish from the indole produced as a co-product in the biotransformation with 2-AEA. To enable the use of this strain for such high-throughput screening, it was first necessary to remove tryptophanase from the host strain. Therefore, Dr Michael Sharkey used the in-frame gene deletion method of Link *et al.*,<sup>209</sup> whereby the gene open reading frame is removed in its entirety. This methodology ensures no antibiotic resistance marker remains, meaning that any protein expression plasmid may be subsequently used for the generation of variant ATA libraries.

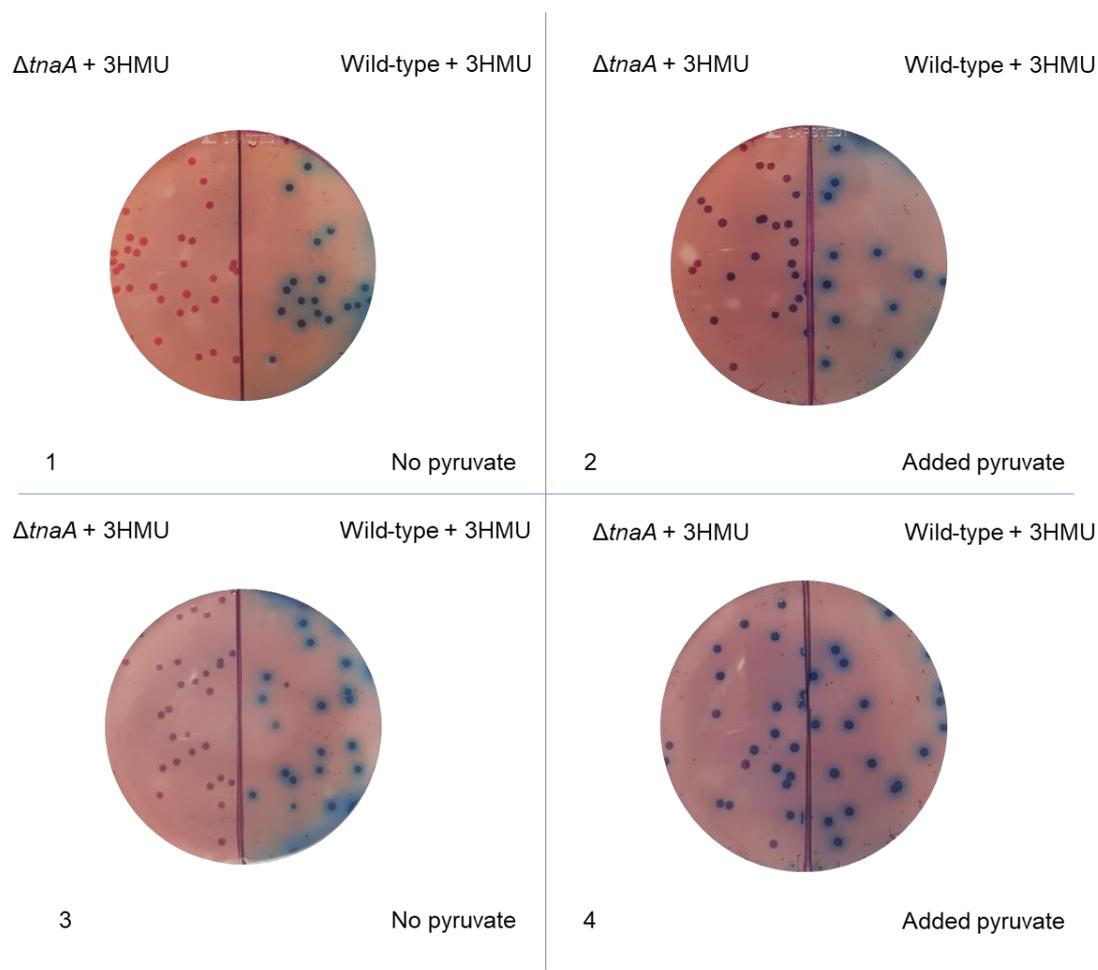
Colony-based screening, which relies on the detection of indole, typically substitutes 4-DMAB with *p*-dimethylaminocinnamaldehyde (DMACA), and this was used during the assay development.<sup>199</sup> Interference from pyruvate naturally produced within *E. coli* was noted as limiting the length of time the biotransformation can be performed. However, a pre-incubation step recently reported by the Paradisi group, for use with the *o*-xylylenediamine screen, was utilised to overcome this problem.<sup>197</sup> The principle of this pre-incubation step involves the addition of an alternative scavenging amine donor, such as (S)-MBA, prior to addition of the 'smart amine donor'. This operationally simple step results in the transamination of cellular pyruvate, using the scavenger donor, and hence the intracellular concentration is depleted. Further to this, the cells were also frozen overnight to significantly reduce their metabolism and ability to produce pyruvate (see **Figure 19** for illustrated protocol).



**Figure 19** – Illustrated diagram of the solid phase screening protocol. Cells are initially grown o/n on LB plates containing the necessary antibiotic. They are then induced with 1 mM IPTG for 8 hours before overnight (o/n) dialysis in Tris (10 mM, pH 7.5). This is then followed by a pyruvate depletion step for 1 hour where the membrane is soaked in 10 mM (S)-mba before further dialysis. The membrane is then soaked in a solution of the acceptor to be screened, pyruvate (11 mM) in this case, for up to 1 h. This is followed by application of the indole spot test with p-dimethylaminocinnamaldehyde.

BL21(DE3) and BL21(DE3) $\Delta$ *tnaA* were transformed with a pET22b vector harbouring the 3HMU gene, as it had previously been noted as accepting 2AEA as a substrate (during development of the liquid assay). Upon exposure of 2-AEA, and in the absence of an external pyruvate source, the BL21(DE3) (indole positive) colonies rapidly turned blue (Plate 1), whilst BL21(DE3) $\Delta$ *tnaA* (indole negative) remained pink (**Figure 20**). However, in the presence of pyruvate as a substrate, the  $\Delta$ *tnaA* colonies turned blue, due to the production of the indole co-product, which reacts with DMACA/conc. HCl (Plate 2). After 1-hour incubation (Plates 3 & 4), no significant background was evident in the knockout (KO) strain. These results clearly demonstrate the applicability of this

solid-phase assay for the high-throughput screening of large ATA variant libraries using BL21(DE3) $\Delta$ *tnaA*.



**Figure 20** - Solid-phase screen employing *E. coli* BL21(DE3)  $\Delta$ *tnaA* (indole negative tryptophanase KO strain), alongside BL21(DE3) (indole positive expression strain). 1) 2-AEA (11mM), 30 min incubation at r.t. 2) 2-AEA (11 mM), pyruvate (10 mM), 30 min incubation 3) 2-AEA (11 mM), 1 h incubation at r.t. 4) 2-AEA (11 mM), pyruvate (10 mM), 1 h incubation at r.t. See experimental (**Section 3.5**) for colour development procedure.

### 3.4 Conclusion

This work outlines the development of a fully comprehensive, high-throughput amine acceptor screen that works in both liquid phase, and on colonies arrayed on agar. Furthermore, the liquid-phase assay described is atypically both colorimetric and quantitative. This required the utilisation of a novel amine donor, 2-aminoethylaniline (2-AEA), which can be utilised to screen a panel of carbonyl acceptors/ATAs. Work in parallel to produce an amine donor screen

using 4-dimethylaminobenzaldehyde as an amine acceptor, allowed for a screening protocol that can determine the enantioselectivity of ATAs, which is also highly sensitive and fully quantitative. This assay has enormous potential to simplify and significantly increase the output of ATA evolution endeavours, especially in the absence of expensive robotics platforms, and become the screening method of choice for the identification of novel substrates/characterisation of novel enzymes.

### **3.5 Experimental**

#### **General Methods and Materials**

NMR spectra were recorded on a Bruker Avance 400 spectrometer ( $^1\text{H}$  400 MHz,  $^{13}\text{C}$  100 MHz) and are referenced internally according to residual solvent signal. Chiral GC analysis was performed using a ThermoFisher 1310 chromatograph equipped with a flame ionising detector, an AI 1310 autosampler and an Agilent Technologies CP-Chirasil-Dex-CB column (25 m x 0.25 mm x 0.36 mm) using helium as the carrier gas. The front inlet temperature was set to 230 °C and the front detector was set to 250 °C. Split flow was set to 170 mL.min<sup>-1</sup> and the helium gas was set to a constant flow of 1.7 mL.min<sup>-1</sup>. Temperature profiles for individual experiments are detailed in the relevant sections.

#### **Materials**

Commercially available reagents and solvents were purchased from Acros Chemicals, Fluorochem, Sigma Aldrich and Thermo Fisher Scientific. Commercially available transaminase ATA113 was purchased from Codexis in the form of lyophilised cell extracts.

## Protein Expression

Plasmids containing the genes for  $\omega$ -transaminase enzymes HEWT, 3HMU, CV, MycVan, AspOry, HypNep and 3FCR were used to transform *E. coli* BL21(DE3) competent cells for gene expression. They were grown on solid LB agar containing  $100 \mu\text{g}\cdot\text{mL}^{-1}$  ampicillin in all cases, except for the CV clone, which was cultured with  $30 \mu\text{g}\cdot\text{mL}^{-1}$  kanamycin. Subsequent liquid cultures contained the same concentration of the appropriate antibiotic.

A culture of *E. coli* BL21(DE3) containing the HEWT expression vector was inoculated into ZYP-5052 auto-induction medium (containing LB as the base) with antibiotic and grown at  $37^\circ\text{C}$  for 24 h. The remaining clones were grown in 300 mL antibiotic-containing LB broth at  $37^\circ\text{C}$  until the  $\text{OD}_{600}$  reached  $\sim 0.7$ . Expression of 3HMU, CV and 3FCR was then induced with 0.1 mM IPTG, while 0.02% w/v rhamnose was used for MycVan, AspOry and HypNep (all at  $20^\circ\text{C}$  for 18 h). Cells were harvested by centrifugation at 4000 rpm ( $3220 \times g$ ) for 15 min in an Eppendorf model 5810R centrifuge, and the cell pellets were resuspended in 0.1 M HEPES buffer, pH 7.5, containing 0.1 mM PLP. Cells were disrupted by ultrasonication on ice for 10 cycles (30s on, 30s off) using a QSonica model Q55 sonicator. Extracts were then clarified by centrifugation at 15000 rpm ( $21,130 \times g$ ) for 30 min using an Eppendorf 5424R microcentrifuge. The presence of over-expressed protein was confirmed by SDS-PAGE, and activity in cell extracts was compared to published data using the acetophenone assay (analysing on an EPOCH2 microplate reader with (S)- or (R)-MBA (2.5 mM), pyruvate (2.5 mM) and enzyme. The cell free lysate was then used for subsequent biotransformations.

### **Ketone screen**

Commercially available (*S*)-selective ATA 113 (2.5 mg/mL) was rehydrated in 0.1 M HEPES, pH 7.5, containing 1 mM PLP and 2-(2-aminoethyl)aniline (2-AEA) (10 mM from a 500 mM stock in DMSO). To this was added the ketone substrate (5 mM from a 200 mM stock in DMSO). The reaction mixture was incubated for 24 h at 30 °C, 200 rpm in a shaking incubator. Negative controls were set up in the absence of ketone acceptor or enzyme.

The colorimetric screen was performed by adding a sample of the biotransformation (10 µL) to a well containing 90 µL of a mixture of DMSO:H<sub>2</sub>O/9:1 and mixed thoroughly. To this was added 100 µL of a solution of p-dimethylaminobenzaldehyde (0.67 M in EtOH:c.HCl/1:1) and mixed thoroughly. The colour was allowed to develop for 30 minutes and a spectral scan of the plate was performed on an EPOCH 2 microplate reader (BioTek Instruments) from 470-600 nm. The maximum absorption was recorded as 572 nm.

The remainder of the biotransformation reactions were basified (pH 13), extracted with EtOAc (750 µL), derivatised with 15 µL triethylamine and 15 µL acetic anhydride and analysed by GC-FID using the following temperature program: 40 °C hold for 2 minutes followed by 20 °C.min<sup>-1</sup> temperature rise to 150 °C and then a hold for 5 minutes followed by a 30 °C.min<sup>-1</sup> temperature rise to 200 °C and a further hold for 3 minutes. Percent conversion values were calculated by expressing the area under the curve of the product peak as a percentage of the sum of the product + substrate peak areas.

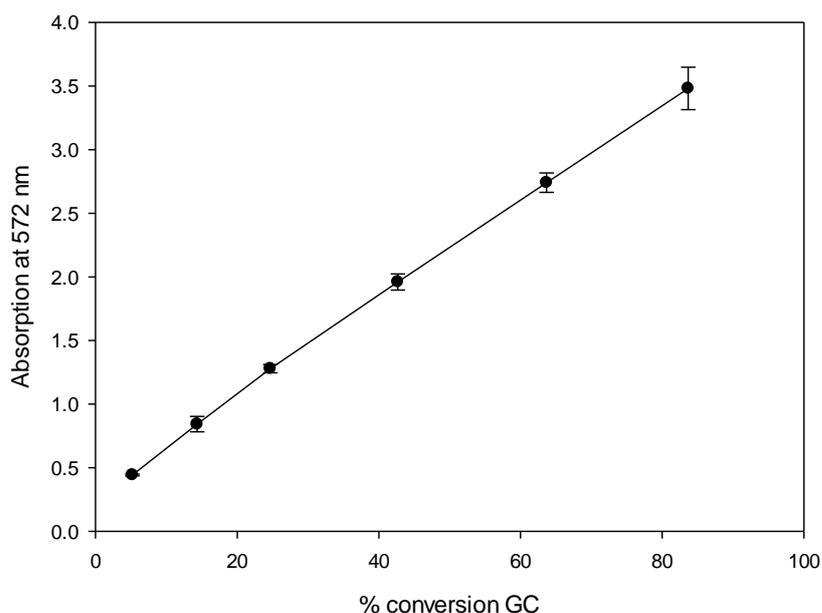
### **Acceptor screen**

Transaminases 3HMU, HEWT and CV (20  $\mu$ L cell-free extract) were diluted in HEPES buffer (1 mL, 100 mM, pH 7.5) containing PLP (1 mM) and 2-(2-aminoethyl)aniline (15 mM from a 500 mM stock in DMSO). To this was added 4-fluorophenylacetone (5 mM from a 200 mM stock in DMSO). (*S*)-selective ATA 113 (2.5 mg.ml<sup>-1</sup>) was rehydrated in 1 mL 0.1 M HEPES buffer, pH 7.5, containing 1 mM PLP and 2-(2-Aminoethyl)aniline (15 mM from a 500 mM stock in DMSO). To this was added 4-fluorophenylacetone (10 mM from a 200 mM stock in DMSO). Reaction mixtures were incubated at 30 °C, 200 rpm in a shaking incubator for up to 24 hours.

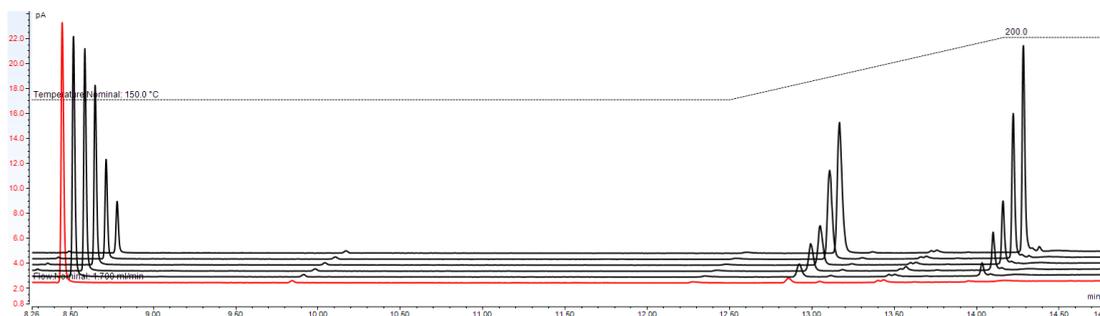
Negative controls were analogous but set up in the absence of enzyme.

The colorimetric screen was performed as detailed above for the ketone screen, and samples prepared and analysed by GC\_FID in the same way. A standard curve drawn to relate concentration of indole to absorbance at 572 nm is linear, with an R<sup>2</sup> value of 0.9744 (**Figure 15, section 3.3.2**). A linear plot of % conversion against absorption at 572 nm was then plotted (**Figure 21**).

Figure 22 shows traces from a 1 to 24-hour time-course experiment. Percent conversion values were calculated by expressing the area under the curve of the product peak as a percentage of the sum of the product + substrate peak areas. GC results showed that conversions from 25-84% were achieved during these time periods.



**Figure 21** - Plot of % conversion vs absorption at 572 nm for the ATA113 acceptor screen showing a linear relationship between the concentration of indole product and absorption.



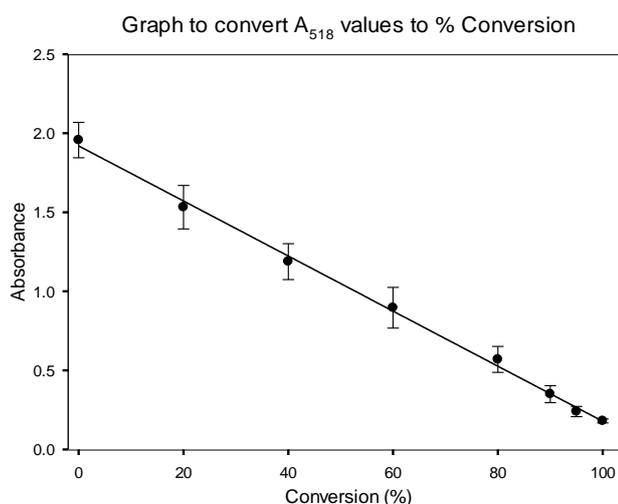
**Figure 22** - Overlay of the GC chromatograms from the acceptor screen with ATA113 showing the disappearance of 4-fluorophenylacetone (8.5 mins) with time.

## Donor screen

1 mL biotransformations were carried out at 30 °C for varying lengths of time from 6 min to 24 hr. Reactions were initiated by adding a 50  $\mu$ L solution of HEPES (0.1M, pH 7.5, PLP (1 mM), ATA-256 (20 mg/mL) ATA-256 to a 950  $\mu$ L solution of HEPES (0.1 M, pH 7.5), PLP (1 mM), (S)-MBA (5.26 mM) and 4-DMAB (5.26 mM) to give an initial substrate concentration of 5 mM, and an enzyme concentration of 1 mg/mL.

Reactions were performed in triplicate, with staggered starting times for different reaction lengths, so that reactions from all incubation times could be terminated simultaneously. At the end of the incubation period, 10  $\mu\text{L}$  samples were transferred to microtitre wells containing 50  $\mu\text{L}$  indole (50 mM (in DMSO)), to which was added 40  $\mu\text{L}$  HEPES/PLP [HEPES (0.1 M, pH 7.5), PLP (1 mM)], followed by 100  $\mu\text{L}$  DMSO/HCl (1:1 (v/v)). The plates were then exposed to high-intensity white LED light for 3-10 min. Absorbance was then measured at 518 nm using an Epoch 2 plate spectrophotometer (BioTek Instruments). A set of standards was included in each plate, consisting of 10  $\mu\text{L}$  substrate solution diluted appropriately with HEPES (0.1 M, pH 7.5), PLP (1 mM), in place of the biotransformation samples.

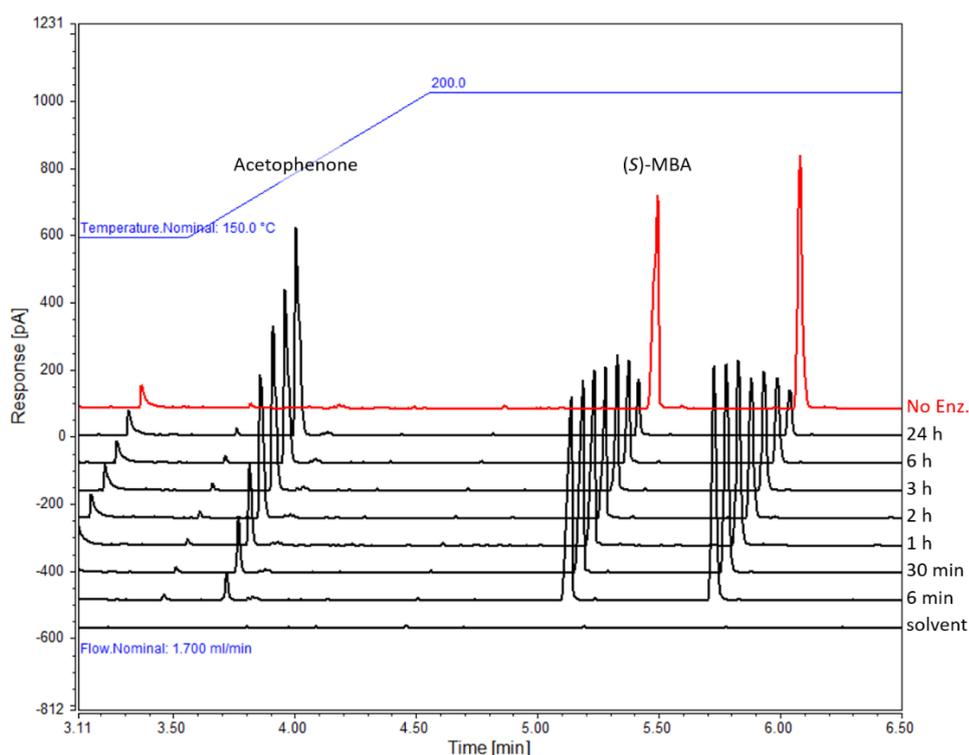
A standard curve drawn to relate concentration of DMAB to absorbance at 518 nm was linear, with an  $R^2$  value of 0.998 (**Figure 17, section 3.3.3**). This was used to convert absorbance values of biotransformation reactions into conversion values, which were plotted against time of incubation to show the progress of the reactions over time (**Figure 23**).



**Figure 23** - A standard graph, using values from figure X, that can be used to convert absorbance values directly into % conversion.

In parallel to the sampling of biotransformation reactions for the colorimetric screen, the remaining reaction contents were immediately basified by the addition of NaOH (50  $\mu$ L, 10 N) to stop the enzymatic reaction, and extracted with EtOAc (0.8 mL). After mixing and centrifuging, the organic layer was transferred to an autosampler vial, and derivatised for GC-FID analysis by the addition of 15  $\mu$ L acetic anhydride and 15  $\mu$ L triethylamine.

GC analysis using a chiral column was performed on a ThermoFisher 1310 chromatograph using the following temperature program: 40  $^{\circ}$ C hold for 0.5 min followed by a 50  $^{\circ}$ C. $\text{min}^{-1}$  ramp to 150  $^{\circ}$ C, a hold at this temperature for 5 min, followed by a 50  $^{\circ}$ C. $\text{min}^{-1}$  ramp to 200  $^{\circ}$ C and a further hold for 3 min.



**Figure 24** - GC-FID traces from Donor Screen biotransformations carried out for varying lengths of time, showing the progressive conversion of (S)-MBA to acetophenone.

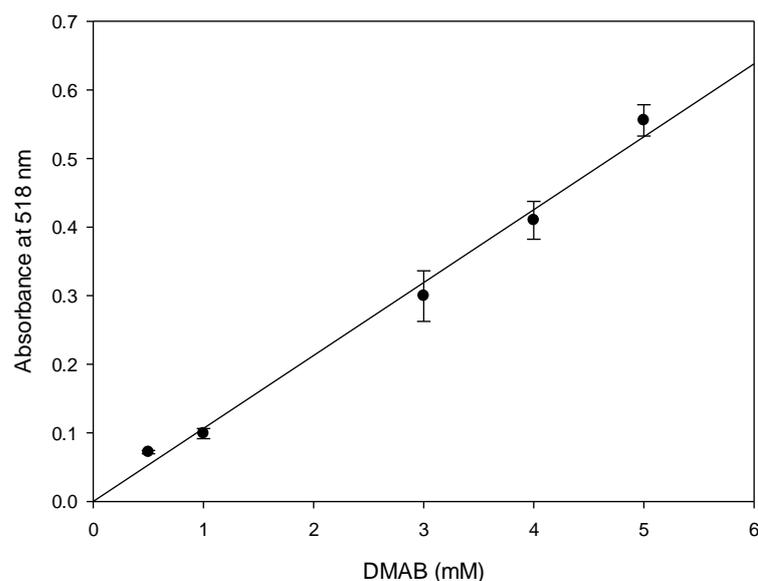
## Enantiomeric screen

AspOry, HypNep, MycVan, 3FCR, HEWT and CV (50  $\mu$ L cell free lysate) were diluted in KPi buffer (1 mL, 100 mM, pH 8) containing PLP (1 mM), with either (*S*)-(-)- $\alpha$ -methylbenzylamine or (*R*)-(+)- $\alpha$ -methylbenzylamine (5 mM from a 500 mM stock in DMSO). To this was added *p*-dimethylaminobenzaldehyde (5 mM from a 500 mM stock in DMSO). The reaction mixture was incubated at 30  $^{\circ}$ C, 200 rpm in a shaking incubator. The biotransformations were left for 24 hours.

The colorimetric screen was performed by diluting the biotransformations (40  $\mu$ L) in KPi buffer (100 mM, pH 8) containing PLP (1 mM) (160  $\mu$ L). 50  $\mu$ L of the resulting solution to a well containing 50  $\mu$ L of a solution of indole (50 mM) in DMSO and mixed thoroughly. To this was added 100  $\mu$ L of a mixture of DMSO:conc.HCl (1:1) and mixed thoroughly. The colour was developed for 8 minutes under white light whilst rotating the plate 90 $^{\circ}$  2 two minutes to ensure even light coverage. A spectral scan of the plate was then performed on an EPOCH2 microplate reader from 500-530 nm on each well. The maximum absorption was recorded at 518 nm.

Standards were prepared in analogous manner to the biotransformations with the absence of enzyme and a gradient of DMAB/(*S*)-(-)- $\alpha$ -methylbenzylamine from 5 mM to 0 mM to create a linear standard curve (**Figure 25**).

The conversions were measured by GC-FID, the reactions were basified (pH 13), extracted with EtOAc (750  $\mu$ L), derivatised with 15  $\mu$ L triethylamine and 15  $\mu$ L acetic anhydride and analysed by GC-FID.



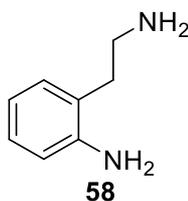
**Figure 25** - Plot of DMAB standard (mM) against absorbance at 518 nm showing a clear linear relationship ( $R^2=0.99$ ).

### Solid phase screen

50  $\mu$ L of *E. coli* BL21(DE3)  $\Delta$ *tnaA* or *E. coli* BL21(DE3) electro-competent cells were transformed with 250 ng of 3HMU pET22b plasmid. 150  $\mu$ L of the transformed cells diluted 1:10 with LB media were spread onto LB agar plates (100  $\mu$ g.mL<sup>-1</sup> ampicillin) and incubated o/n at 37 °C. A single colony was used to inoculate LB (5 mL, 100  $\mu$ g.mL<sup>-1</sup> ampicillin) and incubated for 5 h at 37 °C and 200 rpm. The cells were harvested by centrifugation at 4000 rpm for 15 min and the cell pellets were resuspended in LB/glycerol/d.H<sub>2</sub>O (5:3:2) (1 mL) from which 50  $\mu$ L cell stocks were prepared and frozen at -80 °C. The cell stocks were diluted 2000-fold in LB and 150  $\mu$ L was spread onto a nitrocellulose membrane (Protran BA85, GE Healthcare) overlaid on LB agar containing ampicillin (100  $\mu$ g.mL<sup>-1</sup>) and incubated o/n at 37 °C. Protein expression was then induced by transferring the nitrocellulose membrane onto an LB-ampicillin plate containing 1 mM IPTG and incubated for 8 h at 30 °C. The membrane was subsequently transferred to a dialysis plate (filter paper

soaked with 10 mM Tris.HCl, pH 8, and 0.1 mM PLP) o/n at 4 °C. Background depletion was achieved by transferring the membrane to a soaked filter paper (KPi buffer 50 mM, pH 8, (S)-(-)-1-phenylethylamine 10 mM) for 1 h at room temperature.<sup>197</sup> The membrane was then transferred to a second soaked filter paper disc (phosphate buffer 50 mM, pH 8) and subsequently stored o/n at -20 °C. The screening step was performed by placing the membranes on soaked filter paper (phosphate buffer 50 mM, pH 8, 2-(2-aminoethyl)aniline 11 mM, pyruvate 10 mM) and left at room temperature for up to 1 hour. The colour was developed by transferring the membranes to another soaked filter paper (*p*-dimethylaminocinnamaldehyde 57 mM, conc.HCl:H<sub>2</sub>O/1:9).

### Synthesis of 2-(2-aminoethyl)aniline (2-AEA) 58

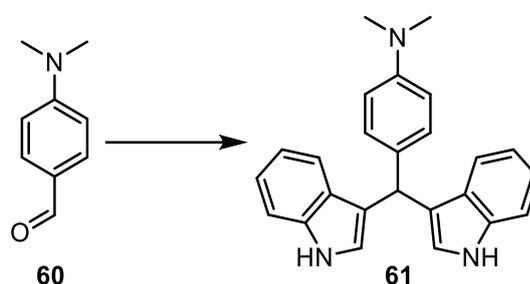


To a solution of 2-cyanomethylnitrobenzene (1.5 g, 9.26 mmol) in anhydrous THF (25 mL) under N<sub>2</sub> was added BH<sub>3</sub>.DMS solution (2M in toluene, 10.6 mL, 21.3 mmol) dropwise at 0° C. After addition, the mixture was stirred at 60 °C for 18 h. The solution was cooled to 0 °C and quenched with CH<sub>3</sub>OH (15 mL). The solvent was removed *in vacuo* and the residue purified by silica gel chromatography, eluting in CH<sub>2</sub>Cl<sub>2</sub>:CH<sub>3</sub>OH, affording a brown oil (685 mg). The oil was dissolved in EtOH (4 mL) and 10% Pd/C (59 mg, 10 wt. %) was added. The mixture was stirred under an atmosphere of H<sub>2</sub> for 18 h at room temperature. The solution was diluted with CH<sub>3</sub>OH (30 mL), filtered and concentrated under reduced pressure. The resulting residue was filtered

through Celite® and the solvent was removed *in vacuo* to afford 2-(2-aminoethyl)aniline (626 mg, 49.7%) as a brown oil.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.10 – 7.01 (m, 2H), 6.79 – 6.68 (m, 2H), 3.02 (t, *J* = 6.5 Hz, 2H), 2.70 (t, *J* = 6.6 Hz, 2H). **<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 145, 130, 127, 125, 119, 116, 42, 35. **HRMS** (EI) *m/z*: Calculated C<sub>8</sub>H<sub>13</sub>N<sub>2</sub> [M+H] 137.1073; found 137.1072.

### Synthesis of 4-(di(1*H*-indol-3-yl)methyl)-*N,N*-dimethylaniline 61

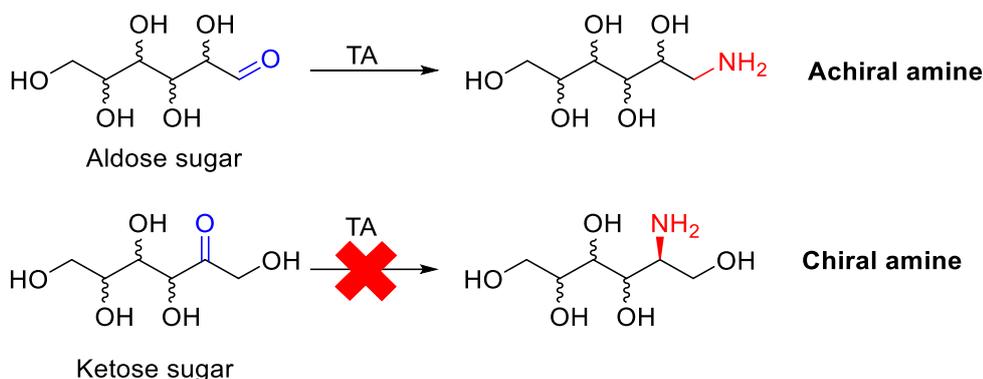


Indole (1.17 g, 10 mmol) was added to a solution of 4-dimethylaminobenzaldehyde (745 mg, 5 mmol) in THF (50 mL). To this conc. HCl (1 mL) was added. This suspension was stirred at room temperature for one hour and subsequently filtered and the solid washed with H<sub>2</sub>O (20 mL) and acetone (20 mL) and dried *in vacuo* to yield the product as a pink powder (1.36 g, 74% yield). **m.p** 207-211 °C **<sup>1</sup>H NMR** (400 MHz, DMSO-*d*<sub>6</sub>) δ 10.93 (s, 1H), 10.93 (s, 1H), 7.62 – 7.54 (m, 4H), 7.57 – 7.48 (m, 2H), 7.42 – 7.29 (m, 2H), 7.34 – 7.26 (m, 2H), 7.11 – 6.98 (m, 2H), 6.94 – 6.83 (m, 4H), 5.91 (s, 1H), 3.07 (s, 6H). **<sup>13</sup>C NMR** (101 MHz, DMSO-*d*<sub>6</sub>) δ 142.3, 137.0, 130.1, 126.9, 124.0, 121.3, 120.5, 120.2, 119.5, 118.6, 117.9, 112.2, 45.5, 39.7. **HRMS** (EI) *m/z*: Calculated C<sub>25</sub>H<sub>24</sub>N<sub>3</sub> [M+H] 366.1892; found 366.1891. **IR (ATR)** 3246, 2531, 2440, 1710, 1658, 1415, 1311 cm<sup>-1</sup>

## 4.0 Molecular modelling of monosaccharides in $\omega$ -ATAs and the semi-rational engineering of 3HMU towards ketose activity

### 4.1 Introduction

Chapter 2 describes the development of a protocol for the direct amination of monosaccharides by ATAs, which provided access to a panel of amino alcohols. However, the sugars identified as substrates for these ATAs were aldoses (aldehydes), rather than ketoses (ketones), and therefore the generated amine was achiral in each case (**Scheme 22**). It has been estimated that at least 40% of pharmaceuticals contain at least one chiral amine and their use as chiral auxiliaries, or advanced intermediates constitutes 15 % of the market for chemicals used in the life science sector alone.<sup>210</sup> The generation of aminopolyols from ketoses would clearly be preferable for industrial uses and lead to the greater valorisation of this sustainable feedstock.



**Scheme 22** - Schematic showing the transamination of an aldose sugar to generate an achiral amine, as demonstrated in chapter 2. Also shown is the transamination of a ketose sugar towards a chiral amine, which was not possible using the ATAs explored in chapter 2.

Work by Surbrizi *et al* published shortly after our work (Chapter 2) demonstrated the amination of both aldoses and ketoses using (*R*)-selective ATAs, with up to 45% conversion to product recorded with D-fructose.<sup>211</sup> Within

in our group, only (*S*)-selective ATAs were identified as active on aldose sugars. The sole use of *o*-xylylenediamine to screen for activity likely hindered the identification of (*R*)-selective ATAs as it is commonly a poor donor for these enzymes. However, this work by Subrizi *et al* clearly demonstrates that this enzyme class can perform the desired amination of ketose sugars to chiral aminopolyols. Furthermore, the (*S*)-selective ATAs we identified are active on both aldehydes and ketones, therefore, we hypothesised their inability to convert ketoses must be down to sterics/ligand-residue interactions in the active site.<sup>180,184</sup> An obvious explanation could be the addition of the large alcohol group into the small pocket when using ketoses. The limitation of many ATAs to a small side chain no larger than an ethyl group is known and several groups have demonstrated it is possible to engineer the active site to accommodate bulky substituents.<sup>212–218</sup>

Computational methods to aid engineering can be grouped into three broad categories: molecular modelling, bioinformatics, and *de novo* design. There is a recent trend to combine these approaches in a complementary manner for structural analysis and directed evolution.<sup>219</sup> Protein structures provide valuable information about the molecular basis of their function, which can aid rational redesign. Methods such as, nuclear magnetic resonance spectroscopy (NMR), cryo-electron microscopy (cryo-EM) and X-ray crystallography can precisely determine the structure of proteins and associated complexes.<sup>220–223</sup> These structures are then commonly uploaded to the protein data bank (PDB) as open-source files. The availability of such structures allows for increasingly accurate *in silico* modelling where sequence/function relationships can be determined to assist with rational/semi-rational design.<sup>219</sup> The protein/ligand

complex, often in the presence of cofactor, is commonly used for such molecular modelling. Several methods for docking/screening have been released, such as: *AutoDock*, *AutoDock Vina*,<sup>224</sup> *Glide*,<sup>225</sup> *DOCK*,<sup>226</sup> *RosettaDock*<sup>227</sup> and *Surflex*.<sup>228,229</sup>

These protocols use a combination of a search algorithm and a scoring function to give the resulting docking assessment.<sup>230</sup> The scoring function will represent the thermodynamics of the protein/ligand interaction system to adequately distinguish the most probable binding modes from all possibilities.<sup>231</sup> Many approximations are made here to significantly reduce the level of computation required to make these calculations feasible. This can reduce the accuracy of the docking assessment. In fact, the absence of a scoring function that can be both highly accurate and rapid is the major bottleneck in such protocols.<sup>232</sup> The search algorithm aims to give enough degrees of freedom to the protein/ligand system to sample the space including the true binding modes.<sup>233,234</sup> Its success in achieving this is measured in terms of root-mean-square deviation (RMSD) between the experimentally observed heavy-atom positions of the ligand and those predicted by the algorithm.

*AutoDock* is a successful protein/ligand docking program developed in the 1990's with subsequent updates being released thereafter.<sup>235</sup> It has assisted the development of several drug molecules.<sup>236,237</sup> Furthermore, it has been described as the most cited docking program.<sup>238</sup> *AutoDock Vina* was developed as an update to improve accuracy and performance.<sup>224,239</sup> It can achieve an increase in speed of two orders of magnitude when compared to *AutoDock*, whilst simultaneously improving the binding mode prediction accuracy. *AutoDock Vina* is a much more user friendly and intuitive interface

as it automatically calculates and displays grid maps/clusters.<sup>239</sup> After docking it is then up to the user to conclude what binding mode is best based on known structural data and other relevant information (e.g. enzymatic mechanism). If a suitable binding mode is not found, then the parameters must be changed, and the experiment repeated until a good fit is found.

This docking protocol is commonly followed by molecular dynamics simulation to obtain the lowest local energy minima of a given binding mode. These simulations work by moving each atom randomly in the field of all others and to represent protein/ligand flexibility more accurately.<sup>240</sup> This is normally achieved by introducing random changes to several parameters, such as van der Waals radius or atomic charge. Minimizations are performed to ensure all forces responsible for the docked structure are fully considered. Therefore, a stable energy minima is obtained that can be used for protein design. Multiple software packages exist to perform these simulations, such as ROSETTA,<sup>241</sup> ORCA,<sup>242</sup> PREFMD,<sup>243</sup> and YASARA.<sup>244</sup>

## **4.2 Aims and objectives**

The work presented in this chapter aimed to engineer an (S)-selective ATA to accept ketose activity, specifically D-fructose. This has been attempted by modelling D-deoxyribose into the active site of 3HMU, HEWT and 3I5T. The first two of which are known to accept the monosaccharide as a substrate and the last of which has been shown by the O'Reilly group to be inactive towards sugars. This is followed by modelling D-fructose, a ketose sugar, in these enzymes and comparing these docking results to identify hot spots in the active site for rational design. Single site and combinatorial mutagenesis have been

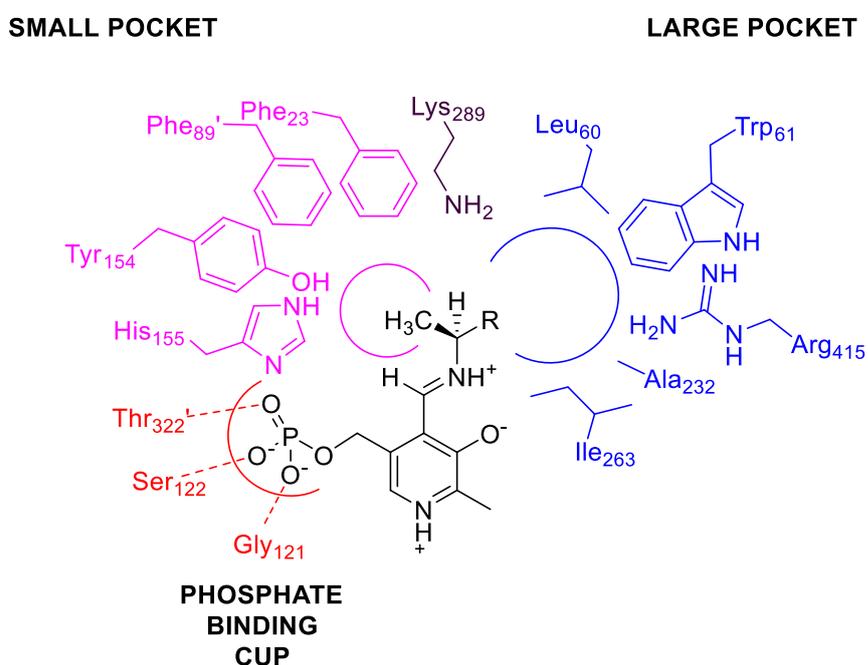
trials on 3HMU to increase its activity towards D-fructose and the results are shown herein.

### 4.3.0 Results and discussion

#### 4.3.1 Molecular modelling of monosaccharides in $\omega$ -ATAs

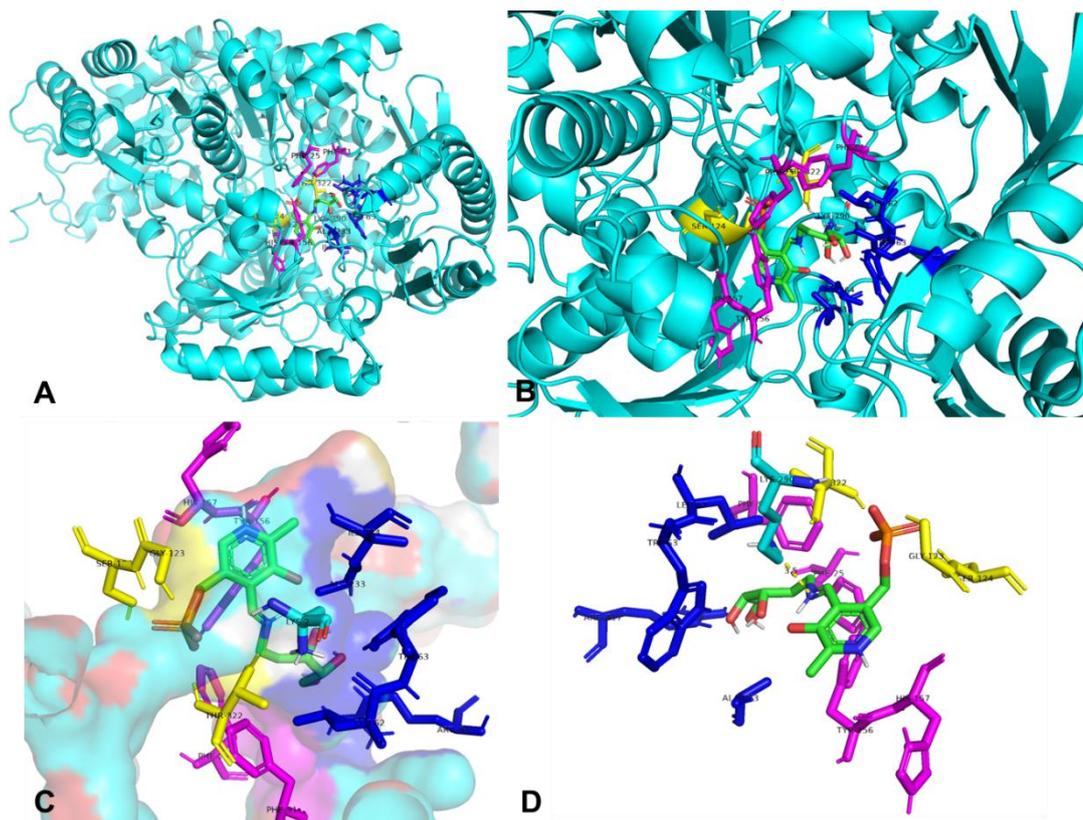
Interactive PyMOL sessions for all models shown are available at a link found in appendix 21.

The configuration of active site residues surrounding the PLP-substrate adduct has been comprehensively understood and described for *V. fuvialis* (PDB: 4E3Q) (**Figure 26**).<sup>245</sup> Many of these residues are highly conserved amongst  $\omega$ -ATAs and several groups have used this model, alongside sequence alignment, to aid with protein engineering/characterisation.<sup>246,247</sup> Therefore, the molecular modelling herein is based on the alignment of 3HMU, HEWT and 3I5T with this model.



**Figure 26** – The active site of *V. fuvialis* (PDB: 4E3Q) containing a PLP-substrate adduct. The small pocket, large pocket, and phosphate binding cup residues are shown.

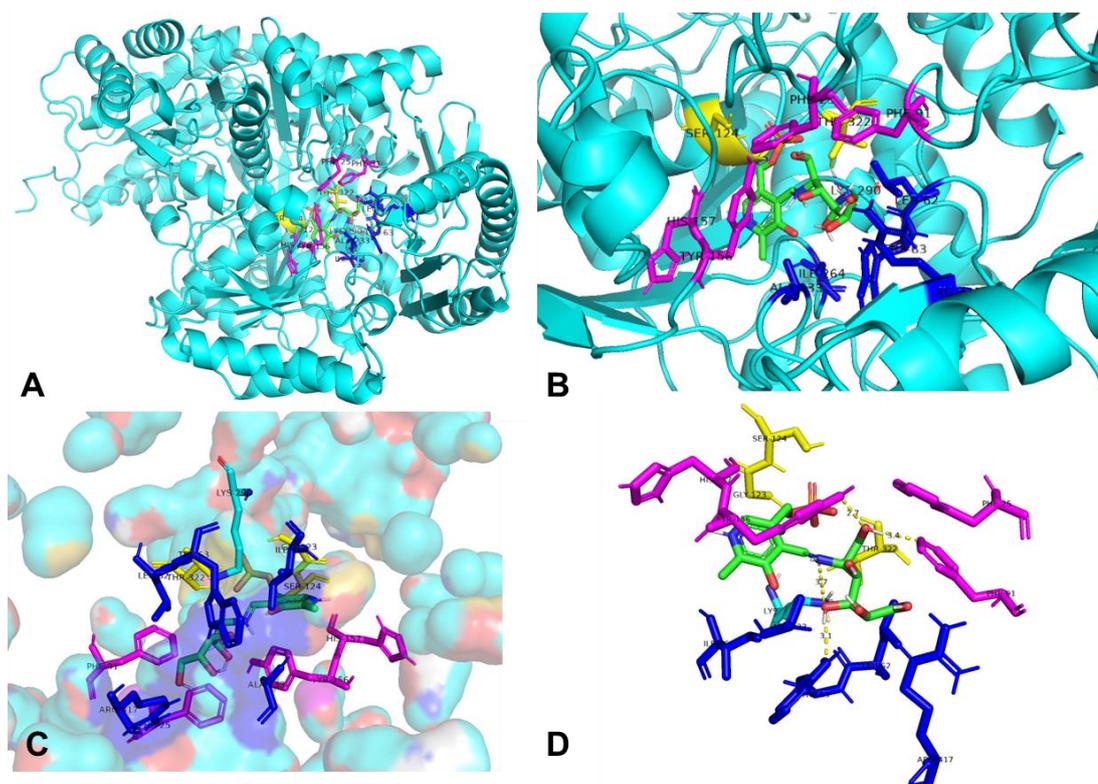
Modelling of D-deoxyribose in the active site of 3HMU revealed a PLP-substrate adduct configuration closely aligned to that predicted from the literature (**Figure 27**). Where there are no obvious clashes between large pocket residues and the polyol chain, and furthermore, the catalytic lysine is oriented towards the imine of the external aldimine at a distance of 3.5 Å. This proximity is essential as the next step in the catalytic cycle involves the nucleophilic attack of this amine onto the imine, binding the enzyme to the adduct and eventually causing dissociation of the amino product and the regeneration of PLP. As this is an aldose sugar, only a hydrogen lies in the small pocket, so no negative clashes are predicted. Therefore, D-deoxyribose could be expected from this model to be a substrate for 3HMU, as confirmed by experiment (chapter 2).



**Figure 27** - Computational model of D-deoxyribose in the active site of 3HMU. A) Cartoon of the whole protein with substrate docked in the active site. B) Close-up cartoon of the substrate within the active site. C) Illustration of the protein surface, with the substrate within the open pocket of the active site. D) The docked substrate with key active site residues highlighted. Key: Large pocket residues are highlighted in dark blue, small pocket residues in magenta, phosphate binding cup residues in yellow and the catalytic lysine in cyan. See appendix 1-4 for larger pictures.

The subsequent analogous modelling of D-fructose revealed a hypothesis for the enzyme's inactivity towards this and other ketose substrates (**Figure 28**). The introduction of a hydroxy group into the small pocket creates the possibility of several negative interactions between the substrate and the enzyme. In the model, this OH group is orientated towards the OH group of Tyr156 at 2.7 Å, suggesting hydrogen bonding between the substrate and enzyme. However, the proximity between this group and the aromatic system on Tyr would introduce strong repulsive forces. A less bulky, polar residue could be beneficial here. Furthermore, this OH group is located 3 Å from the non-polar F91, and as alcohol groups are polar/non-favourable interactions would arise between this and the non-polar phenyl residue. In the large pocket, another

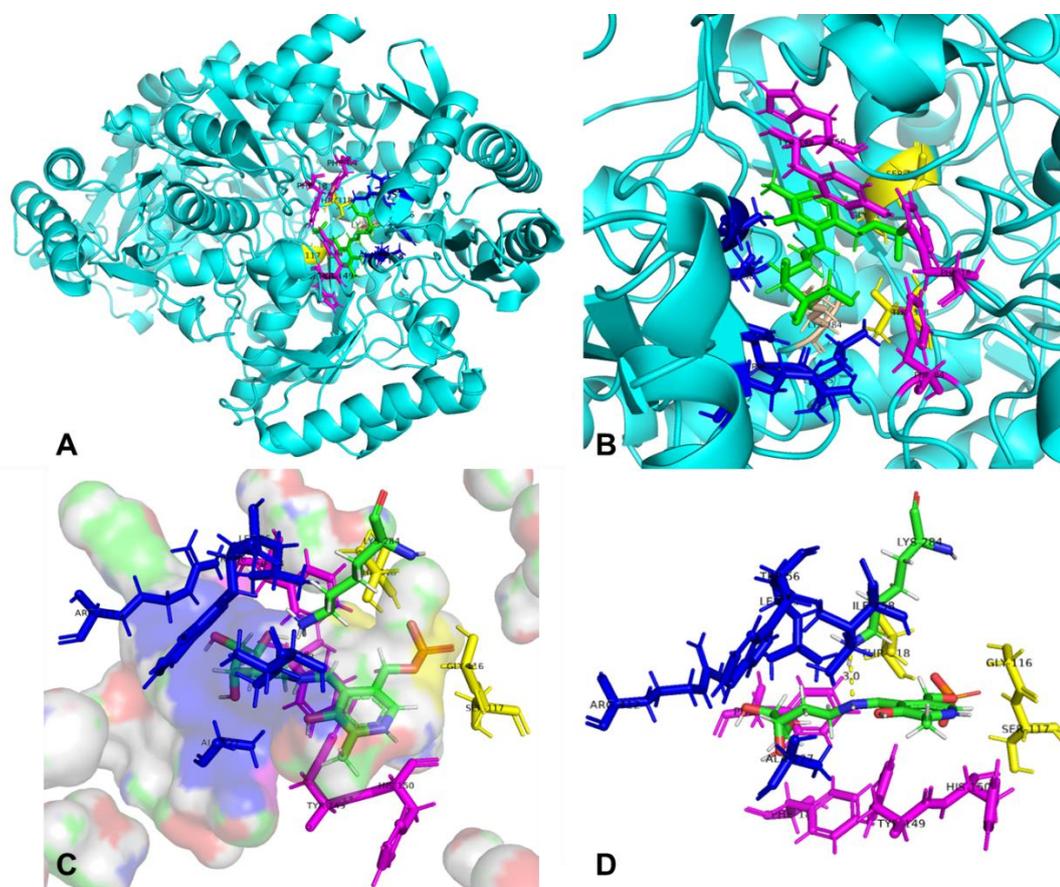
OH group is situated  $<3 \text{ \AA}$  from Trp63, another non-polar residue. Additional negative interactions can be expected here. It is also noted that there is significant unoccupied space in the large pocket between Leu62, Ile264 and the substrate where mutagenesis could introduce bulkier and preferably polar residues to increase the number of favourable interactions. The similar proximity and orientation of the catalytic lysine residue towards the imine, compared to D-deoxyribose, suggests activity can be achieved by removal of negative interactions between the substrate and enzyme.



**Figure 28** - Computational model of D-fructose in the active site of 3HMU. A) Cartoon of the whole protein with substrate docked in the active site. B) Close-up cartoon of the substrate within the active site. C) Illustration of the protein surface, with the substrate within the open pocket of the active site. D) The docked substrate with key active site residues highlighted. Key: Large pocket residues are highlighted in dark blue, small pocket residues in magenta, phosphate binding cup residues in yellow and the catalytic lysine in cyan. See appendix 5-8 for larger images.

Modelling D-deoxyribose in the active site of HEWT revealed an analogous model to that seen with 3HMU (**Figure 29**), with the catalytic lysine residue situated  $3 \text{ \AA}$  from the imine. Again, as there is only a hydrogen situated in the

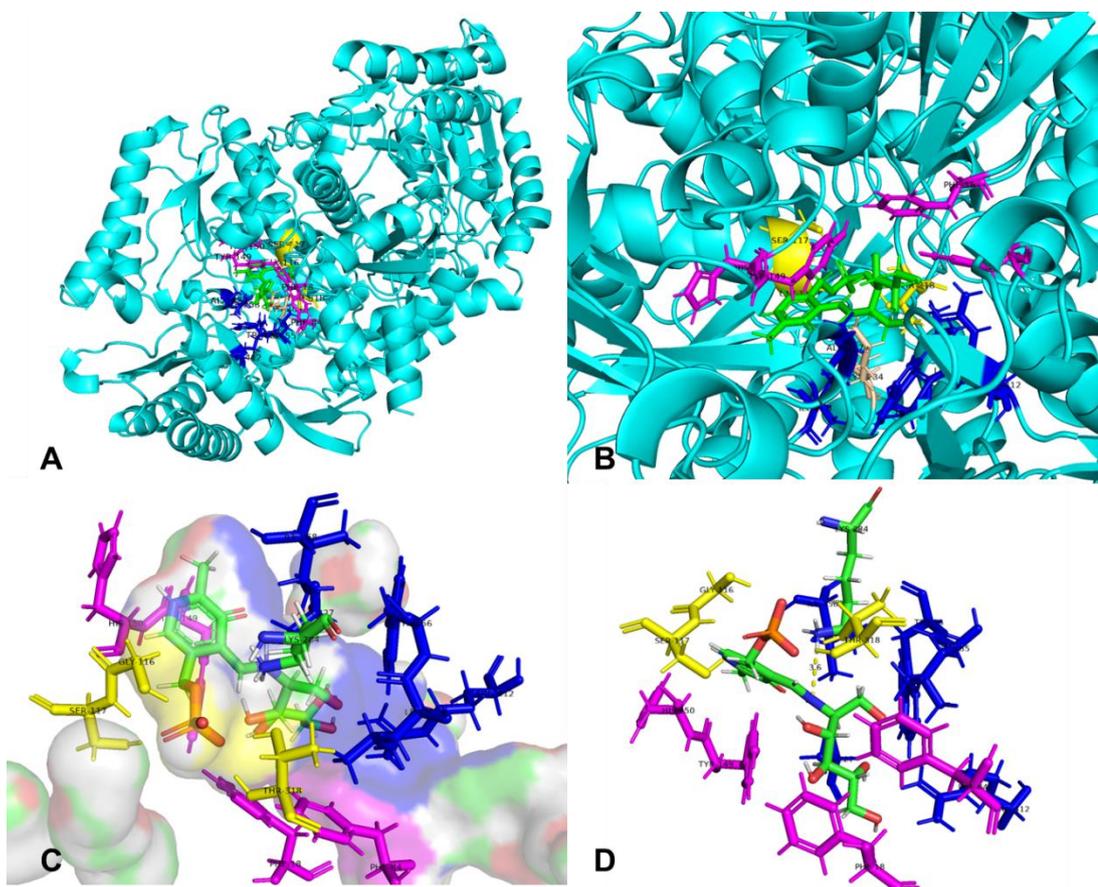
small pocket no negative interactions are expected here. Therefore, it is not surprising that this substrate is very well accepted by the enzyme.



**Figure 29** – Computational model of D-deoxyribose in the active site of HEWT. A) Cartoon of the whole protein with substrate docked in the active site. B) Close-up cartoon of the substrate within the active site. C) Illustration of the protein surface, with the substrate within the open pocket of the active site. D) The docked substrate with key active site residues highlighted. Key: Large pocket residues are highlighted in dark blue, small pocket residues in magenta, phosphate binding cup residues in yellow and the catalytic lysine in green. See appendix 9-12 for larger images.

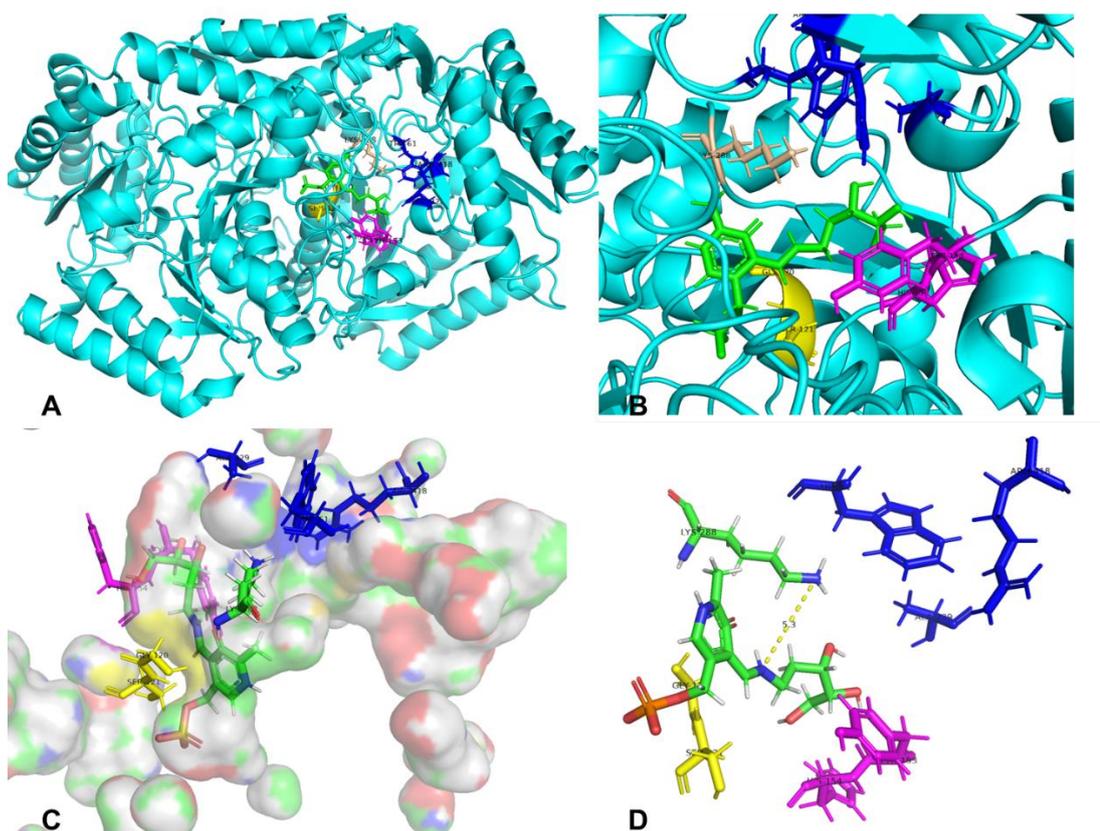
Further modelling of D-fructose in the active site of HEWT suggests a similar introduction of negative interactions for ketose sugars as seen with 3HMU (**Figure 30**). Here, the Tyr149 residue is 3 Å from the OH group in the small pocket. This polar OH group is also situated 3.3 Å from the non-polar Phe18, introducing non-favourable interactions in the small pocket, as seen in 3HMU. Therefore, increasing the size of the small pocket in HEWT, alongside the introduction of polar residues could increase the number of favourable

substrate-enzyme interactions. In the large pocket, further negative interactions are predicted between the OH group on C-3 and the non-polar residues Leu55 (3.3 Å) and Trp56 (3.0 Å), suggesting mutagenesis at these positions could be beneficial. As in 3HMU, significant space is unoccupied around Ile251 in this model. The introduction of a bulkier, polar region here has the potential to increase the number of positive interactions. Again, the close proximity (3.0 Å) of the catalytic lysine to the imine suggests catalysis is possible with this substrate if more favourable substrate-active site interactions are introduced.



**Figure 30** -Computational model of *D*-fructose in the active site of HEWT. A) Cartoon of the whole protein with substrate docked in the active site. B) Close-up cartoon of the substrate within the active site. C) Illustration of the protein surface, with the substrate within the open pocket of the active site. D) The docked substrate with key active site residues highlighted. Key: Large pocket residues are highlighted in dark blue, small pocket residues in magenta, phosphate binding cup residues in yellow and the catalytic lysine in green. See appendix 13-16 for larger images.

D-Deoxy-ribose was also modelled into the active site of 3I5T (**Figure 31**). This was chosen as it was shown not to accept aldose sugars as substrates. Therefore, this model could demonstrate that certain active sites do not accommodate this bulky/polar substrate class. Further, the variant could prove a useful template for hypothesising the origins of activity towards aldoses. Manual docking of the substrate in the active site made it clear that the volume of the open pocket in 3I5T is significantly smaller than the previous enzyme. It was not possible to fit the entire adduct into the active site without side chains coming into contact with the surface of the protein. Computation revealed a docked adduct that was significantly distorted away from that predicted by literature. Here, the substrate is largely found in the small pocket where many steric clashes are predicted. Several more clashes are found in the phosphate binding cup between the phosphate group and Ser121, as this severe distortion reduces their separation to 2.5 Å. Furthermore, the catalytic lysine residue is found 5.5 Å away from the imine making nucleophilic attack unlikely. This suggests the volume of the active site is key for acceptance of these substrates due to the bulky nature of the OH groups.



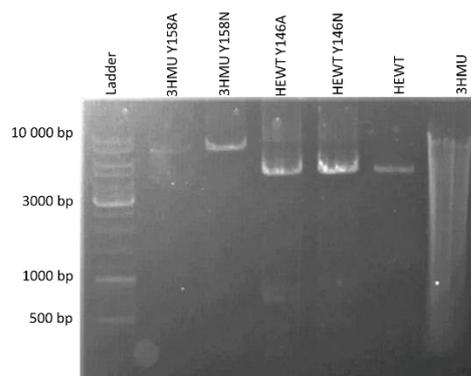
**Figure 31** – Computational model of *D*-deoxyribose in the active site of 3I5T. A) Cartoon of the whole protein with substrate docked in the active site. B) Close-up cartoon of the substrate within the active site. C) Illustration of the protein surface, with the substrate within the open pocket of the active site. D) The docked substrate with key active site residues highlighted. Key: Large pocket residues are highlighted in dark blue, small pocket residues in magenta, phosphate binding cup residues in yellow and the catalytic lysine in green. See appendix 17-20 for larger images.

Generally, it appears ketose activity could be introduced to these ATAs by increasing the volume of the small pocket and establishing polar residues within the active site. Park *et al.* describe in detail the molecular determination for the commonly observed substrate specificity of  $\omega$ -ATAs.<sup>248</sup> They identify 6 key residues based on the modelling of an ATA from *Pseudomonas putida*; Tyr23, Trp60, Phe88', Tyr156, Ile262 and Arg414. All six of these residues are highly conserved amongst studied  $\omega$ -ATAs. Therefore, engineering efforts focused on these residues could significantly alter substrate selectivity. Indeed, the small pocket Phe and Tyr residues, along with the large pocket Ile and Trp residues, have been independently highlighted in these docking

experiments as potential sites of negative adduct-enzyme interactions with D-fructose.

#### **4.3.2 Site directed mutagenesis of 3HMU and HEWT to increase the size of the small pocket**

Initial engineering efforts focused on increasing the size of the small pocket to accommodate the bulky alcohol group on the side chain of ketose sugars. The tyrosine residue of this pocket was targeted, as this was the bulkiest residue observed from the computational studies. It also appeared that this is the most likely cause of negative clashes between the bulky side chain and the enzyme. Therefore, rational mutations were added to this position in both 3HMU and HEWT. Alanine was selected to replace the tyrosine residue as it is the smallest chiral amino acid and therefore would create the largest space in the small pocket, whilst maintaining chirality. Asparagine was also chosen as it is a polar residue like tyrosine, but, considerably less bulky. This polarity should allow for favourable interactions with the polar alcohol group of D-fructose, whilst also increasing space in the small pocket. Primers were designed to introduce these mutations *via* PCR and an agarose gel was performed on the PCR product subsequent to DPN1 digestion to confirm a successful reaction (**Figure 32**).

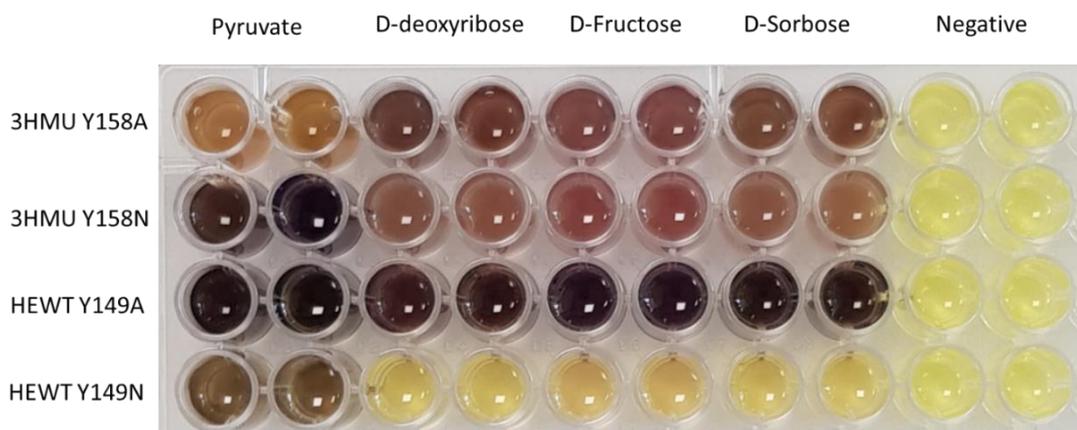


**Figure 32** – Agarose gel of the PCR products for the rational mutations of Tyr156 on 3HMU and HEWT. Controls are wild-type DNA.

The mutations were confirmed by sequencing and the variants were then tested for activity with pyruvate and deoxyribose (known substrates of the WT) and two ketose sugars, D-fructose and D-sorbose (**Figure 33**). In the three variants 3HMU Y158N/Y158A and HEWT Y149N, none of the wells turned black suggesting these mutations substantially reduce the activity of the protein. This was evident when using pyruvate as the acceptor, where conversions >99% can be expected with the WT. This is a highly sensitive screen, and, on this scale, wells can be expected to turn a deep black with once 4-5% conversion is reached. Therefore, dark brown/purple wells suggest conversion lower than this.

In the case of HEWT Y149N, activity towards pyruvate appears to be maintained, with the wells turning black. However, activity towards D-deoxyribose is reduced with the wells only turning brown whereas 66% conversion to product is observed with the wild-type. Interestingly, the wells also turn purple/brown for D-Fructose and D-sorbose suggesting very low levels of conversion to amino alcohol product. Quantitative NMR studies (see chapter 2) were used to detect the presence of product, with ~3% amino alcohol observed for D-deoxyribose. Unfortunately, no evidence of product was

observed for the ketose sugars, suggesting that conversion to product was minimal. However, this amino acid position appeared to be promising for further mutation as these results indicate amino acid changes to this position can affect the selectivity towards this substrate class.



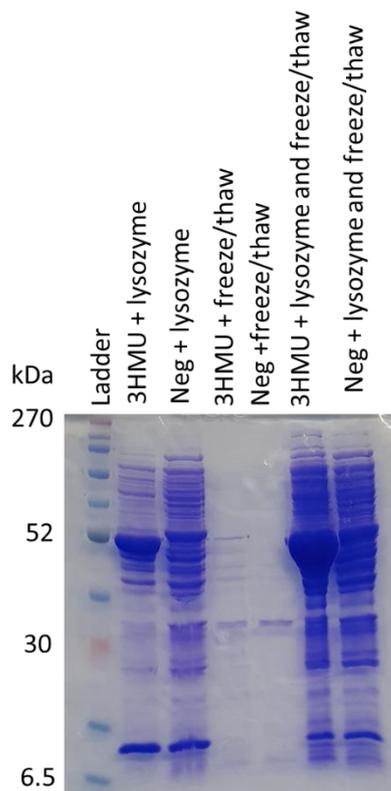
**Figure 33** – Screening for mutant activity towards ketoses. 11 mM *o*-xylylenediamine, 10 mM acceptor, KPi buffer 100 mM, PLP 1 mM, 25% lysate, 30C, 24 h.

Deszcz *et al*, describe this residue as being highly conserved amongst  $\omega$ -ATAs, and it appears from computational modelling that the aromatic system could be involved in  $\pi$ - $\pi$  stacking with the PLP/PMP co-factor.<sup>249</sup> This would suggest mutation at this position would be detrimental to activity. Nevertheless, they demonstrated that single point mutations at this position can change the substrate specificity. This was achieved using the (*S*)-selective enzyme *Chromobacterium violaceum* and creating an NNK variant library at this position. Up to 4.5x increase in activity towards serine was recorded. Therefore, it is clear this site is important for substrate specificity and further work focused on saturating this position to test a wider variety of amino acid environments.

### 4.3.3 Optimisation of protein isolation from 48-well plate

The amine acceptor screen developed in Chapter 3 was to be used to screen for mutant activity with 3HMU. This screen was chosen as it is quantitative and would allow for the selection of the best variants during each round of evolution, and additionally, its use with 3HMU had already been demonstrated. The over expression of 3HMU in pET-22b(+) is already well established. However, the common lysis method involves sonication, which is not practically feasible when using deep well plates to screen >100 variants at once. Therefore, alternative methods were trialled in deep well plates. These included using lysozyme from hen eggs, an overnight freeze/thaw cycle of the cells suspended in buffer and finally, the application of lysozyme followed by an overnight freeze/thaw cycle.

SDS-page analysis (**Figure 34**) demonstrated that the cells remained mostly intact upon freezing and after the use of lysozyme many of the cells became disrupted as protein is clearly visible in the lysis buffer. The best results were observed when combining the lysozyme with the freeze-thaw cycle. As *E. coli* is a gram-negative bacterium the lysozyme disintegrates only the inner membrane, leaving the outer membrane intact, so without disruption the proteins can remain largely within the cell.<sup>250</sup> The penetration of ice crystals appears to be enough to satisfactorily disrupt the outer membrane and release large quantities of protein. The activity of the WT enzyme was subsequently assessed and compared to that of protein lysed by sonication.

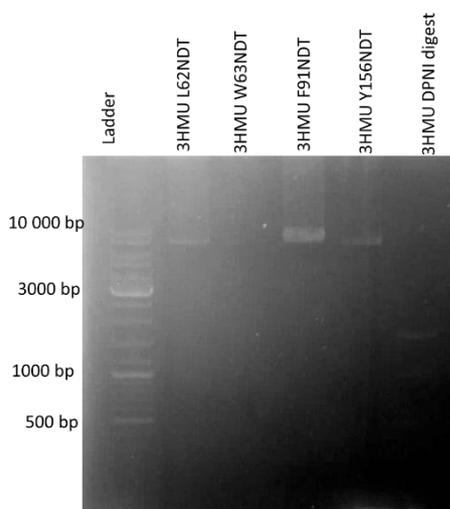


*Figure 34 – SDS-page showing protein in the lysate after several lysis methods. Lysozyme only, overnight freeze/thaw and lysozyme plus freeze thaw. See experimental for conditions.*

#### 4.3.4 Combinatorial mutagenesis of 3HMU

Four active site residues in 3HMU were identified for mutagenesis after computational analysis for combinatorial mutagenesis, these were L62, W63, F91 and Y156. Forward and reverse primers were designed that were complementary to the wild type. To reduce the number of mutants that would need to be screened during each round an NDT codon was placed at each position, as this codon only codes for 12 amino acids. Importantly, it still allows for a wide variety of polar/non-polar/aromatic/aliphatic residues to be screened whilst substantially reducing the size of the library. PCR was performed using Phusion® high-fidelity polymerase and the reaction was checked by agarose gel after DPN1 digestion (**Figure 35**). The libraries were plated, and several mutants were selected and sent for sequencing, however, only ~1/8 colonies had the desired mutations. The rest were WT DNA. The presence of WT DNA

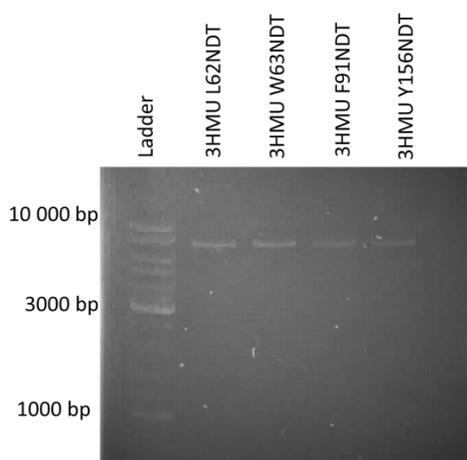
is often due to the presence of undigested methylated DNA, however, a control DPNI digest of WT 3HMU plasmid shows the DNA is entirely digested during this protocol. Furthermore, transformation of digested plasmid into *E. coli* DH5 $\alpha$  revealed no colonies, indicating full digestion of the DNA. It is not feasible to screen for novel activity with such an inefficient protocol as the number of colonies that need to be screened to sample 95% of the available amino acids becomes too large. This protocol was attempted several times and the results remained the same, with most of the colonies containing WT DNA.



**Figure 35** - Agarose gel of the PCR products of L62NDT, W63NDT, F91NDT and Y156NDT. 3HMU WT after DPNI digestion is shown to demonstrate its complete digestion.

Site-directed, Ligase-Independent Mutagenesis (SLIM) is a PCR-mediated mutagenesis methodology developed by Chiu *et al.*<sup>251</sup> This protocol involves a PCR amplification of the template by two tailed long primers and two short primers in one step. The long primers comprise the mutation on complementary overhangs. PCR products are denatured and reannealed, leading to heteroduplex formation between the mixed PCR products to give the preferred mutated plasmid. The authors describe this methodology as

being highly robust and suitable for use in high-throughput engineering/library constructions, the efficiency of this protocol at obtaining the desired mutation is described as 95%. The presence of PCR product was again confirmed by agarose gel (**Figure 36**), however, sequencing revealed only 20% of the 'mutant' colonies contained the desired mutation. This was a slight improvement compared to the previous protocol, but, still too inefficient for this screening methodology. Several problems could give rise to this scenario, including poor temperature design or sub-optimal reaction conditions (temperature). Due to time constraints this could not be investigated further.



**Figure 36** - Agarose gel of the SLIM PCR products of L62NDT, W63NDT, F91NDT and Y156NDT.

#### **4.4 Conclusions and future work**

This chapter outlines the computational modelling of 3 ATAs; 3HMU, HEWT and 3I5T with D-deoxyribose and D-fructose. This led to the hypothesis that ketose sugars were not substrates of these (S)-selective enzymes due to clashes with active site residues that have previously been described as important for determining substrate specificity, particularly those in the small pocket. Four active residues of 3HMU were identified for rational/semi-rational engineering towards D-fructose activity, these were L62, W63, F91 and Y156.

In the first instance, the residue Y156 was rationally engineered to increase the size of the small pocket by replacing the Tyr residue with both Ala and Asn. This was done in both 3HMU and HEWT. Evidently this residue is essential for determining specificity as activity towards the controls was significantly reduced. However, it appears the mutation Y149A in HEWT allows for the conversion to trace amounts of the amino polyol from D-fructose. Therefore, the introduction of further mutations at this position is likely to increase the activity towards ketose sugars.

Finally, a combinatorial approach was attempted, where NDT variant libraries were created for these four residues for successive rounds of evolution. However, the PCR protocols used were inefficient at generating the desired mutants and the WT DNA was largely observed. This prevented the screening of variants as the number of colonies needed would be too large with the assay methodologies available. This work was not explored further due to time constraints. Future work should focus on the optimisation of the PCR protocol by modifying primer design and reaction conditions. This will allow for the screening of variants using the amine acceptor screen developed in chapter 3. At the end of each round, the best variant(s) should be selected, and further rounds of combinatorial mutagenesis can then be applied until a variant with the desirable activity is found.

#### **4.5 Experimental**

Commercially available reagents and solvents were purchased from Acros Chemicals, Fluorochem, Sigma Aldrich and Thermo Fisher Scientific.

Plasmid isolation was achieved using the GeneJET Plasmid Miniprep Kit from thermo scientific. Sequencing was done by Eurofins Genomics. Plasmids were sent for sequencing at a concentration of 50-200 ng/μl. Sequence analysis was done in Benchling by alignment with the original DNA sequence.

### **General Procedure for Transformation**

Transformations were performed using 50 μL chemically competent *E. coli* Top10/DH5α/BL21(DE3) cells thawed on ice and incubated on ice for 40 min with 1 μL plasmid solution. Cells were incubated at 42 °C for 45 s and immediately chilled on ice for 2 min. 450 μL LB medium was added. The transformed cells were shaken at 200 rpm for 45 minutes at 37 °C. 150 μL was spread on solid LB medium containing the appropriate antibiotic. Cells were grown o/n at 37 °C.

### **Molecular Docking**

The substrate was docked as the external aldimine. Docking was performed by manually overlaying the specific substrate in the active site of the protein using PyMOL comparing to literature. In cases where the cofactor was resolved in the crystal structure, this was used to orient the substrate. This was done to assist the program with docking in the subsequent step.

Specifically, this allowed for the minimisation of the 'Grid Box', which has a critical role in the speed of calculations. Molecular docking was done using Autodock Vina Vina (PyRx 0.8).<sup>252–254</sup> AutoDock Vina Python 3.6 was used as the programming language to perform calculations.

This included the creation of pdbqt files of each respective substrate and protein. The degrees of freedom for the substrate were then set (always to

maximum possible). The coordinates of the Grid Box were customized for each specific complex. In each case the size of the Grid Box was 30 x 30 x 30. This information was then saved in the corresponding configuration file. The AutoDock Vina file was then used to perform the calculations.

Docking results including the lowest binding energy and mean binding energy were obtained from the docking log (dlg) file. The docked substrates configurations were then compared in PyMOL. The best docked substrates were chosen and separate saved in a new pdb file. The substrates were then integrated into the protein in PyMOL. This model was finally refined by YASARA and subsequent visualization performed in PyMOL.

### **Single point mutations of 3HMU and HEWT**

Forward and reverse Oligo-primers containing the desired codon mutations were ordered from Sigma Aldrich. As a template the vector pET22b containing the 3HMU gene or pMP89a containing the HEWT gene. PCR Hi-Fi DNA polymerase was used to polymerise the nucleotides. Different annealing temperatures were trialled to find the best conditions in each case. The reactions were hot-started by heating to 98 °C for 2 min, then cooling to 85 °C before the DNA polymerase was added. They were then subjected to 35 cycles of 95 °C for 15 s, 55 °C for 30 s and 72 °C for 3.5 min, with a final 5 min extension step at 72 °C. The PCR products were checked by Agarose-Gel. The blunt ends of the plasmids were ligated through by using the Quick Ligation Kit (New England BioLabs). 10 µL of the PCR product was added to 10 µL of the 2 x Quick Ligation Buffer added, followed by 1 µL Quick T4 DNA Ligase. The mixture was incubated by RT for 15 min. 0.5 µL DpnI was added to the sample

and incubated for 2 h and subsequently denatured at 80 °C for 20 min. Afterwards 1 µL of the mixture was transformed into *E. coli* DH5α cells and the resultant colonies sent for sequencing.

#### High-Fidelity DNA polymerase

5 µL PCRBio reaction buffer (5x)

1.25 µL Forward Primer

1.25 µL Reverse Primer

1 µL Plasmid DNA

11.25 µL Nuclease Free Water

5 µL dNTP Mix (1.25 mM)

0.25 µL Polymerase

---

Total: 25 µL

#### **Gene Expression/Protein Production**

Plasmids containing the genes for the mutated ω-transaminase (3HMU/HEWT) was used to transform *E. coli* BL21(DE3) competent cells for gene expression. They were grown on solid LB agar containing 100 µg/mL ampicillin. Subsequent liquid cultures contained the same concentration of the appropriate antibiotic.

The clones were grown o/n by shaking at 200 rpm at 37 °C in 5 mL antibiotic-containing LB. 3 mL of the o/n culture was added to 300 mL antibiotic-containing autoinduction media. And shaken at 200 rpm and 37 °C for 24 h. Cells were harvested by centrifugation at 4000 rpm (3220 x g) for 15 min in an Eppendorf model 5810R centrifuge, and the cell pellets were resuspended in

0.1 M KPi buffer, pH 7.5, containing 0.1 mM PLP. Cells were disrupted by ultrasonication on ice for 10 cycles (30s on, 30s off) using a QSonica model Q55 sonicator. Extracts were then clarified by centrifugation at 15000 rpm (21,130 x g) for 30 min using an Eppendorf 5424R microcentrifuge. The presence of over-expressed protein was confirmed by SDS-PAGE.

### **Biotransformations**

Transaminases (50  $\mu$ L cell-free extract) were diluted in KPi buffer (1 mL, 100 mM, pH 7.5) containing PLP (1 mM) and amine donor (15 mM from a 500 mM stock). To this was added amine acceptor (5 mM from a 200 mM stock). Reaction mixtures were incubated at 30 °C, 200 rpm in a shaking incubator for up to 24 hours. Conversion measured by NMR.

### **Optimisation of protein isolation from 48-well plate**

3HMU containing pet28b was cloned into BL21(DE3) as described above. Single colonies were then used to inoculate each well of the 48 well plate containing 3 mL LB and ampicillin (100  $\mu$ L/mL) and the cultures were shaken at 200 rpm, 37 °C o/n. 30  $\mu$ L of o/n culture was then used to inoculate 3 mL of autoinduction media in each well. The plates were shaken at 200 rpm, 37 °C for 24 h. The cells were pelleted by centrifugation at 4000 rpm (3220 x g) for 15 min in an Eppendorf model 5810R centrifuge, and the cell pellets were resuspended in 0.05 M Tris buffer, pH 8, EDTA 1 mM, 0.1 mM PLP (1 mg/mL lysozyme from egg white was added where this method was tested). The plates containing lysozyme were shaken at 37 °C, 200 rpm for 45 minutes. The plates were then frozen o/n and thawed the following day. SDS-pages were used to compare the level of protein in the lysis buffer. Freeze-thaw only,

lysozyme only and freeze-thaw plus lysozyme was tested. The activity in cell extracts was assessed using the acetophenone assay analysing on an EPOCH2 microplate reader with (S)- or (R)-MBA (2.5 mM), pyruvate (2.5 mM) and enzyme. The cell free lysate was then used for subsequent biotransformations.

### **Combinatorial mutagenesis of 3HMU**

The initial protocol was analogous to the protocol described under 4.8.2, here an NDT codon was used at the codon site for mutagenises. This was to reduce the size of the library required for screening. ~80% of mutants were the wild type after sequencing using this protocol.

Site-directed, Ligase-Independent Mutagenesis (SLIM) was then attempted.<sup>251</sup> Short and long forward/reverse primers were designed with an overhang of ~18 bases. Here the longer tailed primers had the same 3' sequence as the shorter primers, but contained the region to be inserted (NDT codon) into the gene on their 5' ends.

**SLIM PCR:** A single PCR was performed for each site and contained the following reagents:

5.0 µL Phusion® HF buffer (5x)

1.0 µL Long forward Primer

1.0 µL Short forward Primer

1.0 µL Long reverse Primer

1.0 µL Short reverse Primer

0.5 µL Plasmid DNA

11.0 µL Nuclease Free Water

4  $\mu$ L dNTP Mix (20 mM)

0.5  $\mu$ L Polymerase Phusion®

---

Total: 25  $\mu$ L

The reactions were hot started by heating to 98 °C for 2 min, then cooling to 85 °C before the DNA polymerase was added. They were then subjected to 26 cycles of 95 °C for 15 s, 59 °C\* for 30 s and 72 °C for 3.5 min, with a final 5 min extension step at 72 °C. 0.5  $\mu$ L of DpnI was added to the sample and incubated for 2 h and subsequently denatured at 80 °C for 20 min.

\*Different annealing temperatures were trialled to find the best conditions were used.

**SLIM hybridization:** Hybridization was performed using two cycles of 65 °C for 5 min followed by 30 °C for 15 min. An aliquot of 20  $\mu$ l was then used to transform 50  $\mu$ L chemically competent TOP10 cells as per the above protocol. The resultant colonies were then sequenced by Eurofins genomics.

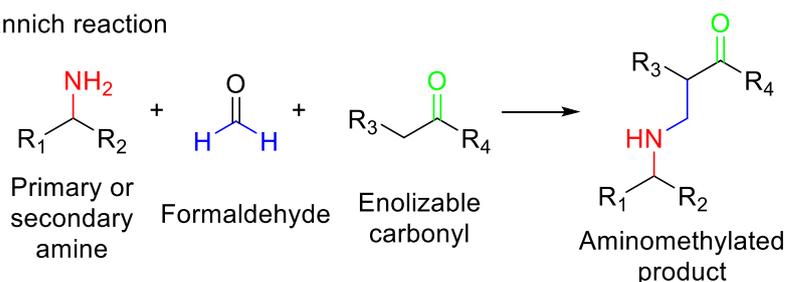
## 5.0 A Transaminase-Mannich Cascade Towards the Generation of Complex Alkaloids

The work presented in this chapter is unpublished but is currently under review in the form of a communication.

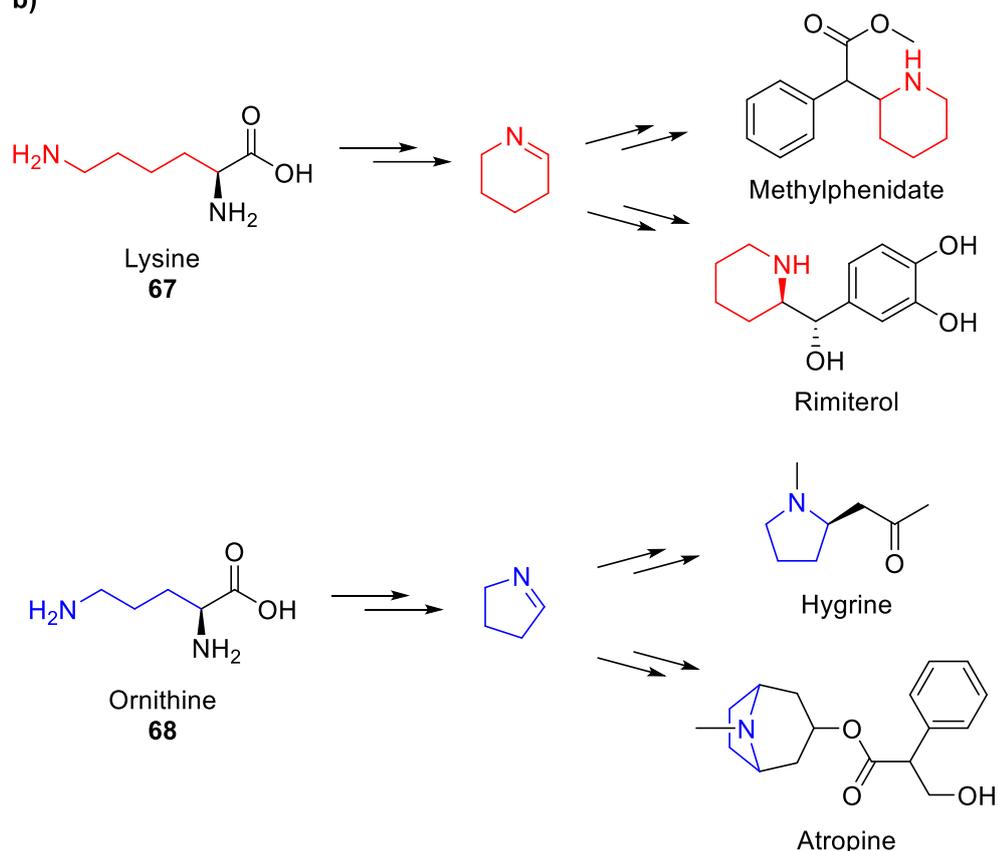
### 5.1 Introduction

The Mannich reaction is a multi-component condensation involving formaldehyde, an enolizable carbonyl and either a primary or secondary amine (**Scheme 23a**), which can allow access to aminomethylated compounds. This reaction is a key step in the biosynthesis of many natural alkaloids, particularly heterocycles derived from the amino acids lysine (**67**) and ornithine (**68**) (**Scheme 23b**).<sup>255</sup> In nature, these are decarboxylated and cyclised to form 5 or 6-membered heterocyclic imines and this is followed by Mannich-type nucleophilic addition to generate 2-substituted heterocyclic alkaloids. These structural motifs frequently appear in pharmaceutical products, such as psychostimulants, antidepressants, and adrenergic drugs.<sup>256,257</sup> The propensity of the Mannich reaction to generate powerful pharmacophores has led to its extensive investigation in total synthesis, and notably, many of these synthetic approaches rely on the use of organocatalysts.<sup>258–261</sup>

a) Mannich reaction



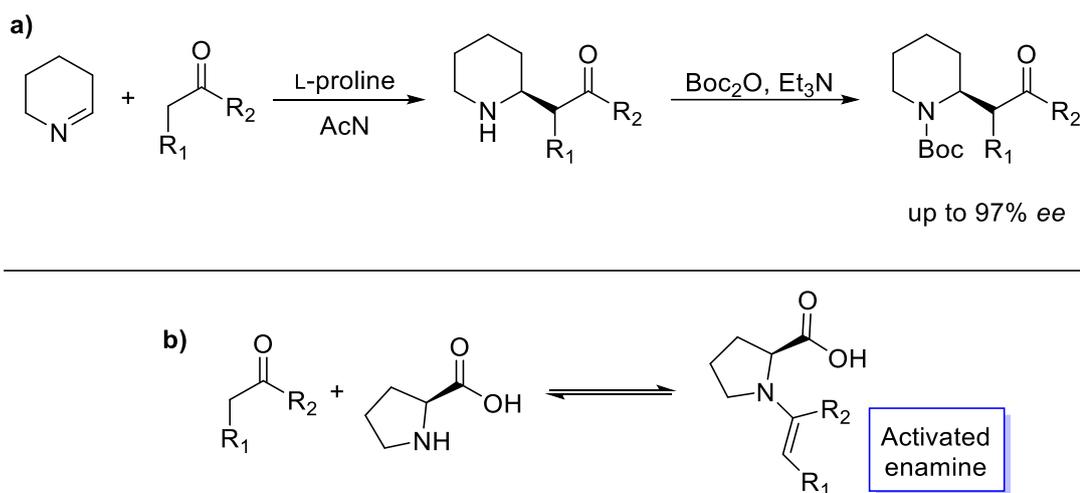
b)



**Scheme 23** – a) General scheme for the Mannich reaction. A multi-component condensation towards aminomethylated compounds involving a non-enolizable aldehyde, an enolizable carbonyl and either a primary or secondary amine. b) A selection of pharmaceuticals/natural products that can be derived from the decarboxylation, and subsequent Mannich-type reaction, of the amino acids lysine and ornithine.

Monaco *et al.* investigated the biomimetic organocatalytic asymmetric synthesis of 2-substituted alkaloids using a variety of catalysts and solvent systems.<sup>262</sup> They report highest yields using L-proline as the catalyst (**Scheme 24a**). Proline works by activation of the carbonyl through the generation of an enamine, and the protic nature of proline further activates this intermediate (**Scheme 24b**).<sup>263</sup> While the use of this chiral ligand also allows for excellent

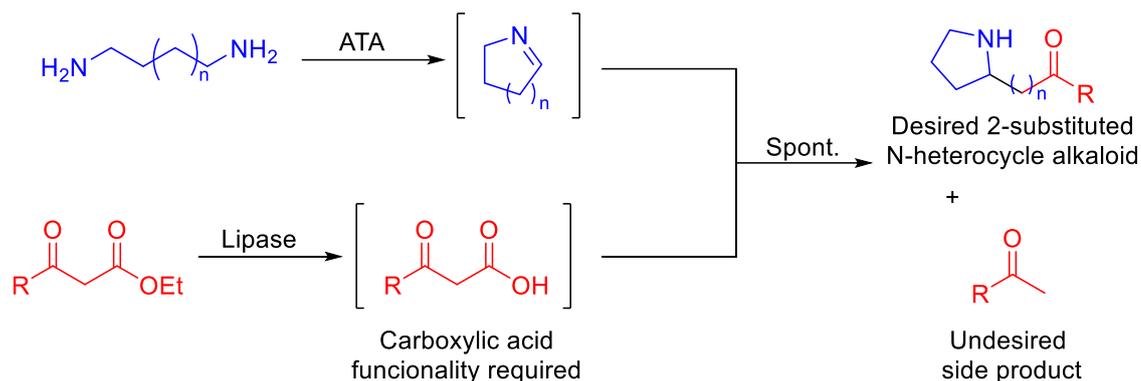
enantioselectivity (up to 97% ee), racemisation of these products in certain solvent systems, particularly under basic conditions, is noted.<sup>262</sup> The generation of these chiral alkaloids, such as pelletierine, could provide a starting point for the asymmetric synthesis of more complex alkaloid natural products. This includes vertine, an anti-inflammatory, which has thus far been synthesised from racemic pelletierine in 11 steps.<sup>264</sup>



**Scheme 24** – a) The biomimetic organocatalytic asymmetric synthesis of 2-substituted alkaloids using L-proline as a catalyst. b) The activation of an enolizable carbonyl using L-proline to generate an enamine.

Galman *et al.* took inspiration from this work to design a biomimetic route to these compounds using a one-pot transaminase-lipase coupled cascade involving a Mannich-type reaction (**Scheme 25**).<sup>265</sup> The transamination of alkyl diamines leads to the generation of cyclic imines that undergo a Mannich-type reaction with a carboxylic ketone, which is produced *via* the enzymatic hydrolysis of the corresponding ethyl ester. While the approach enables access to a variety of 2-substituted N-heterocycle alkaloids, the authors note the necessity of an intermediate with carboxylic acid functionality, as without this, very low conversion to product was recorded. This is further noted when the carboxylic acid is used as a starting material, as it spontaneously

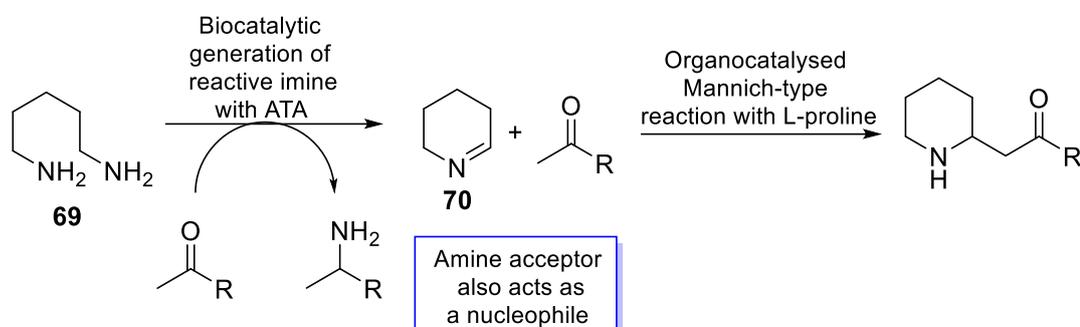
decarboxylases and the resulting ketone/enone does not readily react. For this reason, the authors designed the transaminase-lipase coupled system to generate the carboxylic acid *in situ*. Additionally, the compounds were isolated as racemates due to the predilection of 2-substituted N-heterocycle alkaloids to racemise at basic pH, *via* a retro aza-Michael reaction.



**Scheme 25** – The biomimetic synthesis of 2-substituted N-heterocycle alkaloids using a one-pot transaminase-lipase coupled cascade.

We propose introducing the organocatalyst L-proline to negate the requirement for carboxylic acid functionality in the generation of these complex alkaloids, starting from an alkyl diamine and a simple methyl ketone (**Scheme 26**). The transamination of cadaverine (**69**) will lead to the spontaneous cyclisation to  $\Delta^1$ -piperidine (**70**), a reactive intermediate, that can participate in a cascade towards 2-substituted piperidine alkaloids. Proline will activate the ketone to allow the use of simple methyl ketones, rather than acids, as previously described above by Monaco *et al.* We also anticipate that the use of a chiral ligand will allow for significant enantioselectivity. Our proposed approach would represent a significantly simplified route compared to the two-enzyme approach developed by Galman *et al.* It is envisaged that our methodology will require the use of a single enzyme (the ATA) and allow the ketone to be employed as both the amino acceptor and the nucleophile for the Mannich-

type reaction. Additionally, such methyl ketones are a more readily available starting material than their corresponding ethyl ester and may increase the scope and utility of the methodology.



**Scheme 26** – Proposed scheme for the generation of 2-substituted piperidine alkaloids via a Mannich-type reaction between simple methyl ketones and *in situ* generated  $\Delta^1$ -piperidine.

A key aspect of our proposed methodology is the application of a single ATA for complexity generation. ATAs, like other transferases and widely used biocatalysts, such as ADHs, are typically employed for functional group interconversions. Only in rare examples do these enzymes lead to significantly more complex products, compared to their substrates. Our group have a continued interest in developing methodology where the ATA is responsible for generating reactive intermediates *in situ*, which react further to build complexity. We believe this would be an excellent addition to our repertoire as an example of a one-pot cascade mediated by chemo and enzymatic catalysis, under mild conditions, to generate complex bioactive molecules using cheap and readily available starting materials.

## 5.2 Aims and Objectives

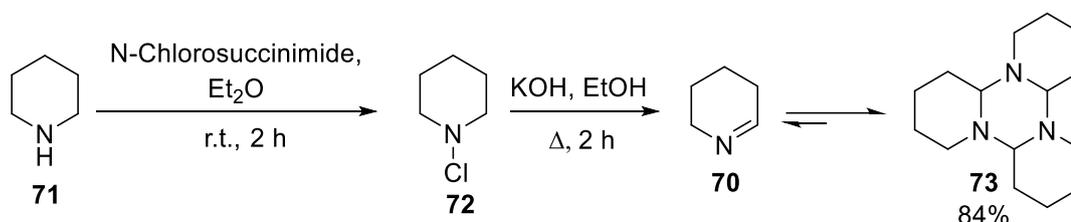
The work shown within this chapter aimed to develop a one-pot organo-biocatalytic cascade for the preparation of 2-substituted N-heterocyclic alkaloids by expanding on work by Galman *et al.* The use of an organocatalyst,

L-proline, has been trialed to access higher conversions and enantioselectivity using simple ketones lacking the previously essential carboxylic functionality. Further work was performed to assess the practicality of a whole cell process towards these alkaloids, whereby an engineered *C. glutamicum* strain produces the cadaverine starting material and ATA/ketone is added exogenously to access the product. Finally, preparative scale reactions towards these alkaloids have been attempted to demonstrate the applicability of this chemistry on scale.

### 5.3.0 Results and discussion

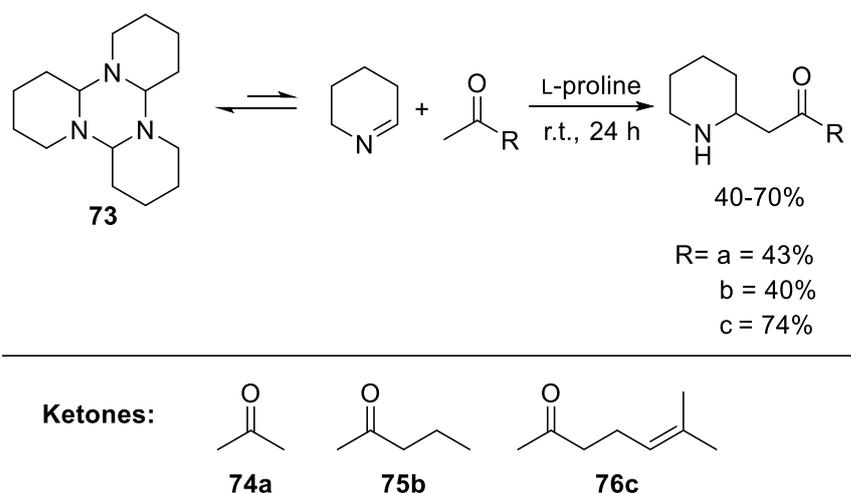
#### 5.3.1 Synthesis of Alkaloid Chemical Standards

The target 2-substituted N-heterocycle alkaloids were first chemically synthesised, to ensure accurate biotransformation conversions could be recorded by using these compounds as standards for GC analysis. A two-step synthesis towards these compounds and their analogues was reported by Monaco *et al.*<sup>262</sup> The first step involved oxidation of piperidine (**71**) to  $\Delta^1$ -piperidine (**70**) by chlorination of the nitrogen using N-chlorosuccinimide, with subsequently refluxing under basic conditions (**Scheme 27**). **70** spontaneously trimerises to the more thermodynamically stable  $\alpha$ -tripiperidine (**73**) and high yield (84%) was recorded in agreement with the literature (76%).<sup>262</sup>



**Scheme 27** – Synthesis of  $\alpha$ -tripiperidine (**73**) from piperidine (**71**) and N-chlorosuccinimide followed by reflux under basic conditions.

The work reported by Monaco *et al* describes the subsequent Mannich reaction with a variety of catalysts and solvent systems to maximise yield and enantioselectivity. Optimal results were observed using L-proline in acetonitrile at room temperature (**Scheme 28**). It is noteworthy that this group did not attempt to isolate the alkaloid product, and instead, boc-protected these compounds *in situ* prior to isolation. However, such protection was not desirable for this work, and it proved possible to isolate these 2-substituted N-heterocycle alkaloids directly from the reaction mixture in moderate-good yields (40-74%).

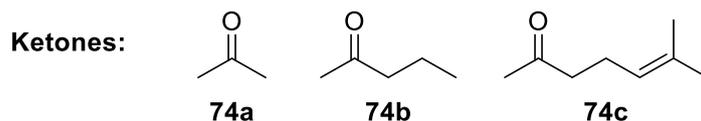
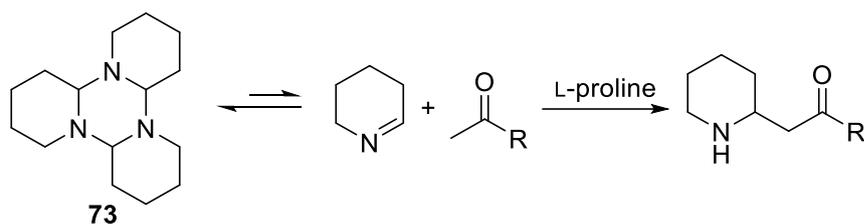


**Scheme 28** – Synthesis of 2-substituted N-heterocycle alkaloids from  $\alpha$ -tripiperidine (**73**) using L-proline as a catalyst.

### 5.3.2 Mannich Reaction Under Biotransformation Conditions

Several Mannich reactions were performed using  $\alpha$ -tripiperidine (**73**) in an analogous manner to the synthesis of the chemical standards, however, this time the chemistry was performed in aqueous media, rather than acetonitrile. This was carried out to assess whether the desired chemistry could take place under biocatalytic conditions, after the biocatalytic generation of  $\alpha$ -tripiperidine

(73). Previous work by our group has shown that **73** is in equilibrium with  $\Delta^1$ -piperidine and that high pH (10-11) is necessary to generate this cyclic imine from cadaverine when using ATA-256.<sup>188</sup> Furthermore, higher pH favours enolate formation, so a basic pH of 10 was chosen for these reactions to drive the equilibrium towards product formation. The results of these transformations are shown in Table 7.



Entry	Ketone	L-Proline (mM)	Conv. (%)
1	74a	0	13
2	74b	0	11
3	74c	0	0
4	74a	100	85
5	74b	100	28
6	74c	100	11

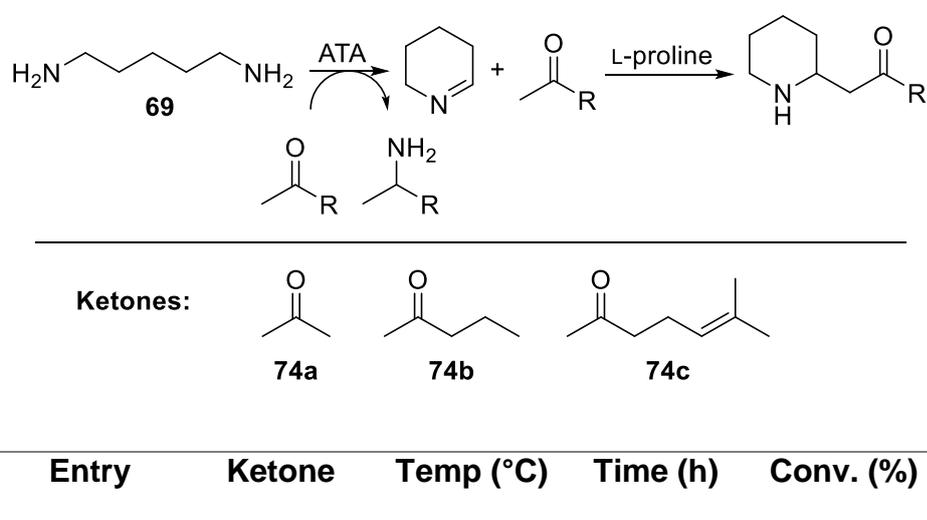
**Table 7** – Mannich reaction towards 2-substituted N-heterocycle alkaloids using  $\alpha$ -tripiperidine, under biotransformation conditions.  $\alpha$ -tripiperidine (3.33 mM), ketone (200 mM), L-proline (0 or 100 mM), HEPES (100 mM), DMSO (10 % v/v), 37 °C, 200 rpm. Conversion was measured by GC-FID. Results are the average of 3 replicates.

When no proline is added to the reaction, only low conversion to product is observed for all substrates, likely due to the poor availability of piperidine as the equilibrium will lie towards trimer formation. However, there is a significant increase in conversion (up to 85%) with the addition of proline, as it activates

the ketone by forming an enamine. Notably, the conversion drops as the length of the alkyl chain increases from acetone (**74a**) to methylheptenone (**74c**). More work is needed to reveal definite trends; however, the 2,3-enamine is expected to be more favourable than the 1,2-enamine with pentanone (**74b**) and methylheptenone (**74c**). Crucially, only the 1,2-enamine would lead to product formation.

### 5.3.3 Time and Temperature Screen

This work was then extended to include the biocatalytic formation of  $\Delta^1$ -piperidine from cadaverine (**69**). The previous conditions, established in the optimisation phase (**section 5.4/Table 7**), were replicated with the replacement of  $\alpha$ -tripiperidine (**73**) with cadaverine (**69**) and the addition of the commercially available ATA256. The reactions were performed over 24 and 48 h as previous work has shown 48 h is required for the maximum conversion of cadaverine (**69**) to piperidine.<sup>188</sup> Additionally, the temperature was also increased to 50 °C as the enzyme has been engineered to be active at higher temperature and it was thought higher temperature may increase conversion. Results are shown in table **8**.



<b>1</b>	<b>74a</b>	37	24	73
<b>2</b>	<b>74a</b>	37	48	75
<b>3</b>	<b>74a</b>	50	24	68
<b>4</b>	<b>74a</b>	50	48	71
<b>5</b>	<b>74b</b>	37	24	21
<b>6</b>	<b>74b</b>	37	48	30
<b>7</b>	<b>74b</b>	50	24	23
<b>8</b>	<b>74b</b>	50	48	23
<b>9</b>	<b>74c</b>	37	24	11
<b>10</b>	<b>74c</b>	37	48	12
<b>11</b>	<b>74c</b>	50	24	10
<b>12</b>	<b>74c</b>	50	48	10

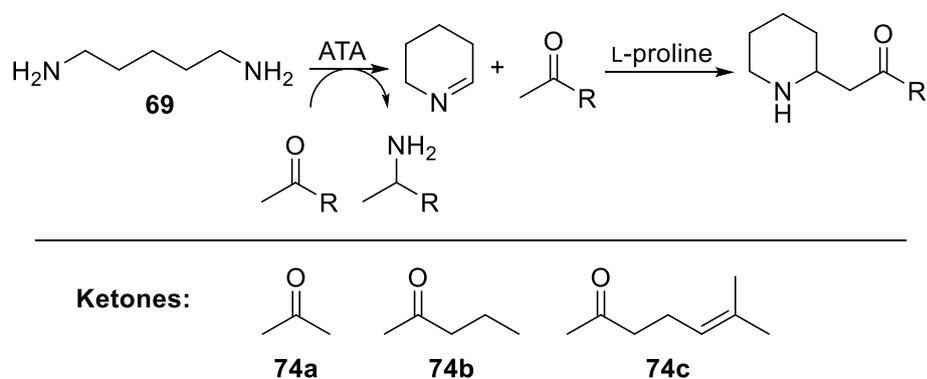
**Table 8** – Analytical scale reaction towards 2-substituted N-heterocycle alkaloids – temperature and time screen. Cadaverine (**69**) (10 mM), ketone (200 mM), L-proline (100 mM), ATA-256 (5 mg mL<sup>-1</sup>), HEPES (100 mM, PLP (1 mM)), DMSO (10 % v/v), 37 or 50 °C, 200 rpm. Conversion was measured by GC-FID. Results are the average of 3 replicates.

In each case, increasing the temperature has an insignificant impact on the conversion to product, suggesting high levels of  $\Delta^1$ -piperidine formation is achieved at 37 °C and the Mannich reaction is perhaps hindered by kinetic rather than thermodynamic influences. Additionally, with acetone (**74a**) and methylheptenone (**74c**), increasing the reaction time leads to no significant change in conversion, suggesting maximum levels of  $\Delta^1$ -piperidine are achieved in 24 h. However, a significant increase in conversion is recorded with pentanone (**74b**) after 48 h when compared to 24 h, further implying there is a kinetic factor limiting the Mannich reaction. Overall, the conversion to alkaloid product aligns well with conversions recorded performing this chemistry organocatalytically with  $\alpha$ -tripiperidine (**section 5.4**), confirming the

transamination step is not rate limiting when performing such biotransformations.

### 5.3.4 Amine Donor Concentration Screen

Reactions up to this point were limited to 10 mM of cadaverine (**69**) and, hence, the maximum concentration of product possible was also 10 mM. Additionally, 20 eq of ketone had been used, leading to poor atom economy. Clearly, better conditions are required to make this an economically feasible process at scale. Higher concentrations of amine were trialed to access higher concentrations of the alkaloid and improve the atom economy (**Table 9**). It is worth noting that 2 eq. of ketone is essential to reach the theoretical maximum yield of 100%, as 1 eq. is used as the amine acceptor and the second eq. is then available for the Mannich reaction.



Entry	Ketone	Substrate	Conv.	Product
		concentration (mM)	(%)	concentration (mM)
1	74a	10	75	8
2	74a	50	65	32
3	74a	100	37	37
4	74a	150	18	27

<b>5</b>	<b>74b</b>	10	30	3
<b>6</b>	<b>74b</b>	50	20	10
<b>7</b>	<b>74b</b>	100	11	11
<b>8</b>	<b>74b</b>	150	8	12
<b>9</b>	<b>74c</b>	10	12	1
<b>10</b>	<b>74c</b>	50	5	3
<b>11</b>	<b>74c</b>	100	3	3
<b>12</b>	<b>74c</b>	150	2	3

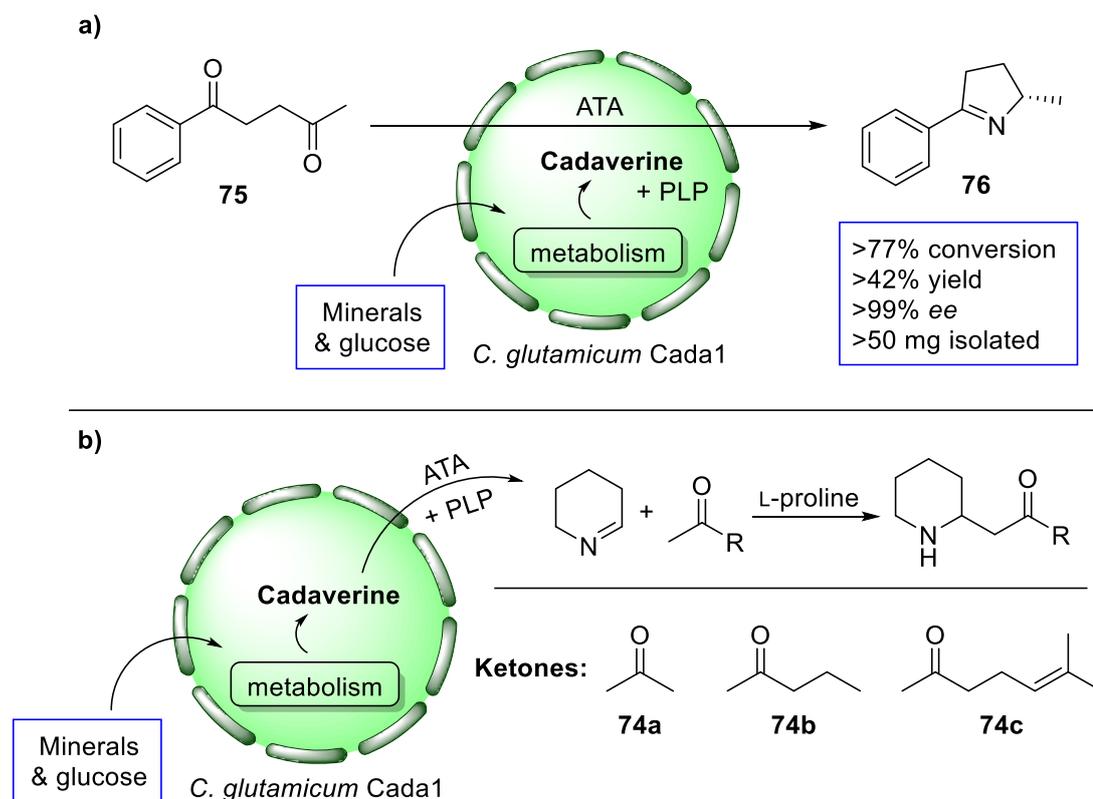
**Table 9** – Amine concentration screen. Cadaverine (**69**) (10, 50, 100 or 150 mM), ketone (200 mM), L-proline (100 mM), ATA-256 (5 mg mL<sup>-1</sup>), HEPES (100 mM, pH 10, PLP (1 mM)), DMSO (10 % v/v), 37 °C, 200 rpm. Conversion was measured by GC-FID. Results are the average of 3 replicates.

Increasing the concentration of cadaverine to 50 mM leads to a significant improvement in the atom economy as the eq. of ketone drops from 20 to just 4 equivalents. While this increase in concentration causes a drop in conversion across all three substrates, up to 65% is achieved when using acetone (**74a**), allowing access to >30 mM pelletierine, making this process more feasible on a larger scale. Increasing the concentration of cadaverine to 100 mM, with two eq. of ketone, shows a much more significant drop in conversion (**Table 9, entry 3**). However, the concentration of product increases, indicating that the transamination of cadaverine (i.e generation of  $\Delta^1$ -piperidine) is still successful at high concentrations and that it is the availability of ketone that is preventing product formation. It therefore appears that an excess of ketone is essential for this chemistry. The concentration of cadaverine was finally increased to 150 mM, with 1.3 eq. of ketone. In each case, this leads to a significant drop in conversion and no increase in N-heterocycle alkaloid concentration was observed. This shows that accessing higher concentrations

of  $\Delta^1$ -piperidine does not always lead to higher concentrations of product, likely due to the low eq. of ketone (0.65) remaining after transamination.

### 5.3.5 Whole Cell Biocatalysis Using the Cadaverine Producing *C. glutamicum* Strain Cada1

Previous work by our group demonstrated the use of *C. glutamicum* as a biofactory for self-sufficient transamination reactions,<sup>189</sup> where the cells generate the amine donor (cadaverine) and produce the ATA and the ketone is added exogenously (**Scheme 29a**). However, we envisaged a one-pot system whereby the ketone and commercial ATA could be added exogenously to the cells producing cadaverine, hence, generating the 2-substituted N-heterocycle alkaloids (**Scheme 29b**).



**Scheme 29** – a) Previous work demonstrating a whole-cell biotransformation, using engineered *C. glutamicum* cells that produce cadaverine to generate (S)-pyrroline (**76**) from 1-phenyl-1,4-pentanedione (**75**). b) General scheme for the whole cell synthesis of 2-substituted N-heterocycle alkaloids using cadaverine producing *C. glutamicum*.

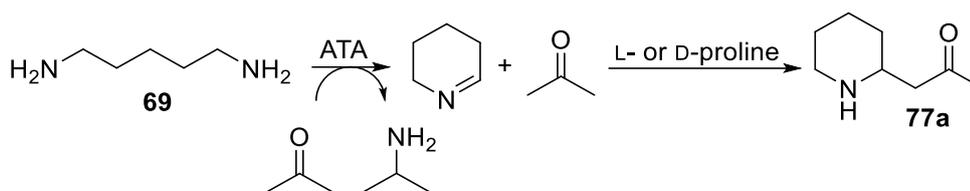
Biotransformations were performed using 200 mg/mL of wet cells in either minimal media or HEPES buffer, as previously described by our group.<sup>189</sup> The use of minimal media, in theory, allows for the continual generation of cadaverine from glucose and ammonium salts. 1-Phenyl-1,4-pentanedione was used as a positive control as up to 77% conversion to the corresponding cyclic amine (**76**) is reported with these cells and an  $\omega$ -ATA from *P. chlororaphis* ssp. *Aerofaciens* (PcATA).<sup>189,266</sup> Furthermore, ATA-256 readily accepts this substrate. Using analogous conditions to that described by Grigoriou et al, replacing PcATA with exogenously added ATA-256 enabled conversions of 37 %. L-proline was added to confirm it was not toxic to the cells and an insignificant drop in conversion to 35 % was recorded. This is a significant (40%) drop in conversion compared to the use of endogenously produced PcATA. The underperformance of ATA-256 at pH <10 has been described and is probable that the use of pH 8 hindered conversion to product with this enzyme.<sup>188</sup>

*C. glutamicum* has a preference for slightly basic conditions (7-8.5), however, increasing the pH beyond this quickly inhibits growth.<sup>267,268</sup> Indeed, no growth is observed at pH >10. Therefore, the optimal pH for the cells and the ATA differs significantly. So, a range of pHs was tested. The biotransformations were performed analogous to those from the analytical scale reactions above, with the addition of *C. glutamicum* cells rather than cadaverine. pH 7-10 was tested, however, no product was detected with any of the three ketone substrates at any of these pHs. It is noteworthy that previous work by the group restricted ketone concentrations to 5 mM as conversions beyond this were limited, suggesting that even under optimal conditions the concentration of

cadaverine within the cells remains very low. A range of concentrations of acetone (20-200 mM) at pH 7-10 was explored, however, no product was detected. It is likely that the reaction conditions are too unfavourable for product formation due the low levels of cadaverine present and the incompatibility of the cells/ATA optimal pH ranges. ATAs that can perform this chemistry at lower pHs have not been explored here, however, future work should focus on identifying these ATAs and extending their use to whole-cells.

### **5.3.6 Preparative Scale Biotransformation towards a Complex Alkaloid Using L- and D-Proline.**

To demonstrate the applicability of this work on preparative scale, biotransformations were carried out to access the natural product pelletierine (**77a**) (**Scheme 30**). The optimal conditions for maximum atom economy revealed from the analytical scale reactions were used (**Table 9**). This included using 50 mM cadaverine (**69**) and 4 eq. of acetone in 20 mL of HEPES buffer where 65% conversion was recorded. The use of both L- and D-proline as the catalyst was trialled to ascertain whether the chirality of the catalyst effects the enantioselectivity of the Mannich reaction. Purification by silica gel column chromatography was necessary after the biotransformation to remove unreacted  $\alpha$ -tripiperdine generated from the transamination of cadaverine (**69**). After purification, 80 mg (57%) and 85 mg (60%) of pelletierine was obtained using L- and D-proline, respectively. HPLC analysis revealed the products were racemic suggesting the presence of chiral ligands does not prevent the retro Michael reaction under the basic conditions used. Therefore, enzymes that can perform this chemistry at lower pH should be a focus of future work.



**Scheme 30** – General scheme for the preparative scale biocatalytic synthesis of pelletierine (**77a**) from cadaverine. Cadaverine (**69**) (50 mM), acetone (200 mM), L- or D- proline (100 mM), ATA-256 (5 mg mL<sup>-1</sup>), HEPES (100 mM, pH 10, PLP (1 mM)), DMSO (10 % v/v), 37 °C, 200 rpm.

## 5.4 Conclusions and Future Work

The work shown herein expands the scope of hybrid chemo/enzymatic catalytic cascade processes for the synthesis of valuable targets. This has been demonstrated using ATAs to generate reactive intermediates *in situ* that undergo further chemistry to generate significant complexity. The successful transamination of cadaverine using the commercially available ATA-256, and a variety of ketones as the amine acceptor, led to the production of  $\Delta^1$ -piperidine that could then further react with the ketone starting material, under the presence of proline, to generate 2-substituted N-heterocycle alkaloids. Conversions of up to 75% were recorded using this methodology. Using this one-pot chemoenzymatic cascade, a preparative scale up of these reactions enabled access to the natural product pelletierine (60% yield), which further demonstrates the synthetic utility of designing hybrid cascade processes. However, the alkaloids generated were racemic due their known propensity to racemise *via* a retro Michael reaction at basic pH. Therefore, future work should focus on the identification of ATAs that can perform this chemistry at lower pH as the use of chiral proline should allow for the isolation of chiral alkaloids. Further work should also be done to broaden the

substrate scope of this chemistry, particularly using cyclic/aromatic ketones to further increase the complexity of the products.

## 5.5 Experimental

### General Methods

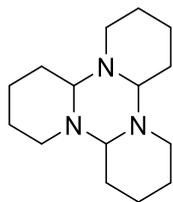
NMR spectra were recorded on a Bruker Avance 400 spectrometer ( $^1\text{H}$  400 MHz,  $^{13}\text{C}$  100 MHz) and are referenced internally according to residual solvent signal.

Analytical samples were prepared for GC analysis by basifying (pH ~13) with 50  $\mu\text{L}$  NaOH (10 M) and extracted with EtOAc (800  $\mu\text{L}$ ). GC analysis was performed using a Nexis GC-2030 chromatograph equipped with a flame ionising detector, an AOC-20s autosampler and an achiral SHIMADZU SH-Rxi-5ms Crossbond® column (30 m x 0.25 mm x 0.25  $\mu\text{m}$ ). The front inlet temperature was set to 230  $^\circ\text{C}$  and the front detector was set to 270  $^\circ\text{C}$ . Split flow was set to 158.2  $\text{mL}\cdot\text{min}^{-1}$  and the nitrogen gas was set to a constant flow of 1.92  $\text{mL}\cdot\text{min}^{-1}$ . Temperature program: 40  $^\circ\text{C}$  hold for 2 minutes followed by 20  $^\circ\text{C}\cdot\text{min}^{-1}$  temperature rise to 150  $^\circ\text{C}$  and then a hold for 5 minutes followed by a 30  $^\circ\text{C}\cdot\text{min}^{-1}$  temperature rise to 270  $^\circ\text{C}$  and a further hold for 10 minutes.

### Materials

Commercially available reagents and solvents were purchased from Acros Chemicals, Fluorochem, Sigma Aldrich and Thermo Fisher Scientific. Commercially available transaminase ATA256 was purchased from Codexis in the form of lyophilised cell extracts.

### Synthesis of $\alpha$ -tripiperidine (73)<sup>262</sup>



73

N-Chlorosuccinimide (8.23 g, 61.6 mmol) was suspended in Et<sub>2</sub>O and cooled to 0 °C before piperidine (5.0 g, 5.8 mL, 58.7 mmol) was added dropwise and the suspension was then stirred at r.t. for 2 hours. The resultant mixture was filtered and washed with Et<sub>2</sub>O before the filtrate was washed with water (30 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo* (**without heating**)\* to give crude 1-chloropiperide as a yellow oil. Crude 1-chloropiperide was dissolved in a solution of EtOH (60 mL) and KOH (3.50 g) and the solution was refluxed for 2 h, filtered and washed with EtOH (20 mL x 3). The filtrate was concentrated *in vacuo* to ~20 mL and Et<sub>2</sub>O (150 mL) and water were added, and the solution washed with Et<sub>2</sub>O (3 x 20 mL). The organic fractions were combined and dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo* to give α-tripiperidine as a yellow oil (4.09 g, 84%). **<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δH 3.10 - 3.03 (3H, m), 2.79 - 2.72 (3H, m), 2.02 - 1.91 (3H, m), 1.74 - 1.56 (9H, m), 1.55 - 1.46 (6H, m), 1.31 - 1.15 (3H, m); **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δC 81.8 (CH), 46.3 (CH<sub>2</sub>), 29.3 (CH<sub>2</sub>), 25.9(CH<sub>2</sub>), 22.2 (CH<sub>2</sub>); **HRMS-ESI (m/z):** C<sub>15</sub>H<sub>28</sub>N<sub>3</sub><sup>+</sup> [M+H]<sup>+</sup> theoretical 250.2278, found 250.2278; **IR (ATR)** 3350, 2924, 2851, 1445, 1378, 1238, 1112. Data is consistent with literature.<sup>262</sup>

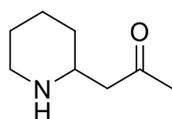
*\*Chloramines are known to be unstable, and heating is likely to result in decomposition of the intermediate.*

### General procedure for Mannich product chemical standards

$\alpha$ -Triperidine (0.82 g, 3.28 mmol), L-proline (188 mg, 1.64 mmol) and ketone (**74a-74c**) (19.7 mmol) were dissolved in acetonitrile (50 mL) before being stirred at r.t. for 24 h. The reaction mixture was diluted in DCM (70 mL) and washed with brine (3 x 40 mL). The organic layer was dried over MgSO<sub>4</sub> and concentrated *in vacuo*. The crude product was purified *via* column chromatography and concentrated *in vacuo* to give product as an oil.

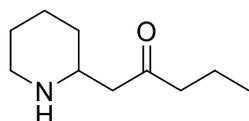
**Note:** 2-substituted N-heterocycles are associated with acute toxicity, handle with extreme care.

### Mannich products



**77a**

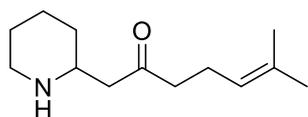
Derived from ketone **74a**. Yellow oil (209 mg, 43%) eluted in DCM/MeOH (90:10). **<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>)  $\delta$ H 3.02 – 2.95 (1H, m, CH<sub>2</sub>), 2.94 – 2.88 (1H, m, CH), 2.64 (1H, td,  $J = 11.9, 2.9$  Hz, CH<sub>2</sub>), 2.48 (2H, d,  $J = 6.4$  Hz, CH<sub>2</sub>), 2.12 (3H, s, CH<sub>3</sub>), 1.88 (1H, s, broad, NH), 1.78 – 1.69 (1H, m, CH<sub>2</sub>), 1.60 – 1.50 (2H, m, CH<sub>2</sub>), 1.41– 1.30 (2H, m, CH<sub>2</sub>), 1.17 – 1.06 (1H, m, CH<sub>2</sub>); **<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>)  $\delta$ C 208.4 (C=O), 52.3 (CH), 50.8 (CH<sub>2</sub>), 46.8 (CH<sub>2</sub>), 32.5 (CH<sub>2</sub>), 30.6 (CH<sub>3</sub>), 26.0 (CH<sub>2</sub>), 24.6 (CH<sub>2</sub>); **HRMS-ESI (m/z):** C<sub>8</sub>H<sub>16</sub>NO<sup>+</sup> [M+H]<sup>+</sup> theoretical 142.1226, found 142.1228; **IR (ATR)** 3325, 2925, 2852, 1707, 1331, 1165, 1077. Data consistent with literature.<sup>262</sup>



**77b**

Derived from ketone **74b**. Yellow oil (222 mg, 40%) eluted in DCM/MeOH (95:5). **<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>)  $\delta$ H 3.01 – 2.94 (1H, m, CH<sub>2</sub>), 2.94-2.87 (1H, m, CH), 2.65 – 2.56 (m, 1H, CH<sub>2</sub>), 2.48 – 2.40 (2H, m, CH<sub>2</sub>),

2.30 (2H, t,  $J = 7.3$  Hz,  $\text{CH}_2$ ), 1.73 – 1.65 (1H, m,  $\text{CH}_2$ ), 1.57 – 1.46 (4H, m,  $\text{CH}_2$ ), 1.44 – 1.24 (2H, m,  $\text{CH}_2$ ), 1.19 – 1.10 (1H, m,  $\text{CH}_2$ ), 0.84 (3H, t,  $J = 7.4$  Hz,  $\text{CH}_3$ );  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  210.7 ( $\text{C}=\text{O}$ ), 52.5 ( $\text{CH}$ ), 49.4 ( $\text{CH}_2$ ), 46.7 ( $\text{CH}_2$ ), 45.4 ( $\text{CH}_2$ ), 32.2 ( $\text{CH}_2$ ), 25.6 ( $\text{CH}_2$ ), 24.4 ( $\text{CH}_2$ ), 17.0 ( $\text{CH}_2$ ), 13.7 ( $\text{CH}_3$ ); **HRMS-ESI (m/z):**  $\text{C}_{10}\text{H}_{20}\text{NO}^+$  170.1545  $[\text{M}+\text{H}]^+$ , found 170.1542. Data consistent with literature.<sup>265</sup>



**77c**

Derived from ketone **74c**. Brown oil (507 mg, 74%)

eluted in DCM/MeOH/ $\text{Et}_3\text{N}$  (95:5:0.1).  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  H 5.06 – 5.00 (1H, m,  $\text{CH}$ ), 3.02 – 2.95 (1H, m,  $\text{CH}_2$ ), 2.95 – 2.88 (1H, m,  $\text{CH}$ ), 2.64 (1H, td,  $J = 11.7, 2.8$  Hz,  $\text{CH}_2$ ), 2.47 -2.38 (3H, m,  $\text{CH}_2$ ), 2.27 – 2.19 (2H, m,  $\text{CH}_2$ ), 1.78 – 1.72 (1H, m,  $\text{CH}_2$ ), 1.66 (3H, s,  $\text{CH}_3$ ), 1.60 (3H, s,  $\text{CH}_3$ ), 1.57 – 1.51 (2H, m,  $\text{CH}_2$ ), 1.45 – 1.26 (3H, m,  $\text{CH}_2$ ), 1.19 – 1.07 (1H, m,  $\text{CH}_2$ );  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  210.4 ( $\text{C}=\text{O}$ ), 132.7 ( $\text{CH}=\text{C}(\text{CH}_3)_2$ ), 122.6 ( $\text{CH}$ ), 52.3 ( $\text{CH}$ ), 49.9 ( $\text{CH}_2$ ), 46.7 ( $\text{CH}_2$ ), 43.4 ( $\text{CH}_2$ ), 32.5 ( $\text{CH}_2$ ), 26.0 ( $\text{CH}_2$ ), 25.6 ( $\text{CH}_3$ ), 24.7 ( $\text{CH}_2$ ), 22.4 ( $\text{CH}_2$ ), 17.6 ( $\text{CH}_3$ ); **HRMS-ESI (m/z):**  $\text{C}_{13}\text{H}_{24}\text{NO}^+$  210.1858  $[\text{M}+\text{H}]^+$ , found 210.1853; **IR (ATR)** 3318, 2925, 2853, 1707, 1638, 1440, 1376, 1287, 1121.

### Mannich reaction under biotransformation conditions

A solution of  $\alpha$ -tripiperidine (3.33 mM, 1 mL) in HEPES buffer (100 mM), containing L-proline (0 or 100 mM) and ketone (**74a-74c**) (200 mM) in 10% DMSO was prepared. The microfuge tubes were then incubated at 37 °C, 200 rpm, for 24 h. The reaction was basified with NaOH (50  $\mu\text{L}$ ) and ethyl acetate (800  $\mu\text{L}$ ) was added, the mixture was centrifuged (13000 rpm, 2 min) and the organic layer was analysed by GC-FID.

## Analytical scale biotransformations towards Mannich products

A 500  $\mu\text{L}$  solution of cadaverine dihydrochloride (10, 50, 100 or 150 mM) in HEPES buffer (100 mM, pH 10, PLP (1 mM)) containing L-proline (100 mM) and ketone (**74a-77c**) (200 mM) in 10% DMSO was prepared. Solution is pH adjusted to pH 10. The final volume was adjusted to 1 mL with an enzyme solution containing the commercially available ATA256 (10 mg/mL), HEPES buffer (100 mM, pH 10, PLP (1mM)) and 10% DMSO. The mixture was incubated at 200 rpm, 37 or 50 °C for 24 or 48 h. The reaction was basified with NaOH (50  $\mu\text{L}$ ) and ethyl acetate (800  $\mu\text{L}$ ) was added, the mixture was centrifuged (13000 rpm, 2 min) and the organic layer was analysed by GC-FID.

## Preparation of Cada1 cells

Minimal media<sup>269</sup> recipe (per litre): 20 g  $(\text{NH}_4)_2\text{SO}_4$ , 5 g urea, 1 g  $\text{KH}_2\text{PO}_4$ , 1 g  $\text{K}_2\text{HPO}_4$ , 13.25 mg  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 0.25 g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 10 mg  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 10 mg  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ , 0.02 mg  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ , 0.313 mg  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 1 mg  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.2 mg biotin, 42 g 3-(*N*-morpholino)-propane-sulphonic acid (MOPS) and 4% (w/v) glucose.

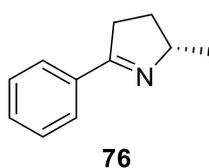
*C. glutamicum* GRLys1 containing the lysine decarboxylase plasmid (Cada1) (see appendix for strain details) was spread over LB agar plates containing kanamycin (50  $\mu\text{g}/\text{mL}$ ) and incubated at 30 °C overnight. Several colonies were used to inoculate 25 mL brain heart infusion media enriched with sorbitol (5% w/v) (BHIS), containing kanamycin (50  $\mu\text{g}/\text{mL}$ ), and incubated at 30 °C overnight (200 rpm). Cell pellets were then harvested *via* centrifugation (4000 rpm, 10 mins), washed with minimal media (3 x 20 mL) and used to inoculate

a 50 mL of minimal media containing kanamycin (50  $\mu\text{g/mL}$ ). The culture was instantly induced with IPTG (1 mM) and incubated at 30  $^{\circ}\text{C}$  (200 rpm). The cells were harvested *via* centrifugation (4000 rpm, 10 mins) after 24 h.

### Analytical scale whole cell biotransformations

A 500  $\mu\text{L}$  solution of cadaverine dihydrochloride (10 mM) in HEPES buffer (100 mM, PLP (0.1 mM)), or minimal media, containing L-proline (100 mM) and ketone (**74a-74c**) (200 mM) in 10% DMSO was prepared. The volume was made up to 1 mL with an enzyme solution containing the commercially available ATA256 (10 mg/mL), 200 mg/mL wet cells in HEPES buffer (100 mM, pH 7-10, PLP (0.1 mM)) or minimal media and 10% DMSO. The mixture was incubated at 200 rpm, 30  $^{\circ}\text{C}$  for 24 or 48 h. The reaction was basified with NaOH (50  $\mu\text{L}$ ) and ethyl acetate (800  $\mu\text{L}$ ) was added. The mixture was then centrifuged (13000 rpm, 2 min) and the organic layer was analysed by GC-FID.

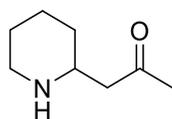
### Preparative scale biocatalytic synthesis of (S)-pyrroline (76)



A 20 mL solution of 1-phenyl-1,4-pentanedione (50 mM) in HEPES buffer (100 mM, pH 8, PLP (1 mM)) containing isopropylamine (200 mM) and ATA256 (5 mg/mL) in 10% DMSO was prepared. The solution was incubated at 37  $^{\circ}\text{C}$ , 200 rpm for 48 h. This solution was acidified with HCl (pH 1.0) and washed with Et<sub>2</sub>O (3 x 20 mL). The organic phases were discarded, and the aqueous phase was basified with NaOH (pH 12.0) and washed with Et<sub>2</sub>O (3 x 20 mL). The combined organic extracts were dried over MgSO<sub>4</sub> and

concentrated *in vacuo* to afford **X** as a light-yellow oil (110 mg, 69% yield). **<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δH 7.85-7.82 (m, 2H), 7.42 – 7.37 (m, 3H), 4.34 – 4.24 (m, 1H), 3.06 (dddd, *J* = 16.9, 9.9, 4.8, 2.2 Hz, 1H), 2.88 (dddd, *J* = 16.9, 9.7, 7.7, 1.8 Hz, 1H), 2.25 (dddd, *J* = 12.5, 9.8, 7.6, 4.8 Hz, 1H), 1.60 – 1.50 (m, 1H), 1.36 (d, *J* = 6.8, 3H); **<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 171.6, 134.5, 130.0, 128.2, 127.4, 68.2, 35.0, 30.5, 22.0; **HRMS-ESI (m/z):** C<sub>11</sub>H<sub>14</sub>N<sup>+</sup> 160.1126 [M+H]<sup>+</sup>, found 160.1126. Data consistent with literature.<sup>189</sup>

### Preparative scale biocatalytic synthesis of **77a**



**77a**

A 20 mL solution of cadaverine dihydrochloride (50 mM) in HEPES buffer (100 mM, PLP (1 mM)) containing acetone (200 mM), L- or D-proline (100 mM) and ATA256 (5 mg/mL) in 10% DMSO was prepared. The solution was incubated at 37 °C, 200 rpm for 48 h. This solution was basified with NaOH (pH 12.0) and washed with Et<sub>2</sub>O (3 x 20 mL). The combined organic extracts were dried over MgSO<sub>4</sub> and concentrated *in vacuo*. The resultant oil was purified *via* column chromatography (90:10/DCM:MeOH) to afford **77a** as a light-yellow oil. L-Proline (80 mg, 57%), D-proline (85 mg, 60%). See *general procedure for the synthesis of the chemical standard for the full characterisation of 77a (Page 149)*.

### GC traces and standard curves

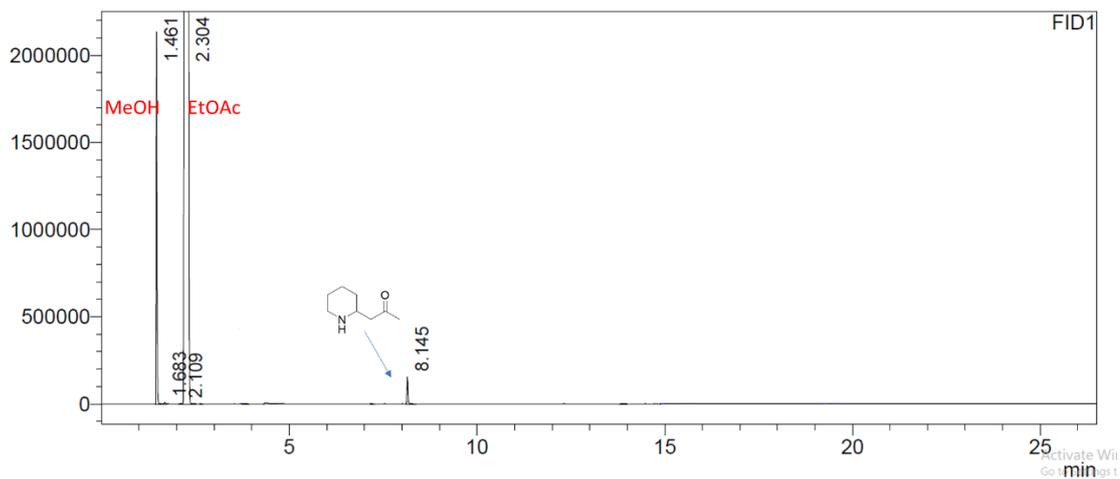


Figure 37 - Representative GC trace of **77a** using an achiral SHIMADZU SH-Rxi-5ms Crossbond® column.

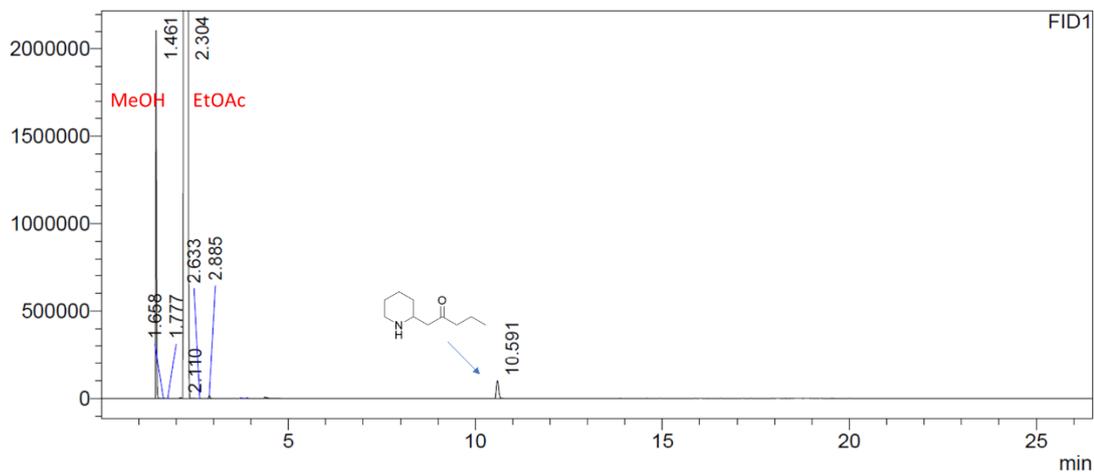
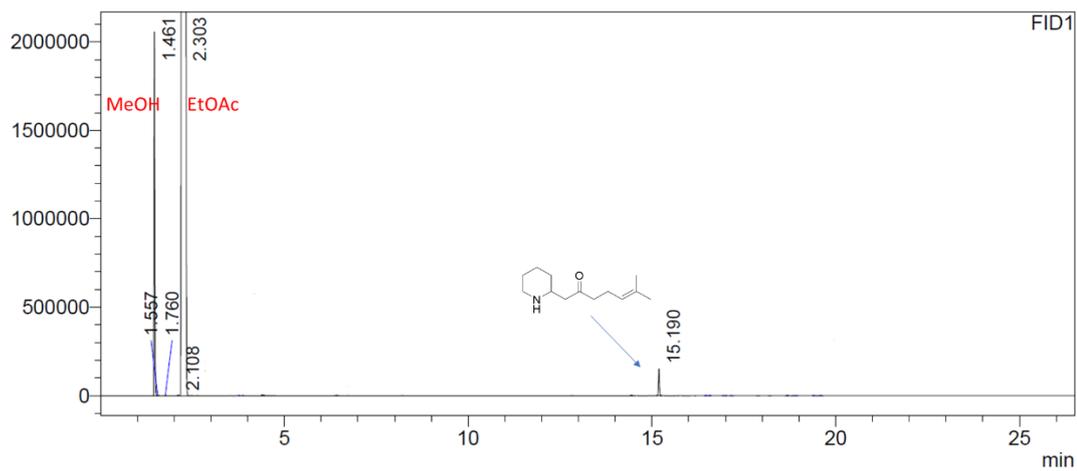
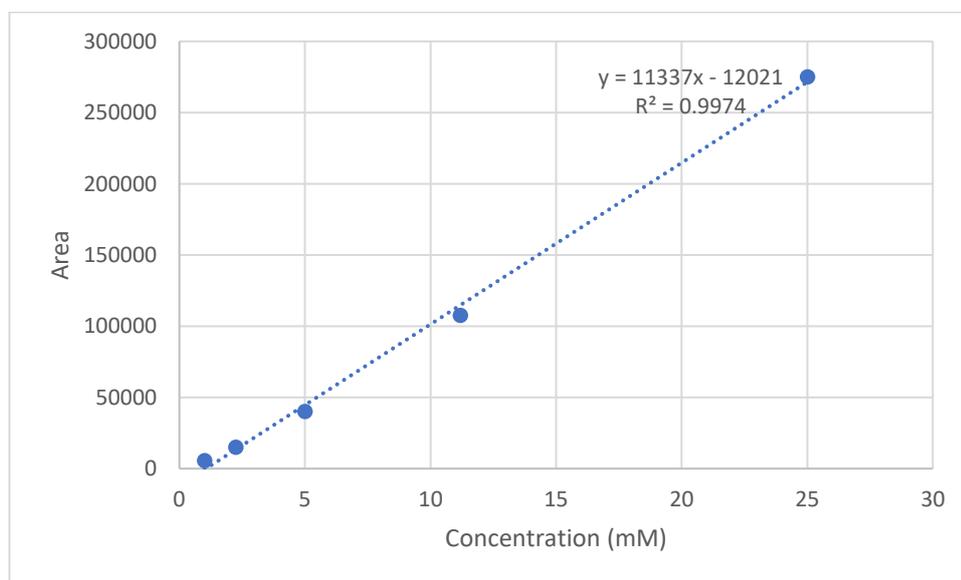


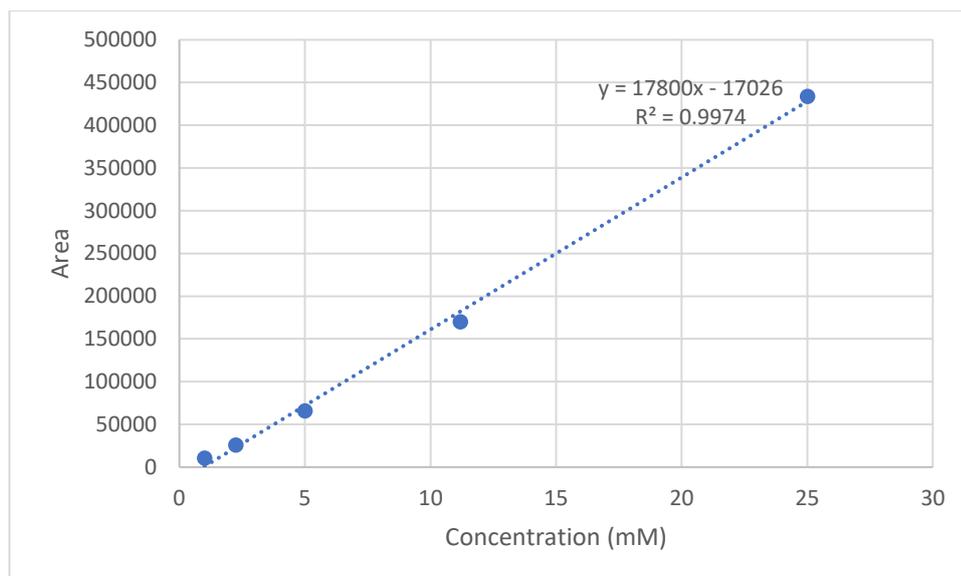
Figure 38 - Representative GC trace of **77b** using an achiral SHIMADZU SH-Rxi-5ms Crossbond® column.



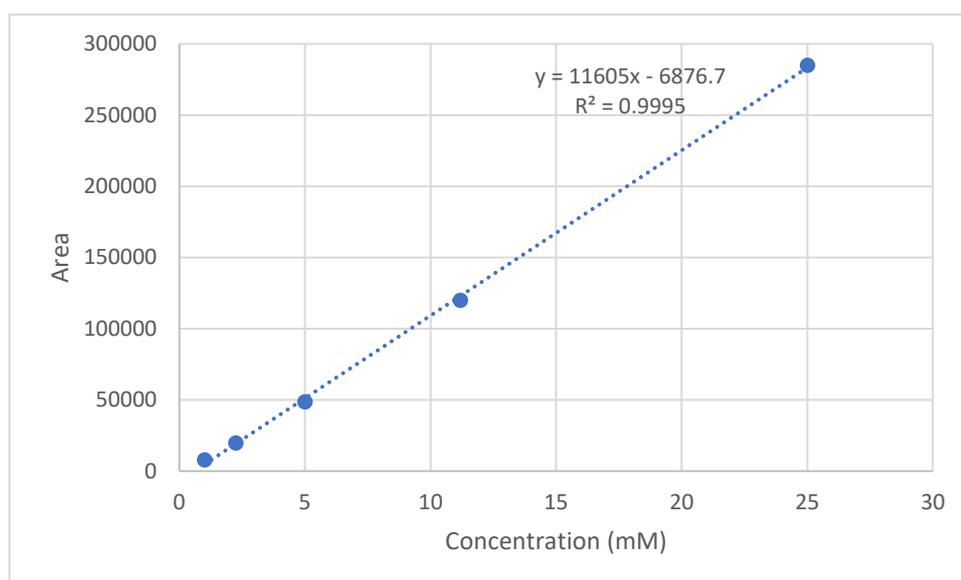
**Figure 39** - Representative GC trace of **77c** using an achiral SHIMADZU SH-Rxi-5ms Crossbond® column.



**Figure 40** - Concentration of **77a** (mM) vs peak area. Standard curve to calculate conversion to **77a**.



**Figure 41** - Concentration of **77b** (mM) vs peak area. Standard curve to calculate conversion to **77b**.



**Figure 42** - Concentration of **77c** (mM) vs peak area. standard curve to calculate conversion to **77c**.

## 6.0 Final Conclusions and Future Work

Initial work demonstrated the first reported example of the biocatalytic conversion of simple, cyclic monosaccharides to the corresponding  $\omega$ -aminopolyols using ATAs. The screening of a large ATA library for activity towards these sugars using a colorimetric assay employing *o*-xylylenediamine suggests these substrates are not widely accepted by this enzyme class, and as such, no ATAs were identified that accepted ketose sugars. However, three ATAs were selected for analytical-scale reactions and activity towards a small panel of aldoses has been documented. This allowed for the optimal conditions with these substrates to be determined. ATAs identified included two wild-type enzymes, 3HMU (from *Rugeria pomeroy*) and HEWT (from *Halomonas elongata*), and the commercially available ATA-256. Further to this, a selection of amino alcohols were synthesized on a preparative scale using ATA-256. This methodology allowed for the facile attainment of up to 475 mg of amino alcohol product, which clearly demonstrates the scalability of this approach giving it the potential to become the primary route for the synthesis of such aminopolyols.

While this work represented a significant expansion to the substrate panel available to these enzymes, only aldoses were identified as substrates and there was a desire to expand this methodology to enable the direct conversion of ketoses, where the enzyme mediates the introduction of a chiral amine. It was hoped that this could be achieved by semi-rational protein engineering, which typically necessitates high-throughput screening methodologies. Upon commencement of this project there was not a suitable protocol available. Hence, work is reported that outlines the development of a fully

comprehensive, high-throughput amine acceptor screen that works in both liquid phase, and on colonies arrayed on agar. Furthermore, the liquid-phase assay described is atypically both colorimetric and quantitative. This required the utilisation of a novel amine donor, 2-aminoethylaniline (2-AEA), which can be utilised to screen a panel of carbonyl acceptors/ATAs. Work in parallel to produce an amine donor screen using 4-dimethylaminobenzaldehyde as an amine acceptor, allowed for a screening protocol that can determine the enantiopreference of ATAs, which is also highly sensitive and fully quantitative. This assay has enormous potential to simplify and significantly increase the output of ATA evolution endeavours, especially in the absence of expensive robotics platforms. However, a drawback of this methodology is that it is an end-point assay that prevents continuous monitoring of conversion to product. Future work should concentrate on developing a continuous assay, which could be developed by the introduction of an oxygenase to the biotransformation to oxidise indole to indigo, a highly coloured compound.<sup>270</sup> This would significantly reduce the operational complexity of the process.

The fourth chapter outlined the computational modelling of 3 ATAs; 3HMU, HEWT and 3I5T with D-deoxyribose and D-fructose. A hypothesis was developed that ketose sugars were not substrates of these (S)-selective enzymes due to clashes with active site residues that have previously been described as important for determining substrate specificity, particularly those in the small pocket. Four active residues of 3HMU were identified for rational/semi-rational engineering towards D-fructose activity, these were L62, W63, F91 and Y156. Initially, the residue Y156 was rationally engineered to increase the size of the small pocket by replacing the Tyr residue with both Ala

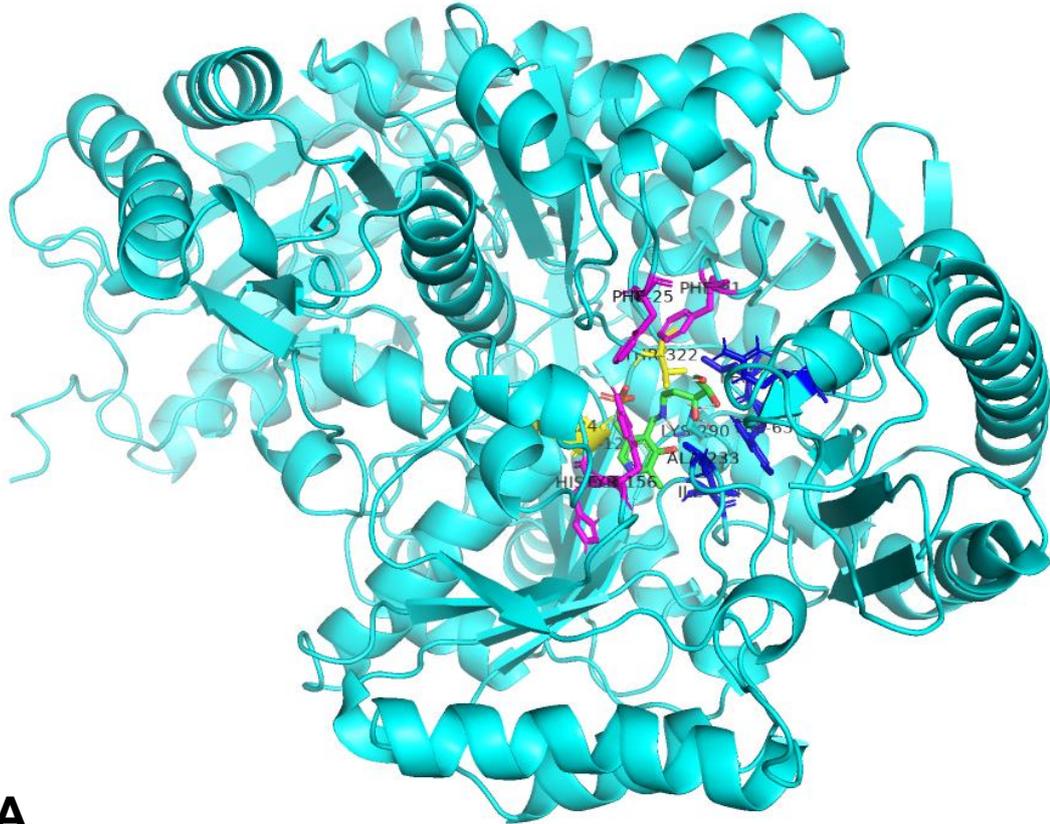
and Asn. This was done in both 3HMU and HEWT. Evidently this residue is essential for determining specificity as activity towards the controls was significantly reduced. However, it appears the mutation Y149A in HEWT allows for the conversion to trace amounts of the amino polyol from D-fructose. Therefore, future work should aim to introduce further mutations at this position, preferably by site saturation mutagenesis.

A combinatorial approach was also attempted, where NDT variant libraries were created for each of the four residues for successive rounds of selection. However, the introduction of nucleotide changes by PCR proved unsuccessful and this work was not explored further due to time constraints. Future work should focus on the optimisation of the PCR protocol by modifying primer design and reaction conditions. This will allow for the screening of variants using the amine acceptor screen developed in chapter 3. At the end of each round, the best variant(s) should be selected, and further rounds of combinatorial mutagenesis can then be applied until a variant with the desirable activity is found.

The final research chapter strays from the previous work and expands the scope of hybrid chemo/enzymatic catalytic cascade processes for the synthesis of valuable targets. This has been demonstrated using ATAs to generate reactive intermediates in situ that undergo further chemistry to generate significant complexity. The successful transamination of cadaverine using the commercially available ATA-256, and a variety of ketones as the amine acceptor, led to the production of  $\Delta^1$ -piperidine that could then further react with the ketone starting material, under the presence of proline, to generate 2-substituted N-heterocycle alkaloids. Conversions of up to 75%

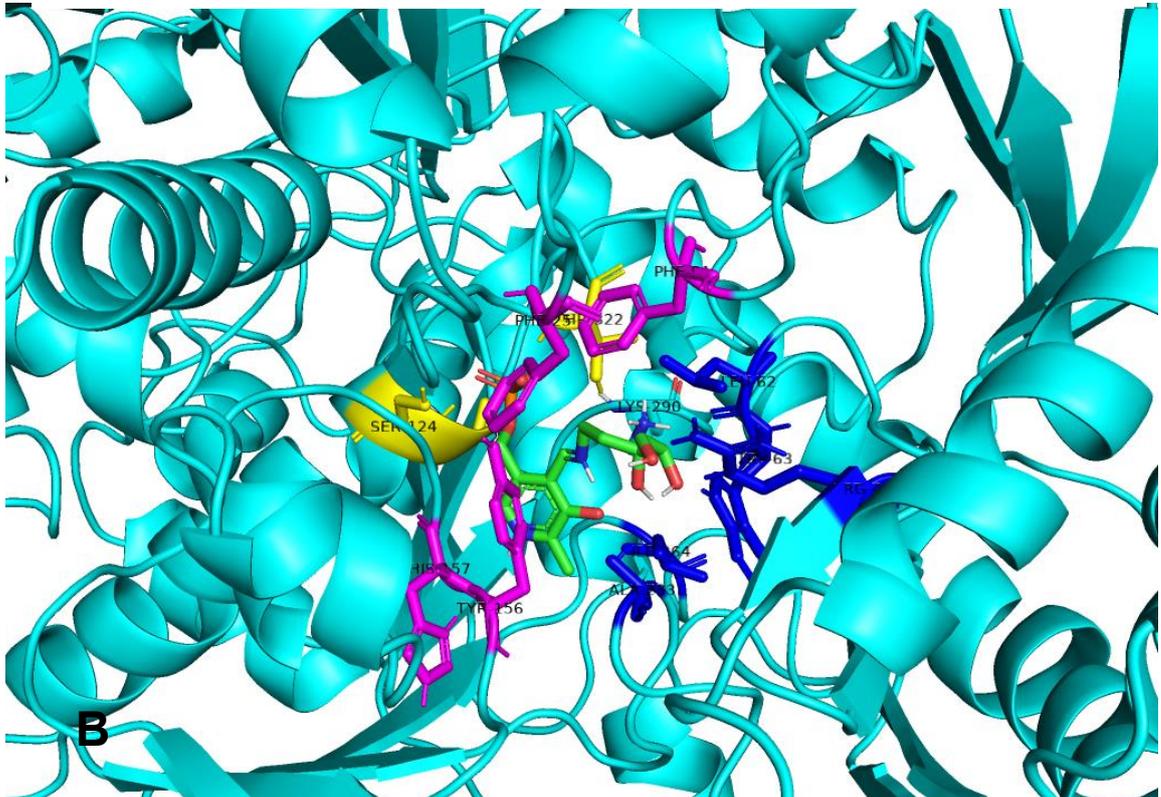
were recorded using this methodology. Using this one-pot chemoenzymatic cascade, a preparative scale up of these reactions enabled access to the natural product pelletierine (60% yield), which further demonstrates the synthetic utility of designing hybrid cascade processes. However, the alkaloids generated were racemic, presumably due to the deprotonation of L-proline in the aqueous environment preventing the transition state hydrogen bond formation critical for stereo control. Future work should aim to investigate alternative organocatalysts that can mediate a stereoselective retro-Michael reaction in water. This has been described using prolinamido-glycosides, where the glucoside moiety acts as a substitute for the carboxyl group.<sup>271,272</sup> Further work should also be done to broaden the substrate scope of this chemistry, particularly using cyclic/aromatic ketones to further increase the complexity of the products.

## 7.0 Appendix



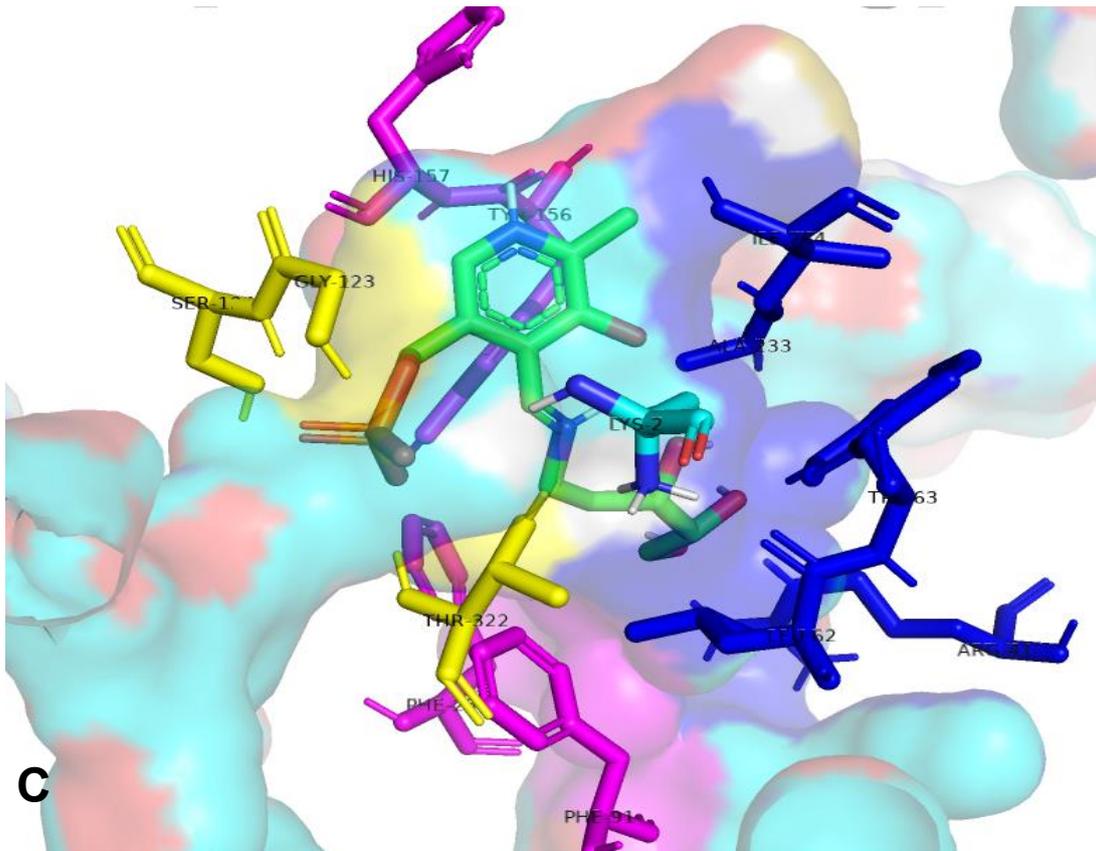
**A**

*Appendix 1 - Computational model of D-deoxyribose in the active site of 3HMU. A) Cartoon of the whole protein with substrate docked in the active site.*



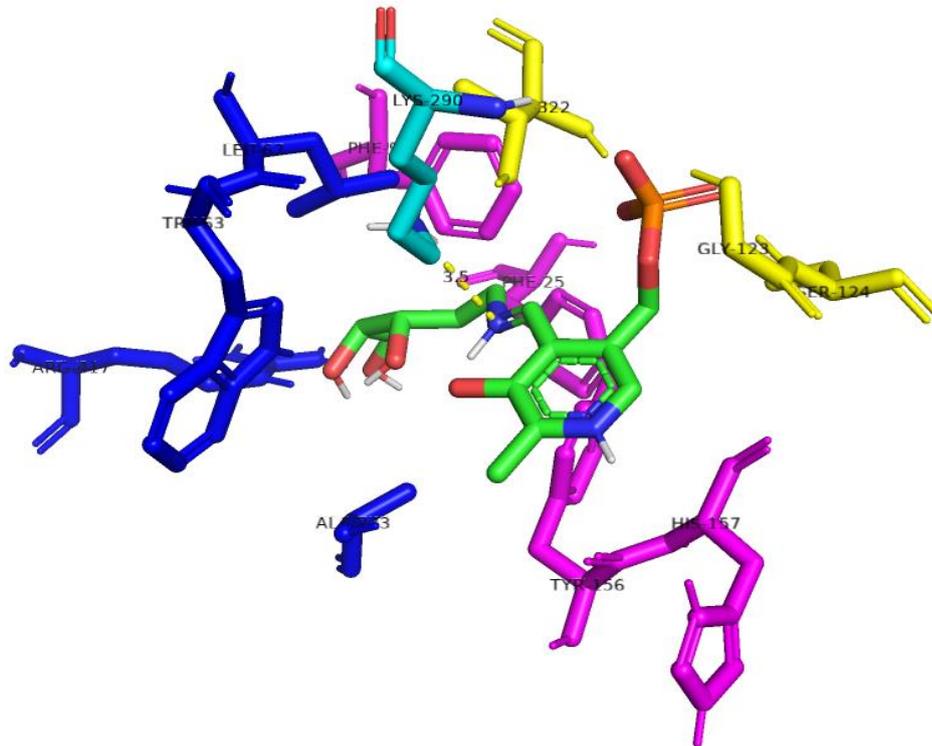
**B**

*Appendix 2 - Computational model of D-deoxyribose in the active site of 3HMU.) Close-up cartoon of the substrate within the active site.*



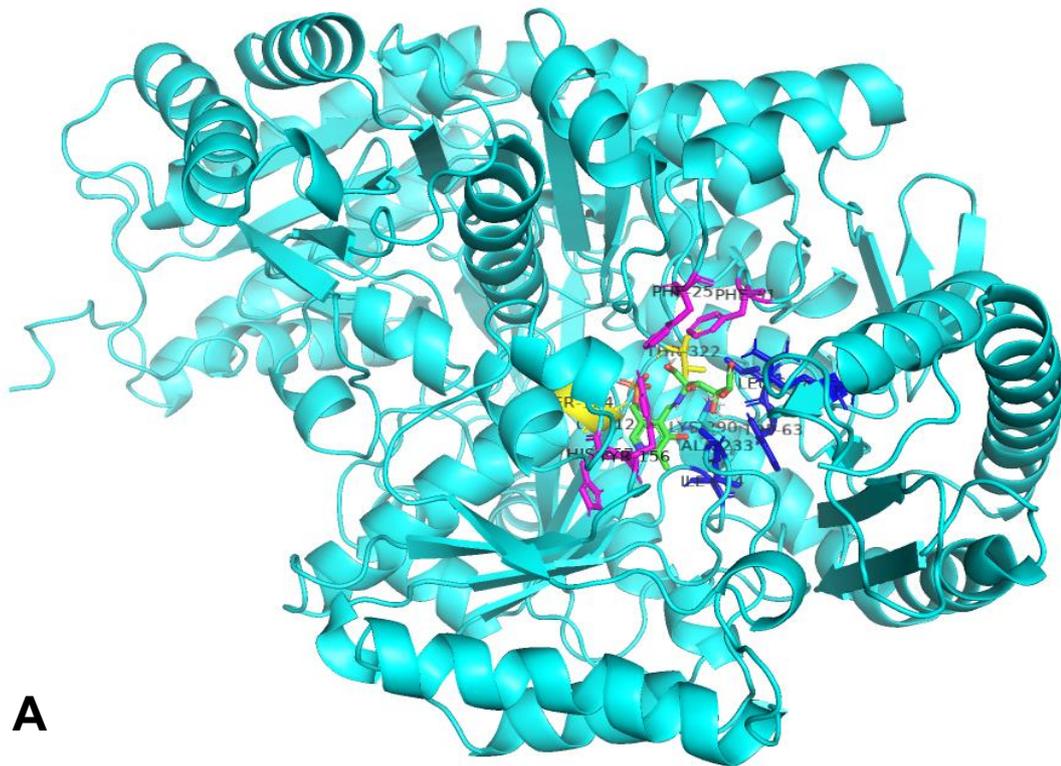
**C**

**Appendix 4** - Computational model of D-deoxyribose in the active site of 3HMU. C) Illustration of the protein surface, with the substrate within the open pocket of the active site.



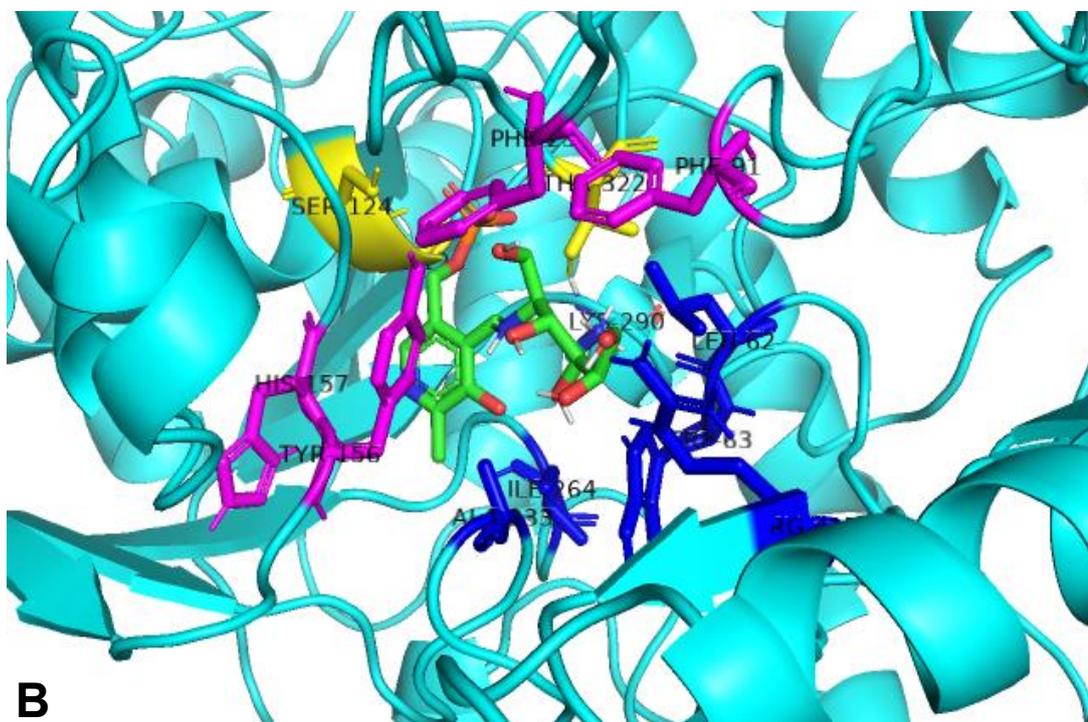
**D**

**Appendix 3** - Computational model of D-deoxyribose in the active site of 3HMU. D) The docked substrate with key active site residues highlighted.



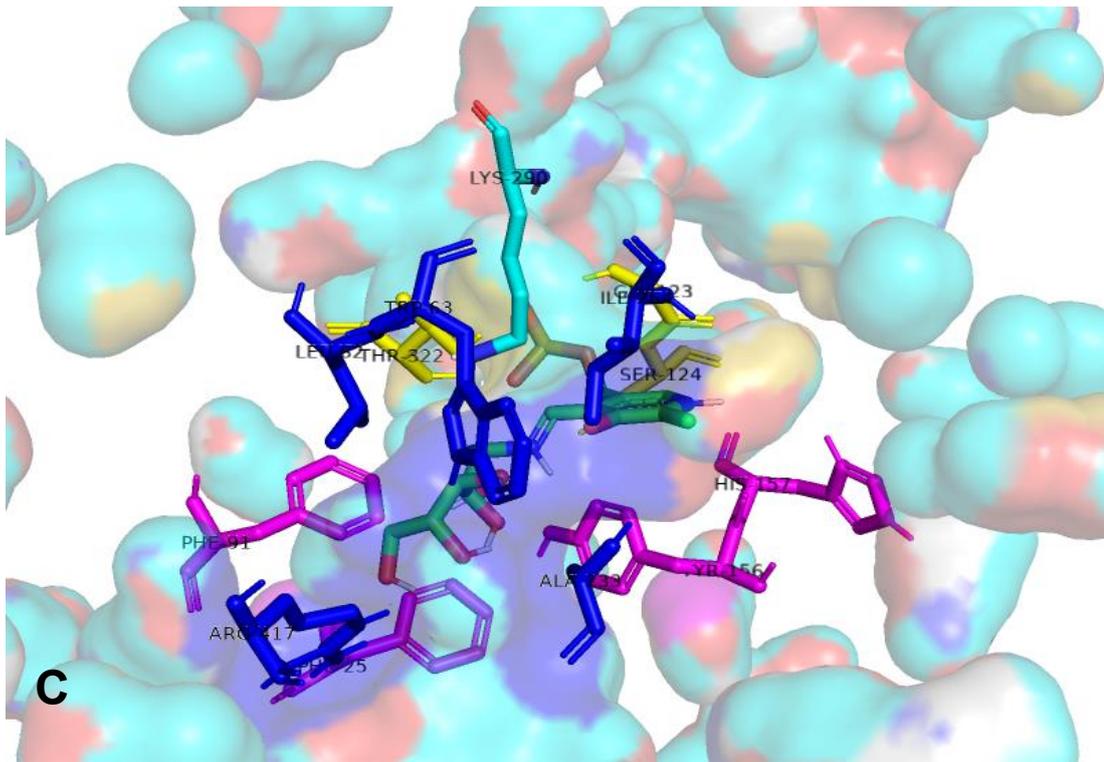
**A**

*Appendix 5 - Computational model of D-fructose in the active site of 3HMU. A) Cartoon of the whole protein with substrate docked in the active site.*

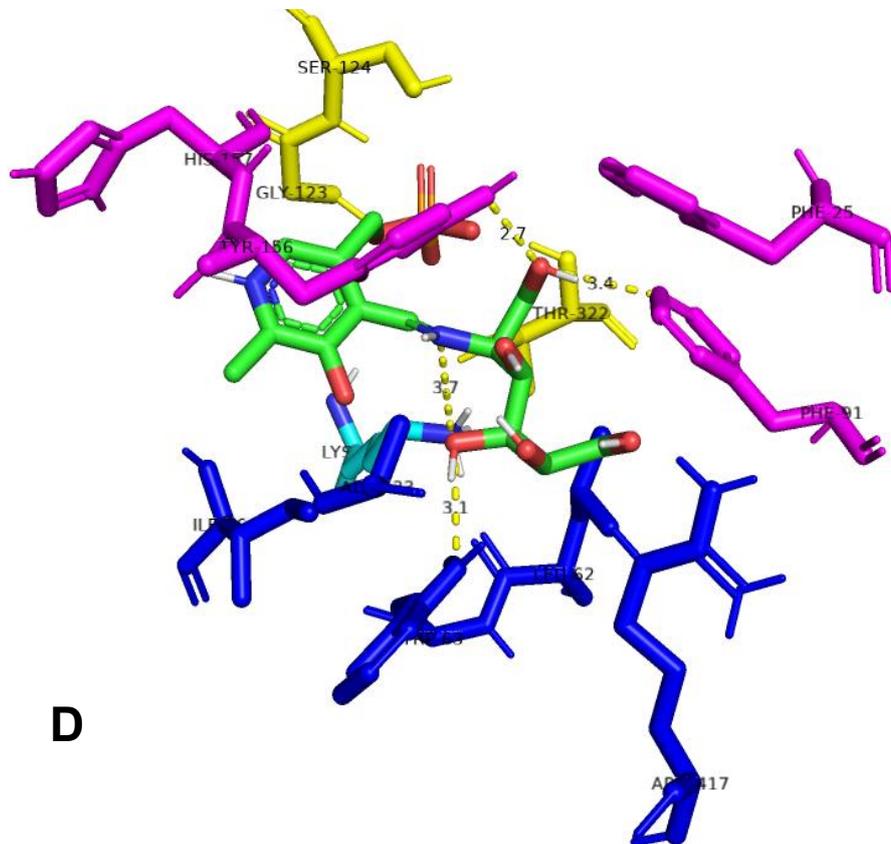


**B**

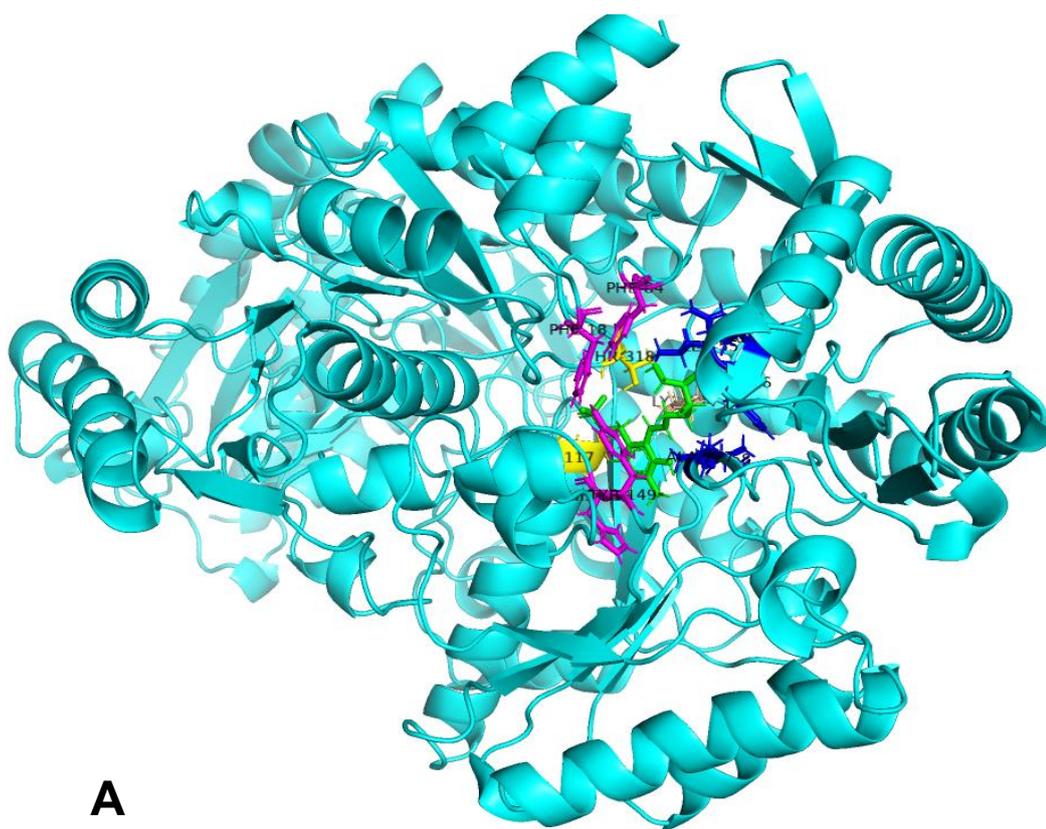
*Appendix 6 - Computational model of D-fructose in the active site of 3HMU. B) Close-up cartoon of the substrate within the active site.*



**Appendix 7** - Computational model of D-fructose in the active site of 3HMUC) Illustration of the protein surface, with the substrate within the open pocket of the active site.

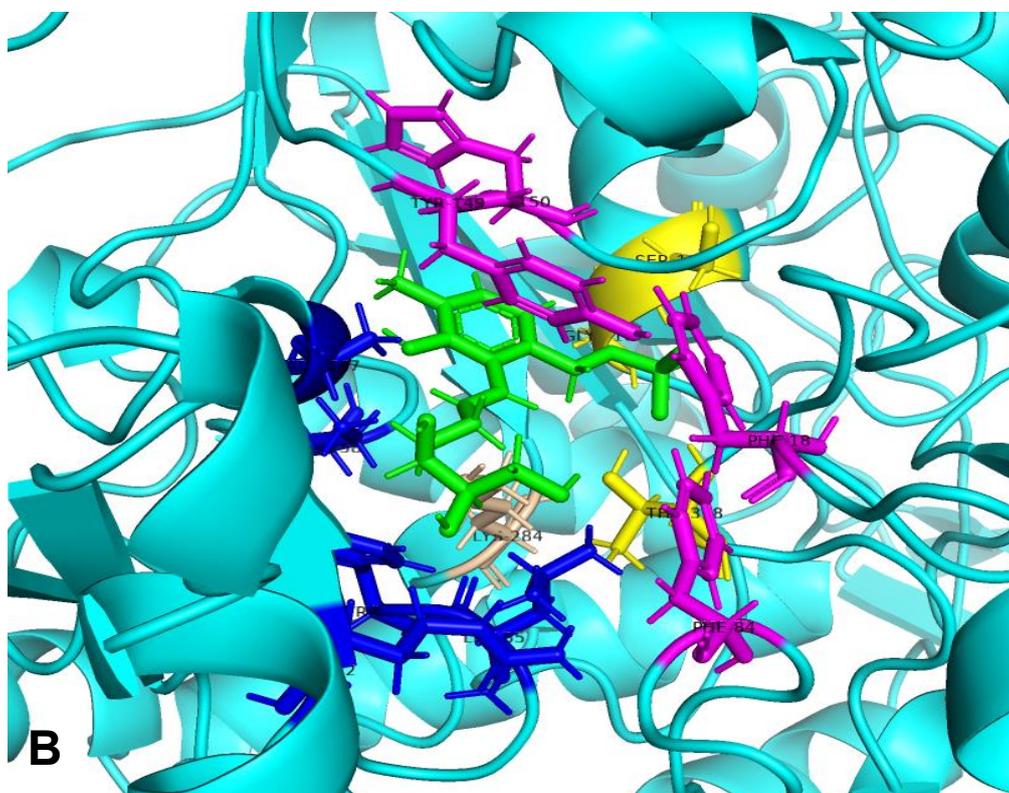


**Appendix 8** - Computational model of D-fructose in the active site of 3HMUD) The docked substrate with key active site residues highlighted.



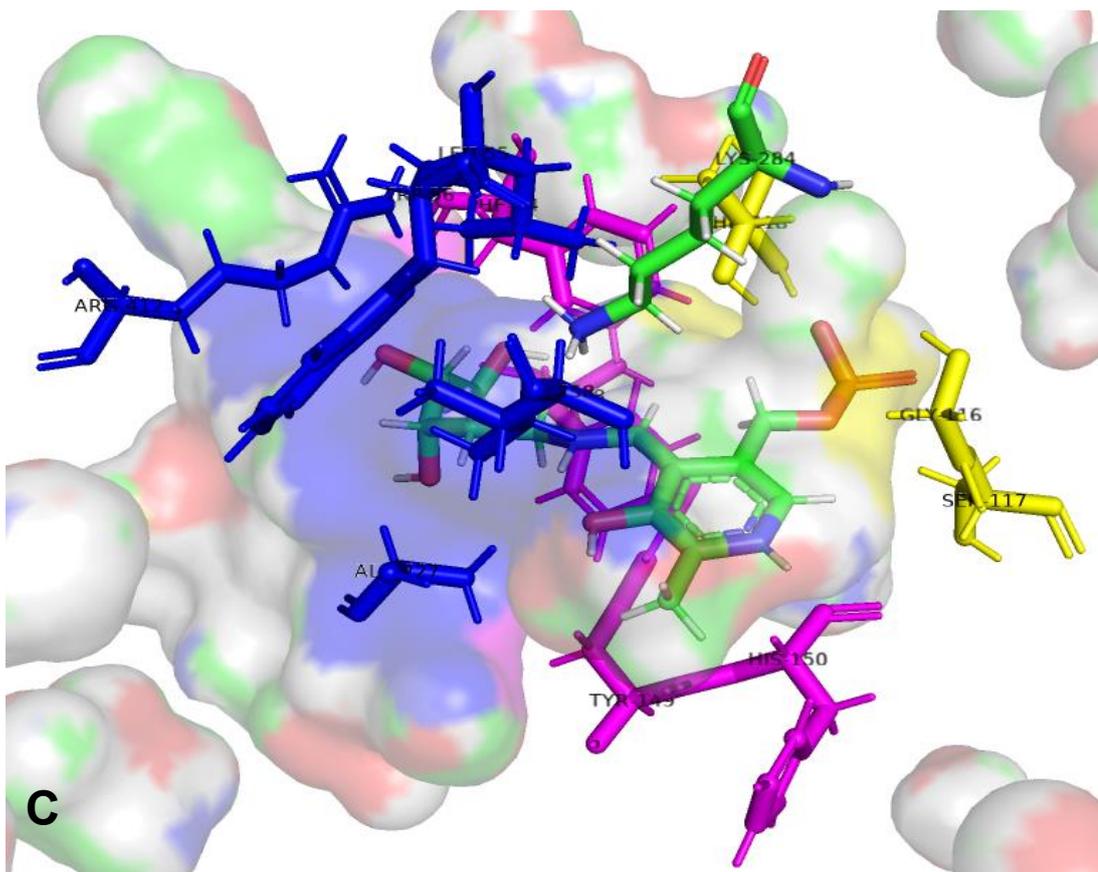
**A**

*Appendix 9 - Computational model of D-deoxyribose in the active site of HEWT. A) Cartoon of the whole protein with substrate docked in the active site.*

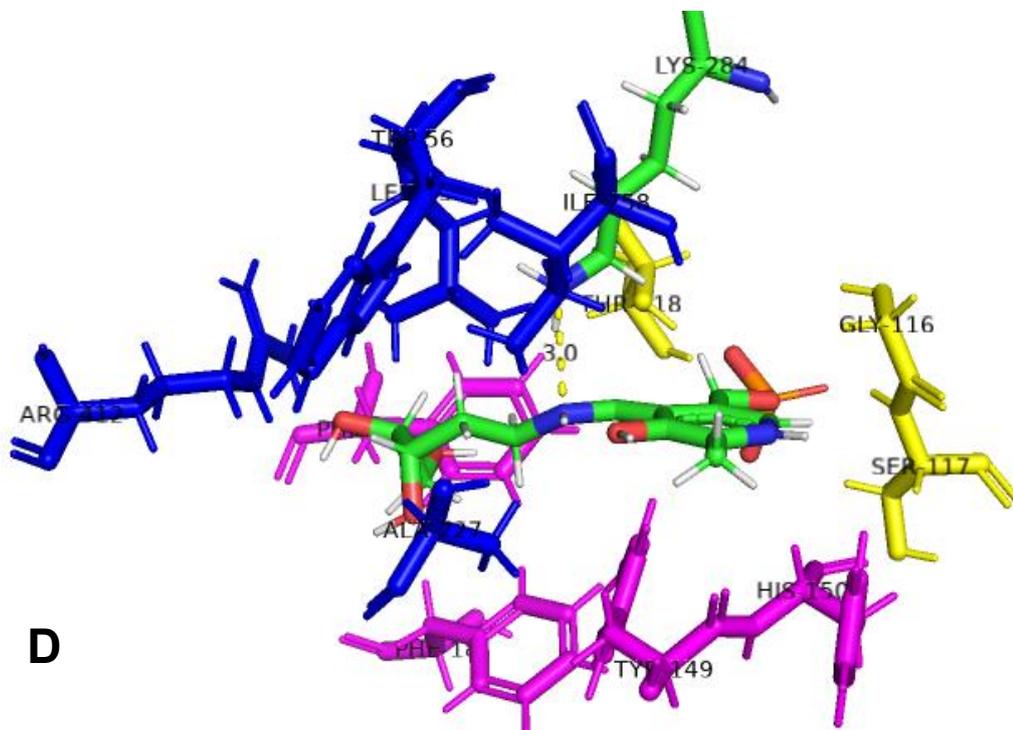


**B**

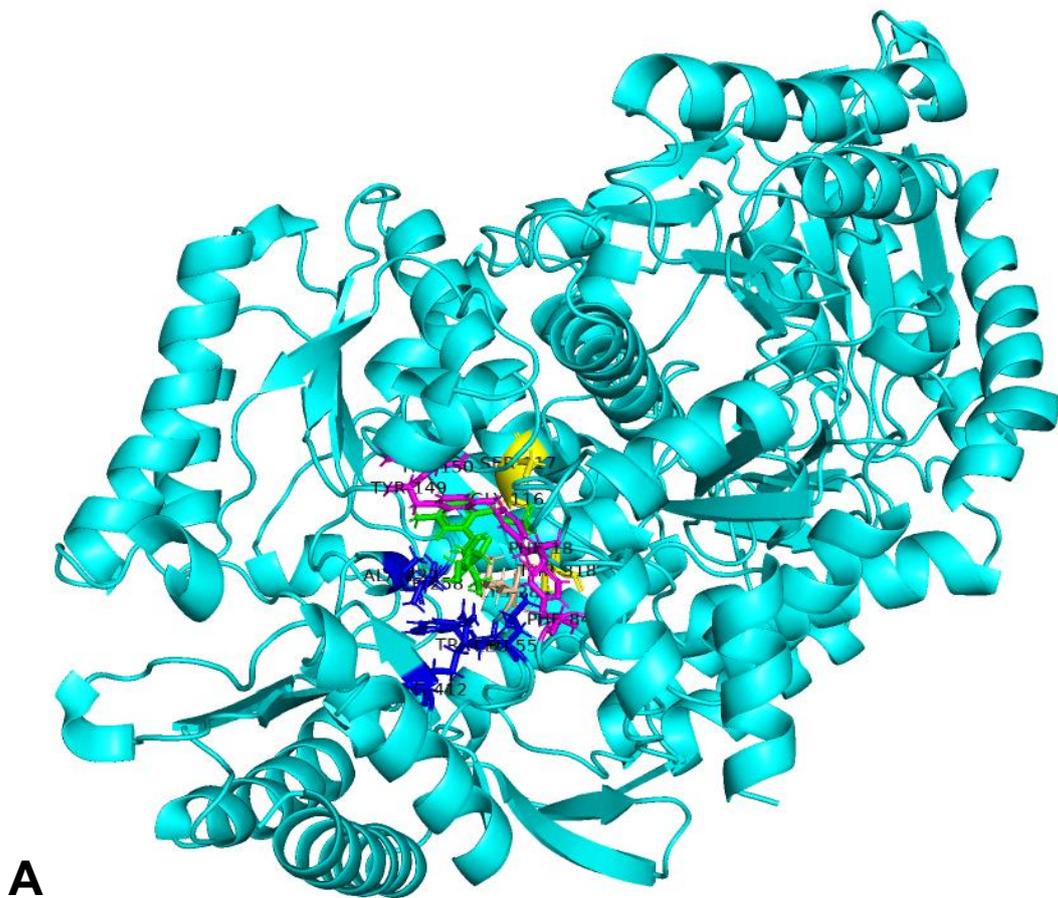
*Appendix 10 - Computational model of D-deoxyribose in the active site of HEWT. B) Close-up cartoon of the substrate within the active site.*



**Appendix 12** - Computational model of D-deoxyribose in the active site of HEWT. C) Illustration of the protein surface, with the substrate within the open pocket of the active site.

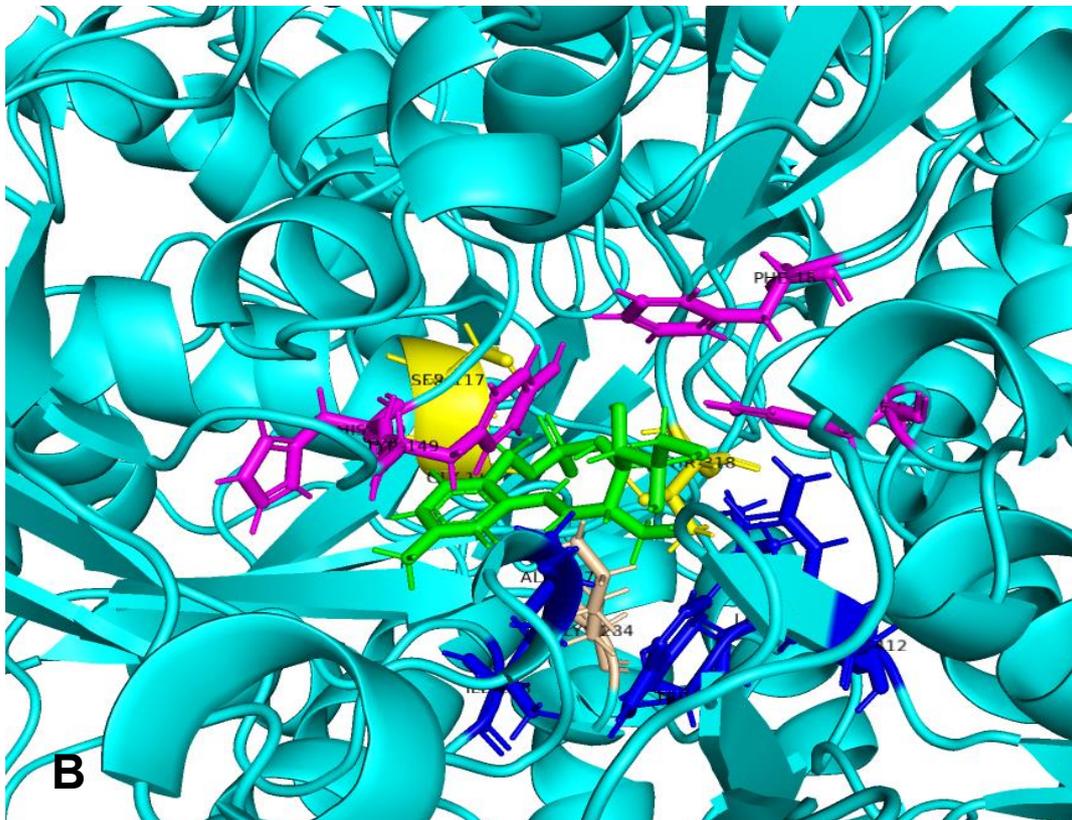


**Appendix 11** - Computational model of D-deoxyribose in the active site of HEWT. D) The docked substrate with key active site residues highlighted.



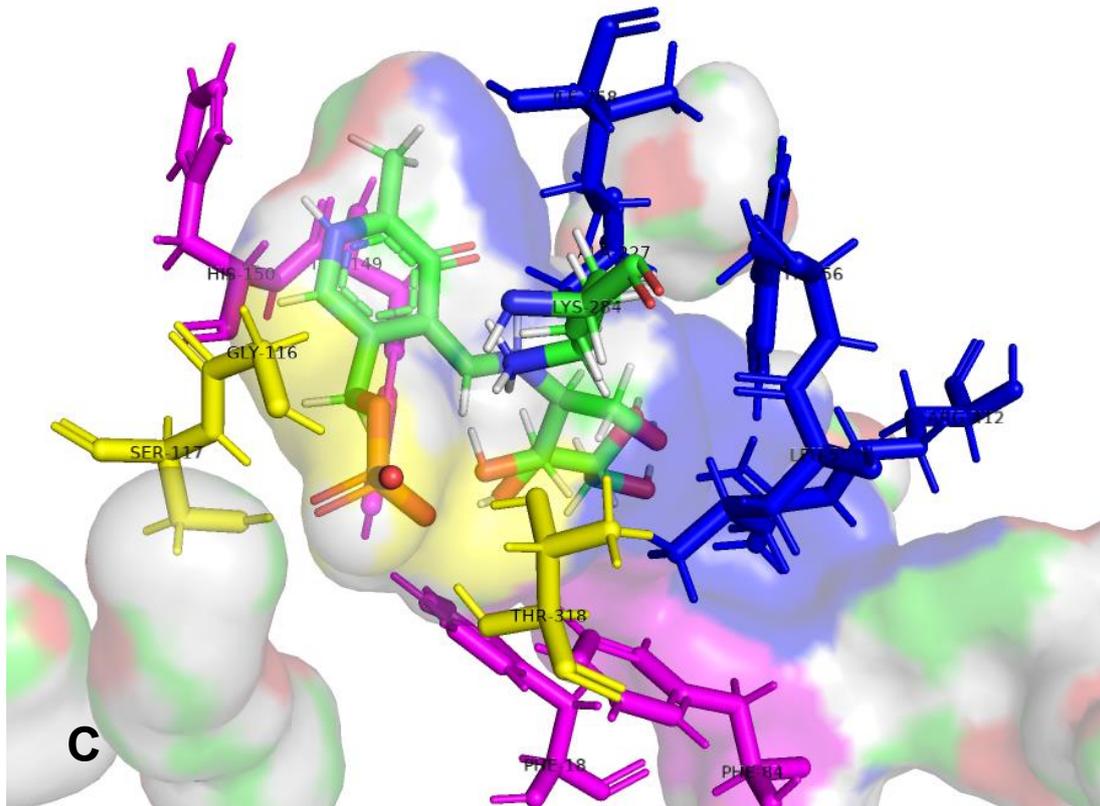
**A**

*Appendix 13 - Computational model of D-fructose in the active site of HEWT. A) Cartoon of the whole protein with substrate docked in the active site.*

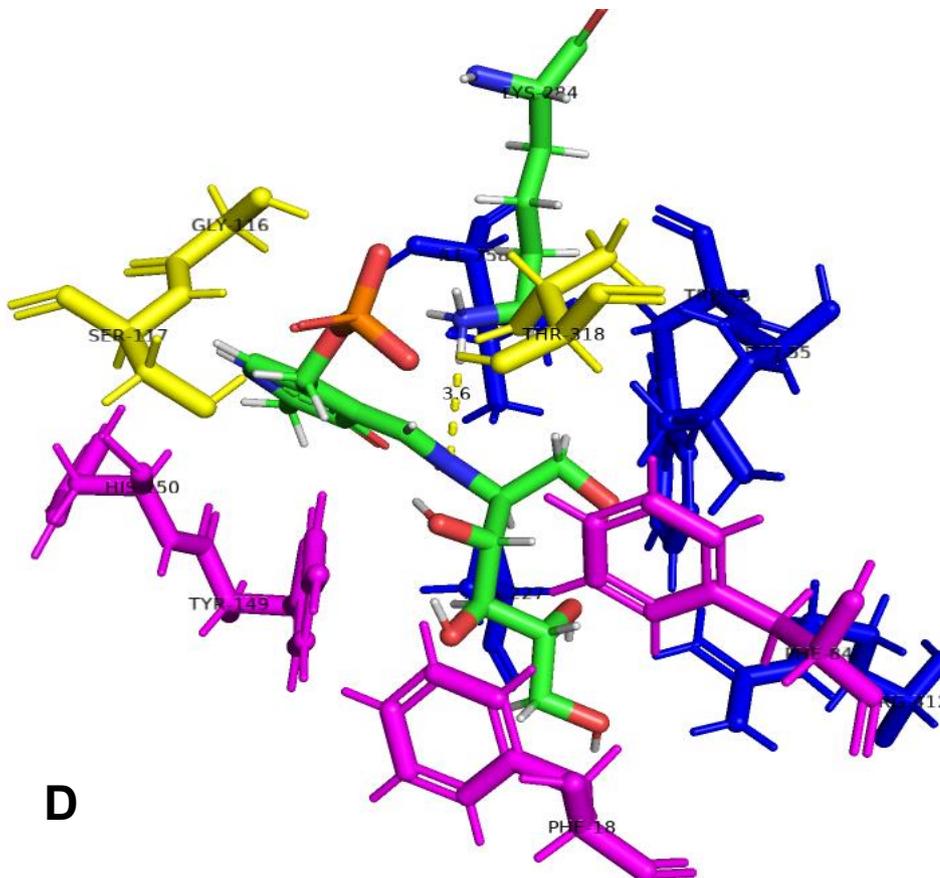


**B**

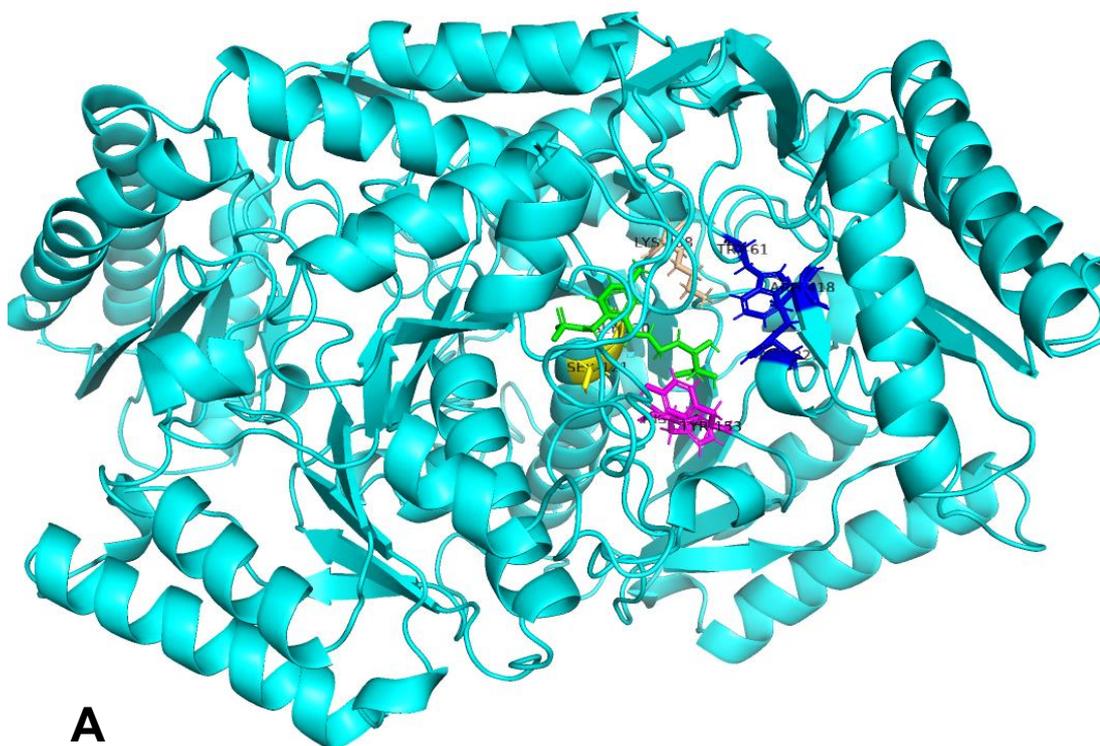
*Appendix 14 - Computational model of D-fructose in the active site of HEWT. B) Close-up cartoon of the substrate within the active site.*



**Appendix 16** - Computational model of *D*-fructose in the active site of HEWT. C) Illustration of the protein surface, with the substrate within the open pocket of the active site.

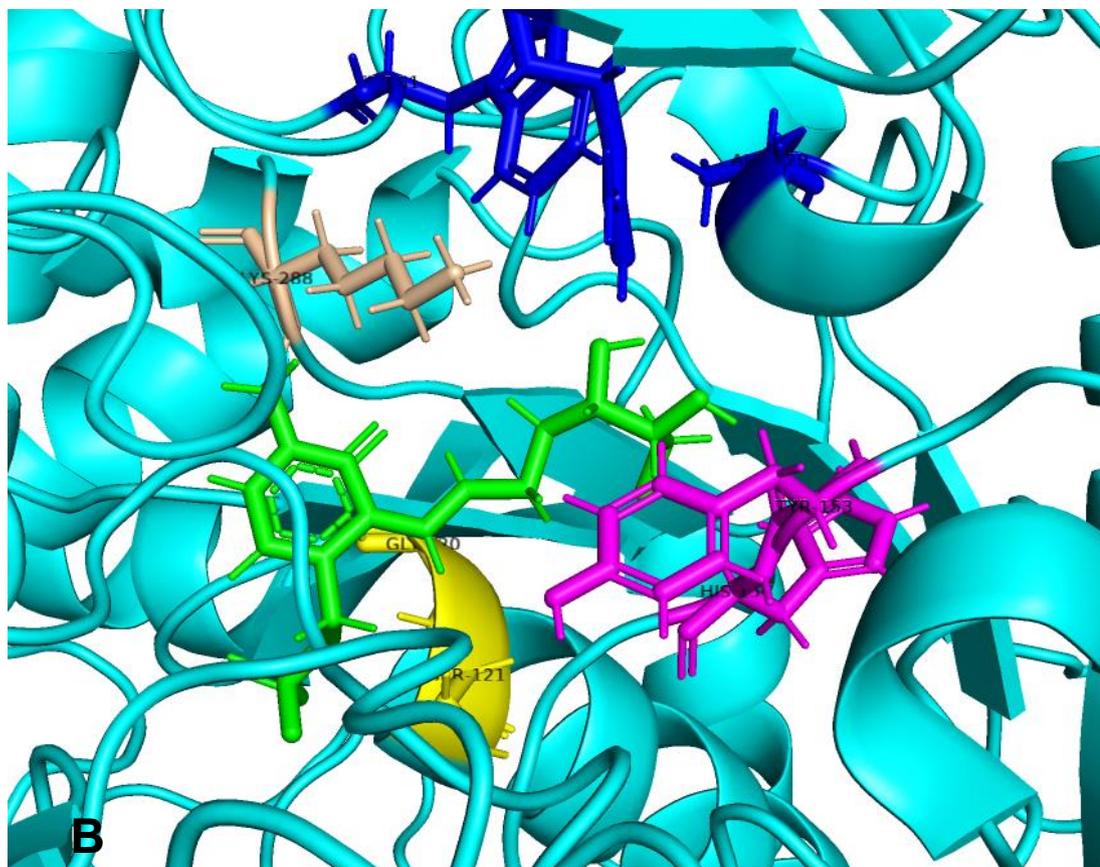


**Appendix 15** - Computational model of *D*-fructose in the active site of HEWT. D) The docked substrate with key active site residues highlighted.



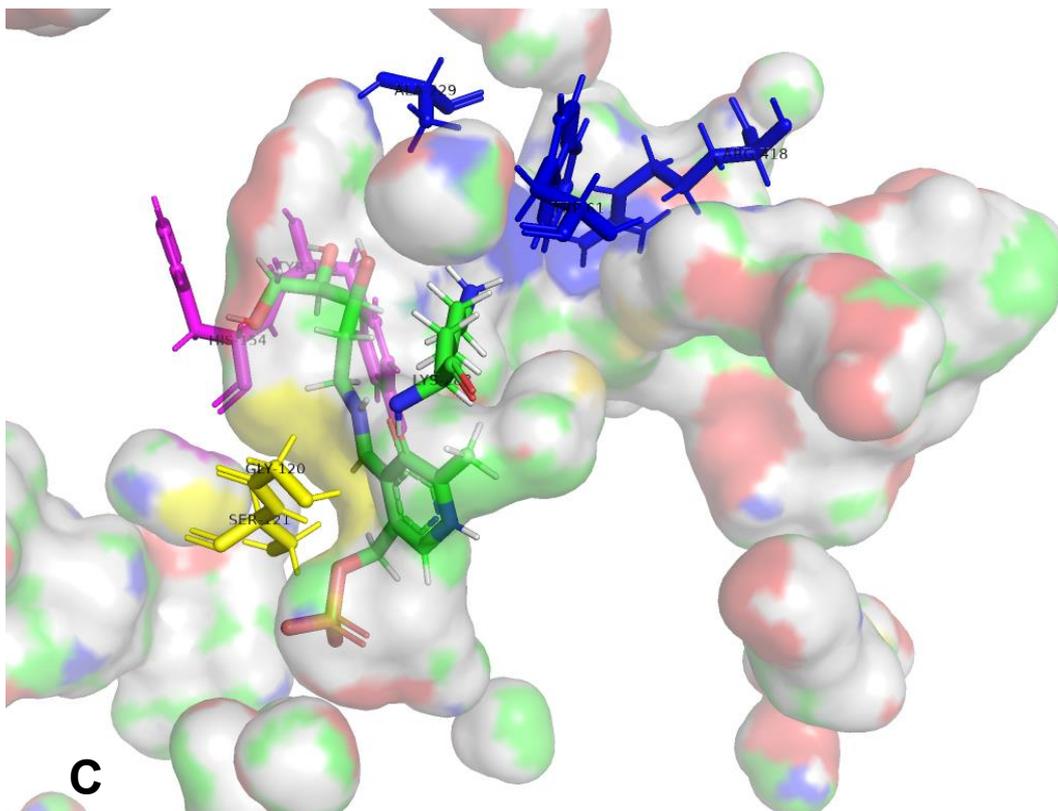
**A**

*Appendix 17 - Computational model of D-deoxyribose in the active site of 3I5T. A) Cartoon of the whole protein with substrate docked in the active site*

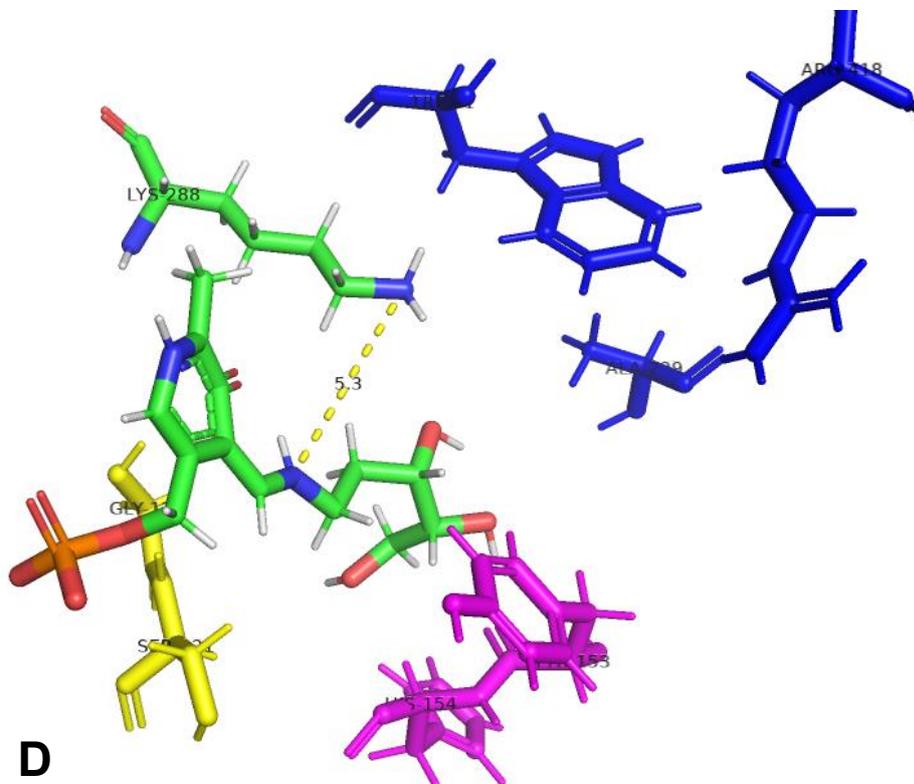


**B**

*Appendix 18 - Computational model of D-deoxyribose in the active site of 3I5T. B) Close-up cartoon of the substrate within the active site.*



**Appendix 20** - Computational model of D-deoxyribose in the active site of 3I5T. C) Illustration of the protein surface, with the substrate within the open pocket of the active site.



**Appendix 19** - Computational model of D-deoxyribose in the active site of 3I5T. D) The docked substrate with key active site residues highlighted.

**Appendix 21** – Public link to downloadable interactive active Pymol sessions of all molecular models described:

[https://uniofnottm-my.sharepoint.com/:f/g/person/ryan\\_cairns\\_nottingham\\_ac\\_uk/EoG9rbEP7vVKpAqiiEodCk4BKMWeRXvBZ6prZWSABjz5KQ?e=CjacNw](https://uniofnottm-my.sharepoint.com/:f/g/person/ryan_cairns_nottingham_ac_uk/EoG9rbEP7vVKpAqiiEodCk4BKMWeRXvBZ6prZWSABjz5KQ?e=CjacNw)

<b>Strain</b>	<b>Relevant characteristic</b>	<b>Reference or source</b>
<i>E. coli</i> DH5a	F <sup>-</sup> thi-1 endA1 hsdR17 (r m) supE44 DlacU169 (F80lacZDM15) recA1 gyrA96 relA1	New Englands Biolabs
<i>E. coli</i> BL21	F <sup>-</sup> ompT gal dcm lon hsdS <sub>B</sub> (r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB <sup>+</sup> ] <sub>k-12</sub> (λ <sup>S</sup> )	New Englands Biolabs
<i>C. glutamicum</i> ATCC 13032 (Wild type)	Wild type strains naturally overproducing L-lysine	American Type Culture Collection
<i>C. glutamicum</i> GRLys1	WT carrying the pVWEx1-ldcC plasmid	(189)

*Appendix 22 - Details of strains used.*

## 8.0 References

- 1 G. Hughes and J. C. Lewis, *Chem. Rev.*, 2018, **118**, 1–3.
- 2 J. M. Woodley, *Curr. Opin. Green Sustainable Chem.*, 2020, **21**, 22–26.
- 3 S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius and U. T. Bornscheuer, *Angew. Chem. Int. Ed.*, 2021, **60**, 88–119.
- 4 C. M. Heckmann and F. Paradisi, *ChemCatChem*, 2020, **12**, 6082–6102.
- 5 R. A. Sheldon and J. M. Woodley, *Chem. Rev.*, 2018, **118**, 801–838.
- 6 H. Renata, Z. J. Wang and F. H. Arnold, *Angew. Chem. Int. Ed.*, 2015, **54**, 3351–3367.
- 7 A. S. Wells, G. L. Finch, P. C. Michels and J. W. Wong, *Org. Process Res. Dev.*, 2012, **16**, 1986–1993.
- 8 S. Raveendran, B. Parameswaran, S. B. Ummalyima, A. Abraham, A. K. Mathew, A. Madhavan, S. Rebello and A. Pandey, *Food Technol. Biotechnol.*, 2018, **56**, 16–30.
- 9 H. P. Meyer, E. Eichhorn, S. Hanlon, S. Lütz, M. Schürmann, R. Wohlgemuth and R. Coppolecchia, *Catal. Sci. Technol.*, 2013, **3**, 29–40.
- 10 M. Hosseini, *Advanced Bioprocessing for Alternative Fuels, Biobased Chemicals, and Bioproducts*, 2019.
- 11 H. P. Austin, M. D. Allen, B. S. Donohoe, N. A. Rorrer, F. L. Kearns, R. L. Silveira, B. C. Pollard, G. Dominick, R. Duman, K. El Omari, V. Mykhaylyk, A. Wagner, W. E. Michener, A. Amore, M. S. Skaf, M. F. Crowley, A. W. Thorne, C. W. Johnson, H. Lee Woodcock, J. E. McGeehan and G. T. Beckham, *Proc. Natl. Acad. Sci. U.S.A.*, 2018, **115**, E4350–E4357.
- 12 V. Tournier, C. M. Topham, A. Gilles, B. David, C. Folgoas, E. Moya-Leclair, E. Kamionka, M.-L. Desrousseaux, H. Texier, S. Gavalda, M. Cot, E. Guémard, M. Dalibey, J. Nomme, G. Cioci, S. Barbe, M. Chateau, I. André, S. Duquesne and A. Marty, *Nature*, 2020, **580**, 216–219.
- 13 Knowledge Sourcing Intelligence LLP, *Specialty Enzymes Market Size: Industry Report, 2019 - 2024*, 2020.
- 14 L. A. Nguyen, H. He and C. Pham-Huy, *Int. J. Biomed. Sci.*, 2006, **2**, 85–100.
- 15 G. Mion and T. Villevieille, *CNS Neurosci. Ther.*, 2013, **19**, 370–380.
- 16 T. C. Nugent, *Chiral amine synthesis : methods, developments and applications*, Wiley-VCH, 2010.
- 17 D. J. C. Constable, P. J. Dunn, J. D. Hayler, G. R. Humphrey, J. L. Leazer, R. J. Linderman, K. Lorenz, J. Manley, B. A. Pearlman, A. Wells, A. Zaks and T. Y. Zhang, *Green Chem.*, 2007, **9**, 411–420.
- 18 J. P. Adams, M. J. B. Brown, A. Diaz-Rodriguez, R. C. Lloyd and G. D. Roiban, *Adv. Synth. Catal.*, 2019, **361**, 2421–2432.
- 19 M. D. Truppo, *ACS Med. Chem. Lett.*, 2017, **8**, 476–480.
- 20 X. Huang, M. Cao and H. Zhao, *Curr. Opin. Chem. Biol.*, 2020, **55**, 161–170.

- 21 M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. McLaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M. D. Truppo, A. M. Whittaker, D. Verma, L. Xiao, Y. Xu and H. Yang, *Science*, 2019, **366**, 1255–1259.
- 22 J. H. Schrittwieser, S. Velikogne, M. Hall and W. Kroutil, *Chem. Rev.*, 2018, **118**, 270–348.
- 23 S. P. France, L. J. Hepworth, N. J. Turner and S. L. Flitsch, *ACS Catal.*, 2017, **7**, 710–724.
- 24 S. Gandomkar, A. Żądło-Dobrowolska and W. Kroutil, *ChemCatChem*, 2019, **11**, 225–243.
- 25 R. Abu and J. M. Woodley, *ChemCatChem*, 2015, **7**, 3094–3105.
- 26 B. A. Sandoval and T. K. Hyster, *Curr. Opin. Chem. Biol.*, 2020, **55**, 45–51.
- 27 A. S. Bommarius, *Annu. Rev. Chem. Biomol. Eng.*, 2015, **6**, 319–345.
- 28 V. Uppada, S. Bhaduri and S. B. Noronha, *Current Science*, 2014, **106**, 946–957.
- 29 S. Kara, J. H. Schrittwieser, F. Hollmann and M. B. Ansorge-Schumacher, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 1517–1529.
- 30 B. Lin and Y. Tao, *Microb. Cell Fact.*, 2017, **16**, 106.
- 31 R. Yuryev, S. Strompen and A. Liese, *Beilstein J. Org. Chem.*, 2011, **7**, 1449–1467.
- 32 J. Britton, S. Majumdar and G. A. Weiss, *Chem. Soc. Rev.*, 2018, **47**, 5891–5918.
- 33 E. Buchner, *Ber. Dtsch. Chem. Ges.*, 1897, **30**, 117–124.
- 34 C. Wandrey, *World Biotechnol. Rep.*, 1984, **1**, 391–404.
- 35 C. Wandrey, E. Fiolitakas, U. Wichmann and M. -R Kula, *Ann. N. Y. Acad. Sci.*, 1984, **434**, 091–094.
- 36 V. Köhler, Y. M. Wilson, M. Dürrenberger, D. Ghislieri, E. Churakova, T. Quinto, L. Knörr, D. Häussinger, F. Hollmann, N. J. Turner and T. R. Ward, *Nat. Chem.*, 2013, **5**, 93–99.
- 37 N. Oberleitner, A. K. Ressmann, K. Bica, P. Gärtner, M. W. Fraaije, U. T. Bornscheuer, F. Rudroff and M. D. Mihovilovic, *Green Chem.*, 2017, **19**, 367–371.
- 38 A. Brahma, B. Musio, U. Ismayilova, N. Nikbin, S. B. Kamptmann, P. Siegert, G. E. Jeromin, S. V. Ley and M. Pohl, *Synlett*, 2016, **27**, 262–266.
- 39 R. O. M. A. de Souza, L. S. M. Miranda and U. T. Bornscheuer, *Chem. Eur. J.*, 2017, **23**, 12040–12063.
- 40 A. P. Green and N. J. Turner, *Perspect. Sci.*, 2016, **9**, 42–48.
- 41 N. J. Turner and E. O’Reilly, *Nat. Chem. Biol.*, 2013, **9**, 285–288.
- 42 E. J. Corey, *Angew. Chem. Int. Ed.*, 1991, **30**, 455–465.
- 43 J. Barluenga, B. Baragana and J. M. Concellon, *J. Org. Chem.*, 2002, **60**, 6696–6699.

- 44 A. S. M. De Miranda, R. C. Simon, B. Grischek, G. C. De Paula, B. A. C. Horta, L. S. M. De Miranda, W. Kroutil, C. O. Kappe and R. O. M. A. De Souza, *ChemCatChem*, 2015, **7**, 984–992.
- 45 G. Horne, F. X. Wilson, J. Tinsley, D. H. Williams and R. Storer, *Drug Discov. Today*, 2011, **16**, 107–118.
- 46 R. J. Nash, A. Kato, C. Y. Yu and G. W. Fleet, *Future Med. Chem.*, 2011, **3**, 1513–1521.
- 47 US Pat., US6683098B1, 2000.
- 48 US Pat., US9079856B2, 2012.
- 49 Y. SUZUKI, *Proc. Jpn. Acad., Ser. B, Phys. Biol.*, 2014, **90**, 145.
- 50 A. Esposito, D. D'Alonzo, M. de Fenza, E. de Gregorio, A. Tamanini, G. Lippi, M. C. Dechecchi and A. Guaragna, *Int. J. Mol. Sci.*, 2020, **21(9)**, 3353–3367.
- 51 J.-Q. Fan, *Iminosugars: from synthesis to therapeutic applications*, J. Wiley, Weinheim, Germany, 2007.
- 52 F. Hossain and P. R. Andreana, *Pharmaceuticals*, 2019, **12**, 84–102.
- 53 T. Wrodnigg, A. Steiner and B. Ueberbacher, *Anti-Cancer Agents Med. Chem.*, 2008, **8**, 77–85.
- 54 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.
- 55 L. A. Sorbera, J. Castañer and L. García-Capdevila, *Drugs Future*, 2005, **30**, 545–552.
- 56 D. S. Alonzi, K. A. Scott, R. A. Dwek and N. Zitzmann, *Biochem. Soc. Trans.*, 2017, **45**, 571–582.
- 57 G. Horne, *Curr. Top. Med. Chem.*, 2014, **12**, 23–52.
- 58 S. V. Glavey, D. Huynh, M. R. Reagan, S. Manier, M. Moschetta, Y. Kawano, A. M. Roccaro, I. M. Ghobrial, L. Joshi and M. E. O'Dwyer, *Blood Rev.*, 2015, **29**, 269–279.
- 59 R. Saul, R. J. Molyneux and A. D. Elbein, *Arch. Biochem. Biophys.*, 1984, **230**, 668–675.
- 60 K. Whitby, T. C. Pierson, B. Geiss, K. Lane, M. Engle, Y. Zhou, R. W. Doms and M. S. Diamond, *J. Virol.*, 2005, **79**, 8698–8706.
- 61 K. Whitby, D. Taylor, D. Patel, P. Ahmed and A. S. Tyms, *Antiviral Chem. Chemother.*, 2004, **15**, 141–151.
- 62 A. A. Watson, G. W. J. Fleet, N. Asano, R. J. Molyneux and R. J. Nash, *Phytochemistry*, 2001, **56**, 265–295.
- 63 M. Yagi, T. Kouno, Y. Aoyagi and H. Murai, *Nippon Nogei Kagaku Kaishi*, 1976, **50**, 571–572.
- 64 M. Shibano, Y. Fujimoto, K. Kushino, G. Kusano and K. Baba, *Phytochemistry*, 2004, **65**, 2661–2665.
- 65 S. Onose, R. Ikeda, K. Nakagawa, T. Kimura, K. Yamagishi, O. Higuchi and T. Miyazawa, *Food Chem.*, 2013, **138**, 516–523.
- 66 R. Zelli, J. F. Longevial, P. Dumy and A. Marra, *New J. Chem.*, 2015, **39**, 5050–5074.

- 67 F. Massicot, R. Plantier-Royon, J. L. Vasse and J. B. Behr, *Carbohydr. Res.*, 2018, **464**, 2–7.
- 68 P. Compain, V. Chagnault and O. R. Martin, *Tetrahedron Asymmetry*, 2009, **20**, 672–711.
- 69 T. Shintani, *Fermentation*, 2019, **5(2)**, 47–51.
- 70 H. Kohls, F. Steffen-Munsberg and M. Höhne, *Curr. Opin. Chem. Biol.*, 2014, **19**, 180–192.
- 71 W. Kroutil, H. Mang, K. Edegger and K. Faber, *Advanced Synthesis & Catalysis*, 2004, **346**, 125–142.
- 72 Y. Zhang, D. R. Wu, D. B. Wang-Iverson and A. A. Tymiak, *Drug Discov. Today*, 2005, **10**, 571–577.
- 73 J. Lee, Y. K. Choi, J. Park and M. J. Kim, *Green Biocatalysis*, wiley, 2016, pp. 115–147.
- 74 D. Ghislieri and N. J. Turner, *Top. Catal.*, 2014, **57**, 284–300.
- 75 E. O’Reilly and N. J. Turner, *Perspect. Sci.*, 2015, **4**, 55–61.
- 76 G. Grogan, *Curr. Opin. Chem. Biol.*, 2018, **43**, 15–22.
- 77 J. Duan, B. Li, Y. Qin, Y. Dong, J. Ren and G. Li, *Bioresour. Bioprocess.*, 2019, **6**, 1–11.
- 78 V. F. Batista, J. L. Galman, D. C. Pinto, A. M. S. Silva and N. J. Turner, *ACS Catal.*, 2018, **8**, 11889–11907.
- 79 E. O’Reilly, C. Iglesias, D. Ghislieri, J. Hopwood, J. L. Galman, R. C. Lloyd and N. J. Turner, *Angew. Chem.*, 2014, **126**, 2479–2482.
- 80 D. Ghislieri, A. P. Green, M. Pontini, S. C. Willies, I. Rowles, A. Frank, G. Grogan and N. J. Turner, *J. Am. Chem. Soc.*, 2013, **135**, 10863–10869.
- 81 C. J. Dunsmore, R. Carr, T. Fleming and N. J. Turner, *J. Am. Chem. Soc.*, 2006, **128**, 2224–2225.
- 82 S. Herter, F. Medina, S. Wagschal, C. Benhaïm, F. Leipold and N. J. Turner, *Bioorg. Med. Chem.*, 2018, **26**, 1338–1346.
- 83 T. Li, J. Liang, A. Ambrogelly, T. Brennan, G. Gloor, G. Huisman, J. Lalonde, A. Lekhal, B. Mijts, S. Muley, L. Newman, M. Tobin, G. Wong, A. Zaks and X. Zhang, *J. Am. Chem. Soc.*, 2012, **134**, 6467–6472.
- 84 M. Sharma, J. Mangas-Sanchez, N. J. Turner and G. Grogan, *Adv. Synth. Catal.*, 2017, **359**, 2011–2025.
- 85 T. Knaus, W. Böhmer and F. G. Mutti, *Green Chem.*, 2017, **19**, 453–463.
- 86 G. P. Anderson, *Life Sci.*, 1993, **52**, 2145–2160.
- 87 F. Campos, M. P. Bosch and A. Guerrero, *Tetrahedron Asymmetry*, 2000, **11**, 2705–2717.
- 88 C. Lu, Q. Zhou, J. Yan, Z. Du, L. Huang and X. Li, *Eur. J. Med. Chem.*, 2013, **62**, 745–753.
- 89 J. Mangas-Sanchez, S. P. France, S. L. Montgomery, G. A. Aleku, H. Man, M. Sharma, J. I. Ramsden, G. Grogan and N. J. Turner, *Curr. Opin. Chem. Biol.*, 2017, **37**, 19–25.

- 90 M. Lenz, N. Borlinghaus, L. Weinmann and B. M. Nestl, *World J. Microbiol. Biotechnol.*, 2017, **33**, 199.
- 91 N. Borlinghaus and B. M. Nestl, *ChemCatChem*, 2018, **10**, 183–187.
- 92 P. Yao, Z. Xu, S. Yu, Q. Wu and D. Zhu, *Adv. Synth. Catal.*, 2019, **361**, 556–561.
- 93 S. Gagliardi, G. Dondio and G. A. M. Giardina, *IDrugs : the investigational drugs journal*, 2003, **6**, 129–37.
- 94 G. A. Aleku, S. P. France, H. Man, J. Mangas-Sanchez, S. L. Montgomery, M. Sharma, F. Leipold, S. Hussain, G. Grogan and N. J. Turner, *Nat. Chem.*, 2017, **9**, 961–969.
- 95 S. P. France, R. Howard, J. Steflík, N. Turner, *ChemCatChem*, 2017, **10**, 510-514.
- 96 Z. S. Seddigi, M. S. Malik, S. A. Ahmed, A. O. Babalghith and A. Kamal, *Coord. Chem. Rev.*, 2017, **348**, 54–70.
- 97 C. -S Chen and C. J. Sih, *Angew. Chem. Int. Ed.*, 1989, **28**, 695–707.
- 98 A. S. de Miranda, L. S. M. Miranda and R. O. M. A. de Souza, *Biotechnol. Adv.*, 2015, **33**, 372–393.
- 99 A. M. Klibanov, *Acc. Chem. Res.*, 1990, **23**, 114–120.
- 100 A. J. Burke, C. S. Marques, N. J. Turner and G. J. Hermann, *Active pharmaceutical ingredients in synthesis: Catalytic processes in research and development*, 2018.
- 101 H. A. Lindner, A. Alary, M. Wilke and T. Sulea, *Biochemistry*, 2008, **47**, 4266–4275.
- 102 W. S. Fones and M. Lee, *J. Biol. Chem.*, 1953, **201**, 847–56.
- 103 M. C. Flickinger, T. Sato and T. Tosa, in *Encyclopedia of Industrial Biotechnology*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010, pp. 1–20.
- 104 Y. P. Xue, Y. G. Zheng, Z. Q. Liu, X. Liu, J. F. Huang and Y. C. Shen, *ACS Catal.*, 2014, **4**, 3051–3058.
- 105 R. Percudani and A. Peracchi, *BMC Bioinf.*, 2009, **10**, 273.
- 106 J. Liang, Q. Han, Y. Tan, H. Ding and J. Li, *Front. Mol. Biosci.*, 2019, **0**, 4.
- 107 J. F. Rocha, A. F. Pina, S. F. Sousa and N. M. F. S. A. Cerqueira, *Catal. Sci. Technol.*, 2019, **9**, 4864–4876.
- 108 P. K. Mehta, T. I. Hale and P. Christen, *Eur. J. Biochem.*, 1993, **214**, 549–561.
- 109 F. Steffen-Munsberg, C. Vickers, H. Kohls, H. Land, H. Mallin, A. Nobili, L. Skalden, T. van den Bergh, H. J. Joosten, P. Berglund, M. Höhne and U. T. Bornscheuer, *Biotechnol. Adv.*, 2015, **33**, 566–604.
- 110 T. Oshima and N. Tamiya, *The Biochemical journal*, 1961, **78**, 116–119.
- 111 K. Yonaha, S. Toyama and H. Kagamiyama, *J. Biol. Chem.*, 1983, **258**, 2260–2265.
- 112 T. Börner, S. Rämisch, E. R. Reddem, S. Bartsch, A. Vogel, A. M. W. H. Thunnissen, P. Adlercreutz and C. Grey, *ACS Catal.*, 2017, **7**, 1259–1269.

- 113 A. Łyskowski, C. Gruber, G. Steinkellner, M. Schürmann, H. Schwab, K. Gruber and K. Steiner, *PLoS ONE*, 2014, **9**, e87350.
- 114 M. S. Humble, K. E. Cassimjee, M. Håkansson, Y. R. Kimbung, B. Walse, V. Abedi, H. J. Federsel, P. Berglund and D. T. Logan, *FEBS J.*, 2012, **279**, 779–792.
- 115 J. S. Shin and B. G. Kim, *J. Org. Chem.*, 2002, **67**, 2848–2853.
- 116 H. Land, F. Ruggieri, A. Szekrenyi, W. D. Fessner and P. Berglund, *Adv. Synth. Catal.*, 2020, **362**, 812–821.
- 117 S. W. Han, E. S. Park, J. Y. Dong and J. S. Shin, *Appl. Environ. Microbiol.*, 2015, **81**, 6994–7002.
- 118 F. Guo and P. Berglund, *Green Chem.*, 2017, **19**, 333–360.
- 119 P. Tufvesson, J. Lima-Ramos, J. S. Jensen, N. Al-Haque, W. Neto and J. M. Woodley, *Biotechnology and Bioengineering*, 2011, **108**, 1479–1493.
- 120 M. Fuchs, J. E. Farnberger and W. Kroutil, *European J. Org. Chem.*, 2015, **32**, 6965–6982.
- 121 J. S. Shin and B. G. Kim, *Biotechnol. Bioeng.*, 1999, **65**, 206–211.
- 122 M. D. Truppo, J. D. Rozzell, J. C. Moore and N. J. Turner, *Org. Biomol. Chem.*, 2009, **7**, 395–398.
- 123 A. Gomm, W. Lewis, A. P. Green and E. O'Reilly, *Chem. Eur. J.*, 2016, **22**, 12692–12695.
- 124 I. Slabu, J. L. Galman, N. J. Weise, R. C. Lloyd and N. J. Turner, *ChemCatChem*, 2016, **8**, 1038–1042.
- 125 A. P. Green, N. J. Turner and E. O'Reilly, *Angew. Chem. Int. Ed.*, 2014, **53**, 10714–10717.
- 126 A. Gomm, S. Grigoriou, C. Peel, J. Ryan, N. Mujtaba, T. Clarke, E. Kulcinskaja and E. O'Reilly, *Eur. J. Org. Chem.*, 2018, **2018**, 5282–5284.
- 127 A. W. H. Dawood, M. S. Weiß, C. Schulz, I. V. Pavlidis, H. Iding, R. O. M. A. de Souza and U. T. Bornscheuer, *ChemCatChem*, 2018, **10**, 3943–3949.
- 128 P. Kelefiotis-Stratidakis, T. Tyrikos-Ergas and I. V. Pavlidis, *Org. Biomol. Chem.*, 2019, **17**, 1634–1642.
- 129 W. Neto, M. Schürmann, L. Panella, A. Vogel and J. M. Woodley, *J. Mol. Catal. B Enzym.*, 2015, **117**, 54–61.
- 130 M. Planchestainer, M. L. Contente, J. Cassidy, F. Molinari, L. Tamborini and F. Paradisi, *Green Chem.*, 2017, **19**, 372–375.
- 131 M. L. Contente and F. Paradisi, *ChemBioChem*, 2019, **20**, 2830–2833.
- 132 A. Gomm and E. O'Reilly, *Curr. Opin. Chem. Biol.*, 2018, **43**, 106–112.
- 133 S. A. Kelly, S. Pohle, S. Wharry, S. Mix, C. C. R. Allen, T. S. Moody and B. F. Gilmore, *Chem. Rev.*, 2018, **118**, 349–367.
- 134 M. D. Patil, G. Grogan, A. Bommarius and H. Yun, *Catalysts*, 2018, **8**, 254.
- 135 I. Slabu, J. L. Galman, R. C. Lloyd and N. J. Turner, *ACS Catal.*, 2017, **7**, 8263–8284.

- 136 C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, *Science*, 2010, **329**, 305–309.
- 137 J. Ryan, M. Šiaučiulis, A. Gomm, B. Maciá, E. O'Reilly and V. Caprio, *J. Am. Chem. Soc.*, 2016, **138**, 15798–15800.
- 138 F. Taday, J. Ryan, S. P. Argent, V. Caprio, B. Maciá and E. O'Reilly, *Chem. Eur. J.*, 2020, **26**, 3729–3732.
- 139 T. Sehl, H. C. Hailes, J. M. Ward, R. Wardenga, E. Von Lieres, H. Offermann, R. Westphal, M. Pohl and D. Rother, *Angew. Chem. Int. Ed.*, 2013, **52**, 6772–6775.
- 140 S. Wu, Y. Zhou, T. Wang, H. P. Too, D. I. C. Wang and Z. Li, *Nat. Commun.*, 2016, **7**, 11917.
- 141 P. Gruber, F. Carvalho, M. P. C. Marques, B. O'Sullivan, F. Subrizi, D. Dobrijevic, J. Ward, H. C. Hailes, P. Fernandes, R. Wohlgemuth, F. Baganz and N. Szita, *Biotechnol. Bioeng.*, 2018, **115**, 586–596.
- 142 M. Höhne, S. Schätzle, H. Jochens, K. Robins and U. T. Bornscheuer, *Nat. Chem. Biol.*, 2010, **6**, 807–813.
- 143 K. Steiner and H. Schwab, *Comput. Struct. Biotechnol. J.*, 2012, **2**, e201209010.
- 144 M. Mettlen, P.-H. Chen, S. Srinivasan, G. Danuser and S. L. Schmid, *Annu. Rev. Biochem.*, 2018, **86**, 637–657.
- 145 F. H. Arnold, *Angew. Chem. Int. Ed.*, 2018, **57**, 4143–4148.
- 146 A. J. Meyer, J. W. Ellefson and A. D. Ellington, *Current Protocols in Molecular Biology*, 2013, **105**, Unit.
- 147 L. Pritchard, D. Corne, D. Kell, J. Rowland and M. Winson, *J. Theor. Biol.*, 2005, **234**, 497–509.
- 148 Press release: The Nobel Prize in Chemistry 2018 - NobelPrize.org, <https://www.nobelprize.org/prizes/chemistry/2018/press-release/>, (accessed April 18, 2020).
- 149 R. A. Chica, N. Doucet and J. N. Pelletier, *Curr. Opin. Biotechnol.*, 2005, **16**, 378–384.
- 150 M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha and A. Vogel, *Angew. Chem. Int. Ed.*, 2005, **44**, 4192–4196.
- 151 M. T. Reetz and J. D. Carballeira, *Nat. Protoc.*, 2007, **2**, 891–903.
- 152 S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz and J. P. Acevedo, *ACS Synth. Biol.*, 2013, **2**, 83–92.
- 153 M. T. Reetz, D. Kahakeaw and R. Lohmer, *ChemBioChem*, 2008, **9**, 1797–1804.
- 154 U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, *Nature*, 2012, **485**, 185–194.
- 155 M. D. Patil, G. Grogan, A. Bommarius and H. Yun, *Catalysts*, 2018, **8**, 254.
- 156 E. S. Park and J. S. Shin, *Enzyme Microb. Technol.*, 2011, **49**, 380–387.

- 157 A. Nobili, F. Steffen-Munsberg, H. Kohls, I. Trentin, C. Schulzke, M. Höhne and U. T. Bornscheuer, *ChemCatChem*, 2015, **7**, 757–760.
- 158 A. R. Martin, R. DiSanto, I. Plotnikov, S. Kamat, D. Shonnard and S. Pannuri, *Biochem. Eng. J.*, 2007, **37**, 246–255.
- 159 A. Telzerow, J. Paris, M. Håkansson, J. González-Sabín, N. Ríos-Lombardía, M. Schürmann, H. Gröger, F. Morís, R. Kourist, H. Schwab and K. Steiner, *ACS Catal.*, 2019, **9**, 1140–1148.
- 160 M. Svedendahl, C. Branneby, L. Lindberg and P. Berglund, *ChemCatChem*, 2010, **2**, 976–980.
- 161 L. Skalden, C. Peters, J. Dickerhoff, A. Nobili, H. J. Joosten, K. Weisz, M. Höhne and U. T. Bornscheuer, *ChemBioChem*, 2015, **16**, 1041–1045.
- 162 H. Xiao, Z. Bao and H. Zhao, *Ind. Eng. Chem. Res.*, 2015, **54**, 4011–4020.
- 163 C. K. Longwell, L. Labanieh and J. R. Cochran, *Curr. Opin. Biotechnol.*, 2017, **48**, 196–202.
- 164 L. Ye, C. Yang and H. Yu, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 559–567.
- 165 G. W. Jack and P. C. McMahon, *BBA - Enzymology*, 1978, **523**, 344–357.
- 166 E. Shaibe, E. Metzger and Y. S. Halpern, *J. Bacteriol.*, 1985, **163**, 933–7.
- 167 R. Voellmy and T. Leisinger, *J. Bacteriol.*, 1976, **128**, 722–729.
- 168 M. Chen, L. Rong and X. Chen, *RSC Adv.*, 2015, **5**, 103557–103562.
- 169 S. Schätzle, M. Höhne, E. Redestad, K. Robins and U. T. Bornscheuer, *Anal. Chem.*, 2009, **81**, 8244–8248.
- 170 J. Hopwood, M. D. Truppo, N. J. Turner and R. C. Lloyd, *Chem. Commun.*, 2011, **47**, 773–775.
- 171 S. C. Willies, J. L. Galman, I. Slabu and N. J. Turner, *Philos. Trans. R. Soc. A*, 2016, **374**.
- 172 M. D. Truppo and N. J. Turner, *Org. Biomol. Chem.*, 2010, **8**, 1280–1283.
- 173 D. Baud, N. Ladkau, T. S. Moody, J. M. Ward and H. C. Hailes, *Chem. Commun.*, 2015, **51**, 17225–17228.
- 174 M. Tamura, R. Tamura, Y. Takeda, Y. Nakagawa and K. Tomishige, *Chem. Commun.*, 2014, **50**, 6656–6659.
- 175 S. L. Shi, Z. L. Wong and S. L. Buchwald, *Nature*, 2016, **532**, 353–356.
- 176 D. Srimani, Y. Ben-David and D. Milstein, *Angew. Chem. International Edition*, 2013, **52**, 4012–4015.
- 177 S. S. V. Ramasastry, H. Zhang, F. Tanaka and C. F. Barbas, *J. Am. Chem. Soc.*, 2007, **129**, 288–289.
- 178 D. J. Ager, I. Prakash and D. R. Schaad, *Chem. Rev.*, 1996, **96**, 835–875.
- 179 S. M. Lait, D. A. Ranric and B. A. Keay, *Chem. Rev.*, 2007, **107**, 767–796.
- 180 L. Cerioli, M. Planchestainer, J. Cassidy, D. Tessaro and F. Paradisi, *J. Mol. Catal. B Enzym.*, 2015, **120**, 141–150.

- 181 N. D. Vetter, D. M. Langill, S. Anjum, J. Boisvert-Martel, R. C. Jagdhane, E. Omene, H. Zheng, K. E. van Straaten, I. Asiamah, E. S. Krol, D. A. R. Sanders and D. R. J. Palmer, *J. Am. Chem. Soc.*, 2013, **135**, 5970–5973.
- 182 S. Singh, Y. Kim, F. Wang, L. Bigelow, M. Endres, M. K. Kharel, G. Babnigg, C. A. Bingman, A. Joachimiak, J. S. Thorson and G. N. Phillips, *Proteins*, 2015, **83**, 1547–1554.
- 183 B. Y. Hwang, H. J. Lee, Y. H. Yang, H. S. Joo and B. G. Kim, *Chem. Biol.*, 2004, **11**, 915–925.
- 184 F. Steffen-Munsberg, C. Vickers, A. Thontowi, S. Schätzle, T. Tumlrirsch, M. SvedendahlHumble, H. Land, P. Berglund, U. T. Bornscheuer and M. Höhne, *ChemCatChem*, 2013, **5**, 150–153.
- 185 S. K. Bharti and R. Roy, *Trends in Anal. Chem.*, 2012, **35**, 5–26.
- 186 E. M. Dangerfield, C. H. Plunkett, A. L. Win-Mason, B. L. Stocker and M. S. M. Timmer, *J. Org. Chem.*, 2010, **75**, 5470–5477.
- 187 D. Monti, M. C. Forchin, M. Crotti, F. Parmeggiani, F. G. Gatti, E. Brenna and S. Riva, *ChemCatChem*, 2015, **7**, 3106–3109.
- 188 A. Gomm, W. Lewis, A. P. Green and E. O'Reilly, *Chem. Eur. J.*, 2016, **22**, 12692–12695.
- 189 S. Grigoriou, P. Kugler, E. Kulcinskaja, F. Walter, J. King, P. Hill, V. F. Wendisch, E. O'Reilly and E. O'Reilly, *Green Chem.*, 2020, **22**, 4128–4132.
- 190 W. Pigman and H. S. Isbell, *Adv. Carbohydr. Chem. Biochem.*, 1968, **23**, 11–57.
- 191 S. J. Cortes, R. L. van Etten and T. L. Mega, *J. Org. Chem.*, 1991, **56**, 943–947.
- 192 H. E. Schoemaker, D. L. Mink and M. G. Wubbolts, *Science*, 2003, **299**, 1694–1697.
- 193 R. Chowdhury and C. D. Maranas, *AIChE J.*, 2020, **66**, e16847.
- 194 M. Ali, H. M. Ishqi and Q. Husain, *Biotechnol. Bioeng.*, 2020, **117**, 1877–1894.
- 195 H. Leemhuis, R. M. Kelly and L. Dijkhuizen, *IUBMB Life*, 2009, **61**, 222–228.
- 196 M. S. Weiß, I. v. Pavlidis, C. Vickers, M. Hohne and U. T. Bornscheuer, *Anal. Chem.*, 2014, **86**, 11847–11853.
- 197 M. Planchestainer, E. Hegarty, C. M. Heckmann, L. J. Gourlay and F. Paradisi, *Chem. Sci.*, 2019, **10**, 5952–5958.
- 198 M. Knowlton, F. C. Dohan and H. Sprince, *Anal. Chem.*, 1960, **32**, 666–668.
- 199 J. M. Miller and J. W. Wright, *J. Clin. Microbiol.*, 1982, **15**, 589–592.
- 200 G. L. Lombard and V. R. Dowell, *J Clin Microbiol.*, 1983, **18**, 609–613.
- 201 A. C. Lamb, R. A. Federico-Perez and Z. L. Xue, *Anal. Biochem.*, 2015, **484**, 21–23.
- 202 R. A. Federico-Perez and Z. L. Xue, *Anal. Biochem.*, 2018, **557**, 104–110.
- 203 M. McMullan, A. García-Bea, P. Miranda-Azpiazu, L. F. Callado and I. Rozas, *Eur. J. Med. Chem.*, 2016, **123**, 48–57.

- 204 R. Cairns, A. Gomm, C. Peel, M. Sharkey and E. O'Reilly, *ChemCatChem*, 2019, **11**, 4738–4743.
- 205 U. Kaulmann, K. Smithies, M. E. B. Smith, H. C. Hailes and J. M. Ward, *Enzyme Microb. Technol.*, 2007, **41**, 628–637.
- 206 M. S. Weiß, I. v. Pavlidis, C. Vickers, M. Hohne and U. T. Bornscheuer, *Anal. Chem.*, 2014, **86**, 11847–11853.
- 207 W. A. Wood, I. C. Gunsalus and W. W. Umbreit, *J. Biol. Chem.*, 1947, **170**, 313–321.
- 208 M. C. Deeley and C. Yanofsky, *J. Bacteriol.*, 1981, **147**, 787–96.
- 209 A. J. Link, D. Phillips and G. M. Church, *J. Bacteriol.*, 1997, **179**, 6228–37.
- 210 D. Ghislieri and N. J. Turner, *Top. Catal.*, 2014, **57**, 284–300.
- 211 F. Subrizi, L. Benhamou, J. M. Ward, T. D. Sheppard and H. C. Hailes, *Angew. Chem.*, 2019, **131**, 3894–3898.
- 212 H. Land, F. Ruggieri, A. Szekrenyi, W. D. Fessner and P. Berglund, *Adv. Synth. Catal.*, 2020, **362**, 812–821.
- 213 S. W. Han, E. S. Park, J. Y. Dong and J. S. Shin, *Appl. Environ. Microbiol.*, 2015, **81**, 6994–7002.
- 214 W. MS, P. IV, S. P, H. SP, W. B, I. H and B. UT, *Chembiochem*, 2017, **18**, 1022–1026.
- 215 I. v. Pavlidis, M. S. Weiß, M. Genz, P. Spurr, S. P. Hanlon, B. Wirz, H. Iding and U. T. Bornscheuer, *Nat. Chem.*, 2016, **8**, 1076–1082.
- 216 A. Nobili, F. Steffen-Munsberg, H. Kohls, I. Trentin, C. Schulzke, M. Höhne and U. T. Bornscheuer, *ChemCatChem*, 2015, **7**, 757–760.
- 217 M. Genz, O. Melse, S. Schmidt, C. Vickers, M. Dörr, T. van den Bergh, H. J. Joosten and U. T. Bornscheuer, *ChemCatChem*, 2016, **8**, 3199–3202.
- 218 K. S. Midelfort, R. Kumar, S. Han, M. J. Karmilowicz, K. McConnell, D. K. Gehlhaar, A. Mistry, J. S. Chang, M. Anderson, A. Villalobos, J. Minshull, S. Govindarajan and J. W. Wong, *Protein Eng. Des. Sel.*, 2013, **26**, 25–33.
- 219 J. Damborsky and J. Brezovsky, *Curr. Opin. Chem. Biol.*, 2014, **19**, 8–16.
- 220 M. Bonomi, R. Pellarin and M. Vendruscolo, *Biophys. J.*, 2018, **114**, 1604–1613.
- 221 A. Srivastava, T. Nagai, A. Srivastava, O. Miyashita and F. Tama, *Int. J. Mol. Sci.*, 2018, **19**, 3401.
- 222 A. Ilari and C. Savino, *Methods Mol. Biol.*, 2008, **452**, 63–87.
- 223 L. E. Kay, *J. Magn. Reson.*, 2011, **213**, 477–491.
- 224 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2009, **31**, NA-NA.
- 225 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.

- 226 P. T. Lang, S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James and I. D. Kuntz, *RNA*, 2009, **15**, 1219–1230.
- 227 I. W. Davis and D. Baker, *J. Mol. Biol.*, 2009, **385**, 381–392.
- 228 A. N. Jain, *J. Med. Chem.*, 2003, **46**, 499–511.
- 229 N. S. Pagadala, K. Syed and J. Tuszynski, *Biophys. Rev.*, 2017, **9**, 91–102.
- 230 J. Biesiada, A. Porollo, P. Velayutham, M. Kouril and J. Meller, *Hum. Genomics*, 2011, **5**, 497–505.
- 231 J. Li, A. Fu and L. Zhang, *Interdiscip. Sci.-Comput. Life Sci.*, 2019, **11**, 320–328.
- 232 I. A. Guedes, F. S. S. Pereira and L. E. Dardenne, *Front. Pharmacol.*, 2018, **9**, 1089.
- 233 K. K. Chaudhary and N. Mishra, *JSM Chem.*, 2016, **4**, 1029.
- 234 T. J. A. Ewing and I. D. Kuntz, *J. Comput. Chem.*, 1997, **18**, 1175–1189.
- 235 D. S. Goodsell, G. M. Morris and A. J. Olson, *J. Mol. Recognit.*, 1996, **9**, 1099–1352
- 236 R. Madaj, A. Sanker, B. A. S. Geoffrey, H. A. David, S. Verma, J. Gracia, A. I. Faleti and A. H. Yakubu, *bioRxiv*, 2020, 2020.07.07.192518.
- 237 J. R. Schames, R. H. Henchman, J. S. Siegel, C. A. Sottriffer, H. Ni and J. A. McCammon, *J. Med. Chem.*, 2004, **47**, 1879–1881.
- 238 S. F. Sousa, P. A. Fernandes and M. J. Ramos, *Proteins*, 2006, **65**, 15–26.
- 239 S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell and A. J. Olson, *Nat. Protoc.*, 2016, **11**, 905–919.
- 240 D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, Wiley-VCH, 1999.
- 241 J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, B. Basanta, B. J. Bender, K. Blacklock, J. Bonet, S. E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B. E. Correia, B. Coventry, R. Das, R. M. de Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A. S. Ford, B. Frenz, D. Y. Fu, C. Geniesse, L. Goldschmidt, R. Gowthaman, J. J. Gray, D. Gront, S. Guffy, S. Horowitz, P. S. Huang, T. Huber, T. M. Jacobs, J. R. Jeliazkov, D. K. Johnson, K. Kappel, J. Karanicolas, H. Khakzad, K. R. Khar, S. D. Khare, F. Khatib, A. Khramushin, I. C. King, R. Kleffner, B. Koepnick, T. Kortemme, G. Kuenze, B. Kuhlman, D. Kuroda, J. W. Labonte, J. K. Lai, G. Lapidoth, A. Leaver-Fay, S. Lindert, T. Linsky, N. London, J. H. Lubin, S. Lyskov, J. Maguire, L. Malmström, E. Marcos, O. Marcu, N. A. Marze, J. Meiler, R. Moretti, V. K. Mulligan, S. Nerli, C. Norn, S. Ó'Conchúir, N. Ollikainen, S. Ovchinnikov, M. S. Pacella, X. Pan, H. Park, R. E. Pavlovicz, M. Pethe, B. G. Pierce, K. B. Pilla, B. Raveh, P. D. Renfrew, S. S. R. Burman, A. Rubenstein, M. F. Sauer, A. Scheck, W. Schief, O. Schueler-Furman, Y. Sedan, A. M. Sevy, N. G. Sgurakis, L. Shi, J. B. Siegel, D. A. Silva, S. Smith, Y. Song, A. Stein, M. Szegedy, F. D. Teets, S. B. Thyme, R. Y. R. Wang, A. Watkins, L. Zimmerman and R. Bonneau, *Nat. Methods*, 2020, **17**, 665–680.
- 242 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
- 243 L. Heo and M. Feig, *Bioinformatics*, 2018, **34**, 1063–1065.
- 244 H. Land and M. S. Humble, *Methods Mol. Biol.*, 2018, **1685**, 43–67.

- 245 J. S. Shin and B. G. Kim, *J. Org. Chem.*, 2002, **67**, 2848–2853.
- 246 B. Guidi, M. Planchestainer, M. Letizia Contente, T. Laurenzi, I. Eberini, L. J. Gourlay, D. Romano, F. Paradisi and F. Molinari, *Scientific Reports*, 2018, **8**, 34434-34444.
- 247 J. L. Galman, D. Gahloth, F. Parmeggiani, I. Slabu, D. Leys and N. J. Turner, *Front. Bioeng. Biotechnol.*, 2018, **6**, 205-210.
- 248 E. S. Park, M. Kim and J. S. Shin, *Appl. Microbiol. Biotechnol.*, 2012, **93**, 2425–2435.
- 249 D. Deszcz, P. Affaticati, N. Ladkau, A. Gegel, J. M. Ward, H. C. Hailes and P. A. Dalby, *FEBS J.*, 2015, **282**, 2512–2526.
- 250 A. Pellegrini, U. Thomas, P. Wild, E. Schraner and R. von Fellenberg, *Microbiol. Res.*, 2000, **155**, 69–77.
- 251 J. Chiu, P. E. March, R. Lee and D. Tillett, *Nucleic Acids Res.*, 2004, **32**, 174-179.
- 252 D. Seeliger and B. L. de Groot, *J. Comput.-Aided Mol. Des.*, 2010, **24**, 417–422.
- 253 S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell and A. J. Olson, *Nat. Protoc.*, 2016, **11**, 905–919.
- 254 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2009, **31**, NA-NA.
- 255 P. Dewick, *Medicinal Natural Products: A Biosynthetic Approach*, Wiley-VCH, 3rd edn., 2009.
- 256 E. Vitaku, D. T. Smith and J. T. Njardarson, *J. Med. Chem.*, 2014, **57**, 10257–10274.
- 257 S. Bala, N. Sharma, A. Kajal, S. Kamboj and V. Saini, *Int. J. Med. Chem.*, 2014, **2014**, 1–15.
- 258 S. G. Subramaniapillai, *J. Chem. Sci.*, 2013, **125**, 467–482.
- 259 J. M. M. Verkade, L. J. C. van Hemert, P. J. L. M. Quaedflieg and F. P. J. T. Rutjes, *Chem. Soc. Rev.*, 2008, **37**, 29–41.
- 260 J. Iwanejko, E. Wojaczyńska and T. K. Olszewski, *Curr. Opin. Green Sustainable Chem.*, 2018, **10**, 27–34.
- 261 S. Saranya, N. A. Harry, K. K. Krishnan and G. Anilkumar, *Asian J. Org. Chem.*, 2018, **7**, 613–633.
- 262 M. R. Monaco, P. Renzi, D. M. Scarpino Schietroma and M. Bella, *Org. Lett.*, 2011, **13**, 4546–4549.
- 263 N. Zotova, A. Moran, A. Armstrong and D. G. Blackmond, *Adv. Synth. Catal.*, 2009, **351**, 2765–2769.
- 264 L. Chausset-Boissarie, R. Àrvai, G. R. Cumming, C. Besnard and E. P. Kündig, *Chem. Commun.*, 2010, **46**, 6264–6266.
- 265 J. L. Galman, I. Slabu, F. Parmeggiani and N. J. Turner, *Chem. Commun.*, 2018, **54**, 11316–11319.
- 266 J. L. Galman, I. Slabu, N. J. Weise, C. Iglesias, F. Parmeggiani, R. C. Lloyd and N. J. Turner, *Green Chem.*, 2017, **19**, 361–366.

- 267 M. Follmann, I. Ochrombel, R. Krämer, C. Trötschel, A. Poetsch, C. Rückert, A. Hüser, M. Persicke, D. Seiferling, J. Kalinowski and K. Marin, *BMC Genomics*, 2009, **10**, 1–23.
- 268 J. Guo, Z. Ma, J. Gao, J. Zhao, L. Wei, J. Liu and N. Xu, *World J. Microbiol. Biotechnol.*, 2019, **35**, 192.
- 269 S. Unthan, A. Grünberger, J. van Ooyen, J. Gätgens, J. Heinrich, N. Paczia, W. Wiechert, D. Kohlheyer and S. Noack, *Biotechnol. Bioeng.*, 2014, **111**, 359–371.
- 270 Q. Ma, X. Zhang and Y. Qu, *Microbiol.*, 2018, **9**, 2625-2630.
- 271 D. Miura and T. Machinami, *MRC*, 2015, **4**, 20-27.
- 272 A. Tsutsui, H. Takeda, M. Kimura, T. Fujimoto and T. Machinami, *Tetrahedron Lett.*, 2007, **48**, 5213-5217.