Structural Insights into Short Chain (SDR) and Medium Chain (MDR) families of Alcohol Dehydrogenases for Biocatalytic Applications

By

Henry Wing-Hong Man

Thesis submitted for MSc by Research

The University of York

The Department of Chemistry

October 2013

Abstract

Current abiotic synthesis is very effective in yielding enantiomerically pure compounds with high yields, however the route taken may not be very efficient nor green with the use of solvents, heavy metals and toxic chemicals. Alcohol dehydrogenases (ADHs) are industrially useful biocatalysts in providing synthetically useful building blocks or chiral alcohols as flavour or fragrance chemicals. In this thesis, the X-ray crystal structures have been determined for the ADHs from *Ralstonia* sp. (RasADH), *Sphingobium yanoikuyae* (SyADH), *Thermus* sp. (TADH) and *Candida parapsilosis* (CPCR2).

Datasets for the *apo*- and NADPH complex for the gene encoding RasADH, an ADH from the short-chain family, from *Ralstonia* sp. DSM 6428 were obtained at 1.5 Å and 2.9 Å respectively. The models for the *apo*- and NADPH complex were refined to R_{cryst}/R_{free} values of 15.8%/18.8% and 26.8%/29.3% respectively.

The dataset for the gene encoding SyADH, an ADH from the short-chain family, from *Sphingobium yanoikuyae* DSM 6900 in complex with NADPH was obtained at 2.5 Å. The model was refined to an R_{cryst}/R_{free} value of 23.4%/25.1%.

The dataset for the gene encoding TADH, an ADH from the medium-chain family, from *Thermus* sp. ATN1 was obtained at 2.7 Å with three of four active sites in complex with NADH. The model was refined to an R_{cryst}/R_{free} value of 19.6%/24.6%.

The dataset for the gene encoding CPCR2, an ADH from the medium-chain family, from *Candida parapsilosis* DSM 70125 was obtained at 2.0 Å in complex with NADH. The model was refined to an R_{cryst}/R_{free} value of 20.5%/24.1%.

These crystal structures provide a robust platform for rational engineering to improve the enzymes' substrate specificity, enantioselectivity, thermostability and solvent tolerance. As such, ADHs as biocatalysts will inevitably provide a greener alternative for the synthesis of chiral alcohols over abiotic synthesis.

Table of Contents

List of Figures

Acknowledgments

I would like to thank Dr. Gideon Grogan for giving me the opportunity to work on this project and proofreading my thesis. I would also like to thank everyone in my group: Annika Frank, Chantel Jensen, Claudia Spandolf, Ben Summers and Elizabeth Wells, who have each helped me on numerous occasions. I also thank Simon Grist and Sally Lewis for maintaining the wet lab, also Sam Hart and Johan Turkenburg for all the crystal data collection.

Without the collaborators, I would not have this project. As such, I would like to thank Serena Gargiulo, Kinga Kędziora, Justyna Kulig, Christoph Loderer and their respective supervisors Frank Hollmann from Delft University of Technology in Delft – The Netherlands, Iván Lavandera and Vincente Gotor-Fernández from Universidad de Oviedo in Oviedo - Spain, Dörter Rother from Institute of Bio- and Geosciences in Jülich - Germany, and Marion Ansorge-Schumacher from Technische Universität Dresden in Dresden - Germany.

Last but not at all least, I would like to thank my family and girlfriend, Sherry, for supporting me through out the year.

It has been a pleasure working at YSBL, which is such a friendly and intellectually stimulating environment. I now have the opportunity to stay for another 3 years for my PhD.

Declaration

X-ray data collection at Diamond Light Source Synchrotron (Didcot – Oxford) was performed by Dr. Johan Turkenburg and Sam Hart. All other work mentioned in this thesis was performed by myself unless stated otherwise.

The RasADH and SyADH projects have resulted in the following publication:

Structures of alcohol dehydrogenases from *Ralstonia* and *Sphingobium* spp. reveal the molecular basis for their recognition of 'bulky-bulky' ketones.
Man, H., Kędziora, K., Kulig, J., Frank, A., Lavandera, I., Gotor-Fernández, V., Rother, D., Hart, S., Turkenburg, J., Grogan, G., *Topics in Catalysis*, **2013**, in press.

The CPCR2 project have resulted in the following publication:

Structure of NADH-dependent carbonyl reductase (CPCR2) from *Candida parapsilosis* provides insights into improved process suitability achieved through mutagenesis.
Man, H., Loderer, C., Ansorge-Schumacher, M., Grogan, G., *ChemCatChem*, **2013**, submitted.

1. Introduction

1.1. Alcohol Oxidation and Carbonyl Reduction

The general oxidation of alcohols and reduction of carbonyl groups is a relatively simple reaction in organic chemistry and the mechanism is well-known. There are many different forms of oxidation and reduction reactions; for example Jones Reagent (1) or acidified chromate (2), can be used to oxidise an alcohol to its corresponding aldehyde or ketone (**Figure 1**) and sodium borohydride (3), can be used to effect the reverse reaction (**Figure 2**).

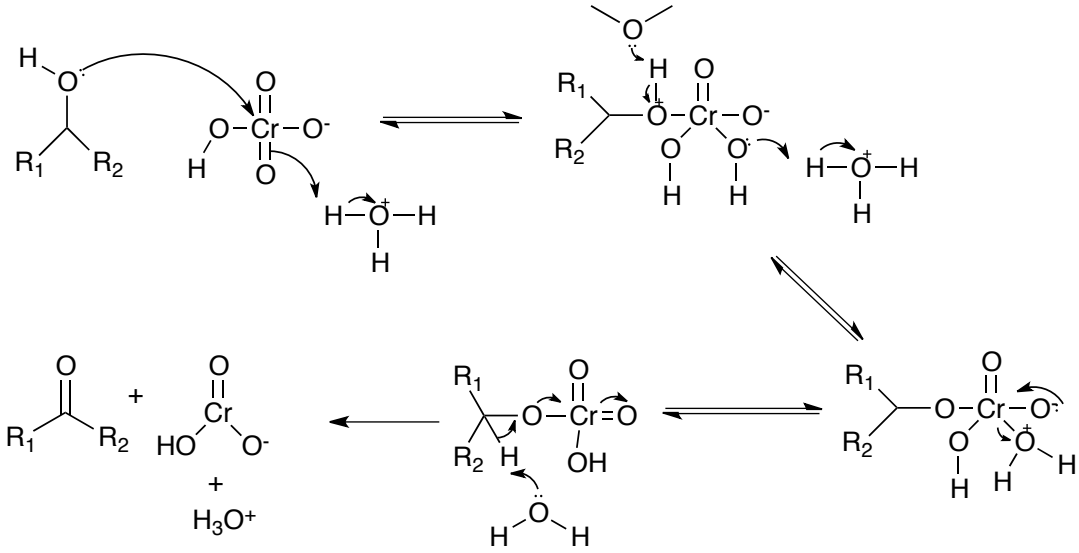


Figure 1 – Mechanism shown for potassium dichromate oxidation of a simple generic alcohol where chromate (VI) is reduced to chromate (III).

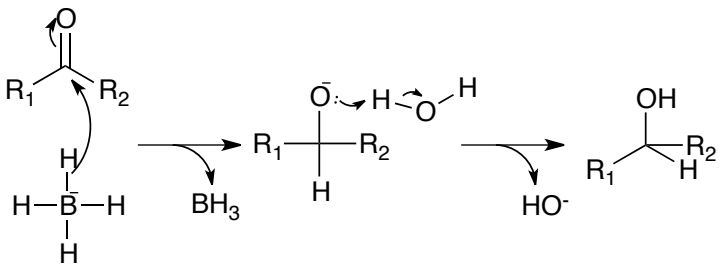


Figure 2 – Mechanism shown for sodium borohydride reduction of a generic ketone. The reaction is driven forward by the formation of borane.

In more complex cases of alcohol oxidation, protecting groups may be required if more than one hydroxyl groups is present and different strategies for oxidation may

be required. The reduction of carbonyl groups also presents an interesting problem from the perspective of selectivity. The carbonyl group is planar, which means it can be attacked from either the *re*- or *si*-face. The topos-selective, asymmetric reduction of carbonyls has often been very difficult to accomplish, however Ryōji Noyori and co-workers developed BINAP (4,5) which utilises ruthenium, rhenium or palladium (**Figure 4** - right) to catalyse the asymmetric hydrogenation of carbonyls with high enantiomeric excess (*e.e.*). For example, (-)-menthol could be synthesised using Rh-BINAP and hydrogen to asymmetrically hydrogenate the terpene precursor (**Figure 3**) with 99% yield and 98% *e.e*. Chiral alcohols themselves are often flavour or fragrance chemicals, such as (-)-menthol, and they can also be used as synthetically useful chiral building blocks.

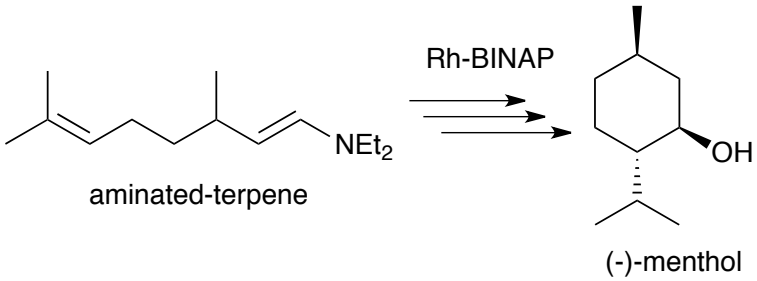


Figure 3 – Scheme of the asymmetric hydrogenation step using Rh-BINAP in the synthesis of (-)-menthol. Adapted from Akutagawa (6).

BINAP is chiral, as it has no plane of symmetry due to restricted rotation about the biphenyl single bond. As seen in **Figure 4**, due to this restriction, the binaphthyl groups create faces where potential aromatic groups, alkenes, alkynes and carbonyls will potentially π-stack and fix in place for attack on a specific face. This potential π-stacking effect leads to very high *e.e*.s.

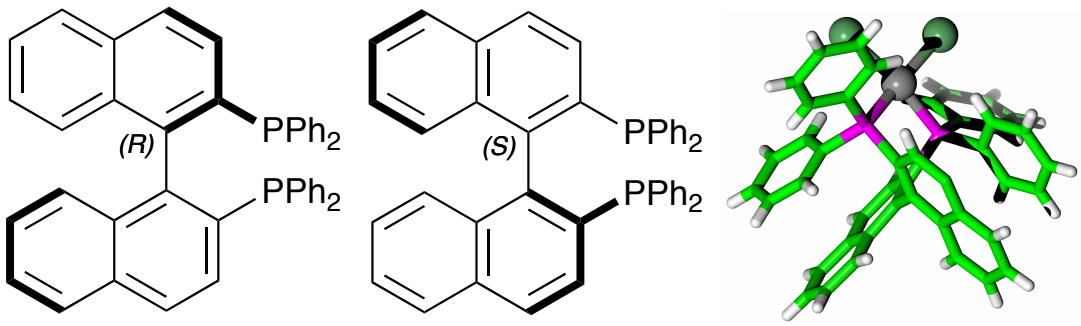


Figure 4 - The chiral catalysts BINAP developed by Noyori and co-workers shown above with both enantiomers: *(R)*-BINAP (left), *(S)*-BINAP (middle) and 3D representation of *(R)*-BINAP, CCDC entry: 879636 (right). The carbon atoms are represented in green, hydrogen atoms in white, phosphorous atoms are in magenta, chloride atoms are in lawn green and palladium metal ion in grey.

Although there has been some progress with abiotic asymmetric catalysis for chiral reduction utilising heavy metals, the ligands are expensive and require gaseous H_2, which is a potential hazard. Greener solutions are required, such as using enzymes as catalysts.

1.2. Enzymes as Catalysts

Enzymes perform powerful functional group transformations and can offer a cleaner and greener alternative to traditional organic chemistry as biocatalysts, as well as performing asymmetric synthesis with incredibly high enantio-, regio- and chemoselectivity. There are still many desirable natural products that are still not available by the means of abotic synthesis. Certain reagents and solvents required in synthesis are toxic and carcinogenic, for example the use of pyridinium chlorochromate (PCC) in the selective oxidation of alcohols (7). The work-up reactions, using reagents such as PCC, are also not trivial. Hence product recovery is still an issue.

When using biocatalysts, solvents can be used in an integrated biphasic organic-aqueous system (8), to separate the desired product from the enzyme or whole cell reaction. Biphasic systems can cause less stress to purified enzyme or whole cell systems as the desired product may be cytotoxic (9) and will often preferentially diffuse into the organic layer.

Enzymes often perform reactions by kinetic resolution (KR), this implies the differentiation of two enantiomers from a racemic mixture using chiral catalysts or reagents; however KRs may not always fully resolve a racemate to >99% *e.e*. If an enzyme is not able to perform the desired reaction, it is possible to reengineer the enzyme to improve its KR properties using techniques such as site-directed mutagenesis (SDM) and Directed Evolution; however, mutations can be time consuming and expensive with chances that the mutant will still not perform the desired reaction. An alternative to engineering is to make the KR a dynamic kinetic resolution (DKR) or a dynamic kinetic asymmetric transformation (DYKAT) (10). DKRs and DYKATs are dynamic systems where the minor product or by-product are transformed back to the starting material, thereby forming a continuous system to produce the desired product in the theoretical yield of 100% rather than the 50% achieved in KRs. While it is often difficult to find a stereo-complementary enzyme or a racemase with the desired reaction and substrate specificity for DKRs or a suitable substrate for DYKAT, when parameters have been optimised the results are very

promising. One example is the production of D-alanine (**Figure 5**) from L-alanine amide using D-Aminopeptidase (11).

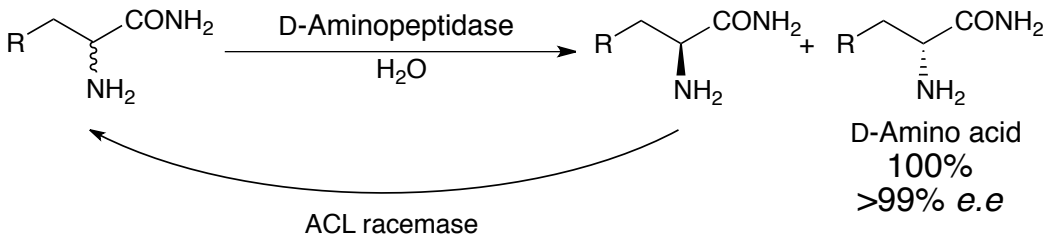


Figure 5 - Production of D-Alanine from L-Alanine amide employing a DKR. The D-Aminopeptidase kinetically resolves the amino amide to the L- and D-amino acid, however l-α-amino-ε-caprolactam (ACL) racemase will transform the by-product back to the starting material, therefore 100% yield of the desired product can be achieved. Modified from Yamaguchi and co-workers (11).

1.3. Alcohol Dehydrogenases (ADHs)

ADHs have evolved to be able to oxidise alcohols and reduce aldehyde/ketones (**Figure 6**). By exploiting this reactivity, they can be used to perform these desirable reactions on substrates of choice. ADHs utilise the nicotinamide cofactor [NAD(P)], which are donors or acceptors of hydride, to perform the reduction or oxidation like many other dehydrogenases. ADHs are able to perform biotransformations on substrates with very high enantioselectivity, selectively oxidising or reducing the functional group. They are able to perform these reactions much more elegantly than using abiotic synthesis, however work-up and product recovery can still be an issue.

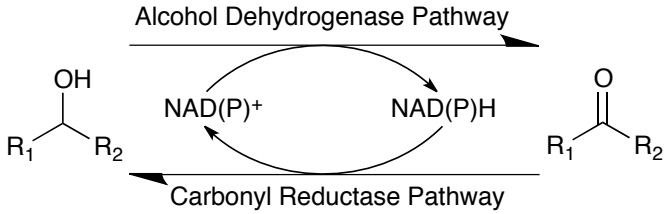


Figure 6 - The two pathways that alcohol dehydrogenases can perform on a generic alcohol and ketone aided by the nicotinamide cofactor

There are three classes of ADHs: the long-chain family (LDR), the medium-chain family (MDR) and the short-chain family (SDR). Each class of ADHs has evolved to oxidise or reduce a specific set of substrates, thus, ADHs often have very narrow substrate specificities.

1.3.1. Previously studied ADHs

ADHs have been extensively studied as a family of enzymes not only as biocatalysts, but as potential targets for inhibition in the treatment of diseases such as Alzheimer's (12), Parkinson's (13) and heart (14) and liver (15) related diseases. Many of these medically-related investigations are inhibition studies of ADHs.

Due to the usefulness of ADHs, many have been become commercially available. The medium chain alcohol dehydrogenase from *Thermoanaerobium brockii*'s (TbADH) (16) substrate versatility and thermostability allows it to be a powerful

biocatalyst. The mechanism of MDRs was initially investigated using Horse Liver ADH (HLADH) (17) and this has since been the model for the medium chain family.

Many enzymes, including thermostable enzymes, are not able to tolerate high concentrations of solvent. Investigations in solvent tolerant ADHs were conducted by Karabec and co-workers in 2010 (18) with ADH-'A' from *Rhodococcus ruber*; ADH-'A' is tolerant up to 80% v/v isopropanol and still active in 99% v/v hexane (19). Solvent tolerance is very useful as hydrophobic substrates are not water-soluble.

1.3.2. Short-Chain Alcohol Dehydrogenases (SDRs)

SDRs are composed of approximately 250 amino acids with a Rossmann Fold nucleotide binding motif GxGxxG (20) and therefore utilise nicotinamide cofactors assisting catalysis. SDRs have a highly conserved catalytic triad (**Figure 7**) in the active site (21) consisting of, for example, Ser138, Tyr151 and Lys155 in *Drosophila* ADH (DADH), PDB entry: 1B2L.

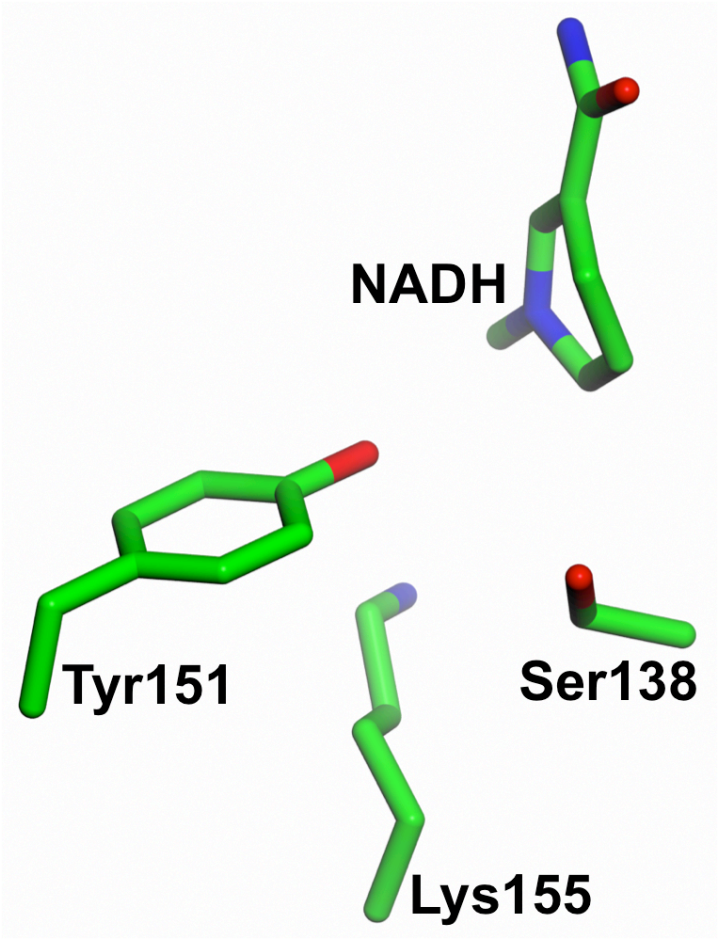


Figure 7 – Catalytic triad (Ser138-Tyr151-Lys155) with nicotinamide cofactor shown for SDRs as proposed by Jörnvall (22) obtained from the crystal structure of DADH (PDB entry: 1B2L).

The mechanism of SDR action (**Figure 8**) was proposed by Jörnvall and co-workers (22). The reduction mechanism is driven by regaining aromaticity of the nicotinamide ring, which is followed by the substrate receiving a proton from the proton donor Tyr151. The oxidation mechanism is driven by Tyr151's reprotonation from the alcohol.

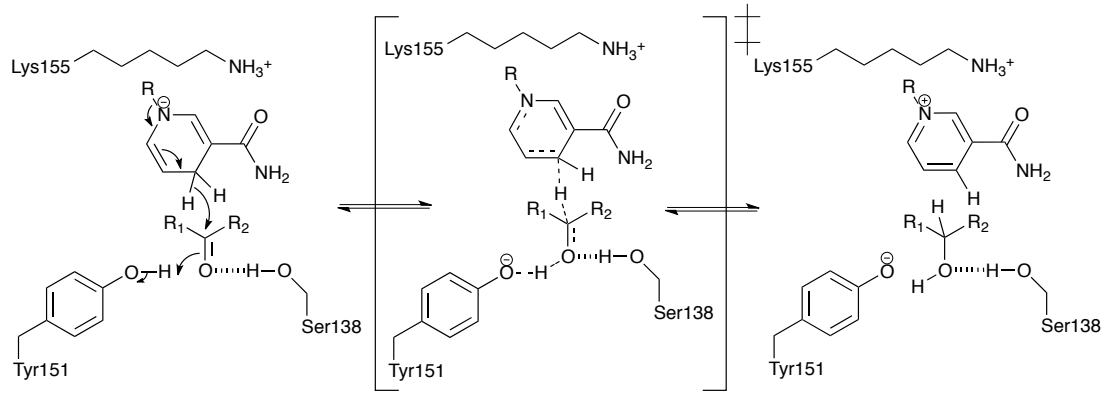


Figure 8 – The SDR mechanism as proposed by Jörnvall and co-workers (22) in DADH. Catalytic triad Ser-Tyr-Lys with NAD(P)H showing the carbonyl reductase mechanism. Regaining aromaticity of the nicotinamide ring facilitates the carbonyl reductase pathway. Tyrosine regaining its proton will facilitate the alcohol dehydrogenase pathway.

Members of the SDR family include RasADH (from *Ralstonia sp.* DSM 6428) and SyADH (from *Sphingobium yanoikuyae* DSM 6900), each of which has been a target for the studies in this thesis.

1.3.2.1. The SDR from Ralstonia sp. (RasADH)

RasADH has previously been extensively characterised by Lavendera and co-workers in 2008 (23,24) and, more recently, by Cuetos and co-workers (25) and Kulig and co-workers in 2012 (26) and 2013 (27). RasADH is an NADPH dependent, 27 kDa enzyme of the SDR family with 249 amino acids from *Ralstonia* sp. DSM 6428. RasADH was found to be capable biotransforming so-called 'bulky-bulky' ketones (**Figure 9**) such as *n*-pentyl phenyl ketone (PPK), with a high preference for the (*S*)-alcohol of up to 99% *e.e*. The addition of calcium ions was observed to stabilise the protein in solution (27).

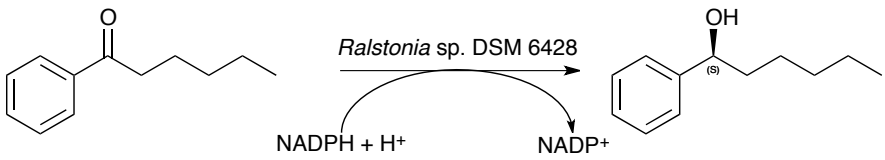


Figure 9 - Representative of a 'bulky-bulky' ketone, *n*-pentyl phenyl ketone (PPK), reduced to the (*S*)-alcohol by RasADH.

The crystal structure for RasADH was recently solved by Lerchner and co-workers in 2013 (28) and it did not reveal any calcium ions present. RasADH has a broad

substrate range for sterically demanding ketones with incredibly high *e.e.*s (>99%) and more useful diastereomeric excess (*d.e.*s > 99%) with respect to the reduction of β-hydroxy ketones to 1,2-diols. It has therefore been shown that RasADH is a very useful biocatalyst for the asymmetric reduction of β-hydroxy ketones.

1.3.2.2. The SDR from Sphingobium yanoikuyae (SyADH)

SyADH is also an NADPH dependent, 29 kDa enzyme of the SDR family with 272 amino acids from *Sphingobium yanoikuyae* DSM 6900. SyADH is also able to catalyse the formation of enantiomerically pure 'bulky-bulky' alcohols from the respective ketone as with RasADH; althought, interestingly with different diasteroselectivity (**Figure 10**). The biochemical characteristics of SyADH have not been extensively investigated as much as for RasADH.

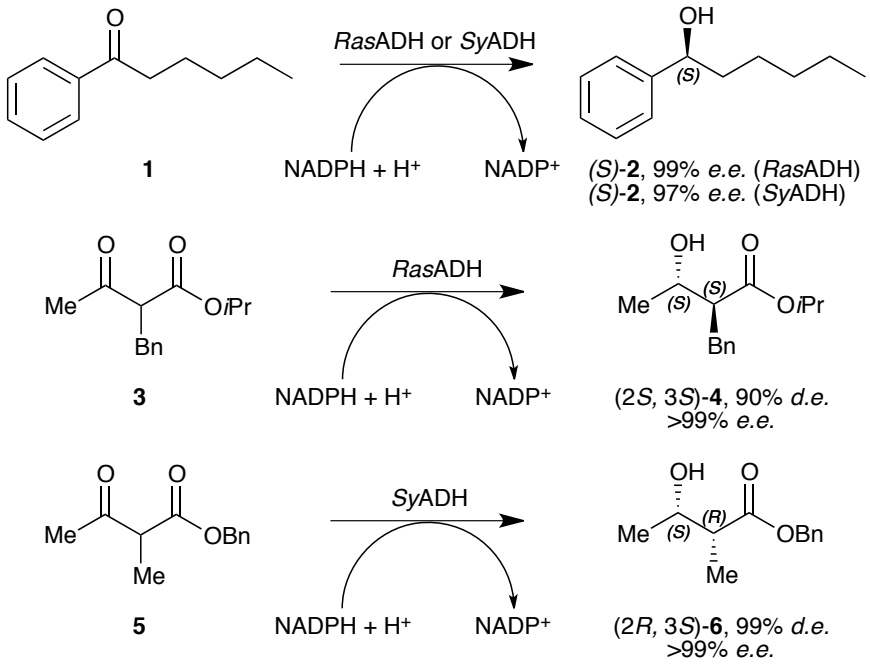


Figure 10 – Representative asymmetric reductions of bulky-bulky ketones catalysed by ADHs from RasADH and SyADH. Adapted from Man and co-workers(29).

However, it was hypothesised by Lavandera and co-workers in 2008 (30) that, as SyADH was able to reduce 'bulky-bulky' ketones its active site would be quite large and was therefore able to non-selectively oxidise racemic 'small-bulky' secondary alcohols (**Figure 11**). When used with a co-substrate for cofactor regeneration such

as acetone, a KR of 50% *e.e*. was achieved with *rac*-2-octanol. However when α-halo ketones were used, such as chloroacetone, complete conversion (>99% yield) could be achieved.

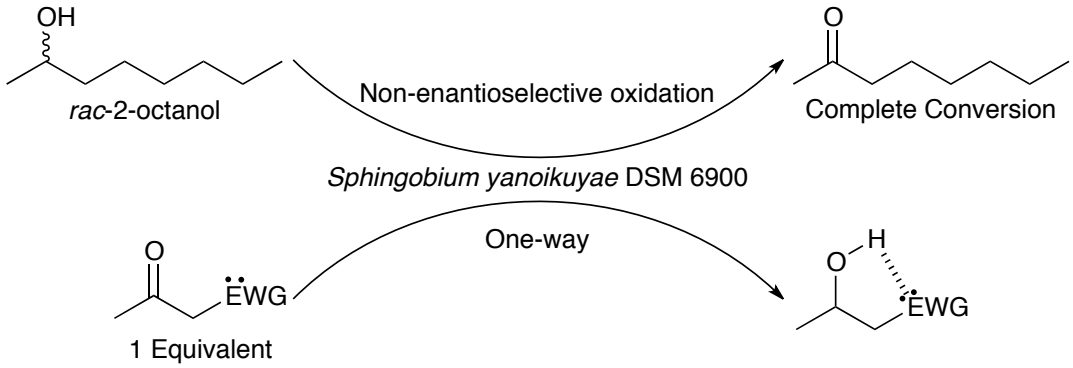


Figure 11 – Modified from Lavandera and co-workers (30). Schematic of the quasi-irreversible, complete oxidation of secondary alcohols with the use of α-halo ketones as hydrogen acceptors for biocatalytic hydrogen transfer reaction using SyADH.

The improved activity towards α-halo ketones is a consequence of exploiting the chemistry of halohydrins. It is more energetically favourable for the hydrogenation of α-halo ketones than the formation of iso-propanol. Furthermore, stability gained from the new hydrogen bond formed from the alcohol's proton and the α-halo group caused SyADH to be unable to reoxidise the halohydrin. Therefore a quasi-irreversible system was formed where complete oxidation of racemic alcohols was achieved.

1.3.3. Medium Chain Alcohol Dehydrogenases (MDRs)

MDRs are composed of approximately 350 amino acids with a Rossmann nucleotide binding motif for NAD(P)H. The enzymes typically contain two zinc atoms in each monomer; utilising zinc both catalytically and structurally. MDRs can also perform oxidoreductase reactions similar to that of SDRs. However the active site of MDRs does not consist of a catalytic triad as seen with SDRs. Catalysis is aided by a zinc atom, NADH and a catalytic serine (**Figure 12**) as seen in the active site of horse liver ADH (HLADH), PDB entry: 6ADH.

The mechanism of action proposed initially by Eklund (31) for HLADH involves a static zinc atom coordinated by Cys46, His67, Cys174 and the hydroxyl group on the

bound substrate, dimethyl sulfide (DMS), in a tetrahedral geometry (**Figure 12**). The residues surrounding the catalytic zinc are not strictly conserved.

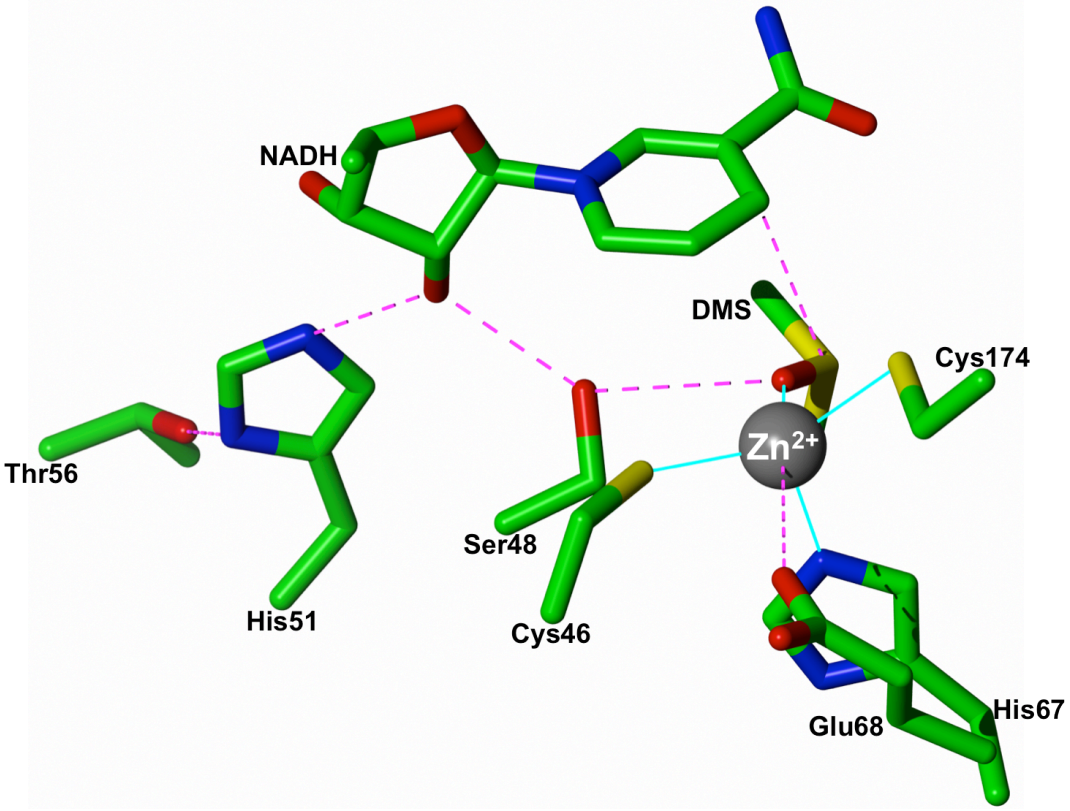


Figure 12 – The key catalytic residues shown for the medium chain alcohol dehydrogenase mechanism as proposed by Eklund (31) as in HLADH. Cyan vectors represent the main catalytic interactions arranged tetrahedrally with the zinc atom. Magenta dashed-vectors represent other potential catalytically involved interactions. Dimethyl sulphide (DMS) is bound in the active site.

The reduction mechanism (**Figure 13**) is again driven by the nicotinamide ring regaining aromaticity; the nicotinamide hydride attacks the carbonyl carbon shifting electron density to the oxygen allowing the oxygen to take the proton from Ser48. After protonation from Ser48, a cascade of proton transfers occurs for the MDR to adopt the most stable form; Ser48 reprotonates itself from the proton of the hydroxyl group on the 2' position of the ribose ring bound to the nicotinamide group, followed by 2' hydroxyl of the reprotonation from the neighbouring protonated N3-His51. Finally His51 reprotonates the N1 nitrogen by proton abstraction from Thr56.

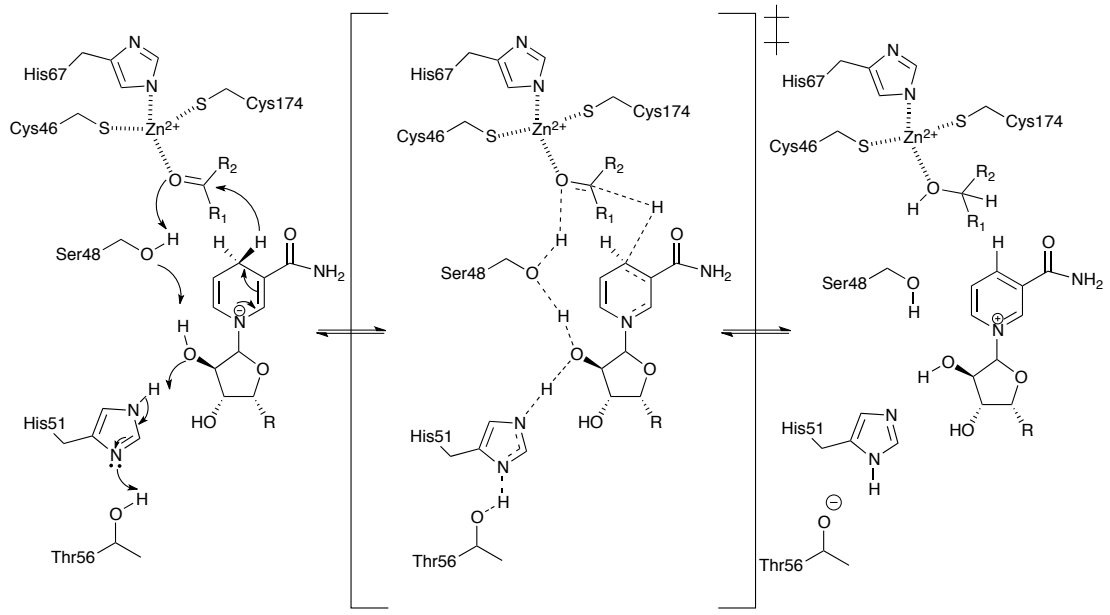


Figure 13 - The medium chain alcohol dehydrogenase mechanism as proposed by Eklund (31) as in HLADH. Again, regaining aromaticity of the nicotinamide ring facilitates the carbonyl reductase pathway.

In more recent studies of the MDR active site by Meijers and co-workers in 2001 (32) and Baker and co-workers in 2009 (33), it has been noted that the catalytic zinc atom is not static as previously thought. Catalysis is potentially water-mediated and Glu68 in Liver ADH (LADH), PDB entries: 1HEU, 1HF3 and 1HET, is a potential ligand for zinc; where zinc is now coordinated tetrahedrally with Cys46, His67, Glu68 and Cys174. Whether or not the catalysis is water-mediated could be due to the size of the active site; water may be required to complete the reaction if the substrate is held too far for coordination (**Figure 14**) with zinc as in glucose dehydrogenase (33).

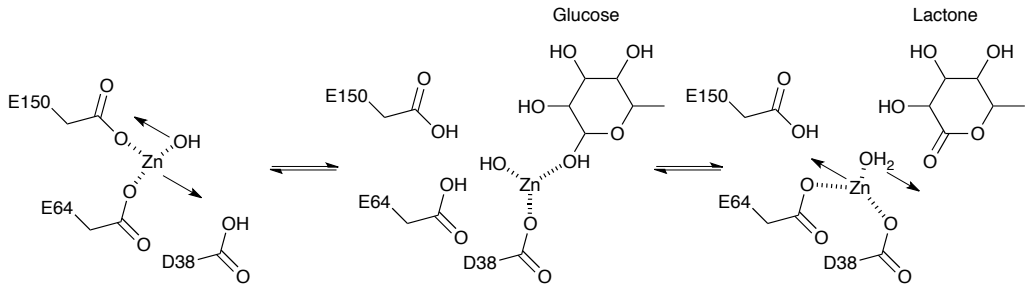


Figure 14 – Schematic diagram of water mediated catalysis from Baker 2009 (33), viewed down the His 63 NE2–Zn bond, of the zinc ligands in the NADP(H)/Zn complex (*Left*), $NADP^+$/Zn/glucose complex (*Centre*), and $NADP^+$/Zn/gluconolactone complex (*Right*). The arrows give the direction of the movement of the zinc and water between the complexes. PDB entries: 2VWP, 2VWQ, 2VWH and 2VWG.

1.3.3.1. The MDR from Thermus sp. (TADH)

Thermophilic and hyperthermophilic biocatalysts have been of particular interest to industry. Their robust characteristics allow them to be used under reaction conditions suitable for industrial processes and to remain stable. Enzymes that are stable at high temperatures are often relatively stable in high concentration of solvents. The use of high solvent concentration may not necessarily be required with high temperature reactions, as higher temperatures may allow for increased substrate solubility and rate of diffusion (34). However, the major downfall of many MDRs, including thermophilic MDRs, is the limited substrate scope.

TADH is an NADH dependent, 37 kDa thermophilic MDR from *Thermus* sp. ATN1 with 347 amino acids. TADH has a remarkably wide substrate spectrum and highest activities were observed with primary alcohols and aldehydes (35). Activity loss was observed when acetone was used as a co-substrate for cofactor regeneration; however, this was still preferable over introducing a second enzyme system. Interestingly TADH displays high dismutation activity, the simultaneous oxidation and reduction of a single species with aldehydes, therefore TADH can be used as a biocatalytic tool for accessing carboxylic acids (**Figure 15**). Due to the irreversible conversion from aldehyde to carboxylic acid, complete conversion can be achieved over 24 hours.

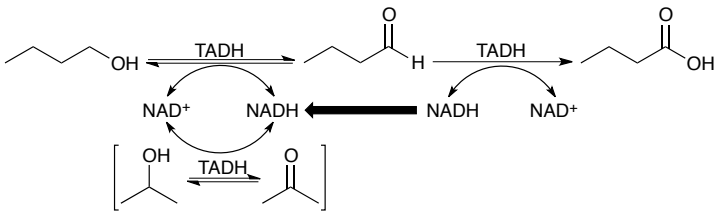


Figure 15 – Modified from Hollmann and co-workers (35). Scheme of aldehyde dismutation reaction catalysed by TADH. Acetone was used for cofactor regeneration.

Not only is TADH thermostable, but it is even active in 10% (*v*/*v*) SDS and activity was doubled in 10% (*v*/*v*) Triton-X100 with cyclohexanol. TADH's overall stability has also allowed it to be coupled with a hydrogenating ruthenium based organometallic catalysts to regenerate the nicotinamide cofactor (36). Activity assays performed at relatively high temperatures of 60°C, allowed for the organometallic catalyst to efficiently regenerate the cofactor.

1.3.3.2. The MDR from Candida parapsilosis (CPCR2)

Many ADH substrates of interest are quite often very hydrophobic and will precipitate out of solution on contact with aqueous buffer due to limited water solubility. This issue can occasionally be solved with organic co-solvents or biphasic systems, with the trade-off of reduced activity and stability. In the case of ADH-'A', despite it tolerating neat solvent or co-substrate, much of its activity was diminished.

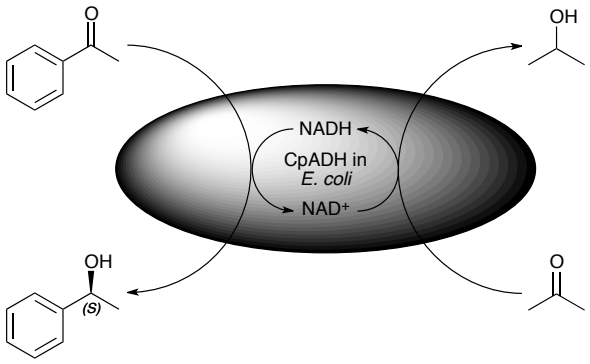


Figure 16 - Modified from Jakoblinnert and co-workers (37). Biocatalytic reduction of acetophenone using lyophilised *E. coli* cells with overexpressed CPCR2 in neat substrate with iso-propanol as a co-substrate.

CPCR2 is an NADH dependent, 36 kDa MDR with 336 amino acids from *Candida parapsilosis* DSM 70125. CPCR2 has a broad substrate spectrum producing enantiomerically pure (*S*)-alcohols. Purified CPCR2 is particularly unstable and rapidly degrades, however efforts to improve solvent and thermal stability by Site-Directed Mutagenesis (SDM) have been successful (38). CPCR2 has been applied as a whole cell catalyst (**Figure 16**) due to the lack of additional expense required with implementing cofactor regeneration systems, simplified product isolation and increased solvent tolerance. Lyophilised *E. coli* cells with CPCR2 overexpressed are capable of retaining all of its activity in neat substrate (37), as in Figure 14. Therefore, CPCR2 has opened up new biocatalytic routes to the production of synthetically useful building blocks with substrates previously not accessible.

1.4. Nicotinamide Cofactors

ADHs utilise nicotinamide cofactors (**Figure 17**), as mentioned in section 1.3. ADHs have a nucleotide recognition site, a Rossmann Fold (GxGxxG) allowing them to bind NAD(P)H. The reduced form of the cofactor is used for the carbonyl reductase reaction.

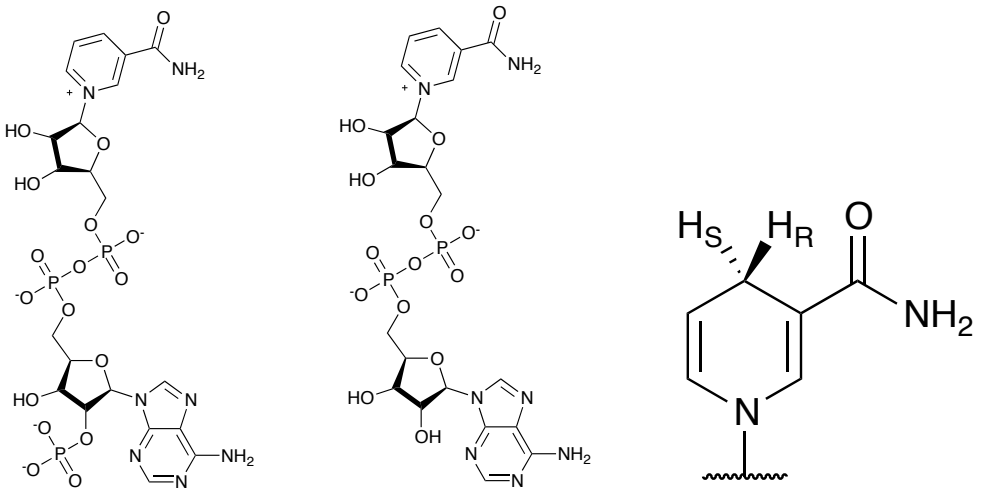


Figure 17 - Chemical structures of the nicotinamide cofactors, shown above are: NAD+ (left), NADP+ (middle) and an expanded view of the reduced form of the nicotinamide group with the prochiral hydrogens stereospecifically labelled (right).

The reduced form of the nicotinamide group follows the Cahn-Ingold-Prelog (CIP) priority rules (39,40) for α,β-unsaturated carbonyls (enone) about the Cα (**Figure 18** - left).

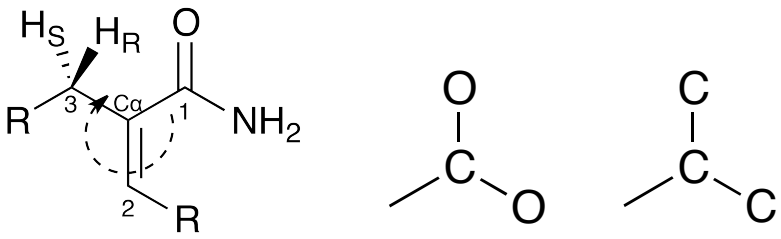
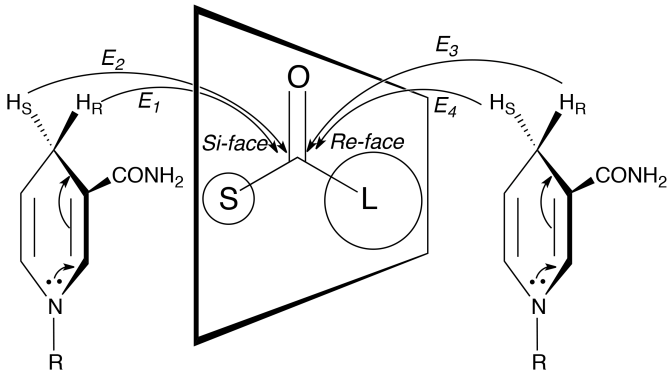


Figure 18 – Schematic of a generic enone showing the CIP priority resulting in the prochirality of the H_R and H_S hydrogens (left). Schematics of CIP equivalents of a carbonyl (middle) and an alkene (right).

Due to the CIP rules, assuming the substituent R's priority is less than oxygen (i.e. carbon or nitrogen), Cα-C=O has a higher priority than Cα=C-R; this is as the carbonyl equates to a carbon being bound to two oxygens, whereas an alkene equates to a carbon being bound to two carbons (**Figure 18** - middle and left). The CIP priority rules for enones give rise to prochirality around the hydrogens, previously been

mentioned by Prelog in 1946 (41), which describes the interesting stereochemical outcome for ADHs and their substrates. It was previously identified, by Bradshaw and co-workers in 1991 (42), that the different stereospecificities of alcohol dehydrogenases arise from how the nicotinamide cofactor is bound; therefore a specific hydride is delivered to a specific face (**Figure 19**) of the carbonyl (either the *re-* or *si-face*).



E1: *pro-R*/*si*-face, *Pseudomonas* sp. and *Lactobacillus kefir* ADH
E2: *pro-S*/*si*-face, *Mucor jávanicus* ADH
E3: *pro-R*/*re*-face, yeast ADH, HLADH and TbADH
E4: *pro-S*/*re*-Face, unknown (in 1991)

Figure 19 - Modified from Bradshaw and co-workers (42). Stereospecificity of ADHs where 'S' is a small group and 'L' is a large group.

Enzymes that utilise these cofactors are often selective for one or the other. Enzymes that are selective towards NADPH often have an arginine residue near the phosphate group at the 2' position of the adenine-ribose ring; the positively charged residue is effectively hugging the negatively charged phosphate group as with Arg200 in TbADH, PDB entry: 1KEV. Enzymes that are selective towards NADH have a glutamate or aspartate residue near the 2' position of the adenine-ribose ring; this would repel any negatively charged phosphate groups from cofactor binding pocket as with Asp223 in HLADH. Cofactor preference and comparison can be seen for TbADH and HLADH (**Figure 20**). However, there are a few cases where enzymes are able to utilise both cofactors. This is often due a glutamine residue (43) near the 2' position of the adenine-ribose ring; which, due to its neutral charge, permits the binding of both cofactors.

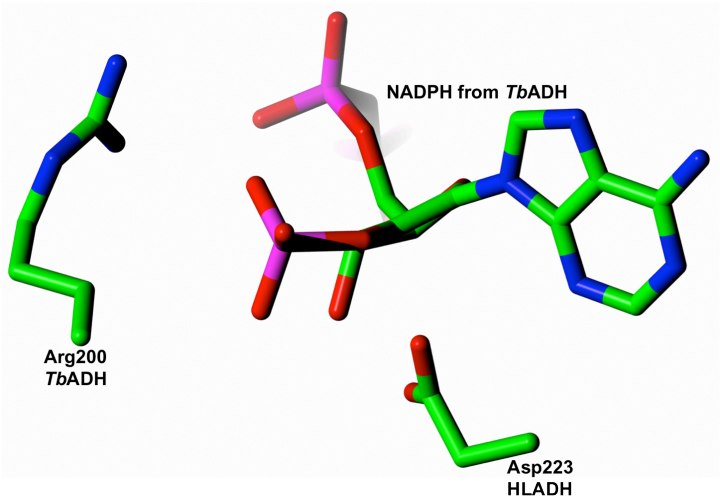


Figure 20 – Residues responsible for cofactor preference for TbADH (PDB entry: 1KEV) and HLADH (PDB entry: 6ADH). TbADH and HLADH are superposed on their cofactors and only the cofactor from TbADH is shown. HLADH would not allow NADPH to bind due to repulsion from Asp223, however in TbADH the phosphate and Arg200 makes a favourable interaction, therefore allowing binding.

Nicotinamide cofactors, despite their uses, are relatively expensive; however NADH is a tenth of the price of NADPH and NAD^+ is a fifth of the price of $NADP^+$. As such, attempts to reengineer cofactor specificity preference towards NADH have been attempted for ADHs (44) and other oxidoreductases (45,46), however it has often proven no simple task to simply mutate one residue to create an unfavourable interaction for one cofactor over the other. Reaction and conversion rates may be significantly lowered simply due to the cofactor no longer binding in the same position, creating a disturbance in preferential binding interaction between the cofactor, enzyme and substrate.

If changing the cofactor specificity has not been successful, it is still possible to use either cofactor efficiently at catalytic quantities with cofactor regeneration (47). Regeneration strategies often involve two enzyme systems or if available co-substrates are used to regenerate the cofactor (**Figure 21**).

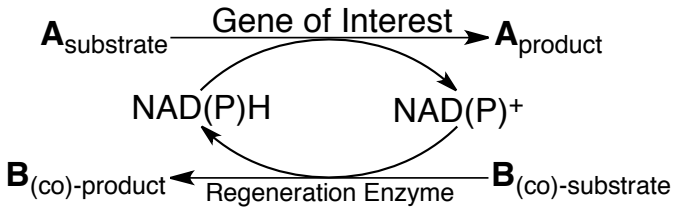


Figure 21 - Generic cofactor regeneration strategies for the reductive reaction of a gene of interest.

In the case of some ADHs it may be applicable to use ethanol/ethanal or iso-propanol/acetone as the co-substrate. If the ADH is solvent tolerant such as ADH-'A'

(18), co-solvents can be used in high concentrations which may also help solubilise hydrophobic substrates and benefit from cofactor regeneration. When co-substrate or co-solvent is not available, a second enzyme (**Figure 22**) is used in the reaction, such as Formate Dehydrogenase for NAD^+/NADH or Glucose-6-phosphate Dehydrogenase for $NADP^+$/NADPH.

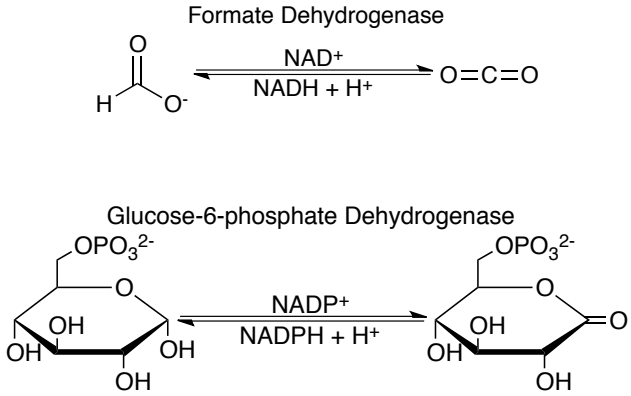


Figure 22 – Reactions of the two typically used recycling systems. Formate Dehydrogenase (top) and Glucose-6-phosphate Dehydrogenase (bottom).

1.5. Aim of the Project

The aim of this project is to gain an insight into the crystal structures of various alcohol dehydrogenases received from collaborators that have been demonstrated to have distinctly useful properties. This would enable us to rationalise the enzymes substrate specificity and allow for further protein engineering.

The gene encoding the SDR from *Ralstonia* sp., RasADH, was previously cloned, expressed and purified by Justyna Külig from The Institute of Bio- and Geosciences in Jülich, Germany. Some crystallisation screening and optimisation had been performed but with limited success. The first aim was to re-establish expression and purification protocols in a hope to crystallise the protein with more success and use the crystal structure to attempt to engineer novel substrate specificity.

The gene encoding the SDR from *Sphingobium yanoikuyae*, SyADH, was cloned, expressed, purified and crystallised by Kinga Kedziora from Universidad de Oviedo in Oviedo, Spain. A dataset was obtained from the optimised crystals. The second aim was to build the crystal structure and perform structural analysis, comparisons and ligand docking with the structure of RasADH.

The gene encoding the MDR from *Thermus* sp., TADH, was cloned, expressed, purified and crystallised by Serena Gargiulo from Delft University of Technology in Delft, The Netherlands. A dataset was obtained from the optimised crystals. The third aim was to build the crystal structure and perform structural analysis and comparisons with other thermophilic and mesophilic MDRs.

The gene encoding the MDR from *Candida parapsilosis*, CPCR2, was previously cloned (into pET-22b Strep-tag vector), expressed and purified by Christoph Loderer from Technische Universität Dresden in Dresden, Germany. Crystallisation screening and optimisation had been performed on the Strep-tag construct, however limited success was observed. The fourth aim was to subclone CPCR2 into the pET-YSBLIC3C vector (48), re-establish expression and purification protocols and re-crystallise the protein to obtain a dataset and crystal structure with NADH.

In gaining the crystal structure for each of these ADHs, they will provide the basis for rational engineering, to further improve the substrate specificity,

enantioselectivity, cofactor preference and enzyme stability for the potential use of industrial biotechnology.

2. Methods

2.1 Cloning

The Structural Proteomics In Europe (SPINE) consortium was a collaboration between 14 laboratories across Europe that developed a whole range of ligation independent cloning (LIC) and In-Fusion compatible *E. coli* vectors using high-throughput cloning methods, with a variety of cleavable and non-cleavable tags (49,50). Through SPINE, the pET-YSBLIC and pET-YSBLIC3C was developed. Both vectors were adapted from pET-28a, where a LIC site was incorporated to allow for cloning without the use of restriction endonucleases or DNA ligases. LIC allows for a faster and higher success rate of cloning.

Genes of interest were routinely cloned only using LIC (**Figure 23**). The template gene of interest received from collaborators was initially amplified with appropriate primers designed using the HiTel Primer Design (http://bioltfws1.york.ac.uk/cgi-bin/primers.cgi) for LIC.

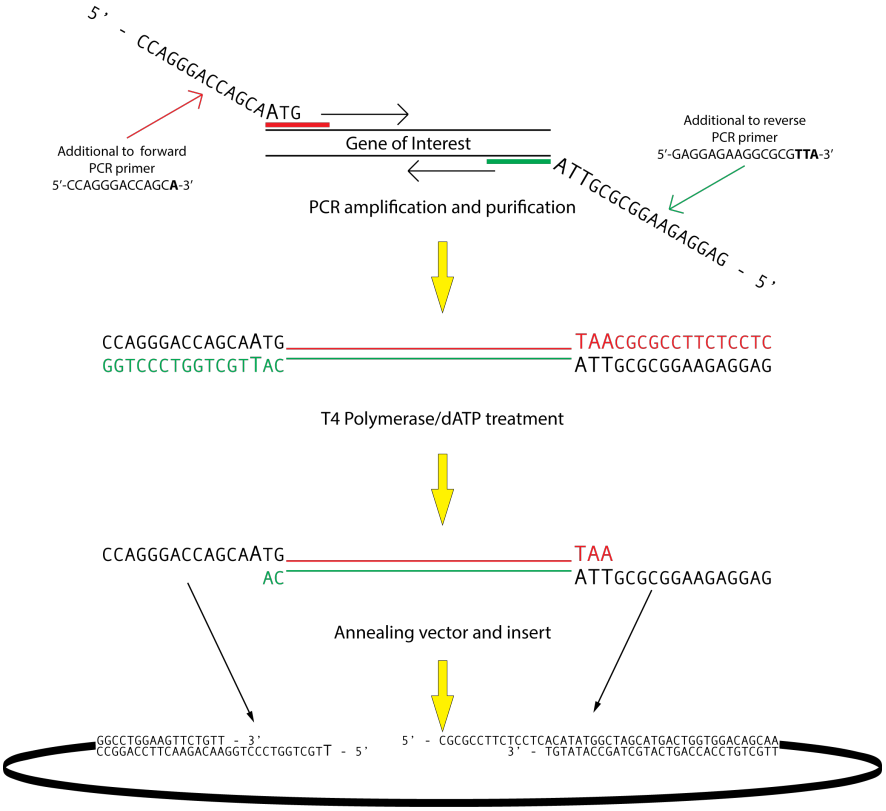


Figure 23 – The schematic for preparation of the PCR fragment and annealing to the pET-YSBLIC3C vector. Adapted from Fogg and Wilkinson (48).

2.1.1. Polymerase Chain Reaction (PCR)

The gene of interest was amplified by polymerase chain reaction using a PCR thermal cycler. The following components were used in the PCR:

Component	DMSO Reaction	Reaction
Template (50-100 ng/µL)	1 µL	1 µL
For Primer (20 pmol/µL)	1 µL	1 µL
Rev Primer (20 pmol/µL)	1 µL	1 µL
dNTPs (2 mM)	5 µL	5 µL
$MgSO_4$	3 µL	3 µL
KOD Polymerase Buffer (10x)	5 µL	5 µL
DMSO	1 µL	0 µL
KOD Polymerase	1 µL	1 µL
ddH_2O	32 µL	33 µL

The reaction was performed in a PCR thermal cycler using the following steps/cycles:

Temperature	Time	Step	Cycles
94°C	2 min	Initial denature	1
94°C	30 s	Denature	35
50°C	30 s	Anneal	
72°C	30 s	Extend	
72°C	3 min	Final extension	1
4°C	∞	Cool	

2.1.2. Fragment Analysis by Agarose Gel Electrophoresis

Agarose gels are commonly used to analyse DNA fragments where the separation is based on size. However for very large proteins > 200 kDa it may be appropriate to use agarose gel electrophoresis (51).

To 0.6 g agarose, 60 mL TAE buffer was added and the mixture was microwaved at maximum heat for 1 min. Once cool enough to handle, 1 µL SYBR Safe was added to the mixture and swirled. The agarose was then poured into a caster, where a 6-well comb was inserted and allowed to set for 1 h.

After the gel was set, the comb was removed and the gel was placed in a gel tank where sufficient buffer was added to submerge the gel. Samples were added to the wells formed by the comb and the gel was run for 1 h at 100 V.

2.1.3. Preparing Samples for Agarose Gel

Prior to loading samples to the agarose gel, dye was added to the samples to be able to image the sample on the gel.

1.5 µL of 1 kb DNA ladder was prepared by adding 3 µL ddH_2O and 3 µL 6x gel loading dye. 10 µL gel loading dye was added to 25 µL reaction sample.

1 kb DNA Ladder and 6x (blue) gel loading dye was purchased from New England BioLabs.

2.1.4. DNA Purification: Gel Extraction (Spin-Down Method)

After a PCR has been completed, the amplified DNA must be purified. Purification methods include PCR clean-up and gel extraction. Gel extraction was the only method used.

For the following protocol, all centrifugation were performed at 13.3k rpm:

Successful DNA fragments analysed on agarose gel are excised and weighed. The gel was dissolved with 3 gel volumes of solubilising solution, where 1 gel volume is 100 µL solution for every 100 mg agarose. The mixture may be incubated with warm water and vortexed to ease dissolution. Once dissolved the gel mixture should be yellow. 1 gel volume of 100% isopropanol was added to the gel mixture and was mixed.

The binding column was prepared whilst the gel is dissolving. 500 µL column preparation solution was added to the binding column and was centrifuged for 1 min. The flow through was discarded.

DNA was then bound to the column by loading the dissolved mixture to the column and centrifuging for 1 min with discarding of the flow through. This was repeated until all the mixture has eluted. After binding, the bound DNA was washed with 700 µL wash solution and centrifuged for 1 min.

The binding column was transferred to a fresh collection tube with 50 µL ddH_2O (@ 60°C) added directly onto the membrane for 30 min. After 30 min, DNA was eluted with centrifugation for 1 min. DNA concentration should be checked after elution.

2.1.5. T4 DNA Polymerase Digest Reaction with Insert

After a PCR clean up, the inserts' overhangs were digested to create sticky ends to allow for annealing to the desired vector. To 0.2 pmol of insert, the following components were added:

Volume	Component
2 µL	T4 Buffer (10x)
2 µL	dATP (25 mM)
0.5 µL	DTT (100 mM)
0.5 µL	T4 DNA Polymerase
12 µL	ddH_2O

The reaction was incubated at 20°C for 30 min, followed by 75°C for 20 min.

2.1.6. Annealing Insert to Vector (pET-YSBLIC3C)

Once the sticky ends were formed from the digest reaction, the insert can be annealed to a linearised vector with complementary stick ends. LIC does not require DNA Ligase for a ligation step as required with standard molecular cloning.

3 µL of insert was incubated with 3 µL linearised pET-YSBLIC3C vector for 10 min at room temperature. Following that, 1.5 µL EDTA (25 mM) was added to the mixture and left for another 10 min at room temperature.

2.1.7. Preparing agar plates

Agar plates are routinely used to selectively grow microorganisms. There is a range of media available for agar plates, however only LB-agar plates are used.

200 mL LB-agar was microwaved at maximum heat in 1 min bursts with swirling in-between until all was dissolved. It was then allowed to cool until it could be handled by hand, then antibiotics (i.e. at 34 mg/mL kanamycin and or chloramphenicol) were added in a 1 in 1000 fold dilution and the agar was swirled to mix. The agar could then be poured onto petri dishes using aseptic technique and was allowed to set for approximately 1 h.

2.1.8. Transformation into Cloning Strains

Competent cells are routinely used for transformation – the uptake of DNA into cells. There are many different strains of (*E. coli*) cells commercially available, however the only cloning strain used were NovaBlue Singles competent cells.

Once the annealing reaction was complete, it was transformed into cloning strain cells. 6 µL annealing reaction was added to 25 µL cells (i.e. NovaBlue Singles) and was left on ice for 5 min. The mixture was heat shocked at 42°C for 45 s. Following the heat shock, the cells were left on ice for another 5 min. 150 µL SOC (52) medium was added to the mixture and left to incubate at 37°C for 1 h in an orbital shaker. The cells were spread, using aseptic technique, on kanamycin-LB-agar plates and was left to incubate in a 37°C oven overnight.

2.1.9. Starter Culture

A starter culture is routinely used prior to inoculating media for further experimentation or for plasmid extraction. This is as bacteria utilise majority of the nutrients available in the initial starter culture reaching the exponential phase. Therefore a growing culture plucked from a single colony in a flask with 1 L of LB medium would not only take a very long time to reach a satisfactory OD_{600} for induction, also a lot of the nutrients in the media would have been already used up, thus expression of the desired protein is not as optimal as it could be. Upon scaling up culture growth, starter cultures can be made in a 1 in 100 fold dilution of the media volume; i.e. 10 mL starter culture for an intended 1 L culture.

The following was performed with aseptic technique:

5 mL of autoclaved LB (53) medium with antibiotics (i.e. kanamycin at 34 mg/mL) added in a 1 in 1000 fold dilution was added to a 50 mL Falcon tube. A colony from a successful transformation on an agar-plate was plucked using a pipette tip and was ejected into the Falcon tube. The Falcon tubes were then incubated in an orbital shaker at 37°C overnight.

2.1.10. Plasmid Extraction: Mini-Prep

Overnight cultures from the plasmid in the cloning strain are used to generate more copies of the plasmid. The plasmids are then extracted from the cells and purified, this allows the plasmid to be stored for future use. This is often performed using a MiniPrep kit which is commercially available.

For the following protocol, all centrifugations were performed at 13.3k rpm:

The starter culture was initially pelleted into a collection tube with the supernatant discarded. The pellet was resuspended with 200 µL resuspension

solution and vortexing. 200 µL lysis solution was added to the resuspended pellet and was gently mixed by inverting the collection tube 4-6 times. The lysate was allowed to clarify for 5 min. After the lysate was clarified, 350 µL neutralising solution was added and was gently mixed again by inversion. The lysate was then centrifuged for 10 min to pellet the cell debris.

The binding column was prepared while the cell debris was pelleted; 500 µL column preparation solution was added to the column in a collection tube and was centrifuged for 1 min with the flow through discarded. The clear lysate was transferred to the column and centrifuged for 1 min with the flow through discarded. The column was centrifuged again for another 1 min to dry the column. The column was transferred to a fresh collection tube and 50 µL ddH_2O (at 60°C) was added directly to the membrane for 30 min. Finally after 30 min, the plasmid was eluted by centrifuging for 1 min. dsDNA concentration was checked after this.

2.1.11. Double Digest

Double digests are often performed to check whether the insert has annealed to the vector or not. The vector codes specific sites for cleavage by a specific enzyme. As long as the gene of interest does not code for the same cleavage site, it can be used for digestion. There are many resources available to check for these sites, such as Serial Cloner by SerialBasics. If the double digest is successful, it can be sent for sequencing for sequence analysis.

After plasmid extraction 7 µL plasmid was required for the double digest. 1 µL 5' restriction digest enzyme (i.e. Nco 1), 1 µL 3' restriction digest enzyme (i.e. Nde 1) and 1 µL suitable buffer (i.e. Buffer 4) was added to the plasmid. The reaction mixture was then incubated for 1-3 h (no longer than 3 h) at 37°C. The samples were prepared for DNA fragment analysis as mentioned in section 2.1.3. then checked on a 1% agarose gel as mentioned in section 2.1.2.

2.2. Site-Directed Mutagenesis (SDM)

SDM is a commonly used technique, for example, to perform a substitution of an amino acid in the active site. As well as eliminating activity, SDM can be used to improve activity and or modify the size of the existing active site to specifically accommodate a desired substrate. SDM also uses PCR, however it differs in that it does not only amplify the gene or site of interest. SDM uses very specific primers to amplify the whole plasmid. Therefore, the original plasmid is the template.

The mutant primers were designed using the Primer X (http://www.bioinformatics.org/primerx/index.htm). Protein-based primer design with default options was selected for the mutation performed and mentioned in the latter results.

2.2.1. Mutagenesis-PCR

For mutagenesis-PCR, the following reaction components were used:

Component	Reaction
Template (50-100 ng/µL)	1 µL
For Primer (20 pmol/µL)	0.5 µL
Rev Primer (20 pmol/µL)	0.5 µL
dNTPs (2 mM)	5 µL
Pfu Polymerase Buffer (10x)	5 µL
DMSO	1 µL
Pfu DNA Polymerase	1 µL
ddH_2O	36 µL

The reaction was performed in a PCR thermal cycler using the following steps/cycles:

Temperature	Time	Step	Cycles
95°C	5 min	Initial denature	1
95°C	1 min	Denature	30
50°C	1 min	Anneal	
72°C	10 min	Extend	
95°C	72 s	Final denature	1
72°C	5 min	Final extension	1
4°C	∞	Cool	

The reaction components and reaction parameters were modified based on the protocol for Pfu DNA polymerase.

2.2.2. Fragment Analysis by Agarose Gel Electrophoresis

Agarose gels can be used to analyse the mutagenesis-PCR products. It can be an indication to whether the primers or PCR conditions were suitable.

10 µL of the mutagenesis-PCR products were taken for agarose gel analysis. The gel was prepared as in (Section 2.1.2.) and the samples prepared as in (Section 2.1.3.).

2.2.3. Dpn1 Digest

After the mutagenesis-PCR is complete, there is still a mixture of template plasmid along with the mutant plasmid. DNA produced from *E. coli* (the template plasmid) is methylated on the N6 position of adenine (m6A) by enzymes in *E. coli* such as methyltransferases. However the mutant plasmid produced in the PCR will not be methylated, as the DNA polymerase used will not methylate DNA. Therefore restriction digest enzymes such as Dpn1 can be used to digest the methylated template plasmid. The DNA recognition site for Dpn1 is Gm6A-TC, where Dpn1 cleaves between N6-methyladenine and thymine.

40-50 µL PCR product was taken for Dpn1 digest. 1.5 µL Dpn1 restriction enzyme was added to the product and was incubated at 37°C for up to 6 h.

2.2.4. Transformation of Cloning Strains by Mutant Plasmids

Typically 4 h is enough time for the Dpn1 digest reaction to digest the template. However it can be difficult to tell if the template plasmid has been sufficiently

digested. Therefore, performing transformations at 4 hours and 6 hours will give a better indication of sufficient template digestion.

Kanamycin-agar plates were prepared as in (Section 2.1.7.). 1.5 µL digested PCR product was taken to transform NovaBlue Single Cells, as in (Section 2.1.8.), at 4 hours and 6 h of the digest reaction.

2.2.5. Starter Culture

Starter cultures of the mutant plasmid in NovaBlue Single Cells were made as in (Section 2.1.9.) from successfully grown colonies from (Section 2.2.4.).

2.2.6. Plasmid Extraction: Mini-Prep

The plasmids from the starter cultures made in (Section 2.2.5) were extracted using a MiniPrep kit following the protocol in (Section 2.1.1.10.).

The extracted mutant plasmid can now be made up to the appropriate concentration for gene sequencing analysis and transformations in expression strains.

2.2.7. Double Digest

A double digest can be optionally performed with the template plasmid as a control and the mutant plasmid as in (Section 2.1.11.) with fragment analysis by agarose gel electrophoresis (Section 2.1.2.) and sample preparation (Section 2.1.3.).

2.3. Gene Expression

Expression of the gene of interest may often not be a simple matter; either low yielding expression or majority of the expressed protein is insoluble. Proteins overexpressed in *E. coli* are often insoluble and are contained in inclusion bodies; it is possible to solubilise and re-fold the protein, however this is often time consuming. There are many other expression systems available, such as baculovirus expression in insect cells for large protein complexes, *Pichia pastoris* and *Bacillus subtilis* that allow for protein secretion into the media; these are only a few examples of the many different options available, however each expression system has its own advantages and disadvantages.

2.3.1. Transformation of Expression Strains

Transformations into expression strains such *E. coli* are similar to that of transformation into cloning strains (Section 2.1.8.), however expression strains are optimised for expression. As such, the protocol differs slightly:

1 µL plasmid was added to 25 µL cells, such as BL21(DE3) or Rosetta 2(DE3), and was left on ice for 20 min. The mixture was heat shocked at 42°C for 45 s. Following the heat shock, the cells were left on ice for 1 min. 1 mL LB medium was added to the mixture and left to incubate at 37°C for 1 h in an orbital shaker. The cells were pelleted by centrifuging at 13.3k rpm for 1 min, 800 µL media was discarded and the cells were resuspended in the remaining 200 µL media. 100 µL cells were taken to spread, using aseptic technique, on each appropriate antibiotic-LB-agar plates and was left to incubate in a 37°C oven overnight.

2.3.2. Buffer Components

Buffers can dramatically change the activity and stability of proteins. Certain buffers are known to change pK_a on increasing or decreasing temperature and are also known to inhibit certain enzymes activity and are seen in the crystal structure of an enzymes active site. With very sensitive and unstable proteins, it may be necessary to screen a range of buffers and optimise conditions prior to protein purification.

Tris buffers were used for all enzymes purified as it has a very large effective buffering range between pH 7.1 and 9.0. Glycerol is added to the buffers to help stabilise sensitive proteins; glycerol is generally thought of to mimic the natural viscous environment inside a cell.

Buffer A was used as the resuspension buffer; 20 mM imidazole was incorporated to decrease non-specific binding to IMAC columns when loading the cell lysate. A gradient between buffer B and C was used to elute the protein bound to the IMAC column. The following components were used for all buffers:

	Buffers		
Component	**A**	**B**	**C**
NaCl	500 mM	500 mM	500 mM
Tris-HCl pH 8.0	50 mM	50 mM	50 mM
Imidazole	20 mM	0 mM	500 mM
Glycerol	10% v/v	10% v/v	10% v/v

1 mM calcium chloride was added to all buffers for RasADH, as in Kulig's study (27), calcium chloride was seen to stabilise and increase activity for RasADH.

2.3.3. Protein Analysis by Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis

Proteins are routinely denatured by Sodium dodecylsulfate (SDS) and are analysed by SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE); which proteins are separated on the basis of charge. SDS binds to the protein backbone unfolding and

therefore denaturing the protein and in high concentrations micelles would be formed around the protein backbone. When SDS is bound to the protein, it causes the overall charge of the protein to be negative; this allows the denatured protein to migrate by electrophoresis. The principle is similar to DNA fragment analysis by agarose gel, however protein is too small be analysed by agarose due to larger pore sizes in agarose. SDS-PAGE is an easy method to analyse protein purity from fractions obtained from purification.

The mixture for a 12% SDS-PAGE gel is shown in the following table:

12% Polyacrylamide Gel

Resolving Gel Mixture	**Stacking Gel Mixture**
2.5 mL Resolving gel buffer	1.3 mL Stacking gel buffer
4.2 mL Acrylamide Stock	0.5 mL Acrylamide stock
3.2 mL ddH2O	3.2 mL ddH2O
16 uL TEMED	8 uL TEMED
100 uL 10% APS	100 uL 10% APS

Initially a sandwich of glass plates with plastic gel spacers and aluminum backing was made and mounted vertically into the gel stand. It was checked for leaks by filling the assembly with ethanol between the glass plates. Once it was satisfied that there were no leaks, ethanol was removed and the assembly was dried with nitrogen air source. The resolving gel mix was poured between the glass plates approximately two-thirds full, the top was evened out with butanol and the gel was left to set for 10 min. Once set, the butanol was removed and dried with a nitrogen air source, the remainder of the glass plates was filled with stacking gel mix and the plastic gel comb (10 or 15 well) was immediately inserted. The stacking gel was left to set for 10 min. The solidified gel was removed from the caster and placed into a gel tank where SDS running buffer was filled to submerge the gel. The plastic comb was removed and samples were loaded into the wells formed by the comb. The power pack was set to 200 V and the gel was run for 50 min.

After the gel had finished running, the gel sandwich was removed from the apparatus and the stacking gel was cut off with the spacer. The resolving gel was placed in a plastic box, 100 mL Coommassie Brilliant Blue were added to the gel and the dye was microwaved at maximum heat for 2 min to stain. The dye was removed

and the gel was rinsed with cold water. 100 mL destain solution was added to the gel and was microwaved for 2 min. Tissue was placed on top of the gel to help absorb the dye from the gel for 1 h before imaging.

2.3.3.1 Sample preparation for SDS-PAGE analysis

Prior to loading the samples into the wells in the gel, dye must be added to the protein samples in order to be able to image the protein and the samples must be denatured for SDS-PAGE analysis.

10 µL of 2x SDS loading buffer with dye was added to each 10 µL protein sample in PCR tubes. Samples were boiled for 5 min at 94°C using a PCR thermal cycler, including a 10 µL solution of low molecular weight ladder from Bio-Rad. The samples should be centrifuged at 13.3k rpm for 30 seconds to collect the entire denatured sample. The ladder was added to the first well on the left of the gel using a HPLC syringe. The syringe was washed thoroughly with SDS running buffer and 10 µL of the first protein sample was transferred to the well adjacent to the ladder. This was repeated for the remaining protein samples. Finally the lid was placed on top of the apparatus and the gel was ran for 50 min at 200 V.

2.3.4. Optimising Gene Expression

In cases where over expression yields low amounts of protein, it is common to tests different expression strains at different temperatures. Once a satisfactory strain and growth temperature has been found, it may be required to perform an induction test to further optimise protein expression. It is often suggested to induce the culture at OD_{600} ≈ 0.6 Abs, as it is said to be the mid-log phase, which should yield greatest expression; however this is not necessarily the case, in well-controlled conditions with plenty of aeration, such as using fermenters, inducing much later at OD_{600} ≈ 4.0 Abs yields best results.

2.3.4.1. Expression Test

An expression test is a commonly performed experiment to optimise protein expression in different strains and at different temperatures. It should be performed before a scale up above 10 mL is attempted, such that optimal results are obtained.

Prior to this experiment, a 5 mL starter culture was made as described in section 2.1.9.

With aseptic technique: four 50 mL Falcon tubes with 5 mL autoclaved LB media with antibiotics (in 1 in 1000 fold dilution) were prepared. Each Falcon tube was inoculated with 50 µL starter culture (1 in 100 fold dilution of the volume) and incubated at 37°C until $OD_{600} \approx 0.6$. Once a satisfactory optical density is achieved the control was not induced and was incubated overnight at 16°C. The remaining three Falcon tubes were induced with 1 mM final concentration IPTG and incubated at 16°C, 30°C and 37°C overnight.

On the following day each sample was pelleted by centrifugation at 13.3k rpm for 1 min into separate 2 mL Eppendorf tubes. Each pellet was resuspended by vortexing with 500 µL buffer A and was lysed by sonication whilst on ice throughout lysis. The cell debris was then pelleted by centrifugation at 13.3k rpm for 5 min and each clarified lysate (soluble fraction) is transferred to a new 2 mL Eppendorf tube. The remaining cell debris was resuspended by vortexing with 500 µL buffer A. 10 µL of each fraction at each temperature was prepared and analysed by SDS-PAGE analysis as described in section 2.2.3. and 2.2.3.1.

2.3.4.2. Induction Test

After a suitable expression strain has been found and gene expression yields are still not satisfactory, testing different points for induction can be performed to further optimise gene expression without reengineering the gene.

Prior to this experiment, one 5 mL starter culture of the desired expression strain was made as described in section 2.1.1.9.

With aseptic technique: twelve 50 mL Falcon tubes with 5 mL autoclaved LB media with antibiotics (in 1 in 1000 fold dilution) were prepared. Each Falcon tube was inoculated with 50 µL starter culture (1 in 100 fold dilution of the volume) and incubated at 37°C. Each culture was induced with 1 mM final concentration IPTG at the following OD_{600}: Negative control (no induction), positive control (induction upon inoculation), 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8 and 2.0. Testing beyond OD_{600} = 2.0 may be detrimental to the growth and expression of cells in a culture flask. After induction, the culture samples were incubated overnight in an orbital shaker at the optimal temperature found in the expression test as mentioned in section 2.2.4.1.

On the following day each sample was pelleted by centrifugation at 13.3k rpm for 1 min into separate 2 mL Eppendorf tubes. Each pellet was resuspended with a lysis solution, i.e. BugBuster Protein Extraction Reagent, and the cell debris is then pelleted by centrifugation at 13.3k rpm for 5 min. 10 µL of each sample was prepared and analysed by SDS-PAGE analysis as described in section 2.2.3. and 2.2.3.1.

2.3.5. Scaling Up Protein Production

Once optimised expression protocols have been established, experiments that require larger quantities of protein can be performed by scaling up the culture volume. However behaviour in a 5 mL culture can be quite different to a 1 L culture.

For 1 L of LB media, 25 g LB media were added to 1 L ddH_2O in a 2 L Erlenmeyer flask. The flasks were stopped with foam bungs and covered with foil.

The flasks were then autoclaved and the flasks should be sufficiently cooled prior to use with bacterial growth. For each flask of LB, starter cultures with 1 in 100 fold dilution of the LB volume were made as described in section 2.1.1.9.

Antibiotics (at a 1 in 1000 fold dilution) were added to the flasks and were inoculated with the starter cultures using aseptic technique. The flasks were the incubated at 37°C until a satisfactory OD_{600} is achieved and induced with 1 mM final concentration IPTG. The culture is then incubated at the optimal temperature as investigated in section 2.2.4.1.

2.3.6. Cell Lysis

There are many methods available for lysing cells; such as sonication or French pressure cell press, which are mechanical lysis, whereas BugBuster Protein Extraction Reagent and buffers with lysozyme and DNAse are chemical methods of lysing. The method of lysis may be chosen based on the vector, as if the protein has the PelB promoter it may only be necessary to lyse the cells with chemical methods if the protein is secreted in between the cell membrane. However if protein expression is within the cell cytoplasm a mechanical method may be required. Certain organisms chosen for expression may not require cell lysis if the protein is secreted into the media.

Prior to lysing, cells are pelleted (i.e. 8k rpm for 15 min on a JLA-8.1000 rotor) to remove the excess media.

2.3.6.1. Sonication:

Each cell pellet formed from 1 L culture was resuspended in 30 mL buffer A. Resuspended cells were sonicated using a suitable probe, between 12-14 amplitude microns whilst on ice to rapidly cool with the following procedure:

Time	Step	Cycles
45 secs 30 secs	Sonicate Rest on Ice	10 Cycles
1 min	Sonicate	Once

Following sonication, cell debris was pelleted with high-speed centrifugation (i.e. 20k rpm for 15 min on a SS34 rotor).

2.3.6.2. Lysozyme:

Each cell pellet formed from 1 L culture was resuspended in 30 mL buffer A with the addition of 15 mg Lysozyme and 1.5 mg DNAse (per 30 mL of buffer). Cell suspension was incubated in an icebox in an orbital shaker – alternatively in an orbital shaker at 4°C. Following lysis, cell debris was pelleted at high-speed centrifugation (i.e. 20k rpm for 15 min on a SS34 rotor).

2.4. Protein Purification

Depending on the protein, purification can be very challenging. The protein could be sensitive to high concentrations of elution buffer causing protein aggregation. High protein concentration could also cause protein aggregation. There are many techniques available for protein purification, however due to the development of fusing hexa histidine-tags (His-tags) to the protein on the DNA level (54), this has greatly simplified the technique. As well as His-tags there are many other tags available, which can be fused to the protein on either N-terminus or C-terminus. These tags have greatly eased and simplified purification, however achieving satisfactory purification may still be a problem and other purification methods and strategies may be required. The method for purifying tagged proteins are known as affinity chromatography; as the tagged-protein binds highly specifically to an immobilised matrix and are eluted by an increasing concentration of a buffer that will out compete the binding of the protein.

2.4.1. Stripping and Recharging an IMAC Column

Prior to reusing a column, it should be stripped and recharged to ensure any impurities bound to the charged resin media will be removed. Columns such as HisTrap Crude FF, HisTrap HP, HisTrap FF, HisTrap excel, HiTrap Chelating HP, HiTrap IMAC HP or HiTrap IMAC Sepharose FF columns can be used for immobilised metal affinity chromatography (IMAC).

The columns are sequentially washed with 5 column volumes (CVs) of: ddH_2O, 0.1 M EDTA with 0.5 M NaCl, H_2O, 0.1 M $NiSO_4$ or 0.1 M $CoSO_4$ and buffer A (or 20% ethanol for storage).

2.4.2. Loading and Recycling Protein

Protein lysate can be loaded onto a column using different methods, however loading the lysate onto the column only once may not maximise the binding efficiency. Therefore recirculating/recycling the lysate should maximise protein binding to the column. Clarified lysate can be loaded onto a column with a peristaltic pump, Superloop, injecting using a loop, using an AKTA sample valve or with a combination of available methods.

Clarified cell lysate was cooled throughout on ice and was loaded onto a column (i.e. 5 mL HisTrap Crude FF) using a peristaltic pump at 5 mL/min. To maximise binding of protein; the lysate was allowed to recirculate back into the lysate and recycle onto the column (**Figure 24** - left), for every 60 mL of lysate would equate to 45 min recirculation.

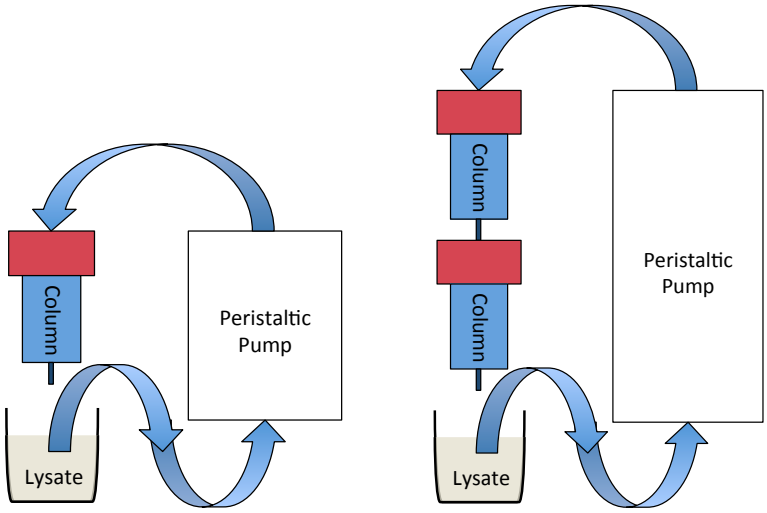


Figure 24 - Schematic of recirculating protein onto a column (left) and schematic of 'daisy-chaining' columns with recirculation (right).

If more of than 3 L of culture were grown, it is possible that the binding capacity of the column is exceeded, ~ 40 mg of protein per ml of medium (i.e. ~ 200 mg for a 5 mL column). Therefore it may be applicable to 'daisy-chain' columns and recycle onto two columns as in the schematic (**Figure 24** - right). After loading, the columns are purified separately on an AKTA, not to increase the void-volume of the column and to keep the highest resolution possible. Not only does 'daisy-chaining' increase

the potential protein yield, it also allows for fewer impurities to bind to a single column, therefore protein purity is significantly increased (**Figure 25**).

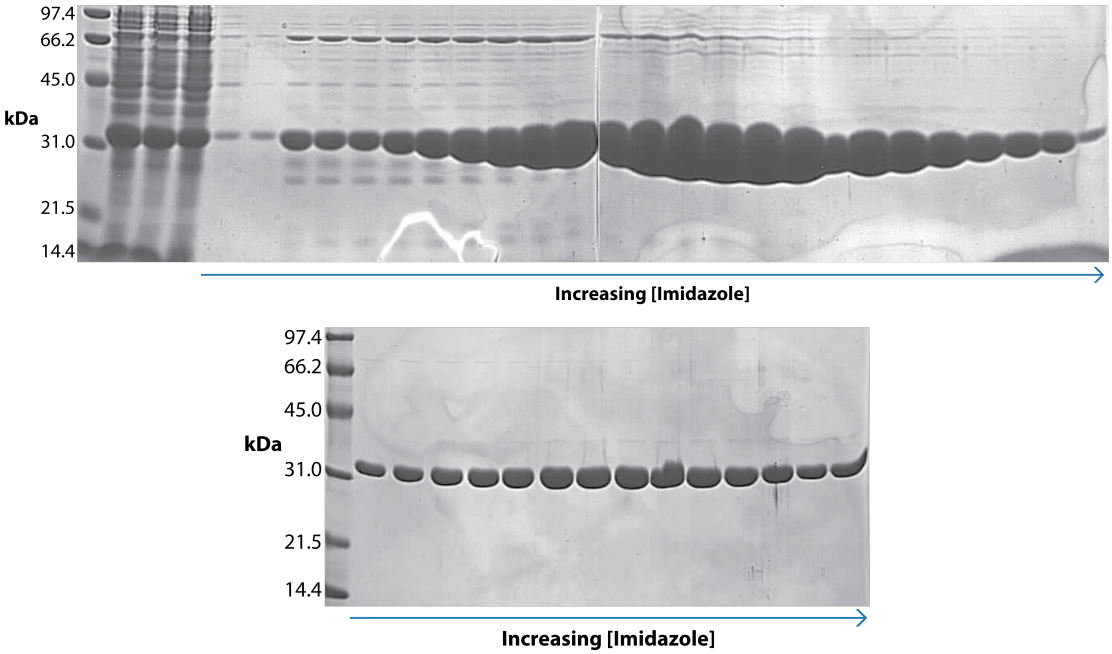


Figure 25 - 12% SDS-PAGE analysis of different purification strategies of nickel purification, both from 180 mL lysate of the same construct: Single column recycling and purification (top) and 'Daisy-chaining' two columns recycling and purification (bottom)

After recycling, the protein may be loaded onto the column again via the sample valve/pump on the AKTA Explorer to further maximise any potential binding. This is effectively the same as a Superloop, however the flow rate for protein binding to the column is controlled by the AKTA better than a Superloop, preventing any pressure issues.

2.4.3. Cobalt/Nickel Purification

By out-competing the protein and impurities bound to the charged resin matrix with a suitable buffer, it is possible to purify the protein. Therefore the standard method for IMAC purification, are wash steps, followed by a gradient and then a final wash step.

Buffers B (0 mM imidazole) and C (500 mM imidazole) are used to purify metal affinity columns (**Figure 26**). Typically, the first two wash steps are collected in 5 CV

fractions in Falcon tubes and the rest is collected in 96 well plates. The fraction size for the elution depends on the desired resolution and flow rate.

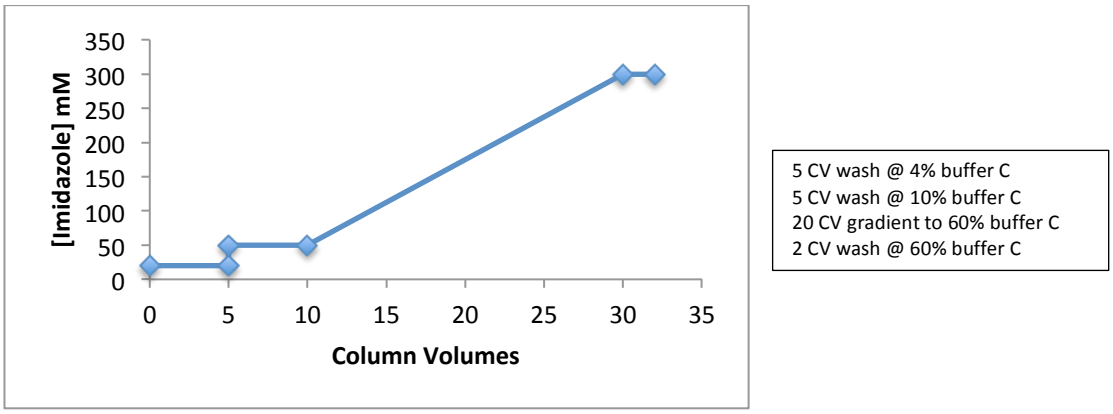


Figure 26 – Schematic for IMAC purification strategy that was used on AKTA's. Two initial washes are performed by a step gradient to wash away any impurities, while the gradient purifies protein.

2.4.4. Size Exclusion Chromatography (SEC)

Size exclusion chromatography or gel filtration is based on separation of size. This is a native technique, therefore if available, a multiangle laser light scattering (MALLS) module can be added to an AKTA to gain more information about the protein in solution whilst eluting, such as molecular mass. This can be useful if a calibration curve is known for a SEC column and the protein is eluting in a region heavier than it should be. This is quite common if the protein exists as an oligomer in solution (ADHs are commonly known to be dimers and tetramers). Even after SDS-PAGE analysis, it is possible to still see bands double the molecular weight of the monomer as the dimerisation or oligomerisation interaction may be very strong. Occasionally monomers can be separated from the oligomers during SEC and they can appear in very low concentrations, however, the monomer may not be a stable form and may oligomerise or form aggregates.

Prior to use, the SEC column must be equilibrated the night before. SEC columns are stored in 20% ethanol after use, therefore 1 CV wash of water, followed by 1 CV buffer A was required to equilibrate the column (**Figure 27**). To store the column in 20% ethanol, the same program was used, except 1 CV 20% ethanol was used rather

than 1 CV water. HiLoad 16/60 Superdex 75 PrepGrade column was used for SEC for all proteins purified.

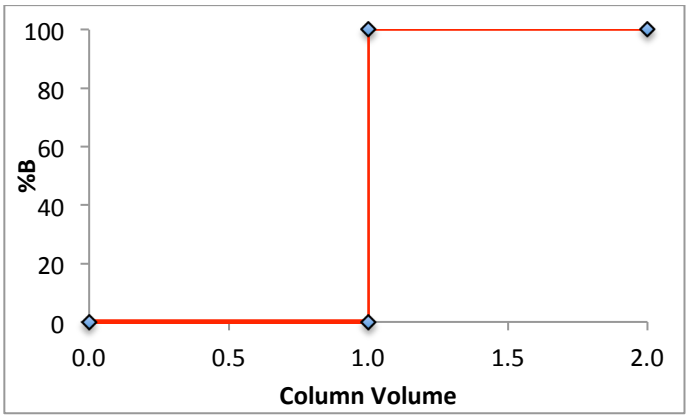


Figure 27 – Schematic of a generic equilibration of SEC columns. Typically 1 column volume of water is required, followed by 1 column volume of a buffer with 0 mM imidazole.

The maximum load capacity for the HiLoad 16/60 Superdex 75 PrepGrade column is ideally ≤ 2 mL; loading is typically performed via injecting protein into a 2 mL loop. The purification was performed by eluting with 1 CV of buffer A at 0.8 mL/min flow rate, however if the column or AKTA presents pressure issues, the flow rate can be reduced. The first 20 mL were collected in a Falcon tube, anything that elutes at this region are usually aggregates or species larger than 75 kDa which was exceeding the column's limits of 75 kDa. After 20 mL, fraction sizes are chosen based on flow rate and desired resolution, typically 1-2 mL fraction sizes are used.

2.4.5. HisTag Removal Purification

The pET-YSBLIC3C vector has a cleavable his-tag that can be digested by 3C Protease. The vector encodes for additional amino acids at the N-terminus as the hexa his-tag:

-MetGlySerSerHisHisHisHisHisHisSerSerGly**LeuGluValLeuPheGln-GlyPro**Ala-

The sequence in bold and underlined is the site that the 3C Protease (a serine protease) recognises and it cleaves between glutamine and glycine. The digest can

be performed after SEC or after the initial IMAC, however if it was performed after the initial IMAC, the samples imidazole should be dialysed out with the protease and a 0 mM imidazole buffer or buffer exchanged into a 0 mM imidazole buffer prior to the digest. The digest should be carried out at 4°C overnight with minimal shaking.

The purification uses buffer B and buffer C again (**Figure 28**). An IMAC column was equilibrated with 5 CV buffer B, the digested protein sample was concentrated to ~ 2 ml and was injected via a loop onto an IMAC column. 1 ml fractions were collected from the start for 10 CV, and then larger fraction sizes (~ 2-4 mL) were collected on the gradient.

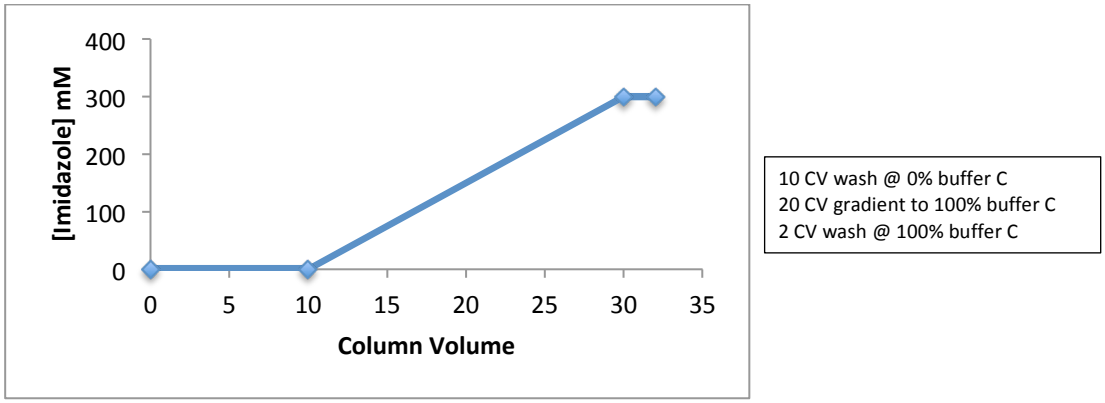


Figure 28 - Strategy for purifying proteins after cleaving the his-tag

The protein of interest no longer has a his-tag and should no longer bind to the IMAC column, therefore it should quickly elute shortly after the void-volume. Fortunately the 3C Protease is his-tagged, thereby binding to the column and eluting on the gradient. Another benefit of this is that, any impurities that still remain with natural poly-histidines will continue to bind to the column and elute on the gradient, thus the protein of interest that has eluted within the first few fractions may improve or retain its purity from the previous purification.

2.5. Protein Characterisation

There are many techniques and methods available for protein characterisation to probe its biochemical or biophysical properties. However, ADHs are enzymes with potential application for industrial biotechnology, it is important to investigate enzyme kinetics and substrate specificity to determine their uses, as there are many ADHs already commercially available.

2.5.1. Enzyme Kinetics

Enzymes such as ADHs can be characterised in solution by UV-Vis spectrometry. Methods for investigation by UV-Vis have already been well established, such as measuring the depletion or increase of absorbance for the nicotinamide cofactor at a wavelength of 340 nm and both cofactors absorption extinction coefficient (ε) is known; where ε = 6220 $cm^{-1}mol^{-1}$ for NAD(P)H at 340 nm. Therefore, the parameters can be applied to the Beer-Lambert law, A = εcl (where A is the absorbance, c is the concentration and l is the pathlength), to investigate the Michaelis-Menten kinetics (**Figure 29**) parameters K_m and V_{max} for the enzymes and substrate of interest.

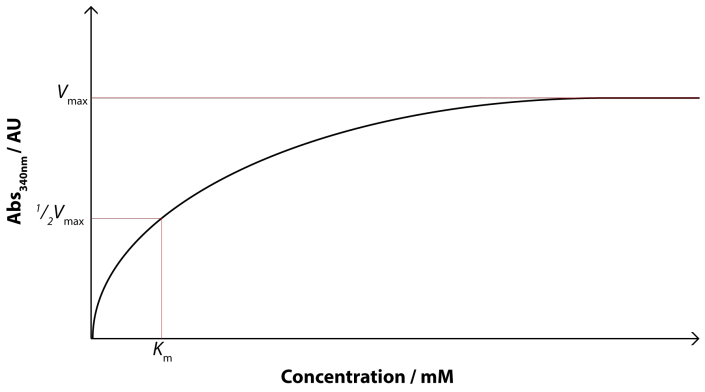


Figure 29 – A schematic of a Michaelis-Menten kinetics curve where K_m is found at ½V_{max}.

The V_{max} for an enzyme and a novel substrate of interest is unknown. Typically a wide range of substrate concentration is probed until saturation (V_{max}) or when substrate concentration is limited by solubility. The substrate concentrations investigated at final concentration were: 0 mM (control), 0.1 mM, 0.2 mM, 1 mM, 2 mM, 4 mM, 8 mM and 10 mM. Many substrates are only partially soluble in water,

therefore at higher concentrations dissolution in water-miscible solvent is required. Ethanol is known to be a substrate for many ADHs, therefore it was not used for dissolving substrates due to potential false readings. As such, DMSO was the solvent of choice.

The total reaction volume was 1 mL and the kinetics buffer used was a solution of 50 mM Tris at pH 7.5 (hereon the kinetics buffer). 50 µL of enzyme at 1 mg/mL and 1 mM cofactor final concentration was used in each reaction. Even if the substrate was dissolved in neat DMSO, it may not be enough to keep hydrophobic or heavily aromatic substrates in solution for the reaction. Despite potential loss of activity with the addition of solvents; DMSO was added from 5%-10% *v.v.*, to attempt to keep the substrate in solution.

An absorbance-time scan at 340 nm was performed to analyse the reaction. The reference cell comprised 1 mL of kinetics buffer. The reaction cell with the kinetics buffer was measured for 30 s, after which the substrate was pipetted and monitored again for 30 s. Following the substrate, the cofactor was pipetted and monitored for another 30 s, finally the enzyme was pipetted into the cell where the reaction was then monitored for 5 min. This protocol was performed on all substrate concentrations, where the reaction curve will be similar to the schematic (**Figure 30**).

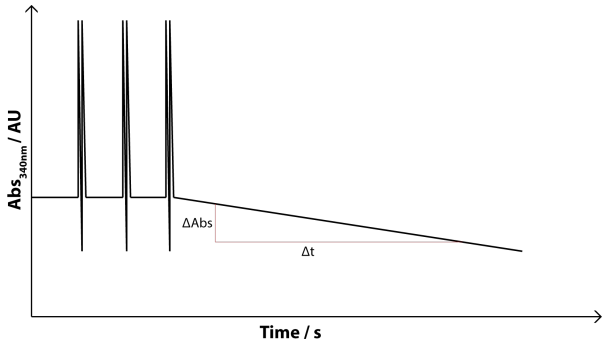


Figure 30 – Schematic of kinetics by UV-Vis analysis, where the depletion of the nicotinamide cofactor results in the desired change in absorbance and change in time.

Once each reaction has been performed, the depletion of cofactor will yield the change in absorbance over a period of time. Through rearranging the Beer-Lambert law, the change in concentration and therefore the velocity can be found. The velocity is then plotted against substrate concentration as in (**Figure 29**) to find the V_{max} and K_m for a particular substrate.

2.5.2. Biotransformations

Enzyme substrate specificity can be analysed by gas chromatography (GC) where biotransformations are performed on a whole range of substrates. The reaction is then quenched and extracted with a volatile solvent, such as ethylacetate to be analysed by GC. The interaction between the GC column and the substrates is based on polarity. This is a good analytic technique for investigating ADH substrate specificity as alcohols and carbonyl compounds are easily distinguished by GC. Alcohol's are more polar than aldehydes and ketones due to greater intermolecular hydrogen bonding. Therefore, alcohols interact more with the GC column and elute with a later retention time than aldehydes and ketones (**Figure 31**).

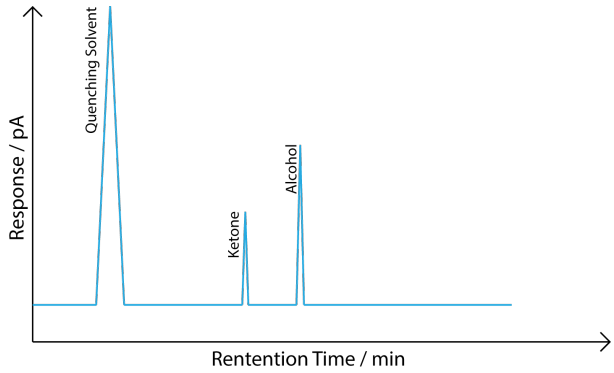


Figure 31 – Schematic of GC analysis for alcohols and ketones. Typically the solvent peak (ethylacetate) elutes early on, with carbonyl compounds such as acetone eluting before its alcohol counterpart, isopropanol, due to alcohols exhibiting a higher degree of polarity from greater intermolecular hydrogen-bonding.

The total reaction volume for a biotransformation was 5 mL. The buffer chosen for biotransformations was the kinetics buffer. The reaction comprised of 1 mL of enzyme at 5 mg/mL, substrate and cofactor at 1 mg/mL concentration and the relevant cofactor recycling system. Biotransformations were only performed with the SDRs, therefore a NADP(H) system. As such, glucose-6-phosphate dehydrogenase (0.14 mg/mL) and glucose-6-phosphate (0.52 mg/mL) were used as the cofactor recycling system. Controls were setup up, where the reaction conditions were the same, however no enzyme of interest was present. Additional reactions containing 5% *v.v.* DMSO were setup up to aid the low solubility of certain substrates.

The biotransformation was quenched at 0 h, 0.5 h, 1 h, 2 h, 4 h, 6 h and 24 h. Where 500 µL of the biotransformation was extracted into an Eppendorf tube with 600 µL ethylacetate. The mixture was vortexed to homogenise the solution and

centrifuged at 13.3K rpm for 2 min to separate the two layers. Finally, the top organic layer was extracted and pipetted into a GC vial, which was stored at 4°C.

Once all reactions had been quenched and collected, the substrate at 1 mg/mL final concentration in ethylacetate was analysed on GC as the standard. Appropriate temperature and time parameters were established. These parameters were used for isothermal analysis. All reaction samples were analysed for the same amount of time and at the same temperature.

2.6. X-ray Crystallography

Structure determination and characterisation of small molecules and protein by X-ray crystallography is technique that is now routinely used. The experimental data obtained from the diffraction pattern, the electron density, can reveal information about the crystallised protein or small molecule to the atomic level, such as disordered regions of a molecule, bond lengths and at atomic resolutions (> 1.00 Å) deformation of the electron density can be observed, serving as visual evidence for polarised bonds. The electron density for atoms are visualised as spheres (**Figure 32** - left) for oxy-myoglobin at 1.00 Å resolution (PDB entry: 1A6M). This is known as the spherical-atom approximation and is based on the Kappa-Formalism to simplify calculations for electron density. The spherical-atom approximation is a very good approximation, however this breaks down at > 0.7 Å when density can be observed for hydrogen atoms due to density deformation (**Figure 32** - right) for crambin at 0.48 Å resolution (PDB entry: 3NIR). Hydrogen atoms are usually not observed in the electron density as it diffracts X-rays very weakly with one electron.

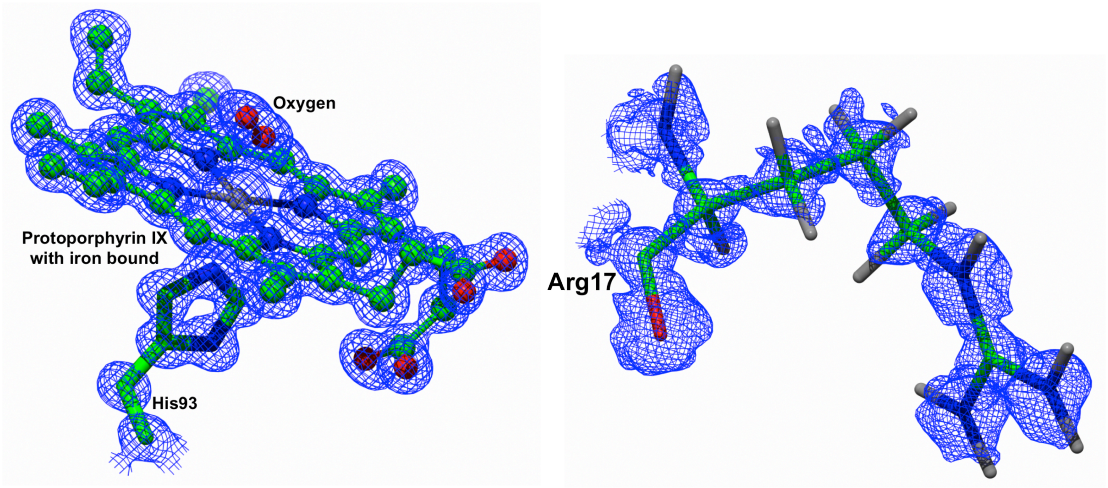


Figure 32 – (Left) spherical-atom approximation at 1.00 Å resolution of the active site in oxy-myoglobin (PDB entry: 1A6M). (Right) deformation of electron density at 0.48 Å resolution of the Arg17 residue in crambin (PDB entry: 3NIR).

Single molecules diffract very weakly. Proteins in solution are not ordered and are constantly moving; therefore X-ray diffraction analysis for structure determination. Crystals are the most ordered form of matter. This is as the proteins or small molecules in crystal are packed in a repeating and well-ordered array. The protein or small molecule crystal behaves as an amplifier for the protein or small

molecule suitable for single-crystal X-ray diffraction; hence, crystals are used for structural determination. The diffraction obtained from the single crystal is an average image of the array of packed molecules; therefore if the crystal packing is well ordered with very low mosaicity (**Figure 33** – left), atomic resolution can be achieved and well-defined diffraction spots will be observed (**Figure 34** – left)

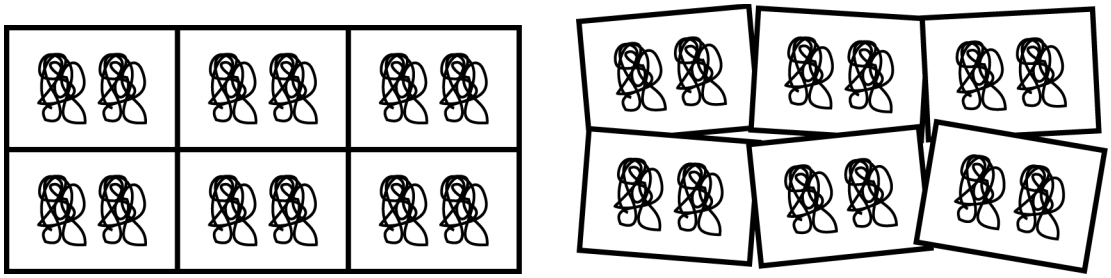


Figure 33 – Schematic of 2-demensional crystal packing with zero mosaicity (left) and a higher degree of mosaicity (right).

High mosaicity (**Figure 33** – right) causes issues when integrating the diffraction data, as the imperfections in crystal packing causes the diffraction pattern to become smeared. It becomes very difficult to define the intensity of a smeared spot (**Figure 34** – middle) or in extreme cases (**Figure 34** – right). Therefore, for optimal data quality well-ordered crystals are required for single crystal X-ray diffraction analysis.

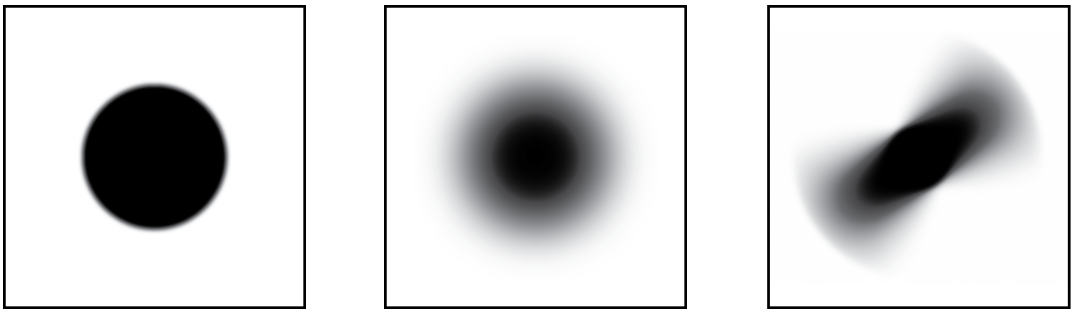


Figure 34 – Schematic of diffraction spots. Well-defined spot from a crystal with negligible mosaicity – left. Poorly defined spot from a crystal with low mosaicity – middle. Very poorly defined smeared spot from a crystal with high mosaicity – right.

Through the development of the Protein Data Bank (PDB) came the development of the .pdb file format. A .pdb file contains information on the protein, such as its basic function, experimental data, authors, refinement parameters and the coordinate of the model. The PDB is not limited to structures by X-ray crystallography, but also Nuclear Magnetic Resonance (NMR), electron microscopy and hybrid solutions. The universal file format allows for the ease of viewing with any molecular graphics suite of choice.

Structures obtained from X-ray crystallography are not only important in unveiling the active site of a protein for revealing the mechanism of a biocatalyst. Structures also serve as a basis for further computational work such as ligand-docking studies, fragment based screening for designing new drugs against diseases in the pharmaceutical industry.

2.6.1. Protein Crystallisation

Once satisfactory protein purity had been achieved and confirmed by SDS-PAGE analysis (2.2.3.), the protein is screened for a condition that it will crystallise in. Currently there is no precise method for crystallising a novel protein, therefore screening is required. INDEX from Hampton Research, PACT (55) and CSS 1 + 2 (56) from Molecular Dimensions crystallisation screens were used with the mosquito® Crystal (from TTP Labtech) to screen for protein crystals in a 96-well sitting drop format. The 300 nL drops were setup in a 1:1 ratio of protein and mother liquor.

Once successful hits were observed from the screens, the conditions were scaled up and optimised by varying precipitant concentrations in 24 well Linbro dishes by hanging drop (vapour diffusion) method. The final volume of the mother liquor in the well was 1 mL and the 2 µL drop suspended from siliconised glass coverslips was set up in a ratio of 1:1 ratio of protein and mother liquor.

2.6.2. Initial Testing and Data Collection

Once protein crystals had been successful grown, they were tested in-house using a Rigaku Micromax-007HF fitted with Osmic multilayer optics and a Marresearch MAR 345 image plate detector. Initial testing was used to test whether complex diffraction was observed and if a cryoprotectant was required due to ice-rings.

Cryoprotectants were composed of the same conditions as the mother liquor with additional glycerol or ethylene glycol ranging from 10-30% *v*/*v*. If a complex

with the appropriate nicotinamide cofactor was required, cofactor was added to a final concentration of 20 mM for a soak or cryoprotectant solution. Co-crystallisation with cofactor had lead to no crystallogenesis or crystals with very poor diffraction. The crystals were soaked for up to 5 min prior to testing or flash cooling in liquid nitrogen.

Only crystals that diffracted greater than 3.5 Å were sent to Diamond Light Source synchrotron in Oxford, Didcot. At the Diamond, the crystals were screened using a grid scan or line scan (**Figure 35**) for the best diffraction quality and the data collection strategy was estimated and calculated by the automatic pipeline software EDNA (57) with less than 20 images.

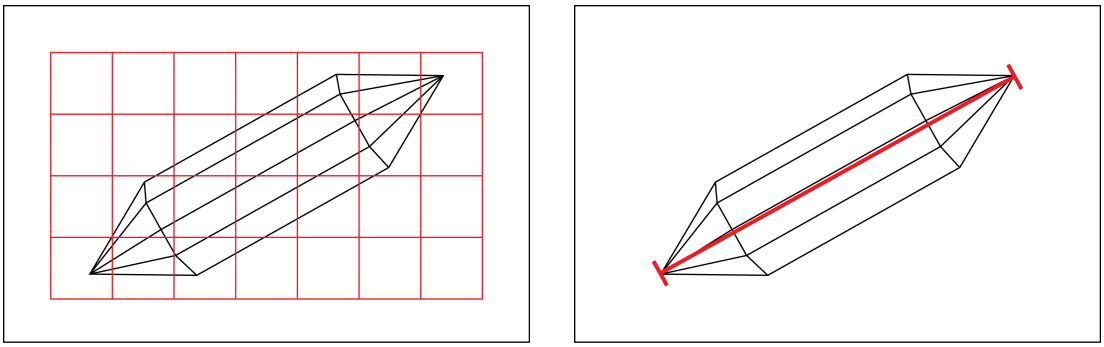


Figure 35 – Schematic of grid scan (left) and line scan (right) screening on a protein crystal. The defined area is tested for best diffraction quality.

Once the optimal position for diffraction quality was established, a full data set can be collected with the strategy suggested by EDNA. The automatic software pipeline allows for the data to be automatically processed by XIA2 (58) which uses the program XDS (59) within the pipeline for data processing. For all four enzymes, the automatically processed data was used for structure solution and data refinement.

2.6.3. Structure Solution

Atomic resolution is required for small molecule X-ray crystallography and direct methods, such as the Patterson method (60), can be used to solve the crystal structure. However, this is mostly not applicable to macromolecular or protein crystallography as the Patterson method is limited to very small proteins and very

high resolution. In the case of protein crystallography solving the phases can be an issue. In an X-ray diffraction experiment the most vital information, the phase, is not collected. Therefore calculating the structure factor is not possible without the phases; this is known as the phase problem. With small molecules this is not so much of an issue, as brute force calculations can be used to solve the phases. However, there are methods for solving the phase problem for macromolecules and proteins.

Typically, isomorphous replacement, anomalous scattering or molecular-replacement are used to solve the phases. Anomalous scattering is a more commonly used technique over isomorphous replacement as it is more convenient. However for a superfamily of enzymes that has been extensively investigated, such as the ADHs, molecular-replacement, using the program MOLREP (61), can be an option due to high sequence-structure conservation. Therefore, models from chainsaw modelling generated from available PDB structures with high sequence similarity (percentage homology), models generated from homology modelling or models generated from secondary structure prediction servers could be used to find some correct phases to solve the structure. The program CHAINSAW (62) generates a chainsaw model that is suitable for molecular replacement, by aligning a target and model sequence and modifies the models .pdb file by cropping non-conserved residues. Alternatively the program BALBES (63), molecular-replacement pipeline, could be used. BALBES will attempt to generate a structure solution from structures deposited in the PDB database, automatically generating the relevant chainsaw models and splicing appropriate domains without the users intervention.

2.6.4. Structural Refinement

Once the structure solution is found, structure building with COOT (64) can begin by initially amending the model with the relevant sequence, checking for side chains and backbone are in the correct position relative to the electron density and are chemically correct in terms of potential neighbouring interactions. Once the initial model is built, the reliability of the model, the reliability factor (R-factor) is calculated by comparing the structure factors from the experimental X-ray data and the

structure factors calculated from the model. The refinement program REFMAC 5 (65) was used; REFMAC 5 attempts to minimise the structure factor difference from the experimental X-ray data and the model by performing rigid body, restrained or unrestrained refinement against the experimental data and the model. Iterative cycles of model building and refinement are performed (**Figure 36**) to minimise the R-factor until satisfactory data is obtained.

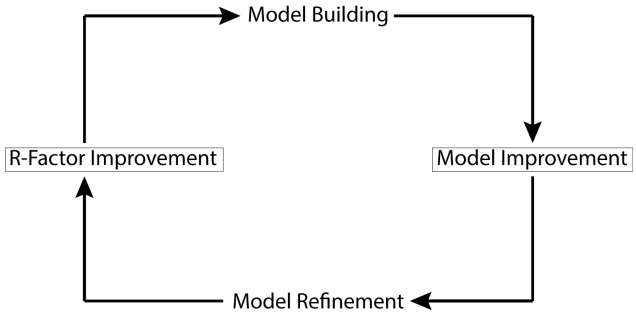


Figure 36 – Schematic of the refinement with iterative cycles of model building and model refinement.

The ideal R-factor is zero; this assumes it is a perfect model that matches the experimental data perfectly. Only data small molecule crystallography will be close to zero. The initial model for small molecules will have an R-factor of 0.05 (5%) – 0.10 (10%), once hydrogens atoms have been added with appropriate B-factors and final refinement performed, the R-factors will typically be between 0.02 (2%) – 0.04 (4%) for a very good dataset. This is only possible with atomic resolution crystal structures and small molecule crystals that diffract to 1.5 Å are normally rejected; this is the reason for requiring atomic resolution structures for small molecule crystallography.

The R-factors for proteins will vary. For a well diffracting protein crystal, 1.5 Å, a fully refined finished structure could have R-factors of 10%. However due to disordered regions such as loops and density that cannot be rationally assigned or built, the R-factors for a 1.5 Å structure could range from 12% to as low as 20% for incomplete structures due to disorder. Despite this, the structures are still acceptable for deposition and further analysis.

2.7. Computational Modelling

Co-crystallisation with ligands can often prove unsuccessful, as can soaking with ligands, therefore obtaining a crystal structure with the desired ligand for mechanistic or inhibition studies can be very difficult. Computational modelling use to require supercomputers for data processing, calculations and simulations, however technology has greatly advanced in that the processing power of a standard desktop computer can perform these simulations and calculations. Therefore these programs have now become accessible to everyone. There are many different search algorithms available for ligand docking, however the Lamarckian Genetic algorithm is commonly used for most docking programs as it provides the most efficient search method based on conformation (66).

2.7.1 Ligand Building

The CCP4 ligand library has a premade dictionary for common ligands and cofactors such as the nicotinamide cofactor. However for novel ligands and ligands that are not available in the CCP4 ligand library (http://www.ccp4.ac.uk/html/refmac5/dictionary/list-of-ligands.html), they will have to be built and energy minimised. Ligands are initially built using the COOT Ligand Builder; the ligand is then saved as a .mol file and is imported into the program PRODRG (67) in the CCP4 interface (68), where the .pdb coordinates and .cif topology files are generated (**Figure 37**).

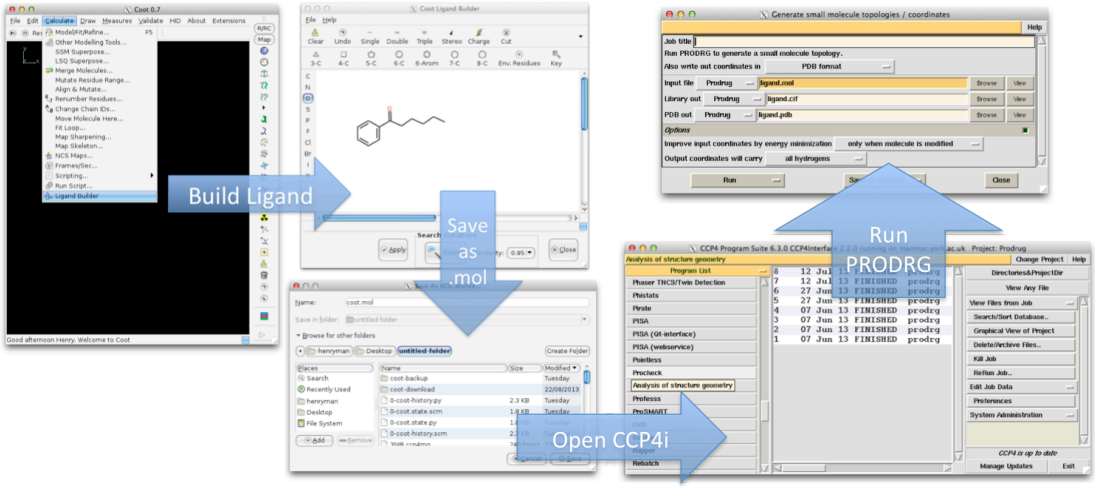


Figure 37 – Schematic for ligand building for ligands, such as PPK, not available in the CCP4 ligand library.

PRODRG carries out the necessary restraints and energy minimisation on the ligand, which allow for successful refinement in REFMAC5 or ligand docking.

2.7.2. Ligand Docking

AutoDock suite (69) was used to dock ligands into the active site of the protein. AutoDock requires PDBQT files, which is an extension of the PDB file format. As well as the normal PDB file headers, PDBQT includes information for polar hydrogen atoms, partial charges, atom types and information on the articulation of flexible molecules. One of the topology files that PRODRG can generate for ligands (as mentioned in section 2.5.1.) are PDBQT files suitable for AutoDock. However they can also be generated within AutoDockTools.

The first step was to prepare the protein's coordinate file in AutoDockTools, by removing all water molecules and adding hydrogens to the protein. The ligand was prepared (**Figure 38**) by detecting the torsional root, choosing the torsions, then setting the number of torsions; the ligand was then saved as a PDBQT file. The protein was chosen to be a rigid receptor which merges non-polar hydrogen atoms and charges, assign aromatic carbons and allow the protein to be saved as a PDBQT.

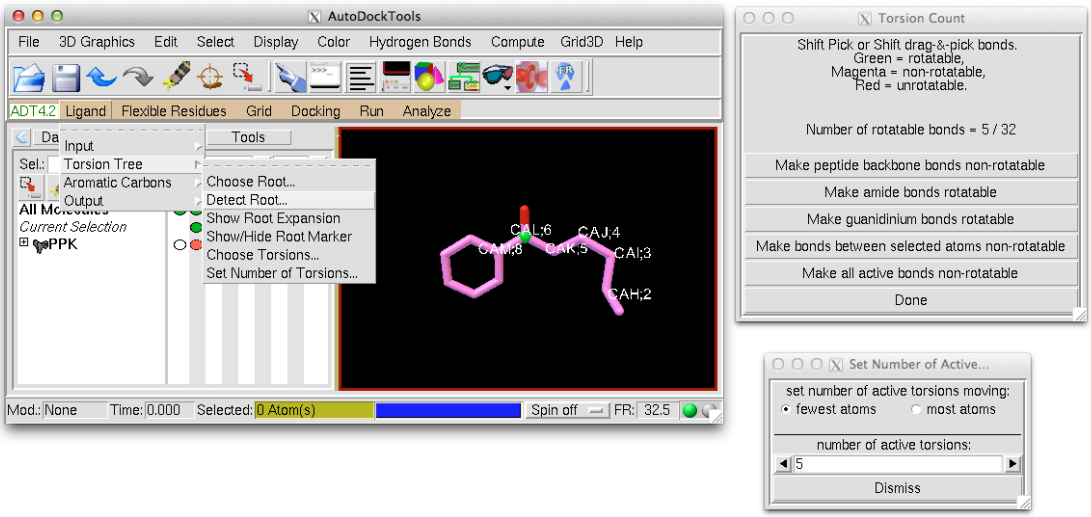


Figure 38 – Preparing the ligand PPK in Autodock Tools, with auto detection of the torsional root.

Prior to docking, AutoDock requires pre-calculated grid maps, one for each atom type present in the ligand being docked. These maps are calculated by AutoGrid (**Figure 39**), which is part of the AutoDock suite. A grid map consists of a three-dimensional lattice of regularly spaced points, which can be either entirely or partly centred on a region of interest of the macromolecule/protein.

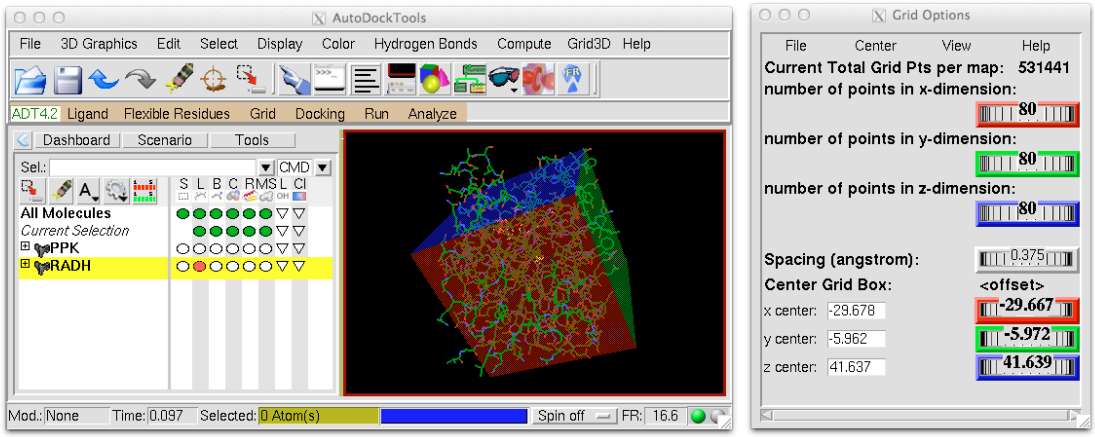


Figure 39 – Defining the area in the protein to be dock with ligand using AutoGrid. Either the whole protein or just the active site can be defined for docking.

After a grid box has been specified, the ligand can be docked with AutoDock. AutoDock can use the Lamarckian Genetic algorithm to explore the conformational state is a flexible ligand, using the maps generated by AutoGrid to evaluate the ligand-protein interaction at each point in the docking simulation (**Figure 40**).

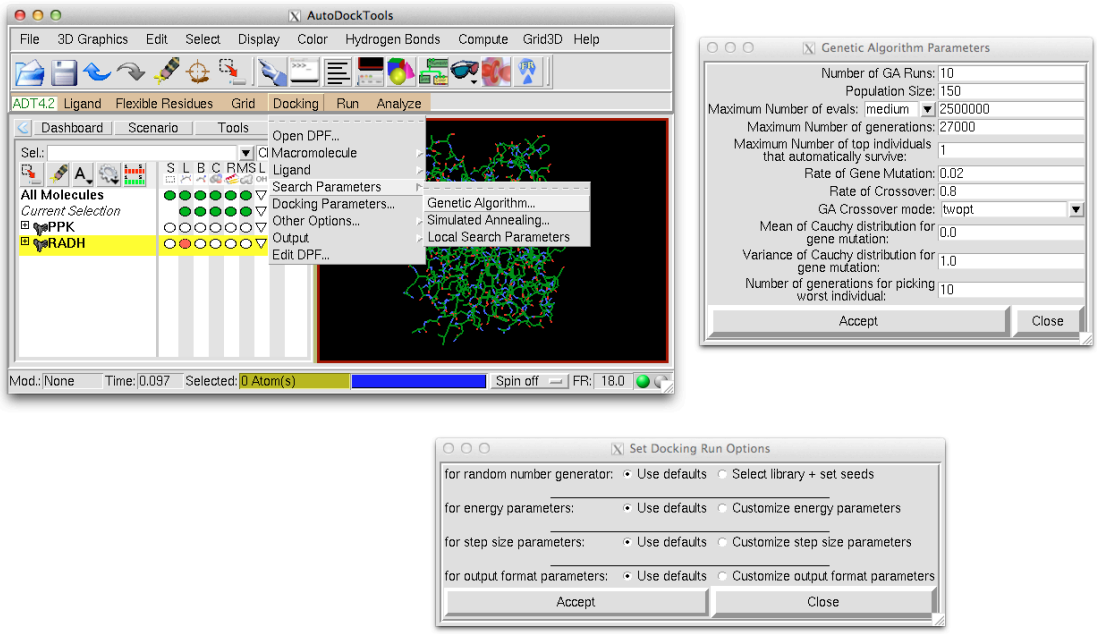


Figure 40 – Defining the docking algorithm to be used with the relevant docking parameters.

Instead of using AutoDock, AutoDockVina can be used, which utilises a multicore processor computer and docks significantly faster with greater accuracy than AutoDock. Another benefit is that the files generated by AutoDockVina can assess neighbouring residues and interactions; AutoDockVina calculates these interactions and therefore they provide a less biased assessment of the neighbouring residues rather than picking residues by eye (**Figure 41**).

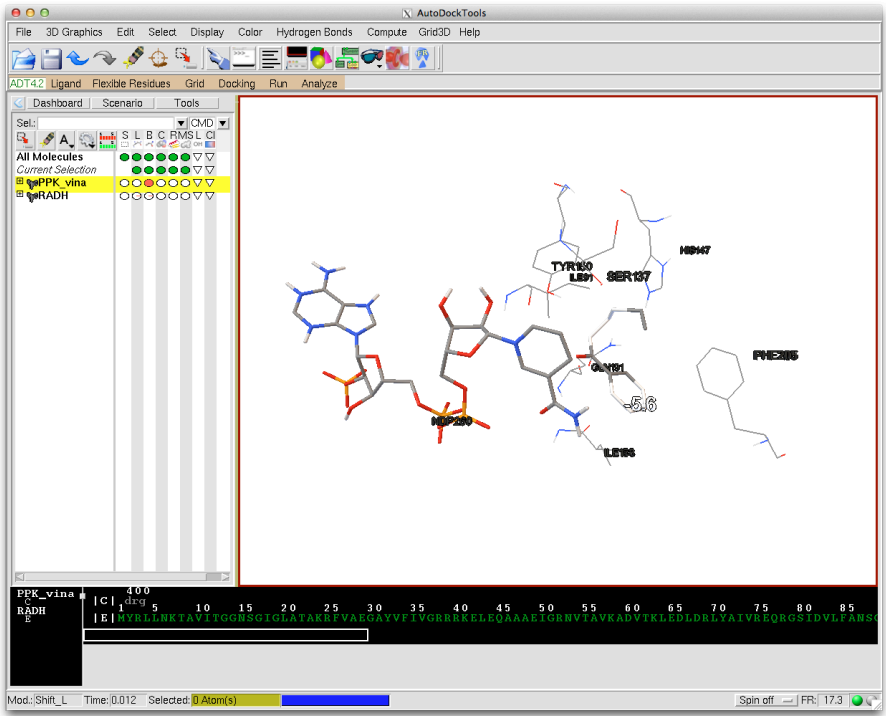


Figure 41 - The docking outcome with AutoDock vina showing potential interacting residues and cofactor, with the ligand PPK being ranked in vina energy.

3. Results

3.1. RasADH

RasADH is an SDR from *Ralstonia* sp. Justyna Kulig from Institute of Bio- and Geosciences in Jülich cloned the gene encoding RasADH into the pET-YSBLIC3C vector, expressed and purified the enzyme but had little success with crystallisation. RasADH was found to be able to accept 'bulky-bulky' ketones (24) and reduces them selectively to the (*S*)-alcohol with *e.e.*s ≥99% (26) utilising NADPH as the cofactor. Many drug compounds have bioactive benzyl alcohol derivatives (**Figure 42**); hence RasADH could be used to accept bulky ketones as the precursors (23). Despite RasADH being a SDR, it was seen to have an affinity for calcium for activity and protein stabilisation (27) and thus, not dependent on divalent metal ions for activity.

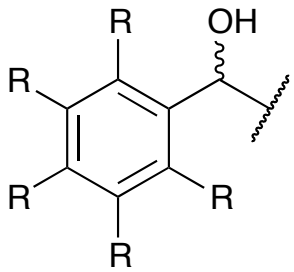


Figure 42 - Generic benzyl alcohol derivative

3.1.1. Expression

An induction test (**Figure 43**) was performed in order to optimise expression (Section 2.3.4.2). The gel suggests later induction at OD_{600} ≥ 1.4 yields more soluble protein. Upon scale-up for purification, the culture was induced at an OD_{600} ≥ 1.5.

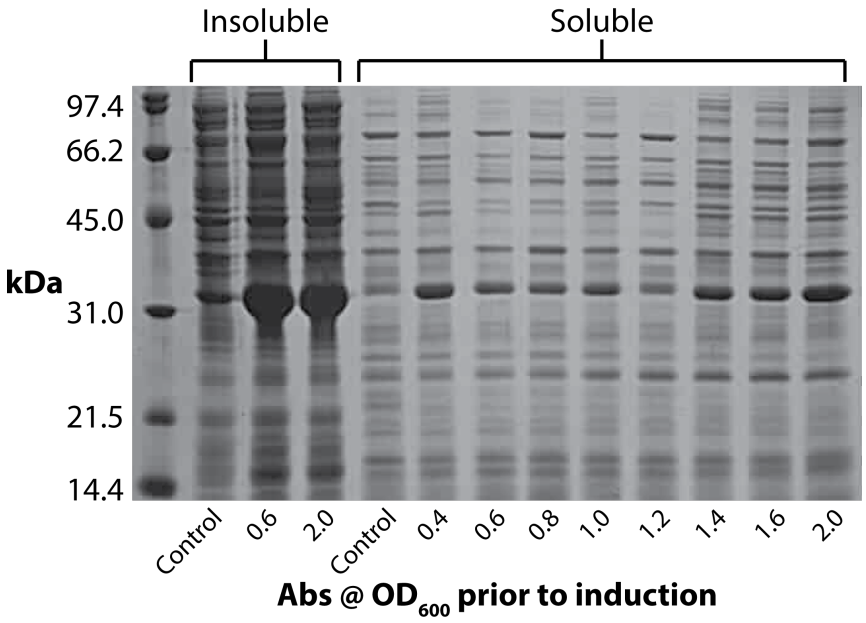


Figure 43 - Induction Test for RasADH to optimise protein production analysed on a 12% SDS-PAGE gel as described in section 2.2.3. Lane 1 is the Low Molecule Weight Ladder from BioRad. Lanes 2-4 are the insoluble fractions for the control, induction at 0.6 Abs and 2.0 Abs respectively. Lanes 5-13 are the soluble fractions for the absorbance's chosen for testing.

3.1.2. Purification

Prior to purification, the clarified cell lysate (from 6 L of culture) was recirculated onto two 5 mL HisTrap Crude FF columns, charged with nickel and joined end-on-end using a peristaltic pump (2.4.2.). RasADH benefited from this purification strategy significantly, as previous attempts had proven difficult and even after SEC, the purity was not satisfactory. Each column was purified separately. Following nickel purification, purity was analysed by SDS-PAGE (**Figure 44**).

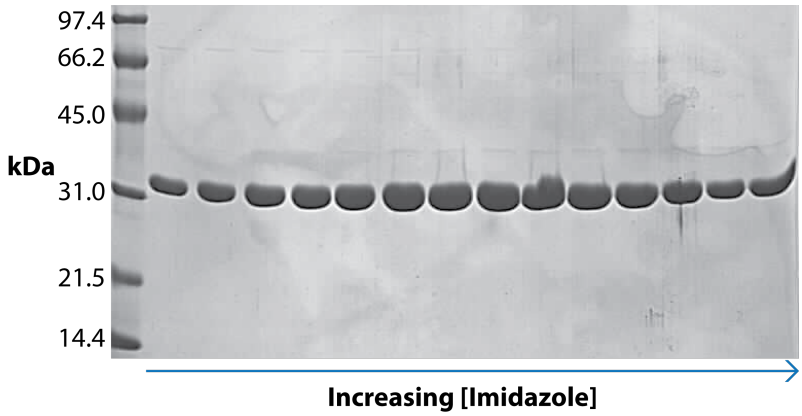


Figure 44 - SDS-PAGE analysis for the nickel purifiaction of RasADH. The SDS-PAGE gel and samples were prepared as described in sections 2.2.3. and 2.2.3.1. Lane 1 is the Low Molecular Weight Ladder from BioRad and lanes 2-25 are the samples obtained from the elution gradient of the purification.

RasADH could only be concentrated to 5 mg/mL before the protein began to aggregate. Previous attempts to dialyse the protein into buffer B were unsuccessful; dialysis overnight caused the protein to precipitate. The appropriate fractions were pooled and concentrated for SEC (**Figure 45**) and purified in batches.

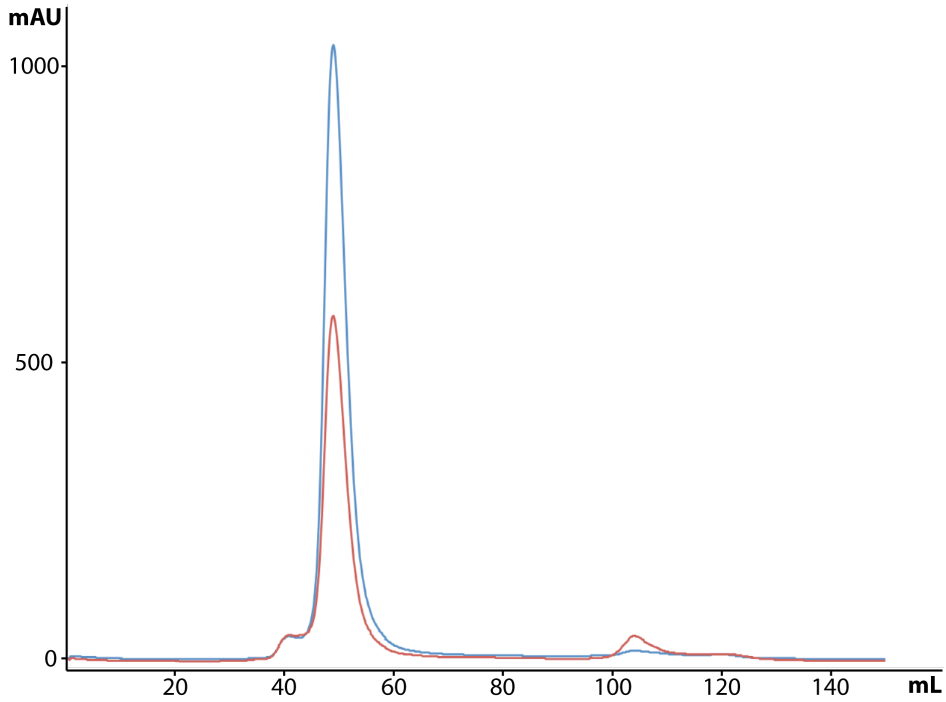


Figure 45 - Chromatagram for the SEC of RasADH after nickel affinity purification. Samples from 40-60 ml of the SEC were taken for analysis on SDS-PAGE.

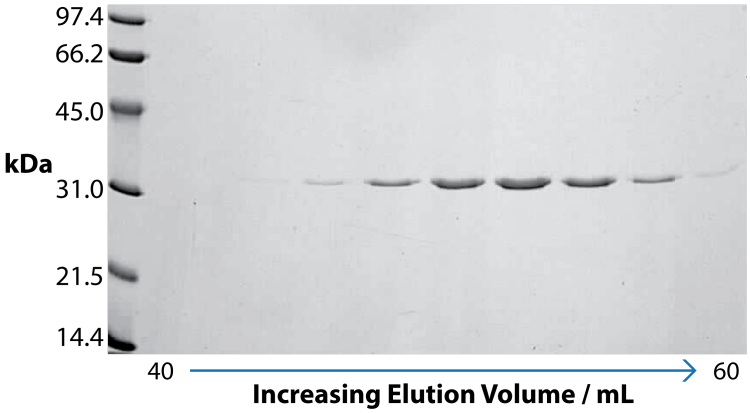


Figure 46 – Protein purity analysis by SDS-PAGE after SEC. Samples and SDS-PAGE gel were prepared as described in sections 2.2.3.1. and 2.3.3. Lane 1 of the SDS-PAGE gel is a Low Molecular Weight Ladder from BioRad, lanes 2-10 were samples every second fraction of the peak between 40-60 ml.

The protein analysed by SDS-PAGE (**Figure 46**) was sufficiently pure after SEC to proceed with crystallisation screening.

3.1.3. Crystallisation

After RasADH was purified, it could be concentrated above 5 mg/mL without any signs of precipitation. Initial crystallisation screen of RasADH was performed at 5 mg/mL in INDEX, PACT and CSS 1 + 2. The most promising hits were observed in PACT, however these were only needle-like crystals. RasADH was rescreened in PACT at 21 mg/mL, with the most promising crystals appearing in PACT-G4 with rod-like morphology. These rod-like crystals were successfully optimised using hanging drop vapour diffusion method (Section 2.6.1.) with additives producing much larger rods with less striation than the crystals from the screen. The best crystals (**Figure 47** – left) were obtained in 0.1 M Bis-tris propane pH 8.0, containing 16% (*w/v*) PEG 3350, 0.2 M potassium isothiocyanate and 5% (*w/v*) ethylene glycol at a protein concentration of 24 mg/mL. The crystals were soaked for 5 min with a cryoprotectant comprised of the mother liquor with 20% (*v.v.*) ethylene glycol and 10 mM NADPH, attempting to obtain a NADPH complex. The crystals were subsequently flash cooled in liquid nitrogen prior to in-house (Section 2.6.2) diffraction analysis (**Figure 47** – Right). Once satisfactory diffraction had been observed in-house, the NADPH complex (*holo*-) dataset was collected at Dimond, Didcot, on the I04 beamline on 2nd February 2013.

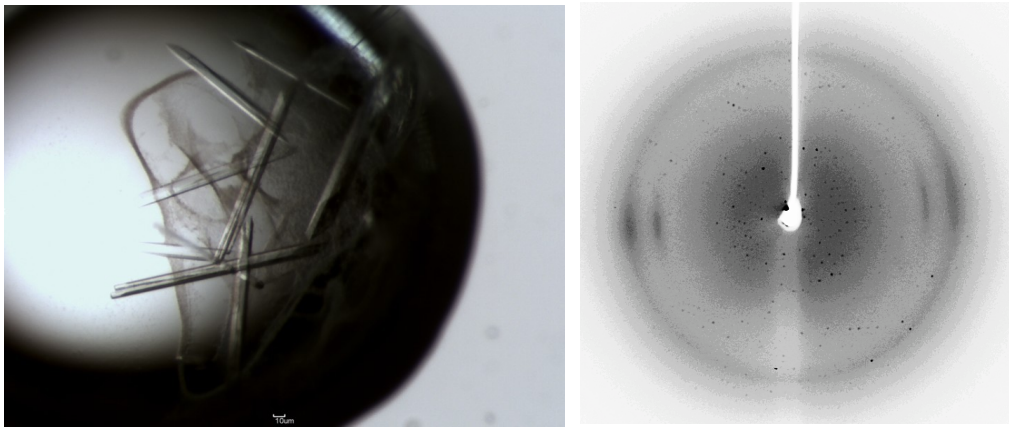


Figure 47 - RasADH crystals obtained from PACT-G4 additive optimisation (left) with its diffraction pattern in-house (right).

More crystals had formed at a much later date in the same PACT rescreen at 21 mg/mL. The crystal hit was in PACT-H10 and it did not have the same rod-like morphology has mentioned previously, however the conditions were similar to those in PACT-G4. The crystals had a shard-like morphology (**Figure 48** – left). The crystals

in PACT-H10 were successfully optimised by the hanging-drop method (Section 2.6.1.). The best crystals were obtained in 0.1 M Bis-tris propane pH 7.0 containing 20% (*w/v*) PEG 3350 and 0.02 M sodium-potassium phosphate at a protein concentration of 24 mg/mL. The crystals were soaked in a solution of mother liquor with additional 20% (*v.*v) ethylene glycol as a cryoprotectant. The crystals were flash cooled in liquid nitrogen prior to in-house diffraction analysis (**Figure 48** – right). In-house observed diffraction was already at a much higher resolution than the previous collected NADPH dataset. The *apo*-crystals were sent to Diamond, Didcot, and a full dataset was collected on the I02 beamline on 28th February 2013.

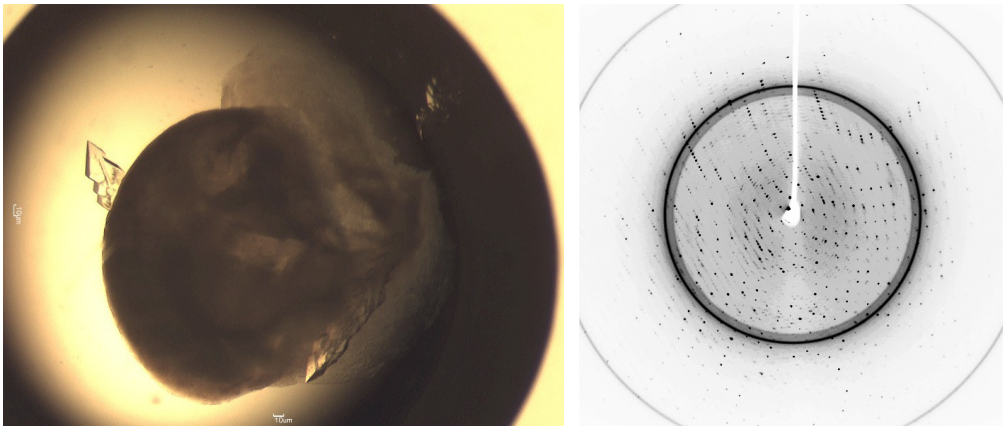


Figure 48 - RasADH crystals obtained from well PACT-H10 optimisation (left) with its diffraction pattern in-house (right).

Data collection and refinement statistics for the *apo*- and NADPH complex of RasADH (29) are shown below (**Table 1**).

	RasADH (*apo*)	**RasADH (NADPH complex)**
Beamline	Diamond I02 on 28/2/2013	Diamond I04 on 2/2/2013
Wavelength (Å)	0.97950	0.97950
Resolution (Å)	60.86-1.52 (1.57-1.52)	74.54-2.89 (2.93-2.89)
Space Group	$C2_1$	$C2_1$
Unit Cell (Å)	a = 136.5; b = 52.5; c = 151.5; α = γ = 90.0; β = 116.6	a = 192.3; b = 135.6; c = 93.6; α = γ = 90.0; β = 100.1
No. of molecules in the asymmetric unit	4	6
Unique reflections	149646 (45160)	52465 (2880)
Completeness (%)	97.7 (96.8)	99.2 (98.8)
R_{merge} (%)	0.03 (0.25)	0.18 (0.73)
$R_{p.i.m.}$	0.03 (0.23)	0.16 (0.64)
Multiplicity	3.3 (3.5)	4.2 (4.2)
$\langle I/\sigma(I) \rangle$	15.6 (4.9)	7.4 (2.0)
$CC_{1/2}$	1.00 (0.95)	0.98 (0.74)
Overall *B* factor from Wilson plot (Å^2)	24	41
R_{cryst}/R_{free} (%)	15.8/18.8	26.8/29.3
No. protein atoms	6942	11081
No. water molecules	846	37
r.m.s.d. 1-2 bonds (Å)	0.02	0.01
r.m.s.d. 1-3 bonds (°)	1.69	1.6
Avge main chain B (Å^2)	20	35
Avge side chain B (Å^2)	23	36
Avge water B (Å^2)	33	14
Avge ligand B (Å^2)	-	32
PDB entry	4BMN	4BMS

Table 1 - Data collection and Refinement Statistics for RasADH apo and NADPH complex with numbers in brackets referring to data in the highest resolution shell.

3.1.4. Structure Solution and Model Building

The *holo*-dataset was solved with MOLREP, using a monomer model derived from the tetrameric structure of *Rhizobium etli* CFN 42 (PDB entry: 4FGS). The solution contained six molecules in the asymmetric unit, representing one and a half tetramers (**Figure 49** – left). The *holo*-structure was built and refined to R_{cryst} and R_{free} values of 26.8% and 29.3% respectively. Subsequently, a monomer model derived from the *holo*-model was used to solve the *apo*-structure (**Figure 49** – right). The *apo*-structure was built and refined to R_{cryst}/R_{free} values of 15.8%/18.8%. Despite ambiguity present with the lower quality NADPH complex model (R_{cryst}/R_{free} = 26.8%/29.3%), much of the ambiguity can be clarified when comparing and superposing with the higher quality *apo*-model.

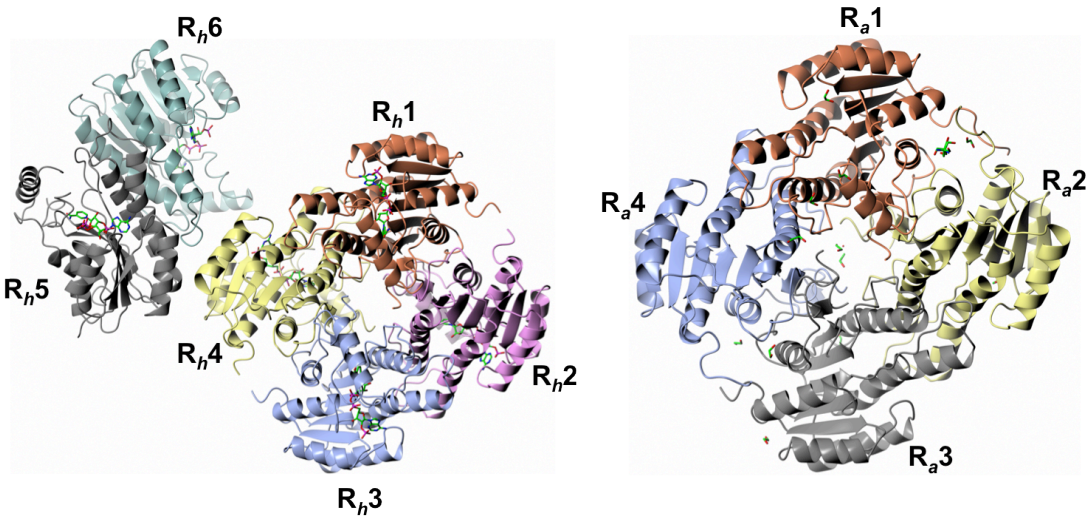


Figure 49 - RasADH *holo*-asymmetric unit (left) and *apo*-asymmetric unit (right). *Holo*-subunits labelled R_h, apo-subunits labelled R_a.

A monomer of RasADH is made up of a classical Rossman fold, as seen with subunit R_h1 (**Figure 50**), with a central β-sheet composed of seven β-strands: β1 (residues 8-13), β2 (33-37), β3 (54-58), β4 (83-87), β5 (131-135), β6 (174-180) and β7 (239-242). The β-sheet was encompassed by seven alpha helices: α1 (residues 17-30), α2 (40-50), α3 (64-77), α4 (101-126), α5 (148-168), α6 (195-208) and α7 (217-228).

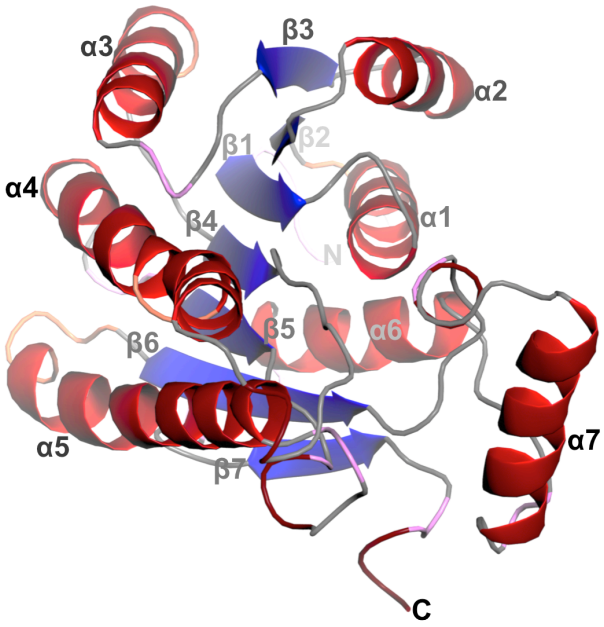


Figure 50 - RasADH Rossman fold represented by subunit R_h1. The helicies are represented by α1-7 and the β-sheets are represented by β1-7.

The majority of the secondary structure is conserved between the *holo*- and *apo*-structure. However, the most noticeable difference was seen is highlighted in the green box (**Figure 51**). In the *apo*-structure, there is no electron density present between the loop region (λ2). Furthermore, when this region is superposed with the *holo*-structure, it becomes apparent that it is a domain shift. The domain movement of the helix (α6) and loop (λ1) suggests that helix (α6) is a 'lid' for cofactor binding.

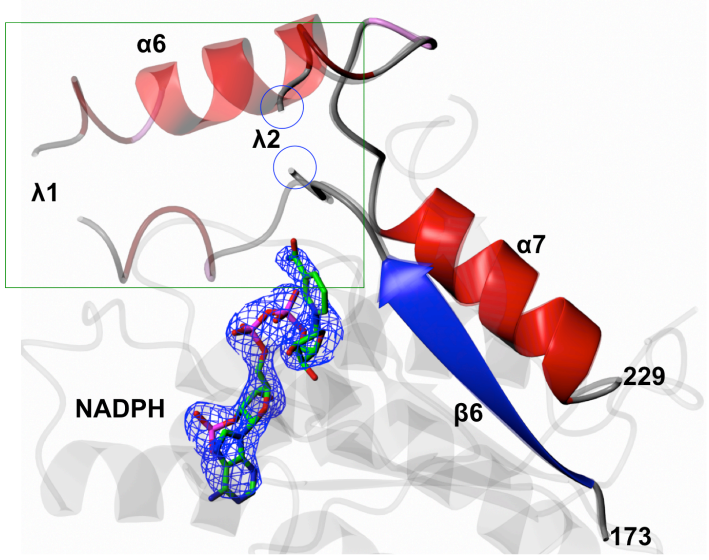


Figure 51 - RasADH overlay of *apo*- and *holo*-structure, residues 173-229 shown with the rest of the protein greyed out. Loops are denoted as λ. α1 is no longer present in the *apo*-structure as it becomes λ2. The electron density for NADPH is generated from the omit map (F_o-F_c) contoured to a level of 3σ which was obtained after refinement in the absence of NADPH.

The active site of RasADH (**Figure 52**) is formed at the top of the central β-sheet of the Rossman fold that is partially covered by the helix α6. The most noticeable feature is the active site is lined by residues Tyr150 (the likely proton donor), Ser137, Phe205, Leu144, Leu142, Leu201, Val138, Ile91, His147 and Gln191 forming a hydrophobic tunnel. The nicotinamide ring of NADPH sits at the base of this hydrophobic tunnel. Residues Arg38 and Arg39 govern RasADHs' strict specificity towards NADPH by close interactions with the 2'-phosphate on the adenine-ribose of NADPH, as well as Arg38 π-stacking interactions with adenine ring.

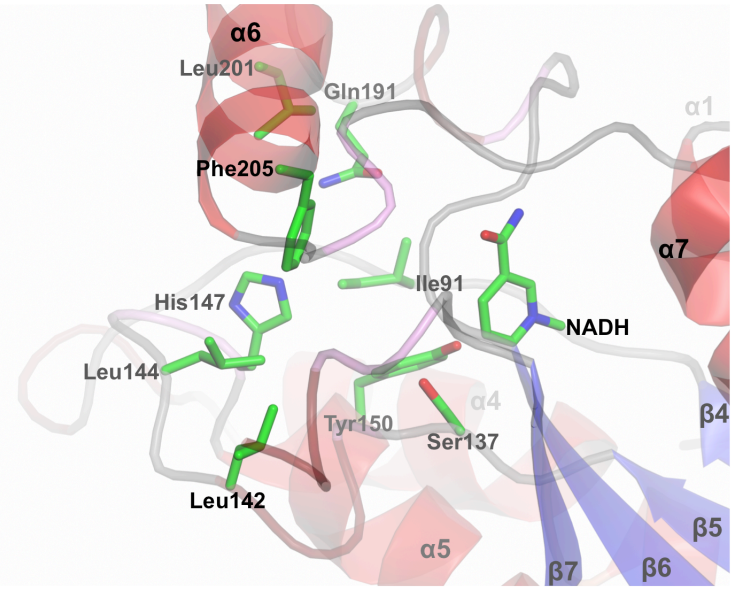


Figure 52 - Hydrophobic active site of RasADH represented by residues: by residues Tyr150, Ser137, Phe205, Leu144, Leu142, Leu201, Val138, Ile91, His147 and Gln191.

Despite the difference in cofactor binding, upon superposing the *apo*-structure and NADPH complex, the structures are highly conserved with an RMSD of only 0.41 Å over 458 atoms. PISA (70) analysis suggests that there is little difference between any of the dimer interfaces for potential interactions in RasADH. Further evidence was observed with the little difference in Gibbs free energy for any dimerising interface. This leads to suggestions that RasADH could predominantly exist as a tetramer over the more common dimer like many ADHs.

Other than the structure of RasADH solved by Lerchner (PDB entry: 4I5E) and the model used for the structure solution (PDB entry: 4FGS); using the *apo*-structure, DALI (71) server identified 3-oxoacyl-[acyl-carrier protein] reductase (FabG) from *Listeria monocytogenes* (PDB entry: 4JRO) which is also a tetramer in crystallographic space and has a sequence similarity of 36% and an RMSD of 1.1 Å over 247 residues.

Also the 3-oxoacyl-[acyl-carrier protein] reductase from *Synechococcus elongates* PCC 7942 (PDB entry: 4DMM) which is a dimer in crystallographic space has a sequence similarity of 35% and an RMSD of 1.3 Å over 237 residues. The two 3-oxoacyl-[acyl-carrier protein] reductase identified by DALI exhibit several similarities to SDRs, in that they are dimers or tetramers and utilise NADPH as the cofactor which most likely display general dehydrogenase or reductase activity.

Using the *holo*-structure for DALI analysis, DALI further identified a putative tetrameric oxidoreductase from *Sinorhizobium meliloti* 1021 (PDB entry: 4ESO) which has a sequence similarity of 40% and an RMSD of 1.4 Å over 251 residues. This was not identified by DALI with the *apo*-structure. This is most likely due to the missing lid domain in the absence of the cofactor. The tetrameric pyridoxal 4-dehydrogenase from *Mesorhizobium loti* (PDB entry: 3RWB) was also identified when using the *holo*-structure; the pryidoxal 4-dehydrogenase has a sequence similarity of 31% and an RMSD of 1.3 Å over 247 residues. Both of the targets identified by DALI also display similar properties to SDRs, exhibiting similar enzyme classification and Rossmann fold for the nicotinamide cofactor binding. Hence, they also display dehydrogenase activity.

Interestingly at higher resolutions, buffer components and crystallisation precipitants like Tris and ethyleneglycol can be seen, however they are not bound in the active sites. Despite RasADH having an affinity for calcium, it could not be rationally assigned and refined with COOT and REFMAC5, even at 1.4 Å. However it was to be expected that calcium would not be present in the active site as an SDR. Most likely the addition of calcium chloride would be a 'salting-in' effect, seeing as protein aggregation is observed at lower salt concentrations (sodium chloride). RasADH appeared most soluble at higher salt concentrations (500 mM sodium chloride). Therefore it is possible that calcium chloride behaves as a kosmotrope stabilising any intermolecular interactions.

3.2. SyADH

SyADH is an SDR from *Sphingobium yanoikuyae*. Kinga Kedziora from Universidad de Oviedo in Oviedo cloned the gene encoding SyADH into pET-YSBLIC3C, expressed and purified the protein and crystallised it. Like RasADH, SyADH is also able to accept 'bulky-bulky' ketones and selectively reduce them to the (*S*)-alcohol with *e.e.*s ≥99%; however SyADH differs in diastereoselectivity to RasADH. Thus, SyADH and RasADH are diastereocomplementary.

The data collection and refinement statistics for SyADH in complex with NADPH (29) are shown below (**Table 2**):

	SyADH
Beamline	Diamond I02
Wavelength (Å)	0.97949
Resolution (Å)	139.0-2.5 (2.56-2.50)
Space Group	$P2_1$
Unit Cell (Å)	a = 144.9; b = 86.8; c = 155.6; α = γ = 90.0; β = 106.4
No. of molecules in the asymmetric unit	10
Unique reflections	133601 (10900)
Completeness (%)	99.7 (99.8)
R_{merge} (%)	0.14 (0.54)
$R_{p.i.m.}$	0.12 (0.46)
Multiplicity	4.2 (4.3)
$\langle I/\sigma(I) \rangle$	8.2 (3.0)
$CC_{1/2}$	0.99 (0.82)
Overall *B* factor from Wilson plot (Å^2)	43
R_{cryst}/R_{free} (%)	23.4/25.1
No. protein atoms	18948
No. water molecules	290
r.m.s.d. 1-2 bonds (Å)	0.01
r.m.s.d. 1-3 bonds (°)	1.6
Avge main chain B (Å^2)	35
Avge side chain B (Å^2)	37
Avge water B (Å^2)	20
Avge ligand B (Å^2)	30
PDB entry	4BMV

Table 2 – The data collection and refinement statistics for SyADH in complex with NADPH.

3.2.1. Structure Solution and Model Building

The structure of SyADH was solved using the automated molecular replacement pipeline BALBES, which derived a monomer model from the NADPH-dependent blue fluorescent protein from *Vibrio vulnificus* (PDB entry: 3P19). The structure solution contained five dimers in the asymmetric unit. The structure was built and refined with iterative cycles of COOT and REFMAC5 with local NCS restraints. Once the model for SyADH was built and refined to a satisfactory level, the omit map, revealed residual electron density for NADPH to be rationally assigned and refined in all ten active sites. SyADH was built (**Figure 53**) and refined to R_{cryst}/R_{free} values of 26.8%/29.3%.

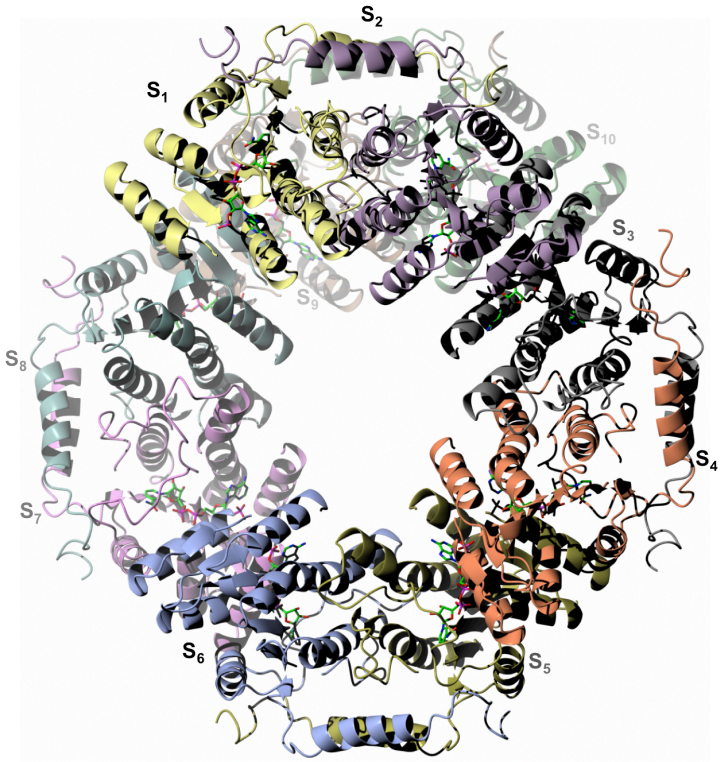


Figure 53 - The structure of SyADH with 10 subunits in the asymmetric unit, each subunit is denoted as S. SyADH asymmetric unit forms a quinary structure resembling a helix.

The monomer of SyADH, subunit S_1 (**Figure 54**), features a Rossman fold with a central β-sheet of seven strands: β1 (residues 6-11), β2 (31-34), β3 (56-58), β4 (82-88), β5 (134-138), β6 (177-181) and β7 (224-225). The central β-sheet is again surrounded by seven α-helicies: α1 (residues 15-27), α2 (38-52), α3 (66-78), α4 (102-130), α5 (151-171), α6 (190-194) and α7 (207-220). SyADH differs to a normal Rossman fold in that an additional helix α8 (234-248) extending to the C-terminus.

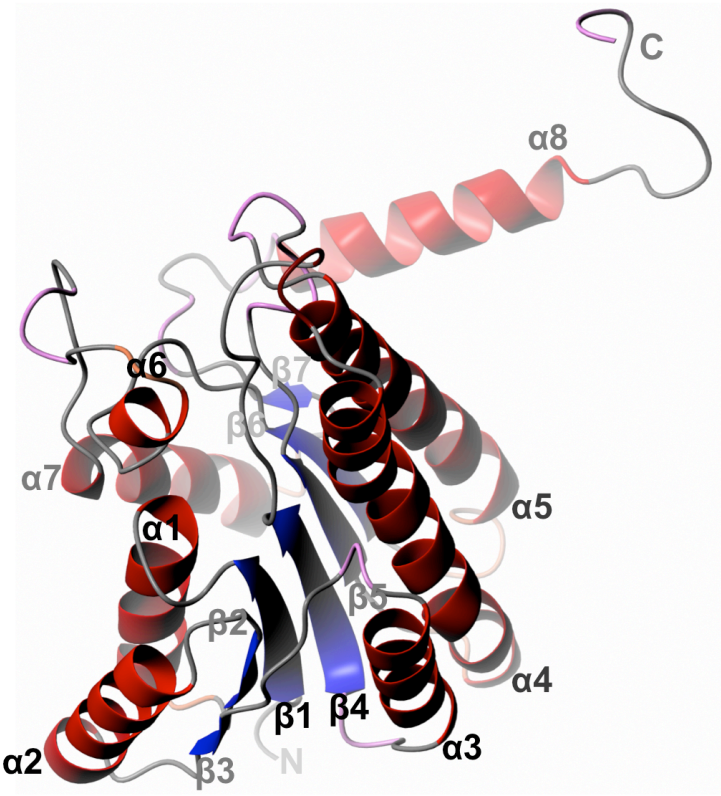


Figure 54 - A monomer of SyADH, represented by subunit S_1, displaying helicies α1-8 and β-sheets β1-7.

The C-terminal helix, α8 (**Figure 54**), participates in monomer-monomer interaction aiding the formation of a dimer (**Figure 55**). Not only are there interlocking and crossing-over C-termini helicies (δ1) and loops (δ2-δ4), but also these interactions extend as far as contributing to the binding pocket as seen with the pink residues (**Figure 56**).

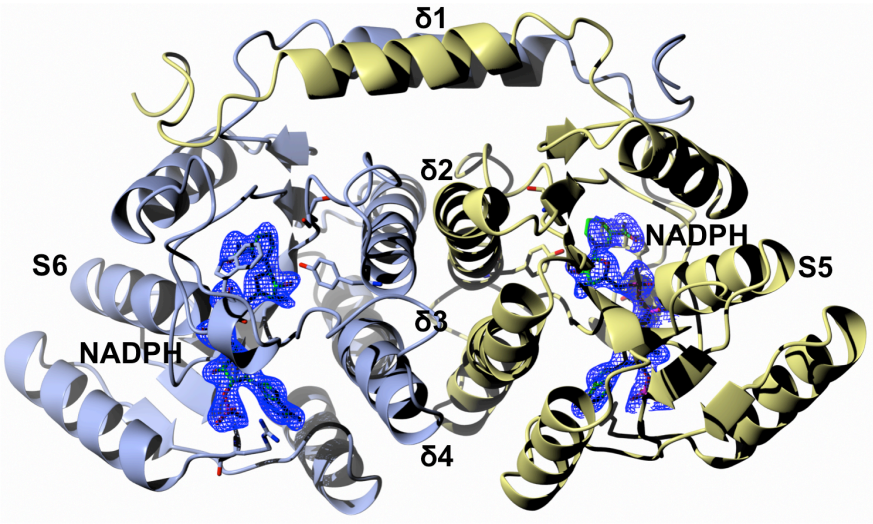


Figure 55 – Subunits 5 and 6 representing the major dimer interface of SyADH. Dimer interfaces denoted as δ; δ1 are the interlocking N-termini helicies providing the predominant interaction, δ2-δ4 illustrate weaker dimer interactions between helicies and loops. The electron density for NADPH is generated from the omit map (F_o-F_c) contoured to a level of 3σ which was obtained after refinement in the absence of NADPH.

Residues B/Gln251 and B/Ile249 (**Figure 56**) from the adjacent dimer aid the formation of a hydrophobic pocket with Ala92, Val142, Phe148 and Trp191 as well as the conserved potential catalytic residues Ser140, Try153 and Lys157. Arg36 in SyADH mirrors the location and function of Arg 38 in RasADH, however Arg39 in RasADH is replaced by Asp37 in SyADH that points away from the 2'-ribose phosphate in NADPH, as such, the cofactor specificity for SyADH does not appear as strict as RasADH.

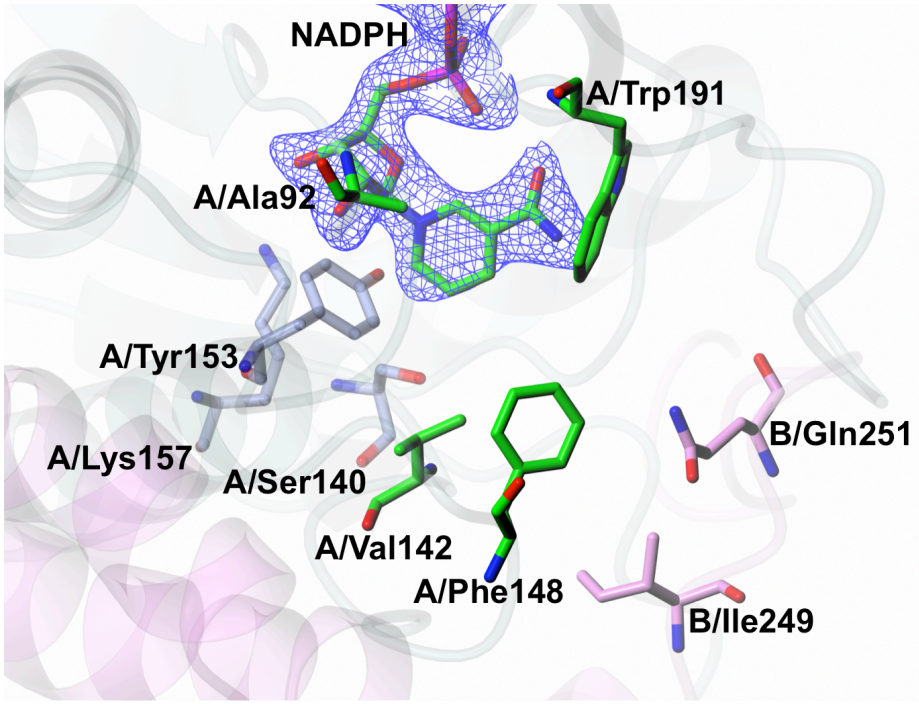


Figure 56 - The binding pocket/groove. Catalytic triad in ice blue: Ser-Try-Lys other residue of the monomer contributing to the binding pocket are in Green. Residues from the monomer are denoted as "A". Residues from the adjacent subunit of a dimer denoted as "B" and in pink are gln and ile, which are from the N-terminus helix. The electron density for NADPH is generated from the omit map (F_o-F_c) contoured to a level of 3σ which was obtained after refinement in the absence of NADPH.

Further evidence for the extensive contribution is observed with PISA analysis. For the major dimer interface, potentially 43 hydrogen bonds and 14 salt bridges contribute to the dimerising interaction. This is even further supported by the free energy for the dimerisation, where ΔG = -53.9 kcal/M.

DALI analysis shows that there are structures available with only limited sequence homology for SyADH. Even the model for the structure solution derived from the NADPH-dependent blue fluorescent protein (PDB entry: 3P19) has a relatively low sequence similarity of 29% and an RMSD of 2.0 Å over 239 residues. The tetrameric sepiaterin reductase from *Chlorobium tepidum* (PDB entry: 2BD0) is

also a NADPH-dependent oxidoreductase that has a low sequence similarity of 25% with an RMSD of 2.0 Å over 240 residues. Many of the targets identified by DALI have similar function to the SDR family, in that they have evolved to accept very bulky substrates such as steroids, like the 11β-hydroxysteroid dehydrogenase type 1 from *Mus musculus* (PDB entry: 3GMD). However, the 11β-hydroxysteroid dehydrogenase only has a sequence similarity of 18% and an RMSD of 3.0 Å over 264 residues. As such, the fold that SyADH exhibits appears to resemble the ability to accept bulky substrates.

The superimposition of active sites (**Figure 57**) for RasADH (ice blue) and SyADH (coral) illustrate the side-chain components of the hydrophobic tunnel in both active sites. NADPH sits at the base of the tunnel for both enzymes with residues sharing significant homology: Tyr153 (RasADH-Tyr150), Ser140 (Ser137), Met150 (His147) and Val141 (Val138) are conserved with respect to steric bulk. Interestingly in both cases, phenylalanine and leucine appears to play a role in closing the top of the hydrophobic tunnels. However they are mirrored in each enzyme, with Phe205 on the left and Leu144 on the right in RasADH, whereas Phe148 is on the right and Leu201 is on the left in SyADH.

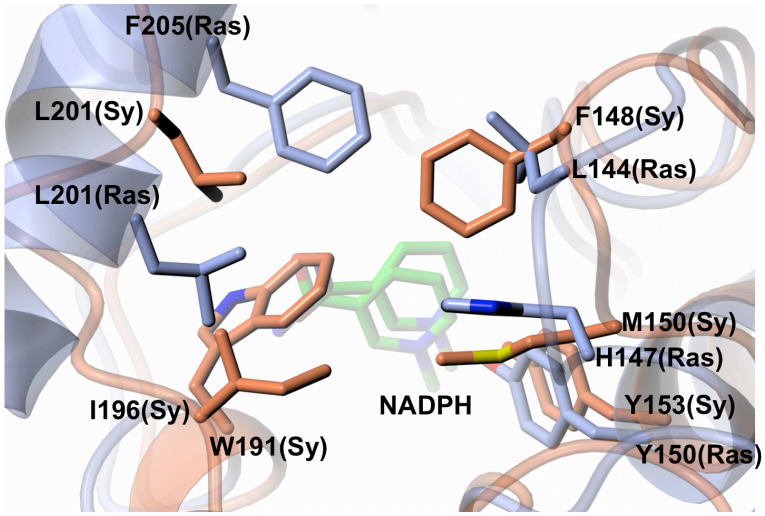


Figure 57 – Comparison of active site with RasADH (ice blue) and SyADH (coral). Majority of the residues share significant homology, however residues Phe205 and Leu144 in RasADH are opposite locations to Phe148 and Leu201 in SyADH.

Despite these similarities in the active site, there are three residues that give rise to larger steric bulk in RasADH (**Figure 58**): Ala92 (equivalent to Ile91 in RasADH), Ala146 (Leu142) and Ala194 (Gln191). This suggests, at least superficially, that SyADH

has the greater active site volume and would bind substrates less tightly than RasADH which could be the basis for observing lower enantioselectivies.

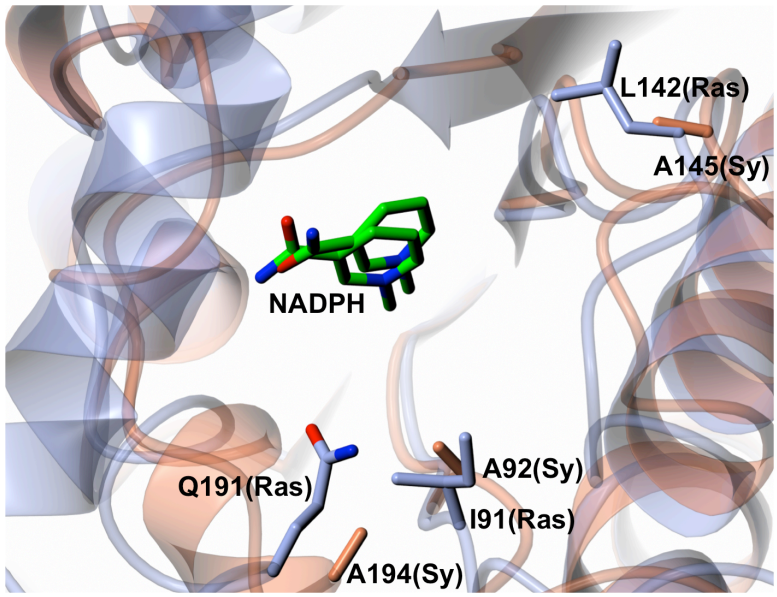


Figure 58 - Steric bulk differences in the active sites of RasADH (ice blue) and SyADH (coral). Where RasADH has larger steric bulk with residues Gln191, Ile91 and Leu142 compared to Ala194, Ala92 and Ala 145 in SyADH.

3.3. Modelling with AutoDock4

Both RasADH and SyADH were subjected to modelling (Section 2.7.) with AutoDock4 and AutoDock Vina in an attempt to rationalise the molecular basis for recognition of 'bulky-bulky' ketones. The ligand PPK (**Figure 59**) was used as the model to represent a 'bulky-bulky' ketone, where both enzymes convert PPK to the (*S*)-alcohol.

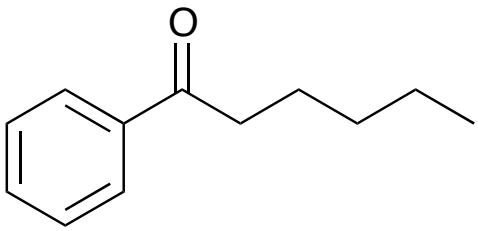


Figure 59 – Due to the solubility in water of the substrate *n*-Phenyl pentyl ketone (PPK), it was not possible to use for co-crystallisation experiments. As such, PPK was modelled into the active site using the docking suite AutoDock (Section 2.7.2.).

The ligand was built using methods described in (Section 2.7.1.) to generate the appropriate coordinate files and REFMAC5 dictionary. RasADH, SyADH and PPK were prepared in AutoDockTools (Section 2.7.2.), and then docking into the active site was attempted. AutoDock Vina had yielded more accurate results than AutoDock4 as well as providing additional information on potential interacting residues. The docking results are shown below (**Figure 60** and **Figure 61**):

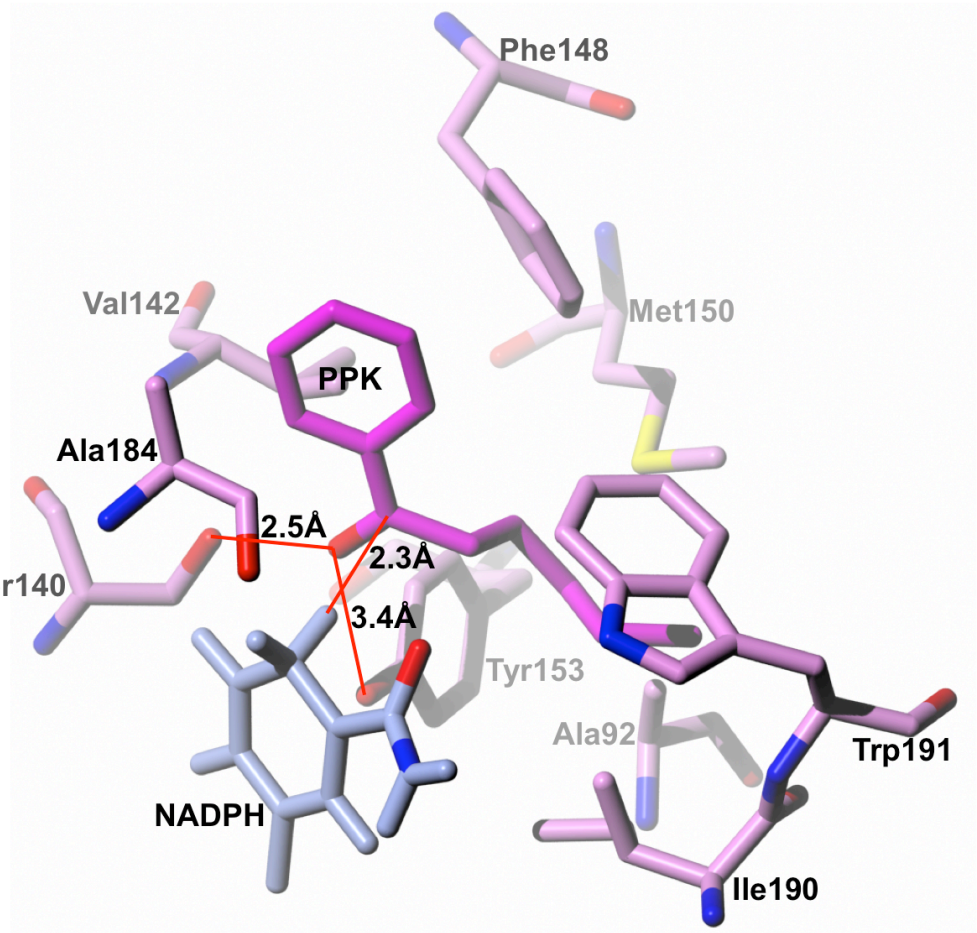


Figure 60 - SyADH docking with PPK. Key catalytic residue distances are shown with closely interacting residues as calculated from AutoDockVina. Again, PPK bound in a hydrophobic pocket, however PPK is able to fully extend its alkyl chain.

The most noticeable feature is that PPK is bound in a hydrophobic pocket in both enzymes with residues of similar function. SyADH's active site (**Figure 60**) exhibits potential π-stacking interaction with PPK and Phe148 which would allow aromatic substrates to bind in a fixed conformation. Where as in RasADH (**Figure 61**), the main driving force for binding PPK, besides the catalytic residues, is due to the potential hydrophobic interactions with the binding pocket. In both cases, the alkyl chain of PPK is fixed in the hydrophobic tunnel allowing for the *re*-face to be presented to the active residues resulting in the (*S*)-alcohol product. However SyADH presents a more favourable hydrophobic interaction between the alkyl chain and Trp191.

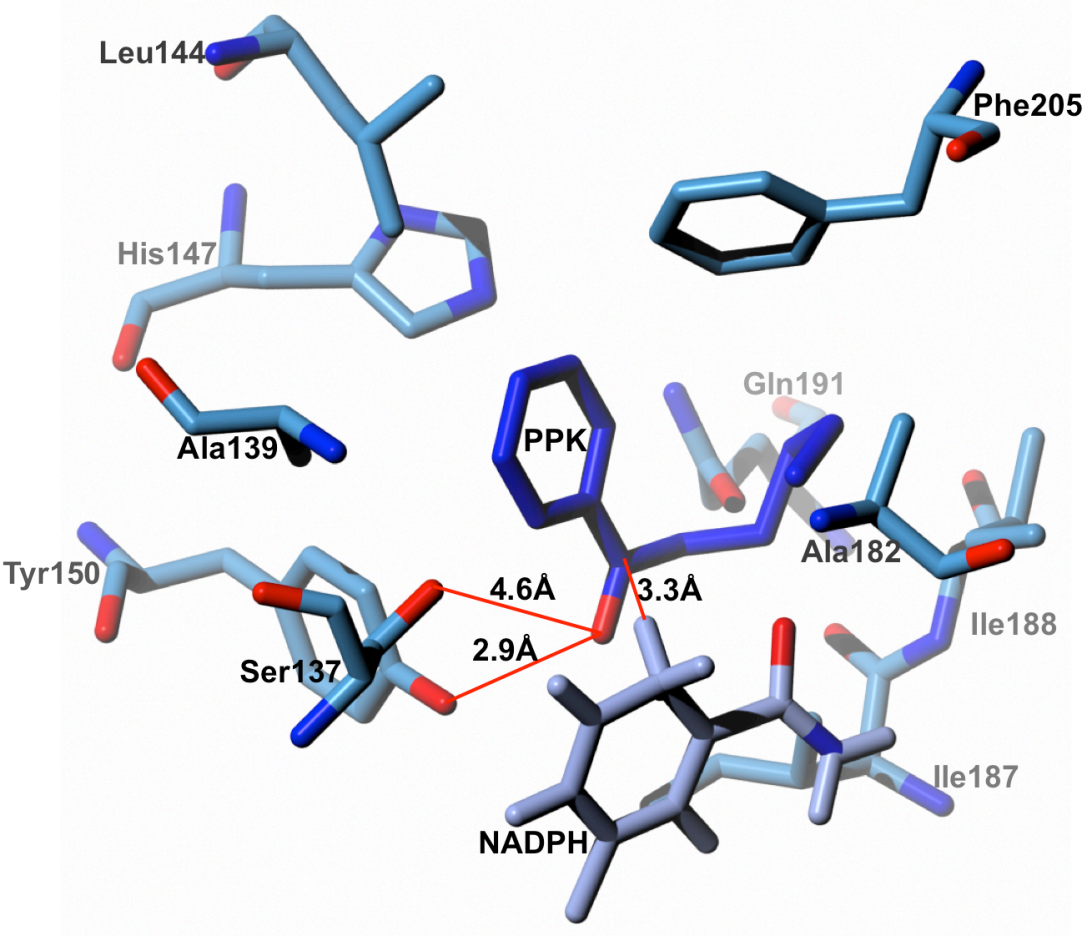


Figure 61 - RasADH docking with PPK. Key catalytic residue distances are shown with closely interacting residues as calculated from AutoDockVina. Most noticeably is PPK bound in a hydrophobic pocket also unable to fully extend the alkyl chain.

Superposing both docking results upon NADPH (**Figure 62**) enables comparisons of binding between the two enzymes. **Figure 62** clearly illustrates why the alkyl chain is unable to fully extend in RasADH, as Gln191 is effectively pushing the chain away. This poses as structural evidence for RasADH being unable to accommodate less flexible, aromatic, 'bulky-bulky' substrates, supporting substrate specificity studies by Kulig and co-workers in 2012 (26). The larger hydrophobic tunnel of SyADH may accommodate such substrates favourably interacting with Trp191. Despite a more favourable interaction between the alkyl chain and Trp191 in SyADH, the larger hydrophobic tunnel may hinder the enantioselectivity for PPK, 99% and 97% *e.e.* for RasADH and SyADH respectively. This also suggests a basis for SyADH's non-selectivity towards smaller prochiral ketones, as hypothesised by Lavandera and co-workers in 2008 (30).

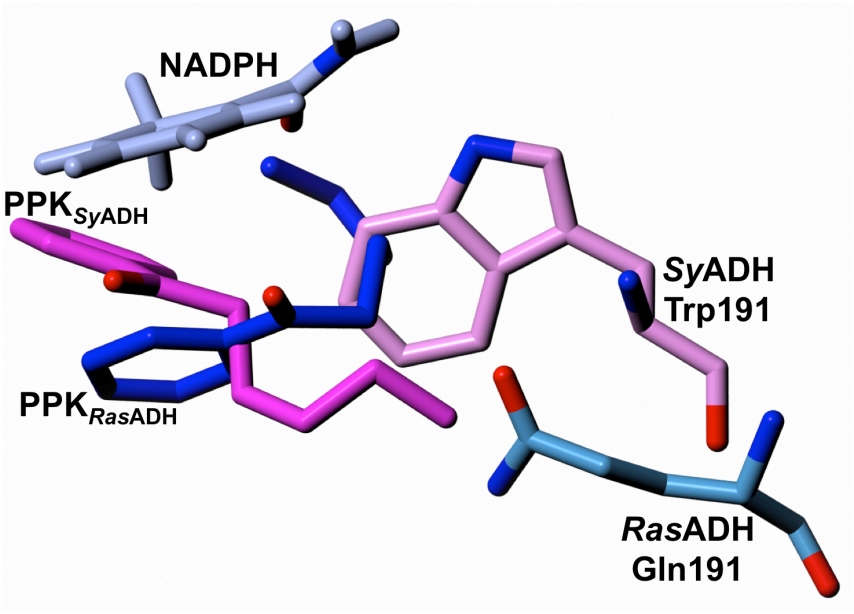


Figure 62 – RasADH-Gln191 and SyADH-Trp191 are potential residues for differentiating substrate specificities. Conformational differences of PPK result from residues SyADH-Trp191 and RasADH-Gln191. PPK is able to slot into the pocket due to Trp191 in SyADH. PPK is forced to curl-up due to Gln191 in RasADH.

Both models (**Figure 60** and **Figure 61**) propose that NADPH provides its *pro-S* hydride to attack the *re*-face of PPK's prochiral carbonyl. This suggests that both enzymes reactions proceed by the 'E4-pathway' (42). Previously incorrect assignment by Lerchner and co-workers in 2013 (28) suggested that the *pro-R* hydride from NADPH was delivered to the *re*-face of a prochiral ketone yielding the (*S*)-alcohol, suggesting for reactions proceeding by the 'E3-pathway'. Other ADHs that follow the selectivity from the 'E3-pathway' are yeast ADH (YADH), HLADH and TbADH (PDB entries: 2HCY, 6ADH and 1KEV respectively) from the MDR family, where the opposite face of the nicotinamide ring is presented to the active site, through rotation about the nicotinamide-ribose bond, relative to the conformation observed with the SDRs.

3.4. Structure-Guided Mutation: RasADH I187A Mutant

The elucidation of the crystal structure of RasADH has allowed for further structure-guided engineering. Kulig and co-workers recently showed that RasADH has a very broad substrate spectrum, however there were still some substrates such as benzoin (**Figure 63**) and its derivatives that had negligible or no detected activity with RasADH (26). Thus, the appropriate structure-guided mutations for RasADH could be made for RasADH to be a biocatalytic tool to oxidise or reduce sterically demanding substrates like benzoin.

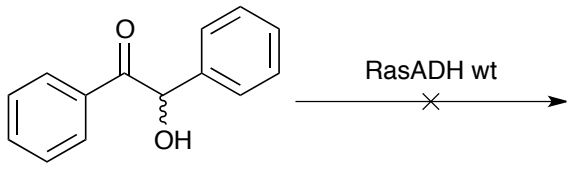


Figure 63 – The desired reaction for RasADH. The reduction or oxidation of benzoin like substrates.

The docking with PPK (**Figure 61**) presented the potential interactions with the catalytic residues. Using the same interactions as seen with PPK, benzoin was manually docked into the active site (**Figure 64**), such that the *pro-S* nicotinamide hydride was presented to the *re*-face.

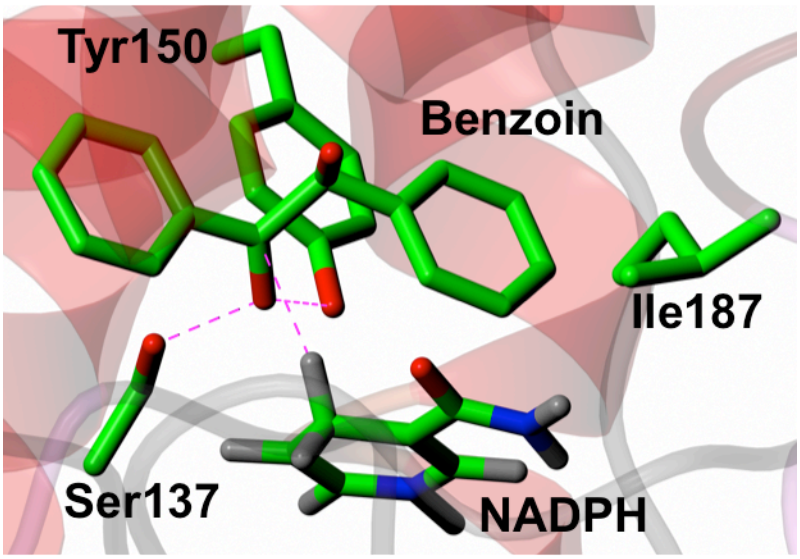


Figure 64 - Manual dock of benzoin in the active site of RasADH to rationalise why the active site could not accommodate substrates like benzoin.

3.4.1. Site-Directed Mutagenesis

It can be seen in the manual dock (**Figure 64**) that Ile187 may be causing a steric clash with the substrate. As such, if a smaller residue was present at position 187, it may allow for benzoin and its derivatives to fit in the active site. Hence, the mutation proposed, I187A (**Figure 65**), would potentially allow the active site to accommodate larger substrates.

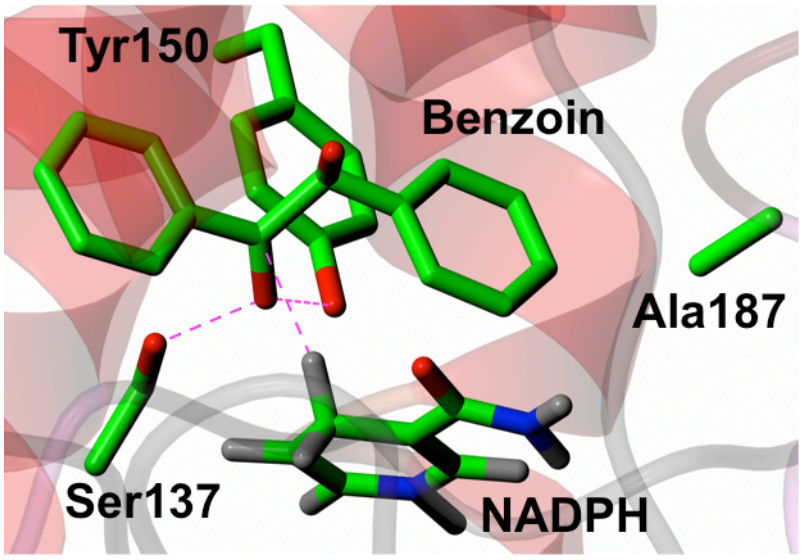


Figure 65 – The proposed mutation for RasADH at residue position 187. Mutating Ile187 to Ala187.

The proposed mutant I187A was successfully performed (Section 2.2. and 2.2.1.) and confirmed by gene sequence analysis.

3.4.2. Expression Test

An expression test (Section 2.3.4.1.) was performed to analyse the solubility of the mutant in solution. BL21 (DE3) was chosen as the expression strain. Surprisingly the mutant expresses much better and in much higher soluble yields than the wild type (**Figure 66**). The expression test samples were analysed by SDS-PAGE (Section 2.3.3. and 2.3.3.1.). Expression at 16°C overnight after induction yielded best results.

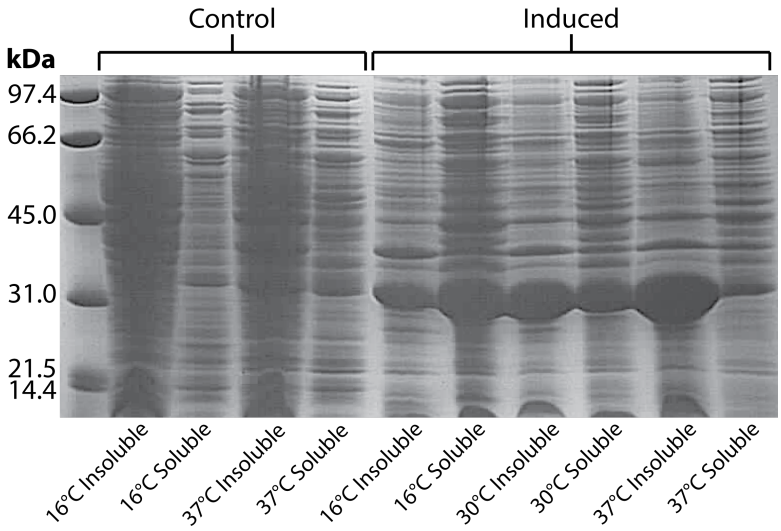


Figure 66 – Expression test of RasADH I187A in BL21 (DE3) expression strain. The mutant yielded better expression as well as more soluble protein.

3.4.3. Purification

The same purification protocol that was used to purify RasADH wt (Section 3.1.2) was used to successfully purify RasADH I187A by nickel IMAC (Section 2.4.3.). After nickel purification, the mutant was concentrated above 5 mg/mL without signs of aggregation. As such, the mutant was much more stable in solution compared to the wild type.

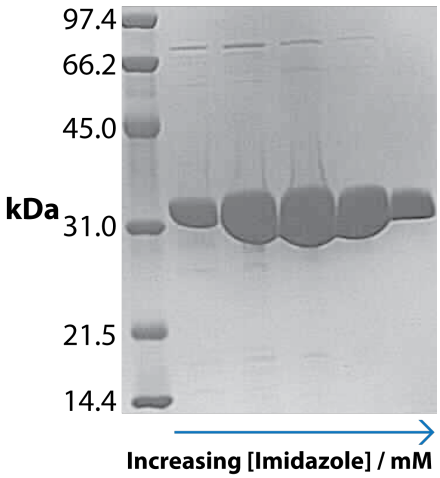


Figure 67 – SDS-PAGE analysis of RasADH I187A after nickel purification.

Following nickel purification (**Figure 67**), RasADH I187A was purified by SEC (Section 2.4.4.). Interestingly, the second peak at approximately 100 mL elution volume was nearly undetectable (**Figure 68**).

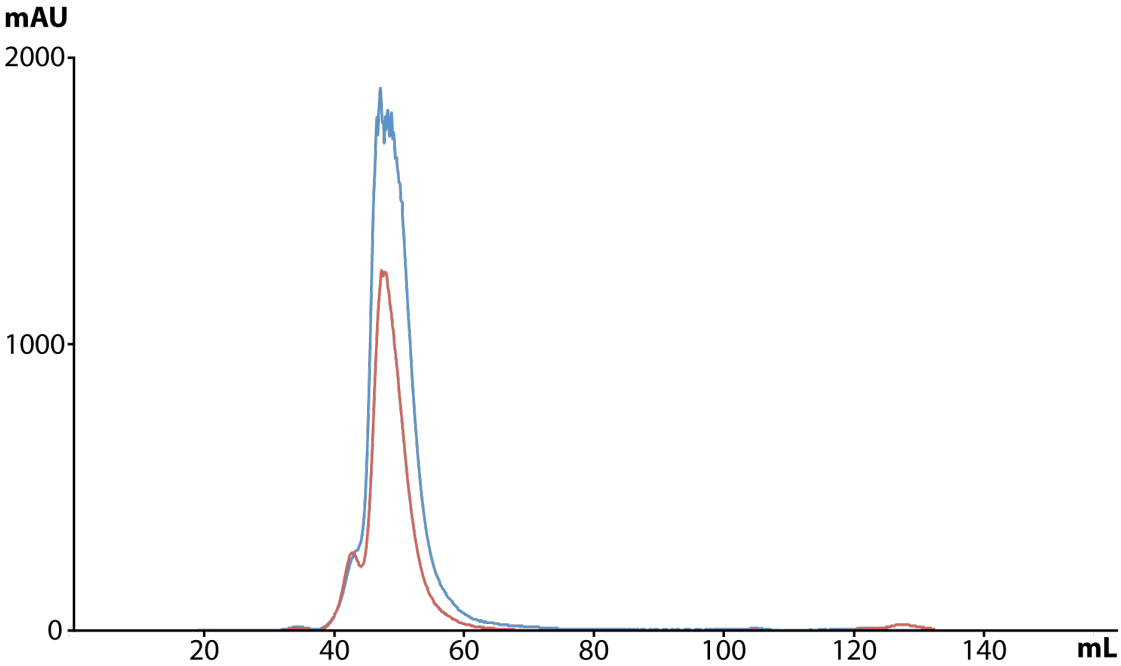


Figure 68 – Chromatagram for RasADH I187A purification by SEC. The peak between 40-60 mL represents the dimeric form of the mutant.

Protein purity after SEC was confirmed by SDS-PAGE analysis (Section 2.3.3. and 2.3.3.1). The protein was sufficiently pure (**Figure 69**) for further characterisation experiments by UV-VIS kinetics and biotransformations.

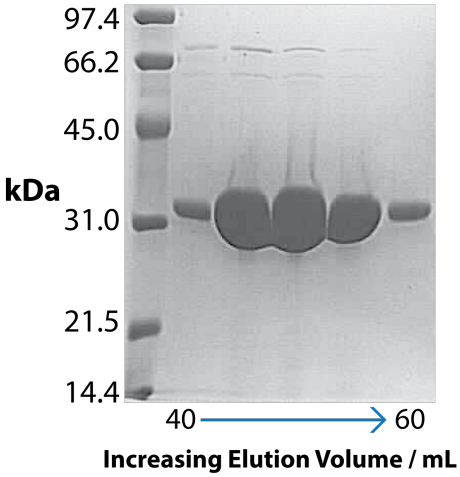


Figure 69 – SDS-PAGE analysis of RasADH I187A after SEC. The gel suggests the protein is sufficiently pure for further characterisation.

3.4.4. Kinetics

The mutant I187A was predicted to reduce the carbonyl with NADPH (**Figure 70**). Using UV-VIS spectroscopy, the reaction was monitoring the depletion of NADPH at 340 nm by an absorbance-time scan.

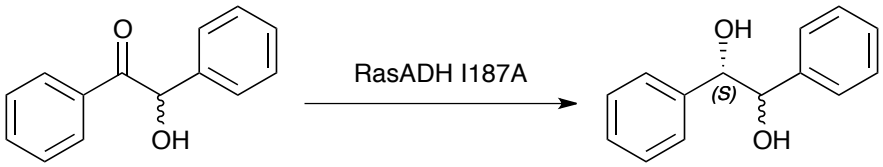


Figure 70 – Schematic of the predicted reaction for the RasADH 187A mutant.

However, due to the very limited solubility of benzoin in buffer, the reaction could not be monitored in solution. Despite optimising the buffer with additional water miscible solvents, phase separation would occur with the substrate forming an oil drop and the reaction cell becoming turbid.

3.4.5. GC Analysis

GC analysis was also attempted to monitor the kinetics parameters for RasADH I87A with the substrate benzoin. However, the results for the control conditions without enzyme were identical to the reaction with enzyme. Therefore, the I187A mutant was not sufficient enough to accommodate large substrates like benzoin.

3.5. TADH

TADH is an MDR from *Thermus* sp. Serena Gargiulo from Delft University of Technology in Delft had cloned the gene encoding TADH, expressed, purified and crystallised the protein. Purification was performed in one step by denaturing the lysate except for the desired protein due to its thermophilic properties and purifying with SEC. TADH has major benefits as a biocatalyst in that it utilises NADH (over NADPH), has a very broad substrate specificity producing enatiomerically pure (*S*)-alcohols, such as (S)-2-pentaol (>99% *e.e.*) from 2-pentanone and it is also able to tolerate up to 10% *v.v* water-miscible solvents.

The data collection and refinement statistics for TADH are shown below (**Table 3**):

	TADH
Beamline	Diamond I04-1
Wavelength (Å)	0.91999
Resolution (Å)	48.22 - 2.74 (2.91-2.74)
Space Group	P3121
Unit Cell (Å)	a = b = 87.36; c = 337.56; α = β = 90; γ = 120
No. of molecules in the asymmetric unit	4
Unique reflections	40559 (6444)
Completeness (%)	99.9 (99.4)
R_{merge} (%)	0.09 (0.50)
$R_{p.i.m.}$	0.03 (0.15)
Multiplicity	12.2 (11.9)
$\langle I/\sigma(I) \rangle$	22.2 (4.9)
$CC_{1/2}$	0.99 (0.81)
Overall *B* factor from Wilson plot (Å^2)	54
R_{cryst}/R_{free} (%)	19.6/24.6
No. protein atoms	10147
No. water molecules	165
r.m.s.d. 1-2 bonds (Å)	0.01
r.m.s.d. 1-3 bonds (°)	1.53
Avge main chain B (Å^2)	56
Avge side chain B (Å^2)	60
Avge water B (Å^2)	44
Avge ligand B (Å^2)	65
PDB entry	To be deposited

Table 3 - The data collection and refinement statistics for TADH.

3.5.1. Structure Building

The structure of TADH was solved using the molecular replacement program MOLREP. A monomer model was derived from *Pseudomonas putida* formaldehyde dehydrogenase (PDB entry: 1KOL). The structure solution contained four subunits in the asymmetric unit assembling a homotetramer. The structure was built and refined with iterative cycles of COOT and REFMAC5 with local NCS restraints. Once the model for TADH was built and refined to a satisfactory level, the omit map, revealed residual electron density for NADH to be rationally assigned and refined in three of the four active sites. Additionally, two zinc sites per subunit could be assigned for the active site and the structural site. TADH was built (**Figure 71**) and refined to R_{cryst}/R_{free} values of 20.2%/24.2%.

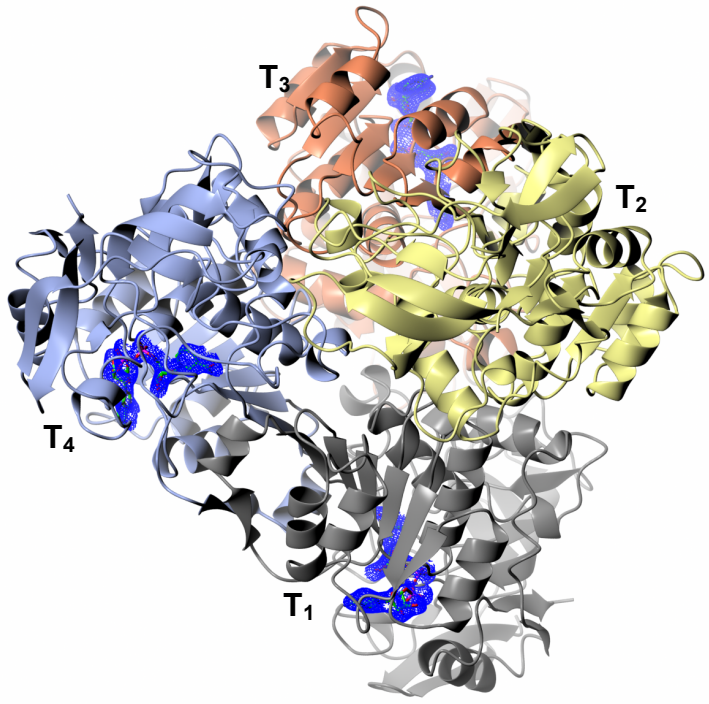


Figure 71 - The asymmetric unit for TADH is a tetramer with NADH bound in 3 of 4 active sites. Each subunit for TADH is denoted as T. The electron density for NADH is generated from the omit map (F_o-F_c) contoured to a level of 3σ which was obtained after refinement in the absence of NADH.

A monomer of TADH, represented by subunit T_4 (**Figure 72**) is comprised of two distinct domains, a metal binding domain and a nucleotide Rossman fold domain. The metal binding domain is supported by an array of β-sheet core formed by β1 (residues 2-6), β2 (12-16), β3 (27-36), β4 (55-56), β5 (61-68), β6 (80-82), β7 (125-129), β8 (126-127), β16 (319-322) and β17 (341-344) along with helicies along the

peripheries: α1 (42-45), α2 (93-97), α3 (131-134), α4 (144-147) and α12 (326-335). At the centre of TADH, α11 (residues 296-308) lies at the base of the cleft forming an axis between the two domains. The nucleotide binding domain is formed by a classic Rossman fold with a central β-sheet formed from six alternating helicies and β-sheets by β9 (residues 172-175), β10 (196-200), β11 (215-218), β12 (239-242), β13 (262-266), β15 (286-290) along with β14 (273-277) on the periphery. The helicies encompassing the β-sheet is formed by α5 (153-164), α6 (179-191), α7 (203-212), α8 (224-231), α9 (247-256) and α10 (279-284).

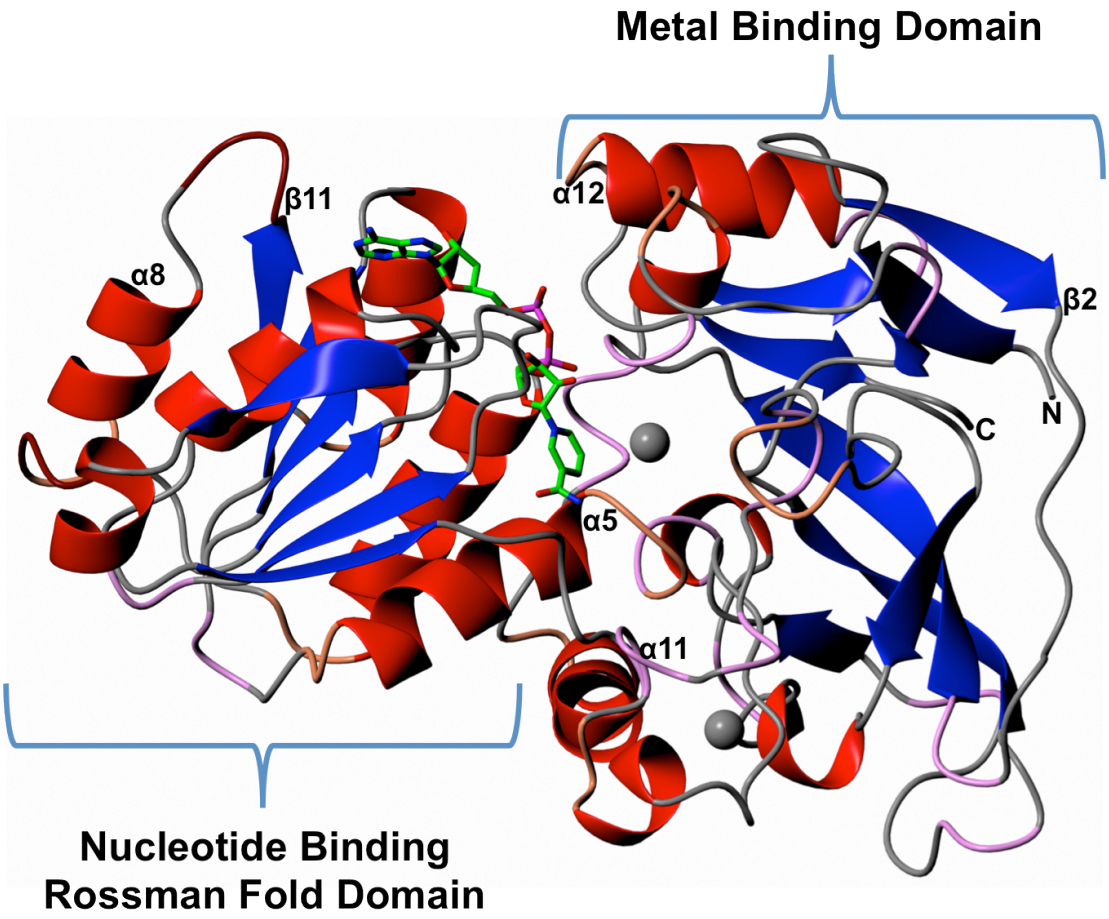


Figure 72 - A monomer of TADH, represented by T4, displaying the Rossman fold and metal binding domains.

Only three of the four active sites of TADH are bound to NADH. As such, when superimposing subunits T_2 (*apo*) in coral and T_4 (*holo*) in ice blue (**Figure 73**), it appears α11 is the axis upon which the two domains may close in for cofactor binding. It is apparent that upon cofactor binding, secondary structure is conserved however there is a small degree of displacement (RMSD of 0.86 Å over 346 residues).

The nicotinamide cofactor resides in the centre of the cleft of a TADH monomer with strict selectivity towards NADH due to Asp200 repelling any potential phosphate groups bound to the 2' position of the ribose-adenine ring. However Arg201 is near-by allowing π-stacking interaction with the adenine ring of the cofactor.

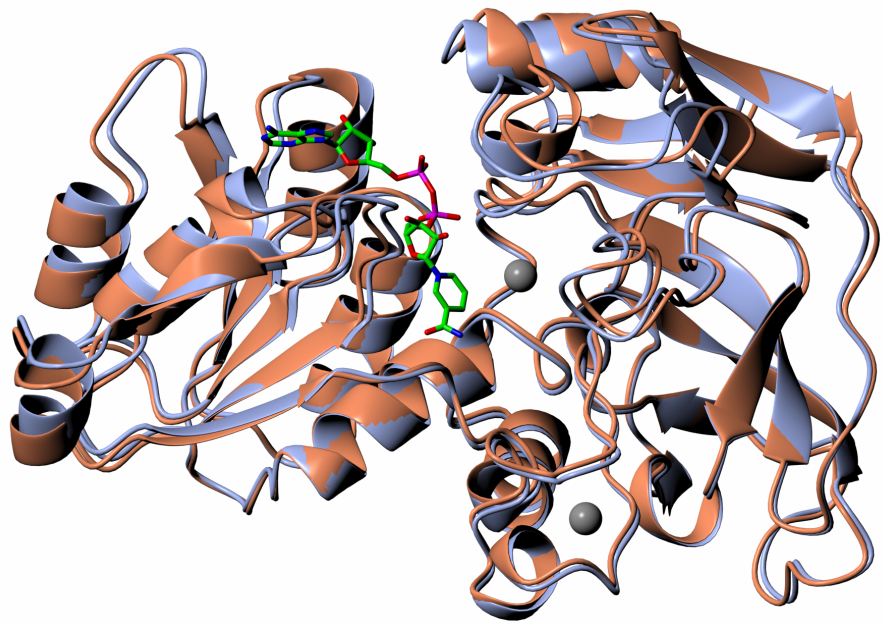


Figure 73 - Subunits T_2 (*apo*) coral and T_4 (*holo*) ice blue displaying displacement in cofactor binding. The two domains close in when the cofactor is bound.

The potential ligands (Asp152, His53 and Cys38) for the catalytic zinc (**Figure 74**) are similar to that of HLADH as studied by Eklund (31). It appears that the zinc cofactor is highly conserved throughout MDRs not only with its catalytic function, but also its structural role. In the case of the reduction of 2-pentanone to (*S*)-2-pentanol, the *re*-face of the prochiral carbonyl must be attacked. The positioning of the cofactor suggests for the delivery of the *pro-R* hydride to the relevant substrate, in this case 2-pentanone. Therefore TADH reactions should proceed by the 'E3-pathway' similar to YADH, HLADH and TbADH.

The active site for TADH is formed in the cleft between the two domains giving rise to a hydrophobic tunnel formed by residues: Ser40 (the potential proton donor), Val108, Leu292, Val268, Ile49, Tyr111 and Phe115. However due to the inherent steric bulk exhibited by the active site residues, the hydrophobic tunnel appears very narrow. Nevertheless, it is possible the active site could be more flexible than it

seem allowing for the accommodation of larger substrates due to the axis between the two domains and the substrate specificity is still relatively broad for MDRs.

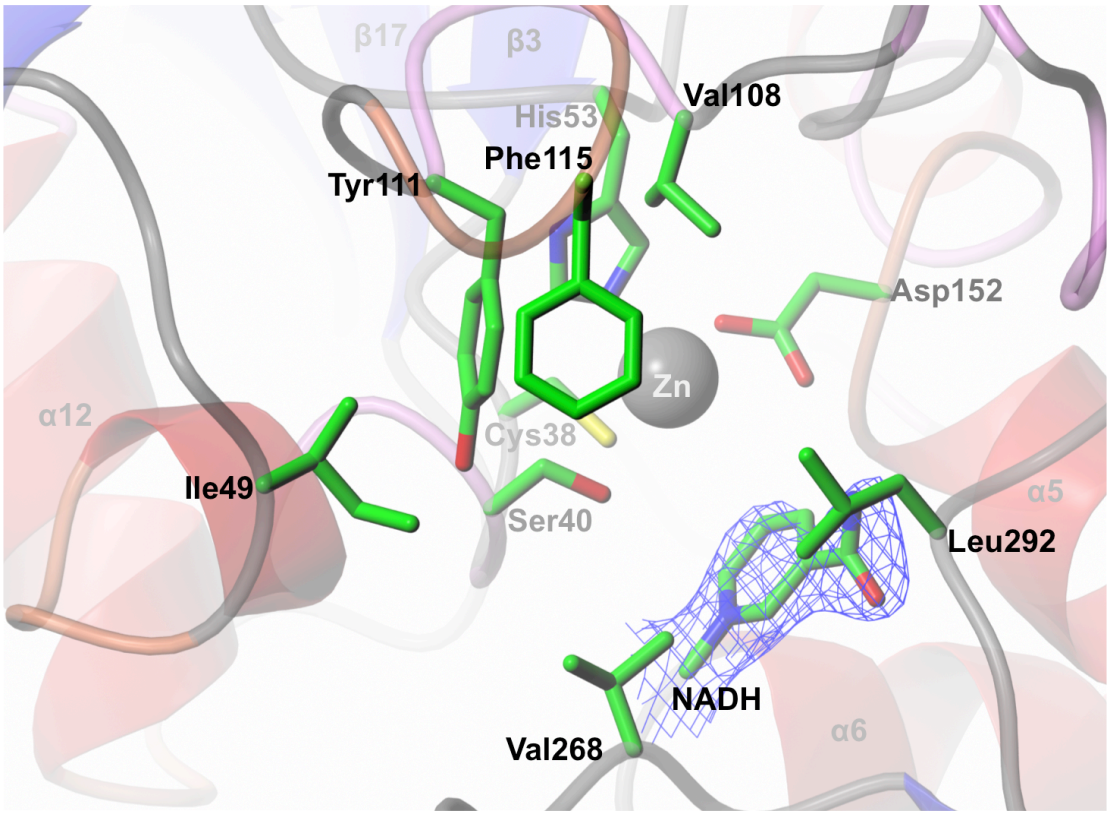


Figure 74 – The active site of TADH in subunit T_4, where zinc is coordinated to three protein ligands. The electron density for NADH is generated from the omit map (F_o-F_c) contoured to a level of 3σ which was obtained after refinement in the absence of NADH.

Analysis with PISA shows the major dimer interface, represented by subunits T_4 in ice blue and T_1 in black (**Figure 75**), is formed from 32 hydrogen bonds and 14 salt bridges. The Gibbs free energy for the dimer formation is relatively low, where ΔG = -13.1 kcal/M. Despite many hydrogen bonds and salt bridges, the Gibbs free energy for dimerisation is unusually low for MDRs. The free energy may suggest TADH has a preference towards the monomeric form, however this is very unlikely due to the favourable interactions between the dimer interface.

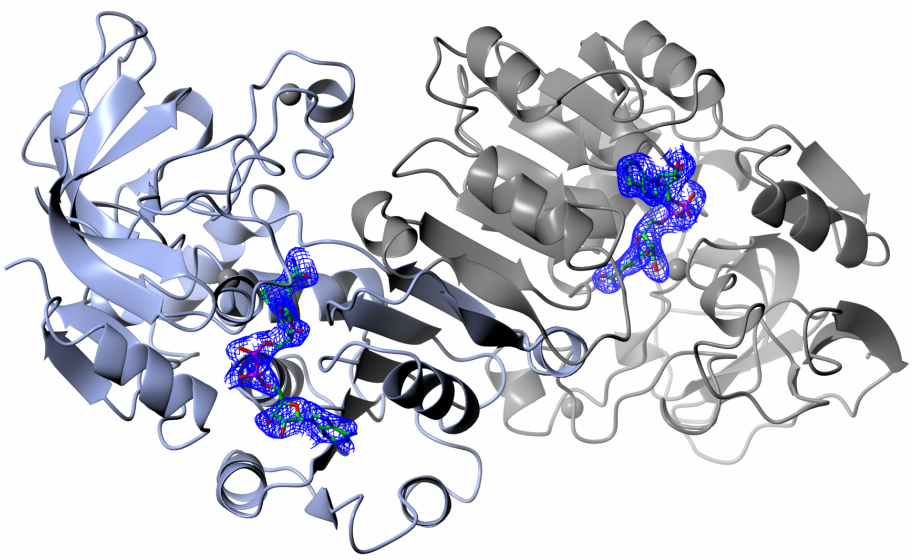


Figure 75 – The major dimer interface for TADH represented by subunits T_4 (blue) and T_1 (black). The electron density for NADH is generated from the omit map (F_o-F_c) contoured to a level of 3σ which was obtained after refinement in the absence of NADH.

Aside from the model derived from formaldehyde dehydrogenase (PDB entry: 1KOL) that was used to solve the structure, DALI identified several metal-dependent MDRs. Such as the tetrameric formaldehyde dismutase from *Pseudomonas putida* (PDB entry: 2DPH) with a sequence similarity of 28% and an RMSD of 1.6 Å over 398 residues. The formaldehyde dismutase also utilises zinc and NADH as the cofactors with oxidoreductase activity. DALI also identified the dimeric hypoxanthine phosphorisosyltransferase from *Trypanosomea cruzi* (PDB entry: 1P18) which has sequence similarity of 26% and an RMSD of 1.8 Å over 356 residues. Interestingly the transferase utilises magnesium rather than zinc for transferase activity, where two magnesium atoms binds the diphosphate substrate in an octahedral geometry, rather than tetrahedrally with zinc in MDRs.

3.6. CPCR2

CPCR2 is an MDR from *Candida parapsilosis*. Christoph Loderer from Technische Universität Dresden in Dresden had previously cloned the gene encoding CPCR2 into a pET-22b C-terminus Strep-tag vector, expressed and purified and attempted to crystallise the Strep-tag construct. However, there were no successes with previous crystallisation experiments.

The pET-22b vector allowed protein secretion in the periplasm due to a PelB leader sequence. This can often lead to an increase in protein stability, which allowed for chemical lysis with lysozyme and DNAse rather than mechanical methods. However the protein yield from chemical cell lysis is not always be reproducible. Despite periplasm secretion, CPCR2 degraded in solution within 48 hours of cell lysis at 4°C.

CPCR2 has very valuable properties as a biocatalyst. Such as, exhibiting a wide substrate spectrum, nearly exclusively producing enantiomerically pure 'bulky-small' (*S*)-alcohols and its ability to retain all activity in neat substrate as lyophilised cells. Mutational studies have been performed based on a high quality homology model from the ADH from the archaeon *Sulfolobus solfataricus* (SsADH). However, SsADH only has 34% homology, therefore many regions may not be conserved. As such, a crystal structure would be very valuable for further mutagenesis studies.

3.6.1. Cloning

The gene encoding CPCR2 was sub-cloned into the pET-YSBLIC3C vector, which is based on pET-28a, promoting cytoplasmic expression. The vector fuses a cleavable N-terminal hexa His-tag allowing for a much cheaper purification by IMAC. Strep-tag purification requires elution by biotin on a streptavidin column, where biotin is more than three times the price of imidazole.

The template, received as a gift from Prof. Ansorge-Schumacher, was successfully amplified (Section 2.1.1.), with the appropriate primers (**Figure 76** - left). After purification and T4 digest, the insert was annealed to the pET-YSBLIC3C vector and

transformed into Nova Blue cells for plasmid extraction (Sections 2.1.4 - 2.1.10.). A double digest (**Figure 76** - right), with *Nco*1, *Ndo*1 and buffer 4 was performed with the plasmid after plasmid extraction (Section 2.1.11.) to evaluate the presence of the insert.

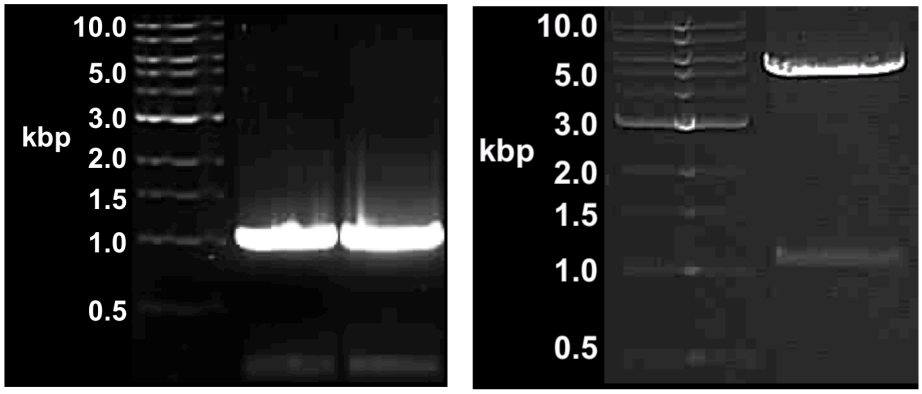


Figure 76 – CPCR2 1% agarose gel of PCR products (left) and double digest (right) of the plasmid with *Nde*1 and *Nco*1. The gels and samples were prepared as described in sections 2.1.2 and 2.1.3. The 1 kb ladder was obtained from New England BioLabs.

3.6.2. Expression

The plasmid was transformed into BL21 (DE3) and Rosetta 2 (DE3) for expression testing (Section 2.3.4.1.), as seen in **Figure 77**. Cells were lysed by sonication (Section 2.3.6.1.). The observed overexpression in both strains was similar, however BL21 (DE3) resulted in greater amount of soluble protein at 16°C.

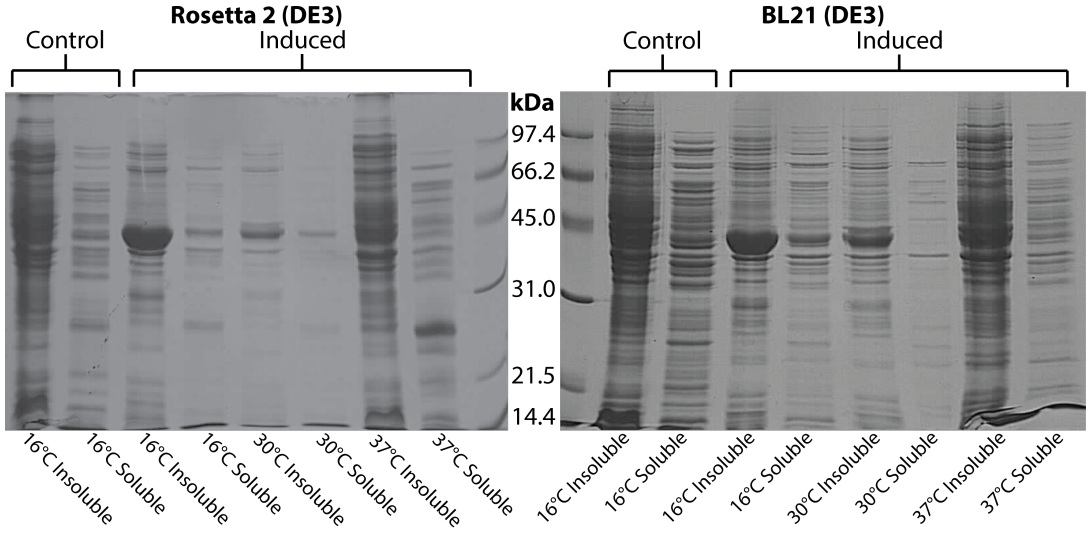


Figure 77 - Expression test: Rossetta 2 (left) and BL21 (right) performed using the methods described in section 2.2.4.1. Results show that greater soluble expression is achieved with BL21 (DE3) strain.

3.6.3. Purification

CPCR2 was known to be a very sensitive and difficult to handle protein in the pET-22b strep-tag vector; after the strep-tag purification protein concentration was limited to 5 mg/mL and after SEC protein concentration was limited to 10 mg/mL. Exceeding those limits would cause protein aggregation.

Cobalt was chosen as the divalent metal for IMAC. Cobalt is known to have higher selectivity for histidine, however the binding affinity is lower than nickel. The benefit of using cobalt was to reduce non-specific binding and allow the protein to elute at lower concentrations of imidazole (**Figure 78**). Purification by nickel IMAC was attempted, however CPCR2 did not bind well to the immobilised nickel.

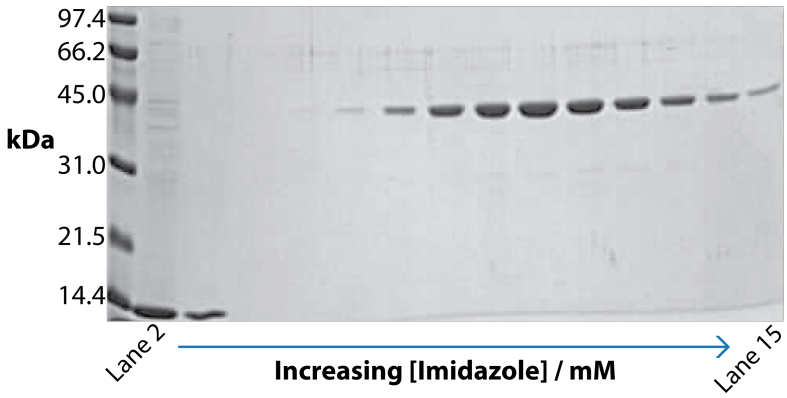


Figure 78 – SDS-PAGE analysis of CPCR2 purified by cobalt IMAC. Lane 1 is a Low Molecular Weight Ladder from BioRad. Lane 2 and 3 are samples from 20 mM and 50 mM imidazole wash steps. Lane 4-15 are samples from the elution gradient. Gel and sample preparation were performed as described in Section 2.3.3. and 2.3.3.1.

The His-tag construct was noticeably more soluble; after cobalt purification it was possible to concentrate the protein beyond 100 mg/mL with no signs of protein aggregation.

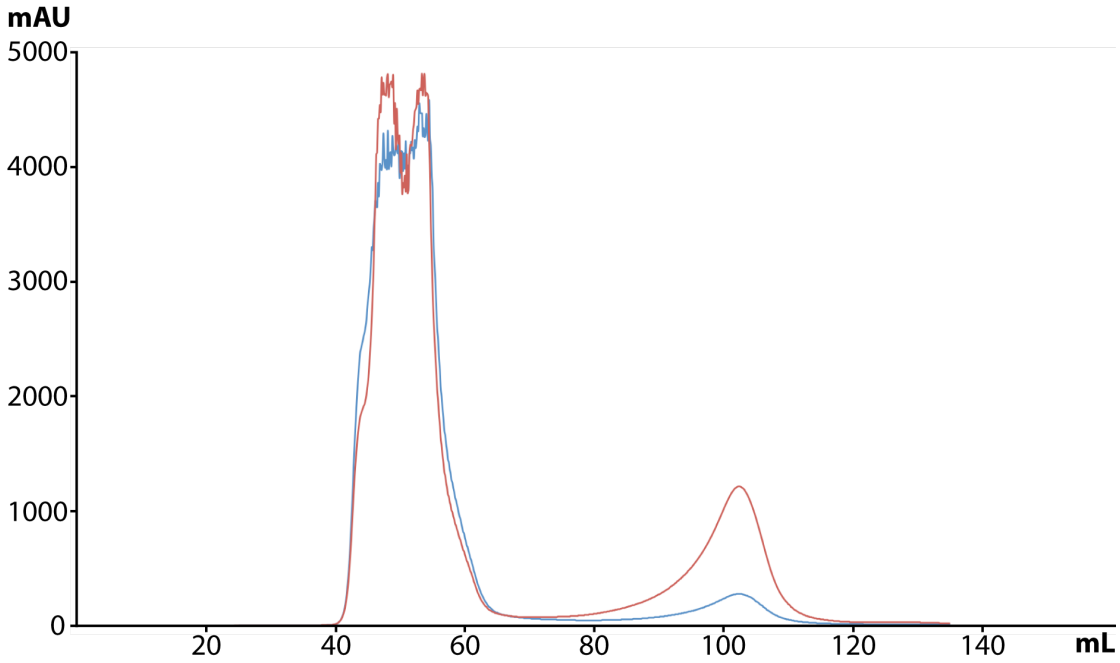


Figure 79 - Chromatagram of the SEC purification of CPCR2 from a 6 L culture preparation. Elution between 40-60 mL would represent the dimer form of CPCR2, elution at 100 ml would be the monomer.

CPCR2 was further purified by SEC (**Figure 79**). Comparing with the calibration curve for the Hiload 16/60 Superdex 75 Prep Grade, the first peak between 40-60 mL elution volume is believed to be the dimeric form. The second peak at 100 ml elution volume is most likely the monomeric form. This suggests CPCR2 primarily exists as a dimer in solution.

3.6.4. Crystallisation

CPCR2 was screen with INDEX, PACT and CSS 1 + 2 crystallisation screens at 21 mg/mL. The most promising hits were found in INDEX after 24 hours, however the crystals formed had very thin needle like morphology. The condition was successfully optimised using hanging drop vapour diffusion method (Section 2.6.1) with the addition of additives producing much larger crystals (**Figure 80** – left). The best crystals were formed in 0.1 M Tris-HCl pH 8.0, 30% (*w*/*v*) PEG 3350 and 2-methyl-1,4-pentanediol (40% *v*/*v*) at 69 mg/mL protein concentration. For an *apo*-structure, crystals were flash cooled in liquid nitrogen prior to in-house diffraction analysis. For NADH and NAD^+ complexes, crystals were soaked in solutions of mother liquor with additional 20 mM NADH and NAD^+ respectively prior to flash cooling and in-house

diffraction analysis (**Figure 80** – right). The best diffracting crystals were sent to Diamond, Dicot for full dataset collection on the I03 beamline on 2nd May 2013, I02 beamline on 22nd June 2013 and I04-1 beamline on 14th July for the *apo*, NADH and NAD^+ datasets respectively.

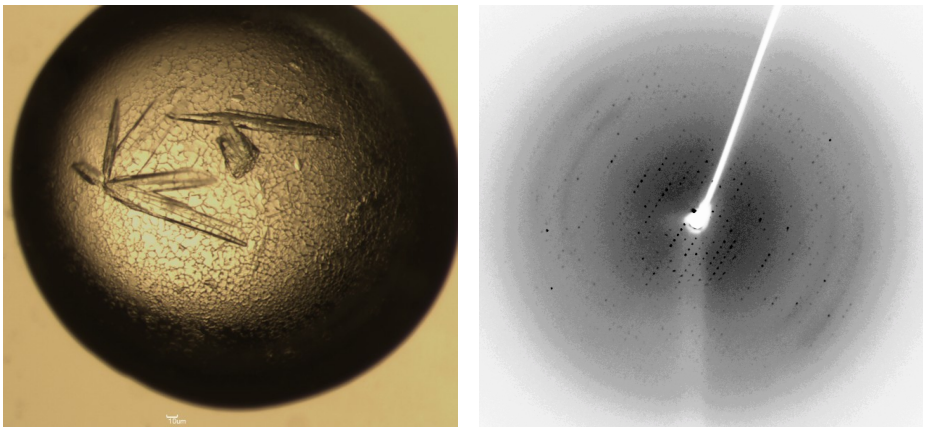


Figure 80 – Protein crystals (left) of CPCR2 in conditions containing 0.1 M Tris-HCl pH 8.0, 30% (*w*/*v*) PEG 3350 and 2-methyl-1,4-pentanediol (40% *v*/*v*) at 69 mg/mL protein concentration with its in-house diffraction pattern (right).

The data collection and refinement statistics for CPCR2 in complex with NADH (72) is shown below (**Table 4**):

	CPCR2 (NADH complex)
Beamline	Diamond I02
Wavelength (Å)	0.97949
Resolution (Å)	58.13-2.05 (2.13-2.05)
Space Group	P211
Unit Cell (Å)	a = 66.72; b = 88.84; c = 118.20; α = γ = 90.0; β = 100.4
No. of molecules in the asymmetric unit	4
Unique reflections	85098 (6264)
Completeness (%)	99.9 (100)
R_{merge} (%)	0.04 (0.58)
$R_{p.i.m.}$	0.04 (0.49)
Multiplicity	4.1 (4.1)
$\langle I/\sigma(I) \rangle$	18.7 (2.1)
$CC_{1/2}$	0.99 (0.67)
Overall *B* factor from Wilson plot (Å^2)	34
R_{cryst}/R_{free} (%)	20.5/24.1
No. protein atoms	9259
No. water molecules	243
r.m.s.d. 1-2 bonds (Å)	0.018
r.m.s.d. 1-3 bonds (°)	1.99
Avge main chain B (Å^2)	46
Avge side chain B (Å^2)	49
Avge water B (Å^2)	42
Avge ligand B (Å^2)	49
PDB entry	4C4O

Table 4 - The data collection and refinement statistics for CPCR2 in complex with NADH.

3.6.5. Structure Solution and Model Building

The initial crystal structure for the 2.8 Å *apo*-dataset, was solved with the automated molecular replacement program BALBES. A dimer model was derived from the tetramer structure of the ADH from *Bacillus stearothemophilus* (PDB entry: 1RJW). The structure solution revealed four subunits assembling as a tetramer in the asymmetric unit. The *apo*-structure was built and refined with iterative rounds of building and refinement with COOT and REFMAC5 to R_{cryst}/R_{free} values of 22.0%/26.1%. Subsequent 2.0 Å NADH and 2.3 Å NAD^+ datasets were solved using a monomer model of the refined *apo*-structure, both revealing a tetramer in the asymmetric unit. Once both complex models had been built to a satisfactory level, the omit map revealed density for the nicotinamide cofactor in all four active sites. However, higher resolution (>1.5 Å) is required to differentiate whether NADH or NAD^+ is bound, as the nicotinamide ring is bent in NADH due to the loss in aromaticity. The two complexes were identical, as such, only the NADH complex was fully built and refined. The NADH complex was built and refined to R_{cryst}/R_{free} values of 20.6%/24.5% (**Figure 81**).

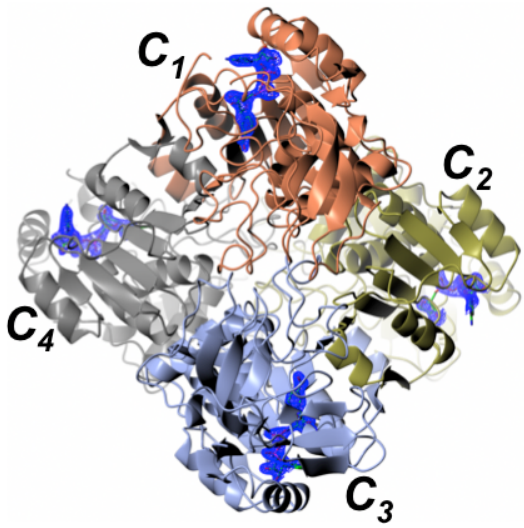


Figure 81 – The tetrameric model of NADH-complex in the asymmetric unit. Each subunit for CPCR2 is denoted as C. The electron density for the cofactor was generated from the omit map (F_o-F_c) contoured to 3σ.

The overall structural fold of CPCR2 is similar to many other MDRs, with a monomer represented by subunit C_1 consisting of a metal binding catalytic domain with a Rossman fold nucleotide binding domain (**Figure 82**). The metal binding domain is β-strand rich featuring ten strands: β1 (residues 6-7), β2 (9-13), β3 (17-21),

β4 (24-25), β5 (34-42), β6 (60-62), β7 (67-74), β8 (85-89), β9 (128-131), β10 (137-139) and a helix from the N-terminus α1 (47-52). The metal binding domain is connected by helix α2 (142-166) leading into the nucleotide binding domain consisting of a typical Rossman fold of a central β-sheet in the domain including β11 (173-177), β12 (217-219), β13 (233-236), β14 (256-260), β15 (266-270) and β16 (280-282) which is encased by five helices: α3 (181-193), α4 (204-213), α5 (241-251), α6 (241-251) and α7 (271-277). The helix α8 (289-301) in the cleft between two domains forms a pivot point for the two domains to close in on nucleotide binding as mentioned earlier with TADH. Following the pivot, two peripheral β-strands [β17 (308-312) and β18 (331-335)] are accompanied by helix α9 (316-325) completing the catalytic metal binding domain.

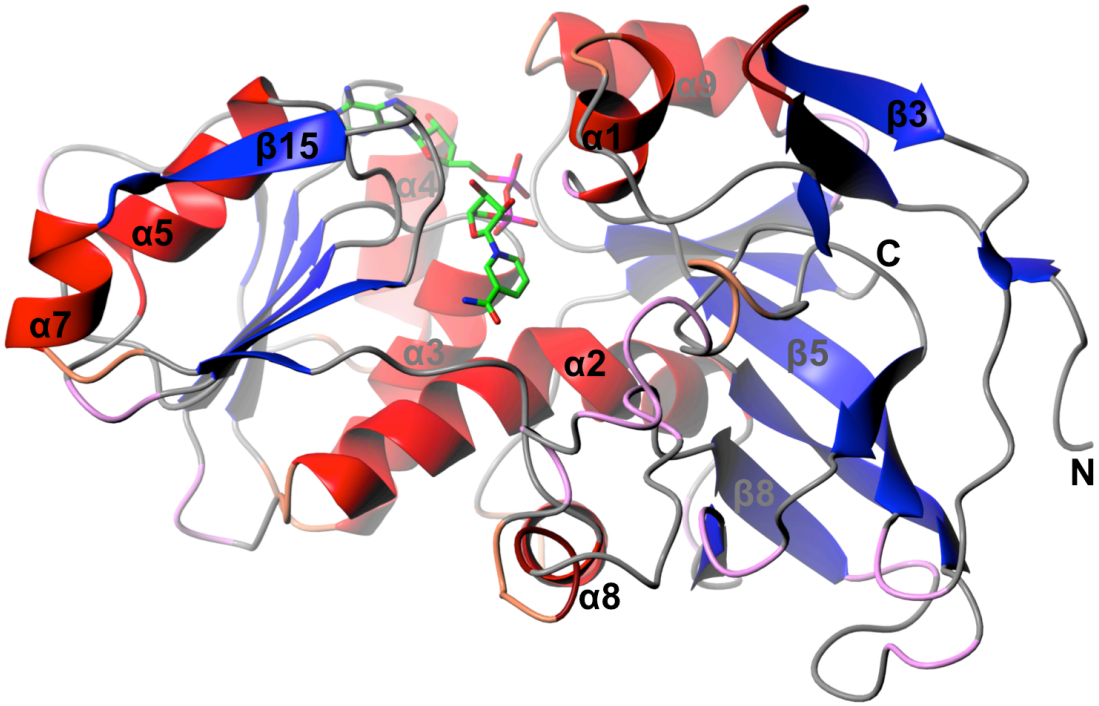


Figure 82 - A monomer of CPCR2 represented by subunit C_1 displaying the catalytic metal binding domain and the nucleotide binding domain.

NADH is bound between the cleft of the two domains with Lys202 providing π-stacking interactions with adenine group and ribose hydroxyl groups anchored by interactions with Asp201. The acidic residue dictates the strict specificity of the non-phosphorylated cofactor in CPCR2. Further interactions are made with His49 and Ser46, the potential proton donor, to the 2'-hydroxyl on the ribose adjacent to the nicotinamide group. The cofactor lies at the base of the cleft in a very hydrophobic

pocket featuring residues: Leu55, Val50, Leu119, Trp116, Pro92, Trp286, Phe285 and Leu262. Despite the inherent steric bulk displayed by the residues in the binding pocket, the active site is exposed and therefore relatively large compared to TADH.

The nicotinamide ring of NADH is bound in a conformation that would present the *pro-R* hydride of NADH to the active site (**Figure 83**). CPCR2 is predominantly (*S*)-selective, where the nicotinamide hydride is delivered to the *re*-face of the carbonyl giving the (*S*)-alcohol. As such, CPCR follows the 'E3-pathway' like many other MDRs such as YADH (PDB entry: 2HCY) and HLADH (PDB entry: 6ADH). However in cases such as 4-chloro-3-oxobutanoate, (*R*)-selectivity is observed however this is due to a switch in CIP priority rules.

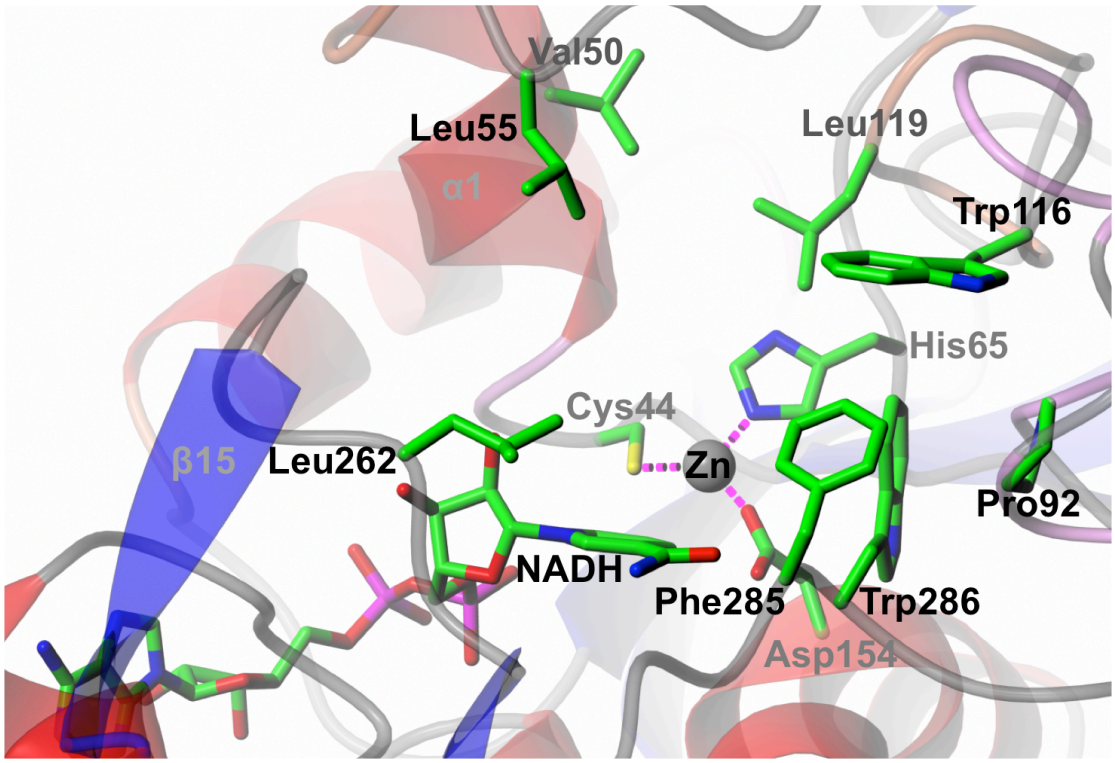


Figure 83 - The active site CPCR2 as in subunit C_1. NADH is bound in the base of the cleft with surrounding hydrophobic residues.

Previous studies on MDRs, such as HLADH (PDB entry: 6ADH), had shown that the catalytic zinc is coordinated in a tetrahedral geometry with three amino acid side-chain ligands (Cys46, His67 and Cys174 in HLADH, PDB entry: 6ADH) and either a water molecule or a substrates hydroxyl or carbonyl group. More recently in 2009, Baker and co-workers (33) have shown that the coordination sphere around the catalytic zinc is dynamic and changes in response to the binding and release of substrates. This particular study was performed on the MDR glucose dehydrogenase

from *Haloferax mediterranei* (*Hm*GDH), where Glu64 (equivalent to Glu68 in HLADH and Glu66 in CPCR2) was observed to be a ligand for the catalytic zinc at certain points to fulfil the tetrahedral coordination. These observations with *Hm*GDH were also seen and supported by the structure of CPCR2. In subunits C_1, C_3 and C_4 the catalytic zinc is coordinated by His65, Asp154 and Glu66 (homologous to Glu68 in HLADH) as well as a water molecule in a tetrahedral coordination sphere.

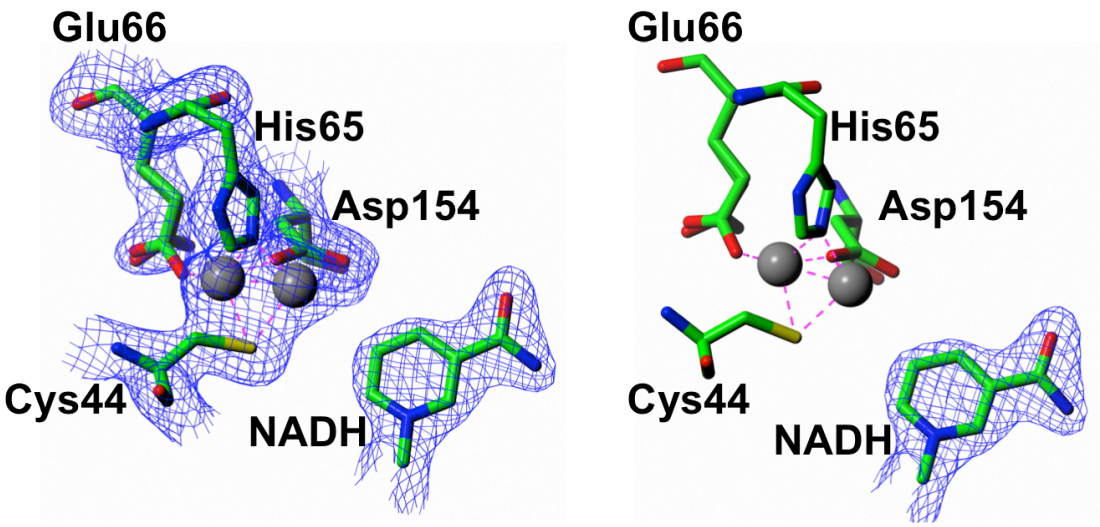


Figure 84 – Split conformation of zinc at the active site with (left) and without density (right), depicting activation and inactivation represented by subunit C_1. The electron density for the active site was obtained from $2F_o$-F_c and cofactor was generated from the omit map (F_o-F_c) contoured to 3σ.

In subunit C_3, Cys44 is not coordinated by the catalytic zinc (**Figure 85** - left), in certain cases of low resolution, the electron density for the loop region may resemble this residue. However as in Baker and co-workers study, this region is a flexible loop that is thought to close over the active site in MDRs upon substrate binding, as such in some cases the electron density is mostly poor or absent (**Figure 85** - right). In subunit C_1, there is excellent density for this loop region and the side-chain for Cys44 is present (**Figure 84**). Also the catalytic zinc in subunit C_1 has alternate conformations and two corresponding peaks in the difference map that can be modelled as two zinc atoms at 0.5 occupancy each. In one conformation is the tetrahedral coordination as seen with subunits C_1, C_3 and C_4 with zinc coordinated to His65, Asp154 and Glu66 as ligands; the second features zinc

displaced by 2.3 Å and now tetrahedrally coordinated with His65, Asp154 and Cys44 from the flexible loop. As such, Glu66 is observed in two conformations and modelled with 0.5 occupancy; one with Glu66 coordinating with the catalytic zinc and another pointing away from the metal. Hence, the involvement of Glu66 in zinc binding is clear and the observed changes in the coordination of zinc overall, provides further evidence of a dynamic coordination sphere for zinc and MDR-catalysed reactions.

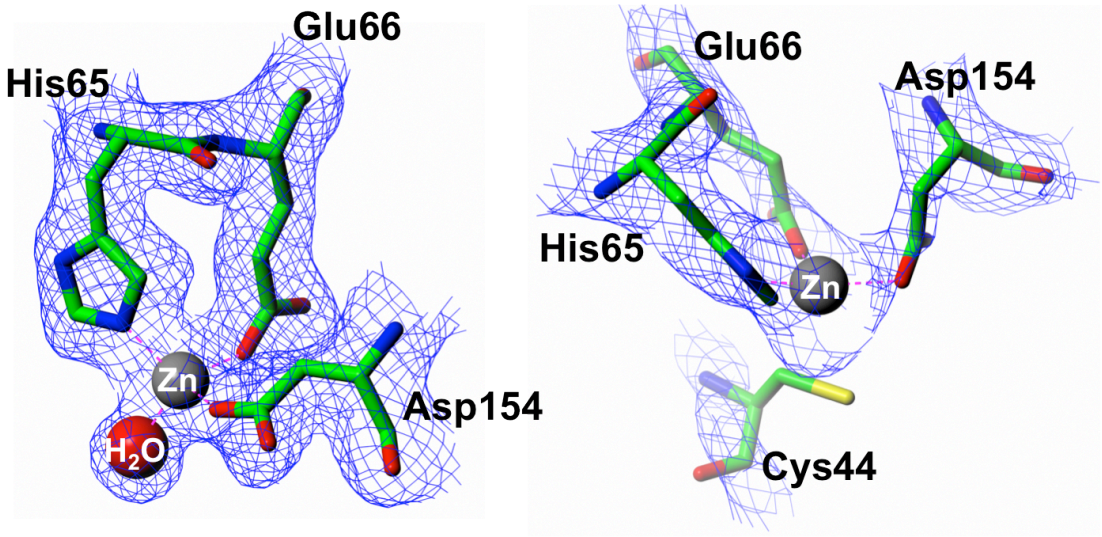


Figure 85 – Accurate interpretation of the active site of CPCR2 at the higher resolution NADH complex (left) represented by subunit C_3 where the electron density was obtained from $2F_o$-F_c contoured to 3σ. At lower resolution, the active site of the *apo*-CPCR2, the electron density was obtained from from the omit map (F_o-F_c) contoured to 3σ.

In solution, CPCR2 is predominately a dimer as seen with subunits C_2 and C_3 (**Figure 86**), however very small quantities of the monomer can be isolated. As such, PISA analysis shows that the major dimer interface potentially consists of 19 hydrogen bonds and 8 salt bridges with a ΔG value of -20.6 kcal/M. The relatively low number of interactions could contribute to its overall low stability. Superposing the major dimer interface for the *apo*- and NADH complex shows that there is little secondary structure conformational change upon cofactor binding. The forms of the dimer superpose with an RMSD of 1.06 Å over 642 atoms.

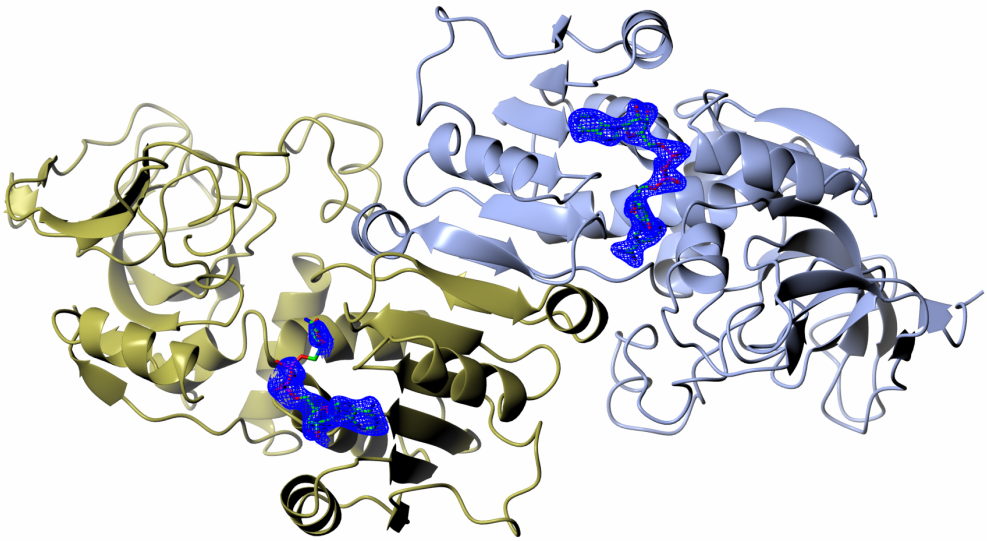


Figure 86 - The major dimer interface for CPCR2 represented by subunits C_2 (yellow) and C_3 (ice blue). The electron density for the cofactor was generated from the omit map (F_o-F_c) contoured to 3σ.

Comparing the major dimer interface of the homology model of CPCR2 in magenta with the crystal structure in cyan (**Figure 87**), despite high secondary structure conservation with MDRs, much of the models' secondary structure is displaced. With advanced homology modelling, the structure of some loop regions can be predicted. However, the actual positions of many atoms of the model deviate greatly from the crystal structure, as seen when superposing the dimers, the model has an RMSD of 2.52 Å over 590 atoms. Whereas, the *apo*-structure and NADH complex superpose with a better RMSD over 642 atoms.

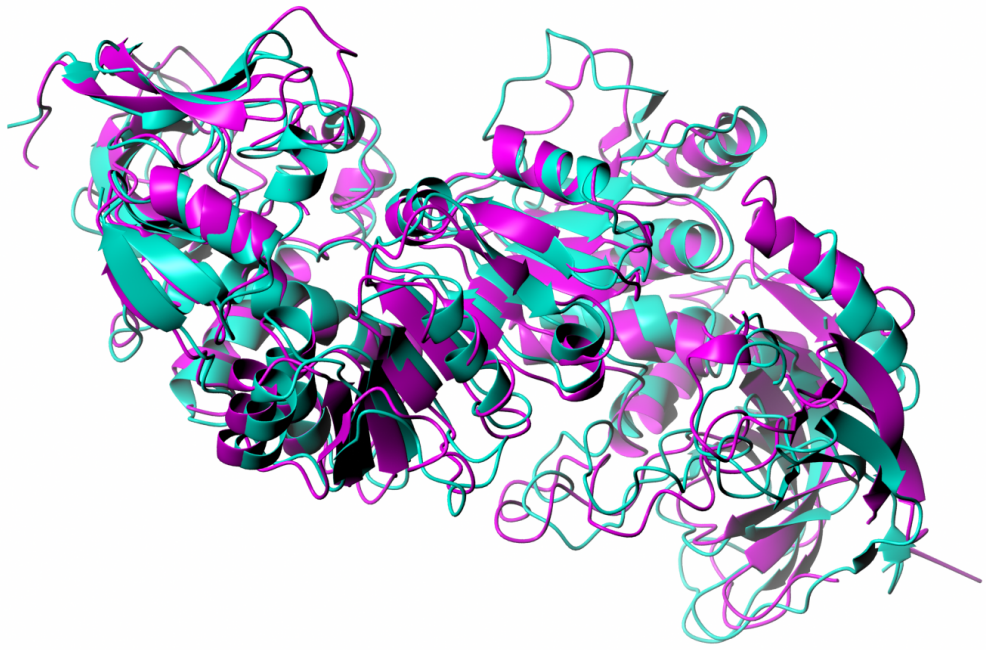


Figure 87 – SSM superpose of the homology model (magenta) with the crystal structure (cyan). Despite conservation of secondary structure, there is still a lot of difference with the crystal structure.

Using the *apo*-structure of CPCR2, many other MDRs were identified by DALI including the tetrameric MDR from *Bacillus stearothermophilus* strain LLD-R (PDB entry: 1RJW), where a model was derived to solve the structure. The MDR from *Bacillus stearothermophilus* has a sequence similarity of 30% and an RMSD of 1.8 Å over 339 residues. Using the NADH-complexed structure of CPCR2, again different results were yielded from DALI. Upon cofactor binding, CPCR2 now bears closer resemblance to the tetrameric MDR from *Pseudomonas aeruginosa* (PDB entry: 1LLU) which has 29% sequence similarity and an RMSD of 1.7 Å over 341 residues. Many of the targets identified by DALI are MDR with relatively low sequence similarity.

Despite CPCR2 (CCR) in ice blue and TADH (TAD) in coral belonging to the MDR family, they superimpose (**Figure 88**) relatively poorly with an RMSD of 1.66 Å over 299 residues. The overall secondary structures are similar, yet there is noticeable difference and displacement; such as helix (α8) in TADH is equivalent to a loop region in CPCR2. Aside from high conservation with the potential proton donor Ser46 (equivalent to Ser40 in TADH) and the protein ligands Cys44 (Cys38), His65 (His53) and Asp154 (Asp152) bound to catalytic; there is limited homology shared between CPCR2 and TADH. Trp286 and its equivalent in TADH-Phe85 are on opposite sides to each other, however it is suspected that both residues provide steric shielding from the active site towards the rest of the protein. The remaining residues: Phe285 (Leu292 in TADH), Leu119 (Tyr111) and Trp116 (Val108) share little homology to their TADH counterpart. This is mostly due to loops being heavily displaced not being conserved, as seen with TADH-Phe115 has no equivalent in CPCR2. However Leu55 (I49 in TADH) share some homology contributing towards the hydrophobicity in the active site.

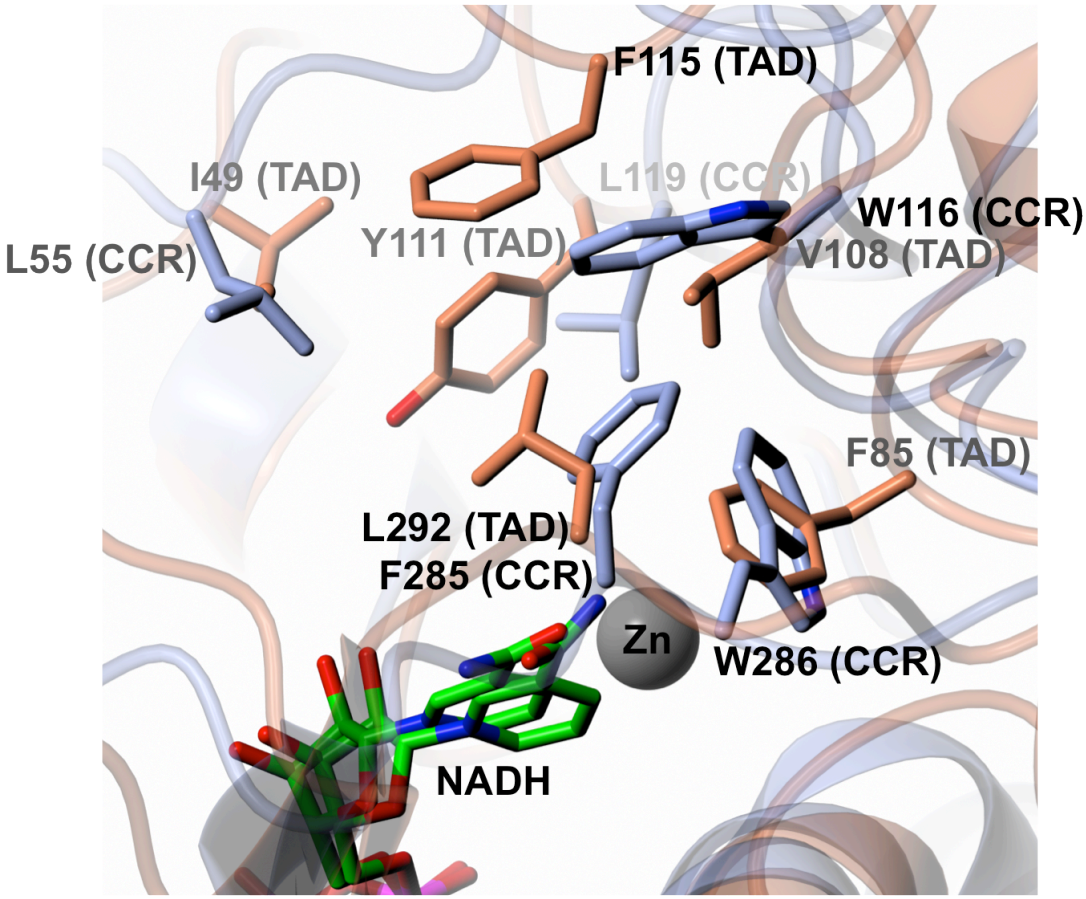


Figure 88 - Active site comparison between CPCR2 (CCR) in ice blue and TADH (TAD) in coral.

4. Conclusion

ADHs are a family of enzymes that have been thoroughly investigated as biocatalysts for academic and industrial purposes. As such, there are ADHs commercially available. Many ADHs have had their crystal structures determined which have lead to structure guided mutants with vastly improved activity. However the downfall of many ADHs is the limited substrate scope.

The elucidation of the crystal structures of RasADH and SyADH has revealed the molecular basis for the recognition of 'bulky-bulky' substrates of both enzymes, which could potentially widen up the substrate specificity of ADHs as a whole. Due to the low solubility of certain substrates, it is often not possible to obtain a ligand complex with these types of substrates. However, using the crystal structures of RasADH and SyADH, ligand docking has enabled insight into the molecular basis for the recognition of 'bulky-bulky' ketones. These dockings may also serve as a basis for future mutants for docking substrates that currently do not currently fit in the active site.

Previous mutant studies with CPCR2 were based on a homology model using SsADH, which only had 34% homology. Hence, many of the mutations made were inaccurate. Both TADH and CPCR2 have very wide substrate spectrum for 'small-bulky' ketones. The crystal structures of both have provided insight into the difference between the dimerising interface of thermophilic and mesophilic ADHs. Where TADH has shown a considerable number of potential interactions between the major dimer interface compared to that of CPCR2. These interactions most likely contribute towards TADHs thermostability and could be used as a basis to engineer these interactions into CPCR2 for better overall stability.

The mechanism for catalysis of MDRs is well-known and yet it is still a debated over. Recent investigations have shown that the catalytic zinc is not as static as previously thought. As such, the crystal structure for CPCR2 has provided supporting evidence of the dynamics in the active site. These dynamics captured in CPCR2 are not easily modelled by computational methods and may provide insights into dynamics for not just MDRs but other metalloproteins.

The basis of the work performed in this thesis was to provide crystal structures of each enzyme to allow for rational structure-guided mutations for the improvement of selectivity, substrate specificity and enzyme stability. As such, these objectives have been satisfied and the crystal structures of RasADH, SyADH, TADH and CPCR2 can be applied as a basis for future mutational studies for the benefit of providing a green alternative to abiotic asymmetric alcohol synthesis.

Abbreviations

ADH	Alcohol Dehydrogenase
APS	Ammonium Persulfate
BINAP	2,2'-bis(diphenylphosphino)-1,1'binaphthyl
CPCR2	*Candida parapsilosis* Alcohol Dehydrogenase
DKR	Dynamic Kinetic Resolution
DMS	Dimethyl Sulfide
DMSO	Dimethyl Sulfoxide
DYKAT	Dynamic Kinetic Asymmetric Transformation
e.e	Enantiomeric Excess
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny Broth
LIC	Ligation Independent Cloning
MDR	Medium Chain Alcohol Dehydrogenase/Carbonyl Reductase
MPD	2-Methyl-2,4-Pentanediol
PCC	Pyridinium chlorochromate
PCR	Polymerase Chain Reaction
PPK	Phenyl Pentyl Ketone
RasADH	*Ralstonia* Alcohol Dehydrogenase
SDM	Site Directed Mutagenesis
SDR	Short Chain Alcohol Dehydrogenase/Carbonyl Reductase
SEC	Size Exclusion Chromatography
SOC	Super Optimal broth with Catabolite repression
SyADH	*Sphingobium* Alcohol Dehydrogenase
TADH	*Thermus* Alcohol Dehydrogenase
TAE	Tris-Acetate-EDTA
TEMED	TetraMethylEthyleneDiamine
Tris	Tris(hydroxmethyl)aminomethane

References

1. Bowden, K., Heilbron, I. M., Jones, E. R. H., and Weedon, B. C. L. (1946) *J Chem Soc* **0**, 39-45
2. Sandborn, L. T. (1929) *Org Synth* **9**, 52-53
3. Schlesinger, H. I., Brown, H. C., Finholt, A. E., Gilbreath, J. R., Hoekstra, H. R., and Hyde, E. K. (1953) *J Am Chem Soc* **75**, 215-219
4. Mashima, K., Akutagawa, T., Zhang, X. Y., Takaya, H., Taketomi, T., Kumobayashi, H., and Akutagawa, S. (1992) *J Organomet Chem* **428**, 213-222
5. Ohta, T., Miyake, T., Seido, N., Kumobayashi, H., Akutagawa, S., and Takaya, H. (1992) *Tetrahedron Lett* **33**, 635-638
6. Akutagawa, S. (1997) *Top Catal* **4**, 271-274
7. Corey, E. J., and Suggs, J. W. (1975) *Tetrahedron Lett* **16**, 2647-2650
8. Ütkür, F. Ö., Thanh Tran, T., Collins, J., Brandenbusch, C., Sadowski, G., Schmid, A., and Bühler, B. (2012) *J Ind Microbiol Biotechnol* **39**, 1049-1059
9. Gong, J.-S., Lu, Z.-M., Li, H., Shi, J.-S., Zhou, Z.-M., and Xu, Z.-H. (2012) *Microb Cell Fact* **11**, 142
10. Steinreiber, J., Faber, K., and Griengl, H. (2008) *Chemistry* **14**, 8060-8072
11. Yamaguchi, S., Komeda, H., and Asano, Y. (2007) *Appl Environ Microbiol* **73**, 5370-5373
12. Yao, J., Du, H., Yan, S., Fang, F., Wang, C., Lue, L. F., Guo, L., Chen, D., Stern, D. M., Gunn Moore, F. J., Xi Chen, J., Arancio, O., and Yan, S. S. (2011) *J Neurosci* **31**, 2313-2320
13. Tan, E. K., Nagamitsu, S., Matsuura, T., Khajavi, M., Jankovic, J., Ondo, W., and Ashizawa, T. (2001) *Neurosci Lett* **305**, 70-72
14. Crabb, D. W., Matsumoto, M., Chang, D., and You, M. (2004) *Proc Nutr Soc* **63**, 49-63

15. Vidal, F., Perez, J., Morancho, J., Pinto, B., and Richart, C. (1990) *Gut* **31**, 707-711

16. Bogin, O., Peretz, M., and Burstein, Y. (1997) *Protein Sci* **6**, 450-458

17. Theorell, H., and McKinley McKee, J. S. (1961) *Nature* **192**, 47-50

18. Karabec, M., Lyskowski, A., Tauber, K. C., Steinkellner, G., Kroutil, W., Grogan, G., and Gruber, K. (2010) *Chem Comm* **46**, 6314-6316

19. de Gonzalo, G., Lavandera, I., Faber, K., and Kroutil, W. (2007) *Org Lett* **9**, 2163-2166

20. Rao, S. T., and Rossmann, M. G. (1973) *J Mol Biol* **76**, 241-256

21. Ghosh, D., Wawrzak, Z., Weeks, C. M., Duax, W. L., and Erman, M. (1994) *Structure* **2**, 629-640

22. Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalezduarte, R., Jeffery, J., and Ghosh, D. (1995) *Biochemistry* **34**, 6003-6013

23. Lavandera, I., Oberdorfer, G., Gross, J., de Wildeman, S., and Kroutil, W. (2008) *Eur J Org Chem*, 2539-2543

24. Lavandera, I., Kern, A., Ferreira-Silva, B., Glieder, A., de Wildeman, S., and Kroutil, W. (2008) *J Org Chem* **73**, 6003-6005

25. Cuetos, A., Rioz-Martínez, A., Bisogno, F. R., Grischek, B., Lavandera, I., de Gonzalo, G., Kroutil, W., and Gotor, V. (2012) *Adv Synth Catal* **354**, 1743-1749

26. Kulig, J., Simon, R. C., Rose, C. A., Husain, S. M., Haeckh, M., Luedeke, S., Zeitler, K., Kroutil, W., Pohl, M., and Rother, D. (2012) *Catal Sci Technol* **2**, 1580-1589

27. Kulig, J., Frese, A., Kroutil, W., Pohl, M., and Rother, D. (2013) *Biotechnol Bioeng,* **7**, 1838-1848

28. Lerchner, A., Jarasch, A., Meining, W., Schiefner, A., and Skerra, A. (2013) *Biotechnol Bioeng*, **11**, 2803-2814

29. Lavandera, I., Kern, A., Resch, V., Ferreira-Silva, B., Glieder, A., Fabian, W. M. F., de Wildeman, S., and Kroutil, W. (2008) *Org Lett* **10**, 2155-2158

30. Man, H., Kędziora, K., Kulig, J., Frank, A., Lavandera, I., Gotor-Fernández, V., Rother, D., Hart, S., Turkenburg, J., Grogan, G. (2013) *Top Catal,* in press

31. Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I., and Åkeson, Å. (1976) *J Mol Biol* **102**, 27-59

32. Meijers, R., Morris, R. J., Adolph, H. W., Merli, A., Lamzin, V. S., and Cedergren-Zeppezauer, E. S. (2001) *J Biol Chem* **276**, 9316-9321

33. Baker, P. J., Britton, K. L., Fisher, M., Esclapez, J., Pire, C., Bonete, M. J., Ferrer, J., and Rice, D. W. (2009) *Proc Natl Acad Sci* **106**, 779-784

34. Littlechild, J. A., Guy, J., Connelly, S., Mallett, L., Waddell, S., Rye, C. A., Line, K., and Isupov, M. (2007) *Biochem Soc Trans* **35**, 1558-1563

35. Hoellrigl, V., Hollmann, F., Kleeb, A. C., Buehler, K., and Schmid, A. (2008) *Appl Microbiol Biotech* **81**, 263-273

36. Hollmann, F., Kleeb, A., Otto, K., and Schmid, A. (2005) *Tetrahedron: Asymmetry* **16**, 3512-3519

37. Jakoblinnert, A., Mladenov, R., Paul, A., Sibilla, F., Schwaneberg, U., Ansorge-Schumacher, M. B., and de Maria, P. D. (2011) *Chem Comm* **47**, 12230-12232

38. Jakoblinnert, A., van den Wittenboer, A., Shivange, A. V., Bocola, M., Heffele, L., Ansorge-Schumacher, M., and Schwaneberg, U. (2013) *J Biotechnol* **165**, 52-62

39. Cahn, R. S., Ingold, C., and Prelog, V. (1966) *Angew Chem Int Ed* **5**, 385-415

40. Prelog, V., and Helmchen, G. (1982) *Angew Chem Int Ed* **21**, 567-583

41. Prelog, V. (1964) *Pure Appl Chem* **9**, 119-130

42. Bradshaw, C. W., Fu, H., Shen, G. J., and Wong, C. H. (1992) *J Org Chem* **57**, 1526-1532

43. Jensen, C. N., Cartwright, J., Ward, J., Hart, S., Turkenburg, J. P., Ali, S. T., Allen, M. J., and Grogan, G. (2012) *Chembiochem* **13**, 872-878

44. Machielsen, R., Looger, L. L., Raedts, J., Dijkhuizen, S., Hummel, W., Hennemann, H.-G., Daussmann, T., and van der Oost, J. (2009) *Eng Life Sci* **9**, 38-44

45. Dudek, H. M., Pazmino, D. E. T., Rodriguez, C., de Gonzalo, G., Gotor, V., and Fraaije, M. W. (2010) *Appl Microbiol Biotechnol* **88**, 1135-1143

46. Kamerbeek, N. M., Fraaije, M. W., and Janssen, D. B. (2004) *Eur J Biochem* **271**, 2107-2116

47. Chenault, H. K., and Whitesides, G. (1987) *Appl Biochem Biotechnol* **14**, 147-197

48. Fogg, M. J., and Wilkinson, A. J. (2008) *Biochem Soc Trans* **36**, 771-775

49. Stuart, D. I., Jones, E. Y., Wilson, K. S., and Daenke, S. (2006) *Acta Crystallogr Sect D-Biol Crystallogr* **62**, ii-i

50. Alzari, P. M., Berglund, H., Berrow, N. S., Blagova, E., Busso, D., Cambillau, C., Campanacci, V., Christodoulou, E., Eiler, S., Fogg, M. J., Folkers, G., Geerlof, A., Hart, D., Haouz, A., Herman, M. D., Macieira, S., Nordlund, P., Perrakis, A., Quevillon-Cheruel, S., Tarandeau, F., van Tilbeurgh, H., Unger, T., Luna-Vargas, M. P. A., Velarde, M., Willmanns, M., and Owens, R. J. (2006) *Acta Crystallogr Sect D-Biol Crystallogr* **62**, 1103-1113

51. Smisek, D. L., and Hoagland, D. A. (1989) *Macromolecules* **22**, 2270-2277

52. Hanahan, D. (1983) *J Mol Biol* **166**, 557-580

53. Bertani, G. (2004) *J Bacteriol* **186**, 595-600

54. Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D. (1988) *Nat Biotechnol* **6**, 1321-1325

55. Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I., and Perrakis, A. (2005) *Acta Crystallogr Sect D-Biol Crystallogr* **61**, 1426-1431

56. Brzozowski, A. M., and Walton, J. (2001) *J Appl Crystallogr* **34**, 97-101

57. Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N., and Svensson, O. (2009) *J Synchrotron Radiat* **16**, 872-879

58. Winter, G. (2010) *J Appl Crystallogr* **43**, 186-190

59. Kabsch, W. (2010) *Acta Crystallogr Sect D-Biol Crystallogr* **66**, 125-132

60. Patterson, A. L. (1935) *Z Kristallogr* **90**, 517-542

61. Vagin, A., and Teplyakov, A. (2010) *Acta Crystallogr Sect D-Biol Crystallogr* **66**, 22-25

62. Stein, N. (2008) *J Appl Crystallogr* **41**, 641-643

63. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) *Acta Crystallogr Sect D-Biol Crystallogr* **64**, 125-132

64. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr Sect D-Biol Crystallogr* **60**, 2126-2132

65. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr Sect D-Biol Crystallogr* **53**, 240-255

66. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) *J Comput Chem* **19**, 1639-1662

67. Schuttelkopf, A. W., and van Aalten, D. M. (2004) *Acta Crystallogr Sect D-Biol Crystallogr* **60**, 1355-1363

68. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) *Acta Crystallogr Sect D-Biol Crystallogr* **59**, 1131-1137

69. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) *J Comput Chem* **30**, 2785-2791

70. Krissinel, E., and Henrick, K. (2007) *J Mol Biol* **372**, 774-797

71. Holm, L., and Rosenström, P. (2010) *Nucleic Acids Res* **38**, 545-549

72. Man, H., Loderer, C., Ansorge-Schumacher, M., Grogan, G. (2013) *ChemCatChem*, submitted