Enzyme Engineering for the Synthesis of Industrially Important Chiral Precursors

A Thesis

Submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Sarvesh Sable

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Date: April, 2025

Under the guidance of

Supervisor: Dr. Mahesh Patil

Scientist, CSIR-National Chemical Laboratory

From May 2024 to Mar 2025

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled Enzyme Engineering for the Synthesis of Industrially Important Chiral Precursors towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Sarvesh Sanjay Sable at the CSIR-National Chemical Laboratory under the supervision of Dr. Mahesh Patil Scientist and Asst. Professor, Department of Chemical Engineering and Process Development(CEPD), during the academic year 2024-2025.

Dr. Mahesh Patil

Committee:

Dr. Mahesh Patil

Supervisor
Scientist
CSIR-National Chemical Laboratory

Sarvesh Sable
Reg No: 20191124

Dr. Parveen Goyal

Co-Supervisor
Senior Scientist
CSIR-National Chemical Laboratory

Dr. Saikrishnan Kayarat

Expert
Professor
Dept. of Biology
IISER Pune

This thesis is dedicated to my Family

Declaration

I hereby declare that the matter embodied in the report entitled “Enzyme Engineering for the Synthesis of Industrially Important Chiral Precursors” are the results of the work carried out by me at the Department of Chemical Engineering and Process Development, CSIR-National Chemical Laboratory, Pune, under the supervision of Dr. Mahesh Patil and Dr. Parveen Goyal, and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Sarvesh Sanjay Sable

Roll No: 20191124

Dr. Mahesh Patil

Scientist

CSIR-NCL

Table of Contents

List of Tables

List of Figures

Abstract

Despite advancements in drug manufacturing and production, the purity of manufactured drugs and/or chemical compounds remains a critical challenge that requires further investigation. Enzyme engineering has emerged as a powerful tool for enhancing biocatalytic efficiency, stability, and substrate specificity, addressing key challenges in industrial and pharmaceutical applications. In this study, we tried to engineer AaDH to improve its catalytic performance in the amination reaction of industrially important chiral precursors (chemical compounds that are used as a building block for complex drugs) at a specific position in the chemical structure, such that the enzyme gives a specific enantiomer (enantioselective) as the product with very high purity that can be used in the industry. We used computational techniques like structural modelling and advanced techniques like Molecular Docking and Molecular Dynamics to study the enzyme's active site and its probability of accepting our substrate of interest. The structural analysis using these computational techniques gave us critical insights into the active site of our wild-type enzyme and the mutations in the wild-type enzyme that are necessary to change the substrate scope of the enzyme. Further structural revealed that there is a list of residues that, when mutated, can increase the catalytic activity of the enzyme. Finally, we tried to clone and express the enzyme in *Escherichia coli* so that it could be tested and optimized further. We have proposed a final set of mutations that are necessary to increase the catalytic activity of AmDH.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my research supervisor Dr. Mahesh Patil, co-supervisor, Dr. Parveen Goyal and expert, Dr. Saikrishnan Kayarat, for the continuous support of my Master's thesis research and for their patience, motivation, enthusiasm, and immense knowledge.

Besides my research supervisors, I would like to thank my fellow labmates Pradhyumna Umbarkar, Dyli Likha, Aruna Chaughule, Kajal Patel, Yogesh Korde, Dinisha Joshi, Riya Sajnani for the stimulating discussions, sleepless nights we were working together before deadlines in last one year. I also thank all my batchmates and friends Shantanu, Mehul, and Aniket for supporting me constantly for one year in every situation.

I also thank the teaching and non-teaching staff at the CSIR-National Chemical Laboratory and Indian Institute of Science Education and Research for their support throughout the course. Especially the whole CDAC team managing the Param Brahma cluster.

Last but not least, I would like to thank the Almighty and my family for believing in me and supporting me throughout my life.

Contributions

Contributor name	Contributor role
Dr. Mahesh Patil, Dr. Parveen Goyal	Conceptualization Ideas
Dr. Mahesh Patil, Dr. Parveen Goyal	Methodology
Sarvesh Sable	Software
Dr. Mahesh Patil, Dr. Parveen Goyal, Sarvesh Sable	Validation
Dr. Mahesh Patil, Dr. Parveen Goyal, Sarvesh Sable	Formal analysis
Dr. Mahesh Patil, Dr. Parveen Goyal, Sarvesh Sable	Investigation
Dr. Mahesh Patil, Dr. Parveen Goyal	Resources
Dr. Mahesh Patil, Dr. Parveen Goyal, Sarvesh Sable	Data Curation
Sarvesh Sable	Writing - original draft preparation
Dr. Mahesh Patil, Dr. Parveen Goyal, Sarvesh Sable	Writing - review and editing
Sarvesh Sable	Visualization
Dr. Mahesh Patil, Dr. Parveen Goyal	Supervision
Dr. Mahesh Patil, Dr. Parveen Goyal	Project administration
Dr. Mahesh Patil, Dr. Parveen Goyal	Funding acquisition

Chapter 1 Introduction

Enzymes serve as nature's catalysts, enabling precise and efficient biochemical reactions. However, their native properties often limit their industrial applicability. At the same time, some properties like enantioselectivity are necessary from the industry's perspective, necessitating enzyme engineering to enhance stability, activity, and selectivity. From the perspective of industrial applications, enzymes can be obtained from different organisms, as the properties of enzymes obtained may depend on the environment in which they live and proliferate; if an enzyme is obtained from an organism residing in a thermal vent, any enzyme obtained from such an organism might have the thermal stability required in industrial processes.

Amines with chiral centres are essential building blocks in pharmaceuticals, agrochemicals, and fine chemicals, playing a crucial role in the synthesising active pharmaceutical ingredients (APIs) and bioactive molecules(Tseliou *et al.*). Chiral amines can only be used as a precursor for synthesising pharmaceuticals if they are highly enantiopure, which means they are pure concerning the enantiomer required in synthesising the pharmaceutical compound or drug. Obtaining such enantiopure compound for drug manufacturing in such a large quantity is a difficult task, be it chemical synthesis or purification from any native source. Often, these chiral amines are paired with a ketone or an ester group, making synthesising such compounds even more difficult. In such scenarios, asymmetric reductive amination (ARA) is employed. A Roundtable on Pharmaceuticals organised by The American Chemical Society Green Chemical Institute in 2007 ranked direct, asymmetric reductive amination between ammonia and carbonyls as one of the most highly desired transformations(Bommarius, 2019). ARA can be of two types: Conventional chemical and Biocatalytic. The Conventional chemical strategy is just like any other chemical reaction, done using standard chemicals and transition metal catalysts. On the other hand, in biocatalytic strategy, Asymmetric Reductive Amination (ARA) is a key strategy for synthesising chiral primary amines directly from ketones using ammonia as an amine donor(Wang *et al.*, 2021). This reaction is highly valuable in the pharmaceutical, agrochemical, and other chemical industries due to its ability to produce optically pure amines with high efficiency and sustainability(Mayol *et al.*, 2019). As the name suggests, a biocatalyst or an enzyme derived from an organism is used in the biocatalytic approach. The biocatalytic method gives an edge over

chemical synthesis methods for the production of chiral amines or any chiral compounds, as the enzymes used in such scenarios are particular with the transformation they carry out, unlike the chemical synthesis method, which might create a racemic mix or any byproduct that increase the purification and production cost.

- **Conventional Chemical ARA vs. Biocatalytic ARA**

Traditionally, ARA has been performed using **transition metal-based catalysts** (e.g., ruthenium, iridium, or palladium complexes). However, these methods suffer from several drawbacks:

- **Low Enantioselectivity:** Many metal-catalyzed processes require additional chiral ligands, which complicate reaction setups.(Mayol *et al.*, 2019)
- **Harsh Reaction Conditions:** Conditions involving high temperatures and hydrogen gas with high pressure are often needed, limiting substrate compatibility.(Tseliou *et al.*)
- **Environmental Concerns:** Metal residues are difficult to remove, making these methods less sustainable.(Lai *et al.*, 2024)

Enzymes used in the context of Asymmetric Reductive Amination reactions in the industry or are in the pipeline for application in industry are Imine Reductases (IREDs), Reductive Aminases (RedAms), Transaminases (TAs), or Opine Dehydrogenases (OpDHs), out of these enzymes, two class of enzymes, Imine Reductases (IREDs) and Transaminase are of greater importance to the industry.

Imine Reductases (IREDs)

Imine Reductases (IREDs) is a wide class of NADPH-dependent oxidoreductases that catalyse the enantioselective reduction of imines to chiral amines. They have emerged as valuable biocatalysts for the asymmetric synthesis of optically active amines(Mangas-Sanchez *et al.*, 2017). IREDs were first reported in 2010 following the screening of bacterial strains capable of reducing 2-methyl-1-pyrroline into (R)- or (S)-2-methylpyrrolidine(Cárdenas-Fernández *et al.*, 2023). The emergence and development of bioinformatics tools and searchable databases have led to the identification of a new and diverse range of IRED biocatalysts that have been characterised and employed in various synthetic processes(Cárdenas-Fernández *et al.*, 2023). Imine Reductases (IREDs) are NADPH-dependent enzymes that catalyse

the reduction of imines to amines with high enantioselectivity, making them highly relevant for biocatalysis. The IRED from Amycolatopsis orientalis (AoIRED) has been structurally characterised, and computational studies have provided insights into its catalytic mechanism and stereoselectivity(Prejanò *et al.*, 2022). IREDs can sometimes catalyse side reactions such as imine disproportionation or undesired secondary amine formation, making them not the most suitable candidate for industrial application.

Transaminases (TAs or ω-TAs)

Transaminases are a crucial group of enzymes that mediate the transfer of amino groups between molecules, playing a vital role in amino acid metabolism and biocatalytic applications. Among them, ω-transaminases (ω-TAs) have gained significant attention due to their ability to catalyze the stereoselective amination of ketones, making them valuable for producing chiral amines in pharmaceutical and industrial settings(Shin *et al.*, 2018). A notable ω-transaminase, derived from Vibrio fluvialis JS17 (Vfat), exhibits high specificity toward arylic chiral amines and operates with strict enantioselectivity. However, its structure and functional dynamics, particularly in the absence of its essential cofactor, pyridoxal 5′-phosphate (PLP)(Shin *et al.*, 2018). PLP being one of the essential cofactors for the ARA reaction performed by ω-transaminase, discourages its use in the industrial setting, although a large number of enzymes from this class are being used as biocatalysts in various pharmaceutical and other chemical industries, the cost associated with purification of the product persists. However, transaminase-catalysed reactions suffer from equilibrium limitations and often substrate and product inhibition(Koszelewski *et al.*, 2010).

As a greener alternative, amine dehydrogenases (AmDHs) have emerged as powerful biocatalysts that catalyze Asymmetric Reductive Amination using ammonia as the nitrogen source and NAD(P)H as a cofactor, with water as the sole byproduct. (Tseliou *et al.*). Amine Dehydrogenase, or AmDH, were explored using a metagenomics approach by Gideon Grogan's group at the University of York, UK(Mayol *et al.*, 2019). These naturally occurring enzymes demonstrated activity on alicyclic and aliphatic ketones and aldehydes(Bommarius, 2019). This class of enzyme have the potential to be used in the industrial drug manufacturing process if their activity and stability are improved. Currently, there are 42 known naturally

occurring and engineered AmDH enzymes(Tseliou *et al.*). AmDHs have an advantage over other known enzymes in the fact that they require only Ammonia(NH_3) as the amine donor instead of a primary or secondary amine as the amine donor, this reduces the process and production cost attracting industries for the uptake of AmDH in their process development for manufacturing pharmaceutical compounds.

Two Types of AmDH

Native Amine Dehydrogenases:

Through bioinformatic screening and enzyme discovery, a novel family of native AmDHs (nat-AmDHs) was identified. These enzymes exhibited significant catalytic activity toward aliphatic and cyclic ketones, expanding the biocatalytic toolbox available for chiral amine synthesis(Mayol *et al.*, 2019). Native Amine Dehydrogenases are naturally occurring AmDHs. There are 6 known naturally occurring AmDHs, with 38-90% sequence identity suggesting that AmDHs are evolutionarily conserved(Bommarius, 2019). All six Nat-AmDHs are promiscuous with respect to nicotinamide cofactors NADH or NADPH(Bommarius, 2019). The structure of four Nat-AmDHs from *Petrotoga mobili*s, *Mycobacterium smegmatis*, *Cystobacter fuscus* and *Neosartorya fumigata* have been solved. Enzyme kinetic analysis data revealed a strong preference for cycloalkanones and linear aliphatic ketones, particularly cyclohexanone and isobutyraldehyde, with activity up to 851.3 mU mg^{-1}(Mayol *et al.*, 2019). These enzymes are yet to be used in the industrial setting as the required turnover rate for an enzyme in an industrial setting is very high compared to the actual turnover rate of these enzymes. Native AmDH can be R-selective or S-selective.

Engineered Amine Dehydrogenases:

Given their potential in pharmaceutical and industrial applications, extensive enzyme engineering efforts have been directed at enhancing their substrate scope, catalytic efficiency, and stability(Ye *et al.*, 2015). Early AmDHs were derived from amino acid dehydrogenases, such as leucine and phenylalanine dehydrogenases, through targeted mutagenesis to alter their substrate specificity(Bommarius *et al.*, 2014a). However, these initial variants exhibited low activity and narrow substrate acceptance, necessitating further optimization(Wu *et al.*, 2023). Protein engineering

strategies such as rational design, directed evolution, and domain shuffling have been employed to address these limitations. For example, chimeric AmDHs, created by combining substrate-binding and cofactor-binding domains from different parent enzymes, have shown improved activity toward structurally diverse ketones(Bommarius *et al.*, 2014b). Similarly, site-directed mutagenesis has been used to reshape the active site, increasing enzyme efficiency and allowing for the aminated conversion of bulky and heterocyclic ketones(Wu *et al.*, 2023). Additionally, cascade reactions integrating engineered AmDHs with chemical catalysts have demonstrated enhanced productivity and enantioselectivity (>99% ee), enabling the synthesis of novel non-α-amino acids(Lai *et al.*, 2024).

Further advancements include the evolution of thermostable AmDHs, which can withstand industrial reaction conditions while maintaining high activity. Recent studies on AmDHs from *Bacillus badius* and *Caldalkalibacillus thermarum* have provided insights into the relationship between enzyme structure and thermostability, leading to variants with increased catalytic turnover rates and higher substrate load compatibility(Pushpanath *et al.*, 2017). In this study we have attempted to do the same. These innovations in AmDH engineering are paving the way for efficient, scalable, and sustainable biocatalytic routes to chiral amine production, offering significant advantages over traditional chemical methods(Bommarius *et al.*, 2014a).

Amino Acid Dehydrogenase:

Amino acid dehydrogenases (AADHs) are a class of enzymes that catalyse the reversible reductive amination of α-keto or keto acids with ammonia, utilising NADH or NADPH as cofactors. Among these, enzymes such as glutamate dehydrogenase, leucine dehydrogenase, and phenylalanine dehydrogenase have been extensively studied due to their high enantioselectivity and atom economy, making them valuable in chiral compound synthesis(Zhou *et al.*, 2022). Mutations in key residues within the active site, entrance tunnel, and hinge regions have been shown to enhance activity toward a broader range of aliphatic, aromatic, and cyclic substrates(Zhou *et al.*, 2022). A notable breakthrough in AADH engineering has been the conversion of leucine dehydrogenase (LeuDH) and phenylalanine dehydrogenase (PheDH) into amine dehydrogenases (AmDHs), enabling them to catalyse direct reductive amination of ketones instead of α-keto acids. This transformation has significantly expanded their utility, particularly for the synthesis of high-value pharmaceutical

intermediates such as (R)-amphetamine and (S)-2-amino-1-butanol(Zhou *et al.*, 2022).

In this work, we attempted to engineer the Amino Acid Dehydrogenase to Amine Dehydrogenase so that they catalyse the transformation of non-alpha keto esters to non-alpha amino acids. These non-alpha amino acids are a group of chiral amines that are of great importance in the pharmaceutical industry, as these compounds, just like the class chiral amines, are used as intermediates in the synthesis of many important classes of drugs that are used in the treatment of Diabetes. Drugs like Sitagliptin that are used for the treatment of Type-2 Diabetes require a long and multistep process of manufacturing as it is a large molecule with no precursor such that it can be synthesised in one step often this takes a longer duration of synthesis per cycle, increasing the overall cost of the drug. Sitagliptin is a DPP-4 inhibitor used for the treatment of Type 2 diabetes and is a chiral β-amino acid derivative with significant pharmaceutical importance. Due to its widespread use and high commercial demand, developing an efficient, cost-effective, and sustainable method for its synthesis has been a key focus in the pharmaceutical industry(Ye *et al.*, 2021). A study carried out by Ye et al. focused on the synthesis of Sitagliptin Phosphate and explored two distinct methodologies, both are chemical synthesis, chemical resolution and asymmetric hydrogenation. The chemical resolution approach involves a five-step synthesis from commercially available precursors, using sodium borohydride ($NaBH_4$) for enamine reduction followed by diastereomeric resolution with (−)-di-p-toluoyl-L-tartaric acid. The asymmetric hydrogenation route employs a rhodium-based chiral catalyst to stereoselectively hydrogenate a β-ketomide intermediate, forming sitagliptin with high enantiomeric excess. While this approach offers high selectivity and efficiency, the reliance on precious metal catalysts and high-pressure hydrogenation equipment presents economic and operational challenges(Ye *et al.*, 2021). Starting from ethyl 3-oxo-4-(2,4,5-trifluorophenyl)butanoate, the process follows a 12 step to reach the final product, Sitagliptin ((R)-3-amino-1-(3-(trifluoromethyl)-5,6-dihydro-[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl)-4-(2,4,5-trifluorophenyl)butan-1-one).

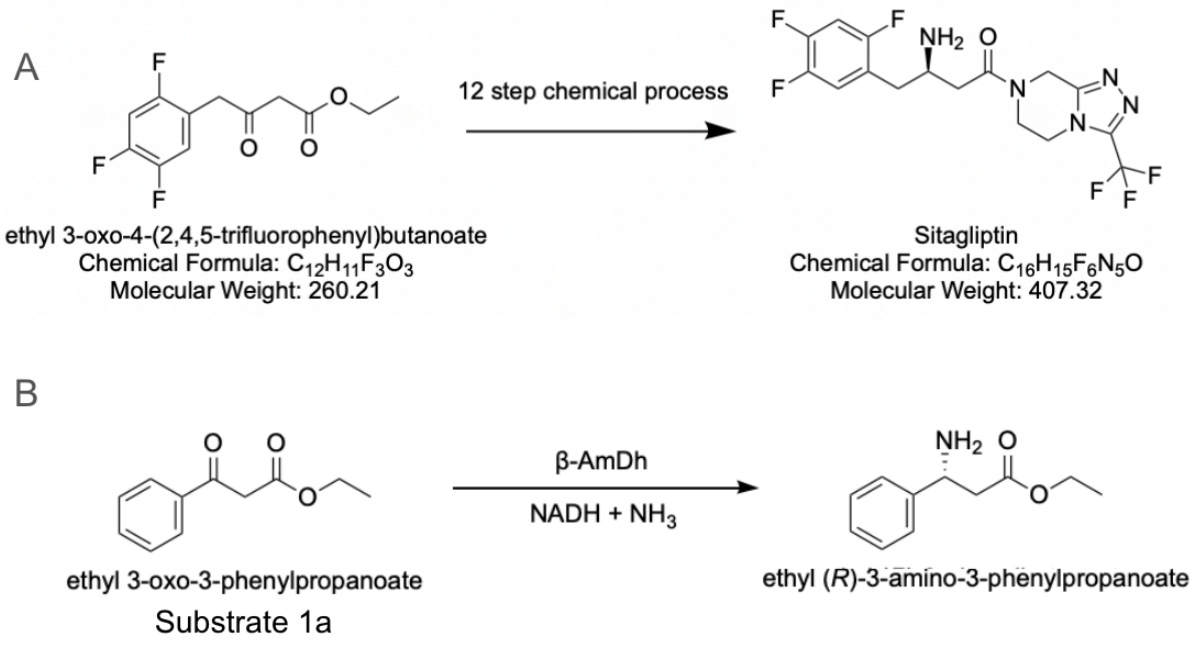

Figure 1: (A) A schematic representation of the starting chemical structure and final Sitaglyptin structure after 12 chemical synthesis steps. (B) Scheme of our intended reaction by converting AADH1 or AmDH1 to β-AmDH enzyme. This step can be used as the initial setp for drug synthesis.

In this work, we started out with an almost similar chemical structure as the initial step to engineer an Amino Acid Dehydrogenase(AADH1) and an Amine Dehydrogenase(AmDH1) using rational design and Computational Approach(in silico design) methods of enzyme engineering. The AmDH1 is already an engineered AmDH and is active towards benzylic ketones, this enzyme also had some activity toward acetophenone. On the other hand, AADH1 that we chose had some activity towards larger substrates, and had a thermostable structure. We have followed a rational design approach for engineering these enzymes, notably structural analysis and in silico mutational studies. The structure of both the enzymes mentioned in this study is not solved by any experimental means. It is to be noted that in this study, we are trying to study the reverse reaction of AADH, the amination step, instead of the forward reaction, the deamination of the amino acid step. AADH1 had already been engineered in other studies for the amination of alpha ketones but not towards non-alpha keto acids. We have also incorporated substrate engineering, along with enzyme engineering, to solve the issue of stability regarding the substrate, as most non-alpha ketones are not stable at room temperature, it was decided that the keto acids of the non-alpha ketones to be used instead of just the ketones itself. A similar approach of substrate had already been applied to engineering another AmDH,

FBb-AmDH, by Zhoufan et al. Employing ester groups to shield carboxyl-induced polarity has proved to be a strategy in substrate engineering, necessitating the search for enzymes capable of catalysing keto-ester reductive amination(Lai *et al.*, 2024).

Conserved Structure of AADHs and AmDHs:

X-ray crystallographic studies of phenylalanine dehydrogenase (PheDH) from *Rhodococcus M sp4*. and leucine dehydrogenase (LeuDH) from *Bacillus badius* species have provided structural insights into the key residues involved in substrate recognition and catalysis(Brunhuber *et al.*, 2000). Amine dehydrogenases (AmDHs) and amino acid dehydrogenases (AADHs) share a conserved structural architecture, reflecting their evolutionary relationship as members of the NAD(P)-dependent oxidoreductase family(Vanhooke *et al.*, 1999). These enzymes are typically composed of two distinct domains: a C-terminal Rossmann fold domain responsible for binding the NAD(P)H cofactor and an N-terminal substrate-binding domain, which governs substrate specificity and catalytic activity(Yousefi *et al.*, 2017). The active site is present within a cavity-like structure between these two domains, facilitating the efficient transfer of a hydride ion during the reductive amination or oxidative deamination processes(Brunhuber *et al.*, 2000). These conserved structural features provide a foundation for enzyme engineering efforts, allowing researchers to tailor AmDHs and AADHs for broader industrial applications. By modifying active site residues, cofactor interactions, and domain flexibility, scientists have successfully developed engineered AmDHs capable of catalyzing asymmetric reductive amination of diverse ketones, expanding their utility in biocatalysis and pharmaceutical synthesis(Brunhuber *et al.*, 2000). Both the AADH and AmDH family of enzymes have two conformations, one open and inactive and the other closed and active, a study on AmDH from *Petratoga mobilis* revealed two conformations of the enzyme(Mayol *et al.*, 2019).

Following a rational approach, we have used different structural prediction techniques for predicting the structures of AmDH1 and AADH1, as none of these enzymes has an already predicted structure. Followed by docking simulation studies to get an idea of the mutations that needed to be done in order to increase the activity of these enzymes. A more advanced structural bioinformatic approach was employed, including Molecular Dynamic Simulations to study the dynamic nature of the interactions between the substrate, the enzyme and the co-factor NADH. In this

attempt, we have successfully figured out the necessary mutations to both the enzymes AADH1 and AmDH1 in order to increase their activity toward non-alpha keto acids for the amination reaction.

Based on the docking simulations, we decided to perform the Molecular Dynamics simulations on the mutations that have a positive outcome from the Docking simulations. After this sequential screening, we are able to decide the necessary mutations to change the substrate scope of the AADH1 and AmDH1. Parallelly, to test our hypothesis, we have initiated the cloning of both AmDH1 and AADH1 genes in the Pet28 vector and tested their expression and if the cloned enzymes can be purified using His-tag affinity chromatography for using them in the reaction as a biocatalyst. We are successful in establishing *E. coli* cell lines with cloned AADH1 and AmDH1 enzymes in BL21-DE3 cells for expression and DH5α cells for cloning.

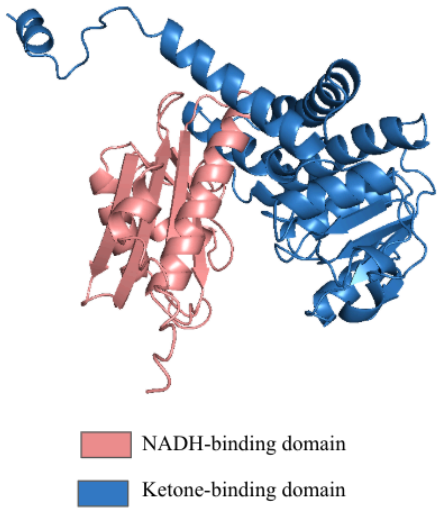

Figure 2: Two Distinct domains of AmDH and AADH class of enzymes

The AADH1 is of greater importance as it is derived from a bacterial species that resides and proliferates in thermal vents, and it has already been shown that this enzyme has a special property of thermostability at temperatures that most chemical process development industries are able to cater to even as a biocatalyst(Pushpanath *et al.*, 2017). The AmDH1 is an already engineered enzyme with two distinct domains from two Amino Acid Dehydrogenase, one Leucine Dehydrogenase and the other Phenylalanine dehydrogenase, giving a substrate acceptance towards both aliphatic ketones and aromatic ketones, thus giving an increased substrate scope(Bommarius *et al.*, 2014b). Since the structure of the AmDH and AADH family of enzymes is conserved, we thought of extrapolating the

mutations that we have found on these two enzymes to other phenylalanine dehydrogenases and leucine dehydrogenases.

Chapter 2 **Materials and Methods**

Structure Prediction using AlphaFold2:

To predict the structure of our AADH1 and AmDH1 models, Alphafold 2 was used which is based on the Google collaboratory server(AlphaFold2.ipynb - Colab). Initially, structure prediction was without a template, and later, it was done using a template of PDI Id: 1C1D. The structures were then validated by plotting a Ramachandran plot in Chimera and by uploading the structure on the SAVES server. The Multiple Sequence Alignment mode(msa_mode) was set as mmseqs2_uniref_env, and the pairing mode was set as unpaired_paired, and the num_recycles parameter was set at 3. This server predicted a total of five structures, starting from rank 1 to 5 for each wild type and its mutant. All five structures were analysed based on their IDDT score for the residues surrounding the active site K80 for AADH1 and K89 for AmDH1 and the overall pLDDT of the structure. The best structure was considered for further analysis.

Structure Prediction using AF-Cluster:

To predict a closed conformation of both the enzymes AADH1 and AmDH1, we used the AF-Cluster Google notebook available at (https://colab.research.google.com/github/HWaymentSteele/AF_Cluster/blob/main/AF_cluster_in_colabdesign.ipynb). All the parameters for AlphFold2 were same as earlier, only the custom MSA options were used, with a sequence convergence of 90 and sequence identity of 90 and query identity was set at 20, this gave us the total number of 94 clusters, these values were obtained after optimisation of the number of clusters that were given as output. Each cluster had a minimum of 3 sequences in it. No template was provided during these predictions. Only the rank 1 models predicted from the Alphafold were used for analysis. Along with this, some changes were made in order to allow the analysis part to accept 6G1M chains A and B as the comparing standards. all the clusters generated were divided into a set of 30-35 clusters each, so that the prediction can be run without disconnecting to the server due to limited runtime. Each set took around 40 minutes to complete the projections and analyse the structures. The analysis part gave a comparative graph with the TM scores on the x and y-axis. All the shortlisted structures were validated using the SAVES server for their accuracy.

Homology Modeling and Molecular Docking Using Swiss-Model Server

To investigate structural variations in the enzyme mutants with respect to our intended substrates, homology models were generated using the SWISS-MODEL server (https://swissmodel.expasy.org/). The template structure was selected based on the crystal structure of RS-AmDH, the PDB-Id 1C1D chain B, which provided a reliable framework for modelling mutant structures. To ensure accurate structural representation, the modelled mutant enzymes were aligned with the experimentally resolved homologous complex structure (PDB ID: 1C1D).

For docking studies, the three-dimensional (3D) structures of ligands were drawn on MarvinSketch, which gave us the 3D conformation of the ligand structure, followed by energy minimization and addition of Hydrogen atoms or protnation was done in the AutoDock Vina (Eberhardt *et al.*, 2021) within the Dockingpie 2.0 (Rosignoli and Paiardini, 2022) software. Protein-ligand docking simulations were performed using AutoDock Vina within the DockingPie v (Rosignoli and Paiardini, 2022). Since the experimental conditions involved a pH of 9.5, all the protein structures were protonated accordingly using the H++ server (http://biophysics.cs.vt.edu/H++)(Anandakrishnan *et al.*, 2012) (Myers *et al.*, 2006) (Gordon *et al.*, 2005). All the docking experiments were replicated three times to check for any errors and uncertainties. The catalytic triad residues (K77, K89, N270) were designated as flexible residues to allow conformational adjustments during docking. The active site cavity of AmDH1 and AADH1 and its mutants was enclosed within a 40 × 40 × 40 Å cuboid grid box, ensuring a comprehensive search space for optimal substrate-binding modes. The DockingPie software gave us the binding energies for all the binding poses, and the lowest-energy complex was selected for further structural analysis and molecular dynamics (MD) simulations.

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were conducted using GROMACS 2019.6 software to analyse the enzyme-ligand complex's structural stability and dynamic behaviour(Abraham *et al.*, 2015). The CHARMM force field was applied to model the protein system, ensuring an accurate representation of atomic interactions. The protein structures were solvated in a cubic simulation box filled with TIP3P water molecules, with an additional 10 Å buffer region to prevent edge effects. To maintain

the charge neutral, sodium ions were introduced as counterions, balancing the system electrostatically. The parameters for substrate 1a, the enzyme's substrate, were retrieved from the CgenFF force field generator and incorporated into the system. Once parameterization was completed, energy minimisation was performed to eliminate steric clashes and ensure a stable initial conformation. The system was then gradually heated from 0 K to 303 K using an NVT ensemble over a period of 50 ps, maintaining a constant number of particles, volume, and temperature. Subsequently, an NPT equilibration phase was carried out for another 50 ps, where Langevin thermostat dynamics with a collision frequency of 2 ps^{-1} and Parrinello-Rahman barostat ensured the system reached equilibrium at 1 atm pressure. For the production phase, an all-bonds restrained 100 ns MD simulation was performed under NPT conditions, allowing natural protein fluctuations to be observed. System stability was evaluated using root-mean-square fluctuations (RMSF) analysis. The simulation trajectory was recorded at intervals of 5000 steps, with an integration time step of 2 fs, and the resulting data was processed using a standard trajectory analysis protocol. The production run was performed using the CDAC ParamBrahma High-Performance Computer or Cluster Computer facility established at IISER Pune.

Experimental

Cloning of AADH1 and AmDH1 in Pet28

The DNA fragments containing 6XHis-tag of AADH1 and AmDH1 were ordered from Eurofins pvt. ltd. The choice for the site for cloning was between the restriction enzyme sites for NdeI and XhoI. The restriction enzymes that were used are from Thermo Scientific Pvt. Ltd. For the restriction digestion, around 700ng of DNA for both the vector and the gene of interest, AADH1 and AmDH1, were taken according to their respective DNA concentration in a 1.5ml Eppendorf. To it the enzyme's compatible buffer, 3µl Buffer O was added, and 3µl NdeI restriction enzyme was added, and the total reaction mixture was made up to 30µl using MilliQ water. This reaction mixture was incubated at 37°C for 16 hours or overnight. After this step, a PCR preparation was performed on the reaction mix, to remove all the salts and buffers in it. The second restriction enzyme XhoI was used for the second restriction reaction, after this step, In this step, the DNA from the first restriction reaction was taken in the appropriate volume. To it, the enzyme's compatible buffer, 5µl Buffer R was added, and 5µl XhoI restriction enzyme was added, and the total reaction

mixture was made up to 50µl using DNAse, RNAse free MilliQ water. This reaction mixture was incubated at 37℃ for 16 hours or overnight. In the next experiment for ligating the fragments, the vector is to the fragment of AmDH1 and AADH1, respectively, with a molar ratio of 1:3, and to this mix, 0.5µl T4 DNA ligase along with 2µl T4 DNA ligase buffer was added, and ATP as a cofactor was added for the reaction to start. The reaction mixture was incubated at 20℃ for two hours, and the ligation mixture was directly transformed into competent DH5α cells.

Preparing competent DH5 α cells

The calcium chloride technique was used to create the DH5α competence cell. The CaCl2 stock solution, 10X, was prepared by adding 11.1 g of anhydrous CaCl2 by dissolving it in 100 mL of distilled water. A 0.22µm syringe filter was used to sterilize the stock solution. The 0.1 M CaCl2 wash solution was created by diluting it in a 1:10 ratio. 9 mL of sterile glycerol, 45 mL of dH2O, and 6 mL of 1M CaCl2 were combined to create another working solution. On day one, DH5α competence was inoculated in a conical flask containing freshly prepared, sterilized 1 mL LB broth. The flask was then incubated for 12 to 16 hours at 37°C and 200 rpm on an orbital shaker. The DH5α cell is subcultured on day 2 by mixing 1 mL of the overnight culture with 99 mL of fresh LB broth. (Ensure that no antibiotics are added to the culture.) Again placed in the orbital shaker at 37°C and 200 rpm until OD reaches 0.4 absorbance. Optical density was measured by the double-beam spectrophotometer. Once the O.D. reached 0.4, the culture was separated into multiple centrifuge tubes, placed on the ice for 20 minutes, and centrifuged at 4 °C at 4000 rpm for 10 minutes. After centrifugation, the supernatant was discarded, and the pellet was resuspended with 20 mL of ice-cold 0.1 M CaCl2. Incubate it on the ice for 30 minutes. Again centrifuged it at 4°C at 4000 rpm for 10 minutes. The supernatant was discarded, and the pellet was resuspended in 5 mL ice-cold 0.1 M CaCl2 with 15% glycerol. Some samples were stored in a -80°C freezer, and others were used for downstream transformation.

Transformation of AmDH1 and AADH1 in DH5α:

The freshly prepared DH5α competence cell was thawed on ice. The AmDH1 and AADH1 plasmid (5µL) was added to the DH5α competence cell. This mixture was incubated on ice for 30 minutes. Heat shock treatment was provided by placing it in a 42°C water bath for exactly 45 seconds. Again, place it on ice for 5 minutes. The

pre-warmed 1 mL LB broth was added to it, and it was incubated at 37 °C, 200 rpm, for 1 hour for outgrowth. The inoculum was spread on the LB agar plate containing Kanamycin antibiotics. The agar plate was incubated at 37°C for 12–16 hours. Isolated colonies were observed. An isolated colony from the transformed plate was chosen. The loop was sterilized in the flame, and after cooling the loop gently, an E. coli colony was touched, and streaking was performed on the LB agar plate. The plate was covered and incubated at 37°C for 12 to 16 hours. Isolated colonies were observed.

Primary cell culture:

A colony was picked from a freshly streaked culture plate and used as a starter culture to inoculate 10 mL of LB broth. 50 micrograms of kanamycin (50 mg/mL) were added to the LB media. The culture was incubated at 37 °C for 16 hours in an orbital shaking incubator.

Plasmid isolation:

Tye Hi-Media Plasmid prep kit was used to isolate the plasmid. The overnight-grown culture of about 2 mL was filled in the microcentrifuge tube, bacterial cell harvesting was performed by centrifugation at 13000 rpm for 1 min, and the supernatant was discarded. Cell resuspension was performed by resuspending the bacterial pellet in 250 µl of Resuspension Solution (HP1) (DS0020) and mixing well with gentle pipetting till no clumps were visible. RNase was added to Resuspension Solution (HP1) (DS0020). Then, cells were lysed by adding 250 µL of Lysis Solution (HP2) (DS0021) mixed thoroughly by gently inverting the tube 4-6 times. This lysis reaction was performed for 5 minutes. After lysis, neutralize the cell by adding 350 µl of Neutralization Solution (HN3) (DS0022) and mix it by gently inverting the tube 4-6 times. After neutralization, the solution becomes cloudy. Centrifugation was performed at 13000 rpm for 10 min to get a compact white pellet. Lysate was loaded on HiElute Miniprep Spin Column (capped) [DBCA03] and centrifuged at 13000 rpm for a minute, and flow through was discarded. The first wash was performed by adding 500 µl of Wash Solution (HPB) (DS0032), and centrifugation was performed at 13000 rpm for a minute. Flow through was discarded, and a second wash was performed by adding 700 µL of diluted Wash Solution (HPE) (DS0024) and again centrifuged at 13000 rpm for a minute. Flow through was discarded, and an empty column was spun to remove any traces of the wash solution. DNA elution was

performed by transferring the column to a clean 2.0 mL uncapped collection tube, and 50 µl of the elution buffer (ET) (DS0040) was added. It was allowed to stand for 1 minute and then centrifuged at 13000 rpm. The elute was transferred to a fresh, uncapped 2 mL collection tube for long storage, and it was stored at -80 °C. A NanoDrop spectrophotometer was used to measure the concentration of plasmid DNA. The upper and lower surfaces of the optical surface were cleaned with the help of nuclease-free water. To initialize the spectrophotometer, nuclease-free water was placed on the surface, and the start button was pressed. Blank measurement was performed by elution buffer (ET) (DS0040). After the blank, the plasmid sample (2µL) was loaded, and the plasmid concentration was measured. Plasmid concentration was recorded, and A260/A280 was also recorded.

PCR:

In a PCR tube of 200 µl, the following items are added:

1. 38 µl sterile water.
2. 0.2 µl of forward primer (10 µM).
3. 0.2 µl of reverse primer (10 µM).
4. 1 µl of dNTPs (50 µM).
5. 0.5 µl Phusion DNA polymerase
6. 1 µl of DNA template (100 ng/µL).

This mixture was pipetted out gently for good homogenization. PCR tubes spin on the centrifuge. And tubes were inserted in the thermocycler. The thermocycler step was set as per the steps of DNA denaturation, annealing, and extension.

PCR purification:

For the purification of PCR products, the G2P Spin Clean PCR Cleanup Kit was used, which employs spin-column technology and a silica membrane to selectively bind DNA while removing contaminants. The process began by mixing the PCR product with an equal volume of Buffer PCA, facilitating DNA binding. The sample was then loaded onto a spin column and centrifuged at 10,000 × g, allowing the DNA to adhere to the membrane while unwanted impurities were discarded. Following this, Buffer D (Wash Buffer) was applied, and the column was subjected to multiple wash and spin cycles to ensure complete removal of residual contaminants. After drying the membrane with an additional spin, the purified DNA was eluted using Buffer E (Elution Buffer) or nuclease-free water (pH 8.0). The eluted DNA was

collected in a fresh microcentrifuge tube and stored for downstream applications such as sequencing.

Gel Electrophoresis:

A 0.8% agarose gel was prepared in an Erlenmeyer flask. TAE buffer (40 mM Tris-acetate, 1 mM EDTA) and TBE (45 mM Tris-borate, 1 mM EDTA) were added to the agarose gel. The mixture was melted by heating in the microwave. 2 µL SYBR Safe DNA Gel stain was added to the mixture. A gel tray was placed in the casting apparatus, and a comb was placed in the apparatus so as to create the appropriate well. Molten agarose gel was poured into the gel mould. The gel was kept for solidification. Once the gel was set, the comb was removed, and the gel was placed in the gel box. Samples were prepared by adding the loading dye 6X concentration (0.25% bromophenol blue, 0.25% xylene cyanol, and 30% glycerol) and filling the gel box with the running buffer. DNA ladder and samples were loaded into the gel, and an electric supply was provided at about 100 V. Once the electrophoresis was completed, DNA bands were observed in the BioRad gel dock apparatus.

Sanger sequencing:

The plasmid samples were amplified with the help of PCR. These PCR samples were purified with the help of BDX terminator purification. Samples were loaded on 96-well plates with the forward and reverse primers in two different wells. Sanger sequencing was performed by the capillary electrophoresis Sanger sequencing method. Hitachi 3500xl genetic analyzer was used for Sanger sequencing.

AmDH1 Expression and Purification

Preparing competent BL21(DE3) cells

The calcium chloride technique was used to create the *E. Coli* BL21(DE3) competence cell. The CaCl2 stock solution, 10X, was prepared by adding 11.1 g of anhydrous CaCl2 by dissolving it in 100 mL of distilled water. A 0.22µm syringe filter was used to sterilize the stock solution. The 0.1 M CaCl2 wash solution was created by diluting it in a 1:10 ratio. 9 mL of sterile glycerol, 45 mL of dH2O, and 6 mL of 1M CaCl2 were combined to create another working solution.

On the first day, *E. Coli* BL21(DE3) competence was inoculated in a conical flask containing freshly prepared, sterilized 1 mL LB broth. The flask was then incubated for 12 to 16 hours at 37°C and 200 rpm on an orbital shaker. The *E. Coli* BL21(DE3)

cell is subcultured on day 2 by mixing 1 mL of the overnight culture to 99 mL of fresh LB broth. (Ensure that no antibiotics are added to the culture.) Again, place in the orbital shaker incubator at 37°C and 200 rpm until OD reaches 0.4 absorbance. Optical density was measured by the double-beam spectrophotometer. Once the O.D. reached 0.4, the culture was separated into multiple centrifuge tubes, placed on the ice for 20 minutes, and centrifuged at 4 °C at 4000 rpm for 10 minutes. After centrifugation, the supernatant was discarded, and the pellet was resuspended with 20 mL of ice-cold 0.1 M CaCl2. Incubate it on the ice for 30 minutes.

Again centrifuged it at 4°C at 4000 rpm for 10 minutes. The supernatant was discarded, and the pellet was resuspended in 5 mL ice-cold 0.1 M CaCl2 with 15% glycerol. Some samples were stored in a -80°C freezer, and others were used for downstream transformation.

Transformation of AmDH1 and AmDH1 for expression:

The freshly prepared E. Coli BL21(DE3) competence cell was thawed on ice. The AmDH1 plasmid (5µL) was added to the E. Coli BL21(DE3) competence cell. This mixture was incubated on ice for 30 minutes. Heat shock treatment was provided by placing it in a 42°C water bath for exactly 45 seconds. Again, place it on ice for 5 minutes. The pre-warmed 1 mL LB broth was added to it, and it was incubated at 37 °C, 200 rpm, for 1 hour for outgrowth. The inoculum was spread on the LB agar plate containing kanamycin antibiotics. The agar plate was incubated at 37°C for 12–16 hours. Isolated colonies were observed. An isolated colony from the transformed plate was chosen. The loop was sterilized in the flame, and after cooling the loop gently, an E. coli colony was touched, and streaking was performed on the LB agar plate. The plate was covered and incubated at 37°C for 12 to 16 hours. Isolated colonies were observed. A colony was picked from a freshly streaked culture plate and used as a starter culture to inoculate 10 mL of LB broth. 50 micrograms of kanamycin (50 mg/mL) were added to the LB media. The culture was incubated at 37 °C for 16 hours in an orbital shaking incubator.

Secondary cell culture:

In an autoclaved flask, 490 mL LB medium was inoculated with 10 mL of freshly cultured primary cells. 250 µL of Kanamaycin (50 mg/mL) was added to the LB media. The culture was incubated at 37 °C for 3-4 hours in an orbital shaking incubator until the O.D. reached 0.4. Then IPTG (isopropyl

β-D-1-thiogalactopyranoside) was added, and the temperature of the orbital shaker was decreased to 20 °C. Incubate it for 16–18 hours.

Affinity chromatography:

The cell mass was separated from the liquid media by centrifugation at 13000 rpm at four °C for 15 minutes, and finally, we got a cell pellet. After centrifugation, the cell pellet was resuspended in buffer A (50 mM potassium phosphate, 100 mM NaCl, and 20 mM imidazole; pH 7.5), followed by an ultrasonic cell disruption system and followed by centrifugation at 13000 rpm, 4 °C, for 30 minutes to remove the cell debris.

Later, it passes through a 0.22 µm filter with a load onto a Ni NTA affinity column, pre-equilibrate with buffer A(50 mM potassium phosphate, 100 mM NaCl, and 20 mM imidazole; pH 7.5), and then elute with buffer B (50 mM potassium phosphate, 100 mM NaCl, and 200 mM imidazole; pH 7.5). For purification, initially, after loading the protein into the column, a standard gradient protocol was employed from the highest imidazole concentration of 500mM, which gave us the elution concentration of imidazole for AADH1 and AmDH1. The purity of the eluted portion was determined by SDS-PAGE.

SDS-PAGE:

Separation gel was prepared by mixing Separating Gel Buffer (1.5 M Tris-HCl pH 8.8) with 0.4% SDS, Acrylamide 30% solution, Ammonium Persulfate (APS) (20 mg/ml), and TEMED. Pippet gel was cast; ensure its height was 1 cm below the gel comb. After adding the gel, fill the remaining cast with isopropanol. Prepared the staking gel by making one change in the buffer, which is 0.5 M Tris-HCl pH 6.8 with 0.4% SDS, and a comb was inserted into it, allowing it to harden for 45 minutes. Sample buffer was prepared by 10 w/V SDS, 10 mM beta-mercaptoethanol, 20% v/v Glycerol, 0.2 M Tris-HCl, pH 6.8, and 0.05% bromophenol blue. A sample was prepared by adding 20 µL of protein and 5 µL of sample buffer. The gel was mounted into the tank, the comb was removed, and the chamber was filled with running buffer. 2µL of ladder and 10µL of sample. The power supply was connected and ran at 100 V for 60 min until the blue line almost reached the bottom of the gel. Once the electrophoresis is done, place the gel into the staining solution containing 0.1% coomassie blue in 40% ethanol and 10% acetic acid. After staining, the gel was put in a destaining solution containing 10% ethanol and 7.5% acetic acid. After destaining, a band was observed in the gel dock apparatus.

Chapter 3 Results

Structure prediction of AmDH1 and AADH1 Using AlphaFold 2:

Enzyme engineering involves modifying enzymes to improve their stability, activity, specificity, or selectivity for various industrial and biotechnological applications. The main strategies used in enzyme engineering can be broadly classified into **rational design**, **directed evolution**, and **computational approaches**. But for any enzyme to be engineered for our benefit, it must have its exact protein structure solved using structural biology techniques or any other bioinformatic technique, as the rational design approach involves substituting an amino acid residue within the active site of the enzyme, an accurate structure of that enzyme is necessary for such approach(Kim *et al.*, 2024b).

Prediction of enzyme structure using AlphaFold Colab Server:

Since the structure of both enzymes is not solved using conventional structural biology techniques, the Alphafold Colab server, which uses the AlphaFold2 algorithm to predict protein structures, is open-accessible (Kim *et al.*, 2024a). The structure for both enzymes was predicted with very high confidence (> 90) by the AlphaFold Colab server.

The Multiple Sequence Alignment (MSA) coverage plot generated from the AlphaFold Colab server provides insights into the sequence conservation and diversity within homologous sequences of the query protein. The x-axis represents the sequence positions, while the y-axis denotes the number of homologous sequences aligned at each position. The colour gradient, ranging from red (low sequence identity) to green/blue (high sequence identity), indicates the degree of conservation across different regions of the protein. From the MSA plot, it is evident that. Core regions (middle of the sequence) exhibit high sequence identity (green/blue regions), and Terminal regions (beginning and end) show lower conservation (red/yellow regions), indicating possible flexibility or variability, which might be due to differences in evolutionary adaptations across homologs. for both the enzymes AADH1 and AmDH1, the confidence metrics and MSA plots, along with the IDDT plots, gave very high confidence in the predicted structures, with a confidence level greater than 90, indicating that the structure can be used for docking analysis(Scardino *et al.*, 2023). The Figure 3 and Figure 4 describe the

confidence PLDDT along with the generated Predicted Aligned Error plots of all 5 structures from rank 1 to 5 for AmDH1. In Figure 3 (A), the plot demonstrates the Multiple sequence alignment of the sequences that were considered by AlphaFold 2 for predicting the structure. The black line shows how conserved each amino acid is in all the sequences. Figure 3(C) and 4(C) have lDDT plots for all 5 structures predicted for AADH1 and AmDH1, respectively.

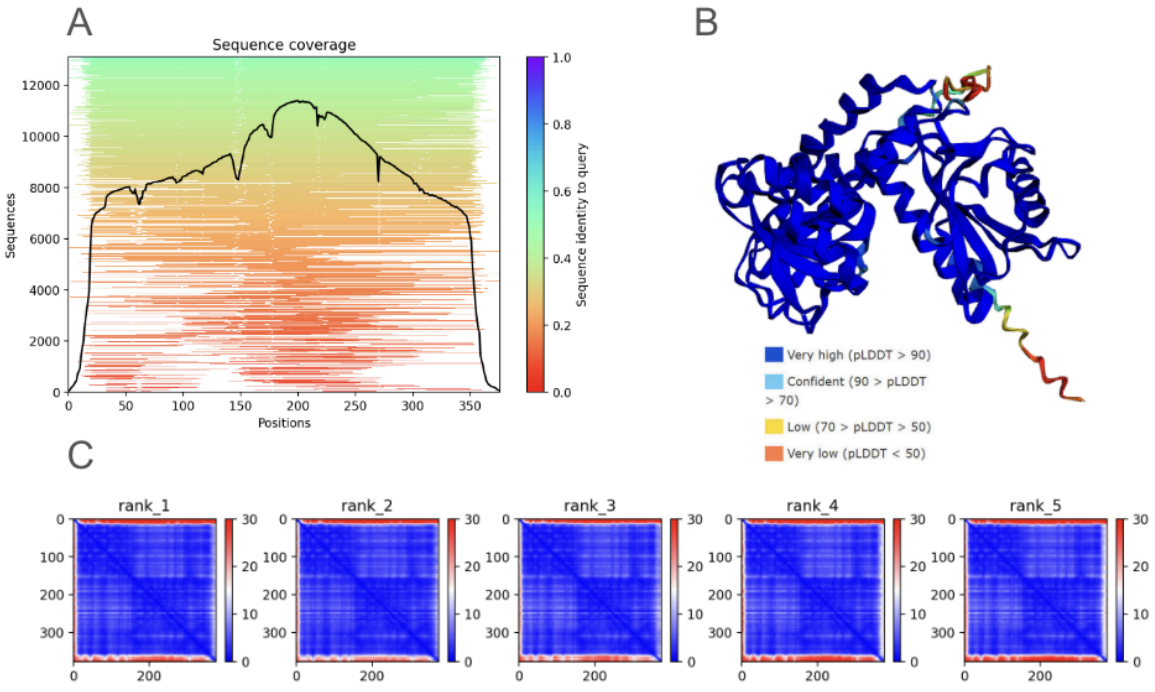

Figure 3: (A) The Multiple Sequence Alignment (MSA) coverage plots generated from the AlphaFold Colab server for AmDH1, The black line represents the number of sequences covering each position. (B) The predicted structure of AmDH1 using the AlphaFold Colab server, the colour depicting the confidence metric given by AlphaFold. (C) Predicted Aligned Error plots of all 5 structures from rank 1 to 5 for AmDH1.

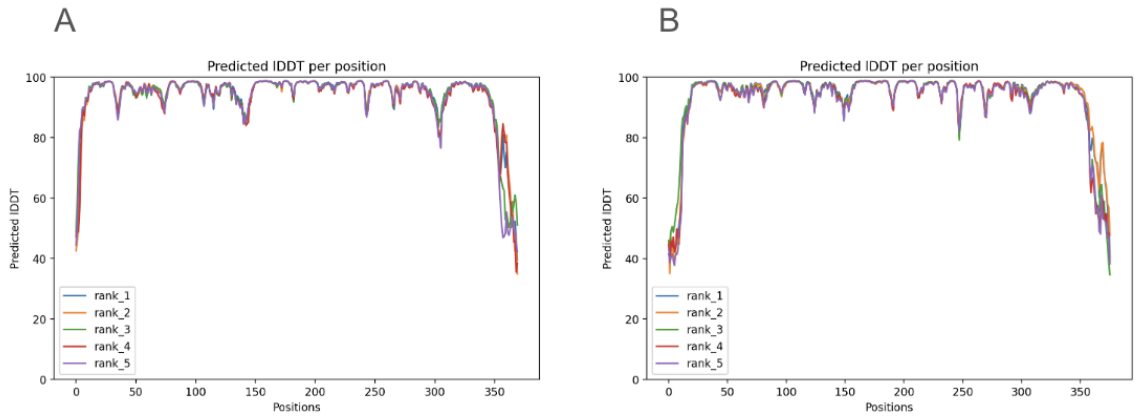

Figure 4: lDDT plots for (A) AADH1 and (B) AmDH1.

Similar results were obtained for AADH1 for the structure predicted using AlphaFold Colab sever. The AmDH1 has two necessary mutations that are derived from the parent enzymes K70S and N270L. As AmDH1 is a chimeric AmDH, that means it is constructed using the domains from two enzymes: the ketone binding domain is contributed by FBb-AmDH from a Phenylalanine Dehydrogenases from *Bacillus badius*, and the NADH binding domain is from Leucine dehydrogenase from *Geobacillus stearothermophilus* (Bacillus stearothermophilus)(Abrahamson *et al.*, 2013). The catalytic residue was determined to be K89 using the sequence alignment of two enzymes used to construct it, as shown in Figure 6 and Figure 7, respectively. The IDDT score for the catalytic residue The IDDT (Interface-Dependent Distance Test) score in AlphaFold is a confidence metric used to evaluate the reliability of a predicted protein structure(Jumper *et al.*, 2021). K89 had an IDDT score between 90-95, indicating that the structure is suitable for docking analysis, shown in Figure 4(B). Similarly, K80 is the catalytic residue for AADH1, which had an IDDT value of 92 as per the plot in Figure 4(A), and hence the position of catalytic residues in both the structures was valid, and both the structures were valid for docking analysis of the substrates.

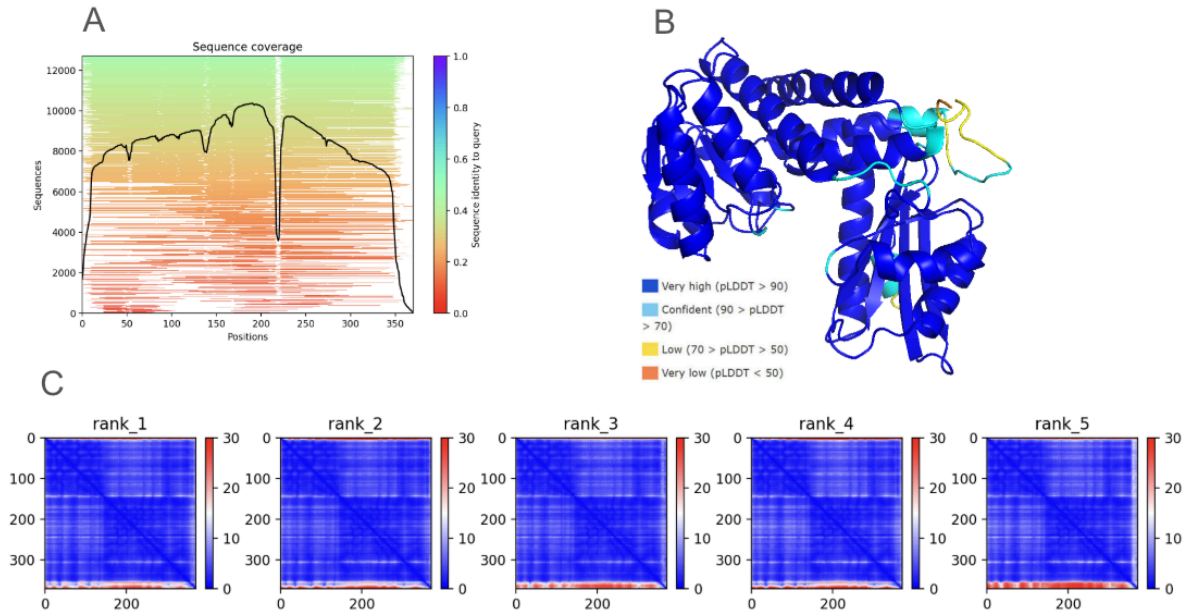

Figure 5: (A) The Multiple Sequence Alignment (MSA) coverage plot generated from the AlphaFold Colab server for AADH1. (B) The predicted structure of AADH1 using the AlphaFold Colab server, the colour depicting the confidence metric given by AlphaFold. (C) Predicted Aligned Error plots of all 5 structures from rank 1 to 5 for AADH1.

But AmDH and AADHs, according to their studied mechanism, exist in two conformations, according to the studies done on an S-selective AmDH from *Petrotoga mobilis* species, one the open conformation and the other closed conformation (Mayol *et al.*, 2019). The structure of AmDH1 and AADH1, when predicted using Alphafold 2 server, gave us an open conformation of the structure that is not the active form. This was confirmed when compared with the open and closed form of the already predicted *Petrotoga mobilis* structure. The rank 1 predicted structure of AmDH1, when compared to the open conformation of PM-AmDH4(chain A), gave a Root Mean Square Deviation(RMSD) value of 0.936 Å and when compared with the closed conformation of PM-AmDH4(chain B) gave a value of 0.940 Å. A lesser RMSD value, when compared with the open conformation, suggested that the predicted structure of AmDH1 is an open conformation. Even after a trial, we had a template of PDB id: 1C1D, an AADH from *Rhodococcous sp*. M4 was used as input, and AlphaFold gave the same open structure. Hence before proceeding to the docking analysis, we decided to predict

the structures using a different method.

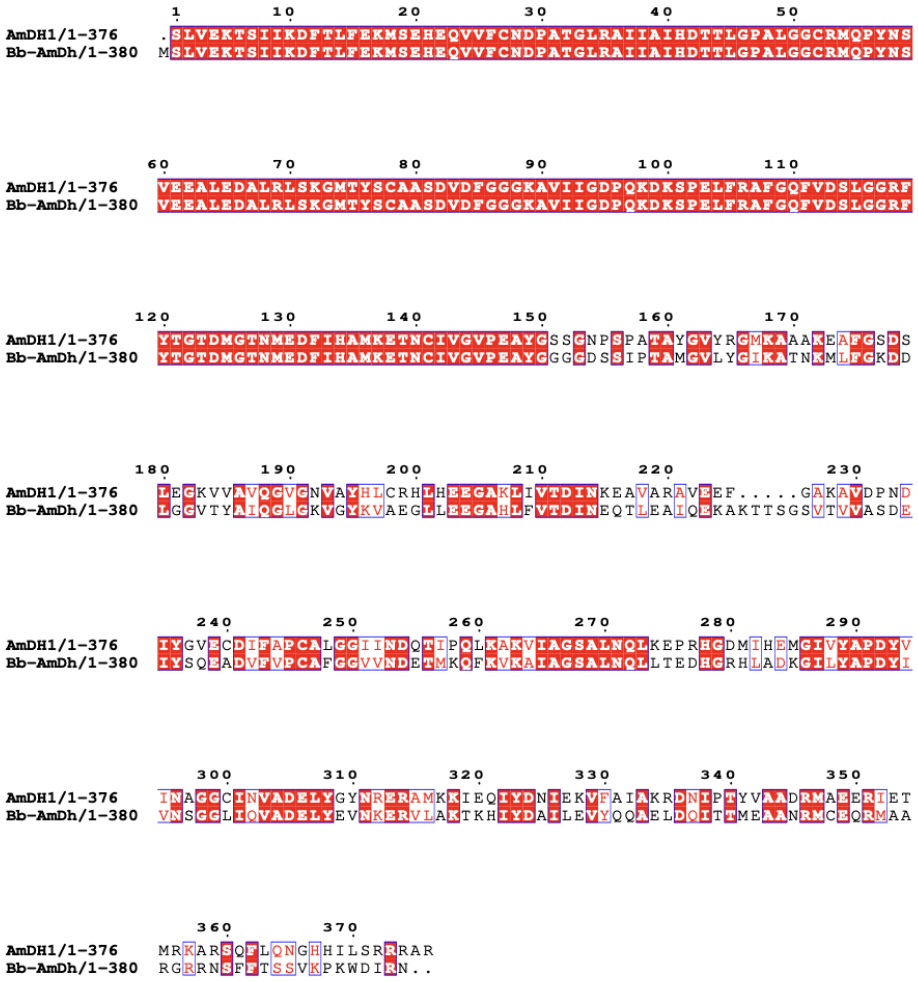

Figure 6: Sequence alignment of Bb-AmDH and AmDH1, 1-149 residues of AmDH1 are contributed by Bb-AmDH.

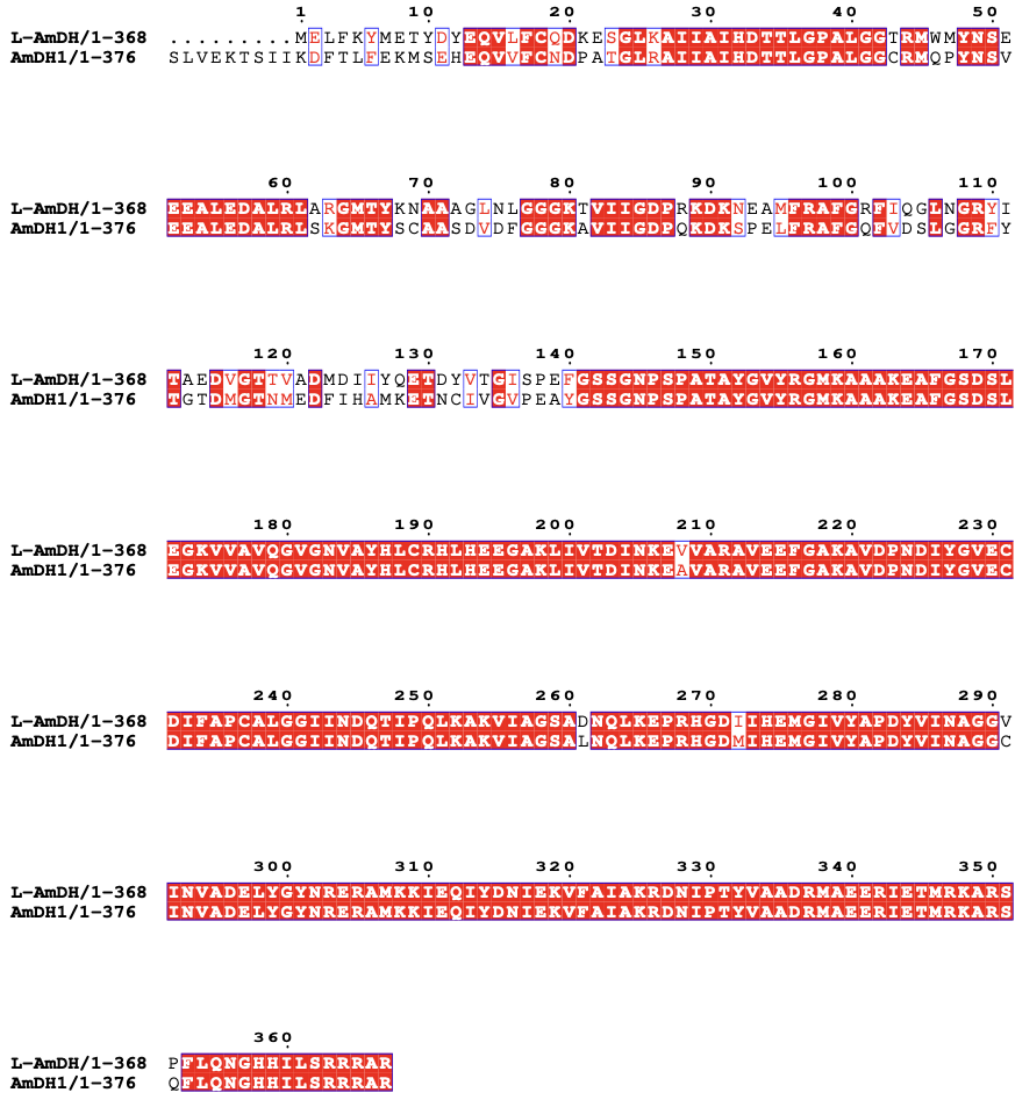

Figure 7: Sequence alignment of Bb-AmDH and AmDH1, 150-370 residues of AmDH1 are contributed by L-AmDH.

Prediction of AmDH1 and AaDH1 structure using AF Cluster Algorithm.

Alpha fold transformed the protein structure prediction domain, but it could not predict the same proteins in different conformations; some conformations of proteins are active, while others are not; therefore, it becomes essential to predict the right conformation of the protein, especially when you are relying on the structural analysis for protein engineering, by following the rational design approach. Hence, we took the help of the AF cluster algorithm developed by Hannah et. al. This algorithm has already been used to predict an open and closed conformation of KaiB protein from *Rhodobacter sphaeroides,* so we decided to predict the closed conformation of AADH1 using AF cluster so that we can get the active conformation of AADH1 for our rational design approach. The AF cluster algorithm for protein structure prediction is similarly available as a Google Colab notebook. AF-Cluster is

a method that enhances AlphaFold2's ability to predict alternative conformations by clustering multiple sequence alignments (MSA) based on evolutionary relationships(Wayment-Steele *et al.*, 2024). Altogether, the Multiple sequence alignment generated by the mmseqs2 algorithm used by AF Cluster gave us around 467 clusters. Due to less wall time on the Google servers, we decided to divide the clusters again into one set of 50 clusters and three sets of 45 clustered MSA each, which gave predictions each so that we could predict the structures without disconnecting from the server. For the reference closed and open structure, we decided to use the PDB id: 6G1M, as it had an AmDH structure predicted in the open and closed conformation. Chain A is the open conformation of Pet-AmDH, and chain B is the closed conformation of the same enzyme, this helped us to compare the Template Modeling Score (TM-Score) for the predicted structures using the clustering algorithm paired with AlphaFold 2. The output was all the PDB files and a graph with all the TM-scores and the confidence pLDDT (Predicted Local Distance Difference Test) for each structure predicted using different clusters clustered based on evolutionary relationship. To select from either an open or closed structure, we need to select the structure from either the left or right side of the diagonal of the graph intersecting at zero; we decided to filter out the structures based on the confidence pLDDT and the position of the point on the ™-score graph. From Set 1, we decided to pick three structures of clusters 4, 10 and 27. The structure predicted from cluster 27 was towards the left of the diagonal of the graph and was the highest point on the left side, indicating that the structure is a closed conformation, in other words, TM-score, when matched with any one of the chains, should be greater. The TM-score for the structure predicted by cluster 27 with 6G1M chain B, that is, the closed conformation, is 0.1839, which means this structure might be a closed conformation of the AADH1; similarly, structures from all 4 sets were sorted and compared for their conformations. The Figure 8 illustrates all the scores for structure from Set 1.

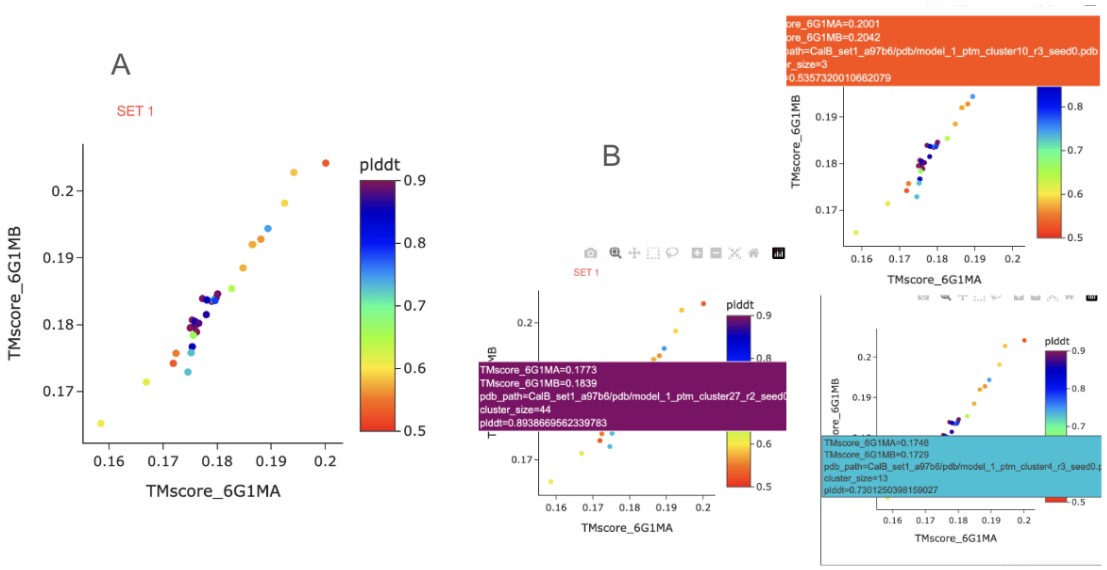

Figure 8: (A) The graph of ™ scores for all the structures predicted by the AF Cluster algorithm for SET 1. Comparison of the predicted structure by clustering the sequences. The colour of the point represents the pLDDT score, and the position on the graph represents global similarity with the closed or open conformation. (B) graph with clusters numbers 4, 10 and 27 highlighted with their TM scores when matched with the open and closed conformation of the A and B chains from the PDB id 6G1M.

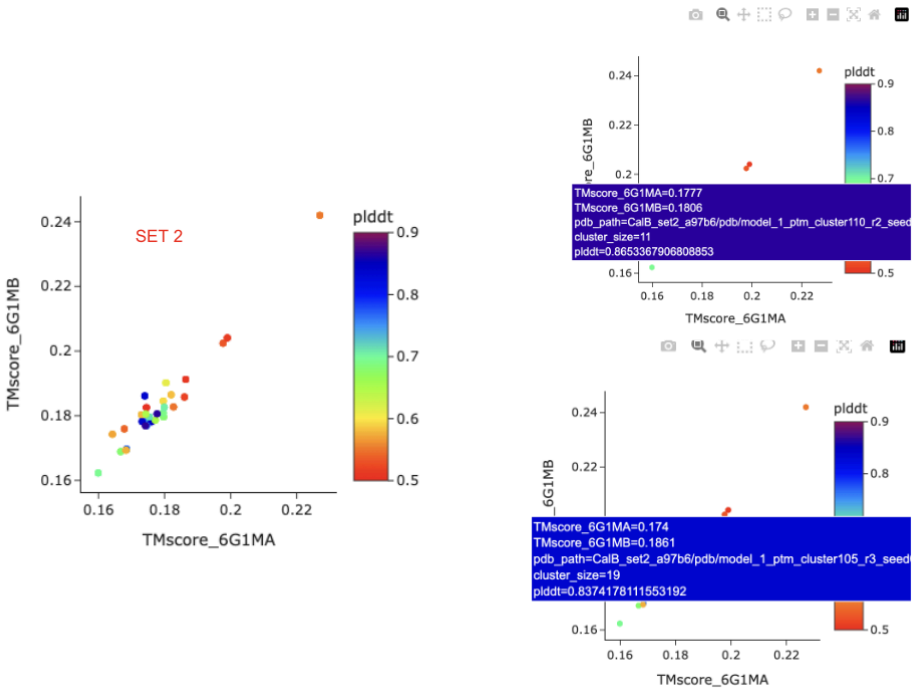

Figure 9: (A) The graph of TM- scores for all the structures predicted by the AF Cluster algorithm for SET 2. In comparison for set 2, the points should lie either on the left side or the right side of the graph, indicating the open or closed conformation of the CalB enzyme. (B) graph with clusters numbers 105 and 110 highlighted with their TM scores when matched with the open and closed conformation of the A and B chains from the PDB id 6G1M.

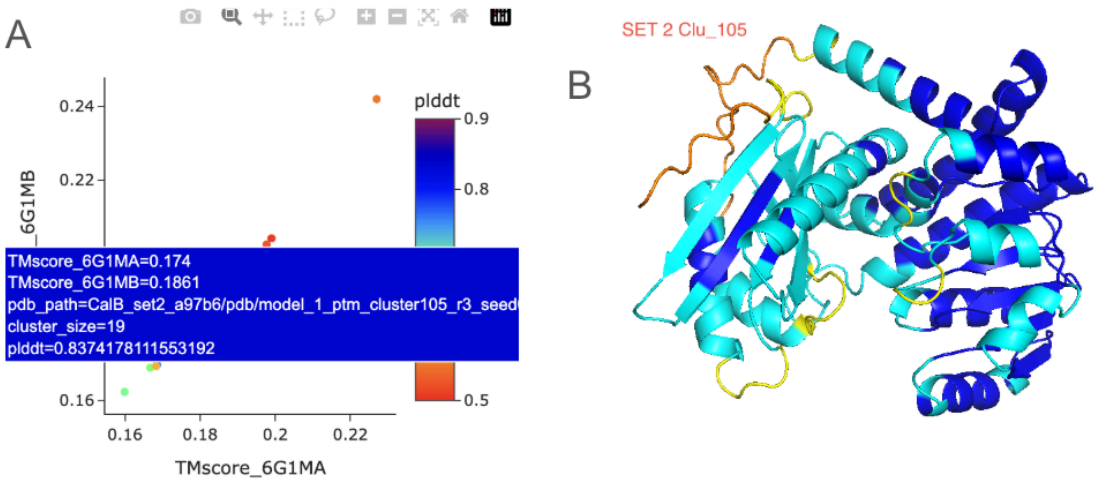

Figure 10: (A): The position of the structure from cluster 105 on the TM-score graph. (B) the overall confidence pLDDT for the same structure from cluster 105.

Upon sorting all the structures, from all 4 sets, we figured out that the structure of AADH1 from cluster 105 is the best with a TM score with the B chain of Pet-AmDh is 0.186, indicating that this is the closed structure for the AADH1. The IDDT of the residue K80, the active site of AADH1, is also between 87-90, indicating that this structure is the best suited for docking analysis. A similar structure prediction was made for the AmDH1, and a closed conformation of the structure was predicted. The Figure 11(A) describes the overlaid structure of AmDH1 predicted by the AF cluster, the closed form, and the structure predicted using AlphaFold 2 native software.

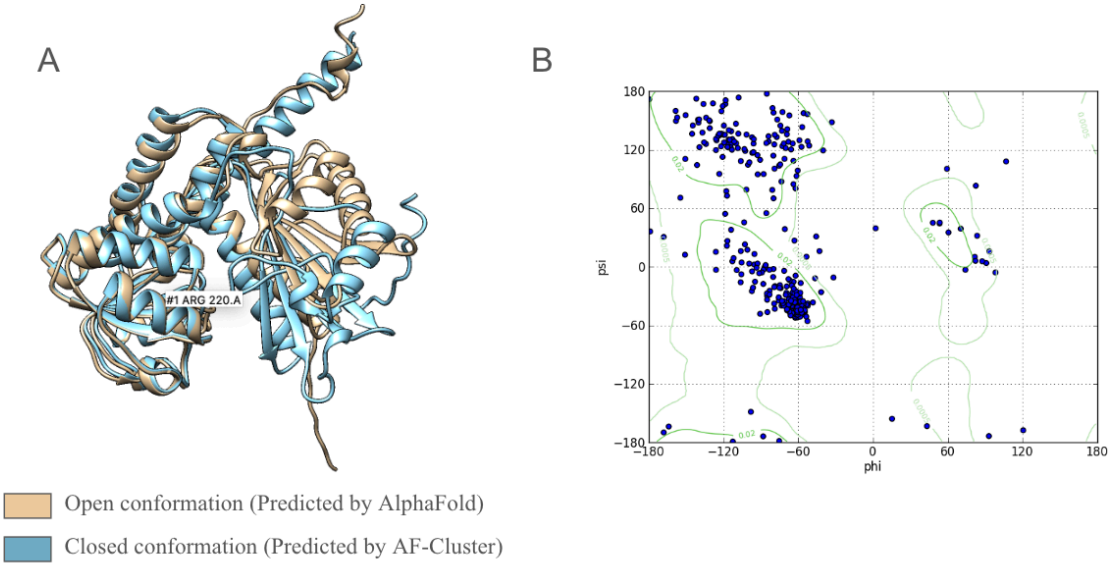

Figure 11: (A)The closed structure of AmDH1 overlaid on the open structure predicted by the AlphaFold. (B) The Ramachandran plot for the closed structure.

A similar approach was employed, and the structure of a mutated enzyme with mutations AmDH1_T114A/D115A was predicted and used for the actual docking

analysis. For all the structures shortlisted for closed conformation, a Ramachandran plot was generated to see for outliers, but very few outliers were observed, just like in the case in Figure 11(B), done for AmDH1.

Docking simulations for the rational design of the enzyme:

Docking simulations were carried out on the original structure of the AmDH1 and AADH1 using the Autodock Vina algorithm as a plugin for Pymol. The plugin used was DockingPie2.0(Rosignoli and Paiardini, 2022). The sequence for docking involved first docking of the substrate 1a and then docking of the cofactor NADH, as the substrate is towards the internal side of the active site cavity. All the docking simulations were repeated 3 times to remove any ambiguity. The substrate made a hydrogen bond with Leu49 in the original structure of AmDH1. As described in Figure 12, the initial docking pose and position for AmDH1. The initial docking pose for substrate 1a did not have any hydrogen bonds in the structure, and the initial distance between the intended β-ketone position and the catalytic residue was 6.837 Å and an average binding affinity of -4.8 kcal/mol. To find the residues. Similarly, docking simulations for AADH1 were performed for the closed form predicted using AF Cluster, which gave similar results, but a peculiar thing to be noticed is that in the AADH1 structure, the substrate 1a was better accepted with a more binding affinity of -5.2 kcal/mol., but the distance between the active site Lys80 and the β-ketone position was 8.707 depicted in Figure 13, indicating that there won't be any interaction between the catalytic residue and the β-ketone position of the substrate. However, according to the proposed reaction mechanism of the AmDH or AADH, hydride transfer between the catalytic residue and substrate needs to be achieved (Tseliou *et al.*). Hence, the enzyme needs to be mutated or engineered so that the substrate scope of the enzyme can be changed.

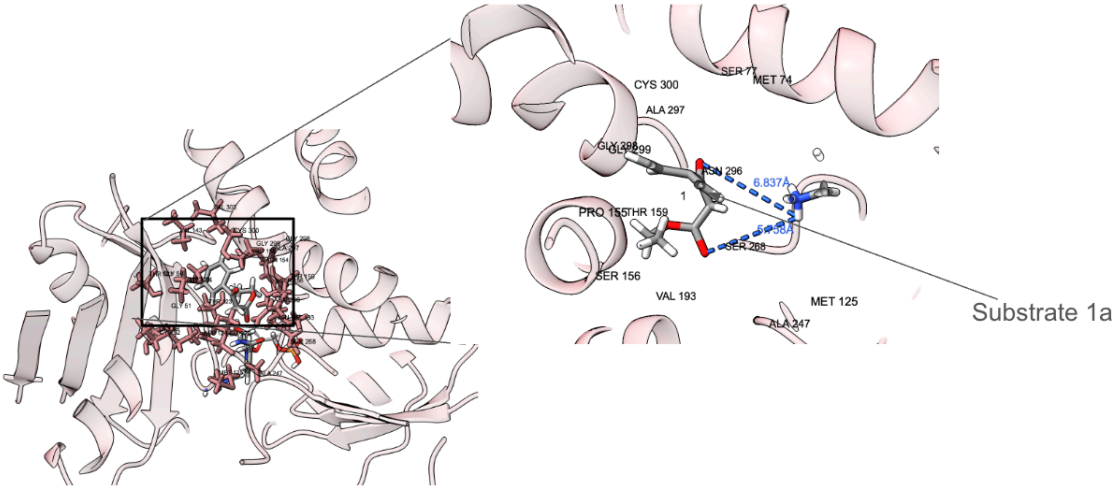

Figure 12: Depicting the pose and position of the substrate into the active site of the enzyme, the catalytic site is labelled as Lys89. The distance between the β-ketone position and the catalytic residue is labelled in blue. The binding affinity for this pose is -4.8 kcal/mol.

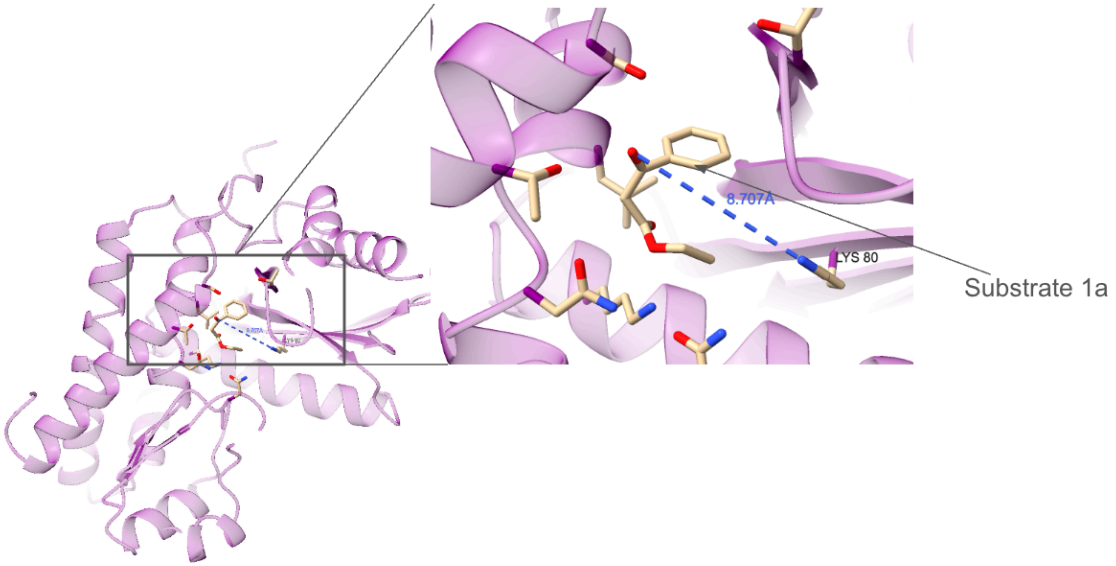

Figure 13: The docking pose for AADH1 and the distance between the β-ketone position and the active site Lys80.

Screening for residues in the AADH and AmDH that need to be mutated:

To make the enzyme accept our substrate of interest, mutations in the structure of AmDH1 and AADH1 are needed. To find this, we employed the strategy of residue screening in the structure of both enzymes. To do this, we compiled the list of

residues within the 6 Å radius sphere of the docked position of substrate 1a and analysed them systematically to look out for possible mutations. Figure 14(A) and (B) describe the residues selected for screening from AmDH1 to change the substrate scope of the enzyme. 6 Å radius of the sphere was selected as an ideal distance as there are no interactions left between the substrate and any residue at this distance.

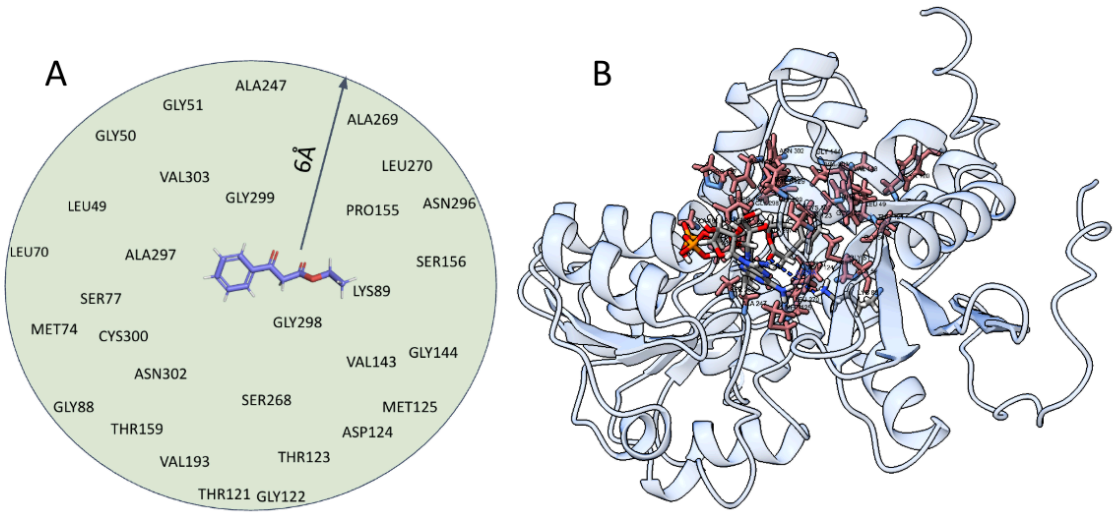

Figure 14: (A) A simplified depiction of the residues that are within the 6 Å of the docked substrate 1a (B) Actual representation of the residues that are present in the 6 Å radius sphere of the docked substrate.

The list of residues that we found is very large, and was almost impossible to screen all the residues in a combinatorial fashion, so we decided to sort them using some general logic to reduce the list of the residues to be screened. From the literature referred, the active site of these AmDH1 and AADH1 enzymes are divided into two aromatic substrates more, comparatively to the distinct groups of residues, outer hydrophilic and inner relatively hydrophobic; this division is more evident in the phenylalanine dehydrogenases that we are using. This structured cavity of hydrophobic and hydrophilic residues is what stabilises the Linear substrate, like the aromatic ring, has a more optimum environment in which to get stabilised within the pocket. Figures 15(A) and 15(B) describe the active site cavity of the AADH1 enzyme. As reported in the literature, the work done on FbB-AmDh had this differentiation of active site cavity in it, and when substitution mutations were done in the hydrophobic part of the cavity with another hydrophobic residue, it had more potential to change the substrate scope or the activity of the enzyme(Wu *et al.*,

2021). A similar approach was employed in our study on AADH1, and only the hydrophobic residues were decided to be screened for further study.

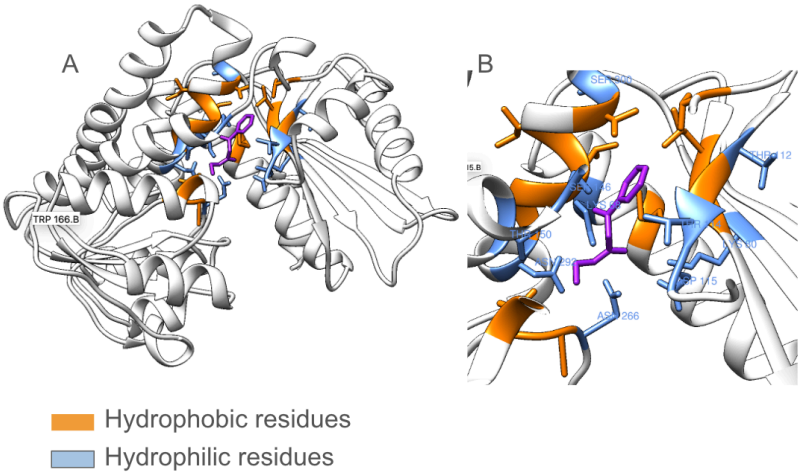

Figure 15:(A) The active site cavity of the AADH1 enzyme, the internal(orange) residues are hydrophobic residues and the outer(blue) are the hydrophilic residues. (B) zoomed in active cavity of the enzyme AADH1, with pink as the substrate.

Alanine screening of the Residues in the active site cavity:

The list of residues within the 6 Å of the substrate position in the wildtype AmDH1 was systematically studied and analysed to figure out the necessary residue substitutions. In the wildtype enzyme, the oxygen at the β-position ketone in the substrate interacts with ASN296; this was mutated to ALA296 to remove and screen for the other possible interactions. A similar screening is done for other amino acid residues interacting with the substrate other than the active site residues. A radius of 6 Å is considered for finding the necessary interacting residues. To figure out the contribution of each residue within the active cavity, we decided to do an in silico alanine screening of all the hydrophobic residues within the active site of AADH1. In this analysis, we mutated only the specific residue in question and did docking simulations to check if the distance between the β-position ketone and the catalytic residue K80 increased or decreased. The distance for the β-ketone position of the substrate and the catalytic residue K80 in AADH1 was 6.6 Å in the wildtype AADH1. As seen in Table, the distance between the catalytic residue K80 and the β-ketone position decreased when bulky residues like Met65 and Leu296 were substituted with alanine, indicating that these residues provided a steric hindrance to the substrate in the question. Hence, we shortlisted these residues in the first step and proceeded to the next step.

Residue	Mutant	Dist to K80(Å)	Docking score
WT AADH1	-	6.6	6.4
LEU40	ALA40	6.5	6.5
MET65	ALA65	6.3	6.5
VAL134	ALA134	6.6	6.4
VAL184	ALA184	6.5	6.4
LEU296	ALA296	6.2	6.4
ILE297	ALA297	6.7	6.4
VAL299	ALA299	7.1	6.5

Table1: A summary of the Alanine screening for all the hydrophobic residues of AADH1.

In the second step of our residue screening, we referred to the literature concerning the engineering of various dehydrogenase enzymes that had similar reactions or activity, or the reaction mechanism of the enzymes is the same. We screened the literature for imine reductases for further screening, as imine reductases are capable of accepting larger substrates for the catalysis reaction. However, the cavity size of the imine reductases and the shape were very different; imine reductases had a larger cavity compared to AADHs or AmDHs, with a conserved pocket structure having a larger pocket or cavity for the ketone substrate(Zheng *et al.*, 2024). Hence, we decided to check for other AADHs and AmDHs that were engineered for the same or similar substrate, such as a benzene ring and a methyl or ethyl ester. For the second round of the screening, we decided to align our AADH1 and AmDH1 with that of the other engineered AMDHs like the FBb-AmDH and Rs-AmDH or AADH; we figured out that the residues that are necessary for the conversion of AADH to AmDH that is K68S and N266L(AADH1 numbering) were necessary for the AADH1 for acceptance for ethyl3-oxo-3-phenylpropanoate as the substrate—indicating that first, we need to convert the AADH1 to its corresponding AmDH by mutating K68S and N266L. According to the structural analysis, these two mutations are necessary because AADHs prefer Amino acids with a carboxyl group similar to the ester group in our substrate. Hence, the substrates made a hydrogen bond with these substrates, directing the β-keto group away from the catalytic residue K80 in AADH1. After the first round of in-silico mutagenesis, it was found that the carboxyl group of

the substrate interacts with Lys66 and Asp266 within the enzyme, as shown in the Figure 16.

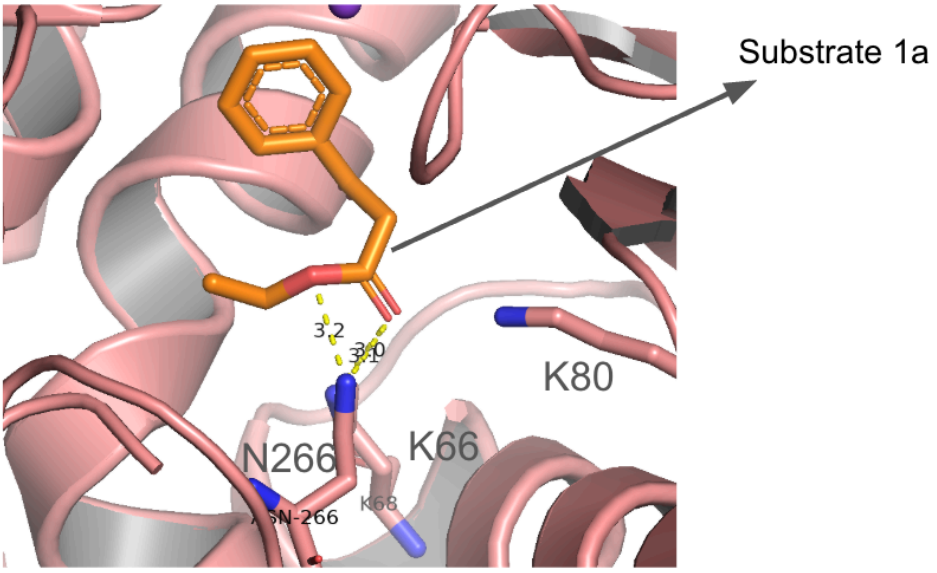

Figure 16: The aromatic substrate 1a having H-bonds with K66 and N266, directing the beta-keto position(adjacent to the six-member ring, behind the plane).

Prediction of AADH1 mutated structure using Homology modelling for rapid screening:

The screening method for residues using Alphafold or AF-Cluster was very accurate, according to the structure of the enzyme, but in the context of enzyme engineering, these methods are very time-consuming and resource-consuming, especially with reference to the computational resource, and when the requirements of a large number of mutated protein structures that are required. In such instances, enzyme engineering requires a less accurate but more resource-sustainable technique for structure prediction or model prediction. Hence we shifted from structure prediction to modelling of the mutated enzyme structure for our in silico analysis. We preferred the Swiss model homology modelling software, which was available via the Expasy web server(Waterhouse *et al.*, 2018). The second and further rounds for in-silico rational structural analysis. The B chain of PDB id 1C1D, an L-Phenylalanine Dehydrogenases from *Rhodococcus sp. M4* was used as the template for making the model, the sequence identity between the template and the query sequence(AADH1) was less than 40%, but this is the only available structure for

R-selective AADH. All the structures generated using the Swiss-model servers using the template were validated using the SAVES server. The Ramachandran plot and Errat plot of the structure were generated to validate the structures. Additionally, Procheck was used on the same server to validate the structure. Figure 17 describes the Ramachandran plot for the wildtype AADH1 enzyme and statistics related to it. As seen, as many as 91.7 residues are within the most favoured regions, indicating that the structure is of good quality and can be used for docking simulations and further analysis. To check if the mutations that we have performed in silico, do they even translate to the change in the activity of the enzyme in a dynamic situation.

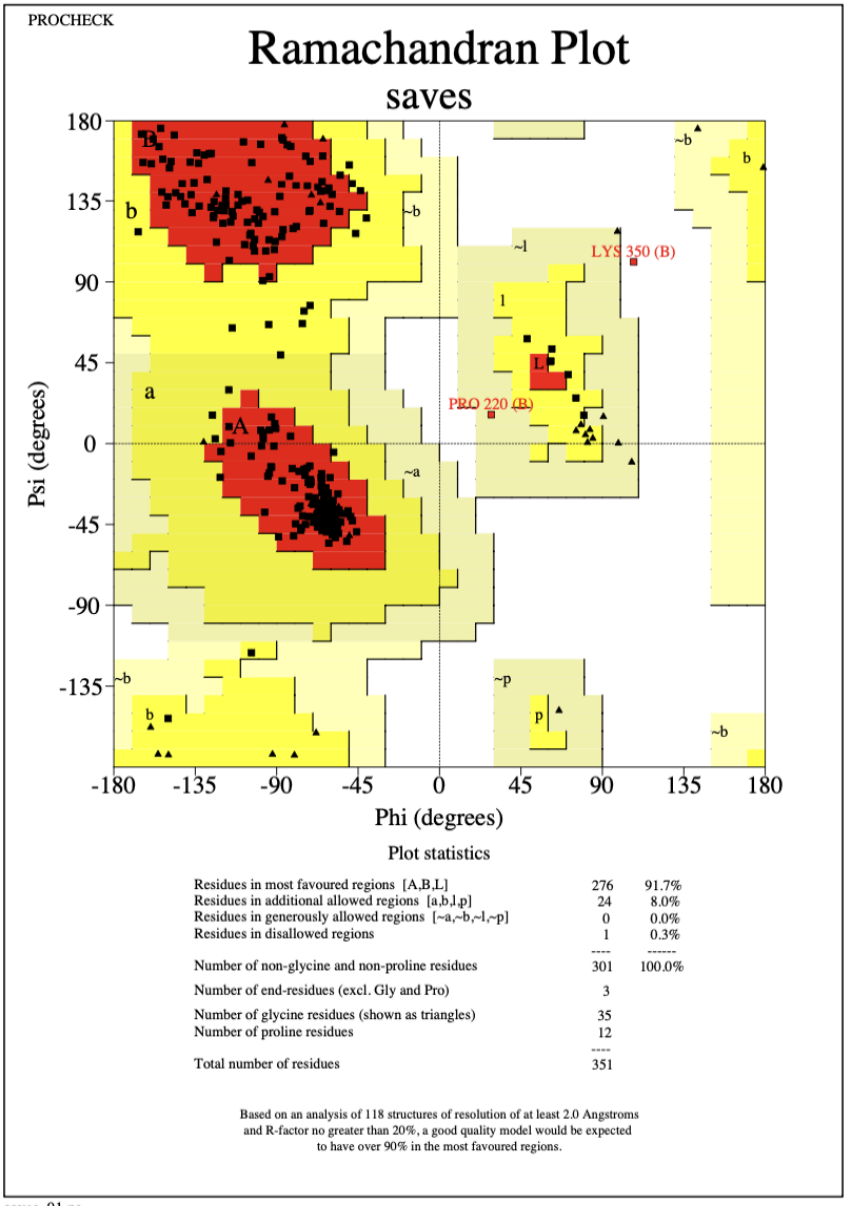

Figure 17: Ramachandran plot for wildtype AADH1, modelled using Homology modelling

Molecular Dynamics Simulations for AADH1:

Even if the modelled structure of the enzyme is accurate, the structure is always stationary, along with the substrate, and the cofactor is stagnant. This is not the situation in natural systems, be it within the organism or outside it. Many times the structure that is used for structural bioinformatic analysis, in the context of biocatalysis, that structure might not be active or not even stable in the reaction conditions with all the buffers, and the solvent in which the substrate or the cofactor or the product is stable. To check if the mutations that we have done in silico have really translated into changing the substrate preference for the enzyme, we decided to do Molecular Dynamics Simulations for all the mutations in the third round along the wildtype AADH1 enzyme. The Root Mean Square Fluctuations for all the Cα atoms for each residue in the protein structure were plotted from the Molecular Dynamics simulations data that were run for 100ns. It is evident from the Figure 18 graph that the mutated structure is more stable with respect to substrate 1a, indicating that the substrate has more stable interactions with the protein overall, especially with the catalytic site K80. In total, six residues were shortlisted for substitution in AADH1.

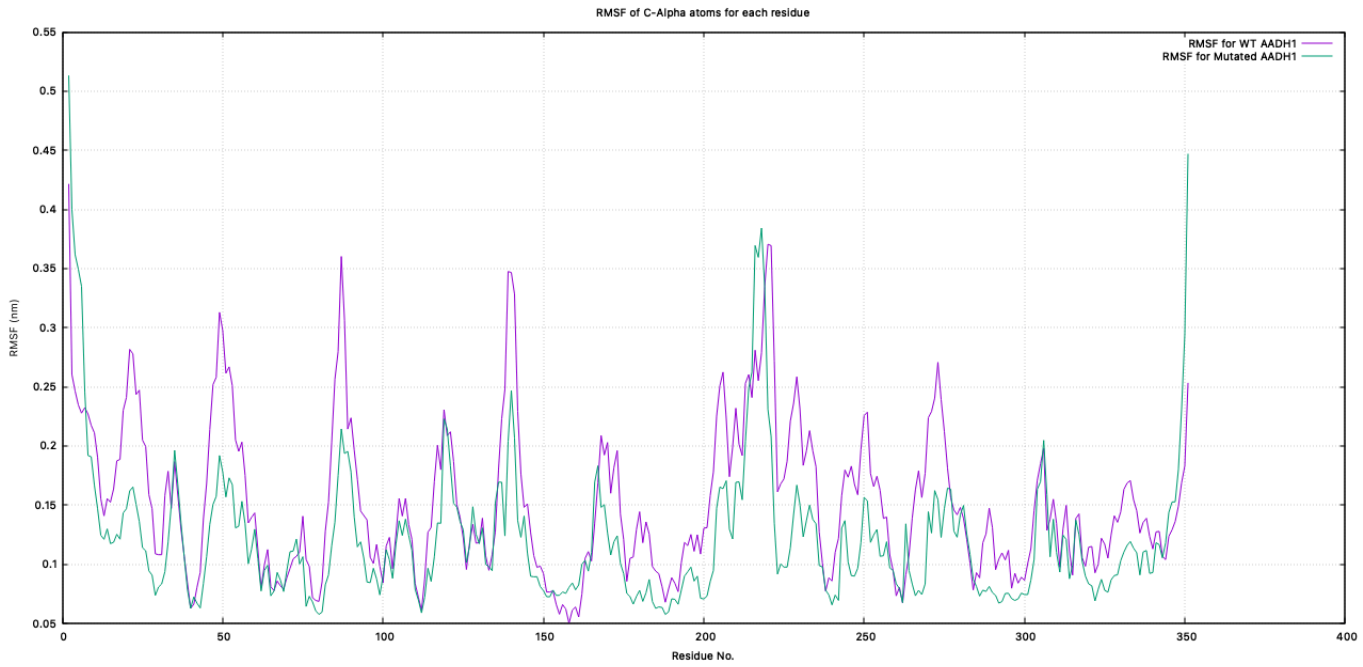

Figure 18: The RMSF values for each residue compared for the Wildtype AADH1 and mutated AADH1 after three rounds of mutations.The mutated AADH1 has a lesser overall RMSF indicating that these mutations make the enzyme stable. Both the structures had the same 1a as the substrate.

Cloning, expression and purification of AmDH1:

The AmDH1 enzyme was cloned into the Pet28 vector within the restriction sites NdeI and XhoI and transformed into *E. coli* DH5α cell lines, five positive colonies from the Petri plate were taken for inoculating and sequencing. A Sanger sequencing for these positive clones was carried out, and one of the colonies gave the expected sequence for AmDH1, the cloned sequence had an N-terminal 6XHis tag in it. *E. coli* DH5α cells were used for cloning the gene, and *E. coli* BL21-DE3 cells were used to express the AmDH1. Nickel-NTA(Ni-NTA) column was used for affinity chromatography; The Figure 19 bhas an image of SDS gel for the colonies expressing AmDH1 and the purified AmDH1. The protein was eluted at 200mM Imidazole concentration. The Figure 19 clearly shows a thick band for elution 1 and a slightly

thin bad for elution 2 at 41kDa, which is the estimated molecular weight of AmDH1 protein.

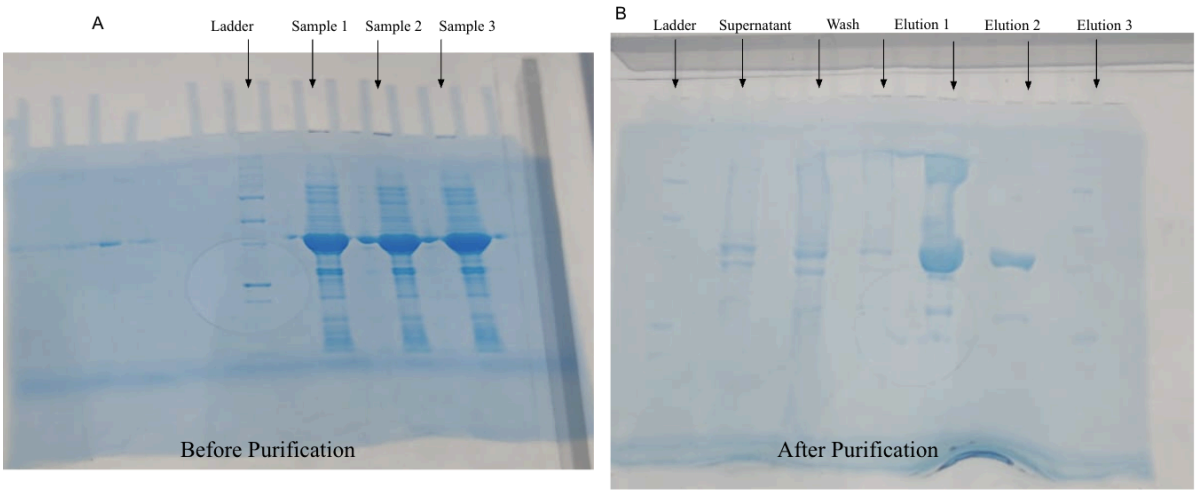

Figure 19: (A)Expression studies of AmDH1 and (B) Purification of AmDH1 using His-tag Affinity Chromatography.

Chapter 4 Discussion

The accurate determination of protein structure is critical for enzyme engineering and functional characterisation. Since experimental structural biology techniques such as X-ray crystallography and cryo-electron microscopy (cryo-EM) are resource-intensive and time-consuming, computational structure prediction using AlphaFold2 has emerged as a powerful alternative(Yousefi *et al.*, 2017). In this study, AlphaFold2 Colab Server was utilized to predict the structures of AmDH1 and AADH1, both of which displayed high confidence scores (pLDDT > 90), indicating structural reliability for further docking and rational enzyme design. The Multiple Sequence Alignment (MSA) coverage plots generated by AlphaFold2 provided insights into sequence conservation. The core regions exhibited high sequence identity, whereas terminal regions displayed greater variability, suggesting potential flexibility or disorder. These findings align with previous studies, where well-conserved core regions were found to be functionally significant for substrate binding and catalysis.

Additionally, IDDT confidence scores confirmed that the catalytic residues (K89 in AmDH1 and K80 in AADH1) were predicted with high accuracy, reinforcing the structural integrity of the modelled enzymes. Similar structural validation approaches have been reported for R-selective AmDHs from Petrotoga mobilis and phenylalanine dehydrogenases from Rhodococcus sp. M4, confirming that computational predictions can provide reliable models for enzyme engineering(Yousefi *et al.*, 2017). Structural analysis revealed that AlphaFold2-predicted structures corresponded to the “open” conformation of both AmDH1 and AADH1. This was confirmed by comparing RMSD values between AmDH1 and the known open and closed forms of Petrotoga mobilis AmDH4 (PDB: 6G1M). The lower RMSD (0.936 Å), when compared to the open form, indicated that the modelled AmDH1 was in an inactive conformation. Since enzyme function is conformation-dependent, obtaining the closed (active) conformation was essential for docking and substrate interaction studies. To address this, the AF-Cluster algorithm was employed, which leverages sequence clustering to predict multiple conformations. This method successfully generated a closed structure for AADH1, confirmed by its higher TM-score (0.186) with the closed state of 6G1M Chain B, indicating significant structural similarity. The importance of predicting multiple

conformations has been highlighted in previous structural studies, where enzyme activity was linked to conformational switching. For example, KaiB, a circadian clock protein, was only functionally active in its closed state, which was accurately modelled using AF-Cluster(Wayment-Steele *et al.*, 2024). Similarly, our findings emphasise that AlphaFold2 alone may not always predict biologically relevant conformations, necessitating complementary approaches like AF-Cluster. The lower confidence levels for the predicted closed conformation might be due to the AlphaFold 2 scoring method, which is still there in the AF-Cluster.

Docking simulations were conducted using AutoDock Vina to assess substrate compatibility. The binding affinity of substrate 1a to AmDH1 was calculated as -4.8 kcal/mol, with a β-ketone distance of 6.837 Å from the catalytic residue K89. Similarly, AADH1 displayed a stronger binding affinity of -5.2 kcal/mol but with a larger β-ketone distance of 8.707 Å from K80, indicating suboptimal positioning for hydride transfer. These results suggest that although AADH1 demonstrated better substrate affinity, its active site geometry was not ideal for catalysis. A similar phenomenon was observed in engineered AmDHs, where binding affinity alone did not always correlate with catalytic efficiency(Ye *et al.*, 2015). Instead, optimal positioning of the substrate relative to the catalytic lysine was a more critical factor for effective amination. From the previous work, it is already clear that the hydride transfer range for AADHs and AmDHs is slightly more than what is naturally observed, even for their native substrate amino acids(Tseliou *et al.*).

To improve enzyme-substrate interactions, residue screening was performed within a 6 Å radius of the docked substrate, identifying critical residues influencing substrate positioning. Among these, Met65 and Leu296 in AADH1 contributed to steric hindrance, preventing optimal substrate alignment. Alanine substitutions of these residues resulted in decreased β-ketone distance, supporting their role in active site accommodation. Previous studies on FbB-AmDH and Rs-AmDH demonstrated that modifying hydrophobic residues within the active site cavity significantly altered substrate scope(Wu *et al.*, 2021). Similarly, our findings indicate that hydrophobic interactions are key determinants of substrate compatibility in AADH1.

Furthermore, sequence alignment with engineered AmDHs revealed that K68S and N266L mutations were essential for converting AADH1 into an AmDH-like enzyme, as these residues influenced amine vs. carboxylate substrate preferences. This

aligns with findings in imine reductase engineering, where active site remodelling enabled broader substrate acceptance

Given the computational expense of AlphaFold2 and AF-Cluster, an alternative homology modelling approach was employed for mutated structures using the Swiss-Model server. The B-chain of Rhodococcus sp. M4 AADH (PDB: 1C1D) was selected as a template despite a moderate sequence identity (<40%). Structural validation using Ramachandran plots (91.7% favoured residues) confirmed the quality of the predicted models, making them suitable for docking and further Molecular Dynamics (MD) simulations. This approach provided a faster and resource-efficient alternative to AlphaFold-based predictions, particularly useful when screening multiple mutants

To evaluate the structural stability of mutant enzymes, 100 ns Molecular Dynamics (MD) simulations were conducted. The Root Mean Square Fluctuation (RMSF) values indicated that the mutated AADH1 exhibited lower overall fluctuations, particularly in substrate-interacting regions, suggesting greater stability. MD simulations have been widely used to assess mutational effects on enzyme flexibility and activity. In a study on phenylalanine dehydrogenase, stabilising mutations led to reduced active site fluctuations, enhancing catalytic efficiency(Lai *et al.*, 2024). Similarly, our findings confirm that rationally designed mutations contribute to improved enzyme rigidity and functional stability. To test this hypothesis, the next step is to get those mutations in situ using site-directed mutagenesis. As per the previous work done on engineering the AADHs or AmDHs, most of them are focused on mutating only one or, at the max, two residues within the wild type, but in this work, more mutations are proposed so that the enzyme which also has low promiscuity towards larger substrates, can be increased.

References

Abraham, MJ, Murtola, T, Schulz, R, Páll, S, Smith, JC, Hess, B, and Lindahl, E (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25.

Abrahamson, MJ, Wong, JW, and Bommarius, AS (2013). The Evolution of an Amine Dehydrogenase Biocatalyst for the Asymmetric Production of Chiral Amines. Adv Synth Catal 355, 1780–1786.

Anandakrishnan, R, Aguilar, B, and Onufriev, AV (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res 40, W537–W541.

Bommarius, AS (2019). Amine dehydrogenases occur in nature. Nat Catal 2, 288–289.

Bommarius, AS, Abrahamson, MJ, and Bommarius, B (2014a). Engineered amine dehydrogenases and methods of use thereof.

Bommarius, BR, Schürmann, M, and Bommarius, AS (2014b). A novel chimeric amine dehydrogenase shows altered substrate specificity compared to its parent enzymes. Chem Commun 50, 14953–14955.

Brunhuber, NMW, Thoden, JB, Blanchard, JS, and Vanhooke, JL (2000). Rhodococcus l-Phenylalanine Dehydrogenase: Kinetics, Mechanism, and Structural Basis for Catalytic Specifity,. Biochemistry 39, 9174–9187.

Cárdenas-Fernández, M, Roddan, R, Carter, EM, Hailes, HC, and Ward, JM (2023). The Discovery of Imine Reductases and their Utilisation for the Synthesis of Tetrahydroisoquinolines. ChemCatChem 15, e202201126.

Eberhardt, J, Santos-Martins, D, Tillack, AF, and Forli, S (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model 61, 3891–3898.

Gordon, JC, Myers, JB, Folta, T, Shoja, V, Heath, LS, and Onufriev, A (2005). H++: a server for estimating p Ka s and adding missing hydrogens to macromolecules. Nucleic Acids Res 33, W368–W371.

Jumper, J, Evans, R, Pritzel, A, Green, T, Figurnov, M, Ronneberger, O, Tunyasuvunakool, K, Bates, R, Žídek, A, Potapenko, A, *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

Kim, G, Lee, S, Levy Karin, E, Kim, H, Moriwaki, Y, Ovchinnikov, S, Steinegger, M, and Mirdita, M (2024a). Easy and accurate protein structure prediction using ColabFold. Nat Protoc, 1–23.

Kim, S, Ga, S, Bae, H, Sluyter, R, Konstantinov, K, Shrestha, LK, Kim, YH, Kim, JH, and Ariga, K (2024b). Multidisciplinary approaches for enzyme biocatalysis in pharmaceuticals: protein engineering, computational biology, and nanoarchitectonics. EES Catal 2, 14–48.

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

Lai, Z, Wu, T, Zhou, F, Xu, Y, and Mu, X (2024). Construction of an enzymatic-chemical cascade involving engineered amine dehydrogenases for the synthesis of chiral non-α-amino acids by substrate engineering and enzyme evolution. Mol Catal 559, 114099.

Mangas-Sanchez, J, France, SP, Montgomery, SL, Aleku, GA, Man, H, Sharma, M, Ramsden, JI, Grogan, G, and Turner, NJ (2017). Imine reductases (IREDs). Curr Opin Chem Biol 37, 19–25.

Mayol, O, Bastard, K, Beloti, L, Frese, A, Turkenburg, JP, Petit, J-L, Mariage, A, Debard, A, Pellouin, V, Perret, A, *et al.* (2019). A family of native amine dehydrogenases for the asymmetric reductive amination of ketones. Nat Catal 2, 324–333.

Myers, J, Grothaus, G, Narayanan, S, and Onufriev, A (2006). A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. Proteins Struct Funct Bioinforma 63, 928–938.

Prejanò, M, Sheng, X, and Himo, F (2022). Computational Study of Mechanism and Enantioselectivity of Imine Reductase from Amycolatopsis orientalis. ChemistryOpen 11, e202100250.

Pushpanath, A, Siirola, E, Bornadel, A, Woodlock, D, and Schell, U (2017). Understanding and Overcoming the Limitations of Bacillus badius and Caldalkalibacillus thermarum Amine Dehydrogenases for Biocatalytic Reductive Amination. ACS Catal 7, 3204–3209.

Rosignoli, S, and Paiardini, A (2022). DockingPie: a consensus docking plugin for PyMOL. Bioinformatics 38, 4233–4234.

Scardino, V, Di Filippo, JI, and Cavasotto, CN (2023). How good are AlphaFold models for docking-based virtual screening? iScience 26, 105920.

Shin, Y-C, Yun, H, and Park, HH (2018). Structural dynamics of the transaminase active site revealed by the crystal structure of a co-factor free omega-transaminase from Vibrio fluvialis JS17. Sci Rep 8, 11454.

Tseliou, V, Masman, MF, Knaus, T, and Mutti, FG Current Status of Amine Dehydrogenases: From Active Site Architecture to Diverse Applications Across a Broad Substrate Spectrum. ChemCatChem n/a, e202400469.

Vanhooke, JL, Thoden, JB, Brunhuber, NMW, Blanchard, JS, and Holden, HM (1999). Phenylalanine Dehydrogenase from Rhodococcus sp. M4: High-Resolution X-ray Analyses of Inhibitory Ternary Complexes Reveal Key Features in the Oxidative Deamination Mechanism,. Biochemistry 38, 2326–2339.

Wang, D-H, Chen, Q, Yin, S-N, Ding, X-W, Zheng, Y-C, Zhang, Z, Zhang, Y-H, Chen, F-F, Xu, J-H, and Zheng, G-W (2021). Asymmetric Reductive Amination of Structurally Diverse Ketones with Ammonia Using a Spectrum-Extended Amine Dehydrogenase. ACS Catal 11, 14274–14283.

Waterhouse, A, Bertoni, M, Bienert, S, Studer, G, Tauriello, G, Gumienny, R, Heer, FT, de Beer, TAP, Rempfer, C, Bordoli, L, *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46, W296–W303.

Wayment-Steele, HK, Ojoawo, A, Otten, R, Apitz, JM, Pitsawong, W, Hömberger, M, Ovchinnikov, S, Colwell, L, and Kern, D (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. Nature 625, 832–839.

Wu, T, Mu, X, Xue, Y, Xu, Y, and Nie, Y (2021). Structure-guided steric hindrance engineering of Bacillus badius phenylalanine dehydrogenase for efficient l-homophenylalanine synthesis. Biotechnol

Biofuels 14, 207.

Wu, T, Wang, Y, Zhang, N, Yin, D, Xu, Y, Nie, Y, and Mu, X (2023). Reshaping Substrate-Binding Pocket of Leucine Dehydrogenase for Bidirectionally Accessing Structurally Diverse Substrates. ACS Catal 13, 158–168.

Ye, F, Zhang, Z, Zhao, W, Ding, J, Wang, Y, and Dang, X (2021). Two methods for the preparation of sitagliptin phosphate via chemical resolution and asymmetric hydrogenation. RSC Adv 11, 4805–4809.

Ye, LJ, Toh, HH, Yang, Y, Adams, JP, Snajdrova, R, and Li, Z (2015). Engineering of Amine Dehydrogenase for Asymmetric Reductive Amination of Ketone by Evolving Rhodococcus Phenylalanine Dehydrogenase. ACS Catal 5, 1119–1122.

Yousefi, F, Ataei, F, Arab, SS, and Hosseinkhani, S (2017). Increase of *Bacillus badius* Phenylalanine dehydrogenase specificity towards phenylalanine substrate by site-directed mutagenesis. Arch Biochem Biophys 635, 44–51.

Zheng, X, Dou, Z, Xiang, W, Zhang, W, Ni, Y, and Xu, G (2024). Engineering the enantioselectivity of a novel imine reductase from *Streptomyces viridochromogenes* for the dynamic kinetic reductive amination of cyclic β-ketoester. Mol Catal 559, 114039.

Zhou, F, Xu, Y, Nie, Y, and Mu, X (2022). Substrate-Specific Engineering of Amino Acid Dehydrogenase Superfamily for Synthesis of a Variety of Chiral Amines and Amino Acids. Catalysts 12, 380.

AlphaFold2.ipynb - Colab. Available at: https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb. Accessed March 21, 2025.