Investigation of structure-function relationship in hydroxynitrile lyases and MIO-containing class I lyase like enzymes

Ph. D. Thesis

Zsófia Bata
Biotechnological Engineering MSc

Supervisor
Prof. László Poppe
Ph. D., D. Sc.

Advisor
Prof. Beáta G. Vértessy
Ph. D., D.Sc.

Budapest
2019
My individual role in the preparation of the Ph. D. thesis

All results presented in this Ph. D. thesis are the result of the combined effort of our research groups and collaborators, hence ‘first person plural’ tense is used throughout the text.

The thesis work is based on the scientific publications listed in Chapter 8.1 and on some very recent yet unpublished results. My contributions for the articles are detailed in Table 1.

Table 1. My individual contributions to the scientific publications that the thesis is based on

<table>
<thead>
<tr>
<th>Scientific publication</th>
<th>My contribution for the manuscript</th>
<th>Contributions of major importance of the co-authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones, Bata, Kazlauskas, 2017 ACS Catalysis</td>
<td>Planning, running and evaluation of the molecular dynamics simulations.</td>
<td>Planning, creating and measuring the activity of the hydroxynitrile lyase mutants. All authors contributed equally to the writing of the manuscript.</td>
</tr>
<tr>
<td>Bata et al. 2017, Advanced Synthesis and Catalysis</td>
<td>Production and purification of MIO enzymes. Planning, performing and evaluation of kinetic assays and thermodynamic binding experiments. The manuscript was in most part written by myself and my supervisors, Prof. László Poppe and Prof. Vértessy G. Beáta</td>
<td>Synthesis, purification and x-ray structure evaluation of the aminophosphonic acid inhibitors.</td>
</tr>
<tr>
<td>Varga et al, 2017, Studia Universitatis Babes-Bolyai Chemia</td>
<td>Production and purification of KkPAL. Sequence identification, initial cloning of KkPAL as well as initial production and purification of KkPAL.</td>
<td>Both authors contributed equally to the concept and writing of this article.</td>
</tr>
<tr>
<td>Bata, Bell 2018 Élet és tudomány</td>
<td>Planning, crystallizing, measuring and solving the PcPAL crystal structures.</td>
<td>The dataset for 6F6T was collected by Ibolya Leveles, the dataset for 6H2O was collected by Dr. Veronika Harmat.</td>
</tr>
</tbody>
</table>
# Table of Contents

Acknowledgement.................................................................................................................. 5  
Abbreviations ........................................................................................................................................ 7  
1 Introduction ........................................................................................................................................... 9  
  1.1 Hydroxynitrile lyases ................................................................................................................... 10  
  1.1.1 Plethora of compounds reachable by HNL catalyzed reactions .............................................. 11  
  1.1.2 HNL activity appeared in at least six different enzyme folds by convergent evolution .............................................................................................................................. 12  
  1.1.3 HNL reaction mechanisms in the α/β-hydrolase fold ............................................................ 15  
  1.1.4 Evolutionary relationships of α/β-hydrolase fold HNLs ......................................................... 17  
1.2 MIO-containing class I like lyases ............................................................................................... 19  
  1.2.1 Reaction mechanism of MIO enzymes .................................................................................. 20  
  1.2.2 Structural determinants of ligand binding and substrate specificity of MIO enzymes .................................................................................................................................................................................. 24  
  1.2.3 Structural features differentiating AL and AM activity ......................................................... 28  
2 Aims .................................................................................................................................................... 33  
3 Materials and methods .................................................................................................................. 34  
  3.1 Cloning and mutagenesis ............................................................................................................. 34  
  3.1.1 Cloning ........................................................................................................................................ 34  
  3.1.2 Mutagenesis ............................................................................................................................... 34  
  3.2 Protein expression, purification .................................................................................................... 35  
  3.2.1 Protein expression .................................................................................................................... 35  
  3.2.2 Protein purification ................................................................................................................... 35  
  3.3 Enzyme functional studies ........................................................................................................... 37  
  3.3.1 Steady state kinetic measurements ......................................................................................... 37  
  3.3.2 Isothermal titration calorimetry (ITC) .................................................................................... 38  
  3.4 Protein structure determination .................................................................................................. 39  
  3.4.1 Protein crystallization ............................................................................................................... 39  
  3.4.2 Diffraction data collection ........................................................................................................ 40  
  3.4.3 Diffraction data processing and protein structure refinement ............................................. 40  
3.5 Molecular modeling .................................................................................................................... 40  
  3.5.1 Construction of model structures ............................................................................................ 40  
  3.5.2 Molecular docking .................................................................................................................. 42
3.5.3 Molecular dynamics simulations of hydroxynitrile lyases ............................................. 45
3.5.4 Random acceleration molecular dynamics simulations of MIO enzymes .......... 47
4 Results and Discussion ........................................................................................................ 50
  4.1 Reaction mechanisms in ancestral and modern HNLs ............................................... 50
      4.1.1 HNL phenotypes of the consensus active sites .................................................. 50
      4.1.2 Modeling the structural effect of the mutations ............................................... 52
  4.2 Characterization of aminophosphonic acid inhibitors of PALs ................................. 65
      4.2.1 Steady state inhibition Studies with PcPAL ....................................................... 66
      4.2.2 Slow binding inhibition of PcPAL by (R)-APEP and (S)-APPA ....................... 68
      4.2.3 Inhibition of AvPAL and RtPAL by APEP and APPA ..................................... 70
      4.2.4 Binding thermodynamics ................................................................................. 71
  4.3 Ligand binding in MIO enzymes ................................................................................. 77
      4.3.1 Crystal structures of PcPAL .............................................................................. 77
      4.3.2 Substrate egress dynamics in PAL and PAM .................................................... 82
5 Summary .............................................................................................................................. 93
6 References ............................................................................................................................ 96
7 Thesis points ......................................................................................................................... 108
8 Publication list ...................................................................................................................... 109
  8.1 List of my publications giving the foundation of the thesis points .............................. 109
  8.2 List of my other publications ....................................................................................... 110
  8.3 List of conference presentations related to the thesis ................................................. 110
      8.3.1 Oral presentations at conferences ..................................................................... 110
      8.3.2 Poster presentations at conferences .................................................................. 111
Appendix ................................................................................................................................. 113
Acknowledgement

First and foremost, I would like to express my heartfelt appreciation to my supervisors, Prof. László Poppe, Prof. Beáta G. Vértessey, for their continuous support, motivation, enthusiasm, and immense knowledge. Although the doctoral school does not allow joint supervision, they both contributed equally to the realization of this work by teaching me about their independent expertise. I feel honored that I could work as their colleague in various projects.

I will be eternally grateful for Prof. Romas Kazlauskas and Dr. Bryan Jones for all their help and inspiration during our work together. They taught me to be critical, brave and creative in my work and inspired me to be kind and respectful to the work of others. I must thank the Hungarian-American Fulbright Commission for their financial support during my stay at the University of Minnesota.

I would like to offer my special thanks to all my collaborators: I want to thank Prof. Friedrich Hammershmidt and Dr. Renzhe Qian from the University of Vienna for synthesizing and sending the aminophosphonic acid inhibitors. I am grateful to Prof. Csaba Paizs, Dr. László Csaba Bencze and Dr. Andrea Varga from the Babes-Bolyai University of Cluj-Napoca for their ongoing collaborative work on MIO enzymes. I want to thank Dr. Gergely Nagy for his kind assistance with the isothermal titration calorimetry measurements. I am indebted to Evelin Bell for writing the manuscript in the Élet és Tudomány journal with me. I also want to thank Ibolya Leveles for her advice and assistance in the x-ray crystallographic studies. I owe a lot to all my colleagues for their friendly advice and for creating a supportive environment in the Bioorganic research group and in the Genom Metabolism research group as well, especially to Dr. Kinga Nyíri who was not only a friend and inspiration but also a supportive colleague through these years.

I also wish to thank Lilla Vida, Bence Molnár, and Erzsébet Madaras-Koncz, former and present students, who contributed remarkably to these projects and their constant enthusiasm and diligent work was truly motivating.

I must thank Prof. György Keglevics, the Head of the Organic Chemistry and Technology Department for accepting me to be a Ph. D. student at the department.

I am grateful to Prof. László Nyulászi, the Head of the George Oláh Ph. D. School for supporting my Ph. D. studies.
I would like to thank Prof. László Buday director of the Institute of Enzymology, Prof. György Keserű and Prof. György Pokol former and current directors of the MTA Research Centre for Natural Sciences for allowing me to accomplish my Ph. D. research at the Institute.

I wish to gratefully acknowledge all the funding that I received throughout my Ph. D.: COST grant for a Short Term Scientific Mission in Trieste, registration grant from Biocat2016, travel grants from the Hungarian Academy of Engineering, the Hungarian Chemical Society; the EMBO Travel grant; the FEBS Youth Travel Fund, and the Biophysical Society for the participation at their annual meeting in 2018. I am grateful for the New National Excellence Program (ÚNKP) Pre-Doctoral fellowship of the Ministry of Human Capacities and the Varga József Pre-Doctoral fellowship award for supporting my Ph. D. studies and research.

Finally, I could fill another book with the names and stories of the kind hearted and helpful people that I’ve met, who contributed in their own ways to this work by inspiring me during the five years of my Ph. D.
Abbreviations

ACN         acetone cyanohydrin
AIP         (2-amino-2,3-dihydro-1H-inden-2-yl)phosphonic acid
AL          ammonia-lyase
AM          2,3-aminomutase
AMPP        (1-amino-3-mercaptopropyl)phosphonic acid
AOPP        (S)-2-aminooxy-3-phenylpropanoic acid
APCP        (amino(1-phenylcyclopropyl)methyl)phosphonic acid
APEP        (1-amino-2-phenylethyl)phosphonic acid
APMP        (amino(phenyl)methyl)phosphonic acid
APPA        (1-amino-2-phenylallyl)phosphonic acid
AtHNL       Arabidopsis thaliana HNL
AtPAL       Arabidopsis thaliana PAL
AvPAL       Anabaena variabilis PAL
βME         β-merkaptopoethanol
BmHNL       Baliospermum montanum HNL
CAC         catalytically active conformations
COM         center of mass
CPD         serine carboxypeptidase
DtHNL       Davallia tyermannii HNL
DTT         dithiothreitol
E. coli     Escherichia coli
EDTA        ethylenediaminetetraacetic acid
EST         esterase
ESTHER      ESTerases and alpha/beta-Hydrolase Enzymes and Relatives database
            (http://bioweb.supagro.inra.fr/ESTHER/general?what=index)
FAD         flavin adenine dinucleotide
FePAL       Fagopyrum esculentum PAL
F-OR        FAD dependent oxidoreductase family
GtHNL       Granulicella tundricola HNL
HAL         histidine ammonia-lyase
HbHNL       Hevea brasiliensis HNL
HEPES       2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HNL         hydroxynitrile lyase
InterPro     protein sequence analysis & classification secondary database
            (https://www.ebi.ac.uk/interpro)
IPTG        isopropyl-β-D-thiogalactoside
ITC         isothermal titration calorimetry
KkPAL       Kangiella koreensis PAL
LuHNL       Linum usitatissimum HNL
MD          Molecular Dynamics
MeHNL       Manihot esculenta HNL
MIO         4-methylideneimidazole-5-one
MNN         mandelonitrile
NpPAL       Nostoc punctiforme PAL
OsTAM       Oryza sativa TAM
PaHNL       Prunus amygdalus HNL
PaPAM       Pantoea agglomerans PAM
Pal: phenylalanine ammonia-lyase
Pam: phenylalanine 2,3-aminomutase
PbPam: Planctomyces brasiliensis PAL
PcPam: Petroselinum crispum PAL
Pcr: polymerase chain reaction
Pdb: Protein Data Bank (www.rcsb.org)
Peg: polyethylene glycol (the number after the PEG refers to its average molecular weight)
Phe: phenylalanine
PpPam: Pseudomonas putida HAL
Ramd: random acceleration molecular dynamics
Rmsd: root-mean-square deviation
Rmsf: root-mean-square fluctuation
RsEst: Rauvolfia serpentine polyneuridine-aldehyde esterase
RsTal: Rhodobacter sphaeroides TAL
RtPam: Rhodosporidium toruloides PAL
Sapb2: salicylic acid binding protein 2 from tobacco
SbHnl: Sorghum bicolor HNL
SbPam: Sorghum bicolor PAL
Sds-page: sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sgtal: Streptomyces globisporus TAM
Tal: tyrosine ammonia-lyase
Tam: tyrosine 2,3-aminomutase
Tca: trans-cinnamic acid
Tcep: tris(2-carboxyethyl)phosphine
TcPam: Taxus canadensis PAM
Tev: tobacco etch virus protease
Tris: 2-amino-2-(hydroxymethyl)propane-1,3-diol
Twpam: Taxus wallichiana var. chinensis PAM
Wt: wild-type: used for the reference protein sequence of an organism found in the UniProt database and in other scientific studies
Zn2+-Adh: Zn2+ dependent alcohol dehydrogenase family
# 1 Introduction

During the past two decades, scientific and technological advances have established biocatalysis as an alternative to traditional metallo- and organocatalysis in chemical synthesis on laboratory scale and in the chemical industry, leading to the development of greener, more efficient and more sustainable synthetic processes\(^1\). This impact has been especially prominent in the preparation of pharmaceuticals, agrochemicals and fine chemicals. In 2012, three waves of biocatalysis were discussed\(^1\), however as Figure 1 shows, the fourth wave of biocatalysis is forecasted already\(^2,3\).

![Four waves of biocatalysis.](image_url)

During the first wave of biocatalysis, started more than hundred years ago, scientists recognized that components of living cells could be applied to useful chemical transformations. The main challenge for these applications is the limited stability of the biocatalyst and the large number of side reactions. The second wave, emerging around the 1980s, took advantage of gene technology, which enabled the cloning and expression of the enzyme of interest in a suitable microbial host. Most importantly, researchers could subject the enzyme to site-directed mutagenesis to improve its properties and to enhance biotransformations of non-natural substrates. The third wave started with the work of Pim Stemmer and Frances Arnold in the mid and late 1990s, when they developed methods capable of rapidly and extensively modifying biocatalysts via an \textit{in vitro} version of Darwinian evolution. Techniques included DNA shuffling and error-prone polymerase chain reaction in combination with high-throughput screening methods. “For the directed evolution of enzymes” Frances Arnold received the Nobel Prize in Chemistry in 2018. A recent study compared the efficiency of five different protein engineering methods, and concluded that mutations to the consensus sequence of many homologous proteins provided the balance of success rate, degree of stabilization, and ease of implementation\(^4\).
The fourth wave of biocatalysis, in development today, includes emerging enzyme classes, novel non-natural reactions and metabolically engineered strains that catalyze a whole cascade of reactions within one cell. Learning more details about the structure function relationships in proteins is fundamental for the development of novel biocatalytic processes and substantially facilitates rational enzyme design. Results presented in this work constitute a small part of the effort to understand how protein structure defines substrate binding and how the protein environment lowers the transition state energy barriers hence catalyze chemical transformations.

Reactions catalyzed by enzymes often do not have a counterpart in traditional synthetic organic chemistry. Enzymes with a long shelf life and good activity and stability in organic solvents should help biocatalysis to spread further into laboratory applications and industrial uses. Recently, the role of biocatalysis in modern retrosynthetic analysis has been reviewed in the hope “to stimulate new ways of teaching and thinking about the science of chemical synthesis.” On the long term, training chemists in both biocatalysis and chemocatalysis will help them choose the best solution for their synthetic aim and see enantioselective chemo- and biocatalysis as “Partners in Retrosynthesis”.

Biocatalysis provides an attractive tool for the one step synthesis of enantiomerically pure molecules. However, there are only a few exceptional cases when enzymes catalyzing the formation of both enantiomers occur naturally. Protein structure determines the function and the catalytic mechanism as well as the enantioselectivity, as proteins are made up of chiral amino acid building blocks. The goal of this work was to investigate the molecular mechanisms leading to highly enantioselective enzymes, in two selected cases: hydroxynitrile lyases (HNLs) and phenylalanine ammonia-lyases (PALs). Detailed understanding of these molecular mechanisms will enable more focused rational engineering designs for enhancing or altering enzyme enantioselectivities.

1.1 Hydroxynitrile lyases

Hydroxynitrile lyases (HNLs, EC 4.1.2.X, X= 10, 11, 46, 47) are defense enzymes that catalyze the elimination of hydrogen cyanide from cyanohydrins. The released hydrogen cyanide kills or deters the predators. It is estimated, that 3,000 plant species, from a wide variety of families, utilize cyanogenic defense. For synthetic applications, chemists carry out the reverse reaction, an enantioselective addition of hydrogen cyanide to carbonyl compounds, as shown on Figure 2. A HNL catalyzed reaction; the emulsin (bitter almond extract) catalyzed
generation of HCN from glycoside amygdalin, was one of the very first biocatalytic transformations ever reported in 1837\textsuperscript{7}. The first enantioselective synthesis, carried out by Rosenthaler in 1908, was the emulsin-catalyzed enantioselective addition of HCN to benzaldehyde to prepare optically pure mandelonitrile\textsuperscript{8}.

1.1.1 Plethora of compounds reachable by HNL catalyzed reactions

The nucleophilic attack of the cyanide on the carbonyl is a base-catalyzed equilibrium reaction. Chemical catalysis as well as biocatalysis may play important roles depending on the functional groups connected to the reacting center. The reaction is not stereoselective, hence it needs to be suppressed in order to ensure that only the enantioselective, HNL-catalyzed reaction takes place. Lowering the pH below 5, lowering the reaction temperatures, changing the thermodynamic reaction conditions (i.e. an excess of HCN or \textit{in situ} product removal) or applying organic solvents can influence the yield and the enantiomeric excess positively\textsuperscript{6,9}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Gamut of compounds producible by HNL catalyzed and follow up reactions.}
\end{figure}

Cyanohydrins synthesized using HNLs offer a broad range of possible follow-up reactions to give a wide spectrum of products, including hydroxy acids, primary and secondary $\beta$-hydroxy amines, azidonitriles, halogenonitriles, hydroxy aldehydes or ketones, hydroxyl carboxylic acids and esters, hydroxyl amino acids, acyl cyanides, and aminonitriles as shown on Figure 2\textsuperscript{6}. The follow up reactions may involve the nitrile group, the hydroxyl group or the carbon center. The nitrile group in cyanohydrins can undergo hydrolysis, solvolysis, and reduction. O-protection of the hydroxyl can suppress or avoid instability, degradation and
racemization. Additionally, the hydroxyl group can be used for the inversion of configuration. Inversion of the carbon center also provides a range of opportunities for follow-up reactions for the cyanohydrins.

1.1.2 HNL activity appeared in at least six different enzyme folds by convergent evolution

HNLs occur mainly in plants, but bacteria and arthropods also contain HNLs. Most HNLs are involved in the biodegradation of cyanogenic glycosides, like linamarin, as part of the aldoxime-nitrile pathway. The cyanogenic glycoside degradation starts with hydrolysis by glycosidases to form the aglycons, hydroxynitriles, or cyanohydrins. The unstable aglycons are then cleaved into HCN and the corresponding carbonyl component by HNLs. Mandelonitrile and acetone cyanohydrine have been identified as the natural substrates of HNLs (Figure 3 a.).

**Figure 3. Illustration of the structural diversity of HNLs.** a. HNL catalyze cyanide elimination from two natural substrates, the chiral mandelonitrile and the smaller achiral acetone cyanohydrin. b-f. Crystal structures of HNLs from entirely different protein folds.

HNLs are an example of non-homologous isofunctional enzymes, a group of unrelated proteins that catalyze the same chemical reaction as a result of convergent evolution. HNLs have independently evolved at least in six different protein folds; FAD dependent oxidoreductase (F-OR) family (Figure 3 b.); Zn\(^{2+}\) dependent alcohol dehydrogenase (Zn\(^2+\)-
ADH) family\textsuperscript{12}, Mn\textsuperscript{2+} dependent cupin family (Figure 3 c.) \textsuperscript{13}, Bet v1-like protein fold (Figure 3 d.) \textsuperscript{14}, carboxypeptidase family (Figure 3 e.) \textsuperscript{15}, \(\alpha/\beta\)-hydrolase family (Figure 3 f.)\textsuperscript{16,17}. This suggests that catalysis of the cyanohydrin cleavage is relatively easy; there are many possibilities for decreasing the activation free energy of the transition state using the protein toolbox. Table 2 provides a comparative summary of selected enzymes of each different fold. A number of HNLs, from the FAD dependent oxidoreductase type, the Zn\textsuperscript{2+}-ADH type and the \(\alpha/\beta\)-hydrolase type are available commercially from companies such as AppliChem and Jülich Fine Chemicals (Codexis)\textsuperscript{6}.

Table 2. Overview of HNL activity in different enzyme folds

<table>
<thead>
<tr>
<th>Protein fold</th>
<th>F-OR (\rightarrow) Zn\textsuperscript{2+} ADHs</th>
<th>Cupin</th>
<th>Bet v1</th>
<th>Peptidase (\rightarrow) (\alpha/\beta)-hydrolase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>PaHNL, LuHNL</td>
<td>GtHNL</td>
<td>DtHNL</td>
<td>SbHNL, HbHNL, AtHNL</td>
</tr>
<tr>
<td>Selectivity</td>
<td>(R)</td>
<td>(R)</td>
<td>(R)</td>
<td>(S)</td>
</tr>
<tr>
<td>Natural substrate\textsuperscript{a}</td>
<td>MNN, CNH</td>
<td>MNN</td>
<td>MNN</td>
<td>MNN, CNH</td>
</tr>
<tr>
<td>(K_m) [mM]</td>
<td>0.29, 2.5</td>
<td>0.3-0.7</td>
<td>0.8</td>
<td>(0.7, 1.4)</td>
</tr>
<tr>
<td>(k_{cat}) [1/s]</td>
<td>(~66, \sim35)</td>
<td>144-356</td>
<td>(~206)</td>
<td>10, 2</td>
</tr>
<tr>
<td>Oligomerisation</td>
<td>Monomer, Dimer, Tetramer, Dimer</td>
<td></td>
<td></td>
<td>Monomer, Dimer, Dimer</td>
</tr>
<tr>
<td>MW [kDa]</td>
<td>61, 43, 14, 20</td>
<td>(\alpha-33)</td>
<td>(\beta-20)</td>
<td>29, 29</td>
</tr>
<tr>
<td>PDB ID</td>
<td>3GDP, 4BIF, 5E46, 1GXS, 1YB6, 3DQZ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution [Å]</td>
<td>1.57, -</td>
<td>2.46</td>
<td>2.3</td>
<td>2.50</td>
</tr>
<tr>
<td>Reference</td>
<td>6, 18, 6, 12</td>
<td>13, 14</td>
<td>15.19</td>
<td>20, 21, 22</td>
</tr>
</tbody>
</table>

\textsuperscript{a}MNN: mandelonitrile, CNH: Acetone cyanohydrin

Catalysts in emulsin are FAD dependent HNLs, like almond HNL (\textit{Prunus amygdalus}, PaHNL). Interestingly, in this class of HNLs the FAD cofactor does not directly participate in the \((R)\)-selective cyanohydrin synthesis. FAD is suggested to be merely an evolutionary remainder from the oxidoreductase precursor, and to be only necessary for the structural integrity of FAD-HNLs\textsuperscript{18}. Typically \textit{Prunus} species encode multiple \((R)\)-selective FAD dependent HNL isoenzymes, like black cherry (\textit{Prunus serotina}), which encodes five functionally similar copies\textsuperscript{23}.

HNL from flax (\textit{Linum usitatissimum}, LuHNL) was the first characterized HNL from the Zn\textsuperscript{2+}-ADH family. The enzyme binds 2-4 equivalents of Zn\textsuperscript{2+} ions, that is also necessary for its activity, suggesting that the cyanohydrin cleavage reaction proceeds via similar mechanism as
the proton removal from alcohols in the related ADH enzymes. Naturally occurring substrates of this class of HNLs are the aliphatic acetone cyanohydrin and butan-2-one cyanohydrin, and the enzymes showed very little activity toward aromatic aldehydes\textsuperscript{12}.

The only few known bacterial HNLs belong to the cupin family of proteins. Cupins are metal-binding proteins, in which the metal is usually involved in the enzymatic reaction either directly in the reaction mechanism or at least via an interaction with the substrate. Cupins bind divalent metals, such as iron, zinc, cadmium or manganese. HNL from acidobacterium \textit{Granulicella tundricola} (GtHNL) showed a Mn\textsuperscript{2+} dependent activity \textit{in vitro}. However, its natural metal cofactor might differ \textit{in vivo} due to tight regulation of metal homeostasis. GtHNL showed (\textit{R})-selectivity for mandelonitrile, although a conclusive report of its exact natural substrate and biological role is still missing\textsuperscript{13}.

Four isoenzymes from white rabbit’s foot fern (\textit{Davallia tyermannii} (DtHNL)) showed remarkably high (\textit{R})-selective HNL activities for mandelonitrile. This is the natural cyanohydrin identified in the genus \textit{Davallia} and different other fern genera. Sequence analysis and structure determination referred DtHNL to the Bet v1 superfamily, however the most similar characterized protein, a lachrymatory factor synthase from onion, showed less than 25\% sequence identity\textsuperscript{14}.

\textit{Sorghum bicolor} HNL (SbHNL) shows an overall $\alpha/\beta$-hydrolase fold, but its amino acid sequence presents insignificant similarities to the sequence of any other known HNL of the $\alpha/\beta$-hydrolase fold. In contrast, SbHNL sequence is 73\% similar to that of wheat serine carboxypeptidase (CPD), hence this HNL is grouped to the CPD family. The SbHNL sequence reveals a two-amino acid deletion immediately adjacent to the catalytic Ser158, hindering the positioning of the nucleophilic residue in a sharp turn and thus catalysis by the canonical serine protease mechanism\textsuperscript{15}. Additionally the crystal structure revealed that even the active site is located in a cavity close to the protein surface, that is different from the active site cavity of other CPDs or $\alpha/\beta$-hydrolase fold HNLs, explaining the limited substrate specificity of the enzyme\textsuperscript{15}.

HNL activity developed multiple times within the $\alpha/\beta$-hydrolase fold that is most similar to esterases. Rubber tree (\textit{Hevea brasilensis}, HbHNL) encodes an (\textit{S})-selective HNL\textsuperscript{20}, while mouse-ear cress (\textit{Arabidopsis thaliana} AtHNL) encodes an (\textit{R})-selective HNL\textsuperscript{22}. HbHNL, like most HNLs, is involved in the biodegradation of cyanogenic glycosides\textsuperscript{10}. Other $\alpha/\beta$-hydrolase fold (\textit{S})-selective HNLs are known, (e.g. cassava, \textit{Manihot esculenta}\textsuperscript{24} and wild castor, \textit{Baliospermum montanum}\textsuperscript{25} encode (\textit{S})-HNLs, MeHNL, BmHNL), but AtHNL is the only $\alpha/\beta$-hydrolase fold (\textit{R})-selective HNL known to date. Cruciferous plants, which include
Arabidopsis, lack the metabolic pathway to make cyanogenic glucosides\textsuperscript{26}. The AtHNL gene is transcribed upon leaf damage or senescence\textsuperscript{27}, so the potential role of AtHNL is plant defense, but its exact role is unknown. AtHNL is a good hydroxynitrile lyase with aromatic substrates like mandelonitrile\textsuperscript{28–31}, but does not cleave small aliphatic cyanohydrins. Arabidopsis thaliana does not produce mandelonitrile or acetone cyanohydrin\textsuperscript{26}, but does produce indole cyanohydrin\textsuperscript{32}. AtHNL’s role may be to release hydrogen cyanide by lysis of indole cyanohydrin or by hydrolysis of subsequent metabolites like carbonyl nitriles. High sequence similarity (73\%) demonstrates that HbHNL and AtHNL are evolutionarily closely related, despite the opposite enantioselectivities and differing biological functions.

This short summary of the different classes of naturally occurring HNL enzymes illustrates that cyanohydrin cleavage is a chemically simple task; only a catalytic base and a cyanide-stabilizing residue is required. The cyanide elimination reaction proceeds at 2.4 *10\textsuperscript{-5} l/s at pH 5.0 in the absence of enzymes as well\textsuperscript{28}, and the rate of the chemical elimination can be enhanced greatly by base catalysis. In the HNL reactions the protein environment mostly serves as a scaffold for enantioselectivity. Therefore, it is intriguing how the same α/β-hydrolase protein fold can catalyze the same cyanohydrin cleavage reaction both (R)- and (S)-selectively.

1.1.3 HNL reaction mechanisms in the α/β-hydrolase fold

HNLs in the α/β-hydrolase fold superfamily evolved at least twice and use at least three different mechanisms to catalyze hydroxynitrile lyase cleavage. These elimination mechanisms differ in the identity and location of the cyanide-stabilizing residue and of the catalytic base. First, and most firmly established, is the mechanism for HNL from rubber tree (HbHNL). A combination of x-ray structure analysis, mutagenesis and kinetics\textsuperscript{33,34} revealed that, although the active site contains an esterase-like catalytic triad of Ser-His-Asp, the role of the serine differs. In esterases, the catalytic serine is a nucleophile, but in HbHNL, it transfers protons between the substrate and catalytic histidine, Figure 4. a. The positive charge of Nε of Lys 236 stabilizes the negative charge on the leaving cyanide, thus it is called Lys mechanism. HbHNL catalyzes efficient cleavage of both its natural substrate the achiral acetone cyanohydrin and of aromatic cyanohydrins like mandelonitrile, where it favors the (S)-enantiomer.

HNL from mouse-ear cress (AtHNL) also evolved from esterases, but its mechanism differs from that of HbHNL. Instead of Thr11 and Lys236, which are essential for the HbHNL
mechanism, AtHNL contains Asn11 and Met236\(^*\). Met236 also occurs in esterases, but Asn11 does not. Docking\(^{22}\) and QM/MM\(^{35}\) studies suggest the catalytic histidine serves as the base as in HbHNL, but the oxyanion hole is the cyanide stabilizing group, thus it is called oxyanion hole mechanism, Figure 4. b. Two main chain N–H groups of the oxyanion hole donate hydrogen bonds to the cyanide. One consequence of this different mechanism is AtHNL’s reduced activity towards acetone cyanohydrin, while its activity remains comparable to HbHNL for mandelonitrile. Another consequence is AtHNL’s opposite enantioselectivity, AtHNL favors the (R)-enantiomer.

Substrate binding modes and the protein sequences of HbHNL and AtHNL are highly similar, nonetheless their enantioselectivities are opposite. Additionally, both of them are closely related to esterases, however only AtHNL shows residual esterase activity. Evolutionary relationships in the rich family of esterases and HNLs in the α/β-hydrolase family could reveal the emergence of the new functions.

![Figure 4. Different catalytically competent orientations of mandelonitrile stabilize the leaving cyanide group differently and favor different enantiomers.](image)

\(a\). In the classical mechanism for HbHNL (Lys mechanism), the ε-ammonium group of Lys236 stabilizes the leaving cyanide. This enzyme favors the (S)-enantiomer. \(b\). In the mechanism for AtHNL, two main chain N–H’s from the oxyanion hole stabilize the leaving cyanide (oxyanion hole mechanism). This enzyme favors the (R)-enantiomer. This enzyme cannot use the classical mechanism because it lacks Lys236. \(c\). Different position of the catalytic Ser158 alters the substrate binding in the active site of SbHNL, due to a deletion immediately after the catalytic Ser158. In the SbHNL mechanism the carboxyl terminal Trp270 abstracts the proton from the hydroxyl group of the substrate. This enzyme favors the (S)-enantiomer.

The third mechanism in the α/β-hydrolase-fold superfamily evolved from carboxypeptidases and occurs in HNL from sorghum, SbHNL\(^{15}\). The amino acid sequence of this HNL is most closely related to serine carboxypeptidases in the α/β-hydrolase-fold superfamily, so this HNL likely evolved independently of HbHNL and AtHNL. SbHNL

\(^*\) The N-terminal end of AtHNL is one residue longer than HbHNL. Residue numbers from HbHNL are used throughout the text for clarity.
contains a dipeptide deletion immediately following the catalytic Ser158 (equivalent to Ser80 in HbHNL). The deletion positions the serine 9 Å away from its putative interaction partner His414 (His235 in HbHNL), rendering both HbHNL and AtHNL mechanisms impossible. In the SbHNL mechanism the carboxylate of the C-terminal tryptophan residue is the proposed base, Figure 4. c. The functional groups that stabilize the cyanide ion have not been identified. This SbHNL mechanism will not be further considered, due to the large differences in protein sequence and substrate binding mode from HbHNL or AtHNL.

1.1.4 Evolutionary relationships of α/β-hydrolase fold HNLs

The InterPro database entry of α/β-hydrolase domain, IPR000073, matches about 440 000 proteins to this fold. Despite having the same protein fold and the same core catalytic machinery, these enzymes diverged to catalyze at least 17 different reaction types. They function as hydrolases, lyases, transferases, hormone precursors or transporters, chaperones or routers of other proteins. The ESTHER database, a dedicated database for classification and analysis of the α/β-hydrolase fold, subdivide these proteins into 47 superfamilies and 148 families based on sequence similarities and biological information. The α/β-hydrolase fold HNLs cluster within a larger group of plant esterases (ESTs), suggesting that they diverged from esterases approximately 100 million years ago when flowering plants (angiosperms) and insects diversified. Presumably, a weak promiscuous ability to cleave cyanohydrins first arose serendipitously in an ancient esterase. The new function could have provided some protection against herbivorous insects, therefore subsequent selection for higher activity favored gene duplication, freeing the new copy to functionally diverge into the modern specialized hydroxynitrile lyases.

Such divergent evolution of new enzymes involves catalytically promiscuous intermediates that catalyze both the old and new catalytic activities. Subsequent optimization of separate genes encoding these intermediates yields specialist enzymes for each catalytic activity. The puzzle regarding the evolution of HNLs from ESTs is that the EST and modern (S)-selective HNL (the Lys mechanism of HbHNL) mechanisms are incompatible. Modern ESTs do not show promiscuous lyase activity and modern HNLs do not show promiscuous EST activity. In the first attempt to understand the differences between EST and (S)-HNL functions, site-directed mutagenesis exchanging key active site residues yielded intermediates with both EST and HNL functions, but with low catalytic activity. Two molecular reasons for the incompatibility are that ESTs require access to the oxyanion hole and an uncharged alcohol binding site, while (S)-HNLs block the oxyanion hole and a key Lys converts the uncharged
alcohol binding site into a charged cyanide binding site. While both enzymes contain a Ser-His-Asp catalytic triad, (S)-HNLs also contain the key residues Thr11 and Lys236.

More recently, ancestral enzyme reconstruction identified the last common ancestors of (R)- and (S)-HNLs and closely related plant esterases (Figure 5)\textsuperscript{31}. The (S)-HNLs included in the reconstruction were HbHNL\textsuperscript{20,21}, MeHNL\textsuperscript{24}. The amino acid sequence of (R)-selective AtHNL resembles esterases, it was initially annotated as a methyl esterase EST5 or MES5 (AtEST). AtHNL (AtEST) catalyzes ester hydrolysis as well as the fast cleavage of mandelonitrile\textsuperscript{31}. The esterases included in the reconstruction were salicylic acid binding protein 2 (SABP2) from tobacco (\textit{Nicotiana tabacum})\textsuperscript{43} and polynueirdine-aldehyde esterase from snakeroot (\textit{Rauvolfia serpentina})(RsEST)\textsuperscript{44}. More distantly related α/β-hydrolase-fold enzymes in the ancestral enzyme reconstruction included esterase from \textit{R. communis} (RcEST)\textsuperscript{45}. On the whole, this group of plant esterases and hydroxynitrile lyases share >40% amino acid identity, making it feasible to infer ancestral enzyme sequences, for four putative ancestral proteins; EST1, EST2, EST3 and HNL1 that are at the divergence points for different protein functions (Figure 5)\textsuperscript{31}. Several of the reconstructed ancestors proved to be catalytically promiscuous\textsuperscript{30,31}. HNL1 catalyzed mainly hydroxynitrile cleavage, but also showed low remaining esterase activity. The active site of HNL1 is similar to modern (S)-HNLs with Thr11 and Lys236 as key residues, like HbHNL. More interestingly the older ancestor EST3mL* showed both esterase and HNL activity. EST3mL was 58% identical to HbHNL and 78% identical to AtHNL, including active site residues such as Met236 and Asn11. Similar to AtHNL, this ancestor was also (R)-selective, but only weakly (E = 6.7).

\textbf{Figure 5. Simplified phylogenetic tree of esterases and HNLs identifies different types of ancestral enzymes.} Ancestral enzymes EST1 and EST2 are conserved function enzymes. Those lie along a path from ancient to modern esterases and are expected to be esterases. Ancestral enzyme EST3 is a functional branch point enzyme. It lies between ancient esterases and modern HNL’s and is expected to be an esterase with promiscuous HNL activity. Ancestral enzyme HNL1 is a transitional ancestral enzyme because it leads only to modern HNLs. This is expected to be an HNL, but may show remaining esterase activity.\textsuperscript{31}

* EST3 have been reconstructed by more than one method. EST3mL refers to the reconstruction by the maximum likelihood method.
1.2 MIO-containing class I like lyases

The MIO-containing class I lyase-like enzyme family (MIO enzymes) constitutes two functionally diverse but structurally related enzyme classes, the aromatic amino acid ammonia-lysases (ALs) and the aromatic amino acid 2,3-aminomutases (AMs), as shown on Figure 6. ALs catalyze the reversible ammonia elimination from histidine (HAL, EC 4.3.1.3), phenylalanine (PAL, EC 4.3.1.24) or tyrosine (TAL, EC 4.3.1.23). AMs catalyze the interconversion between α- and β-phenylalanine (PAM, EC 5.4.3.10, 11) or α- and β-tyrosine (TAM EC 5.4.3.6). For their natural substrates AL and AM are highly selective for α-arylanalines, they only cleave or produce L-aromatic amino acids ((S)-enantiomer). The PAM reaction for β-phenylalanine is also highly enantioselective, PAMs from eukaryotic origin usually produce (R)-β-phenylalanine, while PAMs from prokaryotic origin usually produce (S)-β-phenylalanine.

**Figure 6. Reactions catalyzed by MIO enzymes.** ALs catalyze the non-oxidative ammonia elimination from (S)-α-arylanalines; L-histidine, L-phenylalanine or L-tyrosine. AMs catalyze conversion of (S)-α-arylanalines; L-phenylalanine or L-tyrosine to (R) or, (S)-β-arylanalines. Phenylalanine 2,3-aminomutases (PAM) produce (R) or (S)-β-phenylalanine selectively, while tyrosine 2,3-aminomutases (TAM) produce a mixture of (R)- and (S)-β-tyrosine.

A special electrophilic residue 4-methylideneimidazole-5-one (MIO) in the protein active site catalyzes the AL and AM reactions. This residue forms autocatalytically and post-translationally usually from Ala-Ser-Gly, as shown on Figure 7. In a few cases the first residue of the MIO forming triad is Thr, Ser or Cys instead of Ala. MIO is a unique catalytic electrophile in nature. A similar imidazole ring based chromophore is responsible for the fluorescence of GFP, but the formation chemistry and biological roles of the two modified amino acids differ significantly.

* A fourth aromatic ammonia lyase EC category 4.3.1.25 exists. Enzymes having both phenylalanine and tyrosine ammonia lyase activities belong to this family.
The physiological functions of MIO enzymes are diverse. They play key roles in several metabolic pathways. In most kingdoms of living organisms, HAL plays crucial role in histidine metabolism\textsuperscript{54,55}. In plants PALs catalyze the carbon flow from the shikimate pathway to the phenylpropanoid pathway, leading to an enormous array of secondary metabolites like lignins or flavonoids\textsuperscript{56}. PALs are of key current interest due to this fundamental role of the phenylpropanoid pathway in plant metabolism\textsuperscript{56}. Other MIO enzymes produce secondary metabolites; like the enediyne antitumor antibiotic C-1027 synthetized by \textit{Streptomyces globisporus}\textsuperscript{57,58} or the p-hydroxycinnamic acid cofactor of photoactive yellow protein\textsuperscript{59}.

Application possibilities of MIO enzymes are manifold both in biotechnology and medicine, ranging from synthetic biotransformations to direct human therapy. The native promiscuity and wide substrate scope of numerous MIO enzymes expedite their use as biocatalysts on laboratory as well as on an industrial scale\textsuperscript{46,60}. DSM Pharma Chemicals developed a method based on PAL-catalyzed hydroamination of 2′-chlorocinnamic acid for the production of (S)-2-indolinecarboxylic acid on the ton scale\textsuperscript{61}. The enzyme substitution therapy with PAL for the treatment of phenylketonuria represents a further extension of applications\textsuperscript{62}. In 2018, the FDA approved the first such treatment under the name Palynziq by BioMarin Pharmaceutical Inc.\textsuperscript{63}. Development of such important applications requires a comprehensive understanding of the structure–function relationships of MIO enzymes.

### 1.2.1 Reaction mechanism of MIO enzymes

#### 1.2.1.1 Reaction mechanisms of aromatic ammonia-lyase reactions

Structural similarities of the enzymes, the substrates and the presence of the unique MIO catalytic electrophile in the enzyme’s active site suggest that all AL reactions proceed by the same catalytic mechanism. Three mechanisms have been hypothesized for the enzymatic elimination of ammonia in MIO enzymes, Figure 8.
a. in the N-MIO intermediate, a bond is formed between the amino group of the substrate and in the MIO electrophile; b. in the FC intermediate, a bond is formed between the phenyl ring of the substrate and in the MIO electrophile; c. Single-step transition state mechanism, the MIO is only indirectly involved in catalysis.

The N-MIO intermediate mechanism\cite{64,65}, (Figure 8 a) postulated early on, proposes that the amino group of the substrate forms a covalent bond with the electrophilic MIO residue of the enzyme, thus generating a better leaving group. The subsequent elimination reaction proceeds by either a stepwise; through a carbocation (E1) or through a benzylic carbanion (E1cB) or a concerted (E2) mechanism\cite{46}.

Experiments showed that the pro-S β-proton is removed during the elimination reaction\cite{66}, as shown on Figure 8. Abstraction of the non-acidic β-proton (pKa ~ 40\cite{67,68}) was considered to be difficult by an amino acid base, hence a Friedel-Crafts type (FC) mechanism\cite{69-71} was proposed (Figure 8 b). The FC mechanism involves the addition of electrophilic methylidene group of MIO to C-2 of the substrate’s phenyl ring. As a consequence, the β-hydrogens of the substrate are acidified by the positive charge in the forming σ-complex. Thus, the catalytic base (TyrA) can deprotonate the β-position with concomitant elimination of the protonated amino group as NH₃. At last, the aromaticity is restored with release of MIO.

In both N-MIO and FC mechanisms the catalytic base removing the pro-S proton (Hₛ) from the β-position during the elimination reaction is a tyrosine residue (TyrA on Figure 8)\cite{72}. TyrA is part of the flexible inner loop region covering the active site.
A recent computational study proposed a single-step mechanism for the TAL reaction. This mechanism assumes a single transition state (TS) for the deamination reaction without the formation of a covalent bond between the substrate and the MIO group (Figure 8 c)\textsuperscript{73}. However, this hypothetical mechanism suffers from contradictions to many biochemical evidences such as the importance of MIO\textsuperscript{74,75} or the known reversibility of the reaction\textsuperscript{76}.

Numerous independent evidence supports the \textit{N}-MIO mechanism in different MIO enzymes (Figure 8 a). The first experimental evidence for the \textit{N}-MIO mechanism was the crystal structure of \textit{Rhodobacter sphaeroides} TAL (RsTAL). This crystal structure showed the high affinity inhibitor (2-amino-2,3-dihydro-1H-inden-2-yl)phosphonic acid (AIP) to bind via its amino group, and proved the \textit{N}-MIO mechanism for TALs\textsuperscript{74}. QM/MM studies confirmed the \textit{N}-MIO mechanism for RsTAL from a theoretical aspect as well\textsuperscript{72}. Aiming to consolidate the \textit{N}-MIO and FC mechanisms, it was then suggested that TAL/TAM enzymes may use the \textit{N}-MIO mechanism while PAM/PAL enzymes may use the FC mechanism\textsuperscript{68}. However, later experimental results contradict the hypothesis of different mechanisms for PALs and TALs and indicate that all MIO enzymes proceed by the \textit{N}-MIO mechanism. An original study provided evidence to the \textit{N}-MIO mechanism in \textit{Petroselinum crispum} PAL (PcPAL), by showing that PcPAL catalyzed ammonia elimination from a non-aromatic substrate, propargylglycine, which is only possible by the \textit{N}-MIO mechanism\textsuperscript{76}.

Today the \textit{N}-MIO mechanism is mostly established for the first step of the AL reactions, the question remaining is how the \(\beta\)-proton abstraction occurs from the intermediate shown on Figure 8 a. The E1 mechanism involving a carbocation intermediate, is extremely unlikely, due to the high energy of the \(\alpha\)-carbocation intermediate and because it is incompatible with the relatively large positive charge originating from the orientation of helix dipoles, as observed in the crystal structures\textsuperscript{46}. Kinetic analysis\textsuperscript{65} suggests that ammonia elimination from the covalent intermediate proceeds by a stepwise process via a carbanion intermediate (E1cB mechanism). Assuming that the elimination step is rate-limiting the same experiments excluded the E2 mechanism\textsuperscript{65}. A computational study on the mechanism of RsTAL indicated that this may not be the case\textsuperscript{72}. Calculations suggested that such proton transfer between the amino group of the substrate, TyrB (Figure 8) and the MIO may be the rate-determining step of the reaction, possibly justifying why no significant 2H or 15N isotope effects could be observed experimentally.
1.2.1.2 Reaction mechanisms of aromatic aminomutase reactions

All hypotheses on mechanisms for the AM reactions postulate the existence of the unsaturated carboxylic acid intermediate\(^{46}\) (e.g. \(E\)-cinnamate (TCA) intermediate for PAMs, Figure 9). A common feature of all AM reactions is the release of this unsaturated carboxylic acid intermediate as by-product\(^{77,78}\), assumedly as a result of the intermediate egress from the active site prior to re-amination. Accordingly, the first step of the AM reaction is the AL reaction, and the main difference between AL and AM is that the former ones only catalyze the first deamination step of the reaction while the later ones catalyze the re-amination of arylalanines in the \(\beta\)-position.

The first deamination step of the AM reaction is suggested to also proceed by the \(N\)-MIO mechanism. First, kinetic isotope effects suggested this mechanism for \((S)\)-\(\beta\) selective PAM\(^{50}\). Then the crystal structure of \((S)\)-\(\beta\) selective *Pantoaea agglomerans* PAM (PaPAM) directly showed \(\alpha\) and \(\beta\)-phenylalanine bound to the MIO catalytic electrophile by their amino group, thus excluding any dilemma on the \(N\)-MIO mechanism of \((S)\)-\(\beta\) selective PAM reactions\(^{79}\).

---

**Figure 9. Suggested reaction mechanisms for AM reactions illustrated on phenylalanine.**

PAM reactions lead to different enantiomers of the \(\beta\)-amino acid via “inversion of configuration” and “retention of configuration” routes. The routes are named after the pro-\(R\) \(\beta\)-hydrogen (in green). During the AM reaction the pro-\(S\) \(\beta\)-hydrogen (in orange) is shifted to the \(\alpha\)-carbon atom.

Different reaction routes in PAM reactions result in mirror image enantiomers of \(\beta\)-phenylalanine (\(\beta\)-Phe) (Figure 9). Eukaryotic PAMs convert \(L\)-\(\alpha\)-phenylalanine (\(L\)-Phe) to
enantiopure (R)-β-Phe via “retention of configuration” route\(^47\) (referring to the pro-R β-hydrogen of the substrate, depicted in green on Figure 9). In contrast, prokaryotic PAMs transform L-Phe to enantiopure (S)-β-Phe via “inversion of configuration” route\(^50\) (Figure 9). Independently from their origin, TAMs exhibited no strict enantiopreference towards either enantiomer of β-tyrosine (β-Tyr)\(^80,81\). The structural and mechanistic similarities between AL and AM raise the question how the structure of various MIO-enzymes modulate the substrate and reaction specificity.

### 1.2.2 Structural determinants of ligand binding and substrate specificity of MIO enzymes

Early biochemical studies showed that MIO enzymes function as homotetramers (Figure 10 a.)\(^82\). The crystal structures confirmed the size exclusion chromatography results and found that burial of large hydrophobic surfaces is probably the driving force of tetramerization. For example, in PcPAL each subunit has a surface of 30 330 Å\(^2\) and buries 29% of it in contacts of 4215 Å\(^2\), 3660 Å\(^2\), and 880 Å\(^2\) with the other three subunits\(^83\). Crystal structures revealed that the tetramers contain four active sites, and that three monomers participate in the creation each of the active sites\(^55\) (Figure 10). MIO enzymes from eukaryotes are approximately typically 200 residues longer at their C-terminus as compared to prokaryotic MIO enzymes. Currently 39 structures are available in the PDB for 11 different MIO enzymes including HAL\(^55,84\), PALs\(^85-87\), TAL\(^74\), PAM\(^79\) and TAM\(^81,88,89\) of prokaryotic origin and PALs\(^62,83,90,91\) and PAMs\(^78,92-94\) from eukaryotes.

The active site of MIO enzyme can be divided into three regions, as shown on Figure 10 b. Hydrogen bond donating residues Gln348*\(^85\), Asn384 and a positively charged residue Arg354 from two different chains form the carboxyl binding region. Additionally, the nearby Phe400 forms a flat floor to the active site. These four residues are conserved in MIO enzymes, with the exception of Gln348, that is Glu in (S)-PAMs. Conserved residues Tyr110 (TyrA of Figure 8), MIO203, Asn260, Tyr 351 (TyrB Figure 8) form the central catalytic region of the active site. Mutational studies in *Pseudomonas putida* HAL (PpHAL)\(^95\), PcPAL\(^75\) and RsTAL\(^96\) confirmed the essential nature of these residues. Residues Phe116, Gly117, Leu134, Phe137, Leu138, Leu206, Leu256, Val259, Lys456, Ile460, Glu484, Gln488 from two adjacent chains form the aromatic binding region of MIO enzymes, lining the binding cavity for the aromatic

---

*Residue numbers from PcPAL are used throughout the text concerning the MIO enzymes.*
side chain of the substrates. These residues are variable and are mostly responsible for the substrate specificity of MIO enzymes.

![PcPAL structure](image)

**Figure 10. Important structural features of PcPAL.** a. Overview of the biologically active tetrameric structure of PcPAL. b. Three regions form the active site of MIO enzymes. Cyan lines show the carboxyl binding residues at the bottom of the active site from this view. Green sticks show the catalytic residues at the central part of the active site. Yellow lines show residues forming the hydrophobic binding pocket. Black, purple and orange colors of the one letter code labels with residue numbers of PcPAL indicate different chains, equivalent to the gray, purple and orange chains of the tetramer. (Image of PcPAL with (S)-APP A bound, PDB 6F6T)

Naturally occurring MIO enzymes are quite specific for their individual substrates. *Rhodobacter capsulatus* TAL, for example, has a 150-fold higher catalytic efficiency for tyrosine compared to phenylalanine, while PcPAL1 showed a 10,000-fold higher catalytic efficiency for phenylalanine than tyrosine, and similar results were observed with *Arabidopsis thaliana* PAL (AtPAL).

Complementary studies identified a critical amino acid in substrate selection in aromatic ammonia-lyases. A single amino acid, Phe137, in the aromatic binding region proved to be a critical position, in switching between PAL and TAL activities. PALs like PcPAL encode Phe at position 137 followed by Leu138, in contrast TALs, like RsTAL encode the hydrogen bond donating His at position 137 followed by Leu138 as well. Mutation H89F (137 in PcPAL) in RsTAL abolished all TAL activity while it increased tenfold the catalytic efficiency for Phe.
The reverse mutation in AtPAL F114H decreased the catalytic efficiency for Phe eightyfold while increasing it for Tyr eighteenfold. HALs contain a signature Ser-His dipeptide at the substrate selection position. Nonetheless, there are a number of exceptions to these observations, like Rhodosporidium toruloides PAL (RtPAL) that encodes His-Gln in the equivalent positions. This non-typical motif forecasts unusual substrate specificity; RtPAL is one of bifunctional PALs that cleave tyrosine and phenylalanine with almost equal efficiency. These bifunctional PALs seem to be common among monocotyledonous plants, such as maize or rice, as well as yeasts. Even a new enzyme class, 4.3.1.25 was created to distinguish the monofunctional PALs and TALs from the bifunctional P/TALs. Recently, a study described an unusual MIO enzyme from Pseudomonas fluorescens (PfXAL), that catalyzed ammonia elimination from histidine, phenylalanine and tyrosine as well.

The synthetic application of MIO enzymes is most importantly based on reverse reaction yielding L-Phe derivatives from achiral precursors or the kinetic resolution of racemic amino acids. HALs only act on a narrow substrate range, besides their natural substrate they only accept L-4-fluorohistidine, L-4-nitrohistidine. However several heteroarylalanines and heteroarylacrylates could interact with HAL as strong inhibitors. PALs and TALs have a much broader substrate scope, which has been reviewed recently. Therefore, only mutations modulating the substrate scope will be discussed here in brief.

Electron rich arylalanines (e.g. MeO, Me, MeS substituted phenylalanine) and bulky arylalanines are poor substrates of wild-type PALs. Enzyme discovery yielded novel PALs with measurable amination activities of electron rich ring-substituted phenylacrylates, and Planctomyces brasiliensis PAL (PbPAL) showed the highest conversions for these substrates initially. Following mutagenesis of the selectivity residues Phe137 and Leu138 in PbPAL aiming to widen the active site, increased conversions of electron rich ring-substituted phenylacrylates. An independent study showed, that enlarging the active site of PcPAL by F137V mutation enabled the elimination reaction of bulky substrates such as styrylalanines. However even this more spacious active site in PcPAL was unable to catalyze ammonia elimination from L-tryptophan. So far, only one report is available in literature concerning a putative tryptophan ammonia lyase (WAL) from Rubrivivax benzoatilyticus, showing low sequence similarities to MIO enzymes. A follow up study showed that an additional mutation to create PcPAL-F137V/I460V enabled the catalysis of the amination reaction of styrylalanines as well. The intriguing finding that different active site geometries (F137V or F137V/I460V)
were required for the stabilization of the transition states of the elimination and addition reactions suggest that the two reactions might be uncoupled.

Based on their natural function ALs are assumed to be selective to the proteinogenic amino acids. The idea was confirmed as recombinant PpHAL was competitively inhibited by D-His \((K_i = 10 \, \text{mM})^{111}\). PAL catalyzed ammonia elimination from L-Phe more than 4500 times faster than from D-Phe\(^99\). The same study found D-Phe to be a competitive inhibitor\(^99\), proving that the D-enantiomer also binds to PAL. A recent study, involving three PALs, suggested that those are able to catalyze the formation of the D-enantiomers of substrates with an electron-deficient aromatic moiety by a slower and MIO-independent pathway\(^112\). The promiscuous activity could also be enhanced by mutation of H389Y/K (H359Y/K in \textit{Anabaena variabilis} PAL, AvPAL) and exploited for the enantioselective synthesis of D- and L-phenylalanine derivatives\(^113\).

**Table 3.** Structure and inhibition constant of the most potent PAL inhibitors.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>inhibition constant (K_i) (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOPP</td>
<td>![AOPP Structure]</td>
<td>(K_i = 0.38 , \text{nM}^{114,115}) PcPAL</td>
</tr>
<tr>
<td>AOPP (Fagopyrum esculentum PAL)</td>
<td>![AOPP (FePAL) Structure]</td>
<td>(K_i = 90 , \text{nM}^{117})</td>
</tr>
<tr>
<td>AIP</td>
<td>![AIP Structure]</td>
<td>(K_i = 25 , \text{nM}^{115}) PcPAL</td>
</tr>
<tr>
<td>(R)-AEP</td>
<td>![AEP Structure]</td>
<td>(K_i = 1.5 , \mu\text{M}^{118}) FePAL</td>
</tr>
<tr>
<td>(R)-AEP (of PcPAL)</td>
<td></td>
<td>(K_i = 1.9 , \mu\text{M}^{115}) of PcPAL</td>
</tr>
<tr>
<td>(R)-AEP (FePAL)</td>
<td></td>
<td>(K_i = 1.95 , \mu\text{M}^{117}) FePAL</td>
</tr>
</tbody>
</table>

\(^*\) Inhibition constant values differ slightly with different PALs: \(K_i = 56 \, \text{nM}^{119}\) \textit{Cistanche deserticola} PAL, \(K_i = 80 \, \text{nM}^{119}\) FePAL, \(K_i = 115 \, \text{nM}^{120}\) \textit{Lactuca sativa} L. PAL, \(K_i = 1.91 \, \mu\text{M}^{121}\) \textit{Streptomyces maritimus} PAL (PAM), \(K_i = 16.3 \, \mu\text{M}\) RsTAL / 0.60 \(\mu\text{M}\) RsTAL H89E\(^74\)

A wide range of aromatic compounds inhibit PALs. Similarities of \(K_i\) value of D-Phe and \(K_m\) value of L-Phe, suggested D-Phe to be a general inhibitor of PALs, by a mirror image
packing competitive binding mode\textsuperscript{122,123}. Interestingly, inhibition by SH group-containing compounds seemed to be a general property of HAL\textsuperscript{111,124} and species specific in PAL\textsuperscript{125}. A wide variety of substrate analogues were studied as inhibitors of PAL \textit{in vitro} and \textit{in vivo} for potential herbicide use\textsuperscript{114–121,126–130}. Exchanging the carboxyl group of Phe to a phosphonic acid group yielded potent competitive inhibitors shown in Table 3, that are transition state analogues or suicide substrates. These inhibitors are either achiral, like AIP or show equivalent configuration as L-Phe, like AOPP. All studies found so far, that the natural like enantiomer showed at least one order of magnitude stronger binding compared to the opposite enantiomer, \textit{e.g.} (R)-APEP that has a $K_i$ of 1.5 $\mu$M of FePAL while (S)-APEP only had a $K_i$ of 11.6 $\mu$M\textsuperscript{118}.

1.2.3 Structural features differentiating AL and AM activity

The active site of MIO enzymes is capped by two flexible loops named the inner and the outer loop. The inner loop sequence is similar within MIO enzymes while the outer loop sequence is more diverse (Table 4). The inner-loop in prokaryotic MIO enzymes is two amino acids longer than that of eukaryotic enzymes. The outer loop is more variable in length, it is typically 22 residues long, but shorter in HAL, and longer in TAL/TAM. Accordingly, inner loop conformations in crystal structures are mostly similar, while outer loop conformations are more diverse, Figure 11.

**Table 4.** Inner and outer loop sequence alignment of MIO containing class I lyase like enzymes with crystal structure available in the PDB.

<table>
<thead>
<tr>
<th>Protein\textsuperscript{*}</th>
<th>Inner loop</th>
<th>Outer loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>PcPAL</td>
<td>GTDSYGVTTGFAGTSHRR--TKQ</td>
<td>GSAYVKA--AQKLHED------DPLKPKDR</td>
</tr>
<tr>
<td>RtPAL</td>
<td>SMVSYGVTGFGGSADTR--TED</td>
<td>GSRFAVHEEVEVK------DDECILKDR</td>
</tr>
<tr>
<td>SbPAL</td>
<td>GGD1YGVTGFGGTSHRR--TKD</td>
<td>GSAMKHA--AKVNEK------DPLKPKDR</td>
</tr>
<tr>
<td>NpPAL</td>
<td>AQP1YGVTSGFAGMADVISRQ</td>
<td>DSLVRE--ELDGKHE------YRDKLIDR</td>
</tr>
<tr>
<td>AvPAL</td>
<td>GEP1YGVTGFGMANVAISRQ</td>
<td>NSQLVRD--ELDGKD------YRDHLEDR</td>
</tr>
<tr>
<td>TcPAM</td>
<td>GAD1YGVTGFGACSSR--TNQ</td>
<td>SSFFQDL--SREYYSI------DKLKKPKDR</td>
</tr>
<tr>
<td>TwPAM</td>
<td>GAD1YGVTGFGACSSR--TNR</td>
<td>SSFFQEL--SREYYSI------DKLKKPKDR</td>
</tr>
<tr>
<td>PaPAM</td>
<td>ERV1YGVNTSMGFVNYIYPIAK</td>
<td>DSSLAVNEHEEKLIAE--EMGGLVKASHQIE</td>
</tr>
<tr>
<td>RsTAL</td>
<td>ARVHGLYTGFPLNRLISGEN</td>
<td>GSARVVRHVAERRLDA------GDGTEFEAQQA</td>
</tr>
<tr>
<td>SgTAM</td>
<td>NIP1YGVTGEGMYIMYQVQKSK</td>
<td>GSGLTVHAFLREELQKDKEAGDVQRSEIYFKA</td>
</tr>
<tr>
<td>PpHAL</td>
<td>DRTAYGINTGFLLASTRIASHD</td>
<td>DSSEVSLSH-------------KNCDELVQDP</td>
</tr>
</tbody>
</table>

\textsuperscript{*}PcPAL: \textit{Petroselinum crispum} PAL P24484, RtPAL: \textit{Rhodospirillum rubrum} PAL P11544, SbPAL: \textit{Sorghum bicolor} PAL CSXST8, NpPAL: \textit{Nostoc punctiforme} PAL B2J528, AvPAL: \textit{Anabaena variabilis} PAL Q3M5Z3, TsPAM: \textit{Taxus canadensis} PAM Q6GZ04, TwPAM: \textit{Taxus wallichiana var. chinesis} PAM Q68G84, PaPAM: \textit{Pantoea agglomerans} PAM Q84FL5, RsTAL: \textit{Rhodobacter sphaeroides} TAL Q3IWB0, SgTAM: \textit{Streptomyces globisporus} TAM Q8GGM0, PpHAL: \textit{Pseudomonas putida} HAL P21310.

\textsuperscript{*}The sequence of \textit{Taxus Canadensis} PAM Q6GZ04 and \textit{Taxus Wallichiana var. chinesis} PAM Q68G84 is 97 % identical. Differences are located far from the active site, therefore structural and functional results can be confidently transferred between the two enzymes.
Both the inner and the outer loops are sequentially distant from the MIO catalytic electrophile, the outer loop folds onto the active site from an adjacent monomer. Large crystallographic B factors suggested high flexibility of these two loops, and due to their spatial proximity to the active site those were implicated in substrate binding and catalysis. The inherent flexibility of the loops blurs their electron density in the crystal structures of certain MIO enzymes. Figure 12 shows the electron densities for one representative example for each type of MIO enzyme. Inner loop residues could not be resolved in structures: 2NYF of Nostoc punctiforme PAL (NpPAL), 2NYN of Anabaena variabilis PAL (AvPAL), 4BAB, 4C6G, 4V2Q, 4V2R of Taxus wallichiana var. chinensis PAM (TwPAM), 1Y2M, 1T6J, 1T6P of Rhodosporidium toruloides PAL (RtPAL) and in 6AT7 of Sorghum bicolor PAL (SbPAL). The other loop residues could not be resolved in structures: 2NYF of NpPAL, 2NYN of AvPAL, 4V2R of TwPAM, 1Y2M, 1T6J, 1T6P of RtPAL, and in 6AT7 SbPAL. Additionally, in a number of structures inner and/or outer loops were modeled to residual electron densities as shown on Figure 12 b for both loops and g, f for only a short section of the outer loop.
**Figure 12.** Electron densities of the inner (green) and outer loop (purple) regions of representative structures of the different classes of MIO enzymes. Electron densities (2Fo-Fc maps) are shown by blue mesh at 1 σ level.  

Protein structures shown on a, c, d, e show clear electron densities for both inner and outer loops. Only residual electron densities are present for inner and outer loops in 1W27 (b) structure for PcPAL. Parts of the outer loop electron densities are absent in 2QVE (f) structure of SgTAM and in 1GKM (g) structure of PpHAL.

**Table 5.** Representative crystal structures of different classes of MIO enzymes.

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB</th>
<th>Resolution</th>
<th>Length</th>
<th>Inner-loop</th>
<th>Outer-loop</th>
<th>Origin</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvPAL</td>
<td>3CZO</td>
<td>2.2</td>
<td>565</td>
<td>74-96</td>
<td>291-313</td>
<td>Prokaryotic</td>
<td>86</td>
</tr>
<tr>
<td>PcPAL</td>
<td>1W27</td>
<td>1.7</td>
<td>719</td>
<td>106-126</td>
<td>328-350</td>
<td>Eukaryotic</td>
<td>83</td>
</tr>
<tr>
<td>PaPAM</td>
<td>3UNV</td>
<td>1.54</td>
<td>547</td>
<td>74-96</td>
<td>299-319</td>
<td>Prokaryotic</td>
<td>79</td>
</tr>
<tr>
<td>TwpAM</td>
<td>2YII</td>
<td>2.18</td>
<td>677</td>
<td>76-96</td>
<td>299-321</td>
<td>Eukaryotic</td>
<td>93</td>
</tr>
<tr>
<td>RsTAL</td>
<td>2O6Y</td>
<td>1.5</td>
<td>523</td>
<td>56-78</td>
<td>271-299</td>
<td>Prokaryotic</td>
<td>74</td>
</tr>
<tr>
<td>SgTAM</td>
<td>2QVE</td>
<td>2</td>
<td>539</td>
<td>59-81</td>
<td>274-321</td>
<td>Prokaryotic</td>
<td>88</td>
</tr>
<tr>
<td>PpHAL</td>
<td>1GKM</td>
<td>1</td>
<td>509</td>
<td>50-72</td>
<td>264-280</td>
<td>Prokaryotic</td>
<td>84</td>
</tr>
</tbody>
</table>

The flexibility of the inner loop was implicated to be the key element differentiating ALs from AMs. Increasing the reaction temperature of the *Pantoea agglomerans* PAM (PaPAM) catalyzed reaction not only increased the reaction rate, but also shifted it towards the elimination reaction from the aminomutase reaction. Reaction rate dependence on the viscosity of the reaction medium suggested conformational changes to be rate limiting for the (S)-PAM reactions, besides the cleavage of the Cβ-H shown by kinetic isotope effects previously.
Experimental results of PaPAM and the crystal structure of PcPAL in the loop-out conformation inspired the “loop flexibility hypothesis”. The hypothesis states, that the inner loop in TcPAM – being more rigid than the one in PcPAL – confines the TCA–amino-enzyme intermediate within the active site during isomerization enabling re-amination of the captured TCA from the Si side of the β-position. Mutation of several inner loop residues in TcPAM to their equivalents in PcPAL resulted in alteration of the AM phenotype to AL. Computational study on the loop flexibility hypothesis assumed large conformational change during the substrate access/product release within the catalytic cycle of TcPAM. During the study, the “loop-in conformation” (Figure 13 a) was opened up to adopt a so-called “loop-out conformation” (similar to the one modeled for the crystal structure of PcPAL, Figure 13 b).

![Image](image-url)

**Figure 13. Inner loop conformations in MIO enzymes.** a. Catalytically competent loop-in conformation of the inner loop in Taxus canadensis PAM (TcPAM) structure (3NZ4). b. Loop-out conformation of the inner loop in Petroselinum crispum PAL (PcPAL) structure (1W27). The catalytic base TyrA points away from the active site, thus unable to fulfill its role.

Although a single structure with “loop-out conformation” was published for PcPAL, the poor fit of the modeled loop to the electron density map (Figure 12) made the existence of this “loop-out conformation” in MIO enzymes questionable. Existence of the “loop-out conformation” in any MIO enzyme does not hold up under thorough scrutiny. Based on relevant interrogation of the PDB database, including electron density data, all selected representatives of each classes of MIO enzymes show well-ordered inner loop conformations in a “loop-in” state (Figure 11 and Figure 12). The only exceptions are eukaryotic PALs, suggesting that the high inner-loop flexibility is specific to eukaryotic PALs (Figure 12). Furthermore, as
soon as two years after publication of the 1W27 structure\textsuperscript{83}, molecular dynamics simulation suggested the loop-out conformation to be a catalytically unproductive conformation\textsuperscript{132}.

Loop flexibility differences, however, can only partially explain the differences between ALs and AMs. There are many accounts of modulating the ammonia lyase activity, increasing either the $\alpha$ or $\beta$ selectivity of PAM or changing the substrate specificity of TAM. A single mutation in TcPAM R122S (R92S in TcPAM sequence) increased 44-fold the reaction rate of $\beta$-Phe deamination and increased amination rates of TCA\textsuperscript{131}. The authors explained the experimental result by changes in the inner loop flexibility due to the loss of a salt bridge between Arg92 and Glu542 (Asp 580 in PcPAL). This explanation presumes that substrate release must be the rate determining step in the reaction, however a precise kinetic study of the PAM reaction is still missing due to the difficulty of product detection and the multi-step nature of the reaction. Mutation of the carboxyl binding residue Glu348 (319 in TcPAM) to Met also resulted in 88% $\beta$-regioselective enzymes compared to the 49% $\beta$ selectivity to wt-TcPAM\textsuperscript{93}. Modification of the selectivity switch residue 137 in Oryza sativa TAM (OsTAM)\textsuperscript{133} from Tyr (125 in OsTAM sequence) to the equivalent Cys in TcPAM resulted in a PAL phenotype instead of the expected PAM phenotype\textsuperscript{134}. Further attempts to create a PAM from OsTAM_Y125C by attempting to increase the rigidity of the inner-loop were unsuccessful\textsuperscript{135}. A captivating study showed that prokaryotic AvPAL demonstrates a promiscuous AM activity at 24-144 h reaction time, converting the already synthetized $\alpha$-amino acids to ($R$)-$\beta$-amino acids in equilibrium reactions\textsuperscript{136}, the authors conclude their results by writing “The difference in selectivity between AvPAL variants and closely related enzymes with aminomutase activity … further highlights the poorly understood mechanisms of mutase–lyase selectivity.”
2 Aims

The three-dimensional structure of enzymes determines their function, their catalytic mechanism as well as their enantioselectivity. The objective of this research was to investigate the structure-function relationship in two selected enzyme cases: hydroxynitrile lyases (HNLs) and phenylalanine ammonia-lyases (PALs).

The goal of the research was to identify the α/β-hydrolase fold ancestral enzyme EST3 mL’s mechanism for HNL cleavage. The hypothesis was that the (R)-selective ancestral enzyme EST3 mL followed the (R)-selective AtHNL mechanism and subsequent optimization switched the mechanism to the (S)-selective HbHNL mechanism. To understand the differences between the two types of HNL mechanisms in HbHNL and AtHNL, we attempted to exchange their catalytic abilities by exchanging active-site residues. The choice of residues to exchange was guided by similarities to ancestral enzyme EST3 mL.

We aimed to answer several different questions related to the enantioselectivity and regioselectivity of MIO enzymes. First, we investigated if aminophosphonic acids synthetized by our collaborators at the University of Vienna (Figure 22) inhibited PALs. When the compounds inhibited PcPAL, we set out to determine the mechanism of inhibition and to compare their kinetic inhibition constants and their thermodynamic binding constants. Strong binding to PcPAL suggested to investigate if the inhibitors are general PAL inhibitors.

Thorough scrutiny of the PDB database suggested that the currently available crystal structure for PcPAL might contain modelling errors. Therefore, we aimed to experimentally determine the catalytically active conformation of PcPAL and investigate how the addition of the strong binding inhibitor influence its strucutre PcPAL. Having strong binding inhibitors from both enantiomers, we aimed to probe, how the active site geometry of PcPAL influences the ligand binding conformation and the enantioselectivity of the enzyme.

The catalytically competent crystal structure showed a loop-in conformation, hence questioned the attractive loop flexibility hypothesis for explaining the differences between AL and AM reactions. Therefore, we carried out molecular modeling simulation to investigate if inner loop dynamics may explain reaction specificity. The same simulations also allowed for investigating whether the substrate access/product release pathways are identical in AL and AM. We continued our studies to see if these release pathways are conserved in all MIO-enzymes. Sequence analysis and structure comparison suggested that substrate release pathways might be closed in prokaryotic (S)-PAMS. Therefore, we aimed to find a candidate PAL for modifying its product release pathways and turn a PAL into a PAM for the first time.
3 Materials and methods

3.1 Cloning and mutagenesis

Expression vectors of PcPAL-H10, AvPAL-H10 and RtPAL-H10 have been created previously. Expression vectors of PcPAL-H6 and KkPAL-H10 were created by molecular cloning. Mutants PcPAL-Y110F were created by mutagenesis. TEV protease encoding gene was a gift from the Protein Research Group of the Institute of Enzymology. All protein encoding genes were verified by DNA sequencing by Microsynth AG and by Sequencing Service of Genomed, using the T7-promoter forward and T7-terminator reverse primers (Appendix Table A1).

3.1.1 Cloning

The gene encoding KkPAL was codon optimized for better expression in E. coli and synthesized. The synthetic KkPAL gene (produced by Genscript) was cloned in pUC57 production vector and later sub-cloned into the pET19b expression vector. Restriction sites for NdeI, Ncol and BamHI were added to the protein coding sequence allowing directional cloning into the expression vector. Primers KkPAL_for and KkPAL_rev were used to amplify the synthetic gene from the pUC57 cloning vector, followed by restriction cloning to the expression vector using NdeI and BamHI enzymes and T4 DNA ligase. The pET19b vector contains an N-terminal His10-tag attached through enterokinase cleavage site to the inserted sequence, facilitating protein purification.

To allow removal of the His-tag, PcPAL was cloned to a modified pET15b-TEV vector that contained an N-terminal His6-tag followed by a Tobacco Etch Virus protease (TEV) cleavage site. The original vector, encoding PcPAL-His10 did not contain suitable restriction sites, therefore Nhel and SacI sites were introduced by a PCR reaction to the PcPAL encoding DNA sequence by PcPAL_for and PcPAL_rev primers (Appendix Table A1). The PCR product was introduced to the target vector by restriction cloning using Nhel and SacI enzymes by the same procedure as described for KkPAL.

3.1.2 Mutagenesis

The Y110F mutant of PcPAL was created by site-directed mutagenesis following the protocol described by Naismith and Liu, using suitable mutagenic primers (Appendix Table A1) from the original PcPAL-H10.
3.2 Protein expression, purification

3.2.1 Protein expression

3.2.1.1 MIO enzymes

All MIO enzymes were produced according to the procedure described previously\textsuperscript{137}, with small changes detailed below. Briefly, protein was produced in \textit{E. coli} BL21 Rosetta (pLysS) cells (Novagen) in Luria broth medium. Cultures were grown until exponential phase (OD\textsubscript{600}~ 0.6-0.8), then protein production was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside (IPTG) at 28 °C for 6 h. Longer induction times (up to 16 h) resulted in approximately the same amount of soluble protein. Increasing or lowering the temperature during the induction period decreased the soluble protein yield. Cells were harvested by centrifugation for 20 min at 8000 g. Although cell pellets may be stored at -80 °C for later purification, the best yields were obtained when samples were processed instantaneously.

3.2.1.2 TEV protease

TEV protease was produced in \textit{E. coli} BL21 Rosetta (pLysS) cells (Novagen) in Luria broth medium. Cells were grown until late exponential phase (OD\textsubscript{600}~ 1.3-1.5), then protein production was induced by the addition of 0.4 mM IPTG at 18 °C for 16 h. Cells were harvested by centrifugation for 20 min at 8000 g, and then frozen to -20 °C for at least 24 h.

3.2.2 Protein purification

3.2.2.1 MIO enzymes

MIO enzymes were purified as described previously\textsuperscript{137}. After harvesting and cell lysis by sonication the supernatant was separated from the cell debris by centrifugation for 45 min at 16000 g at chilled to 4 °C. Next the target protein was separated from the rest of the supernatant by gravitational Ni-NTA chromatography. Proteins with the His6-tag were eluted with 250 mM imidazole solutions while 500 mM imidazole solution was used for the elution of proteins with the His10-tag. Imidazole was removed by 16 h dialysis against 10-fold excess volume of 50 mM 2-amino-2-(hydroxymethyl)propane-1,3-diol (TRIS) pH 8.0 buffer. TEV protease (1:10 molar ratio) was added to the PcPAL-His6 preparations during dialysis to remove the His6-tag. Tris(2-carboxyethyl)phosphine (TCEP) was added to the protein solution as soon as possible after dialysis, in 1 mM final concentration. The uncleaved PcPAL-His6, TEV and the cleaved His6-tag was separated from the target PcPAL by gravitational Ni-NTA chromatography.
Figure 14 shows a representative example of a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) experiment analysing the protein content of the different fractions collected during the PcPAL purification. Kinetic assays ensured that the cloning had no effect on PcPAL activity. Both PcPAL-His10 and the tag-free PcPAL were concentrated to approximately 10 mg/ml with Amincon Ultra-15 Centrifugal Filter units. Aliquots of 50-100 µl of protein solutions were frozen in liquid N₂ and stored at -80 °C until use.

![Figure 14. Representative example SDS-PAGE result showing the protein content of fractions collected during Ni-NTA purification of MIO enzymes. The different lanes in order contain the MW ladder, the supernatant after sonication and centrifugation, the precipitate collected after sonication, the flow through of the column, the salt-wash fractions pooled together, the aspecific binder 25 mM imidazole wash, the 500 mM imidazole elution fraction and the 1 M imidazole column purification fraction.](image)

### 3.2.2.2 TEV protease

Cell pellet from 0.5 L culture containing expressed TEV was suspended in 25 ml 50 mM Na₂HPO₄.2H₂O pH 8.0, 300 mM NaCl, 20 mM imidazole, 2 mM phenylmethane sulfonyl fluoride (PMSF), 1 mM benzamidine (BA), 2 mM β-merkaptoethanol (βME), 2 µg/ml Dnase. The cell suspension was sonicated (5x 60 s with 60 s chilling intervals between) on ice, and centrifuged (20 000 g, 45 min, 4 °C). The supernatant was purified with gravitational Ni-NTA chromatography. The resin was washed with 4 bed volumes of wash 1 (50 mM Na₂HPO₄.2H₂O pH 8.0, 300 mM NaCl, 40 mM imidazole, 2 mM βME) and wash 2 buffers (20 mM TRIS, 1 M NaCl, 20 mM imidazole, 2 mM βME pH 8.0), then the TEV protease was
eluted with 50 mM Na$_2$HPO$_4$.2H$_2$O pH = 8.0, 300 mM NaCl, 400 mM imidazole, 1 mM TCEP. TEV was frozen in 1 ml aliquots to -80 °C before use.

3.2.2.3 Protein quality control and quantification

All protein samples were analyzed by SDS-PAGE. Only protein solutions of at least 90% purity were used in experiments, like lane 7 (500 mM Im) of the SDS-PAGE example shown on Figure 14. Protein concentrations were measured using a NanoDrop 2000 UV-Vis spectrophotometer. The concentration was calculated from the absorbance at 280 nm and the theoretical extinction coefficient of the protein sequence obtained with ProtParam$^{141}$. The extinction coefficients and complete protein sequences are listed in Appendix Table A2.

3.3 Enzyme functional studies

3.3.1 Steady state kinetic measurements

Initial reaction rates of the PcPAL-H10 catalyzed reaction of L-Phe were determined in 100 mM TRIS, pH 8.8 at 30 °C, using 0.01 mg/ml PcPAL by measuring the TCA formation at 290 nm in a UV/Vis spectrophotometer (Specord 200; Analytic Jena AG, Jena, Germany). The extinction coefficient for TCA was determined experimentally by 7-point calibration using TCA solutions in 100 mM TRIS (pH 8.8 at 30 °C). Initial reaction velocities ($v_0$) measured at nine substrate concentrations between 75 µM and 10 mM determined the Michaelis constant ($K_m$) and the turnover number ($k_{cat}$).

Inhibition constants for the aminophosphonic acids APEP and APPA (Figure 22) were determined by measuring the initial reaction velocities at the same substrate concentrations in the presence of variable concentrations of the inhibitor. A regression coefficient ($R^2$) of the linear fit lower than 0.999 for the progress curves of the initial reaction velocity suggested slow binding kinetics. In this case the combination of a linear and an exponential equation was fit to describe the production curves. Steady state velocities were used for determining the inhibition constant ($K_i$) values. All measurements were performed in duplicates. Non-inhibited kinetic parameters were obtained by fitting the Michaelis–Menten equation to all data points using the nls (non-linear least squares) fit in program R$^{142}$ (with keeping all settings at their default values). The mechanism of inhibition was determined by fitting the competitive, non-competitive and uncompetitive inhibition models to the measured data. The lowest value of the residuals indicated the best fitting model, and thus the mechanism of the inhibition. Visual
comparison of the fitted curves to the experimental data confirmed in all cases the choice of mechanism.

### 3.3.2 Isothermal titration calorimetry (ITC)

ITC is known to be a powerful tool to characterize binding behavior of small molecules to enzymes, to our knowledge, no direct calorimetric measurements were performed previously for any MIO enzymes. During an ITC measurement, a concentrated protein solution is titrated with small portions of the ligand solution in a way that heat change is measurable for each injection producing a curved thermogram\textsuperscript{143}. At the beginning of the titration, practically all molecules of the ligand in the first injection aliquots bind to the protein, thus \( \Delta H \) of the binding could be determined. As by addition of further portions of ligand, saturation is reached, the released heat decreases. The rate of decrease provides the association constant (\( K_a \)) of the binding from which the binding and dissociation constants and \( \Delta G \) of the binding can be calculated. The \( C \) parameter determining the curvature (\( C = [E] \times K_a^{-1} \)) is suggested to be between 10–1000 for efficient determination of the thermodynamic parameters. In the present titration experiments sufficient heat was generated in all cases, however in some cases weak binding of the ligands limited the accurate determination of the binding enthalpy. Due to precipitation, the protein concentration was limited to 10 mg/ml (125 \( \mu \)M of monomers). Thus, accurate determination of all thermodynamic parameters was limited to compounds with specific dissociation constant being less than 12.5 \( \mu \)M (corresponding to \( C = 10 \)). Although binding entropies could not be determined reliably for \( C < 10 \), ITC can still be used for determining dissociation constants and free energies of binding.

ITC measurements were performed in a micro calorimeter (MicroCal 200; GE Healthcare, Chicago, USA). Protein solutions (100 \( \mu \)M monomer units of PcPAL-H10, 200 \( \mu \)l cell volume) were titrated with 5 times more concentrated ligand solutions using 20–25 injections (2 \( \mu \)l/injection). Initial delay, and time between two injections was set to 180 s. Binding was measured at 30 °C in TRIS (50 mM, pH 8.0, containing 1mM of TCEP). Titration curves were analyzed using the software and methods described by Brautigam \textit{et al.}\textsuperscript{144}. Thermograms were integrated using NITPIC\textsuperscript{145,146}. Binding thermodynamic data were determined by fitting the A+B heteroassociation model in SEDPHAT\textsuperscript{147}. Confidence intervals for \( \Delta H, K_d \) and the stoichiometry of the binding were calculated at p values corresponding to one \( \sigma \) level. The confidence interval of \( K_d \) determined the one for \( \Delta G \), while the minimal and maximal differences between the \( \Delta H \) and \( \Delta G \) confidence intervals yielded the ones for \(-T\Delta S\). Data were plotted using GUSSI\textsuperscript{148}.  

38
3.4 Protein structure determination

3.4.1 Protein crystallization

Protein aliquots stored at -80 °C (10 mg/ml, 50-100 µl) were ultra-centrifuged at 230 000g for 20 min after thawing prior to crystallization to remove any aggregates. First, initial protein crystallization conditions were tested by sitting drop vapor diffusion method, of three commercial crystallization screens (Molecular Dimensions): JCSG+, PACT and SG1. Plates of 96 different precipitants were set up with mosquito® crystal (TTP labtech) liquid handling robot, by mixing together 200 nl of protein solution and 200 nl of the precipitant. Two parallel experiments were carried out on one plate, varying the drop volume from 100 nl to 200 nl or adding ten fold excess of the APEP or APPA inhibitors to the protein. Plates were inspected 7, 14 and 28 days after initial setup, manually on a Leica DM500 optical microscope.

3.4.1.1 PcPAL-H10

Extensive trials of crystallization of PcPAL-His10 in the apo form were unsuccessful. Therefore, protein solutions were supplemented with 10-fold molar excess of (S)-(1-amino-2-phenylallyl)phosphonic acid ((S)-APPA) prior to crystallization, hoping that the strong binding ligand will decrease protein flexibility and enhance its crystallizability. The 288 initial conditions provided one hit (HEPES 0.1 M pH 7.0, PEG 6000 10 V/W%). Consecutively, three sets of 24 different precipitant conditions were assayed in 2 µl drops in hanging drop vapor diffusion setups, before geometrically suitable crystals were obtained for data collection. Crystals grew within one week of setting up drops, from the final condition; HEPES 0.15 M pH 7.0, PEG 6000 14 V/W% as well. Crystals from PcPAL-H10 bound to (S)-APPA yielded the structure accessible under the PDB code: 6F6T.

3.4.1.2 PcPAL-H6

Removal of the His-tag significantly enhanced crystallizability. Protein crystals appeared in 19 precipitant mixes after two weeks of PcPAL using the same initial screens, where no crystallization was observable of PcPAL-H10. A total of 48 new precipitant mixes were created based on two of the initial observations, and crystallization trials were set up by both sitting drop (400 nl protein+ 400 nl precipitant) and hanging drop (1 µl protein+ 1 µl precipitant) vapor diffusion methods. Crystals for the diffraction experiments grew using 20-26 W/V% PEG 3350, potassium formate 0.1-0.3 M as precipitant, and yielded the structure accessible under the PDB code: 6F6T.
code: 6H2O. Co-crystals with (R)-APEP grew from the same conditions as the apo protein. The structure accessible under the PDB code: 6HQF.

3.4.2 Diffraction data collection

Diffraction dataset for 6F6T was collected at DESY Hamburg beamline MX2-P14, on 06/12/2016. Diffraction images of 0.9762 wavelength x-rays were collected by 1° angle rotations, rotating the crystal 180°. Dataset for 6H2O was collected in ELETTRA beamline 5.2R XRD1, on 19/07/2017. Diffraction images of 1.0 wavelength x-rays were collected by 1° angle rotations, rotating the crystal 180°. Dataset for 6HQF was collected in ESRF on beamline ID30A-3, on 02/03/2018. Diffraction images of 0.9677 wavelength x-rays were collected by 1° angle rotations, rotating the crystal 180°. Table 15 shows detailed collection statistics.

3.4.3 Diffraction data processing and protein structure refinement

XDS\textsuperscript{149} was used for data processing. Manual data processing was used for datasets yielding structures 6F6T and 6H2O. Datasets were cut at CC\textsubscript{1/2} 50 values. Automatically processed data at the synchrotron by the XDSApp 2.0\textsuperscript{150} was used for the protein structure refinement resulting in structure accessible in the PDB by the code 6HQF. Table 15 shows detailed processing statistics. Structures were solved by molecular replacement using Phaser\textsuperscript{151}, with 1W27\textsuperscript{83} as a starting model for 6F6T, and with 6F6T as a starting model for 6H2O and 6HQF. Models were refined by PHENIX\textsuperscript{152} and manually adjusted in Coot\textsuperscript{153}. Table 15 shows detailed refinement statistics.

3.5 Molecular modeling

3.5.1 Construction of model structures

3.5.1.1 Hydroxynitrile lyases

Models of two mutant proteins (composite active sites: A9-H7 and A10-H6, see Figure 18 and Appendix Table A3.), showing surprising activity and enantioselectivity, were created using Maestro\textsuperscript{154} in both AtHNL and HbHNL scaffold. Modified residues were modeled by mutating them in the crystal structures of the wild-type proteins (AtHNL: 3DQZ\textsuperscript{22}, and HbHNL: 1YB6\textsuperscript{21}). Main chain conformations were kept the same as in their respective crystal structures except for residues 121-128 in AtHNL scaffold, for which coordinates were taken from HbHNL, after structure alignment. (See Figure 15 for details.) Trp128 is important for substrate...
binding and release in HbHNL so accurate modeling of this loop conformation is important. Five residues of the loop 121-128 of the consensus active sites are identical to those in the wild-type sequence, but the identity of these residues differ, sequences are shown in Figure 15. A10-H6 and A9-H7 active sites contain large, hydrophobic residues identical to HbHNL, shown in red sticks. AtHNL contains a Pro at 125 (trans conformation -176°), while HbHNL contains a Phe at 125. While the hydrophobic phenyl group of Phe points to the interior of the protein (SASA =18 Å²) and participates in the shaping of the active site, proline is entirely solvent exposed (SASA = 133 Å²) and lies further than 6 Å from the active site. Composite active sites are more likely to have HbHNL like conformations at residues 121-128 than AtHNL like conformations.

Side chain conformation of the mutated residues were set to match as closely as possible the conformation in the original crystal structures. Hydrogens were added to the model using the Protein Preparation Wizard, and the protonation states were determined using PROPKA, at the experimentally applied pH 5.5. All hydrogen orientations were optimized using the Interactive Hydrogen bond Optimizer, except for the catalytic His235, which was manually set to the neutral tautomer having a hydrogen on Nδ. Any remaining steric clashes were removed using the automatic energy minimization method in Prime with VSBG 2.0 solvation. The thus obtained structures were used as starting structures for the MD simulations of the apo enzymes, and for docking of the ligands. The OPLS_2005 force field was used throughout the modeling.

<table>
<thead>
<tr>
<th></th>
<th>121</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtHNL</td>
<td>YMEMPGGL</td>
<td></td>
</tr>
<tr>
<td>A10-H6, A9-H7</td>
<td>LMEMFGGW</td>
<td></td>
</tr>
<tr>
<td>HbHNL</td>
<td>LMEVFPDW</td>
<td></td>
</tr>
</tbody>
</table>

Figure 15. Comparison of loop 121-128 in AtHNL and HbHNL. The conformations of residues 121 to 128 (shown in sticks) differ in wt-HbHNL (yellow) and wt-AtHNL (cyan) due to different sequences. Starting positions for residues 121-128 in the AtHNL consensus active sites were modeled based on the HbHNL crystal structure. The two structures were aligned to
each other, then resides 121-128 were copied to the AtHNL structure from HbHNL. The active site Ser80 and His235 are also shown in stick representation to facilitate orientation.

3.5.1.2 MIO enzymes

All MIO enzymes exist and function as homotetramers, where three monomers participate in forming each one of the four active sites. Therefore, it is imperative to model them in their biologically functional homotetrameric form. The 6F6T crystal structure served as the basis for modeling of PcPAL. The tetrameric model was constructed using crystal symmetry data, as the asymmetric unit contains only two chains. Missing loop regions (A: 337-346, 549-557; B: 333-346, 551-554) were modeled based on the homology model created by the intensive algorithm of Phyre2\textsuperscript{162}. The Phyre2 homology modelling webserver uses a complex modeling algorithm based on Modeller\textsuperscript{163}, that uses multiple templates and ab initio modelling to create a complete and reliable model for protein sequences. Models of TcPAM were built similarly to those of PcPAL, based on the crystal structure 2YII\textsuperscript{93}, that was also used for previous molecular dynamics studies\textsuperscript{78}. The coordinates for the missing regions (A: 114-119, 567-569, 606-616, 677-686; B: 55-56, 114-119, 569-571, 606-616, 678-686; C: 115-119, 567-571, 605-616, 677-686; D: 115-119, 568-576, 605-616, 678-686) were also modeled based on a Phyre2\textsuperscript{162} homology model. Hydrogens were added to the model using the Protein Preparation Wizard\textsuperscript{156}, and the protonation states were determined using PROPKA\textsuperscript{157}, at the experimentally applied pH 8.8\textsuperscript{164}.

3.5.2 Molecular docking

3.5.2.1 Docking of mandelonitrile to the HNL’s

The binding of mandelonitrile to the wild-type AtHNL and HbHNL was modeled using Glide\textsuperscript{159}. Docking was carried out without energy minimization, to a rigid protein. The center of mass of the ligand was restricted to a 10*10*10 Å box centered on the midpoint between Ser80 and His235 and the entire ligand was restricted to a larger 25*25*25 Å box centered at the same midpoint. The van der Waals radii of the hydrophobic residues (partial charge less than 0.2) were scaled down to 80%. This smaller size approximated adjustments that could occur if the enzyme were flexible. (R)- and (S)-mandelonitrile were flexibly docked by extra precision docking\textsuperscript{165}. The van der Waals radii of mandelonitrile were not scaled. For other docking settings the default values were kept. Ligand poses that encountered steric clashes, defined as a sum of Coulomb and VdW interactions energies >0 kcal/mol, were discarded. Duplicate poses, defined by an RMS deviation of the substrate heavy atoms less than 1.0 Å or by a maximum atomic displacement less than 1.5 Å from existing poses, were also discarded.
The binding of mandelonitrile to the mutant proteins was modeled using a modified induced fit method\textsuperscript{166} to include the uncertainty related to the side chain conformations of the modified residues. First, a standard precision Glide docking was performed in a similar way as for the wild-type enzymes, however 100 kcal/mol Coulomb-VdW energy cutoff filter was used to increase number of potential binding poses. Second, for a maximum of 25 potential candidates, residues within 5.0 Å of the docked position were refined using Prime\textsuperscript{159}. All non-mutated residues were constrained by 1.0 kcal/(mol Å) during refinement. Out of the 25 minimized structures, the ones within 15 kcal/mol of the lowest energy result, or a maximum of 20 structures, were carried forward. Next, mandelonitrile was redocked into the minimized receptors using extra precision Glide\textsuperscript{165} docking without scaling of the van der Waals radii. The induced fit docking poses were visually screened according to chemical criteria required for catalysis: the hydroxyl group of mandelonitrile must donate a hydrogen bond and the nitrogen atom of the cyano group must be within 2.5 Å of an atom with a partial charge of at least +0.25. For duplicate poses the one with the lower docking score was retained. All docking results were further compared, and ranked by the combined Molecular Mechanical Generalized Born Surface Area (MM-GBSA) binding energy calculations. This fast force-field based method computes the free energy of binding from the difference between the free energies of the protein, ligand, and the complex\textsuperscript{167}. Prior to binding energy calculations, residues within 6.0 Å of the docked substrate, excluding the catalytic Ser80 and His235, were minimized by Prime\textsuperscript{159}.

3.5.2.2 Docking in MIO enzymes, starting structures for RAMD simulations

To obtain the most reliable initial positions for the substrates L-Phe and TCA for the RAMD simulations different methods were used to create the substrate bound protein models. As there is no direct crystallographic data available for L-Phe binding, docking of the zwitterionic form of L-Phe with Glide\textsuperscript{165} determined its initial positions in active sites A and B of the PcPAL and TcPAM structures prepared as described in chapter 3.5.1.2. Identical docking methods were used as the ones described for the wt-HNLs above. The initial L-Phe position was selected from the docking results by visual comparison to the crystal structure bound (R)-APEP in 6HQF and considering the docking score.

Binding conformation of TCA is known from a crystal structure of TcPAM, 4CQ5\textsuperscript{94}, therefore these coordinates were used for the initial TCA position. Experimental results\textsuperscript{76} suggest that amino group remains bound to the MIO, while TCA exits from the active site. Therefore, TCA exit from the active site of PcPAL and TcPAM was modeled with an amino-bound MIO. The amino group position was modeled based on the N-MIO intermediate present.
in the 6F6T structure. Figure 16 shows the initial substrate positions of the four models within the active sites of chain A, where the additional force was applied during the RAMD simulations.

Figure 16. Initial substrate binding conformations in PcPAL and TcPAM models. a. L-Phe binding in PcPAL, b. TCA binding in PcPAL, c. L-Phe binding in TcPAM, d. TCA binding in TcPAM.

Table 6. Active site configurations of the RAMD models for PcPAL and TcPAM#

<table>
<thead>
<tr>
<th>Protein</th>
<th>Active site A</th>
<th>Active site B</th>
<th>Active site C</th>
<th>Active site D</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>PcPAL</td>
<td>L-Phe – RAMD</td>
<td>L-Phe</td>
<td>Empty</td>
<td>Empty</td>
<td>Pc_L-Phe</td>
</tr>
<tr>
<td></td>
<td>MIO</td>
<td>MIO</td>
<td>MIO</td>
<td>Empty</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TYD-110</td>
<td>TYD-110</td>
<td>TYD-351</td>
<td>TYD-110</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCA – RAMD</td>
<td>TCA</td>
<td>Empty</td>
<td>Empty</td>
<td>Pc_TCA</td>
</tr>
<tr>
<td></td>
<td>NH2-MIO</td>
<td>NH2-MIO</td>
<td>NH2-MIO</td>
<td>NH2-MIO</td>
<td></td>
</tr>
<tr>
<td>TcPAM</td>
<td>L-Phe – RAMD</td>
<td>L-Phe</td>
<td>Empty</td>
<td>Empty</td>
<td>Tc_L-Phe</td>
</tr>
<tr>
<td></td>
<td>MIO</td>
<td>MIO</td>
<td>MIO</td>
<td>MIO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TYD-110</td>
<td>TYD-110</td>
<td>TYD-351</td>
<td>TYD-110</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCA – RAMD</td>
<td>TCA</td>
<td>Empty</td>
<td>Empty</td>
<td>Tc_TCA</td>
</tr>
<tr>
<td></td>
<td>NH2-MIO</td>
<td>NH2-MIO</td>
<td>NH2-MIO</td>
<td>NH2-MIO</td>
<td></td>
</tr>
</tbody>
</table>

*NH2-MIO: amino-bound MIO residue. TYD: deprotonated tyrosine residue.
Modeling PcPAL and TcPAM as tetramers results in very large systems. Its downside is the considerable computational cost, however it also allows running four simultaneous experiments in the four active sites. With careful considerations, control simulations can be performed in the same model as the enhanced sampling simulations; Table 6 shows the setup of the different active sites of the four models.

3.5.3 Molecular dynamics simulations of hydroxynitrile lyases

The flexible behavior of AtHNL wild-type, HbHNL wild-type, and A9-H7 and A10-H6 composite active sites of both scaffolds was simulated by molecular dynamics using the Desmond engine\textsuperscript{168}, in apo form and with (\(R\))- and (\(S\))-mandelonitrile (MNN) bound in the active site. All proteins were simulated as their physically relevant dimers. The protein was solvated in TIP3P water molecules in an orthorhombic box with 10 \(\text{Å}\) of solvent in each direction beyond the protein. The system was neutralized using Na\(^+\) ions corresponding to 0.05 M NaCl were added, to model the experimental condition. Using a combined steepest descent and conjugate gradient, the geometry of the systems was optimized until the energy changed by less than 1.0 kcal/(mol Å). None of the models showed significant movement at this step. The models were gradually heated to the simulation temperature (303 K) and equilibrated before the molecular dynamics simulations, using the default five-step relaxation procedure in Multisim workflow of Desmond\textsuperscript{168}. Following the equilibration, three repeats of 10 ns unrestrained molecular dynamics simulations were carried out, with different random starting velocities, under NPT conditions, using default settings in Maestro\textsuperscript{154}.

3.5.3.1 Analysis of simulation results

Energy conservation, constant temperature and constant pressure of the simulations were verified with the Simulation Quality Analysis module in Maestro. The simulations were stable and convergent as the root-mean-square deviation (RMSD) of the backbone protein atoms fixed around a constant value compared to the starting structure. Ligand positions during the molecular dynamics simulation were analyzed by assigning each frame to one of four binding modes: Lys binding, Oxyanion binding, Asn binding, and non-catalytically active binding. The assignment was based on the hydrogen bonding of the hydroxyl group and on the interaction of the cyano group. Data for the assignment were obtained by measuring atomic distances, angles, and dihedral angles between atoms in the protein and the ligand the Simulation Event Analysis engine in Maestro. Table 7 contains a complete list of the measured properties.
Table 7. List of properties measured for characterizing ligand binding

<table>
<thead>
<tr>
<th>DONOR</th>
<th>ACCEPTOR</th>
<th>PROPERTIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNN - OH</td>
<td>His235 – SCH</td>
<td>d, α, β, in plane dihedral</td>
</tr>
<tr>
<td>MNN - OH</td>
<td>Ser80 – SCH</td>
<td>d, α, β</td>
</tr>
<tr>
<td>SER80 - SCH</td>
<td>His235 – SCH</td>
<td>d, α, β, in plane dihedral</td>
</tr>
<tr>
<td>ASN11/THR11 - SCH</td>
<td>MNN – OH</td>
<td>d, α, β</td>
</tr>
<tr>
<td>MNN - OH *</td>
<td>Thr11 – SCH</td>
<td>d, α, β</td>
</tr>
<tr>
<td>THR11 - SCH</td>
<td>Ser80 – SCH</td>
<td>d, α, β</td>
</tr>
<tr>
<td>OXYANION HOLE</td>
<td>MNN – N</td>
<td>d, α, β; d, α, β</td>
</tr>
<tr>
<td>(ALA12 - MCH, PHE81 - MCH)</td>
<td>MNN – N</td>
<td>d, α, β</td>
</tr>
<tr>
<td>ASN11/THR11 - SCH</td>
<td>MNN - N</td>
<td>d, α, β</td>
</tr>
<tr>
<td>LYS236 - SCH</td>
<td>MNN - N</td>
<td>d, α, β; d, α, β; d, α, β</td>
</tr>
</tbody>
</table>

SCH: Side chain, MCH: Main chain

* For any given hydrogen bond (D–H - A–C) the listed parameters are defined as: d: H–A distance. α: D–H -A angle, β: H- -A–C angle. In plane dihedral angles are measured for hydrogen bonds donated to His235: H- -N–C– N dihedral angle.

* Residues 11, and 236 are different in wt-HbHNL than other investigated proteins. Thr11 is a potential hydrogen bond acceptor and Lys236 may donate hydrogen bonds. Asn11 and Met236 cannot fulfill the same properties. The properties marked are only measured in the wt-HbHNL MD simulations.

Residue positions were used to characterize the shape of the active sites. The distances between the centers of mass of residues listed in Appendix Table A4 and Ser80 were measured in each frame for all trajectories with a custom tcl script in vmd. Ser80 occupies a central position in the active site, and a hydrogen bond to His235 minimized movement of the Ser80 side chain. The distance between the centers of mass of MNN and Ser80 was also measured in each frame. All data analyses were performed using R software. The median distance between Ser80 and each residue measured the residue’s position, while the median absolute deviation (MAD), which is defined as

\[
\text{MAD} = \text{median}(X_i - \text{median}(X))
\]

measured the normal movement of each residue during the simulation. The use of the median, instead of the mean, minimized the effect of outlier distances. In a few cases, the side chains adopted two conformations yielding a bimodal distribution, which was verified using the diptest package. Two normal distributions were fit to the data using the expectation maximization algorithm implemented in the mixtools package. The more frequent of the two conformations was used in the analysis.

If the median residue positions in a composite active site differed in the two scaffolds by more than the median movement of these residues (MAD), then the positions of these residues were considered to differ in the two scaffolds. These differences may contribute to the different catalytic activity of the same composite active site within different scaffolds.
3.5.4 Random acceleration molecular dynamics simulations of MIO enzymes

3.5.4.1 Molecular dynamics simulations

Short molecular dynamics (MD) simulations ensured the equilibration of the two PcPAL and the two TcPAM models prior to the random acceleration MD (RAMD) simulations and determined their equilibrium behavior. Parameters for unconventional residues: MIO, Amino-MIO, L-PHE, TCA and deprotonated tyrosine were determined using Antechamber\textsuperscript{172}. The automatically assigned parameters from GAFF\textsuperscript{173} were compared to the geometric values observed in the crystal structures, and corrected when necessarily.

Amber ff14SB\textsuperscript{174} force filed was used to describe protein residues. Proteins were solvated in a truncated octahedral box with TIP3P\textsuperscript{175} water molecules with 10 Å padding from the protein atoms, to minimize the size of the system. Na\textsuperscript{+} ions neutralized the systems, and 0.1 M NaCl modeled the ionic strength of the experimental conditions\textsuperscript{164}.

All MD simulations were run with NAMD 2.11\textsuperscript{176}, as it is the only MD engine capable of running RAMD simulations. First 5000 step minimization removed clashes from the water box, with all protein and ligand atoms constrained using 10 kcal/(mol Å) force constant. Second the whole system was minimized for 10 000 steps. Third, 100 ps NVT simulations relaxed the water box (T=303 K, Langevin dynamics\textsuperscript{177}), with all protein and ligand atoms constrained using 10 kcal/(mol Å) force constant. Finally, 20 ns long NPT simulations (p=1 atm, T=303 K) probed the equilibration behavior of the modeled systems and relaxed the models before the RAMD simulations. All bonds involving hydrogen atoms were constrained during all simulations to permit increasing the time step to 2 fs. Electrostatic interactions were cutoff at 12.0 Å, long-range interactions were treated with PME method\textsuperscript{178}. Constant pressure and temperature were achieved with the Noosé-Hoover Langevin piston method\textsuperscript{177,179}. Energy conservation, constant temperature and constant pressure of the simulations were verified after each simulation.

Convergence of backbone root-mean-square deviation (RMSD) showed that systems reached equilibrium after approximately 2 ns, Figure 17. Snapshots extracted from the MD simulations at 10 ns, 15 ns and 20 ns served as starting conformations for the RAMD simulations.
Figure 17. Root-mean-square deviation (RMSD) calculated for all protein backbone atoms. The RMSD of all simulation increases sharply at the beginning of the trajectories and then levels to an equilibrium value of 1.5-2 Å. This value is reached after approximately 2 ns of simulation, suggesting that 10-20 ns simulation time is sufficient to reach full equilibrium of the model systems.

3.5.4.2 Random acceleration molecular dynamics simulations

The RAMD\textsuperscript{180} enhanced sampling method applies an artificial force to the substrate in a random orientation, accelerating the dissociation process. The first step of setting up RAMD simulation is to find optimal values for the acceleration of the substrate and the minimum distance that the substrate must travel to retain the direction of the acceleration. Excessive acceleration or short distances will produce biologically irrelevant exit paths, while no exit will be observed during the desired time scale if the acceleration is too low. In order to obtain comparable results in the four models (Pc\textsubscript{L}-Phe, Pc\textsubscript{TCA}, Tc\textsubscript{L}-Phe, Tc\textsubscript{TCA}) test simulations were run from all twelve starting structures (three for each model) with the different RAMD settings. Acceleration and displacement settings were optimized to achieve exit times between 0.1 and 2 ns. Accelerations varied from 0.07 to 0.15 kcal/(mol Å) by 0.1 steps during the optimization. For the accelerations 0.10, 0.11 and 0.12 kcal/(mol Å) three displacement settings (0.1, 0.5, 1 Å) were also assayed. Table 8 lists the parameter optimization results. Finally, 15 independent RAMD simulations were run from each of the three starting structures extracted from the MD simulations with 0.11 kcal/(mol Å) acceleration and 0.05 Å displacement. This resulted in 45 RAMD simulations for each of the four models (Pc\textsubscript{L}-Phe, Pc\textsubscript{TCA}, Tc\textsubscript{L}-Phe, Tc\textsubscript{TCA}), a total of 180 simulations.
Table 8. Ligand exit times [ps] depending on acceleration and displacement settings of the RAMD simulations.

<table>
<thead>
<tr>
<th>Acceleration [kcal/(mol Å)]</th>
<th>Displacement [Å]</th>
<th>TePAM</th>
<th>TCA</th>
<th>PcPAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1-Phe</td>
<td>TCA</td>
<td>1-Phe</td>
</tr>
<tr>
<td>0.15</td>
<td>0.05</td>
<td>44/56/91</td>
<td>78/25/46</td>
<td>164/149/40</td>
</tr>
<tr>
<td>0.14</td>
<td>0.05</td>
<td>38/21/57</td>
<td>81/44/92</td>
<td>40/63/50</td>
</tr>
<tr>
<td>0.13</td>
<td>0.05</td>
<td>52/102/116</td>
<td>196/82/211</td>
<td>286/309/41</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>671/314/253</td>
<td>1881/2000/382</td>
<td>216/464/225</td>
</tr>
<tr>
<td>0.12</td>
<td>0.05</td>
<td>117/97/240</td>
<td>77/53/182</td>
<td>88/820/279</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>369/479/577</td>
<td>388/136/790</td>
<td>716/185/367</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>2000/80/145</td>
<td>72/1361/2000</td>
<td>2000/104/432</td>
</tr>
<tr>
<td><strong>0.11</strong></td>
<td><strong>0.05</strong></td>
<td><strong>164/182/668</strong></td>
<td><strong>391/162/232</strong></td>
<td><strong>225/411/791</strong></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>1096/868/1226</td>
<td>85/121/1210</td>
<td>1337/1921/2000</td>
</tr>
<tr>
<td>0.10</td>
<td>0.05</td>
<td>815/361/319</td>
<td>194/88/506</td>
<td>1038/2000/222</td>
</tr>
</tbody>
</table>

*The three values in the table correspond to exit times from three different equilibrated starting model structures. The selected setting is highlighted in red.

3.5.4.3 Analysis of RAMD simulations

Visual Molecular Dynamics\textsuperscript{169} package was used for visual trajectory analysis. Custom tcl scripts measured geometric parameters and movements in the equilibration and RAMD trajectories. Data generated by the tcl scripts were analyzed using R\textsuperscript{142}. 

---

49
4 Results and Discussion

4.1 Reaction mechanisms in ancestral and modern HNLs

4.1.1 HNL phenotypes of the consensus active sites

Twenty-two (22) amino acid residues lie within 6 Å of the substrate mandelonitrile in HbHNL (PDB: 1YB6) or in the superimposed structure of AtHNL (PDB: 3DQZ, chain A). Six of these residues are the same in both proteins (Leu157, His14, Leu146, and Tyr158, as well as the residues Ser80 and His235 from the catalytic triad), while 16 residues differ between the two enzymes (Figure 18). The hypothesis was that exchanging active site residues between AtHNL and HbHNL would exchange their reaction mechanisms.

![Figure 18](image)

**Figure 18. Four different active sites that were examined are laid out horizontally.** Active site residues that differ between the enzymes are listed in yellow if HbHNL-like and cyan if AtHNL-like. Residues within 6 Å of the active site are shown by filled circles, while the second shell residue is represented by an empty circle. Each of these four active sites were made in both HbHNL (top row) and AtHNL (bottom row). Completely swapping the active sites failed to produce soluble protein (red cross). The two composite active sites resulted in soluble protein in either scaffold, and the enantiomeric ratios are indicated in the boxes. Yellow indicates the HbHNL-like (S)-selective enzyme, while cyan indicates the AtHNL-like (R)-selective enzyme. Enantioselectivities were determined by enzymatic synthesis of mandelonitrile and analysis on chiral HPLC.

Various active-site exchanges created 36 protein variants, of which approximately half were soluble and catalytically active. Two variant pairs were particularly interesting, because the equivalent active site in either HbHNL or AtHNL produced soluble, active protein. Starting
with HbHNL, substitution of either nine or ten positions (11, 12, 79, 81, 106, 148, 152, 210, and 236, with or without 178) with the corresponding residues in AtHNL yielded Hb-A9-H7 and Hb-A10-H6, respectively. Similarly, starting with AtHNL, substitution of the remaining six or seven amino acid residues (13, 121, 125, 128, 131, and 209, with or without 178) with the corresponding residues in HbHNL yielded At-A9-H7 and At-A10-H6, Figure 18.

These four proteins create two pairs with identical amino acid residues within the active site, but differing residues outside the active site. In these composite sites, all residues in the catalytic center (non-hydrogen atom closer than 5.4 Å to either the substrate oxygen, chiral C2 carbon, or cyanide: PDBs 1YB6 or 3DQZ) are AtHNL-like, including the catalytically important Asn11 and Met236. The nonpolar substrate binding pocket contains predominantly HbHNL-like residues, but some AtHNL-like residues are also present. The HbHNL-like residues are closer to the phenyl carbons of the substrate, except Cys13, which is closer to the chiral C2 carbon, but is still far away (distance between heavy atoms 6.2 Å). These four proteins had similar catalytic activity toward mandelonitrile (0.4–1.9 s⁻¹), which is significantly lower than the catalytic activity of HbHNL (25.6 s⁻¹) or AtHNL (27.0 s⁻¹) (Appendix Table A3).

Surprisingly, the two proteins with “A10-H6” composite active sites still showed opposite enantioselectivity, corresponding to the favored enantiomer of the starting protein (Figure 18 and Appendix Table A3). Wild-type HbHNL favors (S)-mandelonitrile (E > 39) and is inhibited by (R)-mandelonitrile. Hb-A10-H6 retains this (S)-preference, but also catalyzed cleavage of (R)-mandelonitrile, lowering the enantioselectivity (E = 3.3). Wild-type AtHNL favors (R)-mandelonitrile (E > 39) and At-A10-H6 enzyme retained the (R)-preference, also with lower enantioselectivity (E = 5.2). Thus, the same A10-H6 active site shows opposite enantioselectivity in the different scaffolds, demonstrating that residues outside the active site affect enantioselectivity strongly enough to reverse the favored enantiomer.

Changing phenylalanine 178 to leucine converted the “A10-H6” site to the other composite active site, “A9-H7”, which was (S)-selective in both enzyme scaffolds (Figure 18 and Appendix Table A3). The enantioselectivity in the HbHNL scaffold increased from E = 3.3 (S) in Hb-A10-H6 to E >39 (S) in Hb-A9-H7. The enantioselectivity in the AtHNL scaffold also shifted toward the (S)-enantiomer from E = 5.3 (R) in At-A10-H6 to E = 1.6 (S) in At-A9-H7. Thus, the single substitution of F178L shifts enantioselectivity toward (S)-mandelonitrile in both scaffolds.
These composite active sites resemble the active site of ancestral enzyme EST3 mL. Several substitutions to make this variant more like wild-type HbHNL (Met236 back to Lys or Phe79 back to Glu) yielded only insoluble proteins. Similarly, additional substitutions that made it more like wild-type AtHNL (Phe125 to Pro or Trp128 to Leu) also yielded insoluble proteins. Although we did not identify a path of single substitutions yielding soluble and catalytically active variants at each step, we did not test all possible combinations of these 16 substitutions.

4.1.2 Modeling the structural effect of the mutations

Molecular modeling sought structural explanation on a molecular level for the protein functions observed experimentally.

4.1.2.1 Modeling of static ligand binding

Docking followed by MM-GBSA binding energy calculations generated the starting structures of mandelonitrile enantiomers bound to the active sites for the molecular dynamics simulations. Docking generated multiple binding poses with similar docking scores. Subsequent MM-GBSA calculations refined the ranking of the poses (Table 9). Both enantiomers of mandelonitrile were modeled in every enzyme as they both likely bind. Variants with low enantioselectivity (At-A10-H6, At-A9-H7 and Hb-A10-H6) must bind both enantiomers since both enantiomers react. Variants with high enantioselectivity also likely bind the slow enantiomer, as that inhibited the wild-type proteins.

As expected, (R)- and (S)-MNN docked in a conformation prearranged for the Lys mechanism in wt-HbHNL (2.4–2.7 Å distance between the ε-amino group of lysine and the N atom of the cyano group, Table 9). Docking of (S)-mandelonitrile yielded similar poses (RMSD$_{S-MNN}$: 0.4 Å) to the crystal structure (PDB: 1YB6), confirming that the docking procedure was adequate for generating reliable models for the substrate binding conformations. Nonetheless, the crystal structure pose was the starting structure for the simulations of wt-HbHNL and (S)-MNN.

---

* Following the initial unsuccessful alterations of the AtHNL and HbHNL active sites leading to catalytically inactive and insoluble protein, EST3 mL sequence guided the selection for the composite active site residues, as it was known to be active and to express well in *E. Coli*. 
Docking poses with the lowest MM-GBSA values are prearranged for the oxyanion hole mechanism in wt-AtHNL, the cyano group interacts with the main chain amino group of Ala12. Glide scores indicated a second binding pose for both (R)- and (S)-MNN, where the cyano group interact with the amino group of Asn11, however MM-GBSA calculations indicated these poses to be very high in energy, therefore improbable.

Docking suggested two possible binding conformations for (R) and (S)-MNN in the composite proteins Hb-A9-H7 and Hb-A10-H6. In the first the cyano group fits into the oxyanion hole, like in wt-AtHNL, while in the second it interacts with Asn11. MM-GBSA calculations showed that this second binding conformation is lower in energy in the composite proteins Hb-A9-H7 and Hb-A10-H6, therefore more probable.
Docking suggested two possible binding conformations for \((R)\) and \((S)\)-MNN in the composite protein At-A9-H7, but only one for At-A10-H6. MM-GBSA calculations suggested that the wt-AtHNL like binding, when the cyano group residues in the oxyanion hole, is lower in energy, thus more probable for these two active sites.

Docking pose with the lowest MM-GBSA value was the starting structure for all the molecular dynamics simulations, except for wt-HbHNL and \((S)\)-MNN. MM-GBSA values are commonly used for estimating ligand binding affinities and ligand binding free energies, however due to the crude approximations of the method it is more suitable for ranking binding conformations relative to each other, then for determining the absolute values of binding free energy\(^{181}\).

### 4.1.2.2 Catalytically active conformations indicate reaction mechanisms in wild-type and composite enzymes

Molecular dynamics simulations of mandelonitrile bound to the active site identified reactive conformations, qualitatively explained the unexpected differences in enantioselectivity and identified differences in shapes of the active sites with identical residues. Molecular dynamics generated 6000 possible binding conformations. From these, we defined catalytically active conformations (CAC, Table 10) using two criteria. The first was that the hydroxyl group of mandelonitrile donated a hydrogen bond to a proton acceptor in the enzyme. The ultimate proton acceptor was the catalytic histidine, but, in some cases, Ser 80 and/or Thr 11 created a hydrogen bond network between the mandelonitrile hydroxyl and the catalytic histidine. The second criteria for CAC was that the cyano group of mandelonitrile simultaneously accepted a hydrogen bond from the enzyme. The hydrogen bond donor could be either the lysine ε-amino group, the main chain amide N–H’s in the oxyanion hole or the asparagine δ-amide N–H (Figure 19). Hydrogen bonding criteria were permissive to find all likely CAC.

Molecular dynamics simulations identified catalytically active conformations (CAC) of wt-HbHNL as similar to the x-ray structure. For the favored \((S)\)-mandelonitrile binding corresponding to the Lys mechanism was present in 99% of the CAC (46% of the simulation time), Table 10, Figure 19 a. Although the hydroxyl of mandelonitrile hydrogen bonds to Ser80 in the x-ray structure, only 37% of the CAC retained this hydrogen bond during the molecular dynamics. In the rest of the CAC (63%), the hydrogen bond moved to Thr11. In both cases hydrogen bonding network is present from Thr11 or Ser80 to His235, hence we suggest His235 to be the ultimate proton acceptor.
Table 10. Catalytically active\(^a\) conformations (CAC) from molecular dynamics simulations\(^*\).

<table>
<thead>
<tr>
<th></th>
<th>(R)-mandelonitrile</th>
<th>(S)-mandelonitrile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAC [%] CN-partner</td>
<td>% of CAC OH-partner</td>
</tr>
<tr>
<td>wt-HbHNL</td>
<td>20 Thr11</td>
<td>88 Ser80</td>
</tr>
<tr>
<td>wt-AHNL</td>
<td>44 Oxyanion hole</td>
<td>100 Ser80</td>
</tr>
<tr>
<td>At-A10-H6</td>
<td>35 Oxyanion hole</td>
<td>100 Ser80</td>
</tr>
<tr>
<td>At-A9-H7</td>
<td>54 Oxyanion hole</td>
<td>97 Ser80</td>
</tr>
<tr>
<td>Hb-A10-H6</td>
<td>13 Asn11</td>
<td>82 His235</td>
</tr>
<tr>
<td>Hb-A9-H7</td>
<td>13 Asn11</td>
<td>100 His235</td>
</tr>
</tbody>
</table>

\(^a\)A conformation is considered catalytically active if the hydroxyl group of mandelonitrile forms a hydrogen bond to a potential proton acceptor, and its cyano-group interacts with a partially positive hydrogen atom. Hydrogen bond criteria (D–H–A–C): D–H–A distance \(\leq 2.5\,\text{Å}\), D–H–A angle \(\geq 90^\circ\), H–A–C angle \(\geq 90^\circ\). Additional criterion was employed for histidine, as the lone electron pair, participating in the hydrogen bond formation, of the sp\(^2\) type nitrogen in the imidazole ring resides in the plane of the ring\(^1\). For a H-bond to His235 the torsion angle H–N–C–N was restricted to 20\(^\circ\) out of the plane angles. For catalytically active binding with Ser80 base (Ser80 → His235) hydrogen bond must also be present.

\(^*\)The favored enantiomers are in bold; for each protein data is from three repeats of 10 ns simulations.

Ligand binding poses during a similar molecular dynamics simulation of the natural substrate, acetone cyanohydrin, bound to HbHNL matched the ones in the x-ray crystal structure\(^3\) and in QM/MM calculations\(^15\). Molecular dynamics simulation starting from the crystal structure (PDB: 1SC9\(^3\)) found acetone cyanohydrin in CAC in 47% of the frames, accepting a hydrogen bond from Lys in 98% of the those, and donating a hydrogen bond to Ser80 in 99% of those. This match between molecular dynamics, the starting crystal structure and the QM/MM modeling validates the molecular dynamics methods.

**Figure 19.** Different catalytically competent orientations of mandelonitrile stabilize the leaving cyanide group differently and favor different enantiomers. Lys mechanism (a) and oxyanion hole mechanism (b) presented on Figure 4 in detail. In the mechanism proposed here for the (S)-enantioselective composite active-site enzymes (c), derived from HbHNL, the \(\delta\)-amide N–H group of Asn11 stabilizes the cyanide leaving group (Asn mechanism). The other two mechanisms are not possible, because Lys236 is missing and because distant residues cause Phe81 to block the oxyanion hole region.
Molecular dynamics simulations confirmed the oxyanion hole mechanism for wt-AtHNL, but suggest Ser80, not His235, as the hydrogen bond acceptor. The cyano group of the favored (R)-mandelonitrile, accepted hydrogen bonds from the oxyanion hole in all of the CAC poses (46% of the total conformations), Table 10, Figure 19 b. The hydrogen bond acceptor was Ser80 in 95% of the CAC. Previous docking\(^2\) predicted His235 as the hydrogen bond acceptor. While our docking also predicted His235 as the hydrogen bond acceptor, Table 9, the MD simulations favored Ser80, Table 10. QM/MM calculations\(^3\) found His235 to be a possible hydrogen bond acceptor, but did not test the Ser80 alternative.

Molecular dynamics simulations predicted the oxyanion hole mechanism for composite active sites with the AtHNL scaffold (Table 10). Simulations show the AtHNL variants bound the cyano group in the oxyanion hole in all of the CAC. This similarity to wt-AtHNL is expected, since all residues in the catalytic center of the composite active sites are AtHNL-like, including the catalytically important Asn11 and Met236. Mandelonitrile was hydrogen-bonded to either Ser80 and His235 in the catalytically active frames, suggesting that either may be the proton accepting base.

Variant HbHNLs with composite active sites (Hb-A9-H7 and Hb-A10-H6) likely use a catalytic mechanism different from either wild-type protein. The nine substitutions that convert wild-type HbHNL to Hb-A9-H7 eliminate the possibility of the (S)-selective catalytic Lys mechanism of HbHNL\(^3\), because these substitutions remove the essential Thr11 and Lys236 residues. While we expected that the addition of Asn11, Ala12, and Phe81 would enable the (R)-selective catalysis of AtHNL\(^2\), HbHNL-A9-H7 remained highly (S)-selective (>39). The ability to maintain ~3% of wild-type activity without either catalytic residue suggests these composite enzymes catalyze (S)-selective mandelonitrile cleavage, using a novel mechanism.

Molecular dynamics simulations predicted a new mechanism for the composite active sites with the HbHNL scaffold. In the new mechanism, the cyano group of mandelonitrile accepts a hydrogen bond from HD21 of Asn11, thus it is called the Asn mechanism (Figure 19 c, Table 10). Different binding from AtHNL is unexpected, as all residues in the catalytic center (as defined above) of the composite active sites are AtHNL-like. CAC are present in 81% and 93% of the simulation time for Hb-A10-H6 and Hb-A9-H7, respectively. The cyano group hydrogen-bonded almost exclusively to Asn11 in the CAC. The partially positively charged HD21 Asn11 (q = +0.38e, OPLS_2005 force field) interacts with the partially negatively charged nitrogen of the cyano group, fulfilling a similar role to Lys236 and the two main-chain amino groups in the other two mechanisms. The stabilizing effect on the forming cyanide leaving group should be proportional to the total positive charge of the protein partner.
Lys236 is positively charged, and the oxyanion hole donates two hydrogen bonds, but asparagine is not charged and donates only one hydrogen bond. This smaller positive charge may explain the decreased reaction rates for the composite active sites with the HbHNL scaffold (Appendix Table A3). The hydroxyl group of mandelonitrile, as in the other two mechanisms, donates a hydrogen bond to either Ser80 or His235. A change in mechanism is also consistent with the “island of activity”; additional amino acid substitutions that make the composite active site proteins more wild-type-like yielded insoluble protein in all experimental results. The inability of Hb-A9-H7 to catalyze cleavage of acetone cyanohydrin also supports the conclusion that it has lost the Lys mechanism.

The new asparagine mechanism proposed in the composite site enzymes may have been an evolutionary stepping stone to modern HNL mechanisms. Both HbHNL and AtHNL evolved from esterases, yet a direct transition between modern esterase and lysine mechanisms HNL requires going through catalytically dead intermediates, because of the need to both block the oxyanion hole and to correctly position a charged lysine to bind the leaving cyanide

However, the reconstruction of the last common ancestor of HbHNL and AtHNL, called EST3 mL (Figure 5 shows a phylogenetic tree and lists the reconstructed ancestors), catalyzed both cyanohydrin cleavage and ester hydrolysis. This ancestor’s active site resembles the composite active-site enzymes created here, including the critical Asn11. All active site residues in the ancestor either match A10-H6 (17 residues) or are residues not found in either modern enzyme (5 residues). EST3 mL shows 2.3% of the HNL activity of HbHNL and shows low enantioselectivity ($R$, 7) similar to the activity (1.6% of HbHNL) and enantioselectivity ($S$, 3.3) of HbHNL-A10-H6. Since HbHNL-A10-H6 likely follows the Asn mechanism, we propose that ancestral enzyme EST3 mL similarly follows the Asn mechanism. This suggests that finding the new asparagine mechanism in the composite enzymes may not be due to serendipity, but instead may be uncovering the ancestral stepping-stone mechanism between ancestral esterases and modern HNLs. A stepping stone mechanism would be compatible with both previous (esterase) and subsequent mechanisms (modern HNL mechanisms), allowing an evolutionary path from one to the other via a promiscuous enzyme, with a functional enzyme at every step. Since both the composite site enzymes and the ancestor EST3 mL are significantly slower than modern, wild-type HNLs, evolutionary pressure could optimize HNL function by converting from an asparagine mechanism to the modern lysine or oxyanion hole mechanisms if a viable evolutionary path exists. A few mutants along a few possible trajectories were tested and found to be dead or insoluble, but these only represent a small
sample of the 9! (>360 000) possible direct evolutionary routes to mutate the 9 residues that differ between Hb-A9-H6 and wild-type HbHNL.

4.1.2.3 CAC qualitatively measure enantioselectivity of wild-type and composite enzymes

The relative numbers of catalytically active frames in the molecular dynamics simulations for the enantiomers of mandelonitrile qualitatively match the enantioselectivity. In wt-HbHNL, the favored (S)-enantiomer adopts CAC corresponding to the Lys mechanism 46% of the time, while non-favored (R)-enantiomer adopts this CAC only 2% of the time, Table 10 and Figure 19a. This result qualitatively matches the high (S)-enantioselectivity of HbHNL. The non-favored (R)-enantiomer also adopts a possibly catalytically active conformation where cyano group accepts a hydrogen bond from the hydroxyl of Thr11 and the mandelonitrile hydroxyl donates a hydrogen bond to the hydroxyl of Ser80, in 19% of the simulation time.

In wt-AtHNL, the favored enantiomer, (R)-mandelonitrile, adopts CAC corresponding to the oxyanion hole mechanism 44% of the time, while the non-favored (S)-enantiomer adopts CAC only 7% of the time, Table 10 and Figure 19 b. The hydrogen bond acceptor differs for the enantiomers. (R)-Mandelonitrile donates a hydrogen bond to the side chain of Ser80 (95% of the catalytically active frames), which in turn hydrogen bonds to NE2 of His235. In contrast, (S)-mandelonitrile donates a hydrogen bond directly to NE2 of His235. Ser80 accepts a hydrogen bond from a wide range of geometries, but His235 requires the proton to point to the lone pair, which lies parallel to the ring. This more stringent geometric requirement of the hydrogen bond to His235 accounts for the most of the difference in the fraction of catalytically active frames. The same preference of the favored enantiomers for serine and slow enantiomers for histidine also occurs in the composite active sites as detailed below.

Modeling of composite active sites with AtHNL scaffold explains the shift from (R) toward (S) enantioselectivity from wt-AtHNL to At-A10-H6, but not the further shift toward (S) in At-A9-H7. In At-A10-H6, both enantiomers bound in CAC (S, 33%; R, 35%), in agreement with the low enantioselectivity. Altered binding of the phenyl group of mandelonitrile accounts for the increase in CAC for the slower (S)-enantiomer, compared to wt-AtHNL. Adding the L178F substitution to make AtHNL-A9-H7 reversed the enantioselectivity to favor (S)-MNN (Figure 18 and Appendix Table A3), but the modeling showed similar binding to wild-type AtHNL and found the slow (R) enantiomer more frequently in CAC. These results suggest that both (R)- and (S)-mandelonitrile may be cleaved by the oxyanion hole mechanism.
Modeling of composite active sites with HbHNL scaffold explains the high enantioselectivity in Hb-A9-H7, but misses the shift from (S) toward (R) in Hb-A10-H6 due to F178L. The favored (S)-MNN binds in CAC in 93% and 81% of the simulation time in Hb-A9-H7, Hb-A10-H6 active sites, compared to only 13% for (R)-MNN. The effect of F178L on the enantioselectivity is missed by the modeling, as substrate binding and the CAC ratios are similar in Hb-A9-H7 and Hb-A10-H6 active sites.

4.1.2.4 Relative residue position differences and their correlations in molecular dynamics

The different reactivity and enantioselectivity of composite active sites, despite identical amino acid residues, means that the structures of the active sites differ. The closest 44 residues to the substrates (Appendix Table A4) consist of 33 identical residues (either conserved or mutated to be equivalent), and 11 residues that differ in the two scaffolds. The first shell residues (16 + 5) are all identical, while the second shell contains 12 identical and 11 differing residues. To identify how the differing residues create different conformations in the identical residues, we compared the positions of these 44 residues during the MD simulations. The median distance between the center of mass (COM) of each residue and the COM of Ser80 at each frame of the simulation indicated the typical position of the residue. The median absolute deviation (MAD) of this distance measured the movement during the simulation, Appendix Table A4.

Presence and the nature of the ligand had no visible influence on the shape of the active site, therefore data from the 3 different simulations (apo, (R)-MNN, (S)-MNN) were merged together. The distances from Ser80 were also computed for the substrates (R)- and (S)-MNNs, from the respective simulations. Median and median absolute deviations (MAD) are used as robust statistical measures of the central tendency and variability of the data.
Figure 20. Among the 44 residues closest to the substrate, the positions of 17 differ between HbHNL and AtHNL scaffolds. Residue positions are defined with respect to Ser80 (red sticks). AtHNL-A10-H6 (cyan) contains (R)-mandelonitrile (light cyan sticks), while HbHNL-A10-H6 (yellow) contains (S)-mandelonitrile (light yellow sticks). a. Overview of the 17 differing residues (red). b. Four residues (stick representation) differ between the scaffolds by more than twice the amount that they move within each protein, while the remaining 13 residues (thin lines) differ by more than they move within each protein. For clarity, this image is rotated ~180°, compared to the other two images. c. Close-up of the four residues with most different positions. The side-chain conformation of Phe81 differs notably between the scaffolds; labels show the average dihedral angles. Representative structures are shown from the last time points of the molecular dynamics simulations here.

Seventeen of the 44 residues’ positions varied more between proteins with the same composite active site than the residues moved within each protein (Figure 20 and Appendix Table A5). Four residue positions (54, 81, 103 and 178) differed between protein scaffolds by more than twice the median absolute deviation within each simulation. Correlation analysis, \textit{in silico} and \textit{in vitro} mutational studies investigated the origins of residue position differences that are suggested to be the cause of different ligand binding and enantioselectivity in the
different protein scaffolds. Position 103 was not investigated further in this study, it lies outside the active site, and as described above, changing His103 from HbHNL to Val similar to Leu in AtHNL, only affects protein stability without influencing activity or enantioselectivity. The modeling used histidine at this location because the x-ray structure showed its side chain conformation, while the position of the valine side chain would have been less certain.

Besides occupying different relative positions compared to Ser80, the locations of residues 54, 81 and 178 strongly correlate to each other in both A9-H7 and A10-H6 active sites. Moreover, the locations correlate to whether mandelonitrile adopts a CAC by the Asn or oxyanion hole mechanisms (Figure 21). Residue 54 lies in the second shell and differs in the two scaffolds; phenylalanine in HbHNL scaffolds and valine in AtHNL scaffolds. Residue 81 lies in the first shell directly in contact with mandelonitrile and is phenylalanine in wt-AtHNL and in the composite active sites; in wt-HbHNL residue 81 is cysteine. Residue 178 also lies in the first shell directly in contact with mandelonitrile. In the A9-H7 active sites and in wt-HbHNL, 178 is phenylalanine, while in the A10-H6 active sites and in wt-AtHNL, 178 is leucine (Appendix Table A5).

The smaller valine at position 54 in AtHNL creates a hydrophobic pocket outside the active site. Phe81 occupies this hydrophobic pocket in wt-AtHNL and in AtHNL composite active sites (Phe81-Ser80: 5.6 Å, N-CA-CB-CG dihedral of Phe81 -175°). The larger phenylalanine at position 54 in the HbHNL scaffold pushes the side chain of Phe81 towards the active site (Phe81-Ser80 4.4 Å, N-CA-CB-CG dihedral of Phe81 -68°) Figure 20, Appendix Table A6. Differences in the side-chain orientation of Phe81 may also contribute to switch of the mechanism by blocking the oxyanion hole and preventing binding of the cyano group in the HbHNL scaffold. In HbHNL variants, pointing towards the active site, the side chain of Phe81 pushes residue 178 away from Ser80, creating differently shaped active sites. In the AtHNL active sites Leu178 is 1.2 Å closer to Ser80 than in the HbHNL ones. When 178 is Phe, this difference increases to 1.7 Å. That the different location of residue 178 contributes to different enantioselectivity in the two scaffolds is consistent with the F178L substitution in the A10-H6 active site shifting enantioselectivity, Figure 18, Appendix Table A3.
Figure 21. Representation of correlation between the center of mass (COM) distances of the active site residues from Ser80, and atomic distances between the cyano group of the ligand and the partially charged hydrogens of Asn11, Ala12 and Phe81. A9-H7 and A10-H6 active sites in the two backbones are compared on panels A and B. Panel C illustrates the effect of F54V mutation on the correlations for the A10-H6 active site. Out of the 44 residues listed in Appendix Table A4 and A5, residues whose positions differed more than the variance in their positions, and residues whose positions correlated with more than 0.6 with each other are included on the graphs. First shell and second shell residues are represented by dark gray and light gray circles respectively ($d_{\text{residue}}$). The yellow circles show the distance between the nitrogen of the cyano group and the hydrogen of Asn11 (XN-NH), and cyan circles represent that between the nitrogen of the cyano group and the main chain hydrogens of Ala12 (XN-AH) and Phe81 (XN-PH). Pairwise correlations of more than 0.6 are shown on the figure. Correlations ±0.6 to ± 0.7 are represented with gray edges. Red edges show negative correlations while positive correlations are shown with blue edges. The thickness of the colored edges shows the strength of the correlation between the distances: thin edges show correlations of 0.7-0.8, medium edges represent correlations of 0.8-0.9 and correlations above 0.9 are indicated by thick edges.

We hypothesized that making second shell residue 54 identical in the composite active sites would eliminate differences in the conformations of the first shell residues 81 and 178, but modeling and experiments indicate that differences persist. As hypothesized, modelling indicated that the F54V mutation changed the side chain of Phe81 from pointing into the active site in Hb-A10-H6 to pointing outside of the active site in Hb-A10-H6-F54V (Phe81-Ser80: 5.6 Å, N-CA-CB-CG dihedral -164°, Appendix Table A6). However, contrary to the hypothesis, modelling indicated that the V54F mutation in the At-A10-H6 did not change the conformation of Phe81. It remained pointing outside the active site (Phe81-Ser80: 5.6 Å, N-CA-CB-CG dihedral: -157°, Appendix Table A6) despite the larger residue at position 54. The conformation of Phe81 in the AtHNL scaffold differs from the prediction because the main chain positions of 54 are further away from Ser80 in AtHNL than in HbHNL. This difference creates a larger hydrophobic pocket outside the active site in AtHNL, allowing both Phe54 and Phe81 to fit into it, see Appendix Table A6. Experimentally, the enantioselectivity Hb-A10-H6_F54V and At-A10-H6_V54F did not change significantly as compared to the A10-H6
proteins, Appendix Table A3. At-A10-H6_V54F was still \((R)\) selective, while Hb-A10-H6_F54V was still \((S)\) selective. Variants with a valine at position 54 were twice as fast (0.9-1.0 sec\(^{-1}\)) as the equivalent enzyme with a phenylalanine at 54 (0.4-0.5 sec\(^{-1}\)). In agreement, modeling found that favored \((S)\)-mandelonitrile bound in CAC in Hb-A10-H6_F54V in 73\% of the simulation time, 99\% by the Asn mechanism. Favored \((R)\)-mandelonitrile bound in CAC in At-A10-H6_V54F in 45\% of the time, mostly referring to the oxyanion mechanism. Changes at position 54 reduced the largest differences in the shape of the consensus active sites, however it could not account for difference in reactivity and enantioselectivity.

Thirteen residues account for the remaining differences in the shapes of the active sites. Their positions differ between the AtHNL and HbHNL scaffolds by 1-2-fold more than the median movement during the simulation. At six of these locations, the amino acids differ between the two scaffolds, but at seven locations they are identical (18, 83, 121, 146, 152, 157, 158). Correlation analysis revealed a second region (besides residues 54-81-178 discussed above) where second shell residues alter positions of first shell residues. The positions of second shell residues Leu/Met153 and Asn/Arg156 positively correlated with the positions of first shell residues Phe152, Leu157 and Tyr158. Although these first shell residues are identical in two scaffolds, their positions differ, which could alter substrate positioning, as such they will be the future candidates for our investigations for altering enantioselectivity. Mutagenesis at second shell position 156 in a homologous HNL increased activity and \((S)\)-enantioselectivity, consistent with the role of this residue in orienting the substrate\(^{13}\). Residue 156 was not mutated in this study.

Overall, three structural elements contribute at least 1 kcal/mol to the activation energy differences between the two enantiomers yielding the observed enantioselectivity of these enzymes. These three structural elements are six active-site residues (13, 121, 125, 128, 131, and 209), residue 178, and residues outside the active site. The enantioselectivity of the two wild-type enzymes is >39, favoring opposite enantiomers. The difference corresponds to a \(\Delta\Delta\Delta G^\pm\) between the two wild-type enzymes of >4.3 kcal/mol\(^a\). AtHNL and At-A10-H6 differ by only six residues (13, 121, 125, 128, 131, and 209) near the aromatic ring binding region. Changing the six residues shifts the enantioselectivity from >39 to 5.3 (\(\Delta\Delta\Delta G^\pm >1.2\) kcal/mol),

\(^a\)The free-energy difference in the enantioselectivity of different enzymes (\(\Delta\Delta\Delta G^\pm\)) was given by \(\Delta\Delta\Delta G^\pm = -RT\ln(E_{wt}/E_{mut})\), where \(R\) is the gas constant, \(T\) the temperature (in Kelvin), \(E_{wt}\) the enantioselectivity of the wild-type enzyme, and \(E_{mut}\) the enantioselectivity of the mutant enzyme. The enantioselectivity values must be for the same enantiomer; thus, a change from 5.3 \((R)\) to 1.6 \((S)\) would be indicated as a change from 5.3 \((R)\) to 1.6 \(^{-1}\) \((R)\). The energy value, \(\Delta\Delta\Delta G^\pm\), has three delta symbols, which signify (1) the Gibbs free energy of activation for the reaction, (2) the difference in reactivity between enantiomers (3) the difference in enantioselectivity between enzymes.
and increased the number of catalytically active conformations found \textit{in silico} by $\sim$5-fold for the slow enantiomers while their numbers decreased slightly for the fast enantiomer. The second feature determining enantioselectivity is residue 178. Changing Phe to Leu at 178 changed enantioselectivity by $\Delta \Delta G^+$ values of 1.3 and $\geq 1.4$ kcal/mol when converting from the A9-H7 to the A10-H6 active sites in the AtHNL and HbHNL scaffolds, respectively. The third feature that determines enantioselectivity is residues outside of the active site. The enantioselectivity differences between identical composite active sites in differing scaffolds correspond to 1.7 kcal/mol in the A9-H7 sites and $\geq 1.9$ kcal/mol in the A10-H6 sites. While 121 residues outside of the active site differ between these two proteins, the overall structure is similar (root-mean-square (RMS) roughness of 0.74 Å between 889 peptide backbone atoms). This finding supports the other results that, even without major structural changes, residues outside of the active site are important for determining enantioselectivity$^{184,185}$. 
4.2 Characterization of aminophosphonic acid inhibitors of PALs

Our collaborators, Renzhe Qian and Friedrich Hammerschmidt in Vienna synthesized the aminophosphonic acids, whose interactions with PcPAL were tested in this study. Hence, their synthesis is only described in brief, with emphasis on the absolute configuration determination, as this will be crucial for interpreting the kinetic and thermodynamic results. A full description of the syntheses can be accessed in the published article.\(^{164}\)

\[\text{Figure 22. Structures of the investigated compounds. Note that the configuration of (R)-APEP and (R)-APPA corresponds to that of L-phenylalanine due to the higher CIP rank of the phosphonic acid moiety compared to the carboxylate group.}\]

First racemates of phosphonic acids shown on Figure 22 were prepared, and then the enantiomers were obtained by HPLC separation on a chiral stationary phase. The enantiomers of the intermediate tert-butyl (1-(diisopropoxyphosphoryl)-2-phenylallyl)carbamate were separated by preparative HPLC on achiral stationary phase (Chiralpak IC column; \(t_R =15.05\) and 30.97 min, both >99.6% enantiomerically pure). Figure 23 a. shows the (S) absolute configuration of the less polar (+)-tert-butyl (S)-(1-(diisopropoxyphosphoryl)-2-phenylallyl)carbamate as determined by single crystal x-ray structure analysis. To eliminate any ambiguity of the chirality of the final compounds, an urea derivative of the more polar enantiomer (-)-tert-butyl (R)-(1-(diisopropoxyphosphoryl)-2-phenylallyl)carbamate was prepared. Figure 23 b. shows the result of the single crystal x-ray structure analysis revealing that the urea substituted diisopropyl (R)-(1-(3-(4-bromophenyl)ureido)-2-phenylallyl)phosphonate had the (R)-configuration. By analogy, this proves, that enantiomers of APEP were correctly assigned as well, as they were prepared and separated by analogous methods.
Figure 23. X-ray crystallographic verification of the absolute configuration synthesis intermediate and derivative of APPA Ellipsoid plot (left) and stick (right) representations for 
a. tert-butyl (S)-(1-(diisopropoxypsonphoryl)-2-phenylallyl)carbamate intermediate and 
b. diisopropyl (R)-(1-(3-(4-bromophenyl)ureido)-2-phenylallyl)phosphonate derivative structures.

4.2.1 Steady state inhibition Studies with PcPAL

Initial reaction rates for deamination of L-Phe catalyzed by PcPAL were determined by UV spectrometry detecting the formation of the TCA product at 290 nm. Initial reaction rates ($v_0$) as a function of the substrate concentration determined the Michaelis constant ($K_m$) and the turnover number ($k_{cat}$) of the enzyme. Kinetic measurements in the presence of inhibitors determined the inhibition constants ($K_i$). The mechanism of inhibition could be determined by comparing the fit of experimental data to different inhibition models. Competitive, uncompetitive and non-competitive inhibition models are amongst the most common, hence we tested these models in our experiments. The structure of the tested compounds suggests similar binding as the substrate, hence competitive inhibition is the most likely mechanism.

As expected, (R)-APEP – the phosphonic acid analogue of the L-Phe – was one order of magnitude stronger inhibitor than the opposite enantiomer (S)-APEP (Table 11.). In close agreement with data from the literature$^{118}$, the competitive inhibition model with $K_i= 0.66 \mu$M fitted best to the experimental data for (R)-APEP. The $K_i= 4.28 \mu$M value for (S)-APEP was also in agreement with the previous results on inhibition of PAL from maize$^{118}$. The enantiomers of aminophosphonic acids APEP and APPA were purified to ≥99.6% ee values, hence inhibition by contaminating opposite enantiomers could be ruled out.
Table 11. Apparent inhibition constants and binding equilibrium constants of the aminophosphonic acids shown on Figure 22.

<table>
<thead>
<tr>
<th>Inh.</th>
<th>Type of inh.#</th>
<th>$K_i$ [$\mu$M]</th>
<th>$K_i_{\text{lit}}$ [$\mu$M]</th>
<th>$K_{d,\text{ITC}}^*$ [$\mu$M]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(R)$-AEP</td>
<td>SB-Comp.</td>
<td>$0.66 \pm 0.11$</td>
<td>$1.5^{118#}$</td>
<td>$2.7$ [1.4 – 5.1]</td>
</tr>
<tr>
<td>$(S)$-AEP</td>
<td>Comp.</td>
<td>$4.28 \pm 0.11$</td>
<td>$11.6^{118#}$</td>
<td>$5.2$ [4.4 – 6.3]</td>
</tr>
<tr>
<td>$(R)$-APPA</td>
<td>Comp.</td>
<td>$0.64 \pm 0.02$</td>
<td>$1.1$</td>
<td>$[0.8 – 1.5]$</td>
</tr>
<tr>
<td>$(S)$-APPA</td>
<td>SB-Comp.</td>
<td>$0.04 \pm 0.01$</td>
<td>$0.04$</td>
<td>[0.02 – 0.07]</td>
</tr>
<tr>
<td>rac-APMP</td>
<td>Comp.</td>
<td>$41.5 \pm 5.6$</td>
<td>$6.5^{127}$</td>
<td>n.d.</td>
</tr>
<tr>
<td>rac-AMPP</td>
<td>No inh.</td>
<td>$&gt;1000$</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>rac-APCP</td>
<td>No inh.</td>
<td>$&gt;1000$</td>
<td>n.d.</td>
<td></td>
</tr>
</tbody>
</table>

# Best fitting inhibition model to the experimental data. Comp.: competitive; SB-Comp: competitive with slow binding of the inhibitor; No inh.: no measurable inhibition.

* Confidence intervals of the binding affinity of the parameter fittings are provided as a measure of uncertainty to the given values. Details of the ITC measurements are described in the next section. n.d.: not determined.

## Interestingly, introduction of a single methylidene group at the C2-position of phosphonic acid analogue AEP resulting in a novel inhibitor APPA, reversed the enantiopreference of inhibition of PcPAL. In fact, (S)-APPA was one order of magnitude stronger inhibitor than (R)-AEP. The inhibition constant of (S)-APPA ($K_i = 0.04$ $\mu$M) with PcPAL was only slightly higher than that of AIP – the most potent aminophosphonic PAL inhibitor ($K_i = 0.025$ $\mu$M)$^{115}$. The fact that (S)-APPA was two orders of magnitude stronger inhibitor than (S)-AEP, suggested that the methylidene group had a significant effect on binding of the (S)-enantiomers. In contrast, the inhibition constant of (R)-AEP and (R)-APPA are similar. This is likely due to steric effects in the enzyme active site, as the electronic effects of the methylidene group should be similar in both enantiomers.

The phenylglycine analogue rac-APMP was a weak competitive inhibitor of PcPAL. Competitive inhibition model with $K_i = 41.5$ $\mu$M, described best the measured initial reaction rates ($v_0$). The almost one order of magnitude difference compared to previous data$^{127}$ is considerable, nonetheless rac-APMP was the weakest one amongst the effective PAL inhibitors. The aromatic binding pocket of PcPAL suited to accommodate a benzyl moiety seems to be too large for the smaller phenyl ring of rac-APMP, hence the hydrophobic interactions with the aromatic binding region of the active site pocket are less pronounced.

In contrast to the preliminary expectations, rac-AMPP and rac-APCP showed no measurable inhibition with PcPAL up to 1 mM concentration. This was in contrast to results on Pseudomonas putida HAL (sequence identity to PcPAL: 21%) which was irreversibly inhibited by L-homocysteine at 9 mM concentration after incubation for 120 min in the presence of oxygen$^{124}$. Initial velocities of the TCA formation with PcPAL were identical when
the reaction was started with the addition of PcPAL to the test in the presence or in the absence of rac-AMPP or rac-APCP or the enzyme was preincubated with rac-AMPP for 60 min prior to activity measurement. In case of rac-APCP, the sterically demanding cyclopropyl substituent at the β-carbon atom presumably hindered the binding to PcPAL.

4.2.2 Slow binding inhibition of PcPAL by (R)-APEP and (S)-APPA

Addition of (S)-APPA to the PcPAL-catalyzed reaction resulted in nonlinear progress curves, under the same conditions where the non-inhibited or (R)-APPA inhibited PcPAL-catalyzed reactions showed linear progress curves, as shown on Figure 24. The nonlinear progress curve corresponds to a change in the enzymatic activity in a time dependent manner. The nonlinear rate curves observed on the steady-state time scale due to transient kinetics suggested slow binding mechanism for the inhibitors.

This behavior may originate from different molecular phenomena: (i) if there is a transition-state-mimicking intermediate, (ii) if the inhibitor exists in multiple forms, only one of which may interact with the enzyme, (iii) if only a rare form of the enzyme binds the inhibitor or (iv) if after an initial fast binding, slow structural rearrangements form a second inhibitory complex. In kinetic analysis of slow binding inhibitors, the integrated rate equation for the product concentration (P) was derived by Frieden (Eq. 1).

\[
P = \frac{v_z - v_s}{\lambda} \left(1 - e^{-\lambda t}\right) + v_s t + d
\]  

This integrated rate equation determines the initial velocities \(v_z\), the steady-state velocities \(v_s\) and the frequency constant of the exponential phase \(\lambda\). Steady state velocities determine the apparent inhibitory constants. The two simplest slow binging inhibition mechanisms – shown as Eq. 2 and Eq. 3 – can be distinguished by the initial velocities as a function of the inhibitor concentration.

\[
\begin{align*}
E + I & \overset{k_1}{\underset{k_{-1}}{\xrightarrow{\text{slow}}} E + I} \\
E & \overset{k_1}{\underset{k_{-1}}{\xrightarrow{\text{slow}}} E + I} \\
E + I & \overset{k_1}{\underset{k_{-1}}{\xrightarrow{\text{slow}}} E + I} \\
E + I & \overset{k_1}{\underset{k_{-1}}{\xrightarrow{\text{slow}}} E + I}
\end{align*}
\]

According to the first mechanism represented by Eq. 2, binding of the inhibitor may be a slow process, hence \(v_z\) is independent of the inhibitor concentration. According to another
slow binding mechanism represented by Eq. 3, an EI complex forms rapidly, which subsequently undergoes changes to form the strongly bound (EI)* complex. Due to the first fast step in the second mechanism, $v_z$ depends on the inhibitor concentration.

![Figure 24](image-url)

**Figure 24.** Production curves of tras-cinnamic acid [TCA] formation from 1mM l-Phe catalyzed by PcPAL in the presence of various amounts of enantiopure (R)- and (S)-APPA. 

- **a.** Progress curves at various concentrations of (R)-APPA for 5 min (for clear visibility only every 20th data point is plotted).
- **b.** Initial part of the progress curves at various concentrations of (R)-APPA (total time 0.3 min). Linear equation was fitted to all the progress curves in **a** and **b.**
- **c.** Progress curves at various concentrations of (S)-APPA for 5 min (for clear visibility only every 20th data point is plotted). The curve in blue with empty dots shows the production of TCA in an assay containing PcPAL preincubated with (S)-APPA (2 µM).
- **d.** Initial part of the progress curves at various concentrations of (S)-APPA (total time 0.3 min, for better visibility only production curves at 10 µM and 50 µM inhibitor concentrations are plotted). Linear equation was fitted to the non-inhibited progress curve ($\blacklozenge$) and the integrated rate equation (Eq. 1) was fitted to the (S)-APPA inhibited progress curves ($\bullet$, $\blacktriangle$, $\blacklozenge$, $\bigcirc$).

Figure 24 b. shows the first 0.3 min of the progress curves of the product formation from l-Phe catalyzed by PcPAL in absence and in presence of (S)-APPA. From the curves it is clearly visible that even the early slope of the curves depended on the inhibition concentrations. Parameter fitting confirmed this observation. Thus the two-step slow binding mechanism described by Eq. 3, where after an initial fast binding process a slow change took place is most likely the origin of the non-linear product formation progress curves. Interestingly, the
frequently used MIO enzyme inhibitor AIP, also showed slow binding characteristics with PcPAL. However, the authors suggested the one-step slow binding mechanism (Eq. 2) for AIP.

The enantiopure (R)-APEP also showed slow binding behavior, but to a smaller extent than (S)-APPA. Interestingly, (R)-APPA was not a slow binding inhibitor. Differences in the binding kinetics indicated that although the inhibition constants were similar ($K_i = 0.64 \mu M$ for (R)-APPA and $K_i = 0.66 \mu M$ for (R)-APEP), the binding mechanism was perturbed by the addition of the methylidene group at the C2 position.

Several possibilities could account for this behavior. One possibility may be that deprotonation of the enzyme or the ligand after the fast binding further enforces the interactions further. A second possibility may be that in the slow change step a nucleophilic part of the ligand forms a Michael adduct with the methylidene group of the MIO. Finally, TyrA of the mobile loop of the enzyme may also attack the methylidene group of the ligand.

Decrease in the product formation rate after the slow binding step might happen due to an irreversible covalent binding. However, the progress curves determined after preincubation of PcPAL with (S)-APPA suggested reversible binding. The blue progress curve with empty circles of Figure 24 a was measured when the reaction was started with addition of L-Phe to the assay after preincubation of PcPAL with (S)-APPA for 5 min. Increase of activity to a steady state level indicated reversible binding of (S)-APPA. Additionally, variation of the preincubation time between 1–60 min did not alter the progress curves suggesting that the fast binding step was completed within 1 min.

Later on (R)-APEP and (S)-APPA bound PcPAL crystal structures showed that both inhibitors bound covalently to the MIO group by their amino group. The phosphonic acid moiety of the inhibitors bind to TyrA with short covalent like hydrogen bond. Formation of either of these two interactions between the protein and the inhibitor might be the origin of the slow binding. Unfortunately, structural data is not yet available for (S)-APEP and (R)-APPA binding, therefore the effect of the methylidene group on substrate binding and the reasons for the reversal of the enantiopreference of PcPAL are not yet fully understood. A more detailed analysis of inhibitor binding conformations in PcPAL will be presented in chapter 4.3.1.2.

4.2.3 Inhibition of AvPAL and RtpAL by APEP and APPA

In order to ascertain that APEP and APPA are general inhibitors of PALs, two other PALs were tested. MIO enzymes from prokaryotic and eukaryotic sources differ structurally. Although MIO enzymes of both origins are tetramers, eukaryotic MIO enzymes contain an
additional domain at their C-terminal, rendering each monomeric chain longer by approximately 200 amino acids. Thus, inhibition of the natural reaction of a further PAL of eukaryotic origin (RtPAL, 34% sequence identity to PcPAL), as well as of another PAL of prokaryotic origin (AvPAL, 27% sequence identity to PcPAL) by APEP and APPA were measured. Experiments demonstrated that enantiomers of both APEP and APPA are general inhibitors of PALs. The enantiopure (R)-APEP and (S)-APEP, as well as (R)-APPA and (S)-APPA inhibited AvPAL and RtPAL to a similar degree as they inhibited PcPAL, Table 12. Additionally, methyldiene group at the β-position reversed the enantiopreference of both RtPAL and AvPAL as well. Hence, probably all PALs generally bind (S)-APPA tighter than (R)-APPA. Thus changes in binding conformation induced by the presence of this methyldiene group could provide key information into the molecular mechanism of PAL enantioselectivity and suggest strategies for engineering PAL enantioselectivity. Moreover, the inhibitors described here might happen to be general, strong MIO enzyme inhibitors, by analogy to AIP. Although AIP has a non-substituted aromatic ring, previous results showed that it could efficiently inhibit TAL besides inhibiting PAL74.

**Table 12. Reduction of reaction rate of RtPAL and PcPAL at 5 mM concentration of l-Phe by the enantiomers of APEP and APPA**

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>$c_{inh}$ [µM]</th>
<th>RtPAL</th>
<th>PcPAL</th>
<th>$c_{inh}$ [µM]</th>
<th>AvPAL</th>
<th>PcPAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)-APEP</td>
<td>200</td>
<td>30%</td>
<td>35%</td>
<td>100</td>
<td>22%</td>
<td>14%</td>
</tr>
<tr>
<td>(S)-APEP</td>
<td>200</td>
<td>40% *</td>
<td>55% *</td>
<td>100</td>
<td>74%</td>
<td>22% &amp;</td>
</tr>
<tr>
<td>(R)-APPA</td>
<td>50</td>
<td>53%</td>
<td>69%</td>
<td>100</td>
<td>28%</td>
<td>23%</td>
</tr>
<tr>
<td>(S)-APPA</td>
<td>50</td>
<td>10%</td>
<td>4%</td>
<td>100</td>
<td>2%</td>
<td>2%</td>
</tr>
</tbody>
</table>

* $c_{inh}$= 1000 µM (S)-2 reduced the velocity to 10%.
* Value measured at $c_{inh}$= 1000 µM inhibitor concentration.
* Value measured at $c_{inh}$= 100 µM inhibitor concentration.
* Value measured at $c_{inh}$= 200 µM inhibitor concentration.
& Value measured at $c_{inh}$= 1000 µM inhibitor concentration.

4.2.4 **Binding thermodynamics**

Binding affinities ($K_d$) measured by ITC agreed well with of the inhibition constants of both enantiomers of the APEP and APPA (Table 11). Representative thermograms of wt-PcPAL with each ligand are shown on Figure 25. The small differences could originate from the different protein concentration used in the kinetics measurements and for ITC, or the slight difference in the pH (8.8 for kinetics and 8.0 for ITC).
Figure 25. Representative ITC thermograms of wt-PcPAL with ligands: a. (R)-APEP, b. (S)-APEP, c. (R)-APPA, d. (S)-APPA, e. TCA, f. D-Phe

Figure 26. Thermodynamic binding energies to wt-PcPAL of each enantiomer of APEP and APPA with measured by ITC. Free energies for binding ($\Delta G$) were calculated from the measured $K_a$, using the equation $\Delta G = -RT\ln K_a$. The enthalpy contributions were obtained by fitting to the titration curves and entropies were calculated by $\Delta G = \Delta H - T\Delta S$. 
Binding of both enantiomers of APEP or APPA was enthalpy driven (Figure 26, Appendix Table A7). Enthalpy contributions of the binding process originate from hydrophobic interactions, hydrophilic interactions and salt bridges between the ligand and the protein. The non-significant binding enthalpy difference between (R)-APEP and (R)-APPA (-0.4 kcal/mol), revealed that the methyldiene group had negligible effect on the binding affinity for molecules with (R) configuration. On the other hand, presence of the methyldiene group at C2 position increased the binding enthalpy of the (S)-enantiomer by 3.1 kcal/mol. This effect is in the range of the formation of an additional hydrogen bond between the protein and the ligand.

On one hand solvent displacement may contribute favorably to the entropy of binding. On the other hand, if an initially less ordered region – like a loop – adopts a more rigid conformation, the entropy of the system decreases. Thus, in such instance the overall binding free energy becomes less negative. ITC measured the lowest entropy contributions for the slow binding inhibitors (R)-APEP and (S)-APPA (-1.2 kcal/mol and -0.6 kcal/mol, respectively, Figure 26, Appendix Table A7). ITC measured a somewhat larger entropy contribution to the binding for (S)-APEP and (R)-APPA (-1.7 kcal/mol and -2.1 kcal/mol, respectively, Figure 26, Appendix Table A7). As all of these compounds occupy the same binding pocket, the solvent entropy contributions should be equivalent for their binding. The smaller entropy contribution with the slow binding inhibitors (R)-APEP and (S)-APPA is therefore likely due to a larger reduction of the entropy of the system by more constraining the flexibility of protein loop(s).

Crystallographic B-factors show\(^{83}\) that the most flexible region of the protein neighboring the active site is the inner loop, suggesting that APEP and APPA forms a direct interaction with the inner loop and constrain its flexibility. To further investigate this hypothesis, additional ITC measurements were carried out with PcPAL-Y110F aimed to determine the contribution of TyrA (Figure 8) to the binding energy.

4.2.4.1 Perturbation of ligand binding to PcPAL by the Y110F mutation

The possible role of TyrA in the elimination reaction is presented on Figure 8. It was shown that mutation of TyrA (at position 110 in PcPAL) to phenylalanine resulted in an inactive protein. By assuming that Y110F mutation has no effect on the overall protein structure, the effect of the mutation on the binding energy can be determined directly by using ITC. As the Y110F mutation lead to catalytically inactive protein\(^{75}\), ITC measurements with this mutant could provide additional details on ligand binding thermodynamics to PcPAL.
Representative thermograms of PcPAL-Y110F with each ligand are shown on Appendix Figure A1.

The Y110F mutation in PcPAL weakened the binding of all investigated molecules (Table 13, Appendix Table A7). Comparison of the binding characteristics of several compounds to wt-PcPAL and PcPAL-Y110F revealed (Table 13) small contribution of Tyr110 to the binding (smaller than required for a weak hydrogen bond, <0.8 kcal/mol). The effect of Y110F on the binding of TCA served as control. The planar orientation of the β-hydrogen on TCA, and its position in the crystal structure suggest that there is no direct interaction between Tyr110 and the TCA. Thus, differences in the binding properties of TCA to the two PcPAL variants quantified the maximum perturbation caused by Y110F mutation in the binding of substrates with no direct interaction with Y110. Hence, effects with ΔΔG <0.6 kcal/mol could be due to small perturbation of the structure and/or natural uncertainty of the experiments.

Table 13. Equilibrium binding constants and binding free energy perturbation measured by ITC for various ligands wt-PcPAL and PcPAL-Y110F.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>PcPAL-wt $K_d$ [µM]</th>
<th>PcPAL-Y110F $K_d$ [µM]</th>
<th>ΔΔG [kcal/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Phe</td>
<td>12.6</td>
<td>29.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>L-Phe</td>
<td>n.a. *</td>
<td>23.0</td>
<td>n.a. *</td>
</tr>
<tr>
<td>(R)-APEP</td>
<td>2.7</td>
<td>8.9</td>
<td>-0.7</td>
</tr>
<tr>
<td>(S)-APEP</td>
<td>5.2</td>
<td>6.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>(R)-APPA</td>
<td>1.1</td>
<td>35.3</td>
<td>-2.1</td>
</tr>
<tr>
<td>(S)-APPA</td>
<td>0.04</td>
<td>0.17</td>
<td>-0.9</td>
</tr>
<tr>
<td>TCA</td>
<td>6.5</td>
<td>16.6</td>
<td>-0.6</td>
</tr>
</tbody>
</table>

* Differences in the binding free energies show the contribution of the hydroxyl group of Tyr110 to the binding energies. ΔΔG = Δ$G_{wt}$ − Δ$G_{Y110F}$.

* n.a. not applicable. The natural substrate L-Phe reacts with wt-PcPAL, therefore binding thermodynamics cannot be measured by ITC.

Shift in the equilibrium binding constants demonstrated that Tyr110 is directly involved in binding of (R)-APEP, but only minor change is observed in binding free energy of (S)-APEP. Comparing the ΔΔG values in Table 13 reveals that (R)-APEP binding is weakened by 0.7 kcal/mol, while it is weakened only by 0.1 kcal/mol for (S)-APEP. However, binding thermodynamics of (S)-APEP are also perturbed by the Y110F mutation. Appendix Table A7 shows that the enthalpy contribution to the binding increases while the entropy contribution decreases resulting in the small change in the overall binding free energy, ΔG. On the whole it is clear that Y110F directly influences the binding of both (R)- and (S)-APEP. Unfortunately the fitted ΔH values are quite unreliable due to the limiting protein concentration and the unusually low molar ratios, therefore drawing any further conclusion from these ITC results to the binding of APEP would be unreasonable.
Shift in the equilibrium binding constants demonstrated that Tyr110 is directly involved in binding of both APPA enantiomers. The perturbations in the binding free energies ($\Delta\Delta G = 2.1$ kcal/mol for $(R)$-APPA and $0.9$ kcal/mol for $(S)$-APPA, Table 13) were in the range of a weak hydrogen bond. First we considered that the hydroxyl group of Tyr110 might form a Michael adduct with the methyldene group of APPA, this potentially slow process might accounts for the slow binding behavior of $(S)$-APPA. Nonetheless the crystal structure disproved the Michael adduct hypothesis, and showed that phosphonic acid moiety of (S)-APPA forms a short covalent like hydrogen bond to Tyr110.

ITC measured a dissociation constant of $12.6 \, \mu M$ for $D$-Phe from wt-PcPAL. This value was in the same order of magnitude as $K_m$ of L-Phe. This argued that the same set of molecular interactions were involved in binding of $D$- and $L$-Phe to PcPAL. This is in full agreement with the early model suggesting mirror image packing of the substrate enantiomers within PAL, and directly proves that $D$-phenylalanine can bind to the active site in a non-reactive conformation. The similarity of equilibrium binding constants to the PcPAL-Y110F mutant of the enantiomers of phenylalanine ($K_d = 29.1 \, \mu M$ for $D$-Phe and $23.0 \, \mu M$ for $L$-Phe, respectively, Table 13) provides strong evidence for their mirror image packing within PcPAL. Small perturbation of the binding free energy of $D$-Phe by the Y110F mutation suggest, in agreement with the mechanism shown on Figure 8, a direct role of Tyr110 (TyrA) in catalysis and a minor role in substrate binding.
4.3 Ligand binding in MIO enzymes

4.3.1 Crystal structures of PcPAL

4.3.1.1 X-ray structure of PcPAL with catalytically competent inner loop conformation.

The crystal structures of PcPAL have been solved in apo form (PDB ID 6H2O) and in complex with two phosphonic acid inhibitors described in the previous chapter\textsuperscript{164}, (S)-APPA (PDB ID 6F6T) and (R)-APEP (PDB ID 6HQF) (Table 14). PcPAL crystallized in the same unit cell as the previously published structure 1W27\textsuperscript{83}. Two monomers reside in the asymmetric unit, and the biologically functional tetramer is observable by applying crystal symmetry operations. The overall protein fold is very similar to the previously solved model, and to other eukaryotic MIO enzymes\textsuperscript{62,78,90-94} (Figure 27a), however key details were discovered in the present structures. Inhibitor binding decreased the flexibility of the inner loop and the experimental electron density maps provided an unequivocal structural model. Table 14 gives an overview of the structures, and Table 15 lists the details of the data collection and structure refinement.

**Table 14. Summary of PcPAL structures**

<table>
<thead>
<tr>
<th>PDB Code</th>
<th>Resolution</th>
<th>Residues resolved by the electron density map</th>
<th>Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>6H2O</td>
<td>1.90</td>
<td>A: 25 74 79 106 127 334 348 549 557 716</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B: 25 75 78 106 127 333 348 549 556 716</td>
<td></td>
</tr>
<tr>
<td>6F6T</td>
<td>1.90</td>
<td>A: 25 337 346 549 557 716</td>
<td>(S)-APPA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B: 25 333 346 551 554 716</td>
<td></td>
</tr>
<tr>
<td>6HQF</td>
<td>1.76</td>
<td>A: 25 334 347 551 557 716</td>
<td>(R)-APEP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B: 25 337 348 551 559 716</td>
<td></td>
</tr>
</tbody>
</table>

The most notable difference between apo (Figure 27 cyan colored) and ligand bound structures (Figure 27 green and purple colored) is in the inner loop conformation. The inner-loop (106-126) directly caps the active site of MIO enzymes (Figure 27 a-b.). The conserved catalytic Tyr110 (TyrA on Figure 8) lies at the first part of the inner loop. The central 8 amino acids of the loop (109-117) are conserved in all MIO enzymes, as demonstrated by the sequence logo presented on Figure 27 b created based on the alignment shown in Table 4. In the apo form electron densities for the inner loop region are absent, as a result coordinates cannot be assigned to these residues, Figure 27 c. Binding of phosphonic acid inhibitors renders electron
densities for the inner loop region clearly visible, as a result coordinates can be assigned to these residues, Figure 27 d-e. This is for the first time that clear electron densities for the entire inner loop of a eukaryotic PAL are observed. The electron densities show Tyr110 to be sufficiently close to the MIO to fulfill its role as a catalytic base, as clearly indicated on Figure 28. The conformation of the inner loop is also very similar to the “loop-in conformation” of other MIO enzymes. Figure 11 a showed a structural overlay of the inner loop regions of representative examples of MIO enzymes of different enzyme classes. Importantly, none of the experimental data presented here supports the previously proposed “loop-out conformation”83, therefore its functional relevance is questionable.

**Figure 27. Global features of the PcPAL structures.** a.) Overlay of the cartoon representation of the three PcPAL structures (6H2O cyan, 6F6T green, 6HQF purple), with a close up view in to the active site highlighting the inner and outer loops covering it. b.) Sequence logos of the inner and outer loop regions created from sequence alignments of all MIO enzymes with known 3D structures, with residue numbers and the sequence of PcPAL explicitly shown. Table 4 shows the alignments in more detail. c-e.) Electron densities for the inner and outer loops in the three structures on the 2Fo-Fc map, contoured at 1σ level. c.) Electron densities are absent for both of the loops in the apo structure, gray cartoons show loop conformations in other MIO enzyme structures (6F6T for the inner loop, 3NZ4 TcPAM 92 for the outer loop). d-e.) The inner loops are visible in the electron density maps of the structures obtained with the inhibitors, while electron densities are absent for the outer loop residues, hence gray cartoons show loop conformations like for the apo structure.
A second notable feature of the structures concerns the outer loop. The outer loop (328-350) caps the inner loop of an adjacent monomer. The only conserved residues are Ser328 at the beginning of the loop and Gln348 and Asp349 at the end of the loop, Figure 27 b. The absence of clear electron densities in all three structures (shown on Figure 27 c-e) for this region of the protein suggest that this loop is highly flexible even after ligand binding. This high flexibility is a general feature in eukaryotic PALs, these residues remain unordered in prokaryotic PALs as well. Figure 12 showed a comparison of the electron densities for outer loop region representative examples of MIO enzymes of different enzyme classes. Structure overlay shown on Figure 11 b highlights the diversity of conformations sampled by outer loops of the different structures of MIO enzymes. Based on the proximity of the outer loop to the active site, a regulatory role in substrate binding was suggested.

Table 15. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>PDB ID Code</th>
<th>PcPAL Apo 6H2O</th>
<th>PcPAL (S)-APPA 6F6T</th>
<th>PcPAL (R)-APEP 6HQF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C 2 2 21</td>
<td>C 2 2 21</td>
<td>C 2 2 21</td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a,b,c(Å)</td>
<td>119.89, 161.12, 141.43</td>
<td>118.74, 161.10, 141.65</td>
<td>118.88, 160.95, 141.58</td>
</tr>
<tr>
<td>α,β,γ(°)</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)#</td>
<td>48.09-1.90</td>
<td>95.58-1.90</td>
<td>38.70-1.72</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>104777 (13880)</td>
<td>106555 (37601)</td>
<td>141190 (9135)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.5 (4.6)</td>
<td>6.8 (6.9)</td>
<td>6.9</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.5 (91.9)</td>
<td>99.9 (99.9)</td>
<td>98.7 (87.0)</td>
</tr>
<tr>
<td>I/σI</td>
<td>8.78 (1.39)</td>
<td>8.88 (2.50)</td>
<td>8.52 (0.60)</td>
</tr>
<tr>
<td>R Work (%)</td>
<td>0.144 (1.485)</td>
<td>0.150 (0.787)</td>
<td>0.161 (3.010)</td>
</tr>
<tr>
<td>CCR2</td>
<td>0.99 (0.52)</td>
<td>1.00 (0.85)</td>
<td>0.99 (0.17)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td>Resolution: 38.70-1.76</td>
</tr>
<tr>
<td>Wilson B-Factor (Å2)</td>
<td>28.0</td>
<td>19.2</td>
<td>30.8</td>
</tr>
<tr>
<td>R Work (%)</td>
<td>0.170</td>
<td>0.192</td>
<td>0.168</td>
</tr>
<tr>
<td>R free (%)</td>
<td>0.206</td>
<td>0.239</td>
<td>0.205</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.0071</td>
<td>0.0083</td>
<td>0.0078</td>
</tr>
<tr>
<td>Bond angle (°)</td>
<td>0.78</td>
<td>0.94</td>
<td>0.82</td>
</tr>
<tr>
<td>Ramachandran*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>98.42</td>
<td>98.58</td>
<td>98.34</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0.08</td>
<td>0.00</td>
<td>0.23</td>
</tr>
<tr>
<td>Clashscore</td>
<td>2.37</td>
<td>2.90</td>
<td>1.80</td>
</tr>
<tr>
<td>Average B-factor</td>
<td>33.0</td>
<td>28.0</td>
<td>36.0</td>
</tr>
</tbody>
</table>

# Statistics for the highest-resolution shell are shown in parentheses.
* Categories were defined by PHENIX152. All non-glycine residues are included for this analysis.
4.3.1.2 Quasi-mirror image binding of inhibitors and its mechanistic consequences

Co-crystallization of PcPAL with (R)-APEP and (S)-APPA probed the reaction mechanism and the molecular basis of the enantioselectivity of the PAL reaction. Presence of methylidene group at the β-position reversed the enantiopreference of the enzyme; the (S)-APPA bound one order of magnitude tighter ($K_i = 0.04 \ \mu M$) to the enzyme than the phosphonic acid analogue ((R)-APEP, $K_i = 0.66 \ \mu M$) of the natural substrate L-Phe.\(^{164}\) Co-crystallization with 10-fold excess of either inhibitor resulted in full occupancy of the active sites. Figure 28 shows the electron densities observed in the active site of PcPAL structures. The apo structure contains two small blobs modeled as water molecules (Figure 28 a). Perfect electron densities define even the chirality of the inhibitors for the complex structures (Figure 28 b-c).

Figure 28. Comparison of the active site electron densities of PcPAL and Rhodobacter sphaeroides TAL (RsTAL)\(^{74}\) structures. Electron densities (blue mesh, 2Fo-Fc map, contoured at 1σ level) are displayed for the catalytically essential residues (Tyr110, Mio203, Tyr351) and for residues in direct interactions with the hydrophilic part of the ligands (Gln348, Arg354, Asn384). a.) apo structure of PcPAL, 6H2O: cyan. b.) (S)-APPA bound PcPAL structure, 6F6T: green c.) (R)-APEP bound PcPAL structure 6HQF: purple. d.) (2-amino-2,3-dihydro-1H-inden-2-yl)phosphonic acid (AIP) bound RsTAL structure 2O7E\(^{74}\): salmon. e.) overlay of the two inhibitor bound PcPAL active sites. Note how the two carbon atoms of the pentene ring of AIP occupy the same position as the β-carbon atom in the (R) and (S) enantiomers of the phosphonic acid inhibitors.
Surprisingly, (S)-APPA and (R)-APEP bind covalently to the MIO with their amino groups, despite only reversibly inhibiting the enzyme. Hence, the covalent binding of the inhibitors to the enzyme must also be reversible. The covalent bond formation may also explain the slow binding kinetics of the inhibition. This finding provides the first direct experimental evidence for the N-MIO mechanism (Figure 8 a) for PAL.

The phosphonic acid inhibitors are structurally very similar to AIP one of the highest affinity inhibitors ($K_i= 25$ nM) of PcPAL known to date$^{82}$. The binding mode of the chiral phosphonic acid inhibitors is very similar to that of AIP observed in Rhodobacter sphaeroides TAL (RsTAL)$^{74}$ (Figure 28 d). Additionally, the longer distance between Tyr60 (Tyr110) and the phosphonic acid moiety (2.9 Å compared to 2.2 Å and 2.4 Å), might explain the larger $K_i$ in RsTAL (16.3 µM)$^{74}$ compared to PcPAL (25 nM).

The inhibitor bound structures also revealed that phosphonic acid moiety of the inhibitors forms a short covalent like hydrogen bond to Tyr110, 2.2 Å in (S)-APPA and 2.4 Å for (R)-APEP (Figure 28 e). This interaction potentially facilitates ordering of the inner loop more rigid so that the related electron densities become visible. This corroborates previous ITC results, which suggested that Tyr110 directly participates inhibitor binding, Table 1$^{164}$. The phosphonic acid moiety also interacts with the carboxylic acid binding residues of the active site; Gln348, Arg354 and Asn384. As a result, the side chain of Gln348 is slightly reoriented (Figure 28 a compared to Figure 28 b-c), and the whole outer loop is slightly stabilized (Table 14 and Figure 27 c-e). As expected, the aromatic moiety of the inhibitors lies in the hydrophobic region of the active site, lined by residues Phe116, Gly117, Leu134, Phe137, Leu138, Leu206, Leu256, Val259, Lys456, Ile460 as shown on Figure 10.

Position of the β-carbon atom differs the most between (S)-APPA and (R)-APEP (Figure 28 e). The two β-carbon atoms of the cyclopentene ring of AIP are exactly at the same position as the β-carbon atoms (S)-APPA and (R)-APEP, substantiating the β-carbon position’s role in enantioselectivity. This difference is likely to be a general feature determining the enantioselectivity of the enzyme, and not an effect of the extra methyldene group of (S)-APPA. Fascinatingly, the binding mode matches precisely the one predicted in 1981 based on kinetic parameters of transition state analogue inhibitors$^{123}$.

The strict L-selectivity of AL have been exploited in many synthetic applications$^{46}$. However, this same enantioselectivity also set back direct asymmetric synthesis of D-amino acids by AL. Kinetic resolution, the stereoselective removal of the L-enantiomer, has been successfully applied to natural$^{189}$ and non-natural amino acids$^{60,76}$, but conversion remains limited to 50%. Deracemization, the two-step conversion of the L-enantiomer to either the
D-enantiomer or the racemic mixture, through a nonchiral intermediate offers 100 % theoretical yields\textsuperscript{113}, but is a more complicated process. Interestingly, PALs exhibit variable enantioselectivity for some non-natural aminoacids, a MIO-independent mechanism, where TyrA (Figure 8) still fulfills the role of catalytic base, is suggested to produce D-enantiomers\textsuperscript{112}. Based on the binding mode of the phosphonic acid inhibitors, crystal structures 6F6T and 6HQF enables us to assume the binding conformations of D- and L-Phe, and thus permit speculations about the molecular mechanisms of enantioselectivity. The most notable difference in the quasi mirror image binding of (S)-APPA (D-Phe like binding) and (R)-APEP (L-Phe like binding) is in the position of the β-carbon atom. Different positioning of the β-carbon atom will hinder TyrA in fulfilling its role as a catalytic base, as a consequence, the spatial proximity of TyrA to the either of the β-protons in the bound conformation of the substrates determines the enantioselectivity of the MIO-enzymes. This suggest two strategies for altering the enantioselectivity of the enzyme: repositioning TyrA by altering the loop conformation or repositioning of the β-carbon atom of the substrate to allow proton removal. This second strategy might have already been employed when lower enantioselectivity were observed for non-natural substrates\textsuperscript{112,113}.

The structure 1W27\textsuperscript{83} has long been the only eukaryotic PAL structure with its inner and outer loops modelled. Early on, it was suggested that the loop-out conformation of its inner loop must be a catalytically non-competent conformation\textsuperscript{132}. Nonetheless, some studies explained effects of mutations on enzymatic activity, by simulating large scale loop opening motions\textsuperscript{78,131}. Crystal structures 6F6T and 6HQF show well defined electron densities in the catalytically competent loop-in conformation, and hopefully they will provide reliable starting points in the future for enzyme engineering and computational QM/MM studies for precise evaluation of the PAL reaction mechanism.

### 4.3.2 Substrate egress dynamics in PAL and PAM

The active site of MIO enzymes is shielded from the solvent by the inner loop, which in turn is partially covered by the outer loop (Figure 11). Protein crystal structures suggested that substrate access to the active site proceeds by a narrow tunnel between the inner and outer loops\textsuperscript{89}. Others suggested that the inner loop transits to the loop-out conformation during substrate binding, and that differences in this motion distinguish between AL and AM activity\textsuperscript{78,131}. RAMD simulations map the ligand egress pathways from buried pockets, and the same simulations may suggest the extent of inner loop opening upon ligand binding. In order to investigate the differences in the inner loop dynamics in AL and AM, we ran the RAMD
simulations with the same setting for the PcPAL and TcPAM, simulating the binding of the substrate L-Phe and the release of the product TCA.

Only one of the large number of protein conformations that exists in even a purified protein sample is present and modeled in the crystal structure. Furthermore, partial homology modeling of the missing loops for the crystal structures assigned only one initial conformation to these flexible protein segments. Finally, modeling of water molecules and ions aimed to reduce the overall energy of the model system to avoid steric clashes and eventually the explosion of the model system. Gradual heating and equilibration procedure prior to the random acceleration MD simulations overcame these hardships. RMSD of main chain atoms, already shown on Figure 17 demonstrated that all simulations reached equilibrium after approximately 2 ns.

RMSF value shows the average root-mean-square deviation (RMSD) observed during the trajectory, as such, it is the measure of protein flexibility in a molecular dynamics simulation, similarly to the B-factors in a crystal structure. However, it is difficult to compare the absolute values of these two measures of flexibility. Therefore, relative values RMSF and B-factors characterized the dynamic properties of PcPAL and TcPAM. Figure 29 demonstrates that crystal structures and molecular dynamic model structures behave similarly, the same parts of the protein show smaller than average and larger than average movement. As expected from the same protein folds and 75% sequence similarity, the dynamics of PcPAL and TcPAM are quite similar.

The RMSD and RMSF values confirm that the equilibration simulation properly sampled the non-perturbed behavior of the model systems, therefore three conformations taken from simulation time points 10 ns, 15 ns and 20 ns are used as starting conformations for the RAMD simulations.
Figure 29. Molecular dynamics simulations as well as B-factor analysis in the crystal structures suggest that dynamics of PcPAL and TcPAM are highly similar. Relative RMSF values and relative B-factors, as these two values correspond to the flexibility of the proteins, are plotted by the consensus positions of the sequence alignment. Similar patterns from the crystal structures and the molecular dynamics indicate that the simulations reproduced well the experimentally observed protein flexibility. Residues with average flexibility are plotted in green, more rigid residues are shown in blue. Yellow indicates residues whose movement differs from the average by the same magnitude as the less flexible residues. Extremely flexible residues are represented by orange and red colors. Insertions in the sequence alignment and missing residues in the crystal structures are represented by white blanks. Average B factors for TcPAM were calculated based on 2YIF93 and 3NZ492 structures.
4.3.2.1 Inner loop motions during ligand egress

Random acceleration molecular dynamics (RAMD)\textsuperscript{180} is a powerful tool for investigating ligand egress routes from buried binding sites\textsuperscript{190,191}, as it requires no prior assumption of the dissociation pathway. This method applies a small, randomly oriented force to the substrate during the MD simulation. If the substrate travels less than a threshold distance within a specified time period (\textit{i.e.} the protein blocks its movement), the direction of the force is randomly reassigned. As a result milliseconds time scale ligand binding/unbinding events can be observed on ps simulation time scales\textsuperscript{180}.

In MIO enzymes, it is often assumed that the inner loop undergoes large conformational change and fully opens upon product release\textsuperscript{78,131}. As experimental evidence only proved the existence of the loop-in conformation in PcPAL, we wanted to investigate the inner loop motions during substrate access and product release. Inner loop flexibility is suggested to switch between PAL and PAM activities\textsuperscript{78}. To compare the behavior of eukaryotic PAL and PAM, RAMD simulations mapped the ligand egress pathways and tested inner-loop dynamics upon ligand egress in full tetrameric models of selected eukaryotic MIO enzymes, PcPAL and of TcPAM. Experiments proved that both PcPAL\textsuperscript{164} and TcPAM\textsuperscript{131} bind and react with L-Phe and TCA. Therefore, RAMD simulations investigated both L-Phe and TCA release, assuming that binding of these substrates occurs via the same paths. Four active sites of the tetramer probed different states of the enzyme; apo, NH\textsubscript{2}-MIO after reaction, ligand bound and ligand releasing, Table 6 defines simulation details. During the RAMD simulations only one active site received the additional acceleration force, the other three served as controls for the ordinary behavior of the enzyme. The distance between the COM of the catalytic tyrosine 110 (80 in TcPAM) and that of the MIO residue measures the inner loop opening, in accordance with Heberling et al.\textsuperscript{78}.
Figure 30. Box plots showing Tyr110(80)-MIO center of mass (COM) distances in the different chains of the tetrameric models. a.) PcPAL model with L-Phe substrate, b.) PcPAL model with TCA product, c.) TcPAM model with L-Phe substrate, d.) TcPAM model with TCA product. Dashed lines show the reference values measured for chain A in the starting structures. Labels show the median distances in the different chains. Exit of TCA was modeled from an active site where the MIO modified residue is bound to the amino group, therefore the COM distances between the catalytic residue and Tyr110 (80) in these models are slightly shorter. Table 6 provides detailed explanation of the configuration of the different sites.

Active site D was modeled equally for L-Phe and TCA ligands, thus served as a control. In both PcPAL (average values 12.0 Å and 12.1 Å) and in TcPAM models (average values 12.4 Å in both models) active site D produced the same results, validating modeling and simulation protocols. Ligand bound B active sites showed increased average Tyr-MIO COM distances compared to the unoccupied C active sites. This increase was greater for L-Phe (1.0 Å for PcPAL and 2.0 Å for TcPAM) than for TCA (0.3 Å for PcPAL and 0.4 Å for TcPAM). The median values are smaller for both PcPAL models then the COM distance in the crystal structure (13.4 Å and 12.9 Å), in contrast median values are larger in almost all TcPAM active sites then the COM distance in the crystal structure (12.3 Å and 11.9 Å). These two observations
indicate that the inner loop of PcPAL in 6F6T is slightly open due to the presence of the phosphonic acid inhibitor, while the inner loop is in closed conformation in 2YII as only a small β-mercaptoethanol molecule is bound there. An increase in the width of the distribution can be observed in active site A in all four models, due to the application of the accelerating force. The difference in the median values compared to the ligand bound B active sites is most likely also due to the accelerating force.

![Figure 31](Image)

**Figure 31. Relationship between inner loop opening (measured by Tyr110(80)-MIO distance) and ligand egress.** The y axis shows the distance between the center of mass (COM) of Tyr110 and MIO, the x axis shows the distance between the COM of the ligand (L-Phe, a,c and TCA b,d) and MIO. a,b show data for PcPAL and c,d for TcPAM. Each plot shows data from 45 RAMD simulations for chain A active sites. Dashed lines show Tyr110(80)-MIO COM distance in the loop-in conformations, seen in 6F6T (PcPAL) and 2YII (TcPAM) structures. Dot-dashed lines show the Tyr110-MIO COM distance in the loop-out conformation seen in 1W27.

Analysis of the COM distances between Tyr110(80) and MIO as a function of the ligand exit measured by the ligand-MIO COM distance, reveals that the inner loop remains in its closed loop-in conformation during ligand egress (Figure 31). Simulations showed no transition to the loop-out conformation, COM distance 27.3 Å in 1W27, or systematic correlation between
ligand egress. Loop opening remains within the regular movement range observable in the other active sites (Figure 30 and Figure 31) for both substrates and for PcPAL and TcPAM models equally.

Figure 32. Root-mean-square fluctuation (RMSF) of backbone atoms of the Inner loop (a) and of the outer loop (b) in the four different simulations (PcPAL with L-Phe and TCA dark green and light green, TcPAM with L-Phe and TCA, yellow and orange curves). Separate RMSF data is calculated for loops covering the RAMD active site for frames when ligand exit occurs (COM MIO-Lig>10 Å, blue, cyan, pink, red).

RMSF values for inner and outer loop residues were investigated to compare the loop motions of the non-perturbed protein and loop motions during ligand egress. Plots on Figure 32 showes RMSF values of the backbone atoms averaged for inner and outer loop residues all four chains during all the RAMD simulations. Should the loops open during ligand egress it will translate to a large increase in RMSF values. The non-perturbed, average inner loop movements are identical in PcPAL and TcPAM (dark and bright green, yellow and orange lines of Figure 32 a). Notable conformational differences occur at residues 116-122 of the inner loop(blue and cyan and pink and red lines of Figure 32 a), which are in direct contact with the outer loop. Average inner loop movements are larger in PcPAL (dark and bright green of Figure 32 b) than in TcPAM (yellow and orange lines of Figure 32 b). Outstanding conformational change occurs
upon L-Phe exit from PcPAL at residues 339-345 (blue line of Figure 32 b), while only minor differences occur in all other cases (cyan, pink and red lines of Figure 32 b). Despite the fact that increased loop movement was observed during ligand egress compared to the non-perturbed models, changes in RMSF originating from major loop conformational changes were absent from all simulations. These observations provide a further argument against the biological relevance of the loop-out conformation.

On the whole four separate aspects of the structural dynamics and biological role of the inner loop of MIO enzymes were investigated. First, reevaluation of the experimental data of 1W27\textsuperscript{83} showed no electron densities for the inner loop residues in the loop-out conformation. Second, no traces of the loop out conformation were seen in any of the novel PcPAL crystal structures. Third, RAMD simulations showed no transition to the loop-out conformation during ligand release. Fourth, RMSF values show identical dynamics for the inner loop residues in PcPAL and TcPAM during normal molecular dynamics simulations as well as during substrate release. As PcPAL structure 1W27\textsuperscript{83} was the only evidence for the loop-out conformation, it is quite probable that these conclusions about inner loop dynamics are valid for all MIO enzymes.

4.3.2.2 Conserved access tunnels in MIO enzymes

Visual clustering of ligand egress paths found four paths in PcPAL (Figure 33 a). The same four paths occurred in TcPAM (Figure 33 b), suggesting that substrate access paths are conserved in eukaryotic PALs and PAMs. Path I (pink) runs between the inner and outer loops. Access to the active site was previously suggested to proceed through this pathway, based on static crystallographic data\textsuperscript{81,83}. Statistical data corroborate the previous hypothesis (Figure 33 c), as this is the second most frequently observed egress path. Surprisingly, the TCA intermediate in the TcPAM model never exited through this channel.

Besides the previously proposed Path I, RAMD simulations suggested three additional possible access paths to the active site. Path II (orange) precedes through a turn in the inner loop, while Path III (gray) proceeds through the hydrophobic part of the binding pocket. These two are seldom taken by the ligands (Figure 33 c), suggesting only a minor biological role. Path IV (blue) proceeds through the multi-helix domain and occurrence frequencies (Figure 33 c) suggest that it is the most frequently employed egress route in PcPAL and TcPAM as well. Although not all four paths are detected in all protein-ligand combinations, it is likely that any ligand may use any of the four paths. Occurrence frequencies show Path I and IV to be the favored egress routes.
Figure 33. Exit paths in PcPAL and TcPAM. One representative path for each observed pathway in the PcPAL (a) and in TcPAM (b). c. number of occurrences of the different paths out of the 45 independent runs for each model. d. Box plot of the distribution of exit times for the different models.

Simulations suggest that similar forces are necessary for ligand exit in PcPAL and TcPAM. Generally, PcPAL showed longer exit times during RAMD simulations (Figure 33 d) and larger force was required for successful exits during force optimizations (Table 8). These two findings suggest that the more flexible inner loop seen in PcPAL crystal structures compared to TcPAM does not facilitate ligand access to the active site.

Inspired by the conservation of the ligand egress pathways in eukaryotic PAL and PAM, we wished to see if these paths might be conserved in all MIO-containing class I lyase-like enzymes. Tunnel analysis by MOLE 2.0 revealed that Path IV coincided with a conserved water channel connecting the active site to the multi-helix domain of the protein (Figure 34 a-b). This corroborates the biological relevance of Path IV and proves that MOLE 2.0 is a suitable tool for finding such paths. Representatives of each type of MIO-containing class I lyase-like enzyme was subjected to MOLE 2.0 analysis with the same settings. Very similar water channels are present in HAL, TAL, TAM and eukaryotic PAL structures (Figure 33 c-f), unmasking a highly conserved water channel connecting the active site to the multi-helix domain.
domain. Coincidentally it is the only water channel with larger than 1.5 Å inner sphere size connecting the active site to the protein surface in all of the investigated enzymes. These findings together with structural observations concerning the inner and outer loops capping the active sites suggest that Path I and Path IV are the major substrate access and product release pathways in all MIO enzymes.

![Image](image.png)

**Figure 34. Water tunnels in MIO enzyme crystal structures.** a. Petroselinum crispum PAL 6F6T, b. Taxus Canadensis PAM 2YII, c. Pseudomonas putida HAL 1GKM, d. Anabaena variabilis PAL 3CZO, e. Streptomyces globisporus TAM 2QVE, f. Rhodobacter sphaeroides TAL 2O6Y. Different shades of blue are used for the different tunnels for visualization.

Interestingly, the active site of PaPAM (prokaryotic S-PAM) is shielded from the surface of the protein; no water channels with larger than 1 Å diameter can be detected in the structure. The conserved water channel found in other MIO enzymes (Figure 33) is closed by Phe487 (455 in PaPAM), that is Asn in all other member of the family. Mutation of Phe487 to Asn in PaPAM resulted in increased AL activity, and the authors attributed this to the substrate orienting effect of this residue\(^{192}\). In view of these new finding it is possible that ligand binding kinetics might have been influenced by the mutation. Mutations are required to put this hypothesis to a test. Large differences between prokaryotic and eukaryotic MIO enzymes render PcPAL unfitting candidate for such studies. A highly stable prokaryotic PAL with large sequence similarity to PaPAM is required instead.
4.3.2.3 *Kangiella koreensis PAL, an ideal candidate for engineering*

PALs as biocatalysts must withstand as high as 6 M ammonia concentrations to achieve high conversions. PALs of marine origin – especially PAL from *Idomarina loihiensis* (IIPAL) – were capable of catalyzing the ammonia addition with high activity at elevated ammonia and substrate concentrations. In frame of our general interest to clone thermotolerant and stable PALs as efficient biocatalysts focusing on enzymes of marine and extremophile origin, we described the molecular cloning, expression and purification of a novel PAL from a marine bacterium *Kangiella koreensis* (KkPAL).

KkPAL showed surprising kinetic behavior, as phenylalanine solubility limited the determination of the enzyme’s Michaelis constant ($K_m$) and turnover number ($k_{cat}$), Figure 35. The hydrophobic binding pocket of KkPAL, similarly to IIPAL is a hybrid between the typical motifs found in TALs and HALs, thus large substrate tolerance is expected. Indeed, these substrate concentrations could not be relevant within the bacterial cell. This observation however raises the question about structural determinants of the kinetic parameters in MIO enzymes that vary on a wide scale: $K_m$ varies between 0.0172 mM in PcPAL to 2.56 mM in AtPAL and $k_{cat}$ varies between 0.1 s$^{-1}$ in AtPAL to 10.11 s$^{-1}$ in *Bambusa oldhamii* PAL.

![Figure 35. Characterization of KkPAL. a. Initial velocity of KkPAL catalyzed ammonia elimination from phenylalanine as a function of variable substrate concentrations. KkPAL active center concentration was 2 µM in the reaction. b. Thermal unfolding curve of KkPAL obtained by nanoDSF measurement.](image)

The KkPAL exhibited outstanding thermal stability as well, as its melting temperature was 81.7 °C (Figure 35). This melting temperature is 10 °C higher than that of the eukaryotic PcPAL, 71 °C. The high melting temperature and its 65% sequence similarity to PaPAM, render KkPAL the perfect candidate for investigating the effects of individual and combined residue mutation on the ligand egress pathways and the regioselectivity of KkPAL.
5 Summary

Scientific and technological advances have established biocatalysis as an alternative to traditional metallo- and organocatalysis in chemical synthesis, leading to the development of greener, more efficient and more sustainable synthetic processes\textsuperscript{1}. In 2012, three waves of biocatalysis were discussed\textsuperscript{1}, already today the fourth wave of biocatalysis is forecasted\textsuperscript{2,3}. Results presented here constitute a small part of the effort to understand how protein structure defines substrate binding and catalyze chemical transformations by lowering the transition state energy barriers in hydroxynitrile lyases and MIO containing class-I lyase like enzymes.

Evolutionarily related hydroxynitrile lyases from rubber tree (HbHNL) and from Arabidopsis thaliana (AtHNL) follow different catalytic mechanisms with opposite enantioselectivity toward mandelonitrile. Reaction mechanisms differ in the identity and location of the cyanide-stabilizing residue and of the catalytic base. First, and most firmly established, is the mechanism for \((S)\)-selective HNL from rubber tree (HbHNL). The active site contains an esterase-like catalytic triad of Ser-His-Asp. The role of the serine differs, in HbHNL as that found in esterases as it transfers protons between the substrate and catalytic histidine. The positive charge of N\(_{\varepsilon}\) of Lys 236 stabilizes the negative charge on the leaving cyanide, thus it is called Lys mechanism. The reaction mechanism of \((R)\)-selective HNL from mouse-ear cress (AtHNL) differs from each other. The catalytic histidine serves as the base as in HbHNL, but two main chain N–H groups of the oxyanion hole are the cyanide stabilizing atoms, thus this mechanism is called the oxyanion hole mechanism. We hypothesized that the HbHNL-like mechanism evolved from an enzyme with an AtHNL-like mechanism. We created ancestor-like composite active sites in each scaffold to elucidate how this transition may have occurred. Surprisingly, a composite active site in HbHNL maintained \((S)\)-selectivity, while the identical set of active site residues in AtHNL maintained \((R)\)-selectivity. Composite active-site mutants that are \((S)\)-selective without the Lys236 and Thr11 that are required for the classical \((S)\)-HNL mechanism suggested a new mechanism. Modeling indicated a possibility for this new mechanism that does not exist in modern enzymes. In the new mechanism, the cyano group of mandelonitrile accepts a hydrogen bond from H\(\delta 21\) of Asn11, thus it is called the Asn mechanism. Accordingly, the last common ancestor of HbHNL and AtHNL may have used an extinct mechanism, not the AtHNL-like mechanism. Multiple mechanisms are possible with the same catalytic residues and correlation analysis of relative residue positions demonstrated how residues outside the active site strongly influence the mechanism and enantioselectivity of the reaction.
Aromatic amino acid ammonia-lyases and aromatic amino acid 2,3-aminomutases contain the post-translationally formed 4-methylideneimidazole-5-one (MIO) catalytic electrophile. MIO enzymes catalyze the stereoselective synthesis of α- or β-amino acid enantiomers, making these chemical processes environmentally friendly and affordable. Characterization of novel inhibitors enables structural understanding of the mechanism of enzymatic catalysis. We found that both enantiomers of the aminophosphonic acid analogue of the natural substrate phenylalanine and a novel derivative bearing a methylidene at the β-position inhibited phenylalanine ammonia-lyases (PAL), representing MIO enzymes. X-ray crystal structure determination unambiguously confirmed the absolute configuration of all tested enantiomers during their synthesis. Enzyme kinetic measurements revealed the enantiomer of the methylidene-substituted substrate analogue as being a mirror image relation to the natural L-phenylalanine as the strongest, and slow binding competitive inhibitor. Isothermal titration calorimetry (ITC) confirmed the binding constants and provided a detailed analysis of the thermodynamic driving forces of ligand binding.

The mirror image binding of (S)-APPA and (R)-APEP designated them to be ideal models for examining the reaction mechanism and the molecular basis of the enantioselectivity of the PAL reaction. Surprisingly, in the crystal structures of PcPAL both inhibitors bound covalently to the MIO catalytic electrophile through their amino groups, despite only reversibly inhibiting the enzyme. This finding provided the first direct experimental evidence for the N-MIO mechanism for PAL. Position of the β-carbon atom differs the most between the two enantiomers, thus this difference is likely a general feature determining the enantioselectivity of the MIO enzymes. Fascinatingly, the observed binding modes match precisely the ones predicated in 1981 based on kinetic parameters of transition state analogue inhibitors

The active site covering inner-loop was thought to exist either in a loop-in or in a loop-out conformation. Based on crystallographic and molecular modeling data three separate arguments disproved the biological relevance of the loop-out conformation in PcPAL. First, reevaluation of the experimental data of 1W27 showed no electron densities for the inner loop residues in the loop-out conformation. Second, no traces of the loop-out conformation were seen in any of the novel PcPAL crystal structures. Third, RAMD simulations showed no transition to the loop-out conformation during ligand release. As a result, is quite probable that these conclusions about inner loop dynamics are valid for all MIO enzymes.

Enhanced sampling simulations of the substrate access/ product release pathways in eukaryotic PAL and PAM indicated two main paths in both enzymes. The first path was previously suggested based on static crystallographic data. The second yet unexplored one,
coincided with a conserved water channel connecting the active site to the multi-helix domain in all known MIO enzymes, except for (S)-selective PAMs, like PaPAM. All known such PAMs encode two different residues compared to the usual MIO enzymes lining these tunnels, each closing one of the egress paths. We have selected, cloned and characterized KkPAL to be the model enzyme for testing the effects of these mutations on the enzymatic activity and attempt to engineer a de novo PAM activity for the first time.
References


17. Nakano, S., Dadashipour, M. & Asano, Y. Structural and functional analysis of hydroxynitrile lyase from Baliospermum montanum with crystal structure, molecular


33. Stranzl, G. R. et al. Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from Hevea brasiliensis explains a large pKa shift of the catalytic


66. Wightman, R. H., Staunton, J., Battersby, A. R. & Hanson, K. R. Studies of enzyme-mediated reactions. Part I. Syntheses of deuterium- or tritium-labelled (3S)- and (3R)-phenylalanines: stereochemical course of the elimination catalysed by L-phenylalanine


82. Appert, C., Logemann, E., Hahlbrock, K., Schmid, J. & Amrhein, N. Structural and catalytic properties of the four phenylalanine ammonia-lyase isoenzymes from parsley
98. Watts, K. T., Mijs, B. N., Lee, P. C., Manning, A. J. & Schmidt-Dannert, C. Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic
115. Appert, C., Zoń, J. & Amrhein, N. Kinetic analysis of the inhibition of phenylalanine


7 Thesis points

I. We created identical active sites in the AtHNL and in the HbHNL scaffolds, and measured different enantioselectivities for these enzymes despite their theoretically identical active sites. Using molecular dynamics, I showed that catalytically active conformations could suggest the reaction mechanism in HNLs. Based on the molecular dynamics simulations I proposed the Asn mechanism for the composite active sites, which we hypothesized to be an ancestral HNL mechanism. [2]

II. By analyzing the relative movement of the active site residues I identified seventeen of the forty-four residues occupy different positions rendering the theoretically identical active sites different in shape and function. I described two groups of residues whose relative positions correlated, and suggested that the second shell residues from these groups are the origins of the active site shape difference. [2]

III. I characterized the binding of five aminophosphonic acid molecules to PcPAL using kinetic assays. Amongst those I characterized the binding of the most potent inhibitors; (1-amino-2-phenylethyl)phosphonic acid (APEP) and (1-amino-2-phenylallyl)phosphonic acid (APPA), in their enantiomerically pure form by kinetic analysis and isothermal titration calorimetry as well. I demonstrated that the presence of the additional methyldiene group in APPA reverses the enantiopreference of PcPAL compared that found for APEP. I showed that (R)-APEP and (S)-APPA show slow binding kinetic characteristics. [1,3]

IV. I measured the inhibition of another eukaryotic (RtPAL) and a prokaryotic (AvPAL) PAL enzymes by enantiomerically pure APEP and APPA, thus showed them to be general PAL inhibitors, and suggested them to be general MIO enzyme inhibitors. [3]

V. I solved the crystal structure of PcPAL in an apo form and in complex with (R)-APEP and with (S)-APPA. I demonstrated that binding of the (R) and (S) enantiomers differs mostly in the position of the β carbon atom and suggested this to be the key element for the enantioselectivity of PAL enzymes.

VI. I found independent arguments indicating that the loop-in conformation is the catalytically competent conformation of the inner loop, that does not open significantly during ligand binding. First, reevaluation of the experimental data of 1W2783 showed no electron densities for the inner loop residues in the loop-out conformation. Second, I observed no indication for the loop out conformation in any of the novel PcPAL crystal structures, but found well defined electron densities in the
loop-in conformation in the inhibitor bound structures. Third, RAMD simulations showed no transition to the loop-out conformation during ligand release. Based on these arguments, I suggested that the loop-in conformation is the catalytically competent conformation of the inner loop, that does not open significantly during ligand binding.

VII. We identified and recombinantly expressed a novel MIO enzyme from *Kangiella koreensis*. Subsequent substrate analysis proved the enzyme to be a PAL. We showed that KkPAL has a high melting temperature, 81 °C. Thanks to the high $T_m$, KkPAL may accommodate numerous potentially destabilizing mutations without hindering its expressibility making it an ideal candidate for enzyme engineering studies. [4]

8 Publication list

8.1 List of my publications giving the foundation of the thesis points

1. Bata Zs., Belle E.:  
   Biokatalizátorok a zöldébb jövőért- enzimek a vegyiparban  
   Élet és tudomány, 2018. 03. 23, LXXIII, 12. (BZs 55%)

2. Jones B.J., Bata Zs., Kazlauskas R. J.:  
   Identical active sites in hydroxynitrile lyases show opposite enantioselectivity and reveal possible ancestral mechanism  
   *ACS Catalysis*, 7, 4221–4229, 2017 (IF 10.614, BZs 100%, FI:1)

   Metylidene group on phosphonic acid analogue of phenylalanine reverses the enantiopreference of phenylalanine ammonia-lyases  
   *Advanced Synthesis and Catalysis*, 359, 2109-2120, 2017 (IF 5.646, BZs 100%, FI:2)

   A novel phenylalanine ammonia-lyase from *Kangiella koreensis*  
   *Studia Universitatis Babes-Bolyai Chemia*, 62(3), 293-308, 2017 (IF: 0.148, BZs 90%)
8.2 List of my other publications


8.3 List of conference presentations related to the thesis

8.3.1 Oral presentations at conferences


Bata Zs., Madaras E., Vertessy B. G., Poppe L.: MIO enzimcsaládában konzervált szubsztrát bekötődési útvonalak feltérképezése változóirányú gyorsítású molekuladinamikai szimulációkkal Kemometria és Molekula Modellezés (KeMoMo) szimpózium, 2018. 05. 24-25, Szeged, Hungary

Bata Zs., Qian R., Roller A., Hammerschmidt F., Paizs Cs., Vertessy B. G., Poppe L.: *Mirror image inhibitor enhance crystallization of the industrially relevant PcPAL* Hungarian Molecular Life Sciences 2017. 03. 31- 04.02, Eger, Hungary

Bata Zs., Jones J. B., Kazlauskas R. J.: *Modeling explanation of altering the enantioselectivity of hydroxynitrilase* BME Oláh György Doktoráns konferencia, 2017. 02. 02, Budapest, Hungary


Bata Zs., Jones B.J., Kazlauskas R.J : *R or S, hogyan döntenek a hidroxinitrilázok?* Kemometria és Molekula Modellezés (KeMoMo) szimpózium, 2016. 05. 12-13, Miskolc, Hungary


**8.3.2 Poster presentations at conferences**

Bata Zs., Madaras E., Leveles I., Hammerschmidt F., Paizs Cs., Poppe L., Vertessy B. G.: *Bioactive 3D structure of phenylalanine ammonia-lyase reveal key insights into ligand binding dynamics* BME VBK 145, 2018. 06. 01, Budapest, Hungary


Bata Zs., Madaras E., Ibolya L., Hammerschmidt F., Paizs Cs., Poppe L., Vertessy B. G. *Catalytically competent novel 3D structure of phenylalanine ammonia-lyase reveal key insights into ligand binding dynamics* Federation of European Biochemical Societies (FEBS) Youth Scientific Forum and Annual meeting, 2017. 09. 7-10 and 2017. 09. 10-14, Jerusalem, Israel
**Bata Zs., Leveles I., Qian R., Hammerschmidt F., Paizs Cs., Poppe L., Vertessy B. G.:**

**Catalytically competent structure of eukaryotic phenylalanine ammonia-lyase**

13th International Symposium on Biocatalysis and Biotransformations, 2017. 07. 09-13, Budapest, Hungary, **Short talk, 5 minutes and poster presentation**

**Bata Zs., Qian R., Roller A., Hammerschmidt F., Paizs Cs., Vertessy B. G., Poppe L.:**

**Experimental and computational characterization of aminophosphonic acid MIO enzyme inhibitors,** Conference Poster price

BME Oláh György Doktoráns konferencia, 2017. 02. 02. Budapest, Hungary

**Bata Zs., Qian R., Roller A., Hammerschmidt F., Paizs Cs., Vertessy B. G., Poppe L.:**

**Experimental and computational characterization of aminophosphonic acid MIO enzyme inhibitors**

8th International meeting on biocatalysis, Biocat2016, 2016. 08. 28 – 09. 01. Hamburg, Germany

**Bata Zs., Hammerschmidt F., Vertessy B. G., Poppe L.:**

**Experimental characterization of aminophosphonic acid MIO enzyme inhibitors**

FEBS advances course: Advanced Methods in Macromolecular Crystallization VII. 2016 06. 27- 07. 02. Nové Hrady, Czech Republic
### Appendix

#### Table A1. List of primers used for mutagenesis and cloning

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7_For</td>
<td>AATACGACTCATAATTAGGGAATTG</td>
</tr>
<tr>
<td>T7_Rev</td>
<td>TGCTAGTTATTGCTCAGCGG</td>
</tr>
<tr>
<td>PcPAL_For</td>
<td>TTATGCTAGCATGAAACGGAACCGGAG</td>
</tr>
<tr>
<td>PcPAL_Rev</td>
<td>TATTGAGCTCTCAGAAATGGCAGAAGG</td>
</tr>
<tr>
<td>KkPAL_For</td>
<td>TAGATAATACCATGGCAGATATG</td>
</tr>
<tr>
<td>KkPAL_Rev</td>
<td>CCAGTTATGGAATCTCGATGATTG</td>
</tr>
<tr>
<td>PcPAL_Y110F_for</td>
<td>CTGACTCTCTCGGAGAAACTACTGGATTCGAGCCACT</td>
</tr>
<tr>
<td>PcPAL_Y110F_rev</td>
<td>GAAGAGTCTAGTCTCTGTTATGGAAGG</td>
</tr>
</tbody>
</table>

#### Table A2. Sequence and extinction coefficients for proteins used in this work

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PcPAL_80</td>
<td>MGGSNHHRHHRHSSSGHIDDDDKMLPLEMLNENGATTGGHVNGNHGDFC MKTEDPLYWGI AAEM TQ</td>
</tr>
<tr>
<td></td>
<td>GMHSHLDEVKVMAYERPKVPGKLETLTISQVAIAARSDGGVTELSEARAGVKAASDVM</td>
</tr>
<tr>
<td></td>
<td>MDSMNGKTDYSYVTGTFGATSRHTKQGGQLKEFLIRLNFAGIFGNSDNTLPHTSANTAAM</td>
</tr>
<tr>
<td></td>
<td>VRLINTLQLQYSIGRELFIEAEIATKLQFLNQNIPTCCLPLRGTITASGDLVLPSYIAGLTLTRGRFNSK</td>
</tr>
<tr>
<td></td>
<td>AVGPTQVOLSPEAKFLAVGSGGFTQPEKLGLVNTAVGSMSMVLFIEANAVLNLAVEV</td>
</tr>
<tr>
<td></td>
<td>MSAIFAVMEQKPEFTDHFLTHKHHFQIEAIAAMIEHLDSAYKVAQLEKHMDPLQFK</td>
</tr>
<tr>
<td></td>
<td>QDVRAYLRPSQWLPQIEVEVRSTKMIEREINSVNDNLIDSVIDVMEAAIGGNFQTEPIGVSM</td>
</tr>
<tr>
<td></td>
<td>DNLTRLAIAQKGFLAQESFNLDYFNSNLGSNGRPSLDYFGKAEIAMYSCLSEQ</td>
</tr>
<tr>
<td></td>
<td>LANPVTNHQSAEPMHDVNLGSHLPFCEKEDLRVLVDCIVLFMAYIFIDDCPSATYPMILQKLQR</td>
</tr>
<tr>
<td></td>
<td>TLEWHEALKGNDERNLSTSIQKIAFEDEKALLPKPEVESARALESLNGPAIPNIREECR</td>
</tr>
<tr>
<td></td>
<td>SYLPLYFKVRLGTEYLTGEVSTPSFEEVFIAMSGKEIIIDPLESLWNGAPLIP</td>
</tr>
<tr>
<td>PcPAL_H6</td>
<td>MGSHHHHHHHSSGHIDDDDKMLPLEMLNENGATTGGHVNGNHGDFC MKTEDPLYWGI AAEM TQ</td>
</tr>
<tr>
<td></td>
<td>GMHSHLDEVKVMAYERPKVPGKLETLTISQVAIAARSDGGVTELSEARAGVKAASDVM</td>
</tr>
<tr>
<td></td>
<td>MDSMNGKTDYSYVTGTFGATSRHTKQGGQLKEFLIRLNFAGIFGNSDNTLPHTSANTAAM</td>
</tr>
<tr>
<td></td>
<td>VRLINTLQLQYSIGRELFIEAEIATKLQFLNQNIPTCCLPLRGTITASGDLVLPSYIAGLTLTRGRFNSK</td>
</tr>
<tr>
<td></td>
<td>AVGPTQVOLSPEAKFLAVGSGGFTQPEKLGLVNTAVGSMSMVLFIEANAVLNLAVEV</td>
</tr>
<tr>
<td></td>
<td>MSAIFAVMEQKPEFTDHFLTHKHHFQIEAIAAMIEHLDSAYKVAQLEKHMDPLQFK</td>
</tr>
<tr>
<td></td>
<td>QDVRAYLRPSQWLPQIEVEVRSTKMIEREINSVNDNLIDSVIDVMEAAIGGNFQTEPIGVSM</td>
</tr>
<tr>
<td></td>
<td>DNLTRLAIAQKGFLAQESFNLDYFNSNLGSNGRPSLDYFGKAEIAMYSCLSEQ</td>
</tr>
<tr>
<td></td>
<td>LANPVTNHQSAEPMHDVNLGSHLPFCEKEDLRVLVDCIVLFMAYIFIDDCPSATYPMILQKLQR</td>
</tr>
<tr>
<td></td>
<td>TLEWHEALKGNDERNLSTSIQKIAFEDEKALLPKPEVESARALESLNGPAIPNIREECR</td>
</tr>
<tr>
<td></td>
<td>SYLPLYFKVRLGTEYLTGEVSTPSFEEVFIAMSGKEIIIDPLESLWNGAPLIP</td>
</tr>
</tbody>
</table>

---

**Note:** Table A1 and Table A2 provide lists of primers used for mutagenesis and cloning, along with sequences and extinction coefficients for proteins used in this work. The tables include columns for Protein, Sequence, and Abs 1%.
Table A3. Mutant enzymes activity and enantioselectivity.

|                | 11  | 12  | 13  | 54  | 79  | 81  | 106 | 121 | 128  | 125 | 128 | 125 | 128 | 125 | 128 | 125 | 128 | 209 | 210 | 208 | 206 | Activity (sec⁻¹) | Enantiosel.\(E\) |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----------------|-----------------|
| **HbHNL**      | T   | I   | C   | F   | E   | C   | V   | L   | F   | W   | T   | L   | L   | I   | F   | K   | 25.6 |        | >39 (S)        |
| **Hb-A9-H7**   | N   | A   | C   | F   | F   | F   | F   | F   | L   | F   | W   | T   | M   | F   | L   | I   | M   | 1.9  |        | >39 (S)        |
| **Hb-A10-H6**  | N   | A   | C   | F   | F   | F   | F   | F   | W   | T   | M   | F   | F   | I   | I   | M   | 0.4  | 3.3 ± 0.5 (S) |
| **Hb-A10-H6_F54V** | N   | A   | C   | V   | F   | F   | F   | F   | L   | F   | W   | T   | M   | F   | F   | I   | I   | M   | 0.9  |        | (S)            |
| **At-A9-H7**   | N   | A   | C   | V   | F   | F   | F   | F   | L   | F   | W   | T   | M   | F   | L   | I   | M   | 0.6  | 1.6 ± 0.1 (S) |
| **At-A10-H6**  | N   | A   | C   | V   | F   | F   | F   | F   | L   | F   | W   | T   | M   | F   | F   | I   | I   | M   | 1.0  | 5.3 ± 1.1 (R) |
| **At-A10-H6_V54F** | N   | A   | C   | V   | F   | F   | F   | F   | L   | F   | W   | T   | M   | F   | F   | I   | I   | M   | 0.5  |        | (R)            |
| **AtHNL**      | N   | A   | Y   | V   | F   | F   | F   | Y   | P   | L   | C   | M   | F   | F   | A   | I   | M   | 27.0 |        | >39 (R)        |

*Extinction coefficient \((g^{-1} \text{dm}^3 \text{cm}^{-1})\)

Each enzyme is on a different row, with the active site residues numbered across the top. Residues listed are those that differ between AtHNL and HbHNL and are within 6 Å of the substrate in HbHNL (PDB:1YB6). Underlined residues indicate they are part of the cap/lid domain. Yellow indicates HbHNL-like, while cyan indicates AtHNL-like. Qualitative expression solubilities were determined by SDS-PAGE analysis comparing crude \(E. coli\) lysate with soluble \(E. coli\) lysate and Ni-NTA purified protein samples. Activity was determined by measuring formation of benzaldehyde from the cleavage of 9 mM racemic mandelonitrile at pH 5.1. Limit of detection varied based upon amount of enzyme used. Enantioselectivity given as enantiomeric ratio with preferred enantiomer indicated. The values were determined by synthesis of mandelonitrile from benzaldehyde and hydrogen cyanide, and chiral-HPLC analysis.
Table A4. Active Site Distances to Substrate*

<table>
<thead>
<tr>
<th>Residue Number</th>
<th>Residue Name</th>
<th>Active Site Distances to Substrate</th>
<th>Substrate Atom</th>
<th>Closest Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>THR</td>
<td>ASN</td>
<td>Oy1 N82</td>
<td>O N</td>
</tr>
<tr>
<td>80</td>
<td>SER</td>
<td></td>
<td>Oy</td>
<td>O</td>
</tr>
<tr>
<td>209</td>
<td>ILE</td>
<td>ALA</td>
<td>C81 C6</td>
<td>C6 C5</td>
</tr>
<tr>
<td>128</td>
<td>TRP</td>
<td>LEU</td>
<td>Cq2 C2</td>
<td>C6</td>
</tr>
<tr>
<td>236</td>
<td>LYS</td>
<td>MET</td>
<td>Nc S6</td>
<td>N</td>
</tr>
<tr>
<td>157</td>
<td>LEU</td>
<td></td>
<td>C2</td>
<td>N</td>
</tr>
<tr>
<td>235</td>
<td>HIS</td>
<td></td>
<td>Nf2 C2</td>
<td>N</td>
</tr>
<tr>
<td>148</td>
<td>LEU</td>
<td>MET</td>
<td>C2 S6 C8</td>
<td>N</td>
</tr>
<tr>
<td>81</td>
<td>CYS</td>
<td>PHE</td>
<td>Sy N</td>
<td>O</td>
</tr>
<tr>
<td>14</td>
<td>HIS</td>
<td></td>
<td>C1</td>
<td>N</td>
</tr>
<tr>
<td>210</td>
<td>PHE</td>
<td>ILE</td>
<td>Cx1 Cy1 C4</td>
<td>O</td>
</tr>
<tr>
<td>152</td>
<td>LEU</td>
<td>PHE</td>
<td>C52 C91 C7</td>
<td>O</td>
</tr>
<tr>
<td>12</td>
<td>ILE</td>
<td>ALA</td>
<td>Cy1 C2 O</td>
<td>C2</td>
</tr>
<tr>
<td>121</td>
<td>LEU</td>
<td>TYR</td>
<td>C32 Oe C5</td>
<td>O</td>
</tr>
<tr>
<td>79</td>
<td>GLU</td>
<td>PHE</td>
<td>Oe Cx</td>
<td>N</td>
</tr>
<tr>
<td>125</td>
<td>PHE</td>
<td>PRO</td>
<td>C52 O C6</td>
<td>O</td>
</tr>
<tr>
<td>146</td>
<td>LEU</td>
<td></td>
<td>C81 C32 C7</td>
<td>O</td>
</tr>
<tr>
<td>150</td>
<td>TYR</td>
<td></td>
<td>Cz1 N</td>
<td>N</td>
</tr>
<tr>
<td>131</td>
<td>THR</td>
<td>CYS</td>
<td>Oy1 Sy C7</td>
<td>O</td>
</tr>
<tr>
<td>153</td>
<td>LEU</td>
<td>MET</td>
<td>C52 Cx N C8</td>
<td>O</td>
</tr>
<tr>
<td>170</td>
<td>LEU</td>
<td>PHE</td>
<td>C52 Cz Cy C4</td>
<td>O</td>
</tr>
<tr>
<td>13</td>
<td>CYS</td>
<td>TYR</td>
<td>C81 C32 C2 C8</td>
<td>O</td>
</tr>
<tr>
<td>82</td>
<td>GLY</td>
<td></td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>122</td>
<td>MET</td>
<td></td>
<td>S6 Cy C6 C5</td>
<td>O</td>
</tr>
<tr>
<td>104</td>
<td>ASN</td>
<td></td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>13</td>
<td>HIS</td>
<td></td>
<td>C</td>
<td>O</td>
</tr>
<tr>
<td>156</td>
<td>ASN</td>
<td>ARG</td>
<td>O Nn2 C5</td>
<td>N</td>
</tr>
<tr>
<td>106</td>
<td>VAL</td>
<td>PHE</td>
<td>Ch1 Cy1 C4</td>
<td>O</td>
</tr>
<tr>
<td>149</td>
<td>GLY</td>
<td></td>
<td>N</td>
<td>C7</td>
</tr>
<tr>
<td>18</td>
<td>ILE</td>
<td></td>
<td>Cy1 C31 N</td>
<td>O</td>
</tr>
<tr>
<td>133</td>
<td>TYR</td>
<td>PHE</td>
<td>Oe Cz C7 C6</td>
<td>O</td>
</tr>
<tr>
<td>173</td>
<td>THR</td>
<td>HIS</td>
<td>Cy2 N81 C8</td>
<td>O</td>
</tr>
<tr>
<td>127</td>
<td>ASP</td>
<td>GLY</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>228</td>
<td>GLN</td>
<td>MET</td>
<td>Ns2 S6 N</td>
<td>N</td>
</tr>
<tr>
<td>207</td>
<td>ASP</td>
<td></td>
<td>Oe1 N</td>
<td>N</td>
</tr>
<tr>
<td>57</td>
<td>TYR</td>
<td></td>
<td>Oe1 O</td>
<td>O</td>
</tr>
<tr>
<td>105</td>
<td>SER</td>
<td>ALA</td>
<td>C4</td>
<td>O</td>
</tr>
<tr>
<td>147</td>
<td>LYS</td>
<td></td>
<td>C</td>
<td>C7</td>
</tr>
<tr>
<td>214</td>
<td>PHE</td>
<td></td>
<td>Cx2 C5</td>
<td>O</td>
</tr>
<tr>
<td>103</td>
<td>HIS</td>
<td>LEU</td>
<td>O N</td>
<td>O</td>
</tr>
<tr>
<td>94</td>
<td>PHE</td>
<td>VAL</td>
<td>Cz2 Cy2 C6</td>
<td>O</td>
</tr>
<tr>
<td>17</td>
<td>TRP</td>
<td></td>
<td>Cq2</td>
<td>O</td>
</tr>
<tr>
<td>208</td>
<td>GLU</td>
<td>LYS</td>
<td>O</td>
<td>C6</td>
</tr>
<tr>
<td>120</td>
<td>LYS</td>
<td>GLY</td>
<td>N</td>
<td>C7</td>
</tr>
<tr>
<td>83</td>
<td>GLY</td>
<td></td>
<td>N</td>
<td>O</td>
</tr>
</tbody>
</table>

*Distances of residues to substrate. The shortest heavy atom distances, in angstroms, for closest 45 residues to (S)-mandelonitrile are shown. Residue number, amino acid identity, and closest atoms of amino acid and substrate are identified. Columns are split when values differ between HbHNL and AtHNL. HbHNL and AtHNL structures (PDB: 1YB6 and 3DQZ chain A) were aligned and both measured against (S)-mandelonitrile from PDB: 1YB6. Table is sorted by increasing closest distance in HbHNL. Residues in bold are the residues mutated in this study. All, except 54, were within 6 angstroms of the substrate in one or both enzymes, and had different residues in the two enzymes.
Table A5. Center of mass (COM) distances between the active site forming residues and Ser80 in the different enzymes.κ

<table>
<thead>
<tr>
<th></th>
<th>wt</th>
<th>A9-H7</th>
<th>A10-H6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HbHNL</td>
<td>HbHNL</td>
<td>A1HNL</td>
</tr>
<tr>
<td></td>
<td>HbHNL</td>
<td>A1HNL</td>
<td>A1HNL</td>
</tr>
<tr>
<td>10 HIS</td>
<td>8.8 ± 0.3</td>
<td>8.4 ± 0.3</td>
<td>8.5 ± 0.3</td>
</tr>
<tr>
<td>11 THR</td>
<td>5.4 ± 0.2</td>
<td>5.7 ± 0.2</td>
<td>5.9 ± 0.2</td>
</tr>
<tr>
<td>12 ILE</td>
<td>7.6 ± 0.2</td>
<td>7.2 ± 0.4</td>
<td>7.3 ± 0.3</td>
</tr>
<tr>
<td>13 HIS</td>
<td>11 ± 0.3</td>
<td>10.8 ± 0.4</td>
<td>10.8 ± 0.4</td>
</tr>
<tr>
<td>14 HIS</td>
<td>9.4 ± 0.4</td>
<td>9.9 ± 0.3</td>
<td>10.3 ± 0.4</td>
</tr>
<tr>
<td>17 TRP</td>
<td>13.7 ± 0.3</td>
<td>13.9 ± 0.3</td>
<td>14.2 ± 0.3</td>
</tr>
<tr>
<td>18 ILE</td>
<td>10.7 ± 0.4</td>
<td>10.9 ± 0.4</td>
<td>11.4 ± 0.3</td>
</tr>
<tr>
<td>54 PHE</td>
<td>12.2 ± 0.3</td>
<td>11.9 ± 0.5</td>
<td>14.1 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>VAL #</td>
<td>12.0 ± 0.5</td>
<td>14.0 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.2 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>57 TYR</td>
<td>12.8 ± 0.3</td>
<td>12.2 ± 0.5</td>
<td>12.5 ± 0.4</td>
</tr>
<tr>
<td>79 GLU</td>
<td>4.0 ± 0.2</td>
<td>4.1 ± 0.2</td>
<td>4.1 ± 0.2</td>
</tr>
<tr>
<td>81 CYS</td>
<td>4.5 ± 0.1</td>
<td>4.4 ± 0.3</td>
<td>5.6 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>PHE #</td>
<td>5.6 ± 0.1</td>
<td>5.6 ± 0.1</td>
</tr>
<tr>
<td>82 GLY</td>
<td>5.8 ± 0.2</td>
<td>5.6 ± 0.3</td>
<td>5.3 ± 0.2</td>
</tr>
<tr>
<td>83 GLY</td>
<td>7.5 ± 0.3</td>
<td>7.1 ± 0.4</td>
<td>6.6 ± 0.2</td>
</tr>
<tr>
<td>103 HIS</td>
<td>7.7 ± 0.4</td>
<td>8.0 ± 0.4</td>
<td>6.9 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>LEU</td>
<td>7.9 ± 0.5</td>
<td>6.9 ± 0.3</td>
</tr>
<tr>
<td>104 ASN</td>
<td>5.9 ± 0.2</td>
<td>6.2 ± 0.3</td>
<td>6.0 ± 0.3</td>
</tr>
<tr>
<td>105 SER</td>
<td>4.8 ± 0.2</td>
<td>4.7 ± 0.2</td>
<td>4.8 ± 0.1</td>
</tr>
<tr>
<td>106 PHE</td>
<td>6.1 ± 0.3</td>
<td>5.4 ± 0.4</td>
<td>5.4 ± 0.4</td>
</tr>
<tr>
<td>121 LEU</td>
<td>10.3 ± 0.5</td>
<td>10.9 ± 0.5</td>
<td>11.4 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>TYR</td>
<td>10.9 ± 0.5</td>
<td>11.4 ± 0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.7 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>122 MET</td>
<td>14.6 ± 1.1</td>
<td>14.6 ± 0.7</td>
<td>14.5 ± 0.7</td>
</tr>
<tr>
<td>125 PHE</td>
<td>13.1 ± 0.4</td>
<td>13.7 ± 0.5</td>
<td>14.1 ± 0.7</td>
</tr>
<tr>
<td>127 ASP</td>
<td>17.5 ± 0.5</td>
<td>18.1 ± 0.6</td>
<td>17.1 ± 1.6</td>
</tr>
<tr>
<td>128 TRP</td>
<td>12.6 ± 0.6</td>
<td>13.4 ± 0.9</td>
<td>14.3 ± 0.9</td>
</tr>
<tr>
<td>129 LYS</td>
<td>17.2 ± 0.4</td>
<td>17.8 ± 0.6</td>
<td>18.9 ± 0.7</td>
</tr>
<tr>
<td>131 THR</td>
<td>16.3 ± 0.4</td>
<td>16.3 ± 0.5</td>
<td>15.9 ± 0.5</td>
</tr>
<tr>
<td>133 PHE</td>
<td>18.0 ± 0.5</td>
<td>18.0 ± 0.6</td>
<td>16.3 ± 0.9</td>
</tr>
<tr>
<td>146 LEU</td>
<td>14.1 ± 0.4</td>
<td>14.1 ± 0.5</td>
<td>13.0 ± 0.5</td>
</tr>
<tr>
<td>147 LYS</td>
<td>17.0 ± 0.6</td>
<td>17.4 ± 0.7</td>
<td>17.0 ± 0.6</td>
</tr>
<tr>
<td>148 LEU</td>
<td>12.8 ± 0.4</td>
<td>12.2 ± 0.5</td>
<td>12.2 ± 0.4</td>
</tr>
<tr>
<td>149 GLY</td>
<td>15.4 ± 0.3</td>
<td>15.6 ± 0.5</td>
<td>16.0 ± 0.5</td>
</tr>
<tr>
<td>152 LEU</td>
<td>13.1 ± 0.4</td>
<td>13.1 ± 0.5</td>
<td>13.9 ± 0.5</td>
</tr>
<tr>
<td>153 LEU</td>
<td>13.3 ± 0.4</td>
<td>13.3 ± 0.6</td>
<td>14.1 ± 0.6</td>
</tr>
<tr>
<td>156 ASN</td>
<td>12.5 ± 0.4</td>
<td>12.7 ± 0.5</td>
<td>13.8 ± 0.6</td>
</tr>
<tr>
<td>157 LEU</td>
<td>9.4 ± 0.4</td>
<td>9.8 ± 0.5</td>
<td>10.4 ± 0.5</td>
</tr>
<tr>
<td>158 TYR</td>
<td>11.8 ± 0.5</td>
<td>12.8 ± 0.5</td>
<td>13.6 ± 0.5</td>
</tr>
<tr>
<td>173 THR</td>
<td>14.6 ± 0.4</td>
<td>14.9 ± 0.4</td>
<td>15.9 ± 0.7</td>
</tr>
</tbody>
</table>

κ The distances are in Ångstroms (Å).
<table>
<thead>
<tr>
<th></th>
<th>wt</th>
<th>A9-H7</th>
<th>A10-H6</th>
<th>wt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HbHNL</td>
<td>HbHNL</td>
<td>AtHNL</td>
<td>HbHNL</td>
</tr>
<tr>
<td>178 LEU PHE</td>
<td>10.8 ± 0.4</td>
<td>10.8 ± 0.5</td>
<td>9.6 ± 0.5</td>
<td>10.9 ± 0.7</td>
</tr>
<tr>
<td>207 ASP</td>
<td>10.2 ± 0.3</td>
<td>10.4 ± 0.4</td>
<td>10.6 ± 0.3</td>
<td>10.4 ± 0.4</td>
</tr>
<tr>
<td>208 GLU LYS</td>
<td>13.4 ± 0.3</td>
<td>13.5 ± 0.4</td>
<td>14.4 ± 0.7</td>
<td>13.5 ± 0.5</td>
</tr>
<tr>
<td>209 ALA ILE</td>
<td>9.0 ± 0.3</td>
<td>8.7 ± 0.5</td>
<td>9.0 ± 0.5</td>
<td>8.9 ± 0.6</td>
</tr>
<tr>
<td>210 PHE ILE</td>
<td>6.5 ± 0.4</td>
<td>7.7 ± 0.4</td>
<td>8.3 ± 0.7</td>
<td>7.6 ± 0.5</td>
</tr>
<tr>
<td>214 PHE</td>
<td>10.7 ± 0.4</td>
<td>11.4 ± 0.4</td>
<td>11.8 ± 0.4</td>
<td>11.3 ± 0.4</td>
</tr>
<tr>
<td>235 HIS</td>
<td>6.6 ± 0.5</td>
<td>6.8 ± 0.3°</td>
<td>7.0 ± 0.3</td>
<td>6.8 ± 0.3</td>
</tr>
<tr>
<td>236 LYS MET</td>
<td>8.3 ± 0.4</td>
<td>9.5 ± 0.4</td>
<td>9.4 ± 0.3</td>
<td>9.5 ± 0.4</td>
</tr>
<tr>
<td>238 GLN MET</td>
<td>10.8 ± 0.4</td>
<td>10.9 ± 0.4</td>
<td>11.2 ± 0.3</td>
<td>10.8 ± 0.4</td>
</tr>
<tr>
<td>R-MNN</td>
<td>5.9 ± 0.6</td>
<td>6.0 ± 0.3</td>
<td>6.1 ± 0.5</td>
<td>6.0 ± 0.3</td>
</tr>
<tr>
<td>S-MNN</td>
<td>5.5 ± 0.5</td>
<td>5.8 ± 0.3</td>
<td>6.5 ± 0.2</td>
<td>5.8 ± 0.3</td>
</tr>
</tbody>
</table>

* Differences in medians larger than the median absolute deviation (MAD) have values shown in dark font, and values at least two time larger are shown in bold. When the two wild-type proteins have different residues, both are listed with HbHNL residues listed first, and the consensus residues are shown in bold.

† Investigating the role of residue 54, mutations to the equivalent residue in the other backbone were modeled. This change only influences the positions of residues 54 and 81. Hb-A10-H6_F54V: Val54 13.5 ± 0.5 Å, Phe81 5.6 ± 0.1 Å, Phe178 9.8 ± 0.6 Å. At-A10-H6_V54F: Phe54 13.1 ± 0.3 Å, Phe81: 5.6 ± 0.1 Å, Phe178 9.5 ± 0.5 Å. The size difference between valine and phenylalanine modified the distances by 1.5 Å and 0.9 Å. AtHNL and HbHNL backbone positions differ at position 54; there is 0.5 Å difference between the average distances in case of valine, and 1.2 Å for phenylalanine. In At-A10-H6_V54F Phe81 remains in its original conformation explaining the experimentally observed enantioselectivity of the V54F mutations.

* Bimodal distribution was observed. The COM of the conformation that is more frequently observed is indicated in the table.
Table A6. Phenylalanine 81 side chain orientation during MD simulation.*

<table>
<thead>
<tr>
<th></th>
<th>APO</th>
<th>(R)- MANDELONITRILE</th>
<th>(S)- MANDELONITRILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hb_wt*</td>
<td>-126 ±44</td>
<td>-148 ±26</td>
<td>-138 ±37</td>
</tr>
<tr>
<td>At_wt</td>
<td>-175 ±7</td>
<td>-175 ±7</td>
<td>-175 ±7</td>
</tr>
<tr>
<td>Hb-A9-H7</td>
<td>-64 ±13</td>
<td>-71 ±10</td>
<td>-99 ±15</td>
</tr>
<tr>
<td>At-A9-H7</td>
<td>-176 ±7</td>
<td>-175 ±7</td>
<td>-174 ±8</td>
</tr>
<tr>
<td>Hb-A10-H6</td>
<td>-68 ±13</td>
<td>-76 ±18</td>
<td>-98 ±18</td>
</tr>
<tr>
<td>At-A10-H6</td>
<td>-176 ±8</td>
<td>-175 ±7</td>
<td>-175 ±7</td>
</tr>
<tr>
<td>Hb-A10-H6_F54V</td>
<td>-170 ±20</td>
<td>-158 ±26</td>
<td>-165 ±10</td>
</tr>
<tr>
<td>At-A10-H6_V54F</td>
<td>-135 ±46</td>
<td>-172 ±14</td>
<td>-165 ±21</td>
</tr>
</tbody>
</table>

* Hb_wt contains cysteine at position 81. The values correspond to the N-Cu-Cβ-S dihedral angle.

The angles (in degrees) correspond to average values for the N-Cu-Cβ-Cy dihedral angle during three repeats of 10 ns MD simulation. Values near -75° point into the active site pocket, while values of -175° are buried in the hydrophobic core. Hb-A9-H7 and Hb-A10-H6 behave similarly, the side chain of Phe81 is pointing towards the interior of the active site. The presence of the mandelonitrile influences slightly the equilibration angle, as shown by the second and third columns, to create space for aromatic ring of the substrate. At-A10-H6 and At-A9-H7 mutants show wild-type-like behavior regardless of the presence of ligand. F54V mutation in Hb-A10-H6 created enough space in the buried region for Phe81 to remain in the AtHNL like conformation. Introduction of the larger phenylalanine at position 54 did not influence the conformation of Phe81, only rare spontaneous reversible conformational changes can be observed during the trajectories, as illustrated by the slight increase in the standard deviation, and the decrease of the average values. The presence of the ligand favors the buried wild-type AtHNL like orientation, as the occurrence of the conformation facing the active site decreases from 30% to 10% of the time of the simulations. Cysteine in the wild-type HbHNL crystal structure is in the exposed conformation (-68°), it moves freely between the two conformations during the trajectories, but adopts a buried conformation in the majority of the time (84% of the time in the presence of the ligand and 66% for the apo structures).

Table A7. Fitted thermodynamic and equilibrium binding parameters for various ligands with wt-PcPAL and PcPAL-Y110F*  

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)-APEP</td>
<td>-6.5 [-7.8 - 5.8]</td>
<td>-1.2 [-2.3 - 0.3]</td>
<td>-7.7 [-8.1 - 7.3]</td>
<td>-8.5 [-9.6 - 7.7]</td>
<td>1.5 [0.5 - 2.9]</td>
<td>-7.0 [-7.2 - 6.8]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S)-APEP</td>
<td>-5.6 [-6.1 - 5.2]</td>
<td>-1.7 [-2.2 - 1.1]</td>
<td>-7.3 [-7.2 - 7.4]</td>
<td>-7.3 [-8.9 - 6.3]</td>
<td>0.1 [-1.2 - 1.2]</td>
<td>-7.2 [-7.5 - 6.9]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)-APP  A</td>
<td>-6.1 [-6.6 - 5.8]</td>
<td>-2.1 [-2.7 - 1.5]</td>
<td>-8.3 [-8.5 - 8.1]</td>
<td>-6.0 [ND - 3.6]</td>
<td>-0.5 [-3.6 - ND]</td>
<td>-6.5 [-7.3 - 3.8]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S)-APP  A</td>
<td>-9.7 [-10.1 - 9.4]</td>
<td>-0.6 [-1.4 - 0.2]</td>
<td>-10.3 [-10.7 - 9.9]</td>
<td>-10.8 [-11.4 - 10.2]</td>
<td>1.4 [-0.4 - 2.3]</td>
<td>-9.4 [-9.8 - 9.1]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Phe</td>
<td>-37.7 [-43.4 - 33.7]</td>
<td>31.2 [37.1 - 27.1]</td>
<td>-6.4 [-6.3 - 6.3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Values in square brackets denote the confidence intervals for the fitted parameters.

* ND: Confidence interval could not be determined for the parameter.
Figure A1. Representative ITC thermograms of PcPAL-Y110F with ligands: a. (R)-APEP, b. (S)-APEP, c. (R)-APPA, d. (S)-APPA, e. TCA, f. D-Phe, g. L-Phe.
NYILATKOZAT

Alulírott Bata Zsófia kijelentem, hogy ezt a doktori értekezést magam készítettel és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelölim.

Budapest, ...............................