

Catalytic Plasticity of the Aspartate/Glutamate Racemase Superfamily

Florian Alexander Fisch

Ph.D. Thesis

University of York, Department of Biology and Department of Chemistry

December 2009

Abstract

The bacterial and archaeal Asp/Glu racemase enzyme superfamily contains a variety of catalytic functions that have great potential for use in industrial biocatalysis. Members of this superfamily include aspartate racemases (AspRs), glutamate racemases (GluRs), hydantoin racemases (HydRs), arylmalonate decarboxylases (AMDs) and maleate *cis-trans* isomerases (MIs). Despite their catalytic diversity, all characterised members share the same protein fold, catalytic cysteine residues and reaction intermediate. Attempts to exploit this evolutionary flexibility for new processes have had limited success so far, showing that the employed mechanisms are not yet fully understood. For example, the well-characterised *Bordetella bronchiseptica* AMD (BbAMD) enantiospecifically decarboxylates a range of arylmalonates but is not able to decarboxylate alkylmalonates despite considerable efforts made by site directed mutagenesis.

In this work an investigation of the sequence diversity of the superfamily was undertaken and a range of BbAMD sequence homologues was tested for both aryl- and alkylmalonate decarboxylation (Chapter 3). However, none of the homologues exhibited decarboxylation activity. Targeted mutation of active site residues in an attempt to introduce decarboxylase activity was also unsuccessful. In an alternative approach to identify new alkylmalonate decarboxylating enzymes, a range of bacterial strains capable of processing alkylmalonates was isolated using selective enrichment from soil samples (Appendix D).

The only characterised superfamily enzymes without a described three dimensional protein structure are MIs. In order to illuminate the distinct mechanism of MIs, the activity of the superfamily member *Nocardia farcinica* MI (NfMI) was characterised (Chapter 4) and its structure was determined by X-ray crystallography (Chapter 5). A potent inhibitor and substrate analogue bromomaleate was found. Mutagenesis of the active site cysteine dyad confirmed its catalytic role and Cys76 was found to be more important than Cys194. The data support a mechanism initiated by nucleophilic attack by Cys76 on the double bond of maleate. Although alternative mechanisms cannot be excluded at present, these findings indicate that the mechanistic chemistry in the superfamily is more adaptable than previously thought.

Table of Contents

Abstract	3
Table of Contents	5
List of Tables	11
List of Figures	13
List of Abbreviations	17
For Family and Friends	21
Thesis made easy	21
I. A World of Molecules	22
II. Enzymes – Machines of Life	23
III. Using Enzymes in Industry	24
Acknowledgments	27
Author's Declaration	29
Chapter 1 General Introduction	31
1.1 Enzymes in Industrial Biocatalysis	31
1.2 Enzymatic Decarboxylation	35
1.2.1 PLP-Dependent Decarboxylation	37
1.2.2 Pyruvoyl-Dependent Decarboxylation	39
1.2.3 ThDP-Dependent Decarboxylation	40
1.2.4 Phosphopantetheine- and Biotin-Dependent Decarboxylation	42
1.2.5 Cofactor-Independent Acetoacetate Decarboxylase	43
1.2.6 Cofactor-Independent Phenolic Acid Decarboxylase	44
1.2.7 Cofactor-Independent Vanillate Decarboxylase	46
1.2.8 Cofactor-Independent Arylmalonate Decarboxylase	46



1.3	Asp/Glu Racemase Enzyme Superfamily	47
1.3.1	Glutamate Racemase	49
1.3.2	Aspartate Racemase	54
1.3.3	Hydantoin Racemase	55
1.3.4	Arylmalonate Decarboxylase	56
1.3.5	Maleate <i>cis-trans</i> Isomerase	61
1.4	Aims	63

Chapter 2 General Methods 65

2.1	Cloning	65
2.1.1	Strategy	65
2.1.2	Construct Preparation	66
2.1.3	Plasmid Transformations	67
2.1.4	Plasmid Purification and Verification	68
2.1.5	Source of DNA Samples	70
2.1.6	Agarose Gel Electrophoresis	70
2.1.7	Site Directed Mutagenesis	70
2.2	Sequence Alignments	71
2.2.1	Determination of Sequence Identity	71
2.2.2	Phylogenetic Tree	71
2.3	Protein Production and Purification	71
2.3.1	Protein Expression	71
2.3.2	Cell Lysis	72
2.3.3	Protein Purification	72
2.3.4	Protein Size and Purity Determination	73
2.3.5	Protein Quantification Assay	74
2.3.6	Analysing Protein Monodispersity	74

2.3.7	Analysing Secondary Structure	74
-------	-------------------------------	----

Chapter 3 Screening of BbAMD Homologues for Decarboxylation of Malonates **75**

3.1	Introduction	75
3.2	Methods	77
3.2.1	Chemical Synthesis of MPM	77
3.2.2	Analysis of Synthesised MPM	78
3.2.3	Strategy for Malonate Decarboxylation Assays	79
3.2.4	Extraction and Normal Phase HPLC Analysis	80
3.2.5	Whole Growing Cell Decarboxylation Assay	80
3.2.6	Whole Resting Cell Decarboxylation Assay	81
3.2.7	Cell Lysate Decarboxylation Assay	81
3.2.8	Pure Enzyme Decarboxylation Assay	81
3.2.9	Spectrophotometric Decarboxylation Assay	81
3.2.10	Homology Modelling of RseC	82
3.3	Results	82
3.3.1	Cloning and Expression of BbAMD Sequence Homologues	82
3.3.2	Decarboxylation Assays with BbAMD	84
3.3.3	Screening for Arylmalonate Decarboxylation in Homologues	86
3.3.4	Screening for Arylmalonate Decarboxylation in Mutants of RseC	88
3.3.5	Screening for Alkylmalonate Decarboxylation	90
3.3.6	Alternative Activities of Pfe	91
3.3.7	Crystallisation Trials with Pfe	93
3.4	Discussion	94
Chapter 4 Biochemical Characterisation of NfMI		97
4.1	Introduction	97

4.2	Methods	99
4.2.1	Cloning and Protein Purification of NfMI	99
4.2.2	Spectrophotometric Maleate Isomerisation Assay	99
4.2.3	HPLC Maleate Isomerisation Assay	100
4.2.4	Determination of Kinetic Constants and Inhibition Constants	101
4.2.5	GC/HPLC Enoate Reductase	102
4.2.6	MS of Enzyme-Ligand Adducts	104
4.3	Results	104
4.3.1	Activity Assays	104
4.3.2	Activity of NfMI	105
4.3.3	Effect of pH on NfMI Activity	106
4.3.4	Effect of Temperature and Redox State on NfMI Activity	107
4.3.5	Kinetics and Inhibition of NfMI Activity	111
4.3.6	Alternative Substrates	115
4.3.7	Active Site Mutants	117
4.3.8	Biotransformation in Combination with Enoate Reductases	121
4.4	Discussion	124
4.4.1	Conditions for Activity and Stability	124
4.4.2	Mechanism	126
4.4.3	Industrial Usefulness	129
Chapter 5	Crystal Structure of NfMI	133
5.1	Introduction	133
5.2	Methods	134
5.2.1	Cloning and Protein Purification of NfMI	134
5.2.2	His-tag Cleavage	134
5.2.3	Crystallisation	135

5.2.4	Data collection	135
5.2.5	Data Processing, Structure Solution and Refinement	135
5.2.6	Structure Validation and Analysis	136
5.2.7	Ligand Modelling and Docking	137
5.3	Results	137
5.3.1	Crystallisation	137
5.3.2	Data Collection, Phasing, Model Building and Refinement	139
5.3.3	Quaternary Structure and Overall Fold	140
5.3.4	Active Site Conformation and Substrate Modelling	144
5.3.5	Oxidation Resistance Involving Met197	149
5.4	Discussion	150
Chapter 6 General Discussion		153
6.1	Evolution of the Asp/Glu Racemase Superfamily	153
6.2	Decarboxylation of Malonates	154
6.3	Reaction Mechanism of NfMI	157
6.4	Outlook	159
Appendix A BbAMD Gene Sequence		163
Appendix B Sequence of pET-YSBLIC3C Plasmid		165
Appendix C Alignment of Asp/Glu Racemase Superfamily Members		169
Appendix D Isolation of Alkylmalonates Decarboxylating Bacteria		173
D.1	Introduction	173
D.2	Methods	174
D.2.1	Collection of Soil Samples	174
D.2.2	Preparation of Minimal Medium	174

	10
D.2.3 Liquid Culture Enrichment	175
D.2.4 In-Soil Enrichment	175
D.2.5 Purification of Agar Plates and Liquid Culture	176
D.2.6 HPLC Analysis	176
D.2.7 16S Sequencing	176
D.3 Results	177
D.3.1 Enrichment of Soil Cultures	177
D.3.2 Screening for Activity of the Isolates	179
D.4 Discussion	182
References	185

List of Tables

Table 2.1. List of primers.	69
Table 3.1. Selected homologues and properties.	83
Table 3.2. Summary of decarboxylating activities of RseC mutants.	90
Table 3.3. Summary of decarboxylating activities of homologues.	94
Table 4.2. Steady state kinetic parameters of NfMI mutants.	121
Table 5.1. Structure validation statistics.	136
Table 5.2. Data collection and refinement statistics.	140
Table 5.2. Coordination of metal ion between molecules A and D.	142
Table 6.1. Summary of tested alternative activities.	156

List of Figures

Figure 1.1. Industrial biocatalytic production of L-aspartate.	32
Figure 1.2. Industrial biocatalytic production of L-alanine.	32
Figure 1.3. Industrial biocatalytic production of D-phenylglycine.	33
Figure 1.4. Industrial biocatalytic production of L-malate.	33
Figure 1.5. Kinetic resolution.	33
Figure 1.6. Methods to produce pure enantiomers.	34
Figure 1.7. Quantitative enantiospecific conversion of 2-methyl-2-phenylmalonate to (<i>R</i>)-2-phenylpropionate.	35
Figure 1.8. Formation of carbanion after decarboxylation.	35
Figure 1.9. Cofactors employed by decarboxylases.	36
Figure 1.10. Stabilisation through Schiff base formation between ornithine and PLP. .	37
Figure 1.11. Binding of PLP to the active site of OrnD.	38
Figure 1.12. PLP and β -decarboxylation of aspartate.	39
Figure 1.13. Pyruvoyl stabilisation of the carbanion.	39
Figure 1.14. Stabilisation of the active carbanion form of ThDP.	40
Figure 1.15. Decarboxylation by ThDP through enamine stabilisation.	41
Figure 1.16. Decarboxylation of malonyl thioester.	42
Figure 1.17. Decarboxylation of malonate.	43
Figure 1.18. Decarboxylation of acetoacetate by Schiff base formation.	44
Figure 1.19. Decarboxylation of <i>p</i> -coumaric acid by PAD.	45
Figure 1.20. Possible mechanism for decarboxylation of vanillate.	46
Figure 1.21. Alignment of selected Asp/Glu racemase superfamily members.	48
Figure 1.22. Structure of the Asp/Glu racemase superfamily members.	49
Figure 1.23. Glutamate racemase activity.	50
Figure 1.24. Mechanism of GluR.	51

Figure 1.25. Residues involved in binding of D-glutamate in BsGluR.	52
Figure 1.26. Overlay of GluR structures.	53
Figure 1.27. Catalytic cascade in BsGluR.	54
Figure 1.28. Aspartate racemase activity.	55
Figure 1.29. Hydantoin racemase activity.....	56
Figure 1.30. Typical reaction of BbAMD.....	56
Figure 1.31. BbAMD activity and selection of representative substrates.	58
Figure 1.32. Structure of BbAMD.	59
Figure 1.33. Mechanism of BbAMD with MPM.	61
Figure 1.34. Cascading aldol reaction triggered by BbAMD.	61
Figure 1.35. Maleate <i>cis-trans</i> isomerase reaction.	62
Figure 2.1. LIC vector containing a gene of interest.	66
Figure 2.2. LIC procedure.	67
Figure 2.3. Typical purification of BbAMD homologues.	73
Figure 3.1. Activity of AMD.....	75
Figure 3.2. MPM production and purification.	79
Figure 3.3. Expression and purification of BbAMD homologues.	84
Figure 3.4. Substrates and products tested on BbAMD.....	85
Figure 3.5. Activity of BbAMD on PM.	86
Figure 3.6. HPLC chromatograms of a typical PM decarboxylation assay.....	87
Figure 3.7. Alignment of active sites of RseC model and BbAMD structure.....	89
Figure 3.8. Decarboxylation assays of malonate derivatives.	91
Figure 3.9. Assays for alternative activities of Pfe.	93
Figure 4.1. Activity of NfMI.	97
Figure 4.2. Expression and purification of NfMI mutants.	99
Figure 4.3. Absorbance spectra of maleate (M) and fumarate (F).	100
Figure 4.4. HPLC chromatograms of maleate and fumarate.	101

Figure 4.5. Reaction time course of NfMI.	106
Figure 4.6. Dependence of NfMI activity on pH.	107
Figure 4.7. Dependence of NfMI activity on temperature.	108
Figure 4.8. Temperature stability of NfMI.	109
Figure 4.9. Effect of the redox state on temperature stability of NfMI.	110
Figure 4.10. Long term stability of NfMI depending on redox state.	111
Figure 4.11. Compounds tested for NfMI inhibition.	112
Figure 4.12. Assay of substrate analogues as potential inhibitors of NfMI.	113
Figure 4.13. Kinetics of bromomaleate inhibition.	114
Figure 4.14. Substrates and products used with NfMI and enoate reductases.	116
Figure 4.15. Assay of potential alternative substrates of NfMI.	117
Figure 4.16. NfMI mutants activity assay.	119
Figure 4.17. Kinetic parameter determination of NfMI mutants.	120
Figure 4.18. Citral assay with NfMI and enoate reductases.	123
Figure 4.19. Maleate analogues assay with NfMI and enoate reductases.	124
Figure 4.20. Possible biotransformations of maleate and fumarate.	130
Figure 4.21. Production of fumarate and maleic anhydride.	131
Figure 5.1. Crystals of NfMI.	138
Figure 5.2. Asymmetric unit.	141
Figure 5.3. Three dimensional fold.	143
Figure 5.4. Active site.	145
Figure 5.5. Active site pocket.	145
Figure 5.6. Overlay of dioxyanion holes and ligands.	147
Figure 5.7. Scenarios of ligand binding.	149
Figure 6.2. Three possible reaction mechanisms.	159
Figure D.1. Selection of isolates on media with MPrM as sole carbon source.	178
Figure D.2. Isolates on media containing alkylmalonates as sole carbon source.	179

Figure D.3. HPLC analysis of isolates cultures growing on MPrM for 4 d.	180
Figure D.4. HPLC analysis of isolates cultures growing on EBM for 4 d.	182

List of Abbreviations

AAD	acetoacetate decarboxylase
ACP	acyl carrier protein
AMD	arylmalonate decarboxylase
ApGluR	<i>Aquifex pyrophilus</i> GluR
APS	ammonium persulfate
AspD	aspartate α -decarboxylase
AspR	aspartate racemase
AspβD	aspartate β -decarboxylase
BbAMD	<i>Bordetella bronchiseptica</i> AMD
BIS-TRIS-propane	1,3-bis(tris(hydroxymethyl)methylamino)propane
BLAST	basic local alignment search tool
BME	β -mercaptoethanol
BSA	bovine serum albumin
BsGluR	<i>Bacillus subtilis</i> GluR
BTB	bromothymol blue
CD	circular dichroism
CoA	coenzyme A
dATP	2'-deoxyadenosine 5'-triphosphate
DLS	dynamic light scattering
DMSO	dimethyl sulfoxide
dNTP	deoxyribonucleotide
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
e.e.	enantiomeric excess
EBM	2-ethyl-2-butylmalonate
EDTA	ethylenediaminetetraacetic acid
EfGluR	<i>Enterococcus faecium</i> GluR
EH	2-ethylhexanoate
ESI	electrospray ionisation

f1 ori	f1 phage origin of replication
FID	flame ionisation detection
GC	gas chromatography
gDNA	genomic DNA
GluR	glutamate racemase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high pressure liquid chromatography
HRV 3C	human rhinovirus 14
HydR	hydantoin racemase
IPTG	isopropyl β -D-1-thiogalactopyranoside
LacI	lactose repressor gene
LacO	lactose operator
LB	lysogeny broth
LfGluR	<i>Lactobacillus fermentum</i> GluR
LIC	ligation-independent cloning
MB	2-methylbutanoate
MEM	2-methyl-2-ethylmalonate
MES	2-(N-morpholino)ethanesulfonic acid
MI	maleate <i>cis-trans</i> isomerase
Mle	<i>Mesorhizobium loti</i> enzyme
MOPS	3-(N-morpholino)propanesulfonic acid
MP	2-methylpentanoate
MPM	2-methyl-2-phenylmalonate
MPrM	2-methyl-2-propylmalonate
MS	mass spectrometry
NAD(H)	nicotinamide adenine dinucleotide
NADP(H)	nicotinamide adenine dinucleotide phosphate
NCS	non-crystallographic symmetry
NfMI	<i>Nocardia farcinica</i> MI
NMR	nuclear magnetic resonance
ORF	open reading frame

ori	<i>Escherichia coli</i> origin of replication
OrnD	ornithine decarboxylase
OYE2	old yellow enzyme 2 from <i>Saccharomyces cerevisiae</i>
PA	2-phenylacetate
PAD	phenolic acid decarboxylase
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD	pyruvate decarboxylase
PEG	polyethylene glycol
Pfe	<i>Pyrococcus furiosus</i> enzyme
PhAspR	<i>Pyrococcus horikoshii</i> AspR
PLP	pyridoxal-5'-phosphate
PM	2-phenylmalonate
PMSF	phenylmethanesulphonylfluoride
PP	2-phenylpropionate
PPT	4'-phosphopantetheine
RBS	ribosome binding site
Rse	<i>Rhodococcus</i> sp. RHA1 enzyme
RT	retention time
Sae	<i>Streptomyces avermitilis</i> enzyme
Scd	<i>Streptomyces coelicolor</i> enzyme
SDS	sodium dodecyl sulfate
Sme	<i>Sinorhizobium meliloti</i> enzyme
Ste	<i>Sulfolobus tokodaii</i> enzyme
T7P	T7 phage transcription promotor
T7T	T7 phage transcription terminator
TAE	TRIS, acetic acid, EDTA
TCEP	tris(2-carboxyethyl)phosphine
TEMED	tetramethylethylenediamine
ThDP	thiaminediphosphate

TLC	thin layer chromatography
TLS	translation libration screw
TOF	time of flight detection
TRIS	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
UV	ultra violet
VD	vanillate decarboxylase
YqjM	old yellow enzyme 2 from <i>Bacillus subtilis</i>

For Family and Friends

Thesis made easy

A PhD thesis is a book of over a hundred pages that only very few people are ever going to read - if at all. Someone apparently put a sentence in the middle of his thesis stating he would pay a pint to anyone who would reach that part. The examiners apparently never claimed their prize.



The purpose of this book is firstly to check whether the author is worthy of progressing to the next level of his scientific career and secondly to be added to long bookshelves of different theses conservation institutions: the library, the supervisor and the family.

To prevent my thesis from being immediately lost in the eternal past of irrelevant human history, I wrote this uncommon introduction. I hope it will prevent the family copy from being used as a uniquely ornamental addition to their household.



Part I

is a reaction to my partner's question: "What is actually a **molecule**?" ...!? This question comes *after* several years in a relationship with a fanatic scientist and after four years of chemistry lessons in her pre-university formation! Part I will reintroduce the molecule to all others who were sleeping during their chemistry lessons too.

Part II

is there to convince you that **enzymes** are more than digestion. After reading it, you should be humbled every time you hear of enzymes as they support every single piece of your body.

Part III

is about how enzymes can be **used for chemical industry** - the ultimate aim of this thesis. (You should always aim high!) There will also be references to the different parts of my thesis.

I. A World of Molecules

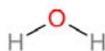
What are things around us made of? A Greek called Leucippus was asking himself this question 2500 years ago. If we divide things in smaller and smaller parts, could we go on forever or would we reach a limit that cannot be passed? Leucippus decided that there must be a moment when we cannot divide things further when we reach the "undividable" (a-tomos). These smallest possible undividable things would have all the properties of the substances they form.

Leucippus idea was very close to what others have shown much later to be true. We know today that things are indeed made of tiny particles. Atoms, however, are not as undividable as their name suggests. They break into even smaller particles when collided at the speed of light. The things that contain the properties of the substances, however, are arrangements of bound atoms also known as molecules. All things around us are made of molecules, except metals and minerals.

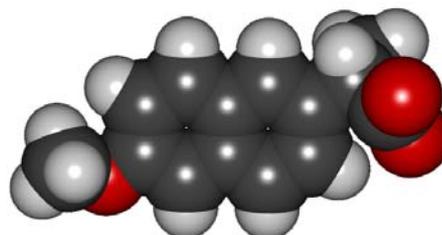
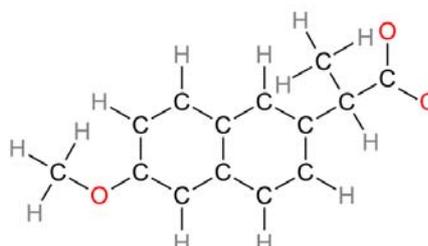
If we want to understand a substance (and chemists want that!) we have to understand the qualities of the molecules the substance is composed of. We have to understand the shape of the molecule, its weight and electrical charge.

Molecules like to stick to each other. That's why water stays in

Water



Naproxen



Two molecules. Water is one of the smallest molecules, made of only three atoms. Naproxen is a painkiller that is made of thirty atoms. Chemists represent molecules as letters and numbers: **Carbon (C)**, **hydrogen (H)** and **oxygen (O)**. To indicate the bonds between atoms lines are drawn or to indicate the size and shape of the molecule atoms are drawn as balls.

drops without the molecules going in all directions. When molecules are cooled down they stick to each other even more and form solid crystals like ice. When they are heated up they lose contact and fly away from each other like vapour.

When two molecules bump into each other they may exchange some atoms. This is called a chemical reaction. By reacting they give themselves new identities and start to be a completely different substance. For example, when six oxygen molecules (O_2) come together with one grape sugar molecule ($C_6H_{12}O_6$) they will react. The result is a process called burning, which results in six water molecules (H_2O) and six carbon dioxide molecules (CO_2) being formed and energy released. This only happens if the molecules are bumping into each other fast enough (by heating) or when enzymes are used.

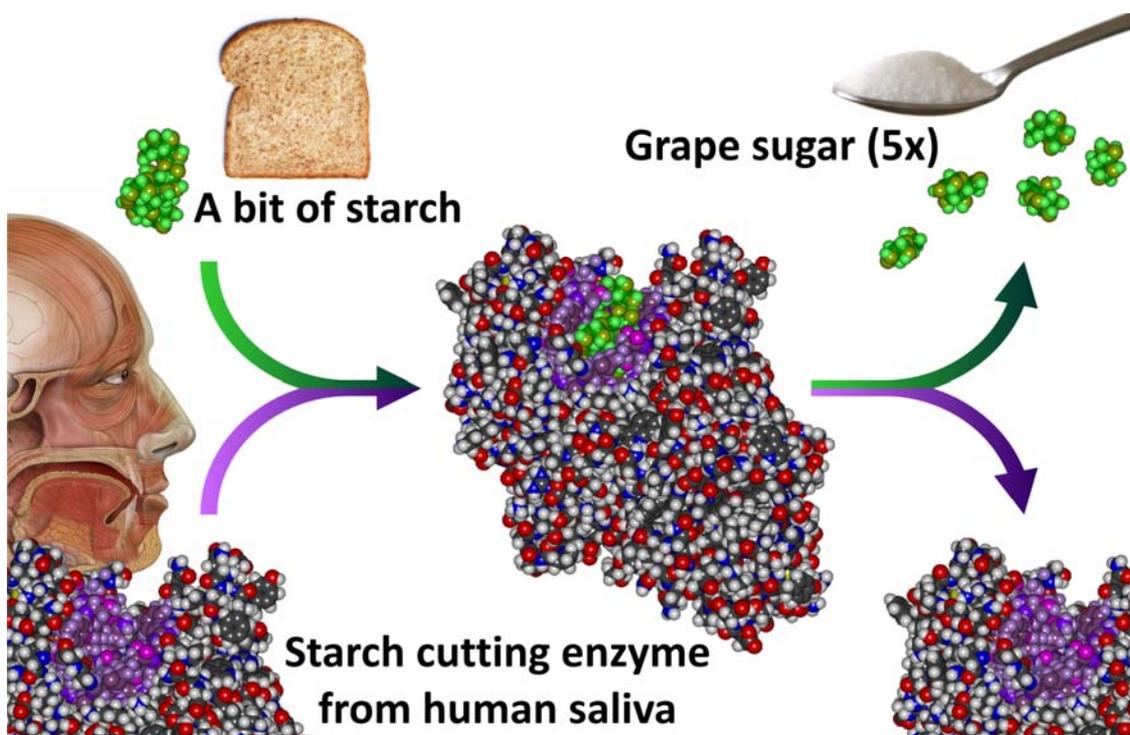
II. Enzymes – Machines of Life

When a tree grows, it has to convert small and simple molecules such as carbon dioxide and water into very large and complex molecules that make its sugars (the basis for wood) and lots of other molecules such as the green colour in the leaves. When in turn, we eat the tree's fruit, we are breaking down all these large molecules again into small and simple molecules to recover the energy that keeps us alive.

All these hundreds of thousands of reactions in the tree, as well as in our body, are done by enzymes. Without the enzymes we would be petrified (dead) in an instant. Our dead bodies wouldn't even decay as the microorganisms degrading the bodies would also be petrified instantly without enzymes.

Enzymes digest our food (large and complex molecules) in the mouth, stomach and gut, they make our muscles contract, our eyes see, our brain think and our body heat. These are chemical reactions where molecules are broken apart, flipped around and stuck together by enzymes. An enzyme can carry out a reaction very efficiently and without the need of intense heat.

An enzyme is a molecule itself (an extremely large and complex one). It contains a cavity where the molecule, which needs to be altered, fits in perfectly. Thanks to the perfect fit, the enzyme can push the atoms of the molecule around easily and thereby change the molecule. This is a bit like a key that fits perfectly into its lock, making it easy to open the door. Without the correct key it is hard to open a locked door. To understand how enzymes change molecules we need to know the precise shape of the enzyme and their cavity.



An enzyme reaction. By chewing a bit of bread for a long time we can feel some sweetness developing slowly. The reason for this is an enzyme in our saliva that cuts the long starch chains from bread, cereals and potatoes into its sugar molecule elements. The enzyme is able to cut fast and precisely because **starch** fits perfectly into the **cavity**. In the cavity, starch is optimally positioned for the enzyme to attack from the right angle. Our gut can only absorb the small sugars but not the large starch.

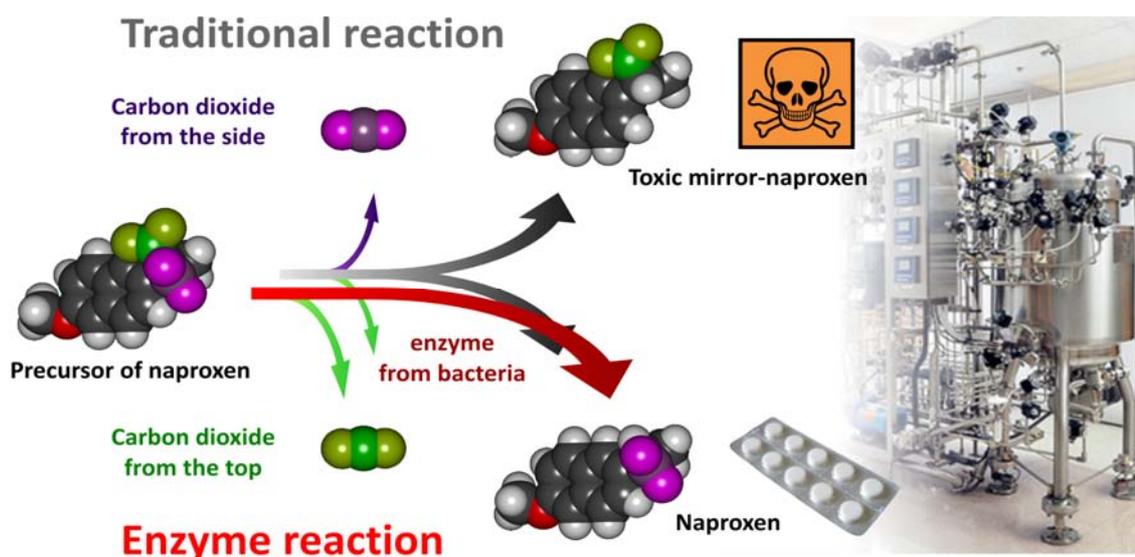
III. Using Enzymes in Industry

People have used enzymes for thousands of years already. Beer brewers use both the starch and the enzymes from barley grains. The enzymes break down starch in the grains when soaked in water. The released sugars are then broken down further by many other enzymes in yeasts to produce alcohol. Cheese is also made with the use of

enzymes, from calf stomach (rennet). The enzymes break down the milk proteins so they stick together and form a flan like texture.

Today, enzymes are also used in washing powders. They break proteins, fats and sugars that are part of the dirt on clothes. Genetic engineering, and thus this thesis, would not have been possible without enzymes. They can cut long DNA molecules in half, stick them together or assemble small molecules to produce a copy of the whole DNA molecule.

Traditional industrial chemistry is very crude compared to the chemistry of life performed by enzymes. Traditionally molecules often have to be dissolved in toxic solvents made from oil, heated to high temperatures to make them react and the result is not very pure. Only recently chemists have started to appreciate enzymes as they are useful for the production of drugs, pesticides, biofuels or new materials. It's not always easy because enzymes were "invented" to support life rather than industry.



Producing a drug with an enzyme. In traditional chemical reactions the carbon dioxides of the precursor are removed at random. The two molecules produced this way are hard to separate chemically as they are mirror images with exactly the same weight, colour and melting point. Nevertheless they have totally different biological effects: one is toxic and the other is a drug. The **carbon dioxide cutting enzyme** has no problem removing the correct carbon dioxide precisely from one side, thus producing pure naproxen painkiller.

My thesis was sponsored by the chemical company BASF who wanted me to look at a carbon dioxide cutting enzyme (see image above). They wanted me to change the enzyme so it would cut the carbon dioxide from different molecules instead.

There I went and checked lots of enzymes I was convinced would cut carbon dioxide (Chapter 3). Unfortunately, they didn't! I had to learn what most PhD students have to learn: coping with frustration. So I changed my approach and looked for completely unknown enzymes in soil bacteria. There are so many different bacteria in the world that there must be one having the enzyme I want (Appendix D). Luckily, there was one! But I found something that interested me even more. One of the enzymes tested for carbon dioxide cutting was instead a maleate switching enzyme. This is interesting because I can learn why the enzyme switches maleate rather than cuts carbon dioxide despite the very similar shape. I therefore first studied its activity in detail (Chapter 4) and then I determined its precise shape (Chapter 5). I don't completely understand the enzyme yet, but that's science: work in progress. And I love it!

Acknowledgments

In contrast to many people's idea, science is not solitary activity, where genius people in their small, hidden garden sheds suddenly discover the theory of everything. In the contrary, modern science stems from the foundation of social groups such as the British Royal Society. Members were gathering, performing experiments and discussing the observations. Today, we still gather in conferences, discuss posters and publish in the hope of being read.

It is my greatest pleasure to thank the people in the research groups **CNAP** and **YSBL** for their crucial contribution to the successful completion of my thesis. In particular I thank the following people:

Prof. Neil Bruce, my biology supervisor, who has kept me on track during the adversities of the research and who encouraged me to define my own direction.

Dr. Gideon Grogan, my chemistry supervisor, who critically challenged my premature thoughts and guided me in choosing, which were the important experiments to perform.

Dr. Carlos Martinez-Fleites, a post-doc, who taught me, during long hours, how to solve protein structures.

Marie Delenne, an exchange master student, who worked hard cloning most of my bacterial genes.

Dr. Rosamond Jackson, a post-doc, who shared her valuable experimental experience and helped me not to get lost in details.

Dr. Nina Baudendistel and **Dr. Bernhard Hauer**, the representatives of my sponsor, who introduced new ideas and gave me the opportunity for a stay at BASF.

Dr. Hazel Housden, **Dr. Joseph Bennett**, **Helen Sparrow** and **Nick Tilney**, who proofread some chapters in this thesis.

Silvia Ursprung, my partner, who helped me to organise myself and get the priorities in life right.

BASF, the chemical company, who financed this work.



**Académie des Sciences,
Paris, 1671**

Author's Declaration

I declare that I am the sole author of the work in this dissertation and that it is original except where indicated by special reference in the text. No part of the dissertation has been submitted for any other degree to any other institution.

Chapter 1

General Introduction

1.1 Enzymes in Industrial Biocatalysis

The production of biologically active compounds, such as pharmaceuticals and agrochemicals, is highly complex. The processes require specific alterations of mostly large molecules, which in addition have to be supplied in high purity. Economic considerations drive the search for fast and inexpensive production of these compounds.

Enzymes are increasingly accepted as an elegant way of meeting the needs of the chemical industry as they are highly chemo-, regio- and enantioselective. Non-enzymatic reactions and chemical catalysis are typically much less selective and therefore require more sophisticated purification processes and chemical protection/deprotection steps. In addition, enzymes are generally more efficient than traditional catalysts, can be combined to perform multiple reactions in one container and are sometimes capable of catalysing reactions that do not have an alternative production pathway in organic chemistry (1,2).

A positive side effect of the utilisation of enzymes is that they represent a step in the path of sustainable development in the chemical industry as they do not require toxic heavy metals or organic solvents and can work efficiently at low temperatures (1-3).

Enzymes are already used in many industrial applications, especially in the pharmaceutical industry, where in 2000 about 80% of the drug targets in the pipeline were chiral. The production of amino acids, carboxylic acids and alcohols make a large portion of the industrial applications (4).

Amino acids for example are synthesised in large quantities for human and animal nutrition. In 2000 amino acids were sold for about 3.5 billion Euros worldwide. L-aspartate is one of the most produced amino acids. It can be produced from fumarate and ammonia using aspartate ammonia lyase (Fig. 1.1, EC 4.3.1.1). In a further transformation, L-aspartate can be converted into L-alanine using the L-aspartate β -decarboxylase (EC 4.1.1.12). (Fig. 1.2) (4).

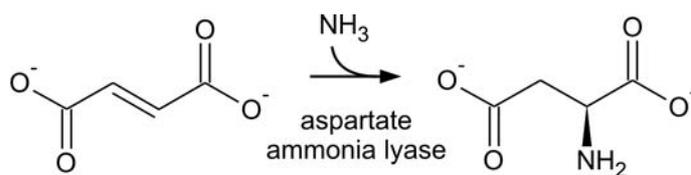


Figure 1.1. Industrial biocatalytic production of L-aspartate.

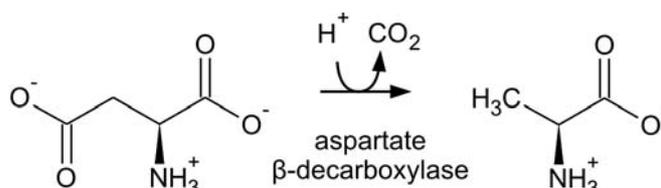


Figure 1.2. Industrial biocatalytic production of L-alanine.

Another, more general, way for the production of amino acids is the production from hydantoins in the “hydantoinase process”. The chemical production of racemic 5-substituted hydantoins is relatively easy. Enantiospecific hydantoin hydrolases (EC 3.5.2.2/X) can be used to open the ring and produce the corresponding carbamoyl. Enantiospecific carbamoyl hydrolases (EC 3.5.1.77/87) further transform them into the corresponding amino acids releasing ammonia and carbon dioxide. A hydantoin racemase (EC 5.1.99.5) can be used to achieve complete conversion. This is a well established process for D-phenylglycine and D-hydroxyphenylglycine used for the production of β -lactam antibiotics (Fig. 1.3) (4).

Carboxylic acids are also important products of the chemical industry. The acidification and stabilisation of fruit juices, for example, is most often achieved by the addition of the diacid L-malate. The enantioselective addition of water to fumarate is used in the production of L-malate performed with fumarate hydratase (EC 4.2.1.2, Fig. 1.4) (4).

Most chemical methods for the production of chiral products produce racemates or mixtures with low enantiomeric excess. The product purification is expensive, laborious and time consuming. Using an enantioselective biocatalyst with a lower catalytic efficiency for one product, significant improvements can be achieved. By stopping the reaction at the maximum enantiomeric excess, kinetic resolution can be achieved. Racemic alcohols are typically acylated in this way, yielding nearly 50% of the desired ester (Fig. 1.5) (5).

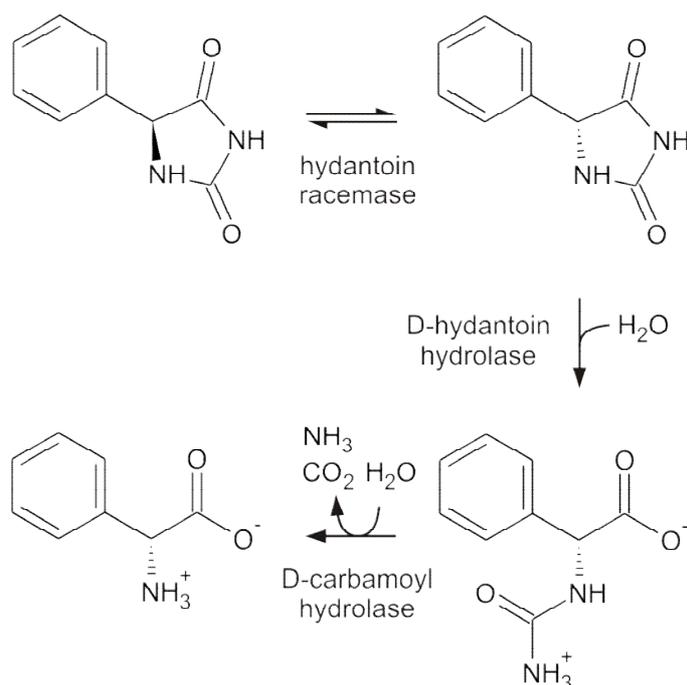


Figure 1.3. Industrial biocatalytic production of D-phenylglycine.

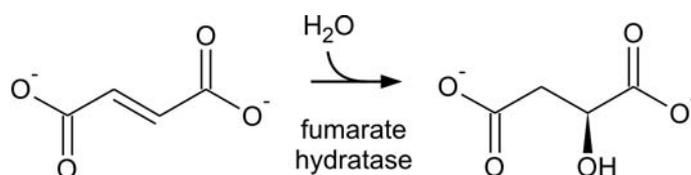


Figure 1.4. Industrial biocatalytic production of L-malate.

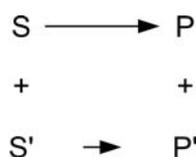


Figure 1.5. Kinetic resolution. S, P: substrate and product in opposite enantiomeric forms.

Although the product purity is increased, the maximum yield for a kinetic resolution process is still as low as 50%, resulting in economic and environmental costs. A number of processes have been developed to overcome this limitation and achieve close to 100% yields with nearly 100% enantiomeric excess. Dynamic kinetic resolution allows the replenishment of the depleted substrate enantiomer through the inclusion of substrate racemisation, which finally channels all the substrate through the faster transformation route (Fig. 1.6, left) (5). The hydantoinase process, mentioned above, is

a typical dynamic kinetic resolution that is applied industrially. Deracemisation of a racemic product is another method. In this process, the addition of an enantiospecific enzyme converts the undesired product enantiomer into a symmetric intermediate. The intermediate is then immediately non-specifically converted back to the product, which allows the correct product enantiomer to be accumulated. A well known example is the enantiospecific oxidation of a D-amino acid to the corresponding imine. The symmetrical imine is then reduced non-specifically with a chemocatalyst, leading to enrichment in L-amino acid. (Fig. 1.6, middle) (6). If enantiospecific enzymes are available to convert symmetrical pro-chiral substrates, full desymmetrisation can be achieved (Fig. 1.6, right) (7). The industrial production of L-aspartate or L-malate from fumarate by aspartate ammonia lyase and fumarate hydratase, mentioned above, are typical desymmetrisations.

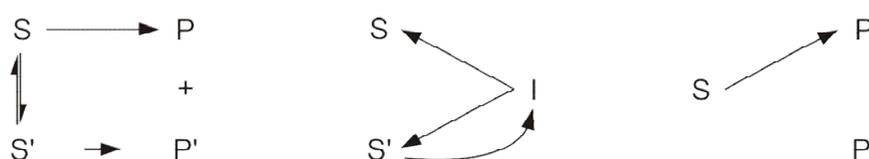


Figure 1.6. Methods to produce pure enantiomers. Dynamic kinetic resolution (left), deracemisation (middle) and desymmetrisation (right). S, P, I: Substrate, product and intermediate with opposite enantiomer if available.

Arylmalonate decarboxylase (AMD) is a prime example of an enzyme achieving desymmetrisation. Decarboxylation is a thermodynamically highly favourable reaction with a yield of nearly 100%. AMD decarboxylates disubstituted aryl- and alkenylmalonates to give apparent 100% enantiomeric excess (Fig. 1.7). It is a potentially useful biocatalyst and has therefore sparked a lot of interest. In particular, the 2-arylpropionates are used for the production of ferroelectric liquid crystals, pyrethroids used as insecticides and non-steroidal anti-inflammatory drugs such as flurbiprofen (3,8-13).

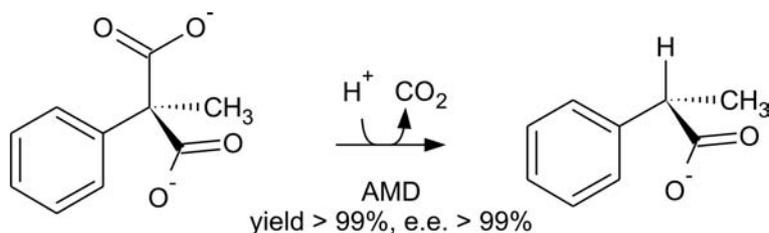


Figure 1.7. Quantitative enantiospecific conversion of 2-methyl-2-phenylmalonate to (*R*)-2-phenylpropionate.

1.2 Enzymatic Decarboxylation

Carboxylations and decarboxylations are crucial processes for life. The biosphere largely depends on the fixation of carbon dioxide from the atmosphere by ribulose biphosphate carboxylase, present in the chloroplasts of plants and algae. The organic molecules formed by photosynthesis are in turn degraded by many heterotrophic organisms through decarboxylation, mainly in the citric acid cycle. Decarboxylations are thermodynamically favourable in physiological conditions, also due to an increase in entropy through the release of carbon dioxide. Therefore carboxylations can drive other unfavourable reactions such as the synthesis of NADH or fatty acids (14).

In contrast to the above, carboxylates are typically very recalcitrant to decarboxylation. This is particularly the case for monocarboxylates such as fatty acids. The carbanion formed by the electron transfer from the oxyanion to the α -carbon is very unstable and needs to be stabilised effectively in order to drive the reaction (Fig. 1.8). The dicarboxylate malonate can be decarboxylated at relatively low temperatures (135 °C) due to the delocalisation of the negative charge between the remaining carbonyl oxygen and the α -carbon, forming an enol (15).



Figure 1.8. Formation of carbanion after decarboxylation.

To stabilise the negatively charged carbanion intermediate, many decarboxylases rely on cofactors such as PLP, ThDP, phosphopantetheine and biotin (Fig. 1.9) (16). However, some decarboxylases are capable of stabilising the intermediate without cofactors. Consequently, the catalytic burden must be carried solely by the enzyme, the substrate or both.

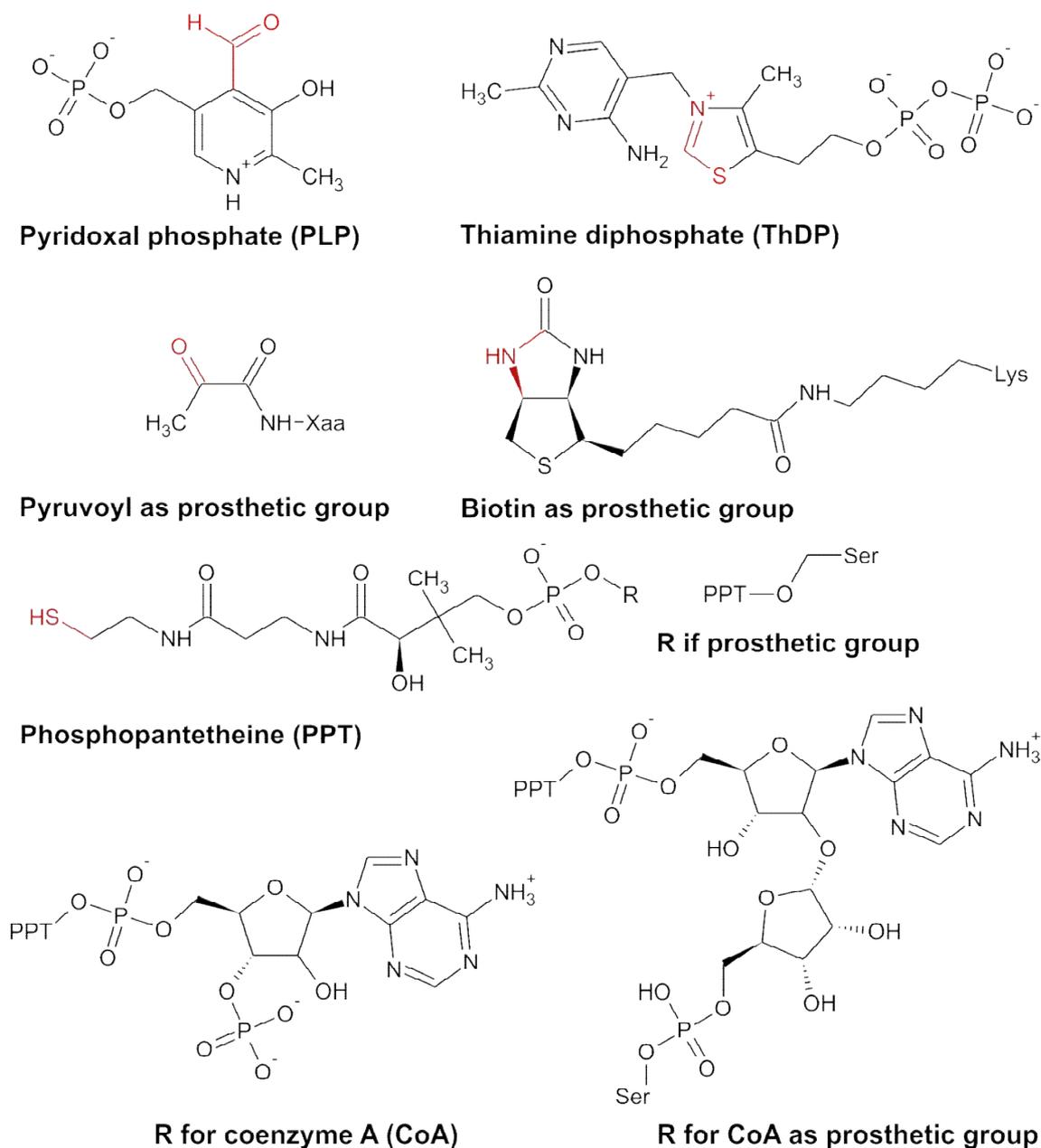


Figure 1.9. Cofactors employed by decarboxylases. The catalytic portions of the cofactors are highlighted in red.

1.2.1 PLP-Dependent Decarboxylation

The cofactor PLP is recruited by enzymes for many reactions involving amino acids. The condensation of the amine with the aldehyde forms a Schiff base. Both the protonated imine nitrogen and the pyridine ring act as efficient electron sinks. The pyridine ring is able to do this with the series of conjugated double bonds allowing the removal of the α -carboxyl of the amino acid. (Fig. 1.10) (16).

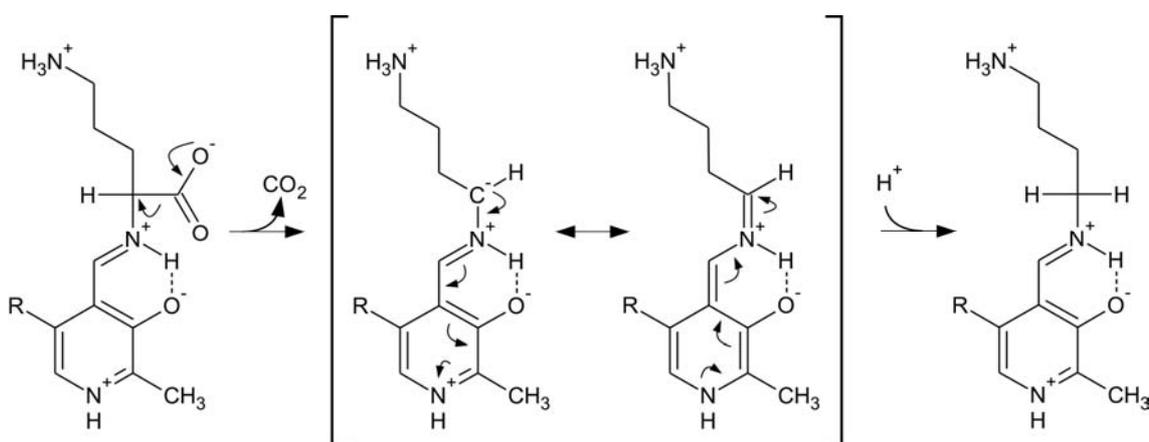


Figure 1.10. Stabilisation through Schiff base formation between ornithine and PLP.

A number of amino acid decarboxylases use that mechanism, including the ornithine decarboxylase (OrnD, EC 4.1.1.17). OrnD produces the diamine putrescine from the amino acid ornithine. The decarboxylation is the first step in the production of polyamines necessary for cell growth and differentiation. OrnD is therefore a target for antibiotics against parasites such as a *Trypanosoma brucei*, the agent of sleeping sickness (17).

There are two folds for OrnD, the α/β -fold for prokaryotes and the α/β -barrel-fold for eukaryotes. Eukaryotic OrnD is closely related to arginine decarboxylases (EC 4.1.1.19) and diaminopimelate decarboxylase (EC 4.1.1.20), which are all distantly related to alanine racemases (EC 5.1.1.1) (18). OrnD is found as a homodimer in solution where the active site is built by the N-terminus of one monomer and the C-terminus the another. Some crucial amino acids have been identified for the reaction mechanism of OrnD (numbering as in *T. brucei*). Lys69 forms a Schiff base with PLP when no substrate is bound. Glu274 stabilises the positive charge on the pyridine nitrogen and Arg277 binds the negatively charged phosphate. Lys69 is also thought to

be responsible for protonation of the intermediate that proceeds through inversion of configuration (Fig. 1.11). In prokaryotic OrnD the configuration is retained (17,19,20).

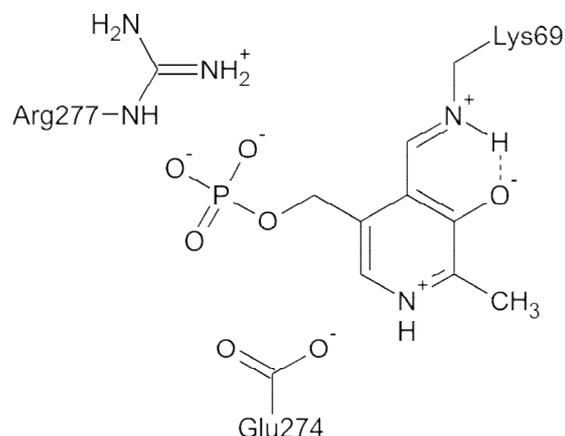


Figure 1.11. Binding of PLP to the active site of OrnD.

Upon substrate binding, first a *gem*-diamine is formed and then Lys69 is released through Schiff base formation with the substrate. The reaction is reversed for product release, which was found to be the rate limiting step through stop flow kinetics experiments. After decarboxylation, the formed quinonoid intermediate stabilises the negative charge before the substrate α -carbon is protonated. All reaction steps, including the non substrate-bound, are tautomers between the charged form having a protonated imine nitrogen and a deprotonated pyrimidine hydroxyl and their uncharged forms (21).

Aspartate β -decarboxylase (Asp β D, EC 4.1.1.12) is another type of PLP-dependent decarboxylase. This enzyme is widely used in industrial biocatalysis for the production of alanine. Asp β D is related to the aspartate transaminase (EC 2.6.1.1) and also uses a similar mechanism. Instead of removing the α -carboxyl it removes the proton at the same position. In *Pseudomonas dacunhae* this is done by Lys315. After the deprotonation β -decarboxylation can occur. In the case of *P. dacunhae* Asp β D the same Lys315 protonates the quinonoid at the carbonyl carbon of PLP. After the decarboxylation at the β -carbon, the negative charge goes to the double bond with the α -carbon and the imine nitrogen is neutralised. An as yet uncharacterised acid protonates the methylene, inverting the stereochemistry. The reaction is then reversed with Lys315 deprotonating the carbonyl and subsequently reprotonating the α -carbon retaining the configuration (Fig. 1.12) (22,23).

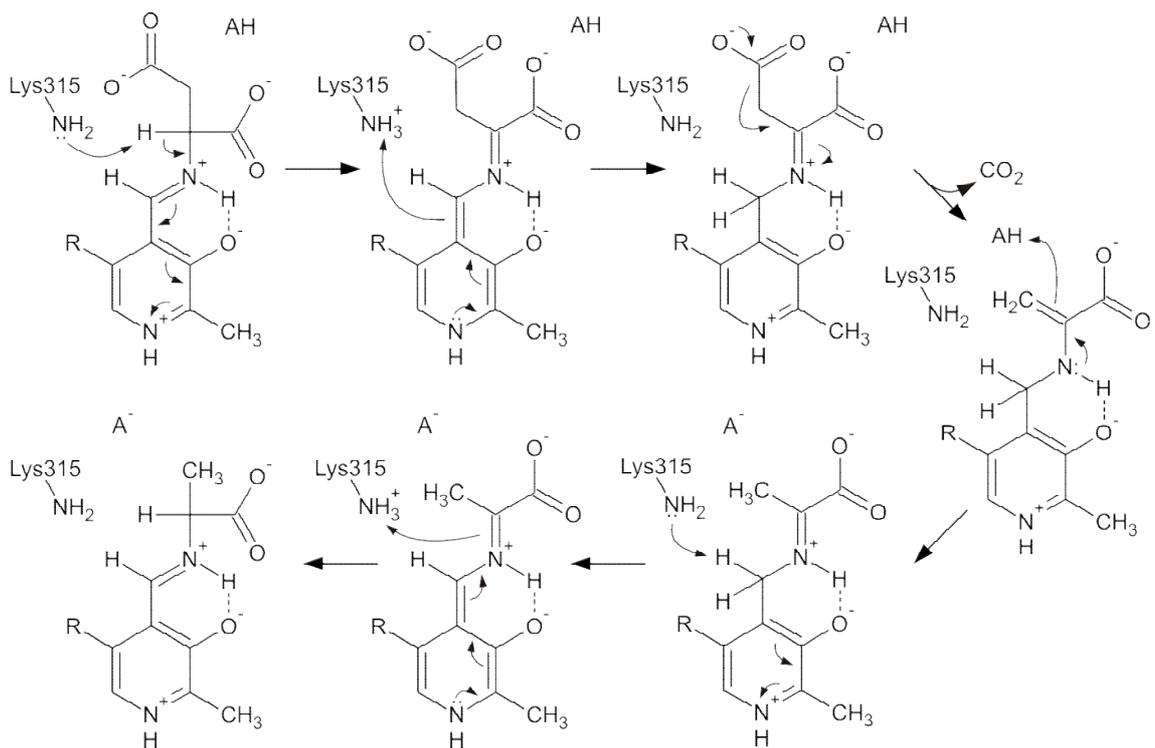


Figure 1.12. PLP and β -decarboxylation of aspartate.

1.2.2 Pyruvoyl-Dependent Decarboxylation

Pyruvoyl cofactor functions in an analogous way to PLP. It forms a Schiff base between its carbonyl and the amino group of the amino acid and stabilises the α -decarboxylation by acting as an electron sink with a series of conjugated double bonds (Fig. 1.13) (24).

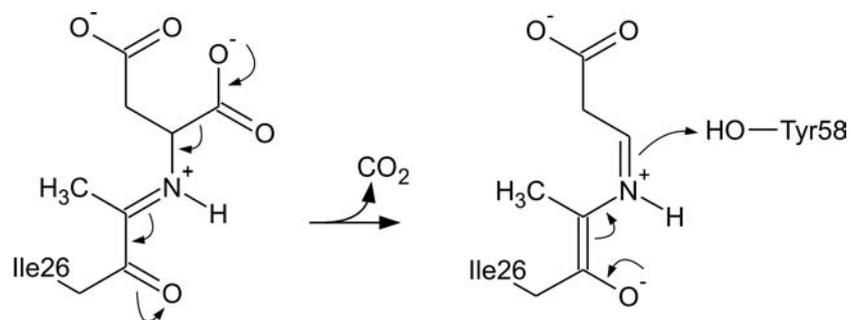


Figure 1.13. Pyruvoyl stabilisation of the carbanion.

Pyruvoyl is typically used by aspartate α -decarboxylase (AspD, EC 4.1.1.11). AspD is a prokaryotic enzyme involved in the production of β -alanine necessary for the formation of phosphopantetheine, a cofactor used for carboxyl transfer (see below). As animals

do not have AspD, it is a good target for antibiotics. Recently, it was proposed as a biocatalyst for the industrial production of nitrogen containing compounds from organic waste material, where the aspartate to β -alanine is a central step in the pathway (25).

Pyruvoyl is a special cofactor as it is formed by an auto-catalytic process, creating and modifying an internal N-terminus. In *Helicobacter pylori* the amide bond between Gly24 and Ser25 is broken transforming Ser25 to a dehydroalanine, which in turn is deaminated. The formed ketone is the catalytic ketone in pyruvoyl cofactor effectively attached to Ile26 (26).

1.2.3 ThDP-Dependent Decarboxylation

The thiazolium C2 carbanion is the active centre of ThDP. The generally unstable carbanion is stabilised by several factors. One factor is the positive charge on the neighbouring tetravalent nitrogen that stabilises the ylid form electrostatically. Another factor is that, when bound to enzymes, ThDP adopts a "V" conformation, not found in free ThDP, that orientates the amine at position 2 of the pyrimidine ring so it is closest to the thiazolium C2 carbon. Furthermore, the aminopyrimidine tautomerises to form an iminopyrimidine, the thiazolium C2 carbon is deprotonated by the imine (Fig. 1.14). This equilibrium is not leaning to the right, but favoured enough to allow reactions to occur at reasonable rates (16,27).

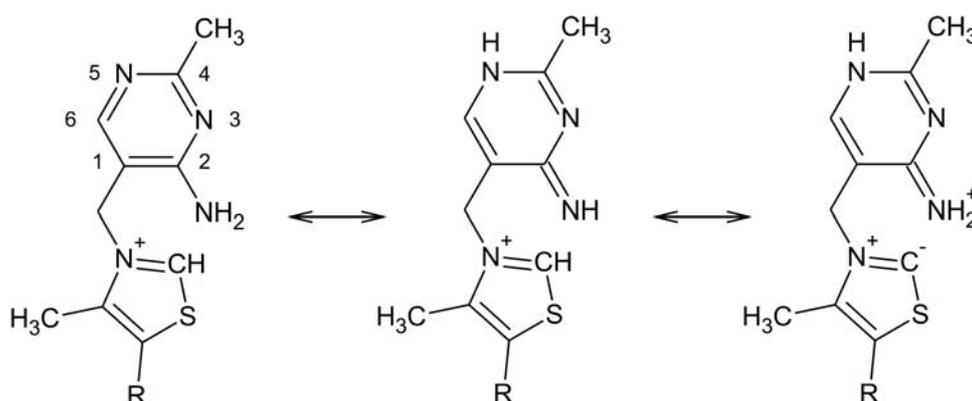


Figure 1.14. Stabilisation of the active carbanion form of ThDP.

Pyruvate decarboxylase from *Saccharomyces cerevisiae* (PD, EC 4.1.1.1) is a typical example of a ThDP-dependent decarboxylase. Although mechanistically very similar, PD is not to be confused with the E1 subunit of the pyruvate dehydrogenase best described in *Escherichia coli* (EC 1.2.4.1) that also relies on ThDP. PD is the first

enzyme in the ethanol production pathway. In PD, the thiazolium carbanion conducts a nucleophilic attack on central carbonyl carbon of pyruvate that is concomitantly protonated. The negative charge left on the decarboxylated intermediate is stabilised in the ThDP thiazolium through the formation of an ylid and through resonance between the carbanion form and the enamine form. The stabilised carbanion is finally protonated and released as acetaldehyde (Fig. 1.15) (16,27).

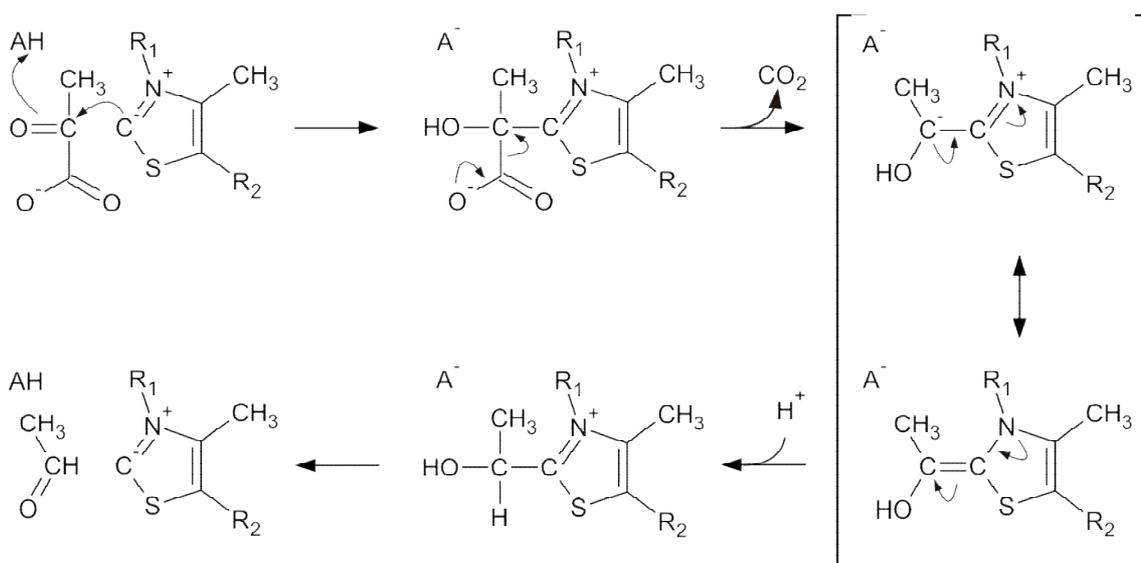


Figure 1.15. Decarboxylation by ThDP through enamine stabilisation.

PD forms a dimer of dimers with two active sites on the tight interfaces of each dimer. The diphosphate moiety of ThDP is bound to PD through the intermediate of a bound Mg^{2+} ion. Ile415 maintains the ThDP in the "V" conformation through its position between the pyrimidine ring and the tetrazolium ring (28). It was shown, mainly by mutagenesis experiments, that most of the catalysis is performed by the cofactor rather than by the protein. No proton transfer was found to be rate limiting and no particular residue was indispensable for catalysis. The role of the protein is mainly to position the reaction partners involved. This involves binding the cofactors and substrate as well as maintaining ThDP in the "V" form. There is a heavy strain imposed onto ThDP by the enzyme. The release of that strain is thought to be the main driving force for the decarboxylation rather than electrostatic repulsion by negatively charged residues. It is also thought that the active site provides an environment with a lower dielectric constant favouring the formation of ylid and thus dramatically decreasing the pK_a of the C2 proton on the thiazolium by 9 to 10 units (27).

1.2.4 Phosphopantetheine- and Biotin-Dependent Decarboxylation

Phosphopantetheine is a cofactor that forms a thioester with malonate. The formation of an ester effectively removes the negative charge on the carboxylate making it easier to stabilise the negative charge left after the decarboxylation. Thioesters can also accommodate the negative charge better than esters (Fig 1.16) (29).

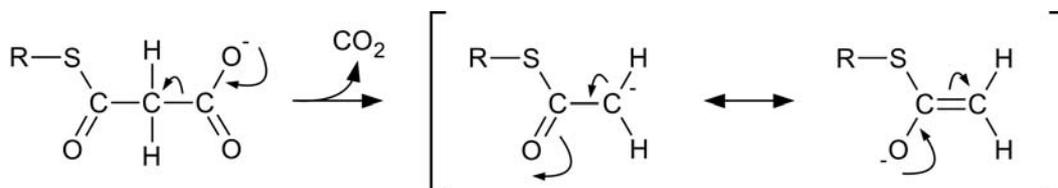


Figure 1.16. Decarboxylation of malonyl thioester. Stabilisation by resonance.

Phosphopantetheine in cells is either linked to a phosphoadenyl to form free CoA or as a prosthetic group via a serine to an acyl carrier protein (ACP). In some prokaryotes the ACPs have a dephosphorylated form of CoA linked via a phosphoribosyl to a serine (Fig 1.9). The thioesters of these phosphopantetheine derived cofactors can readily exchange bound acyl groups and therefore act as acyl carriers. In addition, their elongated form helps to function as flexible linkers presenting the substrates to different enzyme subunits or domains. This is well known for eukaryotic fatty acid syntheses where the phosphopantetheinyl group receives malonyl that subsequently elongates an acyl by two carbons with concomitant decarboxylation. The elongated chain is then reduced in three steps and translocated, which all happens while the group is bound to the phosphopantetheinyl group (14,30).

Biotin is a carrier of activated carbon dioxide. It is a prosthetic group that is linked to a lysine of a protein via an amide bond. As for phosphopantetheine, this makes it a long flexible linker that is able to transport carbon dioxide to different active sites. The carbon dioxide is covalently attached to the amino group of an amide that can be decarboxylated easily as the negative charge is well delocalised and neutralised through protonation of the nitrogen (14).

Malonate is both an energy source and an inhibitor of the citric acid cycle. A number of prokaryotes have evolved a malonate decarboxylase that converts malonate to acetate. They use a phosphoribosyl linked CoA prosthetic group as carrier of the bound

malonyl. The enzyme consists of many subunits. In one group of organisms (*Malonomonas rubra*) the enzyme is membrane bound and biotin dependent (EC 4.1.1.88) whereas in the other group the enzyme is soluble and does not require biotin (EC 4.1.1.89). The soluble proteins can be found in either the periplasm (*Acinetobacter calcoaceticus*) or the cytoplasm (*Pseudomonas ovalis*). In biotin dependent malonate decarboxylases the decarboxylation on ACP occurs via biotin that conserves the energy in order to drive a Na⁺ pump (EC 4.3.99.2, Fig. 1.17) (31-33).

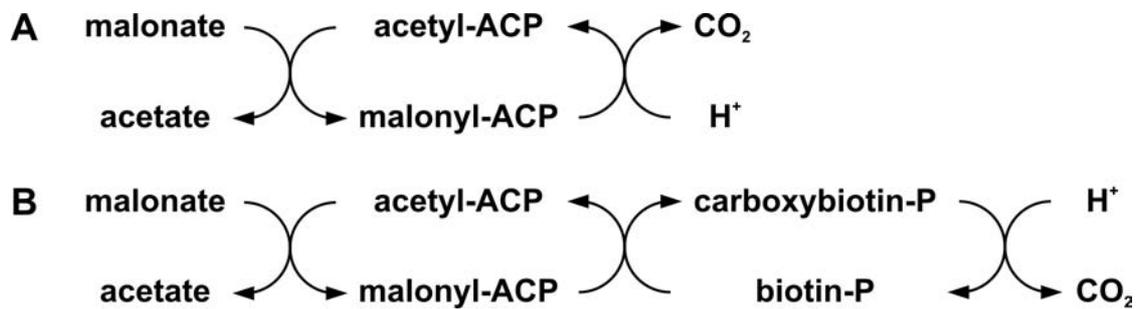


Figure 1.17. Decarboxylation of malonate.

In all cases the first step is a transacylation that releases the acetyl and binds malonyl via the thioester. The decarboxylation of malonyl occurs with retention of the stereochemistry as shown with isotope labelling and substrate analysis with NMR. The malonyl thioester decarboxylation resembles the decarboxylation observed in eukaryotes in respect to retention of configuration and type of cofactor (34,35).

1.2.5 Cofactor-Independent Acetoacetate Decarboxylase

Schiff bases are not only formed with cofactors such as PLP. They are also formed in cofactor-independent proteins through the condensation of lysines with carbonyl groups in ketones or aldehydes. This arrangement equally allows the stabilisation of the negative charge formed by decarboxylation as the positively charged Schiff base acts as an electron sink (Fig. 1.18) (16).

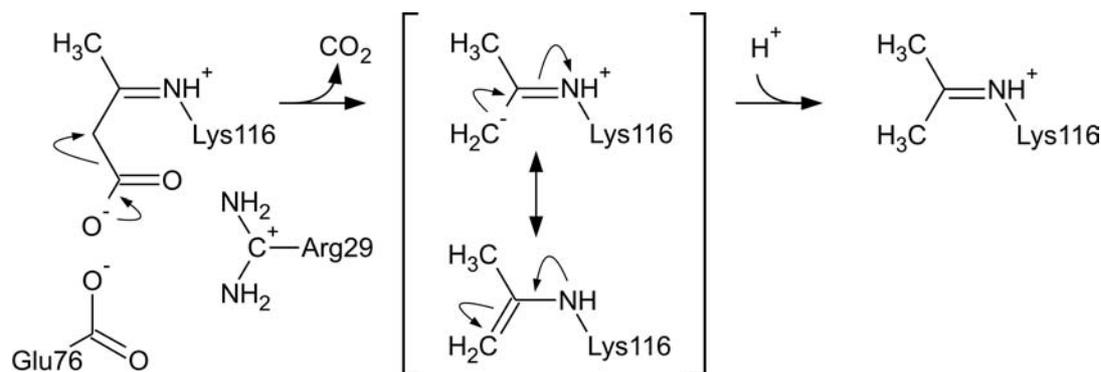


Figure 1.18. Decarboxylation of acetoacetate by Schiff base formation.

Acetoacetate decarboxylase (AAD, EC 4.1.1.4) is an enzyme used by solventogenic bacteria such as *Chromobacterium violaceum*. The conversion of acetoacetate into carbon dioxide and acetone is important in the anaerobic metabolism. The fold of AAD has not been found in other enzymes (PDB: 3bh3). Its central part consists of a β -cone that is exposed to the solvent on the wider end and contains the active site at the narrow end. The substrate is therefore effectively channelled to the active site. The protein forms a dodecamer thought to be involved in maintaining the stability towards low pH and high solvent as well as effectively concentrating the products of the reaction to one localisation (36).

AAD was long seen as the classical example for the microenvironment effect in the active site changing the pK_a by neighbouring residues. In AAD the catalytic Lys116 was shown to have a pK_a of 6 being 4.5 units below the expected pK_a of lysine. Using an optical reporter group Lys115 was found to electrostatically perturb the Lys116 so that it was mainly in the deprotonated state (37,38). The recently solved structure of AAD showed, however, that Lys115 could not be the cause of the pK_a change but that the hydrophobic environment of Lys116 was the cause for charge destabilisation. The observed effect on the reporter group was most probably caused by the neighbouring Arg29 that is involved in binding the carboxylate. Another charged residue, Glu76, is destabilising the negative charge on the carboxylate and the position is maintained by electrostatic repulsion of Glu61 (Fig. 1.18) (36).

1.2.6 Cofactor-Independent Phenolic Acid Decarboxylase

The uncatalysed decarboxylation of *p*-coumarate occurs at 150 °C in a non-stereospecific manner. Whereas, the pyrolytic decarboxylation of [8D]-*p*-coumarate yields both *cis*- and *trans*-[8D]-hydroxystyrene, the phenolic acid decarboxylase (PAD)

from *Klebsiella oxytoca* specifically produces the *trans* form, thus retaining the stereochemistry. PAD does not accept the *cis* isomer of *p*-coumaric acid. Substrates of PAD include ferulate and caffeate, whereas *m*-coumarate is not accepted, suggesting the requirement of a 4-hydroxyl substituent on the aromatic ring. The mechanism of decarboxylation is therefore thought to occur via a *p*-quinone methide intermediate with the substrate acting as its own electron sink (Fig. 1.19, (39,40)).

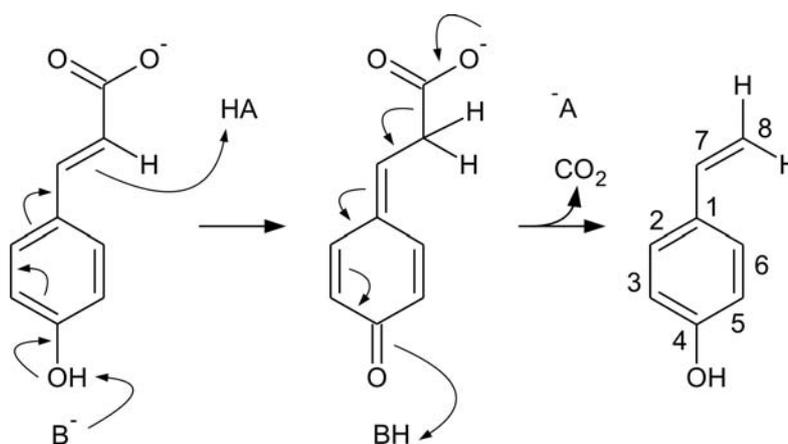


Figure 1.19. Decarboxylation of *p*-coumaric acid by PAD

The structure of *Bacillus subtilis* PAD has recently been solved and released but no accompanying publication exists. The structure forms a dimer of two orthogonal β -sheets that form the active site between the sheets. The structure contains two ethane diols in the active site that span approximately the length of *p*-coumarate. There are many aromatic residues around the active site and its mouth that are candidates for binding the aromatic moiety of the substrate of which Tyr31 is the best positioned. A series of other residues could be involved in protonation and deprotonation of the substrate such as Tyr11 and Tyr13 on one end and Arg41, Glu64 and Thr68 on the other end (PDB: 2p8g, K194Y mutant) (41).

PAD is of economic interest as it is involved in the degradation of cinnamic acids to 4-hydroxyethyl phenols that are components of bad odours in wine (40). The cofactor-independence of PAD also makes it a useful biocatalyst for the production of a range of 4-hydroxystyrenes.

1.2.7 Cofactor-Independent Vanillate Decarboxylase

Vanillate decarboxylase (VD) from *Nocardia* sp. NRRL 5646 was described as a cofactor-independent decarboxylase producing guaiacol. It was found to specifically incorporate solvent deuterium into position 1 of the aromatic ring but not into the hydroxyl. VD is specific for 4-hydroxylated benzoates including vanillate, 4-hydroxybenzoate, protocatechuate, syringate. Substrates having the hydroxyl group in a different position such as isovanillate, 2- or 3-hydroxybenzoate were not substrates. Neither was 4-aminobenzoate or the substrates for PAD *p*-coumarate and ferulate. By analogy to PAD the mechanism is thought to occur via a quinonoid intermediate (Fig 1.20) (42).

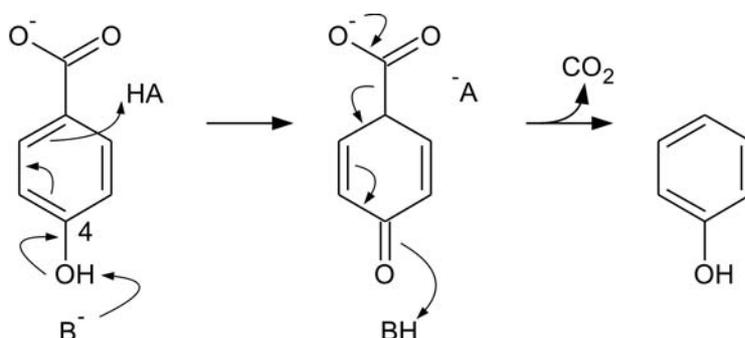


Figure 1.20. Possible mechanism for decarboxylation of vanillate.

VD is a 23 kDa protein found to be dimeric in solution, but no structural data is available for it. It was found to be inhibited by thiol reagents. A 21 residue long N-terminal sequence fragment is identical to superoxide dismutase (EC 1.15.1.1) sequence from *Nocardia asteroides* except for the first residue. The substrate specificity, however, matches with 4-hydroxybenzoate decarboxylase (EC 4.1.1.16) that was reported to depend on metal ions (42-45).

1.2.8 Cofactor-Independent Arylmalonate Decarboxylase

Arylmalonate decarboxylase (AMD, EC 4.1.1.76) is another cofactor-independent decarboxylase. It typically converts 2-methyl-2-phenylmalonate (MPM) into (*R*)-2-phenylpropionate ((*R*)-PP) in a stereospecific manner. As it belongs to the Asp/Glu racemase superfamily, it will be described further in the following section.

1.3 Asp/Glu Racemase Enzyme Superfamily

New enzyme activities evolve from old protein folds resulting in the formation of diverse enzyme families. This may occur through the retention of substrate specificities, protein architecture or mechanistic property. The latter leads to mechanistically diverse superfamilies with low sequence identities. A reaction step or the stabilisation mode of an intermediate is conserved while substrate and product change significantly. The first identified superfamily was the enolase superfamily that conserved the α -proton abstraction of a carboxylate. In 2001 twelve distinct reactions of the enolase superfamily were known, including mandelate racemase, muconate lactonising enzyme and enolase. Other superfamilies have since been identified such as the crotonase, the thyl radical and the amidohyrolase/phosphotriesterase superfamilies (46).

Based on sequence identities the versatile Asp/Glu racemase enzyme superfamily comprising glutamate racemase (GluR, EC 5.1.1.3), aspartate racemase (AspR, EC 5.1.1.13), hydantoin racemase (HydR, EC 5.1.99.5), maleate *cis-trans* isomerase (MI, EC 5.2.1.1) and arylmalonate decarboxylase (AMD, EC 4.1.1.76) activities was defined with an average overall sequence identity of 22% (Fig. 1.21, Pfam: PF01177) (47). This superfamily contains over 1500 known homologous sequences.

The conserved fold of the superfamily is composed of two pseudo-symmetrical domains most probably arising from an ancestral gene duplication. The active site cleft is located between the domains and is buried under two covering loops. Most superfamily members contain two opposed catalytic cysteines, one on each domain (Fig. 1.22). The superfamily is thought to share a catalytic enediolate reaction intermediate (48).

The enzymes of the Asp/Glu racemase superfamily have sparked interest in the chemical industry because they are small proteins mainly from bacterial or archaeal origin and are therefore relatively easy to handle. The main advantage of the superfamily members is their independence from cofactors (49-52).

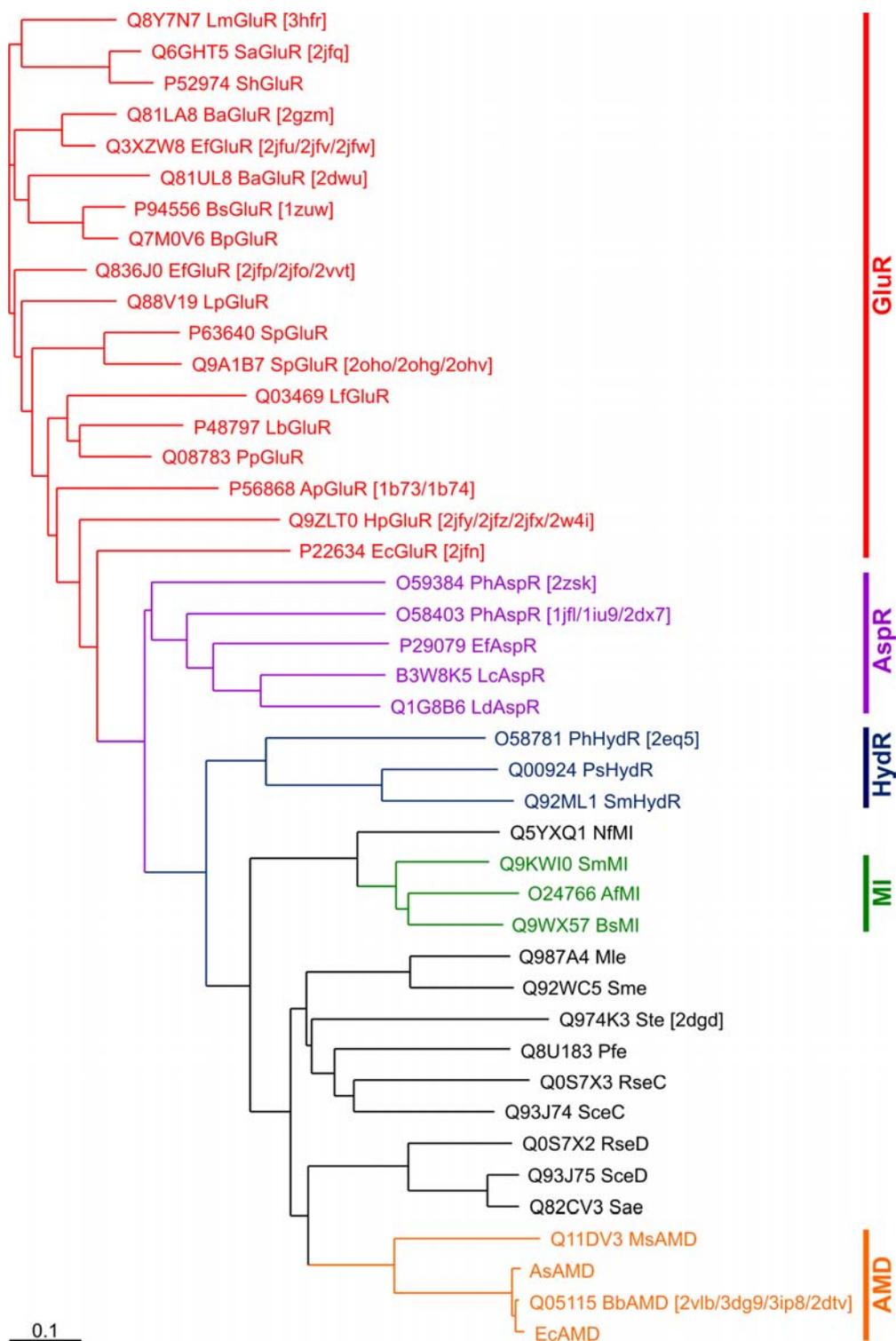


Figure 1.21. Alignment of selected Asp/Glu racemase superfamily members. Sequences of all selected homologues (black, see Chapter 3, Table 3.1), all experimentally confirmed activities and all known structures were aligned using ClustalW2. Four highly conserved stretches from the active site (see Appendix C) were used to calculate a ClustalW2 guide tree based on sequence identity (for details see Chapter 2). UniProt accession codes are indicated followed by the initials of the species, the activity and PDB identifiers in square brackets.

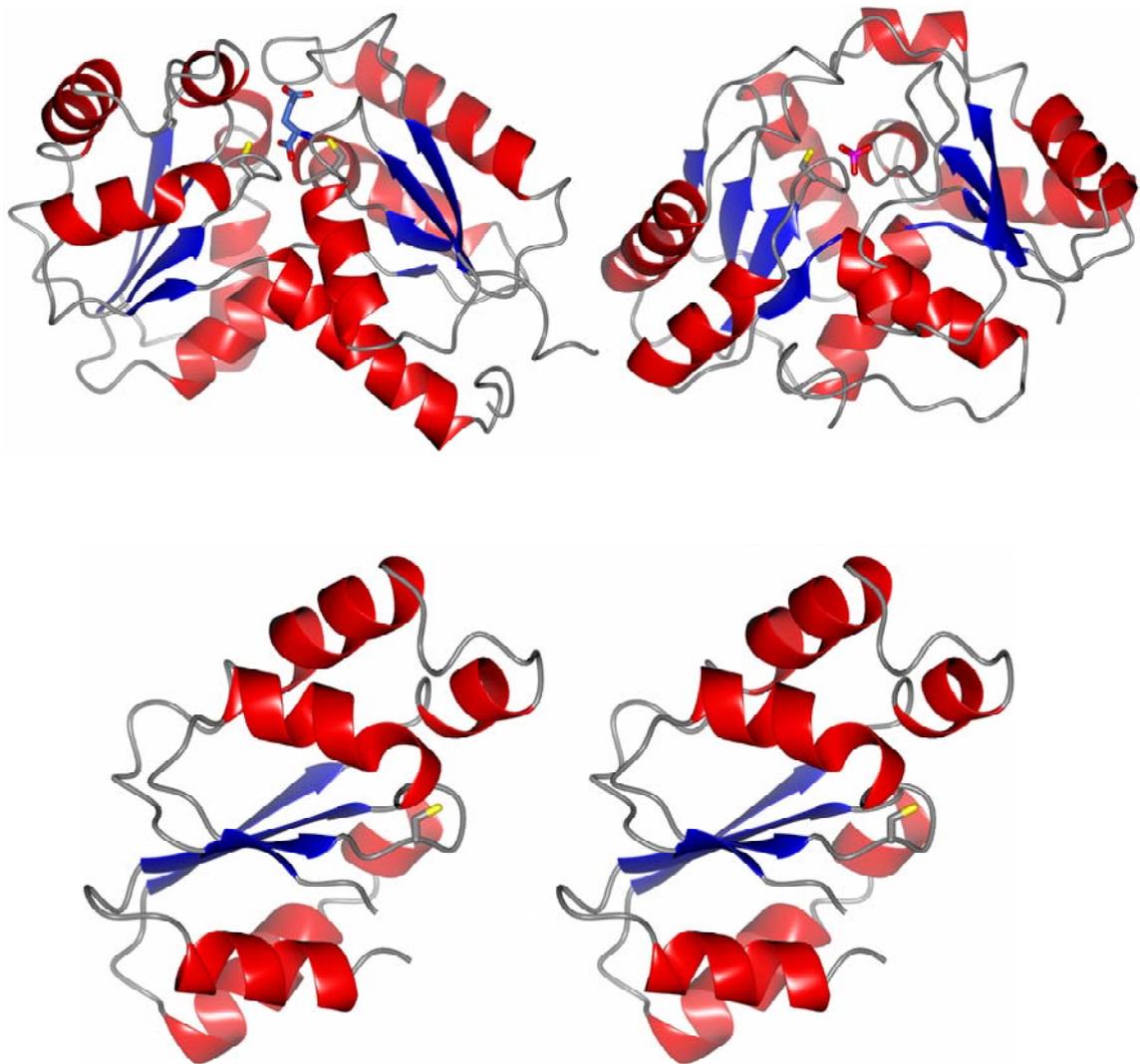


Figure 1.22. Structure of the Asp/Glu racemase superfamily members. The secondary structure is shown for α -helices (red) and β -sheets (blue). Top left: EfGluR with bound D-glutamate (light blue) and two cysteines (PDB: 2vvt). Top right: BbAMD with bound PO_4 (magenta) and one cysteine (PDB: 3dg9). Bottom: Stereoview of the typical Rossmann fold domain of the superfamily corresponding to domain 2 in EfGluR including one cysteine on the loop following the third β -strand.

1.3.1 Glutamate Racemase

The best characterised enzyme in the Asp/Glu racemase superfamily is GluR that catalyses the racemisation of both L- and D-glutamate (Fig 1.23). The enzyme provides D-glutamate necessary for the peptidoglycan cell wall synthesis and is therefore thought to be an interesting target for antibiotics. GluR activity was discovered in lactic acid bacteria as early as 1952 when it was also shown that α -

ketoglutarate was not transformed into glutamate and was, therefore, not an intermediate of the racemisation (53).

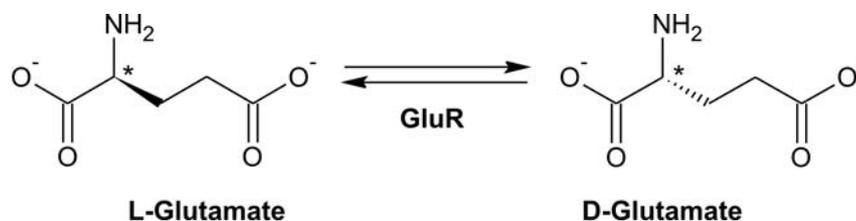


Figure 1.23. Glutamate racemase activity.

Over the years there was some confusion over the cofactor requirement of GluR (54-57). In other amino acid racemases such as alanine, serine and arginine racemases, PLP was found to act as a cofactor forming a Schiff base between the aldehyde and the substrate amine. Later studies on heterologously expressed *Lactobacillus fermentum* GluR (LfGluR, UniProt: Q03469) showed that the typical optical absorption peak for a Schiff base of the covalently attached PLP does not exist and that adding PLP was not necessary for activity. A Pyruvoyl group observed in some decarboxylases and reductases could not be detected by acid or base hydrolysis or specific inhibition by sodium borohydride, hydroxylamine or phenylhydrazine. The requirement of metal cofactors was excluded as the presence of EDTA did not have an effect on the activity of GluR. It was thus concluded that GluR is a cofactor-independent racemase (47).

The mechanism of GluR was studied extensively with isotope labelling. After incubating [^{18}O]glutamate with enzyme in water, no loss of ^{18}O from could be observed by MS, indicating that there was no acyl intermediate formation during the reaction. Incubating the unlabelled glutamate with enzyme in [^2H]water and following the reaction with ^1H -NMR showed that the peak corresponding to the C2 hydrogen disappeared. This demonstrated that racemisation occurs via general acid/base catalysis involving deprotonation and reprotonation at the C2 position. By comparing the k_{cat} of GluR when incubated with pure enantiomers of [$2\text{-}^2\text{H}$]glutamate or [$2\text{-}^1\text{H}$]glutamate, a primary kinetic isotope effect could be observed indicating that deprotonation rather than substrate binding was the rate limiting step. Furthermore, GluR was incubated with either pure D-glutamate or L-glutamate in [^2H]water and the product was analysed by chiral GC/MS. During the initial reactions [^2H] was only incorporated into the product

but not into the substrate showing that the GluR operates with a two-base rather than with a one-base mechanism. Oversaturation could not be observed for GluR indicating that the interconversion of one enzyme species to the other species with opposite cysteines protonated was not rate limiting (Fig. 1.24) (58).

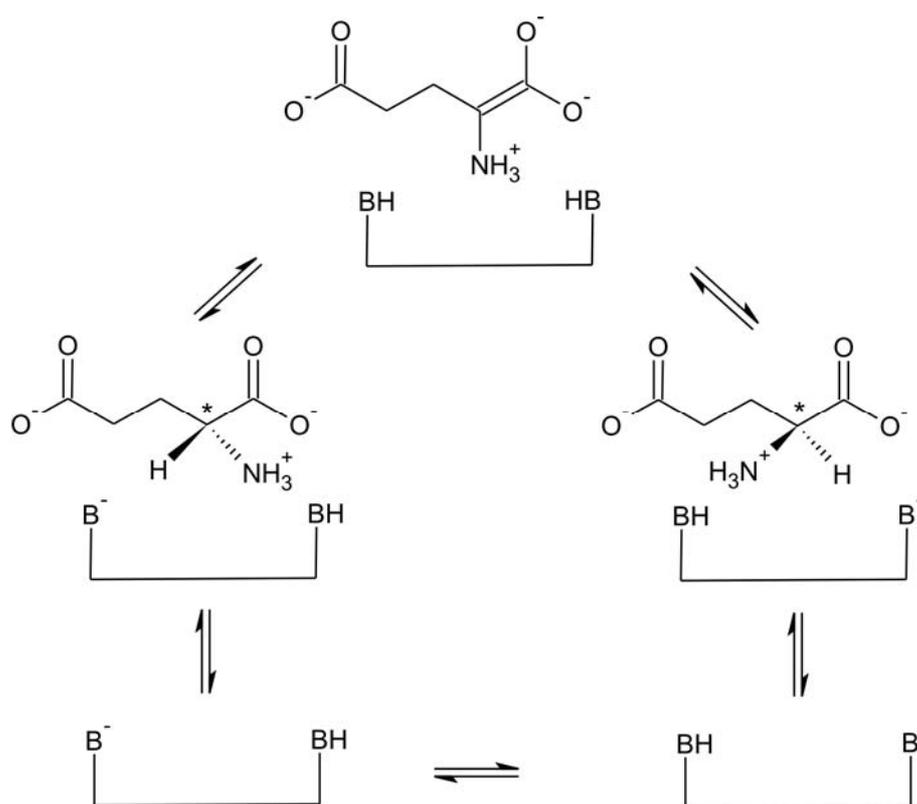


Figure 1.24. Mechanism of GluR. Deprotonation and reprotonation of D-glutamate proceeding through an enediolate intermediate (59).

GluR was found to be stabilised by thiols, inactivated by thiol alkylating agents and two cysteine residues were conserved in all homologues to LfGluR. Mutating the two conserved cysteines to alanines or serines abolished or significantly reduced the activity of the enzyme, indicating that they are indeed the catalytic bases involved in the deprotonation and reprotonation. A series of observations allowed the identification of which of the cysteines deprotonates L-glutamate and which deprotonates D-glutamate. The single cysteine to alanine mutants led to the stereospecific deprotonation of threo-3-chloroglutamate and the single cysteine to serine mutants led to the deprotonation of N-hydroxyglutamate. Moreover, it was shown that the k_{cat}/K_M isotope effect was reversed depending on which cysteine was mutated, indicating that

the activating energy barrier for the serines was higher when they were involved in deprotonation (47,60,61).

A series of structures of GluRs has been determined to date. The first structure from the thermophile *Aquifex pyrophilus* (ApGluR, PDB: 1b73, 1b74) had the weak inhibitor D-glutamine bound ($K_i = 50$ mM) (62) whereas the second structure from *B. subtilis* (BsGluR, PDB: 1zuw) (63) had D-glutamate bound in the active site. Both structures contained the catalytic cysteines in positions flanking the substrate with correct distances. However, the orientation of D-glutamine in ApGluR was upside down exposing C5 to the thiols of the cysteines instead of C2. In addition, the domain 2 of ApGluR was moved by maximally 13 Å largely exposing the bound D-glutamine to the solvent. This probably arises from the fact that ApGluR was soaked in D-glutamine in contrast to the co-crystallisation of BsGluR with D-glutamate. ApGluR seems to resemble the empty form of GluR and BsGluR the substrate-bound form (Fig. 1.25, 1.26) (63).

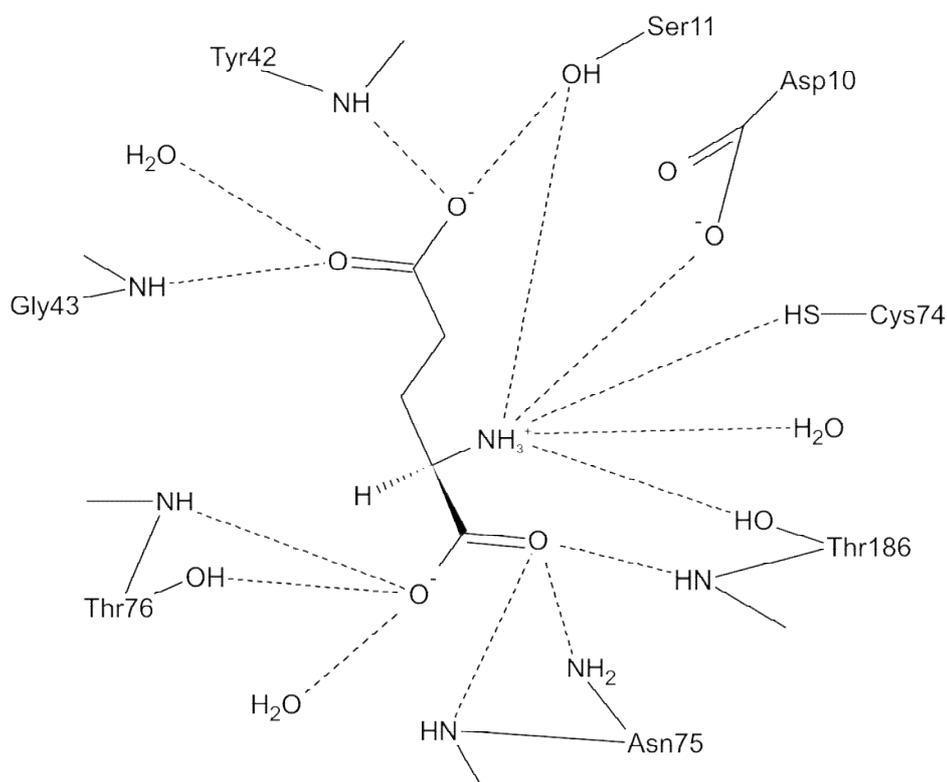


Figure 1.25. Residues involved in binding of D-glutamate in BsGluR. The dotted lines indicate hydrogen bonds that are thought to form at different stages during the reaction.

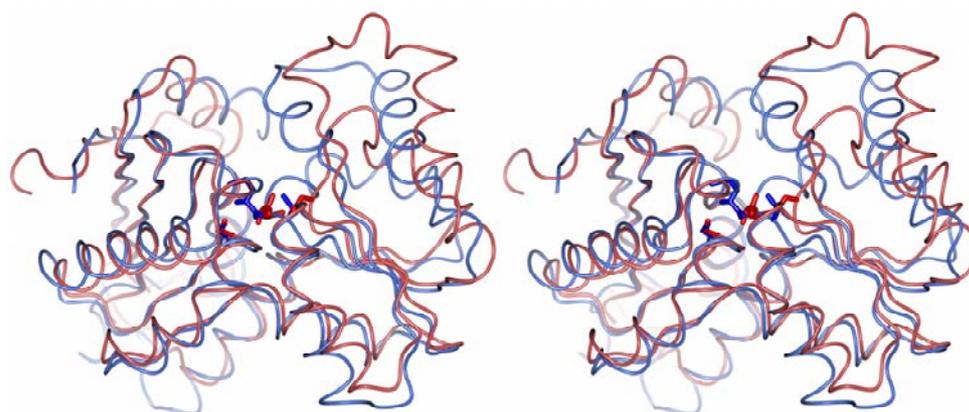


Figure 1.26. Overlay of GluR structures. Stereoview of BsGluR (light blue) and ApGluR (light red) overlaid on domain 1 (left halves of the proteins). D-glutamate and catalytic cysteines of BsGluR (blue) and D-glutamine and catalytic cysteines of ApGluR (red) are shown as stick model.

The findings presented above give a cofactor-independent mechanism for GluR but do not explain how the reaction can be catalysed against the odds of a high pK_a of both the C2 proton and the thiol of the cysteines as well as the low stability of carbanion intermediate with an adjacent negatively charged carboxylate. It was suggested that the cysteines do not have to be deprotonated for activity but that other bases would assist in the deprotonation once the substrate has been bound. Mutations of a conserved aspartate and histidine to asparagine in GluR were found to have significantly lower k_{cat} with an unchanged K_M (64). The residues Asp10 and His187 were adjacent to the catalytic Cys74 and Cys185 respectively (63). In a molecular dynamics approach using D-glutamate in BsGluR this scenario was simulated coming to the conclusion that a deprotonating cascade starting from Asp10 to Cys74, over to C2 of D-Glu and finally to His187 represents a probable pathway (Fig. 1.27). This mechanism would act in a concerted way, going through a single high energy peak, but highly asynchronous as both cysteines are fully protonated in the intermediate stage. It was found that the ammonium group was assisting the reaction by stabilising the deprotonated residue at each stage of the reaction (65). In another molecular dynamic approach based on BsGluR and D-Glu, the behaviour of the unstable carbanion intermediate was simulated in the active site. The conclusion was that the negative charge is destabilised by a range of interactions involving three waters and Ser11, Asn75, Thr76, Thr186 and His187 (66). This is supported by the observation of a dioxanion hole in the homologous BbAMD where the residues equivalent of Asn75, Thr76, Thr186 and a water form six hydrogen bonds with a ligand modelled on the

basis of a benzylphosphonate (67). The dipoles of the N-termini of helices 3 and 8 could provide additional stabilisation of the deprotonated cysteines (63).

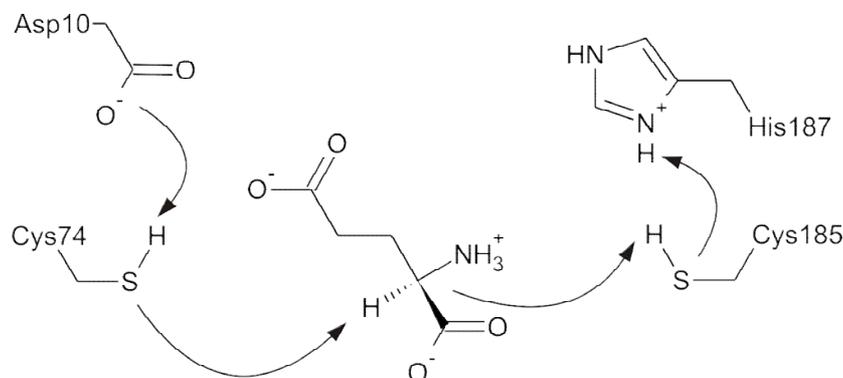


Figure 1.27. Catalytic cascade in BsGluR.

Proline racemase (EC 5.1.1.4) is another well studied cofactor-independent racemase. Interestingly, it employs almost the same mechanism to GluR but adopts a completely different fold. Proline racemase is related to the diaminopimelate epimerase (EC 5.1.1.7). As in GluR, these enzymes recruit a cysteine dyad that deprotonates and reprotonates the bound substrate. In the structure of proline racemase the same aspartate and histidine residues were found adjacent to the catalytic cysteines. It has, therefore, been hypothesised that they would assist in the deprotonation and reprotonation of the cysteine residues (68). However, in a recent molecular dynamics simulation study it was argued that they were tightly bound to the enzyme framework and could not move towards the cysteines. Therefore, the cysteines would have to be deprotonated by water or the substrate rather than by neighbouring residues. Stabilisation of the intermediate in these enzymes would be made by dioxyanion hole formed by three hydrogen bond donors per carboxylate oxygen, equivalent to the proposition made for AMDs (48,68).

1.3.2 Aspartate Racemase

Discoveries of AspR activities were always closely linked to GluR activities. The activity was discovered in lactic acid bacteria and subsequently shown to be an activity of a different enzyme than GluR (Fig. 1.28) (54,56). The enzyme was later purified from *Streptococcus faecalis* and the first AspR gene sequence was cloned from *Enterococcus faecium* (69,70). To date there are three solved structures available from

two genes, both originating from *Pyrococcus horikoshii* OT3 (PhAspR). One structure from the PhAspR C82A mutant has citrate bound in its active site showing similar binding of one carboxyl group to D-Glu in GluRs (UniProt: O58403, PDB: 2dx7) (71).

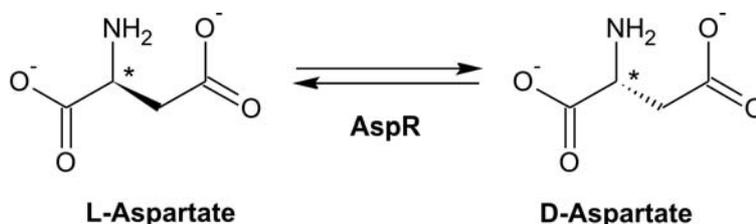


Figure 1.28. Aspartate racemase activity.

The sequence alignment, the protein fold conservation and the substrate similarity clearly indicate that AspR is a member of the superfamily. By analogy the mechanism is thought to be the same as for GluR with deprotonation by one cysteine thiol, stabilisation of the intermediate by hydrogen bonds and reprotonation by the opposite cysteine thiol. An alternative mechanism was also proposed where the same cysteine would deprotonate both aspartate enantiomers, which in turn would be reprotonated by the same cysteine (72).

1.3.3 Hydantoin Racemase

The first HydR activity was discovered in *Pseudomonas* sp. NS671 and found to be on an operon including a hydantoin hydrolase and a carbamoyl hydrolase involved in the degradation of hydantoins into amino acids (Fig. 1.29) (49). This sequence of reactions is also used industrially to produce enantiopure D- and L-amino acids (see first section). Several HydR sequences from various bacterial strains and one structure of *P. horikoshii* OT3 have been determined (UniProt: O58781, PDB: 2eq5). Both the sequence homology and the fold indicate that HydRs are also members of the superfamily.

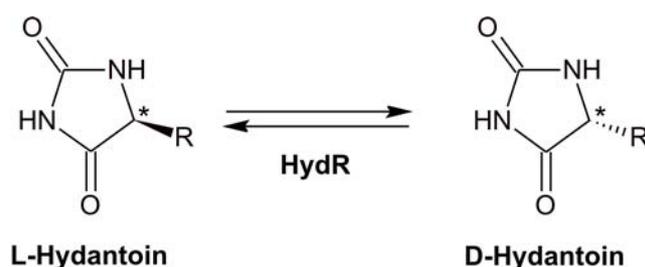


Figure 1.29. Hydantoin racemase activity. R = methyl, ethyl, isopropyl, isobutyl, benzyl, methylthioethyl, phenyl, hydroxyphenyl, 3-methyl-indol.

It was shown that ^2H from the solvent is incorporated specifically at the 5-position confirming the conservation of the deprotonation/deprotonation mechanism found in GluR. By analogy to GluR it was assumed that the intermediate is an enol (73). Through single mutations of the catalytic cysteines to alanine the activity of HydR was completely abolished. Single mutations to serines left some residual activity (100 to 500-fold reduced). As in GluRs, the serines in HydRs have reduced but detectable deprotonation capacity compared to cysteines (74). Binding studies using calorimetry and fluorescence showed that the alanine mutants were able to bind only one of the two enantiomers of the hydantoin suggesting that the proton rather than the side chain of the substrate is determining the binding specificity (75).

1.3.4 Arylmalonate Decarboxylase

A new cofactor-independent decarboxylase was discovered through selective enrichments of soil cultures growing on PM as their sole carbon source. AMD from *B. bronchiseptica* KU 1201 (BbAMD) was specifically and completely decarboxylating MPM yielding pure (*R*)-PP (Fig. 1.30, e.e. > 99%, conversion > 99%) (51,76). The physiological role of AMD was not identified but it was thought to be implicated in the catabolism of special nutrients.

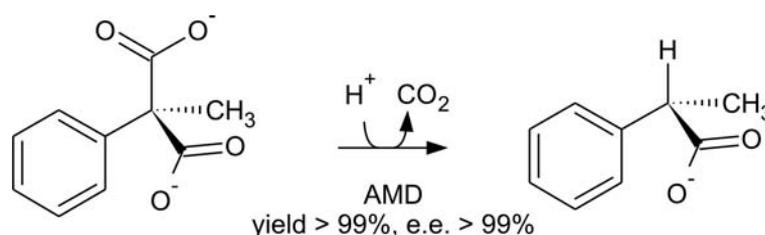


Figure 1.30. Typical reaction of BbAMD.

The enzyme was purified from bacterial cell extract and found to be a monomeric 25 kDa protein with an isoelectric point of 4.7. It had a pH optimum of 8.5 and stable between pH 6.5 and 8.5. It was found to be stable up to 30 °C but inactivated rapidly at 50 °C and could be stored for months at 4 °C. The enzyme was found to be independent of cofactors such as ATP, ADP, CoA or metal ions. BbAMD was effectively inactivated by thiol reagents such as iodoacetate and by Pb^{2+} , Sn^{2+} , Hg^{2+} ,

Hg⁺ and Ag²⁺ metal ions. It was therefore concluded that cysteines are the active site residues (51).

The gene of BbAMD was identified using a genomic fragment library cloned into *E. coli* cells that were selected for alkalinisation of the medium using the pH indicator BTB changing its colour from yellow to blue. The gene was sequenced and found to be 720 bp long (UniProt Q05115). BbAMD was later found to be member of the Asp/Glu racemase superfamily (77-79). Three more sequences of clearly active AMD have been discovered by additional soil culture enrichment and genome mining (48,80,81).

Apart from the achiral substrate PM and the pro-chiral substrate MPM a vast range of other malonates was tested on BbAMD. One substituent had to be an aryl or an alkenyl electron withdrawing group. The other substituent had to be a small hydrogen, methyl, hydroxyl, amine, fluorine or chlorine. The best substrate is PM that yielded achiral PA, whereas interesting and good substrates were MPM, 2-methyl-2-naphthylmalonate, 2-methyl-2-thienylmalonate. It was found that even rather large and complex electron withdrawing groups could be accepted as it was possible to produce flurbiprofen from a dicarboxylic precursor with BbAMD (Fig. 1.31). The more electron withdrawing the group was, however, the more active was the enzyme. The aromatic moiety or the double bond had to be connected directly to the α -carbon as any longer linker abolished any activity. The small groups did not have any electrostatic requirements but had to be small and even 2-ethyl-2-phenylmalonate was not converted by the enzyme. The small substituent was more active the smaller it was (51,67,76,79,82-87).

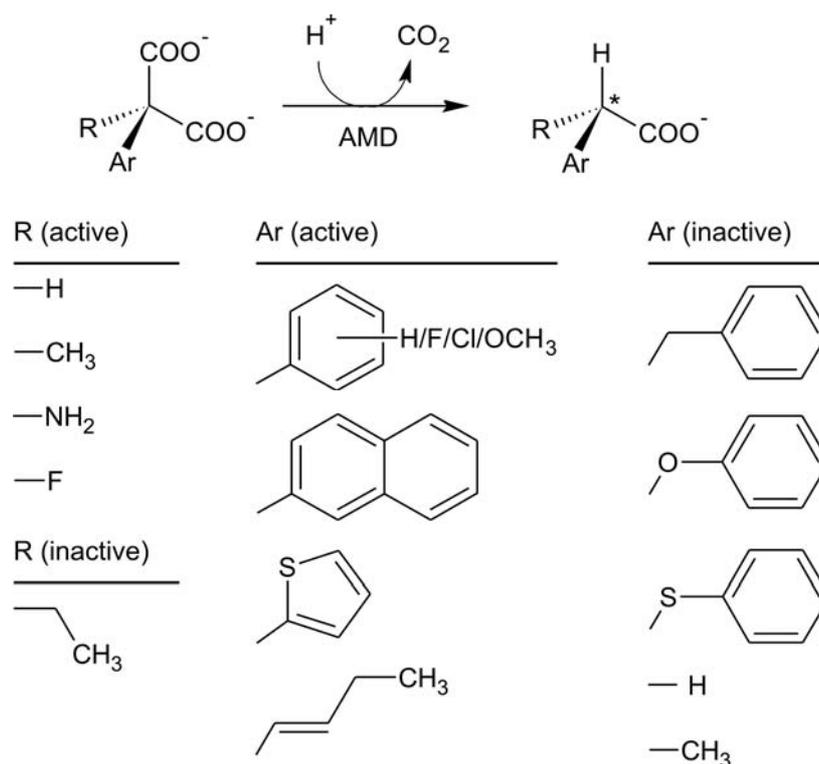


Figure 1.31. BbAMD activity and selection of representative substrates.

The affiliation of BbAMD to the Asp/Glu racemase superfamily was confirmed when the structure was solved and showed that the typical two pseudosymmetric domain structure was retained. As was already seen in sequence alignments, BbAMD only had one catalytic cysteine (Cys188) on domain 2 whereas the other catalytic cysteine on domain 1 is replaced by a glycine (Gly74). There are currently five structural coordinates published in the structure database (PDB: 3dg9, 3ip8, 2vlb, 3eis, 3dtv). Two of them have a phosphate or a benzylphosphonate bound in the active site that represents a model for the decarboxylated intermediate bound in the active site. Both structures are high resolution structures and show the enzyme in a closed form, assumed to be the substrate-bound form. In this form two loops, one from each domain, cover the ligand in the active site completely. The other three structures are lower resolution and show different molecules with different conformations especially in these covering loops probably indicating the flexibility of the unliganded open form (Fig. 1.32) (48,67,88-90).

The benzylphosphonate bound high resolution structure shows distinct pockets for the ligands. The phosphate moiety is bound with two of its oxygen atoms to six hydrogen bonding partners by the hydroxyls of residues Thr75, Ser76 and Tyr126 as well as the

backbone amides of residues Thr75, Ser76 and Gly189. This structure was called the dioxyanion hole. The benzyl moiety is bound in a large aryl binding pocket formed by hydrophobic residues and backbones of Pro14, Val43, Val156, Met159, Gly189 and Gly190. A small binding pocket is formed by Leu40, Val43 and Tyr48, most probably for binding the small substituent and the leaving carboxyl (Fig. 1.32) (67).

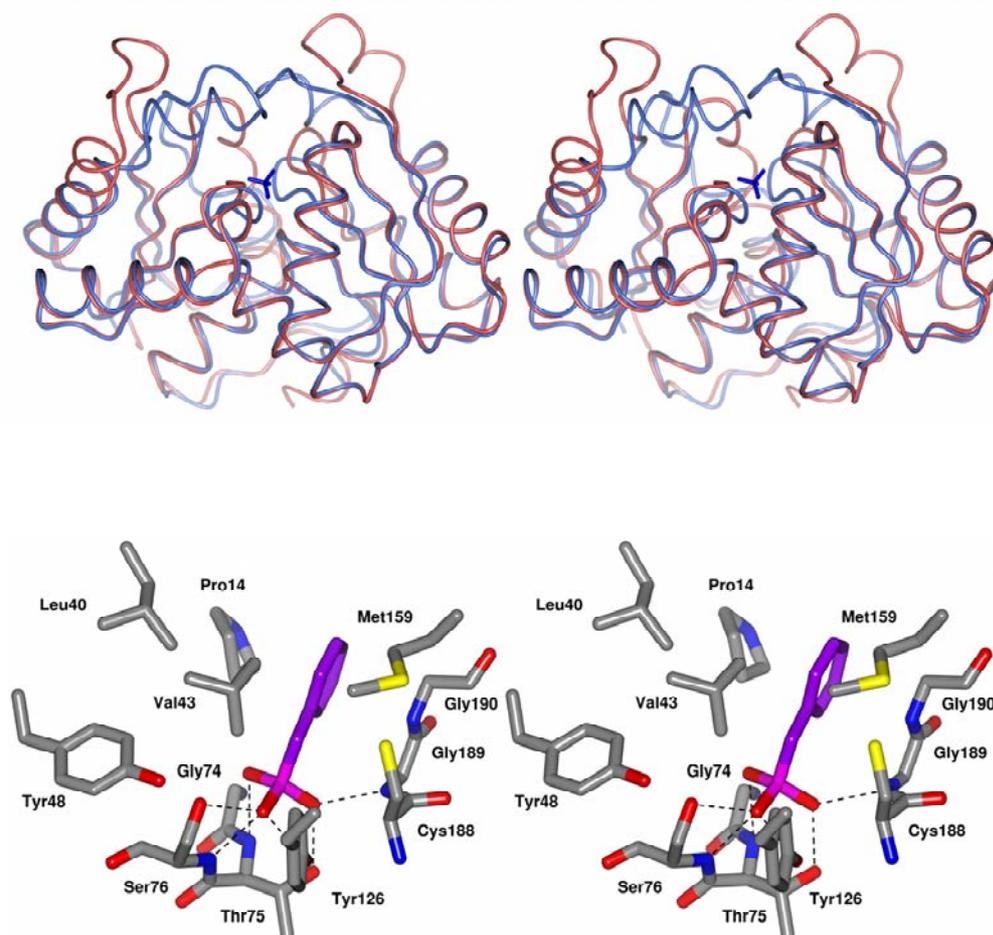


Figure 1.32. Structure of BbAMD. Top: Stereoview of the open, non-liganded structure (light red, PDB: 2vlb) is superposed to the closed, phosphate bound form (light blue, PDB: 3dg9). Bottom: Stereoview of the active site with bound benzylphosphonate (purple, PDB: 3ip8). Six hydrogen bonds (dashed lines) are formed between the active site and the phosphate moiety of benzylphosphonate.

A series of mutagenesis experiments has been produced to confirm the catalytic residue and the stereochemistry of the decarboxylation by BbAMD. The C188S mutant was nearly 600 times less active than wild type and all other cysteine to serine mutations had much smaller effects on the activity. Interestingly, the double mutant

G74C/C188S, although even less active than C188S, showed inversion of the enantioselectivity yielding (*S*)-PP instead of the (*R*)-PP, which was also confirmed for other pro-chiral substrates of BbAMD. Furthermore, the G74C mutant not only lost the stereoselectivity of MPM decarboxylation but was able to racemise (*R*)- and (*S*)-PP, the products of the decarboxylation, although at very low rates. The rates of the enantiospecificity-reversed mutant could be slightly improved 10-fold by the S36N mutation found through a random mutagenesis approach. The improvement had to come from a distant structural effect as Ser36 is located at the solvent exposed part of BbAMD (77,91-95).

The stereospecificity of the decarboxylation and the protonation was determined by decarboxylating MPM carboxyls labelled with ^{13}C or ^{18}O on either the pro-*R* or the pro-*S* carboxyl. NMR was used to analyse ^{13}C and MS was used to analyse ^{18}O . Both cases confirmed that the pro-*R* carboxyl was removed specifically. The product in all cases was (*R*)-PP, indicating an inversion mechanism contrasting with the malonate decarboxylases described in the previous section. The S36N/G74C/C188S mutant was also decarboxylating the pro-*R* carboxyl but showed retention of configuration. The decarboxylation in H_2^{18}O solvent did not show incorporation of ^{18}O into the product as shown by MS. This suggested a mechanism with no thioester intermediate (48,91,96).

In analogy to the related racemases the mechanism is thought to occur through a planar enediolate intermediate that is stereospecifically protonated by a Cys188 from the *si*-face. The doubly negatively charged enediolate intermediate has to be stabilised by the six hydrogen bonds offered by the dioxanion hole (Thr75, Ser76, Tyr126 and Gly189). The pro-*R* carboxyl is bound and destabilised in a hydrophobic pocket made of Leu40, Val43 and Tyr48. By analogy to the racemases the decarboxylation and protonation probably happen in a concerted and asynchronous manner without a clear transition state. This mechanism was further confirmed by the inhibition by 2-bromophenylacetate and benzylphosphonate (Fig. 1.33) (48,67,91).

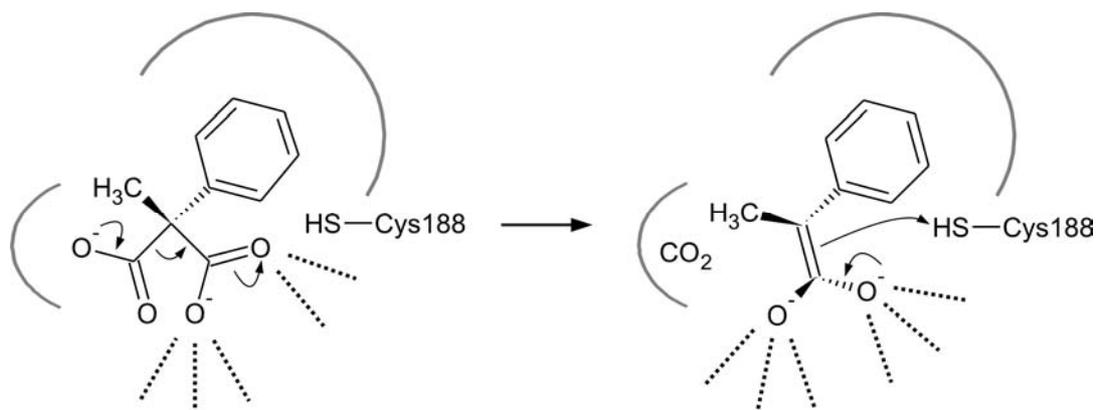


Figure 1.33. Mechanism of BbAMD with MPM. The retained carboxyl is bound in the dioxanion hole with up to six hydrogen bonds acting as an electron sink (dotted lines). The leaving carboxyl is destabilised in the hydrophobic small carboxyl binding pocket (small grey arc). The aryl group is bound in the hydrophobic large aryl binding pocket (large grey arc). Cys188 protonates carbon C2.

The mechanism of AMD could also be exploited to achieve a cascading aldol reaction. BbAMD is able to catalyse the decarboxylation of 2-(2-oxoethyl)phenylmalonate and the subsequently formed enediolate intermediate is able to undergo a cascading aldol reaction to form a 2-hydroxy-2,3-dihydro-1-indenecarboxylate (Fig. 1.34) (86).

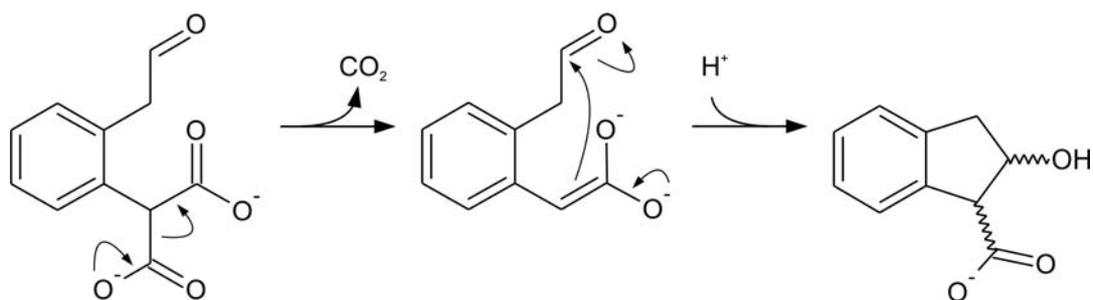


Figure 1.34. Cascading aldol reaction triggered by BbAMD.

1.3.5 Maleate *cis-trans* Isomerase

Maleate isomerase (MI) was discovered in bacteria growing on nicotinic acid. It was involved in the last step of degradation of nicotinic acid after ring cleavage producing a *cis* double bonded compound. The last compound being maleate was isomerised into its *trans* form fumarate that was fed into the citric acid cycle (Fig. 1.35) (97). Since then, a series of MIs from other organisms have been discovered, purified and

characterised (98-102). Only three MI genes have so far been characterised (52,103,104). Sequence alignments and the presence of two conserved cysteine residues clearly indicate that MIs are Asp/Glu racemase superfamily members.

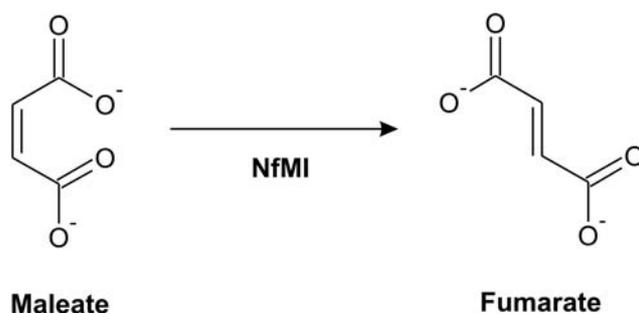


Figure 1.35. Maleate *cis-trans* isomerase reaction.

MIs represent a peculiarity in the Asp/Glu racemase superfamily as they require a fundamentally different reaction mechanism to the simple stereospecific deprotonation/reprotonation. The mechanism must proceed through an intermediate that is able to rotate around the reduced double bond. This requires both another function for the catalytic cysteines as well as flexibility in the active site. Although no protein structure of a MI has been determined to date, the importance of the catalytic cysteine dyad as in the racemases has been confirmed. Direct evidence was given by the inactivation through mutation to alanines (104). Indirect evidence was given by the sensitivity of the enzymes to thiol reagents and heavy metals (100,101). Also the absence of cofactor requirement has been confirmed for MIs (52,101). The enzyme was found to be completely specific for maleate (98,100,101). Despite this restriction, MIs could be useful to industry as fumarate is widely applied in industry for the production of L-aspartate or L-malate (see section above, (105-107).

Maleate and fumarate differ only by their double bond configuration (Fig. 1.35). They are both planar as a result of conjugation between their two carbonyl groups and the central double bond. The difference in configuration has major effects on the properties of the compounds. Maleic acid (*cis* configuration) forms an intramolecular hydrogen bond that fumaric acid (*trans* configuration) does not. Therefore maleic acid is more soluble and more acidic than fumaric acid. Fumarate is also more stable than maleate; however, the rotation around the double bond requires a large activation energy resulting in maleate being relatively stable in physiological conditions. The

isomerisation only occurs at temperatures above 200 °C or through irradiation with UV-light. The *cis-trans* isomerisation can be catalysed by reducing the double bond to a single bond allowing free rotation. This has been achieved with metals, radical addition and nucleophilic attack by thiols and acids (105,108-110).

Some *cis-trans* isomerases involve double bond migration such as linolenate isomerase (EC 5.2.1.5). MI operates through a mechanism without double bond migration as do maleylacetoacetate isomerase (EC 5.2.1.2) and maleylpyruvate isomerase (EC 5.2.1.4) (110). These maleyl decarboxylases are dependent on the coenzyme glutathione that is thought to attack double bond at the 2-carbon (111). Analogously, the catalytic cysteines of MI are thought to act like glutathione, but alternative mechanisms cannot be excluded (104,110).

1.4 Aims

Although many enzymes are already applied in industry, there is still a great need for optimisation and extension of the enzymatic tool kit. Carboxylations/decarboxylations, racemisations and isomerisations are reactions without appropriate biocatalysts that still require development in order to become useful for particular applications (112). Fortunately, enzymes are very flexible catalysts because of the huge number of potential variants and because their sequences can be adapted by directed evolution approaches. With the knowledge of their three dimensional structures more rational approaches can also be applied (113).

The biocatalytic route to the production of chiral alkylalkanoates as chemical intermediates is highly attractive. The decarboxylation of alkylmalonates analogous to the arylmalonate activity of BbAMD seems to be a promising approach. The present work aimed to exploit the potential of the Asp/Glu racemase enzyme superfamily to produce a suitable biocatalyst.

In Chapter 3 the sequence diversity of the Asp/Glu racemase enzyme superfamily was mined in order to identify new AMDs. A range of BbAMD sequence homologues was identified, cloned, expressed and assayed for aryl- and alkylmalonate decarboxylation activity. As no decarboxylation activity could be observed for the tested homologues some homologues were analysed in more depth. The presence of GluR, AspR, MI and PP racemase activity was tested in three homologues. Site directed mutagenesis was applied to change the active site sequence of one homologue to resemble more the

BbAMD sequence. In an alternative approach the microbial diversity in the soil was mined in an attempt to isolate bacterial strains capable of decarboxylating alkylmalonates. Soil samples were selectively enriched in organisms capable of growing on alkylmalonates. Substrate and product quantities in pure cultures were analysed (see Appendix D).

In Chapter 4 BbAMD homologue NfMI was identified to be a MI. MIs are the most closely related enzymes to AMDs in the Asp/Glu racemase superfamily but have a fundamentally different activity. As not much is known about MIs the NfMI was characterised biochemically. A range of substrate analogues were tested in order to find a competitive inhibitor to NfMI. Also the role of the catalytic cysteines was investigated using site directed mutagenesis. In Chapter 5 the structure of NfMI was determined by X-ray crystallography. Based on these findings several hypotheses for a reaction mechanism are discussed which reveal insights into the catalytic plasticity of the Asp/Glu racemase superfamily.

Chapter 2

General Methods

2.1 Cloning

2.1.1 Strategy

For the rapid and partially automated cloning of bacterial genes from gDNA samples the ligation independent cloning (LIC) approach was chosen. LIC uses simple annealing of complementary single strand overhangs of the insert and plasmid, directly followed by the transformation; circumventing an inefficient ligation process and the design of restriction sites.

For the production of long overhangs, the exonuclease proofreading activity of the DNA polymerase was used adding only one type of dNTP to the reaction mix. The exonuclease removes nucleotides from the strand until it reaches a specific nucleotide in the sequence where the polymerase activity predominates. The length of the overhangs was previously determined through the design of the cloning site and is matched by the complementary sequence of specific PCR primer overhangs (114).

The plasmid used was pET-YSBLIC3C that is both a cloning and expression vector based on IPTG inducible pET-28a(+) plasmid (Novagen, Appendix B). The vector contains an N-terminal His-tag to facilitate purification and it is cleavable by the HRV 3C-protease to improve crystallisation if required (Fig. 2.1).

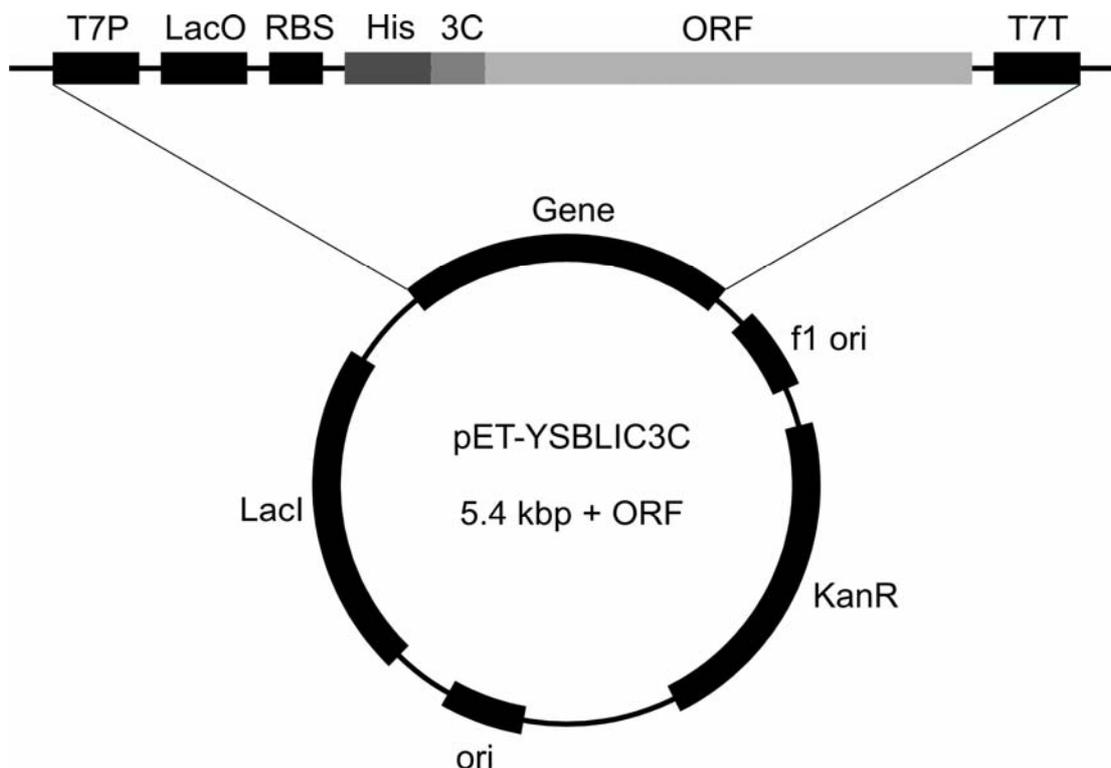


Figure 2.1. LIC vector containing a gene of interest. Based on pET-28a(+) plasmid with an altered cloning site containing the open reading frame (ORF) including a 6x His-tag (His) cleavable at the HRV 3C protease site (3C). The plasmid further contains a T7 promoter (T7P) and terminator (T7T), a Lac operator (LacO) and repressor gene (LacI) for IPTG induction, a ribosome binding site (RBS), a kanamycin resistance gene (KanR) and two origins of replication (ori, f1 ori).

2.1.2 Construct Preparation

Inserts were amplified using 1 U KOD Hot Start DNA Polymerase (Novagen), 1x PCR buffer, 1 mM MgSO₄, 2 mM of each dNTP, 4 μM of each forward and backward primer (Table 2.1, Sigma Genosys), 0.5 ng μL⁻¹ DNA and was completed to 50 μL total volume with water. The temperature cycles were one initial melting step at 94 °C for 120 s followed by 35 times melting at 94 °C for 30 s, annealing at 65 °C for 30 s and elongation at 72 °C for 60 s ending with one final elongation step at 72 °C for 180 s performed on a Px2 Thermal Cycler (Thermo Scientific, Fig. 2.2).

The pET-YSBLIC3C plasmid was cleaved with 20 U of BseR I (NEB), 1x restriction buffer 2 and 5 μg DNA completed to 100 μL total volume with water. The mix was incubated at 37 °C for 110 min, separated from uncleaved plasmid by agarose gel

electrophoresis and the target band was purified using the Wizard SV Gel Clean-Up Kit (Promega).

Single strand overhangs were produced with 0.2 pmol of the PCR product together with 2.5 mM dATP or 0.4 pmol cleaved plasmid together with 2.5 mM dTTP. Both insert and plasmid were mixed with 2 U T4 DNA Polymerase (Novagen), 1x polymerase buffer and 5 mM DTT, then incubated at 22 °C for 30 min and inactivated at 75 °C for 20 min. The plasmid reaction was cleaned using the QIAquick PCR Purification Kit (QIAGEN) whereas the insert was used without cleaning.

Annealing was performed by mixing 2 μ L insert reaction and 1 μ L cleaned plasmid, incubating at room temperature for 10 min. Then 1 μ L 25 mM EDTA was added and incubated at room temperature for another 10 min. *Escherichia coli* NovaBlue Singles were transformed with 2 μ L of the annealing reaction.

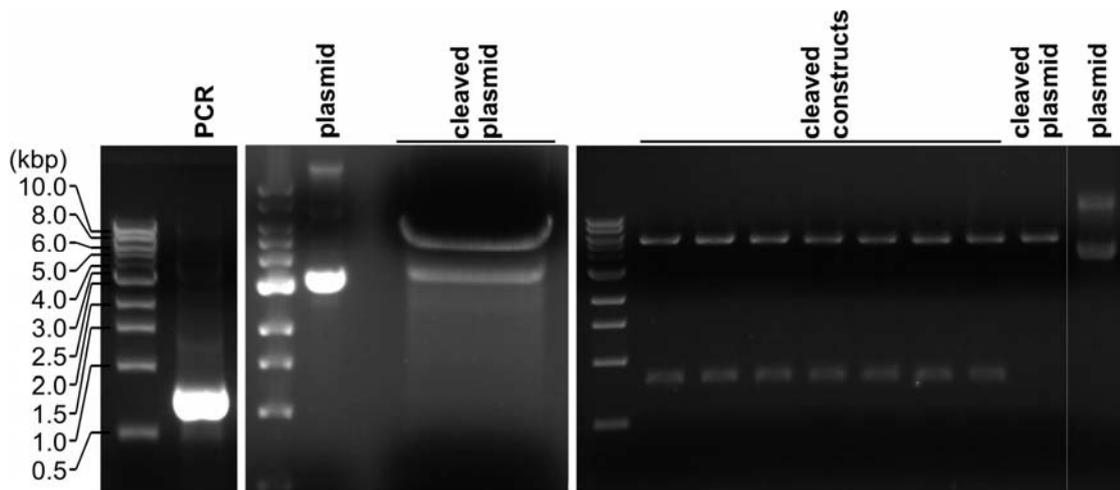


Figure 2.2. LIC procedure. Left: PCR product of BbAMD including LIC overhangs. Middle: Gel purification of cleaved LIC plasmid. Right: Seven clones tested for the presence of the insert.

2.1.3 Plasmid Transformations

E. coli strains used in this study were NovaBlue Singles for cloning, BL21(DE3) for expression of codon optimised genes, Rosetta 2(DE3) for expression of genes containing rare codons. Rosetta 2(DE3) cells contain a pRARE2 plasmid encoding rare *E. coli* tRNA genes and a chloramphenicol resistance gene.

For the production of competent cells, aseptic technique was used throughout the process. 100 mL LB were inoculated with a freshly grown colony of the appropriate strain. 34 $\mu\text{g mL}^{-1}$ chloramphenicol were added if Rosetta 2(DE3) were used. The culture was incubated at 37 °C until an optical density at 600 nm of 0.4-0.5 was reached. From that moment the cells were kept at 0 °C, spun down at 2,000 $\times g$ for 10 min, resuspended in 40 mL sterile 50 mM MgCl_2 , 20 mM CaCl_2 , incubated for 30 min, spun down again and finally resuspended in 4 mL sterile 80 mM CaCl_2 , 20% (v/v) glycerol. The cell suspension was then aliquoted in 50 μL portions, snap frozen in liquid nitrogen and stored at -80 °C.

A microcentrifuge tube containing 50 μL competent cells was thawed at 0 °C and 100 ng plasmid DNA was added. The mix was incubated 5 min, heat shocked at 42 °C for 2 min and cooled down at 0 °C for 5 min. 600 μL LB was added to the suspension and incubated at 37 °C shaking for 1 h. Then 10 μL were spread on LB-agar plates containing 100 $\mu\text{g mL}^{-1}$ kanamycin and in the case of Rosetta 2(DE3) cells additionally 34 $\mu\text{g mL}^{-1}$ chloramphenicol.

2.1.4 Plasmid Purification and Verification

5 mL LB containing the appropriate antibiotic were inoculated with a freshly grown colony, grown at 37 °C shaking at 200 rpm for 16 h and then extracted using the QIAprep Spin Miniprep Kit (QIAGEN). The purified quantity was determined by measuring the absorbance at 260 nm (0.1 units corresponding to 5 $\mu\text{g mL}^{-1}$ DNA).

The presence of the insert with the correct length was verified by double digesting 100 ng plasmid DNA with 3 U of each Nco I and Nde I in 1x restriction buffer 4 and 1 mg mL^{-1} BSA (NEB). The constructs were sequenced by an in-house sequencing service, using Sanger sequencing on an ABI 3130 (Applied Biosystems) with T7P and T7T primers (Table 2.1).

Table 2.1. List of primers.

Name	Sequence	Target
Cloning primers for BbAMD and homologues		
faf1f	CCAGGGACCAGCAATGCAGCAGGCGAGCACCCC	BbAMD
faf2r	GAGGAGAAGGCGCGTTAGCTTTTCATCAAACAGACGGCCATAGC	BbAMD
RseDf	CCAGGGACCAGCAATGACGAGCAGCGCCGACCCCCAC	RseD
RseDr	GAGGAGAAGGCGCGTCAGTCAGAGCCATGACTGTGGCCTTTC	RseD
RseCf	CCAGGGACCAGCAATGTTGGAACCTGAACATTCGGAATTCGAGG	RseC
RseCr	GAGGAGAAGGCGCGTCAGAAGACGTTGCGGAGCCACTTGC	RseC
Nfef	CCAGGGACCAGCAATGGGCATCCGCCGCATCGGGTTG	NfMI
Nfer	GAGGAGAAGGCGCGTCAGCTCGCGGTGACGGCGGAG	NfMI
Saef	CCAGGGACCAGCAATGACCGCACTCGGATTCCTCTACC	Sae
Saer	GAGGAGAAGGCGCGTTACTAGTGGACCTGCACGAGCGGCTC	Sae
Pfef	CCAGGGACCAGCAATGTTTGGATGGAGGGGAAGAATTGGGCTTTTAGTC	Pfe
Pfer	GAGGAGAAGGCGCGTTAGAATCCCTTAGGAGCTTTCCGAGGGTAAG	Pfe
Smef	CCAGGGACCAGCAATGAGAACCCTGTGCGCCTCTCG	Sme
Smer	GAGGAGAAGGCGCGTCACGGCCTTTCCTCCTCGAGTGG	Sme
SceCf	CCAGGGACCAGCAATGGACATCTCCTTTCTCGGCGGACC	SceC
SceCr	GAGGAGAAGGCGCGTCATGTCCACCCTTTCCTGCTGTGTTC	SceC
SceDf	CCAGGGACCAGCAATGACAGCACTGGGATTCCTGTACC	SceD
SceDr	GAGGAGAAGGCGCGTCAGGCCTGCACCACCGGCTC	SceD
Mlef	CCAGGGACCAGCAATGAAACCTTGGCCGAGATCAAGGTCTATCC	Mle
Mler	GAGGAGAAGGCGCGTCAATCGCGCGGAGAGGAGCGTC	Mle
faf15f	CCAGGGACCAGCAATGCCTGGAGGAAGAGGTAGG	Ste
faf16r	GAGGAGAAGGCGCGTTAAAACCTGGAAGTTTTGCTTTTATC	Ste
Site directed mutagenesis primers for RseC		
faf3f	GGTCGTGGCTACCTGGGCACGTCGGGAAGTTTC	C87G
faf4r	GAAACTTCCCACGTCGCCAGGTAGGCCACGACC	C87G
faf5f	CGTTCGACCTGGCGCTCCCGCAGAACTGTGGCGATG	E29P
faf6r	CATCGCCACAGTTCTCGCGGGAGCGCCAGGTCGAACG	E29P
faf7f	CCTGCATCTGGCGGGCAGCCCTACGAACC	R46G
faf8r	GGTTCGTACGGCGTCCCCGCCAGATGCAGG	R46G
faf9f	GTGAGCCTGCATCTGGCGATGACGCCGTACGAACCCGTC	R46M
faf10r	CACGGGTTCGTACGGCGTCATCGCCAGATGCAGGCTCAC	R46M
faf11f	GAGGCGATCTTCGTGAGCTGCGGCAATCTGCCGACGTACGACGTC	T198G
faf12r	GACGTCGTACGTCGGCAGATTGCCGACGTCACGAAGATCGCCTC	T198G
faf13f	GATCTTCGTGAGCTGCACGGGCTGCCGACGTACGACGTC	N199G
faf14r	GACGTCGTACGTCGGCAGGCCGTCAGCTACGAAGATC	N199G
faf17f	GATCTTCGTGAGCTGCGGCGGCTGCCGACGTAC	T189G/N190G
faf18r	GTACGTCGGCAGGCCGCCGACGTCACGAAGATC	T189G/N190G
Site directed mutagenesis primers for NfMI		
faf19f	CTGGTGATCTCCTGCGCGGTGCAGATGCCCTCG	C194A
faf20r	CGAGGGCATCTGCACCGCGCAGGAGATCACCAG	C194A
faf21f	GCTGGTGATCTCCGCGTGCAGATGC	C193A
faf22r	GCATCTGCACGCACGCGGAGATCACCAGC	C193A
faf23f	GTGATCCTCTACGCCAGCCTGGTCGCGGTC	C76S
faf24r	GACCGCGACCAGGCTGGCGTAGAGGATCAC	C76S
faf25f	CGCTGGTGATCTCCAGCGCGGTGCAGATG	C193S/C194A
faf26r	CATCTGCACCGCGCTGGAGATCACCAGCG	C193S/C194A
faf27f	GTGATCTCCGCGAGCGTGCAGATGC	C193A/C194S
faf28r	GCATCTGCACGCTCGCGGAGATCAC	C193A/C194S
Sequencing primers		
T7P	TAATACGACTCACTATAGGG	T7 promoter
T7T	TATGCTAGTTATTGCTCAGCGGT	T7 terminator
fD1	AGAGTTTGATCCTGGCTCAG	16S
rP2	ACGGCTACCTTGTTACGACTT	16S

2.1.5 Source of DNA Samples

BbAMD was received from Dr. Krzysztof Okrasa (University of Manchester) as a codon optimised synthetic gene in a pGA4 plasmid containing an ampicillin resistance gene (GENEART). For the other genes, gDNA samples were acquired: *Nocardia farcinica* IFM 10152 from Dr. Jun Ishikawa (*Nocardia farcinica* Genome Project), *Rhodococcus* sp. RHA1 as a generous gift from Prof. Lindsay Eltis (UBC, Vancouver), *Pyrococcus furiosus* Vc1 and *Streptomyces avermitilis* NRRL 8165 from ATCC (numbers 43587 and 31267 respectively), *Streptomyces coelicolor* A3(2) from DSMZ (number 40783, renamed to *Streptomyces violaceoruber*), *Sulfolobus tokodaii* 7 from NBRC (number 100140), *Sinorhizobium meliloti* 1021 from Delphine Capela (INRA-CNRS, Castanet Tolosan), *Mesorhizobium loti* from Takakazu Kaneko (Rhizobase). The pET-YSBLIC3C plasmid was obtained from Dr. Mark Fogg (YSBL).

2.1.6 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to verify the size of DNA fragments. 1.2% (w/v) agarose was heated to dissolve in 100 mL TAE buffer (40 mM TRIS, 20 mM acetic acid, 50 mM sodium EDTA, pH 8.0) and 1 μL 10 mg mL^{-1} ethidium bromide added. DNA samples of approximately 100 ng were diluted in 4x DNA loading buffer (0.25 mg mL^{-1} bromophenol blue, 0.25 mg mL^{-1} xylene cyanol, 60% (v/v) glycerol, 15 mM TRIS, pH 7.6) and run at 100 V for 30 min together with 1 kbp or 100 bp DNA size ladders (NEB).

2.1.7 Site Directed Mutagenesis

Single amino acid mutants were produced with an adjusted QuickChange protocol (Stratagene, <http://www.stratagene.com/manuals/200518.pdf>) and the primers were designed with PrimerX (<http://www.bioinformatics.org/primerx>). Mutant strands were produced using 2 U Phusion High Fidelity DNA Polymerase, 1x HF buffer (Finnzymes), 0.22 mM of each dNTP, 0.22 μM of both forward and backward primer (Table 2.1, Sigma Genosys), 3.4 ng μL^{-1} original plasmid DNA and was diluted to 50 μL total volume with water. The temperature cycles were an initial melting step at 98 °C for 30 s followed by 16 times melting at 98 °C for 30 s, annealing at 55 °C for 60 s and elongation at 72 °C for 370 s performed on a Px2 Thermal Cycler (Thermo Scientific). The original DNA was digested specifically with 10 U DpnI (NEB) for 3 h at 37 °C and 2 μL were directly transformed into NovaBlue Singles (Novagen). All mutations were checked by sequencing in both directions using the T7P and T7T primers (Table 2.1).

2.2 Sequence Alignments

2.2.1 Determination of Sequence Identity

Pairs of sequences were aligned globally with "EMBOSS::needle" alignment tool on EBI (<http://www.ebi.ac.uk/Tools/emboss/align>) to determine the sequence identity.

2.2.2 Phylogenetic Tree

A selection of Asp/Glu racemase superfamily sequences was created from all homologues chosen for decarboxylation assays (Chapter 3), all sequences with a known structure and all sequences with a published, experimentally determined activity. The sequences were aligned with the ClustalW2 alignment tool (version 1.82) on EBI allowing end gaps (<http://www.ebi.ac.uk/Tools/clustalw2>). For all other parameters the default values were used (gap open penalty = 10, gap extension penalty = 0.2, gap separation penalty = 4, matrix = Gonnet). Obvious misalignments in the regions of the conserved cysteines were corrected by hand. The final alignment is shown in Appendix C.

For the production of the tree, the most conserved stretches were selected that are involved in the active site formation. They contain the positions of each of the conserved cysteines and the conserved dioxyanion hole position (Tyr128 in BbAMD). The stretches were then extracted the corresponding guide tree was calculated with ClustalW2 producing an unrooted tree with Phylip and neighbour joining that is based on pairwise sequence identities (Fig. 1.21). The stretches used for the tree calculation are shown in Appendix C.

2.3 Protein Production and Purification

2.3.1 Protein Expression

All the proteins were expressed heterologously in *E. coli* host strains using the lac operator system of the pET vector induced by IPTG (http://kirschner.med.harvard.edu/files/protocols/Novagen_petsystem.pdf, Novagen). BL21(DE3) strain was used for BbAMD while Rosetta 2(DE3) strain (Novagen) was used for the homologues in pET-YSBLIC3C. Colonies were grown on LB-agar plates containing 100 $\mu\text{g mL}^{-1}$ kanamycin and for Rosetta 2(DE3) containing an additional 34 $\mu\text{g mL}^{-1}$ chloramphenicol. 5 mL LB starter cultures containing the same antibiotics

were inoculated with a fresh colony and incubated at 37 °C shaking at 200 rpm for 16 h. Then 5-2,000 mL main cultures with the same medium were inoculated with 1% (v/v) of the starter culture and incubated in the same conditions. When the optical density at 600 nm reached 0.7-0.9, 1 mM IPTG was added and the culture was incubated at 30 °C shaking for another 21 h before the cells were then spun down at 2,000 × g for 10 min.

2.3.2 Cell Lysis

The cell pastes produced containing the protein were then resuspended in 20 mM TRIS, 150 mM NaCl pH 8.0 supplemented with 200 μM PMSF to inhibit periplasmic proteases, 5 mM DTT to keep the protein reduced and stable, 5 μg mL⁻¹ hen egg white lysozyme to digest the cell walls, 5 μg mL⁻¹ DNase I and 1 mM MgCl₂ to cut the DNA in order to reduce the viscosity of the sample. In some cases of sodium phosphate buffer and 300 mM NaCl was used at the same pH.

The cell suspension was then sonicated using an S-4000 Sonicator (Misonix) at 70% amplitude during for total duration of 4 min in 3 s intervals interrupted by 7 s cooling periods at 0 °C. The lysate was then cleared of cell debris by centrifugation at 45,000 × g for 20 min.

2.3.3 Protein Purification

To purify the protein by its His-tag the clarified cell lysate was filtered through a 0.22 μm membrane (Millipore) and loaded with a syringe onto a 5 mL HisTrap HP nickel column (GE Healthcare). The buffer used was 20 mM TRIS, 150 mM NaCl pH 8.0. The protein was then washed by increasing the imidazole concentration from 0 to 50 mM for 100 mL and eluted by increasing further from 50 to 300 mM for another 100 mL. The fractions corresponding to the peak in absorbance measured at 280 nm were pooled (typically around 130 mM imidazole) and concentrated to 2 mL using Amicon Ultra15 centrifugal filters 10 kDa molecular weight cut off (Millipore). If the protein was only needed for activity assays simple nickel affinity resin batch purification was performed by stepwise adding of higher concentrations of imidazole and spinning down the resin.

Gel filtration was performed to purify the protein further and remove imidazole with the aim of crystallisation. The concentrated protein sample was injected into a HiLoad 16/60 Superdex 75 preparative grade column (GE Healthcare) and eluted with the

same nickel column wash buffer. Again the peak fractions were pooled and concentrated to 5-15 mg mL⁻¹ protein concentration as determined by the Bradford assay. Aliquots of 50 µL were snap frozen in liquid nitrogen and stored at -80 °C (Fig. 2.3).

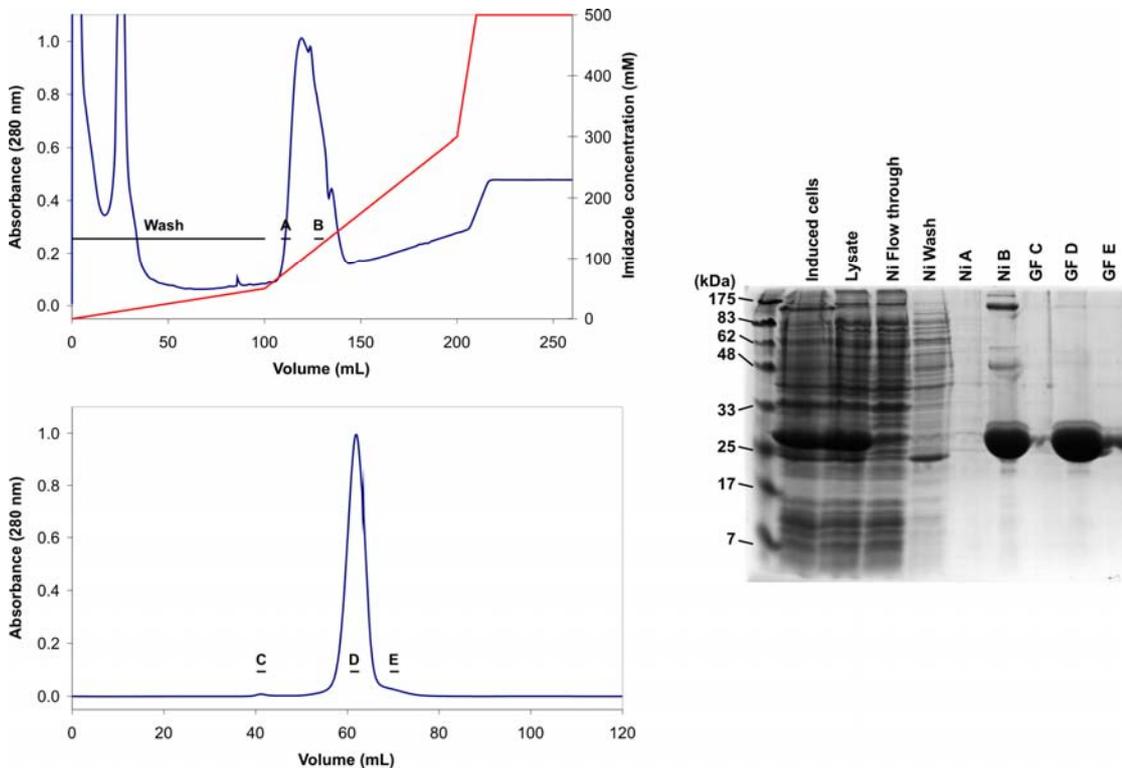


Figure 2.3. Typical purification of BbAMD homologues. Top: Chromatogram of a nickel affinity column purification (Ni). Bottom: Chromatogram of a gel filtration column (GF). Absorbance (blue) and imidazole concentration (red) are shown in chromatograms. Right: SDS-PAGE of different purification steps.

2.3.4 Protein Size and Purity Determination

The proteins were analysed by SDS-PAGE using 0.75 mm, 12% gels. Separating gel (375 mM TRIS pH 8.8, 0.1% (w/v) SDS, 10% (w/v) acrylamide, 0.3% (w/v) N,N-methylene-bis-acrylamide, 0.04% (v/v) TEMED, 0.1% (w/v) APS) was topped by stacking gel (as separating but 125 mM TRIS pH 6.8, 5% (w/v) acrylamide, 0.1% (w/v) N,N-methylene-bis-acrylamide) and immersed in running buffer (25 mM TRIS, 200 mM glycine, 0.1% (w/v) SDS). The samples were diluted 1:4 in loading buffer (2% (w/v) SDS, 0.15% (w/v) bromophenol blue, 62.5 mM TRIS pH 6.8, 40% (v/v) glycerol, 20% (v/v) BME) and unfolded at 95 °C for 2 min.

To analyse whole cell samples including insoluble protein, the pellet of 1 mL cell culture was heated to 95 °C for 10 min in total solubilisation loading buffer (1% (w/v) SDS, 2 M urea, 1.25 (v/v) BME, 2.5% (v/v) glycerol, 15 mM TRIS pH 6.8, 0.15 % (w/v) bromophenol blue) (115). Gels were typically run at 200 V for 45 min and stained using Blue Silver (0.12% (w/v) Coomassie G-250, 10% (w/v) $(\text{NH}_4)_2\text{SO}_4$, 10% (v/v) phosphoric acid, 20% (v/v) methanol) for 1 h and then destained in water for 1 h (116). As a size marker the prestained broad range protein marker (NEB) was used.

2.3.5 Protein Quantification Assay

The Bradford protein quantification assay is an easy to use assay that can be applied to protein mixtures as well as pure protein. A standard curve was produced using Coomassie (Bradford) Protein Assay solution and the BSA standard (Thermo Scientific) using 10 μL samples added to 1 mL of reagent. The absorbance at 595 nm was measured. The approximation formula for the protein concentration in mg mL^{-1} “ $c = 2 \times A - 0.25$ ” was used, where A corresponds to the absorbance. The linearity is maintained for absorbance values between 0.2 and 0.8.

2.3.6 Analysing Protein Monodispersity

Dynamic light scattering (DLS) was used to assay the monodispersity (i.e. non-aggregation) of the protein solution and gain an indication of the multimerisation state of the protein. The protein was diluted to 1 mg mL^{-1} in buffer and centrifuged for 10 min at 18,000 $\times g$ to remove large particles. The solution was tested on a DynaPro particle sizer (Wyatt Technology) at 18 °C. The samples were measured until 15 statistically good data points were recorded (indicated by the software). The polydispersity <25% was considered monodisperse and the corresponding molecular mass, assuming a spherical particle, was read.

2.3.7 Analysing Secondary Structure

Circular dichroism (CD) was used to assess the folding state of the protein by checking the existence of secondary structure elements. Protein solutions of 1 mg mL^{-1} were diluted to 0.2 mg mL^{-1} in water to reduce buffer content and the CD spectrum was measured between 195 nm and 260 nm on a J-810 spectropolarimeter (Jasco). The buffer spectrum was subtracted from the spectrum of the protein in buffer. The presence of properly folded α -helices was verified by the presence of characteristic troughs at 208 and 222 nm.

Chapter 3

Screening of BbAMD Homologues for Decarboxylation of Malonates

3.1 Introduction

BbAMD was isolated from soil cultures growing on phenylmalonate as their sole carbon source and was found to stereospecifically decarboxylate disubstituted malonates (76). BbAMD accepts malonates containing an aryl group as one substituent and hydrogen, methyl or fluorine as the other substituent (51,67,85). BbAMD accepts a large range of aryl substituents including naphthyl, thienyl and phenyl with various substituents on the aromatic rings such as chloro, fluoro, methoxy and methyl groups to alkenyls (51). Special substrates such as [2-(2-oxoethyl)phenyl]propanedionate and 2-(2-fluoro-4-biphenyl)-2-methylmalonate were also reported to be decarboxylated by BbBMD producing an aldol reaction cascade and the drug flurbiprofen respectively (Fig. 3.1) (86,117).

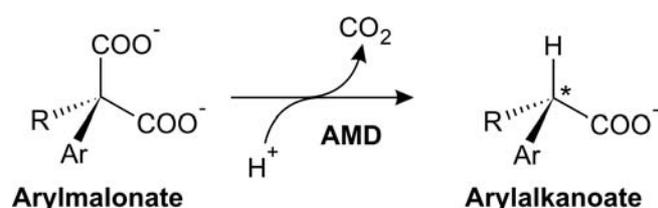


Figure 3.1. Activity of AMD. R = H, CH₃, OH, NH₃. Ar = aryl, alkenyl.

By analogy to AspRs and GluRs, the mechanism of BbAMD is thought to go through an enediolate intermediate resulting from the decarboxylation of the substrate. The intermediate is stabilised by interactions with the active site and is then protonated by Cys188 from the *si*-face of the planar intermediate as shown by deuterium labelling in GluRs and mutation to alanine in BbAMD (58,92). Both the decarboxylation and the protonation have been shown to be stereospecific by experiments involving isotope labelling. The decarboxylation was shown to proceed directly through the loss of the

pro-*R* carboxylate with an overall inversion of stereospecificity by ^{13}C and ^{18}O labelling (48,118).

Recently the structure of BbAMD was solved and found to have the same overall fold as the other known racemase structures (48,67,88,90). The structures have different ligands bound in the active site: BME, phosphate and benzylphosphonate. The binding site of phosphate and the phosphonate revealed six potential hydrogen bonding partners, able to stabilise a doubly negatively charged enediolate intermediate. This dioxanion hole is formed by the hydroxyl of Thr75 and the backbone amides of both Thr75 and Gly189 for one negative charge and for the other negative charge by the backbone amide of Ser76 together with hydroxyls of both Ser76 and Tyr126. The structures also revealed a large and a small binding pocket thought to be responsible for binding the substituents of malonates. By analogy to the binding of the aromatic ring of benzylphosphonate it is thought that the aryl group binds in the large binding pocket whereas the small binding pocket is thought to bind both the cleaved carbon dioxide and the small substituent if present. The overall reaction is ultimately thought to be driven by the destabilisation of the negative charge in the leaving carboxyl group by the hydrophobic environment and the stabilisation of the double negative charge in the carboxyl group bound at the dioxanion hole (48). Gaseous carbon dioxide then leaves the solution driving the equilibrium strongly in the decarboxylation direction.

BbAMD catalyses the efficient production of chiral 2-arylpropionates from 2-methyl-2-arylmalonate substrates. Conversely, the production of neither chiral 2-arylalkanoates with longer aliphatic chains nor fully aliphatic chiral 2-alkylalkanoates has been achieved yet. These chiral intermediates for chemical synthesis would be of commercial interest to industry. The enzymatic decarboxylation of alkylmalonates would represent an easy and efficient route to obtain the intermediates.

There is a range of BbAMD homologues available in the public sequence database UniProt. Testing diverse potential AMDs is interesting in many ways. On one hand, finding new AMDs would allow new patents to be filed for commercial exploitation of the discoveries with engineering methods. On the other hand, the diversity of enzymes might be expected to offer a range of completely new substrate specificities allowing the exploration of the active site flexibility and gaining useful insight for future engineering efforts.

In this chapter, ten sequence homologues of BbAMD were also tested for decarboxylation of PM and a range of other disubstituted malonates. Mutants of an

inactive homologue were produced to increase the resemblance to BbAMD and tested for PM decarboxylation. Another inactive homologue was also tested for alternative activities found in the Asp/Glu racemase superfamily.

3.2 Methods

3.2.1 Chemical Synthesis of MPM

MPM was not available commercially and therefore had to be synthesised in our laboratory. The strategy was to start from readily available dimethylphenylmalonate, methylate the C2 carbon and hydrolyse the methylesters.

The methylation was performed as follows: In 25 mL anhydrous ethanol 1.2 g (17.6 mmol) sodium ethoxide was dissolved at 0 °C and under nitrogen atmosphere. Then 2.73 mL (12.7 mmol) dimethyl-2-phenylmalonate was added and stirred for 20 min at room temperature. Then 3 g (21.1 mmol) iodomethane was added and stirred for 1 h at room temperature. After verifying the completion of the reaction by TLC, 100 mL water was added to stop the reaction. The products were extracted with 400 mL ethyl acetate adding some saturated NaCl solution and methanol, dried with MgSO₄ and filtered. The extract was concentrated by rotary evaporation and purified by silica gel column chromatography with a mobile phase of ethyl acetate/petroleum ether starting with 200 mL 1:19 (v/v) and finishing with 200 mL of 1:9 (v/v). Fractions of 10 mL were collected and analysed by TLC. The fractions containing the product were pooled and the solvent was removed by rotary evaporation. The yield was 2.5 g (11.2 mmol, 88%) dimethyl-MPM.

Reaction products were analysed by TLC with silica gel as the solid phase and ethyl acetate/petroleum ether 1:9 (v/v) as the mobile phase. The plates were stained by immersing in 1% (w/v) KMnO₄ and 2% (w/v) Na₂CO₃, followed by a heat-drying process to reveal the spots (Fig. 3.2, top left).

The hydrolysis was performed as follows: Diethyl-MPM was dissolved in 120 mL ethanol and 110 mL 4M KOH was added and stirred at room temperature for 16 h. The hydrolysis was checked by TLC. Then the product was acidified with concentrated HCl to pH 1 and extracted with 1 L ethyl acetate adding some saturated NaCl and methanol, dried with MgSO₄ and filtered. The extract was concentrated by rotary evaporation and crystallised by adding petroleum ether and cooling it down to 0 °C.

Crystals were filtered with blotting paper. The yield was 0.9 g (4.6 mmol, 41% hydrolysis, 36% overall).

3.2.2 Analysis of Synthesised MPM

The identity and purity of the product were tested by three methods: $^1\text{H-NMR}$, HPLC and MS.

The NMR was conducted by an in-house service using the 500 MHz NMR spectrometer AV500 (Bruker). 20 mg of MPM were dissolved in 750 μL of $\text{CDCl}_3/\text{acetone-D}_6$ 9:1 (v/v). A total of 8 H were expected: A singlet peak equivalent to 3 H with a chemical shift of 1.82 ppm corresponding to the three methyl H, a multiplet peak corresponding to 5 H with a chemical shift of 7.15-7.38 ppm corresponding to the 5 phenyl H, and a broad singlet peak equivalent to 2 H with a chemical shift of 7.55-8.75 ppm corresponding to the 2 carboxylic acid H. These peaks were all visible on the spectrum (Fig. 3.2, top right).

The MS was performed by an in-house service as well using the ABI Qstar ESI-TOF mass spectrometer (Applied Biosystems). 1 mM MPM in methanol was prepared and diluted 1:100 in acetonitrile/water 1:1 (v/v). The m/z of 193.1 Da was detected corresponding to the mass minus 1 H^+ (Fig. 3.2, bottom right).

The HPLC was performed by dissolving MPM in 2-propanol at 100 mM and diluting it 1:20 in eluent and injecting 10 μL as described in HPLC method below. The run resulted in a single and clear peak at RT of 13 min (Fig. 3.2, bottom left)

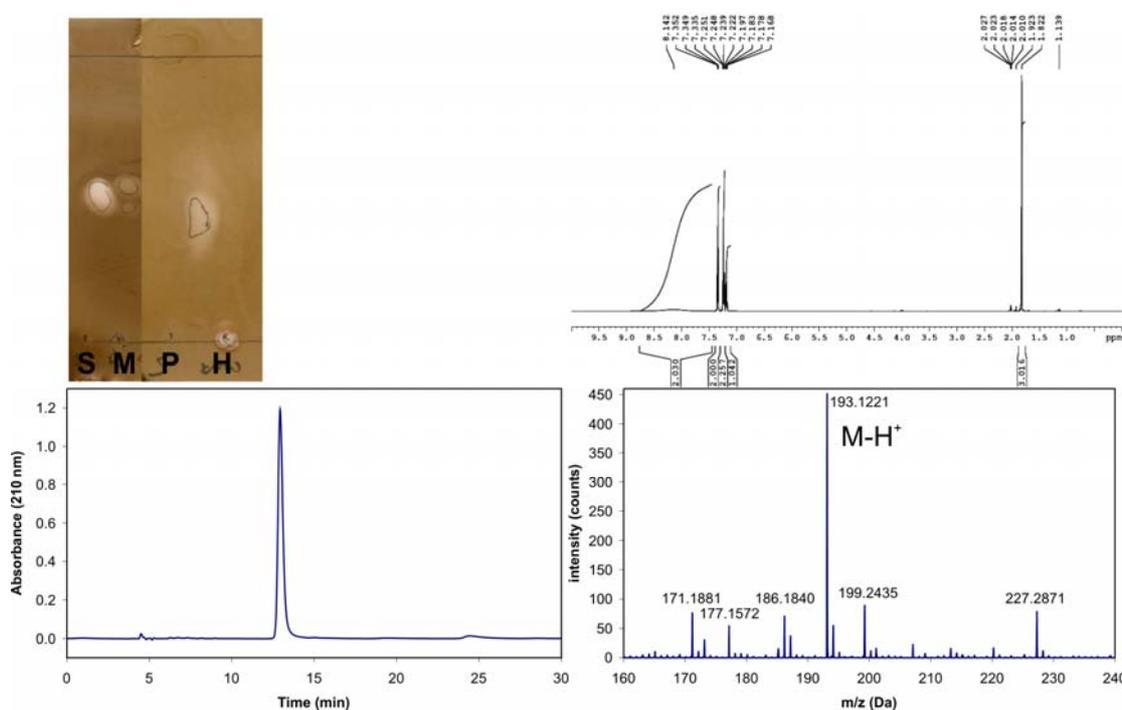


Figure 3.2. MPM production and purification. Top left: TLC of MPM production steps. S: Dimethyl-2-phenylmalonate. M: Methylated S = 2-methyl-dimethyl-2-phenylmalonate. P: Purified M. H: Hydrolysed P = MPM. Top right: ¹H-NMR of MPM. The peaks correspond to carboxyl-Hs at 8.1 ppm, aromatic Hs at 7.1-7.4 ppm and methyl-Hs at 1.8 ppm. Bottom left: HPLC trace of 10 μ L of 5 mM MPM. Bottom right: MS of MPM.

3.2.3 Strategy for Malonate Decarboxylation Assays

Three methods have been described for measuring the decarboxylation activity of disubstituted malonates: HPLC, GC and spectrophotometric detection of a pH change (48,76). HPLC and GC offer the possibility of detecting the stereospecificity of the reaction through chiral columns, whereas the spectrophotometric method allows a real time recording of the reaction and the application of the method to high throughput screening useful for directed evolution approaches.

There are various limitations to the spectrophotometric assay that uses a pH indicator to measure a difference in absorbance. As the pH is used to quantify the turnover it cannot be constant as it should ideally be for enzyme assays. It is also not entirely clear how the reaction and pH change are linked and how solvation and degassing of carbon dioxide indirectly and directly influence the pH during the reaction. In fact, two separate increases of absorbance could be observed upon addition of the enzyme both

not corresponding to the increase observed by HPLC. This method was therefore explored but not applied in the present work.

The HPLC method detects aromatic groups with a high sensitivity at 210 nm and is therefore well suited for the analysis of arylmalonates, but alkylmalonates only have the carboxyl group that only gives a weak signal at 205 - 210 nm. For chiral separations of the products of stereospecific decarboxylations, a normal phase column was used, requiring prior extraction of the reaction products. For the detection of alkylmalonate decarboxylation activities an easier non-chiral reverse phase column was used that was insensitive to the chirality of the compounds.

To increase the sensitivity of the alkylmalonate detection, GC coupled with FID would be a good choice, offering a sensitive and easily quantifiable signal. However, in this case the carboxyl groups would have to be derivatised to become apolar and therefore volatile. The method was not applied in the present work.

3.2.4 Extraction and Normal Phase HPLC Analysis

Samples were saturated with NaCl so non-dissolved residual crystals were visible in the tube and acidified by adding 1/5 volume of 2 M HCl. 1 volume ethyl acetate was added and the samples were vigorously shaken for 30 s. The phases were separated by centrifuging at $18,000 \times g$ for 1 min and the organic layer was removed and dried with $MgSO_4$. A 20 μL sample was added to 180 μL eluent and 10 μL were run on normal phase CHIRALPAK IA, 250 x 4.6 mm (CHIRAL Technologies Europe) HPLC column using n-Heptane/2-propanol/TFA 90:10:0.1% (v/v) as mobile phase at a flow rate of 0.7 mL min⁻¹ on a 510 Pump with 717plus Autosampler (Waters). The peaks were detected at 210 nm using a 2487 Detector (Waters). Retention times were PM 16.7 min, PA 7.6 min, MPM 14.3 min, (R)-PP 7.6 min, (S)-PP 7.8 min, MEM 8.5 min, MPrM 8.7 min, EBM 8.2 min and DMCHDC (4 diastereoisomers) 8.6, 9.0, 9.8, 10.0 min.

3.2.5 Whole Growing Cell Decarboxylation Assay

Escherichia coli Rosetta 2(DE3) cells were grown and induced as described earlier (Chapter 2). During induction 1, 4, 20, 40, 400 or 1000 mM PM, pH 8 were added to the cell culture. The culture was at 60% (alternatively 50%, 75%) of the total reaction volume. The reaction was buffered to pH 8.0 with 20 mM TRIS or pH 6 with 20 mM sodium citrate/sodium phosphate buffer and 150 mM NaCl were added. The mix was

then incubated at 30 °C shaking at 200 rpm for 16 h. The products were extracted and analysed by normal phase HPLC.

3.2.6 Whole Resting Cell Decarboxylation Assay

The proteins were expressed as described earlier but grown for 3 h at 37 °C after induction. Then the cells were spun down and the pellet was resuspended in the reaction mix at 0.5 g mL⁻¹. The mix was 20 mM PM, 150 mM NaCl and 20 mM TRIS pH 8.0. The samples were incubated at 37 °C shaking at 200 rpm for 16 h. The products were extracted and analysed by normal phase HPLC.

3.2.7 Cell Lysate Decarboxylation Assay

The proteins were expressed as described earlier and induced for 6 h at 30 °C or 3 h at 37 °C. The 16 h induction cells were spun down and the equivalent of 3.75 mL culture volume were resuspended in water and 750 µL of 100 mM PM, pH 8.5, 500 µL 1 M TRIS, pH 8.5 were added. The cells were lysed by adding 200 µL BugBuster (Novagen) and the mix was incubated at 37 °C shaking for 24 h.

Alternatively, the cells from the 3 h induction were resuspended in the same volume of 20 mM TRIS, pH 8.0, 150 mM NaCl and 5 µg mL⁻¹ hen egg white lysozyme. The cells were sonicated using S-4000 Sonicator (Misonix) at 70% amplitude during for 4 min total duration in 3 s intervals interrupted by 7 s cooling periods at 0 °C. The cell debris was removed by centrifugation at 2,000 × g for 10 min and 500 µL aliquots were snap frozen in liquid nitrogen and stored at -80 °C. 88 µL of thawed lysate was mixed with 10 µL of water and 2 µL of 1 M PM and incubated at 37 °C shaking at 200 rpm for 16 h. The reaction products were extracted and analysed by normal phase HPLC.

3.2.8 Pure Enzyme Decarboxylation Assay

The enzymes were expressed and purified as described earlier (Chapter 2). 10 mM PM or 15 mM PM, MPM, MEM, MPrM, EBM, DMCHDC were incubated with 1, 2, 20 or 100 µg mL⁻¹ enzyme. Reactions were conducted with 5 mM MOPS pH 7.2 or 100 mM TRIS pH 8.5 at room temperature for 10 min or 3.5 d. The products were extracted and analysed by normal phase HPLC.

3.2.9 Spectrophotometric Decarboxylation Assay

The reaction mix was typically prepared with 10 mM PM, pH 7.2, 5 mM MOPS, pH 7.2 (pK_a of 7.20), 15.5 µM BTB (pK_a of 7.10). The enzyme concentration was typically 0.8

μM . The increase in absorbance due to a BTB colour change was measured at 595 nm indicating change in pH from acidic (yellow) alkaline (blue). The number of OH^- ions produced by the reaction is proportional to the change in absorbance as long as the $\text{p}K_a$ of the buffer and the $\text{p}K_a$ of the indicator are sufficiently close (± 0.1). As the buffer hides part of the produced OH^- from the indicator the ratio between the concentration of the buffer and the indicator is used to relate the change in absorbance to the change in concentration of substrate or product (48,119,120).

3.2.10 Homology Modelling of RseC

The RseC sequence was aligned with BbAMD (PDB: 3dg9), Ste (PDB: 2dgd), PhHydR (PDB: 2eq5), ApGluR (PDB: 1b73) and PhAspR (PDB: 1jfl) using the ClustalW2 alignment tool on EBI allowing end gaps (<http://www.ebi.ac.uk/Tools/clustalw2>). It was then manually adjusted and realigned with the T-Coffee alignment tool on EBI (<http://www.ebi.ac.uk/Tools/t-coffee>). A model was calculated with the software MODELLER (121). The resulting structure was aligned with BbAMD structure on PyMol (DeLano Scientific).

3.3 Results

3.3.1 Cloning and Expression of BbAMD Sequence

Homologues

The protein sequence of BbAMD was used to perform a BLAST (<http://www.expasy.ch/tools/blast>) search against the UniProt database in order to identify known homologues. Ten hits were selected that were more distantly related to BbAMD but still among the top hits (Table 3.1). Ste was included as its crystallographic structure was available. The selection was biased towards easily available gDNA samples and to avoid overlap with the simultaneous process performed by our collaborator Dr. Krzysztof Okrasa from the University in Manchester (personal communication).

All homologues were of similar size (25 - 31 kDa). Despite the relatively low sequence identities (19 - 30% to BbAMD) the homologues contain two conserved sequences around the positions known to contain the active site cysteines in GluR. In GluR both cysteines are located one on each domain opposite each other. The cysteines are conserved in most homologues with some notable exceptions. One exception is

BbAMD, which contains a glycine at position 74 (domain 1). RseD, Sae and SceD contain an aspartate at the position on domain 2 and Ste contains an arginine at the position on domain 1 (Table 3.1, Appendix C). It has to be noted that although the arginine in Ste is a large residue at a conserved position in the active site, the structure of the protein shows that the side chain is bound in the core of the protein exposing only the backbone part to the active site. The impact on the active site therefore resembles the impact of a glycine.

Table 3.1. Selected homologues and properties.

Abbreviation	Organism	Annotation ^a	UniProt ID	Residue ^b		Identity ^c	Gene (bp)	Protein (kDa)
				1	2			
BbAMD ^d	<i>Bordetella bronchiseptica</i>	AMD	Q05115	Gly	Cys	100%	720	24.7
RseD	<i>Rhodococcus</i> sp. RHA1	AMD	Q0S7X2	Cys	Asp	30%	741	25.5
Pfe	<i>Pyrococcus furiosus</i>	MI	Q8U183	Cys	Cys	28%	711	26.6
Sae	<i>Streptomyces avermitilis</i>	Decarboxylase	Q82CV3	Cys	Asp	27%	735	25.9
Sme	<i>Sinorhizobium meliloti</i>	AMD	Q92WC5	Cys	Cys	27%	786	28.2
SceD	<i>Streptomyces coelicolor</i>	Decarboxylase	Q93J75	Cys	Asp	26%	732	25.6
SceC	<i>Streptomyces coelicolor</i>	Decarboxylase	Q93J74	Cys	Cys	26%	903	31.2
RseC	<i>Rhodococcus</i> sp. RHA1	AMD	Q0S7X3	Cys	Cys	26%	744	27.4
Mle	<i>Mesorhizobium loti</i>	AMD	Q987A4	Cys	Cys	25%	780	28.0
NfMI	<i>Nocardia farcinica</i>	MI	Q5YXQ1	Cys	Cys	22%	756	26.5
Ste ^d	<i>Sulfolobus tokodaii</i>	AMD	Q974K3	Arg	Cys	19%	672	25.5

^aThe protein name in the NCBI Entrez Gene database

^bCatalytic residues aligning with the acid/base cysteines in Asp/Glu racemases (see Appendix B)

^cSequence identity determined by pairwise global alignment

^dKnown Structure with PDB ID (BbAMD: 3dg9, 2vlb, 3ips, 3dtv; Ste: 2dgd)

All homologues were cloned from gDNA except BbAMD, which was cloned from a codon optimised sequence provided by Dr. Krzysztof Okrasa. The sequences were cloned using the LIC technique into the pET-YSBLIC3C vector allowing IPTG-inducible protein expression in the heterologous host *E. coli*. The vector encodes an N-terminal His-tag for convenient protein purification.

All proteins expressed well in *E. coli* Rosetta 2(DE3) (Fig. 3.3). With most constructs, soluble expression could be achieved. Large insoluble fractions were obtained only with RseC, Pfe and SceD although later nickel affinity column purification of RseC and Pfe showed that there was a smaller soluble fraction still present, enough to produce 5 - 10 mg of pure protein for further analysis and crystallisation trials.

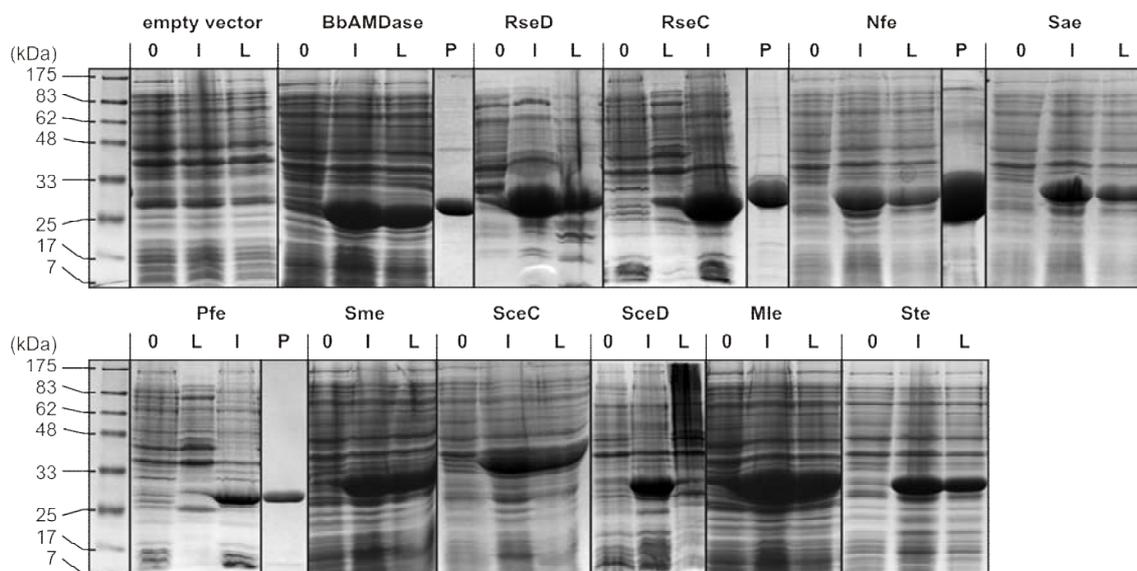


Figure 3.3. Expression and purification of BbAMD homologues. SDS-PAGE of cell culture (0), induced cell culture (I), lysate (L), purified protein (P). The gels were graphically sized to approximately the same scale. All lanes of one homologue are from the same gel except the purified protein.

3.3.2 Decarboxylation Assays with BbAMD

BbAMD was reported to accept a wide range of disubstituted malonates as substrates. For the malonates to be accepted as substrates the portion accepted by the large binding pocket has to be electron withdrawing and the portion accepted by the small binding pocket has to be smaller than an ethyl group. The large pocket binds substrates such as phenyl, naphthyl, thienyl and alkenyls, whereas the small pocket accepts hydrogen, methyl, hydroxyl and amines. The best substrates contain naphthyl or chlorophenyl substituents for the large portion and hydrogen for small portion (Fig. 3.1) (51,67).

From an industrial perspective it would be interesting to develop an AMD that accepts alkylmalonates as substrates that have both larger substituents and no requirement for electron withdrawing potential. A series of substrates was therefore provided by BASF for testing with BbAMD homologues: MEM, MPrM, EBM and DMCHDC (Fig 3.4). The substrate PM and all the products were commercially available. MPM had to be synthesised in-house (see methods above).

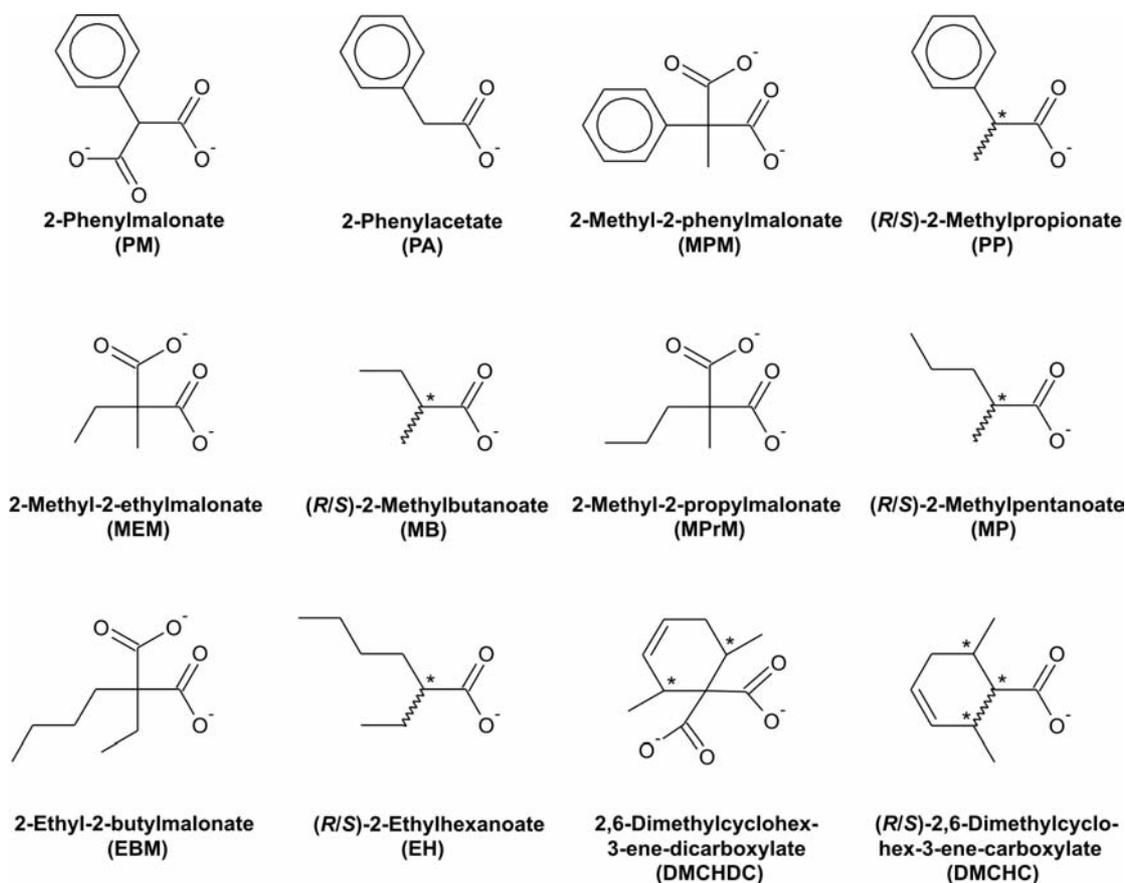


Figure 3.4. Substrates and products tested on BbAMD.

For the analysis of decarboxylation in enzyme assays a quick and easy colorimetric method was proposed (48). However, the signal obtained from a pH indicator could not be linked to the activity and selection. Additionally changes of colony colour on agar plates could never be made a reliable indicator for activity. This method was therefore abandoned in favour of a more laborious HPLC method. Substrate and products were extracted with ethyl acetate and injected into a normal phase chiral HPLC column and detected as absorbance at 210 nm.

BbAMD was established as a positive control, as reported earlier, using purified enzyme together with HPLC analysis (51). The enzyme converted 10 mM of PM completely into PA in less than 10 min. Despite the relatively high experimental error (most of it probably due to the handling of volatile solvent during extraction) the initial activity approximately corresponds to a k_{cat} of 40 s^{-1} . This is approximately nine times lower than the published value. This is not surprising as the assay temperature for the published value was $37 \text{ }^\circ\text{C}$ whereas here the temperature was only $25 \text{ }^\circ\text{C}$ (51).

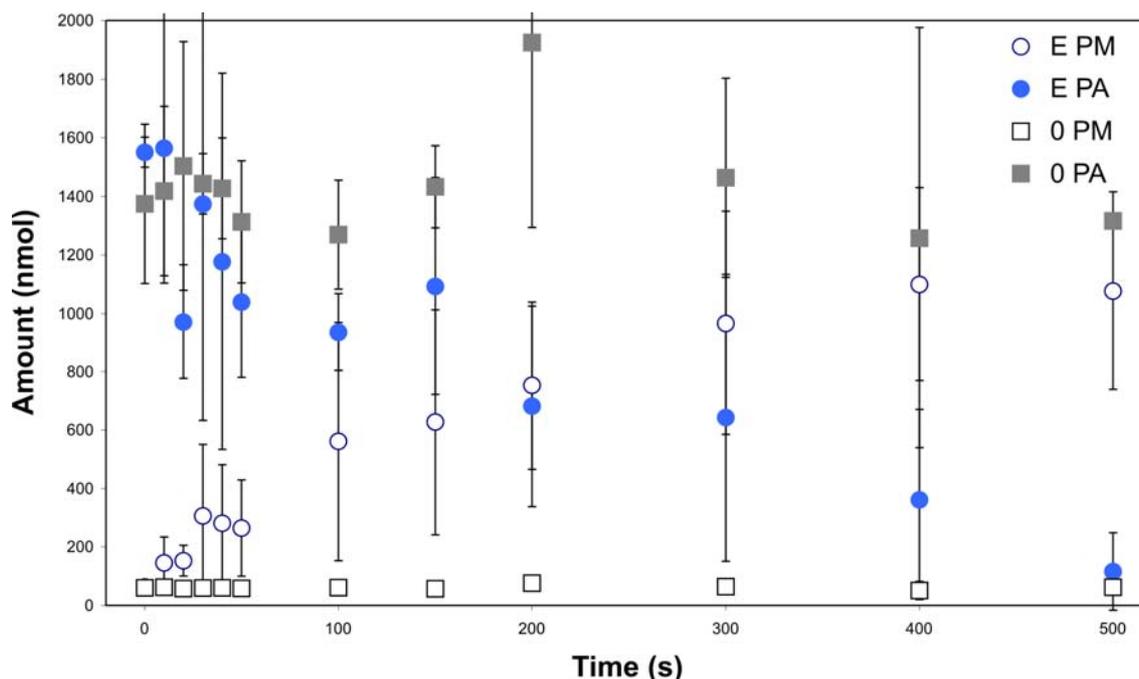


Figure 3.5. Activity of BbAMD on PM. 10 mM of substrate was incubated with $2 \mu\text{g mL}^{-1}$ (E) and without (0) purified AMD at 25°C and pH 8.5. Compounds were extracted with ethyl acetate and $10 \mu\text{L}$ were analysed by HPLC. Mean values are plotted and error bars correspond to standard deviation of experiments performed in triplicate.

3.3.3 Screening for Arylmalonate Decarboxylation in Homologues

All homologues were tested with the whole resting cell assay for conversion of PM to PA and analysed by HPLC. This method keeps the enzymes in their cytoplasmic environment avoiding degradation during the extraction process. Compared to a growing cell approach, which was also used, the resting cell approach allowed the cells to concentrated and control reaction conditions such as temperature and buffer composition. A positive BbAMD control was always performed to show that the substrate and product could diffuse through the bacterial cell wall and membrane.

For all of the homologues analysed, the results remained the same. The positive BbAMD control invariably converted all the substrate into product. The negative control clones with the empty vector showed no conversion. The small PA peak constantly visible on the chromatograms corresponds to substrate contamination also observed in pure standard injections. A typical result, representative of all assays performed this way is shown in Fig. 3.6.

A number of variations of this assay were performed with the same result. Some homologues were tested in growing cell assays, allowing the substrate to interact with the enzyme at different concentrations or different post-translational modification stages during the induction time. Almost all were tested in lysate, which allows tighter control of the reaction conditions such as pH and high temperatures. RseC and Pfe were also tested as purified enzymes. Purified Pfe and Ste lysate, both originating from thermophilic organisms, have been tested at high temperatures. The applied temperatures ranged from 50 °C to 95 °C. No decarboxylation could be observed in these conditions either, although at approximately 80 °C a chemical background decarboxylation appeared but never exceeded the negative control (data not shown).

A summary of the homologues tested and assays performed can be found in the discussion of this chapter (Table 3.3).

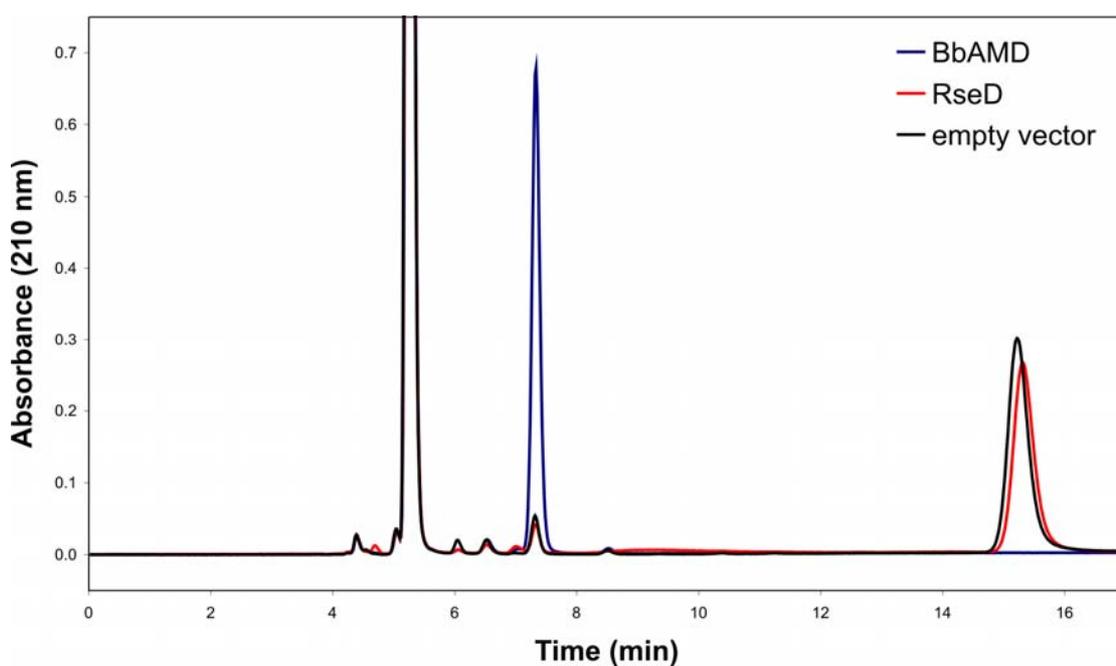


Figure 3.6. HPLC chromatograms of a typical PM decarboxylation assay. In this case it is a whole resting cell assay with 20 mM PM transformed by a 0.5 g mL^{-1} suspension of induced cells incubated overnight at 37 °C and pH 8. The samples were extracted with ethyl acetate and 10 μL were injected into the HPLC run. PM runs at 15.3 min, PA at 7.3 min and ethyl acetate at 5.2 min.

3.3.4 Screening for Arylmalonate Decarboxylation in Mutants of RseC

Given the position of the tested homologues in the phylogenetic tree of the Asp/Glu racemase superfamily, it is rather surprising not to find any decarboxylase activity. In particular, the residues identified as being part of the mechanism of BbAMD are retained in RseC and Pfe as seen in the alignment of the superfamily sequences (Appendix C) (48). In addition, the 26 and 28% overall sequence identities of RseC and Pfe with BbAMD in the background of only 22% identity for the whole superfamily might also suggest at least some residual AMD activity potential in these homologues. In another study, residual activity has been identified in other similarly closely related BbAMD homologues with similar sequences (48).

To check whether one or two crucial mutations in the active site would transform an inactive homologue into an AMD, three dimensional models of RseC and Pfe were produced based on the homology of the superfamily. The modelling was based on the sequence alignment of one enzyme with each activity in the Asp/Glu racemase superfamily including the structure of Ste (PDB: 3dg9 for BbAMD, 2dgd for Ste, 2eq5 for PhHydR, 1jfl for PhAspR and 1b73 for ApGluR).

The resulting models were aligned with the BbAMD structure. It has to be noted that the alignments of the sequences are ambiguous on large stretches and as the models are heavily dependent on the correct alignment they have to be considered carefully. As expected, both models conserved the overall fold. The dioxyanion hole region was very well conserved in these models but large discrepancies with BbAMD were found in external loop regions as well as in the residues forming the small binding pocket. The proline residues of the large binding pocket were conserved but many other residues in RseC were pointing into that pocket introducing charge, polarity or simply volume into the active site. The charged or polar residues are expected in AspRs, GluRs where carboxyl groups would be bound but not in AMDs where the aliphatic aryl group has to be bound.

Site directed mutagenesis was finally used to introduce changes into RseC. Three sites were targeted and mutated into the corresponding residues in BbAMD according to the alignment. The cysteine on domain 1 was changed to glycine (C87G). The threonine and asparagine following the cysteine on domain two were mutated to glycines (T198G/N199G) as these could sterically and electrostatically inhibit binding of the aryl

group. Finally, the glutamate and arginine in the back of the large binding pocket were mutated to aliphatic residues removing the charge (E29P/R46G) that could also potentially hinder aryl group binding. As R46G is a very large change in the volume of the residue at this position a more conservative methionine mutant was introduced (R46M, Fig. 3.7).

All of the single and double mutants described above were tested with the whole resting cell assay. All proteins were expressed successfully as judged by SDS-PAGE (data not shown). However, as with the homologues none of the mutants showed any residual activity (Fig. 3.2).

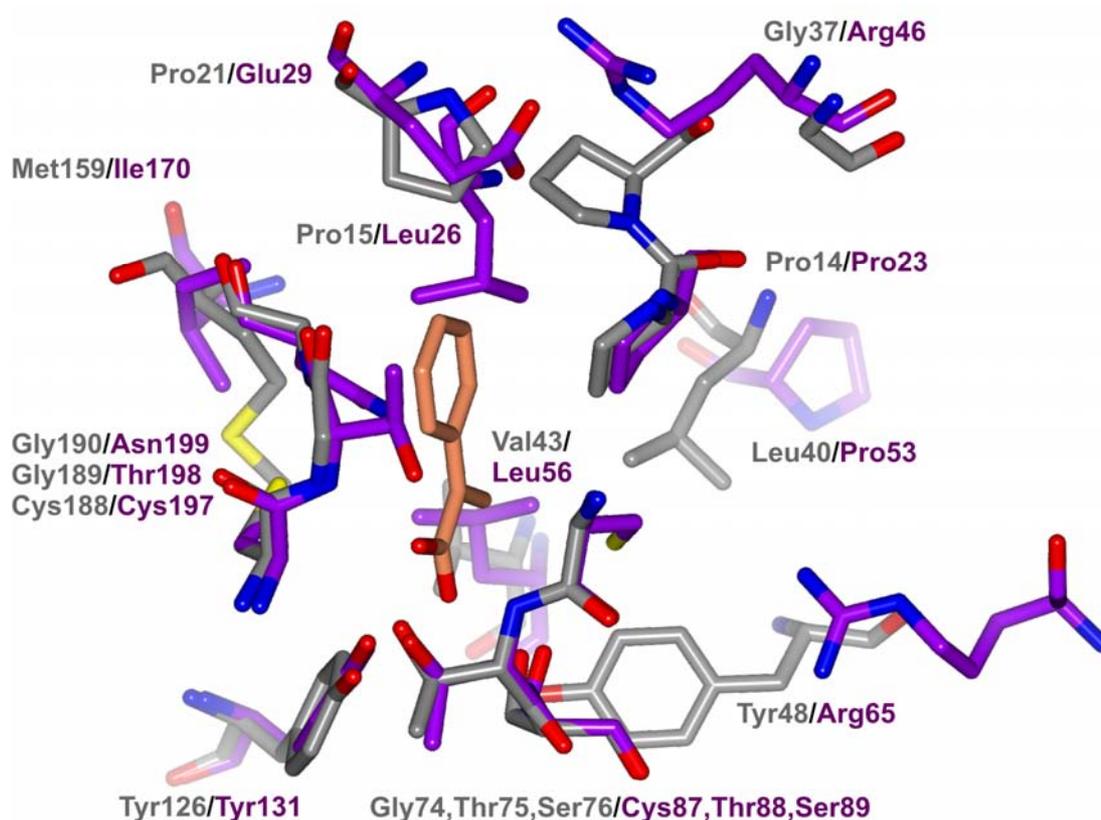


Figure 3.7. Alignment of active sites of RseC model and BbAMD structure. The structure of BbAMD (grey, PDB: 3dg9) aligned to the modelled RseC (purple). The carboxyl group of the intermediate of MPM decarboxylation (orange) is aligned to the PO₄. The active site is seen from the bottom.

Table 3.2. Summary of decarboxylating activities of RseC mutants. BbAMD: *Bordetella bronchiseptica* AMD. RseC: *Rhodococcus* sp. BbAMD homologue containing two conserved cysteines.

Enzyme	Whole resting cells
	PM
BbAMD	+
RseC	0
RseC C87G	0
RseC E29P	0
RseC R46G	0
RseC R46M	0
RseC T198G	0
RseC N199G	0
RseC E29P/R46G	0
RseC E29P/R46M	0
RseC T198G/N199G	0

+ Performed successfully/activity detected
0 No activity detected

3.3.5 Screening for Alkylmalonate Decarboxylation

The ultimate aim of the work on the homologues was to find enzymes active on alkylmalonates. BbAMD and the two homologues, RseC and Pfe, were tested on the alkylmalonate substrates MPM, MEM, MPrM, EBM and DMCHDC. Purified enzyme preparations were used for the assay and the reactions were incubated for several days.

As reported earlier, BbAMD was able to completely convert MPM into (*R*)-PP without producing any detectable (*S*)-PP (76). Conversely, none of the alkylmalonate substrates was converted by BbAMD. The two homologues RseC and Pfe converted neither MPM nor any of the alkylmalonates (data shown for BbAMD and RseC, Fig. 3.8).

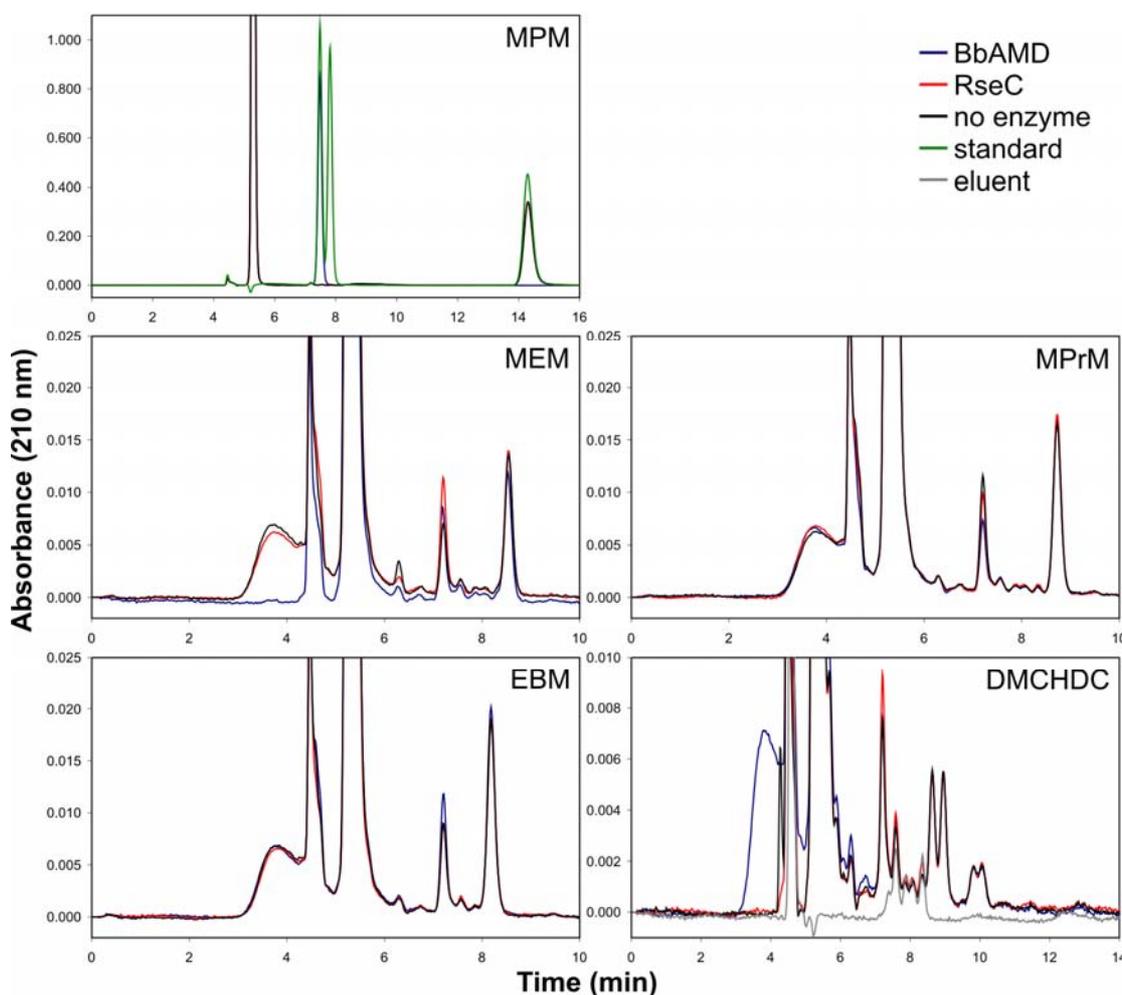


Figure 3.8. Decarboxylation assays of malonate derivatives. 15 mM substrate was incubated for 3.5 d at 25 °C at pH 8.5. Products were extracted with ethyl acetate and 10 μ L were analysed by HPLC. Standard was injected at 20 mM. RTs are 7.5 min for (*R*)-PP, 7.8 min for (*S*)-PP, 14.3 min for MPM, 8.5 min for MEM, 8.7 min for MPrM, 8.2 min for EBM, 5.3 min for ethyl acetate, 7.2 min for an unknown contaminant. The four diastereoisomers of DMCHDC all separated and had RTs of 8.6 min, 9.0 min, 9.9 min, 10.0 min.

3.3.6 Alternative Activities of Pfe

Since Pfe did not convert any of the aryl- or alkylmalonates tested, a series of assays was performed to try to determine its activity. The sequence annotation provided by the UniProt database produced by an automated bioinformatics approach indicated that Pfe was a maleate *cis-trans* isomerase. This was tested using the spectrophotometric approach measuring conversion of maleate to fumarate at 290 nm (see Chapter 4 for method). The enzyme was incubated with 20 mM maleate, 2 mM DTT and 1 mM BME. The thiols are able to isomerase the double bond chemically at a low rate but the

enzyme is expected to convert the substrate at a much higher rate (100). Nevertheless, no activity could be observed above background in this assay (Fig. 3.9, top left). The assays were also performed for 16h and at 50 and 95 °C without any increase in absorbance above the negative control background (data not shown).

It was shown that it is possible to introduce (*R/S*)-PP racemase activity into BbAMD by introducing a single mutation G76C (79). Although this enzyme has a very low racemase activity it exemplifies the mechanistic flexibility of the superfamily. The capability Pfe to racemise PP was tested with 15 mM of either (*R*)- or (*S*)-PP. After 24 h of incubation with the enzyme, the samples were tested using chiral HPLC. Also in this case no trace of the respective opposite enantiomer could be detected in the assays (Fig. 3.9, top right). No racemisation could be observed in RseC either (data not shown).

As Pfe is a member of the Asp/Glu racemase superfamily, its activity towards L-aspartate and L-glutamate was tested using CD. Reactions with 10 mM of the amino acids were followed directly for some time and then incubated over night. The CD of a racemate is 0 mdeg and the observed signal is expected to disappear upon racemisation of the amino acids. No such disappearance could be observed directly or over night. In contrary, there was a small increase in the CD after 5 min of incubation, which was correlated to an increase of the sensitivity of the detector and is thus indicating a change in absorbance. Also after the overnight incubation only a small increase of the same magnitude could be observed, which cannot be explained at present (Fig. 3.9, bottom). The same assay was performed overnight at 95 °C resulting in no detectable racemisation (data not shown). BbAMD and RseC were also tested in overnight incubations with the identical negative result (data not shown).

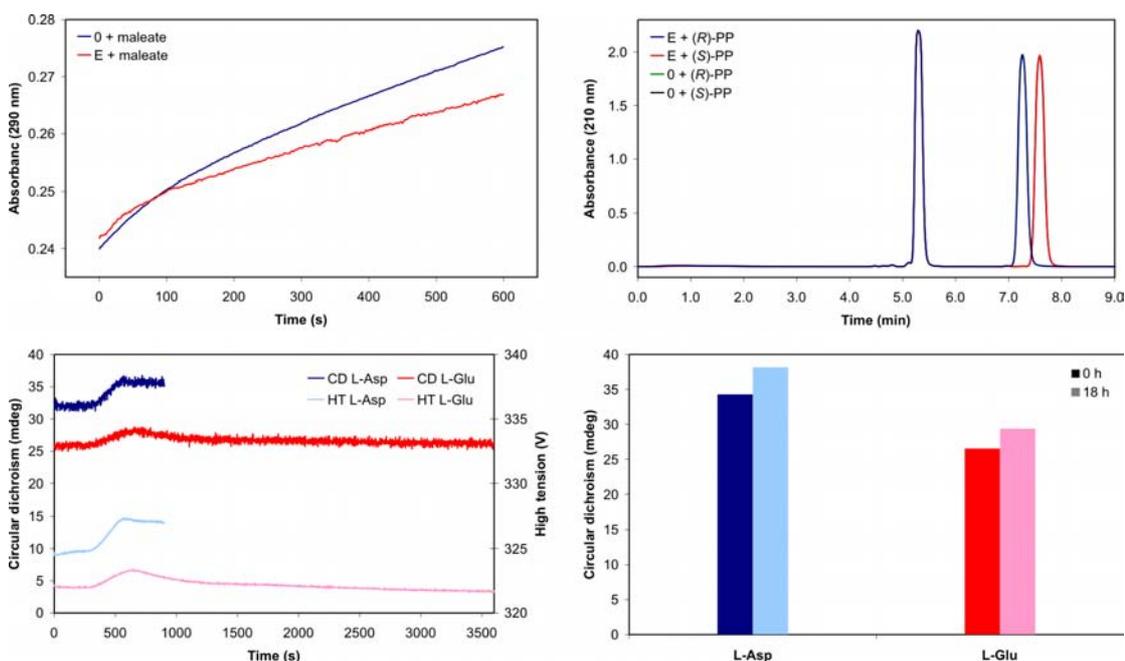


Figure 3.9. Assays for alternative activities of Pfe. Top left: Spectrophotometric assay of maleate *cis-trans* isomerase activity. Purified Pfe was incubated with 20 mM maleate at 25 °C at pH 8.5. Top right: HPLC assay of PP racemase activity. Purified Pfe was incubated with 15 mM (*R*)-PP or (*S*)-PP for 24 h at 25 °C at pH 8.5. RTs are 7.3 min for (*R*)-PP, 7.6 min for (*S*)-PP and 5.3 min for ethyl acetate. Bottom: CD assay for Asp/Glu racemase activity. Purified Pfe was incubated with 10 mM L-amino acid at 25 °C at pH 8.0 including 200 μ M DTT. CD and high tension (sensitivity, roughly inversely proportional to absorbance) were measured continuously (bottom left) and after an overnight incubation (single experiment, bottom right).

3.3.7 Crystallisation Trials with Pfe

In order to understand the significance of the sequence similarities and to explain the absence of any known Asp/Glu racemase activity in Pfe, structural data are required. Crystallisation trials with purified Pfe were performed with the aim of resolving its structure by X-ray crystallography. Pfe was a good candidate as proteins from thermophiles are generally more stable due to their compact fold resulting in a higher probability for it to crystallise. Crystallisation screens and optimisation were performed as described later (see methods of Chapter 5).

Crystals were obtained from Pfe; however, the best ones were still highly amorphous probably, resulting from multiple crystals (Fig. 3.10). The conditions used were: 5 mM of each of NaH_2PO_4 , Na_2HPO_4 , KH_2PO_4 and K_2HPO_4 with 100 mM BIS-TRIS Propane and pH 6.5, 20% (w/v) PEG 1500. Seeding into conditions containing less precipitant

(15% (w/v) PEG 1500 resulted in more structured plates according to the visual aspect (Fig. 3.10). None of the crystals diffracted on the in-house X-ray facility.



Figure 3.10. Crystals of Pfe. Initial Crystals in 20% PEG 1500 (left) were used for seeding in 15% PEG 1500 to produce new crystals (right).

3.4 Discussion

It was possible to clone and express all ten homologues of BbAMD in *E. coli* with at least a fraction as soluble proteins. Despite the automated annotation of most of the homologues as hypothetical AMDs or hypothetical decarboxylases and despite the close clustering of the homologues together with the known decarboxylases, no decarboxylation activity could be detected in any of them (Table 3.3).

Table 3.3. Summary of decarboxylating activities of homologues.

Enzyme	Cloned	Expressed	Purified	Whole cells		Lysate	Purified enzyme					
				growing PM	resting PM		PM	MPM	MEM	MPrM	EBM	DMCHDC
BbAMD	+	+	+	+	+	+	+	+	0	0	0	0
RseC	+	+	+	0	0	0	0	0	0	0	0	0
Pfe	+	+	+	0	0	0	0	0	0	0	0	0
RseD	+	+		0	0	0						
SceD	+	+		0	0							
NfMI	+	+	+		0	0						
Sae	+	+			0	0						
Sme	+	+			0	0						
SceC	+	+			0	0						
Mle	+	+			0	0						
Ste	+	+			0	0						

+ Performed successfully/activity detected

0 No activity detected

It is generally difficult to conclude that an enzyme is not active towards a substrate as there are many variables that have to be considered. On the other hand a lot of arguments support this conclusion. The positive control of BbAMD was always included and was consistently active in all tested conditions. The proteins have all been tested in whole cell (as well as other) conditions where they can be expected to be properly folded and in physiological conditions (apart from their concentration). In addition, the enzymes were of prokaryotic origin (mostly from soil bacteria) where they are generally not posttranslationally modified. Eukaryotic proteins are often activated or inactivated by modifications such as glycosylation, phosphorylation or lipoylations. The absence of the requirement for posttranslational modifications makes it more likely that the BbAMD homologues are in the active form when overexpressed in *E. coli*. None of the described enzymes in the Asp/Glu racemase superfamily is known to need cofactors or any special conditions to be active, apart from requiring being in a reduced state, which is normally the case in the cytoplasm. Also higher temperatures for the enzymes cloned from thermophilic species did not result in decarboxylation activity. It is therefore reasonable to conclude that the homologues are not AMDs.

This result seems to contradict findings that other homologues with similar sequence identities and even the ApGluR would have residual PM decarboxylation activities (48). This could be explained by different reaction conditions or higher concentrations of enzyme. It has to be noted, however, that attempts to replicate the published assay have failed. Both the initial and the late signal measured could not be correlated to the result obtained from HPLC and no solid explanation was available for these signals (for further details see methods).

The attempts to engineer decarboxylase activity in the RseC mutant by simply mutating key residues to match those found in BbAMD failed. Even though the correct folding of the proteins would have to be tested, it is another indication that the activity of the homologues is more different from AMD than the automated annotation suggests.

The homologues do not seem to have any of the other activities found within the superfamily. This suggests that the branches containing the homologues in the phylogenetic tree have different activities from AMDs and MIs that have yet to be determined. Although, it has to be noted that the data obtained from Pfe are preliminary and difficult to establish for a thermophile species. It further has to be noted that Pfe was not tested for HydR activity.

Considering the phylogenetic data (Figure 1.21, Table 3.1) it is possible that there are two different activities corresponding to the homologues with two catalytic cysteines and the ones with a catalytic cysteine and an aspartate. To find the unknown activities of the homologues is a very difficult task that exceeds the scope of this work. Even if structural information of Pfe were available it would not necessarily help to find the activity. For example Ste has a known structure exhibiting an unknown function. The physiological role of AMDs is not known either. There are ectoine utilisation proteins (EutA) in the same superfamily which might suggest that some homologues are used in the catabolism of ectoine (122).

Most of the source organisms of the homologues have been sequenced, which allows the surrounding genetic environment to be analysed. If they lie on an operon with known genes one can speculate as to their natural role. The rhizobial homologues Sme and Mle both lie on an operon with many potential amino acid transporters, a potential threonine dehydratase and a potential ornithine deaminase. These homologues could therefore be involved in the catabolism of amino acids. For the other homologues no clear function can be found for the operon that contains them. It is interesting, however, that all three homologues with the potentially catalytic Asp on domain 2 (RseD, SceD, Sae) are just downstream of their two cysteine paralogue. In addition, as HydRs and MIs are both enzymes involved in the catabolism of natural compounds, a similar role is likely for AMDs and the unknown homologue activities. Another possibility is that these enzymes do not have a biochemical role but are merely products of pseudogenes ready to be recruited for new enzymatic tasks. However, this is unlikely as prokaryotes do not tend to maintain pseudogenes.

BbAMD was successfully confirmed to be an arylmalonate decarboxylase. Its preferred substrate is PM but it also readily accepts MPM and various other arylmalonates as well as alkenylmalonates (51,67). Alkylmalonates, however, were not accepted by BbAMD as was confirmed by experiments conducted by Dr. Krzyszof Okrasa (personal communication). The nature of the substituents on carbon C2 of the malonates is, therefore, crucial for the activity. It is not surprising that the homologues tested unsuccessfully with arylmalonates are not active against alkylmalonates either. This further supports the hypothesis that the electron withdrawing capability of the substrate is essential for the stabilisation of the enediolate intermediate formed by the decarboxylation (67). It would be interesting to know whether other electron withdrawing groups, such as carboxyl or halo groups, would have similar effects.

Chapter 4

Biochemical Characterisation of NfMI

4.1 Introduction

Maleate is a metabolic intermediate appearing in the breakdown of nicotinic acid by certain bacteria, such as *Pseudomonas fluorescens* N-9. It forms after oxidative ring breakage of 2,5-dihydropyridine and subsequent deamidation of maleamate. Maleate *cis-trans* isomerase converts maleate into fumarate, which in turn is used in the general metabolism of the cell (Fig. 4.1). The responsible enzymes are induced when nicotinic acid is supplied as the sole substrate for bacterial growth (97).

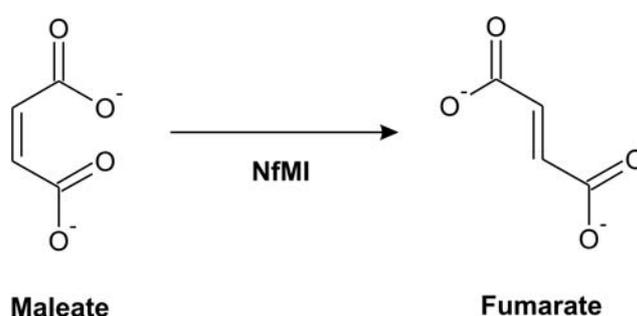


Figure 4.1. Activity of NfMI.

A range of maleate isomerases has been purified from various other bacterial genera such as *Arthrobacter*, *Bacillus*, *Alcaligenes* and *Serratia* and characterised biochemically (98-102). Only three characterised maleate isomerases genes have been sequenced to date including maleate isomerases from *Alcaligenes faecalis* MI, *Bacillus stearothermophilus* and *Serratia marcescens* (52,103,104).

All described maleate isomerases are specific to maleate as substrate. Various other similar compounds have been tested unsuccessfully on maleate isomerases including maleamide, 2,3-dimethylmaleate, *cis*-3-chloroacrylate, citraconate, chloro- and bromomaleate as well as various of their *trans* forms. Fumarate can be a substrate but,

due to the thermodynamic equilibrium of the isomerisation maleate, was only ever detected at amounts lower than 0.2% when the sensitivity of the methods permitted (52,98,100,101).

Most of the maleate isomerases discovered were not dependent on any cofactors. Nevertheless the activity was in some cases reported to be dependent on activation by thiol compounds such as glutathione, cysteine, thioglycerol, DTT or BME. This was concluded as dialysis was inactivating the enzymes and the addition of thiols restored the activity (99,100). Heavy metal ions such as Hg^{2+} and Cu^{2+} and oxidising reagents such H_2O_2 , p-chloromercuribenzoate and NaIO_4 were inhibitors (99,101,103). No competitive inhibitor has been identified so far.

Cysteines were identified to be catalytic residues by mutational studies on *B. stearothermophilus* maleate isomerase. Loss of activity occurred after replacement of cysteines 80 and 198 with serines (103). The activating thiols are thus thought to act through the reduction of the cysteines in the active site. However, until today no three dimensional structure of a maleate isomerase has been published; all structural knowledge originates from superfamily homologues.

All identified maleate isomerases are 25 to 30 kDa proteins. They have pH optima of 8.0 to 8.5 and are active in the range from pH 6.5 to 9.5. *B. stearothermophilus* maleate isomerase has been reported to be stable up to a temperature of 55 °C (104). Kinetic constants of *Alcaligenesa faecalis* maleate isomerase are K_M of 40 μM and k_{cat} of 37 s^{-1} . Compared to other members of the Asp/Glu racemase superfamily this is a rather high activity.

Understanding the biochemical characteristics and eventually the reaction mechanism of NfMI will allow us to understand how the superfamily recruits its fold to perform greatly divergent reactions. This will lead to a better understanding of all superfamily members allowing targeted engineering approaches of the substrate specificities in order to obtain industrially useful activities.

In this chapter the identification of NfMI as a maleate *cis-trans* isomerase will be presented and its biochemical properties will be characterised. Work on the newly discovered competitive inhibitor bromomaleate as well as work on active site mutants will also be presented.

4.2 Methods

4.2.1 Cloning and Protein Purification of NfMI

For cloning and protein expression of wild type and mutant NfMIs as well as for purification of NfMI wild type, see Chapter 2.

Mutants of NfMI were purified in parallel batch purification using Ni-affinity resin. Cell cultures of 50 mL containing overexpressed NfMI mutant proteins were spun down, resuspended in 2.5 mL BugBuster (Novagen) containing 200 μM PMSF, 5 $\mu\text{g mL}^{-1}$ DNase I, 5 $\mu\text{g mL}^{-1}$ hen egg white lysozyme, 1 mM MgCl_2 and 5 mM DTT, and incubated for 15 min at 4 °C shaking. The lysates were diluted with 15 mL buffer (Chapter 2) and added to 1 mL of His-select resin (Sigma). The resin samples with bound protein were washed in three consecutive steps with 15 mL buffer containing 0, 10, 20 mM imidazole and finally eluted with 15 mL 50 mM imidazole. Imidazole was removed by concentrating the sample to 500 μL and diluting in 15 mL buffer without imidazole four times; diluting first with 5 mM DTT, then without reducing agent and finally with 5 mM TCEP. After the final concentration step, the protein sample concentrations were adjusted to 0.1 mg mL^{-1} as determined by the Bradford assay (Chapter 2), the purity analysed by SDS-PAGE (Fig. 4.2) and snap frozen in liquid nitrogen.

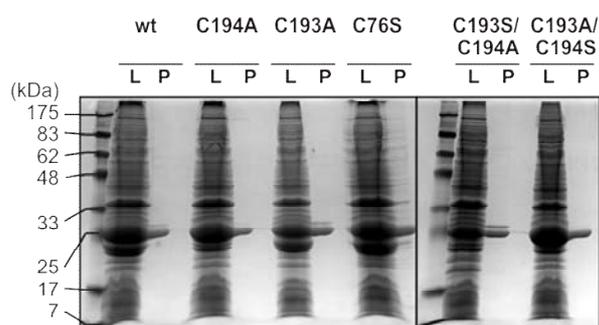


Figure 4.2. Expression and purification of NfMI mutants. SDS-PAGE of lysate (L) and purified protein (P).

4.2.2 Spectrophotometric Maleate Isomerisation Assay

Unless indicated otherwise, all reactions were performed as follows. 10 mM maleic acid was incubated with 20 $\mu\text{g mL}^{-1}$ purified enzyme at room temperature (~ 25 °C) without

shaking. Assays were buffered in 50 mM HEPES at pH 7.5 adjusted with sodium hydroxide and water was added to 1 mL. Reactions were started by the addition of enzyme and stopped, if necessary, by adding 50 μ L of concentrated hydrochloric acid. The formation of fumarate was followed at 290 nm on a Cary 50 UV-Vis spectrophotometer (Varian). Despite a similar shape of the absorbance spectrum fumarate has a higher extinction coefficient than maleate (Fig. 4.3) (97,100). Unless indicated differently assays were performed in triplicate.

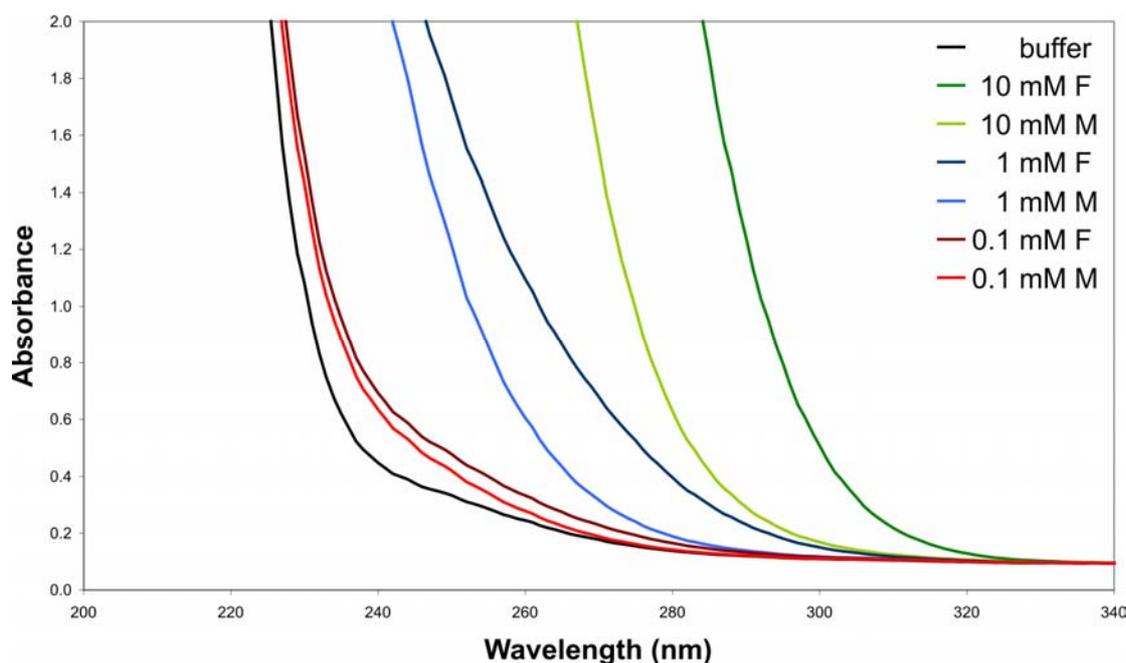


Figure 4.3. Absorbance spectra of maleate (M) and fumarate (F). Spectra were recorded in 50 μ M HEPES pH 7.5.

4.2.3 HPLC Maleate Isomerisation Assay

Unless indicated otherwise, all reactions were performed as follows: 30 μ M maleic acid was incubated with 140 ng mL^{-1} purified NfMI at 30 $^{\circ}$ C for 10 min without shaking. Assays were buffered in 50 mM HEPES at pH 7.5, adjusted with sodium hydroxide and final reaction volumes of 1 mL were attained by adding water. Reactions were started by the addition of enzyme and stopped by adding 50 μ L of concentrated hydrochloric acid. 50 μ L methanol were also added to match the solvent content required for HPLC analysis (52). For long term storage samples were kept at -80 $^{\circ}$ C until analysis. Unless indicated otherwise, assays were performed in triplicate.

10 μL of the samples were injected using a 717plus Autosampler (Waters) into a reverse phase Onyx Monolithic C18, 100 x 4.8 mm column (Phenomenex) and run with a solvent mixture of 25 mM KH_2PO_4 pH 2.5 / methanol 95:5% at a flowrate of 1 mL min^{-1} on a 510 pump (Waters). The compounds were detected by absorbance on a 2487 detector (Waters). Expected RTs were maleate 2.1 min, fumarate 2.6 min (Fig. 4.4), bromomaleic acid 3.0 (contaminating peaks at 2.1 and 2.6 min), citraconate 3.4 min, maleamine 1.8 min (small contaminating peaks 2.0, 2.4 and 2.6 min).

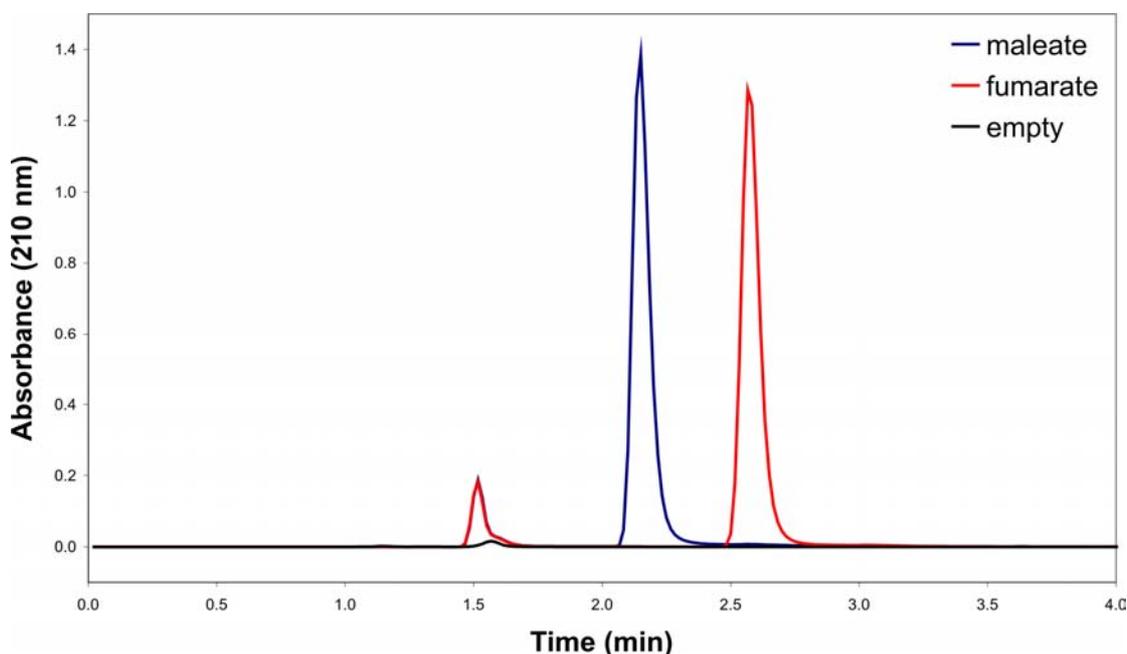


Figure 4.4. HPLC chromatograms of maleate and fumarate. 10 μL of 1 mM samples were injected into a 100 mm monolithic C18 column. As eluent 5% methanol at 1 mL min^{-1} flow rate was used. The samples were detected optically at 210 nm wavelength.

4.2.4 Determination of Kinetic Constants and Inhibition

Constants

HPLC maleate isomerase assays were performed for the determination of kinetic constants. In this assay the following enzyme concentrations were used: wild type NfMI 14 ng mL^{-1} , mutant NfMI 140 ng mL^{-1} for high activity mutants and 4 $\mu\text{g mL}^{-1}$ for low activity mutants. Seven different substrate concentrations (five for inhibition and mutant activity studies) were evenly distributed between 1 and 200 μM . Samples from the same reaction were taken at 5 min and at 10 min and the difference of conversion was

taken as an approximation of initial rates. The conversion rates were below 10%. The experiment was performed in triplicate for wild type NfMI and single experiments were performed for each mutant NfMI. The rate between the two points was plotted and a simple rectangular hyperbola ($y = k_{cat} \cdot x / (K_M + x)$) was fitted using the regression function of SigmaPlot 10.0 (Systat Software). The values were averaged over three (only one for the mutants) experiments. The K_i was determined as $K_i = c_i / (K_M / K_M - 1)$, where “ c_i ” is the concentration of the inhibitor, K_{Mi} the K_M at that inhibitor concentration. The average K_i of four different inhibitor concentrations was averaged. For single each single hyperbolic curve the standard errors were calculated with SigmaPlot 10.0 (Systat Software).

4.2.5 GC/HPLC Enoate Reductase

Two enoate reductases were used, provided as *Escherichia coli* cell extracts by researchers at BASF (Ludwigshafen). The enoate reductase, OYE2 from *Saccharomyces cerevisiae* (Old Yellow Enzyme 2, EC 1.6.99.1, UniProt: Q03558), is NADPH dependent as is its homologue YqjM from *Bacillus subtilis* (EC 1.6.99.1, uniprot: P54550).

The partially purified FactorX, thought to be a *cis-trans* isomerase, was also provided by BASF. FactorX is one of three activities on methyl-pentenal isolated from *Candida rugosa* (BASF Lu905) crude extract. One alcohol dehydrogenation activity produces methylpentenol and a enoate reduction activity produces methylpentenal. When the enoate reductase activity is purified, the incubation yields an excess of (S)-methyl-pentanal. By replacing FactorX into the purified sample an excess of (R)-methyl-pentanal is produced. Although there is no direct evidence, FactorX is hypothesised to be a methyl-pentenal *cis-trans* isomerase. Homologues present in UniProt (Q5AHZ9 from *Candida albicans* and Q6C6X4 from *Yarrowia lipolytica*) determined by peptide fragment sequencing were cloned and expressed in *E. coli* but did not show any activity (personal communication by Andreas Schädler, BASF).

Assays were either performed with *cis/trans* citral or with maleate derivatives (maleate, fumarate, citraconate, mesaconate, dimethylcitraconate). All reactions were carried out in 100 mM KH_2PO_4 buffer at pH 7, 1 mM of both NAD and NADP were added. Cofactor regeneration was achieved by adding 100 mM glucose and a 1:4 dilution of glucose dehydrogenase from cell extract (BASF Lu11589). Volumes were completed with PBS.

Overexpressed OYE2 and YqjM were added from *E. coli* crude extract to 0.1 mg total protein. FactorX was either added as a semi-purified protein solution to 2.2 mg mL⁻¹ or 1.5 mg mL⁻¹ separated from enoate reductase and alcohol dehydrogenase activities by two purification steps or added as crude extract to 5 mg mL⁻¹ total protein (BASF Lu905) containing these two activities. Semi-purified enzyme was used for assays containing citral and crude extract was used for the maleate assays. NfMI was added to 0.015 mg from purified enzyme at 15 mg mL⁻¹ in 20 mM TRIS, 150 mM NaCl, 5 mM DTT, pH 8.0.

Cis/trans-Citral was added as a 10% (w/v) solution in DMSO to a final concentration of 15.8 mM whereas maleate, fumarate, citraconate, mesaconate or dimethylcitraconate were added as 100 mM solutions (pH adjusted to 5-7) to 10 mM. The reactions were started by the addition of substrate and stopped either by adding 250 μ L chloroform (for GC analysis) or 50 μ L of concentrated HCl (for HPLC analysis). All assays were performed at 30 °C shaking at 1,400 rpm for 0 up to 70 h. For GC analysis the chloroform reactions were performed at 25 °C for 10 min at 1,400 rpm and then centrifuged at 18,000 \times g for 3 min. The chloroform extract was dried using molecular sieve 4 Å for another 10 min. For HPLC analysis the reactions were only centrifuged at 18,000 \times g.

The consumption of *cis/trans* citral and the formation of citral derivatives during enoate reductase assays were analysed by GC. 1 μ L of chloroform extract was injected into a 30 m BGM-174, 0.25 mm ID, 0.25 μ m FD column on GC 02 (BASF) and eluted with He at a flow rate of 1 mL min⁻¹. The compounds were analysed by FID and normalised by the total peak area. Expected RTs were (*S*)-citronellal (32.4 min), (*R*)-citronellal (33.4 min), nerol (36.0 min), (*S*)-citronellol (37.0 min), (*R*)-citronellol (37.2 min), geraniol (41.0 min), *cis*-citral (45.2 min), *trans*-citral (50.0 min).

The interconversion of maleate and fumarate, citraconate and mesaconate as well as the consumption of dimethylcitraconate during the enoate reductase assays were analysed by HPLC. 1 μ L of acidified sample was injected into a reverse phase Aqua C18 100 \times 4.6 mm column (Phenomenex) on LC 05 (BASF) and the compounds were detected by absorbance at 210 nm and the identity of the peaks was verified with the absorbance spectrum. Different methods were used for the different compounds (RTs in brackets): Method CMF-A.M (BASF) for maleate (1.93 min), fumarate (2.34 min), method CMF2-A.M (BASF) for citraconate (2.32 min), mesaconate (2.81 min) and method DICSMS-A.M (BASF) for dimethylcitraconate (2.82 min).

Succinate produced in enoate reductase assays was analysed by HPLC. 5 μ L of acidified sample was injected into an Aminex HPX-87H column on LC 09 (BASF) with the Z9-CMF30.M method (BASF) and compounds were detected by refractive index. Expected RTs were succinate (15.2 min), fumarate (19.6 min) and maleate (15.5 min, signal hidden by lactic acid).

4.2.6 MS of Enzyme-Ligand Adducts

The buffer of NfMI was replaced by dilution in 20 mM ammonium acetate (pH 11) and subsequent concentration with Amicon Ultra 15 centrifuge filters with 10 kDa molecular weight cut off (Millipore, USA). 10 mM of the potential adduct was added to the solution and analysed on a ABI Qstar ESI-TOF mass spectrometer (Applied Biosystems, USA).

4.3 Results

4.3.1 Activity Assays

The conversion of maleate to fumarate can be measured in a number of different ways. In the literature two spectrophotometric methods have been described. First, a direct and easy method is to exploit the higher extinction coefficient of fumarate compared to maleate at wavelengths around 290 nm (97). Unfortunately, the difference in absorbance is below the detection limit at low concentrations of substrate ($< 10 \mu$ M). For high concentrations above 0.1 mM the direct spectrophotometric method represents an easy way to follow the reaction in real time and to quickly determine end points (see methods for absorbance spectra).

A second method has been suggested where an excess of fumarate hydratase (fumarase, EC 4.2.1.2) is used to specifically hydrate the double bond of fumarate and measure the disappearance of total double bond concentration at 240 nm (99). This method was not used in the present work as the slight difference of the extinction coefficients of maleate and fumarate would interfere and the substrate would have to be below 1 mM not to saturate the signal.

A third and new spectrophotometric method would be to use an excess of fumarate hydratase in combination with an excess of L-malate dehydrogenase (EC 1.1.1.37). The reaction could be followed by the formation of NADH, the cofactor of L-malate dehydrogenase, at 340 nm. As NADH has a strong chromophore this is potentially a highly sensitive method; however, the activity of the purchased recombinant enzyme

from *E. coli* (Fluka) in this direction was too low at the pH used in this assay. Consequently, the method was not used this project.

To be able to measure below 10 μM and to avoid interference of other chromophores, reverse phase HPLC on a C18 column was used to measure end points of reactions as already described (52). The signal was linear and could be accurately measured between 1 μM and 1 mM (see methods for chromatograms).

4.3.2 Activity of NfMI

As NfMI sequence consistently clustered with the three other known maleate isomerase sequences in phylogenetic analyses, having sequence identities between 44 and 28%, it was tested for activity with 10 mM maleate using the spectrophotometric assay at 290 nm. A small increase in absorbance could be observed in cell lysate corresponding to the formation of fumarate. The same lysate was used as for the arylmalonate decarboxylation assays described in Chapter 3.

The assay was repeated using purified NfMI (see Chapter 3) and the formation of fumarate was followed over time. The fumarate formation rate was constant for 15 min, then slowed down to reach full completion of the reaction after 20 min resulting in a typical enzyme reaction curve. To confirm the enzymatic origin of the activity, boiled enzyme was used in the same assay. As expected, no maleate formation was observed in this case. Furthermore, as in the literature maleate isomerase from *P. fluorescens* was reported to require thiols as activators (100) a mix of DTT and BME was added to the reaction. As suggested for other MIs, thiols did not increase the activity of NfMI. (Fig. 4.5).

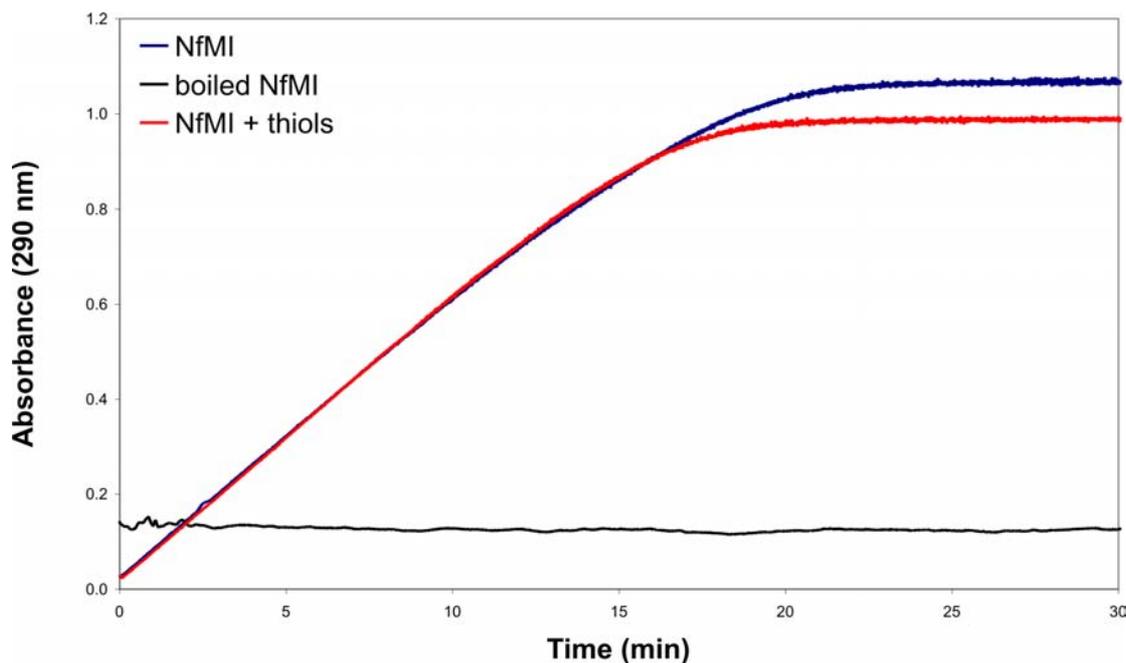


Figure 4.5. Reaction time course of NfMI. Spectrophotometric assay using $20 \mu\text{g mL}^{-1}$ purified NfMI and 10 mM maleate at room temperature and pH 8.5.

4.3.3 Effect of pH on NfMI Activity

To determine the pH optimum of NfMI the amount of fumarate produced from 10 mM maleate after 20 min was measured in three different buffers with pH ranging from 4 to 10 and activity was measured using the spectrophotometric assay at 290 nm. The maximal activity was observed at pH 7.5 being approximately 1 pH point lower than those measured for other known maleate isomerases (52,100,101). At pH 4.5 and 11.0 the activity was almost completely abolished (Fig. 4.6). As a consequence in further experiments the activity was measured at pH 7.5 using HEPES, which has an optimal pK_a .

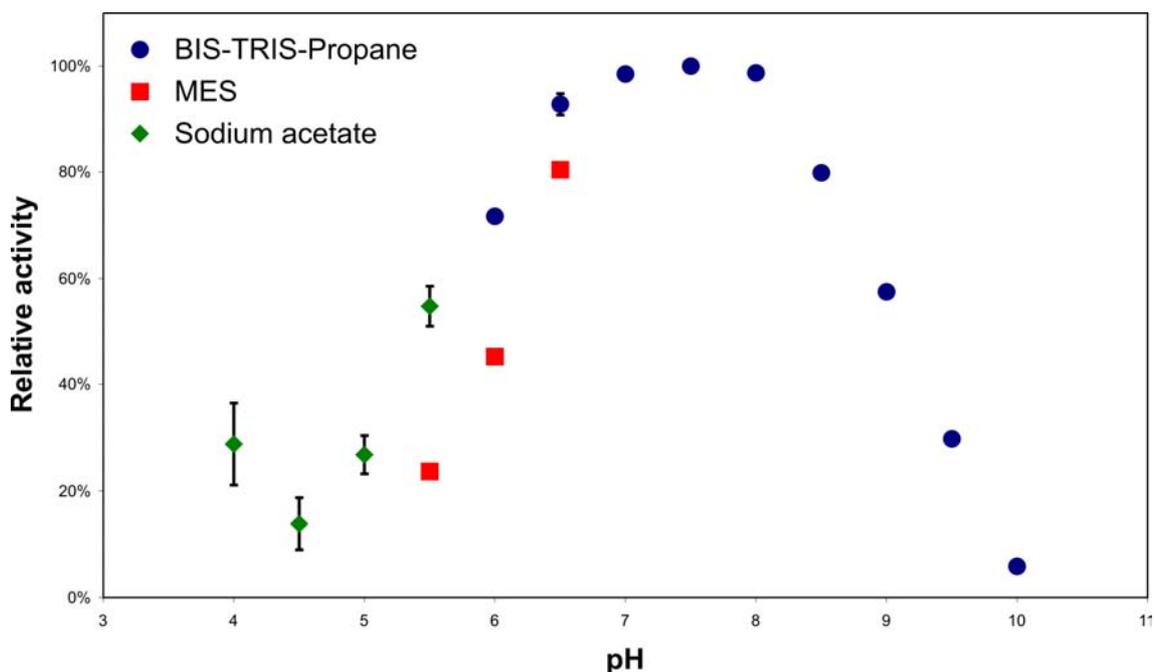


Figure 4.6. Dependence of NfMI activity on pH. Spectrophotometric assay using $20 \mu\text{g mL}^{-1}$ purified NfMI with 10 mM maleate using various buffers. The reaction was performed for 20 min at 37 °C. All experiments were performed in triplicate and the error bars represent standard deviation.

4.3.4 Effect of Temperature and Redox State on NfMI Activity

For further characterisation of NfMI the temperature optimum was determined using the same spectrophotometric end point assay as for pH optimum determination. The enzyme was incubated with 10 mM maleate for 20 min at temperatures ranging from 20 to 60 °C at pH 8.5. Under these conditions the pH optimum was between 35 and 40 °C. At 20 and 50 °C almost no activity was left (Fig. 4.7).

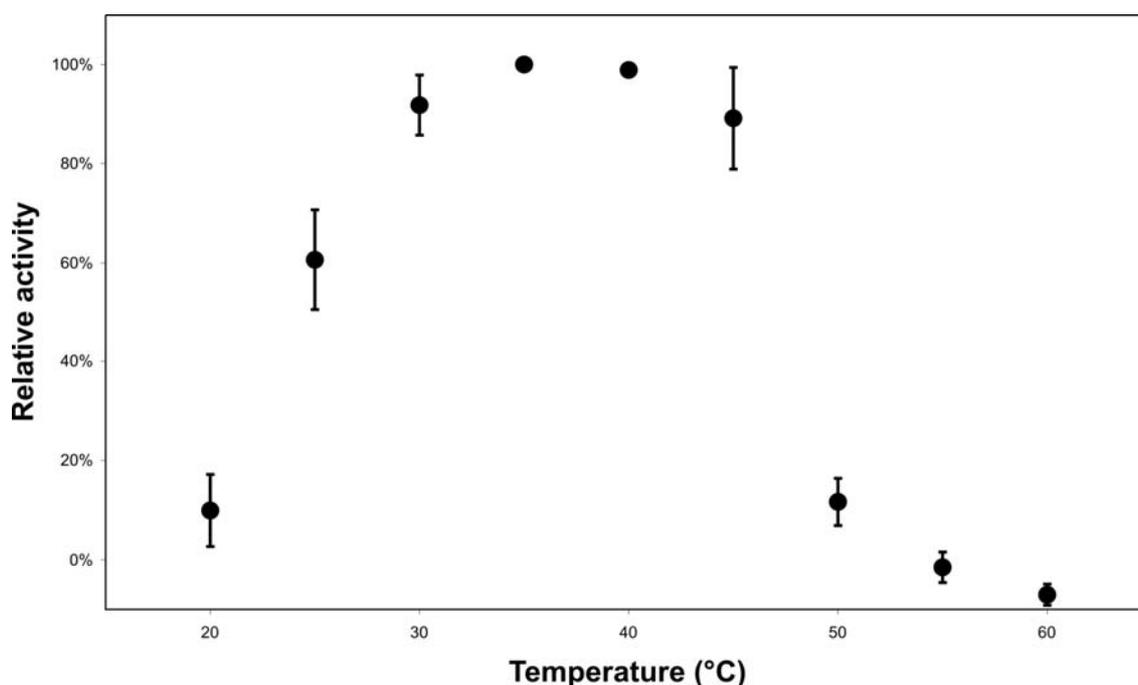


Figure 4.7. Dependence of NfMI activity on temperature. Spectrophotometric assay using $20 \mu\text{g mL}^{-1}$ purified NfMI with 10 mM maleate at pH 8.5. Enzyme stability was not considered. All experiments were performed in triplicate and the error bars represent standard deviation.

As the temperature optimum depends on protein stability and as protein denaturation is a process that is also subject to kinetics, the temperature optimum only applies under precise conditions. For this reason it is also important to know the stability of the enzyme with respect to the temperature. The optimal temperature would then be the highest temperature at which the enzyme is stable over the reaction period.

To determine the enzyme stability NfMI was preincubated for 1 h in the reaction buffer at various temperatures and then assayed with 10 mM maleate at 25 °C for 30 min. The activity was determined using the spectrophotometric assay measuring end points. The data show that the enzyme is stable up to approximately 30 °C and almost completely inactivated at 55 °C under these conditions (Fig. 4.8). Further experiments were therefore performed at 30 °C.

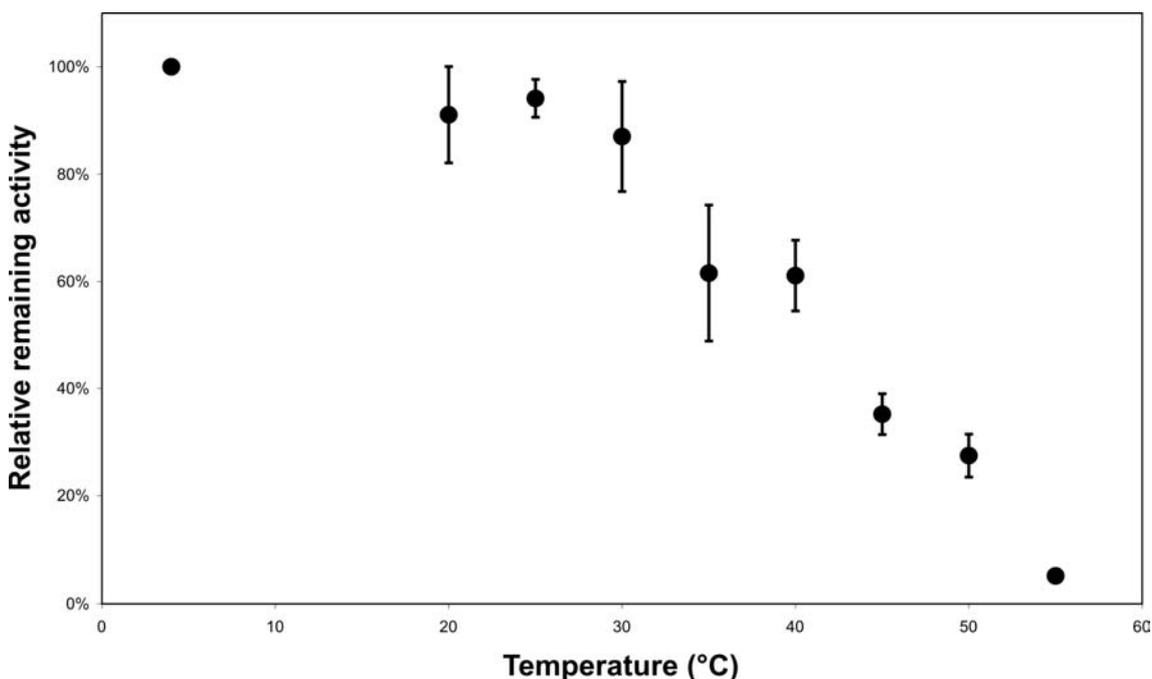


Figure 4.8. Temperature stability of NfMI. Purified NfMI at $20 \mu\text{g mL}^{-1}$ was preincubated 1 h in buffer at pH 7.5. The reaction was then started by adding 10 mM maleate and run for 30 min at 25 °C. The samples were analysed spectrophotometrically. All experiments were performed in triplicate and the error bars represent standard deviation.

As mentioned above, the activity of other maleate isomerases was dependent on thiol activators whereas NfMI is not. It was also reported that maleate isomerases are sensitive to inactivation by oxidation (100,103). Whether thiols act as oxidation protectors was tested in two ways.

The first way was testing whether the temperature instability is due to thermal denaturation of the enzyme or due to increased oxidation at higher temperature. The enzyme was preincubated in the reaction mix for 1 h at various temperatures and subsequently assayed with 10 mM maleate for 30 min. 10 mM DTT was added before the preincubation or before the assay or before the analysis. As DTT interferes with the spectrophotometric assay the end points were analysed with the HPLC.

At 95 °C preincubation, where the enzyme is completely inactivated, there is a non-enzymatic chemical conversion of maleate to fumarate if DTT was present. After accounting for this background, the activity was higher if DTT was present during the reaction. If DTT was already added during the preincubation, the activity was even higher (Fig. 4.9).

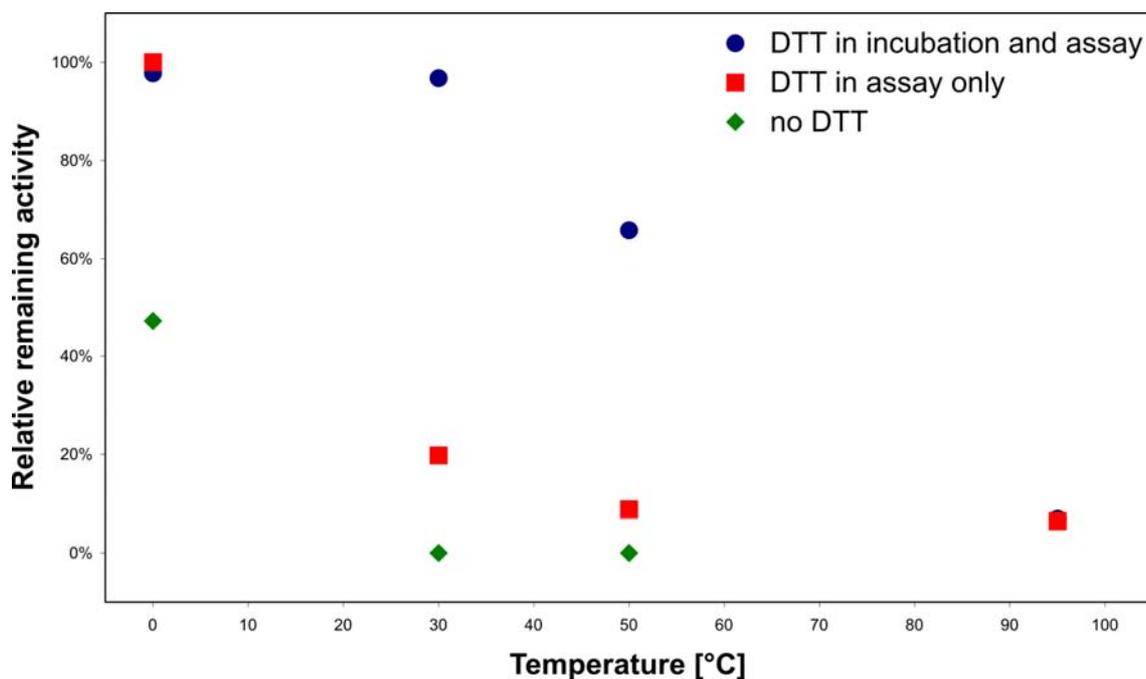


Figure 4.9. Effect of the redox state on temperature stability of NfMI. Purified NfMI at 140 ng mL^{-1} was preincubated 1 h in buffer at pH 7.5 at various temperatures. The incubations were performed by adding 10 mM DTT at different stages and under normal atmosphere. The reaction was started by adding 10 mM maleate and incubated for 30 min at 30 °C. The turnover was measured by HPLC. Single experiments were performed.

The second test of the oxidation protection on thiols was combined with a long term stability assay. Aliquots of concentrated enzyme were incubated with 5 mM DTT in enzyme storage buffer at 4 °C for several days. After incubation reaction mix was added, containing 1 mM maleate and incubated at 30 °C for 10 min. The samples were analysed by HPLC

In both conditions the enzyme lost activity over the course of less than a month. However, the activity of the samples without DTT dropped rapidly below 50% of the initial activity after 1 d and was almost completely inactivated after 14 d. In contrast, the samples containing DTT were more active and dropped below 50% of the initial activity only after 14 d and were inactive after 21 d (Fig. 4.10).

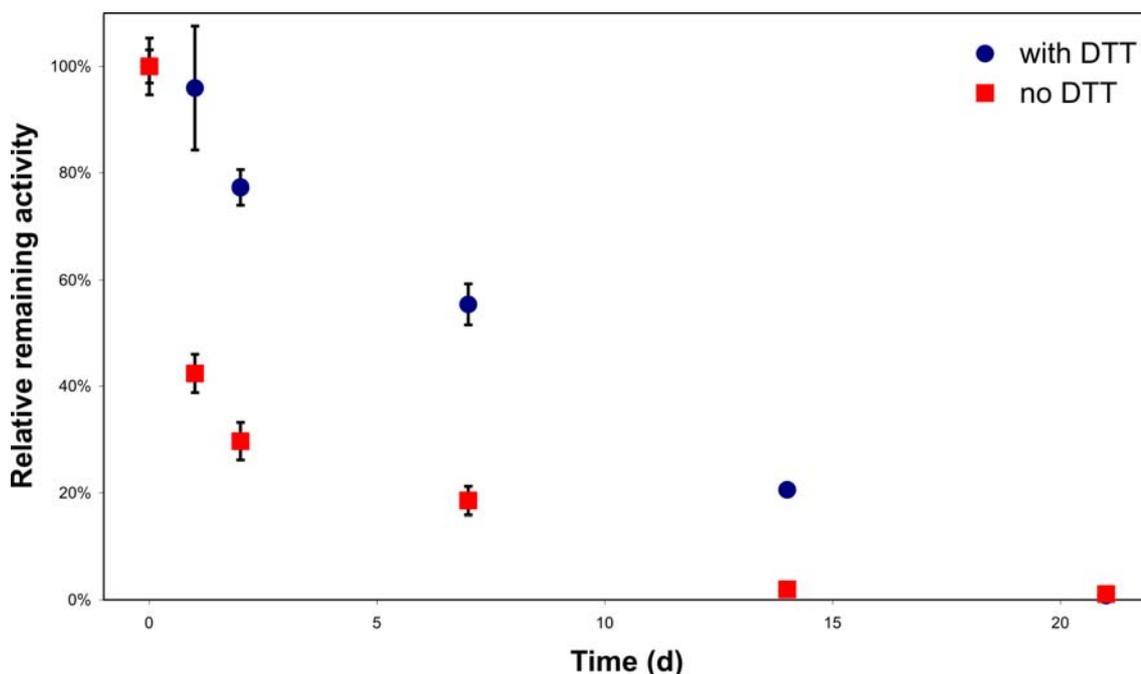


Figure 4.10. Long term stability of NfMI depending on redox state. Purified NfMI at $7 \mu\text{g mL}^{-1}$ was incubated at $4 \text{ }^\circ\text{C}$ for several days with and without 10 mM DTT under normal atmosphere. The reaction was then started by adding the reaction mix containing 1 mM maleate diluting NfMI to 140 ng mL^{-1} and run for 10 min at $30 \text{ }^\circ\text{C}$. The turnover was analysed by HPLC. All experiments were performed in triplicate and the error bars represent standard deviation.

4.3.5 Kinetics and Inhibition of NfMI Activity

The Michaelis-Menten kinetic parameters were determined taking the difference of conversion between 5 and 10 min as an approximation of the initial rate and fitting a square hyperbola onto the data. The assay was performed at $30 \text{ }^\circ\text{C}$ and analysed by HPLC. K_M was $10 \pm 2 \mu\text{M}$ and k_{cat} was $6 \pm 1 \text{ s}^{-1}$. The initial rate was maximal at approximately $100 \mu\text{M}$ and slightly lower at $200 \mu\text{M}$ (data not shown but also visible in Fig. 4.13 and 4.17). This is indicative of substrate inhibition that hitherto was not quantified. As a result k_{cat} is underestimated.

To find competitive inhibitors of NfMI a range of compounds resembling the substrate was identified. These include the product (fumarate), substrate analogues with substituents on the double bond (bromomaleate, citraconate), cyclic non-isomerisable compounds (cyclopentanedione, benzoquinone) and modified carboxylic acids (maleamide, dimethylmaleate) as well as general small carboxylic acids (malonate, succinate, D/L-malate, citrate, formate, acetate) and acidic amino acids (L-aspartate, L-glutamate) (Fig 4.11).

NfMI was incubated with 30 μM maleate well above K_M but close enough to ensure high sensitivity of the assay. The inhibitor was added at 1 μM without DTT and at 60 μM with 5 mM DTT. The conversion was analysed by HPLC.

Most of the compounds tested did not inhibit the activity of NfMI. Only bromomaleate and benzoquinone reduced the conversion rate. When DTT was present, the inhibition of benzoquinone disappeared despite the higher concentration, which is indicative of oxidative inhibition. Of the tested compounds only bromomaleate inhibited NfMI activity when the concentration was increased (Fig. 4.12).

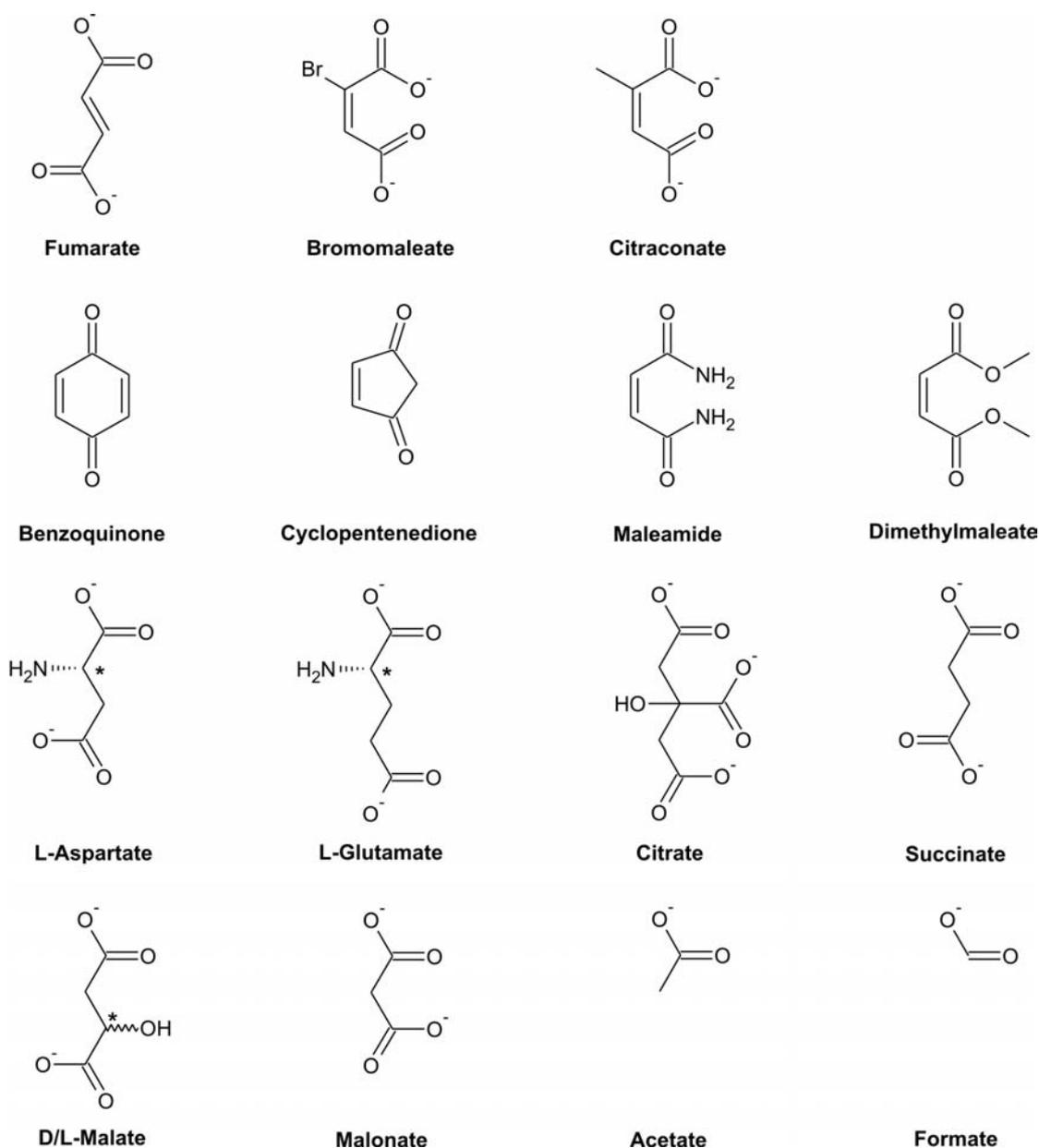


Figure 4.11. Compounds tested for NfMI inhibition.

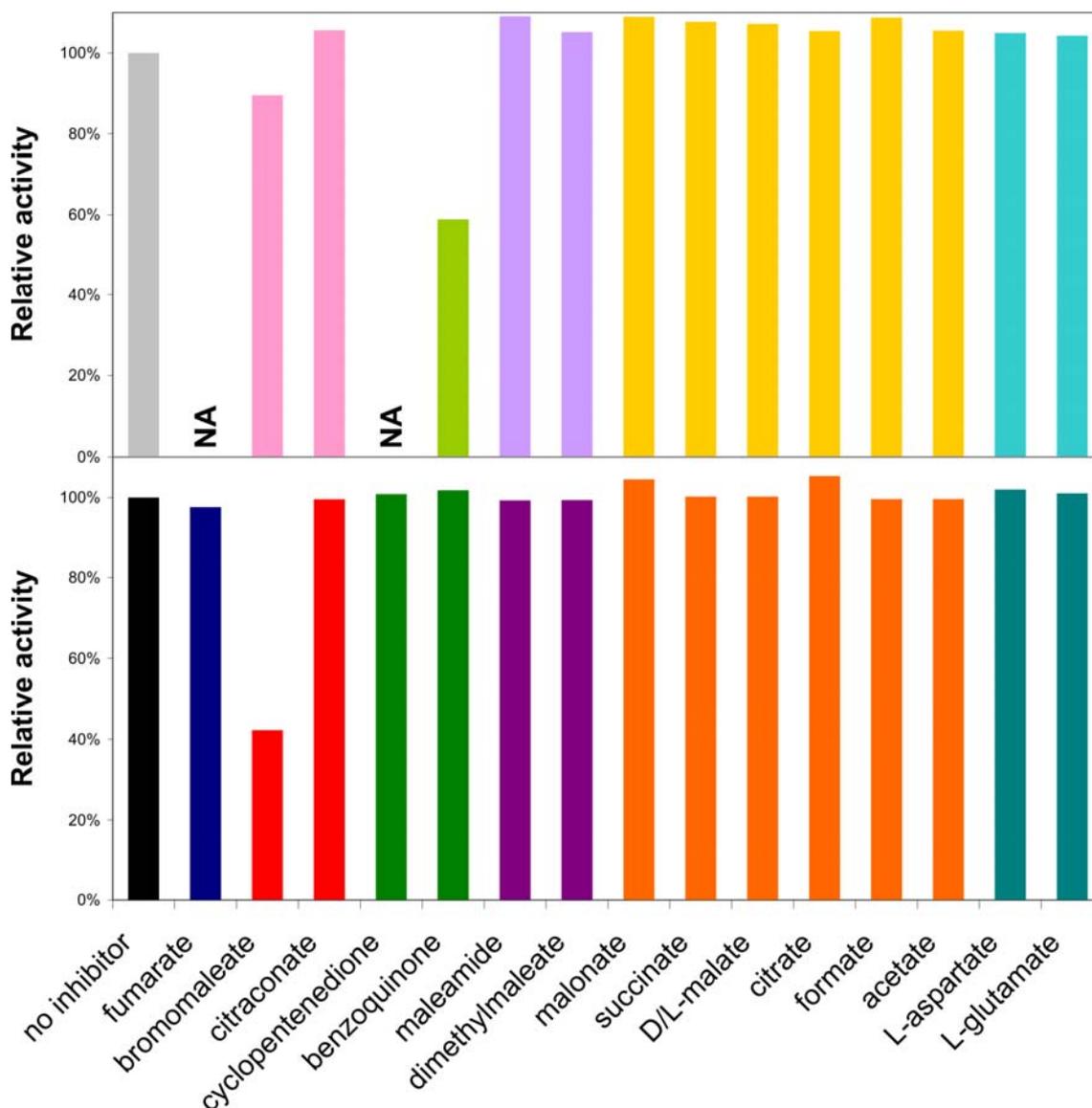


Figure 4.12. Assay of substrate analogues as potential inhibitors of NfMI. Purified NfMI at 140 ng mL^{-1} and $30 \text{ } \mu\text{M}$ maleate were tested with $1 \text{ } \mu\text{M}$ (top) and $60 \text{ } \mu\text{M}$ (bottom) of each compound respectively. The $1 \text{ } \mu\text{M}$ experiments were performed without DTT whereas the $60 \text{ } \mu\text{M}$ experiments contained 5 mM DTT. Substrate analogues with a substituted double bond are shown in red, cyclic compounds in green, ones with derivatised carboxyl groups in purple, general carboxylates in yellow and amino acids in turquoise. NA, not available.

To test whether bromomaleate inhibition was competitive, kinetic parameters were determined for different bromomaleate concentrations. These could be used to calculate the inhibition constant K_i . With increasing concentration of inhibitor the K_M increased as well whereas the k_{cat} remained approximately stable at the same level, which is in accordance with competitive inhibition resulting in a K_i of $0.9 \pm 0.3 \text{ } \mu\text{M}$ (Fig. 4.13, Table 4.1).

To test whether the substrate or the inhibitor bind covalently, MS analysis was performed at pH 11 in the presence of both maleate and bromomaleate. At this pH the enzyme is not active and both cysteines are expected to be deprotonated. A potential nucleophilic attack of one cysteine on the substrate could still occur; however, no covalent adduct was found.

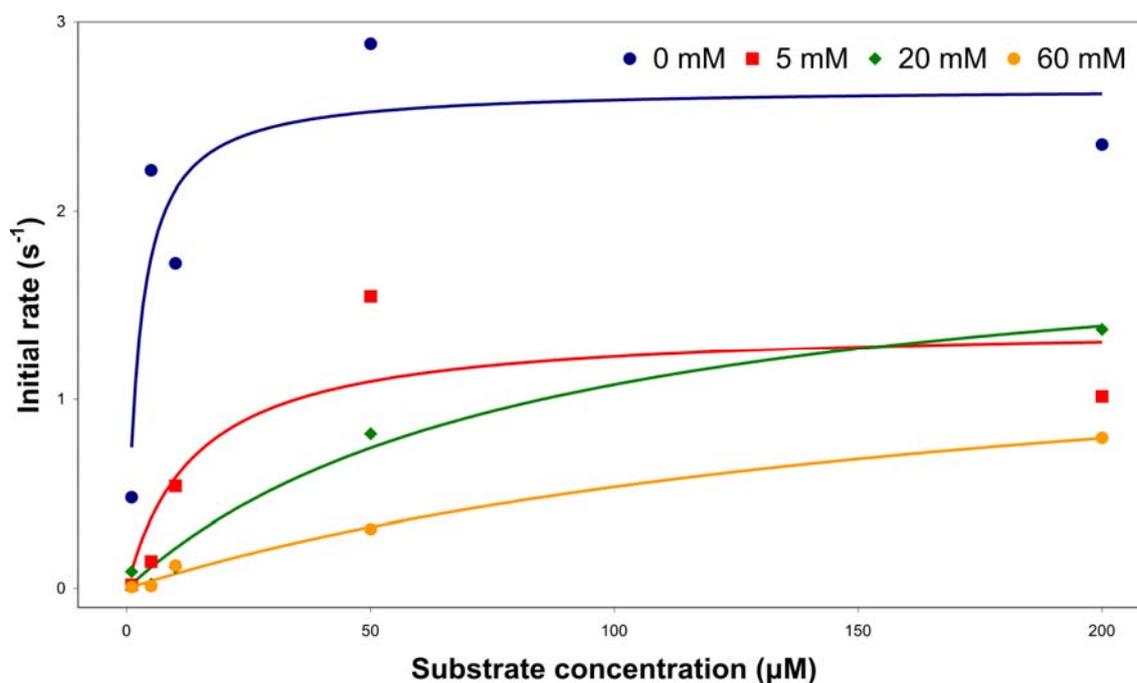


Figure 4.13. Kinetics of bromomaleate inhibition. Purified NfMI at 14 ng mL^{-1} was assayed with different maleate and bromomaleate concentrations for 10 min at $30 \text{ }^\circ\text{C}$. The samples were analysed by HPLC. Simple rectangular hyperbola were ($y = k_{cat} \cdot x / (K_M + x)$) fitted to determine the inhibition constant (see table 4.1).

Table 4.1. Determination of the K_i of bromomaleate. Data see Fig. 4.13. The K_i was determined using the following equation: $K_i = c_i / (K_{Mi} / K_M - 1)$. Standard errors and mean of standard errors are indicated.

Bromomaleate concentration (μM)	K_M (μM)	k_{cat} (s^{-1})	K_i (μM)
0	2.6 ± 1.7	2.7 ± 0.3	NA
5	14 ± 13	1.4 ± 0.4	1.1 ± 1.1
20	82 ± 32	2.0 ± 0.3	0.6 ± 0.2
60	186 ± 57	1.5 ± 0.3	0.8 ± 0.3
mean	NA	1.9 ± 0.3	0.9 ± 0.5

NA = not applicable

4.3.6 Alternative Substrates

To extend the range of known substrates from the single one known to date, various compounds were screened for activity. For that purpose, all tested inhibitors containing an isomerisable double bond (fumarate, bromomaleate, citraconate, maleamide and dimethylmaleate) were incubated at 1 mM at 1h (Fig 4.14). The samples were analysed by HPLC checking the traces for consumption of the substrate.

None of the reactions showed detectable activity towards any alternative substrate. The maleate control was completely converted into fumarate. A chemical background reaction due to DTT could also be detected. The reverse reaction with fumarate could not be detected as the maleate contamination in the used fumarate sample was higher than the expected amount at the thermodynamic equilibrium. Using the peak areas of the maleate and the fumarate conversion in the presence of the enzyme, the equilibrium constant was determined to be 480. Dimethylmaleate did not elute within the analysed elution time (Fig. 4.15).

Furthermore, in an attempt to combine NfMI with enonate reductases (OYE2, YqjM) to change the reaction specificity (see below), it was also incubated with either 16 mM *cis/trans* citral, 10 mM mesaconate or dimethylcitraconate (Fig. 4.14). None of these substrates was transformed by NfMI. Equally in no case did the enoate reductase produce a different product pattern when incubated in combination with NfMI (Fig. 4.18 and Fig 4.19).

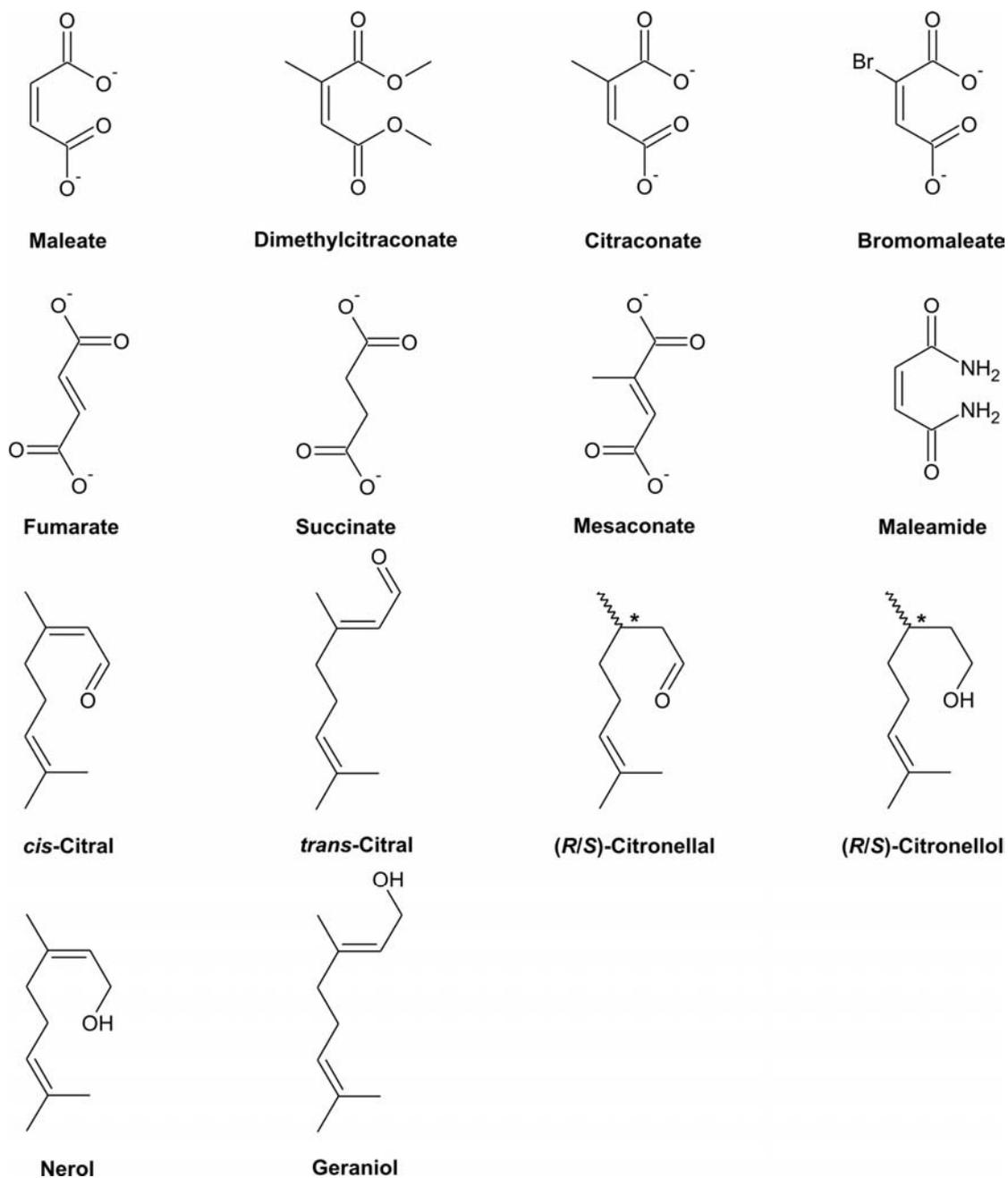


Figure 4.14. Substrates and products used with NfMI and enoate reductases.

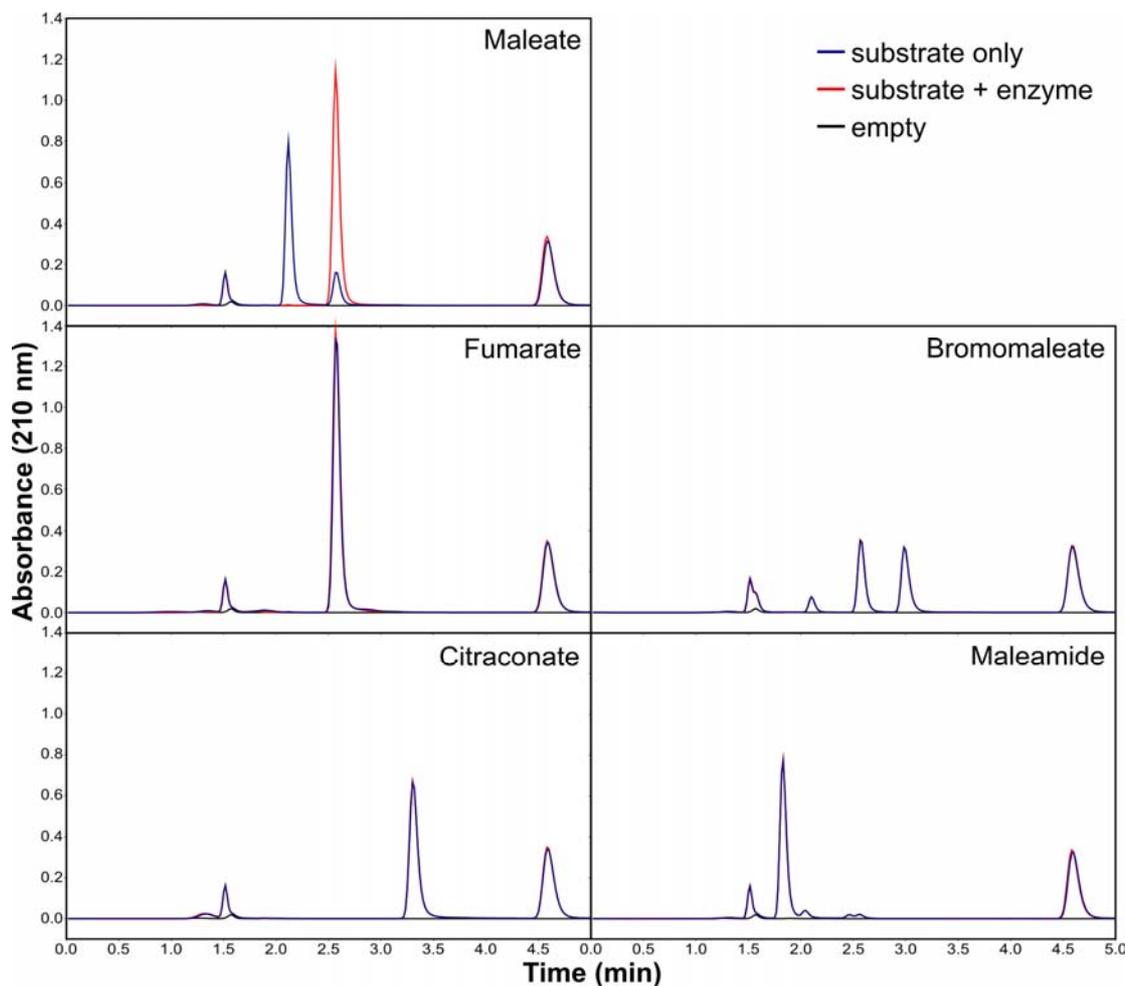


Figure 4.15. Assay of potential alternative substrates of NfMI. Purified NfMI at $2.8 \mu\text{g mL}^{-1}$ and 1 mM substrate were incubated for 1 h at 30 °C together with 5 mM DTT. The samples were analysed by HPLC. The peak at 1.5 min corresponds to HEPES and the peak at 4.6 min corresponds to DTT. Of the three bromomaleate peaks only the one at 3.0 min corresponds to the initial compound.

4.3.7 Active Site Mutants

The cysteine residues in MIs aligning with the catalytic cysteines in racemases of the superfamily were confirmed by serine mutations to be the catalytic residues in a previous study (104). This was expected to be the case for NfMI; however, in an alignment of NfMI with the sequence of the other maleate isomerases there is Cys193 adjacent to the suspected catalytic Cys194, resulting in some uncertainty about the correct attribution of the role of these residues.

To clarify the situation about the correct active site cysteine, the mutants C193A and C194A were produced and tested with 1 mM maleate and the kinetic parameters

determined. The activity of C193A remained largely wild type. Only a slight decrease in k_{cat} accompanied by a three-fold decrease in K_M can be observed in C193A whereas C194A suffers an almost total loss of activity resulting in a 1000-fold reduction in k_{cat} accompanied by a two-fold decrease in K_M (Fig. 4.16, Fig. 4.17 and Table 4.2). This confirms the result of the alignment with Cys194 as the catalytic residue (Appendix C).

Cysteine to serine mutants were used to differentiate between their role as a potential nucleophile and their role as a potential hydrogen bonding partner. Serine generally is a suitable replacement for cysteines if hydrogen bonding is required as it is approximately isosteric but a lot less likely to be a nucleophile as hydroxyl pK_a is much higher than the hydroxyl pK_a of the cysteine thiol. For this reason the mutants C76S, C193S/C194A and C193A/C194S were made and tested with 1 mM maleate and for the determination of kinetic parameters.

The mutants C76S and C193A/C194S resulted in the complete elimination of detectable activity. The mutant C193S/C194A showed very low residual activity of the same magnitude as C194A (1000-fold reduction in k_{cat} and a two-fold decrease in K_M compared to wt, Fig. 4.16, Fig. 4.17 and Table 4.2).

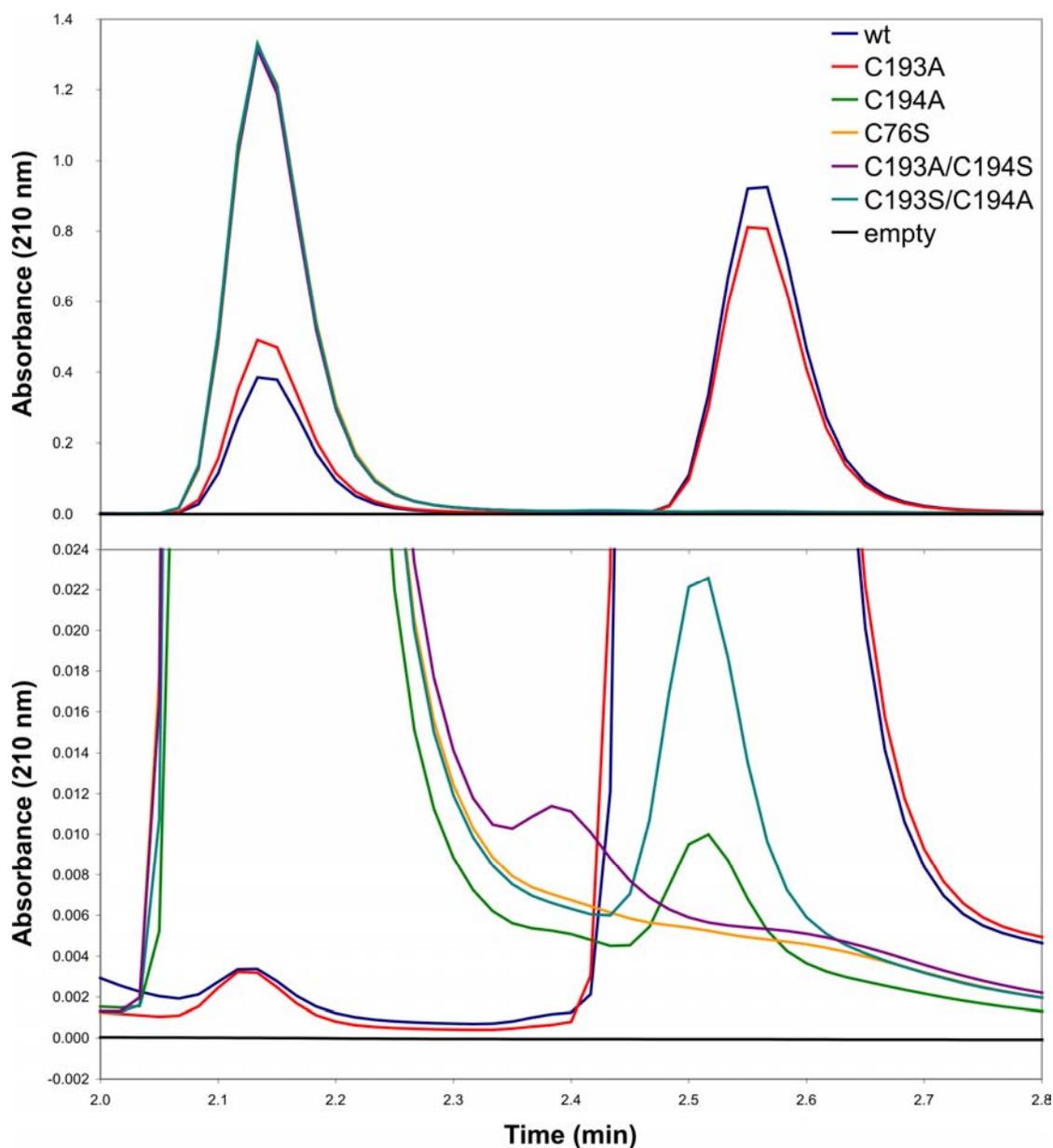


Figure 4.16. NfMI mutants activity assay. HPLC chromatograms of purified NfMI mutant enzymes incubated with 1 mM maleate for 6 min at 140 ng mL^{-1} (top) or 18 h at $9 \text{ } \mu\text{g mL}^{-1}$ (bottom) at $30 \text{ } ^\circ\text{C}$ without DTT. Maleate at RT of 2.15, fumarate at RT of 2.55. The peak at RT of 2.40 corresponds to TCEP.

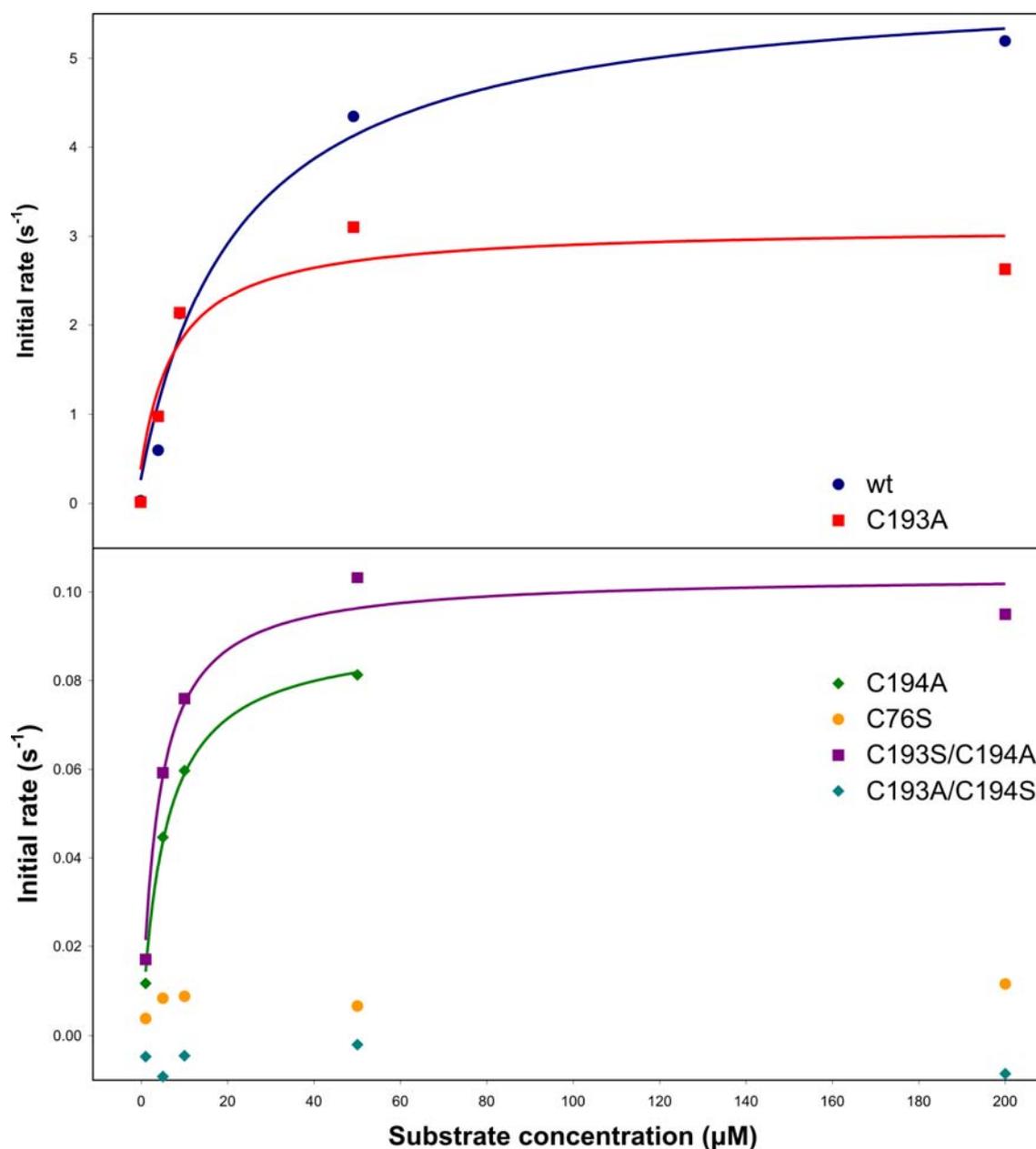


Figure 4.17. Kinetic parameter determination of NfMI mutants. Purified NfMI mutant enzymes assayed at different maleate concentrations for 5 min at 140 ng mL⁻¹ (top) or 60 min at 4 μg mL⁻¹ (bottom) at 30 °C without DTT. Samples were analysed by HPLC. The signals of C76S and CC193AS were too low to be analysed. Simple rectangular hyperbolas were ($y = k_{cat} \cdot x / (K_M + x)$) fitted to determine the kinetic parameters (see Table 4.2).

Table 4.2. Steady state kinetic parameters of NfMI mutants. Data in Figure 4.14. Standard errors are indicated.

Mutant	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu\text{M s}^{-1}$)
wt	21.5 ± 6.3	5.9 ± 0.5	0.27 ± 0.10
C193A	7.2 ± 3.8	3.1 ± 0.4	0.43 ± 0.28
C194A	3.9 ± 0.9	$3.6 \cdot 10^{-3} \pm 0.2 \cdot 10^{-3}$	$9.4 \cdot 10^{-4} \pm 0.3 \cdot 10^{-4}$
C76S	ND	ND	ND
C193S/C194A	5.4 ± 0.6	$3.1 \cdot 10^{-3} \pm 0.1 \cdot 10^{-3}$	$5.9 \cdot 10^{-4} \pm 0.1 \cdot 10^{-4}$
C193A/C194S	ND	ND	ND

ND = not determined

4.3.8 Biotransformation in Combination with Enoate

Reductases

In an attempt to extend the industrial usefulness of NfMI, combined reactions were conducted with enoate reductases from the old yellow enzyme family (OYE2 and YqjM) in collaboration with scientists at BASF (Ludwigsafen). By isomerising the substrate of enoate reductases, the stereospecificity of the reaction is expected to change. Additionally, another potential *cis-trans* isomerase identified in yeast (*C. rugosa*) cell extract (Lu905). FactorX was found to change the product specificity of enoate reductase reactions. As changes in product stereospecificity of enoate reductases resulting from the substrate double bond configuration have been observed, it was assumed that FactorX is a *cis-trans* isomerase.

The enoate reductases were added as cell extracts from overexpressions in *E. coli* whereas FactorX was a sample of partially purified cell extract. All combinations of isomerases with reductases were assayed with either 16 mM *cis/trans* citral or 10 mM of the other substrates at 30 °C. The purity and the amount of enzyme as well as the incubation time varied from experiment to experiment.

In the absence of isomerases, OYE2 specifically reduced citral into (*R*)-citronellal and preferred the *trans* over the *cis* configuration, which corresponds to prior knowledge of this reaction (personal communication by Daniela Burkhardt from BASF). Of the remaining substrates, OYE2 only transformed maleate and dimethylcitraconate. In the case of maleate, as expected, some succinate was produced and surprisingly also some fumarate. YqjM mainly produced geraniol plus small amounts of nerol and (*S*)-citronellal hence the sample contained more alcohol dehydrogenase activity than

enoate reductase activity. YqjM like OYE2 transformed maleate into succinate and fumarate and converted dimethylcitrate. Unlike OYE2, YqjM also had low activity with fumarate, converting it into succinate. The disappearance of citraconate observed in the presence of YqjM yet not in the presence of NfMI and YqjM is highly anomalous and should be considered as an experimental error until confirmed by repetition (Fig 4.18 and Fig. 4.19).

NfMI did not have an effect on the activity of either of the two reductases, except in the case of maleate, which is in accordance with previous experiments, where maleate was the sole substrate. When NfMI was added to the OYE2 maleate reactions the depletion of maleate (and fumarate) is slower. When added to the YqjM maleate reactions, the depletion of maleate and fumarate is not noticeably affected (Fig 4.18 and Fig. 4.19).

The analysis of FactorX is very complex as it was used as different preparations, initially in a partially purified form (FactorX) and when the sample ran out as *C. rugosa* total cell extract (Lu905). It can therefore only be properly analysed within a single experiment. In its freshest and semi-purified form with citral as substrate it produced mainly (*R*)-citronellol and a small amount of nerol and slightly preferred *trans* over *cis*-citral, which is indicative of alcohol dehydrogenase activity but not isomerase activity. When combined with OYE2, surprisingly, the main product is (*S*)-citronellol, not produced by any of the single enzyme controls. Also the appearance of geraniol at lower quantity is a new observation whereas the other products are produced in similar quantities as with both enzymes individually (Fig. 4.18 top). When older FactorX was used in lower quantity but longer incubation time, the result was strikingly different. Instead of (*R*)-citronellol, the main compound now was geraniol with nerol produced at a similar level. The combination of YqjM besides the combination of the individual enzymes only produces a small amount of (*S*)-citronellol (Fig. 4.18 bottom).

Whole cell extract containing Factor X (Lu905) was used with the remaining substrates. Lu905 was highly active on fumarate as well as citraconate and showed some activity on all the other compounds. The combination with the reductases did not result in any difference in depletion of substrate. Because of high amounts of lactate present in the samples, hiding the succinate signals, the latter could not be detected. These results are not very informative, given that we know Lu905 contains enoate reductase activity (personal communication by Andreas Schädler from BASF).

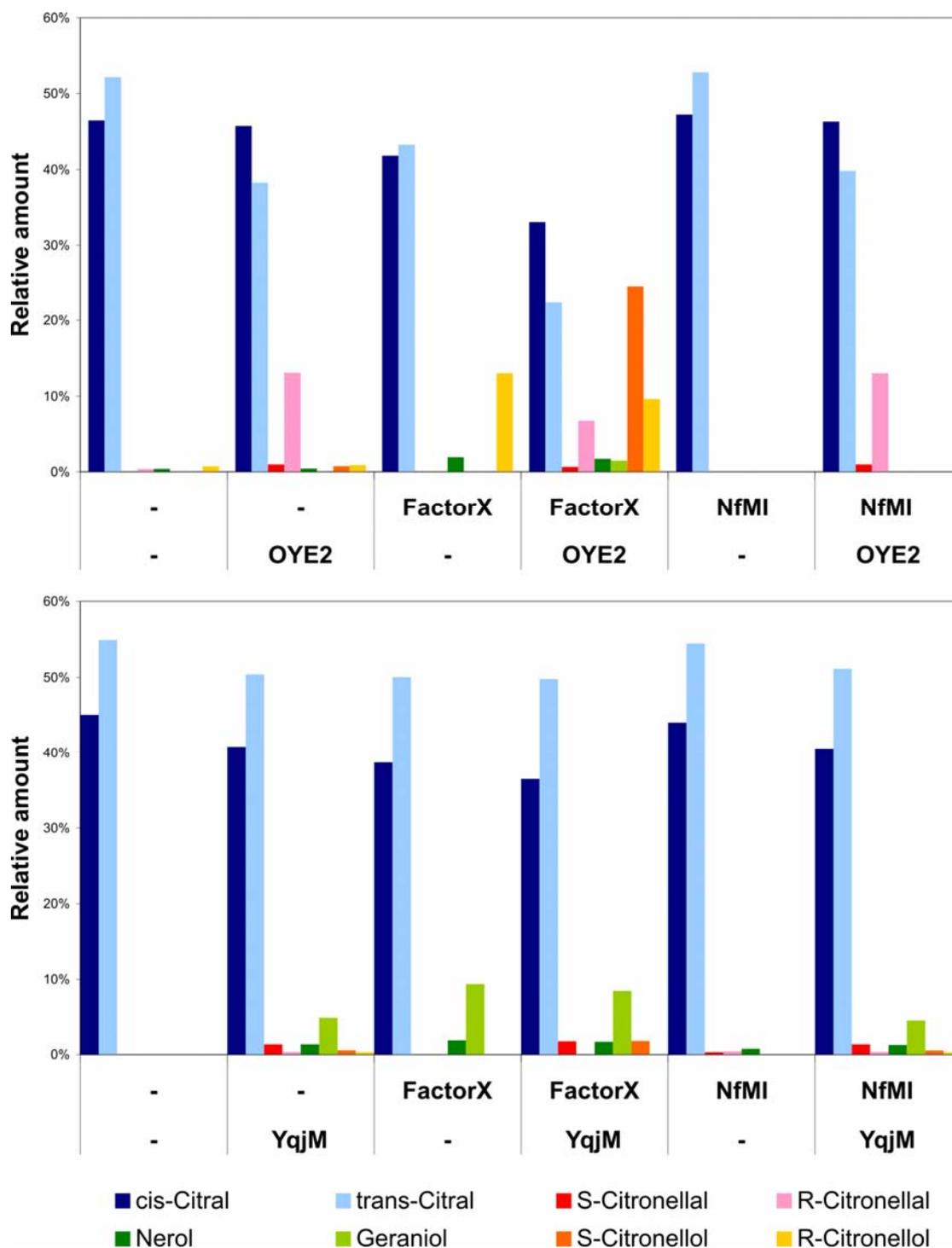


Figure 4.18. Citral assay with NfMI and enoate reductases. 16 mM *cis/trans*-citral was incubated with a combination of an isomerase (15 $\mu\text{g mL}^{-1}$ NfMI or partially purified FactorX, 2.2 mg mL^{-1} with OYE2 or 1.5 mg mL^{-1} with YqjM) and an enoate reductase (OYE2 or YqjM, both 0.1 mg mL^{-1} total protein) at 30 °C for 6 h (top) or 16 h (bottom). The samples were analysed by GC.

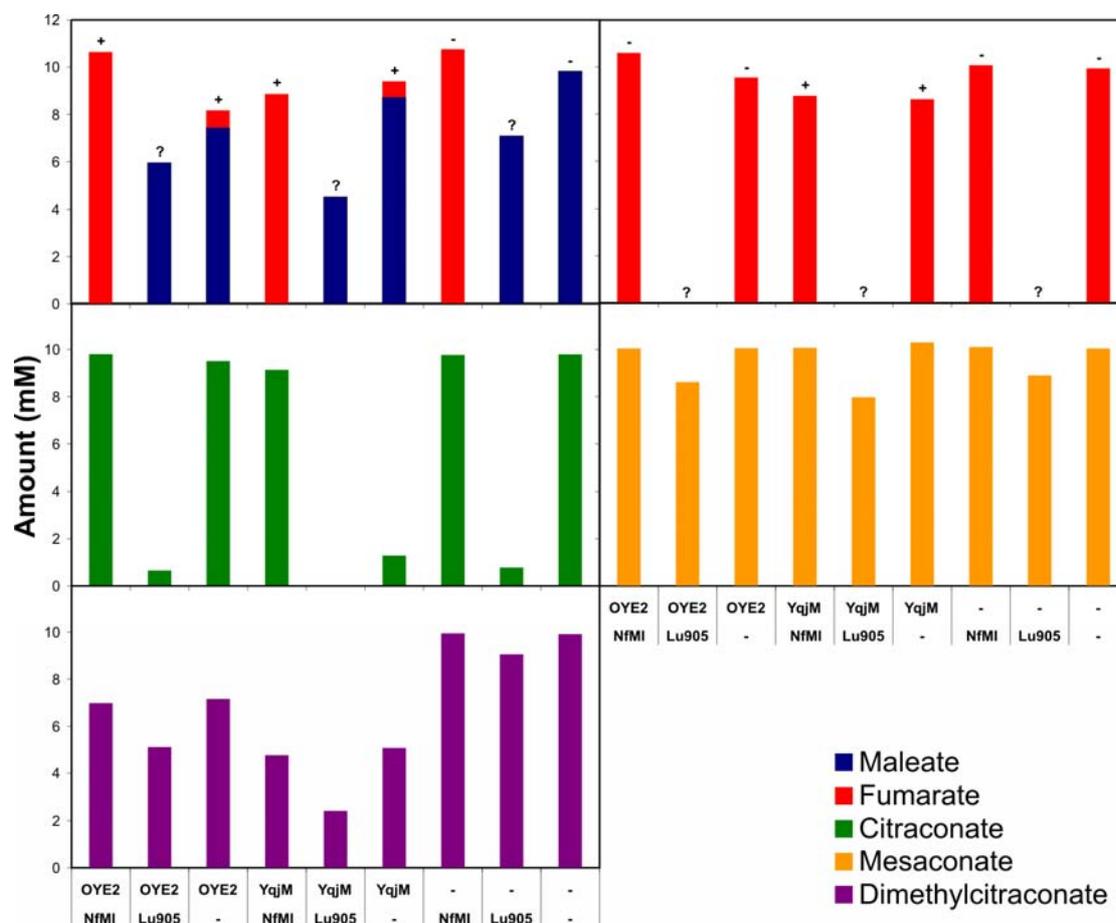


Figure 4.19. Maleate analogues assay with NfMI and enoate reductases. 10 mM of maleate (top left), fumarate (top right), citraconate (middle left), mesaconate (middle right), dimethylcitraconate (bottom) respectively were incubated with a combination of an isomerase ($15 \mu\text{g mL}^{-1}$ NfMI or 5mg mL^{-1} total protein Lu905) and an enoate reductase (OYE2 or YqjM, both 0.1mg mL^{-1} total protein) for 16 h at 30°C . The samples were analysed by HPLC. The presence of succinate in maleate and fumarate samples was determined by another HPLC method. + = succinate detected. - = succinate not detected. ? = succinate peak hidden under a large lactate peak.

4.4 Discussion

4.4.1 Conditions for Activity and Stability

As predicted by phylogenetic analysis of sequence homologues (Fig. 1.21, Chapter 3, sequence alignment in Appendix C) NfMI was experimentally confirmed to be a maleate isomerase. It is only the fourth sequenced enzyme with this associated activity.

The pH optimum of 7.5 is lower than any described MI and is at the lower end in the superfamily. The pH range for activity is similar than in other described enzymes in the superfamily with a slight shift in activity towards lower pH. The K_M of 10 μM is the lowest of any superfamily member with their respective substrates described in the literature to date, whereas the k_{cat} of 6 s^{-1} is lower than for the two maleate isomerases described but in the middle compared to the rest of the superfamily.

The temperature analysis is more complex as the optimum temperature for a reaction also depends on the speed of protein inactivation at high temperatures. Likewise, the oxidation of the enzyme further reduces its activity. In assays without reducing agents the temperature optimum for a 20 min assay was determined to be between 35 and 40 °C but the enzyme was only stable for 1 h up to 30 °C.

P. fluorescens MI was described to be dependent on thiols for activity (100). Like all other described MI activities NfMI does not need this activation. Conversely thiols do stabilise the enzyme and protect it from inactivation at increased temperatures and prolong its long term stability at low temperature. In the literature, the exact role of thiols in the activation or stabilisation has never been established clearly (100,103). The analysis of these factors is complicated by the fact that thiols are able to isomerise maleate chemically. This chemical background activity is still detectable at 5 mM DTT in the enzyme assays.

Combining the work presented here and previously published data, it is reasonable to assume that the cysteines are highly susceptible to oxidation which inactivates the enzyme. This process is reversible through the addition of thiols, capable of specifically reducing the cysteines. The necessity for activation using thiols with *P. fluorescens* is probably due to the oxidation of the enzyme by the extraction process. The inactivation of NfMI after long term storage even in presence of DTT, can be explained by the instability of DTT under the normal atmosphere especially at alkaline pH. This can be tested by rescuing old enzyme preparation by adding fresh DTT, using more stable TCEP instead of DTT or storing the enzyme under inert atmosphere to protect it from oxidation. As TCEP is specific for cysteines, its use makes it possible to distinguish whether enzyme inactivation is due to the oxidation of cysteine or methionine. It was suggested that conserved Met197 makes it susceptible to oxidation in *Alcaligenes faecalis* MI (103).

In addition to increased long term stability, DTT clearly improves the temperature stability of NfMI. In preliminary data more than 60% of the activity was retained at 50 °C

when preincubated with DTT for 1 h. There seem to be two separate effects of thermal inactivation; that are oxidation dependent and oxidation independent. The oxidation independent inactivation is most likely simple denaturation of the protein that cannot be reversed by adding DTT. The protein starts to denature between 30 and 50 °C. The oxidation dependent inactivation is most probably equal to the effect that inactivates the enzyme during long term incubations at low temperature. The oxidation of cysteines is accelerated by higher temperature between 4 and 30 °C. These results have to be considered cautiously here, as the instability of DTT may be a major factor. Quantifying the effect with the use of DTT and H₂O₂ would help to disentangle the roles of oxidation and denaturation.

4.4.2 Mechanism

The reduced mutant activities of C76S, C194A and C193A/C194S compared to the nearly full activity of C193A confirm that the sequence alignment with other known superfamily members as well as the other maleate isomerases was correct. It identifies Cys76 and Cys194 (not Cys193) as the catalytic cysteines. This is also in accordance with reported cysteine to serine mutations in *Bacillus stearothermophilus* MI that lost detectable activity. Unfortunately, the quantification limit reported is above the detection limit of the HPLC assay used in this work. Therefore, possible residual activities in these mutants cannot be compared with those reported here.

The thiol moieties of cysteines are observed to have different roles in enzymes. They can act as acid/base catalysts, nucleophiles or weak hydrogen bonding partners. In GluRs and AMDs the cysteines act as acid/base catalysts deprotonating and protonating the C_α carbons (64,67,123).

Chemical isomerisation was reported to occur with three mechanisms involving a rotation around a single bonded intermediate: radical, acid and thiol catalysis (105,108-110). The radical mechanism can be excluded in this case. In both the acid and thiol catalysis, a nucleophilic attack on a double bond carbon leads to a covalent adduct of water or thiol to maleate to reach a single bonded intermediate. It is therefore possible that one of the cysteine thiols acts as such a nucleophile.

Even after prolonged incubation the C76S mutant does not show any detectable activity, which suggests that the thiol of Cys76 is crucial for activity and cannot be replaced by an alcohol. This finding would be in accordance with Cys76, being a

nucleophile, attacking the double bond carbon and producing an enediolate intermediate observed in all the other Asp/Glu racemase superfamily members.

The analysis of the role of Cys194 is more complex. The complete inactivity of C193A/C194S double mutant suggests that also the second cysteine's thiol group is equally indispensable for activity, yet the C193S/C194A mutant shows some activity, though just detectable. The same residual activity is detectable in the C194A mutant. It therefore seems that role of Cys194 can be partially assumed by a neighbouring serine or cysteine. As C193A/C194S is not active at all and as two neighbouring amino acids are not likely to point into the same direction, its complete inactivation is unlikely to be through removal of the chemical group on the same amino acid but rather through a structural rearrangement that shifts the chemical group from a more distant amino acid away from the correct position. Tyr133 would be a good candidate as it is conserved between MIs and AMDs and plays an important role as part of the dioxyanion hole (see Chapter 5). Cys194 is therefore more likely to be an acid base catalyst or a crucial hydrogen bonding partner than a nucleophile. This coincides with its role in AMDs.

The role of Cys194 and Tyr133 can be tested in several ways. A C194S mutant would be expected to have some residual activity as Cys193 would remain unchanged and keep Tyr133 in place. A Y133F mutant would lose the hydroxyl group in this position and should have a reduced but still detectable activity, similar to C193A. Likewise the same mutation in the background of C194A and C193S/C194A mutants should completely abolish all remaining activity. Other ways, based on structural knowledge of NfMI's three dimensional structure, could be used to induce structural changes on Tyr133 to confirm its role in the active site.

Confirming other reports on MIs from other organisms, NfMI is completely specific to maleate as substrate. The only other reaction detected in a previous study (52) was the reverse reaction that could not be confirmed here as fumarate was contaminated by too high amounts of maleate and as K_{eq} in the measured conditions was 517. Bromomaleate, citraconate and maleamate are not accepted as substrates. These and other substrates including *cis*- and *trans*-3-chloroacrylate, 2-chloromaleate and 2-chlorofumarate, tiglate, *trans*-2-pentenoate and 2,3-dimethylmaleate have also been tested with negative results in previous studies (98,100,101). In this study dimethylcitraconate and citrate have also been shown not to be substrates.

The activity can be affected by different properties of the substrates. Firstly, the thermodynamic equilibrium of the double bond isomers determines that only maleate

reacts in sufficient amounts to be detected. This might explain the inactivity of some substrates. Secondly, the two carboxyl groups are strongly electron withdrawing thus rendering the double bond carbons more positively charged, which makes the possibility of a nucleophilic attack of a cysteine a reasonable mechanism. Thirdly, the substrate has to be able to bind to the active site. With other substrates, the binding might be hindered sterically because of substituents on the double bond or larger chemical groups at the ends. The substrate may also be prevented from binding simply by the absence of hydrogen bonding possibilities with the active site.

For most of these substrates, steric hindrance would be the most straightforward explanation of the inactivity. The methyl or halogen substituents on the double bond or the methylesters all increase the size of the substrate. For maleamate and *cis*-3-chloroacrylate this cannot be the explanation, especially for maleamate, which is approximately isosteric. Although the hydrophilicity of the amide is similar to the carboxyl, the hydrogen bonding and the missing negative charge may strongly reduce the binding. In addition, the amide group is much less electron withdrawing than the carboxyl group, which may also explain the inactivity. In the case of the chloroacrylates the electron withdrawing capacity is maintained. In this case, the reduced binding must be the reason for the inactivity as the chlorines are much smaller than the carboxylates. The hypothesis of steric hindrance can be tested by using fluoromaleate being isosteric to maleate. Testing the methylester of maleate would complete the picture of steric hindrance although it is not expected to be converted by NfMI.

Despite the possibility of steric hindrance affecting substrate binding, it is surprising that chloro- and bromomaleate are not substrates. Their size is still limited and if a nucleophilic attack was to occur, the additional electron withdrawing capacity would favour nucleophilic attack. It cannot be excluded that the loss of symmetry by the addition of a substituent affects the mechanism as the molecule can bind in two ways, one of which may disfavour the reaction. To test this hypothesis, disubstituted compounds like difluoromaleate and dimethylmaleate could be assayed.

A large variety of compounds have been tested as inhibitors on MIs in previous reports (98,100,101). All compounds identified this way were either heavy atoms or oxidative inhibitors such as Hg^{2+} and iodoacetamide, inhibiting the catalytic cysteines. Various other compounds containing mono-, di- and tricarboxylic acids and/or double bonds have been tested including maleamate, *cis*-1,4-butanediol, diethyl- and dimethylmaleate, citraconate, mesaconate, cyclopentenedione, L-malate, L-aspartate,

L-glutamate and EDTA. Apart from L-aspartate and EDTA inhibiting at high concentrations, no competitive inhibitor has been identified so far.

The only strong non-oxidative inhibitor identified was bromomaleate. Although more data have to be collected for more accurate rate determinations, taking into account the substrate inhibition, the data are consistent with the model of competitive inhibition. Irreversible inactivation can be ruled out as the rate would have to drop proportionally to the substrate concentration. Noncompetitive and uncompetitive inhibition would not affect the K_M that is clearly increasing with inhibitor concentration. Mixed inhibition leading to an increase in K_M with a simultaneous decrease in k_{cat} cannot be ruled out with complete certainty; however, the similarity of the structures of the inhibitor and the substrate further support the competitive inhibition model over the mixed inhibition model.

It is surprising that bromomaleate is not a substrate and it would be interesting to see whether fluoromaleate would behave similarly to bromomaleate or whether it would overcome the supposed steric hindrance and be a substrate. Also citraconate being isosteric to bromomaleate is not an inhibitor, suggesting an electrostatic component in the inhibition.

More information about the mechanism could be obtained from structural data showing the binding mode of substrate and/or inhibitor. Also structural data of slightly active and completely inactive mutants are likely to reveal more details about the binding mode and the role of the catalytic residues.

To gain more insight into the role of the cysteines, their pK_a should be determined. The values of active site cysteines can be in a wide range depending on the microenvironment. A plot of the pH versus the catalytic efficiency (k_{cat}/K_M) would show two different inflection points corresponding to the pK_a values of both cysteines as the enzyme is most active when both cysteines are in the correct protonation state.

4.4.3 Industrial Usefulness

Due to the absolute substrate specificity observed for all MIs so far, the use of MI is limited to the production of fumarate from maleate or, if the equilibrium is pushed, also in the reverse direction. Depending on the reasons for the substrate restriction, there is hope for the extension of the isomerisation activity through engineering of the active site. Structural data are crucial for informed decisions on mutations.

NfMI has the lowest pH optimum of all MIs described to date and a large pH range, which could potentially make it more useful. It also has the lowest K_M value with a moderately high k_{cat} . The latter would most probably need to be improved for industrial application but the low K_M provides an interesting addition to the enzymes available for engineering. Also the substrate inhibition is an unwanted property for industrial production that would have to be considered in more detail.

Fumarate, however, with its two distinct chemical functionalities is very versatile chemical intermediate that is widely used for the production of unsaturated polyester resins, paper sizing chemicals, lubricants and lacquers. Fumarate can be converted into malate. Both are used as food and drink acidulants. Fumarate is also converted into aspartate (105). Both L-malate and L-aspartate can easily be produced stereospecifically using the enzymes fumarate hydratase (EC 4.2.1.2) and aspartate ammonia lyase (EC 4.3.1.1) (4). Analogously, maleate hydratase (EC 4.2.1.31) could be used to produce D-malate (Fig. 4.20) (4).

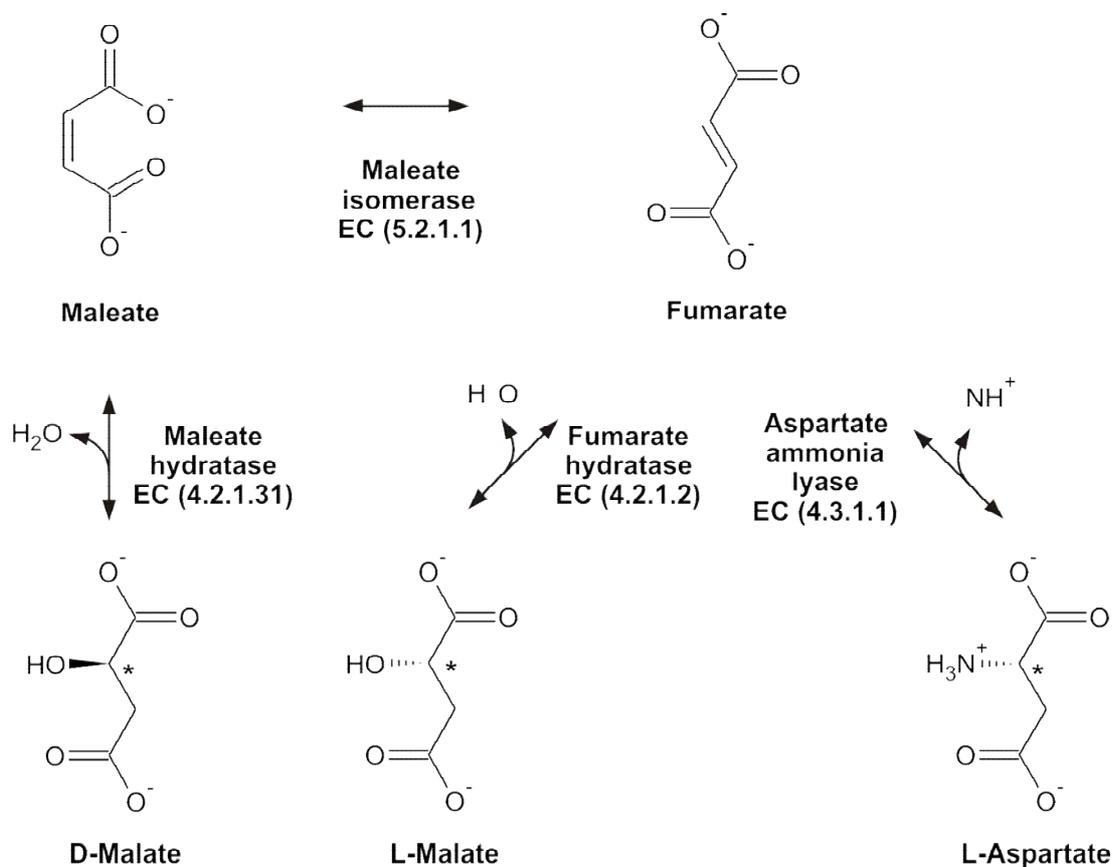


Figure 4.20. Possible biotransformations of maleate and fumarate.

Currently fumarate as well as its precursor maleic anhydride, that is also widely used, are produced from butane that is refined from fossil oil (105). The microbial production of maleate from organic waste materials, originating from food industries and agriculture, provides a possible alternative route in a chemical industry independent of fossil oil. Fumarate could then be produced by the same microbes using maleate isomerase or alternatively, the conversion to fumarate could be interrupted to allow the extraction of maleate that can further be dehydrated to form maleic anhydride (Fig. 4.21.)

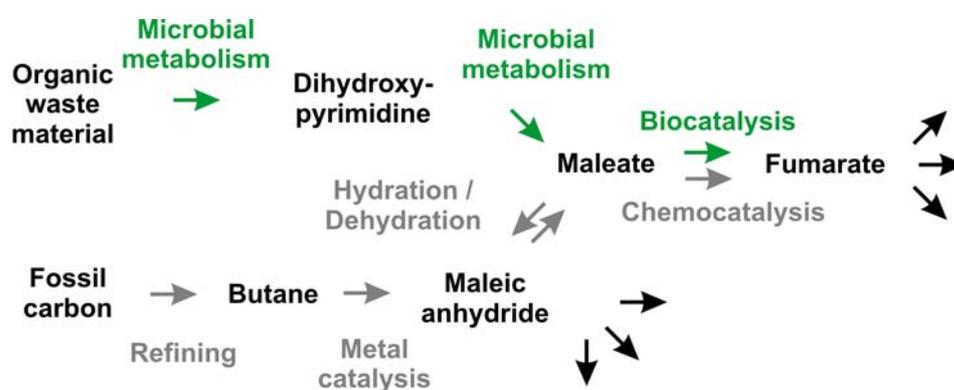


Figure 4.21. Production of fumarate and maleic anhydride. The current fossil fuel based production (grey) and a possible bio based production (green).

Chapter 5

Crystal Structure of NfMI

5.1 Introduction

The biochemical observations of NfMI in the previous chapter have confirmed many characteristics known for other MIs such as the substrate specificity, the sensitivity to oxidation and the presence of the catalytic cysteine dyad. In addition, bromomaleate was found to be a competitive inhibitor of NfMI. Despite all this, characterisation of the precise mechanism of the maleate isomerisation still remains elusive. Structural data are key for a full understanding of an enzymatic mechanism; however, no experimental three dimensional structure of a MI is described so far.

A series of maleate isomerisation activities has been described (see previous chapter) but only three MI sequences have been annotated experimentally (52,103,104). With NfMI a fourth sequence has been added to this collection. Until now, all structural knowledge about MIs had to be inferred from sequence comparisons with known structures in the Asp/Glu racemase superfamily. This has led to the identification of the catalytic cysteines as confirmed by site directed mutations to serine in *Alcaligenes faecalis* MI (104). In the same way Cys76 and Cys194 were found to be the catalytic residues in NfMI (previous chapter).

Although the approach based on sequence comparisons has correctly predicted the conservation of the protein fold in the superfamily, it clearly has its limitations in predicting the function and mechanism. In Chapter 3 it was shown that the automatic annotation of gene functions based on sequence motifs and grouping by similarity still fails to recognise the activity of the homologues that have less than 40% sequence identity. Only NfMI and *Mesorhizobium* sp. BNC1 AMD activities were successfully predicted (48).

The Asp/Glu racemase superfamily is catalytically very flexible and harbours industrially and pharmaceutically interesting activities such as racemisation and decarboxylation. While it was possible to invert the enantiospecificity of the decarboxylation by BbAMD by simply swapping the conserved glycine and cysteine and to obtain racemisation by replacing the glycine with a second cysteine, only low

activities could be obtained (79,93). Also attempts to increase the range of substrates in BbAMD had rather limited success (48,67). This results in an apparent contradiction that despite the large evolutionary flexibility of the superfamily, the individual activities are very specific and cannot be changed by simple mutagenesis in substrate binding residues.

These limitations in our knowledge about the structure function relationship in the Asp/Glu racemase superfamily stress the need for a MI structure. Maleate *cis-trans* isomerisation is substantially different in its mechanism from the other reactions in the superfamily but recruits the same fold and the same catalytic residues. The mechanism most probably has to go through an intermediate that allows free rotation around the double bond. This is a completely new feature in the superfamily and the understanding of the structural requirements will broaden the understanding of catalysis in the whole superfamily and help facilitate the exploitation of the catalytic potential by targeted engineering and the discovery of new activities.

In this chapter, the solution of the X-ray crystal structure of NfMI will be presented. Based on the binding of a TRIS molecule in the active site the possible role of the active site residues will be discussed and different models for substrate binding will be presented. The role of flexibility in the enzyme will also be highlighted.

5.2 Methods

5.2.1 Cloning and Protein Purification of NfMI

See Chapter 2.

5.2.2 His-tag Cleavage

The pET-YSBLIC3C plasmid includes a HRV 3C protease cleavage site (LEVL~~F~~Q↓GP) in front of the insert protein's N-terminus allowing the cleavage of the His-tag. The cleavage was performed on 1 mg mL⁻¹ target protein incubated with 0.01 mg mL⁻¹ of the protease at 4 °C for 14 h in the protein storage buffer (20 mM TRIS, 150 mM NaCl, pH 8.0). The cleavage of the His-tag was checked by SDS-PAGE analysis. The cleaved protein was separated from the uncleaved protein, cleaved His-tag and the protease (containing a His-tag itself) by an additional nickel-affinity column purification step. The purified protein contained three remaining N-terminal amino acids (GPA) not belonging to the native protein.

5.2.3 Crystallisation

Protein samples at 3 - 17 mg mL⁻¹ concentration were prepared freshly or taken from snap frozen -80 °C stocks. Proteins were stored at pH 8 in TRIS or sodium phosphate buffers containing 150 - 300 mM NaCl. Some samples contained additives: 5 mM DTT, 10 mM BME or 10 mM TCEP to prevent oxidation or 100 - 200 µM PMSF to inhibit serine proteases.

Crystallisation trials were performed with the following screens: PACT Premier (Molecular Dimension Ltd), Index (Hampton Research), CSS 1 and 2 (with MES pH 5.6 or 6.0 and TRIS pH 8.0 (124), and PEG Ion/Ion 2 (without buffers, Hampton Research). The screens were prepared on 96-well sitting drop vapour diffusion plates (MRC Wilden) with a reservoir volume of 60 µL. Drops of 150 nL protein solution and 150 or 300 nL reservoir solution respectively were dispensed using a Mosquito robot (TTP Lotech).

Positive hits were scaled up to 1 µL protein solution and 1 or 2 µL reservoir solution and optimised using the hanging drop vapour diffusion technique on siliconised glass cover slides over 24-well cell culture plates (Greiner) using 1 mL reservoir solutions. Drops were incubated from 1 day to 2 weeks at 25 °C.

5.2.4 Data collection

Protein crystals were transferred from the drops into cryoprotectant solution using the mother liquor including 10% 1,2,6-hexanetriol or 30% PEG 3350 and subsequently snap frozen in liquid nitrogen. Diffraction was tested in-house using Micromax 007HF rotating copper anode X-ray generator (Rigaku) and mar345 X-ray detector (marresearch) at 120 K. Diffracting crystals were sent to the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), beamline ID23-1 with a Q315r CCD detector (ADSC). Diffraction images were collected at a wavelength of 0.98 Å at 100 K (Oxford Cryosystems Cryostream). For data collection statistics see Table 5.2.

5.2.5 Data Processing, Structure Solution and Refinement

The diffraction data obtained from the synchrotron were indexed and integrated with imosflm (125) and the intensities were scaled and reduced with SCALA (126). The Matthews coefficient of 2.4 Å³ Da⁻¹ with a corresponding solvent content of 48% indicated four molecules in the asymmetric unit with a probability of 66% (127). The structure phases were solved by molecular replacement using the BbAMD coordinates

(PDB: 3dg9) as a search model using Phaser (128). The phases were improved by NCS averaging and solvent flattening using DM (129).

A model of the well defined regions was built in COOT (130) and refined with Phenix (131) including simulated annealing to reduce the model bias. Several cycles of building additional parts of the structure model and refinement were used to build the remaining fragments. The final model was refined using Phenix introducing TLS corrections taking each polypeptide chain as a group.

5.2.6 Structure Validation and Analysis

The model was validated with Molprobity (Table 5.1) (132) and the secondary structure was assigned with DSSP (133). The areas of interfaces between molecules were calculated with PISA via the EBI web interface (134).

Table 5.1. Structure validation statistics.

Molprobity output	
Clashscore ¹	19.83 (73 rd percentile) ³
Poor rotamers	3.1%
Ramachandran outliers	0.5%
Ramachandran favored	95.7%
C β deviations >0.25Å	0.0%
MolProbity score ²	2.47 (77 th percentile) ³
Residues with bad bonds	0.0%
Residues with bad angles	0.0%

¹ Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

² MolProbity score is defined as the following: $0.42574 \cdot \log(1 + \text{clashscore}) + 0.32996 \cdot \log(1 + \max(0, \text{pctRotOut} - 1)) + 0.24979 \cdot \log(1 + \max(0, 100 - \text{pctRamaFavored} - 2)) + 0.5$

³ 100th percentile is the best among structures of comparable resolution
0th percentile is the worst.

The identity of a metal atom present between molecules was analysed by refining with each of the suspected identities. Subsequently, the σ -value of the corresponding peak in the electron difference map, the resulting B-factor of the metal atom and the distances to the coordinating atoms were determined. The σ -value and the B-factor were compared between the different atoms and the coordination distances were compared to known experimental values in the MESPEUS database (135). An additional X-ray fluorescence scan was performed on another crystal. The presence of a heavy metal was analysed by determining an anomalous difference map.

Images were produced using CCP4mg (136). In the case of TRIS bound to NfMI the angular restraints were relaxed to account for the flexibility of TRIS and the uncertainty of the exact oxygen position.

5.2.7 Ligand Modelling and Docking

The surface and volume of the active site was determined with the CASTp server (137), the surfaces and volumes of TRIS, maleate and fumarate were determined with ChemBioOffice (CambridgeSoft). The volume necessary for a rotation of one carboxyl group from the maleate conformation to the fumarate conformation was approximated by calculating the volumes of the theoretical tori resulting from the rotation of the atoms around the axis. In all cases the molecular surfaces and volume (Connolly's surface and volume) were used.

The observed density in the active sites after refinement that could not be explained by modelling water molecules was attributed to TRIS. The TRIS coordinates were taken from the REFMAC5 monomer library (138). The ligand was build in with COOT and refined again as described above.

Based on the TRIS molecule, maleate and fumarate were modelled into the active site of molecule A by manually overlaying the oxygen atoms of the ligands (also obtained from the REFMAC5 monomer library) with the oxygen atoms of TRIS. Orientations of the molecules that had atoms heavily clashing with the protein atoms were not considered. Alternatively, maleate was docked to the active site of the protein model using AutoDock (139).

5.3 Results

5.3.1 Crystallisation

The full sequence of NfMI was cloned using the LIC technique (see Chapter 2) and expressed by IPTG induction in *Escherichia coli* cells. The protein was purified by its N-terminal His-tag first by nickel-affinity and subsequently by gel filtration and was concentrated to 12-15 mg mL⁻¹. The concentrated sample was run on SDS-PAGE and judged to be sufficiently pure for crystallisation (see Chapter 3). The monodispersity of the protein sample was tested with DLS. With 14% polydispersity the sample was considered monodisperse. The hydrodynamic radius was determined to be 3.2 nm which is equivalent of an approximate molecular mass of 51 kDa corresponding

assuming a spherical protein. This would correspond to a dimer of NfMI with a molecular mass of 53 kDa.

Conditions for crystallisation of NfMI were found through screening of a large number of conditions using commercially available sets of crystallisation solutions in a 96-well format (PACT, Index, PEG Ion I&II, CSS I&II). Conditions found in this way were optimised and upscaled in a larger 24-well format. The best crystals were grown in 300 mM calcium acetate, 100 mM TRIS pH 7.5, 16% PEG 3350, 5 mM maleate using seeding of crushed low quality crystals grown in higher PEG concentrations. If the protein was stored in a buffer containing 10 mM TCEP in addition to 20 mM TRIS pH 8.0, 150 mM NaCl the quality of the crystals increased greatly.

In all conditions containing bulky crystals the shape of the crystals was conserved. They were elongated hexagonal prisms with a slanted end containing a defect in the form of a cavity at the thicker end (Fig. 5.1 left). In an attempt to improve the diffraction power, protein with 3C protease cleaved His-tag was produced. Crystals resulting from these protein samples grew in different conditions (200 mM MgCl₂, 100 mM TRIS pH 8.0, 20% PEG 3350) resulting in different crystal shapes (rhombohedral, Fig. 5.1 right). No diffraction could be achieved in-house with these crystals.

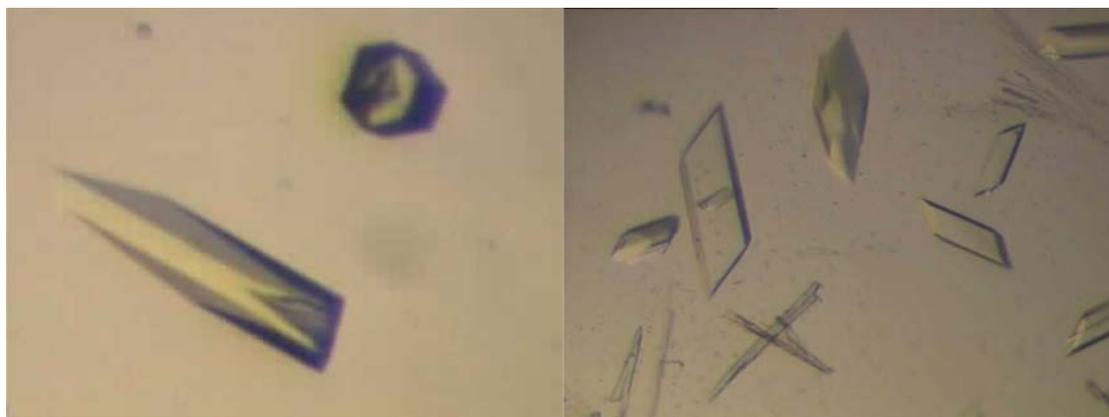


Figure 5.1. Crystals of NfMI. Left: His-tagged protein. Right: Protein with cleaved His-tag.

Co-crystallisation with 1-100 mM maleate (substrate) and fumarate (product), 1 mM bromomaleate (inhibitor) and 1-20 mM of various other substrate analogues such as malonate, succinate and D/L-malate were tried without an improvement of the crystal shapes or diffraction. Co-crystallisations with bromomaleate only diffracted to

approximately 5 Å in-house. A crystal from protein containing its His-tag, co-crystallised with 5 mM maleate was finally used for data collection. It was snap frozen in liquid nitrogen with a total of 30% PEG 3350 as cryoprotectant and diffracted to a maximum of 2.7 Å resolution in-house.

5.3.2 Data Collection, Phasing, Model Building and Refinement

Data were collected at the synchrotron light source ESRF in Grenoble (France). The crystals diffracted to a maximum resolution of 1.9 Å. However, the data were highly anisotropic, which resulted in the final useful resolution limit of only 2.5 Å corresponding to the resolution along the weakly diffracting axis. Crystals belonged to the orthorhombic space group $P2_12_12_1$ with a very long c axis (238 Å) containing four molecules in the asymmetric unit.

The phases could be calculated using molecular replacement based on the Asp/Glu enzyme superfamily homologue representing the highest sequence similarity available in the protein structure database: BbAMD (PDB: 3dg9). The proteins only have a sequence identity of 22%, yet the protein fold is highly conserved in all available structures throughout the whole superfamily. Initial phases yielding an initial model could be determined this way. Through the use of non-crystallographic symmetry (NCS) averaging and density modification through solvent flattening, the map could be improved considerably. This way, it was possible to determine the core of the protein used for the manual extension of the model.

To overcome the model bias caused by the low sequence similarities, refinement with simulated annealing was applied to the molecular replacement solution. This method introduces molecular dynamics into the refinement process meaning that the atoms are randomly assigned a certain velocity resulting in an overall temperature. The method allows the structure to overcome local minima and converge towards the global minimum while the temperature is decreased in each optimisation cycle (140). The same method was used throughout the refinement process. In the end TLS refinement was applied taking each protein molecule as one group. The resulting final R-values were 0.21 (R_{work}) and 0.29 (R_{free}). For additional statistics see Table 5.2.

Table 5.2. Data collection and refinement statistics.

Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
a, b, c (Å)	50.16, 84.51, 238.13
α , β , γ (°)	90, 90, 90
Wavelength (Å)	0.9760
Resolution (Å)	50 – 2.50 (2.61-2.50)*
R _{merge}	0.070 (0.197)*
I/ σ	12.4 (5.3)*
Completeness (%)	99.5 (100.0)*
Redundancy	4.3 (4.4)*
Refinement	
No. reflections	35766
R _{work} / R _{free}	0.21 / 0.29
No. atoms	
Protein	7240
Ligand/ion	18
Water	123
B-factors	
Protein	58
Ligand/ion	49
Water	45
R.m.s. deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.82
Ramachandran statistics	
Favoured (%)	91.4
Allowed (%)	8.5
Generously allowed (%)	0.1
Disallowed (%)	0.0

* Values in parentheses are for highest-resolution shell.

5.3.3 Quaternary Structure and Overall Fold

The asymmetric unit is composed of four molecules. The two central molecules (A and B) form a crystallographic dimer. The molecules other two molecules (C and D) each form another dimer with the corresponding partner in the adjacent asymmetric unit. The dimer contact interface in both AB and CD is formed by α -helices A and C and β -strand 2 on domain 1. The total interface area is 926 Å² as determined with the PISA server.

The two monomers are rotated by approximately 180° and orientated resulting in their active sites facing the other monomer (Fig. 5.2).

The dimers are in contact via a smaller interface of 414 \AA^2 formed by the start of α -helix E and the preceding loop as well as α -helix G and the following loop. Between the two dimers is a metal ion participating in keeping the dimers together. Moreover, the dimers are rotated in respect to each other by 180° forming a long chain of monomers in the crystal linked by inter and intra dimer contacts (Fig. 5.2).

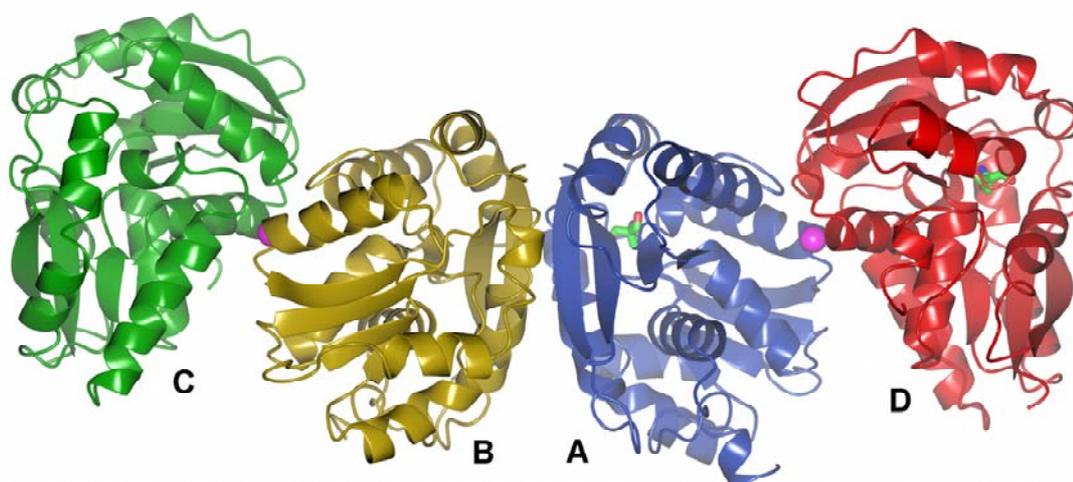


Figure 5.2. Asymmetric unit. Ribbon model of the four molecules. AB forms a dimer with an approximate 180° symmetry axis parallel to the image plane and CD', C'D form equivalent dimers across the asymmetric unit boundaries. Dimers are linked through Ca^{2+} ions (magenta). TRIS molecules (green) are bound in molecule A and molecule D.

The metal ion is coordinated in octahedral fashion but its identity is not entirely clear. The crystallisation conditions contained 300 mM Ca^{2+} and the protein buffer contained 150 mM Na^+ . An X-ray fluorescence scan of a crystal from the same conditions showed a very small peak corresponding to zinc atoms. The metal ions Na^+ , Ca^{2+} and Zn^{2+} were all placed in this position and refined (Table 5.2). The picture resulting from these data is not entirely clear. According to the difference map Zn^{2+} would fit best. The B-factor for Na^+ is the lowest but the distances between the coordinating oxygen atoms and the metal are closest to the expected values for Ca^{2+} . As Ca^{2+} has intermediate values for both the difference map peak and B-factor, it is most likely to be the

observed metal. In addition, no anomalous difference map peak could be observed as would be expected for Zn^{2+} at a wavelength of 0.98 Å.

Table 5.2. Coordination of metal ion between molecules A and D.

Metal ion	σ of difference peak ^a	B-factor	Mean distances		Mean distances	
			$\text{O}\delta_{\text{Glu}}\text{-Me}$		$\text{O}_{\text{water}}\text{-Me}$	
			observed	known ^b	observed	known ^b
Na^+	6.0	54	3.2	2.4	3.3	2.4
Ca^{2+}	2.5	67	2.8	2.4	2.7	2.5
Zn^{2+}	0.5	97	2.9	2.2	2.7	2.2

^adetermined with COOT

^bvalues from MESPEUS database

NfMI consists of two similar Rossmann fold domains ($\beta\alpha\beta$, with parallel β -strands), arranged in a pseudo 2-fold symmetry corresponding to a rotation of about 170 °, forming the active site cleft between the domains. Each domain consists of a central β -sheet formed of four parallel strands (1-4 and 6-8) alternated with one or two α -helices (A-L). β -Strand number 8 and α -helix L are only weakly structured. The two catalytic residues Cys76 and Cys194 are on equivalent positions on the loops between the third β -strand and the following α -helix (3 and D as well as 7 and J respectively) on either side of the active site in positions facing each other. The loops at both ends of the active site cleft, following the second β -strand in each domain (2 and 6 respectively) cover the active site (Fig. 5.3). This fold is shared with all the other known structures of the Asp/Glu racemase superfamily (62,63,71,141-150). The fold is thought to have evolved from by ancestral gene duplication (143).

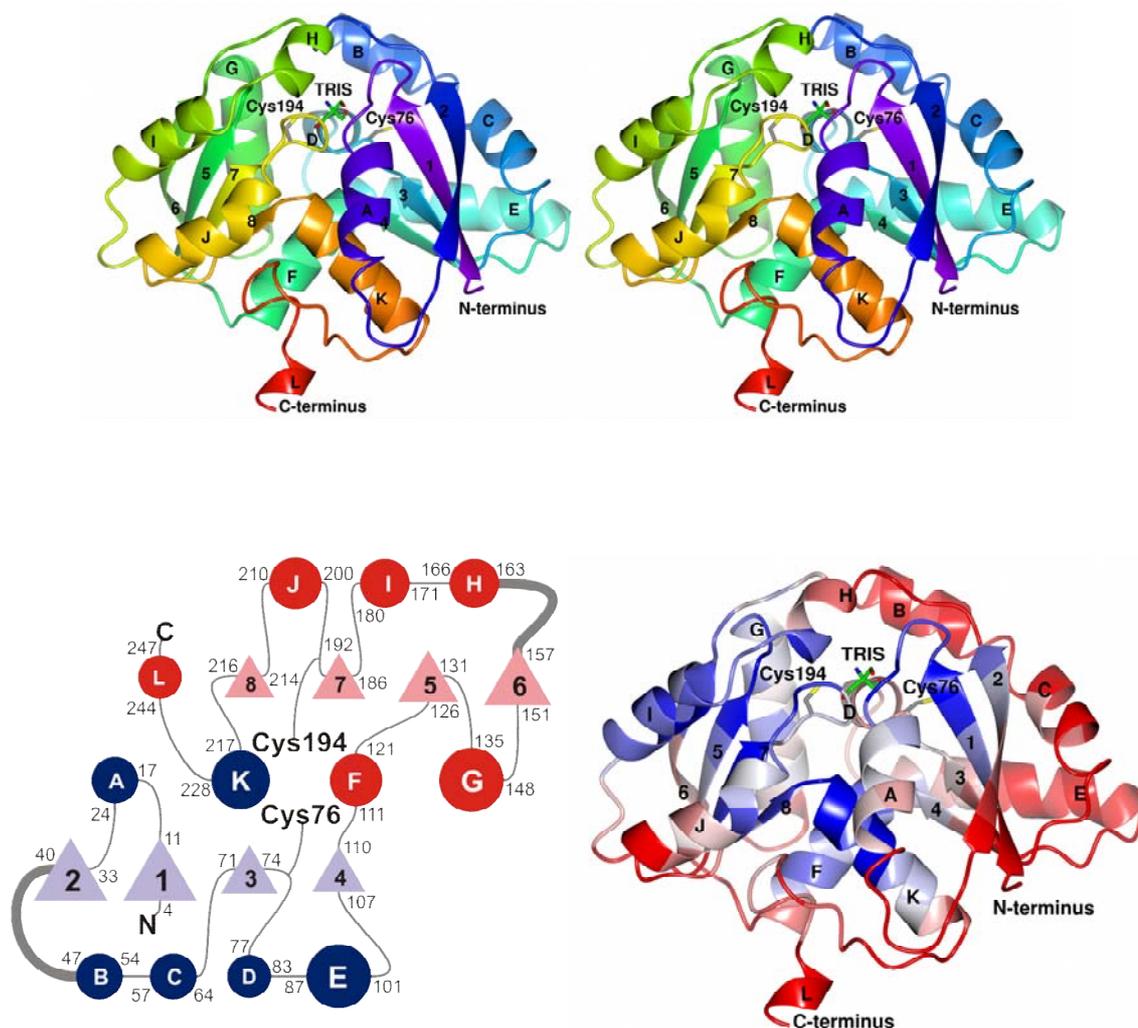


Figure 5.3. Three dimensional fold. Top: Stereoview of ribbon diagram in rainbow colour following the C α -trace. In the active site the catalytic cysteines (grey) and the bound TRIS molecule (green) are shown as stick models. Bottom left: Topology diagram showing domain 1 (blue) and domain 2 (red). β -Strands are shown as triangles (1-8) and α -helices as circles (A-L) with their first and last residue numbered. The two loops covering the active site are shown as thick lines. Bottom right: Ribbon diagram showing temperature factors of 30 and below (blue), 30-50 (grey), 50-70 (orange), 70-90 (red), 90-110 (dark red), 110-130 (brown), 130-150 (black), 150-170 (dark grey), 170-190 (light grey), 190-210 (white), 210-230 (light blue), 230-250 (medium blue), 250-270 (dark blue), 270-290 (very dark blue), 290-310 (black), 310-330 (dark grey), 330-350 (medium grey), 350-370 (light grey), 370-390 (white), 390-410 (light blue), 410-430 (medium blue), 430-450 (dark blue), 450-470 (very dark blue), 470-490 (black), 490-510 (dark grey), 510-530 (medium grey), 530-550 (light grey), 550-570 (white), 570-590 (light blue), 590-610 (medium blue), 610-630 (dark blue), 630-650 (very dark blue), 650-670 (black), 670-690 (dark grey), 690-710 (medium grey), 710-730 (light grey), 730-750 (white), 750-770 (light blue), 770-790 (medium blue), 790-810 (dark blue), 810-830 (very dark blue), 830-850 (black), 850-870 (dark grey), 870-890 (medium grey), 890-910 (light grey), 910-930 (white), 930-950 (light blue), 950-970 (medium blue), 970-990 (dark blue), 990-1010 (very dark blue), 1010-1030 (black), 1030-1050 (dark grey), 1050-1070 (medium grey), 1070-1090 (light grey), 1090-1110 (white), 1110-1130 (light blue), 1130-1150 (medium blue), 1150-1170 (dark blue), 1170-1190 (very dark blue), 1190-1210 (black), 1210-1230 (dark grey), 1230-1250 (medium grey), 1250-1270 (light grey), 1270-1290 (white), 1290-1310 (light blue), 1310-1330 (medium blue), 1330-1350 (dark blue), 1350-1370 (very dark blue), 1370-1390 (black), 1390-1410 (dark grey), 1410-1430 (medium grey), 1430-1450 (light grey), 1450-1470 (white), 1470-1490 (light blue), 1490-1510 (medium blue), 1510-1530 (dark blue), 1530-1550 (very dark blue), 1550-1570 (black), 1570-1590 (dark grey), 1590-1610 (medium grey), 1610-1630 (light grey), 1630-1650 (white), 1650-1670 (light blue), 1670-1690 (medium blue), 1690-1710 (dark blue), 1710-1730 (very dark blue), 1730-1750 (black), 1750-1770 (dark grey), 1770-1790 (medium grey), 1790-1810 (light grey), 1810-1830 (white), 1830-1850 (light blue), 1850-1870 (medium blue), 1870-1890 (dark blue), 1890-1910 (very dark blue), 1910-1930 (black), 1930-1950 (dark grey), 1950-1970 (medium grey), 1970-1990 (light grey), 1990-2010 (white), 2010-2030 (light blue), 2030-2050 (medium blue), 2050-2070 (dark blue), 2070-2090 (very dark blue), 2090-2110 (black), 2110-2130 (dark grey), 2130-2150 (medium grey), 2150-2170 (light grey), 2170-2190 (white), 2190-2210 (light blue), 2210-2230 (medium blue), 2230-2250 (dark blue), 2250-2270 (very dark blue), 2270-2290 (black), 2290-2310 (dark grey), 2310-2330 (medium grey), 2330-2350 (light grey), 2350-2370 (white), 2370-2390 (light blue), 2390-2410 (medium blue), 2410-2430 (dark blue), 2430-2450 (very dark blue), 2450-2470 (black), 2470-2490 (dark grey), 2490-2510 (medium grey), 2510-2530 (light grey), 2530-2550 (white), 2550-2570 (light blue), 2570-2590 (medium blue), 2590-2610 (dark blue), 2610-2630 (very dark blue), 2630-2650 (black), 2650-2670 (dark grey), 2670-2690 (medium grey), 2690-2710 (light grey), 2710-2730 (white), 2730-2750 (light blue), 2750-2770 (medium blue), 2770-2790 (dark blue), 2790-2810 (very dark blue), 2810-2830 (black), 2830-2850 (dark grey), 2850-2870 (medium grey), 2870-2890 (light grey), 2890-2910 (white), 2910-2930 (light blue), 2930-2950 (medium blue), 2950-2970 (dark blue), 2970-2990 (very dark blue), 2990-3010 (black), 3010-3030 (dark grey), 3030-3050 (medium grey), 3050-3070 (light grey), 3070-3090 (white), 3090-3110 (light blue), 3110-3130 (medium blue), 3130-3150 (dark blue), 3150-3170 (very dark blue), 3170-3190 (black), 3190-3210 (dark grey), 3210-3230 (medium grey), 3230-3250 (light grey), 3250-3270 (white), 3270-3290 (light blue), 3290-3310 (medium blue), 3310-3330 (dark blue), 3330-3350 (very dark blue), 3350-3370 (black), 3370-3390 (dark grey), 3390-3410 (medium grey), 3410-3430 (light grey), 3430-3450 (white), 3450-3470 (light blue), 3470-3490 (medium blue), 3490-3510 (dark blue), 3510-3530 (very dark blue), 3530-3550 (black), 3550-3570 (dark grey), 3570-3590 (medium grey), 3590-3610 (light grey), 3610-3630 (white), 3630-3650 (light blue), 3650-3670 (medium blue), 3670-3690 (dark blue), 3690-3710 (very dark blue), 3710-3730 (black), 3730-3750 (dark grey), 3750-3770 (medium grey), 3770-3790 (light grey), 3790-3810 (white), 3810-3830 (light blue), 3830-3850 (medium blue), 3850-3870 (dark blue), 3870-3890 (very dark blue), 3890-3910 (black), 3910-3930 (dark grey), 3930-3950 (medium grey), 3950-3970 (light grey), 3970-3990 (white), 3990-4010 (light blue), 4010-4030 (medium blue), 4030-4050 (dark blue), 4050-4070 (very dark blue), 4070-4090 (black), 4090-4110 (dark grey), 4110-4130 (medium grey), 4130-4150 (light grey), 4150-4170 (white), 4170-4190 (light blue), 4190-4210 (medium blue), 4210-4230 (dark blue), 4230-4250 (very dark blue), 4250-4270 (black), 4270-4290 (dark grey), 4290-4310 (medium grey), 4310-4330 (light grey), 4330-4350 (white), 4350-4370 (light blue), 4370-4390 (medium blue), 4390-4410 (dark blue), 4410-4430 (very dark blue), 4430-4450 (black), 4450-4470 (dark grey), 4470-4490 (medium grey), 4490-4510 (light grey), 4510-4530 (white), 4530-4550 (light blue), 4550-4570 (medium blue), 4570-4590 (dark blue), 4590-4610 (very dark blue), 4610-4630 (black), 4630-4650 (dark grey), 4650-4670 (medium grey), 4670-4690 (light grey), 4690-4710 (white), 4710-4730 (light blue), 4730-4750 (medium blue), 4750-4770 (dark blue), 4770-4790 (very dark blue), 4790-4810 (black), 4810-4830 (dark grey), 4830-4850 (medium grey), 4850-4870 (light grey), 4870-4890 (white), 4890-4910 (light blue), 4910-4930 (medium blue), 4930-4950 (dark blue), 4950-4970 (very dark blue), 4970-4990 (black), 4990-5010 (dark grey), 5010-5030 (medium grey), 5030-5050 (light grey), 5050-5070 (white), 5070-5090 (light blue), 5090-5110 (medium blue), 5110-5130 (dark blue), 5130-5150 (very dark blue), 5150-5170 (black), 5170-5190 (dark grey), 5190-5210 (medium grey), 5210-5230 (light grey), 5230-5250 (white), 5250-5270 (light blue), 5270-5290 (medium blue), 5290-5310 (dark blue), 5310-5330 (very dark blue), 5330-5350 (black), 5350-5370 (dark grey), 5370-5390 (medium grey), 5390-5410 (light grey), 5410-5430 (white), 5430-5450 (light blue), 5450-5470 (medium blue), 5470-5490 (dark blue), 5490-5510 (very dark blue), 5510-5530 (black), 5530-5550 (dark grey), 5550-5570 (medium grey), 5570-5590 (light grey), 5590-5610 (white), 5610-5630 (light blue), 5630-5650 (medium blue), 5650-5670 (dark blue), 5670-5690 (very dark blue), 5690-5710 (black), 5710-5730 (dark grey), 5730-5750 (medium grey), 5750-5770 (light grey), 5770-5790 (white), 5790-5810 (light blue), 5810-5830 (medium blue), 5830-5850 (dark blue), 5850-5870 (very dark blue), 5870-5890 (black), 5890-5910 (dark grey), 5910-5930 (medium grey), 5930-5950 (light grey), 5950-5970 (white), 5970-5990 (light blue), 5990-6010 (medium blue), 6010-6030 (dark blue), 6030-6050 (very dark blue), 6050-6070 (black), 6070-6090 (dark grey), 6090-6110 (medium grey), 6110-6130 (light grey), 6130-6150 (white), 6150-6170 (light blue), 6170-6190 (medium blue), 6190-6210 (dark blue), 6210-6230 (very dark blue), 6230-6250 (black), 6250-6270 (dark grey), 6270-6290 (medium grey), 6290-6310 (light grey), 6310-6330 (white), 6330-6350 (light blue), 6350-6370 (medium blue), 6370-6390 (dark blue), 6390-6410 (very dark blue), 6410-6430 (black), 6430-6450 (dark grey), 6450-6470 (medium grey), 6470-6490 (light grey), 6490-6510 (white), 6510-6530 (light blue), 6530-6550 (medium blue), 6550-6570 (dark blue), 6570-6590 (very dark blue), 6590-6610 (black), 6610-6630 (dark grey), 6630-6650 (medium grey), 6650-6670 (light grey), 6670-6690 (white), 6690-6710 (light blue), 6710-6730 (medium blue), 6730-6750 (dark blue), 6750-6770 (very dark blue), 6770-6790 (black), 6790-6810 (dark grey), 6810-6830 (medium grey), 6830-6850 (light grey), 6850-6870 (white), 6870-6890 (light blue), 6890-6910 (medium blue), 6910-6930 (dark blue), 6930-6950 (very dark blue), 6950-6970 (black), 6970-6990 (dark grey), 6990-7010 (medium grey), 7010-7030 (light grey), 7030-7050 (white), 7050-7070 (light blue), 7070-7090 (medium blue), 7090-7110 (dark blue), 7110-7130 (very dark blue), 7130-7150 (black), 7150-7170 (dark grey), 7170-7190 (medium grey), 7190-7210 (light grey), 7210-7230 (white), 7230-7250 (light blue), 7250-7270 (medium blue), 7270-7290 (dark blue), 7290-7310 (very dark blue), 7310-7330 (black), 7330-7350 (dark grey), 7350-7370 (medium grey), 7370-7390 (light grey), 7390-7410 (white), 7410-7430 (light blue), 7430-7

as molecule C loop and helix B) is mostly based on the symmetry related parts of the other molecules and has to be considered with caution.

There are also regions of higher B-factors than the rest of the molecule that are repeated in all molecules: the active site covering loops on both domains (between strand 2 and helix B and between strand 6 and helix G) and the solvent exposed extremity of domain 2 (loop between helix I and strand 7, end of helix J, Fig. 5.3 bottom right).

5.3.4 Active Site Conformation and Substrate Modelling

The active sites contain unexplained electron density that was modelled by adding a TRIS molecule, present in the crystallisation conditions. The TRIS molecules refined well in the active site of molecule A and D and very badly in the other two (very high B-factors and low quality electron density maps). Therefore TRIS was only modelled into these active sites and the active sites of molecule B and C were modelled with water molecules. As molecule A contains a TRIS in the active site and has the lowest B-factor and the best electron density map, only molecule A will be considered from now onwards (Fig. 5.4).

Oxygen O1 of TRIS is in hydrogen bonding distance to the backbone nitrogen of Val195 and to the hydroxyl oxygen of Tyr133. O1 is also close to sulphur of Cys194 (3.4 Å). Oxygen O2 is not in hydrogen bonding distance to any atom but relatively close to the backbone nitrogen atoms of Leu77 (3.5 Å), Val78 (3.8 Å) and Ala79 (3.8 Å) and could potentially establish transitional bonding given there is enough flexibility in the binding. Oxygen O3 is in hydrogen bonding distance to the amide nitrogen atom of Asn14. Here again the amide nitrogen and oxygen are close (3.6 and 3.8 Å). The amide nitrogen in TRIS is in a completely hydrophobic environment mainly determined by one γ - and the β -carbon atom of Val78 (3.1 and 3.6 Å) as well as the ϵ -carbon of Met42 (4.0 Å, Fig. 5.4).

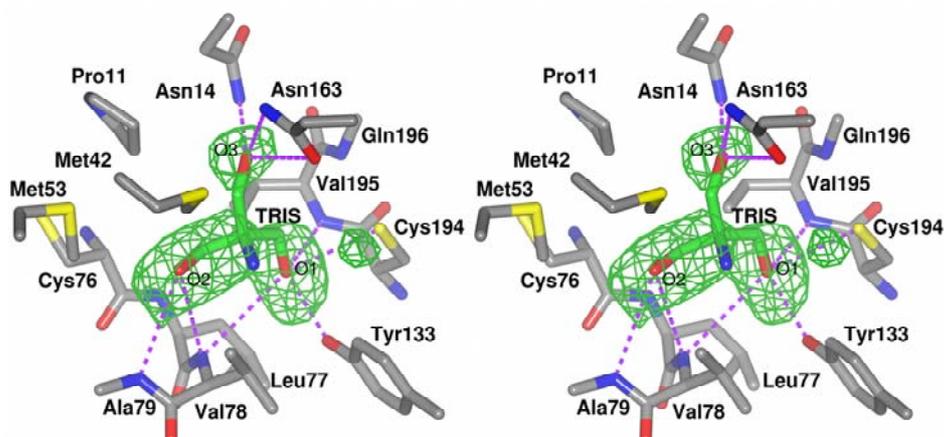


Figure 5.4. Active site. Stereoview of stick model showing active site residues (grey sticks) with the positive difference density map set to 3σ (green wire). The fitted TRIS molecule (green sticks) is shown with hydrogen bonds to active site residues (purple dashed lines).

The presence of the active site pocket was confirmed with the CASTp server that identified it as the biggest pocket measuring 237 \AA^3 having a surface of 260 \AA^2 . This fits well with TRIS that has a volume of 111 \AA^3 and a surface of 197 \AA^2 (Fig. 5.5).

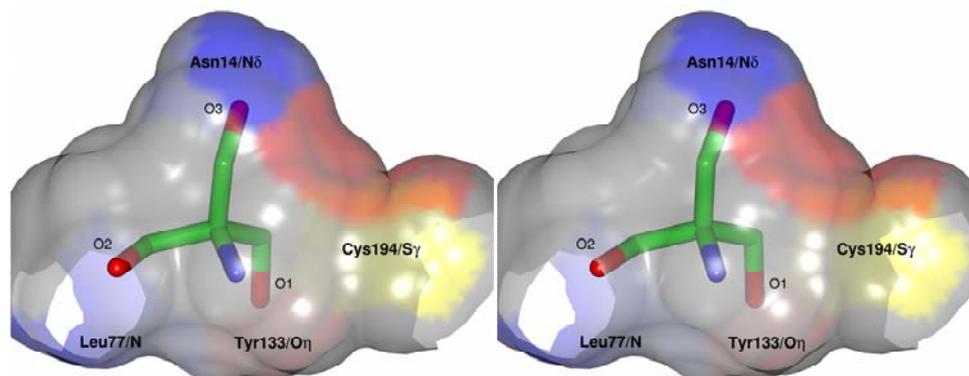


Figure 5.5. Active site pocket. Stereoview of TRIS (green sticks) bound to the active site.

There are currently 29 structures of 16 Asp/Glu racemase superfamily members in the protein structure database, 21 structures (12 sequences) of GluR, 4 structures (1 sequence) of AspR, 1 HydR structure, 5 AMD structures (1 sequence), 1 structure of unknown function. Although the sequence identities within GluRs vary from 36% to

68%, their protein fold is highly conserved. When aligned, the structures around the active site vary more according to the bound ligand than according to their sequence differences. In eleven GluR structures containing D-glutamate in the active site there are only two different conformations of the ligand. The proteins are all well aligned.

The dioxyanion hole has been described as a conserved feature in AspRs, GluRs and AMDs. The two negatively charged oxygen atoms of the enol intermediate (enediolate), formed after deprotonation or decarboxylation, are stabilised by six hydrogen bonds. Three hydrogen bonds are made with backbone amines; two with hydroxyls of threonine and serine or asparagine and the last with water or hydroxyl of tyrosine. These features are conserved in the sequence alignment of these activities. When MIs are aligned, these features are only partially conserved as the hydrophilic side chains of threonine and serine or asparagine are replaced by the hydrophobic side chains Leu77 and Val78 in MI removing two out of six hydrogen bonding possibilities (for alignment see Appendix C).

All the structures of different activities and containing different ligands could be structurally aligned on three conserved atom positions of this dioxyanion hole lying on domain 2. These are; the sulphur atom of Cys194, the nitrogen atom of Val195 and the η -oxygen of Tyr133 (equivalent to the oxygen atom of water between threonine and the ligand in racemases). Through the alignment, almost all ligands aligned at their potential enediolate position.

The structure of NfMI was aligned with one representative of GluRs, AspRs and AMDs containing a ligand. The chosen structures were *Enterococcus faecalis* GluR binding D-glutamate (EfGluR, PDB: 2vvt, resolution = 1.7 Å, $R_{\text{free}} = 0.18$) (151), *Pyrococcus horikoshii* AspR binding citrate (PDB: 2dx7, resolution = 2.0 Å, $R_{\text{free}} = 0.22$) (72) and BbAMD binding PO₄ (PDB: 3dg9, resolution = 1.5 Å, $R_{\text{free}} = 0.19$) (48). Oxygen (O1) of TRIS co-localises with the oxygen of the other ligands binding to nitrogen of Val195 and to the η -oxygen of Tyr133 (Fig. 5.6). As was already predicted from the sequence, the dioxyanion hole is only partially conserved as the two hydrophobic side chains of Leu77 and Val78 remove two possible hydrogen bonds in NfMI. In spite of this, the general structure of this part of the active site remains unchanged.

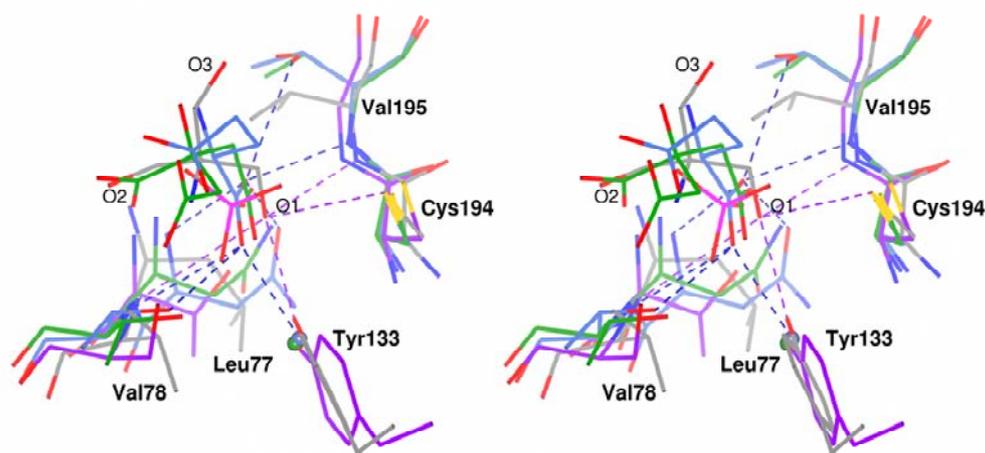


Figure 5.6. Overlay of dioxyanion holes and ligands. Stereoview of NfMI (grey) containing TRIS (grey), EfGluR containing D-Glu (blue, PDB: 2vvt), PhAspR (green, PDB 2dx7) containing citrate and BbAMD containing PO_4 (purple, PDB 3dg9) are superposed onto Tyr133 η -oxygen, Water molecules of GluR and AspR aligning with Tyr133 η -oxygen are shown as small spheres. Cys194 sulphur and Val195 backbone nitrogen. Hydrogen bonds involved in ligand binding are shown for NfMI (green broken line) and for GluR (blue broken line).

As expected, the two cysteines conserved in racemases are also structurally conserved in NfMI. The Cys194 is located on the more stable domain 2, the side of the active site that is structurally strongly conserved, whereas Cys76 is embedded in the more flexible domain 1, which changes according to the bound ligand in GluRs and contains the glycine in AMDs. Furthermore, the position of Cys76 differs considerably from the equivalent positions in the other enzymes. The backbone is about 1.3 Å further away from the Cys194 side, and the side chain is turned approximately 120 ° away from the active site compared to the situation in EfGluR pointing into the protein core. It is bound in a small, largely hydrophobic pocket formed by Val9, Pro11, Met40, Met53 and Tyr74. This results in a large distance of 4.3 Å between the positions of the equivalent sulphur of NfMI and EfGluR.

Based on the conservation of the dioxyanion binding site and the bound TRIS molecule, maleate and fumarate molecules were modelled into the active site. Fumarate and maleate both have a similar molecular volume (80 Å³) and surface (265 and 271 Å²) to TRIS and should therefore fit into the cavity as observed in the structure (Fig. 5.5). TRIS has three oxygen positions of which one is bound to the remnant of the dioxyanion whole (O1), one is close to Cys76 (O2) and one is bound to Asn14 in the back of the active site (O3). Moreover, the distance between two oxygen atoms of TRIS approximately corresponds to the distance between an inner (OA) and an outer

oxygen (OB) of each carboxyl group in maleate. Putting one oxygen atom of maleate into the conserved “dioxyanion” position and aligning it with the TRIS oxygen atoms, four possible positions for maleate remain. However, only two of the four positions do not result in major clashes. One position exposes the double bond carbon (C3) to Cys76 whereas the other exposes the second carboxyl group to Cys76. In both orientations the corresponding fumarate resulting from a simple superposition of the O1 oxygen atoms and the double bond slightly clashes with Met42. This does not take into account other binding possibilities of fumarate or active site rearrangements as there is in principle enough space in the active site and flexibility in the methionine side chain (Fig. 5.7).

As a second substrate modelling approach, docking using the software AutoDock was run. In this approach, no bias from the binding of TRIS is introduced as the substrate is fitted into a space grid defining the active site as probabilities of the presence of certain atom types. The 100 most likely positions of maleate all corresponded within a very low deviation to the orientation having the carboxyl group exposed to Cys76 (Fig. 5.7 top). According to AutoDock, the chance of the orientation having the double bond carbon C3 exposed to Cys76 (Fig. 5.7 bottom) is there for less than 1% (Data not shown). This result has to be considered with care as the flexibility of the protein was not taken into account.

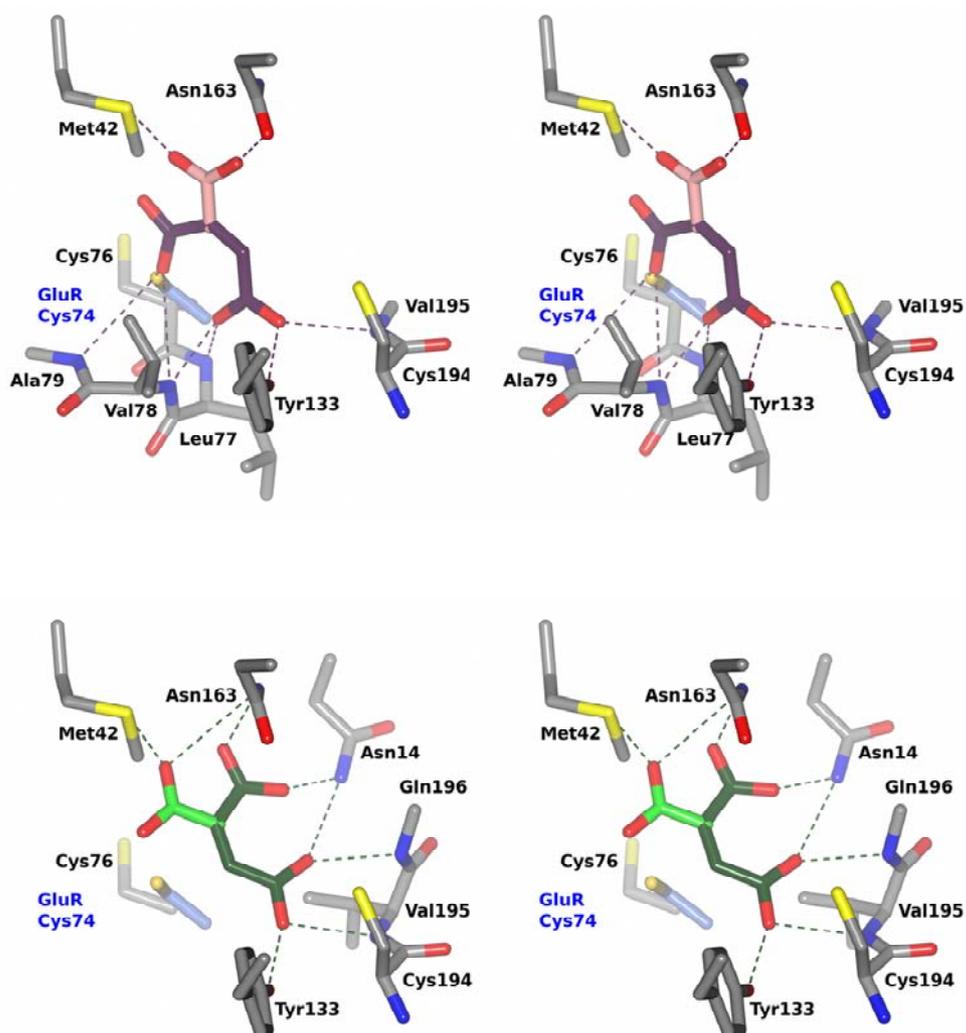


Figure 5.7. Scenarios of ligand binding. Maleate oxygens are superposed onto the TRIS oxygens in two orientations conserving the dioxyanion binding site and minimising clashes between atoms. Subsequently, fumarate molecules (light colours) are superposed onto the maleate molecules (dark colours). The relative position of EfgluR Cys74 is shown in blue. Top: Stereoview of orientation exposing the carboxyl group to Cys76. Bottom: Stereoview of orientation exposing the maleate double bond to Cys76.

5.3.5 Oxidation Resistance Involving Met197

By mutating a conserved methionine close to the active site to cysteine in *Serratia marcescens* MI (SmMI, Met197 in NfMI) it was reported that the enzyme became more resistant to oxidation by H_2O_2 whilst being only slightly less active (103). However, the authors did not give a mechanistic explanation for this. Despite the Val195 and Gln196 between Met197 and the active site Cys194, they are structurally adjacent to each other. Yet, the sulphur atoms are not in contact as Pro132 is intruding the space

between them. These are all residues that are conserved between the different sequences (see Appendix C). The sulphur atom of a cysteine in the same place could be in contact with Cys194 as it would pass below Pro132. Additionally Met197 is also in close proximity of the residues Ile169 between α -helix H and I and Val174 on helix I being close adjacent to the active site covering loop in domain 2 (data not shown).

5.4 Discussion

Despite the low sequence identity to AMDs and racemases, the structure of NfMI clearly shows that the enzyme is part of the Asp/Glu racemase superfamily. It contains its characteristic features like the pseudo dyad of two Rossmann fold domains. Also the two catalytic cysteines are situated at the same positions in small conserved stretches of the protein backbone (48,62,71). Like the other superfamily members, NfMI accepts a small carboxylic acid as substrate (dicarboxylic acid in all cases except for hydantoins). The bound substrate is completely buried under covering loops following the second β -strands on each domain, which shield the reaction from bulk water. In contrast to GluRs, AspRs and AMDs the dioxyanion hole is less pronounced and offers less stabilisation of the intermediate as in the former enzymes. MIs seem to share this loss of dioxyanion hole strength with HydR that even lack the Tyr as in BbAMD or Thr/water as in GluR and AspR for dienolate binding.

The contact area between two monomers is large enough to expect a stable dimer in solution. The crystal structure therefore confirms the dimeric state of NfMI observed by DLS. The interaction between the dimers is most probably an artefact from the crystallisation process that is helped by Ca^{2+} . To confirm more conclusively that this multimeric state is really present in solution, a calibration using gel filtration would be needed. If possible this would have to be done at low concentration more similar to physiological concentrations where the enzymes could still be present as monomers.

The structure allows speculation about the reason for the reduction of oxidation sensitivity resulting from the mutation of conserved Met197 to cysteine. One possibility would be the increase in size of oxidised Met197 interfering with the activity through displacement of the neighbouring catalytic Cys194 or Tyr133 involved in substrate binding. The proximity of Met197 to the covering loop on domain 2 could also mean that it does not properly close anymore. By reducing the size of the amino acid in this position these effects would not occur anymore. Alternatively, the thiol of cysteine

would be able to contact Cys194 below Pro132 allowing it to reduce the catalytically active form if it is inactivated by oxidation.

The high B-factor and poor electron density suggest a high flexibility in various parts of the protein, including the covering loops, the stretch containing Cys76 as well as the whole domain 1. This flexibility might be an explanation for the anisotropy of the crystal and thus the relatively low resolution of the X-ray diffraction. Although this could also simply be due to the elongated shape of the asymmetric unit, the flexibility can also be observed in other superfamily members. In one structure of BbAMD (PDB 2vlb) the four molecules in the asymmetric unit show significant conformational differences (88). The different glutamate racemase adopt different structures according to the binding of different ligands. Also in these enzymes the flexibility is especially pronounced in domain 1. The flexibility could therefore be an inherent property of the whole superfamily.

The proposed mechanisms of the decarboxylase and racemases do not largely depend on flexibility. Only the access of the substrate to the active site requires the covering loops to move. For NfMI, however, the conformation of maleate and fumarate differ substantially requiring a large space when converted, which involves a flip or rotation of 4.6 Å of oxygen OA around the double bond axis. The rotation of the carboxyl group around carbon C2-C3 bond would require an additional volume of 120 Å³, more than doubling the space requirement for the active site. For this movement to be possible, there must either be large space available or the enzyme has to be very flexible. Although the active site volume of 237 Å³ is numerically enough it does not account for shape. The observed flexibility is, therefore, probably crucial for the isomerisation activity. The inherent flexibility of the superfamily may have allowed the evolution of the new activities. Unfortunately, it is very difficult to observe the flexibility in an enzyme in action. Molecular dynamics simulation would be a way of showing possible movements in loops and in the active site.

Mutational studies have shown that the mechanism of isomerisation involves both cysteines (see Chapter 4) (104). This is also underpinned by the fact that the cysteines in the other superfamily members are also the catalytic residues. However, in both models for the possible orientation of maleate in the active site the substrate lacks direct contact with Cys76. Given the apparent flexibility of domain one and Cys76 with its surroundings it is reasonable to imagine that a reorientation occurs upon binding of the correct substrate. In this case, Cys76 would take a position more similar to that in

its homologues. As Cys76 is bound into a small hydrophobic pocket only little energy is necessary for breaking the interaction.

Considering the apparent flexibility of the protein, the models of the substrate binding are only speculative. However, the conservation of the dioxyanion hole conformation with Leu77, Cys194 and Val195 suggests that this binding site is conserved. For the decision on the orientation of maleate and fumarate, clearly more data are needed. The orientation found by the docking experiment strengthens the case for the orientation exposing the carboxyl group to Cy76. Crystals with bromomaleate inhibitor bound or crystals of active site mutants with the substrate bound should give more insight on the mode of substrate binding.

The implications of the structural data of the present chapter and biochemical data of the previous chapter for the reaction mechanism will be discussed in the general discussion (Chapter 6). Different scenarios for the reaction mechanism will be discussed.

Chapter 6

General Discussion

6.1 Evolution of the Asp/Glu Racemase Superfamily

Taken together, the sequence similarity, the conservation of the protein fold and the similarity of the mechanism reconfirm that the HydR, AMDs and MI are part of the Asp/Glu racemase superfamily. Superfamily is the appropriate term as clearly the homology is present, the reaction intermediate seems to be conserved but the activities are various (46).

It therefore becomes interesting to speculate about the evolution of the superfamily, even though this is not the main scope of this study. The phylogenetic tree shows clear grouping of the activities, which allowed the prediction the NfMI function (Fig. 1.21). From the failure of predicting the activity of the other homologues it can be concluded that the assignment of group activities is more detailed than previously thought. As there was no outgroup defined, the tree is not rooted. On the other hand GluRs are, at least in some organisms, essential for cell wall synthesis and had to be present for a long time in evolutionary terms. HydR, MI and AMD activities seem to be involved in the catabolism of special nutrients and are not absolutely required for survival. It is therefore reasonable to set the evolutionary origin of the superfamily at GluRs. This would explain the large sequence variation within GluRs compared to the other activities like in AMDs.

Starting from GluRs, only minor arrangements in the active site would have to be made to evolve into AspR. The catalysis is conserved completely and the substrate binding only changes by a little. To evolve HydR, the deprotonation and reprotonation mechanism would still be conserved but the stabilisation of the intermediate needs adjustments and the formation of the enzyme-substrate complex needs a much larger binding pocket. This change could have allowed the evolution of a very different mechanism involving rotation around a carbon-carbon bond such as with the MI. The resulting increased flexibility of the active site could have lead to the possibility of a

second carboxyl group to binding and the increased space could have allowed an aryl group to bind; the requirements for AMDs. The cysteine on domain 1 would then have lost its function and the mutated glycine allowing for more space. It has to be added that the natural substrate of AMDs is not known at present.

Given the pseudo 2-fold symmetry in the protein structures it has been hypothesised that these arise from an ancient gene duplication (143). No mention has been made of the original gene's function.

The tested homologues in Chapter 3 are not AMDs and Pfe seems not to be a MI. This corresponds well to the details in phylogenetic tree where the homologues form three distinct groups corresponding to the observation of their genomic context. The group of Sme and Mle are both from rhizobial species and embedded into possible amino acid degradation operons. The group with RseD, SceD and Sae containing an aspartate residue at the catalytic position on domain 2 are adjacent to their paralogue with a double cysteine (RseC, SceC and the paralogue of Sae that was not tested). Ste, Pfe, RseC, SceC make the last group. Three more different activities can thus be expected to be present in the Asp/Glu racemase superfamily.

6.2 Decarboxylation of Malonates

Of all the enzymes tested in the present work, only BbAMD was active against arylmalonates. Starting from the sequence annotations in the databases and the conserved active site residues for the active site, this was at first surprising. The fact that it was possible to confer racemase activity to BbAMD by simply mutating Gly74 to cysteine and subsequently change the product enantiospecificity of BbAMD by further replacing Cys188 to serine instigated initial confidence of being able to transform any superfamily activity into another superfamily activity (79).

Reconsidering the case, it appears much less surprising that the homologues lacked AMD activity and that Pfe does not catalyse other reactions (Fig. 6.1). The groups of the phylogenetic tree cluster well (Fig. 1.21). The sequence identities of below 30% are generally considered too low for predictions. The homologues have differences in the amino acids crucial to BbAMD activity. Given the completely different structural requirements of MIs and AMDs that are so close in the phylogenetic tree the failure of transforming RseC into a decarboxylase by one or two mutations is not surprising.

The superfamily enzyme structures seem to be very flexible over a long term evolutionary time scale but the enzymes have no apparent promiscuous activities and cannot be transformed to different catalytic activities by simple point mutations. The binding of the substrate to the active site results in a complete covering of the substrate by the covering loops requiring these loops to be very flexible. This was observed in crystals structures of superfamily members that were present in open and closed forms, had changed loop conformations in different molecules of the asymmetric unit and high B-factors of the loops. The requirements for these loops are difficult to rationalise in terms of point mutations. The superfamily would therefore probably lend itself more to directed evolution approaches than to rational engineering. However, an efficient selection process would be needed that allows for reliable colony screening and/or screening of multi-well plates. A better understanding of the colorimetric decarboxylation assay, not applied in this study, would be useful for high throughput approaches.

Alkylmalonates are neither converted by BbAMD nor by any of the homologues or mutants of RseC and BbAMD. This is surprising in the sense that there is a great variety of sequences and activities in the Asp/Glu racemase superfamily. Thus, one would have expected that there are at least residual activities towards these compounds. On the other hand, the chemistry of decarboxylation in general requires enzymatic mechanisms that involve activation by oxidation or binding to cofactors such as coenzyme A, phosphopantetheine, thiamine diphosphate, pyridoxal phosphate or biotin. Retrospectively it is therefore reasonable to expect some restriction in the range of substrates accepted.

Decarboxylation results in the transfer of one electron to the α -carbon, which has to be stabilised. In AMDs the dioxanion hole together with the hydrophobic pocket for the carboxylate are responsible for the destabilisation of the cleaved carboxylate as well as the stabilisation of the decarboxylated intermediate. However, the additional electron withdrawing capacity of the aryl- or alkenyl group is required. The electron donating effect of alkyl groups would be expected to destabilise the intermediate and therefore to deactivate the leaving carboxyl group making it more difficult to decarboxylate.

Table 6.1. Summary of tested alternative activities.

Enzyme	Arylmalonate decarboxylation	Alkylmalonate decarboxylation	Maleate isomerisation	Asp/Glu racemisation	PP racemisation
BbAMD	+	0		0	
RseC	0	0		0*	0
Pfe	0	0	0	0	0
NfMI	0		+		

+ Performed successfully/activity detected

0 No activity detected

* Only indirect evidence available

These reasons taken together seem to suggest that the only viable routes to the production of chiral alkylalkanoates are substrate engineering or utilisation of a cofactor-dependent enzyme. Substrate engineering would involve the introduction of an electron withdrawing group as substituent. Alkenyl groups seem to be very suitable but a range of other groups could be considered introducing other functions that would need to be converted to aliphatic chains if required (67). The use of cofactor-dependent enzymes is unattractive because of the cost of cofactors; however, the use of whole cell systems using altered metabolic routes in the original host or engineered pathways in heterologous hosts could represent a feasible alternative. The fact that the substrate has no problem entering the cell is, in principle, encouraging in this respect. Whether engineering the whole metabolism is possible depends on the nature of the enzyme system. Factors such as the solubility of the enzymes, requirements for subcellular location and inhibition will play a role.

In another alternative approach to mining known genomes, the microbial diversity of the soil was mined using selective enrichment cultures (Appendix D). This shows that there are indeed organisms capable of growing on alkylmalonates. Because of time constraints, the isolates could not be characterised further to determine the responsible biocatalysts or to determine whether decarboxylation is indeed the first step in the degradation. It was shown that there exist phosphopantetheine-dependent decarboxylase pathways for the decarboxylation of malonate in soil bacteria (30,31,152). It is therefore possible that the identified soil isolates are using the same or a similar pathway. The results show in addition that there are microorganisms capable of growing on these substrates making it more likely to find a microorganism suitable for engineering.

6.3 Reaction Mechanism of NfMI

The *cis-trans* isomerisation of maleate to fumarate must involve a free rotation of one carboxyl group around the bond between carbon C2 and C3. It will therefore involve a reversible saturation of the double bond during the course of the reaction. Cys76 and Cys194 are conserved catalytic residues in the Asp/Glu racemase superfamily and were shown to be indispensable for the activity of NfMI. They can therefore be assumed to be the catalytic residues. At the pH range where the enzyme is active maleate and fumarate are deprotonated.

The racemisation and decarboxylation reaction mechanisms of the superfamily share a common enediolate intermediate that is stabilised by the conserved dioxyanion hole of the active site. A possibility to achieve the formation of an enediolate from maleate through the action of a cysteine would be a nucleophilic attack by the thiolate on one of the double bond carbons. The enediolate intermediate could then be stabilised by four hydrogen bonds of the remnant dioxyanion hole and by the other cysteine (Fig. 6.2 A).

As replacing Cys76 with serine leads to complete loss of activity whereas replacing Cys194 with alanine results in a low residual activity, it is more likely in this case that Cys76 would be performing the nucleophilic attack. The flexibility of Cys76 and that side of the active site would be necessary for the movements involved in rotating the free carboxyl group around the carbon C2 and C3 bond. The result of docking the substrate into the active site led to the preference of the carboxyl group exposed to Cys76 rather than the double bond, which seems to speak against this mechanism. Then again, Cys76 would have to undergo less movement during the rotation by attacking C3 from *re*-face that is exposed in this orientation than by attacking from the *si*-face.

Succinylations of cysteines have been reported as posttranslational modification of proteins (153). It could therefore be imagined that a nucleophilic attack on C3 of maleate could lead to a succinyl intermediate without the need for stabilisation. In this case the carboxyl at the dioxyanion hole would be free to rotate instead of Cys76 turning around with the rest of the molecule. The thiol of Cys194 could be responsible too for the protonation of the intermediate after the nucleophilic attack. The protonation would be favoured as the dioxyanion hole is less electron withdrawing than in decarboxylases and racemases leaving more negative charge at C2 (Fig. 6.2 B).

The fact that maleamide and citraconate are not substrates for NfMI could be explained by the hypothesis of a nucleophilic attack as both methyls and amides are either electron donating or less electron withdrawing than hydrogen and carboxylate. On the other hand, bromomaleate would not be expected to be a competitive inhibitor with this mechanism as bromine is an electron withdrawing group, exposing C3 more to nucleophilic attack than in the case of maleate. This would either result in an improved substrate or an irreversible inhibitor rather than a competitive inhibitor. There is also no evidence by MS for a covalent adduct of the substrate to NfMI during incubation at alkaline pH.

As cysteine is generally protonated at physiological pH a mechanism for deprotonation of the attacking thiol has to be provided. In cofactor-independent racemases this is not fully understood either. It has been proposed that a neighbouring aspartate or the N-terminal dipoles adjacent α -helices would stabilise the deprotonated state (63). There are no adjacent acids to Cys76 but α -helices D and F do point to its position. In addition, Cys76 is bound in a hydrophobic pocket that could provide an environment to lower its pK_a . Similar destabilisation of an active site lysine was observed in acetoacetate decarboxylase (36).

Other than by thiol addition, chemical catalysis can also be achieved through radicals or with acids. Although not related to MIs, the enzymes isopropylmalate isomerase (able to dehydrate malate and hydrate maleate) and fumarate hydratase are catalysing this reaction (154-156). The cysteines of NfMI could therefore also act as general acid and base, activating water, which then performs a nucleophilic attack on the double bond. The other cysteine could then stabilise the negative charge or fully protonate the intermediate (Fig. 6.2 C). The fact that TRIS binds to the active site with three alcohols might indicate the binding of water besides the substrate. In spite of this, there is no evidence for a stable intermediate during the reaction as all maleate was stoichiometrically converted into fumarate without the appearance of other HPLC peaks during the reaction.

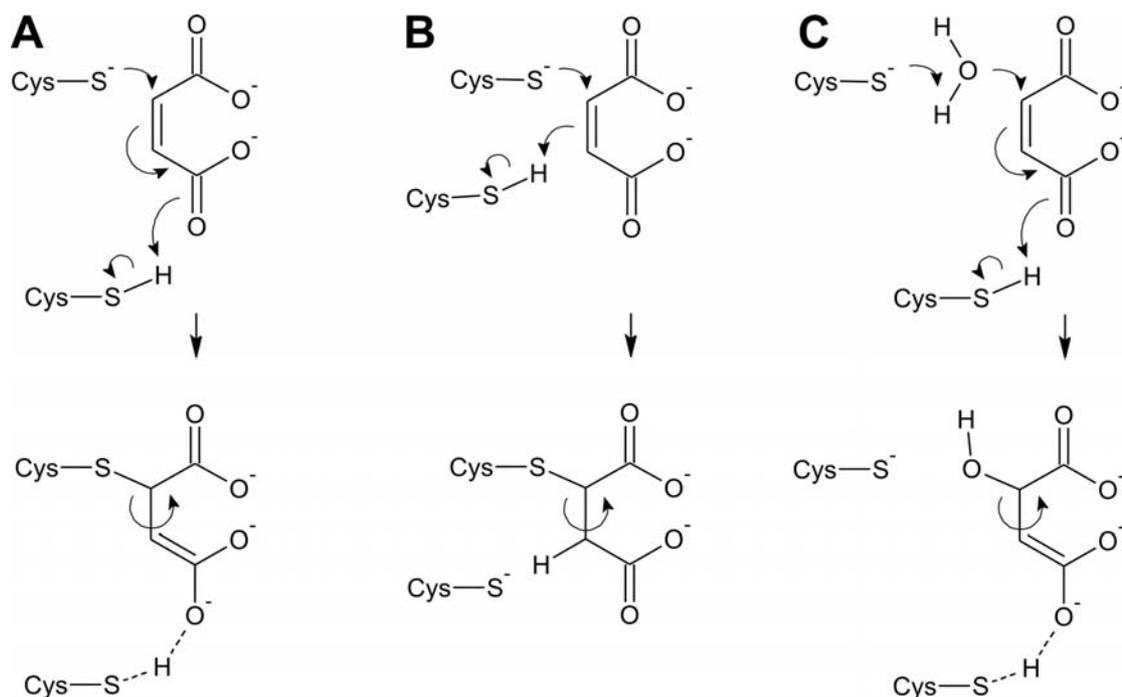


Figure 6.2. Three possible reaction mechanisms. (A) Mechanism conserving the role of the dioxanion hole, initiated by a nucleophilic attack on a double bond carbon. (B) The same as (A) but with a succinyl intermediate independent of dioxanion stabilisation. (C) General acid/base mechanism involving the activation of a water molecule.

The behaviour of bromomaleate is surprising in that it is not a substrate and is an inhibitor. It cannot be determined at the moment whether this is because of a steric effect rather than an electrostatic one. The size of bromine could restrict the access of the cysteine to the carbon or simply interfere with substrate binding. A simple steric effect is unlikely as citraconate is not an inhibitor but it could be a combined effect. Fluoromaleate would be a good candidate to test this as it is isosteric with maleate and isoelectric with bromomaleate. If it is an inhibitor similar to bromomaleate then the electrostatics is everything. If it is a substrate then electrostatics has to be combined with size. Solving the structure of the complex of NfMI bound to bromomaleate could also disentangle the two effects. Some weakly diffracting crystals have already been obtained.

6.4 Outlook

The main target of the work with NfMI is the elucidation of the mechanism as this would improve the understanding of the whole Asp/Glu racemase superfamily. Knowledge of

the mode of ligand binding is crucial to this understanding. It has to be determined whether the ligand binds covalently to the protein and if it binds to the cysteine. The range of possible functions of the cysteines could be narrowed down by determining their pK_a by plotting the pH versus the k_{cat}/K_M . The role of the cysteine could also be narrowed down by complementing an inactive mutant with a soluble thiol. If a thiol complemented a lacking active site cysteine this would show that the thiol acts in a general way not requiring the precise microenvironment and positioning, rather than stabilising an intermediate with precisely orientated hydrogen bonds. The determination of the ligand binding mode by crystallography would also allow differentiating between different attacking points of the active site cysteines with the whole protein environment, including the flexible loops and Cys76. Assisting residues can more easily be identified in substrate/enzyme or inhibitor/enzyme complex, which would further illuminate the mode of action. Most importantly the transition state of the reaction and the mode of stabilisation by the enzyme have to be determined. It is not clear what part of the substrate is rotating and how it is allowed. Molecular dynamics simulations could show more precisely how flexible the covering loops are and what portion of the active site allows the carboxyl to rotate.

Once the mechanism is better understood, it would be possible to focus on an extension of the substrate specificity. Substrates used unsuccessfully in this work, such as citraconate or mesaconate would lend themselves as potential new substrates. An enzyme with an extended substrate range would be highly interesting for industrial applications as it has the potential capacity of changing the specificities of transformations. This could be applied in enzymatic resolution biotransformations. Maleate and fumarate are both symmetrical molecules and in this respect less interesting for industry. Both citraconate and mesaconate are small variations of maleate and fumarate making them asymmetrical. Substituting the double bond could then be performed in an asymmetric manner. An enzyme with a broader substrate range also has the advantage of being suitable for directed evolution approaches as these require residual activities to start with.

Enzymes from the Asp/Glu racemase superfamily are potentially ideal tools for industry as they are small bacterial proteins, easy to produce and independent of cofactors. GluRs and AMDs have already met broad interest. Elucidating the activity of the BbAMD homologues represents an opportunity to find more activities in a group of proteins that are easy to handle. Unfortunately it is a difficult task to test for unknown

activities. Therefore, this highlights the importance of understanding the mechanism of MIs as unusual members of the superfamily.

The common characteristics of all reactions could be taken as starting point for the discovery of new activities. All substrates are small molecules (<200 Da) with a carboxylate at one end, HydR being the exception. All reactions proceed through a stabilised enediolate intermediate that is protonated during the reaction, MIs may be different. If a reaction fulfilling these properties is required, it is worth testing the substrates on the cloned homologues. Other approaches could include screening for ligands by NMR or computationally model ligands into the active site. The computational approach could be supported by restricting the possibilities using the conditions mentioned above. It has been shown that it is possible to engineer completely new activities on suitable protein scaffolds combining the power of computational protein modelling with directed evolution approaches if the appropriate chemical reaction partners and stabilising residues for the reaction are given (157). Reaction partners for maleate isomerisation have been identified for a long time already (108-110) and a nearly perfect protein scaffold is known for NfMI now. It should therefore be possible to perform the reverse process and identify the role of the catalytic residues computationally. Using similar mechanisms, it should also become possible to assign functions to unknown enzymes in the Asp/Glu racemase superfamily.

If the decarboxylation of alkylmalonates rather than harnessing the superfamily is the primary goal, an analysis of the isolates growing on these substrates, described in the Appendix D, would be necessary. An analysis of the pathway could involve feeding of intermediates to test whether they are accepted as substrates or analysing a mutant library able to grow or not on the substrate or intermediates. Against the evidence shown in Appendix D, the activity could still be due to a single soluble enzyme. If this is the case, it could be possible to isolate the activity by targeted purification of the catalyst and sequencing of the protein to design degenerate primers able to amplify the gene from genomic DNA.

Appendix A

BbAMD Gene Sequence

720 bp codon-optimised gene without stop codon.

160
ATGCAGCAGGCGAGCACCCCGACCATTGGCATGATTGTGCCGCCGGCAGCGGGTCTGGTT
CCGGCGGATGGCGCGCGTCTGTATCCGGATCTGCCGTTTATTGCGAGCGGCCTGGGCCTG
GGTAGCGTTACCCCGGAAGGCTATGATGCGGTGATTGAAAGCGTGGTGGATCATGCGCGT
CGTCTGCAGAAACAGGGTGCGGCGGTGGTGAGCCTGATGGGCACCAGCCTGAGCTTTTAT
CGTGGTGCGGCGTTTAAACGCGGCGCTGACCGTGGCGATGCGTGAAGCGACCGGTCTGCCG
TGCACCACCATGTCTACCGCGGTGCTGAATGGTCTGCGTGCGCTGGGTGTGCGTCGTGTT
GCGCTGGCCACCGCGTATATCGATGATGTGAACGAACGTCTGGCCGCGTTTCTGGCCGAA
GAAAGCCTGGTGCCGACCGGTTGTCGTAGCCTGGGCATTACCGGCGTGGAAGCGATGGCG
CGTGTGGATAACCGGACCCTGGTGGATCTGTGCGTGCGTGCCTTTGAAGCGGCACCGGAT
AGCGATGGCATTCTGCTGTCTTGCGGCGGTCTGCTGACCCTGGATGCGATTCGGAAGTG
GAACGTCGTCTGGGCGTGCCGGTGGTTAGCAGCAGCCCGGCAGGCTTTTGGGATGCGGTG
CGTCTGGCCGGTGGTGGTGCGAAAGCGCGTCCGGGCTATGGCCGTCTGTTTGATGAAAGC

Appendix B

Sequence of pET-YSBLIC3C Plasmid

5399 bp sequence based on pET-28a(+) with altered multiple cloning site. The BseRI binding site is underlined. Two nucleotides are replaced by the insert during LIC (between | and |). The added N-terminal His-tag sequence is shown in single letter code including The HRV-3C protease binding site (><).

```

1                                                                 60
CTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTT
CCTTTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAG
GGTTCGGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTT
CACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT
TCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATT
CTTTTGATTTATAAGGGATTTTGCCGATTTTCGGCCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTCAGGTGGCACTT
TTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGT
ATCCGCTCATGAATTAATTCTTAGAAAACTCATCGAGCATCAAATGAAACTGCAATTTA
TTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAA
AACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACT
CGTCCAACATCAATACAACCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAG
AAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAAGTTTATGCATTTCTTTC
CAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAA
CCGTTATTCATTCGTGATTGCGCCTGAGCGGAGACGAAATACGCGATCGCTGTTAAAAGGA
CAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATA
TTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCA
GTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTTCGGAAGAGGC
ATAAATCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTTGGCAACGCTA
CCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCATAACAATCGATAGATT
GTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCC
ATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTTCCCGTTGAATATGGCTCATAACA
CCCCTTGATTACTGTTTATGTAAGCAGACAGTTTTTATTGTTTCATGACCAAAATCCCTTA
ACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG

```

1501

1560

AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGC
GGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAG
CAGAGCGCAGATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
GAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGC
CAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTGCGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTA
CACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT
TCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGGTTTCGCCACCTCTGACTTGA
GCGTCGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
GGCCTTTTTTACGGTTCTTGGCCTTTTTGCTGGCCTTTTTGCTCACATGTTCTTTCTGCGTT
ATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCG
CAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCG
GTATTTTTCTCCTTACGCATCTGTGCGGTATTTACACCGCATATATGGTGCACCTCTCAGT
ACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTCCGCTATCGCTACGTGACT
GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTG
TGCTCCCGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGA
GGTTTTTACCAGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGT
CGTGAAGCGATTACAGATGTCTGCCTGTTTATCCGCTCCAGCTCGTTGAGTTTCTCCA
GAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCTGTT
TGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTATGGGGTAATGATAACCGATGA
AACGAGAGAGGATGCTCACGATACGGTTACTGATGATGAACATGCCCGGTTACTGGAAC
GTTGTGAGGGTAAACAACACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGG
GTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATC
CTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTT
ACGAAACACGGAAACCGAAGACCATTATGTTGTTGCTCAGGTGCGAGACGTTTTGTCAGC
AGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTCACTCTGCTAACCAGTAAGGCAAC
CCCGCCAGCCTAGCCGGTCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCG
CCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGA
AGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTGC
CGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCTTA
CGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCC
ACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCT
AATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAA

3541 3400

ACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTGCGTA
TTGGGCGCCAGGGTGGTTTTTCTTTTCACCCAGTGAGACGGGCAACAGCTGATTGCCCTTC
ACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTTGCCCCAGCAGGCGA
AAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCG
TATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATT
GCGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTC
AGCATTGTCATGGTTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCT
ATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCC
GAGACAGAACTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGA
TGCTCCACGCCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTC
TGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATG
GCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGA
TTGTGCACCGCCGCTTTACAGGCTTCGACGCGCTTCGTTCTACCATCGACACCACCACG
CTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGC
AGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGT
GCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTT
TTCGAGAAACGTGGCTGGCTGGCTTACCACGCGGAAACGGTCTGATAAGAGACACC
GCATACTCTGCGACATCGTATAACGTTACTGGTTTTACATTCACCACCCTGAATTGACTC
TCTTCCGGGCGCTATCATGCCATACCGCGAAAGTTTTTGCGCCATTGATGGTGTCCGGG
ATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAG
GCCGTTGAGCACCGCCGCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCC
CCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTG
GCGAGCCCAGTCTTCCCATCGGTGATGTGCGCGATATAGGCGCCAGCAACCGCACCTGT
GGCGCCGGTATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCG
AAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATA
ATTTTGTTTAACTTTAAGAAGGAGATATAACCATGGGCAGCAGCCATCATCATCATCA

M G S S H H H H H H

CAGCAGCGGCCTGGAAGTTCTGTTCCAGGGACCAGCAA | GG | CGCGCCTTCTCCTCACAT

S S G L E V L F Q <G P A

ATGGCTAGCATGACTGGTGGACAGCAAATGGGTGCGCGGATCCGAATTCGAGCTCCGTCGA
CAAGCTTGCAGCCGCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAA
AGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCT
TGGGGCCTCTAAACGGGTCTTGAGGGTTTTTTGCTGAAAGGAGGAACCTATATCCGGAT

Appendix C

Alignment of Asp/Glu Racemase Superfamily Members

Sequences of all selected homologues (see Chapter 3, Table 3.1), all experimentally confirmed activities and all known structures were aligned using ClustalW2. The highly conserved stretches from the active site used to build the ClustalW2 guide tree (Fig. 1.21) are marked with a line on top of the alignment (for details see Chapter 2).

The positions of the conserved catalytic cysteines are marked with # on top of the alignment. The active site residues of BbAMD are marked in underlined bold face. UniProt accession codes are followed by the name and X if structure coordinates are available. Sequences discussed in results chapters are underlined.

Abbreviations used are: AfMI, *Alcaligenes faecalis*; PhAspR, *Pyrococcus horikoshii*; ApGluR, *Aquifex pyrophilus*; AsAMD, *Achromobacter* sp. KU 1311; BaGluR, *Bacillus anthracis*; BbAMD, *Bordetella bronchiseptica*; BpGluR, *Bacillus pumilus*; BsGluR, *Bacillus subtilis*; BsMI, *Bacillus stearothermophilus*; EcAMD, *Enterobacter cloacae*; EcGluR, *Escherichia coli*; EfAspR, *Enterococcus faecium*; EfGluR, *Enterococcus faecium*; EfGluR, *Enterococcus faecalis*; HpGluR, *Helicobacter pylori*; LbGluR, *Lactobacillus brevis*; LcAspR, *Lactobacillus casei*; LdAspR, *Lactobacillus delbrueckii*; LfGluR, *Lactobacillus fermentum*; LmGluR, *Listeria monocytogenes*; LpGluR, *Lactobacillus plantarum*; Mle, *Mesorhizobium loti*; MsAMD, *Mesorhizobium* sp. BNC1; NfMI, *Nocardia farcinica*; Pfe, *Pyrococcus furiosus*; PhAspR, *Pyrococcus horikoshii*; PhHydR, *Pyrococcus horikoshii*; PpGluR, *Pediococcus pentosaceus*; PsHydR, *Pseudomonas* sp. NS671; RseC, *Rhodococcus* sp. RHA1; RseD, *Rhodococcus* sp. RHA1; Sae, *Streptomyces avermitilis*; SaGluR, *Staphylococcus aureus*; SceC, *Streptomyces coelicolor*; SceD, *Streptomyces coelicolor*; ShGluR, *Staphylococcus haemolyticus*; Sme, *Sinorhizobium meliloti*; SmHydR, *Sinorhizobium meliloti*; SmMI, *Serratia marcescens*; SpGluR, *Streptococcus pneumoniae*; SpGluR, *Streptococcus pyogenes*; Ste, *Sulfolobus tokodaii*.

201 # 300

Q8Y7N7 LmGluR X FVSVVESGSEYKS-AIAKKVVAESLLP--LKSTKIDTILGCTHYPLLPKPIIENFMGDGVAVINS--GEETASEVSALLDYHNLDDATD-EIEHRRFFTTG
Q6GHT5 SaGluR X FVPLVEQMRYSDDPTITSIVIHQTLKR--WRNSEDVTILGCTHYPLLYKPIYDFYGGKKTIVISS--GLETAREVSALLTFSNEHASYT-EHPDHRFFATG
P52974 ShGluR X FVPLVEQMRYSDDPTITSIVIHQTLKQ--WRNTDADTILGCTHYPLLYKPIYDFYGGKKTIVISS--GLETAREVSALLTFSNEHASYT-QHPDHRFFATG
Q81LA8 BaGluR X FVELVESGNFES-EMAYEVVRETLQP--LKNTDIDTILGCTHYPLILGPVIKQVMGDKVQLISS--GDETAREVSTILYHSKMLNEGE-EQSDHLFLTTG
Q3XZW8 EFgluR X FVSVVESNEYS-SVAKKIVAETLAP--LTTKKIDTILGCTHYPLLRPIIQNVMGENVQLIDS--GAETVGEVSMLLDYFNLSNPQNGRTLCOQFYTTG
Q81UL8 BaGluR X LATVVENRLEDT-AYVTQVQKQALLP--LTKEDIDTILGCTHYPLLESYIKKELGEDVTI ISS--AEETAIELSTILQHKGILADNL--NPKHRFFTTG
P94556 BaGluR X LVFPVESGKFLD-KTADLIVKTSLYP--LKDTSIDSLILGCTHYPLKKEAIQRYMGEHVNI ISS--GDETAREVSTILYKGLLNQSP-IAPDHQFLTTG
Q7M0V6 BpGluR X LVFPVESGTFLD-QTAEAVVKSLEP--MKETGIDTILGCTHYPLKKEPIQRFMGSDVSI ISS--GDETAREASTILYKGLLNQSK-ETPVHTFYTTG
Q836J0 EFgluR X FVPIVESNQYRS-SVAKKIVAETLQA--LQLKGLDITILGCTHYPLLRPVIQNVMGSHVTLIDS--GAETVGEVSMLLDYFDIAHTPEAPTQHPHFYTTG
Q88V19 LpGluR X FVPLVESNEYQS-TVAKRVVAETLQ--LKKQDVDTLVLGCTHYPLLRPIIQNVMGPGVTLIDS--GAETVNDVSAVLDYLDIANDRSTKRYPDEYTTG
P63640 SpGluR X FAPLVESGALST-SVTKKVVYETLRP--LVG-KVDSLILGCTHYPLLRPIIQNVMGPKVQLIDS--GAEVCVRDLSVLLNYFENRGRDAGLHHRFFYTTA
Q9A1B7 SpGluR X FVPIVESNEMCS-SIAKKIVYDSLAP--LVG-KIDTLVLGCTHYPLLRPIIQNVMGPKVQLIDS--GAEVCVRDLSVLLNYFENRGRDAGLHHRFFYTTA
Q03469 LfGluR X MVEIVEHGQGTG-AKAQEVVSEQLMT--FKEHPVKTLIMGCTHFPPLAPEISKAVGPTVALVDP--AKETVATAKSWLEQHQAMGNHA--HPNYHLYSTG
P48797 LbGluR X MVTLAEQNLDLT-THAQSVVAANLAS--LMDKKIDTLVGMCTHFPPLRSALQHAVGSOVTLVDP--GLATAEQTVAILKTRGLLNSAT-TRGTAQFFTTG
Q08783 PpGluR X FVETAEKNELHT-TAAQKVMNEKLAE--FRQDQIDTILGCTHFPPLLEEGIQAAVGPVDTLVDP--GVETVQLEIELLTKQALQHAEG-PKADQOYFSTG
P56868 ApGluR X FVPLAEEGLEG--EITRUVVHYHLK--EFGKIDTILGCTHYPLLRKKEIKKPLGDVEVVDSEALSLSLHNNPKDDGSSSLELFPDLSLNLQFLIK
Q9ZLT0 HpGluR X FVPLIEESILEG--ELLETCMHYFTT--PLEILPEVILGCTHFPPLIAQKIEGYFMGHFALPFPPLLIHSGDAIVEYLQKQYALKNNACTFPKVEFHASG
P22634 EcGluR X MVELAEAK-LHGEDVSLDALKRILRPLWRMKEPPDVTVLGCTHFPPLQBELQLPEGTRLVDS--GAAIARRTAWLLHEEAPDAKADANAFPCAMMAT
059384 PhAspR X ELNRIIFEELAFGNLKNKEIWRLEIKYRESEGEVILGCTELPLAIKQ-----GDVSVVEVFD-----AEIHMRKLLI ELASE-----
058403 PhAspR X VMR-GYEGVKAGNLKLGRELLKTKAKILBERGAEIC IAGCTEVSVVLLKQ-----DDLKVPILIDP-----MDVIAEAVAVKVALEK-----
P29079 EFAspR X KINLYIYHEIKESDHLNQELYYEILEEAEVRLNCEKVLSELMLNFAEDN--HYPVIDA--QSI LADRTIERALAEARN-EALDVTSEK-----
B3W8K5 LcAspR X MVTELYKYVKEKGEVNAALYHRLQRMHDFNVNVIIVLGCETSLPAQKAGDH--PVHVIDA--QSI IVDKIALGKAFROSEAGKALLNQMMLAR--
Q1G8B6 LdAspR X MVTELYRDIKEKGIIVDKELYHRLRKMHDYGGCDVILGCTELSLAQQAQPDH--PVQVIDP--QSI IADYSIELLALAIRNGQDPRQAQCKVLYD--
058781 PhHydR X HPSGVSNLTDLLTDWGRREVINAARK--LKEKGVVIALGCTGMSTIGIAPVLEEVEGVPVDP--VIASGAVALHALKRREVKRPFGR-----
Q00924 PsHydR X TTPMGVLDFERDPEAGIEMLRQEGKRAVED--NAEAILGCGAGMAEF--ADSLKELGVPVDP--VVAGVKFAETIVDLGKKTSLKTKYKPEKKEYVG
Q92ML1 RmHydR X AIDLPLVALEEDPQRAERLLLEKEIEIAKAED-GAEAIVLGCAGMSSL--CDRLQKATGVPVDP--VTAAVKMAEALLGAGYATSKVNTYAYPRKAAAG
05YXQ1 NfMI WRALVADNTEVGCIPGEQVMAAARSLLDSE--VDALVISCCVQMPSLPVVETAREFPIVLSA--ATAGAYSILRSLDPLVAVPGAGRLLRQDSAVTAS
Q9KW10 SmMI WRALVADNTEVGCIPGEQVMAAARSLLDSE--VDALVISCCVQMPSLPVVETAREFPIVLSA--ATAGAYSILRSLDPLVAVPGAGRLLRQDSAVTAS
024766 AfMI SISLEVSDNLEVGLLNPENLEHVKRLNHDG--VDAVILSACVQMPSLPAIQRAQDQIGKPVLSA--AVWTVYQMLKNLETRVFNAGHILSGVKQPA--
Q9WX57 BsMI SISLEIPDNLEVGRQDPMRLIEIVKNDVSN--ADAVVLSACVQMPSLPAIQKQVQDQGLPVLISA--ATSTVVKLLKSLNLTYPVNGSLLSGKY--
Q987A4 Mle FTCLGFEDDEREMARI PPAALVDLARQVTDQP--ADALFVSCALR GALAVTGMQEAIGRPVVTSS--NQATAMNCLRLCGDETP--RPFQGLMTLPLPRD
Q92WC5 Sme FTDCLGLTDDREMARISPKIEVAFAKEAAPA--SDALFISCTAVRAAGVIAIEIQAIGKPVVSS--NYATAWACLRLCGDREA--RPQLGRMLMELPLEE
Q92ML1 RmHydR X YDGLGIRGIDISNTPIFTIYRLVKRHLNEVLKADAVYIACLTALSTYEAQVYLEDHLDMPVVS--NAAMAMWALNKLKIKAKLPGF-----
Q8U183 PFe IRLGLIEDNLEIGRLESPYIPYMLAKSMF--MEEADGVFISCTNLRFFETI EKLERDLGIPVVTSS--NQATLWYALREI DVRES-LPLGLKLLREF-----
Q0S7X3 RseC SDHLGLGG--GIWKVNYRTIAERIIAADDP--SEAI FVSCNTLPTVDVIAPLEQLGKPVLTSS--NQLTIWACLGRMKLPMTGPGKWLNRNV-----
Q93J74 SceC CAMFGLTR--HIWKVYPRDVAEMARQAVPAPGAADALFISCTNLPYDVI PQLAEALRIPVLSA--NQVTMWAALRRGLTPAVGFGYQALTDPAARLGPAA
Q0S7X2 RseD MSSAGIDTAAEVGLLSPESVVELAVENDHDP--AEALLIPDTAMRTL GALSTLEQRLGKPVLTSS--NQVTIWEGLRLAGHTAV--HPSLGTLFKEGHSHG
Q93J75 SceD VVSSGIITAAEVGTWTEAEVLALARAADRPD--VEAVLLPDTALHTAAH I PALEKELGKPVLTSS--NQVTIWEGLRLADRRVN--APELGAFTREPVVQ
Q82CV3 Sae WRSSGIITAAEVATWDEEQVLALARAADHPE--AEAVLLPDTALHTASHIQALEAALGKPVLTSS--NQVTVNEALRLADRRVN--APRLGTLFTKEPLVQ
Q11DV3 MsAMD LKSLGITSVADVAGTPAKRLLQLCKEAVHDAGPVDVAVLISCGGLHTLDVIVDVEISTGLPVVTS--ATAGVRGAVGLLKRSAETAHAHATSKFLTGRA--
AsAMD CRSLGITGVAEAMARVDTDTLVDLVCRAFAEAPSDGILLSCGGLLTLDAIPEVERRLGMVVS--SPAGFWDVAVRLAGGGGK-ARPGYGRFLFDES
Q05115 BbAMD X CRSLGITGVAEAMARVDTDTLVDLVCRAFAEAPSDGILLSCGGLLTLDAIPEVERRLGMVVS--SPAGFWDVAVRLAGGGGK-ARPGYGRFLFDES
EcAMD CRSLGITGVADAMARVDTDTLVDLVCRAFAEAPSDGILLSCGGLLTLDAIPEVERRLGMVVS--SPAGFWDVAVRLAGGCEVQ-AAPGYGRLSGVH-----

301 347

Q8Y7N7 LmGluR X STQIFKDIKDWLNMPDMTVEHIKLGK-----
Q6GHT5 SaGluR X DTTHITNIIKEWMLNL-SVMVERISVND-----
P52974 ShGluR X DTVHIKNIILQWLKL-DVEVERISVND-----
Q81LA8 BaGluR X KIGLFKEIASKWFGQP IENVKHIHLEKE-----
Q3XZW8 EFgluR X SAKLFEEIAERDLGIGHLNVEHIELGGK-----
Q81UL8 BaGluR X SVSSFEHIAERWLGY-QISVDCVDLPVKNARICN-----
P94556 BaGluR X ARDQFAKIADDFWGHVGVHCISLQEPKIKR-----
Q7M0V6 BpGluR X QQQNFQNIARDWFGYLPKGVETVLSLEHIYQQ-----
Q836J0 EFgluR X SAKMFEETASSWLG IENLKAQOIHILGNEND-----
Q88V19 LpGluR X AADQFEAIARNWLGQDFHAHQHIDLGSEAND-----
P63640 SpGluR X SSSQFAQIGEEWLEK-EIHVEHVEL-----
Q9A1B7 SpGluR X NPFIQEIASIWLLKQ-KINVEHVTL-----
Q03469 LfGluR X NLPDLRAGVKNWLLSGHFDLGTAIIEEGD-----
P48797 LbGluR X ETDQFDTLASQWLDQOPTPAKHVAIAQLTTPMEVN-----
Q08783 PpGluR X NIKNFEEIARTFLNQDLR-VEEVKID-----
P56868 ApGluR X LILGRDYPVKLAEGVPTH-----
Q9ZLT0 HpGluR X DVIWLERQAKEWLKL-----
P22634 EcGluR X GAEQLLPLVLRQYGFETLEKLAVLG-----
059384 PhAspR X -----
058403 PhAspR X -----
P29079 EFAspR X -----
B3W8K5 LcAspR X -----
Q1G8B6 LdAspR X -----
058781 PhHydR X -----
Q00924 PsHydR X ALENFGRNQTTTK-----
Q92ML1 RmHydR X HKVCA-----
05YXQ1 NfMI -----
Q9KW10 SmMI -----
024766 AfMI -----
Q9WX57 BsMI -----
Q987A4 Mle -----
Q92WC5 Sme RP-----
Q974K3 Ste X -----
Q8U183 PFe -----
Q0S7X3 RseC -----
Q93J74 SceC TGAVGPAVVGPPGVDPAVAGVPVSPAADVMPGVGDVDPDTEQQQEGWT
Q0S7X2 RseD SD-----
Q93J75 SceD A-----
Q82CV3 Sae VH-----
Q11DV3 MsAMD -----
AsAMD -----
Q05115 BbAMD X -----
EcAMD -----

Appendix D

Isolation of Alkylmalonates Decarboxylating Bacteria

D.1 Introduction

The mining of known microbial gene sequences from the Asp/Glu racemase superfamily has proven less effective than assumed as no enzyme active against arylmalonate could be identified (see Chapter 3). In addition, altering the substrate specificity to alkylmalonates through rational engineering of BbAMD proved to be limited. One reason for the limitation lies in the substrate itself. Additional electron withdrawing capability is needed on the malonates, provided by the aryl or alkenyl substituent, allowing the stabilisation of the decarboxylated doubly negatively charged intermediate. Another reason is the fold of the enzyme. The size of the small substrate binding pocket of BbAMD could not be increased to accept substituents larger than methyl. Site saturation mutagenesis of residues in this pocket only yielded inactive enzyme and therefore proved to be more complex than previously assumed (personal communication by Dr. Krzysztof Okrasa) (48,67).

To overcome this apparently intrinsic limitation of AMDs given by their fold and mechanism the search of target enzymes has to be broadened. As no other aryl- or alkylmalonate decarboxylating enzymes are described in literature, mining of the microbial diversity to find an organism that decarboxylates alkylmalonates provides a promising alternative approach. This approach opens the scope for new enzyme families. Only a very small part of the enormous microbial biodiversity has been studied (generally estimated to around 1%), which offers a large pool of potential biocatalysts. Most of the vast number of natural compounds can be degraded by microorganisms.

BbAMD was found in *Bordetella bronchiseptica* isolated through enrichment from soil samples for arylmalonate decarboxylating bacteria by providing arylmalonates as the sole carbon source (76). A different system using an acyl carrier protein to produce acetate by decarboxylating malonate was described *Klebsiella pneumoniae* and *Malonomonas rubra* (30,31,152). Given this diversity of organisms and mechanisms it

is reasonable to assume that there exists a microorganism in nature that contains an enzyme capable of transforming alkylmalonates.

Screening directly for alkylmalonate decarboxylation instead of converting an arylmalonate decarboxylating enzyme also shortcuts the complex process of engineering of the active site to change the enzyme's substrate specificity. It offers the possibility of finding a completely unanticipated mechanism and is not restricted to the architecture of one single active site.

Nonetheless, this approach also has the disadvantage that the number of organisms to screen is practically limited. In order to increase the chances, samples would need to be taken at sites where the substrate is present in high concentrations. Unfortunately, alkylmalonates are not natural compounds and no site polluted by these chemicals is known. The only remaining option is therefore to screen organisms from soil samples taken from random sites.

BASF is interested in the decarboxylation of MPrM and EBM, having slightly longer chains than MEM tested on microorganisms in their own laboratories. In this chapter the isolation of several strains capable of growing on alkylmalonates as their sole carbon source will be presented together with results from preliminary analysis of activity by whole cells.

D.2 Methods

D.2.1 Collection of Soil Samples

Soil samples were collected from various locations around the campus of the University of York (UK) and from a Ministry of Defence demolition site in Longtown (Cumbria, UK). The samples were collected under dense vegetation cover. All soil samples were stored at 4 °C.

D.2.2 Preparation of Minimal Medium

In 971 mL water, 0.1 g NaCl, 1 g (NH₄)₂SO₄, 0.2 g MgSO₄•7H₂O, 0.02 g CaCl₂•2H₂O, 15 g agarose and optionally 0.05 g yeast extract were dissolved and autoclaved. The medium was buffered by adding 20 mL of filter sterilised 10% (w/v) KH₂PO₄ pH 7. The medium was supplemented with 4 mL of filter sterilised 250x trace elements solution produced by dissolving 10 g MgSO₄, 2 g CaCO₃, 5.5 g FeSO₄•7H₂O, 1.44 g

ZnSO₄•7H₂O, 1.11 g MnSO₄•4H₂O, 0.25 CuSO₄•5H₂O, 0.28 g CoSO₄•7H₂O, 0.062 g H₃BO₃ in 50 mL concentrated HCl and water to a total of 1000 mL. The medium was further supplemented with 2 mL of 1000x vitamins solution produced by dissolving 5 mL of 100 mg mL⁻¹ nicotinic acid in 70% (v/v) ethanol, 50 mg myo-inositol calcium pantothenate, 50 mg riboflavin (B2), 50 mg ascorbic acid (C), 50 mg folic acid, 50 mg D-biotin in a total of 50 mL water. An optional 4 mL of filter sterilised 50% (w/v) glucose solution was added together with 20 mL of filter sterilised 25% (w/v) solution of MPrM or EBM adjusted to pH 7.

D.2.3 Liquid Culture Enrichment

Selective enrichments in liquid cultures were performed as described previously (76). A soil sample of 0.5 g was added to 10 mL of minimal medium containing 5 g L⁻¹ of MPrM or EBM, supplemented with yeast extract and glucose. The cultures were incubated in conical flasks at 30 °C shaking for 4 d. Then the cultures were filtered with Whatman 0 filter paper (GE Healthcare) and 200 µL filtrate was subcultured in 10 mL fresh medium supplemented with yeast extract but not glucose. The cultures were incubated for another 4 d before 100 µL samples from both sets of culture were serially diluted from 10⁻¹ to 10⁻⁵ and grown on minimal medium agarose plates containing 5 g L⁻¹ MPrM and EBM with and without the supplementation of glucose and yeast extract. Colonies were visible after incubation for 4 d at 30 °C.

D.2.4 In-Soil Enrichment

Selective enrichments in wet soil were performed according to the method used by BASF for MEM (personal communication by Dr. Nina Baudendistel). Samples of 5 g soil were soaked with minimal medium containing 5 g L⁻¹ of MPrM or EBM, supplemented with yeast extract and glucose. The soil samples were kept wet in open 50 mL centrifuge tubes at 30 °C for 11 d (a total of approximately 5 mL medium was added per sample) before another 5 mL minimal medium was applied, the tubes closed and incubated shaking for another 4 d. Then the cultures were filtered with Whatman 0 filter paper (GE Healthcare) and 100 µL of each culture were serially diluted from 10⁻¹ to 10⁻⁵ and grown on minimal medium agarose plates containing 5 g L⁻¹ MPrM and EBM, without yeast extract, with and without the supplementation of glucose. Colonies were visible after incubation for 4 d at 30 °C.

D.2.5 Purification of Agar Plates and Liquid Culture

Single colonies obtained from enrichment cultures were purified spreading on fresh minimal medium agar plates containing 5 g L⁻¹ MPrM or EBM containing no yeast extract or glucose. This was repeated 2 - 3 times. To exclude the cultures growing on agarose, 12 large colonies for each substrate were picked and 5 mL minimal medium containing 5 g L⁻¹ substrate containing no glucose or yeast extract were inoculated and incubated at 30 °C shaking for 4 d before serial dilutions from 10⁻¹ to 10⁻⁵ were grown on minimal medium agar plates. From the same cultures, 250 µL were taken to test the disappearance of the substrate by reverse phase HPLC. Another 800 µL of growing cultures were mixed with 400 µL of 60% glycerol, snap frozen in liquid nitrogen to store at -80 °C.

D.2.6 HPLC Analysis

Samples originating from bacterial isolates cultures containing alkylmalonates (5 g L⁻¹) as their sole carbon source were centrifuged at 18,000 × g for 5 min and 50 µL were directly injected into Onyx Monolithic C18 100 × 4.8 mm reverse phase HPLC column (Phenomenex). The compounds were run at 45 °C at 2 mL min⁻¹ flow rate with a programme using a 20 mM KH₂PO₄ pH 2.5 buffer solution (A) and methanol (B). The programme sequence was 0.0-5.5 min 50% B, 6.0-7.0 min 70% B and 7.5-10.0 min 50% B. The 2695 Separations Module and the 2996 Photodiode Array Detector (Waters) were used and compounds were detected at 205 nm. Expected retention times were MPrM 1.1 min, MP 1.6 min, EBM 1.3 min, EH 3.9 min.

D.2.7 16S Sequencing

Taxonomic characterisation was performed by sequencing of ribosomal 16S subunit RNA as described earlier (158). A large colony was suspended in 50 µL pure water. The cells were lysed by heating the suspension to 95 °C for 300 s then five cycles of heating to 95 °C for 30 s and cooling to 15 °C for 30 s and finally freezing the cells at -20 °C. The 16S sequences were amplified using 0.625 U GoTaq DNA Polymerase and 1x PCR buffer (Promega), 1.5 mM MgCl₂, 0.25 mM of each dNTP, 0.4 µM of each primer fD1 and rP2 (Table 2.1, Sigma Genosys), 1 µL cell extract and completed to 25 µL total volume with water. The temperature cycles were one initial melting step at 95 °C for 300 s then 30 times melting at 95 °C for 30 s, annealing at 55 °C for 30 s and

elongation at 72 °C for 90 s followed by one final elongation at 72 °C for 300 s performed on a Px2 Thermal Cycler (Thermo Scientific).

D.3 Results

D.3.1 Enrichment of Soil Cultures

There is no known natural source of alkylmalonates. Equally, there are no contaminated sites of alkylmalonates as these compounds were not produced in large quantities by industry. Therefore no known site exists where microorganisms have been exposed to alkylmalonates and have met such selective pressure to evolve enzyme activities that are able to specifically degrade the compounds. The University campus was therefore chosen to take soil samples for its easy accessibility.

Samples were collected from different types of soil under dense vegetation cover in order to maximise the diversity of microorganisms. As an additional source, a microorganism rich soil sample from an explosives contaminated soil on a military training ground previously collected by our laboratory was chosen.

Two methods were used to enrich the soil cultures in alkylmalonate degrading microorganisms. Firstly a soil suspension culture was produced as described previously (76), which involved adding 0.5 g of soil to 10 mL of minimal mineral medium containing no other carbon source than either of MPrM or EBM (Fig. D.1). The cultures were incubated and subcultured for twice 4 d.

Secondly, a different wet soil culture method was applied using a protocol provided by BASF (personal communication with Dr. Nina Baudendistel). In this method 5 g of soil were immersed with the same minimal medium containing either of the two substrates and incubated for 11 d keeping the soil wet all the time. In the end the soil samples were washed with the medium and the liquid collected.

Culture samples from both methods were spread on agarose plates containing the same minimal medium and substrates. Growing colonies were picked and spread on fresh plates to obtain pure cultures. In this process two cultures were growing to thick cultures in the soil suspension containing MPrM. On plates, colonies from these (FIs01 and FIs02) were growing fast if the medium contained substrate but not if the medium contained no substrate (Fig. D.1)



Figure D.1. Selection of isolates on media with MPrM as sole carbon source. Erlenmeyer flasks containing medium with cultures growing (right, Flso1) and not growing (left). Petri dishes with two isolates growing on MPrM (top) and not growing without carbon source (bottom).

Many colonies showed slow growth on the plates and were also observed to grow non-specifically on plates without substrate. To exclude the possibility of them growing on agarose or impurities of the latter, an additional isolate purification step was introduced. This was again performed in liquid medium not containing any agarose. After 4 to 17 d the cultures were checked for signs of microbial growth (turbidity or aggregated cells in flakes). The growing cultures were again spread on selective media.

In total nine isolates were selected in this manner (Fig. D.2). Surprisingly Flso2, selected earlier, did not show repeated growth in the purification procedure. Only isolates growing on MPrM showed significant growth (Flso1, 2-6) whereas colonies growing on EBM (Flso6-9) showed only residual growth that was not significantly more pronounced than on the control plates not containing substrate.

Flso1 and 3 were large reddish colonies whereas Flso4 and 5 were of the same size and aspect but with a more yellowish tint. Flso2 from the previous isolation step showed similar growth to Flso4 and 5. In contrast the colonies of Flso6 were smaller and white-transparent having a very distinct aspect (Fig. D.2).

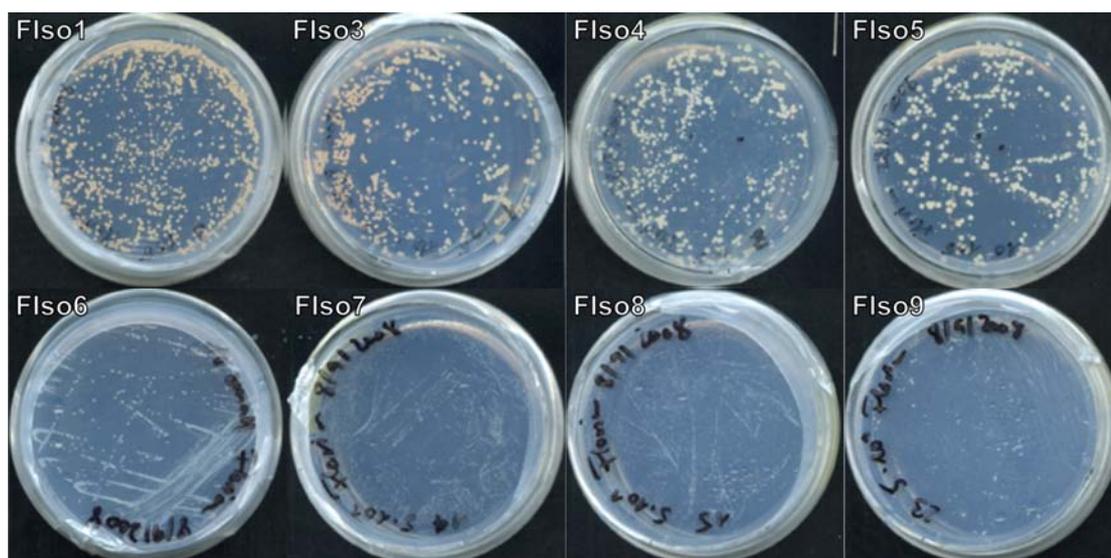


Figure D.2. Isolates on media containing alkylmalonates as sole carbon source. Flso1-6 on MPrM, Flso7-9 on EBM. Flso2 missing.

Colonies of Flso1, 3-6, 9 were analysed by 16S ribosomal subunit sequencing to determine their relationship to known bacterial species. As this was a preliminary experiment, the PCR conditions were not optimised, the sequences were only analysed with one forward primer. As a result PCR products could only be obtained for Flso4-6. The quality of chromatograms was good allowing to read approximately 700 bp for Flso4 and 5 and 500 bp for Flso6. No contamination could be observed in the sequencing chromatograms. A BLAST search using the obtained sequences suggested that Flso4 and 6 related to those from *Acinetobacter* species and in the case Flso5 to those from *Rhodococcus* species.

D.3.2 Screening for Activity of the Isolates

To test whether the bacterial growth on the substrates is the result of the conversion of the alkylmalonates (MPrM and EBM) into alkylalkanoates (MP and EH) the isolates were regrown in liquid cultures with minimal medium with one of each substrate not containing yeast or glucose. After 5 days of growth at 30 °C the media were analysed on HPLC detecting carboxyl groups at 205 nm for removal of substrate and appearance of product.

During the incubation time Flso1 and 3 were using up almost all MPrM, producing a small amount of MP and another compound with a shorter elution time than MP, compound X. In addition there was a small amount of another compound with an even

shorter elution time, compound Y. Flso4 and 5 showed a similar behaviour using up less MPrM and converting in much smaller amounts of the compound X and only very small traces of MP if at all. Flso2 and 6 do not show any degradation of MPrM when compared to the no bacteria control (Fig. D.3). These results confirm previously recorded results with the only exception that at that time Flso6 was a contaminant of Flso5 that was only separated later (data not shown).

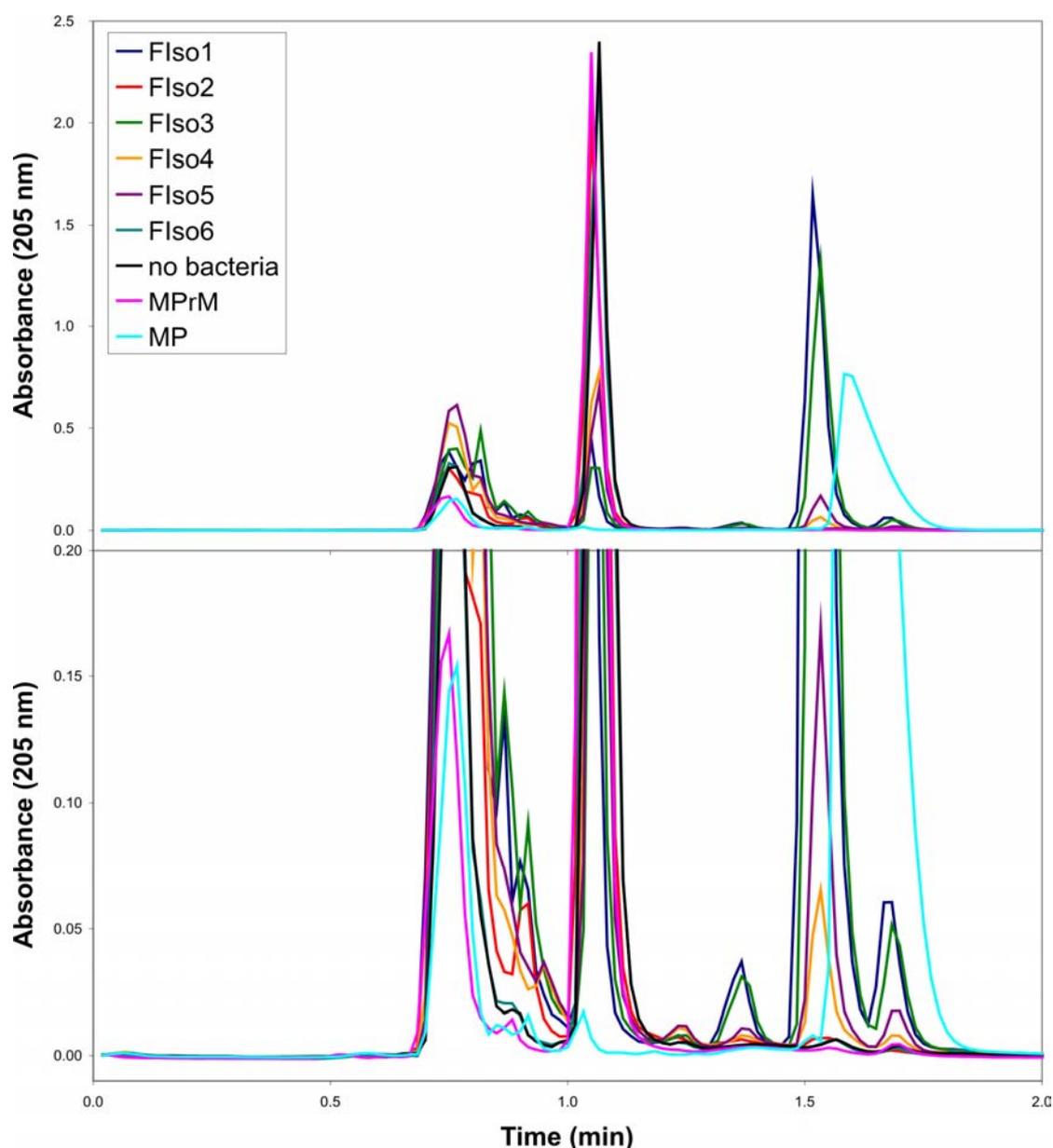


Figure D.3. HPLC analysis of isolates cultures growing on MPrM for 4 d. Full view on top, enlargement on bottom. 50 μL of supernatant from cultures grown on 5 g L^{-1} substrate were injected. RTs are 1.1 min for MPrM, 1.6 min for MP.

In the case of EBM none of the isolates (Flso7-9) degraded any significant amount of the substrate. The peak corresponding to the product EH observed for Flso7 and 8 exactly matches the size of the substrate contamination in the no bacteria control. There are few insignificant peaks not present in the standards but present in the no bacteria control that most probably correspond to chemicals in the medium. Flso9, however, is the only sample where the EH peak completely disappeared (Fig. D.4). These results confirmed previous results with the exception that Flso8 behaved like Flso9 in that experiment (data not shown).

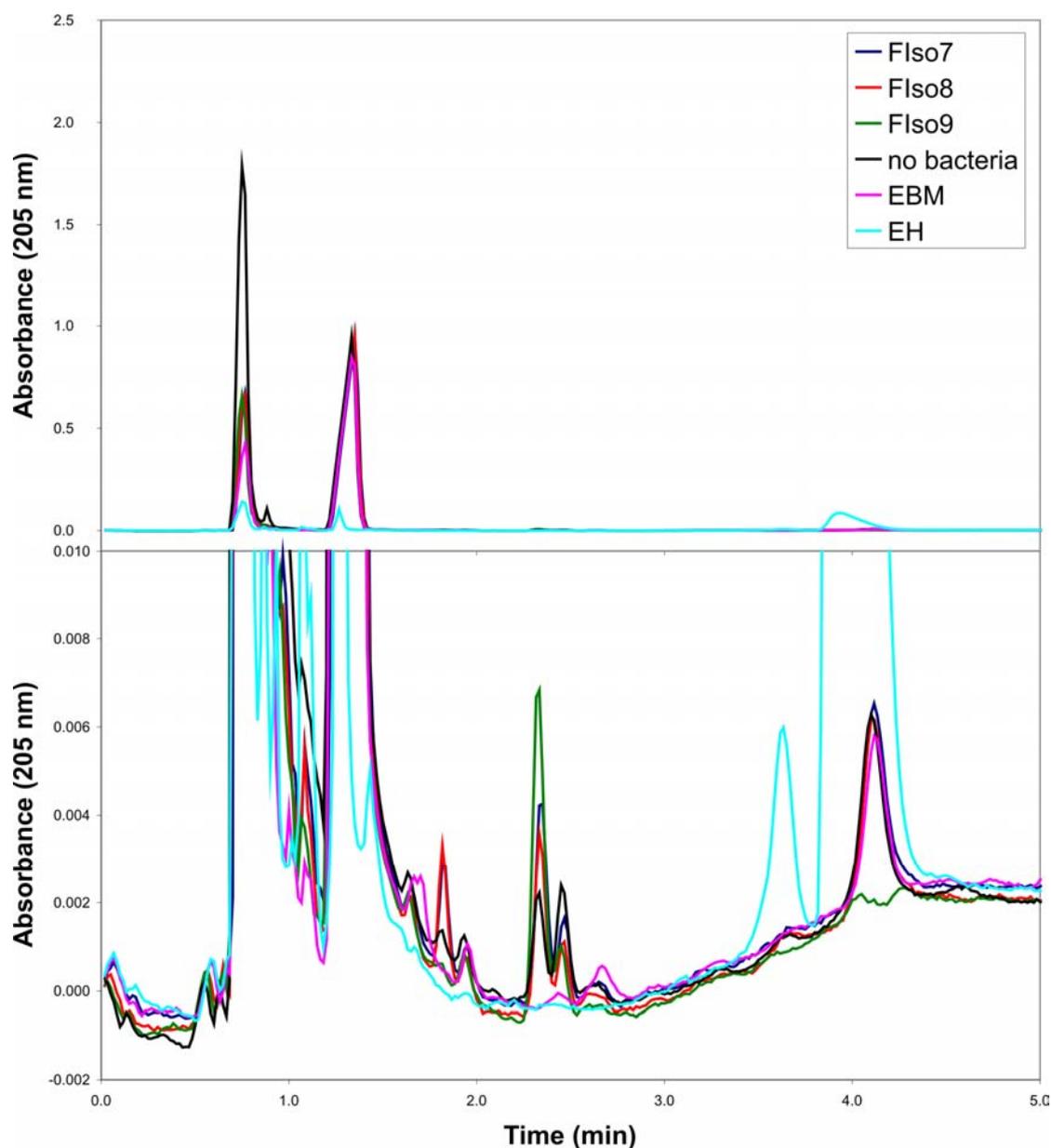


Figure D.4. HPLC analysis of isolates cultures growing on EBM for 4 d. Full view on top, enlargement on bottom. 50 μL of supernatant from cultures grown on 5 g L^{-1} substrate were injected. RTs are 1.3 min for EBM, 3.9 min for EH.

D.4 Discussion

It was previously reported that many bacteria are able to degrade malonate using either of two different acyl carrier protein systems to decarboxylate malonate and produce acetate and carbon dioxide (30,31,152). The enzyme systems are located in the cytoplasm and need transport of the substrate into the cytoplasm. It is likely that these systems are also capable of degrading substituted alkylmalonates.

The present results show that there are bacteria capable of growing on MPrM and degrading it whereas no degraders for EBM were found. Taking the physical aspect of the colonies together with the activity results, one could reasonably conclude that there three strains are able to degrade MPrM. Flso1 and 3 can be considered as the same reddish strain being the most active on MPrM, converting most of it into compound X. Flso4 and 5 can be considered as another yellowish strain being less active on MPrM and consequently producing less compound X. Flso9 seems to be able to degrade EH. All other strains are not active on their respective substrates. For simplification, only Flso1, Flso4 and Flso9 will be discussed further.

MPrM is thus degraded by at least two different bacterial strains; however, it is not clear through what chemical pathway this happens and whether the first step of this pathway really is decarboxylation. Yet, the presence of a small amount of MP in some cultures indicates that decarboxylation is a probable first step in the degradation. Other mechanisms could equally be involved such as a reduction of the carboxylic acid to an aldehyde and subsequent decarbonylation (159,160).

Due to time constraints, the identity of compound X and Y could not be determined. In addition, the optical detection at 205 nm entails the danger of overestimating the amount of compounds as the optical sensitivity depends on the types of bonds contained in the molecules. The introduction of a double bond for example, would greatly increase the response of the detector. To avoid this risk of overestimation, an absorbance independent approach such as GC-FID could be an easy way to circumvent that problem as it results in peak heights proportional to the number of carbon atoms in the compound. Currently, it can only be said that compound X and Y are both more hydrophobic than MPrM but less hydrophobic than MP.

Experiments conducted at BASF by the group of Dr. Nina Baudendistel using the shorter chain length MEM showed very similar results to the experiments with MPrM. MEM was used and converted resulting in a peak appearing just before MB. Efforts to analyse this peak at BASF did unfortunately not result in the identification of the compound corresponding to that peak. This could be an indication of insufficient amount of a double bonded compound giving a disproportionately large signal.

When the cells were lysed at BASF all activity was lost. There could be several explanations for this: the system could be membrane associated, cofactor-dependent or sensitive to the conditions such as the redox potential or pH of the buffer. A multi-subunit acyl carrier protein system as described above would be expected to lose the

activity after cell lysis. It is therefore reasonable to assume that the isolates used in this work and at BASF are indeed acyl carrier protein systems. Still, the lysis has not yet been performed with the isolates in the present work.

Even though Flso9 is probably able to degrade EH no EBM was degraded. The chain length of the malonates therefore is important. This could simply be a membrane transport issue or a limitation by the size of the active site. On the other hand the degraded MEM and MPrM both have a relatively small methyl group as a second substituent on carbon C2 of malonate and EH has an even smaller hydrogen atom at the same position. EBM has an ethyl group which could be specifically excluded by the enzyme. The importance of the size difference between methyl and ethyl groups is evident with BbAMD that accepts MPM but not 2-ethyl-2-phenylmalonate as substrate (67,76).

For any further conclusion the pathway and the enzymes would have to be identified. To analyse the pathway, cell extract could be analysed and the relevant metabolites determined by GC-MS. Feeding the decarboxylated intermediates to the cell cultures would indicate whether these are indeed the intermediates in the degradation pathway. The presence of the known acyl carrier protein degradation system could be confirmed by PCR using degenerate primers.

If the described multi subunit enzyme system was identified as the responsible catalytic system it would make it more complex to use for industrial applications. A possible solution to that problem could be to knock out the downstream processing enzymes in the selected strains and use whole cell transformations instead of engineered enzyme in a cell free system. As the strains can feed on the substrates, the process would lend itself to directed evolution approaches for the engineering of the whole metabolism.

Finally, 16S sequencing showed that there are at least two different strains (*Acinetobacter* sp. and *Rhodococcus* sp.) from two different phyla (Proteobacteria and Actinobacteria). There is supposedly a large diversity of bacteria capable of degrading alkylmalonates and the trait seems to be rather common as they were found in soils not known for containing any special bacteria. More bacteria can therefore be expected to be found using this procedure.

References

1. Faber, K. (2004) *Biotransformations in Organic Chemistry*, 5 ed., Springer, Berlin
2. Grogan, G. (2009) *Practical Biotransformations. A Beginner's Guide*, Wiley, Chichester
3. Erb, R., and Große Ophoff, M. (2007) Cofaktor-unabhängige Decarboxylasen. 13133-79 Ed., Deutsche Bundesstiftung Umwelt (DBU)
4. Breuer, M., Ditrich, K., Habicher, T., Hauer, B., Kessler, M., Sturmer, R., and Zelinski, T. (2004) *Angew. Chem. Int. Ed.* **43**, 788-824
5. Huerta, F. F., Minidis, A. B. E., and Backvall, J.-E. (2001) *Chem. Soc. Rev.* **30**, 321-331
6. Turner, N. J. (2004) *Curr. Opin. Chem. Biol.* **8**, 114-
7. Garcia-Urdiales, E., Alfonso, I., and Gotor, V. (2004) *Chem. Rev.* **105**, 313
8. Fotheringham, I., and Royer, S. (2003) Methods for identifying racemases. Thermogen, Inc., United States
9. Yoshioka, H. (1980) *J. Synth. Org. Chem. Jpn.* **38**, 1151-1162
10. Rieu, J. P., Boucherle, A., Cousse, H., and Mouzin, G. (1986) *Tetrahedron* **42**, 4095-4131
11. Kusumoto, T., Ueda, T., Hiyama, T., Takehara, S., Shoji, T., Osawa, M., Kuriyama, T., Nakamura, K., and Fujisawa, T. (1990) *Chem. Lett.*, 523-526
12. Hauer, B. D., Baudendistel, N. D., Okrasa, K. G., Micklefield, J. G., Leys, D. G., and Levy, C. G. (2009) New Malonate Decarboxylases for Industrial Applications.
13. Ayguen, H., Wojczewski, S., Kircher, M., and Rosmus, S. (2005) Enzymatic asymmetric decarboxylation of disubstituted malonic acids. in *European Patent Office*, Biospring GmbH
14. Stryer, L. (1997) *Biochemistry*, 4th ed., W. H. Freeman & Co., New York
15. Hart, H. (1987) *Organic Chemistry, A Short Course*, 7th ed., Houghton Mifflin Company, Boston
16. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman & Co., New York
17. Grishin, N. V., Osterman, A. L., Brooks, H. B., Phillips, M. A., and Goldsmith, E. J. (1999) *Biochemistry* **38**, 15174-15184
18. Kidron, H., Repo, S., Johnson, M. S., and Salminen, T. A. (2007) *Mol. Biol. Evol.* **24**, 79-89

19. Jhee, K. H., Yoshimura, T., Kurokawa, Y., Esaki, N., and Soda, K. (1999) *J. Microbiol. Biotechnol.* **9**, 695-703
20. Kern, A. D., Oliveira, M. A., Coffino, P., and Hackert, M. L. (1999) *Structure* **7**, 567
21. Brooks, H. B., and Phillips, M. A. (1997) *Biochemistry* **36**, 15147-15155
22. Rosenberg, R. M., and O'Leary, M. H. (1985) *Biochemistry* **24**, 1598
23. Lima, S., Sundararaju, B., Huang, C., Khristoforov, R., Momany, C., and Phillips, R. S. (2009) *J. Mol. Biol.* **388**, 98
24. Lee, B., and Suh, S. W. (2004) *J. Mol. Biol.* **340**, 1-7
25. Konst, P. M., Franssen, M. C. R., Scott, E. L., and Sanders, J. P. M. (2009) *Green Chem.* **11**, 1646-1652
26. Schmitzberger, F., Kilkenny, M. L., Lobley, C. M. C., Webb, M. E., Vinkovic, M., Matak-Vinkovic, D., Witty, M., Chirgadze, D. Y., Smith, A. G., Abell, C., and Blundell, T. L. (2003) *EMBO J.* **22**, 6193
27. Jordan, F. (2003) *Nat. Prod. Rep.* **20**, 184-201
28. Dyda, F., Furey, W., Swaminathan, S., Sax, M., Farrenkopf, B., and Jordan, F. (2002) *Biochemistry* **32**, 6165
29. Castro, E. A. (1999) *Chem. Rev.* **99**, 3505-3524
30. Hoenke, S., Wild, M. R., and Dimroth, P. (2000) *Biochemistry* **39**, 13223-13232
31. Berg, M., and Dimroth, P. (1998) *Arch. Microbiol.* **170**, 464-468
32. Takamura, Y., and Kitayama, Y. (1981) *Biochem. Int.* **3**, 483-491
33. Kim, Y. S., and Byun, H. S. (1994) *J. Biol. Chem.* **269**, 29636-29641
34. Micklefield, J., Harris, K. J., Groger, S., Mocek, U., Hilbi, H., Dimroth, P., and Floss, H. G. (1995) *Journal Of The American Chemical Society* **117**, 1153-1154
35. Handa, S., Koo, J., Kim, Y., and Floss, H. (1999) *Arch. Biochem. Biophys.* **370**, 93-96
36. Ho, M.-C., Menetret, J.-F., Tsuruta, H., and Allen, K. N. (2009) *Nature* **459**, 393
37. Kokesh, F., and Westheimer, F. (2002) *J. Am. Chem. Soc.* **93**, 7270
38. Frey, P., Kokesh, F., and Westheimer, F. (2002) *J. Am. Chem. Soc.* **93**, 7266
39. Hashidoko, Y., and Tahara, S. (1998) *Arch. Biochem. Biophys.* **359**, 225
40. Godoy, L., Martínez, C., Carrasco, N., and Ganga, M. A. (2008) *Int. J. Food Microbiol.*
41. JCSG. (2007) Crystal structure of phenolic acid decarboxylase (2635953) from *Bacillus subtilis* at 1.36 Å resolution (PDB: 2P8G).

42. Dhar, A., Lee, K.-S., Dhar, K., and Rosazza, J. P. N. (2007) *Enzyme Microb. Technol.* **41**, 271
43. Liu, J., Zhang, X., Zhou, S., Tao, P., and Liu, J. (2007) *Curr. Microbiol.* **54**, 102
44. Lack, A., Tommasi, I., Aresta, M., and Fuchs, G. (1991) *Eur. J. Biochem.* **197**, 473-479
45. Lupa, B., Lyon, D., Shaw, L. N., Sieprawska-Lupa, M., and Wiegel, J. (2008) *Can. J. Microbiol.* **54**, 75-81
46. Gerlt, J. A., and Babbitt, P. C. (2001) *Annu. Rev. Biochem.* **70**, 209-246
47. Gallo, K. A., and Knowles, J. R. (1993) *Biochemistry* **32**, 3981-3990
48. Okrasa K, Levy C, Hauer B, Baudendistel N, Leys D, and J, M. (2008) *Chem. Eur. J.* **14**, 6609-6613
49. Watabe, K., Ishikawa, T., Mukohara, Y., and Nakamura, H. (1992) *J. Bacteriol.* **174**, 3461-3466
50. Choi, S.-Y., Esaki, N., Yoshimura, T., and Soda, K. (1992) *J. Biochem. (Tokyo)* **112**, 139-142
51. Miyamoto, K., and Ohta, H. (1992) *Eur. J. Biochem.* **210**, 475-481
52. Hatakeyama, K., Asai, Y., Uchida, Y., Kobayashi, M., Terasawa, M., and Yukawa, H. (1997) *Biochem. Biophys. Res. Commun.* **239**, 74-79
53. Ayengar, P., and Roberts, E. (1952) *J. Biol. Chem.* **197**, 453-460
54. Johnston, M. M., and Diven, W. F. (1969) *J. Biol. Chem.* **244**, 5414-5420
55. Glaser, L. (1960) *J. Biol. Chem.* **235**, 2095-2098
56. Tanaka, M., Kato, Y., and Kinoshita, S. (1961) *Biochem. Biophys. Res. Commun.* **4**, 114
57. Nakajima, N., Tanizawa, K., Tanaka, H., and Soda, K. (1988) *Agric. Biol. Chem.* **52**, 3099-3104
58. Gallo, K. A., Tanner, M. E., and Knowles, J. R. (1993) *Biochemistry* **32**, 3991-3997
59. Tanner, M. E. (2002) *Acc. Chem. Res.* **35**, 237-246
60. Tanner, M. E., Gallo, K. A., and Knowles, J. R. (1993) *Biochemistry* **32**, 3998-4006
61. Glavas, S., and Tanner, M. E. (1999) *Biochemistry* **38**, 4106-4113
62. Hwang, K. Y., Cho, C. S., Kim, S. S., Sung, H. C., Yu, Y. G., and Cho, Y. (1999) *Nat. Struct. Biol.* **6**, 422-426
63. Ruzheinikov, S. N., Taal, M. A., Sedelnikova, S. E., Baker, P. J., and Rice, D. W. (2005) *Structure* **13**, 1707-1713
64. Glavas, S., and Tanner, M. E. (2001) *Biochemistry* **40**, 6199-6204

65. Puig, E., Mixcoha, E., Garcia-Viloca, M., González-Lafont, A., and Lluch, J. M. (2009) *J. Am. Chem. Soc.* **131**, 3509-3521
66. Spies, M., Reese, J., Dodd, D., Pankow, K., Blanke, S., and Baudry, J. (2009) *J. Am. Chem. Soc.* **131**, 5274-5284
67. Okrasa, K., Levy, C., Wilding, M., Goodall, M., Baudendistel, N., Hauer, B., Leys, D., and Micklefield, J. (2009) *Angew. Chem. Int. Ed.* **48**
68. Buschiazzo, A., Goytia, M., Schaeffer, F., Degrave, W., Shepard, W., Grégoire, C., Chamond, N., Cosson, A., Berneman, A., Coatnoan, N., Alzari, P. M., and Minoprio, P. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 1705-1710
69. Lamont, H. C., Staudenbauer, W. L., and Strominger, J. L. (1972) *J. Biol. Chem.* **247**, 5103-5106
70. Yohda, M., Okada, H., and Kumagai, H. (1991) *Biochim. Biophys. Acta* **1089**, 234-240
71. Liu, L., Iwata, K., Kita, A., Kawarabayasi, Y., Yohda, M., and Miki, K. (2002) *J. Mol. Biol.* **319**, 479
72. Ohtaki, A., Nakano, Y., Iizuka, R., Arakawa, T., Yamada, K., Odaka, M., and Yohda, M. (2008) *Proteins* **70**, 1167-1174
73. Wiese, A., Pietzsch, M., Syldatk, C., Mattes, R., and Altenbuchner, J. (2000) *J. Biotechnol.* **80**, 217
74. Andújar-Sánchez, M., Martínez-Rodríguez, S., Heras-Vázquez, F. J. L., Clemente-Jiménez, J. M., Rodríguez-Vico, F., and Jara-Pérez, V. (2006) *Biochim. Biophys. Acta* **1764**, 292
75. Martinez-Rodriguez, S., Andujar-Sanchez, M., Neira, J. L., Clemente-Jimenez, J. M., Jara-Perez, V., Rodriguez-Vico, F., and Las Heras-Vazquez, F. J. (2006) *Protein Sci.* **15**, 2729-2738
76. Miyamoto, K., and Ohta, H. (1990) *J. Am. Chem. Soc.* **112**, 4077-4078
77. Terao, Y., Ijima, Y., Miyamoto, K., and Ohta, H. (2007) *J. Mol. Catal. B: Enzym.* **45**, 15
78. Miyamoto, K., and Ohta, H. (1992) *Appl. Microbiol. Biotechnol.* **38**, 234-238
79. Terao, Y., Miyamoto, K., and Ohta, H. (2006) *Chem. Commun.*, 3600-3602
80. Miyamoto, K., Yatake, Y., Tamura, K., Terao, Y., and Ohta, H. (2007) *J. Biosci. Bioeng.* **104**, 263-267
81. Yatake, Y., Miyamoto, K., and Ohta, H. (2008) *Appl. Microbiol. Biotechnol.* **78**, 793-800
82. Miyamoto, K., Ohta, H., and Osamura, Y. (1994) *Bioorg. Med. Chem.* **2**, 469-475
83. Miyamoto, K., and Ohta, H. (1991) *Biocatalysis* **5**, 49-60

84. Miyamoto, K., Tsuchiya, S., and Ohta, H. (1992) *J. Fluorine Chem.* **59**, 225-232
85. Fukuyama, Y., Matoishi, K., Iwasaki, M., Takizawa, E., Miyazaki, M., Ohta, H., Hanzawa, S., Kakidani, H., and Sugai, T. (1999) *Biosci. Biotechnol., Biochem.* **63**, 1664-1666
86. Terao, Y., Ijima, Y., Kakidani, H., and Ohta, H. (2003) *Bull. Chem. Soc. Jpn.* **76**, 2395-2397
87. Tamura, K., Terao, Y., Miyamoto, K., and Ohta, H. (2008) *Biocatal. Biotransform.* **26**, 253-257
88. Kuettnner, E. B., Keim, A., Kircher, M., Rosmus, S., and Sträter, N. (2008) *J. Mol. Biol.* **377**, 386
89. Nakasako, M., Obata, R., Okubo, R., Nakayama, S., Miyamoto, K., and Ohta, H. (2008) *Acta Cryst.* **F64**, 610-613
90. Nakasako, M., Obata, R., Okubo, R., Miyamoto, K., Ohta, H. (2009) Structural Basis of the Enantioselective Decarboxylation by Arylmalonate Decarboxylase.
91. Matoishi, K., Ueda, M., Miyamoto, K., and Ohta, H. (2004) *J. Mol. Catal. B-Enzym.* **27**, 161-168
92. Miyazaki, M., Kakidani, H., Hanzawa, S., and Ohta, H. (1997) *Bull. Chem. Soc. Jpn.* **70**, 2765-2769
93. Ijima, Y., Matoishi, K., Terao, Y., Doi, N., Yanagawa, H., and Ohta, H. (2005) *Chem. Commun.*, 877-879
94. Terao, Y., Miyamoto, K., and Ohta, H. (2006) *Appl. Microbiol. Biotechnol.* **73**, 647-653
95. Terao, Y., Miyamoto, K., and Ohta, H. (2007) *J. Mol. Catal. B: Enzym.* **48**, 101-102
96. Miyamoto, K., Tsutsumi, T., Terao, Y., and Ohta, H. (2007) *Chem. Lett.* **36**, 656-657
97. Behrman, E. J., and Stanier, R. Y. (1957) *J. Biol. Chem.* **228**, 923-945
98. Kato, Y., Yamagishi, J., and Asano, Y. (1995) *J. Ferment. Bioeng.* **80**, 610
99. Otsuka, K. (1961) *Agric. Biol. Chem.* **25**, 726-&
100. Scher, W., and Jakoby, W. B. (1969) *J. Biol. Chem.* **244**, 1878-1882
101. Takamura, Y., Takamura, T., Soejima, M., and Uemura, T. (1969) *Agric. Biol. Chem.* **33**, 718-&
102. Safronova, I. Y., and Semenova, E. V. (1998) *Microbiology* **67**, 23-27
103. Hatakeyama, K., Goto, M., Kobayashi, M., Terasawa, M., and Yukawa, H. (2000) *Biosci. Biotechnol. Biochem.* **64**, 1477-1485

104. Hatakeyama, K., Goto, M., Uchida, Y., Kobayashi, M., Terasawa, M., and Yukawa, H. (2000) *Biosci. Biotechnol. Biochem.* **64**, 569-576
105. Felthouse, T. R., Burnett, J. C., Horrell, M. J. M., Mummey, M. J., and Kuo, Y.-J. (2001) Maleic Anhydride, Maleic Acid, and Fumaric Acid.
106. Mukouyama, M., Yasuda, S., and Komatsuzaki, S. (2000) *Official Gazette of the United States Patent and Trademark Office Patents* **1239**
107. Waller, A. S. (2001) *Official Gazette of the United States Patent and Trademark Office Patents* **1249**
108. Knox, W. E. (1960) Glutathione. in *The Enzymes*, 2 Ed. pp
109. Topper, Y. J. (1961) Isomerisation Reactions (Survey). in *The Enzymes*, 2 Ed. pp
110. Seltzer, S. (1972) Cis-Trans Isomerisation. in *The Enzymes* (Boyer, P. D. ed.), 3 Ed., Academic Press, New York, London. pp
111. Polekhina, G., Board, P. G., Blackburn, A. C., and Parker, M. W. (2001) *Biochemistry* **40**, 1567
112. Barbara, S., Kurt, F., and Wolfgang, K. (2003) *Adv. Synth. Catal.* **345**, 653-666
113. Farinas, E. T., Bulter, T., and Arnold, F. H. (2001) *Curr. Opin. Biotechnol.* **12**, 545
114. Bonsor, D., Butz, S. F., Solomons, J., Grant, S., Fairlamb, I. J., Fogg, M. J., and Grogan, G. (2006) *Org. Biomol. Chem.* **4**, 1252-1260
115. Lindwall, G., Chau, M. F., Gardner, S. R., and Kohlstaedt, L. A. (2000) *Protein Eng.* **13**, 67-71
116. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G. M., Carnemolla, B., Orecchia, P., Zardi, L., and Righetti, P. G. (2004) *Electrophoresis* **25**, 1327-1333
117. Terao, Y., Miyamoto, K., and Ohta, H. (2007) *Chem. Lett.* **36**, 420-421
118. Miyamoto, K., Tsuchiya, S., and Ohta, H. (1992) *J. Am. Chem. Soc.* **114**, 6256-6257
119. Janes, L. E., Lowendahl, A. C., and Kazlauskas, R. J. (1998) *Chem.-Eur. J.* **4**, 2324-2331
120. Banerjee, A., Kaul, P., Sharma, R., and Banerjee, U. C. (2003) *J. Biomol. Screen.* **8**, 559-565
121. N. Eswar, M. A. M.-R., B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. . (2006) Comparative Protein Structure Modeling With MODELLER. in *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc. pp
122. Jebbar, M., Sohn-Bosser, L., Bremer, E., Bernard, T., and Blanco, C. (2005) *J. Bacteriol.* **187**, 1293-1304

123. Ramamurthy, V., Swann, S. L., Spedaliere, C. J., and Mueller, E. G. (1999) *Biochemistry* **38**, 13106
124. Brzozowski, A. M., and Walton, J. (2001) *J. Appl. Crystallogr.* **34**, 97-101
125. Leslie, A. G. W. (1992) *Joint CCP4 and ESF-EACBM Newsletters on Protein Crystallography* **26**
126. CCP4. (1994) *Acta Cryst.* **D50**, 760-763
127. Kantardjieff, K. A., and Rupp, B. (2003) *Protein Sci.* **12**, 1865-1871
128. A. J. McCoy, R. W. G.-K., P. D. Adams, M. D. Winn, L.C. Storoni and R.J. Read. . (2007) *J. Appl. Cryst.* **40**, 658-674
129. Cowtan, K. D., and Zhang, K. Y. J. (1999) *Prog. Biophys. Mol. Biol.* **72**, 245-270
130. Emsley, P., and Cowtan, K. (2004) *Acta Cryst.* **D60**, 2126-2132
131. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger., T. C. (2002) *Acta Cryst.* **D58**, 1948-1954
132. Simon C. Lovell, I. W. D., W. Bryan Arendall III, Paul I. W. de Bakker, J. Michael Word, Michael G. Prisant, Jane S. Richardson, David C. Richardson. (2003) *Protein Struct. Funct. Genet.* **50**, 437-450
133. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577-2637
134. Krissinel, E., and Henrick, K. (2007) *J. Mol. Biol.* **372**, 774-797
135. Hsin, K., Sheng, Y., Harding, M. M., Taylor, P., and Walkinshaw, M. D. (2008) *J. Appl. Crystallogr.* **41**, 963-968
136. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A., and Noble, M. (2004) *Acta Cryst.* **D60**, 2288-2294
137. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) *Nucleic Acids Res.* **34**, W116-W118
138. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) *Acta Cryst.* **D60**, 2184-2195
139. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) *J. Comput. Chem.* **30**, 2785-2791
140. Brunger, A. T., and Adams, P. D. (2002) *Acc. Chem. Res.* **35**, 404
141. May, M., Mehboob, S., Mulhearn, D. C., Wang, Z., Yu, H., Thatcher, G. R., Santarsiero, B. D., Johnson, M. E., and Mesecar, A. D. (2007) *J. Mol. Biol.* **371**, 1219-1237
142. Kim, K. H., Bong, Y. J., Park, J. K., Shin, K. J., Hwang, K. Y., and Kim, E. E. (2007) *J. Mol. Biol.* **372**, 434-443

143. Liu, L., Iwata, K., Yohda, M., and Miki, K. (2002) *FEBS Lett.* **528**, 114-118
144. Lundqvist, T., Fisher, S. L., Kern, G., Folmer, R. H. A., Xue, Y., Newton, D. T., Keating, T. A., Alm, R. A., and de Jonge, B. L. M. (2007) *Nature* **447**, 817
145. Geng, B., Basarab, G., Comita-Prevoir, J., Gowravaram, M., Hill, P., Kiely, A., Loch, J., MacPherson, L., Morningstar, M., Mullen, G., Osimboni, E., Satz, A., Eyermann, C., and Lundqvist, T. (2009) *Bioorg. Med. Chem. Lett.* **19**, 930-936
146. Kita, A., Tasaki, S., Yohda, M., and Miki, K. (2009) *Proteins* **74**, 240-244
147. Majorek, K. A., Chruszcz, M., Zimmerman, M.D., Klimecka, M.M., Cymborowski, M., Skarina, T., Onopriyenko, O., Stam, J., Otwinowski, Z., Anderson, W.F., Savchenko, A., Minor, W. (2009) Crystal structure of glutamate racemase from *Listeria monocytogenes*. Center for Structural Genomics of Infectious Diseases (CSGID)
148. Majorek, K. A., Chruszcz, M., Skarina, T., Onopriyenko, O., Stam, J., Anderson, W.F., Savchenko, A., Bujnicki, J.M., Minor, W. (2009) Crystal structure of glutamate racemase from *Listeria monocytogenes* in complex with acetate ion. Center for Structural Genomics of Infectious Diseases (CSGID)
149. Mizutani, H., Kunishima, N. (2007) Crystal structure of hydantoin racemase from *Pyrococcus horikoshii* OT3.
150. Tanaka, Y., Sasaki, T., Tanabe, E., Yao, M., Tanaka, I., Kumagai, I., Tsumoto, K. (2007) Crystal structure of ST0656, a function unknown protein from *Sulfolobus tokodaii*.
151. Geng, B., Breault, G., Comita-Prevoir, J., Petrichko, R., Eyermann, C., Lundqvist, T., Doig, P., Gorseth, E., and Noonan, B. (2008) *Bioorg. Med. Chem. Lett.*
152. Dimroth, P., and Hilbi, H. (1997) *Mol. Microbiol.* **25**, 3-10
153. Alderson, N., Wang, Y., Blatnik, M., Frizzell, N., Walla, M., Lyons, T., Alt, N., Carson, J., Nagai, R., Thorpe, S., and Baynes, J. (2006) *Arch. Biochem. Biophys.* **450**, 1-8
154. Drevland, R. M., Waheed, A., and Graham, D. E. (2007) *J. Bacteriol.* **189**, 4391-4400
155. Hill, R. L. T., J.W. (1971) Fumarase and crotonase. in *The Enzymes* (Boyer, P. D. ed.), 3rd Ed. pp
156. Alberty, R. A. (1961) Fumarase. in *The Enzymes* (Boyer, P. D., Lardy, H., Myrbäck, K. ed.), 2nd Ed. pp
157. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) *Nature* **453**, 190

158. Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991) *J. Bacteriol.* **173**, 697-703
159. Venkitasubramanian, P., Daniels, L., and Rosazza, J. P. N. (2007) *J. Biol. Chem.* **282**, 478-485
160. Schneider-Belhaddad, F., and Kolattukudy, P. (2000) *Arch. Biochem. Biophys.* **377**, 341