University of Groningen

Structural and functional characterization of tautomerase and aspartase/fumarase superfamily enzymes

Poddar, Harshwardhan

DOI:

[10.33612/diss.126026140](https://doi.org/10.33612/diss.126026140)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Poddar, H. (2020). *Structural and functional characterization of tautomerase and aspartase/fumarase superfamily enzymes*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.126026140>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

**STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF
TAUTOMERASE AND ASPARTASE/FUMARASE
SUPERFAMILY ENZYMES**

Harshwardhan Poddar

2020

The research described in this thesis was carried out at the Department of Chemical and Pharmaceutical Biology (Groningen Research Institute of Pharmacy, University of Groningen, The Netherlands) and the Department of Biophysical Chemistry (Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands). The research was financially supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no 242293, an ALW grant 820.02.021 from the Division of Earth and Life Sciences of the Netherlands Organisation of Scientific Research, an ECHO grant 700.59.042 from the Division of Chemical Sciences of the Netherlands Organisation of Scientific Research and the European Union 7th framework project Metaexplore (KBBE-2007-3-3-05, grant number 222625).

The research work was carried out in accordance to the requirements of the Graduate School of Science and Engineering, Faculty of Science and Engineering, University of Groningen, The Netherlands.

The University Library and the Graduate School of Science and Engineering, Faculty of Science and Engineering, University of Groningen, The Netherlands have financially supported the printing of this thesis.

ISBN: 978-94-034-2684-6 (Printed)

ISBN: 978-94-034-2685-3 (Electronic)

Layout and Printing: Off Page, Amsterdam

Cover Art: Tetrameric structure of EDDS lyase (Chapter 5)

Copyright © 2020 Harshwardhan Poddar. All rights are reserved. No part of this thesis may be reproduced or transmitted in any form or by any means without the prior permission in writing of the author.

university of
 groningen

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF
TAUTOMERASE AND ASPARTASE/FUMARASE
SUPERFAMILY ENZYMES

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on
Friday 5 June 2020 at 16.15 hours

by

Harshwardhan Poddar

born on 21 April 1986
in Kolkata, India

SUPERVISORS

Prof. Dr. Gerrit J. Poelarends

Prof. Dr. Wim J. Quax

CO-SUPERVISOR

Dr. Andy-Mark W.H. Thunnissen

ASSESSMENT COMMITTEE

Prof. Dr. Dirk J. Slotboom

Prof. Dr. Marco W. Fraaije

Prof. Dr. Willem J.H. van Berkel

*This thesis is dedicated to my parents,
Ashok and Madhu Poddar*

TABLE OF CONTENTS

Aim and outline of this thesis	9
Chapter 1 Introduction	15
Chapter 2 Evidence for the Formation of an Enamine Species During Aldol and Michael-type Addition Reactions Promiscuously Catalyzed by 4-Oxalocrotonate Tautomerase	35
Chapter 3 Using Mutability Landscapes of a Promiscuous Tautomerase to Guide the Engineering of Enantioselective Michaelases	53
Chapter 4 Functional and Structural Characterization of an Unusual Cofactor-Independent Oxygenase	123
Chapter 5 Structural Basis for the Catalytic Mechanism of Ethylenediamine- <i>N</i> , <i>N'</i> -disuccinic Acid Lyase, a Carbon-Nitrogen Bond-Forming Enzyme with a Broad Substrate Scope	169
Chapter 6 Summary and Future Perspectives	203
Nederlandse Samenvatting	211
Acknowledgements	219

AIM AND OUTLINE OF THIS THESIS

Understanding the underlying mechanistic principles of enzymatic reactions is highly desirable. It allows us to better comprehend and identify the key events and players during the catalytic cycle of an enzyme-catalyzed reaction. Using this information it is possible to further improve the catalytic efficiency or substrate scope of an enzyme. One of the most widely used techniques to study the mechanistic details of enzymatic reactions is X-ray crystallography. This powerful technique helps us to visualize structures of biological macromolecules in three-dimensional space. Using X-ray crystallography it is possible to get important insights into the overall fold, oligomeric state and active-site region of an enzyme.

The aim of the work described in this thesis was to provide structural and functional insights into three different enzymes: 4-oxalocrotonate tautomerase (4-OT) and 4-hydroxyphenylenolpyruvate oxygenase (RhCC), which belong to the tautomerase superfamily of enzymes, and ethylenediamine-*N,N'*-disuccinic acid (EDDS) lyase belonging to the aspartase/fumarase superfamily of enzymes. We have extensively used X-ray crystallography to get structural insights into the overall fold and catalytic mechanisms of the above-mentioned enzymes. The results discussed in the following chapters have enhanced our knowledge about the tautomerase and aspartase/fumarase superfamilies and opened new avenues for further research into these three intriguing enzymes.

In **Chapter 1**, we give a brief introduction into both the tautomerase and aspartase/fumarase superfamily of enzymes. The different members of the tautomerase superfamily are involved in catalyzing different reactions, such as tautomerization, dehalogenation and decarboxylation, but they all possess the characteristic β - α - β structural fold. The main catalytic residue, an N-terminal proline (Pro-1), has been found to be instrumental in catalyzing all these reactions. Aspartase/fumarase superfamily members have attracted a lot of interest due to their ability to introduce C-N bonds into the products of their reactions. It has been proposed that members of the aspartase/fumarase superfamily of enzymes employ a common mechanism and utilize a highly flexible loop in the active site for efficient catalysis.

4-OT has been reported to catalyze promiscuous C-C bond-forming aldol and Michael-type addition reactions and these reactions have been proposed to proceed through a nucleophilic enamine intermediate. In **Chapter 2**, we report a high-resolution crystal structure of 4-OT showing a covalent enamine adduct on the Pro-1 residue in the active site. This provides an important confirmation of the first step during the 4-OT catalyzed C-C bond-forming reactions, in which a carbonyl substrate is activated for nucleophilic addition via Pro-1 dependent formation of an enamine intermediate. H-D exchange studies further suggest the importance of the Pro-1 residue for substrate activation through enamine formation.

Chapter 3 highlights some important contributions to the field of laboratory evolution of enzymes. Using the small monomeric size of 4-OT to our advantage, mutability landscapes, reflecting the effect of amino acid substitutions at each position on expression and natural and promiscuous activities of 4-OT, were generated. Remarkably, a single mutant (A33D)

showed improved enantioselectivity and a double mutant (M45Y/F50A) displayed inverted enantioselectivity. We have solved the structure of this double mutant to understand this altered stereoselectivity. The crystal structures with and without substrate highlighted the remodeling of the active site and possibly explain the rationale behind the inverted enantioselectivity displayed by this mutant enzyme.

In **Chapter 4** we describe the structural and functional characterization of a new member of the tautomerase superfamily. This superfamily member, RhCC from *Rhodococcus jostii* RHA1, was found to catalyze a direct reaction between molecular oxygen (O₂) and the substrate 4-hydroxyphenylpyruvate. Interestingly, this oxygenase reaction was catalyzed without the help of any cofactors. This was further confirmed by solving the structure of RhCC, which showed the absence of any bound cofactors to the enzyme, establishing RhCC as a cofactor-independent oxygenase. The identification and characterization of RhCC has added a new facet to the growing repertoire of the tautomerase superfamily and it sets the stage for further investigations into this intriguing reaction.

EDDS lyase, a C-N lyase, has been reported to catalyze the breakdown of the metal chelator ethylenediamine-*N,N'*-disuccinic acid (*S,S*-EDDS) via an unusual sequential two-step reaction mechanism. However, our knowledge about this enzyme remains limited due to the absence of any structural and functional data. To better understand its catalytic mechanism we have characterized EDDS lyase from *Chelativorans* sp. BNC1 (**Chapter 5**). Using a combination of co-crystallization and soaking techniques we were successful in elucidating the structures of EDDS lyase in ligand-free and substrate- and product-bound states. These structures strongly support a general base catalyzed deamination mechanism, which is a characteristic of the aspartase/fumarase superfamily. The structures also provide a good starting point for future enzyme engineering studies. We further explored the substrate scope of this robust biocatalyst and showed its potential to be effectively used in synthesis of difficult aminocarboxylic acids.

The final section (**Chapter 6**) summarizes the work presented in the preceding chapters and provides a perspective for the future directions that can be pursued.

c h a p t e r o n e

INTRODUCTION

SECTION 1

AN INTRODUCTION INTO THE TAUTOMERASE SUPERFAMILY OF ENZYMES

Introduction

Enzymes are nature's catalysts playing a crucial role in chemical transformations. It has been a long held belief that enzymes have evolved to carry out specific reactions on specific substrates. However, in the past few decades, more and more enzymes have been found to display catalytic promiscuity, i.e., the ability of an enzyme to catalyze a chemically distinct biotransformation in addition to its native activity for which it has evolved^{1, 2}. The incidence of catalytic promiscuity is especially pronounced in different enzymes belonging to the same superfamily³. Enzymes belonging to the same superfamily are often characterized as having a similar structural fold and a common catalytic mechanism.

It has been postulated that enzymes belonging to a superfamily have evolved from an ancestral progenitor, which was more of a generalist exhibiting a diverse range of activities, albeit at low levels. In the course of evolution, selection pressure for a particular activity gave rise to specialists but their ability to carry out other activities was not completely lost⁴. Thus, in a superfamily of enzymes it is common to identify a native activity of a family member as a promiscuous activity of other family members. Interestingly, divergent evolution might also lead to the emergence of new activities that were not present in the progenitor generalist, further adding to the catalytic diversity of the superfamily. Therefore, studying different members of a particular superfamily of enzymes enhances our knowledge of how divergent evolution has shaped these individual enzymes⁵. Using tools of directed evolution these enzymes can be further evolved so as to make them more efficient in catalyzing promiscuous reactions.

The tautomerase superfamily is a fine example of a group of enzymes showing catalytic and structural diversity. The small monomeric size and absence of any co-factors or coenzymes together with the inherent stability of these enzymes make them ideal candidates to study the effect of divergent evolution on structure and function within a superfamily of enzymes. This superfamily exhibits remarkable chemical versatility and it is known that the enzymes belonging to this superfamily are proficient in catalyzing varied biotransformations such as tautomerization, isomerization, dehalogenation and decarboxylation, acting on different substrates^{6, 7}. More recently, new promiscuous activities such as Michael-type addition and aldol condensation reactions, dehydration reactions, and co-factor independent oxygenation reactions have also been attributed to some members of this superfamily⁸. All the members of this superfamily share a conserved β - α - β structural fold, yet they show considerable structural diversity and are active as either hexamers, trimers or dimers⁶. A conserved N-terminal proline residue, which is exposed after the initiating methionine is cleaved off during post-translational modification, is the main catalytic residue and is known to act as a catalytic base, nucleophile or an acid, dependent on its protonation state, which in turn is largely influenced by the active

site architecture, i.e., the composition and placement of residues with respect to the N-terminal proline⁷.

The tautomerase superfamily consists of the following known families that have been characterized till date (Table 1); 1) 4-oxalocrotonate tautomerase (4-OT), 2) 5-(carboxymethyl)-2-hydroxymuconate isomerase (CHMI), 3) macrophage migration inhibitory factor (MIF), 4) *cis*-3-chloroacrylic acid dehalogenase (*cis*-CAAD), and 5) malonate semialdehyde decarboxylase (MSAD)⁶. Despite the common features shared by these enzymes there are significant differences and this short review aims to discuss the salient features of the different members of the tautomerase superfamily to highlight their catalytic and structural diversity.

4-Oxalocrotonate tautomerase

4-OT from *Pseudomonas putida* mt-2 is by far the most widely studied member of this superfamily. This enzyme is found in the catabolic pathway for aromatic hydrocarbons

Table 1. Salient features of different tautomerase superfamily members

Enzyme	Monomer length	Oligomeric state	Reaction catalyzed
4-OT	62	homohexamer	2-hydroxymuconate $\rightleftharpoons$ 2-oxohex-3-enedioate
CHMI	125	trimer	5-(carboxymethyl)-2-hydroxymuconate $\rightleftharpoons$ 5-(carboxymethyl)-2-oxohex-3-enedioate
MIF	114	trimer	phenylenolpyruvate $\rightleftharpoons$ phenylpyruvate
<i>cis</i> -CaaD	149	trimer	<i>cis</i> -3-chloroacrylate $\xrightarrow{+H_2O/-HCl}$ malonate semialdehyde
MSAD	129	trimer	malonate semialdehyde $\longrightarrow$ acetaldehyde + CO ₂

where it catalyzes the tautomeric rearrangement of 2-hydroxymuconate to yield 2-oxohex-3-enedioate⁹ (Table 1). It is active as a homo-hexamer where each monomer consists of only 62 residues. The characteristic β - α - β structural motif, which is a signature of this superfamily of enzymes, was first reported in a 4-OT monomer¹⁰. Briefly, the structural fold starts off with an exposed N-terminal proline, which is the first residue of the first β -strand (β_1) (Figure 1A). This is followed by an α -helix and a second β -strand (β_2). Interspersed between these two secondary structural motifs is a 3_{10} -helix turn. The β -strands, β_1 and β_2 , interact in a parallel manner to form a β -sheet, which is flanked by the α -helix on one side. There is also a β -hairpin motif formed by two short β -strands (β_3 and β_4) near the C-terminus. The dimeric association is stabilized by the anti-parallel interactions between the β -sheets and α -helices from the neighboring monomers (Figure 1B). These interactions form a compact four-stranded β -sheet. The coming together of three such dimers, where the hydrophobic β -sheets form the core of the assembly and the α -helices are situated towards the periphery, forms the physiologically active hexamer (Figure 1C).

Figure 1. Cartoon depiction of A) a 4-OT monomer, B) a 4-OT dimer and C) a 4-OT hexamer (PDB id. 4X19)³³; the β -sheets are colored red, the α -helix is shown in cyan and the Pro-1 is shown as balls and sticks. D) A CHMI monomer (same color scheme as 4-OT) and E) a CHMI trimer (PDB id. 1OTG)²² with different monomers shown in blue, pink and green respectively. Trimeric structures of F) MIF (PDB id. 1MIF)²⁴, G) RhCC (PDB id. 4U5P)³⁰ and H) MSAD (PDB id. 2AAG)³⁴.

The C-terminal β -hairpin loops also play a critical role in the assembly, by interacting with the β -sheets from the neighboring dimers, thus stabilizing the hexamer.

The active site pocket with the catalytic Pro-1 is formed at the interface of two adjacent dimers and is composed of residues from three different monomers. In 4-OT, Pro-1 has a pK_a of ~ 6.4 which implies that at cellular pH it exists as a deprotonated species⁹. This allows Pro-1 to function as a general base catalyst allowing it to abstract the 2-hydroxyl proton from the substrate 2-hydroxymuconate and deliver it to the C-5 position, thus effecting the enol-keto tautomerization reaction. Based on a crystal structure obtained with a bound inhibitor, 2-oxo-pent-3-noate, and subsequent mutational analysis, it was proposed that several residues in the active site play a key role during catalysis¹¹. Notably, Arg-11 and Arg-39 (contributed from the adjacent dimer) play a significant role in binding to the C-6 and C-1 carboxylate groups of 2-hydroxymuconate respectively. The hydrophobic residue, Phe-50, also contributes significantly to the overall structural integrity and catalysis.

Another member of the 4-OT family acts as a dehalogenase. The *trans*-3-chloroacrylic acid dehalogenase or CaaD from *Pseudomonas pavonaceae* 170 catalyzes the hydrolytic dehalogenation of *trans*-3-chloroacrylic acid to produce malonate semialdehyde¹² (Table 2). Interestingly, CaaD is a heterohexamer with three α subunits (75 residues per chain) and three β subunits (70 residues per chain). Individual α and β subunits comprise a heterodimer and coming together of three heterodimers results in the formation of a heterohexamer. Different from 4-OT, the pK_a value of the catalytic β Pro-1 in CaaD is ~ 9.3 ; this allows it to function as a catalytic acid at cellular pH¹³. Structural analysis together with mutagenesis data confirmed the importance of β Pro-1, α Arg-8 and α Arg-11 residues for the dehalogenase activity. In 4-OT at the position 8 there is an equivalent Leu residue and as such 4-OT exhibits only low-level dehalogenase activity. Indeed, a Leu-8-Arg mutation in 4-OT results in increased efficiency for the dehalogenase activity¹⁴. This ability to promiscuously catalyze dehalogenation of *trans*-3-chloroacrylic acid is also visible in YwhB, an orthologue of 4-OT from *Bacillus subtilis*. YwhB is also a tautomerase and it acts as a catalyst in conversion of phenylenolpyruvate to phenylpyruvate^{15, 16} (Table 1). It shares roughly 36% identity with 4-OT and shows remarkable similarity in active site composition with all the major residues (Pro-1 and Arg-11) positionally conserved with the exception of Val-39 (equivalent Arg-39 in 4-OT). Notably, both 4-OT and CaaD are also able to catalyze enol-keto tautomerization of phenylenolpyruvate, but to a lesser extent¹⁷ (Table 1).

In addition to the above-mentioned promiscuous activities, 4-OT was reported to catalyze C-C bond-forming reactions (Table 2). At physiological pH it is possible for Pro-1, which exists in a deprotonated state, to function as a nucleophile. Realizing this, Poelarends and co-workers proposed that 4-OT can be used to perform nucleophilic catalysis. They successfully demonstrated that 4-OT with its N-terminal Pro-1 residue was capable of generating an enamine species with a set of small carbonyl-containing compounds (aldehydes and ketones)¹⁸. These intermediate enamine species behave as nucleophiles, which subsequently react with a range of different electrophiles. Utilizing this mechanism it was shown that 4-OT can catalyze the addition of acetaldehyde to benzaldehyde to

give cinnamaldehyde in an aldol condensation reaction¹⁸. Furthermore, the F50A mutant of 4-OT showed a 600-fold increase in catalytic efficiency for this reaction. Interestingly, the proposed intermediate of the aldol coupling reaction, 3-hydroxy-3-phenylpropanal was also shown to be a substrate, which can undergo 4-OT-catalyzed dehydration to form cinnamaldehyde or a retro-aldol reaction to yield acetaldehyde and benzaldehyde¹⁹. Another C-C bond-forming reaction attributed to 4-OT was the asymmetric Michael-type addition of acetaldehyde to *trans*- β -nitrostyrene to make 4-nitro-3-phenylbutanal. This product is a precursor for the synthesis of the antidepressant phenibut and 4-OT produces it with high stereoselectivity (*S* enantiomer, 89% *ee*)²⁰. Mutational analysis revealed that the same set of residues, namely Pro-1, Arg-11 and Arg-39, which have been shown to be important for the natural activity, have a key role to play during promiscuous C-C bond-forming reactions. It is intriguing that using the same active site 4-OT is capable of showcasing such diversity in types of different chemical reactions. The discoveries of these two promiscuous C-C bond-forming reactions are fine examples of important contributions to the field of catalytic promiscuity.

5-(carboxymethyl)-2-hydroxymuconate isomerase

CHMI is an isomerase, using a catalytic mechanism similar to 4-OT, although it shows a preference for a different substrate, 5-(carboxymethyl)-2-hydroxymuconate (Table 1). Compared to 4-OT, which exhibits 62 residues, the CHMI monomer consists of 125 amino acid residues²¹. Wigley and coworkers reported the first crystal structure of CHMI and showed that the biological unit was a trimer instead of a hexamer²². The significant difference in the respective sizes of the monomer and the lack of detectable homology between the sequences of 4-OT and CHMI is not reflected in the overall structure, with the fold of a CHMI monomer closely resembling that of a 4-OT dimer (Figure 1D). Similar to a 4-OT dimer, the CHMI monomer has a core of four β -strands, which forms a compact β -sheet. The two α -helices are positioned at the top of the β -sheet. The whole fold can be sub-divided into two adjacent β - α - β motifs running in opposite directions. Similar to the 4-OT hexameric structure, the CHMI trimer has β -sheets forming the core of the assembly (Figure 1E). Extensive interactions between the secondary structure elements from neighboring subunits provide stability to the overall trimer. Another marked difference between 4-OT and CHMI is the presence of only three active sites in CHMI, although the degree of similarity in the active site pockets is quite remarkable. The positions of Pro-1 and Arg-40 (equivalent Arg-39 in 4-OT) are well conserved in the CHMI active site. It is not surprising that both 4-OT and CHMI catalyze similar reactions employing the same catalytic mechanism as the overall chemistry of the active sites is nearly identical. This similarity, however, is only limited to the active site, as the structures diverge outside of this region. The question whether the striking similarity between these two enzymes of different oligomeric state is a result of convergent or divergent evolution remains unanswered.

Table 2. Promiscuous reactions catalyzed by members of the tautomerase superfamily

Enzyme	Reaction type	Reaction Catalyzed
4-OT	aldol condensation	acetaldehyde benzaldehyde cinnamaldehyde
4-OT F50A mutant	dehydration	3-hydroxy-3-phenylpropanal cinnamaldehyde
4-OT F50A mutant	retro-aldol condensation	3-hydroxy-3-phenylpropanal benzaldehyde acetaldehyde
4-OT	Michael-type addition	acetaldehyde <i>trans</i> -β-nitrostyrene 4-nitro-3-phenylbutanal
CaaD 4-OT MIF	dehalogenation	<i>trans</i> -3-chloroacrylate malonate semialdehyde
RhCC	oxygenation	4-hydroxy-phenyl enolpyruvate 3-hydroxy-(4-hydroxyphenyl)-pyruvate 4-hydroxybenzaldehyde oxalic acid
<i>cis</i> -CaaD MSAD	hydratase	2-oxo-3-pentynoate acetopyruvate

Macrophage migration inhibitory factor

MIF is an important immunoregulatory protein and classified as a cytokine. Studies have shown that MIF also possesses phenylenolpyruvate tautomerase (PPT) activity (Table 1), although this is not believed to interfere with its physiological cytokine activity²³. Crystal structures have shown that MIF is also a trimer with 114 residues per monomer²⁴. The overall trimeric structure of MIF closely resembles that of CHMI with identical interactions between subunits and presence of three active sites characterized by the occurrence of N-terminal Pro-1 residues (Figure 1F). A crystal structure of MIF with bound 4-hydroxyphenylenolpyruvate (4HPP, a substrate) gave important insights into the residues involved in substrate binding and catalysis²⁵. Only Pro-1 was found to be a common residue with CHMI and 4-OT. Lys-32 and Ile-64 were found to make hydrogen bonds with the carboxyl moiety of the substrate. Tyr-95 from the neighboring monomer occupies a similar position as Arg-39 in 4-OT, but appears too far from the substrate HPP, and thus unlikely to play any meaningful role during PPT activity. To date MIF remains the enzyme with the highest PPT activity in the tautomerase superfamily. It has been reported that MIF is also capable of promiscuously catalyzing the dehalogenation of *trans* 3-chloroacrylic acid; this low-level dehalogenase activity can be enhanced by making a few mutations in MIF²⁶.

cis-3-chloroacrylic acid dehalogenase

The first *cis*-CaaD was isolated from coryneform bacterium strain FG41 and shown to carry out the hydrolytic dehalogenation of *cis*-3-chloroacrylic acid²⁷ (Table 1). Both CaaD and *cis*-CaaD catalyze similar dehalogenation reactions but there are significant differences observed that justifies *cis*-CaaD to be designated as a distinct member in the tautomerase superfamily. Compared to a heterohexameric CaaD, *cis*-CaaD functions as a trimer, with each monomer being 149 residues long, and the overall structure closely resembling CHMI and MIF. The second more important difference that explains the substrate specificity of CaaD and *cis*-CaaD is the shape and geometry of their respective active sites. The active site in CaaD has a more elongated shape, which facilitates the positioning of the *trans* isomer, whereas in *cis*-CaaD the active site pocket assumes a U-shaped bent that is consistent with the shape of the *cis* isomer of 3-chloroacrylic acid⁶. A homolog of *cis*-CaaD from *Corynebacterium glutamicum* (Cg10062) was found to accept both isomers of 3-chloroacrylic acid as substrate, with the *cis* isomer preferred more over the *trans* isomer. Interestingly, Cg10062 also shows a low-level tautomerase activity²⁸. However, another homolog from *Mycobacterium smegmatis* MC2 155 (MsCCH2) was found to be a robust tautomerase and exhibiting only low-level dehalogenase activity²⁹. Both these homologs were found to carry out a promiscuous hydratase activity converting 2-oxo-3-pentynoate into acetopyruvate (Table 2).

More recently, Baas *et al.* identified and characterized a homolog of Cg1002 from *Rhodococcus jostii* RHA1 (RhCC)³⁰. This enzyme was found to carry out none of the known reactions catalyzed by the other members of the superfamily. Intriguingly, RhCC was

found to catalyze an oxygenation reaction. The enzyme accepted 4HPP as a substrate and using molecular oxygen as a co-substrate produced 3-hydroxy-(4-hydroxyphenyl)-pyruvate, 4-hydroxybenzaldehyde and oxalic acid as major products of the reaction (Table 2). Under anaerobic conditions the turnover was not observed. Structural analysis of RhCC showed a similar trimeric arrangement of the monomers, as seen in the other members of the superfamily. The only exception was the presence of a metal binding site at the non-crystallographic three-fold symmetry axis near the C-terminus of the enzyme (Figure 1G). Careful inspection suggested that the bound metal was Mg^{2+} and it was approximately 15 Å away from the N-terminal Pro-1. The metal was stabilized by interactions with three carboxylates from Asp106 residues of the three monomers. A D106A mutant of RhCC showed the loss of the metal binding site but it did not affect the oxygenation reaction. As such, RhCC is the first cofactor-independent oxygenase in the tautomerase superfamily and this is the first instance when the β - α - β structural fold has been shown to host oxygenation chemistry.

Malonate semialdehyde decarboxylase

Poelarends *et al.* identified the first member of the tautomerase superfamily that can catalyze the decarboxylation of malonate semialdehyde to yield acetaldehyde and CO_2 ³¹ (Table 1). This enzyme (MSAD) also shows a familiar trimeric structure (Figure 1H) and absence of any metal-ion in the Pro-1 pocket. The pK_a for Pro-1 was determined to be 9.2, which would allow it to remain protonated at cellular pH³². An Asp-37 residue, which participates in hydrogen bonding interaction with Pro-1 via a water molecule, was suggested to be an important residue during catalysis. In addition to decarboxylation, MSAD also displays hydratase activity on 2-oxo-3-pentynoate^{32, 34}.

Concluding remarks

The tautomerase superfamily of enzymes is a set of homologous proteins that have evolved to carry out different enzymatic reactions using a highly similar structural fold and an N-terminal Pro-1 residue as a key catalytic residue. These enzymes are ubiquitous in nature and found to carry out important functions in various organisms, ranging from unicellular bacteria to higher eukaryotic life forms. Extensive studies on different members of the superfamily have revealed that they share promiscuous activities, which strongly suggests that these enzymes have undergone divergent evolution from a common ancestor. The promiscuous activities can be seen as the vestige of the functions of the common progenitor. Furthermore, discovery of new promiscuous functions, in 4-OT especially, has opened new avenues for application of tautomerase superfamily members in the field of biocatalysis. It is an excellent example of how in-depth understanding of reaction mechanisms and chemical characteristics of key residues can further help in discovery of new functions, which can be successfully employed in generation of compounds of commercial value. In the versatile β - α - β structural motif, nature has discovered a simple yet a robust and sturdy scaffold, which can be evolved to host different chemistries.

SECTION 2

AN INTRODUCTION INTO THE ASPARTASE/FUMARASE SUPERFAMILY OF ENZYMES

Introduction

C-N lyases are enzymes that are involved in the cleavage of C-N bonds to release ammonia or amines. This reaction is important for the production of intermediates of different metabolic pathways and hence these enzymes are found ubiquitously in microbes, plants and animals³⁵. Enzymes with C-N lyase activity have been identified in different enzymatic classes exhibiting diverse structural folds and catalytic mechanism for the cleavage of C-N bonds^{36, 37}. From a commercial point of view, the reverse reaction (*i.e.*, the addition of ammonia and amines to various acceptor substrates) is of particular interest, because introducing C-N bonds in various organic substrates, in a chemo- and stereoselective manner, is still a formidable challenge for the synthetic chemists³⁸. Moreover, the reactions that are generally employed by the organic chemists involve the use of expensive catalysts and are often conducted under harsh conditions³⁹. C-N lyases present themselves as attractive and milder catalysts for the formation of C-N bonds with both regio- and stereoselectivity. It has been demonstrated that different C-N lyases can be used for the successful production of enantiomerically pure amino acids and amines⁴⁰. These speciality chemicals are important building blocks for pharmaceuticals, agrochemicals and food additives and there is a significant demand for such chemicals in the industry. Thus a lot of research has been focused on identifying novel C-N lyases and characterizing them both functionally and structurally.

The aspartase/fumarase superfamily is one such group of homologous enzymes, which have been of interest and studied in great details for their ability to catalyze C-N bond-forming reactions⁴¹. Most members of this superfamily reversibly process succinyl-containing substrates to generate fumarate as one of the common products. The different members include aspartate ammonia lyase (also known as aspartase), fumarase, argininosuccinate lyase (ARL, which also includes d-crystallin), adenylosuccinate lyase (ADL) and 3-carboxy-*cis,cis*-muconate lactonizing enzyme (CMLE). Note that the CMLE-catalyzed reaction results in the formation of a lactone instead of fumarate (Scheme 1).

C-N bond forming reactions catalyzed by different superfamily members

Aspartase catalyzes the reversible amination of fumarate to produce L-aspartic acid⁴² (Scheme 1A). ADL and ARL reversibly convert adenylosuccinate and argininosuccinate to produce adenosine monophosphate and arginine respectively, together with the release of fumarate⁴³. Fumarase, on the other hand is the only member of the superfamily, which catalyzes the reversible addition of water to fumarate to yield L-malate⁴⁴. The CMLE-catalyzed reaction (Scheme 1B) is another exception, which produces muconolactone as a product⁴⁵. More recently, EDDS lyase, a new member from the superfamily was characterized structurally and functionally and found to reversibly degrade the metal

chelator ethylenediamine-*N,N'*-disuccinic acid (*S,S*-EDDS) (Scheme 1C) to form fumarate and ethylenediamine in an unusual sequential two-step reaction⁴⁶. Of these enzymes, aspartases and EDDS lyase have been of particular interest and studied in great details for their ability to produce compounds of commercial value^{37, 46}.

A **B** **C**

Scheme 1. A) Reversible conversion of different succinyl-containing substrates catalyzed by members of the aspartase/fumarase superfamily of enzymes. B) CMLE catalyzed formation of muconolactone from 3-carboxy-*cis,cis*-muconate. C) Reversible two-step breakdown of *S,S*-EDDS catalyzed by EDDS lyase to generate two molecules of fumaric acid and one molecule of ethylenediamine.

A common structural fold

All the members in the superfamily have been structurally characterized and found to exhibit a conserved structural fold^{41, 44, 47-51}. The sequences share as little as ~15% identity, yet the overall structural architecture and assembly shows remarkable conservation. Each monomeric unit, ~500 residues long, can be subdivided into three domains: an N-terminal domain, a central-helix domain and finally a C-terminal domain (Figure 1A). All the three domains are largely composed of α -helices motifs. The physiologically active entity, a tetramer, is formed by the interaction between α -helices from the central-domain to form a compact 20 helical bundle, which consists of four composite active sites (Figure 1B).

Figure 1. A) Cartoon representation of an aspartase (AspB from *Bacillus* sp. YM55-1, PDB id 3R6Q)⁴⁷ monomer, showing N-terminal domain, central domain and C-terminal domain in blue, green and pink respectively. The SS loop is highlighted in red. B) The functional AspB tetramer forms a compact structure. The other three monomers are shown in light grey. C) Superposition of ligand-free (yellow) and L-aspartate bound (cyan) AspB (PDB id 3R6V)⁴⁷ illustrating the movement of SS loop upon substrate binding. The ligand is shown in ball and sticks and the serine residues as sticks.

The residues forming the active site are contributed by three conserved regions (C1, C2 and C3), which are all located in the central-domain of each monomer. The C3 conserved region has a signature sequence GSSXXPKXNPXXXE and it forms a loop in the active site. This loop, named SS loop after the two serine residues found in this structural motif, plays an important role during catalysis (Figure 1A,C).

The SS loop and its role in substrate binding

Crystal structures have shown the crucial role this flexible SS loop plays during catalysis. Structures of AspB, an aspartase, were solved both in the presence and absence of substrate L-aspartate⁴⁷. Comparison of both the structures revealed that the SS loop, which adopts an open conformation in an unliganded state, moves in and closes over the active site when the substrate binds. This transition from an open to closed state is one of the key events during catalysis (Figure 1C). In the occluded active site the substrate is stabilized by extensive hydrogen-bonding interactions with residues from all the three-conserved regions⁴¹. This tight binding allows L-aspartate to assume a high-energy enediolate-like conformation. More importantly, the substrate adopts a favorable position, which allows the first serine residue (Ser318 in AspB) from the SS loop to act a catalytic base and abstract the *pro-R* proton from the *C β* position of the substrate.

A general acid-base catalytic mechanism

It has been proposed that all the members of this superfamily employ a common acid-base catalytic mechanism to carry out an *anti* 1,2-addition-elimination reaction^{41, 52-54} (Scheme 2). The first step of the reaction entails the abstraction of proton from the substrate

Scheme 2. General acid-base catalytic mechanism for the *anti* 1,2-addition-elimination reaction (adapted from ref. 41).

by a residue acting as a base (serine in this case). The resultant carbanion is stabilized in the active site as an enediolate intermediate, which collapses to release fumarate and the corresponding leaving group. The rate-limiting step for this reaction is the cleavage of the C-N bond, which may be facilitated by a residue in the active site playing the role of a general acid and donating a proton to the leaving group⁵⁵. A highly conserved histidine residue (from C2 region) has been proposed to play this role of catalytic acid, however not all family members are dependent on this residue for efficient catalysis⁴¹.

Concluding remarks

The use of enzymes to carry out challenging chemistries is an attractive prospect. C-N lyases are of special importance as they have the ability to form novel compounds with relative ease when they are working in the reverse mode. The aspartase/fumarase superfamily of enzymes harbors a unique group of enzymes employing an identical catalytic mechanism on different substrates to achieve reversible cleavage of C-N bonds. Some members of this superfamily have demonstrated their potential as C-N bond-forming biocatalysts; however, the restricted and narrow substrate scope remains a bottleneck. In such a scenario, laboratory evolution of these enzymes becomes extremely important. Discovery of new superfamily members with better stability and expanded substrate scope is also highly desirable.

REFERENCES

1. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* 6:R91–105
2. Copley, S. D. (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* 7:265–72
3. Khersonsky, O., Roodveldt, C., and Tawfik, D. S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* 10:498–508
4. Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* 30:409–25
5. Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79:471–505
6. Poelarends, G. J., Puthan Veetil, V., and Whitman, C. P. (2008) The chemical versatility of the β - α - β fold: Catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life Sci.* 65:3606–3618
7. Whitman, C. P. (2002) The 4-oxalocrotonate tautomerase family of enzymes: how nature makes new enzymes using a beta-alpha-beta structural motif. *Arch. Biochem. Biophys.* 402:1–13
8. Baas, B.-J., Zandvoort, E., Geertsema, E. M., and Poelarends, G. J. (2013) Recent advances in the study of enzyme promiscuity in the tautomerase superfamily. *ChemBioChem*, 14(8): 917–926
9. Whitman, C. P., Aird, B. A., Gillespie, W. R., and Stolowich, N. J. (1991) Chemical and enzymic ketonization of 2-hydroxymuconate, a conjugated enol. *J. Am. Chem. Soc.* 113:3154–3162
10. Subramanya, H. S., Roper, D. I., Dauter, Z., Dodson, E. J., Davies, G. J., Wilson, K. S., and Wigley, D. B. (1996) Enzymatic ketonization of 2-hydroxymuconate: specificity and mechanism investigated by the crystal structures of two isomerases. *Biochemistry* 35: 792–802
11. Taylor, A. B., Czerwinski, R. M., Johnson Jr., W. H., Whitman, C. P., and Hackert, M. L. (1998) Crystal structure of 4-oxalocrotonate tautomerase inactivated by 2-oxo-3-pentynoate at 2.4 Å resolution: analysis and implications for the mechanism of inactivation and catalysis. *Biochemistry* 37: 14692–14700
12. Poelarends, G. J., Saunier, R. and Janssen, D. B. (2001) trans-3-Chloroacrylic acid dehalogenase from *Pseudomonas pavonaceae* 170 shares structural and mechanistic similarities with 4-oxalocrotonate tautomerase. *J. Bacteriol.* 183:4269–4277
13. Azurmendi, H. F., Wang, S. C., Massiah, M. A., Poelarends, G. J., Whitman, C. P. and Mildvan, A. S. (2004) The roles of active-site residues in the catalytic mechanism of trans-3-chloroacrylic acid dehalogenase: a kinetic, NMR, and mutational analysis. *Biochemistry* 43, 4082–4091
14. Poelarends, G. J., Almrud, J. J., Serrano, H., Darty, J. E., Johnson, Jr.W.H., Hackert, M. L. and Whitman, C. P. (2006) Evolution of enzymatic activity in the tautomerase superfamily: mechanistic and structural consequences of the L8R mutation in 4-oxalocrotonate tautomerase. *Biochemistry* 45: 7700–7708
15. Wang, S. C., Johnson, Jr., W. H., and Whitman, C. P. (2003) The 4-oxalocrotonate tautomerase- and YwhB-catalyzed hydration of 3E-haloacrylates: implications for the evolution of new enzymatic activities *J. Am. Chem. Soc.* 125:14282–14283
16. Wang, S. C., Johnson, Jr., W. H., Czerwinski, R. M., Stamps, S. L., and Whitman, C. P. (2007) Kinetic and stereochemical analysis of YwhB, a 4-oxalocrotonate tautomerase homologue in *Bacillus subtilis*: Mechanistic implications for the YwhB- and 4-oxalocrotonate tautomerase-catalyzed reactions. *Biochemistry* 46:11919–11929
17. Poelarends, G. J., Johnson, Jr., W. H., Serrano, H., and Whitman, C. P. (2007) The phenylpyruvate tautomerase activity of trans-3-chloroacrylic acid dehalogenase: Evidence for an enol intermediate in the dehalogenase reaction *Biochemistry* 46: 9596– 9604

18. Zandvoort, E., Baas, B.-J., Quax, W. J., and Poelarends, G. J. (2011) Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: Enamine formation and aldolase activity. *ChemBioChem*, 12(4): 602-609
19. Zandvoort, E., Geertsema, E. M., Quax, W. J., and Poelarends, G. J. (2012) Enhancement of the promiscuous aldolase and dehydration activities of 4-oxalocrotonate tautomerase by protein engineering. *ChemBioChem*, 13(9):1274-1277
20. Zandvoort, E., Geertsema, E. M., Baas, B.-J., Quax, W. J., and Poelarends, G. J. (2012) Bridging between organocatalysis and biocatalysis: Asymmetric addition of acetaldehyde to β -nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angewandte Chemie-International Edition*, 51(5):1240-1243
21. Whitman, C. P., Hajipour, G., Watson, R. J., Johnson, Jr. W. H., Bembenek, M. E. and Stolowich, N. J. (1992) Stereospecific ketonization of 2-hydroxymuconate by 4-oxalocrotonate tautomerase and 5-(carboxymethyl)-2-hydroxymuconate isomerase. *J. Am. Chem. Soc.* 114:10104–10110
22. Subramanya, H. S., Roper, D. I., Dauter, Z., Dodson, E. J., Davies, G. J., Wilson, K. S. and Wigley, D. B. (1996) Enzymatic ketonization of 2-hydroxymuconate: specificity and mechanism investigated by the crystal structures of two isomerases. *Biochemistry* 35:792–802
23. Rosengren, E., Aman, P., Thelin, S., Hansson, C., Ahlfors, S., Bjork, P., Jacobsson, L. and Rorsman, H. (1997) The macrophage migration inhibitory factor MIF is a phenylpyruvate tautomerase. *FEBS Lett.* 417:85–88
24. Sun, H.-W., Bernhagen, J., Bucala, R. and Lolis, E. (1996) Crystal structure at 2.6 Å resolution of human macrophage migration inhibitory factor. *Proc. Natl. Acad. Sci.* 93, 5191–5196
25. Lubetsky, J. B., Swope, M., Dealwis, C., Blake, P., and Lolis, E. (1999) Pro-1 of macrophage migration inhibitory factor functions as a catalytic base in the phenylpyruvate tautomerase activity. *Biochemistry* 38:7346-7354
26. Wasiel, A. A., Baas, B.-J., Zandvoort, E., Quax, W. J., and Poelarends, G. J. (2012) Dehalogenation of an anthropogenic compound by an engineered variant of the mouse cytokine macrophage migration inhibitory factor. *ChemBioChem*, 13(9):1270-1273
27. de Jong, R. M., Brugman, W., Poelarends, G. J., Whitman, C. P., and Dijkstra, B. W. (2004) The X-ray structure of trans-3-chloroacrylic acid dehalogenase reveals a novel hydration mechanism in the tautomerase superfamily. *J Biol. Chem.* 279:11546 - 11552
28. Poelarends, G. J., Serrano H., Person, M. D., Johnson, Jr., W. H., and Whitman, C. P. (2008) Characterization of Cg10062 from *Corynebacterium glutamicum*: Implications for the evolution of cis-3-chloroacrylic acid dehalogenase activity in the tautomerase superfamily *Biochemistry* 47:8139-8147
29. Baas, B.-J., Zandvoort, E., Wasiel, A. A., Quax, W. J., and Poelarends, G. J. (2011) Characterization of a newly identified Mycobacterial tautomerase with promiscuous dehalogenase and hydratase activities reveals a functional link to a recently diverged cis-3-chloroacrylic acid dehalogenase. *Biochemistry* 50:2889-2899
30. Baas, B.-J., Poddar, H., Geertsema, E. M., Rozeboom, H. J., de Vries, M., Permentier, H. P., Thunnissen, A.-M. W. H., and Poelarends, G. J. (2015) Functional and structural characterization of an unusual cofactor-independent oxygenase. *Biochemistry*, 54:1219-1232
31. Poelarends, G. J., Johnson, Jr. W. H., Murzin, A. G., and Whitman, C. P. (2003) Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: Identification of a new activity in the tautomerase superfamily. *J. Biol. Chem.*, 278: 48674-48683
32. Poelarends, G. J., Serrano, H., Johnson, Jr. W. H. and Whitman, C. P. (2005) Inactivation of malonate semialdehyde decarboxylase by 3-halopropiolates: evidence for hydratase activity. *Biochemistry* 44:9375–9381

33. Poddar, H., Rahimi, M., Geertsema, E., Thunnissen, A., and Poelarends, G. J. (2015) Evidence for the formation of an enamine species during aldol and Michael-type addition reactions promiscuously catalyzed by 4-oxalocrotonate tautomerase. *ChemBioChem*, 16(5):738-741
34. Almrud, J. J., Poelarends, G. J., Johnson Jr., W. H., Serrano, H., Hackert, M. L., Whitman, and C. P. (2005) Crystal structures of the wild-type, P1A mutant, and inactivated malonate semialdehyde decarboxylase: A structural basis for the decarboxylase and hydratase activities. *Biochemistry* 44: 14818-14827
35. Wu, B., Szymański, W., Crismaru, C. G., Feringa, B. L., and Janssen, D. B. (2012) C-N lyases catalyzing addition of ammonia, amines and amides to C=C and C=O bonds. *Enzyme Catalysis in Organic Synthesis* (3rd Edition) 749–778
36. Parmeggiani, F., Weise, N. J., Ahmed, S. T., and Turner, N. J. (2018) Synthetic and therapeutic applications of ammonia lyases and aminomutases. *Chem. Rev.* 118:73-118
37. De Villiers, M., Puthan Veetil, V., Raj, H., De Villiers, J., and Poelarends, G. J. (2012) Catalytic mechanisms and biocatalytic applications of aspartate and methylaspartate ammonia lyases. *ACS Chem. Biol.* 7:1618–1628
38. Hili, R., and Yudin, A. K. (2006) Making carbon-nitrogen bonds in biological and chemical synthesis. *Nat. Chem. Biol.* 2:284-287
39. Bariwal, J., and Van der Eycken, E., (2013) C-N bond forming cross-coupling reactions: an overview. *Chem. Soc. Rev.* 42:9283-9303
40. Turner, N. J. (2011) Ammonia lyases and aminomutases as biocatalysts for the synthesis of α -amino and β -amino acids. *Curr. Opin. Chem. Biol.* 15:234-240
41. Puthan Veetil, V., Fibriansah, G., Raj, H., Thunnissen, A.-M. W. H., and Poelarends, G. J. (2012) Aspartase/Fumarase superfamily: A common catalytic strategy involving general base-catalyzed formation of a highly stabilized aci-carboxylate intermediate. *Biochemistry* 51:4237–4243
42. Viola, R. E. (2000) L-Aspartase: New tricks from an old enzyme. *Adv. Enzymol. Relat. Areas Mol. Biol.* 74:295–341
43. Ratner, S. (1972) Argininosuccinate lyases and adenylosuccinate lyases. In *The Enzymes* (Boyer, P., Ed.) Vol. 7:167–197, Academic Press, New York
44. Weaver, T. M., Levitt, D. G., Donnelly, M. I., Stevens, P. P., and Banaszak, L. J. (1995) The multisubunit active site of fumarase C from *Escherichia coli*. *Nat. Struct. Biol.* 2:654–662
45. Williams, S. E., Woolridge, E. M., Ransom, S. C., Landro, J. A., Babbitt, P. C., and Kozarich, J. W. (1992) 3-Carboxy-*cis,cis*-muconate lactonizing enzyme from *Pseudomonas putida* is homologous to the class II fumarase family: A new reaction in the evolution of a mechanistic motif. *Biochemistry* 31:9768–9776
46. Poddar, H., De Villiers, J., Zhang, J., Puthan Veetil, V., Raj, H., Thunnissen, A. M. W. H., and Poelarends, G. J. (2018) Structural basis for the catalytic mechanism of ethylenediamine-N,N'-disuccinic acid lyase, a carbon-nitrogen bond-forming enzyme with broad substrate scope. *Biochemistry* 57:3752-3763
47. Fibriansah, G., Puthan Veetil, V., Poelarends, G. J., and Thunnissen, A. M. W. H. (2011) Structural basis for the catalytic mechanism of aspartate ammonia lyase. *Biochemistry* 50:6053–6062
48. Toth, E. A., and Yeates, T. O. (2000) The structure of adenylosuccinate lyase, an enzyme with dual activity in the de novo purine biosynthetic pathway. *Structure* 8:163–174
49. Sampaleanu, L. M., Vallée, F., Slingsby, C., and Howell, P. L. (2001) Structural studies of duck δ 1 and δ 2 crystallin suggest conformational changes occur during catalysis. *Biochemistry* 40:2732–2742

50. Yang, J., Wang, Y., Woolridge, E. M., Arora, V., Petsko, G. A., Kozarich, J. W., and Ringe, D. (2004) Crystal structure of 3-carboxycis, cis-muconate lactonizing enzyme from *Pseudomonas putida*, a fumarase class II type cycloisomerase: Enzyme evolution in parallel pathways. *Biochemistry* 43:10424–10434
51. Sampaleanu, L. M., Vallée, F., Thompson, G. D., and Howell, P. L. (2001) Three-dimensional structure of the argininosuccinate lyase frequently complementing allele Q286R. *Biochemistry* 40:15570–15580
52. Yoon, M. Y., Thayer-Cook, K. A., Berdis, A. J., Karsten, W. E., Schnackerz, K. D., and Cook, P. F. (1995) Acid-base chemical mechanism of aspartase from *Hafnia alvei*. *Arch. Biochem. Biophys.* 320:115–122
53. Blanchard, J. S., and Cleland, W. W. (1980) Use of isotope effects to deduce the chemical mechanism of fumarase. *Biochemistry* 19:4506–4513
54. Garrard, L. J., Bui, Q. T., Nygaard, R., and Raushel, F. M. (1985) Acid-base catalysis in the argininosuccinate lyase reaction. *J. Biol. Chem.* 260:5548–5553
55. Nuiiry, I. I., Hermes, J. D., Weiss, P. M., Chen, C. Y., and Cook, P. F. (1984) Kinetic mechanism and location of rate-determining steps for aspartase from *Hafnia alvei*. *Biochemistry* 23:5168–5175

c h a p t e r t w o

EVIDENCE FOR THE FORMATION OF AN ENAMINE SPECIES DURING ALDOL AND MICHAEL-TYPE ADDITION REACTIONS PROMISCUOUSLY CATALYZED BY 4-OXALOCROTONATE TAUTOMERASE

H. Poddar*
M. Rahimi*
E. M. Geertsema
A.-M. W. H. Thunnissen
G. J. Poelarends

*These authors contributed equally to this work

Published in ChemBioChem 2015, 16, 738-741

ABSTRACT

The enzyme 4-oxalocrotonate tautomerase (4-OT), which has a catalytic N-terminal proline residue (Pro-1), can promiscuously catalyze various carbon-carbon bond-forming reactions, including the aldol condensation of acetaldehyde with benzaldehyde to yield cinnamaldehyde and the Michael-type addition of acetaldehyde to a wide variety of nitroolefins to yield valuable γ -nitroaldehydes. To gain insight into how 4-OT catalyzes these unnatural reactions, we carried out both exchange studies in D_2O and X-ray crystallography studies. The former establishes that H-D exchange within acetaldehyde is 4-OT-catalyzed and that the Pro-1 residue is crucial for this activity. The latter shows that Pro-1 of 4-OT has reacted with acetaldehyde to give an enamine species. These results provide evidence for a mechanism of the 4-OT-catalyzed aldol and Michael-type addition reactions in which acetaldehyde is activated for nucleophilic addition via Pro-1 dependent formation of an enamine intermediate.

4-oxalocrotonate tautomerase (4-OT) is a member of the tautomerase superfamily, a group of homologous proteins that share a characteristic β - α - β structural fold and a unique catalytic N-terminal proline (Pro-1).^[1,2] 4-OT catalyzes the conversion of 2-hydroxy-2,4-hexadienedioate (**1**) to 2-oxo-3-hexenedioate (**2**) (Scheme 1) as part of a catabolic pathway for aromatic hydrocarbons in *Pseudomonas putida* mt-2.^[3,4] The Pro-1 residue acts as a general base that abstracts the 2-hydroxyl proton of **1** for delivery to the C-5 position to yield **2**. Pro-1 can function as a general base because the prolyl nitrogen has a pK_a of ~ 6.4 and exists largely as the uncharged species at cellular pH.^[5]

In addition to its natural tautomerase activity, 4-OT can promiscuously catalyze various carbon-carbon bond-forming reactions, including the aldol condensation of acetaldehyde (**3**) with benzaldehyde (**4**) to yield cinnamaldehyde (**6**) and the Michael-type addition of acetaldehyde (**3**) to a variety of nitroalkenes (**7**) to yield chiral γ -nitroaldehydes (**8**) (Scheme 2).^[6-13] γ -Nitroaldehydes are versatile and practical precursors for chiral γ -aminobutyric acid (GABA) analogues such as marketed pharmaceuticals Baclofen, Pregabalin, Phenibut and Rolipram.^[14-19]

Site-directed mutagenesis and labeling experiments suggested a key catalytic role for Pro-1 in the 4-OT-catalyzed carbon-carbon bond-forming reactions.^[6,7,12] Although NaCNBH_3 trapping suggested that a Schiff base can form between 4-OT's Pro-1 residue and acetaldehyde, this observation does not rule out the possibility that Pro-1 acts as a catalytic base (like in 4-OT's natural tautomerase activity). Hence, compelling evidence for the precise mechanistic role of Pro-1 in the 4-OT-catalyzed carbon-carbon bond-forming reactions is still lacking. To gain further insight into how 4-OT promiscuously catalyzes aldol and Michael-type addition reactions, we carried out both exchange studies in D_2O and X-ray crystallography studies. The former establishes Pro-1 dependent deprotonation of acetaldehyde; the latter reveals formation of an enamine species between acetaldehyde and Pro-1.

Given that Pro-1 has the correct protonation state ($pK_a \sim 6.4$) to be able to act as a base or nucleophile at pH 7.3, we anticipated that 4-OT would initiate catalysis by formation of either an enolate or enamine intermediate (Scheme 3). Both mechanisms

Scheme 1. Natural proton-transfer reaction catalyzed by 4-OT.

Scheme 2. Aldol condensation (A) and Michael-type addition (B) reactions promiscuously catalyzed by 4-OT.

involve polarization of the carbonyl group of acetaldehyde, which will lower its $\text{p}K_{\text{a}}$ value (~ 17 in aqueous solutions), and deprotonation at C2 (Scheme 3). To evaluate whether one (or all three sequentially) of the C2 hydrogens can be removed as a proton by 4-OT during catalysis, we investigated the ability of wild-type 4-OT (WT 4-OT) and the Pro-1-Ala mutant (4-OT P1A)^[20] to catalyze hydrogen-deuterium (H-D) exchange within acetaldehyde. Accordingly, WT 4-OT and 4-OT P1A (0.73 mol% compared to **3**) were incubated with 20 mM acetaldehyde in 20 mM NaD_2PO_4 buffer (pD 7.5, which corresponds to pH 7.3), and the progress of the reactions was followed by ^1H NMR spectroscopy (Figures 1 and S1). A control experiment, in which **3** was incubated in 20 mM NaD_2PO_4 buffer (pD 7.5) in the absence of enzyme, was also performed. Notably, for each reaction mixture, an equilibrium between the hydrated (59%) and unhydrated (41%) form of **3** was reached in the time between mixing all reaction components and recording the first ^1H NMR spectrum. ^1H NMR spectroscopic signals of the unhydrated (i.e., acetaldehyde: 2.24 and 9.67 ppm) and hydrated (i.e., ethane-1,1-diol- d_2 : 1.32 and 5.25 ppm) form of **3** are shown in Figure 1.

Interestingly, the acidic protons of substrate **3**, which are located at the C2 position (marked with b in Figure 1), were almost completely exchanged (94%, 24 h) with deuterium

Scheme 3. Proposed mechanisms for the 4-OT-catalyzed hydrogen-deuterium exchange within **3**, with a catalytic role for Pro-1 as a base (A) or as a nucleophile (B).

in the reaction mixture incubated with WT 4-OT (Figure 1, spectrum E; Figure S1). The exchange most likely takes place at C2 of the unhydrated form of **3** (i.e., acetaldehyde) and not at C2 of the hydrated form (i.e., ethane-1,1-diol-*d*₂) since the protons at C2 of the latter are not acidic. However, since the rate for reaching equilibrium between unhydrated and hydrated form is relatively high compared to the rate of H-D exchange, the vanishing of signals b (protons at C2 of unhydrated form of **3**) and d (protons at C2 of hydrated form of **3**) was witnessed in equal proportion (spectrum E, Figure 1). A relatively low rate of H-D exchange was found for the control sample without enzyme (11%, 24 h) and the sample incubated with 4-OT P1A (19%, 24 h) (Figures 1 and S1). These data indicate that the H-D exchange within **3** is enzyme-catalyzed and that the Pro-1 residue is essential for catalysis.

The H-D exchange activity indicates that WT 4-OT can indeed deprotonate acetaldehyde, thereby providing evidence for a mechanism for the 4-OT-catalyzed aldol and Michael-type addition reactions in which acetaldehyde is activated for nucleophilic addition via Pro-1 dependent formation of an enolate or enamine intermediate. To distinguish between these two possible intermediates, we determined the crystal structures of native

Figure 1. Stack plot of ^1H NMR spectra. A) ^1H NMR spectrum of acetaldehyde (**3**) incubated in 20 mM NaD_2PO_4 buffer at pD 7.5 for 1 h. Equilibrium is reached between unhydrated and hydrated forms of **3**. Their signals are marked with a,b and c,d respectively; B) Acetaldehyde (**3**) incubated in 20 mM NaD_2PO_4 buffer at pD 7.5 with WT 4-OT for 1 h; C) Acetaldehyde (**3**) incubated in 20 mM NaD_2PO_4 buffer at pD 7.5 with 4-OT P1A mutant for 1 h; D) Acetaldehyde (**3**) incubated in 20 mM NaD_2PO_4 buffer at pD 7.5 for 1 d; E) Acetaldehyde (**3**) incubated in 20 mM NaD_2PO_4 buffer at pD 7.5 with WT 4-OT for 1 d. Acidic protons of **3** (marked with b and d in spectrum A) are completely exchanged with deuterium; F) Acetaldehyde (**3**) incubated in 20 mM NaD_2PO_4 buffer at pD 7.5 with 4-OT P1A for 1 d.

4-OT and 4-OT in complex with **3** (in the absence of NaCNBH_3). Homohexameric 4-OT from *Pseudomonas putida* mt-2 has been crystallized before, in complex with the inhibitor 2-oxo-3-pentynoate, and the structure was solved to 2.4 Å resolution (PDB code 1BJP).^[21] We have crystallized native 4-OT in a new space group (P2₁) and solved its structure to a resolution of 1.94 Å (Figure 2A). The crystallographic R-factor for the final model is 24.8% ($R_{\text{free}} = 28.9\%$, Table S1). The somewhat high values for the R-factors are most probably due to problems with the data from ice-ring interference. Overall, the native 4-OT amino acid residues are well defined in the electron density maps, including those located at the active sites. Structure validation further confirmed the reliability of the refined model.

Figure 2. Hexameric structure of A) native 4-OT (1.94 Å) and B) acetaldehyde-bound 4-OT (1.70 Å). Close-up of the N-terminal proline of C) native 4-OT and D) acetaldehyde-bound 4-OT. Individual chains are depicted in different colours. The grey mesh depicts the composite omit $2F_o - F_c$ maps (contoured at 1.0 σ).

Co-crystallization experiments with substrate **3** resulted in a crystal which belonged to space group C2 with 15 chains of 4-OT in the asymmetric unit (2.5 hexamers, solvent content 40%). The structure was solved to a resolution of 1.70 Å (Figure 2B) and refined to R and R_{free} values of 19.7% and 22.7%, respectively, with excellent geometry (Table S1).

Analysis of the electron densities of the N-terminal proline residues in the structure of 4-OT complexed with **3** evidently indicates that a covalent modification has taken place. Of the 15 active sites present in the asymmetric unit, 10 clearly show extra electron density protruding from the amino group of the Pro-1 pyrrolidine ring (Figure 2D). This additional electron density was not visible in the structure of native 4-OT (Figure 2C). It is known that secondary amines react with carbonyl compounds to preferably form enamines.^[22] Accordingly, reaction of acetaldehyde with Pro-1 of 4-OT would result in an ethylene moiety bound to the nitrogen atom of Pro-1. Therefore, ethylene covalently linked to Pro-1 in an enamine conformation was used as a model to account for the extra electron density. Subsequent refinements of this model support the presence of this enamine species. It should be noted, though, that the extra electron densities found at the Pro-1 residues that have reacted with **3** are not very well defined, most likely because these

prolines are not fully modified. This is apparent from the significant higher B-factors of the N-linked ethylene atoms, as compared to the atoms in the pyrrolidine ring. As a result, the conformation of the enamine adduct cannot be unambiguously defined, especially with respect to the position of the terminal methylene group. It is important to emphasize that non-covalently bound acetaldehyde (or the corresponding enolate anion) was not observed in the structure of 4-OT complexed with acetaldehyde.

C α -backbone superposition of the structure of native 4-OT with that of 4-OT in complex with **3** resulted in a root-mean-square deviation of only 0.25 Å (Figure 3A,B).

Figure 3. Stereo view of C α backbone superposition of native and acetaldehyde-bound 4-OT as A) dimer or B) hexamer (which is a trimer of dimers). C) Superposition of active-site residues of native 4-OT and acetaldehyde-bound 4-OT. The structure of native 4-OT is depicted in green whereas the acetaldehyde-bound 4-OT structure is shown in orange. The residues are depicted as sticks and the apostrophes indicate that the residues are from the neighboring chains.

Residues lining the Pro-1 pocket adopt similar conformations in both the structures with the only exception of Arg-11 from the neighboring chain which seems to be flexible and favors two alternative conformations (Figure 3C). This shows that modification of Pro-1 by acetaldehyde does not result in any significant structural change in the vicinity of this N-terminal residue. To the best of our knowledge, this is the first reported structure from a tautomerase superfamily member with an enamine adduct on the N-terminal proline residue.

In summary, we provide evidence that the 4-OT-catalyzed C-C bond-forming aldol and Michael-type addition reactions proceed through an enamine intermediate. Hence, these reactions are initiated by nucleophilic attack of Pro-1 on the carbonyl carbon of **3** to give an iminium ion, which upon deprotonation leads to the formation of an enamine intermediate (Scheme 3B). A reaction between this nucleophilic intermediate and an electrophilic substrate such as **4** or **7** results in carbon-carbon bond formation. While the proposed mechanism of the reaction mimicks that used by proline-based organocatalysts,^[23,24] it clearly differs from that used by class I aldolases.^[25] Indeed, class I aldolases use the primary amine of a lysine to form enamines with carbonyl substrates, whereas 4-OT appears to be unique in using the secondary amine of a proline as the nucleophile catalyst to form enamines with carbonyl substrates.

AUTHOR CONTRIBUTION

H.P and M.R. designed and performed all aspects of protein crystallography and NMR experiments respectively. E.M.G., A.M.W.H.T. and G.J.P supervised the scientific work. All authors contributed to writing the paper.

DECLARATION

Dr. Mehran Rahimi also used this paper as a chapter in his PhD thesis entitled Catalytic promiscuity of a proline-based tautomerase – Aldolase activities and enzyme redesign (University of Groningen, 2016, ISBN: 978-90-367-8841-0).

ACKNOWLEDGMENTS

This research was financially supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no 242293 to G.J.P. The authors wish to thank the staff at the ESRF (Grenoble) for providing facilities and assistance for data collection.

REFERENCES

1. G. J. Poelarends, V. Puthan Veetil, C. P. Whitman, *Cell. Mol. Life Sci.* **2008**, *65*, 3606–3618.
2. C. P. Whitman, *Arch. Biochem. Biophys.* **2002**, *402*, 1–13.
3. S. Harayama, M. Rekik, K. L. Ngai, L. N. Ornston, *J. Bacteriol.* **1989**, *171*, 6251–6258.
4. C. P. Whitman, B. A. Aird, W. R. Gillespie, N. J. Stolowich, *J. Am. Chem. Soc.* **1991**, *113*, 3154–3162.
5. J. T. Stivers, C. Abeygunawardana, A. S. Mildvan, G. Hajipour, C. P. Whitman, *Biochemistry* **1996**, *35*, 814–823.
6. E. Zandvoort, E. M. Geertsema, B.-J. Baas, W. J. Quax, G. J. Poelarends, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 1240–1243.
7. Y. Miao, E. M. Geertsema, P. G. Tepper, E. Zandvoort, G. J. Poelarends, *ChemBioChem* **2013**, *14*, 191–194.
8. E. M. Geertsema, Y. Miao, P. G. Tepper, P. de Haan, E. Zandvoort, G. J. Poelarends, *Chem. Eur. J.* **2013**, *19*, 14407–14410.
9. B.-J. Baas, E. Zandvoort, E. M. Geertsema, G. J. Poelarends, *ChemBioChem* **2013**, *14*, 917–926.
10. E. M. Geertsema, G. J. Poelarends, in *Science of Synthesis: Biocatalysis in Organic Synthesis*, Vol. 2 (Eds: K. Faber, W. D. Fessner, N. Turner), Thieme Chemistry, Stuttgart, Germany, in press.
11. E. M. Geertsema, Y. Miao, G. J. Poelarends, *Practical Methods in Biocatalysis and Biotransformations*, Vol. 3 (Eds: J. Whittall, P. Sutton, W. Kroutil), Wiley, UK, in press.
12. E. Zandvoort, B.-J. Baas, W. J. Quax, G. J. Poelarends, *ChemBioChem* **2011**, *12*, 602–609.
13. E. Zandvoort, E. M. Geertsema, W. J. Quax, G. J. Poelarends, *ChemBioChem* **2012**, *13*, 1274–1277.
14. J. Vicario, D. Badía, L. Carrillo, *Synthesis* **2007**, 2065–2092.
15. H. Gotoh, H. Ishikawa, Y. Hayashi, *Org Lett.* **2007**, *9*, 5307–5309.
16. S. B. Tsogoeva, *Eur. J. Org. Chem.* **2007**, 1701–1716.
17. S. Sulzer-Mossé, A. Alexakis, *Chem. Commun.* **2007**, *14*, 3123–3135.
18. F. Felluga, V. Gombac, G. Pitacco, E. Valentin, *Tetrahedron: Asymmetry* **2005**, *16*, 1341–1345.
19. O. V. Maltsev, A. S. Kucherenko, I. P. Beletskaya, V. A. Tartakovsky, S. G. Zlotin, *Eur. J. Org. Chem.* **2010**, 2927–2933.
20. Before applying the 4-OT P1A mutant, which has essentially no H-D exchange activity, we first confirmed that purified 4-OT P1A was catalytically active by measuring its promiscuous oxaloacetate decarboxylase activity; see: A. Brik, L. J. D'Souza, E. Keinan, F. Grynszpan, P. E. Dawson, *ChemBioChem* **2002**, *3*, 845–851.
21. A. B. Taylor, R. M. Czerwinski, W. H. Johnson, C. P. Whitman, M. L. Hackert, *Biochemistry* **1998**, *37*, 14692–14700.
22. P. Y. Bruice *J. Am. Chem. Soc.* **1989**, *111*, 962–970.
23. B. List, *Acc. Chem. Res.* **2004**, *37*, 548–57.
24. D. A. Bock, C. W. Lehmann, B. List, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20636–20641.
25. T. D. Machajewski, C.-H. Wong, *Angew. Chem. Int. Ed.* **2000**, *39*, 1352–1374, and references therein.

SUPPORTING INFORMATION

General methods

Protein was analyzed by polyacrylamide gel electrophoresis (PAGE) using sodium dodecyl sulfate (SDS) gels containing polyacrylamide (10%). Coomassie brilliant blue was used to stain the gels. Protein concentrations were determined using the Waddell method.^[1] Enzymatic assays were performed on a V-650 or V-660 spectrophotometer from Jasco (IJsselstein, The Netherlands). ¹H NMR spectra were recorded on a Varian Inova 500 (500 MHz) spectrometer using a pulse sequence for selective presaturation of the water signal. Chemical shifts for protons are reported in parts per million scale and are referenced to H₂O (δ = 4.80). Dynamic light scattering (DLS) experiments were performed using a DynaPro MS800TC instrument (Wyatt Technology Corporation, Santa Barbara, CA) at 20 °C. DLS data were processed and analyzed with Dynamics software (Wyatt Technology Corporation).

Expression and purification of WT 4-OT and 4-OT P1A

The enzymes WT 4-OT and 4-OT P1A were produced in *E. coli* BL21 (DE3) as native proteins without His-tag using the pET20b(+) expression system as described before.^[2] The construction of the expression vectors and the purification procedure for WT 4-OT and 4-OT P1A were reported previously.^[2]

Crystallization of WT 4-OT and 4-OT in complex with acetaldehyde

Purified WT 4-OT was concentrated to 8 mg/mL in 20 mM Tris-HCl buffer (pH 7.5), containing 200 mM NaCl. DLS experiments indicated that the enzyme exists as a hexamer in solution. Attempts to crystallize native 4-OT by using the crystallization condition previously determined for 4-OT inactivated by 2-oxo-3-pentynoate were unsuccessful.^[3] New crystallization conditions were determined by vapour-diffusion sitting-drop experiments in MRC 96-well crystallization plates (Molecular Dimensions), using a Mosquito crystallization robot (TTP Labtech) for automatic drop-dispensing. Protein drops were prepared by mixing 150 nL protein solution with an equal volume of precipitant reservoir solution. Initial crystallization leads were observed in the Pact Premier screen (Molecular Dimensions), and were optimized manually using the vapour-diffusion sitting-drop method. The final crystallization condition was as follows: 0.1 M Bis-Tris propane (pH 8.5), 0.2 M potassium thiocyanate, 20 mM hexaammine cobalt chloride, 20-22% polyethylene glycol 3350 (PEG 3350). Thin rod-shaped crystals appeared within 16 h and grew within 4 d to a maximum size of 250 × 75 × 50 μm^3 . WT 4-OT was also co-crystallized in the presence of 50 mM acetaldehyde to get the structure of 4-OT in which Pro-1 is covalently modified by acetaldehyde (4-OT-enamine).

Data collection, processing and structure determination

Prior to X-ray diffraction data collection, the crystals were transferred to a drop containing the mother liquor supplied with 15% glycerol, and subsequently flash-frozen in liquid nitrogen. The datasets for WT 4-OT and 4-OT-enamine were collected at the ID29 and ID23-2 beamlines, respectively, at the European Synchrotron Radiation Facility (Grenoble, France). Diffraction data were processed, scaled and merged using the programs XDS^[4] and SCALA.^[5] The native WT 4-OT crystal belonged to space group $P2_1$ whereas the structure of 4-OT-enamine belonged to space group C2. A summary of the data collection and model refinement statistics is given in Supplementary Table S1.

Phaser^[6] was used to calculate initial phases by the molecular replacement method using the published structure of 4-OT inactivated by 2-oxo-3-pentynoate (PDB code 1BJP) as a search model. Iterative cycles of refinement in Refmac5, together with manual model building in COOT, were used to improve the structures. The models were completed by performing final cycles of refinement using the program phenix.refine from the Phenix software suite.^[7] The coordinates and topology for the enamine moiety (Pro-1 modified by acetaldehyde) were generated using the Prodrp server.^[8]

Crystal structure analysis

Molprobit was used to validate the stereochemical quality of the models.^[9] Superpositions and calculation of $\text{C}\alpha$ -backbone rmsd values were performed using the protein structure comparison service Fold at the European Bioinformatics Institute (PDBeFold).^[10] PyMOL^[11] (Schrödinger) was used for structure analysis and figure preparations. Coordinates for the structures of native 4-OT and 4-OT modified by acetaldehyde have been deposited with the Protein Data Bank (accession codes 4X19 and 4X1C, respectively).

Preparation of NaD_2PO_4 buffer

NaH_2PO_4 buffer (20 mL, 20 mM; pH 7.3) was lyophilized. Subsequently, the residue was dissolved in D_2O (2 mL), and stirred for 60 min. This solution was again lyophilized, dissolved in D_2O (2 mL), and stirred for 60 min. The resulting mixture was lyophilized once more, after which the residue was dissolved in D_2O (20 mL) yielding a stock solution of NaD_2PO_4 (20 mM; pD 7.5).

Redissolving WT 4-OT and 4-OT P1A mutant in NaD_2PO_4 buffer

A VIVASPIN concentrator (from Sartorius Stedim Goettingen, Germany) with a cut-off filter of 5000 Da was washed four times with H_2O by centrifugation (4000 rpm, 20 min). Subsequently, the concentrator was charged with a solution of 4-OT (either WT or mutant; 300 μL with a concentration of ~ 10 mg/mL in 20 mM NaH_2PO_4 buffer, pH 7.3) and centrifuged (4000 rpm, 30 min). The enzyme was retained on the filter and redissolved in NaD_2PO_4 (200 μL , 20 mM; pD 7.5) and centrifuged (4000 rpm, 30 min). Once more,

the remaining enzyme on the filter was redissolved in NaD_2PO_4 (300 μL , 20 mM; pD 7.5) after which the final enzyme concentration was determined.

^1H NMR spectroscopy assay for H-D exchange within acetaldehyde (3)

In separate experiments, WT 4-OT and 4-OT P1A (145 μM each) were incubated with **3** (20 mM) and 18-crown-6 ether (internal standard; 2.5 mM) at room temperature in NaD_2PO_4 buffer (20 mM; pD 7.5, final volume of 650 μL in an NMR tube). A control sample was prepared containing all the reaction components except for enzyme. The progress of the reactions was followed by ^1H NMR spectroscopy.

Internal standard 18-crown-6 (1,4,7,10,13,16-hexaoxacyclooctadecaan)

^1H NMR (500 MHz, 20 mM NaD_2PO_4 ; pD 7.5): δ 3.68 (s, 24H)

Acetaldehyde (3)

^1H NMR (500 MHz, 20 mM NaD_2PO_4 ; pD 7.5): δ 9.68 (q, $J=3.0$ Hz, 1 H), 2.24 (d, $J=3.0$ Hz, 3H).

Ethane-1,1-diol- d_2

^1H NMR (500 MHz, 20 mM NaD_2PO_4 ; pD 7.5): δ 5.25 (q, $J=4.8$ Hz, 1 H), 1.32 (d, $J=4.8$ Hz, 3 H).

Figure S1. The progress curve of the H-D exchange within 3 in control (□), WT 4-OT (●) and 4-OT P1A samples (▼).

Table S1. Crystallographic data collection and refinement statistics

	WT 4-OT	4-OT-enamine
Data collection		
Spacegroup	P2 ₁	C2
Unit cell dimensions, a, b, c (Å)	58.4, 88.8, 169.8	55.2, 85.1, 170.9,
	$\beta = 94.5^\circ$	$\beta = 96.9^\circ$
Resolution range (Å)	48.7 - 1.9	46.1 - 1.7
	(2.05 - 1.94)	(1.79 - 1.70)
R _{sym}	0.122 (0.54)	0.041 (0.557)
Nr of measurements	407554	270007
Nr of unique reflections	118466	84590
Completeness (%)	93 (96.3)	97.7 (95.6)
Average I/ σ	5.2 (1.9)	12.5 (2.1)
Refinement		
R/ R _{free}	0.248/0.289	0.197 / 0.227
Composition of asymmetric unit (AU)	30 chains	15 chains
total number of atoms	13511	7012
solvent molecules	449	305
others	1 Hexaammine cobalt (III) chloride	2 Hexaammine cobalt (III) chloride
Average B (Å ²)	38.9	42.4
R.m.s.d bonds (Å) /angles (°)	0.006/ 1.0	0.005 / 0.93
Ramachandran plot		
% most favored	99.8	99.4
% allowed	0.2	0.6
% disallowed	0.0	0.0
PDB code	4X19	4X1C

REFERENCES

1. W. J. Waddell, *J. Lab. Clin. Med.* **1956**, *48*, 311-314.
2. E. Zandvoort, B.-J. Baas, W. J. Quax, G. J. Poelarends, *ChemBioChem* **2011**, *12*, 602–609.
3. A. B. Taylor, R. M. Czerwinski, W. H. Johnson, C. P. Whitman, M. L. Hackert, *Biochemistry* **1998**, *37*, 14692–14700.
4. W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 133-144.
5. P. Evans, *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 72-82.
6. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
7. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta. Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 213-221.
8. W. Schuttelkopf, D. M. F. van Aalten, *Acta. Crystallogr. D. Biol. Crystallogr.* **2004**, *60*, 1355-1363.
9. V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12-21.
10. E. Krissinel, K. Henrick, *Acta. Crystallogr. D. Biol. Crystallogr.* **2004**, *60*, 2256-2268.
11. The PyMOL molecular graphics system, version 1.6. Schrodinger, LLC.

c h a p t e r t h r e e

USING MUTABILITY LANDSCAPES OF A PROMISCUOUS TAUTOMERASE TO GUIDE THE ENGINEERING OF ENANTIOSELECTIVE MICHAELASES

J.-Y. van der Meer
H. Poddar
B.-J. Baas
Y. Miao
M. Rahimi
A. Kunzendorf
R. van Merkerk
P. G. Tepper
E. M. Geertsema
A.-M. W. H. Thunnissen
W. J. Quax
G. J. Poelarends

ABSTRACT

The Michael-type addition reaction is widely used in organic synthesis for carbon-carbon bond formation. However, biocatalytic methodologies for this type of reaction are scarce, which is related to the fact that enzymes naturally catalyzing carbon-carbon bond-forming Michael-type additions are rare. A promising template to develop new biocatalysts for carbon-carbon bond formation is the enzyme 4-oxalocrotonate tautomerase, which exhibits promiscuous Michael-type addition activity. Here we present mutability landscapes for the expression, tautomerase and Michael-type addition activities, and enantioselectivity of 4-oxalocrotonate tautomerase. These maps of neutral, beneficial, and detrimental amino acids for each residue position and enzyme property provide detailed insight into sequence-function relationships. This offers exciting opportunities for enzyme engineering, which is illustrated by the redesign of 4-oxalocrotonate tautomerase into two enantiocomplementary 'Michaelases'. These 'Michaelases' catalyze the asymmetric addition of acetaldehyde to various nitroolefins, providing access to both enantiomers of γ -nitroaldehydes, which are important precursors for pharmaceutically active γ -aminobutyric acid derivatives.

Chiral γ -nitroaldehydes are valuable precursors for pharmaceutically active γ -aminobutyric acid (GABA) derivatives. For example, marketed pharmaceuticals Phenibut (anxiolytic), Pregabalin (anticonvulsant), Baclofen (muscle relaxant), and Rolipram (antidepressant) can be readily obtained from distinct chiral γ -nitroaldehydes by two simple chemical steps.^{1,2} Preparation of the appropriate γ -nitroaldehyde precursors for these GABA derivatives would require the Michael-type addition of acetaldehyde to diverse nitroalkene acceptors. The Michael-type addition of unmodified aldehydes to nitroalkenes has recently become viable by the development of proline- and peptide-based organocatalysts.³⁻⁵ However, examples including acetaldehyde as donor substrate are scarce and 10-20 mol% of organocatalyst in organic solvent is typically applied.⁴⁻⁹ Therefore, there is great interest in the development of alternative biocatalytic procedures for the enantioselective synthesis of γ -nitroaldehydes. Unfortunately, enzymes that naturally catalyze carbon-carbon bond-forming Michael-type additions are extremely rare.¹⁰ A few elegant studies on promiscuous enzyme-catalyzed carbon-carbon bond-forming Michael-type additions have been reported, but most of these reactions proceed in organic solvent with low or moderate stereocontrol, and do not involve acetaldehyde as donor substrate.¹¹⁻¹⁸

The enzyme 4-oxalocrotonate tautomerase (4-OT) is composed of six identical subunits of only 62 amino acid residues each.^{19,20} It belongs to the tautomerase superfamily, a group of homologous proteins that share a unique catalytic amino-terminal proline (Pro-1) and a characteristic β - α - β structural fold.²¹⁻²³ 4-OT takes part in a catabolic pathway for aromatic hydrocarbons in *Pseudomonas putida* mt-2 (**Fig. 1a**), where it catalyzes the tautomerization of 2-hydroxymuconate (**1**) to yield 2-oxohex-3-enedioate (**2**).²⁴ Residue Pro-1 functions as a general base ($pK_a \approx 6.4$) that transfers the 2-hydroxyl proton of **1** to the C5-position to give **2**.²⁵ Two other key catalytic residues are Arg-11 and Arg-39. Arg-39 is proposed to interact with the 2-hydroxyl group of **1** and a C-1 carboxylate oxygen, whereas Arg-11 is proposed to interact with the C-6 carboxylate group in a bidentate fashion.²⁶ The latter interaction may draw electron density toward C-5 to facilitate protonation by Pro-1.^{21, 26, 27}

Inspired by the versatile success of proline and its derivatives as organocatalysts, we anticipated and found the proline-based enzyme 4-OT to promiscuously catalyze carbon-carbon bond-forming reactions such as aldolizations and Michael-type additions.²⁸⁻³¹ These include Michael-type additions of linear aldehydes, such as acetaldehyde (**3**) or butanal (**4**) (**Fig. 1b**), to a variety of nitroalkenes (**5a-5g**) to yield enantioenriched γ -nitroaldehydes (**6a-6g** and **7**), which are important precursors for GABA analogues such as the above-mentioned pharmaceuticals.³¹ The proposed catalytic mechanism involves formation of an enamine intermediate of the aldehyde donor with the active site Pro-1 residue of 4-OT, reminiscent of the mechanism of proline-based organocatalysts.^{28,32}

The discovery of the promiscuous Michael-type addition activity of 4-OT is an important step on the way to develop new biocatalysts for synthetically useful Michael-type reactions. Accordingly, it would be very attractive to enhance the Michael-type addition activity of 4-OT and improve its enantioselectivity. Because of the strong advantage of working with a protein with a small monomer size, we describe herein the construction of entire

Figure 1. Natural and promiscuous reactions catalyzed by 4-OT. **a**, 4-Oxalocrotonate tautomerase (4-OT)-catalyzed tautomerization of 2-hydroxymuconate (**1**) to yield 2-oxohex-3-enedioate (**2**) as part of a degradative pathway for aromatic compounds. **b**, 4-OT-catalyzed Michael-type additions of acetaldehyde (**3**) or butanal (**4**) to nitroalkenes **5a–5g** to give γ -nitroaldehydes **6a–6g** or **7**. **c**, 4-OT-catalyzed tautomerization of phenylenolpyruvate (**8**) to yield phenylpyruvate (**9**).

mutability landscapes for the expression, tautomerase and Michael-type addition activities, and enantioselectivity of 4-OT. These landscapes may be assessed and reasoned about as a whole, providing detailed insight into sequence-function relationships. Simultaneously, intelligent mutability-landscape navigation, coupled to combinatorial mutagenesis, offers exciting opportunities for rapid enzyme engineering.^{33, 34} This is illustrated here by the generation of 4-OT variants with improved activity and enantioselectivity, as well as inverted enantioselectivity, in Michael-type additions. The usefulness of these designer ‘Michaelases’ is demonstrated by their application in the asymmetric addition of acetaldehyde to various nitroalkenes, providing convenient access to both enantiomers of chiral γ -nitroaldehydes, which are valuable precursors for GABA derivatives, several of which represent marketed pharmaceuticals.

RESULTS

Collection of single 4-OT mutants

To chart the mutability landscapes of 4-OT, a collection of 4-OT genes encoding nearly all possible single variants of 4-OT was constructed. This unique collection covered at least 15 of the 19 possible variants at each residue position, from Ile-2 to the carboxy-terminal Arg-62. Single mutants of the amino-terminal proline residue (Pro-1) were not included in the collection, because Pro-1 is a key catalytic residue and mutations at this position lead to incorrect demethionylation of the protein.^{35,36} The 4-OT genes did not contain any affinity tag, and were individually cloned, stored, and expressed in *E. coli* BL21(DE3). Importantly, because all 4-OT variants are stored and analyzed separately, no oversampling is required.

Mutability landscape of 4-OT for protein expression

A mutability landscape for protein expression was generated by determining the effect of each mutation on the production of soluble 4-OT protein. From the alignment of this landscape with the secondary structure elements of 4-OT (**Fig. 2**), it becomes apparent that many mutations in the α_1 -helix result in decreased production of soluble protein. Most notably, the introduction of a proline residue at any given position in the α_1 -helix has a pronounced effect on soluble 4-OT production. It seems likely that the introduction of a proline disrupts this α_1 -helix, resulting in an insoluble 4-OT mutant. The β -hairpin element (Gly-51 to Leu-56) is thought to play a crucial role in the stabilization of the hexameric structure of 4-OT.^{20,21} Although most mutations in this region do not seem to affect the amount of soluble protein produced, replacing the negatively charged Glu-55 with a positively charged residue (Lys or Arg) decreases the amount of soluble 4-OT protein produced to a level that is below the detection limit. The average effect of each amino-acid substitution on the amount of soluble protein produced reveals that substitutions to Trp, Arg and Pro are the most unfavorable ones (**Fig. 2b**). A likely explanation for this is that these residues have the biggest impact on the protein structure because of their volumetrically large side chain, charge or conformational rigidity, respectively. The finding that these amino-acid substitutions are highly unfavorable for soluble protein production is consistent with the results of a recently published systematic mutagenesis study on a protein-binding domain.³⁷ On the other hand, some mutations seem to increase the amount of soluble protein produced. These mutations can be found mainly at positions His-6, Ile-7 and Lys-47. Gratifyingly, the majority of the mutants was produced in sufficient amounts to enable activity assays.

Mutability landscape of 4-OT for tautomerase activity

A mutability landscape for the tautomerase activity of 4-OT was generated by determining the effect of each mutation on the ability of the enzyme to ketonize phenylenolpyruvate (**8**) to phenylpyruvate (**9**) (**Fig. 1c**). Screening assays with the native substrate of 4-OT,

Figure 2. Mutability landscape of 4-OT for protein expression. **a**, The horizontal axis of the data matrix depicts the wild-type sequence of 4-oxalocrotonate tautomerase (4-OT) and the vertical axis depicts the 20 possible amino acid residues. The wild-type amino acid residue at each position is indicated by bold squares, and white boxes represent mutants which are not present in the collection. The color indicates the concentration of soluble 4-OT variants in cell-free extract (CFE) in mg ml⁻¹, which was determined by quantitative densitometry on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-page) gels (**Supplementary Figure 1**; details on this procedure can be found in the Methods section). The detection limit of this method is 0.5 mg ml⁻¹. The depicted data is an average of two separate experiments ($n=2$). **b**, The average effect of each amino acid substitution across the entire protein on the expression of soluble protein. **c**, The secondary-structure elements of 4-OT.²⁰

2-hydroxymuconate (**1**), were not possible due to the high activity ($k_{\text{cat}} \sim 3500 \text{ s}^{-1}$) of the enzyme with this substrate.³⁸ Hence, activity profiling required very low enzyme concentrations, likely leading to hexamer dissociation, yielding unreproducible results. Therefore, the alternative substrate **8** ($k_{\text{cat}} = 73 \text{ s}^{-1}$) was used to monitor the tautomerase activity of 4-OT, which enabled a practical screening assay with reproducible results.³⁹

The mutability landscape for 4-OT's tautomerase activity demonstrates the mutational robustness of the enzyme for this activity (**Fig. 3a**). This is apparent from mutations in various regions of the enzyme that show no or little effect on activity, including the regions Ser-12 to Arg-29 and Gly-54 to Arg-62. It has been hypothesized that such neutral mutations can have an important role in natural enzyme evolution because they may result in 'neutral drift'.⁴⁰⁻⁴² Whitman and co-workers demonstrated that the two active-site arginines (Arg-11 and Arg-39) of 4-OT are required for the tautomerization of dicarboxylate substrates, because they interact with the carboxylate moieties of the substrate.^{26,27} Fully consistent with this finding, we observed that only one active-site arginine (Arg-11) is essential for the tautomerization of the monocarboxylate substrate **8**. In addition to Arg-11, also Gly-10 appears to be important for the tautomerase activity. As mutations at Gly-10 have no

significant effect on enzyme expression (**Fig. 2**), it is unlikely that this residue is required for correct overall folding of the enzyme. It is therefore tempting to speculate that this residue is required for securing the correct architecture of the active site.

Interestingly, five positions at which single mutations result in significantly improved tautomerase activity (>5-fold) were identified. These include mutations at positions Ile-2, Gln-4, Leu-8, Ser-37 and Phe-50. Whereas Ile-2, Leu-8, Ser-37 and Phe-50 are lining the Pro-1 pocket, the distance between Gln-4 and Pro-1 is ~11 Å, illustrating that both close and distant mutations can improve activity.⁴³ These findings suggest that 4-OT could be readily engineered to become a more efficient catalyst for the tautomerization of **8**.

Mutability landscape of 4-OT for the addition of **3** to **5a**

To generate a mutability landscape for the promiscuous 'Michaelase' activity of 4-OT, the effect of each mutation on the ability of the enzyme to catalyze the addition of acetaldehyde (**3**) to *trans*- β -nitrostyrene (**5a**) was determined. This activity was monitored (using UV spectroscopy) by following the depletion of substrate **5a** in the presence of **3** and enzyme. For each mutant, control experiments confirmed that the depletion of **5a** was dependent on the presence of both **3** and enzyme. This indicates that the conversion of **5a** was indeed the result of an enzyme-catalyzed Michael-type addition of **3** to **5a** (rather than the enzyme-catalyzed addition of water to **5a**).

The mutability landscape shows that the majority of mutations does not have a significant effect on this 'Michaelase' activity of 4-OT (**Fig. 3b**). This demonstrates the mutational robustness of the enzyme for this promiscuous activity. The key functionally important residue that can be identified from the mutability landscape is Arg-39, which is in accordance with the proposed mechanism of this Michael-type addition reaction, in which Arg-39 serves as a catalytic acid.²⁹ The observation that the activity is partially restored when Arg-39 is replaced by a positively charged residue that could function also as an acid (Lys or His) further strengthens this proposed mechanistic role of Arg-39. Interestingly, the other active-site arginine (Arg-11) appears to be less important for this promiscuous 'Michaelase' activity (**Fig. 3b**), whereas it was shown to be essential for the phenylenolpyruvate tautomerase activity of 4-OT (**Fig. 3a**). Residues around and in the β -hairpin element (Gly-51 to Leu-56) also appear to be important for this Michael-type addition activity of 4-OT. Especially mutations at the positions of Phe-50 and the two glycine residues in the middle of the β -hairpin seem to have a large functional effect.

Most notably, four mutations at position Ala-33 result in an improved 'Michaelase' activity of 4-OT. When this residue was replaced by negatively charged residues (Asp or Glu), or the corresponding polar residues (Asn or Gln), the activity was significantly enhanced. Of these mutants, variant A33D had a 3.5-fold increase in specific activity, making it the most active single 4-OT mutant for this Michael-type addition reaction. In contrast, the phenylenolpyruvate tautomerase activity of variant A33D was reduced ~8-fold when compared to that of wild-type 4-OT.

Figure 3. Mutability landscapes of 4-OT for tautomerase and promiscuous ‘Michaelase’ activities. The horizontal axes of the data matrices depict the wild-type sequence of 4-oxalocrotonate tautomerase (4-OT) and the vertical axes depict the 20 possible amino acid residues. The wild-type amino acid residue at each position is indicated by bold squares. Grey boxes represent mutants which were not produced above the detection limit (0.5 mg ml⁻¹ in the cell-free extract [CFE]) and white boxes represent mutants which were not present in the collection. The color indicates the specific activity of each mutant relative to that of wild-type 4-OT ($[U_{mut} \cdot \text{mg}^{-1}] / [U_{wt} \cdot \text{mg}^{-1}]$). All activities were determined using UV-spectroscopy (details can be found in the Methods section), and enzyme concentrations

- in CFE were individually determined for each 4-OT variant using quantitative densitometric analysis of sodium dodecyl sulfate (SDS) gels. Each screening plate contained samples of wild-type 4-OT of which the average specific activity was used to calculate the relative specific activities of all 4-OT variants on that same plate. **a**, The mutability landscape of 4-OT for tautomerase activity using substrate **8** ($n=4$). **b**, The mutability landscape of 4-OT for the Michael-type addition of **3** to **5a** ($n=2$). **c**, The mutability landscape of 4-OT for the Michael-type addition of **4** to **5a** ($n=2$). **d**, The secondary-structure elements of 4-OT.²⁰

Figure 4. Enzyme-catalyzed transformations. **a**, Progress curves of the Michael-type addition of acetaldehyde **3** (50 mM) to *trans*-nitrostyrene **5a** (2 mM, 18 mg) catalyzed (0.7 mol%) by wild-type 4-oxalocrotonate tautomerase (4-OT) or 4-OT mutants A33D or M45Y/F50A. The reactions were carried out in buffer [20 mM NaH₂PO₄ / 10% EtOH (v/v)] at pH 7.3. **b**, Progress curves of the Michael-type addition of butanal **4** (50 mM) to *trans*-nitrostyrene **5a** (2 mM, 18 mg) catalyzed (0.7 mol%) by wild-type 4-OT or 4-OT mutants R39E, H6M/A33E/F50V or M45Y/F50A. The reactions were carried out in buffer [20 mM NaH₂PO₄ / 10% EtOH (v/v)] at pH 5.5. These progress curves are derived from the preparative scale reactions listed in **Table 1**.

The preparative usefulness of the A33D mutant was compared to that of wild-type 4-OT by transformations using 0.7 mol% biocatalyst and a 25-fold excess of **3** (50 mM) over **5a** (2 mM). Analysis of the progress curves of these reactions confirmed that mutant A33D is more efficient in the addition of **3** to **5a** than wild-type enzyme (**Fig. 4a**), with the A33D-catalyzed reaction being completed within just 90 min at room temperature. The products were purified and characterized using ¹H NMR spectroscopy, chiral phase HPLC, and optical rotation measurements (**Supplementary Figures 2 and 3** and **Supplementary Table 1**). This analysis revealed that in addition to enhanced activity, mutant A33D also has improved enantioselectivity compared to wild-type 4-OT (**Table 1**, compare entries **1** and **2**), producing the 3*S* enantiomer of 4-nitro-3-phenylbutanal (**6a**) with an enantiomeric ratio (e.r.) of 99:1. The isolated product yield could be improved by conducting the enzymatic reaction at pH 6.5 instead of pH 7.3 (**Table 1**, entries **4** and **5** and **Supplementary Figures 4-6**), lowering the amount of byproducts formed, which result from the inherent

Table 1. Preparative scale Michael-type addition reactions of aldehydes (**3** or **4**) to *trans*-nitrostyrene (**5a**) catalyzed by wild-type 4-OT or 4-OT mutants (0.7 mol%)

Entry	Substrates	Catalyst	Solvent pH	Reaction time (h)	Yield (%)	d.r. ^[a] (<i>syn:anti</i>)	e.r. ^[b]	Abs. ^[c] conf.	Product
1	3 and 5a	WT 4-OT	7.3	6	36 ^[d]	-	93 : 7	3S	6a
2	3 and 5a	A33D	7.3	1.5	28 ^[d]	-	99 : 1	3S	6a
3	3 and 5a	M45Y/F50A	7.3	3	30 ^[d]	-	94 : 6	3R	6a
4	3 and 5a	WT 4-OT	6.5	1.8 ^[e]	94	-	95 : 5	3S	6a
5	3 and 5a	A33D	6.5	0.7 ^[e]	94	-	99 : 1	3S	6a
6	3 and 5a	M45Y/F50A	5.5	21	65	-	96 : 4	3R	6a
7	4 and 5a	WT 4-OT	5.5	23	67	91 : 9	57 : 43	2R3S	7
8	4 and 5a	H6M/A33E/F50V	5.5	1	76	97 : 3	77 : 23	2S3R	7
9	4 and 5a	R39E	5.5	23	67	93 : 7	95 : 5	2R3S	7
10	4 and 5a	M45Y/F50A	5.5	2.5	86	96 : 4	96 : 4	2S3R	7

^[a]The d.r. values were determined by ¹H NMR spectroscopy (**Supplementary Figure 7**). ^[b]The e.r. values were determined by HPLC on a chiral stationary phase (**Supplementary Figures 3, 6 and 8**). ^[c]The absolute configuration of the major enantiomers was determined by comparison of chiral phase HPLC data and optical rotation data with literature data (**Supplementary Table 1**). ^[d]At pH 7.3, the formation of undefinable side products, resulting from the inherent tendency of acetaldehyde to form oligomers,⁴⁴ requires product purification by flash column chromatography, which lowers the isolated yield of the desired product **6a**. ^[e]1.4 mol% catalyst loading was used. The progress curves for these preparative scale reactions are given in **Fig. 4** and **Supplementary Figure 4**.

tendency of acetaldehyde to rapidly react with itself, forming oligomers.⁴⁴ This makes the A33D mutant a promising biocatalyst for application in the asymmetric synthesis of enantiopure **3S-6a**.

Synthetic usefulness of the A33D mutant enzyme

To further demonstrate the synthetic usefulness of the A33D mutant enzyme, we used this biocatalyst in the Michael-type addition of acetaldehyde (**3**) to a series of aromatic and aliphatic nitroolefin acceptors (**5b-5g**, **Table 2**). Nitroalkene (2-5 mM), acetaldehyde (50-150 mM), and A33D enzyme (0.5-5.3 mol%) were incubated in an appropriate solvent system, and reactions were followed by UV spectroscopy (**Supplementary Figures 9-14**). After complete conversion of the respective nitroalkene, standard workup and purification procedures were carried out, which yielded γ -nitroaldehydes **6b-6g** as confirmed by ¹H NMR spectroscopy (**Supplementary Figures 15-20**). Analysis of products **6b-6g** by HPLC or GC on a chiral stationary phase (**Supplementary Figures 21-26**) revealed that the A33D-catalyzed Michael-type addition reactions are highly enantioselective, producing these γ -nitroaldehydes with excellent e.r. values between 95:5 and >99:1 (**Table 2**). These results clearly demonstrate the potential of the A33D mutant enzyme for application in enantioselective synthesis of various γ -nitroaldehydes. The discovery of the A33D mutant illustrates the opportunities for protein improvement afforded by experimental protein mutability landscapes; this single mutant would likely not have been generated by using an error-prone PCR approach because the codon for Ala-33 in 4-OT (GCG) requires two point mutations to produce an aspartate (GAT or GAC).

Mutability landscapes of 4-OT for the addition of **4** to **5a**

To gain further insight into which residues govern the activity and enantioselectivity of 4-OT in Michael-type addition reactions, we determined the effect of each mutation on the ability of 4-OT to catalyze the addition of butanal (**4**) to *trans*- β -nitrostyrene (**5a**), and to yield enantioenriched product **7**. The addition of **4** to **5a** was chosen as a model reaction for determining the mutational effect on stereoselectivity because the diastereo- and enantiopurity of product **7** can be determined directly by HPLC on a chiral stationary phase (**Supplementary Figure 27**, **Supplementary Table 2**), whereas the separation of the enantiomers of product **6a** (from the enzymatic addition of **3** to **5a**) by chiral phase HPLC requires derivatization.^{29,30} To prevent epimerization of product **7** and minimize the formation of side-products, reactions were performed at pH 5.5 instead of pH 7.3.³⁰ Under these conditions, however, slight protein precipitation was observed with some mutants, precluding accurate activity measurements.

Nevertheless, the mutability landscape clearly demonstrates the mutational robustness of the C-terminal domain (Ala-57 to Arg-62) of 4-OT for the Michael-type addition of **4** to **5a** (**Fig. 3c**). This robustness of the C-terminal domain has also been observed for the other two activities analyzed (**Fig. 3a,b**), which may suggest a minor contribution of

Table 2. 4-OT(A33D)-catalyzed acetaldehyde addition to nitroalkenes **5a-5g**

acetaldehyde + nitroalkene $\xrightarrow{4\text{-OT(A33D)}}$ γ -nitroaldehyde

Entry	Nitro-alkene	R-group	Product (γ -nitroaldehyde)	A33D (mol%) ^a	t (min)	Yield (%)	e.r. ^b	Abs. Conf.
1	5a	Ph		6a 1.4 ^g	40	94	99:1	S ^c
2	5b	<i>p</i> -Cl-C ₆ H ₄		6b 2.8	70	81	98:2	S ^c
3	5c	<i>p</i> -F-C ₆ H ₄		6c 1.5	60	63	99:1	S ^c
4	5d	<i>p</i> -HO-C ₆ H ₄		6d 0.5 ^g	150	44	95:5	S ^d
5	5e	3-H O - 4 - MeO-C ₆ H ₃		6e 1.8 ^g	40	94	99:1	S ^c
6	5f	3-c-PentO-4-MeO-C ₆ H ₃		6f 3.7	50	65	>99:1	S ^c
7	5g	isobutyl		6g 5.3	50	60	>99:1	R ^{e,f}

^aCompared to nitroalkene. The amounts of applied A33D enzyme (0.5 – 5.3 mol%) were adjusted such that conversions of **5a-5g** were all completed within 2.5 h. For conversion of **5g** into **6g**, 5.3 mol% of enzyme was required to outcompete non-enzymatic water addition to **5g**.³¹ ^bDetermined by HPLC or GC analysis on a chiral stationary phase. ^cThe absolute configuration was determined by chiral phase HPLC and comparison with literature data. ^dThe absolute configuration of the major enantiomer could be tentatively assigned based on analogy and the sign of optical rotation. ^eThe absolute configuration of the major enantiomer was determined by chiral phase GC and comparison with literature data. ^fThe deviant configuration of **6g** is due to different prioritization of the substituents at the chiral center relative to **6a-f**. ^gThe reaction was performed at pH 6.5 instead of pH 5.5 (full details on the reaction conditions are given in the Methods section).

this domain to catalysis. The overall mutational robustness of 4-OT for the Michael-type addition of **4** to **5a**, however, seems less when compared to the Michael-type addition activity of **3** to **5a**. This may be related to the difference in the pH of the assay buffer (pH 5.5 versus pH 7.3).

Importantly, the mutability landscape (**Fig. 3c**) revealed four residue positions at which mutations significantly improved the specific activity of 4-OT for the Michael-type addition of **4** to **5a**. The higher number of favorable mutations found for this reaction, when compared to the Michael-type addition of **3** to **5a**, could be the result of the additional ethyl group on the aldehyde substrate which could serve as an extra handle for the enzyme to coordinate substrate binding. The first position where beneficial single mutations (Val, Ile, Leu or Met) were found is His-6. For this position, the best mutant enzyme was H6M, which has an ~3-fold increased specific activity compared to that of wild-type 4-OT. The second position is Ala-33, which upon mutation to a Glu improved the specific activity ~4-fold compared to that of wild-type 4-OT (a Gln at this position improved the activity slightly). Notably, the A33E mutation also significantly improved the activity of 4-OT for the Michael-type addition of **3** to **5a** (**Fig. 3b**). The third position is Met-45; at this position four single mutations improved the activity, including mutations to Ile (2.1-fold), Tyr (2.3-fold), His (3.6-fold), and Thr (2-fold). Finally, at position Phe-50 mutations to Leu and Val improved the specific activity 2.5- and ~5-fold, respectively, compared to that of wild-type 4-OT. This makes the F50V enzyme the most active single 4-OT mutant for the Michael-type addition of **4** to **5a** (**Supplementary Figure 28, Supplementary Table 3**). These identified hotspots are good targets to further enhance 4-OT's promiscuous 'Michaelase' activity, as shown by combinatorial mutagenesis, which yielded a triple mutant, H6M/A33E/F50V, with strongly enhanced activity for the Michael-type addition of **4** to **5a** (**Fig. 4b, Table 1**, see Supplementary Discussion for full details).

No significant changes in the diastereoselectivity were observed as all active mutant enzymes, like wild-type 4-OT, produced the *syn* diastereomer of **7** in large excess (d.r. $\geq$ 72:28). This observation is consistent with the postulated topological rule that explains the *syn*-selectivity in diastereoselective Michael-type additions of enamines to nitroalkenes.⁴⁵ Wild-type 4-OT has enantiopreference towards the 2R3S enantiomer of the *syn* diastereoisomer, producing **7** with an e.r. of 57:43 (2R3S:2S3R) (**Table 1, entry 7**). The effect of each mutation on the ability of 4-OT to produce enantioenriched product **7** is shown in **Fig. 5**. The mutability landscape shows that many single mutations have a significant effect on 4-OT's enantioselectivity. The residue positions where single mutations led to the most pronounced improvement in enantioselectivity are Ala-33, Arg-39, Ala-57 and Arg-61. A comparison of the mutability landscapes suggests that in contrast to the mutational robustness of the C-terminal domain for the natural and promiscuous activities of 4-OT, this domain appears to contribute largely to the enantioselectivity of 4-OT in the Michael-type addition of **4** to **5a**.

The single mutants with the best enantioselectivity at each of these residue positions (A33D, R39E, A57Y and R61M) were purified and characterized for their ability to produce

enantioenriched products in the Michael-type addition of **4** to **5a** as well as **3** to **5a** (**Supplementary Table 4**). Fully consistent with the mutability landscape shown in **Fig. 5**, the enzyme R39E has the highest enantioselectivity for the Michael-type addition of **4** to **5a**, producing γ -nitroaldehyde **7** with an e.r. of 94:6 (2R3S:2S3R). Although mutant R39E also displayed significantly improved enantioselectivity in the Michael-type addition of **3** to **5a**, producing 3S-**6a** with an e.r. of 96:4 (**Supplementary Table 4**), mutant A33D has the highest enantioselectivity in this reaction (vide supra).

The preparative usefulness of the R39E mutant was compared to that of wild-type 4-OT by transformations using 0.7 mol% biocatalyst and a 25-fold excess of **4** (50 mM) over **5a** (2 mM). Analysis of the progress curves of these reactions showed that R39E exhibits similar activity compared to WT 4-OT (**Fig. 4b**). However, in contrast to wild-type 4-OT, R39E afforded highly enantioenriched product **7** with an e.r. of 95:5 (**Table 1, compare entries 7 and 9**). This makes the R39E mutant a promising biocatalyst for application in the asymmetric synthesis of enantioenriched 2R3S-**7**.

Interestingly, the mutability landscape (**Fig. 5**) further shows that several single variants of 4-OT have inverted enantioselectivity, producing the 2S3R enantiomer of **7** in excess over the 2R3S enantiomer. Mutations at positions His-6, Arg-11, Met-45, Phe-50 and

Figure 5. Mutability landscape of 4-OT for enantioselectivity in the Michael-type addition of **4** to **5a**. **a**, The horizontal axis of the data matrix depicts the wild-type sequence of 4-oxalocrotonate tautomerase (4-OT) and the vertical axis depicts the 20 possible amino acid residues. The wild-type amino acid residue at each position is indicated by bold squares, and white boxes represent mutants which are not present in the collection. The color indicates the enantiomeric ratio (2R3S:2S3R) of the major syn diastereoisomer of **7** as produced by the 4-OT variants. The e.r. values were determined by high-performance liquid chromatography (HPLC) on a chiral stationary phase. Grey boxes represent mutants which did not produce sufficient amounts of **7** to give an adequate UV-signal during HPLC analysis. As expected, these grey boxes mainly coincide with the grey and dark blue boxes in **Fig. 3c** which indicate mutants that either did not express or lack significant ‘Michaelase’ activity. **b**, The secondary-structure elements of 4-OT.²⁰

Gly-54 had the most pronounced effect on inversion of enantioselectivity. Three of these positions (His-6, Met-45 and Phe-50) overlap with the positions at which single mutations were identified which improve 4-OT's 'Michaelase' activity for addition of **4** to **5a** (Fig. 3c). The best mutant at each of these positions (H6I, R11I, M45H, F50A and G54E) was purified and characterized for its ability to produce enantioenriched products in the Michael-type addition of both **3** to **5a** and **4** to **5a**. At position Met-45, the second-best mutant was chosen for analysis because the best mutant at this position (M45Y) could not be purified to homogeneity due to insufficient expression (consistent with the mutability landscape shown in Fig. 2a). Strikingly, all five mutants did not only have inverted enantioselectivity for the Michael-type addition of **4** to **5a** but also for the Michael-type addition of **3** to **5a** (Supplementary Table 5). Of these mutants, the enzyme F50A has the highest enantioselectivity producing both 3R-**6a** and 2S3R-**7** with an e.r. of 93:7, demonstrating its suitability for the preparation of the opposite enantiomers (as compared to those produced in the same reactions catalyzed by A33D, R39E or wild-type 4-OT) of γ -nitroaldehydes **6a** and **7**.

Engineering 4-OT variants with inverted enantioselectivity

To further improve the selectivity of 4-OT towards the 2S3R enantiomer of **7**, we generated double and triple mutants based on all possible combinations of the best single 4-OT mutants (i.e. H6I, M45H or M45Y, and F50A). Using analytical scale assays, these double and triple mutant enzymes were evaluated for their enantioselectivity in the Michael-type addition of **4** to **5a** yielding **7**. Three (H6I/M45H, H6I/M45Y and H6I/M45H/F50A) of the seven constructed mutant enzymes displayed either low or no detectable 'Michaelase' activity. Gratifyingly, the other four mutants (H6I/F50A, M45Y/F50A, M45H/F50A, and H6I/M45Y/F50A) displayed respectable activity and produced **7** with an e.r. >96:4 (2S3R:2R3S) (Supplementary Table 6). This product enantiopurity is higher than that obtained with the best single mutant (F50A), indicating an additive effect of the mutations.

The preparative usefulness of the mutant with the best activity and enantioselectivity, M45Y/F50A, was analyzed by transformations using 0.7 mol% biocatalyst and a 25-fold excess of aldehyde (**3** or **4**) over **5a**. The progress curves of these reactions demonstrate that this mutant, in addition to inverted enantioselectivity, also has significantly enhanced 'Michaelase' activity for the addition of **3** to **5a** (Fig. 4a) as well as **4** to **5a** (Fig. 4b) when compared to wild-type 4-OT. The M45Y/F50A-catalyzed Michael-type addition of **4** to **5a** gave 2S3R-**7** in excellent yield (86%) and with high enantiopurity (96:4) (Table 1, entry 10). Interestingly, the M45Y/F50A mutant also displayed high enantioselectivity in the Michael-type addition of **3** to **5a**, yielding 3R-**6a** with an e.r. of 94:6 and a yield of 30% (Table 1, entry 3). The product yield could be improved to 65% by conducting the enzymatic reaction at pH 5.5 (Table 1, entry 6), which lowers the amount of side-products formed.

Synthetic usefulness of the M45Y/F50A mutant enzyme

To further demonstrate the usefulness of the M45Y/F50A mutant for synthesis of the opposite enantiomers of γ -nitroaldehydes **6b-6g**, as compared to those afforded by mutant A33D (**Table 2**), we also used this engineered enzyme in the Michael-type addition of acetaldehyde (**3**) to the nitroolefin acceptors **5b-5g** (**Table 3**). After complete conversion of the respective nitroalkene (**Supplementary Figures 9-14**), standard workup and purification procedures were carried out, affording γ -nitroaldehydes **6b-6g** (**Supplementary Figures 15-20**). Analysis of these enzymatic products by chiral phase HPLC or GC (**Supplementary Figures 21-26**) revealed that the M45Y/F50A-catalyzed Michael-type addition reactions indeed yield the opposite enantiomers of γ -nitroaldehydes **6b-6g** with e.r. values between 62:38 and 97:3 (**Table 3**). These results clearly demonstrate the potential of the M45Y/F50A enzyme for application in the synthesis of enantioenriched **3R-6b-6f** and **3S-6g**, which are valuable precursors for pharmaceutically active GABA analogues, and further illustrate the exciting opportunities for enzyme engineering afforded by experimental protein mutability landscapes for enantioselectivity.

Structural consequences of the M45Y and F50A mutations

To obtain insight into the structural consequences of the M45Y and F50A mutations, we determined crystal structures of mutant M45Y/F50A in an unliganded state and in complex with substrate *trans*- β -nitrostyrene (**5a**) at 2.7 Å and 2.3 Å resolution, respectively (**Supplementary Table 7 and Supplementary Figure 29**). The overall structure of the M45Y/F50A mutant did not reveal significant changes in backbone conformation compared to wild-type 4-OT. Hexameric C α backbones of mutant and wild-type 4-OT superimpose with an average root-mean-square deviation (rmsd) of 0.32 Å (**Supplementary Figure 30**). However, the M45Y and F50A mutations did have a pronounced effect on the active-site geometry near Pro-1 (**Fig. 6a, Supplementary Figure 31a,b**). In the wild-type 4-OT structure, the side chains of Met-45 and Phe-50 point towards Pro-1, restricting access to a small hydrophobic pocket at the back of the active site. In the structure of mutant M45Y/F50A, the side chain of Tyr-45 is rotated away from Pro-1 towards the back of the active site, where it forms a hydrogen bond with His-6 (**Supplementary Figure 31c**). Combined with the replacement of the bulky Phe-50 with an alanine, the M45Y mutation results in a significant opening and enlargement of the hydrophobic pocket near Pro-1 (**Fig. 6a, Supplementary Figure 31a,b**). This enlarged pocket in the M45Y/F50A mutant is able to accommodate the phenyl group of **5a**, as revealed in the crystal structure of mutant M45Y/F50A complexed with this substrate (**Fig. 6b**). The phenyl group of **5a** makes van der Waals contacts with residues Pro-1, His-6, Ile-7, Leu-8 and Ile-41 in the hydrophobic pocket, while the nitro group is in van der Waals contact with Ser-37 and forms a salt bridge with the guanidinium group of Arg-39. The central ethylene group of **5a** is only ~3 Å away from the amino group of Pro-1. The observed binding mode of **5a** in the structure of mutant

Table 3. 4-OT(M45Y/F50A)-catalyzed acetaldehyde addition to nitroalkenes **5a-5g**

acetaldehyde (3) + nitroalkene (5) $\xrightleftharpoons{4\text{-OT(M45Y/F50A)}}$ γ -nitroaldehyde (6)

Entry	Nitro-alkene	R-group	Product (γ -nitroaldehyde)	M45Y/F50A (mol%) ^a	t (min)	Yield (%)	e.r. ^b	Abs. Conf.
1	5a	Ph		6a 0.7	o.n. ^g	65	96:4	R ^c
2	5b	<i>p</i> -Cl-C ₆ H ₄		6b 2.8	90	60	62:38	R ^c
3	5c	<i>p</i> -F-C ₆ H ₄		6c 1.5	180	47	86:14	R ^c
4	5d	<i>p</i> -HO-C ₆ H ₄		6d 0.5	o.n. ^g	48	82:18	R ^d
5	5e	3-HO-4-MeO-C ₆ H ₃		6e 1.8	100	73	97:3	R ^c
6	5f	3-c-PentO-4-MeO-C ₆ H ₃		6f 3.7	90	70	68:32	R ^c
7	5g	isobutyl		6g 5.3	70	60	90:10	S ^{e,f}

^aCompared to nitroalkene. The amounts of applied M45Y/F50A enzyme are identical to those of applied A33D enzyme (Table 2), enabling a direct comparison of catalytic performance. All reactions were performed at pH 5.5 (full details on the reaction conditions are given in the Methods section). ^bDetermined by HPLC or GC analysis on a chiral stationary phase. ^cThe absolute configuration was determined by chiral phase HPLC and comparison with literature data. ^dThe absolute configuration of the major enantiomer could be tentatively assigned based on analogy and the sign of optical rotation. ^eThe absolute configuration of the major enantiomer was determined by chiral phase GC and comparison with literature data. ^fThe deviant configuration of **6g** is due to different prioritization of the substituents at the chiral center relative to **6a-f**. ^go.n., overnight (entry 1, ~21 h; entry 4, ~16 h).

M45Y/F50A is not possible in the wild-type 4-OT structure due to steric hindrance by the Phe-50 and Met-45 residues.

In conclusion, our results reveal that the two mutations (M45Y and F50A) unexpectedly created a new substrate-binding pocket. This active-site remodeling is most likely responsible for the observed differences in activity and enantioselectivity between mutant and wild-type 4-OT. However, these findings *per-se* do not explain the enantiopreference of mutant M45Y/F50A. The electron density for bound **5a** does not indicate a preference of this substrate to bind with an orientation relative to Pro-1 that is consistent with the formation of the (*R*)-product; the other orientation differing by a rotation of ~ 180 degrees around the longitudinal axis of the substrate molecule, and consistent with formation of the (*S*)-product, fits equally well (**Supplementary Figure 31d**). In addition to highlighting that active-site remodeling may lead to inverted enantioselectivity, our crystallographic results provide an important guide for future engineering experiments.

DISCUSSION

In summary, we provide evidence that a catalytically promiscuous tautomerase is a good starting point to develop novel enantiocomplementary biocatalysts for unnatural Michael-type additions of acetaldehyde (or butanal) to various nitroalkenes. These enzyme-catalyzed reactions provide ready access to both enantiomers of γ -nitroaldehydes, given that the product yields may be further improved by optimizing the reactions conditions and purification protocols. This new biocatalytic procedure provides an attractive complement to existing organocatalytic methodologies. Chiral γ -nitroaldehydes are valuable building

Figure 6. Structural characteristics of the M45Y/F50A mutant. **a**, Superposition of the residues lining the hydrophobic Pro-1 pocket in wild-type 4-oxalocrotonate tautomerase (orange) and the M45Y/F50A mutant (green). **b**, Surface and stick representation of the Pro-1 pocket in the M45Y/F50A mutant, depicting the bound *trans*- β -nitrostyrene (orange) as observed in the crystal structure. The grey mesh shows the $2F_o - F_c$ electron density contoured at 1.0σ . Residues from the neighboring chains are shown in different colors and labelled with apostrophes. The hydrogen bonding interactions (distances in Å) are shown as black dashed lines.

blocks for GABA derivatives and their rapid synthesis should facilitate the development of new pharmaceuticals.

The applied systematic mutagenesis strategy allows the important discrimination between beneficial mutations and those that are neutral or detrimental, providing detailed insight into sequence-function relationships. The results provide support for the notion that the generation of large-scale mutational data, revealing the functional consequences of a great number of protein variants, may have important implications for understanding and engineering proteins.⁴⁶⁻⁴⁸ Interestingly, the mutability landscapes of 4-OT demonstrate that in contrast to the mutational robustness of the C-terminal domain for 4-OT's different activities, this domain appears to contribute largely to the enantioselectivity of the enzyme in the Michael-type addition reaction. Hence, this domain may represent a good target for future mutagenesis experiments, with the aim to optimize the enantioselectivity of 4-OT in Michael-type additions using different combinations of aldehyde and nitroalkene. We also aim to exploit the large amount of mutational data generated within this study, which can be used as a unique training set, and machine learning algorithms to generate mutation prediction models for 4-OT as well as closely related tautomerase superfamily members.

METHODS

Mutant collection preparation

To create mutability landscapes of 4-oxalocrotonate tautomerase (4-OT), a defined collection of 1040 single mutant 4-OT genes was purchased from DNA2.0 (Menlo Park, CA). This collection covered at least 15 of the 19 possible variants on each amino acid residue position ranging from Ile-2 to the carboxy-terminal Arg-62. In total 90% of all possible single 4-OT mutants is present in this collection. Each mutant gene was individually cloned in a pJexpress 414 vector, sequenced and transformed into *E. coli* DH10B by DNA2.0. Each DH10B culture, which carried a mutated 4-OT gene in a pJexpress 414 vector, was separately grown in liquid lysogeny broth (LB) containing 100 µg/ml ampicillin (amp). All plasmids, each containing a single mutant 4-OT gene, were isolated from these cultures using the NucleoSpin® 96 Plasmid Core Kit (Bioké, Leiden, NL). An aliquot of each isolated plasmid was transformed individually into chemically competent *E. coli* BL21 (DE3). The transformants were selected at 37°C in liquid LB containing 100 µg/ml amp. Each *E. coli* BL21 (DE3) transformant harboring a pJexpress 414 vector with a unique 4-OT gene was stored at -80°C until further use.

CFE production and enzyme concentration assessment

The expression levels, activities and enantioselectivities of all members of the 4-OT mutant collection were determined using cell-free extracts (CFEs) of cultures each expressing a different 4-OT mutant. To inoculate these cultures, aliquots of the -80°C stock of the transformed *E. coli* BL21 (DE3) cells were placed into wells of a 2.2-ml 96-deep-well plate (Greiner Bio-one, 96 well Masterblock®). Each mutant was placed into

two wells as a duplicate. Each well contained 1.25 ml LB, supplemented with 100 µg/ml amp and 100 µM isopropyl-β-D-1-thiogalactopyranoside (IPTG) as inducer. The deep-well plates with inoculated LB medium were sealed with sterile, gas-permeable seals (Greiner Bio-one, BREATHseal™) and incubated overnight at 37°C with shaking at 250 rpm. After the incubation, the cultures were pelleted at 3,500 RPM for 30 min at 4°C. The pellets of the duplicates were pooled and lysed with 375 µl BugBuster™ (Novagen) which was supplemented with 25 U/ml benzonase nuclease. The cell lysates were incubated with vigorous shaking at room temperature for 20 min. The cell lysate was cleared by centrifugation at 4,000 RPM for 20 min at 4°C, after which the CFE was obtained as the supernatant. In order to assess the 4-OT concentration in each CFE, 2 µl of each CFE was mixed with 18 µl of protein sample buffer [containing: 9% (m/v) sodiumdodecylsulfate (SDS), 40% (v/v) glycerol, 20% (v/v) β-mercaptoethanol, 0.1% (m/v) bromophenol blue and 250 mM tris(hydroxymethyl)aminomethane (tris) pH 6.8]. Of the final mixture, which contained CFE and protein sample buffer, 10 µl was loaded on a 26-wells pre-cast 10% polyacrylamide gel (NuPAGE® Novex® 10% Bis-Tris). Four samples with standard amounts of 0.5, 1, 2.5 and 5 µg of homogenous wild-type 4-OT were loaded on the gel besides the CFE samples and served as calibration samples for the quantification of the 4-OT mutant protein in the CFE. The purification method of wild-type 4-OT is described elsewhere.²⁸ The concentration of the purified wild-type 4-OT was determined by the Waddell method.⁴⁹ After electrophoresis (at 150V for 1 h), the gels were stained using the Coomassie-based stain InstantBlue™ (Expedeon Ltd.). A digital image of the gel was recorded using the Chemi Genius² Bio Imaging System (Syngene, Cambridge, UK). A typical picture of an SDS-gel which was used in this densitometric concentration assessment is present in **Supplementary Figure 1**. The concentrations of the 4-OT mutants in the CFE were quantified based on the size and intensity of the 4-OT bands on the digital image of the gel, relative to the 4-OT bands used for calibration. The calibration samples were present on each gel and were only used to quantify 4-OT concentrations of the samples that were loaded on that same gel. The size and intensity of the protein bands were quantified by using the software GeneTools (version 4.02, Syngene, Cambridge, UK) and were used to make a calibration curve. The average R² value of the 122 calibration curves, used to determine all expression levels, was 0.97. This indicates the accuracy of this quantification method. This concentration assessment preceded each activity screening of each CFE sample. The CFE was stored at room temperature (<5 h) until further use. The mutability landscape of 4-OT for expression of soluble protein was generated using the data obtained from these concentration assessments. The mutability landscape was generated using Microsoft Excel 2010.

Mutability landscapes of 4-OT for activity

All enzymatic activities of the 4-OT mutants were determined by UV-spectroscopy. The UV-spectroscopic measurements were done in 96-wells micro titer plates (MTP) (UV-star

µclear, Greiner Bio-one). The volume of the reaction mixtures in all wells was 100 µl and the temperature in the UV-plate reader was set to 25°C for all analyses.

Mutability landscape of 4-OT for tautomerase activity, using phenylenolpyruvate (8)

The tautomerase activity of the 4-OT mutants was assessed by determining their ability to ketonize phenylenolpyruvate (**8**). The final reaction mixtures consisted of the following: CFE (0.2% v/v), phenylenolpyruvate [400 µM (two times K_m of wild-type 4-OT)³⁹], and EtOH (10% v/v) in 10 mM NaH₂PO₄ buffer (pH 7.3). First, the CFE and buffer were added to each well of the MTP, using a Packard Multiprobe® II HT EX 8 tip robotic liquid handling system. The assay was initiated by the addition of 10 µl a stock solution of phenylenolpyruvate (4.0 mM in EtOH) to each well of the MTP which already contained the appropriate CFE and buffer mixture, using the integrated reagent injector on the SPECTROstar Omega plate reader (BMG LABTECH, Isogen Life Science, de Meern, NL). After adding the substrate to one well, the plate was briefly shaken (2 s at 500 rpm) before that well was analyzed for 40 s at 283 nm with a 0.26 s data interval, following the depletion of **8**. After finishing the measurement in one well of the MTP, the substrate was added to the next well to initiate the assay in that well. These reaction conditions enabled accurate rate determination in the range of 5-fold decreased to 5-fold increased activity compared to wild-type 4-OT. This activity assay was done in quadruplicate.

Mutability landscape of 4-OT for the Michael-type addition of acetaldehyde (3) to trans-β-nitrostyrene (5a)

The reaction mixtures used to monitor the 4-OT-catalyzed Michael-type addition of **3** to **5a** consisted of the following: CFE (40% v/v), substrate **5a** [500 µM (two times K_m of wild-type 4-OT)²⁹], substrate **3** (50 mM), and EtOH (5% v/v) in 10 mM NaH₂PO₄ buffer (pH 7.3). First, the appropriate amounts of CFE and buffer were added to each well of the MTP, using a Packard Multiprobe® II HT EX 8 tip robotic liquid handling system. The assay was initiated by adding 10 µl of a stock solution of **3** [500 mM in 10 mM NaH₂PO₄ buffer (pH 7.3)] and 5 µl of a stock solution of **5a** (10 mM in EtOH) to each well of the screening plate, which already contained the appropriate CFE and buffer mixture. To ensure proper mixing of the reagents, the plate was shaken (30 s at 500 rpm) before the reaction rates were recorded by monitoring the depletion of **5a** at 320 nm (λ_{max} **5a** = 320 nm) for 40 min with a 60 s data interval. In order to observe possible precipitation or other atypical changes in the reaction mixture, the UV-spectrum (220 nm - 500 nm) of each well was recorded before and after the reaction. The activity assay was done in duplicate. Separate control experiments in which the depletion rate of **5a** was assessed in the presence of each 4-OT mutant but in the absence of aldehyde, indicated that none of the 4-OTs in the mutant collection could catalyze the conversion of **5a** at a significant rate in the absence of **3**.

*Mutability landscape of 4-OT for the Michael-type addition of butanal (**4**) to **5a***

The reaction mixture used to monitor the 4-OT-catalyzed Michael-type addition of **4** to **5a** consisted of the following: CFE (20% v/v), substrate **4** (32.5 mM), substrate **5a** (650 μ M), and EtOH (5% v/v) in 50 mM NaH₂PO₄ buffer (pH 5.5). First, the appropriate amounts of CFE and buffer were added to each well of the MTP, using a Packard Multiprobe® II HT EX 8 tip robotic liquid handling system. The assay was initiated by adding 32.5 μ l of a stock solution of **4** [100 mM in 50 mM NaH₂PO₄ (pH 5.5)] and 5 μ l of a stock solution of **5a** (13 mM in EtOH) to each well of the MTP, which already contained the appropriate CFE and buffer mixture. To minimize evaporation of solvent and aldehyde **4**, the MTPs were sealed with UV-transparent plate seals (VIEWseal™, Greiner Bio-one). To ensure proper mixing of the reagents, the plate was shaken (30 s at 500 rpm) before the reaction rates were recorded by monitoring the depletion of **5a** at 320 nm for 80 min with a 60 s data interval. In order to observe possible precipitation or other atypical changes in the reaction mixture, the UV-spectrum (220 nm - 500 nm) of each well was recorded before and after the reaction. The buffer pH of 5.5 was used to prevent epimerization of the product (**7**). The activity assay was done in duplicate.

Data analysis

The duplicates of the substrate depletion curves obtained in the activity assays were combined, and an average substrate depletion curve was plotted. The initial substrate depletion rates were determined from the slopes of the linear section of these curves using MARS data analysis software [version 2.40 (BMG LABTECH, Isogen Life Science, de Meern, NL)]. The background substrate depletion rates were determined separately for each assay by assessing the average initial rate of substrate depletion in reaction mixtures containing CFE of *E. coli* BL21 (DE3) cultures carrying an empty pJexpress 414 vector (EV). This background substrate depletion rate was subtracted from all other substrate depletion rates determined for 4-OT variants on that same MTP. When the resulting enzyme activity (U) was zero the sample was marked accordingly. The earlier established enzyme concentrations in the CFEs were used to calculate the specific activities (U/mg) of each mutant. When an enzyme concentration in the CFE was lower than 0.5 μ g/ μ l (i.e. the detection limit), the sample was marked accordingly. An average of the wild-type 4-OT specific activities was calculated for each MTP, and all the other specific activities of the mutants on the same MTP were divided by this value to yield the relative specific activities $[(U_{\text{mut}} \cdot \text{mg}^{-1}) / (U_{\text{wt}} \cdot \text{mg}^{-1})]$ of all mutants. The EV and wild-type 4-OT controls were exclusively used for calculations of relative activities of mutants on that same MTP. The data was graphically represented in the mutability landscapes for activity, which were generated using Microsoft Excel 2010.

Mutability landscape of 4-OT for enantioselectivity

Reverse phase HPLC method development for chiral separation of stereoisomers of 7

Racemic 2-ethyl-4-nitro-3-phenylbutanal (**7**) was prepared according to a literature procedure.³⁰ This racemic compound was used to test the ability of a Chiralpak® ID column (150 mm × 4.6 mm, Daicel) to separate the 2R3S-**7** and 2S3R-**7** enantiomers, with the aim of establishing a rapid reverse phase HPLC method which could be used to generate the mutability landscape of 4-OT for enantioselectivity. All four stereoisomers of **7** could be separated using this column with an isocratic mobile phase (MeCN:H₂O 31:69) at a flow rate of 1.5 ml/min (**Supplementary Figure 27**). The chromatogram was recorded using a diode array detector (210 or 220 nm). The absolute configuration of the *syn*-enantiomers was determined by injecting **7** that was enzymatically produced by wild-type 4-OT, which produces the 2R3S-**7** enantiomer in excess.³⁰ The in-house HPLC apparatus consisted of the following components: Shimadzu LC-10 AT pump, a Shimadzu SIL-20A auto injector and a Shimadzu SPD-M10A VP diode array detector. The HPLC chromatographic data were analyzed using data processing software (LC Solutions) obtained from Shimadzu.

Supplementary Figure 27a demonstrates the chiral separation of racemic **7** using the above mentioned HPLC conditions. As wild-type 4-OT has been reported to produce **7** with a d.r. of 89:11 (*syn:anti*) and an e.r. of 69:31 (2R3S:2S3R),³⁰ the absolute configuration of the *syn*-enantiomers could be assigned based on the HPLC chromatogram of enzymatically produced **7** (**Supplementary Figure 27b** and **Supplementary Table 2**). As the enantioselectivity of wild-type 4-OT is unknown for the *anti*-enantiomers, their absolute configuration cannot be assigned. Therefore the peaks corresponding to the enantiomers of the *anti*-diastereoisomer are labelled as '*anti*-1' and '*anti*-2' based on the order of elution (**Supplementary Figure 27**).

Generating the mutability landscape of 4-OT for enantioselectivity

The same reaction mixtures that were used to monitor the Michael-type addition of **4** to **5a** catalyzed by the 4-OT variants were also used to evaluate the ability of each 4-OT variant to produce enantioenriched **7**. The duplicate reaction mixtures were pooled and filtered using 96-well ultrafiltration plates (AcroPrep™ Advance 96 Filter Plate 3K Omega, Pall® Life Sciences). The samples were passed through the filters by applying vacuum. The filtrate, which was now free of debris and protein, was analyzed for enantioenriched **7** by reverse-phase chiral HPLC, based on the above mentioned protocol. All four stereoisomers of the product **7** were separated using a Chiralpak® ID column (150 mm × 4.6 mm, Daicel) as shown in **Supplementary Figure 27**. To decrease the retention times, the flow rate was increased to 1.8 ml/min for the HPLC analyses used to generate the mutability landscape of 4-OT for enantioselectivity. The injection volume of the samples (i.e. the filtered reaction mixtures) was 20 µl. The e.r. values were calculated based on the peak integrations which were determined at 210 nm. This data was graphically represented in the mutability landscape for enantioselectivity, which was generated using Microsoft Excel 2010.

*Analytical scale determination of the ability of purified 4-OT variants to produce enantioenriched **7***

To measure the ability of purified 4-OT enzymes to produce enantioenriched **7**, analytical scale experiments were conducted. The reaction mixtures for this consisted of the following: **4** (32.5 mM), **5a** (0.65 mM), 65 μ M 4-OT (monomer concentration), and EtOH (5% v/v) in 20 mM NaH_2PO_4 buffer (pH 5.5); the final volume of this reaction mixture was 2 ml. The reaction was followed by monitoring the depletion of **5a** with UV-spectroscopy at 320 nm. Aliquots of 200 μ l were taken from this reaction mixture, cleared by ultrafiltration using Nanosep® 3K omega centrifugal filters (Pall® Life Sciences), and analyzed by chiral reverse phase HPLC using a Chiralpak ID column as described above. The obtained e.r. values are depicted in **Supplementary Table 4** and **5**.

*Assessment of the ability of purified 4-OT variants to produce enantioenriched **6a***

The reaction conditions used to assess the ability of purified 4-OT variants to produce enantioenriched **6a** was based on a previously described method.³⁰ The reaction mixture consisted of the following: 4-OT [14.7 μ M (0.7 mol%)], **3** (50 mM), **5a** (2.0 mM), and EtOH (10% v/v) in 20 mM NaH_2PO_4 buffer (pH 5.5). The reaction volume was 20 ml. UV-spectra (200 nm - 400 nm) were recorded at 30-min intervals to monitor the depletion of **5a**. The reaction was completed when the absorbance at 320 nm (λ_{max} **5a** = 320 nm) vanished. The enzyme was removed from the reaction mixture by ultrafiltration [Vivaspin centrifugal concentrator (MWCO 5000 Da)]. The filtrate was collected and extracted with ethyl acetate (3 $\times$ 10 ml) using a separatory funnel. The organic phases were combined, dried with MgSO_4 and evaporated *in vacuo* yielding a yellowish oil. The product was dissolved in CDCl_3 and a ^1H NMR spectrum was taken to confirm the formation of **6a**. All ^1H NMR spectra were recorded on a Bruker DRX-500 (500 MHz) spectrometer, and were referenced to residual CHCl_3 (δ = 7.26). After confirming that product **6a** was obtained, it was converted to its corresponding ethylene glycol acetal according to a literature procedure.²⁹ The e.r. of the ethylene glycol acetal of **6a** was determined based on a literature procedure, using reverse phase HPLC with a Chiralpak® AD-RH column (150 mm $\times$ 4.6 mm, Daicel) and an isocratic mobile phase ($\text{MeCN}:\text{H}_2\text{O}$ 70/30) at 0.5 ml/min.³⁰ The obtained e.r. values are depicted in **Supplementary Table 4** and **5**.

Combinatorial mutagenesis of ‘hotspots’

*Construction and screening of the focused library to further improve the Michael-type addition (**4** to **5a**) activity of 4-OT*

The mutability landscape revealed four residue positions at which mutations improved the activity of 4-OT for the Michael-type addition of **4** to **5a**. To further improve this activity, a 4-OT mutant library was constructed in which the most favorable mutations at these four positions were combined. The degeneracy at the positions His-6, Ala-33 and Phe-50 was limited to the following amino acid residues: His-6 : H, I, L, M and V; Ala-33:

A, D, E and Q; and Phe-50: F, V, L and A. Complete degeneracy was used at position Met-45 by using an NNK-codon. The number of possible variants in this library equaled 1,600. To construct this library, an equimolar mixture of the genes coding for A33D-, A33E-, A33Q- and wild-type 4-OT in the pJexpress 414 plasmid was used as a template in the PCR. The forward primers were an equimolar mixture of oligonucleotides carrying a mutagenic codon at position 6 (entry 7 – 11, **Supplementary Table 8**) and the reverse primers were an equimolar mixture of oligonucleotides carrying a mutagenic codon at position 50 and an NNK-codon at position 45 (entries 3 – 6, **Supplementary Table 8**). In total, 35 PCR cycles were done with the following temperatures and times: denaturation at 95°C (30 s), annealing at 60°C (60 s) and elongation at 72°C (60 s). The PCR product was gel purified, digested with *NdeI* and *BamHI* and cloned into an empty pJexpress 414 plasmid, which was compatibly digested and treated with alkaline phosphatase. After the ligation, the plasmid was transformed into chemically competent *E. coli* DH5 α . The transformants were selected at 37°C on LB agar plates containing 100 μ g/ml amp. Plasmid DNA was extracted from a few randomly selected transformants and the entire 4-OT gene was sequenced to ensure that the desired mutations were introduced. After establishing that the library was of good quality, the library was collected by pooling ~5,200 DH5 α colonies, an aliquot of which was used for plasmid DNA isolation. An aliquot of this isolated plasmid mixture was used to transform chemically competent *E. coli* BL21 (DE3) cells. Transformants were selected at 37°C on LB agar plates containing 100 μ g/ml amp. Of these transformants, 1,600 were randomly chosen and grown in 1.25 ml liquid LB/amp medium. These cultures were used to produce CFE according to the above mentioned protocol. The 4-OT mutants in the CFE were screened for their activities for the Michael-type addition of **3** to **5a** and **4** to **5a**. The same assay conditions as described above were used but without assessing the 4-OT expression levels. After screening the 1,600 members, 55 were selected for their pronounced activity. These mutants were again assayed for their Michael-type addition activities (**3** to **5a** and **4** to **5a**) as well as tested in a control assay monitoring the depletion of **5a** in the absence of aldehyde. Also the expression level of each of the mutant enzymes was assessed using the densitometric assay as described above. The most active mutant enzymes were identified based on their specific activities relative to wild-type 4-OT [$(U_{mut} \cdot \text{mg}^{-1}) / (U_{wt} \cdot \text{mg}^{-1})$] for the Michael-type additions. The plasmid DNA of the corresponding transformants was extracted and the entire 4-OT gene was sequenced to identify the mutations.

4-OT purification method

All 4-OT enzymes were purified using a procedure described elsewhere.²⁸ This procedure required adjustments for the purification of the 4-OT mutants H6M/A33E/F50V and H6L/A33E/F50V. In these cases the elution of enzyme from the DEAE-sepharose column was done using 0.1 M Na₂SO₄ (instead of 0.5 M). The protein precipitation was conducted with 1.6 M (NH₄)₂SO₄ in accordance with the protocol. However, after the precipitation step

the concentration of $(\text{NH}_4)_2\text{SO}_4$ was reduced to 1.0 M by adding the appropriate amount of 10 mM NaH_2PO_4 buffer (pH 7.3). The solution was loaded on a phenyl-sepharose column (equilibrated with 1.0 M $(\text{NH}_4)_2\text{SO}_4$) and washed with 1.0 M $(\text{NH}_4)_2\text{SO}_4$ (instead of 1.6 M). These adjustments were required because of the higher binding affinity of H6M/A33E/F50V and H6L/A33E/F50V for the phenyl-sepharose column material compared to wild-type 4-OT. Purified H6M/A33E/F50V and H6L/A33E/F50V enzymes were obtained in the flow through and wash fractions during this column chromatography step. The subsequent desalting, concentrating and concentration assessment of the purified enzyme was done according to a literature procedure.²⁸ All purified 4-OT variants were analyzed by electrospray ionization mass spectrometry (ESI-MS) to confirm the correct mass of the proteins.

Characterization of H6L/A33E/F50V and H6M/A33E/F50V

To confirm the improved "Michaelase" activity of the engineered 4-OT variants H6L/A33E/F50V and H6M/A33E/F50V, a progress curve analysis was done using these biocatalysts in the Michael-type addition of **4** to **5a** (**Supplementary Figure 28**). The reaction mixture for this analysis consisted of the following: **4** (100 mM), **5a** (2 mM), 4-OT (0.7 mol% relative to **5a**, based on monomer concentration), and EtOH (5% v/v) in 20 mM NaH_2PO_4 (pH 5.5); the final volume of this reaction mixture was 20 ml. Samples were taken from the reaction mixture and analyzed by UV-spectroscopy (320 nm) at timely intervals to monitor the depletion of **5a**. After the reaction was completed, product **7** was extracted with ethyl acetate (3 × 25 ml). The combined organic layers were dried using anhydrous MgSO_4 and concentrated *in vacuo* to yield **7**. The product was analyzed by chiral HPLC to establish the e.r. values (**Supplementary Table 3**).

Construction and testing of double and triple 4-OT mutants for improved enantioselectivity towards 2S3R-7

Double and triple mutants at positions His-6, Met-45, and Phe-50 were constructed by PCR using the appropriate combinations of template and primers. The gene coding for either M45Y, M45H or wild-type 4-OT in the pJexpress 414 vector was used as a template in the PCR in combination with either reverse primer "Rev. WT 4-OT" or "Rev. F50A" and forward primer "Fwd. WT4-OT" or "Fwd. H6I" (**Supplementary Table 8**, entries: 1, 2, 7 and 8). The PCR product was gel purified, digested with *Nde*I and *Bam*HI and cloned into an empty pJexpress 414 plasmid, which was compatibly digested and treated with alkaline phosphatase. After the ligation, the plasmid was transformed into chemically competent *E. coli* DH5 α cells. The transformants were selected at 37°C on LB agar plates containing 100 $\mu\text{g/ml}$ amp. Plasmid DNA was extracted from randomly picked transformants and the entire 4-OT gene was sequenced to ensure that the desired mutations were introduced. An aliquot of this plasmid solution was used to transform chemically competent *E. coli* BL21 (DE3) cells. CFEs prepared from cultures of these cells, expressing the 4-OT mutant proteins,

were used to determine the ability of these 4-OT mutants to produce enantioenriched **7** in the Michael-type addition of **4** to **5a** according to the earlier described protocol (see above). The e.r. values of **7** produced by these mutants are listed in **Supplementary Table 6**, depicting only those mutants which were active.

Semi-preparative scale synthesis experiments

Semi-preparative scale synthesis of 6a

To demonstrate the synthetic usefulness of the 4-OT mutants A33D and M45Y/F50A for the production of enantioenriched **6a**, relative to wild-type 4-OT, semi-preparative scale synthesis experiments were conducted using these three enzymes. For this, an amount of enzyme (6 or 12 mg) was added to a 100 ml round bottom flask, after which 20 mM NaH₂PO₄ buffer (pH 7.3, 6.5 or 5.5) was added, adjusting the final volume to 54 ml. The reaction was initiated by the addition of **3** (170 μ l) and 6 ml of a stock solution of **5a** (20 mM in EtOH). The final concentrations were: **5a** (2 mM), **3** (50 mM), 4-OT (14.7 or 29.4 μ M, based on monomer concentration), and 10% EtOH (v/v) in 20 mM NaH₂PO₄ buffer. Samples were taken from the round bottom flask at timely intervals for UV-spectrophotometric analysis, to monitor the reaction progress by following the depletion of **5a** (**Fig. 4a** and **Supplementary Figure 4**). After each measurement, the sample was recombined with the original reaction mixture in the round bottom flask. When the depletion of **5a** was complete, the product (**6a**) was extracted with ethyl acetate (3 $\times$ 40 ml) using a separatory funnel. The organic phases were combined, dried with MgSO₄ and evaporated *in vacuo* yielding a yellowish oil. The crude oil obtained after a reaction at pH 7.3 was purified using flash column chromatography (silica gel, EtOAc: n-Heptane 1:4 (v/v), 3 ml/min flow); the fractions containing **6a** were identified using TLC (EtOAc: n-Heptane 1:2 (v/v), visualization: KMnO₄) and combined. The product was concentrated *in vacuo*, redissolved in CDCl₃ and a ¹H NMR spectrum was recorded (**Supplementary Figure 2**) on a Bruker DRX-500 (500 MHz) spectrometer; all spectra were referenced to residual CHCl₃ (δ = 7.26). The crude oil obtained after the reactions at pH 5.5 and pH 6.5 did not require flash chromatography as the ¹H NMR spectra (**Supplementary Figures 2b** and **5**) indicated that **6a** was obtained in high purity. After confirming by ¹H NMR analysis that product **6a** was obtained, the yields were determined (wild-type 4-OT: 8.2 mg, 4.2 $\times$ 10⁻² mmol, 36%; M45Y/F50A: 6.9 mg, 3.6 $\times$ 10⁻² mmol, 30%; M45Y/F50A (pH 5.5): 15.2 mg, 7.9 $\times$ 10⁻² mmol, 65%; A33D: 6.5 mg, 3.4 $\times$ 10⁻² mmol, 28%; wild-type 4-OT (pH 6.5): 22 mg, 1.1 $\times$ 10⁻¹ mmol, 94%; A33D (pH 6.5): 22 mg, 1.1 $\times$ 10⁻¹ mmol, 94%. Aliquots of **6a** were converted to their corresponding ethylene glycol acetals according to a literature procedure.²⁹ The e.r. of the derivatized **6a** was determined using reverse phase HPLC with a Chiralpak® AD-RH column (150 mm $\times$ 4.6 mm, Daicel) (see **Supplementary Figures 3** and **6** for the chromatograms).

*Semi-preparative scale synthesis of **7***

To demonstrate the synthetic usefulness of the 4-OT mutants H6M/A33E/F50V, R39E and M45Y/F50A for the production of enantioenriched **7**, relative to wild-type 4-OT, semi-preparative scale synthesis experiments were conducted using these four enzymes. For this, 6 mg of enzyme was added to a 100 ml round bottom flask, after which 20 mM NaH₂PO₄ buffer (pH 5.5) was added, adjusting the final volume to 54 ml. The reaction was initiated by the addition of **4** (540 μ l) and 6 ml of a stock solution of **5a** (20 mM in EtOH). The final concentrations were: **5a** (2 mM), **4** (100 mM), 4-OT (14.7 μ M, based on monomer concentration), and 10% EtOH (v/v) in 20 mM NaH₂PO₄ buffer (pH 5.5). Samples were taken from the round bottom flask at timely intervals for UV-spectrophotometric analysis, to monitor the reaction progress by following the depletion of **5a** (Fig. 4b). After each measurement, the sample was recombined with the original reaction mixture in the round bottom flask. When the depletion of **5a** was complete, the product (**7**) was extracted with ethyl acetate (3 $\times$ 40 ml) using a separatory funnel. The organic phases were combined, dried with MgSO₄ and evaporated *in vacuo* yielding a yellowish oil. The crude oil was dissolved in CDCl₃ and a ¹H NMR spectrum was recorded (Supplementary Figure 7), using a Bruker DRX-500 (500 MHz) spectrometer (all spectra were referenced to residual CHCl₃ (δ = 7.26)). After confirming that product **7** was obtained, the yields were determined and corrected for the presence of small amounts of 2-nitro-1-phenylethanol (**10**) which is the product of the non-enzymatic water addition to **5a** (see Supplementary Figure 7). The corrected yields were: wild-type: 17.7 mg, 8.0×10^{-2} mmol, 67%; H6M/A33E/F50V: 20.2 mg, 9.1×10^{-2} mmol, 76%; R39E: 17.6 mg, 8.0×10^{-2} mmol, 66%; and M45Y/F50A: 22.7 mg, 10.2×10^{-2} mmol, 86%. The e.r. was determined using reverse phase HPLC with the Chiralpak® ID column (150 mm $\times$ 4.6 mm, Daicel) (see Supplementary Figure 8 for the chromatograms).

*Optical rotation analysis of **6a** and **7***

The optical rotations of enzymatically produced **6a** and **7** were measured in CHCl₃ on a Schmidt + Haensch polarimeter (polartronic MH8) with a 10 cm cell (c is given in g/100 ml). The data is presented in Supplementary Table 1.

*Semi-preparative scale synthesis of **6b-g** using A33D and M45Y/F50A*

To investigate the synthetic usefulness of 4-OT A33D and 4-OT M45Y/F50A for the enantioselective synthesis of γ -nitroaldehydes **6b-g**, semi-preparative scale reactions were conducted using these engineered 4-OT variants. The reaction conditions of these semi-preparative scale reactions and the enantiomeric ratio determinations were based on procedures described in earlier reports.^{29,31} The general procedure of these experiments was as follows: the appropriate amount of biocatalyst, **3** and co-solvent were diluted in buffer (20 mM NaH₂PO₄, pH 5.5). The reactions were initiated by the addition of the nitroalkene (**5b-g**) and the reaction progress was monitored using UV-spectrophotometric analysis.

After the UV-signal of the nitroalkene vanished, the products (**6b-g**) were extracted from the reaction mixtures and characterized by ^1H NMR spectroscopy and chiral HPLC or chiral GC, and the yields were determined. One adjustment on the earlier reported reaction conditions was made in the buffer pH of the A33D catalyzed reactions yielding compounds **6d** and **6e**. In these reactions EtOH was used as a co-solvent instead of DMSO, and the A33D enzyme appeared to precipitate under these conditions. To reduce protein precipitation, the buffer pH was increased to pH 6.5. More details on the reaction conditions (**Supplementary Table 9**), work-up procedures and enantiomeric ratio determinations are listed below.

3-(4-chlorophenyl)-4-nitrobutanal (**6b**)

The 4-OT catalyzed Michael-type addition of acetaldehyde (**3**) to *trans*-4-chloro- β -nitrostyrene (**5b**), using 4-OT A33D or 4-OT M45Y/F50A, was conducted under earlier reported optimized reaction conditions.³¹ The reaction mixtures consisted of the following: 4-OT (2.8 mol%, relative to **5b**), **3** (65 mM), **5b** (1.3 mM), and DMSO (45% v/v) in 20 mM NaH_2PO_4 buffer (pH 5.5); the final reaction volume was 50 ml (see **Supplementary Table 9**). After UV-spectrophotometric analysis of the reaction mixtures revealed full conversion of the nitroalkene (**Supplementary Figure 9**), the reaction mixtures were diluted with H_2O till a final DMSO concentration of 10% (v/v) was reached. Then compound **6b** was extracted with ethyl acetate (3×60 ml). The combined organic layers were washed with H_2O (3×60 ml) to eliminate traces of DMSO, and dried with brine (25 ml) and anhydrous Na_2SO_4 . The dried organic layer was concentrated *in vacuo* to yield 3-(4-chlorophenyl)-4-nitrobutanal (**6b**) (A33D: 12 mg, 5.3×10^{-2} mmol, 81% and M45Y/F50A: 9 mg, 3.93×10^{-2} mmol, 60%) as a colorless oil. The ^1H NMR spectroscopic data of **6b** (**Supplementary Figure 15**) are in agreement with previously published data.^{1,31} Enantiomeric ratios were determined by reverse phase HPLC using a Chiralpak® AD-RH column (150 mm $\times$ 4.6 mm, Daicel) (MeCN/water 32:68, 25°C) at 0.5 ml/min. UV detection at 220 nm: t_{R} : 3*R*-**6b** = 50.8 min, 3*S*-**6b** = 54.6 min, led to the following enantiomeric ratios (3*S*:3*R*): racemic **6b** = 50:50, 4-OT A33D **6b** = 98:2 and 4-OT M45Y/F50A **6b** = 38:62 (**Supplementary Figure 21**). The assignment of the absolute configuration was based on earlier reported chiral HPLC-data.³¹

3-(4-fluorophenyl)-4-nitrobutanal (**6c**)

The 4-OT catalyzed Michael-type addition of **3** to *trans*-4-fluoro- β -nitrostyrene (**5c**), using 4-OT A33D or 4-OT M45Y/F50A, was conducted under earlier reported optimized reaction conditions.³¹ The reaction mixtures consisted of the following: 4-OT (1.5 mol%, relative to **5c**), **3** (50 mM), **5c** (2.0 mM), and DMSO (40% v/v) in 20 mM NaH₂PO₄ buffer (pH 5.5); the final reaction volume was 60 ml (see **Supplementary Table 9**). After UV-spectrophotometric analysis of the reaction mixtures revealed full conversion of the nitroalkene (**Supplementary Figure 10**), the reaction mixtures were diluted with H₂O till a final DMSO concentration of 10% (v/v) was reached. Then compound **6c** was extracted with ethyl acetate (3 $\times$ 60 ml). The combined organic layers were washed with H₂O (3 $\times$ 60 ml) to eliminate traces of DMSO and dried with brine (25 ml) and anhydrous Na₂SO₄. The dried organic layer was concentrated *in vacuo* to yield 3-(4-fluorophenyl)-4-nitrobutanal (**6c**) (A33D: 16 mg, 7.6×10^{-2} mmol, 63% and M45Y/F50A: 12 mg, 5.8×10^{-2} mmol, 47%) as a colorless oil. The ¹H NMR spectroscopic data of **6c** (**Supplementary Figure 16**) are in agreement with previously published data.^{31,50} Enantiomeric ratios were determined by reverse phase HPLC using a Chiralpak® AD-RH column (150 mm $\times$ 4.6 mm, Daicel) (MeCN/water 30:70, 25°C) at 0.5 ml/min. UV detection at 220 nm: t_R : 3*R*-**6c** = 37.4 min, 3*S*-**6c** = 39.0 min, led to the following enantiomeric ratios (3*S*:3*R*): racemic **6c** = 49:51, 4-OT A33D **6c** = 99:1 and 4-OT M45Y/F50A **6c** = 14:86 (**Supplementary Figure 22**). The assignment of the absolute configuration was based on earlier reported chiral HPLC-data.³¹

3-(4-hydroxyphenyl)-4-nitrobutanal (**6d**)

The 4-OT catalyzed Michael-type addition of **3** to *trans*-4-hydroxy- β -nitrostyrene (**5d**) using 4-OT A33D and 4-OT M45Y/F50A was conducted based on earlier reported reaction conditions.²⁹ The reaction mixtures consisted of the following: 4-OT (0.5 mol%, relative to **5d**), **3** (50 mM), **5d** (2.0 mM), and EtOH (10% v/v) in 20 mM NaH₂PO₄ buffer (pH 6.5) for the A33D catalyzed reaction and in 20 mM NaH₂PO₄ buffer (pH 5.5) for the M45Y/F50A

catalyzed reaction. The final reaction volume was 60 ml (see **Supplementary Table 9**). After UV-spectrophotometric analysis of the reaction mixtures revealed almost complete nitroalkene conversion (**Supplementary Figure 11**), compound **6d** was extracted with ethyl acetate (3 × 40 ml). The combined organic layers were dried with brine (25 ml) and anhydrous Na₂SO₄, and concentrated *in vacuo* to yield 3-(4-hydroxyphenyl)-4-nitrobutanal (**6d**) (A33D: 11 mg, 5.2 × 10⁻² mmol, 44% and M45Y/F50A: 13 mg) as a yellowish oil. ¹H NMR spectroscopic analysis of the obtained product from the 4-OT M45Y/F50A catalyzed reaction, revealed incomplete conversion (~92%) of **5d** (**Supplementary Figure 17**). The yield of the product obtained after the M45Y/F50A catalyzed reaction was therefore corrected for this contamination of 6% (w/w) **5d**. The corrected yield of product **6d** from the M45Y/F50A catalyzed reaction was 12 mg (5.8 × 10⁻² mmol, 48%). The ¹H NMR spectroscopic data of **6d** (**Supplementary Figure 17**) are in agreement with previously published data.²⁹ The aldehyde functionality of **6d** was derivatized into a cyclic acetal to determine the enantiomeric ratios by reverse phase HPLC using a Chiralpak® AD-RH column (150 mm × 4.6 mm, Daicel) (MeCN/water 67:33, 25°C) at 0.5 ml/min. UV detection at 210 nm: *t_R*: deriv.-3R-**6d** = 5.8 min, deriv.-3S-**6d** = 8.0 min, led to the following enantiomeric ratios (3S:3R): racemic **6d** = 51:49, 4-OT A33D **6d** = 95:5 and 4-OT M45Y/F50A **6d** = 18:82 (**Supplementary Figure 23**). As a different stationary phase was used compared to literature, the optical rotations of the enzymatically obtained product **6d** were recorded for assigning the absolute configuration of the product. The optical rotations were A33D **6d**: = -2.6° (c=0.8, CHCl₃) and YA **6d**: = 11.0° (c=1.3, CHCl₃). The absolute configuration of the major enantiomers was tentatively assigned based on the sign of the optical rotation compared to other γ-nitroaldehydes, and based on the assumption that A33D and M45Y/F50A generate Michael-type addition products with a consistent geometry at the chiral center (see **Table 2 and 3**).³¹

3-(3-hydroxy-4-methoxyphenyl)-4-nitrobutanal (**6e**)

The 4-OT catalyzed Michael-type addition of **3** to *trans*-4-methoxy-3-hydroxy-β-nitrostyrene (**5e**) using 4-OT A33D or 4-OT M45Y/F50A was conducted under earlier reported optimized reaction conditions.³¹ The reaction mixtures consisted of the following: 4-OT (1.8 mol%, relative to **5e**), **3** (50 mM), **5e** (2.0 mM), and EtOH (10% v/v) in 20 mM NaH₂PO₄ buffer (pH 6.5) for the A33D catalyzed reaction and in 20 mM NaH₂PO₄ buffer (pH 5.5) for the M45Y/F50A catalyzed reaction. The final reaction volume was 60 ml (see **Supplementary Table 9**). After UV-spectrophotometric analysis of the reaction mixtures

revealed full conversion of the nitroalkene (**Supplementary Figure 12**), compound **6e** was extracted from the reaction mixture with ethyl acetate (3 × 25 ml). The combined organic layers were dried with brine (2 × 25 ml) and anhydrous Na₂SO₄, and concentrated *in vacuo* to yield 3-(3-hydroxy-4-methoxyphenyl)-4-nitrobutanal (**6e**) (A33D: 27 mg, 11.2 × 10⁻² mmol, 94% and M45Y/F50A: 21 mg, 8.8 × 10⁻² mmol, 73%) as a colorless oil. The ¹H NMR spectroscopic data of **6e** (**Supplementary Figure 18**) are in agreement with previously published data.³¹ The aldehyde functionality of **6e** was derivatized into a cyclic acetal to determine the enantiomeric ratios by reverse phase HPLC using a Chiralpak® AD-RH column (150 mm × 4.6 mm, Daicel) (MeCN/water 33:67, 25°C) at 0.8 ml/min. UV detection at 220 nm: *t_R*: deriv.-3*R*-**6e** = 20.6 min, deriv.-3*S*-**6e** = 31.2 min, led to the following enantiomeric ratios (3*S*:3*R*): racemic **6e** = 49:51, 4-OT A33D **6e** = 99:1 and 4-OT M45Y/F50A **6e** = 3:97 (**Supplementary Figure 24**). The assignment of the absolute configuration was based on earlier reported chiral HPLC-data.³¹

3-(3-(cyclopentyloxy)-4-methoxyphenyl)-4-nitrobutanal (**6f**)

The 4-OT catalyzed Michael-type addition of **3** to *trans*-4-methoxy-3-cyclopentyloxy-β-nitrostyrene (**5f**) using 4-OT A33D and 4-OT M45Y/F50A was conducted under earlier reported optimized reaction conditions.³¹ The reaction mixtures consisted of the following: 4-OT (3.7 mol%, relative to **5f**), **3** (50 mM), **5f** (2.0 mM), and DMSO (40% v/v) in 20 mM NaH₂PO₄ buffer (pH 5.5); the final reaction volume was 60 ml (see **Supplementary Table 9**). After UV-spectrophotometric analysis of the reaction mixtures revealed full conversion of the nitroalkene (**Supplementary Figure 13**), the reaction mixtures were diluted with H₂O till a final DMSO concentration of 10% (v/v) was reached. Then compound **6f** was extracted with ethyl acetate (3 × 60 ml). The combined organic layers were washed with H₂O (3 × 60 ml) to eliminate traces of DMSO and dried with brine (25 ml) and anhydrous Na₂SO₄. The dried organic layer was concentrated *in vacuo* to yield 3-(3-(cyclopentyloxy)-4-methoxyphenyl)-4-nitrobutanal (**6f**) (A33D: 24 mg, 7.8 × 10⁻² mmol, 65% and M45Y/F50A: 26 mg, 8.5 × 10⁻² mmol, 70%) as a colorless oil. The ¹H NMR spectroscopic data of **6f** (**Supplementary Figure 19**) are in agreement with previously published data.^{4,31} The aldehyde functionality of **6f** was derivatized into a methyl ester (see 6.4.1) to determine the enantiomeric ratios by normal phase HPLC, using a Chiralpak® IB column (150 mm × 4.6 mm, Daicel) (*n*-heptane/*i*-PrOH 95:5, 25°C) at 1 ml/min. UV detection at 220 nm: *t_R*: deriv.-3*R*-**6f** = 12.0 min, deriv.-3*S*-**6f** = 13.5 min, led to the following enantiomeric ratios (3*S*:3*R*): racemic **6f** = 51:49, 4-OT A33D **6f** = >99:1 and 4-OT M45Y/F50A **6f** = 32:68

(**Supplementary Figure 25**). The assignment of the absolute configuration was based on earlier reported chiral HPLC-data.³¹

5-methyl-3-(nitromethyl)hexanal (**6g**)

3

The 4-OT catalyzed Michael-type addition of **3** to (*E*)-4-methyl-1-nitropent-1-ene (**5g**) using 4-OT A33D or 4-OT M45Y/F50A was conducted under optimized reaction conditions as earlier reported.³¹ The reaction mixtures consisted of the following: 4-OT (5.3 mol%, relative to **5g**), **3** (150 mM), **5g** (3.0 mM), and DMSO (40% v/v) in 20 mM NaH₂PO₄ (pH 5.5); the final reaction volume was 12.8 ml (see **Supplementary Table 9**). After UV-spectrophotometric analysis of the reaction mixtures revealed full conversion of the nitroalkene (**Supplementary Figure 14**) (A33D: 50 min and M45Y/F50A: 70 min), compound **6g** was extracted from the reaction mixture with toluene (3 × 10 ml) and the combined organic layers were washed with dH₂O (3 × 5 ml) to eliminate traces of DMSO. The organic layer was dried with brine (5 ml) and anhydrous Na₂SO₄ and concentrated *in vacuo* to yield 5-methyl-3-(nitromethyl)hexanal (**6g**) (A33D: 4 mg, 2.3 × 10⁻² mmol, 60% and M45Y/F50A: 4 mg, 2.3 × 10⁻² mmol, 60%) as a colorless oil. The ¹H NMR spectroscopic data of **6g** (**Supplementary Figure 20**) are in agreement with previously published data.^{1,31} Enantiomeric ratios were determined by GC using a HP chiral 20% permethylated beta-cyclodextrin column (20 m, 90°C isocratic, 1.5 ml/min). Flame ionization detection: *t*_R: 3*S*-**6g** = 65.6 min and 3*R*-**6g** = 67.5 min, led to the following enantiomeric ratios (3*R*:3*S*): racemic **6g** = 50:50, 4-OT A33D **6g** = >99:1 and 4-OT M45Y/F50A **6g** = 10:90 (**Supplementary Figure 26**). The assignment of the absolute configuration was based on earlier reported chiral GC-data.³¹

DERIVATIZATION OF 6A, 6D, 6E AND 6F FOR ENANTIOMERIC EXCESS DETERMINATION

Acetalysation of 6a, 6d and 6e

The aldehyde functionality of **6a**, **6d** and **6e** was derivatized into a cyclic acetal, according to a literature procedure.^{29, 31} The γ -nitroaldehyde (5 mg), ethylene glycol (25.6 mg, 0.4 mmol) and *p*-TsOH (5 mol%, 0.2 mg, 1.0 × 10⁻³ mmol) were stirred in chloroform (800 μ l)

under a nitrogen atmosphere for 1 d at room temperature to yield the cyclic acetal of the γ -nitroaldehyde.

Esterification of **6f**

The aldehyde functionality of **6f** was converted into a methyl ester according to a literature procedure.³¹ A solution of KH_2PO_4 (1 mg, 7.35 μmol), NaClO_2 (0.65 mg, 7.2 μmol) and MeOH (75 μl , 16.3 μmol) in dH_2O (75 μl) were cooled down to 0°C, and **6f** (5 mg, 2.3×10^{-2} mmol) in acetonitrile (75 μl) was added. After the addition of H_2O_2 (35% solution, 8.5 μl) the mixture was stirred at room temperature for 2 h. The pH was adjusted to 3 with 0.1 M HCl and saturated Na_2SO_3 solution (200 μl) was added. The product was extracted with CH_2Cl_2 (3×1 ml). The combined organic layers were washed with dH_2O (500 μl), dried on Na_2SO_4 and concentrated *in vacuo*. The residue was dissolved in toluene (85 μl) and methanol (165 μl) and cooled down to 0°C. To this cooled solution trimethylsilyl diazomethane (20 μl , 40 μmol , 2.0 M in n-hexane) was added and stirred at 0°C for 5 min. The solution was stirred for additional 15 min at room temperature and the reaction quenched with concentrated AcOH. The solvents were evaporated under vacuum to yield methyl 3-(3-(cyclopentyloxy)-4-methoxyphenyl)-4-nitrobutanoate.

Synthesis of nitroalkenes **5a-g**

Nitroalkenes **5a-d** were commercially available while **5e-g** were synthesized and purified according to literature procedures.^{31, 51-54}

Synthesis of racemic **6a-6g**

The racemic products **6a-6g**, which served as references in the chiral-HPLC and chiral-GC analyses, were synthesized and purified according to literature procedures.³¹

Structural analysis of M45Y/F50A

Crystallization of M45Y/F50A

Prior to crystallization, purified enzyme was subjected to size-exclusion chromatography using 20 mM Tris-HCl (pH 7.5), containing 200 mM NaCl, as an eluent. Fractions containing 4-OT M45Y/F50A were pooled and concentrated to 10 mg/ml. Dynamic light scattering (DynaPro NanoStar, Wyatt Technology, CA) confirmed that the mutant existed as a hexamer in solution. Mass spectrometric analysis indicated that the enzyme had been correctly processed with the N-terminal methionine being cleaved off, leaving an exposed Pro-1. Attempts to reproduce the previously reported condition of crystallization for wild-type

4-OT were unsuccessful.³² Therefore, a screening for new conditions was carried out at room temperature in 96-well sitting drop crystallization plates using various available commercial screens. Drops (300 nl) were prepared with a nanodispenser (Mosquito, TTP Labtech) by mixing protein and reservoir solutions at a 1:1 ratio. The structure of M45Y/F50A was solved from a crystal that appeared in the following crystallization condition: 0.2 M sodium formate, 0.1 M bis-tris propane pH 8.5 and 20% PEG 3350 (w/v). Small crystals appeared after 3 days and grew to a size of approximately $100 \times 75 \times 75 \mu\text{m}^3$ within 7 days. Additionally, M45Y/F50A was co-crystallized in the presence of acetaldehyde (**3**) and *trans*- β -nitrostyrene (**5a**) with the aim of trapping a substrate or product bound structure. A stock solution of 100 mM **5a** was prepared in DMSO. Before setting up drops, M45Y/F50A was incubated with 50 mM **3** and 10 mM **5a** for 30 min. A screening was conducted to find new conditions as described above. The final concentration of DMSO in the drops was 5% (v/v). A single crystal with dimensions of $150 \times 100 \times 100 \mu\text{m}^3$ appeared in a condition consisting of 0.1 M PCTP buffer pH 7.0 and 25% (w/v) PEG 1500 after 2 weeks.

Data collection, processing and structure determination

Before proceeding with data collection, crystals were transferred to a drop containing the mother liquor together with 15% glycerol (v/v). Crystals were directly frozen in the liquid N₂ stream of the X-ray data collection equipment. Datasets for both crystals were collected in house at 110 K, using a Microstar rotating anode (Cu, wavelength of 1.5418 Å) X-ray source (Bruker AXS GmbH), coupled with Helios optics (Incoatec GmbH) and a MAR345dtb detector (Marresearch GmbH). Diffraction datasets were processed, scaled and merged using XDS⁵⁵ and Aimless.⁵⁶ The M45Y/F50A structure belonged to the space group C2, whereas the M45Y/F50A structure in complex with substrate **5a** crystallized in the P1 space group. A summary of the data collection and model refinement statistics is given in **Supplementary Table 7**.

Phaser MR⁵⁷ from the CCP4 software suite⁵⁸ was used to calculate initial phases. The structure of wild-type 4-OT (PDB ID: 4X19)³² was used as a search model to perform molecular replacement for the M45Y/F50A structure. Subsequently, multiple cycles of refinement were performed using Refmac⁵⁹, together with manual model building in COOT⁶⁰ to improve the structure. The final cycles of refinement were carried out in phenix.refine available from the Phenix software suite.⁶¹ The coordinates and topology for **5a** were generated using the PRODRG server.⁶² Target values for the bond lengths and bond angles, used in refinement, were derived from the crystal structure of **5a**⁶³, deposited in Cambridge Structural Database (CSD entry: KANWOL10).

Crystal structure analysis

Calculation of C α -backbone rmsd values, and superpositions of structures were performed using the protein structure comparison service Fold at the European Bioinformatics Institute (PDBeFold).⁶⁴ PyMol (Schrödinger)⁶⁵ was used to prepare figures and carry out

analysis of the structures. Molprobity, available as a part of the phenix.refine module, was used to validate the stereochemical quality of the models.⁶⁶ Coordinates for the free and nitrostyrene-bound structure of M45Y/F50A have been deposited with the Protein Data Bank (accession codes 5CLN and 5CLO, respectively).

ACKNOWLEDGEMENTS

The authors thank A. van Dam and M. Jeronimus-Stratingh for their expert assistance in acquiring ESI-MS data, L. Biewenga for his assistance in enzyme purification, and S. Heida for his assistance in characterizing single 4-OT mutants with inverted enantioselectivity. This research was financially supported by ALW grant 820.02.021 from the Division of Earth and Life Sciences of the Netherlands Organisation of Scientific Research (to W.J.Q. and G.J.P.), and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement n° 242293 (to G.J.P.).

AUTHOR CONTRIBUTIONS

B.J.B. and J.Y.M. developed the mutation strategy. J.Y.M., Y.M., M.R. and P.G.T. developed the screening assays. J.Y.M. performed all screening experiments. R.v.M. automated data analysis and robotic liquid handling. J.Y.M. performed combinatorial mutagenesis. J.Y.M. and A.K. performed preparative scale reactions and product analysis. H.P. performed X-ray crystallography experiments. E.M.G., A.M.W.H., W.J.Q. and G.J.P. supervised scientific work. All authors contributed to writing the paper.

DECLARATION

Dr. Jan-Ytzen van der Meer also used this paper as a chapter in his PhD thesis entitled Mutability-landscape guided enzyme engineering: Improving the promiscuous C-C bond-forming activities of 4-oxalocrotonate tautomerase (University of Groningen, 2016, ISBN: 978-90-367-8980-6).

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

ACCESSION CODES

Coordinates for the free and nitrostyrene-bound structure of M45Y/F50A have been deposited with the Protein Data Bank (PDB accession codes 5CLN and 5CLO, respectively).

REFERENCES

1. Gotoh, H., Ishikawa, H. & Hayashi, Y. Diphenylprolinol silyl ether as catalyst of an asymmetric, catalytic, and direct Michael reaction of nitroalkanes with α,β -unsaturated aldehydes. *Org. Lett.* **9**, 5307-5309 (2007).
2. Palomo, C. et al. Water-compatible iminium activation: organocatalytic Michael reactions of carbon-centered nucleophiles with enals. *Angew. Chem. Int. Ed. Engl.* **46**, 8431-8435 (2007).
3. Wiesner, M., Revell, J., D. & Wennemers, H. Tripeptides as efficient asymmetric catalysts for 1,4-addition reactions of aldehydes to nitroolefins- a rational approach. *Angew. Chem. Int. Ed. Engl.* **47**, 1871-1874 (2008).
4. García-García, P., Ladépêche, A., Halder, R. & List, B. Catalytic asymmetric Michael reactions of acetaldehyde. *Angew. Chem. Int. Ed. Engl.* **47**, 4719-4721 (2008).
5. Hayashi, Y., Itoh, T., Ohkubo, M. & Ishikawa, H. Asymmetric Michael reaction of acetaldehyde catalyzed by diphenylprolinol silyl ether. *Angew. Chem. Int. Ed. Engl.* **47**, 4722-4724 (2008).
6. Alza, E. & Pericas, M. A. A highly selective, polymer-supported organocatalyst for Michael additions with enzyme-like behavior. *Adv. Synth. Catal.* **351**, 3051-3056 (2009).
7. Riente, P., Mendoza, C. & Pericás, M. A. Functionalization of Fe_3O_4 magnetic nanoparticles for organocatalytic Michael reactions. *J. Mater. Chem.* **21**, 7350-7355 (2011).
8. Jentzsch, K. I., Min, T., Etcheson, J. I., Fetting, J. C. & Franz, A. K. Silyl fluoride electrophiles for the enantioselective synthesis of silylated pyrrolidine catalysts. *J. Org. Chem.* **76**, 7065-7075 (2011).
9. Qiao, Y., He, J., Ni, B. & Headley, A. D. Asymmetric Michael reaction of acetaldehyde with nitroolefins catalyzed by highly water-compatible organocatalysts in aqueous media. *Adv. Synth. Catal.* **354**, 2849-2853 (2012).
10. Geertsema, E. M. & Poelarends, G. J. in *Science of Synthesis : Biocatalysis in Organic Synthesis* (eds Faber, K., Fessner, W. D. & Turner, N.) (Thieme Chemistry, Stuttgart, Germany, 2014).
11. Xie, B., Guan, Z. & He, Y. Promiscuous enzyme-catalyzed Michael addition: synthesis of warfarin and derivatives. *J. Chem. Technol. Biotechnol.* **87**, 1709-1714 (2012).
12. Chai, S. et al. One-pot synthesis of spirooxindole derivatives catalyzed by lipase in the presence of water. *Adv. Synth. Catal.* **353**, 371-375 (2011).
13. Xu, J., Zhang, F., Liu, B., Wu, Q. & Lin, X. Promiscuous zinc-dependent acylase-mediated carbon-carbon bond formation in organic media. *Chem. Commun.*, 2078-2080 (2007).
14. Xu, J., Zhang, F., Wu, Q., Zhang, Q. & Lin, X. Hydrolase-catalyzed Michael addition of 1,3-dicarbonyl compounds to α,β -unsaturated compounds in organic solvent. *J. Mol. Catal. B.* **49**, 50-54 (2007).
15. Cai, J., Guan, Z. & He, Y. The lipase-catalyzed asymmetric C-C Michael addition. *J. Mol. Catal. B.* **68**, 240-244 (2011).
16. Wu, M. et al. A novel enzymatic tandem process: utilization of biocatalytic promiscuity for high stereoselective synthesis of 5-hydroxyimino-4, 5-dihydrofurans. *Tetrahedron* **67**, 2681-2688 (2011).
17. Strohmeyer, G. A. et al. Investigation of lipase-catalyzed Michael-type carbon-carbon bond formations. *Tetrahedron* **65**, 5663-5668 (2009).
18. Xu, K., Guan, Z. & He, Y. Acidic proteinase from *Aspergillus usarii* catalyzed Michael addition of ketones to nitroolefins. *J. Mol. Catal. B.* **71**, 108-112 (2011).
19. Chen, L. H. et al. 4-Oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer. *J. Biol. Chem.* **267**, 17716-17721 (1992).
20. Taylor, A. B., Czerwinski, R. M., Johnson, W. H., Jr, Whitman, C. P. & Hackert, M. L. Crystal structure of 4-oxalocrotonate tautomerase inactivated by 2-oxo-3-pentynoate at 2.4

- A resolution: analysis and implications for the mechanism of inactivation and catalysis. *Biochemistry* **37**, 14692-14700 (1998).
21. Whitman, C. P. The 4-oxalocrotonate tautomerase family of enzymes: how nature makes new enzymes using a beta-alpha-beta structural motif. *Arch. Biochem. Biophys.* **402**, 1-13 (2002).
 22. Poelarends, G. J., Puthan Veetil, V. & Whitman, C. P. The chemical versatility of the β - α - β fold: Catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life. Sci.* **65**, 3606-3618 (2008).
 23. Baas, B. J., Zandvoort, E., Geertsema, E. M. & Poelarends, G. J. Recent advances in the study of enzyme promiscuity in the tautomerase superfamily. *ChemBioChem* **14**, 917-926 (2013).
 24. Whitman, C. P., Aird, B. A., Gillespie, W. R. & Stolowich, N. J. Chemical and enzymic ketonization of 2-hydroxyruconate, a conjugated enol. *J. Am. Chem. Soc.* **113**, 3154-3162 (1991).
 25. Stivers, J. T. et al. Catalytic role of the amino-terminal proline in 4-oxalocrotonate tautomerase: affinity labeling and heteronuclear NMR studies. *Biochemistry* **35**, 803-813 (1996).
 26. Harris, T. K. et al. Kinetic, stereochemical, and structural effects of mutations of the active site arginine residues in 4-oxalocrotonate tautomerase. *Biochemistry* **38**, 12343-12357 (1999).
 27. Czerwinski, R. M. et al. Effects of mutations of the active site arginine residues in 4-oxalocrotonate tautomerase on the pKa values of active site residues and on the pH dependence of catalysis. *Biochemistry* **38**, 12358-12366 (1999).
 28. Zandvoort, E., Baas, B. J., Quax, W. J. & Poelarends, G. J. Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: enamine formation and aldolase activity. *ChemBiochem* **12**, 602-609 (2011).
 29. Zandvoort, E., Geertsema, E. M., Baas, B. J., Quax, W. J. & Poelarends, G. J. Bridging between organocatalysis and biocatalysis: asymmetric addition of acetaldehyde to beta-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angew. Chem. Int. Ed. Engl.* **51**, 1240-1243 (2012).
 30. Miao, Y., Geertsema, E. M., Tepper, P. G., Zandvoort, E. & Poelarends, G. J. Promiscuous catalysis of asymmetric Michael-type additions of linear aldehydes to beta-nitrostyrene by the proline-based enzyme 4-oxalocrotonate tautomerase. *ChemBioChem* **14**, 191-194 (2013).
 31. Geertsema, E. M. et al. Biocatalytic Michael-type additions of acetaldehyde to nitroolefins with the proline-based enzyme 4-Oxalocrotonate Tautomerase yielding enantioenriched gamma-nitroaldehydes. *Chem. Eur. J.* **19**, 14407-14410 (2013).
 32. Poddar, H., Rahimi, M., Geertsema, E. M., Thunnissen, A. M. W. H. & Poelarends, G. J. Evidence for the formation of an enamine species during aldol and Michael-type addition reactions promiscuously catalyzed by 4-Oxalocrotonate Tautomerase. *ChemBioChem* **16**, 738-741 (2015).
 33. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567-573 (2009).
 34. Chaparro-Riggers, J. F., Polizzi, K. M. & Bommaris, A. S. Better library design: data-driven protein engineering. *Biotechnol. J.* **2**, 180-191 (2007).
 35. Baas, B. J., Zandvoort, E., Wasil, A. A. & Poelarends, G. J. Demethionylation of Pro-1 variants of 4-oxalocrotonate tautomerase in *Escherichia coli* by co-expression with an engineered methionine aminopeptidase. *FEBS Open Bio* **4**, 651-658 (2014).
 36. Czerwinski, R. M., Johnson, W. H., Jr & Whitman, C. P. Kinetic and structural effects of mutations of the catalytic amino-terminal proline in 4-oxalocrotonate tautomerase. *Biochemistry* **36**, 14551-14560 (1997).
 37. McLaughlin, R. N., Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138-142 (2012).

38. Poelarends, G. J. et al. Evolution of enzymatic activity in the tautomerase superfamily: mechanistic and structural consequences of the L8R mutation in 4-Oxalocrotonate Tautomerase. *Biochemistry* **45**, 7700-7708 (2006).
39. Burks, E. A., Fleming, C. D., Mesecar, A. D., Whitman, C. P. & Pegan, S. D. Kinetic and structural characterization of a heterohexamer 4-oxalocrotonate tautomerase from *Chloroflexus aurantiacus* J-10-fl: implications for functional and structural diversity in the tautomerase superfamily. *Biochemistry* **49**, 5016-5027 (2010).
40. Bloom, J. D., Romero, P. A., Lu, Z. & Arnold, F. H. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct* **2**, 17 (2007).
41. Aharoni, A. et al. The 'evolability' of promiscuous protein functions. *Nat. Genet.* **37**, 73-76 (2005).
42. Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572-582 (2010).
43. Morley, K. L. & Kazlauskas, R. J. Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **23**, 231-237 (2005).
44. Fan, X., Rodríguez-Eschrch, C., Sayalero, S. & Pericàs, M. A. Paraldehyde as an acetaldehyde precursor in asymmetric Michael reactions promoted by site-isolated incompatible catalysts. *Chem. Eur. J.* **19**, 10814-10817 (2013).
45. Seebach, D. & Goliński, J. Synthesis of open-chain 2,3-disubstituted 4-nitroketones by diastereoselective Michael-addition of (E)-enamines to (E)-nitroolefins. A topological rule for C, C-bond forming processes between prochiral centres. *Helv. Chim. Acta* **64**, 1413-1423 (1981).
46. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat Methods* **11**, 801-807 (2014).
47. Fowler, D. M., Stephany, J. J. & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc.* **9**, 2267-2284 (2014).
48. Hecht, M., Bromberg, Y. & Rost, B. Better prediction of functional effects for sequence variants. *BMC Genomics* **16 Suppl 8**, S1-2164-16-S8-S1. Epub 2015 Jun 18 (2015).
49. Waddell, W. J. A simple ultraviolet spectrophotometric method for the determination of protein. *J. Lab. Clin. Med.* **48(2)**, 311-314 (1956).
50. Wang, Y., Li, P., Liang, X., Zhang, T. Y. & Ye, J. An efficient enantioselective method for asymmetric Michael addition of nitroalkanes to α , β -unsaturated aldehydes. *Chem. Commun.*, 1232-1234 (2008).
51. Milhazes, N. et al. β -Nitrostyrene derivatives as potential antibacterial agents: A structure–property–activity relationship study. *Bioorg. Med. Chem.* **14**, 4078-4088 (2006).
52. Hynes, P. S., Stupples, P. A. & Dixon, D. J. Organocatalytic asymmetric total synthesis of (R)-rolipram and formal synthesis of (3S,4R)-paroxetine. *Org. Lett.* **10**, 1389-1391 (2008).
53. Li, Z., Luo, S., Guo, Y., Xia, A. & Xu, D. The highly enantioselective Michael addition of ketones to nitrodienes catalyzed by the efficient organocatalyst system of pyrrolidinyl-thioimidazole and chiral thioureido acid. *Org. Biomol. Chem.* **8**, 2505-2508 (2010).
54. Bassas, O., Huuskonen, J., Rissanen, K. & Koskinen, A. M. A simple organocatalytic enantioselective synthesis of pregabalin. *Eur. J. Org. Chem.* **2009**, 1340-1351 (2009).
55. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133-144 (2010).
56. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204-1214 (2013).
57. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).

58. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760-763 (1994).
59. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184-2195 (2004).
60. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
61. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
62. Schüttelkopf, A. W. & Van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355-1363 (2004).
63. Pedireddi, V. R., Sarma, J. A. R. P. & Desiraju, G. R. Crystal engineering and solid state chemistry of some β -nitrostyrenes. *J. Chem. Soc., Perkin Trans.*, 311-320 (1992).
64. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256-2268 (2004).
65. The PyMOL molecular graphics system, version 1.6. Schrodinger, LLC.
66. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21 (2009).

SUPPLEMENTARY FIGURES

Supplementary Figure 1. Picture of a typical SDS-page gel used in the quantitative densitometric analysis to determine the 4-OT concentration in the cell-free extracts. Lanes 1, 2, 4 and 5 contain calibration samples of known amounts of purified wild-type 4-OT (resp. 0.5, 1, 2.5 and 5 μg). Lane 3 contains the PageRuler™ prestained protein ladder (Thermo Scientific). Lanes 7-24 contain CFE samples of *E. coli* BL21 (DE3) cultures, each expressing a different 4-OT variant. The blue boxes represent lanes and dashed orange lines indicate an area in which a protein band is detected by the program 'GeneTools'. This program quantifies each detected protein band, based on the size and intensity.

Supplementary Figure 2. ^1H NMR spectra of enzymatically obtained **6a**. Product **6a** was obtained from the semi-preparative scale reactions listed in **Table 1** and shown in **Fig. 4a**. Spectra a, c and d were obtained from the reactions performed at pH 7.3; these products were purified using flash chromatography before the ^1H NMR spectra were recorded. Spectrum b was obtained from a reaction performed at pH 5.5 with 4-OT mutant M45Y/F50A; no flash purification step was performed on this sample before the ^1H NMR spectrum was recorded. The spectroscopic data are in agreement with the literature.¹

Supplementary Figure 3. HPLC chromatograms of derivatised racemic **6a** and derivatised enzymatically obtained **6a**. The enzymatically obtained **6a** was obtained after the semi-preparative scale reactions listed in Table 1 and shown in Fig. 4a. The ^1H NMR spectra of the products are shown in Supplementary Figure 2.

Supplementary Figure 4. Progress curves of the Michael-type addition of **3** (50 mM) to **5a** (2 mM, 18 mg) catalyzed (1.4 mol%) by wild-type 4-OT or 4-OT A33D. The reactions were carried out in buffer [20 mM NaH₂PO₄ / 10% EtOH (v/v)] at pH 6.5. These progress curves are derived from the preparative scale reactions listed in **Table 1** (entries 4 and 5).

Supplementary Figure 5. ¹H NMR spectra of enzymatically obtained **6a**. Product **6a** was obtained from the semi-preparative scale reactions listed in **Table 1** (entries 4 and 5) and shown in **Supplementary Figure 4**. No flash purification step was performed on these samples before the ¹H NMR spectra were recorded. The spectroscopic data are in agreement with the literature.¹

Supplementary Figure 6. HPLC chromatograms of derivatised racemic **6a** and derivatised enzymatically obtained **6a**. The enzymatically obtained **6a** was obtained after the semi-preparative scale reactions (conducted at pH 6.5) listed in **Table 1** (entries 4 and 5) and shown in **Supplementary Figure 4**. The ^1H NMR spectra of the products are shown in **Supplementary Figure 5**.

Supplementary Figure 7. ¹H NMR spectra of enzymatically obtained **7**. The spectra of product **7**, which were obtained from the semi-preparative scale reactions listed in **Table 1** and shown in **Fig. 4b**, are in agreement with literature data.^{2,3} The peaks assigned with a' and f' correspond to the anti-diastereoisomer of **7**. 2-Nitro-1-phenylethanol (**10**) is the product of non-enzymatic water addition to **5a**.⁴

Supplementary Figure 8. Chromatograms for e.r. determination of enzymatically produced **7**. The enzymatically obtained **7** was obtained after the semi-preparative scale reactions listed in Table 1 and shown in Fig. 4b. The ^1H NMR spectra of the products are shown in Supplementary Figure 7.

Supplementary Figure 9. UV-spectra showing the conversion of the nitroalkene **5b**, upon incubation with **3** and either 4-OT A33D (left) or 4-OT M45Y/F50A (right) in 20 mM NaH_2PO_4 /45% DMSO (v/v) at pH 5.5.

Supplementary Figure 10. UV-spectra showing the conversion of the nitroalkene **5c**, upon incubation with **3** and either 4-OT A33D (left) or 4-OT M45Y/F50A (right) in 20 mM NaH_2PO_4 /40% DMSO (v/v) at pH 5.5.

Supplementary Figure 11. UV-spectra showing the conversion of the nitroalkene **5d**, upon incubation with **3** and either 4-OT A33D (left) or 4-OT M45Y/F50A (right) in 20 mM NaH_2PO_4 /10% EtOH (v/v). The pH of the buffer was pH 6.5 for the reaction with 4-OT A33D and pH 5.5 for the reaction with 4-OT M45Y/F50A.

Supplementary Figure 12. UV-spectra showing the conversion of the nitroalkene **5e**, upon incubation with **3** and either 4-OT A33D (left) or 4-OT M45Y/F50A (right) in 20 mM NaH_2PO_4 /10% EtOH (v/v). The pH of the buffer was pH 6.5 for the reaction with 4-OT A33D and pH 5.5 for the reaction with 4-OT M45Y/F50A.

Supplementary Figure 13. UV-spectra showing the conversion of the nitroalkene **5f**, upon incubation with **3** and either 4-OT A33D (left) or 4-OT M45Y/F50A (right) in 20 mM NaH_2PO_4 /40% DMSO (v/v) at pH 5.5.

Supplementary Figure 14. UV-spectra showing the conversion of the nitroalkene **5g**, upon incubation with **3** and either 4-OT A33D (left) or 4-OT M45Y/F50A (right) in 20 mM NaH_2PO_4 /5% DMSO (v/v) at pH 5.5.

Supplementary Figure 15. ^1H NMR spectra of **6b** obtained with 4-OT A33D (top) or 4-OT M45Y/F50A (bottom).

Supplementary Figure 16. ^1H NMR spectra of **6c** obtained with 4-OT A33D (top) or 4-OT M45Y/F50A (bottom).

Supplementary Figure 17. ^1H NMR spectra of **6d** obtained with 4-OT A33D (top) or 4-OT M45Y/F50A (bottom); **5d** is residual starting material.

Supplementary Figure 18. ^1H NMR spectra of **6e** obtained with 4-OT A33D (top) or 4-OT M45Y/F50A (bottom).

Supplementary Figure 19. ¹H NMR spectra of **6f** obtained with 4-OT A33D (top) or 4-OT M45Y/F50A (bottom).

Supplementary Figure 20. ¹H NMR spectra of **6g** obtained with 4-OT A33D (top) or 4-OT M45Y/F50A (bottom).

Supplementary Figure 21. HPLC chromatograms of racemic **6b** and enzymatically obtained **6b**.

Supplementary Figure 22. HPLC chromatograms of racemic **6c** and enzymatically obtained **6c**.

3

Supplementary Figure 23. HPLC chromatograms of derivatised racemic **6d** and derivatised enzymatically obtained **6d**.

Supplementary Figure 24. HPLC chromatograms of derivatised racemic **6e** and derivatised enzymatically obtained **6e**.

Supplementary Figure 25. HPLC chromatograms of derivatised racemic **6f** and derivatised enzymatically obtained **6f**.

Supplementary Figure 26. GC chromatograms of racemic **6g** and enzymatically obtained **6g**.

Supplementary Figure 27. Chiral phase HPLC analysis of **7** and assignment of absolute configuration. **a**, Chromatogram of chemically synthesized racemic **7**. **b**, Chromatogram of enzymatically produced **7**.

Supplementary Figure 28. Progress curves of the Michael-type addition of **4** to **5a** catalysed by wild-type 4-OT and 4-OT variants.

Supplementary Figure 29. Stereo diagrams of the electron density at the Pro-1 pocket in the 4-OT mutant structures. **a**, The Pro-1 pocket in the substrate-free M45Y/F50A mutant (PDB code – 5CLN) and **b**, in nitrostyrene-bound M45Y/F50A (PDB code – 5CLO). The $2F_o - F_c$ electron density maps have been depicted as a grey mesh and contoured at 1.0 σ . Protein residues and *trans*- β -nitrostyrene are shown as sticks. The apostrophes in the labels denote that the residues are from neighbouring chains. Two binding modes of *trans*- β -nitrostyrene are shown (using different colors) which fit the electron density equally well.

Supplementary Figure 30. Stereo view of C α -backbone superpositions of wild-type 4-OT and mutant M45Y/F50A. Wild-type 4-OT is depicted in orange whereas the mutant enzyme is shown in green.

Supplementary Figure 31. Surface representation of the Pro-1 pocket in **a**, wild-type 4-OT and **b**, M45Y/F50A. **c**, Close-up view of the active site showing the orientation of Tyr-45 in the mutant M45Y/F50A. The hydrogen bonding interaction is shown with a black dashed line and the distance is in Å. The M45Y/F50A structure is shown in green and the wild-type structure in orange. **d**, The two conformations of *trans*- β -nitrostyrene (**5a**), in the M45Y/F50A structure, fitting the electron density map. The grey mesh portrays the composite omit $2F_o - F_c$ map at 2.3 Å resolution (and contoured at 1.0 σ).

SUPPLEMENTARY TABLES

Supplementary Table 1. Optical rotations of the enzymatically produced **6a** and **7**

Biocatalyst	Optical rotation ([α] _D ²⁵)	Product	Abs. conf. ^a	e.r. ^b
WT 4-OT ^c	-6.7° (c= 0.78, CHCl ₃) ^c	6a	3S	90 : 10 ^c
A33D	-4.8° (c= 0.60, CHCl ₃)	6a	3S	99 : 1
M45Y/F50A	+5.0° (c=1.3, CHCl ₃)	6a	3R	94 : 6
WT 4-OT	+5.8° (c= 1.3, CHCl ₃)	7	2R3S	57 : 43
R39E	+45.4° (c= 1.8, CHCl ₃)	7	2R3S	95 : 5
H6M/A33E/F50V	-53.0° (c= 2.2, CHCl ₃)	7	2S3R	77 : 23
M45Y/F50A	-73.2° (c= 2.2, CHCl ₃)	7	2S3R	96 : 4

^a The absolute configuration is of the major enantiomer. ^b The e.r. was determined using chiral phase HPLC, and the absolute configurations were assigned based on literature data and confirmed by optical rotations.^{2,5-7} ^c Data reported previously.⁷

Supplementary Table 2. Peak integrations corresponding to the chromatograms in **Supplementary Figure 27**.

entry	Sample name	e.r. (2R3S:2S3R)
1 ^a	Rac- 7	50 : 50
2 ^b	WT 4-OT	70 : 30

^a Corresponding to the chromatogram of racemic **7** (**Supplementary Figure 27a**). ^b Corresponding to the chromatogram of enzymatically produced **7** (**Supplementary Figure 27b**).

Supplementary Table 3. Michael-type addition reactions of **4** to **5a** catalysed by wild-type 4-OT and 4-OT variants

Entry	Biocatalyst	Catalyst loading (mol%)	Reaction time (h) ^a	e.r. (2R3S:2S3R) ^b
1	WT	0.7	o.n.	60 : 40
2	H6L/A33E/F50V	0.7	1.5	11 : 89
3	H6M/A33E/F50V	0.7	1	23 : 77
4	F50V	0.7	5	22 : 78

^a Data obtained from **Supplementary Figure 28**. ^b The e.r. values were obtained with chiral phase HPLC analysis of samples of purified **7**, obtained from the same reactions as depicted in **Supplementary Figure 28**.

Supplementary Table 4. The e.r. values of **7** and **6a** produced by 4-OT mutants with a selectivity towards **2R3S-7** and **3S-6a**

Biocatalyst	e.r. of syn- 7 ^a (2R3S : 2S3R)	e.r. of 6a ^b (3S : 3R)
WT 4-OT	66 : 34	89 : 11
A33D	82 : 18	98 : 2
R39E	94 : 6	96 : 4
A57Y	84 : 16	84 : 16
R61M	86 : 14	88 : 12

All depicted e.r. values are obtained from experiments using purified biocatalysts. ^a The e.r. values were determined by HPLC analysis using a chiral stationary phase (see Methods). ^b The e.r. values of derivatised **6a** were determined by HPLC analysis using a chiral stationary phase (see Methods).

Supplementary Table 5. The e.r. values of **7** and **6a** produced by 4-OT mutants with a selectivity towards **2S3R-7** and **3R-6a**

Biocatalyst	e.r. syn- 7 ^b (2S3R : 2R3S)	e.r. of 6a ^c (3R : 3S)
H6I	92 : 8	79 : 21
R11I	75 : 25	53 : 47
M45H ^a	80 : 20	87 : 13
F50A	93 : 7	93 : 7
G54E	77 : 23	87 : 13

All depicted e.r. values are obtained from experiments using purified biocatalysts. ^a The mutant which had the best enantioselectivity towards **2S3R-7** at position Met-45 was the Tyr mutant. However, this mutant could not be purified as it expressed poorly. Therefore, M45H was used instead, as it was the second-best mutant at this residue position in terms of enantioselectivity, based on the data from the mutability landscape. ^b The e.r. values were determined by HPLC analysis using a chiral stationary phase (see Methods). ^c The e.r. values of derivatised **6a** were determined by HPLC analysis using a chiral stationary phase (see Methods).

Supplementary Table 6. The e.r. values of product **7** produced by double and triple mutants of 4-OT

Entry	Mutant ^a	e.r. (2R3S : 2S3R) ^b
1	H6I/M45Y/F50A	4 : 96
2	M45H/F50A	2 : 98
3	M45Y/F50A	1 : 99
4	H6I/F50A	4 : 96

^a The other double and triple mutants (H6I/M45H/F50A, H6I/M45Y and H6I/M45H) did not have any detectable 'Michaelase' activity. ^b The e.r. values were determined by HPLC analysis using a chiral stationary phase (see Methods).

Supplementary Table 7. Crystallographic data collection and refinement statistics

	M45Y/F50A	M45Y/F50A-Nitrostyrene
Data collection		
Space group	C2	P1
Unit cell dimensions, <i>a</i> , <i>b</i> , <i>c</i> (Å)	87.1, 87.2, 97.2	56.7, 61.1, 75.8,
α , β , γ (°)	90.0, 94.5, 90.0	98.2, 95.1, 91.1
Resolution range (Å)	57 – 2.7	56 – 2.3
	(2.84 – 2.71) *	(2.38 – 2.30)
R_{merge}	0.148 (0.714)	0.074 (0.414)
Completeness (%)	99.4 (95.8)	94.7 (93.8)
Average <i>I</i> / σ	6.9 (1.8)	10.7 (2.0)
Redundancy	4.2	2.0
Refinement		
Resolution range	57 – 2.7	44 – 2.3
Nr of reflections	18204	42246
$R_{\text{work}} / R_{\text{free}}$	0.232/0.263	0.236/0.270
Composition of asymmetric unit (AU)	12 chains	18 chains
total number of atoms	5222	8027
solvent molecules	50	244
others	--	1 <i>trans</i> - β -nitrostyrene
Average B (Å ²)		
protein	40.7	33.7
ligand	---	41.9
solvent molecules	32.4	28.7
R.m.s.d bonds (Å) / angles (°)	0.005/ 0.988	0.004/ 0.876
Ramachandran plot		
% most favoured	97.4	98.2
% allowed	2.6	1.8
PDB code	5CLN	5CLO

* Highest resolution shell values are shown in parentheses. Single crystals of the YA mutant and of the mutant crystallised in the presence of nitrostyrene were used to collect data sets for structure determination.

Supplementary Table 8. The sequence of the used mutagenic primers

Entry	Name	Sequence 5'-3'
1	Rev. WT 4-OT	ATGTTATGGATCCTCAGCGTCTGACCTTGCTGGCCAGTTTCGCCGCGCGATGCCGAAGTGGCCCTT
2	Rev. F50A	ATGTTATGGATCCTCAGCGTCTGACCTTGCTGGCCAGTTTCGCCGCGCGATGCCGGCGTGCCCTT
3	Rev. 45NNK	ATGTTATGGATCCTCAGCGTCTGACCTTGCTGGCCAGTTTCGCCGCGCGATGCCGAAGTGCCCTTGGCMNNNCTCCGT
4	Rev. F50A/45NNK	ATGTTATGGATCCTCAGCGTCTGACCTTGCTGGCCAGTTTCGCCGCGCGATGCCGGCGTGCCCTTGGCMNNNCTCCGT
5	Rev. F50V/45NNK	ATGTTATGGATCCTCAGCGTCTGACCTTGCTGGCCAGTTTCGCCGCGCGATGCCGACGTTGGCCCTTGGCMNNNCTCCGT
6	Rev. F50L/45NNK	ATGTTATGGATCCTCAGCGTCTGACCTTGCTGGCCAGTTTCGCCGCGCGATGCCCAGGTTGGCCCTTGGCMNNNCTCCGT
7	Fwd. WT 4-OT	ATAGCAGGTACATATGCGCTATTGCCCAGATCCACAT
8	Fwd. H6I	ATAGCAGGTACATATGCGCTATTGCCCAGATCATTTATCCTTGAAGGC
9	Fwd. H6L	ATAGCAGGTACATATGCGCTATTGCCCAGATCCTTATCCTTGAAGGC
10	Fwd. H6M	ATAGCAGGTACATATGCGCTATTGCCCAGATCATGATCCTTGAAGGC
11	Fwd. H6V	ATAGCAGGTACATATGCGCTATTGCCCAGATCGTCATCCTTGAAGGC

Mutagenic codons are underlined and in bold, NdeI restriction sites (CATATG) are represented with a dashed underline and BamHI restriction sites (GGATCC) are represented with a wave underline.

Supplementary Table 9. Summary of the reaction conditions for the semi-preparative scale synthesis of γ -nitroaldehydes **6b-g**.

Product	[Nitroalkene] (mM)	[Acetaldehyde] (mM)	Mol% 4-OT ^a	Co-solvent (% v/v)	Reaction volume (ml)
6b	1.3	65	2.8	DMSO 45%	50
6c	2.0	50	1.5	DMSO 40%	60
6d	2.0	50	0.5	EtOH 10%	60
6e	2.0	50	1.8	EtOH 10%	60
6f	2.0	50	3.7	DMSO 40%	60
6g	3.0	150	5.3	DMSO 5%	12.8

These reaction conditions were based on earlier reported optimized reaction conditions.^{7, 8} ^a Relative to the nitroalkene.

SUPPLEMENTARY DISCUSSION

Engineering 4-OT variants with further enhanced 'Michaelase' activity

To investigate whether the positions at which single mutations improved the activity of 4-OT in the Michael-type addition of **4** to **5a** (i.e. His-6, Ala-33, Met-45 and Phe-50) are good targets to further enhance 4-OT's promiscuous 'Michaelase' activity, these 'hotspot' positions were subjected to combinatorial mutagenesis (see Methods). The resulting library was screened for Michael-type addition activity of **3** or **4** to **5a**. As might be expected from the mutability landscape shown in **Fig. 3b** (main text), this library did not contain any mutants which were more active in the Michael-type addition of **3** to **5a** than the single mutant A33D. On the other hand, screening of the library resulted in the identification of two triple mutants (H6M/A33E/F50V and H6L/A33E/F50V) with pronounced activity for the Michael-type addition of **4** to **5a**. Surprisingly, these mutants did not have a mutation at position Met-45, even though this position was randomized to all possible amino acids. This may suggest that mutations at position Met-45 do not contribute to the favourable effect of the mutations at the other positions in either an additive or synergistic manner. The progress curves of the Michael-type addition of **4** to **5a** catalyzed by the purified enzymes clearly demonstrate the strongly enhanced activities of these triple mutants when compared to wild-type 4-OT and the best single mutant F50V (**Supplementary Figure 28**).

The synthetic potential of the best triple mutant, H6M/A33E/F50V, was further compared to that of wild-type 4-OT using 0.7 mol% enzyme and a 25-fold excess of **4** (50 mM) over **5a** (2 mM). Under these optimized conditions, the triple mutant is more efficient in the addition of **4** to **5a** than wild-type 4-OT, with the H6M/A33E/F50V-catalyzed reaction being completed within 1 h at room temperature (**Fig. 4b**). Product analysis (**Supplementary Figures 7 and 8**) revealed that mutant H6M/A33E/F50V exhibits inverted enantioselectivity (compared to wild-type 4-OT) for the Michael-type addition of **4** to **5a**, producing the 2S3R enantiomer of 2-ethyl-4-nitro-3-phenylbutanal (**7**) with an e.r. of 77:23 (**Table 1, entry 8**). Thus, combinatorial mutagenesis of the identified 'hotspot' positions

indeed yielded a highly active biocatalyst for the Michael-type addition of **4** to **5a**, yielding enantioenriched γ -nitroaldehyde **7**.

SUPPLEMENTARY REFERENCES

1. Gotoh, H., Ishikawa, H. & Hayashi, Y. Diphenylprolinol silyl ether as catalyst of an asymmetric, catalytic, and direct Michael reaction of nitroalkanes with α,β -unsaturated aldehydes. *Org. Lett.* **9**, 5307-5309 (2007).
2. Betancort, J. M. & Barbas, C. F., Catalytic direct asymmetric Michael reactions: taming naked aldehyde donors. *Org. Lett.* **3**, 3737-3740 (2001).
3. Patora-Komisarska, K., Benohoud, M., Ishikawa, H., Seebach, D. & Hayashi, Y. Organocatalyzed Michael addition of aldehydes to nitro alkenes - Generally accepted mechanism revisited and revised. *Helv. Chim. Acta* **94**, 719-745 (2011).
4. Purkharthofer, T. et al. A biocatalytic Henry reaction--the hydroxynitrile lyase from *Hevea brasiliensis* also catalyzes nitroaldol reactions. *Angew. Chem. Int. Ed. Engl.* **45**, 3454-3456 (2006).
5. Alexakis, A. & Andrey, O. Diamine-catalyzed asymmetric Michael additions of aldehydes and ketones to nitrostyrene. *Org. Lett.* **4**, 3611-3614 (2002).
6. Miao, Y., Geertsema, E. M., Tepper, P. G., Zandvoort, E. & Poelarends, G. J. Promiscuous catalysis of asymmetric Michael-type additions of linear aldehydes to beta-nitrostyrene by the proline-based enzyme 4-oxalocrotonate tautomerase. *ChemBioChem.* **14**, 191-194 (2013).
7. Geertsema, E. M. et al. Biocatalytic Michael-type additions of acetaldehyde to nitroolefins with the proline-based enzyme 4-oxalocrotonate tautomerase yielding enantioenriched gamma-nitroaldehydes. *Chem. Eur. J.* **19**, 14407-14410 (2013).
8. Zandvoort, E., Geertsema, E. M., Baas, B. J., Quax, W. J. & Poelarends, G. J. Bridging between organocatalysis and biocatalysis: asymmetric addition of acetaldehyde to beta-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angew. Chem. Int. Ed. Engl.* **51**, 1240-1243 (2012).

c h a p t e r f o u r

FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF AN UNUSUAL COFACTOR-INDEPENDENT OXYGENASE

B.-J. Baas
H. Poddar
E. M. Geertsema
H. J. Rozeboom
M. P. de Vries
H. P. Permentier
A.-M. W. H. Thunnissen
G. J. Poelarends

Published in Biochemistry 2015, 54, 1219-1232

ABSTRACT

The vast majority of characterized oxygenases use bound cofactors to activate molecular oxygen to carry out oxidation chemistry. Here, we show that an enzyme of unknown activity, RhCC from *Rhodococcus jostii* RHA1, functions as an oxygenase using 4-hydroxyphenylenolpyruvate as substrate. This unique and complex reaction yields 3-hydroxy-3-(4-hydroxyphenyl)-pyruvate, 4-hydroxybenzaldehyde, and oxalic acid as major products. Incubations with H_2^{18}O , $^{18}\text{O}_2$, and a substrate analogue, suggest that this enzymatic oxygenation reaction likely involves a peroxide anion intermediate. Sequence similarity and the crystal structure of RhCC (solved at 1.78 Å resolution) reveal that this enzyme belongs to the tautomerase superfamily. Members of this superfamily typically catalyze tautomerization, dehalogenation or decarboxylation rather than oxygenation reactions. The structure shows the absence of cofactors, establishing RhCC as a rare example of a redox metal- and coenzyme-free oxygenase. This sets the stage to study the mechanistic details of cofactor-independent oxygen activation in the unusual context of the tautomerase superfamily.

The oxygen-rich environment of the earth's biosphere has provided nature with the opportunity to evolve a vast number of enzymes that have the ability to utilize the strong oxidizing power of molecular oxygen (dioxygen, O_2).^{1,2} These enzymes possess the capacity to activate molecular oxygen in order to oxidize organic substrates. In order to achieve this, oxidases and oxygenases are dependent on redox cofactors such as flavin, pterin, or coordinated transition metals such as Fe^{2+} or Cu^+ .³ In recent years, however, a few O_2 -utilizing enzymes have been discovered that stand apart from classical oxidases and oxygenases as they possess the remarkable ability to activate molecular oxygen without any cofactor, a highly unusual property which has evoked great scientific interest.⁴

Cofactor-independent oxidases and oxygenases appear to catalyze a direct reaction between molecular oxygen and an organic substrate, which is both thermodynamically unfavorable as well as chemically forbidden.²⁻⁴ The reason for this is that molecular oxygen in its ground state exists as a triplet-biradical, meaning that its electronic structure has two unpaired electrons in its highest occupied molecular orbitals.³⁻⁵ This results in a total spin quantum number $S=1$ (triplet-state), whereas the electronic structure of organic substrates is fully composed of paired electrons and hence have a spin-number $S=0$ (singlet-state).³⁻⁵ The direct reaction of a triplet-state molecule with a molecule in the singlet-state is so-called 'spin-forbidden' as it violates the Wigner-rule of spin conservation.⁶ Therefore, standard two-electron reaction mechanisms that are generally valid for reactions involving singlet-state molecules do not apply to molecular oxygen. Cofactors can overcome this barrier by providing spin-allowed reaction mechanisms.^{3,5} Activated oxygen species are generated of which the reaction with organic substrates is thermodynamically favorable, thus providing the driving force for these reactions.

Cofactor-independent oxidases and oxygenases must therefore rely on a different mechanism in order to gain access to the oxidizing power of molecular oxygen. Evidence for common mechanistic features is slowly emerging and involves base-catalyzed formation of a substrate anion and a phenomenon termed 'substrate-assisted catalysis'.⁷⁻¹⁰ The anionic form of the substrate is able to reduce molecular oxygen via single-electron transfer, by which molecular oxygen is converted into its superoxide anion ($O_2^{\cdot-}$) form. The one-electron oxidized substrate anion now resides in the form of a radical cation, where the stabilization of the remaining unpaired electron by the electronic structure of the substrate itself is a determining feature. Although the few cofactor-independent oxidases and oxygenases that have been identified to date seem to share mechanistic features, they appear not to be related in sequence or structure as they belong to distinct enzyme (super)families.⁴ The identification of novel cofactor-independent oxidases or oxygenases from different protein superfamilies, and characterization of the structural and mechanistic features that underlie catalysis, would provide valuable insights into the mechanistic requirements for cofactor-independent activation of molecular oxygen.

In this work, we report that an enzyme of unknown activity, RhCC from *Rhodococcus jostii* RHA1, functions as a unique cofactor-independent oxygenase utilizing 4-hydroxyphenylenolpyruvate as substrate. This complex enzymatic reaction likely

involves a peroxide anion intermediate and yields 3-hydroxy-3-(4-hydroxyphenyl)-pyruvate, 4-hydroxybenzaldehyde, and oxalic acid as major products. Sequence and structure comparisons demonstrate that this enzyme belongs to the tautomerase superfamily, the members of which are characterized by a structural β - α - β -fold and a catalytic amino-terminal proline.¹¹⁻¹³ RhCC is the first identified oxygenase within this superfamily, providing a framework to investigate cofactor-independent oxygenation within a structural fold generally employed to catalyze tautomerization, dehalogenation or decarboxylation reactions.

EXPERIMENTAL PROCEDURES

Materials

The sources of the components of buffers and media, PCR purification, gel extraction and miniprep kits, pre-packed PD-10 Sephadex G-25 columns, and the oligonucleotides, enzymes and reagents used in the molecular biology procedures are reported elsewhere.¹⁴ All other chemicals and biochemicals, including 4HPP and phenylenolpyruvate (**9**), were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) unless stated otherwise. (*E*)-2-fluoro-*p*-hydroxycinnamic acid (**10**) was a kind gift from Prof. Christian P. Whitman (University of Texas, Austin, TX). The genomic DNA of *R. jostii* RHA1 (GenBank sequence: CP000431) was kindly supplied by Dr. Robert van der Geize (University of Groningen, The Netherlands).

General methods

Techniques for restriction enzyme digestions, ligation, transformation, and other standard molecular biology manipulations were based on methods described elsewhere or as suggested by the manufacturer.¹⁵ The PCR was carried out in a DNA thermal cycler (model GS-1) obtained from Biolegio (Nijmegen, The Netherlands). DNA sequencing was performed by ServiceXS (Leiden, The Netherlands) or Macrogen (Seoul, Korea). Protein was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in gels containing 10% polyacrylamide. The gels were stained with Coomassie brilliant blue. Protein concentrations were determined by the method of Waddell.¹⁶ Kinetic data were obtained on a V-660 spectrophotometer from Jasco (IJsselstein, The Netherlands). The kinetic data were fitted by nonlinear regression data analysis using the Grafit program (Erithacus, Software Ltd., Horley, U.K.) obtained from Sigma Chemical Co. Protein ESI-MS spectra were recorded using a Sciex API 3000 triple quadrupole mass spectrometer (AB Sciex, Concord, Ontario, Canada). ¹H NMR spectra were recorded on a Bruker DRX-500 (500 MHz) spectrometer. Chemical shifts for protons are reported in parts per million scale (d scale) and are referenced to H₂O (d = 4.80). ¹³C NMR, COSY, ¹H,¹³C-HSQC, and ¹H,¹³C-HMBC spectra were recorded (CryoNMR) using a Bruker 500 MHz Avance III system (Bruker Biospin, Rheinstetten, Germany). Dynamic light scattering (DLS) experiments were performed using a DynaPro MS800TC instrument (Wyatt Technology Corporation, Santa

Barbara, CA) at 20 °C. DLS data were processed and analyzed with the Dynamics software (Wyatt Technology Corporation).

Cloning of the gene encoding RhCC

The gene coding for RhCC was amplified using two synthetic primers, genomic DNA of *R. jostii* strain RHA1, and the PCR reagents supplied in the Phusion DNA Polymerase system. The forward primer 5'- A TAG CAG GTA **CAT ATG** CCG TAC TGG GAG ATC TTC ACG CCC-3' contains an *NdeI* restriction site (in bold) followed by 24 bases corresponding to the coding sequence of the gene. To clone the gene encoding RhCC in frame with the sequence that codes for the polyhistidine region of the expression vector pET20b(+), the reverse primer 5'- G TGA TGT TAT **AAG CTT** CGA GGC AGC GAC GTC GTA CGG A -3', which contains a *HindIII* restriction site (in bold), was used. The amplification mixture was made up in Phusion High-Fidelity buffer and contained 250 µM of each dNTP, 150 ng of each primer, 250 ng of genomic DNA, and 1 U of Phusion DNA polymerase. The cycling parameters were 95 °C for 5 min followed by 35 cycles of 98 °C for 30 s, 45 °C for 30 s and 72 °C for 20 s, with a final elongation step of 72 °C for 10 min. The resulting PCR product and the pET-20b(+) vector (Novagen) were digested with *NdeI* and *HindIII* restriction enzymes, after which the vector was dephosphorylated with alkaline phosphatase. Following purification, the PCR product and vector were ligated using T4 DNA ligase. An aliquot of the ligation mixture was transformed into competent *Escherichia coli* BL21 (DE3) cells. Transformants were selected at 37 °C on LB/ampicillin plates. Plasmid DNA was isolated from several colonies and analyzed by restriction analysis for the presence of the insert. The cloned gene was sequenced to verify that no mutations had been introduced during the amplification of the gene. This newly constructed expression vector was named pRhCC.

4

Expression and purification of RhCC

The RhCC protein was produced constitutively in *E. coli* BL21(DE3) using the T7 expression system. Fresh BL21(DE3) transformants containing vector pRhCC, were collected from a LB/ampicillin plate and used to inoculate a preculture (LB/ampicillin medium, 5 mL). After growth for 3-5 h at 37 °C, the preculture was used to inoculate 1 L auto-induction (ZYM) medium (10 g/L tryptone, 5 g/L yeast extract, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 5 mM Na₂SO₄ (pH 6.7), supplemented with 0.5% (v/v) glycerol, 0.05% (w/v) glucose, MgSO₄ (2 mM), ampicillin (100 µg/mL), and 0.2% (w/v) lactose) in a 3 L Erlenmeyer flask to a starting OD₆₀₀ of about 0.01. After overnight growth at 30 °C with vigorous shaking, cells were harvested by centrifugation (10 min at 2300g) and resuspended in 10 mM NaH₂PO₄ buffer (pH 7.3, buffer A) to a total volume of about 10 mL. Protease inhibitors (Complete Mini, Roche, Mannheim, Germany) were added and cells were disrupted by sonication for 5 ´ 1 min (with 5 min rest in between each cycle) at 50% duty cycle/50% output in a Branson Sonifier 450 (Branson Ultrasonics Corporation, Danbury, CT). Unbroken cells and debris

were removed by centrifugation (45 min at 18000g). The supernatant was filtered through a 0.2 μm pore diameter filter, made 10 mM in imidazole (from a 1 M stock solution in water, pH adjusted to 7.3), and applied to a gravity flow column containing 0.5 mL Ni-Sepharose 6 Fast Flow resin. After overnight incubation at 4 °C, while mixing at 20 rpm on a rotor, the nonbound proteins were removed from the column by gravity flow. The column was washed with 20 mM imidazole in buffer A, followed by 40 mM imidazole in buffer A, after which retained proteins were eluted with 1 mL of 250 mM imidazole in buffer A. Subsequently, the protein sample was applied to a PD-10 gel filtration column, which was previously equilibrated in buffer A, and proteins were eluted with buffer A. Fractions (1 mL) were analyzed by SDS-PAGE, and those containing highly purified RhCC were combined and concentrated to a protein concentration of 5-20 mg/mL using a Vivaspin centrifugal concentrator equipped with a 5000 Da molecular weight cut-off filtration membrane (Sartorius Stedim Biotech S.A., France).

Crystallization of RhCC

Initial crystallization conditions were determined by vapour-diffusion sitting-drop experiments in MRC 96-well crystallization plates (Molecular Dimensions), using a Mosquito crystallization robot (TTP Labtech) for automatic drop-dispensing. Protein drops were prepared by mixing 200 nL protein solution (12.5 mg/mL in 20 mM Tris/HCl buffer, pH 7.5) with an equal volume of precipitant reservoir solution. Reservoir solutions were obtained from various commercial crystallization screens, and equilibration was performed at room temperature. The crystallization lead thus determined was reproduced and optimized manually by vapour-diffusion hanging-drop experiments (drops containing 2 μL protein and 2 μL reservoir solution), resulting in the following final crystallization condition: 5–12.5 mg/mL protein, 24–28% (w/v) polyethylene glycol 3350 (PEG 3350) and 100 mM Bis-Tris-propane (pH 8.0–8.5). Crystals grew within one week to an average size of 200 x 100 x 50 μm^3 .

Data collection, processing and RhCC structure determination

Prior to X-ray diffraction data collection, crystals were soaked briefly in a cryoprotectant solution containing 30% PEG 3350 and 15% glycerol in 100 mM Bis-Tris-propane (pH 8), and subsequently flash-cooled in liquid nitrogen. A 1.78 Å data set was collected at beamline ID14-4 at the European Synchrotron Radiation Facility (Grenoble, France). Diffraction data were processed, scaled and merged using the programs iMosflm and Aimless from the CCP4 software suite.¹⁷⁻¹⁹ Crystals belonged to space group C22₁ and contained a protein trimer in the asymmetric unit (solvent content 47%). A summary of the data collection statistics is given in **Supplementary Table 5**. Initial phases were calculated by the molecular replacement method with the help of the program Phaser.²⁰ The unpublished structure of AU4130/APC7354, a putative tautomerase from the thermophilic fungus *Aspergillus fumigatus* (PDB code 3C6V) served as molecular

replacement search model. The RhCC structure was improved and completed by iterative cycles of refinement in Refmac5, alternated by manual model building in Coot. The final cycles of refinement were performed using the program phenix.refine from the Phenix software suite.²¹ At the early stages of model building, a bound metal ion was identified at the surface of the RhCC trimer, positioned at the non-crystallographic 3-fold symmetry axis. Based on the metal-coordination geometry, the absence of anomalous difference electron density, and analysis of the B-factors in the later stages of refinement, the bound metal was identified as a Mg²⁺ ion.

Construction of the D106A mutant

The D106A variant of RhCC was constructed by overlap extension PCR using the Phusion DNA polymerase system. Synthetic primers 5'-G GAG ATC CAC ATC GCC GAG ACC CCC ATG GAC C-3' and 5'-G GTC CAT GGG GGT CTC GGC GAT GTG GAT CTC C-3' were used to introduce the desired mutation (underlined). These primers were used in combination with the general T7-forward and T7-reverse primers of the pET-system. The mutant gene was cloned into the pET-20b(+) expression vector as described above for the wild-type gene. The cloned gene was sequenced to verify that only the desired mutation had been introduced.

Purification, crystallization and structure determination of the D106A mutant

The D106A mutant protein was produced and purified following identical procedures as used for wild-type RhCC, except that cells were grown in auto-induction medium at 20 °C for 36 hours (instead of overnight at 30 °C) to maximize the amount of soluble protein. From 1 L of cell culture, ~30 mg of homogeneous protein was obtained, which was concentrated to 12 mg/mL in 10 mM NaH₂PO₄ buffer (pH 7.5). Dynamic light scattering analysis indicated that the mutant enzyme formed a trimer in solution. Crystallization screening resulted in growth of a single crystal (300 x 150 x 100 μm³), which was obtained after four weeks in a sitting-drop vapour diffusion experiment from condition B9 of the Morpheus screen (Molecular Dimension). With this crystal, 1.55 Å X-ray diffraction data were collected in-house at 110 K using a Microstar rotating anode X-ray source (Bruker AXS GmbH) in combination with Helios optics (Incoatec GmbH) and a MAR345dtb detector (Marresearch GmbH). X-ray data were integrated using the program XDS and further processed with the program Aimless, as described for data obtained with wild-type RhCC.²² The crystal belonged to space group C2 and contained a protein trimer in the asymmetric unit (solvent content 49%). The wild-type RhCC structure was used as a molecular replacement search model to obtain an initial structure. Subsequent model building and refinement of the D106A structure was performed with the programs Coot, Refmac5 and phenix.refine.²¹ A summary of the data collection and model refinement statistics is given in **Supplementary Table 5**.

Crystal structure analysis

Molprobit was used for validating the stereochemical quality of the models.²³ Structure-based sequence alignments were carried out using the protein structure comparison service Fold at the European Bioinformatics Institute (PDBeFold)²⁴ and visualized using Jalview.²⁵ Superpositions and calculation of C α -backbone rmsd values were performed using PDBeFold. The CheckMyMetal web server was used to validate the metal-binding site.²⁶ PyMOL (Schrödinger) was used for structure analysis and figure preparations. Coordinates for the structures of native RhCC (1.78 Å) and mutant D106A (1.55 Å) have been deposited with the Protein Data Bank (accession codes 4U5P and 4U5R, respectively).

Enzyme assays

The conversion of 4HPP by RhCC was monitored in 10 mM Na₂HPO₄ buffer (pH 7.3) at 22 °C. The reaction was initiated by the addition of 4HPP from a stock solution to yield a final concentration of 100-200 μ M in a cuvette containing 1 mL of 10 mM Na₂HPO₄ buffer (pH 7.3) and 10-20 μ g of RhCC. Stock solutions of 4HPP were made by dissolving the appropriate amount of the free acid in absolute ethanol. The crystalline free acid of 4HPP is exclusively in the enol form. UV-VIS spectra were recorded between 200 nm and 400 nm with time intervals of 60 or 120 seconds. Decrease of the absorbance at 290 nm indicates depletion of 4HPP (λ_{max} = 290 nm). The reaction performed to examine the spontaneous non-enzymatic ketonization of 4HPP was executed as described above, except that no enzyme was added to the reaction buffer.

For kinetics assays, an appropriate amount of enzyme (from a 10 mg/mL stock solution) was diluted in 50 mL of 10 mM Na₂HPO₄ buffer (pH 7.3) or 50 mL of 10 mM Na₂HPO₄ buffer (pH 8.3), after which the mixture was allowed to equilibrate for one hour at 22 °C. Substrate concentrations ranged between 0.025 and 2.75 mM. At each substrate concentration, the rate of spontaneous non-enzymatic ketonization was determined and subtracted from the observed rate of depletion of 4HPP in the presence of enzyme.

The RhCC-catalyzed conversion of 4HPP under anaerobic conditions, using glucose and glucose oxidase, was performed as follows. A small scale reaction was performed at 22 °C in a 25 mL glass reaction phial equipped with a septum and a magnetic stirrer bar, to which was added 10 mL of 10 mM Na₂HPO₄ buffer (pH 7.3) supplemented with 100 μ g/mL RhCC. Molecular oxygen was purged from the reaction buffer by applying a constant flow of nitrogen gas through the reaction phial for two hours. Subsequently, in order to ensure and maintain reaction conditions devoid of molecular oxygen, glucose was added to a final concentration of 15 mM from a 1 M stock solution in 10 mM Na₂HPO₄ buffer (pH 7.3), and glucose oxidase was added to a final concentration of 25 μ g/mL from a 2 mg/mL stock solution in H₂O. Next, 4HPP was added to the reaction mixture to a concentration of 100 μ M from a 100 mM stock solution in absolute ethanol. A 1.5 mL sample was drawn from the reaction mixture directly after the addition of the substrate, and transferred to a 1 cm quartz cuvette which was sealed air tight with a lid. The volume of the sample made

sure that the cuvette was completely filled with sample, thereby minimizing remaining air bubbles after sealing the cuvette. The reaction was monitored by UV-VIS spectroscopy, and spectra were recorded every 2 min.

HPLC detection of enzymatic products

Separation of the various products of the RhCC-catalyzed conversion of 4HPP was achieved by a reverse-phase HPLC method. HPLC was carried out on an analytical HPLC equipped with a Shimadzu LC-10 AT pump and a Shimadzu SPD-M10A diode array detector. Separation of the compounds in the reaction mixture was achieved using an Agilent Eclipse XDB-C18 column (250 × 4.6 mm, 5 µm) as a stationary phase and H₂O (eluent A) and absolute methanol (eluent B), both supplemented with 1% formic acid, as mobile phase ingredients. The gradient program was composed as follows: 0-7 minutes 100% eluent A, 7-30 minutes linear gradient to 30% eluent B, 30-31 minutes linear gradient to 100% eluent B, 31-37 minutes 100% eluent B, 37-38 minutes linear gradient to 100% eluent A, 38-47 minutes 100% eluent A. The flow rate and temperature were maintained constant, at 800 µL/min and 25 °C, respectively. The sample injection volume ranged between 20-100 µL depending on the concentration of the products in the sample. The products were detected by setting the diode array detector at 230 nm, 250 nm, and 300 nm.

The HPLC chromatogram shown in **Supplementary Fig. 6**, and the data represented by the progress curves (**Supplementary Fig. 7**), were obtained by performing a batch conversion of 4HPP by RhCC and analyzing samples taken at specific time points by the HPLC method described above. The batch conversion of 4HPP by RhCC was performed at 22 °C in 12.0 mL of 10 mM Na₂HPO₄ buffer (pH 7.3), supplemented with 40 µg/mL RhCC. The reaction was initiated by the addition of 60 µL of a stock solution of 4HPP (100 mM in absolute ethanol) yielding a final concentration of 500 µM. Samples (500 µL) were withdrawn at the appropriate time intervals and directly mixed with 500 µL of 100 mM Na₂HPO₄ buffer (pH 6.0) in an Eppendorf tube. This ensured that the pH of the sample was reduced to pH 6.0, at which the activity of RhCC is completely quenched. Samples were flash-frozen in liquid N₂ until further analysis, or directly applied to a Vivaspin centrifugal concentrator equipped with a 5000 Da molecular weight cut-off filtration membrane (Sartorius Stedim Biotech S.A., France) and centrifuged at 4000 rpm for two minutes in order to remove all protein from the sample. The flow-through was subsequently analyzed by HPLC.

Batch conversion of 4HPP by RhCC for analysis by ¹H NMR spectroscopy

Due to the inhibition of RhCC at substrate concentrations exceeding 1 mM at pH 7.3, multiple additions of substrate to a batch reaction were needed in order to achieve a sufficiently high concentration of products to allow detection by ¹H NMR spectroscopy. A batch reaction was typically performed at 22 °C in a 25 mL glass reaction phial equipped with a septum and a magnetic stirrer bar, to which was added 8.25 mL of 10

mM Na₂HPO₄ buffer (pH 9.2), supplemented with 100 µg/mL wild-type RhCC, and 750 µL D₂O. The reaction was initiated by the addition of 1 mL of 10 mM Na₂HPO₄ buffer (pH 7.3) in which 3.6 mg of crystalline 4HPP was dissolved directly before the start of the reaction. The reaction was placed under a constant flow of molecular oxygen through the septum with constant stirring of the reaction mixture, ensuring no oxygen limitation could occur during the course of the reaction. Reaction progress was determined by UV-VIS spectroscopic analysis of 20 µL samples taken from the reaction mixture. Typically, 3.6 mg of 4HPP was fully converted in 30 minutes reaction time. Once the reaction was completed, 1 mL of reaction mixture was extracted from the reaction phial and used to dissolve another amount of 3.6 mg of crystalline 4HPP, and subsequently added back to the reaction system for conversion by RhCC. This process was repeated six times in order to achieve significant concentrations of each product. Upon reaction completion after the addition of the final amount of 4HPP (final pH of the reaction mixture = 7.1), a 600 µL sample was taken and analyzed by ¹H NMR spectroscopy.

The RhCC-catalyzed conversions of 4HPP that were analyzed with HPLC and with ¹H NMR spectroscopy were conducted at different pH values (HPLC: pH = 7.3; ¹H NMR: the initial pH of the reaction mixture was set to 9.2 to take into account the drop in pH upon multiple additions of substrate 4HPP). However, an HPLC chromatogram recorded from the reaction mixture obtained at pH 9.2 showed high similarity with the HPLC chromatogram recorded from the reaction mixture obtained at pH 7.3 demonstrating that identical products are formed at both pH values. The only exception is that the HPLC chromatogram as well as the ¹H NMR spectrum recorded from the reaction mixture obtained at pH 9.2 indicated the absence of the keto-form of 4HPP whereas the HPLC chromatogram recorded from the reaction mixture obtained at pH 7.3 did indicate trace amounts of the keto-form of 4HPP. Control experiments (at both pH 7.3 and pH 9.2) demonstrated that in the absence of enzyme 4HPP undergoes ketonization without detectable formation of oxygenation products.

Structure determination of products 1 and 2 by cryoNMR spectroscopy

The structure of products **1** (3-hydroxy-3-(4-hydroxyphenyl)-pyruvate) and **2** (4-hydroxybenzaldehyde) were determined by cryoNMR spectroscopy (Spinnovention Analytical B.V., Oss, The Netherlands). An amount of 10 mg of 4HPP was converted into products at 22 °C in a reaction mixture consisting of 70 mL of 10 mM Na₂HPO₄ buffer (pH 7.3) supplemented with 30 µg/mL RhCC. Reaction progress was monitored by UV-VIS spectroscopic analysis. After the reaction had reached completion (30 min), RhCC was removed using a Vivaspin centrifugal concentrator equipped with a 5000 Da molecular weight cut-off filtration membrane. The protein free flow-through was collected and subsequently freeze-dried overnight. The resulting lyophilisate was stored on ice prior to further analysis. CryoNMR spectra were recorded using a Bruker 500 MHz Avance III system (Bruker Biospin, Rheinstetten, Germany) equipped with a 5 mm cryo probe,

CPTCI (^1H - $^{13}\text{C}/^{15}\text{N}/^2\text{H}$ + Z-gradients). Separation of the reaction products was achieved using the HPLC protocol described above, carried out on an Agilent 1200 system (Agilent Technology, Santa Clara, CA, USA) equipped with an Agilent Eclipse XDB-C18 column (250 $\times$ 4.6 mm, 5 μm). Fractions containing products **1** and **2** were collected, freeze-dried, and dissolved in D_2O prior to analysis. NMR spectra were recorded at 30 $^\circ\text{C}$, and depending on the complexity and concentration of each product, a set of NMR spectra (^1H -1D, COSY, ^1H , ^{13}C -HSQC, ^1H , ^{13}C -HMBC) were recorded. Full spectroscopic data are given in the **Supplementary Information**.

Incubations with $^{16}\text{O}_2$, $^{18}\text{O}_2$ and H_2^{18}O and high-resolution LC-MS and MS/MS analyses

High-resolution LC-MS and MS/MS analyses were performed on an LTQ-Orbitrap XL (Thermo Scientific, Bremen, Germany) coupled to a Shimadzu Prominence UFLC system (Shimadzu Corp., Kyoto, Japan) equipped with an Agilent Eclipse XDB-C18 column (250 $\times$ 4.6 mm, 5 μm), and using the HPLC method described above.

Samples of a reaction performed in aerated buffer were prepared as follows. A reaction was performed at 22 $^\circ\text{C}$ in a quartz cuvette containing 1 mL of 10 mM Na_2HPO_4 buffer (pH 7.3) supplemented with 50 μg RhCC. The reaction was initiated by the addition of 7.5 μL of a stock solution of 4HPP (100 mM in absolute ethanol), yielding a final concentration of 500 μM , and the reaction was monitored in a UV-VIS spectrophotometer. After a reaction time of ~ 5 min, at which $\sim 50\%$ of the substrate 4HPP had been consumed, the enzyme was removed using a Vivaspin column (as described above), and a 100 μL sample of the flow through was directly injected into the HPLC system and analyzed by high-resolution LC-MS and MS/MS.

Samples of a reaction performed in the presence of H_2^{18}O were prepared as described above, with the following modification. The reaction buffer (1 mL) was prepared by mixing 500 μL 20 mM Na_2HPO_4 buffer (pH 7.3) with 500 μL H_2^{18}O .

Reactions in the presence of $^{18}\text{O}_2$ (and in the absence of $^{16}\text{O}_2$) were performed using a closed, air tight, reaction system (**Supplementary Fig. 8**, arrow 1) that was built of metal tubes, fittings and valves (Swagelok Company, Solon, Ohio, USA) to which a reaction container (**Supplementary Fig. 8**, arrow 2), a vacuum pump and a small gas cylinder containing $^{18}\text{O}_2$ (**Supplementary Fig. 8**, arrow 3) were attached. The reaction container consisted of two small flasks connected to one another by a glass tube. A single outlet (**Supplementary Fig. 8**, arrow 4) allowed it to be attached to the air tight reaction system, by which the reaction buffer could be degassed and subsequently equilibrated in $^{18}\text{O}_2$. A reaction in the presence of $^{18}\text{O}_2$ was performed as follows. The reaction buffer (3 mL, 10 mM Na_2HPO_4 buffer, pH 7.3) supplemented with RhCC (50 $\mu\text{g}/\text{mL}$) was added to one of the flasks of the reaction container while 4HPP was added to the second flask in an amount by which it would reach a final concentration of 500 μM in the reaction mixture. The amount of 4HPP was added from a stock solution in absolute ethanol. The ethanol

was evaporated prior to connection of the reaction container to the air tight reaction system. Magnetic stirrer bars were added to each flask to allow for adequate mixing. Once assembled, the reaction system was placed under vacuum (**Supplementary Fig. 8**, arrow 5) for 15 minutes in order to remove air from the system and to purge the reaction buffer containing RhCC from dissolved gasses, including $^{16}\text{O}_2$. Subsequently, a valve that connects the vacuum pump to the reaction system was closed and $^{18}\text{O}_2$ (**Supplementary Fig. 8**, arrow 3) was allowed to fill the reaction system until it had reached atmospheric pressure. The reaction buffer containing RhCC was equilibrated with $^{18}\text{O}_2$ for 30 minutes prior to reaction initiation. The reaction was started by inverting the reaction container so that the reaction buffer containing RhCC could flow into the second reaction flask containing solid 4HPP, allowing 4HPP to dissolve in the reaction buffer. After 10-15 minutes reaction time, the reaction container was disconnected from the reaction system and a sample of the reaction mixture was prepared for LC-MS and MS/MS analysis as described above.

Product **3** (oxalic acid) could be detected by high-resolution LC-MS using the HPLC method described above. Due to the high polarity of **3**, it ended up in the void volume of the reverse phase C18 column, but was thus separated from the other reaction products. In order to further establish the identity of **3** as oxalic acid, a second HPLC method was developed using an Waters XBridge amide column (4.6 x 150 mm, 3.5 μm ; Waters Corporation, Milford, MA, USA). An isocratic flow of 20 mM NH_4HCO_3 buffer (pH 7.0) containing 50 % (v/v) acetonitrile using a flow rate of 0.5 mL/minute allowed oxalic acid to be retained on the column resulting in an appropriate retention time. The composition of the eluent allowed for the detection of oxalic acid on the LTQ Orbitrap XL system.

RESULTS

Purification of RhCC

By sequence similarity, the hypothetical protein YP_702633.1 (referred to as RhCC) of *Rhodococcus jostii* RHA1 most resembles the members of the *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD) family within the tautomerase superfamily.²⁷⁻²⁹ Despite this relationship to enzymes with known activity, the substrate preference of RhCC is anything but clear. To determine its structure and function, RhCC was produced as a C-terminal His₆-tagged protein in *E. coli* BL21(DE3) and purified to near homogeneity using a Ni-based immobilized metal affinity chromatography procedure (**Supplementary Results, Supplementary Fig. 1**). Analysis of RhCC by electrospray ionization-mass spectrometry (ESI-MS) showed one major protein species with a mass of 18489 ± 1 Da. A comparison of this value to the calculated subunit mass (18620 Da) indicates that the translation initiating methionine is removed during posttranslational processing, which results in a mature protein of 145 amino acids with an N-terminal proline. The native molecular mass of RhCC was estimated by dynamic light scattering to be ~53 kDa, which suggests that the native protein exists as a homotrimer in solution.

Overall structure of native RhCC

To enhance our understanding of the relationship between RhCC and members of the *cis*-CaaD family within the tautomerase superfamily, we have determined the crystal structure of native RhCC. The RhCC structure was solved at 1.78 Å resolution by molecular replacement methods and refined to R and R_{free} values of 15.2% and 18.7%, respectively. The asymmetric unit in the crystal contains one homotrimer and has a solvent content of 47%. The overall fold of the individual subunits of the trimer, and their mode of association, are very similar to those observed in the structures of *cis*-CaaD and Cg10062 (a *cis*-CaaD homologue from *Corynebacterium glutamicum*, PDB entry 3N4D), despite the low sequence identity between RhCC and these two proteins of 14-16% (**Figs. 1 and 2, Supplementary Fig. 2**).^{30,31} Residues 1-123 form the core of the individual RhCC subunits, containing the two β - α - β structural motifs which are characteristic of the tautomerase superfamily fold.¹³ Strands $\beta 1$, $\beta 2$, $\beta 5$ and $\beta 6$ of the two β - α - β structural motifs associate into a four-stranded β -sheet within each RhCC subunit. The three additional β -strands ($\beta 3$, $\beta 4$ and $\beta 7$) are important for subunit-subunit interactions, by complementing the central β -sheets of neighboring subunits in the RhCC trimer. Residues 124-142 form a short α -helical extension at the C-terminus of each RhCC subunit, which is also present in the *cis*-CaaD and Cg10062 structures, but with more structural divergence as compared to the N-terminal cores of these proteins. $\text{C}\alpha$ -backbone superpositions of the RhCC trimer with its structural homologs resulted in a root-mean-square deviation (rmsd) of 2.5 Å with *cis*-CaaD (PDB entry 2FLZ, 334 equivalent $\text{C}\alpha$ -atoms) and 2.4 Å with Cg10062 (PDB entry 3N4D, 334 equivalent $\text{C}\alpha$ -atoms). Significantly higher structural similarity was observed between RhCC and the structure of a putative tautomerase from the thermophilic fungus *Aspergillus fumigatus* (PDB code 3C6V), which share a sequence identity of 40% (**Fig. 2**) and an rmsd in $\text{C}\alpha$ -backbone conformation of 1.4 Å (for 398 equivalent $\text{C}\alpha$ atoms within the trimers). The implications of this structural similarity for the function of RhCC are unknown due to lack of functional data for the *A. fumigatus* protein.

Metal binding site

Unlike the structures of *cis*-CaaD, Cg10062 and the *A. fumigatus* protein, the RhCC crystal structure reveals the presence of a metal binding site on the surface of the trimer. The metal binding site is positioned at the non-crystallographic 3-fold symmetry axis, near the C-terminal ends of the three trimer-related $\beta 6$ strands, with residue Asp106 from each subunit acting as metal-coordinating ligand (**Supplementary Fig. 3a**). Three carboxylate oxygen atoms of the trimer-related aspartate residues and three water molecules coordinate the metal ion in an octahedral geometry. The high resolution X-ray data allowed an accurate determination of oxygen-metal bond distances of 2.1-2.2 Å, which, together with the observed coordination geometry, is consistent with the presence of either Mg^{2+} , Zn^{2+} or Mn^{2+} . Additional analysis of B-factors and anomalous electron density maps, complemented by validation with the CheckMyMetal web server, strongly suggest

Figure 1. Overall structure of RhCC and its comparison with *cis*-CaaD. **(a)** Ribbon representation (green) of a single RhCC subunit. Pro-1 is drawn as sticks (yellow), the Mg^{2+} ion is colored violet. Secondary structure elements are labeled. **(b)** Ribbon representation (blue) of a single *cis*-CaaD subunit (PDB entry 2FLZ). **(c)** Stereo view of the RhCC trimer in ribbon representation, with the three subunits drawn in different colors. The Pro-1 residues are drawn as sticks (yellow), while the Mg^{2+} ion (violet sphere) at the centre of the trimer is shown together with its coordinating ligands: three Asp-106 residues (sticks, different colors) and three water molecules (red spheres). Also shown is the difference electron density (grey mesh, $F_o - F_c$) of the unidentified ligand bound at the three putative active sites.

that the bound metal ion is Mg^{2+} .²⁶ The distance between the metal ion and the N-terminal proline residues is 15 Å.

Putative active site

Based on the structural similarities with *cis*-CaaD and Cg10062, we expect the RhCC trimer to contain three active sites, defined by the locations of the N-terminal proline residues (**Fig. 1** and **Supplementary Fig. 2**). The observation of extra electron densities near the three N-terminal proline residues indicates the presence of a bound ligand (**Figs. 1c** and **3**). Unfortunately, the features of the electron densities are insufficiently detailed to allow identification of the bound ligand. Residues surrounding the unidentified ligand in each RhCC subunit are Pro-1; Tyr-28, Tyr-31, Val-32 and Leu-34 ($\alpha 2$ helix); His-68

Figure 2. Structure-based multiple sequence alignment of RhCC, 3C6V, *cis*-CaaD and Cg10062. The alignment has been produced using the programs T-coffee²⁴ and JalView.²⁵ Residues belonging to the common structural core of the proteins are shown in capital fonts. The secondary structure topology is derived from the RhCC structure. Red and black asterisks anotate residues surrounding the unidentified, bound ligand at the putative active site of RhCC, and the active site residues of *cis*-CaaD, respectively (see also **Figure 3**). The metal-coordinating residue of RhCC (Asp-106) is labeled with a red "M".

($\beta 5$ strand); Arg-71 and Met-73 ($\beta 5$ - $\alpha 2$ loop); Val-78 and Met-82 ($\alpha 2$ helix). Although the location of this putative active site in a RhCC subunit coincides well with that observed in the *cis*-CaaD and Cg10062 subunits, the surrounding residues are very different. In *cis*-CaaD and Cg10062, five conserved active site residues, in addition to the N-terminal proline, were shown to be critical for the hydrolytic dehalogenation reaction catalyzed by these enzymes (in *cis*-CaaD these residues are His-28, Arg-70, Arg-73, Tyr-103' and Glu-114, with the apostrophe signifying that the residue comes from a neighboring subunit). None of these five residues are conserved in RhCC (**Fig. 2**). In addition, the conformation of the $\beta 5$ - $\alpha 2$ loop, which is a critical component of the active sites in *cis*-CaaD and Cg10062, is significantly different in RhCC, due to the insertion of two extra residues. In *cis*-CaaD and Cg10062, the $\beta 5$ - $\alpha 2$ loop contains the two catalytically important arginine residues, Arg-70 and Arg-73. In RhCC, the single arginine residue present in the $\beta 5$ - $\alpha 2$ loop, Arg-71, takes up a position that is somewhat in between those of Arg-70 and Arg-73 in the *cis*-CaaD structure. However, the side chain of Arg-71 in RhCC points away from the putative active site and towards the C-terminal helix, where it is stabilized by a salt bridge interaction with Glu-128 (**Fig. 3**). Overall, the putative active site in RhCC seems more hydrophobic than the active sites in *cis*-CaaD and Cg10062. Also the direct environment of the N-terminal proline in RhCC is more hydrophobic than in *cis*-CaaD and Cg10062, with Ile-69 and Trp-113 in RhCC occupying similar 3D positions as Tyr-103' and Glu-114 in *cis*-CaaD and Cg10062. The prolyl nitrogen atom of Pro-1 in RhCC forms two hydrogen bond interactions with surrounding residues: one direct hydrogen bond with the main chain carbonyl oxygen of

Tyr-38 and a water-bridged hydrogen bond with the main chain amide nitrogen atom of Ile-69.

Unusual enzyme activity

Having established that RhCC belongs to the tautomerase superfamily, the enzyme was assayed for activity towards several available substrates of previously characterized members of this superfamily. Unlike its structural homologs *cis*-CaaD and Cg10062, RhCC does not catalyze the hydrolytic dehalogenation of *cis*-3-chloroacrylic acid (or the corresponding *trans*-isomer), consistent with the lack of conservation of key catalytic residues of *cis*-CaaD and Cg10062 in RhCC (**Figs. 2 and 3**). However, during the course of activity screening, RhCC was found to accept 4-hydroxyphenylenolpyruvate (4HPP) as a substrate. 4HPP is a known substrate of the well-characterized tautomerase superfamily member macrophage migration inhibitory factor (MIF), also known as phenylpyruvate

Figure 3. Stereo view of the putative active site region of RhCC and its comparison with the active site of inactivated *cis*-CaaD. The putative active site region of RhCC is shown as ball-and-sticks with yellow carbons, whereas the active site of inactivated *cis*-CaaD (PDB entry 2FLT) is shown as ball-and-sticks with white carbons. The covalently bound (*R*)-2-hydroxypropanoate adduct on Pro-1 of *cis*-CaaD, which results from the inactivation of the enzyme by oxirane-2-carboxylate, is shown. Also shown, as a grey mesh, is the difference electron density ($F_o - F_c$) for the unidentified bound ligand in RhCC (1.78 Å resolution, contoured at 2.5σ). Oxygen and nitrogen atoms are colored red and blue, respectively. Residues of RhCC are indicated with black, underlined labels, residues of *cis*-CaaD with grey labels. The apostrophe in Y103' signifies that this residue belongs to a neighboring subunit.

tautomerase. MIF catalyzes the enol-keto tautomerization of 4HPP, utilizing its amino-terminal proline (Pro-1) as the catalytic base (**Supplementary Scheme 1**).³²

To probe the nature of the reaction catalyzed by RhCC, the enzyme was incubated with 4HPP and the conversion was monitored by UV-VIS spectroscopy (**Fig. 4**). The spectra indicate that 4HPP ($\lambda_{\text{max}} = 290 \text{ nm}$) is rapidly consumed by RhCC during the course of the reaction (**Fig. 4a,b**). However, to our surprise, the RhCC-catalyzed conversion of 4HPP did not result in formation of the keto-form of 4HPP, but resulted in the formation of a mixture of products with a characteristic UV-VIS spectrum, which is clearly different from that of the keto-form of 4HPP (**Fig. 4c**). Control experiments indicated that the keto-form of 4HPP, which was prepared by spontaneous non-enzymatic ketonization of 4HPP, is not accepted as a substrate by RhCC. Hence, although RhCC has the catalytic capacity to convert 4HPP, it does not function as an enol-keto tautomerase.

By monitoring the depletion of 4HPP by UV-VIS spectroscopy we were able to gain insight into the kinetics of the reaction catalyzed by RhCC. The enzyme was found to be inactive below pH 6.0 whereas at higher pH values the activity of the enzyme increases. The catalytic efficiency (in terms of apparent k_{cat}/K_m) was estimated to be $2.5 \cdot 10^3 \text{ M}^{-1} \text{ s}^{-1}$ at pH 7.3 and $4.7 \cdot 10^4 \text{ M}^{-1} \text{ s}^{-1}$ at pH 8.3. The enzyme showed inhibition at increasing substrate concentrations, but the exact type of inhibition could not be established; hence the individual values of k_{cat} and K_m could not be determined. The highest observed reaction rates were 1.4 s^{-1} and 6.6 s^{-1} at pH 7.3 and pH 8.3, respectively (**Supplementary Figs. 4a,b**).

Identification of the products of the RhCC-catalyzed conversion of 4HPP

To identify the products of the RhCC-catalyzed conversion of 4HPP, we performed the reaction at semi-preparative scale and under constant flow of molecular oxygen. After the reaction was completed, as indicated by UV-VIS spectroscopic analysis, a sample was taken from the reaction mixture and analyzed by ^1H NMR spectroscopy. The spectrum indicated the formation of a rather complex mixture of products (**Supplementary Fig. 5**). The substrate 4HPP was completely consumed as its characteristic signals at 7.68, 6.94, and 6.36 ppm (in D_2O) could not be observed. The keto form of 4HPP was present in the reaction mixture (7.16, 6.91, and 4.03 ppm). However, its presence is not caused by the action of RhCC but by non-enzymatic ketonization of 4HPP as verified by control experiments (**Fig. 4b,d**).

To identify the products that give rise to the major peaks observed in the ^1H NMR spectrum, a reverse phase HPLC method was set up with the aim to separate and isolate the various reaction products for full characterization by NMR and mass spectroscopy (**Supplementary Fig. 6**). The eluents in the HPLC-method were chosen such that they are compatible with high-resolution LC-MS and MS/MS analysis. This enabled us to determine the exact mass and elemental composition of the major products, while product ion scans at high resolution provide additional insight into the molecular structures. The exact mass

Figure 4. UV-VIS spectroscopic analysis of the unusual enzymatic activity of RhCC. **(a)** UV-VIS spectra monitoring the RhCC-catalyzed conversion of 4HPP (100 mM) in aerated 10 mM Na₂HPO₄ buffer (pH 7.3). Spectra were recorded every two minutes. These spectra clearly indicate that the enol-form of 4HPP ($\lambda_{\text{max}} = 290$ nm) is rapidly consumed during the course of the reaction, resulting in a product mixture with a characteristic UV-VIS spectrum ($\lambda_{\text{max}} = 257$ nm). **(b)** UV-VIS spectra monitoring the spontaneous non-enzymatic ketonization of 4HPP (100 mM) in aerated 10 mM Na₂HPO₄ buffer (pH 7.3). Spectra were recorded every two minutes. **(c)** A comparison of the UV-VIS spectrum of the keto-form of 4HPP (in red) and that of the reaction mixture after completion of the reaction of RhCC with 4HPP (in blue). The keto-form of 4HPP was obtained by spontaneous non-enzymatic ketonization of 4HPP (200 mM) in 10 mM Na₂HPO₄ buffer, pH 7.3 (120 min reaction time). The reaction mixture of the RhCC-catalyzed conversion of 4HPP was obtained by the addition of 4HPP (200 mM) to 10 mM Na₂HPO₄ buffer, pH 7.3, supplemented with RhCC (2 mM). The spectrum was recorded after 10 min reaction time. **(d)** UV-VIS spectra monitoring the RhCC-catalyzed conversion of 4HPP (100 mM) under anaerobic conditions in 10 mM Na₂HPO₄ buffer (pH 7.3) in the presence of 15 mM glucose and 25 mg/mL glucose oxidase. Spectra were recorded every two minutes.

data related to the most prominent peaks in the HPLC chromatogram and the NMR data of lyophilized fractions, corresponding to those peaks, allowed us to identify 3-hydroxy-3-(4-hydroxyphenyl)-pyruvate (**1**, **Supplementary Fig. 5**) and 4-hydroxybenzaldehyde (**2**) as two major products of the RhCC-catalyzed conversion of 4HPP (see **Supplementary Results**). Products **1** and **2** eluted at 4.3 and 31.5 min, respectively, and ¹H NMR

data obtained from these fractions, after lyophilization, exactly match with the most intensive signals in the ^1H NMR spectrum recorded from the crude reaction mixture (**Supplementary Fig. 5**).

The products responsible for peaks A and B in the HPLC chromatogram of the crude reaction mixture (**Supplementary Fig. 6**) could not be identified. The lyophilized fraction corresponding to peak A did not provide coherent spectroscopic data indicating that the product responsible for peak A is unstable. The ^1H NMR spectrum of the lyophilized fraction corresponding to peak B showed signals at ~ 6.6 , ~ 6.8 and 6.95 ppm, which match with minor absorptions in the ^1H NMR spectrum of the crude reaction mixture (**Supplementary Fig. 5**). High-resolution LC-MS and MS/MS data related to peaks A and B (**Supplementary Tables 1a** and **1b**) also did not allow us to unambiguously identify the products corresponding to these peaks. We decided not to pursue identification of the products responsible for peaks A and B any further since progress curves, that were generated for all major products that are visible in the HPLC chromatogram (**Supplementary Fig. 7b-e**), show an initial lag phase for the formation of the products corresponding to peaks A and B indicating they are not initial reaction products. In contrast, the progress curves of products **1** and **2** do not show an initial lag phase, which suggests that both **1** and **2** are initial products formed during the reaction or both originate from a relatively quickly formed unstable intermediate, which undergoes rapid decay. Accordingly, the progress curve of substrate 4HPP shows initial linear decrease over time (**Supplementary Fig. 7a**).

Importantly, the identity of **2** implied the presence of yet another product based on a two-carbon unit, originating from the C1 and C2 atoms of 4HPP, which are evidently missing in product **2**. Indeed, high-resolution LC-MS data revealed the presence of oxalic acid (**3**, **Scheme 1**) in the reaction mixture, the identity of which could be confirmed by analytical high-resolution LC-MS using two different LC protocols and oxalic acid as an authentic standard (**Supplementary Table 1c**). The ability to detect **3** by high-resolution LC-MS allowed us to generate a progress curve for this product (**Supplementary Fig. 7f**), which, consistent with that for product **2**, shows no initial lag phase.

The identification of **1**, **2** and **3** as the major products of the RhCC-catalyzed conversion of 4HPP indicates that this enzyme catalyzes the incorporation of at least one atom of oxygen into substrate 4HPP, the C-C double bond of which is fully cleaved during the formation of **2** and **3**. This immediately raises questions about the source of the oxygen introduced into the substrate and the type of reaction that is catalyzed by RhCC.

RhCC is an oxygenase

As the conversion of 4HPP by RhCC is performed in aerated buffer, either water or molecular oxygen (O_2) is the source of incorporated oxygen in products **1-3**. To elucidate the origin of the incorporated oxygen, reactions of RhCC with 4HPP were performed in the presence of H_2^{18}O or $^{18}\text{O}_2$, using an experimental setup that was built in house (**Supplementary Fig. 8**), and the resulting product mixtures were analyzed by high-resolution LC-MS and MS/

Table 1. ^{18}O -incorporation in products **1-3** as determined by high-resolution LC-MS and MS/MS analysis of reaction mixtures obtained from RhCC-catalyzed conversion of 4HPP in aerated buffer supplemented with H_2^{18}O or buffer equilibrated with $^{18}\text{O}_2$.

Product	$\text{H}_2^{18}\text{O}^a$	$^{18}\text{O}_2^b$
3-hydroxy-3-(4-hydroxyphenyl)-pyruvate (1)
4-hydroxybenzaldehyde (2)
Oxalic acid ^c (3)

^a The reaction was performed in 10 mM Na_2HPO_4 buffer, pH 7.3, containing 50% (v/v) H_2^{18}O . ^b The reaction was performed in 10 mM Na_2HPO_4 buffer, pH 7.3, which was degassed and subsequently placed under $^{18}\text{O}_2$ at atmospheric pressure in a closed reaction system, allowing the buffer to equilibrate with $^{18}\text{O}_2$ prior to reaction initiation and to be kept under an $^{18}\text{O}_2$ atmosphere during the course of the reaction. ^c Control experiments (see **Supplementary Results**) have shown that under the stated reaction conditions no ^{18}O is incorporated in oxalic acid by spontaneous exchange of its carboxylate-oxygens with H_2^{18}O .

MS. High-resolution MS spectral analysis showed that the incorporated oxygen atom in product **1** originates exclusively from molecular oxygen (**Table 1**). Additionally, product ion spectra allowed us to identify the C3-carbon atom of **1** as the site of oxygenation (**Supplementary Table 2**), thus providing insight into the nature of the reaction catalyzed by RhCC. Interestingly, product **3** shows oxygen incorporation both from water and molecular oxygen, while product **2** shows no ^{18}O -isotope labeling under any of the tested conditions (**Table 1**, **Supplementary Table 3**). It was anticipated that product **2** would exchange its aldehyde oxygen rapidly with solvent-derived oxygen under the conditions of LC-MS analysis (pH 1.8). This was validated by incubating compound **2** in H_2^{18}O at pH 1.8. The anticipated almost complete exchange-in of solvent-derived oxygen was detected by high-resolution MS for compound **2** when incubated for 3 min in H_2^{18}O at pH 1.8 (see **Supplementary Results**).

To further establish that RhCC catalyzes an O_2 -dependent reaction with 4HPP, the reaction was performed under anaerobic conditions. Reactions were performed in airtight cuvettes and oxygen was removed from the buffer by first flushing with nitrogen gas and the subsequent addition of glucose and glucose oxidase prior to reaction initiation. Reaction progress was monitored by UV-VIS spectroscopy (**Fig. 4d**). The results clearly show that the unusual catalytic activity of RhCC is dependent on molecular oxygen, with no product formation ($\lambda_{\text{max}} = 257 \text{ nm}$) or rapid loss of substrate detected under anaerobic conditions. Notably, in the absence of molecular oxygen (like in its presence),

Scheme 1. Oxidative substrate cleavage catalyzed by RhCC. A) A possible mechanism for the decomposition of hydroperoxide **4** as proposed by Jefford et al.³³ B) A proposed mechanism for the formation of **2** and **3**, based on hydroperoxide **4** as the presumed initial enzymatic oxygenation product formed.

RhCC does also not catalyze the ketonization of 4HPP, since the rate of consumption of 4HPP is the same as that observed for the non-enzymatic ketonization of 4HPP (**Fig. 4b,d**). Separate control experiments have shown that glucose or glucose oxidase do not affect the substrate itself or the activity of RhCC.

Mutagenesis of the metal binding site

In order to investigate the importance of the bound Mg^{2+} ion for the oxygenase activity of RhCC, a single site-directed mutant was constructed in which the metal-coordinating residue Asp106 (**Supplementary Fig. 3a**) was replaced by an alanine. This mutation is expected to remove the metal binding site, leading to the loss of the Mg^{2+} ion. It was found that the oxygenase activity of the purified D106A mutant (apparent $k_{\text{cat}}/K_m = 2.4 \cdot 10^3 \text{ M}^{-1} \text{ s}^{-1}$ at pH 7.3) was similar to that of wild-type RhCC, indicating that the mutation

had no significant effect on catalytic efficiency. Analysis of the purified D106A mutant by dynamic light scattering revealed that the mutant protein, like wild-type RhCC, exists as a homotrimer in solution. The structural consequences of the D106A mutation were examined by determining the crystal structure of the mutant enzyme to a resolution of 1.55 Å. The crystallographic R-factor for the final model is 16% ($R_{\text{free}} = 17.4\%$). A comparison of the D106A structure with that of wild-type RhCC reveals only local and minor effects resulting from the mutation, leaving the overall architecture otherwise unchanged (**Supplementary Fig. 3c**). The α -carbon backbones of the wild-type and D106A mutant enzyme superimpose with an rmsd of 0.34 Å. An additional short α -helix of about ten residues is visible at the C-termini of the D106A trimer, as a result of better defined electron density. The only other difference between the two structures is the loss of the metal binding site on the surface of the D106A trimer (**Supplementary Fig. 3a,b**). As the D106A mutant exhibits full oxygenase activity but has no metal ion bound, we conclude that the oxygenase activity of RhCC is metal-independent.

DISCUSSION

The radical-initiated non-enzymatic autoxidation of 4HPP has been studied in detail by Jefford et al., which has led to the identification of several reaction products including **2** and **3**.³³ The authors proposed that these are formed by the decomposition of 3-hydroperoxy-3-(4-hydroxyphenyl)-pyruvate (**4**, **Scheme 1A**) as the initial product of 4HPP autoxidation, the inherent instability of which prohibited its isolation. These findings may suggest that the formation of **2** and **3** during the RhCC-catalyzed conversion of 4HPP could also be the result of a decomposition reaction of initially formed hydroperoxide **4**. However, the mechanism proposed by Jefford et al.³³ for the formation of **2** and **3** (**Scheme 1A**) is inconsistent with the results obtained from our ^{18}O -labelling experiments performed with RhCC, since this mechanism would result in ^{18}O -labeled **2** in the presence of $^{18}\text{O}_2$ and would not result in ^{18}O -labeled **3** in the presence of H_2^{18}O , which is the opposite of our findings with RhCC. Even though control experiments validated that any ^{18}O transferred to the aldehyde oxygen of **2** would be washed out under the conditions of LC-MS analysis, the enzymatic formation of ^{18}O -labeled **3** in incubations with substrate in H_2^{18}O (exchange-in of solvent-derived oxygen for product **3** does not occur under the conditions used) suggests that products **2** and **3** may be formed by a different mechanism in the RhCC-catalyzed conversion of 4HPP compared to those proposed for the non-enzymatic autoxidation of 4HPP.

We therefore suggest an alternative mechanism, starting with presumed intermediate **4** or its deprotonated peroxyanionic analog, which could explain the formation of the enzymatic products **2** and **3** (**Scheme 1B**). In the first step of the proposed mechanism, intramolecular attack of the peroxyanion on the carbonyl function of **4** leads to the formation of 4-(4-hydroxyphenyl)-3-oxido-1,2-dioxetane-3-carboxylate (**6**). Secondly, the oxyanion of dioxetane **6** may attack the C1-carbon of the dioxetane-ring, resulting in the formation

of 2-dioxidanide-3-(4-hydroxyphenyl)oxirane-2-carboxylate (**7**), by which the peroxide function of **4** has migrated from the C3- to the C2-position. Subsequently, hydrolysis of the epoxide and protonation of dioxidanide **7** results in the formation of 2-hydroperoxy-2-hydroxy-3-(4-hydroxyphenyl)-3-oxidopropanoate (**8**). Finally, the shift of one of the electron pairs of the 3-oxido function of **8** results in carbon-carbon bond cleavage and the concomitant expulsion of a hydroxyl-ion (or water) from the 2-hydroperoxy group, leading to the formation of **2** and **3**. This proposed mechanism for the formation of **2** and **3** is consistent with the results obtained from our ^{18}O -labelling experiments performed with RhCC.

We propose that product **1** may be formed as a result of the reaction between **4** and 4HPP (**Supplementary Scheme 2**). There is precedent in literature for reactions between organic hydroperoxides and alkenes, resulting in the formation of epoxides and alcohols.³⁴ Accordingly, **4** may react with 4HPP resulting in the formation of **1** and 2-hydroxy-3-(4-hydroxyphenyl)oxirane-2-carboxylate (**5**). Due to the hydroxyl-substituent on the epoxide-ring, **5** is likely to rapidly isomerize resulting in the formation of a second molecule of **1**. This mechanism is in agreement with the observation that the formation of **1** does not involve water and in the presence of $^{18}\text{O}_2$ exclusively results in an ^{18}O -labeled C3-hydroxyl group.

Enzymes that catalyze reactions between organic substrates and molecular oxygen generally are dependent on cofactors such as flavin, pterin or a coordinated transition metal cation.³ However, our crystallographic studies show that RhCC is devoid of any cofactor. The metal ion that is present in RhCC was identified as Mg^{2+} , which, as a redox-inactive metal ion, cannot facilitate the reaction between 4HPP and molecular oxygen.^{35,36} Furthermore, the distance between the Mg^{2+} ion and Pro-1 in the presumed active site is over 14 Å, which would preclude a direct involvement of Mg^{2+} in catalysis. Finally, removal of the Mg^{2+} ion upon mutagenesis of its binding site has no significant effect on the oxygenase activity of RhCC. The combination of these observations lead us to identify RhCC as a rare example of a cofactor-independent oxygenase.

In recent years, a few enzymes have been identified as cofactor-independent oxygenases.⁴ The catalytic mechanisms of these enzymes seem to have two features in common, being general base-catalyzed abstraction of a proton from the substrate, thereby

Scheme 2. Oxygenation of substrate 4HPP catalyzed by RhCC. A proposed mechanism for the RhCC-catalyzed oxygenation of 4HPP to yield hydroperoxide **4**.

creating a substrate anion, in combination with substrate-assisted catalysis.^{4,7,37} Along the reaction pathway, the substrate anion is converted into a radical by single-electron transfer to dioxygen creating a caged “substrate radical–superoxide anion” pair. The ability of the substrate to stabilize the remaining unpaired electron is a determining feature. In effect, the substrate, being able to generate a stable radical, facilitates the activation of molecular oxygen and acts as a cofactor in the reaction. Via this mechanism, the substrate is initially converted into an organic hydroperoxide, which by subsequent reactions leads to the formation of the final product or products.

These common mechanistic features of cofactor-independent oxygenases can be applied to the RhCC-catalyzed conversion of 4HPP to explain the formation of the presumed hydroperoxide intermediate **4** (**Scheme 2**). We propose that the reaction is initiated by proton abstraction from the 2-hydroxyl group of 4HPP, mediated by an active site general base residue, which results in the formation of a carbanion at the C3-carbon of the substrate. Subsequently, the transfer of a single electron from the carbanion to an appropriately positioned dioxygen molecule results in the formation of a triplet-state caged-radical pair, consisting of a superoxide anion and a resonance-stabilized radical cation of 4HPP (**Scheme 2, Fig. 5**). Finally, spin-inversion of one of the unpaired electrons of the superoxide anion in the caged-radical pair yields a singlet-state, which allows C–O bond formation between 4HPP and molecular oxygen to take place, yielding hydroperoxide **4** as the initial product.

Further support for this mechanism comes from activity assays with the substrate analogue phenylenolpyruvate (**9**) (**Supplementary Chart 1**). While the 2-hydroxyl group of **9** could be deprotonated, the lack of a *p*-hydroxyl substituent on the ring makes this compound less suitable for stabilizing a radical cation that would be formed upon single-electron transfer from the substrate anion to dioxygen. Hence, **9** appears to be an interesting probe to investigate the requirement of a feature termed “substrate-assisted

Figure 5. Diagram showing the spin-states of the valence electrons that initially form the lone electron pair at the carbon atom of 4HPP[−] that bears the negative charge and both unpaired electrons of dioxygen during the course of the proposed mechanism.⁵ Transfer of a single electron from the lone pair at the anionic position of 4HPP[−] to dioxygen yields a triplet-state caged-radical pair. Subsequently, one of the electrons of the superoxide anion in the caged-radical pair undergoes spin-inversion, which is a crucial step in the reaction mechanism, and causes the caged-radical pair to be converted into the singlet-state. The two remaining unpaired electrons in the singlet-state caged-radical pair now have an opposite spin-sign, which allows them to readily form a set of paired electrons. This results in C–O bond formation, yielding **4** as the initial product of the enzymatic oxygenation reaction.

catalysis" in the reaction mechanism. As anticipated, RhCC was found not to accept **9** as substrate, fully consistent with the proposed mechanism shown in **Scheme 2**.

In summary, we provide evidence for cofactor-independent oxygenation of 4HPP within a protein architecture (i.e., the β - α - β -fold) generally employed to catalyze tautomerization, dehalogenation or decarboxylation reactions. Unfortunately, the genomic context of the gene encoding RhCC does not provide any clues about the biological function of this protein in *Rhodococcus jostii* RHA1 (the gene coding for RhCC is flanked by genes encoding hypothetical proteins annotated either as enoyl-CoA hydratase-like or α , β -hydrolase-fold-like proteins). In the absence of a pathway context, it is tempting to speculate that the oxidative cleavage of 4HPP, yielding products **2** and **3**, might be the "true" physiological function of RhCC. However, although the β - α - β -fold is used here to host oxygen chemistry, which is proposed to involve hydroperoxide intermediate **4**, it clearly lacks the ability to carefully direct the chemistry of this highly reactive intermediate to prevent the formation of additional products. Indeed, RhCC could be an "accidental" oxygenase that is able to activate 4HPP by deprotonation of its 2-hydroxyl group, but prevents formation of the keto form of the substrate, allowing the 4HPP anion to react with a correctly positioned O₂ molecule. Therefore, the question remains whether the oxygenase activity of RhCC is physiologically relevant or reflects catalytic promiscuity.^{38,39} Nonetheless, this work sets the stage for in-depth mechanistic and structural studies of RhCC, which could identify the residues involved in catalysis as well as the additional features necessary for a highly specific cofactor-independent oxygenase. Work is in progress to determine the crystal structure of RhCC in complex with substrate 4HPP or a 4HPP analog that cannot be converted by RhCC [i.e., **9** or (*E*)-2-fluoro-*p*-hydroxycinnamic acid (**10**, **Supplementary Chart 1**)].

AUTHOR CONTRIBUTION

B.-J.B. performed the identification and characterization of the oxygenation reaction performed by RhCC. H.P. made the D106A mutant. H.P. and H.J.R. performed the crystallization and structure determination of wild type RhCC and D106A mutant. M.P.V. and H.J.P. assisted in high-resolution LC-MS and MS/MS analyses. E.M.G., A.M.W.H.T. and G.J.P supervised the scientific work. All authors contributed to writing the paper.

DECLARATION

Dr. Bert-Jan Baas also used this paper as a chapter in his PhD thesis entitled Exploring chemical versatility within the tautomerase superfamily: Catalytic promiscuity and the emergence of new enzymes (University of Groningen, 2014, ISBN: 978-94-6295-053-5).

ACKNOWLEDGMENTS

We thank the staff of the ESRF (Grenoble) for providing facilities for diffraction measurements and for assistance. We thank C.M. Jeronimus-Stratingh and A. van Dam for their expert assistance in acquiring and interpreting ESI-MS spectra, Dr. R. van der Geize for the kind gift of genomic DNA of *R. jostii* RHA1, and Prof. Dr. C.P. Whitman for the kind gift of (*E*)-2-fluoro-*p*-hydroxycinnamic acid and for critically reading the manuscript.

REFERENCES

1. Kerr, R. A. (2005) The story of O₂, *Science* 308, 1730-1732.
2. Klinman, J. P. (2007) How do enzymes activate oxygen without inactivating themselves?, *Acc. Chem. Res.* 40, 325-333.
3. Bugg, T.D.H. (2003) Dioxygenase enzymes: catalytic mechanisms and chemical models, *Tetrahedron* 59, 7075-7101.
4. Fetzner, S., and Steiner, R. S. (2010) Cofactor-independent oxidases and oxygenases, *Appl. Microbiol. Biotechnol.* 86, 791-804.
5. Minaev, B. F., and Minaeva, V. A. (2008) Spin-dependent binding of dioxygen to heme and charge-transfer mechanism of spin-orbit coupling enhancement, *Ukr. Bioorg. Acta* 2, 56-64.
6. Allen, R. C. (1994) Role of oxygen in phagocyte microbicidal action, *Environ. Health Perspect.* 102 suppl. 10, 201-208.
7. Tseng, C. C., Vaillancourt, F. H., Bruner, S. D., and Walsh, C. T. (2004) DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic pathway, *Chem. Biol.* 11, 1195-1203.
8. Thierbach, S., Bui, N., Zapp, J., Chhabra, S. R., Kappl, R., and Fetzner, S. (2014) Substrate-assisted O₂ activation in a cofactor-independent dioxygenase, *Chem. Biol.* 21, 217-225.
9. Bugg, T. D. H. (2014) How to break the rules of dioxygen activation, *Chem. Biol.* 21, 168-169.
10. Steiner, R. A., Janssen, H. J., Roversi, P., Oakley, A. J., and Fetzner, S. (2010) Structural basis for cofactor-independent dioxygenation of *N*-heteroatomic compounds at the α/β -hydrolase fold, *Proc. Natl. Acad. Sci. U.S.A.* 107, 657-662.
11. Whitman, C. P. (2002) The 4-oxalocrotonate tautomerase family of enzymes: how nature makes new enzymes using a beta-alpha-beta structural motif, *Arch. Biochem. Biophys.* 402, 1-13.
12. Poelarends, G. J., and Whitman, C. P. (2004) Evolution of enzymatic activity in the tautomerase superfamily: mechanistic and structural studies of the 1,3-dichloropropene catabolic enzymes, *Bioorg. Chem.* 32, 376-392.
13. Poelarends, G. J., Puthan Veetil, V., and Whitman, C. P. (2008) The chemical versatility of the beta-alpha-beta fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily, *Cell. Mol. Life Sci.* 65, 3606-3618.
14. Wasiel, A. A., Rozeboom, H. J., Hauke, D., Baas, J. B., Zandvoort, E., Quax, W. J., Thunnissen, A.-M. W. H., and Poelarends, G. J. (2010) Structural and functional characterization of a macrophage migration inhibitory factor homologue from the marine cyanobacterium *Prochlorococcus marinus*, *Biochemistry* 49, 7572-7581.
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
16. Waddell, W. J. (1956) A simple ultraviolet spectrophotometric method for the determination of protein, *J. Lab. Clin. Med.* 48, 311-314.
17. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM, *Acta. Crystallogr. D. Biol. Crystallogr.* 67, 271-281.
18. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution?, *Acta. Crystallogr. D. Biol. Crystallogr.* 69, 1204-1214.
19. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography, *Acta. Crystallogr. D.* 50, 760-763.

20. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Crystallogr.* **40**, 658-674.
21. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive python-based system for macromolecular structure solution, *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 213-221.
22. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 133-144.
23. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography, *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 12-21.
24. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Cryst. D60*, 2256-2268.
25. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench, *Bioinformatics* **25**, 1189-1191.
26. Zheng, H., Chordia, M. D., Cooper, D. R., Chruszcz, M., Müller, P., Sheldrick, G. M., and Minor, W. (2014) Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server, *Nat. Protoc.* **9**, 156-170.
27. Poelarends, G. J., Serrano, H., Person, M. D., Johnson, W. H., Jr., Murzin, A. G., and Whitman, C. P. (2004) Cloning, expression, and characterization of a *cis*-3-chloroacrylic acid dehalogenase: Insights into the mechanistic, structural, and evolutionary relationship between isomer-specific 3-chloroacrylic acid dehalogenases, *Biochemistry* **43**, 759–772.
28. Poelarends, G. J., Serrano, H., Person, M. D., Johnson, W. H., Jr., and Whitman, C. P. (2008) Characterization of Cg10062 from *Corynebacterium glutamicum*: implications for the evolution of *cis*-3-chloroacrylic acid dehalogenase activity in the tautomerase superfamily, *Biochemistry* **47**, 8139-8147.
29. Baas, B. J., Zandvoort, E., Wasiel, A. A., Quax, W. J., and Poelarends, G. J. (2011) Characterization of a newly identified mycobacterial tautomerase with promiscuous dehalogenase and hydratase activities reveals a functional link to a recently diverged *cis*-3-chloroacrylic acid dehalogenase, *Biochemistry* **50**, 2889-2899.
30. De Jong, R. M., Bazzacco, P., Poelarends, G. J., Johnson, W. H., Jr., Kim, Y. J., Burks, E. A., Serrano, H., Thunnissen, A. M. W. H., Whitman, C. P., and Dijkstra, B. W. (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase. Structural basis for substrate specificity and inactivation by (*R*)-oxirane-2-carboxylate, *J. Biol. Chem.* **282**, 2440-2449.
31. Guo, Y., Serrano, H., Johnson, W. H., Jr., Ernst, S., Hackert, M. L., and Whitman, C.P. (2011) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase: Implications for the catalytic and inactivation mechanisms, *Bioorg. Chem.* **39**, 1-9.
32. Lubetski, J. B., Swope, M., Dealwis, C., Blake, P., and Lolis, E. (1999) Pro-1 of macrophage migration inhibitory factor functions as a catalytic base in the phenylpyruvate tautomerase activity, *Biochemistry* **38**, 7346-7354.
33. Jefford, C. W., Knöpfel, W., and Cadby, P. A. (1978) Oxygenation of 3-aryl-2-hydroxyacrylic acids. The question of linear fragmentation vs. cyclization and cleavage of intermediates, *J. Am. Chem. Soc.* **100**, 6432-6436.
34. Oyama, S. T. et al. (2008) Mechanisms in homogeneous and heterogeneous epoxidation catalysis, Ch. 1 (Elsevier, Amsterdam).

35. Tsui, E. Y., Tran, R., Yano, J., and Agapie, T. (2013) Redox-inactive metals modulate the reduction potential in heterometallic manganese-oxido clusters, *Nat. Chem.* 5, 293-299.
36. Kriek, S., Yu, L., Reiher, M., and Westerhausen, M. (2010) Subvalent organometallic compounds of the alkaline earth metals in low oxidation states. *Eur. J. Inorg. Chem.* 2010, 197-216.
37. Widboom, P. F., Fielding, E. N., Liu, Y., and Bruner, S. D. (2007) Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis, *Nature* 447, 342-345.
38. Baas, B. J., Zandvoort, E., Geertsema, E. M., and Poelarends, G. J. (2013) Recent advances in the study of enzyme promiscuity in the tautomerase superfamily, *ChemBioChem.* 14, 917-926.
39. Copley, S. D. (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity, *Curr. Opin. Chem. Biol.* 7, 265-272.

SUPPLEMENTARY RESULTS

Supplementary Figure 1. SDS-PAGE analysis of fractions collected during the purification of His₆-tagged RhCC by column chromatography using Ni-sepharose. Lane 1, flow-through of cell free extract prepared from *E. coli* BL21 (DE3) producing His₆-tagged RhCC. Lane 2, proteins eluted with 10 mM imidazole. Lane 3, proteins eluted with 20 mM imidazole. Lane 4, proteins eluted with 40 mM imidazole. Lane 5, proteins eluted with 250 mM imidazole. As a general procedure, the protein fractions corresponding to lanes 4 and 5 were pooled, imidazole was removed by buffer exchange using a PD10 gel filtration column, the protein was concentrated to 5-10 mg/mL in 10 mM Na₂HPO₄ buffer (pH 7.3) and subsequently frozen using liquid nitrogen and stored at -80 °C until further use.

Supplementary Figure 2. Stereo views of Ca-backbone superpositions of RhCC with its structural homologs. **(a)** Ca-backbone superposition of single subunits. **(b)** Ca-backbone superposition of trimers. Color scheme used: black, RhCC; green, putative tautomerase from the thermophilic fungus *Aspergillus fumigatus* (PDB entry 3C6V); blue, *cis*-CaaD (PDB entry 2FLZ) and orange, Cg10062 (PDB entry 3N4D).

Supplementary Figure 3. Structural comparison of wild-type and D106A mutant RhCC. Close-up views of (a) the metal binding site in the wild-type RhCC trimer and (b) the equivalent, mutated site in the D106A trimer. Residues D106 and A106 are shown as ball-and-sticks, the Mg^{2+} ion (violet) and its coordinating waters (red) as spheres. Metal-ligand distances are indicated in Å. Also shown in (b) as grey mesh is the electron density ($2F_o - F_c$, 1.55 Å resolution, contoured at 1.0 σ) corresponding to the mutated residues. Apostrophes in labels denote the different subunits of the trimers. (c) Stereo view of $\text{C}\alpha$ -backbone superpositions of wild-type RhCC and the D106A mutant. Wild-type RhCC is depicted in orange whereas the mutant D106A is shown in black.

Supplementary Scheme 1. The enol-keto tautomerization of 4HPP catalyzed by MIF.

Supplementary Figure 4a. Plot of the RhCC-catalyzed conversion of 4HPP performed in 10 mM Na_2HPO_4 buffer (pH 7.3). These data were obtained by monitoring the depletion of the enol-form of 4HPP by UV-VIS spectroscopy. Although the shape of the graph suggests substrate inhibition, we were not able to fit the data using the equation that describes substrate inhibition.

Supplementary Figure 4b. Plot of the RhCC-catalyzed conversion of 4HPP performed in 10 mM Na_2HPO_4 buffer (pH 8.3). These data were obtained by monitoring the depletion of the enol-form of 4HPP by UV-VIS spectroscopy. Although the shape of the graph suggests substrate inhibition, we were not able to fit the data using the equation that describes substrate inhibition.

Supplementary Figure 5. Aromatic region of the ^1H NMR spectrum (in 10 mM Na_2HPO_4 buffer with 7.5% D_2O v/v) of the crude reaction mixture of the RhCC-catalyzed conversion of 4HPP. Characteristic signals are observed of major, enzymatically formed products **1** (7.27, 6.95, and 5.80 ppm) and **2** (9.71, 7.88, and 7.00 ppm) and of the chemically formed keto-form of 4HPP [7.16, 6.91 (signal overlaps with signal of product B, see HPLC chromatogram), and 4.03 ppm]. Signals of substrate 4HPP, expected at 7.68, 6.94, and 6.36 ppm (in D_2O) were not observed. Absorptions of the product responsible for peak B in the HPLC chromatogram of the crude mixture (**Supplementary Fig. 6**) are witnessed at ~6.6, ~6.8 and 6.95 ppm (the latter signal overlaps with an absorption of the keto-form of 4HPP).

Supplementary Figure 6. Chromatogram of the analysis of the reaction mixture of the RhCC-catalyzed conversion of 4HPP by reversed-phase HPLC. The chromatogram shows traces recorded at three different wavelengths, 230 nm (black), 250 nm (blue) and 300 nm (red), which clearly indicates that multiple products are present in the reaction mixture. In contrast to the ^1H NMR spectrum shown in **Supplementary Fig. 5**, the HPLC chromatogram indicates the presence of a remainder of substrate 4HPP. This is due to the fact that the ^1H NMR spectrum and the HPLC chromatogram were recorded from different reaction mixtures obtained from different experiments.

Supplementary Figure 7a. Progress curve of the substrate 4HPP as monitored by HPLC.

Supplementary Figure 7b. Progress curve of product 1 as monitored by HPLC.

Supplementary Figure 7c. Progress curve of product 2 as monitored by HPLC.

Supplementary Figure 7d. Progress curve of product **A** as monitored by HPLC.

Supplementary Figure 7e. Progress curve of product **B** as monitored by HPLC.

Supplementary Figure 7f. Progress curve of product **3** as monitored by high resolution LC-MS.

Supplementary Figure 8. The experimental setup that was built in order to be able to perform the RhCC-catalyzed conversion of 4HPP in the presence of $^{18}\text{O}_2$ (and in the absence of $^{16}\text{O}_2$).

Supplementary Scheme 2. A proposed mechanism for the formation of product **1**, based on hydroperoxide **4** as the initial enzymatic product formed by RhCC.

Supplementary Chart 1. Two structural analogs of substrate 4HPP, phenylpyruvate (**9**) and (*E*)-2-fluoro-*p*-hydroxycinnamic acid (**10**).

Supplementary Table 1a. High-resolution LC-MS and MS/MS data (negative mode) obtained for product A observed in reaction mixtures of reactions performed in aerated buffer.

Mass (Da) and elemental composition of parent ion	Observed mass (Da) of principal product ions	Calculated mass (Da) of principal product ions	Elemental composition of product ions
193.01439	167.01942	165.01933	$C_8H_5O_4$
(calcd. 193.01425)	149.02453	149.02442	$C_8H_5O_3$
$C_9H_5O_5$	121.02954	121.02950	$C_7H_5O_2$
	105.03460	105.03459	C_7H_5O
	77.03971	77.03967	C_6H_5

Supplementary Table 1b. High-resolution LC-MS and MS/MS data (negative mode) obtained for product B observed in reaction mixtures of reactions performed in aerated buffer.

Mass (Da) and elemental composition of parent ion	Observed mass (Da) of principal product ions	Calculated mass (Da) of principal product ions	Elemental composition of product ions
165.01941	121.02951	121.02950	$C_7H_5O_2$
(calcd. 165.01933)			
$C_8H_5O_4$			

Supplementary Table 1c. Identification of **3** as a product present in the reaction mixture of the RhCC-catalyzed conversion of 4HPP by high-resolution LC-MS (negative mode) using two different reverse phase LC-methods (XDB Eclipse C18 and XBridge columns) and oxalic acid (calcd. mass = 88.98803 Da) as a reference. The stated retention times indicate the beginning and the end of the observed peak.

Sample	Observed mass (Da)	Retention time (minutes) XDB Eclipse C18 column	Retention time (minutes) XBridge column
Oxalic acid (reference)	88.98807	3.3 – 3.6	13.5 – 14.0
3	88.98802	3.3 – 3.6	13.7 – 13.9

Supplementary Table 2a. High-resolution LC-MS and MS/MS data (negative mode) obtained for **1** and a product ion of Mw 151.04019 Da observed in reaction mixtures of reactions performed in aerated buffer.

Mass (Da) and elemental composition of parent ion	Observed mass (Da) of principal product ions	Calculated mass (Da) of principal product ions	Elemental composition of product ions
195.03001	167.03520	167.03498	C ₈ H ₇ O ₄
(calcd. 195.02990)	151.04025	151.04007	C ₈ H ₇ O ₃
C ₉ H ₇ O ₅	123.04533	123.04515	C ₇ H ₇ O ₂
	93.04361	93.03459	C ₆ H ₅ O
151.04019	123.04526	123.04515	C ₇ H ₇ O ₂
(calcd. 151.04007)	107.05032	107.05024	C ₇ H ₇ O
C ₈ H ₇ O ₃	93.03465	93.03459	C ₆ H ₅ O

Supplementary Table 2b. High-resolution LC-MS and MS/MS data (negative mode) obtained for **1** and a product ion of Mw 153.04439 Da observed in reaction mixtures of reactions performed in buffer equilibrated in ¹⁸O₂.

Mass (Da) and elemental composition of parent ion	Observed mass (Da) of principal product ions	Calculated mass (Da) of principal product ions	Elemental composition of product ions
197.03432	169.03938	169.03923	C ₈ H ₇ O ₃ ¹⁸ O
(calcd. 197.03414)	153.04453	153.04431	C ₈ H ₇ O ₂ ¹⁸ O
C ₉ H ₇ O ₄ ¹⁸ O	125.04952	125.04940	C ₇ H ₇ O ¹⁸ O
	93.03454	93.03459	C ₆ H ₅ O
153.04439	107.05029	107.05024	C ₇ H ₇ O
(calcd. 153.04431)			
C ₈ H ₇ O ₂ ¹⁸ O			

Supplementary Table 3. High-resolution LC-MS data (negative mode) obtained for **2** (calcd. mass = 121.02950 Da) observed in reaction mixtures of reactions performed in aerated buffer supplemented with H₂¹⁸O or buffer equilibrated with ¹⁸O₂.

Aerated buffer		¹⁸ O ₂		H ₂ ¹⁸ O	
Observed mass (Da)	Elem. comp. of parent ions	Observed mass (Da)	Elem. comp. of parent ions	Observed mass (Da)	Elem. comp. of parent ions
121.02956	C ₇ H ₅ O ₂	121.02957	C ₇ H ₅ O ₂	121.02953	C ₇ H ₅ O ₂

On the ^{18}O incorporation observed for oxalic acid (**3**)

The results reported in **Table 1** (see main text) show that partial ^{18}O incorporation was observed for product **3**; ~30% in the presence of 50% H_2^{18}O in the reaction buffer and ~30% under $^{18}\text{O}_2$ atmosphere as calculated from relative peak intensities in the LC/MS spectra. When normalized to 100% H_2^{18}O , still 40% of **3** would be unlabelled. Likewise, ^{18}O labeling of product **3** in the reaction performed under $^{18}\text{O}_2$ shows partial ^{18}O incorporation whereas for product **1** in the same samples labeling was shown to be in excess of 95%, indicating that nearly all $^{16}\text{O}_2$ had been expelled from the reaction system. These observations suggest that wash-out of ^{18}O label from product **3** could occur due to oxygen exchange with H_2O in the solvent. Since the exchange of oxygen with solvent water could have occurred under the used reaction conditions (pH 7.3) or during LC-MS analysis (pH 1.8), we performed control experiments in which we incubated oxalic acid at pH 1.8 and pH 7.3 in H_2^{18}O and followed oxygen exchange by high-resolution MS.

The reaction at pH 1.8 was performed as follows. A 50 mM stock solution of oxalic acid in water was prepared and set to pH 0.8. A 4 mL aliquot was subsequently diluted in 100 mL H_2^{18}O , resulting in a final pH of 1.8. The reaction at pH 7.3 was performed as follows. A 25 mM stock solution of oxalic acid in water was prepared and set to pH 9.3. A 4 mL aliquot was subsequently diluted in 100 mL H_2^{18}O , resulting in a final pH of 7.3. These final solutions were allowed to equilibrate at room temperature for 10 min, 2 h and 48 h and analyzed by high-resolution MS (flow-injection) on an LTQ Orbitrap XL (Thermo Fisher Scientific, Bremen, Germany).

The results are reported in **Supplementary Table 4**. From these results it is evident that no significant (within the detection limit of the LTQ Orbitrap XL) exchange of carboxylate-oxygen of oxalic acid with solvent water occurs at pH 7.3. Oxygen-exchange does occur, but only under acidic conditions and after a lengthy incubation time. We therefore conclude that wash-out of ^{18}O label from oxalic acid is not likely to occur under the enzymatic assay or HPLC conditions used. It thus seems that the proposed mechanism shown in **Scheme 1** (see main text), which would result in fully ^{18}O -labeled **3** both in H_2^{18}O and in $^{18}\text{O}_2$, is not the only mechanism that accounts for the formation of **3**.

On the exchange of the aldehyde oxygen of 4-hydroxybenzaldehyde (**2**) with solvent-derived oxygen

Unlike product **3**, it was anticipated that enzymatic product **2** might exchange its aldehyde oxygen with solvent-derived oxygen under the used reaction (pH 7.3) or HPLC (pH 1.8) conditions. To test this, we performed control experiments in which we incubated **2** at pH 1.8 and pH 7.3 in H_2^{18}O and followed oxygen exchange by high-resolution MS, using similar procedures as described above for oxalic acid. The results (data not shown) indicate that no significant exchange of the aldehyde-oxygen of **2** with solvent-derived oxygen occurs at pH 7.3, whereas complete oxygen exchange occurs at pH 1.8 within 3 min at

Supplementary Table 4. Exchange of carboxylate-oxygen of oxalic acid with H₂¹⁸O at 22 °C at pH 1.8 and pH 7.3 after 10 min, 2 h and 48 h incubation time as analyzed by high-resolution MS.

pH	Incubation time	Observed mass (Da)	Calculated mass (Da)	Elemental composition
1.8	10 min	88.98809	88.98803	C ₂ HO ₄
1.8	2 h	88.98812	88.98803	C ₂ HO ₄
1.8	48 h	88.98817	88.98803	C ₂ HO ₄
		90.99239	90.99228	C ₂ HO ₃ ¹⁸ O
		92.99650	92.99652	C ₂ HO ₂ ¹⁸ O ₂
		95.00070	95.00077	C ₂ HO ¹⁸ O ₃
		97.00495	97.00502	C ₂ H ¹⁸ O ₄
7.3	10 min	88.98807	88.98803	C ₂ HO ₄
7.3	2 h	88.98807	88.98803	C ₂ HO ₄
7.3	48 h	88.98807	88.98803	C ₂ HO ₄

room temperature. Thus, these control experiments demonstrate that any ¹⁸O transferred to the aldehyde oxygen of **2** would be washed out under the conditions of LC-MS analysis.

SPECTROSCOPIC DATA

4HPP

¹H NMR (500 MHz, D₂O, 25°C) δ 7.68 (d, *J* = 7.7 Hz, 2H), 6.94 (d, *J* = 7.7 Hz, 2H), 6.36 (s, 1H); HRMS (ESI): *m/z* = 179.03503 [M-H]⁻ (calcd. 180.04226 for C₉H₈O₄)

4HPP (keto form)

¹H NMR (500 MHz, 10 mM Na₂HPO₄ buffer with 7.5% D₂O v/v, 25°C) δ 7.16 (d, *J* = 8.3 Hz, 2H), 6.91 (d, *J* = 8.3 Hz, 2H), 4.03 (s, 2H); HRMS (ESI): *m/z* = 179.03508 [M-H]⁻ (calcd. 180.04226 for C₉H₈O₄)

3-Hydroxy-3-(4-hydroxyphenyl)-pyruvate (1)

^1H NMR (500 MHz, 10 mM Na_2HPO_4 buffer with 7.5% D_2O v/v, 25°C) δ 7.27 (d, $J = 9.0$ Hz, 2H), 6.95 (d, $J = 9.0$ Hz, 2H), 5.80 (s, 1H); ^{13}C NMR (125 MHz, D_2O , 25°C) δ 202.0, 160.0, 159.2, 132.4, 132.4, 130.4, 118.5, 118.5, 79.1; HRMS (ESI): $m/z = 195.03001$ $[\text{M}-\text{H}]^-$ (calcd. 196.03717 for $\text{C}_9\text{H}_8\text{O}_5$)

^1H , ^{13}C -HSQC NMR (500 MHz, D_2O , 25°C)

^1H , ^{13}C -HMBC NMR (500 MHz, D_2O , 25°C)

4-Hydroxybenzaldehyde (2)

^1H NMR (500 MHz, 10 mM Na_2HPO_4 buffer with 7.5% D_2O v/v, 25°C) δ 9.71 (s, 1H), 7.88 (d, $J = 8.3$ Hz, 2H), 7.00 (d, $J = 8.3$ Hz, 2H); HRMS (ESI): $m/z = 121.02956$ $[\text{M}-\text{H}]^-$ (calcd. 122.03678 for $\text{C}_7\text{H}_6\text{O}_2$)

Supplementary Table 5. Crystallographic data collection and refinement statistics of RhCC and mutant D106A.

	RhCC	Mutant D106A
Data collection		
Wavelength	0.9395	1.5418
Spacegroup	$\text{C}222_1$	C2
Unit cell dimensions, a, b, c (Å)	54.7, 98.6, 188.5	99.7, 57.7, 92.0, $\beta = 90.4^\circ$
Resolution range (Å)*	46 – 1.78 (1.88 – 1.78)	50.0 – 1.55 (1.63 – 1.55)
$R_{\text{merge}}^{\S,*}$	0.084 (0.257)	0.084 (0.255)
Nr of measurements	129204	197991
Nr of unique reflections	42398	74309
Completeness (%)*	86.8 (88.7)	97.5 (92.0)
Average I/σ *	7.9 (2.9)	7.4 (3.5)
Refinement		
R-factor / $R_{\text{free}}^{\S}$	0.152 / 0.187	0.160 / 0.174
Composition of asymmetric unit		
total number of atoms	4138	4208
protein residues (3 chains)	1 – 141	1–151 (chain A,C), 1-152 (chain B)
solvent molecules	542	466
ions	1 Mg^{2+}	None
others	3 bis-tris-propane, 1 glycerol	1 tris
Average B (Å ²)	22.5	19.5
R.m.s.d bonds (Å) / angles (°)	0.008 / 1.1	0.007 / 1.1
Ramachandran plot		
% most favored	98.1	97.6
% allowed	1.9	2.4
% disallowed	0.0	0.0

*Values in parentheses refer to the highest resolution shell

c h a p t e r f i v e

STRUCTURAL BASIS FOR THE CATALYTIC MECHANISM OF ETHYLENEDIAMINE-*N*, *N*'-DISUCCINIC ACID LYASE, A CARBON-NITROGEN BOND-FORMING ENZYME WITH A BROAD SUBSTRATE SCOPE

H. Poddar*
J. de Villiers*
J. Zhang
V. P. Veetil
H. Raj
A.-M. W. H. Thunnissen
G. J. Poelarends

*These authors contributed equally to this work

Published in Biochemistry 2018, 57, 3752-3763

ABSTRACT

The natural aminocarboxylic acid product ethylenediamine-*N,N'*-disuccinic acid [(*S,S*)-EDDS] is able to form a stable complex with metal ions, making it an attractive biodegradable alternative for the synthetic metal chelator ethylenediamine tetraacetic acid (EDTA), which is currently used at large scale in numerous applications. Previous studies have demonstrated that biodegradation of (*S,S*)-EDDS may be initiated by an EDDS lyase, converting (*S,S*)-EDDS via the intermediate *N*-(2-aminoethyl)aspartic acid (AEAA) into ethylenediamine and two molecules of fumarate. However, current knowledge of this enzyme is limited due to the absence of structural data. Here, we describe the identification and characterization of an EDDS lyase from *Chelativorans* sp. BNC1, which has a broad substrate scope, accepting various mono- and diamines for addition to fumarate. We report crystal structures of the enzyme in an unliganded state and in complex with formate, succinate, fumarate, AEAA and (*S,S*)-EDDS. The structures reveal a tertiary and quaternary fold that is characteristic of the aspartase/fumarase superfamily and support a mechanism that involves general base-catalyzed, sequential two-step deamination of (*S,S*)-EDDS. This work broadens our understanding of mechanistic diversity within the aspartase/fumarase superfamily and will aid in the optimization of EDDS lyase for asymmetric synthesis of valuable (metal-chelating) aminocarboxylic acids.

Aminocarboxylic acids that contain several carboxylate groups bound to one or more nitrogen atoms form an important group of metal chelating agents.¹ The synthetic chelator ethylenediamine tetraacetic acid (EDTA) is a well-known example of this group of metal complexing compounds. It is used in large amounts in numerous industrial applications, including soil bioremediation and in the production of paper, textile, cosmetics, detergents and fertilizers.² Although still widely in use, there is a growing concern about the adverse environmental effects of this synthetic compound.²⁻⁵ Its high resistance to biodegradation leads to accumulation in the environment. High concentrations of EDTA and its chelating of heavy metals is feared to have a negative effect on drinking water and aquatic life. Hence, aminocarboxylic acid metal chelators that are readily biodegradable are highly desirable as 'green' alternatives.

The natural aminocarboxylic acid product ethylenediamine-*N,N'*-disuccinic acid [(*S,S*)-EDDS], which is produced by a wide range of bacteria to facilitate the uptake of metal ions, is a structural isomer of EDTA.^{6,7} Interestingly, (*S,S*)-EDDS exhibits similar chelating properties as EDTA. However, in contrast to EDTA, (*S,S*)-EDDS is readily biodegradable and is therefore an attractive alternative with a favourable environmental profile.⁸⁻¹⁰ Previous studies have demonstrated that biodegradation of (*S,S*)-EDDS may be initiated by an EDDS lyase, converting (*S,S*)-EDDS (**1**) via the intermediate *N*-(2-aminoethyl)aspartic acid (AEAA, **3**) into ethylenediamine (**4**) and two molecules of fumarate (**2**) (Scheme 1).⁸⁻¹¹ Mizunashi firstly reported the cloning of a gene (from *Brevundimonas* sp. TN3) coding for an EDDS lyase.¹¹ This *Brevundimonas* enzyme was shown to have potential for use as a biocatalyst in the preparation of EDDS and its derivatives.¹¹

Sequence analysis suggested that EDDS lyase belongs to the aspartase/fumarase superfamily.¹² Members of this superfamily include aspartase, fumarase, argininosuccinate lyase, adenylosuccinate lyase, δ -crystallin, and 3-carboxy-*cis,cis*-muconate lactonizing enzyme. They share a common tertiary and quaternary fold, as well as a similar active site architecture, and process succinyl containing substrates, leading to the formation of fumarate as the common product (except for the CMLE-catalyzed reaction, which results in the formation of a lactone).¹² Current knowledge of the reaction mechanism of EDDS lyase is limited, however, due to the absence of structural data. We have therefore focused our attention on the molecular cloning of an EDDS lyase and initiated structural

Scheme 1. Reversible two-step deamination of (*S,S*)-EDDS catalyzed by EDDS lyase.

studies with the aim of elucidating the details of its unusual two-step addition–elimination reaction mechanism.

Here we describe the identification and cloning of the gene encoding EDDS lyase from the bacterium *Chelativorans* sp. BNC1, which was isolated from industrial sewage receiving EDTA-containing wastewater effluents.¹³ The enzyme has been purified to homogeneity and subjected to functional and structural characterization. It was found to accept a wide range of structurally distinct amines for addition to fumarate. In addition, we have previously determined that this *Chelativorans* enzyme also accepts a wide variety of amino acids with terminal amino groups for selective amination of fumarate, demonstrating its synthetic usefulness for the production of various important metal-chelating aminocarboxylic acids.¹⁴ Crystal structures of the enzyme in an unliganded state and in complex with formate, succinate, fumarate, AEAA and (S,S)-EDDS were determined. These structures confirm a structural fold that is characteristic of the aspartase/fumarase superfamily and support a mechanism that involves general base-catalyzed, sequential two-step deamination of (S,S)-EDDS.

MATERIALS AND METHODS

Materials

Ingredients for buffers and media were obtained from Duchefa Biochemie (Haarlem, The Netherlands) or Merck (Darmstadt, Germany). All other chemicals used in the experiments including the sodium salt of (S,S)-EDDS, fumaric acid and succinic acid were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) unless stated otherwise. Molecular biology reagents, including restriction enzymes, PCR reagents, T4 DNA ligase, and DNA and protein ladders were obtained from Fermentas (ThermoFisher Scientific, Pittsburgh, PA) or Promega Corp. (Madison, WI). PCR purification, gel extraction and Miniprep kits were provided by Macherey-Nagel (Duren, Germany). Ni-Sepharose 6 Fast Flow and pre-packed PD-10 Sephadex G-25 columns were purchased from GE Healthcare Life Sciences (Uppsala, Sweden). Primers for DNA amplification were synthesized by Eurofins MWG Operon (Cologne, Germany).

General Methods

Techniques for restriction enzyme digestions, ligation, transformation, and other standard molecular biology manipulations were based on standard protocols or as suggested by the manufacturer. PCR was carried out in a DNA thermal cycler obtained from Biolegio (Nijmegen, The Netherlands). DNA sequencing was performed by Macrogen (Amsterdam, The Netherlands). Protein was analyzed by polyacrylamide gel electrophoresis under denaturing conditions using sodium dodecyl sulfate (SDS) on precast gels containing 7.5–10% polyacrylamide (Invitrogen). The gels were stained with InstantBlue™ (Expedeon Inc.). Protein concentrations were determined by the Waddell method.¹⁵ Kinetic data were obtained on a V-650 spectrophotometer obtained from Jasco (IJsselstein, The Netherlands).

The kinetic data were fitted by nonlinear regression data analysis using the Graft program (Erithacus, Software Ltd., Horley, U.K.) obtained from Sigma Chemical Co. Dynamic light scattering (DLS) experiments were performed using a DynaPro MS800TC instrument (Wyatt Technology Corporation, USA) at 20 °C and data were processed and analyzed with Dynamics software (Wyatt Technology Corporation).

Enzymatic synthesis of AEAA using MAL-Q73A

The reference compound *N*-(2-aminoethyl)aspartic acid [(S)-AEAA] was synthesized using the previously reported MAL-Q73A enzyme.¹⁶ A solution (80 mL) of fumarate (2g, 17.24 mmol, 215 mM), 1,2-diaminoethane dihydrochloride (32 g, 240 mmol, 3 M) and MgCl₂ (20 mM) was prepared. The pH was adjusted to 9.0 by using aqueous NaOH. The reaction was started by the addition of freshly purified MAL-Q73A enzyme (40 mg, 0.0005 mol%) and the reaction mixture was incubated at room temperature. The progress of the reaction was monitored by UV-Vis spectroscopy. The reaction was stopped after 7 d by incubating the reaction mixture at 100°C for 10 min, and precipitated protein was removed by filtration. Excess of amine was removed from the reaction mixture by the use of a rotary evaporator. The concentrated crude reaction mixture was dissolved in 50 mL 1 N HCl and the desired amino acid product was purified by cation-exchange chromatography by following an earlier described protocol.¹⁶ The purified product was obtained as the bis-ammonium salt and identified as *N*-(2-aminoethyl)aspartic acid by ¹H NMR and ¹³C NMR spectroscopy and HRMS analysis. ***N*-(2-aminoethyl) aspartic acid**: conversion 100% (7 d), yield (61%, 2.2 g). Brown solid. ¹H NMR (500 MHz, D₂O): δ = 2.66 (dd, 1H, *J* = 17.3 Hz, 9.0 Hz, CHCH₂), 2.83 (dd, 1H, *J* = 17.3 Hz, 3.8 Hz, CHCH₂), 3.34-3.46 (m, 4H, NHCH₂CH₂NH₂), 3.83 (dd, 1H, *J* = 9.0 Hz, 3.7 Hz, CHNH₂). ¹H NMR signals are in agreement with the literature data.¹⁶

Cloning of the EDDS lyase gene into an expression vector

The amino acid sequence of the putative EDDS lyase from *Chelativorans* sp. BNC1, which was annotated as an argininosuccinate lyase, was obtained from NCBI protein database under GenBank accession no. ABG61966 (NCBI reference sequence WP_011579909.1). The corresponding DNA sequence was codon-optimized for *Escherichia coli* and synthesized by Eurofins MWG Operon (Ebersberg, Germany). The gene was delivered in the pBSII SK+ vector with restrictions sites for *Nde*I and *Hind*III at the 5' and 3' end of the gene, respectively. The gene was amplified by PCR using primers Ed_fw-*Nde*I: GGAGGAATT**ACATATGA**ACATCAACGTACCGGACGC and Ed_his-st-rv-*Hind*III: CAT**AAGCTT**TATCAATGATGATGATGATGATGGCGCAGATATTTGCGGTCCG (the restrictions sites are depicted in bold), digested with *Nde*I/*Hind*III and cloned into the pBADN-Myc-His expression vector to obtain the pBADN(EDDS-His) construct. The entire gene was sequenced to verify that no mutations were introduced during the cloning procedure.

Expression and purification of EDDS lyase

The His₆-tagged enzyme was overproduced in *E. coli* TOP10 cells using the pBADN(EDDS-His) expression plasmid. Freshly transformed TOP10 cells containing this plasmid were used to inoculate 10 mL of LB/Amp medium. After overnight growth at 37 °C, this culture was used to inoculate 1 L of LB/Amp medium in a 5 L Erlenmeyer flask. Cultures were grown to an A₆₀₀ of 0.4-0.6 at 37 °C with vigorous shaking, induced with arabinose (0.04% w/v) and placed at 20°C for overnight incubation (~16 h). Cells were harvested by centrifugation (6000g, 15 min) and stored at -20 °C until further use.

In a typical purification experiment, cells from 1 L culture (~5.0 g wet weight) were thawed and suspended in 10 mL lysis buffer (50 mM Tris-HCl, 20 mM imidazole, pH 8.0). Cells were disrupted by sonication for 4 x 1 min (with 5 min rest in between each cycle) at a 60 W output, after which unbroken cells and debris were removed by centrifugation (10,000g, 30 min). The supernatant was filtered through a 0.45 µm-pore diameter filter and incubated with Ni-Sepharose (1 mL slurry in a small column at 4 °C for 318 h), which had previously been equilibrated with lysis buffer. The non-bound proteins were eluted from the column by gravity flow. The column was first washed with lysis buffer (10 mL) and then with buffer A (50 mM Tris-HCl, 40 mM imidazole, pH 8.0; 10 mL). Retained proteins were eluted with buffer B (50 mM Tris-HCl, 500 mM imidazole, pH 8.0; 4 mL). Fractions were analyzed by SDS-PAGE on gels containing 10% acrylamide, and those that contained purified EDDS lyase were pooled and the buffer was exchanged against 50 mM Tris-HCl (pH 8.0), 200 mM NaCl using a pre-packed PD-10 Sephadex G-25 gelfiltration column. The purified enzyme (yield ~50 mg) was stored at either 4 °C or -20 °C until further use.

Construction and production of EDDS lyase mutants

The S280A and D290A mutants of EDDS lyase were constructed using the QuikChange site-directed mutagenesis method (Stratagene). The plasmid pBADN(EDDS-His) was used as a template. For the S280A mutation, the following oligonucleotides were used as forward and reverse primers: 5'-GCGGGAACCG**CGC**TGCATCATGCCG-3' and 5'-GCGGCATGATCGA**CGC**GGTTCCCGC-3', respectively, where the mutated codon is depicted in bold. For the D290A mutation, the following oligonucleotides were used as forward and reverse primers: 5'-CGCAGAAGAAGAACCCG**GCT**AGCCTGGAACGTAGTCGC-3' and 5'-GCGACTACGTTCCAGGCT**AGC**GGGTCTTCTTCTGCG-3', respectively, where the mutated codon is depicted in bold. DNA sequencing of the mutant genes was performed to assure that only the desired mutation had been introduced.

The S280A and D290A mutants were produced and purified using similar protocols as those used for wild-type EDDS lyase. The S280A mutant was further purified by gel filtration chromatography with a Superdex 200 10/300 GL column (GE Healthcare) using 0.1 M NaCl in 50 mM Tris-HCl buffer (pH 7.5) as an eluent, whereas the D290A mutant was further purified to homogeneity by gel filtration chromatography with a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) using 20 mM NaH₂PO₄ buffer (pH 8.5) as an

eluent. Activity assays were done with freshly purified proteins (notably, some protein precipitation was observed upon storage of the S280A mutant, indicative of poor stability).

Enzyme assays

Kinetic assays were performed at 25 °C in 50 mM Tris/HCl buffer, pH 8.0, observing the increase in absorbance at 240 nm corresponding to the formation of fumarate ($\epsilon = 2530 \text{ M}^{-1}\text{cm}^{-1}$). An aliquot of EDDS lyase (180 μg) was diluted into buffer (15 mL) and incubated for 30 min at 25 °C. Subsequently, a 1-mL portion was transferred to a 10 mm quartz cuvette and the enzyme activity was assayed by the addition of a small quantity (0.5–20 μL) of EDDS from a stock solution (10 mM). The stock solution was made up in 50 mM Tris/HCl buffer, pH 8.0. The concentrations of EDDS used in the assay ranged from 0.005 to 0.2 mM.

The pH optimum of EDDS lyase was determined in 50 mM phosphate buffers with pH values ranging from 4.4 to 9.2 at 25 °C. A sufficient quantity of enzyme was added (12 $\mu\text{g}/\text{mL}$) and its activity assayed by adding (*S,S*)-EDDS from a stock solution to a final concentration of 0.1 mM, following the increase in absorbance at 240 nm corresponding to the formation of fumarate. The normalized initial reaction rates were plotted against pH.

The temperature optimum was determined in Tris-HCl buffer (50 mM, pH 8.0), using a temperature range of 10 to 80 °C. At each temperature, the pH of the Tris buffer was adjusted to the desired pH 8.0. A 1-mL portion of the buffer was transferred to a 10 mm cuvette, a sufficient quantity of enzyme was added (12 $\mu\text{g}/\text{mL}$), and its activity assayed using 0.1 mM (*S,S*)-EDDS (2 μL of 50 mM stock solution) as the substrate. Substrate stock solutions were made in Tris-HCl buffer (50 mM, pH adjusted to 8.0). The normalized initial reaction rates were plotted against temperature.

Enzymatic synthesis of AEAA and (*S,S*)-EDDS using EDDS lyase

A reaction mixture containing fumaric acid (50 mM) and ethylenediamine (10 mM) was prepared in 20 mM NaH_2PO_4 -NaOH buffer, pH 8.5. EDDS lyase (14 mg, 0.05 mol%) was added to start the reaction, and the reaction volume was immediately adjusted to 50 mL with 20 mM NaH_2PO_4 -NaOH buffer (pH 8.5). The reaction was allowed to proceed at room temperature. At different time points, reaction samples (0.5 mL) were taken from the reaction mixture and boiled for 10 min to inactivate the enzyme. The samples were dried under vacuum and re-dissolved in 0.5 mL D_2O for ^1H NMR measurements.

The ^1H NMR (500 MHz, Deuterium Oxide) signals of (*S,S*)-EDDS are: δ 3.57 (dd, $J = 8.7, 4.3 \text{ Hz}$, 2H), 3.08 – 2.88 (m, 4H), 2.66 (dd, $J = 16.2, 4.3 \text{ Hz}$, 2H), 2.48 (dd, $J = 16.2, 8.8 \text{ Hz}$, 2H). The ^1H NMR signals (500 MHz, Deuterium Oxide) of AEAA are: δ 3.43 (dd, $J = 10.4, 3.7 \text{ Hz}$, 1H), 3.14 – 2.92 (m, 3H), 2.84 – 2.77 (m, 1H), 2.60 (dd, $J = 15.5, 3.7 \text{ Hz}$, 1H), 2.29 (dd, $J = 15.5, 10.4 \text{ Hz}$, 1H). After 24 h, the ratio between AEAA and (*S,S*)-EDDS in the reaction mixture was ~2:1, calculated by integration of the signals at 3.43 ppm and 3.57 ppm, respectively.

Substrate scope and product identification by LC-MS/MS

Various amines were tested as substrates for EDDS lyase in the addition to fumarate. Fumarate (5 mM), amine (400 mM) and EDDS-lyase (0.5 mg/mL, 0.17 mol%) in 50 mM Tris-HCl buffer, pH 8.0, were incubated at room temperature in a 96-well plate (final volume of 150 μ L). Reactions were monitored by UV spectroscopy, following the decrease in absorbance at 240 nm corresponding to the depletion of fumarate.

For several selected amine substrates, reactions were performed with two different substrate ratios and the products identified with LC-MS/MS. Reactions were initially performed with a 20-fold excess of amine. Fumarate (50 mM), amine (1000 mM) and EDDS-lyase (0.5 mg/mL) in buffer (Tris-HCl, pH 8.0) were incubated for 24 h. Reactions were subsequently also performed with a 2-fold excess of fumarate. Fumarate (100 mM), amine (50 mM) and EDDS-lyase (0.1 mg/mL) in buffer (Tris-HCl, pH 8.0) were incubated for 24 h. All samples were prepared for LC-MS/MS analysis as follows. After incubation of the reaction mixture for 24 h, the samples were incubated at 100 °C for 1-2 min to stop the reaction. The precipitated enzyme was removed by centrifugation. The supernatant was filtered (0.42 μ m filter) and subjected to LC-MS/MS to confirm formation of single and/or double addition products. Mass spectrometric analysis was performed by the Mass Spectrometry Facility Core in the Department of Pharmacy at the University of Groningen.

Crystallization

Before setting up crystallization trials, the EDDS lyase protein sample was further purified by gel filtration chromatography with a Superdex 200 (GE Healthcare) column using 0.2 M NaCl in 50 mM Tris-HCl buffer, pH 7.5, as an eluent. The protein eluted as a tetramer with an apparent molecular weight of about 200 kDa, as confirmed by dynamic light scattering analysis. The protein sample was concentrated to 9 mg/mL using a centrifugal concentrator (Vivaspin 15R, 30-kDa molecular weight cut-off, Sartorius Stedim Biotech). A search for crystallization conditions was performed using various commercial crystallization screens. Screening was carried out at room temperature in 96-well sitting-drop crystallization plates using a robot (Mosquito, TTP LabTech) to dispense 300 nL drops (protein:reservoir, 1:1). A single cubic-shaped crystal was obtained directly from the Structure Screen (Molecular Dimensions) with a solution containing 0.1 M HEPES pH 7.5 and 4 M NaCl, but this condition could not be reproduced. Crystals also grew at a condition from Clear Strategy Screen II (Molecular Dimensions), containing 0.1 M sodium cacodylate pH 6.5 and 2.0 M sodium formate. Small crystals would typically appear within an hour after setting up the crystallization plates, reaching a maximum size of about 140 x 90 x 80 μ m³ after overnight growth.

Crystal soaking experiments to trap a substrate, intermediate or product in the active site of EDDS lyase were unsuccessful, probably due to the high sodium formate concentration in the crystallization solution and the presence of bound formate ions in the active site. Thus, an alternative strategy was employed, replacing sodium formate

in the crystallization condition by sodium salts of fumaric acid, succinic acid, (S,S)-EDDS or a bis-ammonium salt of AEAA (synthesized as described above). Diffracting crystals appeared overnight, with all of the above-mentioned salts at a concentration of 0.2-0.3 M, similar as the crystals grown in the presence of sodium formate. As an additional strategy to trap an intermediate-bound state of EDDS lyase, crystals co-crystallized with fumarate were soaked for 30 seconds in drops of mother liquor (0.1 M sodium cacodylate pH 6.5 and 0.3 M fumarate) containing 20 mM ethylenediamine or AEAA, immediately followed by flash cooling in liquid nitrogen.

X-ray data collection and crystal structure determination

Prior to X-ray data collection, crystals were briefly transferred to drops containing mother liquor supplied with 25% PEG 400 and flash-cooled in liquid nitrogen. All the diffraction data were collected in-house at 110 K using a Microstar rotating Cu anode X-ray source (Bruker AXS GmbH) in combination with Helios optics (Incoatec GmbH) and a MAR345dtb detector (MarResearch GmbH). Data sets were integrated and scaled using XDS¹⁷ and merged using the program AIMLESS¹⁸ from the CCP4 software suite.¹⁹ All crystals belonged to the *F*222 space group and contained a single polypeptide chain per asymmetric unit, with a solvent content of about 62%. Relevant data collection and refinement statistics are shown in Table S2. The upper resolution limit of the data sets varied between 2.6 Å and 1.9 Å. It should be noted that some of the data sets were recorded with a non-optimal crystal-to-detector distance, in which cases the actual resolution limit of the diffraction was somewhat higher. Structure determination was started with diffraction data collected from the single crystal obtained at 4 M NaCl. The program PHASER²⁰ from the CCP4 package was used to obtain initial phases by molecular replacement. With the help of the Fold and Function Assignment Server (FFAS),²¹ an ensemble of three homologous protein structures was used as a molecular replacement search model: duck δ crystallin II (28% identity, PDB entry 1TJU),²² argininosuccinate lyase from *Thermus thermophilus* HB8 (30% identity, not published, PDB entry 2E9F) and duck δ crystallin I (29% identity, PDB entry 1U15).²³ Automatic model building was performed using ARP/wARP.²⁴ The programs REFMAC5²⁵ and COOT²⁶ were used for subsequent rounds of refinement and model building, including the placement and validation of water molecules. The final rounds of refinement were performed using phenix.refine from the Phenix software suite.²⁷ All other structures were adapted and refined starting from the apo structure, using identical strategies and software. Coordinates and restraints for formate, fumarate and succinate were readily available from the CCP4 database with the following ligand identifiers, FMT, FUM and SIN. Co-ordinates and restraints for (S,S)-EDDS and AEAA were generated using the PRODRG2 server.²⁸

Crystal structure analysis

Molprobit²⁹ was used for validating the stereochemical quality of the models. Structure-based sequence alignments were carried out using T-coffee³⁰ and visualized using ESPript 3.0 server.³¹ Superpositions and calculation of C α -backbone rmsd values were performed using the protein structure comparison service Fold at the European Bioinformatics Institute.³² PyMOL (Schrödinger)³³ was used for structure analysis and figure preparations. ChemBioDraw 12.0 was used to draw schemes and chemical structures. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (www.rcsb.org), with the following PDB codes, 6G3D, 6G3E, 6G3F, 6G3G, 6G3H, 6G3I.

RESULTS

Identification of an EDDS lyase in *Chelativorans* sp. BNC1

A sequence similarity search in the NCBI microbial database was performed with the BLASTP program using the EDDS lyase amino acid sequence from *Brevundimonas* sp. TN3 as the query. This search yielded several bacterial proteins that shared significant sequence similarity with EDDS lyase. The top hits included a sequence from the bacterium *Chelativorans* sp. BNC1, which was isolated from industrial sewage receiving EDTA-containing wastewater effluents.¹³ This *Chelativorans* protein, which shows 79% sequence identity to EDDS lyase from *Brevundimonas* sp. TN3 (Figure S1),¹¹ was annotated as a putative argininosuccinate lyase and selected for further study.

The gene coding for the EDDS lyase homologue from *Chelativorans* sp. BNC1 was cloned into the expression vector pBADN/Myc-His A, resulting in the construct pBADN (EDDS lyase-His). Using this expression plasmid, the enzyme was produced upon induction with arabinose in *Escherichia coli* TOP10 as a C-terminal hexahistidine fusion protein. The enzyme was purified by a one-step Ni-sepharose affinity chromatography protocol, which typically provides ~30 mg of homogeneous enzyme per liter of culture. Analysis of the purified enzyme by size-exclusion chromatography and dynamic light scattering revealed a native molecular mass of ~200 kDa. A comparison of this value to that of the calculated subunit mass suggests that the enzyme is a homotetrameric protein.

To examine whether the *Chelativorans* enzyme can promote the synthesis of AEAA (**3**) and EDDS (**1**), the enzyme was incubated with fumarate (**2**) and ethylenediamine (**4**) and the reaction was monitored by ¹H NMR spectroscopy (Figure 1). The results showed that the enzyme indeed catalyzes the addition of ethylenediamine to fumarate to give AEAA and (S,S)-EDDS. The enzyme also catalyzes the reverse reaction, that is, the deamination of (S,S)-EDDS to yield AEAA, fumarate and ethylenediamine as determined by UV and ¹H NMR spectroscopy (data not shown). Having established that the *Chelativorans* enzyme exhibits EDDS lyase activity, kinetic parameters (at 25°C) and the optimum pH and temperature were determined. The enzyme catalyzes the deamination of (S,S)-EDDS with a k_{cat} of $6.5 \pm 0.2 \text{ s}^{-1}$ and a K_m of $16 \pm 3 \text{ }\mu\text{M}$, and shows maximum activity at pH 8.0 and at

60°C (Figure S2). Notably, incubation of the enzyme with fumarate and arginine showed that the enzyme displays no argininosuccinate lyase activity.

Substrate scope of EDDS-lyase

It has previously been determined that the *Chelativorans* enzyme accepts a wide variety of amino acids with terminal amino groups for selective addition to fumarate, yielding the natural products Aspergillomarasmine A and Aspergillomarasmine B, as well as various related aminocarboxylic acids (Scheme S1).¹⁴ To further explore the substrate scope of EDDS lyase, the enzyme was incubated with fumarate and different amines and the rate of the reactions was monitored spectrophotometrically by following the depletion of fumarate at 240 nm. The results showed that EDDS lyase has a broad nucleophile scope and accepts various mono- and diamines as unnatural substrates in the amination of fumarate, albeit with lower catalytic efficiency when compared to the reaction with the native substrate ethylenediamine (Table 1). For several selected substrates, the enzymatic addition reactions were performed with either a 20-fold excess of amine or a 2-fold excess of fumarate, and formation of the corresponding single and double addition products was determined by LC-MS/MS (Table S1). The results demonstrate that by using an appropriate molar ratio of starting substrates, different aminocarboxylic acid products, including EDDS derivatives, can be prepared using EDDS lyase as the catalyst. In contrast to its very broad nucleophile scope, the enzyme was found to be highly specific for fumarate, with fumaric acid monomethyl ester, crotonic acid, mesaconic acid, itaconic acid, 2-pentenoic acid and glutaconic acid not accepted as alternative electrophiles.

5

Overall structure of EDDS lyase

The crystal structure of EDDS lyase was determined at 2.2 Å resolution by molecular replacement and refined to a crystallographic R-factor of 18.2% (R_{free} is 22.2%) with good geometry (Table S2). Diffraction data were obtained from a single crystal grown in the presence of 4 M NaCl at pH 7.5. The EDDS lyase crystal structure belongs to the space group *F*222 and contains one monomer per asymmetric unit (labeled A); the other three subunits (B, C and D) of the functional tetramer are “generated” by crystallographic 2-fold axes. The final model consists of 496 residues; only the first five residues at the N-terminus and the last seven residues at the C-terminus, which comprises Arg502 and the (His)₆-tag, could not be modeled due to weak or absent electron density.

The overall fold and topology of EDDS lyase closely resemble those of other members of the aspartase/fumarase superfamily (Figures 2 and 3).¹² The dumbbell-shaped, mainly α -helical subunit can be sub-divided into three domains (Figure 2a): an N-terminal domain (residues 6-111), an elongated central domain (residues 112-351) and a C-terminal domain (residues 352-501). In the functional tetramer (Figure 2b) the central domains of the four subunits interact co-axially to form a tightly packed bundle of 20 α -helices, with the N- and C-terminal domains of neighboring subunits positioned near to each other at the ends.

Figure 1. EDDS lyase-catalyzed addition of ethylenediamine (4) to fumaric acid (2) yielding AEAA (3) and EDDS (1). (A) ¹H NMR spectrum of AEAA (3) in D₂O. (B) ¹H NMR spectrum of the reaction mixture after 24 h of incubation. The reaction was started by addition of EDDS lyase (0.05 mol%) to fumaric acid (50 mM) and ethylenediamine (10 mM) in 20 mM NaH₂PO₄-NaOH buffer, pH 8.5. (C) ¹H NMR spectrum of (S,S)-EDDS (1) in D₂O. The assignments of key signals are highlighted in red color.

Each active site (there are four in the tetramer) is composed of three regions of highly conserved amino acid residues, each of which originates from a different subunit (Figure S3, Figure 2b). These conserved regions are C1 (residues 111-114, for a representative active site we will refer to the subunit as A), C2 (residues 158-165, subunit B) and C3 (also known as the “SS-loop”, residues 279-293, subunit C).¹²

Two striking differences are observed in the overall structure of EDDS lyase compared to other aspartase/fumarase superfamily members (Figure 3). Firstly, the EDDS lyase subunit contains a final C-terminal α-helix (residues 451-500) which is elongated (Figure S3) and runs alongside the entire length of the central domain back to the N-terminal domain. Secondly, within a subunit the first ~30 residues at the N-terminus form a loop (residues 5-15) and a short α-helix (residues 20-31), which fold away from the N-terminal domain to pack against the central domain and SS-loop. Two arginine residues (Arg10 and Arg13) in the N-terminal loop form hydrogen bonds with the main chain carbonyls of Ala274, Ala277 and Gln285, effectively locking the SS-loop in place. In the tetramer, each SS-loop is further stabilized by interactions with residues of the N- and C-terminal domains

Table 1. Relative activities of EDDS lyase towards ammonia and various amines.

Entry	Substrate	Substrate Structure	Relative Activity [%]
1	4a		100 ^a
2	4b		2
3	4c		< 1
4	4d		< 1
5	4e		< 1
6	4f		< 1
7	4g		6
8	4h		3
9	4i		9
10	4j		6
11	4k		34
12	4l		14
13	4m		59
14	4n		< 1
15	4o		< 1
16	4p		1
17	4q		112
18	4r		74
19	4s		23
20	4t		2

^aThe initial rate of the enzyme-catalyzed addition of ethylenediamine to fumarate was assigned as 100% activity.

from two neighboring subunits, resulting in a rigid, and well-defined conformation. In the structures of other aspartase/fumarase superfamily members the SS-loop is intrinsically flexible, and can undergo an open-closed transition upon binding of substrate.^{12, 34-37} In EDDS lyase the occurrence of such a transition of the SS-loop is unlikely: the additional interactions with the N-terminal loop region stabilize a conformation that is highly similar to the catalytically competent closed conformation observed in the structures of other aspartase/fumarase superfamily members, even when no substrate is bound.

Formate-, fumarate-, and succinate-bound structures

During the initial crystallization screening a condition containing 2.0 M sodium formate (pH 6.5) also yielded protein crystals. The crystals belonged to the same space group *F*222

Figure 2. Overall structure of EDDS lyase. a) Cartoon representation of a monomer. The N-terminal domain, central domain and C-terminal domain are depicted in blue, yellow and orange, respectively. The conserved regions C1, C2 and SS-loop are highlighted in red. b) Cartoon representation of the functional tetramer, with the polypeptide chains in different colours. The inset shows the zoomed-in view of an active site (viewing orientation different as for the tetramer). Residues from the conserved regions forming the composite active site are shown in stick representation.

Figure 3. Comparison of the EDDS lyase monomer with other aspartase/fumarase family enzymes. Cartoon representation of apo EDDS lyase (PDB entry: 6G3D), argininosuccinate lyase ($\delta 2C$, PDB entry: 1HY1), adenylosuccinate lyase (ADL, PDB entry: 2PTR), aspartase (AspB, PDB entry: 3R6V), and fumarase (FumC, PDB entry: 1YFE). The N-terminal domain, central domain and C-terminal domain are depicted in blue, yellow and orange, respectively. The conserved regions C1, C2 and SS-loop are highlighted in red. The location of the N- and C-terminus is also shown.

5

as the one obtained with 4 M NaCl, with identical cell parameters, and allowed a structure determination at 1.9 Å resolution. The overall structure of EDDS lyase in the crystals grown with sodium formate is identical to that in the crystal obtained with sodium chloride, including the conformation of the SS-loop (the structures superpose with a r.m.s.d value of only 0.26 Å). Interestingly, however, extra wedge-shaped electron density clearly indicated the presence of three formate ions bound in the active site (Figure 4a, Figure S4a). One formate ion (FMT1) makes hydrogen bonds to Ser111, Arg112 (C1 region) from subunit A and Ser281 (SS-loop) from subunit C, while another formate ion (FMT2) makes hydrogen bonds to Asn113 (C1 region) from subunit A, Thr158 (C2 region) from subunit B and Lys286, Asn288 (SS-loop) from subunit C. The third formate ion (FMT3) is bound further away from the SS-loop making an interaction with Arg112. Based on a comparison with other aspartase/fumarase superfamily structures it was realized that two of these formate ions (FMT1 and FMT2) likely mimic the carboxylate groups of the succinyl moiety of the substrate, or of the product fumarate, when bound to the active site.

To investigate this further, we repeated the crystallization experiments using solutions in which sodium formate was substituted with sodium salts of fumaric acid or succinic acid. Large cubic-shaped crystals appeared overnight at conditions containing 0.2-0.3 M concentration of either salt, allowing the structure determination of fumarate-bound and succinate-bound EDDS-lyase at 2.2 Å and 2.6 Å resolution, respectively (Table S2, Figure 4b,c and Figure S4b,c). The fumarate and succinate molecules are tightly coordinated at the active site with overall identical binding interactions. As expected, the C1 and C4 carboxylate groups of fumarate/succinate occupy identical positions as the formate ions FMT1 and FMT2 discussed above, making similar hydrogen bonds with residues from the C1, C2 and SS-loop region. Most notably, the hydroxyl group of Ser280 in the SS-loop

Figure 4. Figure depicting formate (a), fumarate (b), succinate (c), (*S,S*)-EDDS (d) and AEAA (e) bound in the active site of EDDS lyase. The residues interacting with the ligands are shown as sticks in different colours (same coloring scheme as in Figure 2b). The grey mesh, contoured at 3σ , shows the simulated annealing mFo-DFc omit map.

(subunit C) is positioned at a close distance ($\sim 3 \text{ \AA}$) and proper orientation from the C β -atom in fumarate/succinate, in accordance with its presumed role as catalytic base in the α,β -elimination reaction.

EDDS- and AEAA-bound structures

Using a similar approach as described above, crystals diffracting to $2.2 \text{ \AA}$ resolution were obtained of EDDS lyase grown in the presence of 0.3 M (*S,S*)-EDDS. Since the enzyme is able to convert (*S,S*)-EDDS to fumarate and ethylenediamine, we expected to observe a fumarate ion bound at the active site. Much to our surprise, however, the electron density at the active site was consistent with the presence of an intact (*S,S*)-EDDS molecule, allowing the determination of the substrate-bound structure (Table S2, Figures 4d and 5). Apparently, the conditions for cleavage of (*S,S*)-EDDS in the crystal are not ideal, which may be explained by the non-optimal pH ($\text{pH} = 6.5$) of the crystallization solution (optimal pH for activity is 8) and/or insufficient protein flexibility preventing the deamination step and effectively trapping (*S,S*)-EDDS in the active site. One of the two succinyl-moieties of (*S,S*)-EDDS (we refer to it as the proximal succinyl, based on its close distance to the SS-loop) binds at the active site similar as fumarate or succinate in the other EDDS lyase structures, making identical interactions with the C1 and C2 regions and the SS-loop (Figure 6). The other (distal) succinyl moiety points away from the active site, and is bound

near to the protein surface, with one of the carboxylate groups forming hydrogen bonds with residues Arg294 and Tyr26 from chain C of the functional tetramer. The other largely solvent exposed carboxylate group makes a water-mediated hydrogen bond with Asn113. A comparison of the formate- and EDDS-bound structures shows that although the third formate ion (FMT3) and the carboxylate group of the distal succinyl moiety bind at similar regions of the active site pocket, their actual binding site locations and geometries are quite different (see Figure 4A and 4D). Thus unlike FMT1 and FMT2, FMT3 indeed does

Figure 5. a) Stereo view of the interactions of bound (S,S)-EDDS in the active site. Dashed lines show the hydrogen bonding interactions and the labels show the distances in Å. b) Schematic representation of the interactions between (S,S)-EDDS and active site residues. Hydrogen bonds are represented as dashed lines and letters in the parentheses denote the polypeptide chain identifier. The putative role of Ser280 as a catalytic base, abstracting the proton from the C β atom of (S,S)-EDDS, is shown with an arrow.

Figure 6. Superposition of bound (S,S)-EDDS (wheat), fumarate (light blue), and succinate (pale cyan) in the active site of EDDS lyase. The SS-loop of the (S,S)-EDDS-bound, fumarate-bound and succinate-bound structures, with putative catalytic base Ser280, is shown in green, yellow and magenta, respectively. The distance in Å from the hydroxyl group of Ser280 to the C β atom of the (S,S)-EDDS substrate is shown as dashed line.

not mimic the binding of one of the carboxylates of (S,S)-EDDS. In addition, the internal amino group of (S,S)-EDDS that is directly linked to the proximal succinyl moiety forms a hydrogen bond with Asn113, while the other amino group linked to the distal succinyl moiety makes water-mediated hydrogen bonds to Asn288 and Asp290.

To obtain a full description of all relevant structures in the catalytic cycle of EDDS lyase, crystals were also grown in the presence of the reaction intermediate AEAA. Surprisingly, instead of AEAA a molecule of (S,S)-EDDS was found in the active site bound in an identical conformation as in the structure obtained from a crystal grown in the presence of (S,S)-EDDS (data not shown). A possible explanation for this result is that EDDS lyase in solution will reversibly cleave AEAA to fumarate and ethylenediamine. EDDS lyase likely first crystallizes in a fumarate-bound state, after which (S,S)-EDDS can be produced in the crystal in one step by attack of AEAA. This can happen as long as the concentration of AEAA in the crystallization solution is sufficiently high. Upon formation of (S,S)-EDDS in the crystal, it will remain trapped in the active site (as shown by the co-crystallization with (S,S)-EDDS). In another attempt to obtain an AEAA-bound structure, a fumarate-bound crystal (prepared by crystallizing the enzyme in the presence of 0.3 M fumarate) was briefly soaked for ~30 seconds in mother liquor containing 20 mM ethylenediamine. The crystal was then immediately flash-cooled in liquid nitrogen and used for X-ray diffraction data collection. Difference Fourier analysis of the active site revealed a small adduct at the C α carbon of bound fumarate, most likely resulting from the addition of ethylenediamine to form the intermediate AEAA. Subsequent refinement yielded an AEAA-bound structure at 2.4 Å resolution (Table S2, Figure 4e, Figure S4d). It should be noted that in this structure the ethylenediamine moiety of the bound AEAA molecule is relatively disordered, as indicated by the absence of electron density for the terminal, free amino group. In addition,

a second molecule of fumarate is observed near to AEAA, bound at a location similar to where the distal succinyl moiety of (S,S)-EDDS binds. Possibly, this second molecule of fumarate prevented a quick release of AEAA, thereby inhibiting the full conversion towards (S,S)-EDDS.

Mutagenesis of active site residues

The crystal structure of the enzyme in complex with (S,S)-EDDS (Figure 5) suggests important roles for Ser280 and Asp290 in catalysis and substrate binding, respectively. Ser280 is positioned near the C β proton of the substrate and in a suitable orientation to allow proton abstraction to initiate the deamination reaction. Asp290 forms a water mediated hydrogen bond with the internal amino group connected to the distal succinyl moiety of (S,S)-EDDS, which appears to be an important interaction for binding and positioning of ethylenediamine for addition to fumarate. To study the importance of these residues for catalytic activity, each residue was replaced with an alanine. The mutation of Ser280 to an alanine resulted in an inactive enzyme (data not shown), confirming the essential role of this residue in catalysis. The mutation of Asp290 to an alanine resulted in a mutant enzyme that displayed significantly lower activity than wild-type EDDS lyase for the addition of ethylenediamine to fumarate (Figure S5).

5

DISCUSSION

The crystal structures of EDDS lyase determined in this study confirm a tertiary and quaternary fold that is characteristic of the aspartase/fumarase superfamily and provide the first detailed view of the substrate- and product-bound active site of an EDDS lyase, allowing an understanding of the roles of the various active site residues in substrate binding and catalysis (Figure 7). In all members of the aspartase/fumarase superfamily, a strictly conserved serine residue from the SS-loop has been implicated to act as the catalytic base, abstracting a proton from the C β atom of the substrate to initiate the α,β -elimination reaction.^{12, 34-37} In EDDS lyase, the equivalent SS-loop serine residue, Ser280, most likely functions as the catalytic base. Indeed, in both the (S,S)-EDDS- and AEAA-bound structures Ser280 is positioned near (~ 3 Å) the C β proton of the substrate (Figures 5, 6 and S4d) and in a suitable orientation to allow proton abstraction in the first step of the reaction (Figure 7). As expected, mutation of Ser280 to an alanine resulted in an inactive enzyme. However, for Ser280 to function as the catalytic base, it needs to be activated to form a Ser-O $^-$ oxyanion. Possible mechanisms for activation of the catalytic serine in aspartase/fumarase superfamily members have been suggested previously, including substrate-assisted deprotonation of the serine residue to generate the Ser-O $^-$ oxyanion.¹² The serine oxyanion could be stabilized by main chain amide interactions. In the structure of EDDS lyase, Ser280 is within hydrogen-bonding distance of the neighbouring main-chain backbone amides of Ser281, Ile282 and Met283, which form part of the SS-loop.

The question how Ser280 is activated to function as the catalytic base thus remains to be answered.

The extensive number of hydrogen-bonding interactions observed for the binding of (*S,S*)-EDDS and AEAA (Figures 5 and S4d) forces the substrates to adopt an energetically unfavourable rotamer conformation in which the C α , C β and β -carboxylic atoms are coplanar. This is consistent with a conformation resembling a putative enediolate (*aci*-carboxylate) intermediate during catalysis (Figure 7).^{12,36} The oxygens of the β -carboxylate group of (*S,S*)-EDDS and AEAA form hydrogen bonds with the side chain hydroxyls of Ser111 (subunit A) and Ser281 (subunit C, SS-loop), the guanidinium moiety of Arg112 (subunit A) and main chain amides of Arg112 and Ser281. This extensive hydrogen-bonding network is not only essential for substrate binding but also plays a crucial role in stabilizing the additional negative charge that develops on one of the β -carboxylate oxygens as a result of C β -proton abstraction by Ser280. The presence of a positively charged residue, Arg112, has an additional stabilizing effect on the negatively charged *aci*-carboxylate intermediate. Both (*S,S*)-EDDS and AEAA are expected to undergo similar catalytic mechanisms for the cleavage of the C-N bond (Figure 7). Hence, our results, combined with earlier mechanistic studies on other superfamily members,^{12,34-37} support a mechanism that involves general base-catalyzed formation of a highly stabilized enediolate (or *aci*-carboxylate) intermediate during the sequential two-step deamination of (*S,S*)-EDDS.

The water mediated hydrogen bond between Asp290 and the internal amino group connected to the distal succinyl moiety of (*S,S*)-EDDS (Figure 5) appears to be an important interaction for binding and positioning of ethylenediamine (and other diamine substrates) for addition to fumarate. Indeed, EDDS lyase seems to display higher amination activity using diamines as substrates compared to monoamines (Table 1). The results showed that mutation of Asp290 to an alanine resulted in significant loss of activity of EDDS lyase for the addition of ethylenediamine to fumarate (Figure S5).

The crystal structures may also provide a possible explanation for the very broad substrate scope of EDDS lyase, which accepts a wide variety of amines (Table 1) and amino acids with terminal amino groups¹⁴ for selective addition to fumarate. In the structures of most members of the aspartase/fumarase superfamily the SS-loop is intrinsically flexible and undergoes an open-closed transition upon binding of substrate.^{12, 34-37} This substrate-dependent movement of the SS-loop is crucial because it positions catalytic and substrate-binding residues in a suitable orientation for catalysis, and is accompanied by a closure of the active site pocket. Hence, the rather narrow substrate range of members of this superfamily may mirror this complex step in the catalytic mechanism. In EDDS lyase, however, the occurrence of an open-closed transition of the SS-loop is highly unlikely; additional interactions with the N-terminal loop region stabilize a conformation of the SS-loop that is highly similar to the catalytically competent closed conformation observed in the structures of other superfamily members, even when no substrate is bound. Thus, in contrast to superfamily lyases where substrate-dependent movement of the SS-loop positions catalytic and substrate-binding residues in a suitable orientation

Figure 7. Proposed catalytic mechanism for the deamination of (S,S)-EDDS to yield AEAA and fumarate.

for catalysis and results in closure of the active site pocket, EDDS lyase appears to have a fully competent catalytic machinery in the uncomplexed enzyme, its SS-loop does not seem to undergo conformational changes upon substrate binding, its active site pocket remains more accessible and the pocket for the amine substituent is less defined. It is tempting to speculate that these unique structural properties reflect its very broad amine substrate scope and make EDDS lyase a more promising template for redesign to convert new unnatural substrates.

The crystal structures of EDDS lyase in complex with AEAA and (S,S)-EDDS set the stage for structure-based engineering of this fascinating enzyme. Improved variants of EDDS lyase with enhanced catalytic activity and expanded unnatural substrate scope will provide a powerful synthetic tool for the preparation of diverse complex molecules by C-N bond formation. We have therefore initiated studies aimed at engineering of EDDS lyase, guided by the crystal structures reported here, to increase its synthetic usefulness in the production of metal-chelating aminocarboxylic acids as potential metallo- β -lactamase inhibitors, amino acid precursors for artificial dipeptide sweeteners, and substituted aspartic acids as aspartate/glutamate transporter inhibitors. The results of these studies will be reported in due course.

AUTHOR CONTRIBUTION

H.P. carried out all the crystallography experiments and determination of the different structures. J.dV. performed the substrate scope analysis and J.Z. constructed and characterized the D290A mutant. V.P.V. identified and cloned the gene and H.R. synthesized N-(2-aminoethyl)aspartic acid. A.M.W.H.T. and G.J.P. supervised the scientific work. All authors contributed to writing the paper.

DECLARATION

Dr. Jieli Zhang also used this paper as a chapter in her PhD thesis entitled Biocatalytic asymmetric hydroamination by native and engineered carbon-nitrogen lyases: New enzymes to prepare amino acid precursors to pharmaceuticals and food additives. (University of Groningen, 2019, ISBN: 978-94-034-1845-2).

ACKNOWLEDGEMENTS

The authors thank Marianne de Villiers, Haigen Fu and Niels van Oosterwijk for insightful discussions and Marijke Jansma for assistance with enzyme purification and crystallization. The authors also wish to thank Johan Hekelaar for his assistance with data collection using an in-house X-ray generator. This research was financially supported by ECHO grant 700.59.042 from the Division of Chemical Sciences of the Netherlands Organisation of Scientific Research, the European Union 7th framework project Metaexplore (KBBE-2007-3-3-05, grant agreement number 222625), and by the European Research Council

under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement number 242293. Jieliin Zhang acknowledges funding from the China Scholarship Council.

REFERENCES

1. Bucheli-Witschel, M., and Egli, T. (2001) Environmental fate and microbial degradation of aminopolycarboxylic acids. *FEMS Microbiol. Rev.* 25, 69-106.
2. Oviedo, C., and Rodriguez, J. (2003) EDTA: The chelating agent under environmental scrutiny. *Quim. Nova.* 26, 901-905.
3. Madsen, E. L., and Alexander, M. (1985) Effects of chemical speciation on the mineralization of organic compounds by microorganisms. *Appl. Environ. Microbiol.* 50, 342-349.
4. Allard, A. -S., Renberg, L., and Neilson, A. H. (1996) Absence of $^{14}\text{CO}_2$ from ^{14}C -labelled EDTA and DTPA and the sediment/water partition ratio. *Chemosphere* 33, 577-583.
5. Nörtemann, B. (1999) Biodegradation of EDTA. *Appl. Microbiol. Biotechnol.* 51, 751 -759.
6. Nishikiori, T., Okuyama, A., Naganawa, T., Takita, T., Hamada, M., Takeuchi, T., Aoyagi, T., and Umezawa, H. (1984) Production by actinomycetes of (S,S)-N,N'-ethylenediaminedisuccinic acid, an inhibitor of phospholipase C. *J. Antibiot.* 37, 426-427.
7. Spohn, M., Wohlleben, W., and Stegmann, E. (2016) Elucidation of the zinc-dependent regulation in *Amycolatopsis japonicum* enabled the identification of the ethylenediamine-disuccinate ([S,S]-EDDS) genes. *Environ. Microbiol.* 18, 1249-1263.
8. Witschel, M., and Egli, T. (1998) Purification and characterization of a lyase from the EDTA-degrading bacterial strain DSM 9103 that catalyzes the splitting of [S, S]- ethylenediaminedisuccinate, a structural isomer of EDTA. *Biodegradation* 8, 419-428.
9. Egli, T. (2001) Biodegradation of metal-complexing aminopolycarboxylic acids. *J. Biosci. Bioengi.* 92, 89-97.
10. Schowanek, D., Feijtel, T. C. J., Perkins, C. M., Hartman, F. A., Federle, T. W., and Larson, R. J. (1997) Biodegradation of [S, S], [R, R] and mixed stereoisomers of ethylenediamine disuccinic acid (EDDS), a transition metal chelator. *Chemosphere* 34, 2375-2391.
11. Mizunashi, W. (2001) Protein having ethylenediamine-N,N' disuccinic acid: Ethylenediamine lyase activity and gene encoding the same. United States Patent, Patent no. US 6,168,940 B1.
12. Puthan Veetil, V., Fibriansah, G., Raj, H., Thunnissen, A.-M. W. H., and Poelarends, G. J. (2012) Aspartase/fumarase superfamily: A common catalytic strategy involving general base-catalyzed formation of a highly stabilized aci-carboxylate intermediate. *Biochemistry* 51, 4237-4243.
13. Nörtemann, B. (1992) Total degradation of EDTA by mixed cultures and a bacterial isolate. *Appl. Environ. Microbiol.* 58, 671-676.
14. Fu, H., Zhang, J., Saifuddin, M., Cruiming, G., Tepper, P. G., and Poelarends, G. J. (2018) Chemoenzymatic asymmetric synthesis of the metallo-beta-lactamase inhibitor aspergillomarasamine A and related aminocarboxylic acids. *Nat. Catal.* 1, 186-191.
15. Waddell, W. J. (1956) *J. Lab. Clin. Med.* 48, 311-314.
16. Raj, H., Szymański, W., de Villiers, J., Rozeboom, H. J., Veetil, V. P., Reis, C. R., de Villiers, M., Dekker, F.J., de Wildeman, S., Quax, W. J., Thunnissen, A.-M. W. H., Feringa, B. L., Janssen, D. B., and Poelarends, G. J. (2012) Engineering methylaspartate ammonia lyase for the asymmetric synthesis of unnatural amino acids. *Nat. Chem.* 4, 478-484.
17. Evans, P. R. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* 62, 72-82.
18. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta. Crystallogr. D Biol. Crystallogr.* 69, 1204-1214.
19. Collaborative Computational project, Number 4 (1994) The CCP4 suite: programs for protein crystallography *Acta Crystallogr. D Biol. Crystallogr.* 50, 760-763.
20. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658-674.

21. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005) FFAS03: server for profile-profile sequence alignments. *Nucl. Acids Res.* 33, W284-288.
22. Sampaleanu, L. M., Coddling, P. W., Lobsanov, Y. D., Tsai, M., Smith, G. D., Horvatin, C., and Howell, P. L. (2004) Structural studies of duck delta2 crystallin mutants provide insight into the role of Thr161 and the 280s loop in catalysis *Biochem. J.*, 384, 437-447.
23. Tsai, M., Sampaleanu, L. M., Greene, C., Creagh, L., and Haynes, C. (2004) A duck delta1 crystallin double loop mutant provides insight into residues important for argininosuccinate lyase activity. *Biochemistry* 43, 11672-11682.
24. Langer, G., Cohen, S. X., Lamzin, V.S., and Perrakis, A. (2008) Automated macromolecular model building or X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* 3, 1171-1179.
25. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240-255.
26. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486-501.
27. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution *Acta. Crystallogr. D. Biol. Crystallogr.* 66, 213-221.
28. Schüttelkopf, A. W., and van Aalten, D. M. F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* 60, 1355-1363.
29. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12-21.
30. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205-217.
31. Robert, X., and Gouet, P., (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* 42(W1), W320-W324.
32. Krissinel, E., and Henrick, K., (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta. Crystallogr. D. Biol. Crystallogr.* 60, 2256-2268
33. The PyMOL molecular graphics system, version 1.6. Schrodinger, LLC.
34. Sampaleanu, L. M., Vallee, F., Slingsby, C., and Howell, P. L. (2001) Structural studies of delta 1 and delta 2 crystallin suggest conformational changes occur during catalysis. *Biochemistry* 40, 2732-2742.
35. Tsai, M., Koo, J., Yip, P., Colman, R. F., Segall, M. L., and Howell, P. L. (2007) Substrate and product complexes of *Escherichia coli* adenylosuccinate lyase provide new insights into the enzymatic mechanism. *J. Mol. Biol.* 370, 541-554.
36. Fibriansah, G., Puthan Veetil, V., Poelarends, G. J., and Thunnissen, A.-M. W. H. (2011) Structural basis for the catalytic mechanism of aspartate ammonia lyase. *Biochemistry* 50, 6053-6062.
37. Mechaly, A. E., Haouz, A., Miras, I., Barilone, N., Weber, P., Shepard, W., Alzari, P.M., and Bellinzoni, M. (2012) Conformational changes upon ligand binding in the essential class II fumarase Rv1098c from *Mycobacterium tuberculosis*. *FEBS Lett.* 586, 1606-1611.

SUPPORTING INFORMATION

Scheme S1. Enzymatic synthesis of the natural aminocarboxylic acid products Aspergillomarasmine A and Aspergillomarasmine B using EDDS lyase as the biocatalyst.

Figure S1. Sequence alignment of EDDS lyases from *Chelativorans* sp. BNC1 and *Brevundimonas* sp. TN3. These two EDDS lyase amino acid sequences share ~79% sequence identity.

Figure S2. Graphs showing the pH and temperature optima for the EDDS lyase catalyzed reaction. The initial rates of the EDDS lyase catalyzed deamination of (S,S)-EDDS were plotted against pH (top graph) or temperature (bottom graph).

Figure S3. Structure-based sequence alignment of EDDS lyase and members of the aspartate/fumarate superfamily of enzymes. The conserved regions C1, C2 and SS-loop are shown in red and underlined in the alignment. The secondary structure of EDDS lyase is shown above the alignment. EDDSL – EDDS lyase, d2C – duck delta II crystallin, ADL – *E. coli* adenylosuccinate lyase, AspB – aspartase from *Bacillus* sp. YM55-1, FumC – *E. coli* fumarase.

Figure S4. Schematic representation of the interactions between formate (a), succinate (b), fumarate (c) and AEAA (d) and active site residues of EDDS lyase. Hydrogen bonds are represented as dashed lines and letters in the parentheses denote the polypeptide chain identifier. The putative role of Ser280 as a catalytic base is highlighted with an arrow in (d).

Figure S5. Progress curves of the addition of ethylenediamine (100 mM) to fumaric acid (5 mM) catalyzed by wild-type EDDS lyase and the D290A mutant (each 0.5 mg/ml) in 20 mM NaH_2PO_4 buffer (pH 8.5). Reaction progress was monitored at 270 nm by UV spectrophotometry.

Table S1. Influence of amine/fumarate ratio on product formation of EDDS lyase-catalyzed reactions.

Entry	Substrate	Substrate Structure	Observed Products ^a	
			20-Fold Excess of Amine	2-Fold Excess of Fumaric Acid
1	4a			
2	4k			
3	4l			
4	4m			
5	4q			
6	4r			
7	4s			

^aProducts were identified by LC-MS/MS

Table S2. Data collection and refinement statistics

PDB ID	Native EDDS lyase	Formate-bound	Fumarate-bound	Succinate-bound	EDDS-bound	AEAA-bound
Data collection	6G3D	6G3E	6G3F	6G3G	6G3H	6G3I
Space group	F222	F222	F222	F222	F222	F222
Unit cell <i>a</i> , <i>b</i> , <i>c</i> (Å)	144.4, 144.8, 147.6	143.4, 144.3, 145.8	145.6, 146.0, 152.4	145.1, 145.9, 151.1	144.6, 146.4, 147.1	145.5, 146.4, 152.8
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	51.7 – 2.2	41.7 – 1.9	52.7 – 2.2	52.5 – 2.6	44.3 – 2.2	52.9 – 2.4
R_{merge}	0.151 (0.544)*	0.15 (0.73)	0.035 (0.098)	0.06 (0.19)	0.075 (0.34)	0.115 (0.60)
$\langle I/\sigma \rangle$	9.9 (2.8)	11.2 (2.6)	26.0 (9.5)	28.5 (6.7)	5.1 (1.7)	12.2 (4.1)
Completeness (%)	99.0 (94.4)	99.0 (97.5)	99.3 (95.1)	99.5 (96.5)	98.5 (91.4)	99.9 (99.5)
Redundancy	4.1 (3.9)	6.9 (6.5)	3.7 (3.4)	4.9 (4.7)	5.5 (5.2)	4.1 (4.1)
CC1/2	0.99 (0.66)	0.99 (0.66)	0.99 (0.98)	0.99 (0.96)	0.98 (0.72)	0.99 (0.82)
Refinement						
Resolution (Å)	51.6 – 2.2	19.9 – 1.9	52.6 – 2.2	51.4 – 2.6	43.1 – 2.2	51.6 – 2.4
No. reflections	37577	58476	39635	24456	35533	31534
$R_{\text{work}} / R_{\text{free}}$	18.2 / 22.4	14.5 / 16.6	16.8 / 19.4	19.0 / 22.3	18.1 / 21.5	19.7 / 23.3
No. of residues						
Protein	496	497	497	497	497	496
Water	261	449	259	114	214	167
Ligand/ion	----	3 Formate, 1 Na	1 Fumarate, 8 PEG	1 Succinate, 8 PEG	1 EDDS	1 AEAA, 1 Fumarate
B-Factors (Å²)						
Protein	25.1	17.4	26.1	34.3	41.6	33.3
Ligand	----	22.5	42.7	47.5	36.6	32.2
R.m.s. deviations						
Bond lengths (Å)	0.003	0.004	0.003	0.003	0.003	0.003
Bond angles (°)	0.550	0.917	0.543	0.591	0.579	0.582

Table S2. (continued)

	Native EDDS lyase	Formate-bound	Fumarate-bound	Succinate-bound	EDDS-bound	AEAA-bound
Ramachandran						
Favoured (%)	98.0	98.4	98.0	98.0	96.6	97.2
Allowed (%)	1.8	1.4	1.8	1.6	3.0	2.4
Outlier (%)	0.2	0.2	0.2	0.4	0.4	0.4

*Values in parentheses are for highest-resolution shell

c h a p t e r s i x

SUMMARY AND FUTURE PERSPECTIVES

Nature has evolved enzymes to carry out important biotransformations. The physiological role played by the different enzymes is of utmost importance and all life forms depend on them for their proper functioning. It was a long held belief that enzymes have evolved for a specific function, i.e., they are capable of only catalyzing one type of reaction on a specific substrate. However, decades of studies have somewhat busted this myth and it is now known that although enzymes are very efficient at catalyzing one reaction for which they have evolved, they are also capable of exhibiting catalytic promiscuity. Promiscuity in this context refers to the ability of enzymes to catalyze fortuitous alternative reaction(s) in addition to their main catalytic role. The incidence of catalytic promiscuity is more pronounced in members belonging to the same superfamily, with often one member being able to catalyze the reaction performed by another member. This property of enzymes showing catalytic and substrate promiscuity can be exploited to our advantage to perform complex reactions and generate compounds of economic value. Using enzymes as biocatalysts has advantages over traditional organic synthesis. Firstly, enzymes perform their function in milder conditions, which is both environmentally friendly and also relatively cheaper compared to expensive catalysts that generally work under harsh conditions. Secondly, enzymes catalyze their reactions in a stereo- and regio-selective manner, which is very difficult to control when working with traditional catalysts. However, there are also certain disadvantages when it comes to efficiently using enzymes for synthesis of target compounds. One of the biggest bottlenecks is the low-level turnover, exhibited by enzymes especially for the promiscuous reactions. The narrow substrate scope for any given reaction catalyzed by an enzyme is also its Achilles heel and prevents their large-scale application in industry. Therefore, considerable efforts are directed towards improving an enzyme's catalytic performance by using tools of directed evolution. Laboratory evolution of enzymes is immensely aided by knowledge of its structure.

X-ray crystallography is by far the most widely used technique to study the structure of biological macromolecules. Briefly, X-ray crystallography involves obtaining protein crystals, which have an ordered and repeating arrangement of atoms, and shooting them with X-rays. The resultant diffraction pattern can be used to compute the electron density in the three-dimensional space. Based on the observed densities it is possible to place individual atoms in the three-dimensional space and obtain a complete model for a macromolecule. Phillips *et al.*, reported the first crystal structure of the hen egg-white lysozyme (1965 *Nature* 20:757-761) and a structure in complex with an inhibitor (1965 *Nature* 20:761-763). This opened up new avenues for research into understanding how enzymes work at the molecular level. The structure in complex with the inhibitor provided important insights into the active site of the enzyme and also the crucial role individual residues play in inhibitor binding and possibly during catalysis. More importantly it was proven that X-ray crystallography could be used to obtain snapshots of the events unfolding during catalysis. Since then X-ray crystallography has been the technique of choice for researchers to solve structures of enzymes and gain insights into their catalytic mechanism. This structural knowledge helps the researchers to further improve

the efficiency or substrate scope of enzymes by identifying and rectifying the elements, which act as barriers in proper functioning of these enzymes.

This thesis describes three enzymes belonging to the tautomerase (4-OT and RhCC) and aspartase/fumarase (EDDS lyase) superfamily of enzymes. We have extensively used X-ray crystallography to gain insights into the structures and mechanisms of these three enzymes. 4-Oxalocrotonate tautomerase (4-OT), the most widely studied member of the tautomerase superfamily of enzymes, has attracted a lot of interest for its ability to promiscuously catalyze C-C bond-forming aldol and Michael-type addition reactions. It was proposed that the first step in the catalytic cycle would involve formation of an enamine species between the catalytic Pro-1 residue and the carbonyl substrate. Indeed, using X-ray crystallography we have solved a 1.7Å resolution structure (PDB id 4X1C) and showed the covalent modification of Pro-1 with acetaldehyde to yield an enamine intermediate (**Chapter 2**). This nucleophilic intermediate undergoes reaction with electrophiles such as benzaldehyde and *trans*-β-nitrostyrene, resulting in C-C bond formation. Furthermore, using NMR studies it was shown that H-D exchange within acetaldehyde is catalyzed by 4-OT, and Pro-1 is crucial for this activity. The results discussed in this chapter provided important insights into the mechanism of C-C bond-forming reactions catalyzed by 4-OT. It is the first instance where the structure of a tautomerase superfamily member was reported with an enamine adduct on the Pro-1 residue. This sets the stage for further investigations into the mechanistic intricacies of these important reactions. Similar structures of 4-OT in complex with the electrophiles (e.g. benzaldehyde, *trans*-β-nitrostyrene, etc.) and products (e.g. cinnamaldehyde and 4-nitro-3-phenylbutanal) will help in explaining the spatial positioning of these compounds with respect to Pro-1 and other active site residues in 4-OT during catalysis. This knowledge will help us to improve the activity and specificities of these reactions.

In **Chapter 3**, we report mutability landscapes that were generated for soluble expression, natural tautomerase activity and promiscuous C-C bond-forming Michael-type addition activity of 4-OT. Exploiting the small monomeric size of 4-OT, only 62 residues, a library of genes encoding for almost all possible single mutants of 4-OT was constructed and the effect of individual mutations on the above mentioned activities was studied. A single mutant A33D was identified with significantly improved activity and enantioselectivity compared to wild-type 4-OT for the Michael-type addition reaction. Combining two favorable mutations resulted in a double mutant M45Y/F50A which showed good enantioselectivity for the addition of acetaldehyde to *trans*-β-nitrostyrene, producing the opposite product enantiomer compared to both wild-type and A33D mutant of 4-OT. Intriguingly, the structure analysis of unliganded and *trans*-β-nitrostyrene bound M45Y/F50A (PDB id 5CLO) revealed the significant reorganization of the active site as a result of these two mutations. Most unexpectedly the Tyr45 in the mutant flipped away from the active site, and together with the absence of a bulky side-chain at position Ala50, this allowed the opening of a hydrophobic pocket. It was shown that this new pocket can accommodate the phenyl ring of *trans*-β-nitrostyrene and it is likely that this is responsible

for the altered enantioselectivity and improved activity. The potential to use 4-OT to produce precursors of pharmaceutically active compounds is immense. In conjunction with other enzymes 4-OT can be used in cascade reactions where it can use its ability to introduce C-C bonds to build larger products from simpler precursors. The vast amount of data collected by making mutability landscapes will go a long way in our understanding of sequence function relationships.

Discovery of new members in a superfamily of enzymes opens new research opportunities and gives an opportunity to characterize possible new activities in the superfamily. **Chapter 4** describes a new member of the tautomerase superfamily, RhCC. Mechanistic characterization showed that RhCC acts as an oxygenase, using molecular O₂ as a co-substrate, and incorporating it into substrate 4-hydroxyphenylenol pyruvate (4HPP). Structure elucidation revealed that RhCC possesses the signature β - α - β structural fold, which is characteristic of the tautomerase superfamily of enzymes (PDB id 4U5P). Interestingly, a metal binding site was identified in the enzyme at the non-crystallographic three-fold symmetry axis with the bound metal being Mg²⁺. To prove that this metal does not play any role during catalysis or is needed to maintain structural integrity, the metal binding residue was mutated. The structure of the resultant mutant (D106A) was also solved, which showed absence of any bound metal but no changes in the overall structure or any loss of activity. Therefore, it was suggested that RhCC is the first cofactor-independent oxygenase to be characterized in the tautomerase superfamily of enzymes. The concept of substrate-assisted catalysis where the substrate itself acts as a cofactor during catalysis is also a new feature in this superfamily. The emergence of cofactor-independent oxygenation reaction in the β - α - β structural fold is an exciting development and more research is needed to better understand the mechanism of this reaction. A structure in complex with the substrate 4HPP will be of immense help to understand the events during the catalytic cycle. Engineering of RhCC to limit the production of side products during the reaction cycle is also important. Identification and characterization of RhCC homologues from other microorganisms will also help comprehend the details of this unusual enzymatic reaction.

Members of the aspartase/fumarase superfamily have been of interest due to their ability to introduce C-N bonds into the products. EDDS lyase (**Chapter 5**), a new member of this superfamily, was shown to reversibly degrade ethylenediamine-*N,N'*-disuccinic acid (EDDS, metal chelator) in an unusual two-step elimination reaction to yield fumarate and ethylenediamine as products via the intermediate *N*-(2-aminoethyl)aspartic acid (AEAA). Crystal structures showed a conserved quaternary fold, which is characteristic of the aspartase/fumarase superfamily (PDB id 6G3D). Using co-crystallization and soaking techniques we were able to solve high-resolution structures in complex with EDDS, AEAA and fumarate. This validated the two-step elimination reaction mechanism and also gave important information about the key residues in the active site of EDDS lyase. In addition to gaining structural insights we also explored the substrate scope of EDDS lyase and showed that the enzyme is capable of accepting different mono- and diamines as substrates for amination of fumarate, albeit with lower catalytic efficiency. The structural

and functional characterization of EDDS lyase sets the stage for further engineering of this interesting enzyme to improve its catalytic efficiency and to expand the substrate scope. EDDS lyase can be used in the production of different metal-chelating aminocarboxylic acids, which are used in large variety of domestic products and industrial applications, and chiral building blocks for pharmaceuticals, agrochemicals and food additives.

X-ray crystallography as a technique over the years has seen some radical advancements, which has further eased the process of solving structures. Advent of powerful synchrotron radiation, automation and improvement in screening conditions for crystallogenes together with development of software for data processing has made it possible for researchers to tackle challenging targets. In recent years, relatively new techniques such as X-ray free electron laser (XFEL) and electron microscopy (EM) have also emerged as powerful techniques to study macromolecules but they still have their limitations and can only be used for limited number of targets. It is an exciting time for structural biologists as new developments open new avenues. X-ray crystallography can coexist with the new techniques as a part of an integrated array of different methods, allowing structural biologists to address challenging questions.

NEDERLANDSE SAMENVATTING

In de natuur zijn enzymen essentieel voor het uitvoeren en reguleren van biotransformaties. Enzymen hebben verschillende fysiologische rollen en alle levensvormen zijn van hun juiste werking afhankelijk. Er werd lang aangenomen dat enzymen zijn geëvolueerd voor een specifieke functie, d.w.z. dat ze slechts één type reactie met een specifiek substraat kunnen katalyseren. Decennia van wetenschappelijke studies hebben deze mythe echter enigszins doorgeprikt en het is inmiddels bekend dat hoewel enzymen zeer efficiënt zijn in het katalyseren van één reactie waarvoor ze zijn geëvolueerd, ze ook vaak katalytische promiscuïteit vertonen. Promiscuïteit verwijst in deze context naar het vermogen van enzymen om, naast hun primaire katalytische functie, ook katalyse te vertonen voor een alternatieve reactie of met een alternatief substraat. Katalytische en substraat-promiscuïteit wordt vaker gezien bij leden die tot dezelfde enzym superfamilie behoren, waarbij vaak één lid de reactie van een ander lid kan katalyseren. Deze eigenschap van enzymen kan in ons voordeel worden benut om complexe reacties uit te voeren en om verbindingen met economische waarde te produceren. Het gebruik van enzymen als biokatalysatoren heeft veel voordelen ten opzichte van traditionele organische synthese. Ten eerste vervullen enzymen hun functie in mildere omstandigheden, die zowel milieuvriendelijk als goedkoper zijn in vergelijking met chemische katalysatoren. Ten tweede katalyseren enzymen hun reacties op een stereo- en regio-selectieve manier, iets wat erg moeilijk te realiseren is met traditionele chemische katalysatoren. Er kleven ook nadelen aan het gebruik van enzymen voor de synthese van specifieke chemische verbindingen. Met name met betrekking tot promiscue reacties vertonen enzymen vaak een lage substraat-omzettingssnelheid: dit verhindert hun grootschalige toepassing in de industrie. Om dit probleem aan te pakken wordt er ingezet op het verbeteren van de katalytische prestaties van enzymen met behulp van gerichte evolutie methoden. Kennis van de drie-dimensionale structuur van enzymen is hierbij van onmisbare waarde.

Röntgenkristallografie is verreweg de meest gebruikte techniek om de structuur van biologische macromoleculen te bestuderen. In het kort: röntgenkristallografie omvat het verkrijgen van eiwitkristallen, die een geordende en herhalende rangschikking van atomen hebben, en deze met röntgenstralen te belichten. De resulterende diffractiepatronen kunnen vervolgens worden gebruikt om de elektronendichtheid in de driedimensionale ruimte te berekenen. Op basis van deze electronendichtheden is het mogelijk om de individuele atomen van een eiwitmolecuul in de driedimensionale ruimte te plaatsen en zo een compleet atomair model te bouwen van een macromolecuul. In 1965 publiceerden Phillips et al. de eerste kristalstructuur van een enzym, namelijk van het eiwit lysozym uit kippenei (1965 Nature 20: 757-761). Tegelijkertijd publiceerden ze ook de structuur van lysozym met een gebonden remmermolecuul (1965 Nature 20: 761-763). Dit opende nieuwe wegen voor onderzoek om te begrijpen hoe enzymen op moleculair en atomair niveau werken. De structuur in complex met de remmer bracht het actieve centrum van het lysozym duidelijk in beeld, wat belangrijke inzichten opleverde over de cruciale rol die individuele aminozuren spelen bij de binding van de remmer en mogelijk tijdens de katalyse van de reactie. Vele andere publicaties hebben

vervolgens laten zien dat röntgenkristallografie ook kan worden gebruikt om snapshots te maken van de gebeurtenissen die zich tijdens enzymkatalyse ontfouwen. Inmiddels is röntgenkristallografie een van de favoriete technieken van onderzoekers om structuren van enzymen op te helderen en om inzicht te krijgen in hun katalytische mechanismen. Deze structurele kennis helpt onderzoekers om de efficiëntie in substraatomzetting en/of de breedte in substraat-herkenning van enzymen verder te verbeteren door de elementen die een belemmering vormen te identificeren en via mutagenese te corrigeren .

Dit proefschrift beschrijft drie enzymen die behoren tot de tautomerase (4-OT en RhCC) en aspartase/fumarase (EDDS lyase) enzym superfamilies. In ons onderzoek hebben we uitgebreid gebruik gemaakt van röntgenkristallografie om inzicht te krijgen in de structuren en mechanismen van deze drie enzymen. 4-Oxalocrotonaat-tautomerase (4-OT), het meest bestudeerde lid van de tautomerase superfamilie, heeft veel belangstelling getrokken vanwege zijn promiscue katalytische activiteit ten aanzien van C-C-bindingvormende aldol- en Michael-type additie-reacties. In het veronderstelde reactiemechanisme van dit enzym vormt zich tijdens de eerste reactie stap een enamine verbinding tussen het katalytische Pro-1 aminozuur en het carbonyl-substraat. Met behulp van röntgenkristallografie hebben we de driedimensionale structuur van een intermediaire 4OT reactie toestand opgehelderd bij een oplossend vermogen van 1.7 Å (PDB code 4X1C). Met deze structuur kon de aanwezigheid van een enamine tussenproduct, gevormd door covalente modificatie van Pro-1 door acetaldehyde, worden aangetoond (**hoofdstuk 2**). Dit nucleofiele tussenproduct kan vervolgens reacties aangaan met elektrofielen zoals benzaldehyde en trans- β -nitrostyreen, wat resulteert in C-C-bindingsvorming. Bovendien werd met behulp van NMR-experimenten aangetoond dat H-D-uitwisseling binnen het acetaldehyde door 4-OT wordt gekatalyseerd, en dat Pro-1 cruciaal is voor deze activiteit. De resultaten die in dit hoofdstuk worden besproken geven belangrijke inzichten in het mechanisme van C-C-bindingsvormende reacties die door 4-OT worden gekatalyseerd. Voor het eerst werd de structuur van een tautomerase-superfamilielid gerapporteerd met een enamine adduct op het Pro-1 aminozuur, wat de basis vormde voor verder onderzoek naar de mechanistische details van C-C-bindingsvormende reacties die door een proline worden gekatalyseerd. Vergelijkbare structuren van 4-OT in complex met elektrofielen (bijv. benzaldehyde, trans- β -nitrostyreen) en producten (bijv. cinnamaldehyde en 4-nitro-3-fenylbutanal) kunnen aantonen hoe tijdens de katalyse deze verbindingen zich positioneren ten opzichte van Pro-1 en andere aminozuren in het actieve centrum. Deze kennis zal ons helpen om de activiteit en specificiteit van 4-OT ten aanzien van deze reacties te verbeteren.

In **hoofdstuk 3** beschrijven we hoe zogenaamde enzym 'mutabiliteitslandschappen' werden verkregen met betrekking tot de oplosbare expressie, natuurlijke tautomerase activiteit en promiscue Michael-type additieactiviteit van 4OT. Dankzij de geringe grootte van de 4-OT polypeptide keten, dat slechts 62 residuen omvat, kon een bibliotheek worden geconstrueerd van 4OT genen die coderen voor bijna alle mogelijke enkelvoudige mutaties in het eiwit en werd het effect van individuele mutaties op de bovengenoemde

activiteiten bestudeerd. De enkelvoudige 4OT mutant A33D werd geïdentificeerd als een variant die, vergeleken met het wild-type 4-OT, een significant verbetering in activiteit en enantioselectiviteit voor de Michael-type additiereactie liet zien. Het combineren van twee gunstige mutaties resulteerde in de tweevoudige 4OT mutant M45Y/F50A, die aanzienlijke enantioselectiviteit vertoonde voor de additie van acetaldehyde met trans- β -nitrostyreen, waarbij het tegenovergestelde enantiomere product werd geproduceerd dan met het wild-type en de A33D mutant van 4-OT. Analyse van de niet-gebonden en met trans- β -nitrostyreen gebonden structuur van M45Y/F50A (PDB code 5CLO) toonde aan dat het actieve centrum als gevolg van de twee mutaties een significante reorganisatie ondergaat. Totaal onverwacht draait de zijketen van aminozuur Tyr45 in de mutant weg van het actieve centrum, wat samen met de verandering van een grote naar kleine zijketen door mutatie F50A, de opening van een hydrofobe pocket mogelijk maakt. Deze pocket is groot genoeg om de fenyl-ring van trans- β -nitrostyreen te huisvesten en geeft een verklaring voor de verandering in enantioselectiviteit en verbetering in activiteit. Er zijn legio mogelijkheden om 4-OT te gebruiken voor het produceren van precursors van farmaceutisch actieve verbindingen. In combinatie met andere enzymen kan 4-OT worden gebruikt in cascadereducties waarbij zijn potentie kan worden ingezet om C-C-bindingen te introduceren in eenvoudigere precursor moleculen om op die manier meer complexe verbindingen te synthetiseren. Het genereren van enzym mutabiliteitslandschappen zoals voor 4OT, en de analyse van de enorme schat aan gegevens die hierbij beschikbaar komt, is een belangrijke stap op weg naar het begrijpen en voorspellen van sequentie-functie relaties in enzymen.

De ontdekking van nieuwe leden in een enzym superfamilie opent nieuwe onderzoeksmogelijkheden en biedt een kans om nieuwe activiteiten in de superfamilie te karakteriseren. **Hoofdstuk 4** beschrijft een nieuw lid van de tautomerase superfamilie, RhCC. Er kon worden aangetoond dat dit enzym een oxygenase reactie katalyseert, waarbij moleculair zuurstof wordt ingebouwd in het substraat 4-hydroxyfenylenolpyruvaat (4HPP). De kristalstructuur van RhCC vertoont de kenmerkende β - α - β opvouwing van de tautomerase enzym superfamilie (PDB code 4U5P). Het vormt een trimeer waarin als unieke eigenschap een Mg^{2+} -bindingsplaats werd geïdentificeerd, dat zich op de drievoudige symmetrieas in het trimere grensvlak bevindt. Om te bewijzen dat dit metaal geen rol speelt bij de katalyse of nodig is voor de structurele integriteit van het eiwit, werd het metaalbindende aminozuur aspartaat gemuteerd in een alanine. De kristalstructuur van de mutant (D106A) was identiek aan de structuur van het wild-type RhCC, afgezien van de mutatie in de metaalbindingsplaats en de afwezigheid van een gebonden metaal ion. Bovendien vertoonde de mutant geen enkel verlies in oxygenase activiteit. Het lijkt er dus op dat RhCC de eerste gekarakteriseerde cofactor-onafhankelijke oxygenase is in de tautomerase superfamilie. Het concept van substraatondersteunde katalyse waarbij het substraat zelf tijdens de katalyse als cofactor fungeert, is ook een nieuw kenmerk in deze superfamilie. De ontdekking van een cofactor-onafhankelijke oxygenatiereactie voor een enzym met een β - α - β opvouwing is erg interessant en er is meer onderzoek nodig

om het mechanisme van deze reactie beter te begrijpen. Een structuur in complex met het substraat 4HPP zal hierbij enorm helpen. Engineering van RhCC om de productie van neven-producten tijdens de reactiecyclus te beperken is ook belangrijk. Identificatie en karakterisering van RhCC-homologen uit andere micro-organismen zal er tevens toe bijdragen om de details van deze buitengewone enzymatische reactie te begrijpen.

Leden van de aspartase/fumarase-superfamilie zijn interessant vanwege hun vermogen om C-N-bindingen in de producten te introduceren. EDDS lyase (**hoofdstuk 5**), een nieuw lid van deze superfamilie, bleek ethyleendiamine-N,N'-disuccinezuur (EDDS, metaalchelator) reversibel af te breken in een ongebruikelijke tweestaps eliminatiereactie, waarbij via het intermediair N-(2-aminoethyl)asparaginezuur (AEAA), fumaraat en ethyleendiamine als producten worden gevormd. Eiwitkristallografische analyse toonde aan dat EDDS lyase een geconserveerde quarternaire structuur heeft, die kenmerkend is voor de aspartase/fumarase superfamilie (PDB code 6G3D). Met behulp van co-kristallisatie technieken en methoden waarbij kristallen gedrenkt worden in substraat oplossingen waren we in staat hoge resolutie structuren op te helderen van EDDS lyase complexen met gebonden EDDS, AEAA en fumaraat. Dit valideerde het tweestaps eliminatiereactiemechanisme en gaf ook belangrijke informatie over de belangrijkste aminozuren in de actieve centrum van EDDS-lyase. Naast het verkrijgen van structurele inzichten, hebben we ook de substraatomvang van EDDS-lyase onderzocht en aangetoond dat het enzym verschillende mono- en diamines als substraten voor aminering van fumaraat kan accepteren, zij het met een lagere katalytische efficiëntie. De structurele en functionele karakterisering van EDDS-lyase opent de weg voor verdere engineering van dit interessante enzym met als doel om de katalytische efficiëntie te verbeteren en de substraatomvang uit te breiden. EDDS lyase kan worden ingezet voor de productie van chelerende aminocarbonsuren, die worden toegepast in een grote verscheidenheid aan huishoudelijke producten en industriële processen, en van chirale bouwstenen voor farmaceutica, landbouwchemicaliën en voedseladditieven.

Röntgenkristallografie als techniek heeft in de loop der jaren een aantal radicale verbeteringen doorgemaakt, wat het proces van het oplossen van eiwitstructuren verder heeft vergemakkelijkt. De komst van krachtige synchrotronstraling, automatisering en verbetering van het screeningsproces naar geschikte kristallisatiecondities, samen met de ontwikkeling van software voor dataverwerking, heeft het voor onderzoekers mogelijk gemaakt om uitdagende doelen aan te pakken. In de afgelopen jaren zijn relatief nieuwe technieken zoals de vrije elektronen röntgenlaser (XFEL) en elektronenmicroscopie (EM) ook naar voren gekomen als krachtige technieken om macromoleculen te bestuderen, maar ze hebben nog steeds hun beperkingen en kunnen slechts voor een beperkt aantal doelen worden gebruikt. Het is een spannende tijd voor structurele biologen, omdat nieuwe ontwikkelingen nieuwe wegen openen. Röntgenkristallografie kan in combinatie met de nieuwe technieken een onderdeel vormen van een geïntegreerd scala aan verschillende methoden, waarmee structurele biologen uitdagende vragen kunnen beantwoorden.

ACKNOWLEDGEMENTS

The first time I was in the city of Groningen was in December 2010 for my PhD interview. Although it was bitterly cold and windy (typical of Groningen!), the city looked inviting and its people affable. So when I was offered the position, I didn't think twice and accepted it without any hesitation. Looking back I can confidently say that it was one of the best decisions that I have ever made. During this PhD journey I came to know many wonderful people who have inspired and helped me and to whom I owe deep gratitude.

First and foremost, my heartfelt thanks to my supervisor, **Prof. dr. Gerrit J. Poelarends** for his guidance and support during the project. I couldn't have asked for a better mentor during the formative years of my research career. The independence with which you allowed me to conduct my research activities made it a very enjoyable and satisfying experience. Thanks for involving me in most of the active research lines in your group, which is also reflected in the research presented in this thesis. The different projects (even the failed ones) helped me broaden my knowledge and gave me the opportunity to learn protein crystallography. I am also grateful to my co-supervisor, **Dr. Andy-Mark W. H. Thunnissen** for allowing me to conduct research in his group. I have learnt a lot from your kind guidance and support. You were always available to answer any questions I had and were patient in your explanations. Many thanks for translating the summary and future perspectives section to Dutch.

I would also like to thank **Prof. dr. Wim J. Quax**, my second supervisor, for all his support and suggestions, and nice conversations during lab outings and other social events.

Special thanks to the members of the assessment committee - **Prof. dr. Dirk J. Slotboom**, **Prof. dr. Marco W. Fraaije** and **Prof. dr. Willem J.H. van Berkel** for their evaluation of my thesis.

My sincere thanks to the technicians, **Rita**, **Ronald** and **Pieter** for all their help over the years and also for being such pleasant colleagues. Thanks to **Johan**, for keeping the in-house X-ray generator up and running and teaching me how to collect data. Warm thanks to **Janita** and **Yvonne** for helping me out with all the official paperwork.

I had the good fortune of working with some of the most wonderful colleagues at both the departments. I would forever cherish the camaraderie I shared with all the members. To all the members of Gerrit's group – **Mehran**, **Yufeng**, **Bert-Jan**, **Edzard**, **Hans**, **Vinod**, **Jan-Ytzen**, **Jandre**, **Marianne**, **Ellen**, **Anna**, **Sabry**, **Jielin**, **Zainal**, **Tjie**, **Lieuwe**, thanks for your collaboration, friendship and support. Heartfelt thanks to some other members on the 9th floor, **Robbert** (for all your help with protein purification and also teaching me how to use the AKTA) and **Ykelien** (I had great fun organizing events for the lab with PhFun). I also wish to thank the members of the Biophysical Chemistry Department - **Bauke**, **Anke**,

Tjaard, Henriette, Ilona, Marco, Ali, Paul, for their help in the lab and all the useful suggestions and discussions. **Eswar** and **Niels**, thanks for your friendship, guidance and all the help.

Finally, I would like to express my deepest gratitude to my family. I am grateful to my brother, **Aniruddh** and my parents for always believing in me and encouraging me to pursue my dreams. None of this would be possible without you all. Thanks for always being there.

Harshwardhan Poddar