

# Structural and biochemical characterization of biotechnologically relevant enzymes

Inaugural-Dissertation  
to obtain the academic degree  
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy  
of Freie Universität Berlin

by

Nicole Dimos

2023



This work was carried out in the period of July 2020 to March 2023 under the supervision of Dr. Bernhard Loll at the Institute of Chemistry and Biochemistry, Freie Universität Berlin, Germany.

First Reviewer:

Dr. Bernhard Loll  
Structural Biochemistry  
Institute of Chemistry and Biochemistry  
Freie Universität Berlin  
Takustraße 6  
14195 Berlin, Germany

Second Reviewer:

Prof Markus C. Wahl  
Structural Biochemistry  
Institute of Chemistry and Biochemistry  
Freie Universität Berlin  
Takustraße 6  
14195 Berlin, Germany

Date of defense:

**16.05.2023**





## Selbstständigkeitserklärung

Hierdurch versichere ich, Nicole Dimos, dass ich meine Dissertation selbstständig verfasst und keine anderen als die von mir angegeben Quellen und Hilfsmittel verwendet habe. Geistiges Eigentum anderer Autoren wurde als entsprechend gekennzeichnet. Ebenso versichere ich, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt bzw. die Dissertation in dieser oder anderer Form an keiner anderen Fakultät als Dissertation vorgelegt habe.

Berlin, den 10.03.2023

Nicole Dimos



## Acknowledgement

First, I want to thank Dr. Bernhard Loll for the opportunity to work on so many interesting projects under his great supervision. Without your support I wouldn't have made it so far.

I would also like to thank Prof. Markus Wahl for the chance to work in his lab. I am very thankful to Prof. Thomas Brück and Prof. Robert Kourist for the chance to work on the interesting projects presented in this thesis. Thank you, Dr. Tarek Hilal for the work and support on the Cryo-EM.

Thank you to the whole AG Wahl for the amazing past years, even during corona times. A special thank you to Carl Helmer and Stephanie Himpich for your amazing support on the projects and of course for the fun time in the lab. Thank you, Clemens Langner and Nicole Holton for the great help in the lab, whenever I faced a problem.

Also, I would like to thank Dr. Marion Ringel for the amazing collaboration and fruitful work on the Copu project.

Thanks to our collaborators on projects not related to this thesis: Prof. Dan Major, Prof. Beate Kokschi, Prof Christian Freund, Prof Roderich Süßmuth.

I am very thankful for the access to the MX-beamlines at Bessy II (Berlin) and Petra III at DESY (Hamburg) and the support through the beamline scientists.

I am grateful to the BMBF for funding my research over the last three years!

Thank you to Dr. Giulia Glorani, who has been a great role model and an even better friend to me over the last years.

I would like to thank my parents for being supportive during my studies and my sister for being the most fun distraction whenever I needed her.

Last but not least I would like to thank Benjamin Dimos-Röhl for all the science talk on the way to work, listening to my problems and just being there for me.



## Contents

Acknowledgement .....	7
Contents.....	1
Abstract.....	1
Zusammenfassung .....	3
List of publications .....	5
1 Introduction.....	7
1.1 Terpenes .....	7
1.1.1 Terpenes in white biotechnology.....	9
1.2 Borneol dehydrogenases .....	10
1.3 Terpene synthases .....	11
2 Aims .....	15
3 Results and Discussion.....	17
3.1 Borneol dehydrogenases <i>SrBDH1</i> and <i>SoBDH2</i> .....	17
3.1.1 Crystal Structure of <i>SrBDH1</i> .....	17
3.1.2 Cryo-EM structure of <i>SoBDH2</i> .....	19
3.1.3 Cryo-EM structure of <i>SrBDH1</i> .....	21
3.1.4 Comparison of <i>SrBDH1</i> and <i>SoBDH2</i> with <i>PsBDH</i> .....	22
3.2 Sesquiterpene synthase <i>Copu9</i> .....	25
3.2.1 Overall Structure.....	25
3.2.2 Characterization of <i>Copu5</i> .....	27
3.2.3 Characterization of <i>Copu5</i> and <i>Copu9</i> point-mutants .....	28
4 Conclusion .....	33
References .....	35

## Contents

---

Abbreviations .....	41
Table of Figures .....	42

## Abstract

Climate change, antibiotic resistances and environmental pollution are growing threats. Therefore, finding alternatives for fossil resources and discovery of new pharmaceuticals grows more important every day. Natural compounds and their *in vivo* production pathways proved to be a possible solution to overcome those problems. Optimized microbial hosts can serve as sustainable production platforms for various compounds as it is done for penicillin since many years.

The first research topic of this thesis are borneol dehydrogenases, enzymes which convert borneol to camphor. Enantiomerically pure camphor has numerous applications in cosmetic, pharmaceutical, and chemical industry. Thus, enantioselective borneol dehydrogenases would be an attractive candidate to achieve enantiomerically pure camphor. To better understand the differences of enantioselective and unselective borneol dehydrogenases we solved the structures of two selective borneol dehydrogenases from *Salvia rosmarinus* and *Salvia officinalis* using X-ray crystallography and cryo-electron microscopy. The obtained structures were compared to the previously solved structure of the unselective borneol dehydrogenase of *Pseudomonas sp.* TCU-HL1.

The second focus of this thesis are terpene synthases, a class of enzymes responsible for the cyclization of linear terpene precursors. The products of terpene synthases are interesting candidates for the chemical and pharmaceutical industry due to their diverse characteristics and properties. Latest advances in genome sequencing enabled the discovery of many new and diverse terpene synthases from various organisms. We report on the discovery of two terpene synthases from *Coniophora puteana*, Copu5 and Copu9, that not only have identical product profiles, but also show high yields in an optimized *Escherichia coli* strain. Main product of both enzymes is (+)- $\delta$ -cadinol that has been shown to have cytotoxic effect on MCF7 cells and could be used as a new and sustainable anti-tumor drug. To investigate their properties and gain deeper understanding into their function, we attempted to crystallize and biochemically characterize Copu5 and Copu9.





## Zusammenfassung

Klimawandel, Antibiotikaresistenzen und Umweltverschmutzung sind wachsende Bedrohungen. Daher wird die Suche nach Alternativen für fossile Ressourcen und die Entdeckung neuer Arzneimittel von Tag zu Tag wichtiger. Naturstoffe und ihre *in-vivo*-Produktionswege bieten eine mögliche Lösung dieser Probleme. Optimierte mikrobielle Wirte können als nachhaltige Produktionsplattformen für verschiedene chemische Verbindungen dienen, wie es seit vielen Jahren für Penicillin üblich ist.

Der erste Fokus dieser Arbeit sind Terpensynthasen, eine Klasse von Enzymen, die für die Zyklisierung von linearen Terpenvorläufern verantwortlich sind. Die Produkte der Terpensynthasen sind aufgrund ihrer vielfältigen Eigenschaften interessante Kandidaten für die chemische und pharmazeutische Industrie. Jüngste Fortschritte in der Genomsequenzierung ermöglichten die Entdeckung vieler neuer und vielfältiger Terpensynthasen aus verschiedenen Organismen. Wir berichten über die Entdeckung zweier Terpensynthasen aus *Coniophora puteana*, Copu5 und Copu9, die nicht nur identische Produktprofile aufweisen, sondern auch hohe Ausbeuten in einem optimierten *Escherichia coli*-Stamm zeigen. Hauptprodukt beider Enzyme ist (+)- $\delta$ -Cadinol, das nachweislich eine zytotoxische Wirkung auf MCF7-Zellen hat und als neues und nachhaltiges Antitumormittel eingesetzt werden könnte. Zur Untersuchung ihrer Eigenschaften und um ein tieferes Verständnis ihrer Funktion zu erlangen, haben wir versucht, Copu5 und Copu9 zu kristallisieren und biochemisch zu charakterisieren.

Der zweite Schwerpunkt dieser Arbeit sind Borneoldehydrogenasen, Enzyme, die Borneol zu Kampfer umwandeln. Enantiomerenreiner Kampfer hat zahlreiche Anwendungen in der kosmetischen, pharmazeutischen und chemischen Industrie. Daher wären enantioselektive Borneoldehydrogenasen ein attraktiver Kandidat zur Herstellung von enantiomerenreinem Kampfer. Um die Unterschiede zwischen enantioselektiven und unselektiven Borneoldehydrogenasen besser zu verstehen, haben wir die Strukturen zweier selektiver Borneoldehydrogenasen aus *Salvia rosmarinus* und *Salvia officinalis* mittels Röntgenkristallographie und Kryo-Elektronenmikroskopie gelöst. Die erhaltenen Strukturen wurden mit der zuvor gelösten Struktur der unselektiven Borneol-Dehydrogenase von *Pseudomonas sp.* TCU-HL1 verglichen.



## List of publications<sup>1</sup>

- I. Chánique, Andrea M.\*; Dimos, Nicole\*; Drienovská, Ivana; Calderini, Elia; Pantín, Mónica P.; Helmer, Carl P. O.; Hofer, Michael; Sieber, Volker; Parra, Loreto P.; Loll, Bernhard; Kourist, Robert (2021): A Structural View on the Stereospecificity of Plant Borneol-Type Dehydrogenases. *ChemCatChem* 13 (9), S. 2262–2277. DOI: <https://doi.org/10.1002/cctc.202100110>

\*These authors contributed equally to this work.

ND performed crystallization and diffraction experiments, refined the structure, and contributed to manuscript preparation.

- II. Dimos, Nicole\*; Helmer, Carl P. O.\*; Chánique, Andrea M.; Wahl, Markus C.; Kourist, Robert; Hilal, Tarek; Loll, Bernhard (2022): CryoEM analysis of small plant biocatalysts at sub-2 Å resolution. *Acta Cryst. D* 78 (Pt 1), S. 113–123. DOI: <https://doi.org/10.1107/S205979832101216X>

\*These authors contributed equally to this work.

ND prepared samples for EM measurements, refined the structures and contributed to the manuscript.

- III. Ringel, Marion\*; Dimos, Nicole\*; Himpich, Stephanie; Haack, Martina; Huber, Claudia; Eisenreich, Wolfgang; Schenk, Gerhard; Loll, Bernhard; Brück, Thomas (2022): Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of  $\delta$ -cadinol. *Microb Cell Fact* 21 (1), S. 64. DOI: <https://doi.org/10.1186/s12934-022-01791-8>

\*These authors contributed equally to this work.

ND performed protein purification, crystallization and kinetic experiments, diffraction experiments, refined the structure, and contributed to manuscript preparation.

---

<sup>1</sup> The publications on this page were published under CC-BY-licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Additional publications not included in this thesis:**

Leppkes, Jakob; Dimos, Nicole; Loll, Bernhard; Hohmann, Thomas; Dyrks, Michael; Wieseke, Ariane; Keller, Bettina G.; Kokschi, Beate (2022): Fluorine-induced polarity increases inhibitory activity of BPTI towards chymotrypsin. *RSC Chem. Biol.* 3 (6), S. 773–782. DOI: <https://doi.org/10.1039/d2cb00018k>.

Wehrhan, Leon; Leppkes, Jakob; Dimos, Nicole; Loll, Bernhard; Kokschi, Beate; Keller, Bettina G. (2022): Water Network in the Binding Pocket of Fluorinated BPTI-Trypsin Complexes—Insights from Simulation and Experiment. *J. Phys. Chem. B.* DOI: <https://doi.org/10.1021/acs.jpcc.2c05496>.

Prem, Sophia; Helmer, Carl P. O.; Dimos, Nicole; Himpich, Stephanie; Brück, Thomas; Garbe, Daniel; Loll, Bernhard (2022): Towards an understanding of oleate hydratases and their application in industrial processes. *Microb Cell Fact* 21 (1), S. 58. DOI: <https://doi.org/10.1186/s12934-022-01777-6>.

Raz, Keren; Driller, Ronja; Dimos, Nicole; Ringel, Marion; Brück, Thomas; Loll, Bernhard; Major, Dan Thomas (2020): The Impression of a Nonexisting Catalytic Effect: The Role of CotB2 in Guiding the Complex Biosynthesis of Cyclooctat-9-en-7-ol. *J. Am. Chem. Soc.* 142 (51), S. 21562–21574. DOI: <https://doi.org/10.1021/jacs.0c11348>.

Franz, Andreas; Weber, A. Ioana; Preußner, Marco; Dimos, Nicole; Stumpf, Alexander; Ji, Yanlong; Moreno-Velasquez, Laura; Voigt, Anne; Schulz, Frederic; Neumann, Alexander; Kuropka, Benno; Kühn, Ralf; Urlaub, Henning; Schmitz, Dietmar; Wahl, Markus C.; Heyd, Florian (2023): Branch point strength controls species-specific CAMK2B alternative splicing and regulates LTP. *Life science alliance* 6 (3). DOI: <https://doi.org/10.26508/lsa.202201826>.

## 1 Introduction

In a world with ever growing threats as climate change, depletion of fossil energy sources, water pollution, and antibiotic resistances to name just a few, the need of sustainable solutions is rising more and more. Finding new therapeutics for diseases and sustainable green fuel resources in a CO<sub>2</sub> neutral or even negative manner has been a challenge for quite some time. Moving away from fossil resources is a step that has to be taken. Taking a closer look to nature's solutions is one way to go.

A huge number of natural compounds have been used over centuries as drugs, fragrances, pigments, insect repellents and much more. These natural compounds are often derived from plants, fungi or bacteria. They can be of complex chemical structure and have one or more chiral centers. This can lead to stereoisomers with different properties and effects, making chemical synthesis of pure compounds with good yields rather complicated (Brück et al., 2014). On the other side, direct extraction from their natural sources usually requires large amounts of starting material and high energy consumption for the isolation, leading to insufficient yields. For example the anti-cancer drug Taxol which is extracted from the bark of the Pacific yew with a yield of in average 0.66 g/tree, which leads to the need of 3-10 trees to treat one single patient (Kathiravan et al., 2012).

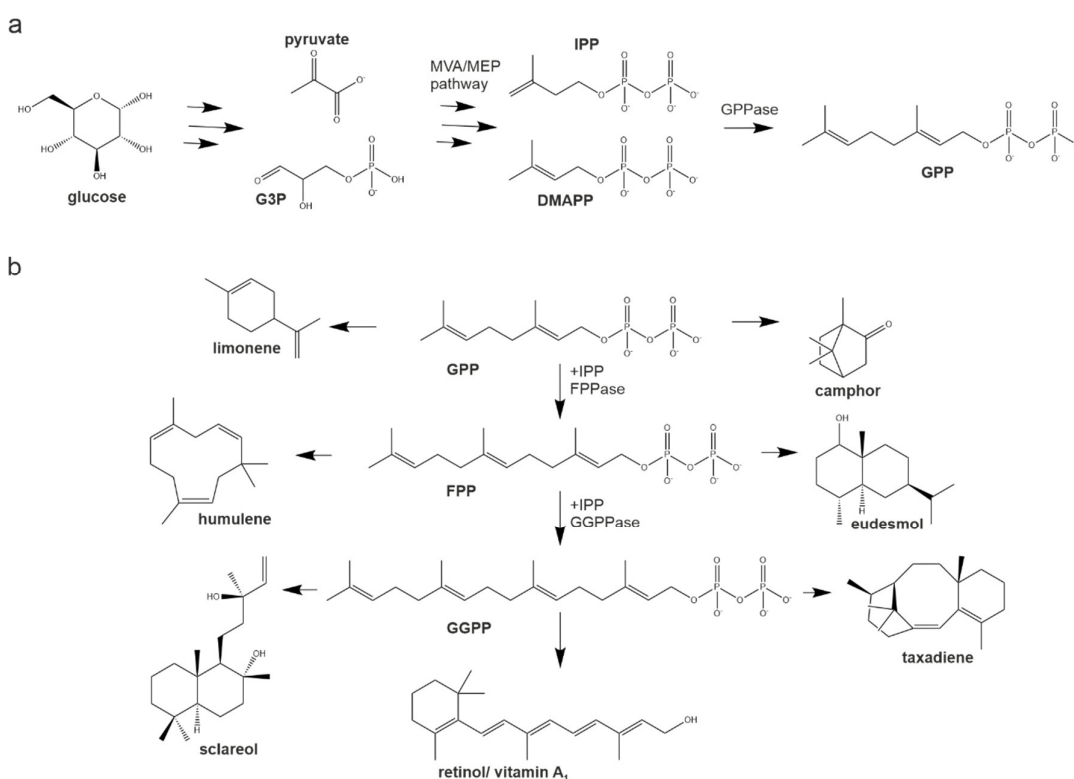
Therefore, the development of biological platforms to produce these compounds, can be a resource-saving option, as it has been done in the past for penicillin. Specifically, non-genetically or genetically modified organisms, as well as purified enzymes, can be used to achieve this goal. The use of nature's toolbox to address drug development and their production can be of help moving away from fossil resources and reducing waste production.

### 1.1 Terpenes

Terpenes are a heterogeneous group of natural products consisting of over 80,000 compounds (Christianson, 2017). The properties of terpenes range from antibiotic, cytotoxic and anti-tumor activity over insecticides to perfumes and flavours, but they can

also serve as building blocks for synthetic polymers (Silvestre and Gandini, 2008). They derive from isoprene ( $C_5H_8$ ) and, depending on the number of their isoprene units, are categorized as monoterpenes ( $C_{10}$ ), sesquiterpenes ( $C_{15}$ ), diterpenes ( $C_{20}$ ), triterpenes ( $C_{30}$ ) and polyterpenes. They are synthesized from linear precursors like geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) by enzymes called terpene synthases or terpene cyclases (Figure 1) (Christianson, 2017).

Bioactive terpenes can be found in essential oils of many plants, which have been used for many centuries to treat different conditions, i.e. eucalyptus oil for scarlet fever and other infectious diseases (Curgenvin, 1891; Masyita et al., 2022). This makes essential oils and the recently elucidated genomes of their origin organisms an interesting starting point for terpene identification and the further investigation of the enzymes involved in their *in vivo* production (Drienovská et al., 2020).



**Figure 1:** Terpene synthesis *in vivo*. **a** Upstream part of terpene synthesis from initial glucose uptake to the linear terpene precursor GPP. Multiple arrows indicate several steps in between. **b** Downstream part of terpene synthesis. Exemplary terpene products are indicated with arrows from their linear precursor. Upstream part of the process is highlighted with a green box, downstream part is highlighted with a blue box. Abbreviations: glyceraldehyde 3-phosphate (G3P); mevalonate (MVA); methylerythritol phosphate (MEP); isopentyl diphosphate (IPP); dimethylallyl diphosphate (DMAPP).

Enzymes involved in downstream terpene metabolism can be divided into two groups, the ones responsible for the carbo-skeleton formation and those taking part in further decoration and functionalization (Brück et al., 2014; Wriessnegger et al., 2014; Dickschat, 2016). In this thesis we focus on representatives of both groups: sesquiterpene synthases that cyclizes FPP and borneol dehydrogenases that convert the bicyclic alcohol borneol to the ketone camphor.

### **1.1.1 Terpenes in white biotechnology**

White biotechnology is the branch of biotechnology that focuses on industrial application of biotechnological techniques. The main goals are the sustainable production of different substances by microorganisms or enzymatic reactions. This is especially interesting for substances with complex chemical structures, such as terpenes. The chemical production of these complex multicyclic compounds often involves many highly complex synthesis steps and yields only minimal amounts of products.

Optimizing microorganisms for the industrial production of these compounds appears to be the logical step considering nature produced these compounds first. Therefore, it makes sense to use nature's toolbox as a starting point. This approach gives the opportunity to minimize waste and save fossil resources. Enzyme engineering is an important part in optimization of these processes as it gives the opportunity to adjust the formed product but also the yield (Brück et al., 2014; Zhang and Hong, 2020).

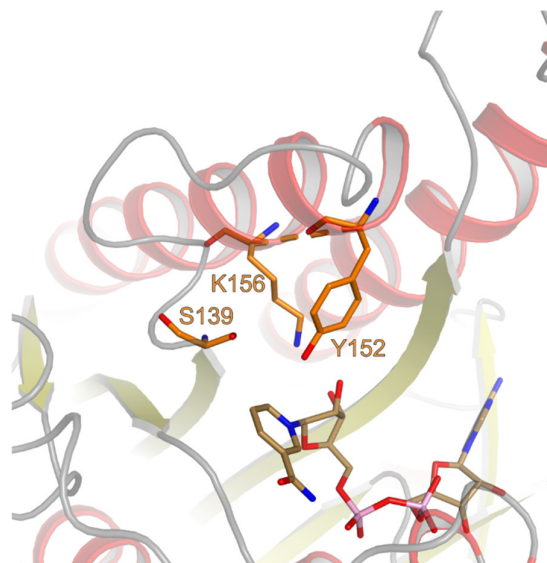
The process of terpene production can be subdivided into an upstream and a downstream part. While the upstream part focuses on the enzyme pathway responsible for the formation of the building blocks for the linear terpene precursors from carbohydrates, the downstream part is centered on the synthesis of the linear precursor generating enzyme and the further synthesis and decoration needed to reach the final product (Figure 1). In this downstream part terpene synthases (TPS) involved in the cyclization of the linear precursors are especially interesting targets as they perform the most challenging chemical reaction. These TPS also offer the great opportunity to shift their product spectra or change their product via site directed mutagenesis (Lodeiro et al., 2004; Janke et al., 2014; Driller et al., 2018).

In addition to exploiting engineered microbial cell factories, purified enzymes can be used in industrial context (i.e.: food industry (amylases), pharmaceutical industry (Penicillin amidase), detergents (proteases)) either in soluble or immobilized form (DiCosimo et al., 2013; Rigoldi et al., 2018). The usage of enzymes could also help producing enantiomerically pure compounds, which have high demand in pharmaceutical and cosmetic industry instead of racemic mixtures that are often hardly or not at all separable.

### 1.2 Borneol dehydrogenases

Borneol dehydrogenases (BDH) belong to the family of short chain dehydrogenases/reductases (SDR). They convert the monoterpene borneol to camphor. SDRs show a classical Rossmann-like fold that harbors the co-factor NAD<sup>+</sup> needed for the oxidation of the substrate (Rossmann et al., 1975). The Rossmann-like fold consists of alternating  $\alpha$ -helices and  $\beta$ -strands that form a central parallel  $\beta$ -sheet consisting of 6 to 7  $\beta$ -strands with three to four  $\alpha$ -helices on each side (Rossmann et al., 1975; Jörnvall et al., 1981; Lesk, 1995). Highly variable C-terminal regions complete the enzyme by forming the substrate binding site (Kavanagh et al., 2008). Most SDRs share a highly conserved catalytic motif YxxxK and a serine in their active sites, where the tyrosine acid/base properties are enhanced by the adjacent lysine (Figure 2) (Fujimoto et al., 2001). The tyrosine abstracts a hydrogen from the substrate, while the serine's role is to stabilize and polarize the carbonyl group of the substrate (Kavanagh et al., 2008).





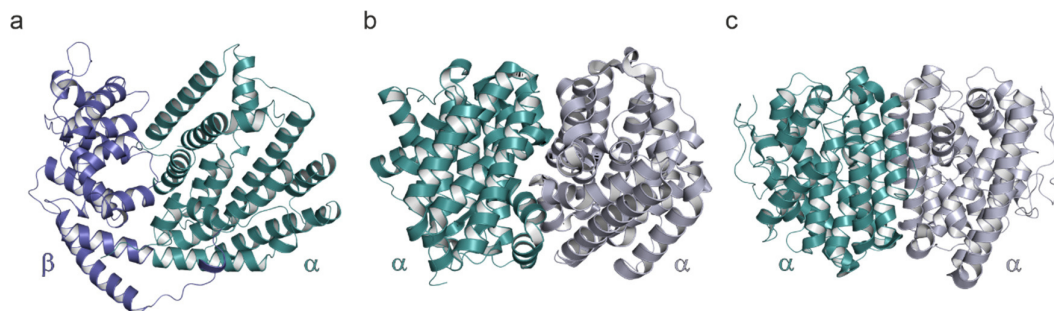
**Figure 2:** Close-up to the catalytic pocket of 3 $\alpha$ /20 $\beta$  hydroxysteroid dehydrogenase (PDB-ID: 2hsd (Ghosh et al., 1994)).  $\beta$ -strands of 3 $\alpha$ /20 $\beta$  HSD are depicted as yellow arrows and  $\alpha$ -helices as red spirals. Catalytic residues are shown as orange sticks and the bound NAD as sand sticks.

The observation of enantiomeric pure camphor and borneol in some plants led to the initial idea of enantiospecific BDHs, that was confirmed in 1978 by Croteau and coworkers with the observation of enantiospecific activity of sage extracts (*S. officinalis*). Many years later, thanks to the advances in genome sequencing and genome mining techniques, the responsible BDH could be identified, together with other enantiospecific BDHs from related organisms (Drienovská et al., 2020).

### 1.3 Terpene synthases

TPS are a group of enzymes responsible for the huge variety of terpenes. TPS cyclize their linear precursors to mono- or polycyclic molecules. They are divided into two classes: class I and class II depending on their reaction mechanism. In class I TPS the initial carbocation is formed through abstraction of the diphosphate group that is bound by 3  $Mg^{2+}$  ions via the DDxD and NSE motif. In contrast, in class II TPS a tertiary carbocation is formed through protonation of an alkene group by the DxDD motif (Christianson, 2008; Rudolf and Chang, 2020). In addition class I TPS have an additional c-terminal WxxxxRY motif that facilitates the folding of the c-terminus above the active site (Baer et al., 2014; Rabe et al., 2016; Driller et al., 2018). These differences are also reflected in their secondary structure and catalytic motifs. Whereas most class I TPS consist of one  $\alpha$ -

helical bundle domain ( $\alpha$ -fold), Class II TPS consist of two  $\alpha$ -helical domains, namely  $\beta$  and  $\gamma$  (Wendt and Schulz, 1998). Although there are examples of monomeric plant class I TPS with additional  $\beta$  and/ or  $\gamma$  domains, for one domain class I TPS parallel or antiparallel dimers are observed (Figure 3).



**Figure 3:** Structural comparison of three class I TPS. **a** Epi-aristochelene synthase from *Nicotiana tabacum* (PDB-ID: 5EAS (Starks et al., 1997)), showing a two domain architecture (  $\alpha$  subunit shown in teal  $\beta$  subunit shown in dark blue). Dimeric terpene synthases **b** CotB2 from *Streptomyces melanosporofaciens* (PDB-ID: 4OMG (Janke et al., 2014)) and **c** SdS of *Streptomyces pristinaespiralis* (PDB-ID: 4OKM (Baer et al., 2014)) exhibiting a one  $\alpha$  domain architecture ( one protomer is shown in teal the second one in gray). CotB2 shows an antiparallel dimeric state while SdS forms a parallel dimer.

One well studied class I TPS is CotB2, a diterpene synthase from *S. melanosporofaciens* which is the first bacterial terpene synthase with a solved crystal structure (Janke et al., 2014). Using a structure-based approach, multiple point mutations of CotB2 in the active site have been tested for their effect on the product as observed already for different TPS (Lodeiro et al., 2004; Hyatt and Croteau, 2005; Wilderman and Peters, 2007). CotB2 single point-mutations lead to many different products (Kim et al., 2009; Görner et al., 2013; Janke et al., 2014; Tomita et al., 2017; Driller et al., 2018). Further structures of CotB2 point mutants and computational approaches supported the opinion that one role of active site decoration is the guidance of the substrate fold and offering a suitable chemical environment for the reaction rather than actively participating in the reaction (Raz et al., 2020b). Since the determination of the CotB2 structure in the closed, active and open, inactive state, further studies have shed light into the reaction mechanism of class I TPS (Janke et al., 2014; Tomita et al., 2017; Driller et al., 2018; Tang et al., 2020). They proposed a structure evidenced timeline that lead from an open apo state in steps to the active closed conformation and a state after the reaction, but before the release of the products and  $Mg^{2+}$  ions (Driller et al., 2018). Briefly, the structures suggest a sequential binding of the first  $Mg^{2+}$  ion, followed by the binding of the substrate, leading

to first conformational changes in the C-terminus (salt-bridge formation between R294 (part of the C-terminal WxxxxRY motif) and D111 (part of the DDXD motif)). After binding of the two remaining  $Mg^{2+}$  ions, the C-terminus folds completely over the active site and CotB2 cleaves the diphosphate moiety of the substrate, starting the cyclization reaction. At the end the substrate and  $Mg^{2+}$  ions are released, leading to the initial open state of CotB2. A CotB2 variant lacking the C-terminus did not show any activity, suggesting its importance in TPS function.

This suggests that TPS offer an extraordinary option for designed terpene production, which can be used in biotechnological industry. In the focus of this thesis are two sesquiterpene synthases of *C. puteana* that share identical product spectra.



## 2 Aims

The aim of this thesis was to gain deeper understanding on biotechnological relevant enzymes by solving their structures in native and substrate/product-bound states. This was attempted by X-ray crystallography and cryo-electron microscopy. Additionally, different biochemical methods were used to gain further insights into their biochemical and catalytic properties. For some enzymes, structure-based mutagenesis was conducted to gain information about the susceptibility to changes in their product spectra. In this thesis the focus lays on the following enzymes:

- The enantiospecific borneol dehydrogenases of *S. rosmarinus* (SrBDH1) and *S. officinalis* (SoBDH2)
- Two sesquiterpene synthases of *C. puteana*, Copu5 and Copu9



## 3 Results and Discussion

### 3.1 Borneol dehydrogenases *SrBDH1* and *SoBDH2*

This section refers to the following publications:

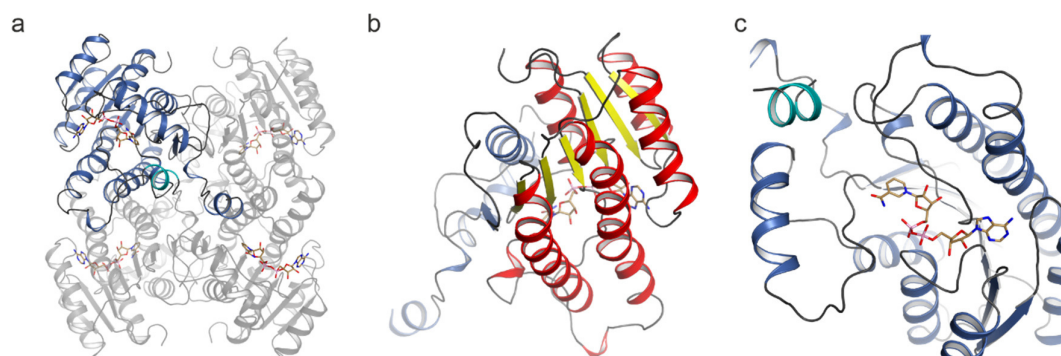
- Chánique, Andrea M.\*; Dimos, Nicole\*; Drienovská, Ivana; Calderini, Elia; Pantín, Mónica P.; Helmer, Carl P. O.; Hofer, Michael; Sieber, Volker; Parra, Loreto P.; Loll, Bernhard; Kourist, Robert (2021): A Structural View on the Stereospecificity of Plant Borneol-Type Dehydrogenases. *ChemCatChem* 13 (9), S. 2262–2277. DOI: <https://doi.org/10.1002/cctc.202100110>
- Dimos, Nicole\*; Helmer, Carl P. O.\*; Chánique, Andrea M.; Wahl, Markus C.; Kourist, Robert; Hilal, Tarek; Loll, Bernhard (2022): CryoEM analysis of small plant biocatalysts at sub-2 Å resolution. *Acta cryst. D* 78 (Pt 1), S. 113–123. DOI: <https://doi.org/10.1107/S205979832101216X>

Previous findings of enantiomeric pure camphor and borneol in *S. rosmarinus* and *S. officinalis* were the first hint towards enantioselective borneol dehydrogenases in these species (Croteau et al., 1978; Satyal et al., 2017). Our cooperation partners from the Technical University of Graz identified several BDHs and tested them for enantioselectivity and specificity. Two of the most promising candidates were *SrBDH1* (BDH from *S. rosmarinus*) and *SoBDH2* (BDH from *S. officinalis*). The lack of a structure for an enantiospecific BDH further motivated us to pursue their structural elucidation.

#### 3.1.1 Crystal Structure of *SrBDH1*

We were able to obtain crystal structures of *SrBDH1* in two different space groups from two different crystallization conditions, one with high concentrations of NaCl at 2.6 Å resolution (PDB-ID: 6ZZT (Chánique et al., 2021)) the other one with 5/4 pentaerythritol propoxylate (5/4 PO/OH) at 2.27 Å resolution (PDB-ID: 6ZYZ (Chánique et al., 2021)). The overall crystal structure of *SrBDH1* is tetrameric, which was also confirmed by SEC-MALS (Figure 4 and Figure 6). Each protomer shows a Rossmann-like fold characteristic for the

family of short-chain dehydrogenases/ reductases (SDR) to which the BDH belongs (Rossman et al., 1975; Jörnvall et al., 1995). This fold harbors the co-factor  $\text{NAD}^+$  which is needed for the oxidation of borneol to camphor as it serves as an oxidizing agent. While only one protomer harbors  $\text{NAD}^+$  in the high salt structure, we observe it in all four in the PO/OH condition. It should be noted that no  $\text{NAD}^+$  was added during purification and crystallization of *SrBDH1*, hence it must have been co-purified with the protein itself. In monomers with  $\text{NAD}^+$  bound we observe the folded loop region from residue V197 to S203, which cannot be observed in monomers lacking  $\text{NAD}^+$ . The presence of  $\text{NAD}^+$  seems to lead to a stabilization of this loop, which then, blocked through a crystal contact, does not allow  $\text{NAD}^+$  diffusion in the high salt structure. In the PO/OH structure we observe density for  $\text{NAD}^+$  in all protomers, although at different sigma levels, which is indicated through different occupancy levels between the protomers of one tetramer.

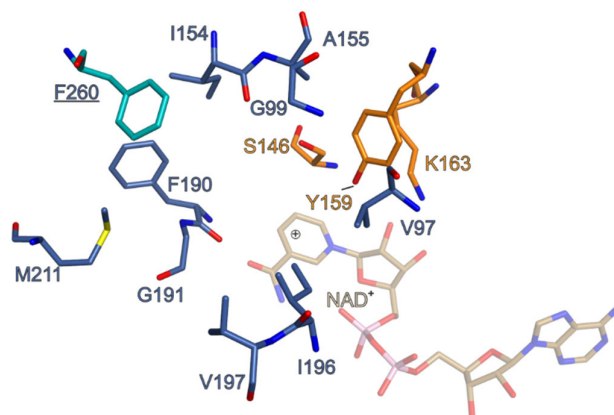


**Figure 4:** Crystal structure of *SrBDH1*. **a** homo tetrameric structure of *SrBDH1* obtained in PO/OH condition (PDB-ID: 6ZYZ). One protomer is highlighted in dark blue.  $\text{NAD}^+$  is depicted as light orange sticks. The diagonal C-terminal  $\alpha$ -helix closing the active site is shown in teal. **b** protomer of *SrBDH1* rotated by  $180^\circ$ . The classical Rossmann-like fold is highlighted in red ( $\alpha$ -helices) and yellow ( $\beta$ -strands). The remaining structure is color coded as in panel a. **c** color coded as panel a. zoom in to the active site showing the C-terminal  $\alpha$ -helix contributing to the active site formation.

Mainly hydrophobic residues line the active site of *SrBDH1* (V97, G99, I154, A155, F190, G191, I196, V197, and M211) and F260 from the C-terminal  $\alpha$ -helix of the diagonal monomer, which completes the active site together with the conserved catalytic motif (S146, Y159, and K163) and the  $\text{NAD}^+$  (Figure 5). It has been shown for other SDRs, that the tyrosine abstracts a proton from the substrate, while the lysine enhances the acid/base properties of the tyrosine. The serine meanwhile stabilizes and polarizes the newly formed carbonyl group of the substrate for the further reaction (Kavanagh et al.,



2008). Loss of activity was observed for S146A and Y159A mutant supporting their importance in the catalysis process.



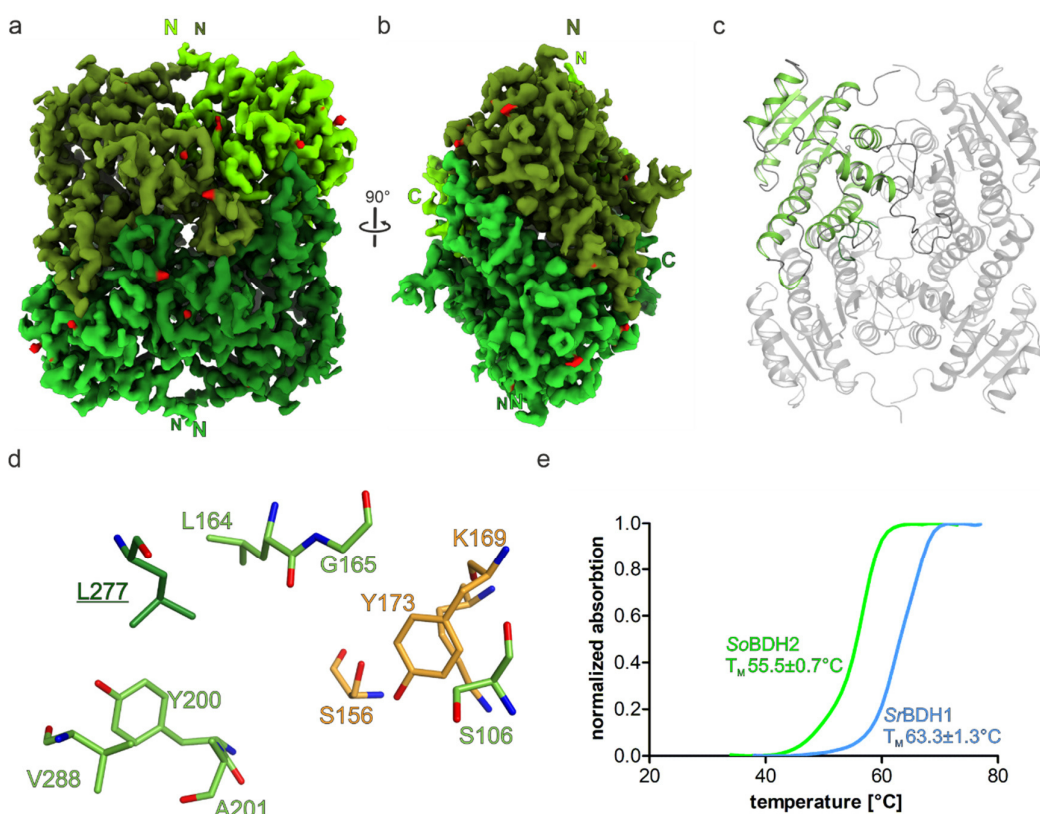
**Figure 5:** *SrBDH1* active site residues. Residues forming the active site are shown as dark blue and teal (originating from another protomer, residue name is underlined) sticks. The catalytic motif is shown as orange sticks. NAD<sup>+</sup> is depicted as sand sticks.

### 3.1.2 Cryo-EM structure of *SoBDH2*

Although *SoBDH2* has 44% sequence identity and 60% sequence similarity to *SrBDH1*, attempt to crystallize it failed. To have a better understanding of the reasons behind this behavior we conducted thermal shift assay (TSA) measurements, comparing the two BDHs. Surprisingly, the melting temperature of *SoBDH2* was around 8 °C lower than of *SrBDH1*, speaking in favor of lower thermostability (Figure 6). Additionally, the multimeric state of *SoBDH2* was analyzed via SEC-MALS. Besides a 123 kDa tetrameric state similarly observed for *SrBDH1*, we identified a smaller fraction that seemingly forms octamers with a 245 kDa apparent molecular weight for the *SoBDH2*. The hint towards higher oligomeric states led us to attempt to solve the aforementioned structure using Cryo-EM. After initial promising negative stain EM trials, we proceeded to Cryo-EM data collection and data processing.

We were able to achieve a resolution of 2.04 Å for the tetrameric *SoBDH2*(EMD-12739; PDB-ID: 7O6P; (Dimos et al., 2022)) exhibiting a *D2* symmetry, similarly to *SrBDH1* and *PsBDH*. Interestingly, we could not observe any octamers. The protomers adopt a Rossmann-like fold which is required for NAD<sup>+</sup> binding, as observed for *SrBDH1*. For the first 12 N-terminal residues, alongside with the preceding 6xHis tag no density was observed, probably due to high flexibility (Figure 6). The density for  $\alpha$ -helix C residues Q52-G65 is weak and fragmented, reflected by elevated B factors (Figure 7). We could

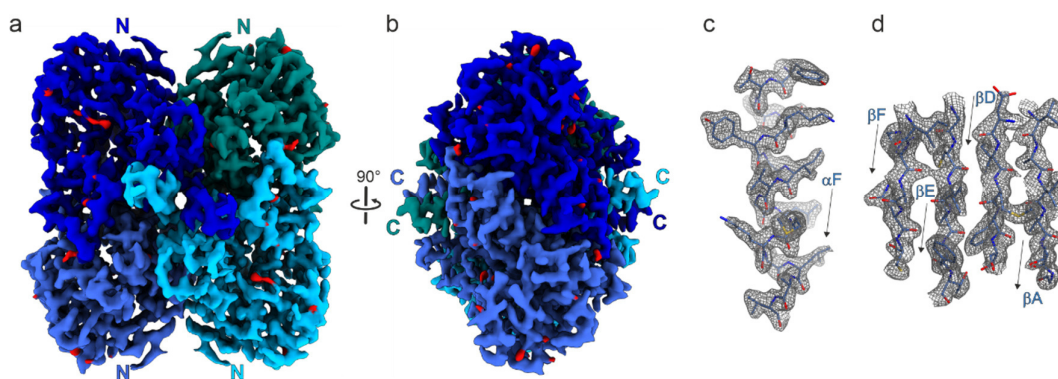
not observe any density for the co-factor NAD<sup>+</sup> in the binding pocket as well as for the residues S205-E218. This observation was not surprising as we only observed a folded loop in *SrBDH1* for protomers with NAD<sup>+</sup> binding. Although there was no need to add NAD<sup>+</sup> to *SrBDH1* to observe it in the crystal structure and an apo structure could only be obtained after co-crystallization trials with the substrate (+)-borneol (PDB-ID: 6ZZ0 (Chánique et al., 2021)), which led to the reduction of NAD<sup>+</sup> and the release of product and cofactor. We could not prevent the loss of the cofactor even when adding 3-fold molar excess to the *SoBDH2* before size exclusion chromatography. The NAD<sup>+</sup> binding site might be artificially stabilized under crystallization conditions and prevent the cofactor loss in contrast to Cryo-EM sample preparation.



**Figure 6:** Cryo-EM structure of *SoBDH2*. **a** Cryo-EM volume of *SoBDH2* (EMD-12739; PDB-ID:7O6P). Each protomer is shown in a different shade of green. Red colored volume indicates positions of water molecules. **b** Side view of **a**. **c** Tetrameric structure of *SoBDH2* as derived from the Cryo-EM volume. One protomer is highlighted in green. The C-terminus of another protomer closing the active site is shown in dark green. **d** Zoom-in to the active site of *SoBDH2*. The active site residues are shown as green and dark green (residue originating from another protomer, residue name underlined) sticks. The catalytic residues are depicted as orange sticks. **e** Exemplary melting curves of *SoBDH2* (green) and *SrBDH1* (blue). *SoBDH2* shows a significantly lower melting temperature compared to *SrBDH1*.

### 3.1.3 Cryo-EM structure of *SrBDH1*

To further evaluate the applicability of Cryo-EM to achieve high resolution structures for small plant enzymes, we opted to structurally analyze *SrBDH1* with Cryo-EM in addition. As mentioned above, the SEC-MALS analysis of *SrBDH1* yielded only one peak corresponding to a tetramer, which we also observed in the crystal structures. We were able to obtain a structure with a resolution of 1.88 Å (EMD-12740; PDB-ID: 7O6Q (Dimos et al., 2022)) (Figure 7). Similar to the structure of *SoBDH2*, no density was observed for the 8 N-terminal residues and the very C-terminal residue as well as for the loop region L193- L205 due to the absence of NAD<sup>+</sup>. Except these parts, almost the entire protein chain could be traced in the Cryo-EM map. This is reflected by an exceptional atom inclusion level at the moderate contour level of 0.3 for 97% of all backbone atoms and 93% of all non-H atoms. The completeness of the Cryo-EM structure is virtually identical to the crystal structures.

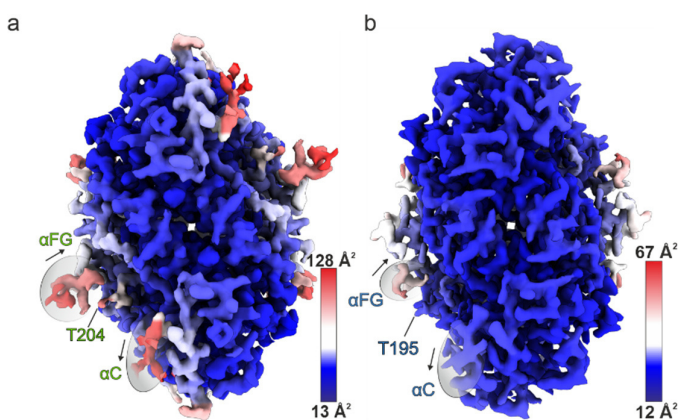


**Figure 7:** Cryo-EM structure of *SrBDH1*. **a** Cryo-EM volume of *SrBDH1* (EMD-12740; PDB-ID: 7O6Q). Each protomer is shown in a different shade of blue. Red colored volume indicates positions of water molecules. **b** Side view of **a**. **c** and **d** show exemplary parts of the model (**c**  $\alpha$ -helix  $\alpha$ F and **d** part of the  $\beta$ -sheet) in the Cryo-EM volume highlighting the high quality of the map.

The apo crystal structure compared to the Cryo-EM model has a root-mean-square deviation (r.m.s.d.) of 0.6 Å (for 982 pairs of C $\alpha$ -atoms), thus are indistinguishable. We identified 399 water molecules uniformly distributed around the protein and in cavities within the protein core, giving a ratio of 0.4 water molecules per residue. This is much lower compared to a crystal structure of comparable resolution, where the expected ratio is one water per residue (Carugo and Bordo, 1999). This might be the effect of missing solvent channels in Cryo-EM and the missing proximity of protein molecules. In

contrast, the observed 34 double conformations of side chains corresponding to 3.5% of all build residues perfectly fit the expected value of 3% for protein crystal structures between 1.0 and 2.0 Å (Miao and Cao, 2016).

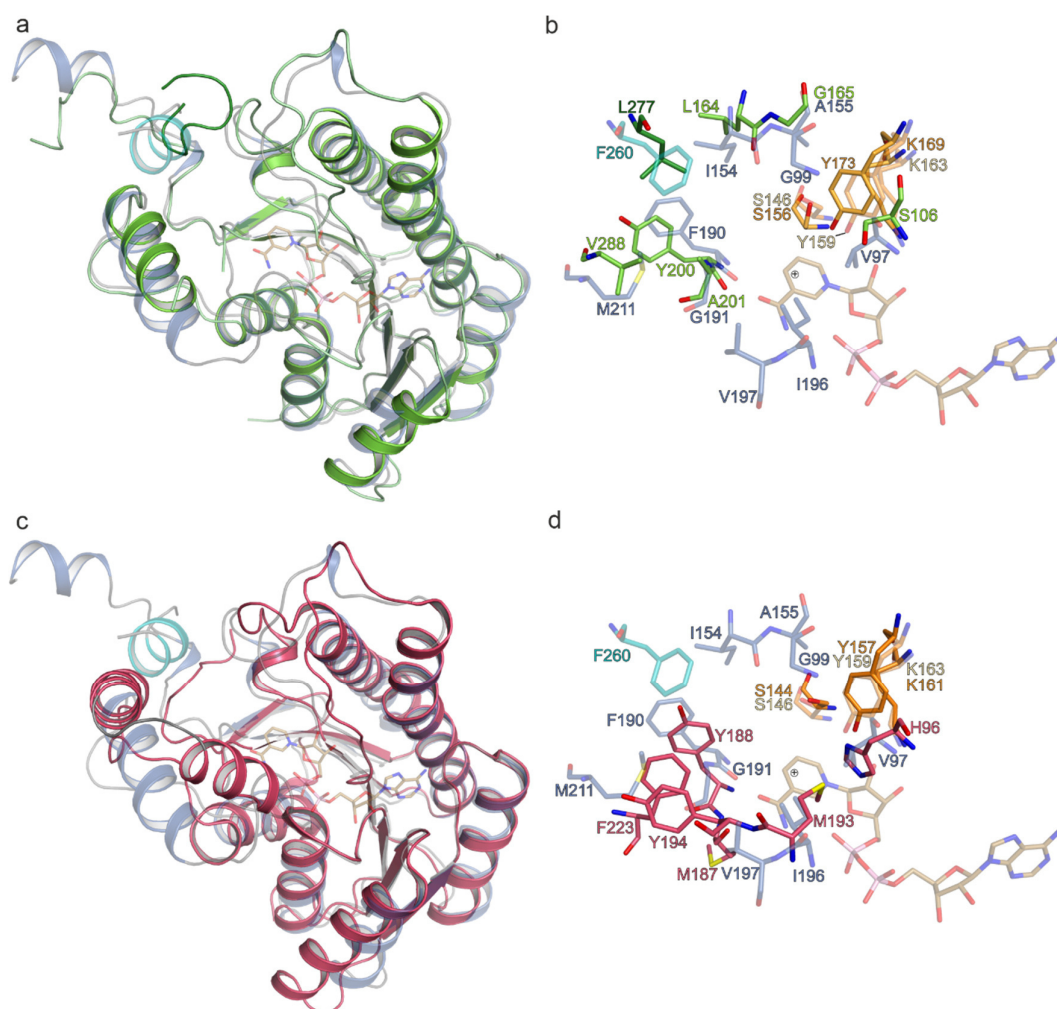
Interestingly, the  $\alpha$ -helix C in *SrBDH1* is much less flexible than in *SoBDH2* (Figure 8). A closer look into the crystal structure shows that it is part of crystal contacts. This might be an explanation on why our attempts to crystalize *SoBDH2* failed.



**Figure 8:** Grouped B factor mapped onto the densities of **a** *SoBDH2* and **b** *SrBDH1*. The color gradient is from blue to red corresponding to increasing B factors. Regions with high B factors are highlighted with gray ellipses and are labeled according to the assigned secondary structure.

### 3.1.4 Comparison of *SrBDH1* and *SoBDH2* with *PsBDH*

The structure of the first unspecific borneol dehydrogenase *PsBDH* originating from *Pseudomonas sp.* TCU-HL1 (PDB ID 6M5N) was recently published by Khine and co-workers making it an interesting target for comparison with the specific BDHs *SrBDH1* and *SoBDH2* (Khine et al., 2020). *PsBDH* shares a sequence identity of 30% with *SrBDH1* and superimposes to *SrBDH1*·NAD<sup>+</sup>/high salt with an r.m.s.d of 1.52 Å (Figure 9). Interestingly, the largest differences are found upstream of  $\beta$ F affecting the potential borneol binding site. In *SrBDH1* the  $\beta$ F strand is connected to  $\alpha$ -helix  $\alpha$ G via a single  $\alpha$ -helix  $\alpha$ FG, the same is observed for *SoBDH2*. In contrast, *PsBDH* has two  $\alpha$ -helices  $\alpha$ FG1 and  $\alpha$ FG2 connecting  $\beta$ F and  $\alpha$ G. While *SrBDH1* has a C-terminal  $\alpha$ -helix harboring F260 which completes the active site of the diagonal protomer, in *SoBDH2* the corresponding residue L277 is part of the C-terminal coiled coil and is absent in *PsBDH*.



**Figure 9:** Comparison of the *So*BDH2 and *Ps*BDH to *Sr*BDH1. **a** Overlay of *So*BDH2 (green) and *Sr*BDH1 (transparent blue). **b** Active site residues of *So*BDH2 and *Sr*BDH1 shown as sticks color-coded as in a. **c** Overlay of *Ps*BDH (Red) and *Sr*BDH1 (transparent blue). **d** Active site residues of *Ps*BDH and *Sr*BDH1 shown as sticks color-coded as in c.

While *Sr*BDH1 and *So*BDH2 have a very similar active site architecture, the *Ps*BDH active site differs. In *Sr*BDH1 and *So*BDH2 the active site is flanked by the single  $\alpha$ -helix  $\alpha$ FG, whereas the corresponding part of *Ps*BDH consists of two  $\alpha$ -helices  $\alpha$ FG1 and  $\alpha$ FG2. This, in addition to the missing C-terminal counterpart from another protomer, leads to a broader active site in *Ps*BDH. This is in agreement with the observation of specific borneol turnover for the plant BDHs with the more restrictive active site and the unspecific turnover for the more open active site. Moreover, the observation of the C-terminus completing the active site of the plant BDHs fits the results from earlier studies that the C-terminus is involved in substrate binding which means variations in C-terminal structure would have consequences on diversity of substrate specificities (Duax et al.,

2000). The differences could be related to the original function of the enzymes. While the bacterial BDH is part of the catabolic pathway, selectivity would restrict the enzyme to a less broad substrate spectrum, thus potential substrates could not be degraded. In contrast in plants BDHs are part of the anabolic pathway, its enzymes being generally more restrictive than their catabolic counterparts.

## 3.2 Sesquiterpene synthase Copu9

This section refers to the following publication:

- Ringel, Marion\*; Dimos, Nicole\*; Himpich, Stephanie; Haack, Martina; Huber, Claudia; Eisenreich, Wolfgang; Schenk, Gerhard; Loll, Bernhard; Brück, Thomas (2022): Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of  $\delta$ -cadinol. *Microbial cell factories* 21 (1), S. 64. DOI: <https://doi.org/10.1186/s12934-022-01791-8>

For the first time two sesquiterpene synthases with the same product spectrum originating from the same organism were discovered. These TPS from *C. puteana*, Copu5 and Copu9, were initially characterized by our collaborators from the technical University of Munich. In addition to identical product spectra, both enzymes show high yields, Copu9 395 mg l<sup>-1</sup>, even higher than Copu5 (225 mg l<sup>-1</sup>).

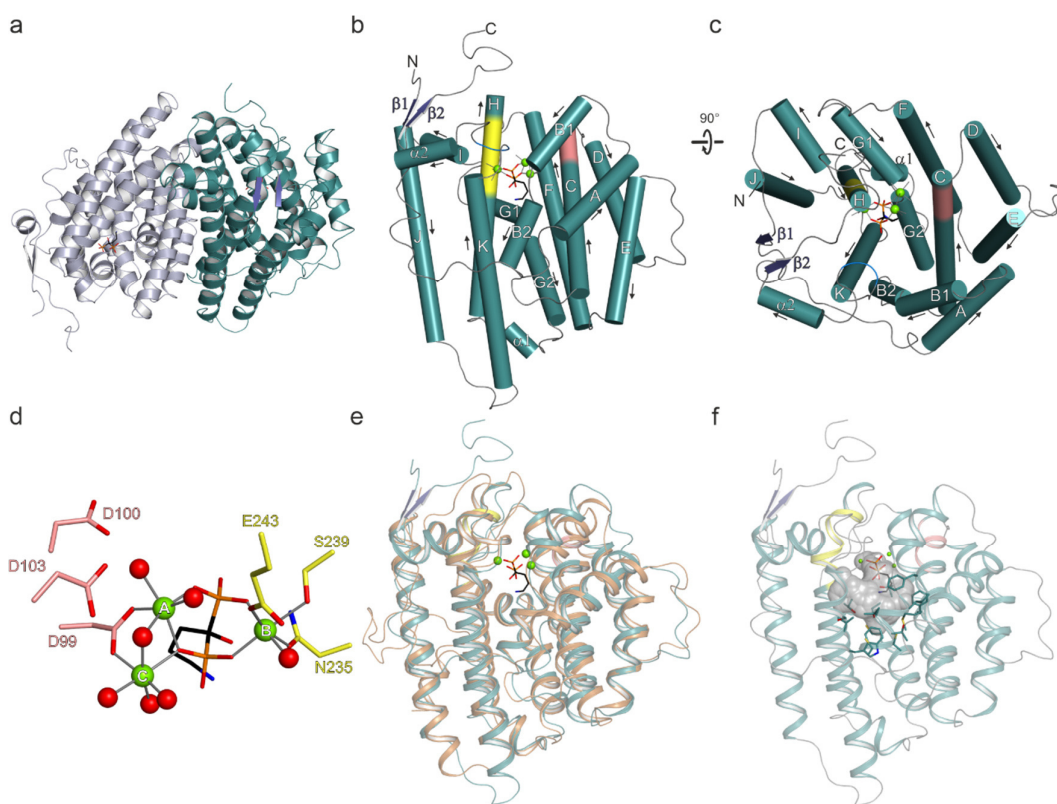
### 3.2.1 Overall Structure

To gain detailed information on the architecture of Copu9 we opted to crystallize it in its apo state and in complex with the non-hydrolysable FPP mimic (4-amino-1-hydroxybutylidene) bisphosphonic acid (alendronate, AHD) to compare the inactive, open with the active, closed conformation. We were not able to crystallize Copu9 in the open form but were successful with the closed AHD bound conformation (PDB-ID: 7OFL (Ringel et al., 2022)).

The 1.83 Å resolution crystal structure revealed that Copu9 forms a dimer, characteristic for class I TPS, *in crystallo* as observed in solution by size exclusion chromatography. Copu9 is a  $\alpha$ -helical bundle protein consisting of 15  $\alpha$ -helices per monomer, but in contrast to other TPS it has two additional  $\beta$ -strands, located at the N- and C-terminus (Figure 10). These two  $\beta$ -strands (T17-L21 and R335-L339) form a short antiparallel  $\beta$ -sheet which might facilitate the closure of the enzyme, as it fixates the C-terminal lid region above the active site. The two protomers of Copu9 show a practically identical fold with a r.m.s.d. of 0.32 Å for 330 C $\alpha$  pairs. The closest structural homologue to Copu9 is the selinadiene synthase SdS (PDB-ID:4OKM (Baer et al., 2014)) according to a DALI



search (Holm, 2020). Superposition of the two structures reveals a r.m.s.d. of 1.45 Å for 296 pairs of C $\alpha$ -atoms showing their high similarity.



**Figure 10:** Crystal structure of Copu9 in complex with AHD. **a** Overall structure of dimeric Copu9 structure is shown as a cartoon. One protomer is highlighted in teal ( $\alpha$ -helices and loops) and dark blue ( $\beta$ -strands). AHD is depicted as black sticks. **b** Same color-coding as in **a**. Mg<sup>2+</sup> ions are shown as green spheres.  $\alpha$ -helices are depicted as cylinders and labeled. The NSE motif is highlighted in yellow and the Asp-rich motif in salmon. **c** Same as in panel **b** but rotated by 90 degrees for a top view. **d** Coordination of AHD and the three Mg<sup>2+</sup> ions by the NSE motif (yellow sticks) and the Asp-rich motif (red sticks). **e** Overlay of Copu9 (colors as before) and SdS (PDB-ID: 4OKM; orange). **f** Active site cavity of Copu9 (colors as before) is shown as a grey negative surface. The residues surrounding this cavity are depicted as teal sticks.

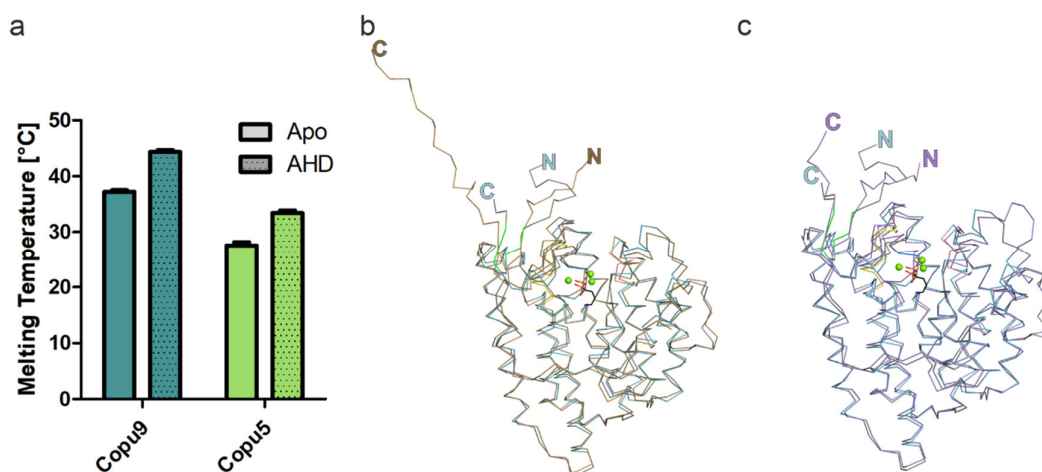
The active site of Copu9 shows the characteristic aspartate rich motif <sup>99</sup>DDWLD<sup>103</sup> on  $\alpha$ -helix C as well as the NSE motif (N235, S239 and E243) on  $\alpha$ -helix H at the opposite side of the active site. These residues coordinate 3 Mg<sup>2+</sup> ions needed for the binding of the diphosphate group of FPP. This is shown in the structure by the coordination of AHD via its diphosphate group. The C-terminal WxxxxRY motif is folded on top of the active site, adopting the same conformation as in the closed conformation of CotB2 (Driller et al., 2018). As a result, R323 and Y324 point towards the active site, R323 forming a salt-bridge with D99 of the aspartate rich motif and Y324 forming a hydrogen bond to the



diphosphate group of AHD. R323 takes the role of the previously described diphosphate sensor by establishing a bidentate salt-bridge to one phosphate group of AHD (Dickschat, 2016). The active site is lined by mostly hydrophobic residues L72, T73, C76, M92, L95, F96, F163, T192, S193, G194, C195, P197, C198, N228, V231, T232, W310 and S314 as already observed for other TPS.

### 3.2.2 Characterization of Copu5

In contrast to Copu9, the purification of Copu5 proved to be much more challenging. After multiple attempts, we were able to obtain pure Copu5 by adding sodium diphosphate to the buffer to mimic the diphosphate of FPP and therefore stabilizing Copu5. TSA buffer screening did give two hits for stabilizing buffers, but unfortunately, neither of them stabilized Copu5 enough to purify it without the addition of diphosphate. Comparison to Copu9 revealed a more than 10 °C lower melting temperature for Copu5 further underlining the observed stability differences (Figure 11).



**Figure 11:** Characterization and computed model of Copu5 and Copu9. **a** TSA results of Copu5 and Copu9. The melting temperature of Copu9 and Copu5 in diphosphate containing buffer is 37.2±0.5 °C/27.6±1.0 °C without AHD respectively and 44.3±0.6 °C/33.5±0.8 °C with addition of 10-fold molar excess of AHD. **b** Copu5(orange) computed model overlaid with the experimental structure of Copu9(teal). **c** Copu9 computed model overlaid (purple) with the experimental structure of Copu9(teal).

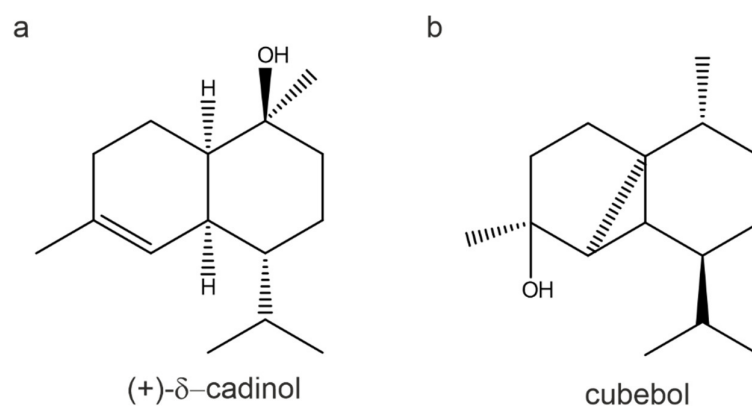
After multiple trials we were not able to crystallize Copu5. To compare the active sites of the two enzymes we used the Robetta server to predict the structure of Copu5 (Baek et al., 2021). As a control we additionally predicted the structure of Copu9 using the

Robetta server (Figure 11). The most striking differences between the modeled Copu9 and the experimental structure are in the N- and C- terminal regions. Also, the NSE and Asp-rich motif are tilted slightly away from the catalytic pocket resulting in a wider active site. The wider active site is a result of the absent  $Mg^{2+}$  ions and the substrate mimic. These are usual differences observed when comparing open, inactive and closed confirmations of TPS (e.g., CotB2). This is also reflected in the model of Copu5.

Interestingly, the residues involved in the formation of the hydrophobic, catalytic pocket are identical between the two TPS. Since they are involved in pre-shaping the geometry of the product or propagation of the carbocation, this might explain the identical product spectrum. To the best of our knowledge, Copu5 and Copu9 are the first reported TPS from the same organism that share identical product profiles and have basically identical active site decoration. It should be noted that differences between the two are observed in the second shell of residues, the ones surrounding the first shell residues. In contrast, comparing the active site residues of Copu5/Copu9 to the cubebol synthase Copu3 (also from *C. puteana* (Mischko et al., 2018)), only 4 of the 12 residues in the active are identical, while the other 8 residues might be critical in guiding the product profiles in different directions.

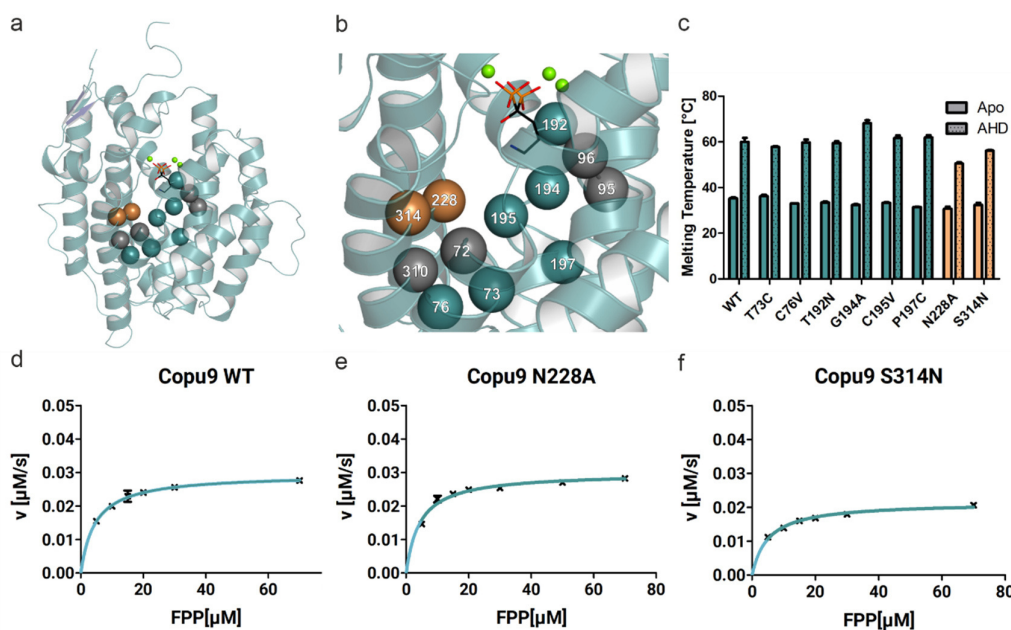
### 3.2.3 Characterization of Copu5 and Copu9 point-mutants

To test the acceptance of point mutants and their effect on the products, we compared active site residues to another sesquiterpene synthase from *C. puteana*, Copu3 (Mischko et al., 2018). The main product of Copu3 is cubebol, which is also produced in minor amounts by Copu5 and Copu9 in addition to their main product (+)- $\delta$ -cadinol (Figure 12). Comparing the 12 identified active site residues from Copu5/Copu9 to the ones of Copu3, 4 residues were identical. The 8 remaining differing residues were chosen to be targeted by a point-mutation study (Figure 13). Initially, our collaborators from the TU Munich analysed the product spectra of each of the 16 point-mutants.



**Figure 12:** Main products of Copu5/ Copu9 and Copu3: **a** (+)-δ-cadinol is the main product of Copu5 and Copu9. **b** Copu3 main product cubebol.

To our surprise, only two point-mutants of Copu9 and 3 point-mutants of Copu5 showed an increase of side product synthesis. Interestingly, these mutants showed a new side product, germacrene D-4-ol, which is one of the Copu3 side products. One point-mutant of Copu5 (P189C) showed reduced product formation. This residue might be essential for catalysis or simply interfere with the active closed conformation of Copu5. As the corresponding point-mutant of Copu9 P197C did not show any change in productivity, the effect observed in Copu5 P189C might also be due to a destabilizing effect on the closed conformation, thus interfering with the active state.



**Figure 13:** Characterization of Copu9 point-mutants. **a** One protomer of Copu9 is shown. Identical amino acids between Copu3 and Copu9 are indicated as grey spheres. Residues that have been subjected to mutagenesis without altering the product are shown as teal-colored spheres. Single amino acid exchanges with an effect on the product profile are shown as orange-colored spheres.

## Results and Discussion

**b** Zoom into the active site. View and colors as in a with residues numbered. **c** TSA results of Copu9 and its variants with and without 10-fold access of AHD showing similar melting temperatures to the wildtype. **d-f** *In vitro* Michaelis–Menten kinetics of Copu9 wildtype and the variant N228A and S314N using the EnzChek™ pyrophosphate Assay. A non-linear regression analysis was performed on the data collected from the time-resolved steady-state kinetic assay.

Next, we conducted *in vitro* kinetic measurements to evaluate the catalytic properties of the point-mutants. While all Copu9 variants showed comparable binding affinities ( $K_m$ ) towards FPP, some slight differences in catalytic turnover could be observed. Variants C76V, N228A and S314N showed slightly decreased turnover compared to the wildtype while variants T73C, T192N, G194A, C195V and P197C retained wildtype activity or showed a slight increase in activity (Figure 13 and Table 1). Variant C76V showed increased substrate specificity although the turnover is minorly reduced. Interestingly, variants N228A and S314N show reduced substrate specificity compared to the wildtype, which might explain the minor shift towards the side products. To exclude the possibility of these differences being an effect of altered stability, we conducted TSA measurements and observed similar melting temperatures for all variants in the apo state suggesting no changes in overall stability. Overall, the observed kinetic parameters are within the range of the previously characterized Cop4 and Cop6 of *C. cinereus* (Lopez-Gallego et al., 2010).

**Table 1:** Steady-state kinetic parameters of Copu9 and its variants calculated from the EnzChek™ pyrophosphate assay. Errors shown are standard errors as calculated with Graphpad Prism 5.0.0.228 (Graph Pad Software Inc.).

Copu9 variant	$k_{cat}$ [ $s^{-1}$ ]	$K_m$ [ $\mu M$ ]	$k_{cat}/K_m$ ( $\times 10^3$ ) [ $s^{-1}M^{-1}$ ]
wild type	$(3.24 \pm 0.03) \times 10^{-2}$	$4.59 \pm 0.17$	7.067
T73C	$(5.02 \pm 0.26) \times 10^{-2}$	$6.28 \pm 1.37$	7.995
C76V	$(2.91 \pm 0.07) \times 10^{-2}$	$3.14 \pm 0.45$	9.252
T192N	$(3.96 \pm 0.08) \times 10^{-2}$	$8.28 \pm 0.60$	4.782
G194A	$(3.60 \pm 0.05) \times 10^{-2}$	$7.00 \pm 0.37$	5.135
C195V	$(3.98 \pm 0.07) \times 10^{-2}$	$8.48 \pm 0.51$	4.692
P197C	$(3.81 \pm 0.10) \times 10^{-2}$	$6.82 \pm 0.77$	5.584
N228A	$(2.62 \pm 0.07) \times 10^{-2}$	$4.42 \pm 0.59$	5.940
S314N	$(1.92 \pm 0.05) \times 10^{-2}$	$5.10 \pm 0.54$	3.755

In contrast to Copu9 variants, Copu5 could only be purified in diphosphate containing buffer due to its stabilizing effect. As the kinetic assay (EnzChek™ diphosphate assay) measures the production of diphosphate and multiple trials to remove the diphosphate from the sample remained unsuccessful, we could not measure the catalytic properties of Copu5 and its variants.

Nevertheless, the effects of the point-mutation were rather mild. None of the variants shifted the product spectrum towards the Copu3 main product cubebol. Likewise, neither of them had drastic effects on the catalytic properties. This contrasts other studies of TPS (i.e., CotB2) where point mutations had drastic effects (Görner et al., 2013; Schrepfer et al., 2016; Driller et al., 2019; Raz et al., 2020a). Overall Copu5 seems to be more receptive to the point-mutations compared to Copu9 that largely compensates the mutations and maintains the wildtype product profile. This might be the effect of general enzyme stability and flexibility but could also point towards the importance of second shell active site residues.



## 4 Conclusion

Over the past decades many concerns have been raised regarding the abuse and depletion of fossil resources in combination with the dramatic effect on climate change. Finding climate friendly production routes for diverse known and new compounds using microbial platforms and enzymatic reactions is a step into the right direction.

We were able to characterize two enantioselective borneol dehydrogenases of plant origin. Despite their total molecular weight of only around 120 kDa we could solve their structures using Cryo-EM to a resolution of at or below 2.0 Å. This result shows how structures of rather small complexes can be solved using Cryo-EM which makes this method an interesting option even for smaller targets as our results show. The low sample consumption makes it a good alternative to crystallization especially for targets that are difficult to produce in large amounts or, as shown in this study, do not crystallize due to higher flexibility. The faster sample preparation and direct structural information make Cryo-EM a valuable tool also for high throughput screening approaches, such as fragment screening.

The rising numbers of fully sequenced genomes enable the identification of to this day uncharacterized TPS from different organisms and with them numerous unknown terpenes with interesting properties. Although the variety of bioactive terpenes produced by fungi is well characterized, their TPS are still underrepresented in the literature (Quin et al., 2014). Therefore, we characterized two TPS originating from the Basidiomycota *C. puteana*, Copu5 and Copu9, that share identical product spectra. Remarkably, this has been the first time that two TPS from the same organism with identical product spectra have been reported.

We were able to solve the first structure of a TPS originating from Basidiomycota. Based on this experimental structure and a predicted structure, we compared both TPS revealing open, inactive conformations for the predicted structures in contrast to the closed structure obtained by crystallization in presence of the substrate mimic AHD. The indistinguishable product spectra seem to be the result of identically decorated active sites. Interestingly, the point-mutations introduced to the active site in this study had a rather mild effect on enzyme activity and stability, which contrasts with other studies

with TPS (i.e., CotB2). It is also fascinating how Copu5 is more susceptible to point-mutations than Copu9, which might be the effect of general enzyme stability and flexibility but could also point towards the importance of second shell active site residues. Previous studies showed the significance of linking residues between the  $\alpha$ -helices G1 and G2 in other TPS (Yoshikuni et al., 2006; Baer et al., 2014; Karunanithi and Zerbe, 2019). While mutations in this region affect the product profile of Copu5, this was not true for Copu9. This linker region is subjected to a conformational change upon substrate binding, triggered by an effector triad at the C-terminal end of  $\alpha$ -helix G1. This triad classically consists of a phosphate sensing arginine, a linking aspartate and the effector carbonyl oxygen of a glycine residue and triggers the ionization of the substrate, thus the start of carbocation cyclization. In the case of Copu5 and Copu9 this glycine residue is exchanged to a serine residue. Mutations of Copu9 in the region of this triad did not have any significant effect on neither the product spectrum nor the catalytic properties underlining its robust structure. This structural robustness of Copu9 could be also responsible for its high yields. This might also suggest that further structural elements influence the catalytic performance and product spectra. This makes Copu5 and Copu9 an interesting model system to gain more mechanistic insights into class I TPS and improve their application in biotechnology.

Since structure and function go hand in hand, the structural elucidation of biotechnologically relevant enzymes remains an important part of their characterization. Structural knowledge of newly discovered enzymes can help improving their stability, selectivity, and productivity by rational design. This in turn will enable faster development of enzymes adjusted to the industrial needs.

The better we understand the underlying mechanisms the better we will be able to optimize them for industrial usage.



---

## References

- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., and Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science (New York, N.Y.)* *373*, 871-876. <https://doi.org/10.1126/science.abj8754>.
- Baer, P., Rabe, P., Fischer, K., Citron, C.A., Klapschinski, T.A., Groll, M., and Dickschat, J.S. (2014). Induced-fit mechanism in class I terpene cyclases. *Angewandte Chemie (International ed. in English)* *53*, 7652-7656. <https://doi.org/10.1002/anie.201403648>.
- Brück, T., Kourist, R., and Loll, B. (2014). Production of Macrocyclic Sesqui- and Diterpenes in Heterologous Microbial Hosts: A Systems Approach to Harness Nature's Molecular Diversity. *ChemCatChem*, n/a-n/a. <https://doi.org/10.1002/cctc.201300733>.
- Carugo, O., and Bordo, D. (1999). How many water molecules can be detected by protein crystallography? *Acta crystallographica. Section D, Biological crystallography* *55*, 479-483. <https://doi.org/10.1107/s0907444998012086>.
- Chánique, A.M., Dimos, N., Drienovská, I., Calderini, E., Pantín, M.P., Helmer, C.P.O., Hofer, M., Sieber, V., Parra, L.P., and Loll, B., et al. (2021). A Structural View on the Stereospecificity of Plant Borneol-Type Dehydrogenases. *ChemCatChem* *13*, 2262-2277. <https://doi.org/10.1002/cctc.202100110>.
- Christianson, D.W. (2008). Unearthing the roots of the terpenome. *Current opinion in chemical biology* *12*, 141-150. <https://doi.org/10.1016/j.cbpa.2007.12.008>.
- Christianson, D.W. (2017). Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews* *117*, 11570-11648. <https://doi.org/10.1021/acs.chemrev.7b00287>.
- Croteau, R., Hooper, C.L., and Felton, M. (1978). Biosynthesis of monoterpenes. Partial purification and characterization of a bicyclic monoterpene dehydrogenase from sage (*Salvia officinalis*). *Archives of Biochemistry and Biophysics* *188*, 182-193. [https://doi.org/10.1016/0003-9861\(78\)90371-5](https://doi.org/10.1016/0003-9861(78)90371-5).
- Curgenvén, J.B. (1891). On the Use of the Oil of Eucalyptus Globulus in Scarlet Fever and Other Infectious Diseases. *Transactions. Epidemiological Society of London* *9*, 93-102.
- Dickschat, J.S. (2016). Bacterial terpene cyclases. *Nat. Prod. Rep.* *33*, 87-110. <https://doi.org/10.1039/C5NP00102A>.
- DiCosimo, R., McAuliffe, J., Poulouse, A.J., and Bohlmann, G. (2013). Industrial use of immobilized enzymes. *Chem. Soc. Rev.* *42*, 6437-6474. <https://doi.org/10.1039/C3CS35506C>.

- Dimos, N., Helmer, C.P.O., Chánique, A.M., Wahl, M.C., Kourist, R., Hilal, T., and Loll, B. (2022). CryoEM analysis of small plant biocatalysts at sub-2 Å resolution. *Acta crystallographica. Section D, Structural biology* 78, 113-123. <https://doi.org/10.1107/S205979832101216X>.
- Drienovská, I., Kolanović, D., Chánique, A., Sieber, V., Hofer, M., and Kourist, R. (2020). Molecular cloning and functional characterization of a two highly stereoselective borneol dehydrogenases from *Salvia officinalis* L. *Phytochemistry* 172, 112227. <https://doi.org/10.1016/j.phytochem.2019.112227>.
- Driller, R., Garbe, D., Mehlmer, N., Fuchs, M., Raz, K., Major, D.T., Brück, T., and Loll, B. (2019). Current understanding and biotechnological application of the bacterial diterpene synthase CotB2. *Beilstein journal of organic chemistry* 15, 2355-2368. <https://doi.org/10.3762/bjoc.15.228>.
- Driller, R., Janke, S., Fuchs, M., Warner, E., Mhashal, A.R., Major, D.T., Christmann, M., Brück, T., and Loll, B. (2018). Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis. *Nat Commun* 9, 3971. <https://doi.org/10.1038/s41467-018-06325-8>.
- Duax, W.L., Ghosh, D., and Pletnev, V. (2000). Steroid dehydrogenase structures, mechanism of action, and disease. *Vitamins and hormones* 58, 121-148. [https://doi.org/10.1016/s0083-6729\(00\)58023-6](https://doi.org/10.1016/s0083-6729(00)58023-6).
- Fujimoto, K., Hara, M., Yamada, H., Sakurai, M., Inaba, A., Tomomura, A., and Katoh, S. (2001). Role of the conserved Ser-Tyr-Lys triad of the SDR family in sepiapterin reductase.
- Ghosh, D., Wawrzak, Z., Weeks, C.M., Duax, W.L., and Erman, M. (1994). The refined three-dimensional structure of 3 alpha,20 beta-hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases. *Structure* 2, 629-640. [https://doi.org/10.1016/S0969-2126\(00\)00064-2](https://doi.org/10.1016/S0969-2126(00)00064-2).
- Görner, C., Häuslein, I., Schrepfer, P., Eisenreich, W., and Brück, T. (2013). Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: A Stereospecific Access to Two New Non-natural Fusicoccane-Type Diterpenes. *ChemCatChem* 5, 3289-3298. <https://doi.org/10.1002/cctc.201300285>.
- Holm, L. (2020). DALI and the persistence of protein shape. *Protein science : a publication of the Protein Society* 29, 128-140. <https://doi.org/10.1002/pro.3749>.
- Hyatt, D.C., and Croteau, R. (2005). Mutational analysis of a monoterpene synthase reaction: altered catalysis through directed mutagenesis of (-)-pinene synthase from *Abies grandis*. *Archives of Biochemistry and Biophysics* 439, 222-233. <https://doi.org/10.1016/j.abb.2005.05.017>.

- Janke, R., Görner, C., Hirte, M., Brück, T., and Loll, B. (2014). The first structure of a bacterial diterpene cyclase: CotB2. *Acta crystallographica. Section D, Biological crystallography* *70*, 1528-1537. <https://doi.org/10.1107/S1399004714005513>.
- Jörnvall, H., Persson, B., Krook, M., Atrian, S., González-Duarte, R., Jeffery, J., and Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). *Biochemistry* *34*, 6003-6013. <https://doi.org/10.1021/bi00018a001>.
- Jörnvall, H., Persson, M., and Jeffery, J. (1981). Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. *Proceedings of the National Academy of Sciences of the United States of America* *78*, 4226-4230. <https://doi.org/10.1073/pnas.78.7.4226>.
- Karunanithi, P.S., and Zerbe, P. (2019). Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Frontiers in plant science* *10*, 1166. <https://doi.org/10.3389/fpls.2019.01166>.
- Kathiravan, G., Sureban, S.M., Sree, H.N., Bhuvaneshwari, V., and Kramony, E. (2012). Isolation of anticancer drug TAXOL from *Pestalotiopsis breviseta* with apoptosis and B-Cell lymphoma protein docking studies. *Journal of Basic and Clinical Pharmacy* *4*, 14-19. <https://doi.org/10.4103/0976-0105.109402>.
- Kavanagh, K.L., Jörnvall, H., Persson, B., and Oppermann, U. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* *65*, 3895-3906. <https://doi.org/10.1007/s00018-008-8588-y>.
- Khine, A.A., Chen, H.P., Huang, K.F., and Ko, T.P. (2020). Structural characterization of borneol dehydrogenase from *Pseudomonas sp.* TCU-HL1. *Acta crystallographica. Section F, Structural biology communications* *76*, 309-313. <https://doi.org/10.1107/S2053230X20008584>.
- Kim, S.-Y., Zhao, P., Igarashi, M., Sawa, R., Tomita, T., Nishiyama, M., and Kuzuyama, T. (2009). Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two p450 hydroxylases. *Chemistry & biology* *16*, 736-743. <https://doi.org/10.1016/j.chembiol.2009.06.007>.
- Lesk, A.M. (1995). NAD-binding domains of dehydrogenases. *Current opinion in structural biology* *5*, 775-783. [https://doi.org/10.1016/0959-440x\(95\)80010-7](https://doi.org/10.1016/0959-440x(95)80010-7).
- Lodeiro, S., Segura, M.J.R., Stahl, M., Schulz-Gasch, T., and Matsuda, S.P.T. (2004). Oxidosqualene cyclase second-sphere residues profoundly influence the product profile. *Chembiochem : a European journal of chemical biology* *5*, 1581-1585. <https://doi.org/10.1002/cbic.200400086>.
- Lopez-Gallego, F., Agger, S.A., Abate-Pella, D., Distefano, M.D., and Schmidt-Dannert, C. (2010). Sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus*: catalytic

- promiscuity and cyclization of farnesyl pyrophosphate geometric isomers. *Chembiochem : a European journal of chemical biology* *11*, 1093-1106. <https://doi.org/10.1002/cbic.200900671>.
- Masyita, A., Mustika Sari, R., Dwi Astuti, A., Yasir, B., Rahma Rumata, N., Emran, T.B., Nainu, F., and Simal-Gandara, J. (2022). Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. *Food Chemistry: X* *13*, 100217. <https://doi.org/10.1016/j.fochx.2022.100217>.
- Miao, Z., and Cao, Y. (2016). Quantifying side-chain conformational variations in protein structure. *Scientific reports* *6*, 37024. <https://doi.org/10.1038/srep37024>.
- Mischko, W., Hirte, M., Fuchs, M., Mehlmer, N., and Brück, T.B. (2018). Identification of sesquiterpene synthases from the Basidiomycota *Coniophora puteana* for the efficient and highly selective  $\beta$ -copaene and cubebol production in *E. coli*. *Microb Cell Fact* *17*, 164. <https://doi.org/10.1186/s12934-018-1010-z>.
- Quin, M.B., Flynn, C.M., and Schmidt-Dannert, C. (2014). Traversing the fungal terpenome. *Nat. Prod. Rep.* *31*, 1449-1473. <https://doi.org/10.1039/C4NP00075G>.
- Rabe, P., Schmitz, T., and Dickschat, J.S. (2016). Mechanistic investigations on six bacterial terpene cyclases. *Beilstein journal of organic chemistry* *12*, 1839-1850. <https://doi.org/10.3762/bjoc.12.173>.
- Raz, K., Driller, R., Brück, T., Loll, B., and Major, D.T. (2020a). Understanding the role of active site residues in CotB2 catalysis using a cluster model. *Beilstein journal of organic chemistry* *16*, 50-59. <https://doi.org/10.3762/bjoc.16.7>.
- Raz, K., Driller, R., Dimos, N., Ringel, M., Brück, T., Loll, B., and Major, D.T. (2020b). The Impression of a Nonexisting Catalytic Effect: The Role of CotB2 in Guiding the Complex Biosynthesis of Cyclooctat-9-en-7-ol. *Journal of the American Chemical Society* *142*, 21562-21574. <https://doi.org/10.1021/jacs.0c11348>.
- Rigoldi, F., Donini, S., Redaelli, A., Parisini, E., and Gautieri, A. (2018). Review: Engineering of thermostable enzymes for industrial applications. *APL Bioengineering* *2*, 11501. <https://doi.org/10.1063/1.4997367>.
- Ringel, M., Dimos, N., Himpich, S., Haack, M., Huber, C., Eisenreich, W., Schenk, G., Loll, B., and Brück, T. (2022). Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of  $\delta$ -cadinol. *Microbial cell factories* *21*, 64. <https://doi.org/10.1186/s12934-022-01791-8>.
- Rossmann, M.G., Liljas, A., Brändén, C.-I., and Banaszak, L.J. (1975). 2 Evolutionary and Structural Relationships among Dehydrogenases. In *Enzymes*, P.D. Boyer, ed. (New York [u.a.]: Acad. Press), pp. 61–102.

- Rudolf, J.D., and Chang, C.-Y. (2020). Terpene synthases in disguise: enzymology, structure, and opportunities of non-canonical terpene synthases. *Natural product reports* 37, 425-463. <https://doi.org/10.1039/c9np00051h>.
- Satyal, P., Jones, T.H., Lopez, E.M., McFeeters, R.L., Ali, N.A.A., Mansi, I., Al-kaf, A.G., and Setzer, W.N. (2017). Chemotypic Characterization and Biological Activity of *Rosmarinus officinalis*. *Foods* 6, 20. <https://doi.org/10.3390/foods6030020>.
- Schrepfer, P., Buettner, A., Goerner, C., Hertel, M., van Rijn, J., Wallrapp, F., Eisenreich, W., Sieber, V., Kourist, R., and Brück, T. (2016). Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proceedings of the National Academy of Sciences of the United States of America* 113, E958-67. <https://doi.org/10.1073/pnas.1519680113>.
- Silvestre, A.J., and Gandini, A. (2008). Chapter 2 - Terpenes: Major Sources, Properties and Applications. In *Monomers, polymers and composites from renewables resources*, M.N. Belgacem and A. Gandini, eds. (Oxford [etc.]: Elsevier), pp. 17–38.
- Starks, C.M., Back, K., Chappell, J., and Noel, J.P. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science (New York, N.Y.)* 277, 1815-1820. <https://doi.org/10.1126/science.277.5333.1815>.
- Tang, X., Zhang, F., Zeng, T., Li, W., Yin, S., and Wu, R. (2020). Enzymatic Plasticity Inspired by the Diterpene Cyclase CotB2. *ACS Chemical Biology* 15, 2820-2832. <https://doi.org/10.1021/acscchembio.0c00645>.
- Tomita, T., Kim, S.-Y., Teramoto, K., Meguro, A., Ozaki, T., Yoshida, A., Motoyoshi, Y., Mori, N., Ishigami, K., and Watanabe, H., et al. (2017). Structural Insights into the CotB2-Catalyzed Cyclization of Geranylgeranyl Diphosphate to the Diterpene Cyclooctat-9-en-7-ol. *ACS Chemical Biology* 12, 1621-1628. <https://doi.org/10.1021/acscchembio.7b00154>.
- Wendt, K.U., and Schulz, G.E. (1998). Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* 6, 127-133. [https://doi.org/10.1016/S0969-2126\(98\)00015-X](https://doi.org/10.1016/S0969-2126(98)00015-X).
- Wilderman, P.R., and Peters, R.J. (2007). A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. *Journal of the American Chemical Society* 129, 15736-15737. <https://doi.org/10.1021/ja074977g>.
- Wriessnegger, T., Augustin, P., Engleder, M., Leitner, E., Müller, M., Kaluzna, I., Schürmann, M., Mink, D., Zellnig, G., and Schwab, H., et al. (2014). Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of *Pichia pastoris*. *Metabolic Engineering* 24, 18-29. <https://doi.org/10.1016/j.ymben.2014.04.001>.

## References

---

Yoshikuni, Y., Martin, V.J.J., Ferrin, T.E., and Keasling, J.D. (2006). Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chemistry & biology* 13, 91-98. <https://doi.org/10.1016/j.chembiol.2005.10.016>.

Zhang, C., and Hong, K. (2020). Production of Terpenoids by Synthetic Biology Approaches. *Front. Bioeng. Biotechnol.* 8, 347. <https://doi.org/10.3389/fbioe.2020.00347>.

## Abbreviations

°C	degrees Celsius
Å	Ångström
AHD	alendronate
BDH	borneol dehydrogenase
Cryo-EM	cryo-electron microscopy
DMAPP	dimethylallyl diphosphate
FPP	farnesyl diphosphate
G3P	glyceraldehyde 3-phosphate
GGPP	geranyl geranyl diphosphate
GPP	geranyl diphosphate
IPP	isopentyl diphosphate
$k_{cat}$	turnover number
kDa	kilodalton
$K_m$	Michaelis-Menten constant
MALS	multi-angle light-scattering
MEP	methylerythritol phosphate
MVA	mevalonate
NAD <sup>+</sup>	nicotinamide adenine dinucleotide
PDB	protein data bank
PO/OH	5/4 pentaerythritol propoxylate
SDR	short-chain dehydrogenase/ reductase
r.m.s.d.	root-mean-square deviation
SEC	size-exclusion chromatography
TPS	terpene synthase
TSA	thermal shift assay

---

## Table of Figures

<b>Figure 1:</b> Terpene synthesis <i>in vivo</i> . .....	8
<b>Figure 2:</b> Close-up to the catalytic pocket of 3 $\alpha$ /20 $\beta$ hydroxysteroid dehydrogenase ..	11
<b>Figure 3:</b> Structural comparison of three class I TPS. ....	12
<b>Figure 4:</b> Crystal structure of <i>SrBDH1</i> .....	18
<b>Figure 5:</b> <i>SrBDH1</i> active site residues. ....	19
<b>Figure 6:</b> Cryo-EM structure of <i>SoBDH2</i> . ....	20
<b>Figure 7:</b> Cryo-EM structure of <i>SrBDH1</i> .....	21
<b>Figure 8:</b> Grouped B factor mapped onto the densities of a <i>SoBDH2</i> and b <i>SrBDH1</i> . ....	22
<b>Figure 9:</b> Comparison of the <i>SoBDH2</i> and <i>PsBDH</i> to <i>SrBDH1</i> . ....	23
<b>Figure 10:</b> Crystal structure of Copu9 in complex with AHD. ....	26
<b>Figure 11:</b> Characterization and computed model of Copu5 and Copu9.....	27
<b>Figure 12:</b> Main products of Copu5/ Copu9 and Copu3.....	29
<b>Figure 13:</b> Characterization of Copu9 point-mutants.....	29



## Publication I

Chánique, Andrea M.; Dimos, Nicole; Drienovská, Ivana; Calderini, Elia; Pantín, Mónica P.; Helmer, Carl P. O.; Hofer, Michael; Sieber, Volker; Parra, Loreto P.; Loll, Bernhard; Kourist, Robert (2021): A Structural View on the Stereospecificity of Plant Borneol-Type Dehydrogenases. *ChemCatChem* 13 (9)  
DOI: <https://doi.org/10.1002/cctc.202100110>



# A Structural View on the Stereospecificity of Plant Borneol-Type Dehydrogenases

Andrea M. Chánique<sup>+, [a, b]</sup> Nicole Dimos<sup>+, [c]</sup> Ivana Drienovská,<sup>[a]</sup> Elia Calderini,<sup>[a]</sup> Mónica P. Pantín,<sup>[a]</sup> Carl P. O. Helmer,<sup>[c]</sup> Michael Hofer,<sup>[d]</sup> Volker Sieber,<sup>[d, e]</sup> Loreto P. Parra,<sup>[f]</sup> Bernhard Loll,<sup>\*, [c]</sup> and Robert Kourist<sup>\*, [a]</sup>

The development of sustainable processes for the valorization of byproducts and other waste streams remains an ongoing challenge in the field of catalysis. Racemic borneol, isoborneol and camphor are currently produced from  $\alpha$ -pinene, a side product from the production of cellulose. The pure enantiomers of these monoterpenoids have numerous applications in cosmetics and act as reagents for asymmetric synthesis, making an enzymatic route for their separation into optically pure enantiomers a desirable goal. Known short-chain borneol-type dehydrogenases (BDHs) from plants and bacteria lack the required specificity, stability or activity for industrial utilization. Prompted by reports on the presence of pure (–)-borneol and (–)-camphor in essential oils from rosemary, we set out to investigate dehydrogenases from the genus *Salvia* and discov-

ered a dehydrogenase with high specificity ( $E > 120$ ) and high specific activity ( $> 0.02 \text{ U mg}^{-1}$ ) for borneol and isoborneol. Compared to other specific dehydrogenases, the one reported here shows remarkably higher stability, which was exploited to obtain the first three-dimensional structure of an enantiospecific borneol-type short-chain dehydrogenase. This, together with docking studies, led to the identification of a hydrophobic pocket in the enzyme that plays a crucial role in the stereo discrimination of bornane-type monoterpenoids. The kinetic resolution of borneol and isoborneol can be easily integrated into the existing synthetic route from  $\alpha$ -pinene to camphor thereby allowing the facile synthesis of optically pure monoterpenols from an abundant renewable source.

## Introduction

The enantiospecificity of many enzymes is a key feature that is currently exploited for numerous biotechnological processes. Often, increase or inversion of the specificity by enzyme engineering is necessary, for which an understanding of the molecular mechanisms of enantiospecificity becomes crucial. Several short-chain dehydrogenases-reductases (SDR) exert an intriguing two-fold stereospecificity in the conversion of chiral monoterpenoids and could potentially be exploited for the synthesis of pure ingredients from essential oils.<sup>[1]</sup> These

enzymes catalyze the kinetic resolution of racemic alcohols and are also capable of forming a stereocenter by the asymmetric reduction of an enantiotopic or diastereotopic keto-group.

Terpenes are a structurally and functionally diverse group of molecules. Their diversity is generated by the outstanding selectivity of the biosynthetic enzymes that participate in both the formation of the hydrocarbon skeleton and its oxyfunctionalization and further decoration.<sup>[2,3]</sup> Among the vast diversity of bioactive terpenoids are the bornane-type bicyclic monoterpenoids borneol (*endo-1a*) and isoborneol (*exo-1a*) and their corresponding ketone, camphor (**1b**) (Scheme 1). They are

[a] A. M. Chánique,<sup>+</sup> Prof. I. Drienovská, Dr. E. Calderini, M. P. Pantín, Prof. R. Kourist  
Institute of Molecular Biotechnology  
Graz University of Technology  
Petersgasse 14  
8010, Graz (Austria)  
E-mail: kourist@tugraz.at

[b] A. M. Chánique<sup>+</sup>  
Department of Chemical and Bioprocesses Engineering  
School of Engineering  
Pontificia Universidad Católica de Chile  
Vicuña Mackenna 4860  
7810000, Santiago (Chile)


[c] N. Dimos,<sup>+</sup> C. P. O. Helmer, B. Loll  
Institute of Chemistry and Biochemistry  
Department of Biology, Chemistry, Pharmacy  
Laboratory of Structural Biochemistry  
Free University of Berlin  
Takustr. 6  
14195 Berlin (Germany)  
E-mail: loll@chemie.fu-berlin.de


[d] Dr. M. Hofer, Dr. V. Sieber  
Fraunhofer Institute for Interfacial Engineering and Biotechnology IGB  
Schulgasse 11a  
94315, Straubing (Germany)

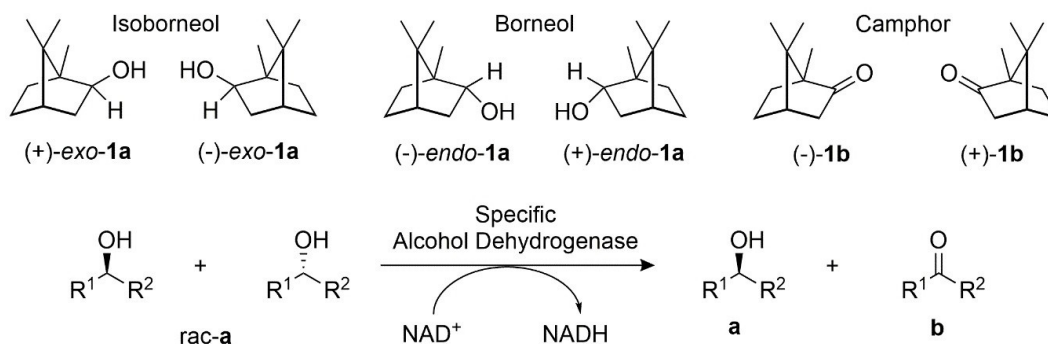
[e] Dr. V. Sieber  
Technical University of Munich  
Straubing Campus for Biotechnology and Sustainability  
Schulgasse 16  
94315 Straubing (Germany)

[f] Prof. L. P. Parra  
Institute for Biological and Medical Engineering  
Schools of Engineering, Medicine and Biological Sciences  
Pontificia Universidad Católica de Chile  
Vicuña Mackenna 4860  
7810000, Santiago (Chile)

[\*] These authors contributed equally to this work.

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cctc.202100110>

 © 2021 The Authors. ChemCatChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



**Scheme 1.** Chemical structures of the terpenoids *exo-1a*, *endo-1a* and **1b** and generic oxidation catalyzed by a specific alcohol dehydrogenase.

found in essential oil extracts from different plants. Their pure isomers, and synthetic mixtures thereof, are widely used in Chinese medicine.<sup>[4]</sup> Different studies have suggested their activity as anti-inflammatory, neuroprotective and vasorelaxant agents, making them valuable ingredients for health-related formulations.<sup>[5]</sup> Derivatives of optically pure isoborneol such as (2*S*)-(–)-3-*exo*-(morpholino)isoborneol and (2*S*)-(–)-3-*exo*-(dimethylamino)isoborneol also find application as chiral ligands in asymmetric synthesis.<sup>[6,7]</sup>

The biosynthesis of **1b** proceeds from the cyclization and subsequent hydrolysis of geranyl diphosphate by borneol diphosphate synthase and borneol synthase.<sup>[8]</sup> Enzymatic oxidation of **1a** by an alcohol dehydrogenase then gives rise to **1b**. This biosynthetic pathway does not necessarily require an enantiospecific borneol dehydrogenase as (+)-*endo-1a* is produced from geranyl pyrophosphate already in optically pure form. In fact, the first recombinantly produced borneol dehydrogenases from the plants *Artemisia annua*<sup>[9,10]</sup> and *Lavandula intermedia*,<sup>[11]</sup> and the bacterium *Pseudomonas* sp. TCU-HL1<sup>[12]</sup> did not show any significant enantiospecificity, which seemed to confirm this notion.

Racemic **1b** can easily be obtained by chemical synthesis from  $\alpha$ - and  $\beta$ -pinene with racemic **1a** as intermediate.<sup>[13]</sup> As  $\alpha$ - and  $\beta$ -pinene are side-streams in the processing of pine trees, intermediates of this route represent inexpensive starting material for the production of optically pure isomers of bornane type monoterpenoids.<sup>[13]</sup> This can be achieved by the kinetic resolution of *rac-1b* using a dehydrogenase that is enantiospecific towards **1b** and diastereoselective for the formation of (+)-*endo-1a*. Using iterative saturation mutagenesis, we recently created the first described variant of a bacterial short-chain dehydrogenase capable of doing this.<sup>[14]</sup> Alternatively, the oxidative kinetic resolution of *rac-exo-1a* as intermediate of the existing chemical route is shorter and would provide a clean alternative to the isolation process from plants.<sup>[15]</sup>

In the 1980's, Croteau and coworkers reported the presence of (+)-specific borneol dehydrogenase activity in sage leaf homogenate (*Salvia officinalis* L.).<sup>[16]</sup> Additionally, the essential oil from the related *Salvia rosmarinus* is reported to contain a high relative content of (–)-*endo-1a* and (–)-**1b**.<sup>[17]</sup> Based on the hypothesis that the synthesis of optically pure **1a** and **1b** isomers in *Salvia* species would require highly enantiospecific

enzymes, we previously identified two dehydrogenases from *S. officinalis* L. that catalyze the specific oxidation of (+)-*endo-1a* with outstanding enantiospecificity ( $E > 200$ ).<sup>[11]</sup> The enantio-preference of these enzymes was at first sight unexpected, as the enzymes preferentially produced (+)-**1b** from racemic *endo-1a*, but (–)-**1b** from racemic *exo-1a*.<sup>[11]</sup> Unfortunately, the low activity and stability of both enzymes represented an obstacle for synthetic application and structure elucidation. In order to find an enzyme with higher stability and to understand whether enantiospecificity is a frequent feature or an exception in this group of dehydrogenases, we continued studying other enzymes from the *Salvia* genus, specifically, *S. rosmarinus* and *S. officinalis*. We also investigated the capability of these plant dehydrogenases to perform the reverse reaction. Herein, we report the specificity, substrate scope and activity of a set of plant borneol-type dehydrogenases, with special emphasis on the highly stable and active borneol dehydrogenase SrBDH1 from *S. rosmarinus*. We also determined the structure of SrBDH1 in complex with NAD<sup>+</sup> at 2.8 Å resolution. This represents the first structure of a selective borneol-type SDR; from which a deeper insight into the molecular basis of the enantiospecificity of the enzyme was obtained.

## Results and discussion

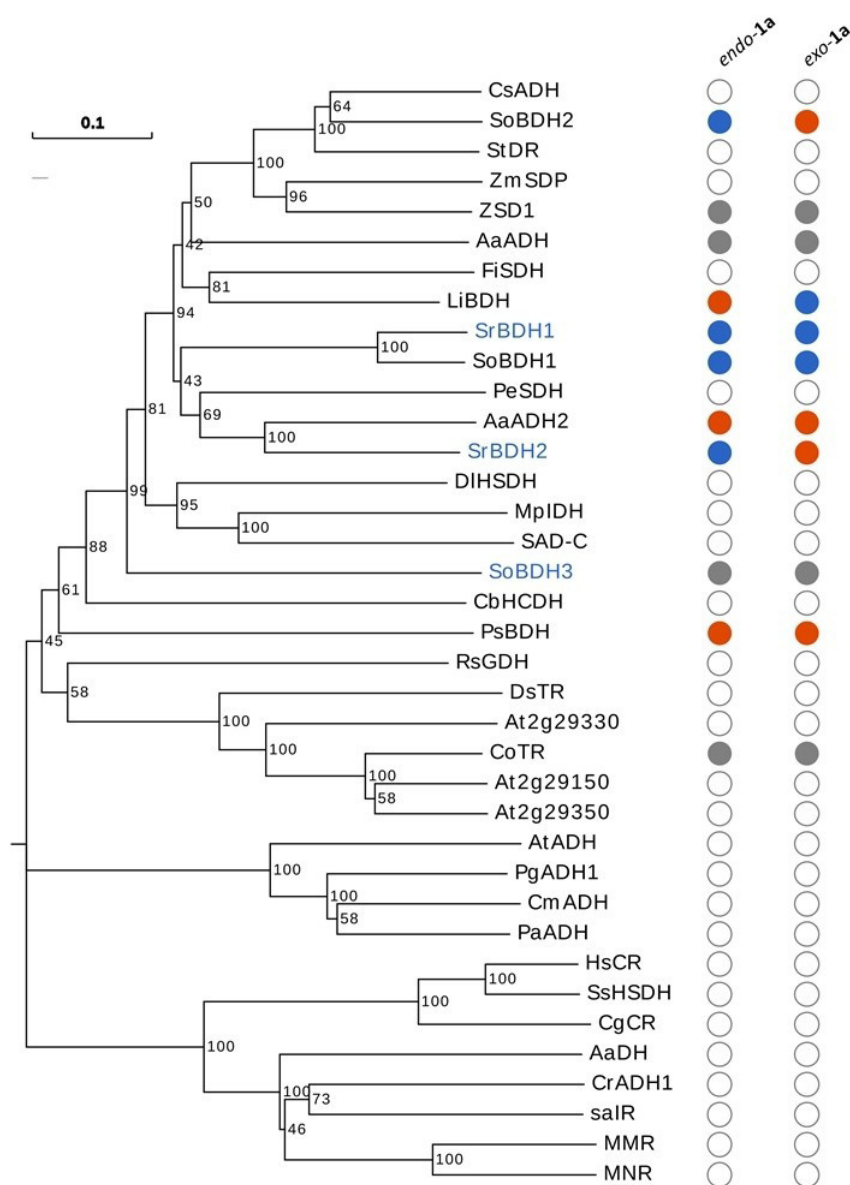
### Sequence analysis

Two putative members of the SDR class from the genome *S. rosmarinus* and one from *S. officinalis* L.<sup>[18]</sup> were identified using the BLASTP server. The percentage of identity with the unselective AaADH2 (*Artemisia annua*)<sup>[10]</sup> ranged between 43% and 53%, making them likely candidates for borneol-converting dehydrogenases. For *S. officinalis* L., the two first hits had already been shown to convert (+)-*endo-1a* in an enantiospecific fashion (SoBDH1 and SoBDH2).<sup>[11]</sup> The three putative dehydrogenases have the typical TGxxx[AG]xG cofactor binding motif and the YxxxK active site motif that is characteristic to the classic SDR family (Figure S1).<sup>[19]</sup> According to the classification suggested by Kallberg *et al.* (2010)<sup>[20]</sup> all of the alcohol dehydrogenases investigated in this paper (except the bacterial borneol dehydrogenase from *Pseudomonas* sp. TCU-HL1

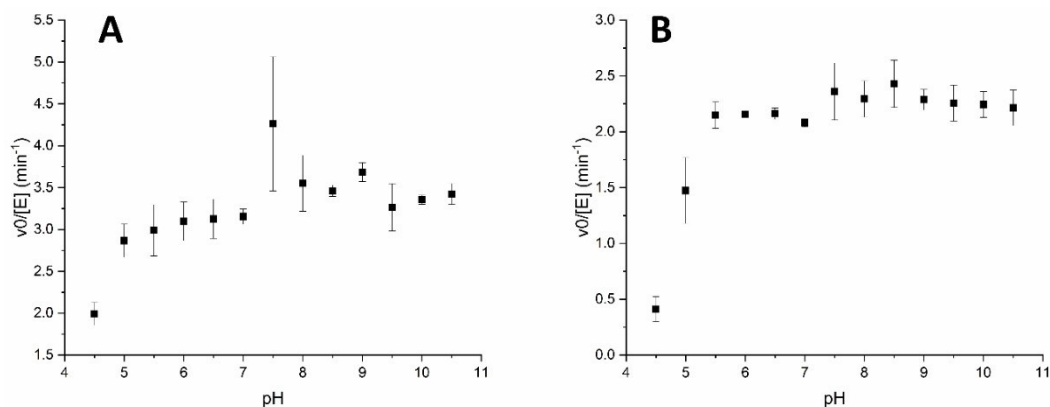
(PsBDH) belong to the SDR110 C subgroup of the SDR superfamily (hereafter, borneol-type dehydrogenases). Other members are the sex determination protein tasselseed-2 from *Zea mays* (ZmSDP) and secoisolariciresinol dehydrogenase from *Podophyllum peltatum* (PeSDH), which show a wide diversity in terms of function and substrate acceptance within the subgroup (Figure 1). Other subgroups displayed in the phylogenetic tree are tropinone reductase-like SDRs (SDR65 C), menthol dehydrogenase-like SDRs (SDR114 C) and carbonyl reductases (SDR21 C).

### Recombinant production and substrate scope

The recombinant production in *E. coli* and subsequent purification of the three enzymes yielded  $\sim 50 \text{ mg L}^{-1}$  for SrBDH1, and values in the same range for SrBDH2 and SoBDH3 (Figure S2). Size exclusion chromatography classifies the enzymes as tetramers (Figure S3). SrBDH1 had a specific activity of  $0.030 \text{ U mg}^{-1}$  towards *rac-endo-1a*, the highest value among borneol-type dehydrogenases from plants. pH stability analysis showed over 50% retention of activity for pH values between 5 and 10.5 after 30 min of incubation at the selected pH (Figure 2A). After 24 h of incubation, the activity showed little variation for pH between 5.5 and 10.5 (Figure 2B). This relatively



**Figure 1.** Phylogram showing evolutionary relationships of different short-chain dehydrogenases. The tree was constructed using Maximum Likelihood method with Mega X software. The three borneol-type dehydrogenases characterized in this paper are highlighted in blue. Specificity for *endo-1a* and *exo-1a* is indicated in circles, with high specificity in blue ( $E > 100$ ) and low specificity in orange. Enzymes described to convert **1a** with no indicated specificity are shown with grey circles and enzymes not described to convert it are shown with white. The bootstrap values are shown next to each branch. For the sequences of short-chain dehydrogenases and their accession numbers, please refer to the Supplementary Information.



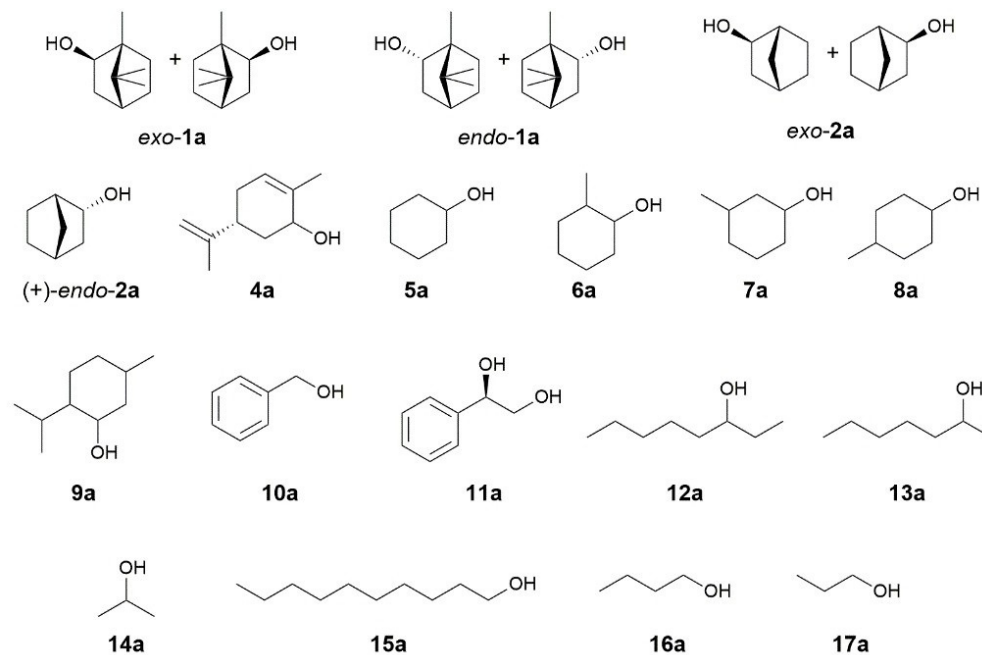
**Figure 2.** Rates for conversion of (+)-*endo*-1a for SrBDH1 after incubation at different pH values. The enzyme was incubated at room temperature for 30 min (A) and 24 h (B) at the indicated pH and then the initial rate of oxidation of (+)-*endo*-1a was determined at pH 9 by following NADH formation at 340 nm.

high stability at a broad pH range is particularly interesting for reduction reactions, which are favored at an acidic pH. It is also observed from Figure 2 that after 24 h of incubation at neutral pH values, the enzyme still retained 50% of its initial activity. This is a very promising starting point for further optimization by protein and reaction engineering. The higher stability of SrBDH1 in comparison to previously investigated enzymes from *S. officinalis* L., also allowed us to successfully elucidate its crystal structure.

Determination of the specific activities of the BDHs in the oxidation of a set of primary and secondary alcohols showed that the enzymes do not exclusively oxidize bicyclic monoterpenes; they also accept monocyclic and linear substrates (Figure 3, Table 1). The dehydrogenases clearly favor secondary

alcohols over primary alcohols. All enzymes had the highest specific activity either for *endo*-1a (SrBDH1 and PsBDH), (–)-carveol (4a) (SrBDH2, SoBDH3 and AaADH2) or 3-methyl cyclohexanol (7a) (SoBDH2).

Interestingly, we noted that none of the enzymes oxidized menthol (9a), despite its structural similarity to 4a. In fact, menthol dehydrogenases appear as a separated clade in the phylogram (Figure 1, MMR and MNR). A BLASTP search using menthol and neomenthol dehydrogenases from *Mentha x piperita* as queries suggests that this subfamily of SDRs is also present in *S. officinalis* and *S. rosmarinus* (best hits with 67% identity for MNR and 68% identity for MMR), leading us to think that *Salvia* plants have independent biocatalysts for the synthesis of menthone-like compounds and camphor-like com-



**Figure 3.** Primary and secondary alcohols tested in the substrate scope study of BDHs.

**Table 1.** Specific activities measured for different borneol-type dehydrogenases.

Substrate	Specific activity [mU/mg] <sup>[a]</sup>		SoBDH3 <sup>[b]</sup>	AaADH2 <sup>[b]</sup>	SoBDH2 <sup>[b]</sup>	PsBDH <sup>[b]</sup>
	SrBDH1 <sup>[b]</sup>	SrBDH2 <sup>[b]</sup>				
exo-1 a	24	5.1	0.3	53	13	115
endo-1 a	<b>30</b>	10.0	n.c. <sup>[c]</sup>	22	18	<b>122</b>
exo-2 a	4.8	1.5	n.c.	29	n.c.	69
(+)-endo-2 a	3.9	5.9	n.c.	23	n.d.	28
endo-3 a	4.6	1.0	n.c.	14	7.1	30
<b>4 a</b>	2.9	<b>16</b>	<b>0.46</b>	<b>88</b>	12	5.0
<b>5 a</b>	3.3	n.c.	n.c.	29	6.7	28
<b>6 a</b>	10	2.5	n.c.	17	18	23
<b>7 a</b>	4.4	n.c.	0.23	21	<b>39</b>	18
<b>8 a</b>	8.3	n.c.	0.21	15	3.8	11
<b>9 a</b>	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
<b>10 a</b>	n.c.	n.c.	n.c.	5.9	n.c.	n.c.
<b>11 a</b>	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
<b>12 a</b>	13	4.4	n.c.	8.7	2.3	43
<b>13 a</b>	9.8	2.0	n.c.	63	6.3	6.2
<b>14 a</b>	n.c.	n.c.	n.c.	10.0	n.c.	n.c.
<b>15 a</b>	n.c.	n.c.	n.c.	n.c.	3.4	n.c.
<b>16 a</b>	n.c.	n.c.	n.c.	4.8	4.7	n.c.
<b>17 a</b>	3.3	n.c.	0.2	n.c.	6.5	3.2

[a] The highest activity for each enzyme is highlighted in bold letters. For the experimental error from technical triplicates, please refer to **Table S1**, [b] AaADH2: alcohol dehydrogenase from *A. annua*<sup>[10]</sup>; SoBDH2: borneol dehydrogenase from *S. officinalis* L.<sup>[11]</sup>; SrBDH1/2: borneol-like dehydrogenases from *S. rosmarinus*; SoBDH3: borneol-like dehydrogenase from *S. officinalis* L.; PsBDH: borneol dehydrogenase from *Pseudomonas* sp. TCU-HL1.<sup>[12]</sup> [c] n.c. = no conversion detected. Reactions with substrates displaying specific activities less than twice the blank are not shown.

pounds. We also noted differences in the specific activities obtained for the unnatural substrate cyclohexanol (**5 a**) and the isomers 2-methylcyclohexanol (**6 a**), 3-methylcyclohexanol (**7 a**) and 4-methylcyclohexanol (**8 a**). For SrBDH1, SrBDH2 and SoBDH2 we observed a higher specific activity for **6 a** in comparison to **5 a**. This situation is similar to what we observed with *endo-1 a* vs *endo-2 a*, where the presence of methyl groups seems to improve the fit in the active site, making the reaction faster. Conversely, the unselective enzymes AaADH2 and PsBDH have a higher specific activity for the alcohol without the methyl group, **5 a**. The specific activities observed for SrBDH1 with *endo-1 a* and *exo-1 a* are the best among the selective enzymes studied and fall within the same range of the unselective borneol dehydrogenase from *A. annua* AaADH2, making SrBDH1 interesting for biocatalytic applications. Cofactor usage was also studied spectrophotometrically. Results showed that AaADH2 was the only enzyme also capable of using NADP<sup>+</sup>, albeit less efficiently than NAD<sup>+</sup>. The other tested enzymes did not show measurable activity for NADP<sup>+</sup> (Figure S4).

### Enantiospecificity and selectivity towards bicyclic alcohols and ketones

To obtain a more systematic overview on the enantiospecificity of the dehydrogenases, we investigated the kinetic resolution of the three bornane type monoterpenols *endo-1 a*, *exo-1 a* and fenchol (*endo-3 a*) and the structurally related *exo-norborneol* (*exo-2 a*) (Table 2, Scheme 2). SoBDH1 and SrBDH1 showed outstanding enantiospecificity ( $E > 200$ ) towards both *endo-1 a* and *exo-1 a*, while SoBDH2 and SrBDH2 showed specificity for *endo-1 a*, but not for *exo-1 a*. The differences in the specificity of

this set of enzymes towards *exo-1 a* and the smaller *exo-2 a* were striking. In particular, SoBDH1 and SrBDH1 were hardly specific towards *exo-2 a*, in contrast with *exo-1 a*, which might be an indication that a precise fit of the substrate is important for the specificity of both enzymes. It should be noted, however, that the activity of SrBDH1 towards *exo-2 a* is substantially lower than towards *exo-1 a*, making comparisons of the specificity difficult. SrBDH1 shows a similar drop in activity and specificity in the conversion of the structural isomer *endo-3 a* (Table 2, Scheme 2). The differences in activity and selectivity of SrBDH1 towards these substrates indicates that the position of the methyl substituents is crucial for the substrate recognition of this enzyme. Therefore, the high specificity of SrBDH1 and SoBDH1 towards both *endo-1 a* and *exo-1 a* is somehow counterintuitive and led us to think that these two enzymes share key active pocket configurations that other BDHs lack. None of the plant dehydrogenases have any noteworthy specificity in the resolution of *endo-3 a*. While the investigated enzymes share the enantiopreference for the same isomer of *endo-1 a*, *exo-1 a* and *exo-2 a*, we noted that the preference for the (+) and (−) isomers of *endo-3 a* differed.

PsBDH displays a different behavior in comparison to plant BDHs, as the bacterial enzyme showed very low specificity towards *endo-1 a*, *exo-1 a* and *exo-2 a*, but it is the most specific of the studied enzymes for the oxidation of *endo-3 a*, with  $> 99\%$  ee, and  $E = 27$ . SoBDH3 was not active towards *endo-1 a* and *exo-2 a* and showed little conversion for *exo-1 a*, leading us to assume that this enzyme has a natural substrate with significant structural differences from *endo-1 a*, which is reflected in the divergent sequence compared to other BDHs from plants (Figure 1).

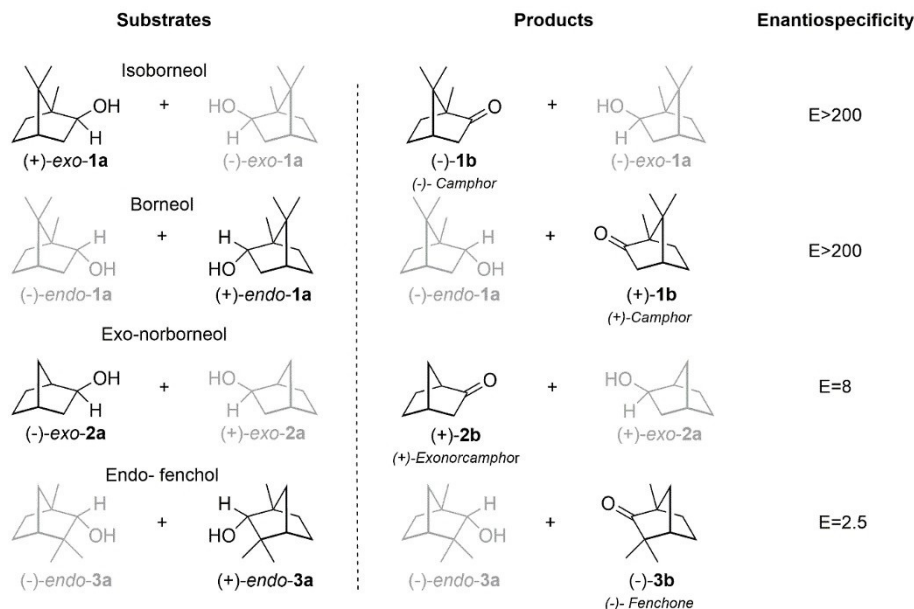
Based on reports describing a short-chain dehydrogenase from the tropinone reductase subfamily capable of catalyzing



**Table 2.** Kinetic resolution of racemic secondary alcohols of the bornane and norbornane types catalyzed by alcohol dehydrogenases at a substrate concentration of 1 mM.

		AaADH2 <sup>[a]</sup>	SoBDH1 <sup>[a]</sup>	SoBDH2 <sup>[a]</sup>	SoBDH3 <sup>[b]</sup>	SrBDH1 <sup>[b]</sup>	SrBDH2 <sup>[b]</sup>	PsBDH <sup>[a]</sup>
<b>exo-1 a</b>	Time (h)	0.5	24	4	48	2	48	0.5
	Specificity	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	%ee <sub>p</sub> <sup>[c]</sup>	4%	99%	13%	> 99%	> 99%	31%	4%
	%ee <sub>s</sub> <sup>[c]</sup>	99%	30%	99%	5%	52%	32%	27%
	%c	96%	28%	88%	5%	34%	51%	87%
E <sup>[f]</sup>	n.d.	> 200	4.6	n.d.	> 200	3	1.3	
<b>endo-1 a</b>	Time (h)	0.5	24	4	n.c. <sup>[d]</sup>	2	48	0.5
	Specificity	(+)	(+)	(+)	n.c.	(+)	(+)	(+)
	%ee <sub>p</sub> <sup>[c]</sup>	46%	> 99%	> 99%	n.c.	> 99%	> 99%	17%
	%ee <sub>s</sub> <sup>[c]</sup>	83%	18%	49%	n.c.	77%	23%	50%
	%c	64%	22%	33%	n.c.	44%	19%	75%
E <sup>[f]</sup>	6.6	> 200	> 200	n.d. <sup>[e]</sup>	> 200	> 200	2.1	
<b>exo-2 a</b>	Time (h)	0.25	48	48	n.c.	48	48	0.25
	Specificity	(-)	(-)	(-)	n.c.	(-)	(-)	(-)
	%ee <sub>p</sub> <sup>[c]</sup>	65%	10%	80%	n.c.	75%	49%	9%
	%ee <sub>s</sub> <sup>[c]</sup>	53%	3%	15%	n.c.	10%	1%	2%
	%c	45%	22%	16%	n.c.	12%	2%	18%
E <sup>[f]</sup>	7.9	1.2	10.3	n.d.	7.7	n.d.	1.2	
<b>endo-3 a</b>	Time (h)	0.25	n.c.	48	48	48	48	48
	Specificity	(+)	n.c.	(-)	n.d.	(+)	(-)	(+)
	%ee <sub>p</sub> <sup>[c,g]</sup>	-	n.c.	-	-	-	-	-
	%ee <sub>s</sub> <sup>[c]</sup>	13%	n.c.	84%	< 1%	26%	6%	> 99%
	%c	67%	n.c.	58%	22%	45%	10%	59%
E <sup>[f]</sup>	1.3	n.d.	10	1	2.5	3.7	27	

[a] Data from literature for *endo-1 a* and *exo-1 a*.<sup>[1]</sup>, AaADH2: alcohol dehydrogenase from *A. annua*.<sup>[10]</sup>; SoBDH1/2: borneol-type dehydrogenases from *S. officinalis* L.<sup>[11]</sup>; PsBDH: borneol dehydrogenase from *Pseudomonas* sp. TCU-HL1.<sup>[12]</sup> [b] SoBDH3: borneol-type dehydrogenases from *S. officinalis*; SrBDH1/2: borneol-type dehydrogenases from *S. rosmarinus*, [c] %ee<sub>s</sub>: Enantiomeric excess of substrates. %ee<sub>p</sub>: Enantiomeric excess of the products. All enantiomeric excess were determined by chiral gas chromatography. [d] n.c.: no conversion, [e] n.d.: not determined, [f] Calculated according to Straathop and Jøgejan (1997).<sup>[53]</sup> [g] %ee<sub>p</sub> not calculated as we did not obtain baseline separation for the fenchone enantiomers in GC analysis.


**Scheme 2.** Preferred substrates (highlighted in black) and enantiospecificity of SrBDH1 in the kinetic resolution of bicyclic secondary alcohols.

the stereoselective reduction of pure **1 b** isomers,<sup>[21]</sup> and an engineered *Pseudomonas* dehydrogenase with high enantiospecificity towards (+)-**1 b**,<sup>[14]</sup> we decided to investigate if borneol-

type dehydrogenases were also capable of catalyzing the reduction. The reduction reaction is favored at acidic pH values, however, we used pH 5.5 as the stability of SrBDH1 was notably



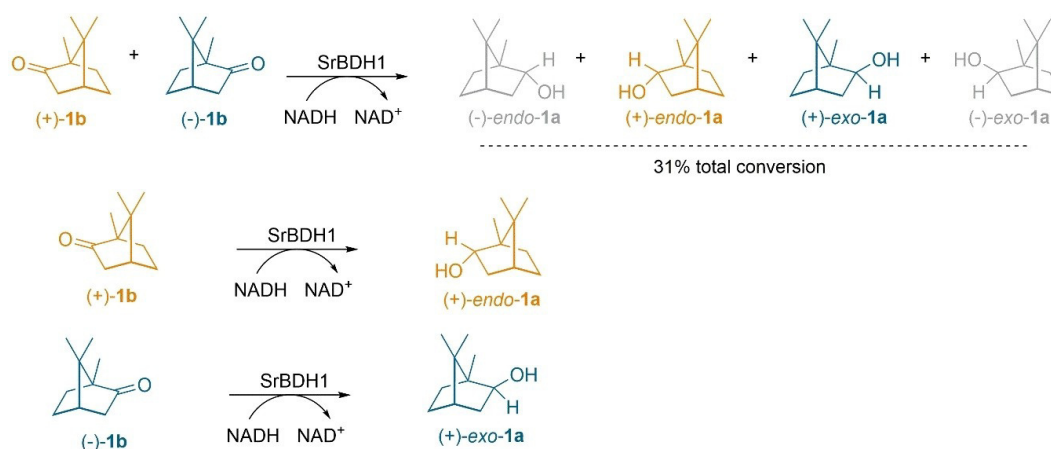
reduced at lower pH values (Figure 2). In fact, all of the enzymes showed some degree of precipitation at pH 5.0. Among borneol-type dehydrogenases, AaADH2 and SrBDH1 were the only ones that catalyzed the reduction of **1b** (Table 3). SrBDH1 exclusively produced (+)-endo-1a from pure (+)-1b and (+)-exo-1a from pure (-)-1b ( $de_p > 99\%$  for both) (Table S2). As the enzyme converts both enantiomers of **1b**, reduction of rac-1b produces a mixture of (+)-endo-1a and (+)-exo-1a in the ratio 91:9. AaADH2 showed lower diastereoselectivity and produced (+)-endo-1a with 89%  $de_p$  and (+)-exo-1b with 94%

$de_p$ , respectively (Scheme 3, Figure S5). Intriguingly, all of the enzymes, except SrBDH1, showed a higher conversion for **2b** in comparison to **1b**, while for the oxidations, we obtained better conversion for endo-1a and exo-1a compared to exo-2a. Also from the DR in the reductions, we observe endo predominance in all cases, leading us to think that endo-2a would be better oxidized than exo-2a for the studied enzymes. In the case of **3b**, the only enzyme catalyzing the reduction was AaADH2. This last result also contrasts with the capacity of catalyzing the oxidation of endo-3a observed in all the studied enzymes, and

**Table 3.** Reduction of racemic bicyclic ketones catalyzed by alcohol dehydrogenases at a substrate concentration of 5 mM.

	AaADH2 <sup>[a]</sup>	SoBDH1 <sup>[a]</sup>	SoBDH2 <sup>[a]</sup>	SoBDH3 <sup>[a]</sup>	SrBDH1 <sup>[a]</sup>	SrBDH2 <sup>[a]</sup>	PsBDH <sup>[a]</sup>
<i>rac-1b</i>	Time (h)	48	n.c. <sup>[e]</sup>	n.c.	n.c.	48	n.c.
	Specificity	(+)	n.c.	n.c.	n.c.	(+)	(+)
	%c	51%	n.c.	n.c.	n.c.	31%	22%
	%ee <sub>s</sub> <sup>[b]</sup>	10%	n.c.	n.c.	n.c.	36%	11%
	%ee <sub>p</sub> <sup>[b]</sup>	10%	n.c.	n.c.	n.c.	80%	37%
	E <sup>[c]</sup>	1.3	n.c.	n.c.	n.c.	12.4	2.4
	DR <sup>[d]</sup>	44:3:51:2	n.c.	n.c.	n.c.	91:0:9:0	n.c.
<i>rac-2b</i>	Time (h)	4	48	48	n.c.	48	48
	Specificity	(+)	(-)	(-)	n.c.	(-)	(-)
	%c	94%	2%	7%	n.c.	30%	2%
	%ee <sub>s</sub> <sup>[b]</sup>	60%	< 1%	69%	n.c.	33%	1%
	%ee <sub>p</sub> <sup>[b]</sup>	4%	19%	85%	n.c.	76%	51%
	E <sup>[c]</sup>	1.6	n.d. <sup>[f]</sup>	5.8	n.c.	10	n.d.
	DR <sup>[d]</sup>	13:41:7:39	36:41:18:4	10:85:0:6	n.c.	6:87:0:6	14:68:7:10
<i>rac-3b</i>	Time (h)	48	n.c.	n.c.	n.c.	n.c.	n.c.
	Specificity	(+)	n.c.	n.c.	n.c.	n.c.	n.c.
	%c	49%	n.c.	n.c.	n.c.	n.c.	n.c.
	%ee <sub>s</sub> <sup>[b]</sup>	9%	n.c.	n.c.	n.c.	n.c.	n.c.
	%ee <sub>p</sub> <sup>[b]</sup>	9%	n.c.	n.c.	n.c.	n.c.	n.c.
	E <sup>[c]</sup>	1.3	n.c.	n.c.	n.c.	n.c.	n.c.
	DR <sup>[d]</sup>	40:54:5:0	n.c.	n.c.	n.c.	n.c.	n.c.

[a] AaADH2: ADH from *A. annua*.<sup>[10]</sup>; SoBDH2: borneol dehydrogenase from *S. officinalis* L.<sup>[11]</sup>; SrBDH1/2: borneol-like dehydrogenases from *S. rosmarinus*; SoBDH3: borneol-like dehydrogenase from *S. officinalis* L. PsBDH: borneol dehydrogenase from *Pseudomonas* sp. TCU-HL1.<sup>[12]</sup> [b] %ee<sub>s</sub>: percentage enantiomeric excess of substrates. %ee<sub>p</sub>: percentage enantiomeric excess of the products. All enantiomeric excess were determined by chiral gas chromatography. The enantiomeric excess of the product was calculated based on the sums of the products resulting from (+)-1b and (-)-1b, (+)-2b and (-)-2b or (+)-3b and (-)-3b, respectively. For instance, the ee<sub>p</sub> for the conversion of 1 is defined as ee<sub>p</sub> = (((+)-exo-1a) + ((-)-endo-1a)) - (((-)-exo-1a) + ((+)-endo-1a)) / (((+)-exo-1a) + ((-)-endo-1a)) + (((-)-exo-1a) + ((+)-endo-1a)), [c] calculated according to Straathop and Jogejan (1997),<sup>[53]</sup> [d] DR: diastereomer ratio: (+)-endo-a : (-)-endo-a : (+)-exo-a : (-)-exo-a. [e] n.c.: no conversion. [f] n.d.: not determined;



**Scheme 3.** Reduction of racemic and pure enantiomers of **1b** by SrBDH1 at pH 5.5 with phosphite dehydrogenase as cofactor-regeneration system. Reaction with 5 mM of substrate, 20  $\mu$ M of NADH, 10 mM of phosphite and 12  $\mu$ M of phosphite dehydrogenase.

could be indicating, for instance, an inhibiting effect caused by *exo-3a*, which we did not analyze for oxidation reactions.

### Kinetic parameters of SrBDH1

The high activity of SrBDH1 stood out amongst this family of enzymes. While the poor solubility of (+)-*endo-1a* limited the determination of initial rates to concentrations up to 6 mM (three-fold of the  $K_M$ -value), the apparent  $k_{\text{obs}}$  obtained corresponds to  $0.20 \text{ s}^{-1}$  (Figure S6). This value is higher by at least two orders of magnitude compared with those of related dehydrogenases from *S. officinalis* L. ( $0.005 \text{ s}^{-1}$ ),<sup>[11]</sup> *A. annua* ( $0.006 \text{ s}^{-1}$ ),<sup>[9]</sup> *L. intermedia* ( $0.0004 \text{ s}^{-1}$ )<sup>[11]</sup> and one order of magnitude compared with tropinone reductase from *Cochlearia officinalis* CoTR ( $0.09 \text{ s}^{-1}$  towards (–)-*endo-1a*).<sup>[12]</sup> The value is comparable to the non-specific bacterial PsBDH from *Pseudomonas* sp. TCU-HL1 ( $0.75 \text{ s}^{-1}$ ).<sup>[12]</sup> The  $K_M$ -value of  $2.02 \pm 0.18 \text{ mM}$  is surprisingly high compared to those of the other plant dehydrogenases (typically  $50 \mu\text{M}$ ). At a non-saturating concentration of (+)-*endo-1a*, the  $K_M$ -value estimated for  $\text{NAD}^+$  was  $100 \pm 26 \mu\text{M}$ , which is comparable to the  $K_M$  for  $\text{NAD}^+$  determined for SoBDH2.<sup>[1]</sup>

### Structure elucidation and comparison with PsBDH of *Pseudomonas* sp. TCU-HL1

SrBDH1 crystals were obtained under two different crystallization conditions. The first one with high concentrations of NaCl and the second one using the polymer pentaerythritol propoxylate (PO/OH). The obtained crystals belong to two different space groups (Table S3) with different crystal packing. The overall architecture of SrBDH1 *in crystallo* is tetrameric (Figure S7), in agreement with SEC/MALS measurements confirming a tetrameric state in solution (Figure S8). The tetramers as obtained from both crystallization conditions are practically undistinguishable with a root mean square deviation (RMSD) of  $0.435 \text{ \AA}$  for 258 pairs of  $C\alpha$  atoms. As a member of the superfamily of short-chain dehydrogenases/reductases (SDR),<sup>[22]</sup> SrBDH1 folds into the characteristic Rossmann-like fold,<sup>[23]</sup> that harbors the cofactor  $\text{NAD}^+$ , and has an additional short  $\alpha$ -helix at the C-terminus (Figure 4, Figure S7). A closer inspection of both structures revealed remarkable differences in the cofactor binding sites. In the structure obtained from crystals under the high salt condition, we saw a defined electron density for the  $\text{NAD}^+$  cofactor in only one monomer (Figure S7). Notably, no  $\text{NAD}^+$  was added during protein purification or crystallization, therefore  $\text{NAD}^+$  was co-purified with SrBDH1. Hence, the structure represents the binary complex (SrBDH1· $\text{NAD}^+$  /high salt). Binding of  $\text{NAD}^+$  leads to a defined folded loop region from residue V197 to S203. The well-defined loop conformation might be a consequence of crystal packing, which hampers dissociation of the cofactor. The apo structure of SrBDH1 (Table S3) lacks electron density for  $\text{NAD}^+$ , and the latter loop region is not defined. The remaining apo structure is practically indistinguishable from the structure of the binary complex.

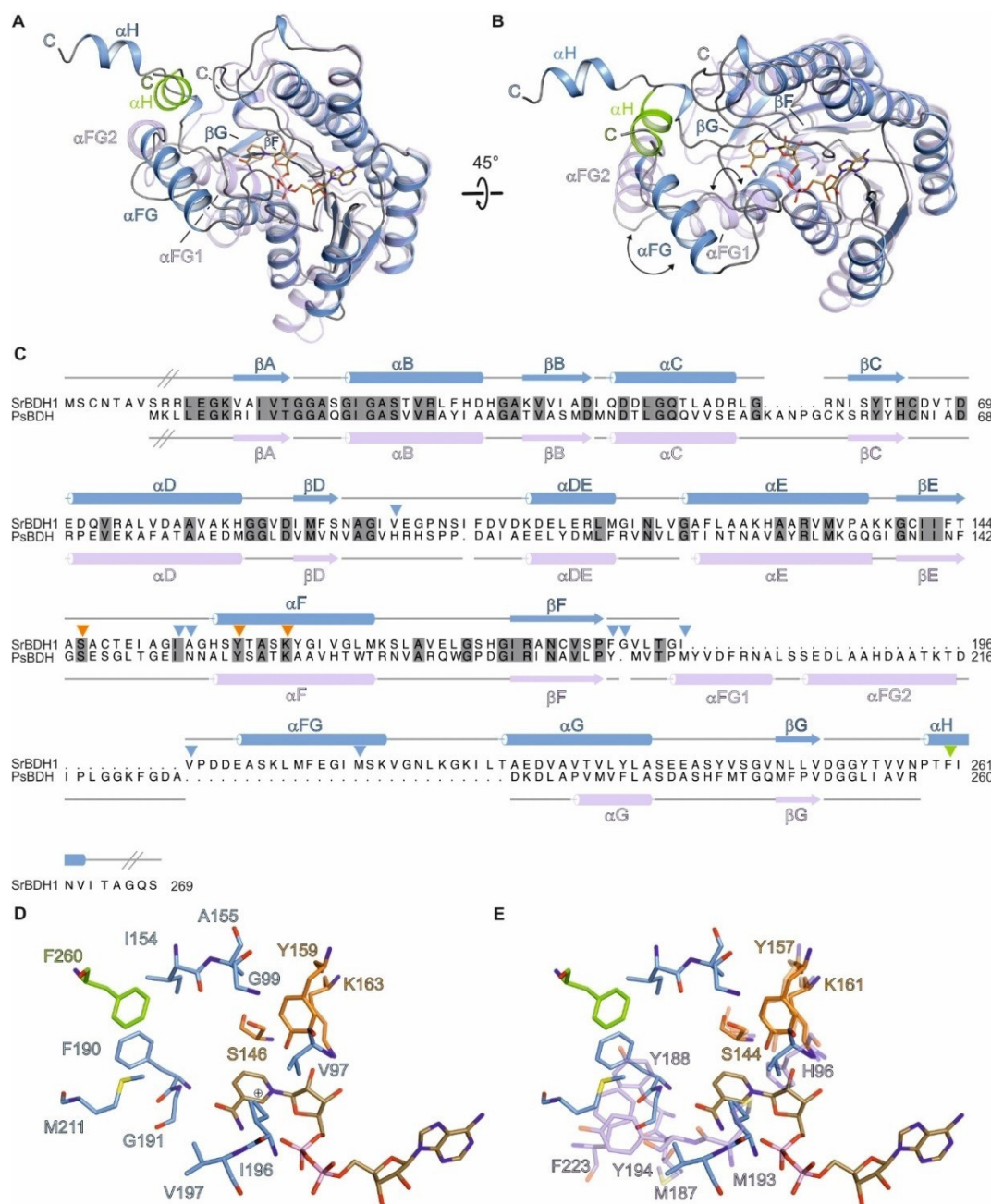
In contrast, we do observe electron density for all four cofactor binding sites in the structure obtained under the PO/OH crystallization condition (SrBDH· $\text{NAD}^+$ /PO/OH) (Figure S7). The electron density clearly shows different sigma levels at the supposed binding groove for  $\text{NAD}^+$ , indicating different occupancies of the  $\text{NAD}^+$  across the four monomers.

The active site is mainly lined by hydrophobic residues (V97, G99, I154, A155, F190, G191, I196, V197, and M211) and is composed of F260 from a symmetry-related molecule (Figure 4C and D and Figure S7A). The  $\text{NAD}^+$ , as well as the strictly conserved catalytic residues S146, Y159, and K163, complete the active site. For other studied SDRs, during oxidation, the hydroxyl group of the tyrosine abstracts a proton from the substrate. The adjacent lysine enhances the acid/base properties of the tyrosine, and the serine stabilizes and polarizes the carbonyl group of the substrate.<sup>[24]</sup> Activity loss of the enzyme after substitution of S146 and Y159 to alanine confirmed their catalytic role.

Recently, the crystal structure of the unspecific PsBDH of *Pseudomonas* sp. TCU-HL1 (PDB ID 6M5N) sharing 30% of sequence identity with SrBDH1, was solved.<sup>[25]</sup> The monomers of PsBDH and SrBDH1· $\text{NAD}^+$ /high salt superimpose with a RMSD of  $1.52 \text{ \AA}$  for 230 pairs of  $C\alpha$  atoms (Figure 4A). The largest differences between both structures are located in the last third of the amino acid sequence upstream of  $\beta$ -strand  $\beta\text{F}$  (Figure 4A and B), affecting the putative borneol binding site. In the structure of SrBDH1, the  $\beta$ -strand  $\beta\text{F}$  is connected by a single  $\alpha$ -helix  $\alpha\text{FG}$  to  $\alpha\text{G}$ , while in the structure of PsBDH we find an insertion of two  $\alpha$ -helices,  $\alpha\text{FG1}$  and  $\alpha\text{FG2}$ . Moreover, SrBDH1 contains an additional C-terminal  $\alpha$ -helix  $\alpha\text{H}$  that is absent in PsBDH (Figure 4A, B, and C). Due to the differences in fold and secondary structure content, amino acids that are flanking the active site differ between both structures (Figure 4C, D, and E). Notably, the active site of SrBDH1 is further shielded by the  $\alpha$ -helix  $\alpha\text{H}$  of a symmetry-related molecule (Figure 4). The location of SrBDH1  $\alpha\text{H}$  is comparable to the  $\alpha$ -helix  $\alpha\text{FG2}$ , which is more distant from the active site (Figure 4A and B), but from the identical monomer. Our observations are in accordance with previous publications stating that the C-terminal portion generally functions in substrate binding, therefore, the obvious structural variation in this region results in the diversity of substrate specificities.<sup>[26]</sup> A more detailed inspection of the active site revealed, that except for I154 of SrBDH1, none of the amino acids present the active site are conserved to PsBDH.

### Probing enantiospecificity by site-directed mutagenesis

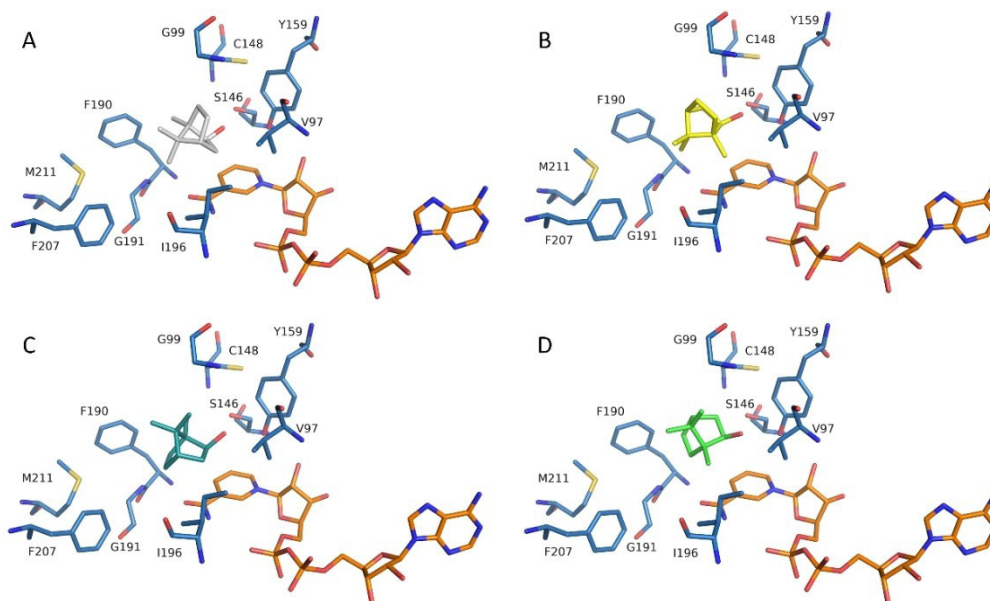
Intrigued by these findings we proceeded to elucidate the origin of specificity in BDH-type enzymes. Since soaking or co-crystallization experiments with substrates failed in yielding a crystal structure with a bound substrate, we performed docking studies using both enantiomers of *endo-1a*, *exo-1a* and **1b** to identify the most probable conformation in the active site. The bicyclic monoterpenols *endo-1a* and *exo-1a* are compact molecules that offer only small differences regarding asymme-



**Figure 4.** (A) In cartoon representation, a superposition of one monomer of SrBDH1 and PsBDH. Identical view as in Figure S7. SrBDH1 is shown in blue and PsBDH in violet. The C-terminal  $\alpha$ -helix ( $\alpha$ H) of a symmetry-related molecule, that shields the active site, is shown in green cartoon representation. The NAD<sup>+</sup> bound to SrBDH1 is shown as brown sticks. (B) View of panel (A) rotated by 45°. (C) Structure-based sequence alignment of SrBDH1 and PsBDH PDB-ID: 6M5N. The secondary structure elements are shown above the alignment for SrBDH1 and below for PsBDH with  $\alpha$ -helices depicted as cylinders and  $\beta$ -strands as arrows. Inclined lines indicated sections of SrBDH1 and PsBDH, that are not included in modelled the crystal structures. The catalytic motif is indicated by orange triangles. Amino acids lining the putative active site of SrBDH1 are indicated by blue triangles and as green triangle if derived for another SrBDH1 monomer within the tetramer. (D) Stick representation of residues lining the active site of SrBDH1 with bound NAD<sup>+</sup>. The active site is defined mainly by hydrophobic amino acid residues from two symmetry-related protein monomers colored in blue and green, respectively. (E) Stick representation of residues lining the active site of chain C with the bound NAD<sup>+</sup> and PsBDH in violet. The active site is defined mainly by hydrophobic amino acid residues from two different protein monomers. Color-coding according to panel A.

tries or polar groups that could facilitate stereodiscrimination by the enzyme. This was reflected in a large number of different poses (8–12) obtained in docking experiments (using a min RMSD of 2 Å for clustering). To identify productive binding modes, we lowered the clustering threshold to get more poses.

We considered a pose to be productive when the hydrogen in  $\alpha$  position from the hydroxyl group was located towards NAD<sup>+</sup> and the hydrogen of the hydroxyl group was in range ( $d < 4$  Å) with the catalytic residues S146 and Y159 (Figure 5).<sup>[24]</sup> The low ratio of productive binding mode is in agreement with the very



**Figure 5.** Docking results of (+)-*endo-1a* (A), (–)-*endo-1a* (manually docked, B), (+)-*exo-1a* (C), (–)-*exo-1a* (D) in the active site of SrBDH1. *Endo-1a* enantiomers seem to be located in the active site in a different orientation compared to *exo-1a* enantiomers, but the most striking difference is the methyl group that points away from the page in accepted molecules (A and B) whereas it is in the opposite direction in non-accepted molecules (B and D).

high  $K_M$  of the enzyme. Docking of *exo-1a* yielded one productive pose for each enantiomer. In the case of *endo-1a*, the preferred (+)-enantiomer led to a productive binding mode, whereas no productive pose was found for the (–)-enantiomer. In this last case, we forced (–)-*endo-1a* in a hypothetical binding mode in the crystallographic structure of the enzyme in order to study how we could enable its conversion (Figure 5B). When (+)-*endo-1a* was soaked into crystals of SrBDH1 prior to structure determination, an additional patch of electron density (data not shown) was revealed. The volume and location of this additional electron density is in good agreement with the modeled position of (+)-*endo-1a*. To further verify that the missing binding mode for (–)-*endo-1a* might be possible, the same ligands were docked in the newly structurally characterized unselective BDH from *Pseudomonas* sp. TCU-HL1 (PsBDH, PDB-ID: 6M5N).<sup>[23]</sup> In this case, a productive binding mode for (–)-*endo-1a* could be found. The PsBDH seems to have more space in the hydrophobic pocket of the active site. However, the significantly different secondary structure in the active-site pocket made it difficult to pinpoint concrete differences that could explain why productive poses for all ligands could be found in this dehydrogenase (Figure 4A and B).

The docking results highlighted two main potential differences between the positioning of the isomers. First, borneol and isoborneol seem to fit differently in the pocket. In particular, the two methyl groups in C7 assume a different position. Second, the methyl group in the chiral C1 in the binding modes of preferred (Figure 5A and C) and non-preferred enantiomers (Figure 5B and D) points in opposite directions.

The crystallographic and docking studies served as a starting point for the selection of active site residues that might have a significant effect on stereospecificity. Furthermore, two strategies were combined to create variants with decreased specificity allowing us to identify residues that are responsible for the excellent enantiodiscrimination in SrBDH1. The first strategy was to introduce residues found in the active site of non-specific BDHs that differed from ones at the same positions in SrBDH1 (Table S4). The second strategy was to use the coupled moves protocol implemented in the Rosetta framework by Ollikainen et al.<sup>[27]</sup> for the four isomers of **1a**. The Rosetta protocol was designed to change substrate specificity by redesigning the active site of a certain enzyme. Specifically, for each isomer we obtained a set of positions in which certain residues could possibly have a positive impact on the binding energy of the enzyme for the isomer. We hypothesized that positions, where the enriched residues were different for each enantiomer, are important for the specificity of SrBDH1. The coupled moves highlighted positions such as V97F/Y, G99T/P, C148 A or I196 L (Figure S9). Table 4 shows the effect of amino acid substitutions in the active site on the specificity towards *endo-1a* and *exo-1a*. Mutagenesis points out the residues V97 and G191 as determinants of enantiospecificity for *exo-1a*. Valine 97 is located in direct vicinity to the cofactor and the substrate and is likely to exert an influence on substrate recognition. The coupled moves protocol suggested multiple mutations at this position, making us think that it might be an important site for enantiospecificity. However, the decision to substitute V97 by proline was made after observing that several non-specific BDHs have a proline in the analogous position. Indeed, V97P has a reduced specificity ( $E = 15$ ), while conversions were similar to those of the wildtype enzyme. In



**Table 4.** Conversion and enantiospecificity for the kinetic resolutions of *endo-1 a* and *exo-1 a* for the point mutants tested. The reaction mix contained 1 mM substrate, 1 mM of NAD<sup>+</sup> and was incubated at 20 °C for 30 min before extraction for GC analysis.

Mutant	<i>endo-1 a</i>		<i>exo-1 a</i>	
	E <sup>[a]</sup>	[%]c <sup>[a]</sup>	E <sup>[a]</sup>	[%] c <sup>[a]</sup>
SrBDH1	> 200	50%	132	53%
V97G <sup>[b]</sup>	> 200	7%	> 200	4%
V97C <sup>[c]</sup>	> 200	50%	47	45%
V97P <sup>[c],[b]</sup>	> 200	41%	15	38%
V97F <sup>[d]</sup>	> 200	10%	14	8%
V97Y <sup>[d]</sup>	> 200	4%	> 200	1%
G99 N <sup>[c]</sup>	> 200	37%	> 200	27%
G99H <sup>[c],[e]</sup>	n.c. <sup>[f]</sup>	n.c.	n.c.	n.c.
G99D <sup>[c]</sup>	> 200	43%	> 200	38%
G99T <sup>[d]</sup>	> 200	50%	76	51%
C148A <sup>[d]</sup>	> 200	50%	> 200	51%
F190Y <sup>[d]</sup>	> 200	47%	> 200	43%
G191M <sup>[c]</sup>	> 200	48%	> 200	49%
G191S <sup>[c]</sup>	> 200	17%	> 200	14%
G191F <sup>[c]</sup>	> 200	11%	4	15%
I196L <sup>[c]</sup>	> 200	20%	> 200	14%
F207W <sup>[d]</sup>	n.c.	n.c.	n.c.	n.c.
M211V <sup>[b],[c]</sup>	> 200	50%	> 200	50%
M211L <sup>[c]</sup>	> 200	50%	> 200	50%

[a] Calculated from ee<sub>p</sub> and ee<sub>s</sub> that were determined by chiral gas chromatography, [b] Selected according to docking results, [c] Selected according to sequence alignment analysis, [d] Selected according to results of Rosetta couple moves protocol, [e] Expressed in insoluble form, [f] n.c. = no conversion detected.

homology models of the non-specific AaADH2, the equivalent proline is in a loop that strongly differs from that of SrBDH1 (Figure S10). As V97P might have consequences on the protein backbone conformation of this loop, its effect on the enantiospecificity is not easy to rationalize. Glycine 191 is positioned at the opposite side of the active site (Figure 5) and its methylene group has a distance of 4 Å to both the nicotinamide moiety of the cofactor and the substrate. Exchanging G191F was done due to the presence of this residue in several non-specific BDHs (Figure S1, Table S4) and led to a considerable reduction of activity towards *exo-1 a*. F191 could undergo  $\pi$ -stacking with F207 and lead to rearrangement of this part of the active-site pocket. However, there is no obvious explanation why the V97P and G191F substitutions increase (–)-*exo-1 a* proportional formation, but do not affect the specificity for (–)-*endo-1 a*. Substitutions in M211 by valine and leucine showed a slight increase in the specificity for *exo-1 a*, while maintaining the high specificity for *endo-1 a*. M211 is located in an  $\alpha$ -helix that has not equivalent structure in PsBDH (Figure 4), at the opposite side of the active site where the hydroxyl group of the substrate is located (Figure 5).

The mechanisms underlying the enantiospecificity of many enzymes are often characterized by well-defined steric and electronic interactions. The different sizes of the two substituents of secondary alcohols guide the stereospecificity of lipases.<sup>[28]</sup> Both (*S*)-specific and (*R*)-specific amine transaminases employ binding pockets for the accommodation of large and medium-sized prochiral ketones, which often results in outstanding specificity.<sup>[29,30]</sup> The stereoselectivity of ketoreductases in the asymmetric reduction of prochiral ketones follows the same principle.<sup>[31]</sup> Many enzyme classes bind their substrates

with a multitude of polar interactions, which allows them to discriminate between different polar groups on the substrate molecule,<sup>[32]</sup> this principle allows carbohydrate-converting enzymes a tremendous specificity toward molecules with several functional groups of similar reactivity.<sup>[32]</sup> In contrast, the stereospecificity of plant SDR for bicyclic monoterpenols *endo-1 a* and *exo-1 a* is a curious case, as their carbon skeleton is rigid and does not have any rotational degrees of freedom. The compact bornane-type structure does not show any obvious steric differences (such as a large and a small substituent) that would facilitate discrimination between both enantiomers. Possibly, *endo-1 a* and *exo-1 a* present minor steric differences in the accessibility of hydrogen in the chiral C1; in *exo-1 a* this hydrogen is on the same plane as the methyl groups in C7 whereas in *endo-1 a* it is on the less sterically-crowded plane. This minor difference might explain the preference of most enzymes for *exo-1 a* in the reverse reaction, here the hydride from NADH should access the less sterically-crowded plane resulting in *exo-1 a*. The positioning of the substrate can be deduced from the required short distances from C<sub>1</sub>-H to the cofactor, and OH-H to S146 and Y159. Other than these, there are no clearly defined interactions between the substrate and the hydrophobic active site of SrBDH1 that could explain its very high specificity towards racemic *endo-1 a* and *exo-1 a*. SrBDH1 preferentially converts (+)-*endo-1 a* and (+)-*exo-1 a* and shows scarce activity towards the (–)-enantiomers. Mutagenesis of 6 of the 8 residues of the active-site pocket (the remaining two being highly conserved) led to the identification of two variants that additionally convert (–)-*exo-1 a*, but, not (–)-*endo-1 a*. This corresponds to the observation that in docking experiments, only (–)-*endo-1 a* did not form a productive binding mode. It is

not clear whether the enantiospecificity towards *endo-1a*, the presumed natural substrate of these SDRs, has any evolutionary advantages for the plant. However, a possible benefit might be the possibility of specific oxidation of (+)-*endo-1a* to (+)-*2a* in the presence of (–)-*endo-1a*. Indeed, the essential oil of *R. officinalis* contains both (–)-*endo-1a* and (–)-*exo-1a*. In view of the very high enantiospecificity of *R. officinalis* towards borneol and the fact that its specificity was not affected by single-site mutagenesis, it is indeed striking that a few of the investigated borneol-dehydrogenases show specificity (Figure 1).

## Conclusion

A selected set of short-chain borneol-type dehydrogenases were characterized in this study in terms of substrate acceptance and specificity. All enzymes converting the substrates showed some extent of preference for (+)-*endo-1a* and (+)-*exo-1a* over their specular images, with the novel SrBDH2 presenting outstanding specificity for (+)-*endo-1a* and SrBDH1 for both of them. The kinetic resolution of *endo-1a* utilizing SrBDH1 produces optically pure (+)-*1b* and allows the isolation of the unreacted (–)-*endo-1a*. In the case of *exo-1a*, SrBDH1 yields (–)-*1b* and (–)-*exo-1a* in optically pure form. The high activity, stability and specificity of SrBDH1, make this enzyme a promising biocatalyst for the preparation of optically pure *endo-1a*, *exo-1a*, (+)-*1b* and (–)-*1b*. Therefore, enzymatic catalysis utilizing SrBDH1 could substitute the currently used extraction from plants, which is unfavorable from both the environmental and economical point of view.

The distribution of enantiospecificity enzymes observed in the phylogenetic tree indicates that either the ability to oxidize borneol evolved independently several times during evolution or that a borneol oxidizing ancestor existed and some descendants lost the affinity for this substrate. Further characterization of different alcohol dehydrogenases belonging to SDR110 C group would be necessary in order to have a more complete and correct interpretation of the evolution of this family.

The structure of SrBDH1 showed a predominantly hydrophobic catalytic pocket. A comparison with the non-specific PsBDH revealed major differences in the structure and amino acids shaping the active site pocket. Docking of the enantiomers of *endo-1a*, *exo-1a* and *1b* was performed with SrBDH1 structure. These results displayed productive binding modes for (+)-*endo-1a*, (+)-*exo-1a* and (–)-*exo-1a*, but not for (–)-*endo-1a*. This is in agreement with the directed mutagenesis study, where the specificity for (+)-*endo-1a* remained high for all mutations, whereas we could significantly reduce it for *exo-1a* with mutations at positions V97, G99 and G191. Despite the high structural similarity between *endo-1a* and *exo-1a*, SrBDH1 can discriminate between the diastereoisomers. In fact, (+)-*endo-1a*, the supposed natural substrate, shares the same backbone with (–)-*exo-1a*, still the found binding mode of (+)-*endo-1a* would be unproductive for (–)-*exo-1a*. Productive binding modes for both substrates seem to depend on different interactions making it difficult to alter the specificity with point

mutations only. This study describes a curious case of enantiospecificity and shows that this feature, in the case of unfunctionalized molecules, can be achieved in nature relying only on multiple weak interactions. Based on our results, the specificity of borneol-type SDRs appears to be more robust to point mutations compared to lipases, transaminases, ketoreductases or carbohydrate-converting enzymes. A combinatorial mutagenesis approach such as CASTing could highlight hotspots with epistatic effects, which are extremely difficult to highlight with point mutations only. Nevertheless, the identification of a highly active, stable and specific SrBDH1 in a still untapped market niche nicely proves once again the potential of biocatalytic applications.

## Experimental section

### Materials

All chemicals were bought from Sigma Aldrich (Germany), except for (+)-*endo-2a* (1S,2R,4R) (AaBlocks, USA) and used without further purification. *E. coli* strain BL21 (DE3) was used for expression.

Protein and DNA concentration were measured using a Nanodrop 2000 UV-Vis spectrophotometer (Thermo Scientific, USA) at 260 and 280 nm respectively. Absorbance at 340 and 550–800 nm was measured using an Eon plate reader (BioTek, USA). The enzymes were purified using ÄKTA pure system (GE Healthcare Life Sciences, Austria).

The genes were ordered at GeneScript (USA), codon-optimized for *E. coli* and cloned into the vector pET15b in frame with an N-terminal poly-histidine tag.

### Phylogenetic tree and alignment

Evolutionary analyses were conducted in MEGA X.<sup>[33]</sup> The evolutionary history was inferred by using the Maximum Likelihood method and Le\_Gascuel\_2008 model.<sup>[34]</sup> The tree with the highest log likelihood is shown in Figure 1. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5189)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.26% sites). Evolview v2 was used for visualization.<sup>[35]</sup>

### Protein expression and purification

Protein expression and purification of AaADH2, PsBDH, SoBDH1 and SoBDH2 were done as described before.<sup>[11]</sup> Briefly, for SrBDH1, SrBDH2 and SoBDH3, *E. coli* BL21 (DE3) chemo-competent cells were transformed with the constructs. Overnight cultures of 12.5 mL were used to inoculate flasks of 500 mL of LB media supplemented with of ampicillin (100 mg L<sup>-1</sup> final concentration). The flasks were shaken at 130 rpm at 37 °C for 3.5 h (until OD600 between 0.6–0.8 was reached). The cultures were then induced with of IPTG (1 mM final concentration) and shaken overnight at 130 rpm at 28 °C.

Cells were harvested by centrifugation (*Beckman Coulter*, USA, JA10 rotor) at 6,000 rpm at 4 °C for 20 min. The pellets were resuspended

in 20 mL of lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM DTT, 10% glycerol, 20 mM imidazole, pH 8). The resuspended cells were sonicated for 6 min, output control 7, duty cycle 70% and then centrifuged at 13,000 rpm for 45 min at 4 °C. The supernatants were recovered, filtered with 0.45 µm filters and loaded into pre-equilibrated His-Trap FF crude 5 mL columns (GE-Healthcare, Austria) for affinity chromatography purification. The loaded columns were washed with 50 mL of binding buffer (20 mM Tris-HCl, 500 mM NaCl, 10% glycerol, 30 mM imidazole, pH 8). The purified enzymes were eluted using 30 mL of elution buffer (Tris-HCl 20 mM, 500 mM NaCl, 10% glycerol, 300 mM imidazole, pH 8) and dialyzed overnight with storage buffer (Tris-HCl 20 mM, 500 mM NaCl, pH 8). If needed, the enzymes were concentrated using Amicon® Ultra 10 K centrifugal filter (Merck KGaA, Germany). The enzymes were aliquoted and stored with a final concentration of 10% glycerol at -20 °C. Following the purification, the fractions were analyzed by SDS-PAGE electrophoresis on 12% polyacrylamide gels (ExpressPlus™ PAGE Gel, Genscript, USA) followed by staining with Coomassie Brilliant Blue. Protein concentration was determined by measuring absorbance at 280 nm. The extinction coefficients for the enzymes were obtained from ExPASy ProtParam Tool (<https://web.expasy.org/protparam/>). The  $\epsilon_{280}$  for the monomers of the enzymes correspond to 9,190 M<sup>-1</sup> cm<sup>-1</sup> for SrBDH1, 10,680 M<sup>-1</sup> cm<sup>-1</sup> for SrBDH2, and 13,200 M<sup>-1</sup> cm<sup>-1</sup> for SoBDH3.

### Size exclusion chromatography

Size exclusion chromatography was carried using a HiLoad™ 16/60 200 Superdex™ column (GE Healthcare). The column was equilibrated overnight with buffer Tris-HCl 100 mM, 500 mM of NaCl, pH 8. 500 µL samples were injected and then eluted at a flow of 1 mL min<sup>-1</sup> in 1.5 column volumes. A calibration curve was elaborated using Gel Filtration Cal Kit High Molecular Weight (GE Healthcare). The molecular mass standards used were Ovalbumin (44 kDa), Conalbumin (75 kDa) Aldolase (158 kDa) and Ferritin (440 kDa).

### Determination of conversion, enantiomeric excess and E value for the oxidative and reductive reactions

For the oxidations, a reaction volume of 1 mL containing the alcohol dehydrogenase (15 µM of the monomer of AaADH2, PsBDH, SoBDH1, SoBDH2, SoBDH3, SrBDH1, SrBDH2), 1 mM of NAD<sup>+</sup>, 1 mM of substrate (*endo-1 a*, *exo-1 a*, *exo-2 a* or *endo-3 a*) in buffer Tris-HCl 100 mM, 500 mM of NaCl, pH 8 at 20 °C, 600 rpm was used. Samples of 200 µL were taken at different time points for gas chromatography analysis. All the reactions were done in duplicate.

A colorimetric screening based in phosphate detection<sup>[36]</sup> was used to identify activity for the reductive reaction of the enzymes. The reactions were prepared in a 96 deep well plate by triplicate. The reaction mix of 1 mL consisted in 15 µM (monomer) of the alcohol dehydrogenase determined by the absorption at 280 nm (AaADH2, PsBDH, SoBDH1, SoBDH2, SoBDH3, SrBDH1 or SrBDH2), 20 µM of NADH, 10 mM of sodium phosphite, 12 µM of phosphite dehydrogenase and 5 mM of substrate (**1 b**, **2 b** or **3 b**) in citrate buffer (50 mM, 500 mM NaCl, pH 5.5). The reactions were incubated at 20 °C, 200 rpm for 48 h. Samples of 200 µL were taken at 0, 24 and 48 h and frozen with liquid nitrogen. For phosphate measurement, a mix of 200 µL of molybdate reagent ((CH<sub>3</sub>CO<sub>2</sub>)<sub>2</sub>Zn 100 mM, ammonium molybdate ((NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>4</sub>·4H<sub>2</sub>O) 10 mM, pH 5 adjusted with HCl), 50 µL of ascorbic acid solution (L-(+)- ascorbic acid 10%, pH 5 adjusted with NaOH) and 20 µL of sample was put together in a 96 well plate. After 30 min at 37 °C, absorbance in the range between 550–800 nm was measured. A calibration curve with known phosphate concentrations from 0 to 10 mM (triplicate) was

used for interpretation of the results. The reactions that after 48 h showed some level of conversion were extracted and analyzed by GC-FID.

### Determination of kinetic parameters and specific activities

Kinetic parameters were determined measuring NADH formation at 340 nm. Reaction mixes were prepared in 96 well plates for UV measurement by triplicate. For (+)-*endo-1 a* kinetic study: 0.5 µM of SrBDH1 (tetramer), 1 mM NAD<sup>+</sup>, (+)-*endo-1 a* in a range between 0.5 and 6 mM, 5% DMSO, buffer Tris-HCl 100 mM pH 9. For NAD<sup>+</sup> kinetic study: 0.5 µM of SrBDH1, NAD<sup>+</sup> in a range between 0.05 and 2 mM, 5 mM of (+)-*endo-1 a*, 5% DMSO, buffer Tris-HCl 100 mM pH 9. All the reactions were started adding NAD<sup>+</sup> to the reaction mix. Absorbance at 340 nm was measured every 15 s for 30 min. The linear range of the curves was used to calculate the initial rates. Origin 2019b (OriginLab Corporation, USA) was used for the nonlinear fitting using Michaelis-Menten model to obtain the kinetic parameters.

Specific activities were obtained in a similar way under the following conditions: 20 µM of the alcohol dehydrogenase (monomer) (AaADH2, PsBDH, SoBDH2, SoBDH3, SrBDH1 or SrBDH2), 2 mM NAD<sup>+</sup>, 2 mM of substrate, 1% DMSO in Tris-HCl 100 mM pH 9.

### Optimum pH determination for SrBDH1

SrBDH1 was incubated for 30 min on the following buffers: citrate buffer 100 mM for pH 4.5, 5 and 5.5; potassium phosphate buffer 100 mM for pH 6, 6.5 and 7; Tris-HCl buffer 100 mM for pH 7.5, 8, 8.5 and 9; carbonate-bicarbonate buffer 100 mM for pH 9.5, 10 and 10.5. After that time, 5 µL of the incubated enzyme were added to a mix consisting in 4 mM of (+)-*endo-1 a*, 5% DMSO, 5 mM NAD<sup>+</sup>, 0.15 µM of SrBDH1 in buffer Tris-HCl 100 mM pH 9 in a final volume of 200 µL. Absorbance at 340 nm was measured every 15 seconds for 45 min. The linear range of the curves was used to calculate the initial rates of reaction. All the reactions were done by triplicate.

### Gas Chromatography-flame ionization detector (GC-FID) analysis

GC-FID analysis was carried using a Shimadzu QP2010 SE GC-FID system. All extractions were performed using 400 µL of DCM and 200 µL of sample at 0, 24 and 48 h. After mixing and discarding the inorganic phase, the samples were dried using Na<sub>2</sub>SO<sub>4</sub>, centrifuged and transferred to 1.5 mL vials with 200 µL inserts. For **1 a**, **1 b**, **exo-2 a** and **2 b**, an Hydrodex β-6TDM chiral column (Macharey-Nagel) (25 m, 0.25 mm I.D., 0.25 µm film thickness) was used with the following program: 60 °C for 8 min, a linear increase of 2 °C/min to 150 °C, a linear increase of 40 °C/min to 200 °C, 200 °C for 2 min. For *endo-3 a* and **3 b**, an Hydrodex β-6TBDAc chiral column (Macharey-Nagel) (50 m, 0.25 mm I.D., 0.25 µm film thickness) was used with the following program: 50 °C for 15 min, a linear increase of 1 °C/min to 110 °C, a linear increase of 20 °C/min to 220 °C, 220 °C for 1 min.

### Protein expression and protein purification for crystallization experiments

*E. coli* BL21-RIL was transformed with pET15a vector containing SrBDH1 fused to an N-terminal hexa-histidine-tag. Protein induction was carried in auto-induction media at 37 °C for 7 h and subsequently cooled down to 16 °C.<sup>[37]</sup> Cells were grown over night and harvested by centrifugation (10 min, 7,000 rpm at 4 °C). The pellets were resuspended with buffer A (20 mM Tris/HCl pH 8.0,

500 mM NaCl). Cells were lysed by homogenization at 4 °C for 7 min after addition of 0.5 mg l<sup>-1</sup> DNase and the lysate was cleared by centrifugation (30 min, 21'500 rpm at 4 °C). Ni<sup>2+</sup>-NTA beads (cv 1 ml, GE Healthcare) were equilibrated with buffer A. SrBDH1 was loaded on the column and washed with 15 cv of buffer A. SrBDH1 was eluted using a linear gradient with increasing imidazole concentration up to 300 mM. Size exclusion chromatography (SEC) was performed with a HighLoad Superdex S200 16/60 column (GE Healthcare), equilibrated with buffer B (20 mM Tris/HCl, pH 8.0, 125 mM NaCl). Pooled protein fractions were concentrated with Amicon-Ultra-15 (Merck KGaA) to 27.3 mg ml<sup>-1</sup> as measured by the absorbance at 280 nm.

### Crystallization

Crystals were obtained by the sitting-drop vapor-diffusion method at 18 °C with a reservoir solution composed of 0.1 M Bis-Tris/HCl pH 5.5 to pH 7.2, and NaCl ranging from 2.7 M to 3.2 M. Crystals were cryo-protected with 25% (v/v) glycerol supplemented to the reservoir resolution and subsequently flash-cooled in liquid nitrogen. A second crystallization condition was obtained with a reservoir solution composed of 0.1 M HEPES/NaOH pH 7.0 to pH 7.8, 5/4 pentaerythritol propoxylate (PO/OH) 25% (v/v), 30%, 35%, 0.1 M and 0.2 M KCl. Crystals were cryo-protected with 15% (v/v) glycerol supplemented to the reservoir resolution and subsequently flash-cooled in liquid nitrogen.

### Soaking and co-crystallization experiments

Crystals were soaked for 70 min in (+)-borneol (80 mM from 400 mM stock in 100% DMSO) containing cryo-protectant and subsequently flash-cooled in liquid nitrogen. For co-crystallization 20 mM (+)-borneol (from 100 mM stock in 100% ethanol) were added to the protein solution and incubated for 1 h on ice prior crystallization. Crystals were cryo-protected as above and subsequently flash-cooled in liquid nitrogen.

### Diffraction data collection, structure determination and refinement

Synchrotron diffraction data were collected at the beamline 14.1 and 14.2 of the MX Joint Berlin laboratory at BESSY (Berlin, Germany). X-ray data collection was performed at 100 K. Diffraction data were processed with XDS<sup>[38]</sup> (Table S3). The structure for the SrBDH1 apo was solved via molecular replacement in PHASER<sup>[39]</sup> by using the structure of the ternary-secoisolariciresinol dehydrogenase from *Podophyllum petatum* (PDB ID 2bgm<sup>[40]</sup>) as search model. Crystals of SrBDH1 apo belong to the space group *P*<sub>4</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub>, with two molecules in the asymmetric unit. Model building and water picking were performed with COOT.<sup>[41]</sup> The structure was initially refined by applying a simulated annealing protocol and in later refinement cycles by maximum-likelihood restrained refinement using PHENIX.refine.<sup>[42,43]</sup> The crystals of SrBDH1 NAD<sup>+</sup> crystallized in space group *P*<sub>6</sub>, and the structure was solved by molecular replacement with the SrBDH1 structure, computed from the crystals of the high salt condition, as search model. Model quality was evaluated with MolProbity<sup>[44]</sup> and the JCSG validation server.<sup>[45]</sup> Figures were prepared using PyMOL (Schrödinger, Inc). Secondary structure elements were assigned with DSSP,<sup>[46]</sup> and ALSRIPT<sup>[47]</sup> was used for secondary structure-based sequence alignments. Structure factor amplitudes and coordinates have been deposited in the ProteinDataBank.

### Multi-angle light scattering (MALS)

MALS experiment was performed at 18 °C. SrBDH1 was loaded onto a Superdex 200 increase 10/300 column (GE Healthcare) that was coupled to a miniDAWN TREOS three-angle light scattering detector (Wyatt Technology) in combination with a RefractoMax520 refractive index detector. For calculation of the molecular mass, protein concentrations were determined from the differential refractive index with a specific refractive index increment (*dn/dc*) of 0.185 ml g<sup>-1</sup>. Data were analyzed with the ASTRA 6.1.4.25 software (Wyatt Technology).

### Docking studies

(+)-*endo-1a*, (-)-*endo-1a*, (+)-*exo-1a*, (-)-*exo-1a*, (+)-*1b* and (-)-*1b* were energy minimized using a MM2-force field for molecular docking into the active site of a monomeric representation of SrBDH1 and PsBDH (PDB-ID: 6M5N). The docking was performed using the AutoDockVina program environment of YASARA Structure.<sup>[48,49]</sup> All 6 substrate structures were docked into a simulation cell (X size = 16 Å, Y size = 16 Å, Z size = 16 Å; angles:  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 90^\circ$ ) extended 1 Å around the residues I96, I154, G195 and I210 in SrBDH1 and the corresponding residues V95, I152, M193 and A208 in PsBDH. For each substrate, 999 docking runs were performed with atoms and bonds of the corresponding substrates set as flexible. Docking of each substrate resulted in one or more clusters using a rmsd cutoff for clustering of 2.0 Å.

### Coupled moves

The SrBDH1 structure was pre-processed by running the Relax protocol with (+)-*endo-1a* in the active site to minimize artefacts in the following protocol. The ligand rotamer library was generated by the free online tool Frog2<sup>[50]</sup> with default options. The coupled moves protocol was run according to published setting<sup>[27]</sup> (command line) as reported in the Supplementary Information.

The resulting output sequences were filtered for redundancy and analyzed using the script analyze\_coupled\_moves.py made available by the Kortemme lab in github.com<sup>[51]</sup> which allows highlighting of the top 10 mutations enriched in the ligand of interest compared to a known accepted ligand and furthermore it allows the visualization of results via WebLogo.<sup>[52]</sup> In all cases the Rosetta version 3.11 was used.

### Point Mutations

The QuikChange method (Agilent Technologies) was used for site-directed mutagenesis. For the complete list of the primers used, please refer to the Supplementary Table S5). After PCR for the insertion of the mutants in pET15b\_SrBDH1 plasmid, 10 U *DpnI* were added to the mix and incubated at 37 °C for 1 h. Chemo-competent *E. coli* Top 10 cells were transformed using 3  $\mu$ L of the mix. The mutants were confirmed by sequencing and retransformed in *E. coli* BL21 (DE3).

Cell-free extract of SrBDH1 mutants was obtained as described before for the wild type. For biotransformations, 1 mM of *endo-1a* or *exo-1a*, 1 mM of NAD<sup>+</sup> and 975  $\mu$ L of cell-free extract were used, for a total reaction volume of 1 mL. Samples of the reaction after 30 min and 24 h were extracted for chiral GC-analysis.



## Author Contributions

R.K., B.L., A.C. and I.D. conceived the project and main conceptual ideas. A.C., N.D., M.P.P. and C.P.O.H. carried out experiments, analyzed results and prepared figures for the paper. E.C. performed molecular modeling. E.C., L.P.P., M.H. and V.S. participated in the analysis of the results. All authors discussed the results and participated in the writing of the manuscript.

## Funding Sources

The authors acknowledge financial support by the German Federal Ministry of Education and Research (BMBF) for the project CbP-camphor based polymers within the bio-economy international program (grant No. 031B050B). R.K. and A.C. also would like to thank the Austrian Science Funds (FWF, P31001-B29) for financial support. C.P.H. is supported by the Hanns Seidel Foundation.

## Abbreviations

SDR	Short-chain dehydrogenase-reductase
SrBDH1	Borneol-type dehydrogenase 1 from <i>Salvia Rosmarinus</i>
SrBDH2	Borneol-type dehydrogenase 2 from <i>Salvia Rosmarinus</i>
SoBDH3	Borneol-type dehydrogenase 3 from <i>Salvia officinalis</i>
SoBDH1	Borneol-type dehydrogenase 1 from <i>Salvia officinalis</i>
SoBDH2	Borneol-type dehydrogenase 2 from <i>Salvia officinalis</i>
PsBDH	Borneol dehydrogenase from <i>Pseudomonas</i> sp. Strain TCU-HL1
AaADH2	Alcohol dehydrogenase 2 from <i>Artemisia annua</i>
CCR2	CC chemokine receptor 2
CCL2	CC chemokine ligand 2
CCR5	CC chemokine receptor 5
TLC	thin-layer chromatography

## Acknowledgements

We accessed beamlines of the BESSY II (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung II) storage ring (Berlin, Germany) via the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Planck-Gesellschaft, the Leibniz-Institut für Molekulare Pharmakologie and Charité – Universitätsmedizin Berlin. Foundation. We are grateful to Y. Huang for support with the SEC/MALS experiments and to C. Langner for help in protein production. We would also like to thank Professor Marco Fraaije (University of Groningen, NL) for kindly providing us with the construct we used for expression of phosphite dehydrogenase enzyme and Leen Assil-Companioni (TU Graz) for helpful corrections to the manuscript.

## Conflict of Interest

There are no conflicts to declare.

**Keywords:** Enantioselectivity · *Salvia rosmarinus* · Terpenoids · Borneol · Short-chain dehydrogenase-reductase (SDR)

- [1] I. Drienovská, D. Kolanovi, A. Châniq, V. Sieber, M. Hofer, R. Kourist, *Phytochemistry* **2020**, *172*, 1–7.
- [2] P. S. Karunanithi, P. Zerbe, *Front. Plant Sci.* **2019**, *10*, 1–23.
- [3] F. Zhou, E. Pichersky, *Curr. Opin. Plant Biol.* **2020**, *55*, 1–10.
- [4] C. Cheng, X. W. Liu, F. F. Du, M. J. Li, F. Xu, F. Q. Wang, Y. Liu, C. Li, Y. Sun, *Acta Pharmacol. Sin.* **2013**, *34*, 1337–1348.
- [5] R. de Cássia da Silveira e Sá, T. C. Lima, F. R. da Nóbrega, A. E. M. de Brito, D. P. de Sousa, *Int. J. Mol. Sci.* **2017**, *18*, DOI 10.3390/ijms18122392.
- [6] Y. K. Chen, S.-J. Jeon, P. J. Walsh, W. A. Nugent, *Org. Synth.* **2005**, *82*, 87.
- [7] J. D. White, K. F. Wardrop, D. J. Sundermann, *Org. Synth.* **2002**, *79*, 130.
- [8] D. A. Whittington, M. L. Wise, M. Urbansky, R. M. Coates, R. B. Croteau, D. W. Christianson, *Proc. Nat. Acad. Sci.* **2002**, *99*, 15375–15380.
- [9] N. Tian, Y. Tang, S. Xiong, D. Tian, Y. Chen, D. Wu, Z. Liu, S. Liu, *Ind. Crops Prod.* **2015**, *77*, 190–195.
- [10] D. R. Polichuk, Y. Zhang, D. W. Reed, J. F. Schmidt, P. S. Covello, *Phytochemistry* **2010**, *71*, 1264–1269.
- [11] L. S. Sarker, M. Galata, Z. A. Demissie, S. S. Mahmoud, *Arch. Biochem. Biophys.* **2012**, *528*, 163–170.
- [12] H. Tsang, J.-L. Huang, Y.-H. Lin, K. Huang, P. Lu, G. Lin, A. A. Khine, A. Hu, H.-P. Chen, *Appl. Environ. Microbiol.* **2016**, *82*, 6378–6385.
- [13] D. Ponomarev, H. Mettee, *Chem. Educ. J.* **2016**, *18*, 2–5.
- [14] M. Hofer, J. Diener, B. Begander, R. Kourist, V. Sieber, *Submitt. Manuscr.* n.d.
- [15] T.-J. Ho, C.-C. Hung, T.-L. Shih, L.-M. Yiin, H.-P. Chen, *J. Food Drug Anal.* **2018**, *26*, 348–352.
- [16] R. Croteau, C. Lee Hooper, M. Felton, *Arch. Biochem. Biophys.* **1978**, *188*, 182–193.
- [17] P. Satyal, T. Jones, E. Lopez, R. McFeeters, N. Ali, I. Mansi, A. Al-kaf, W. Setzer, *Food* **2017**, *6*, 20.
- [18] E. J. Carpenter, N. Matasci, S. Ayyampalayam, S. Wu, J. Sun, J. Yu, F. R. Jimenez Vieira, C. Bowler, R. G. Dorrell, M. A. Gitzendanner, L. Li, W. Du, K. Ullrich, N. J. Wickett, T. J. Barkmann, M. S. Barker, J. H. Leebens-Mack, G. K. S. Wong, *Gigascience* **2019**, *8*, 1–7.
- [19] R. Ladenstein, J. O. Winberg, J. Benach, *Cell. Mol. Life Sci.* **2008**, *65*, 3918–3935.
- [20] Y. Kallberg, U. Oppermann, B. Persson, *FEBS J.* **2010**, *277*, 2375–2386.
- [21] N. Reinhardt, J. Fischer, R. Coppi, E. Blum, W. Brandt, B. Dräger, *Bioorg. Chem.* **2014**, *53*, 37–49.
- [22] H. Jörnvall, M. Krook, B. Persson, S. Atrian, R. González-Duarte, J. Jeffery, D. Ghosh, *Biochemistry* **1995**, *34*, 6003–6013.
- [23] M. G. Rossman, A. Liljas, C. I. Brändén, L. J. Banaszak, *Enzymes* **1975**, *11*, 61–102.
- [24] K. L. Kavanagh, H. Jörnvall, B. Persson, U. Oppermann, *Cell. Mol. Life Sci.* **2008**, *65*, 3895–3906.
- [25] A. A. Khine, H. P. Chen, K. F. Huang, T. P. Ko, *Acta Crystallogr. Sect. F* **2020**, *76*, 309–313.
- [26] W. L. Duax, D. Ghosh, V. Pletnev, *Vitam. Horm.* **2000**, *58*, 121–148.
- [27] N. Ollikainen, R. M. de Jong, T. Kortemme, *PLoS Comput. Biol.* **2015**, *11*, 1–22.
- [28] R. J. Kazlauskas, A. N. E. Weissfloh, A. T. Rappaport, L. A. Cuccia, *J. Org. Chem.* **1991**, *56*, 2656–2665.
- [29] A. Telzerow, J. Paris, M. Håkansson, J. González-Sabín, N. Ríos-Lombardía, M. Schürmann, H. Gröger, F. Moris, R. Kourist, H. Schwab, K. Steiner, *ACS Catal.* **2019**, *9*, 1140–1148.
- [30] I. V. Pavlidis, M. S. Weiß, M. Genz, P. Spurr, S. P. Hanlon, B. Wirz, H. Iding, U. T. Bornscheuer, *Nat. Chem.* **2016**, *8*, 1076–1082.
- [31] D. Zhu, L. Hua, *Pure Appl. Chem.* **2010**, *82*, 117–128.
- [32] P. Wildberger, L. Brecker, B. Nidetzky, *Chem. Commun.* **2014**, *50*, 436–438.
- [33] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
- [34] S. Q. Le, O. Gascuel, *Mol. Biol. Evol.* **2008**, *25*, 1307–1320.
- [35] Z. He, H. Zhang, S. Gao, M. J. Lercher, W. Chen, S. Hu, *Nucleic Acids Res.* **2016**, *44*, 236–241.
- [36] H. M. Dudek, P. Popken, E. Van Bloois, W. A. Duetz, M. W. Fraaije, *J. Biomol. Screening* **2013**, *18*, 678–687.
- [37] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207–234.
- [38] W. Kabsch, *Acta Crystallogr. Sect. D* **2010**, *66*, 125–132.

- [39] 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.
- [40] B. Youn, S. G. A. Moinuddin, L. B. Davin, N. G. Lewis, C. Kang, *J. Biol. Chem.* **2005**, *280*, 12917–12926.
- [41] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501.
- [42] P. D. Adams, P. V. Afonine, G. Bunkóczi, 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, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. Sect. D* **2010**, *66*, 213–221.
- [43] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Crystallogr. Sect. D* **2012**, *68*, 352–367.
- [44] C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, D. C. Richardson, *Protein Sci.* **2018**, *27*, 293–315.
- [45] H. Yang, V. Guranovic, S. Dutta, Z. Feng, H. M. Berman, J. D. Westbrook, *Acta Crystallogr. Sect. D* **2004**, *60*, 1833–1839.
- [46] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577–2637.
- [47] G. J. Barton, *Protein Eng. Des. Sel.* **1993**, *6*, 37–40.
- [48] A. J. Trott, O. Olson, *J. Comput. Chem.* **2019**, *31*, 455–461.
- [49] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- [50] M. A. Miteva, F. Guyon, P. Tufféry, *Nucleic Acids Res.* **2010**, *38*, 622–627.
- [51] Kortemme-Lab, **2019**.
- [52] G. Crooks, G. Hon, J. Chandonia, S. Brenner, *Genome Res.* **2004**, *14*, 1188–1190.
- [53] A. J. J. Straathop, J. A. Jogejan, *Enzyme Microb. Technol.* **1997**, *0229*, 559–571.

---

Manuscript received: January 23, 2021

Revised manuscript received: February 7, 2021

Accepted manuscript online: February 12, 2021

Version of record online: March 10, 2021

# ChemCatChem

Supporting Information

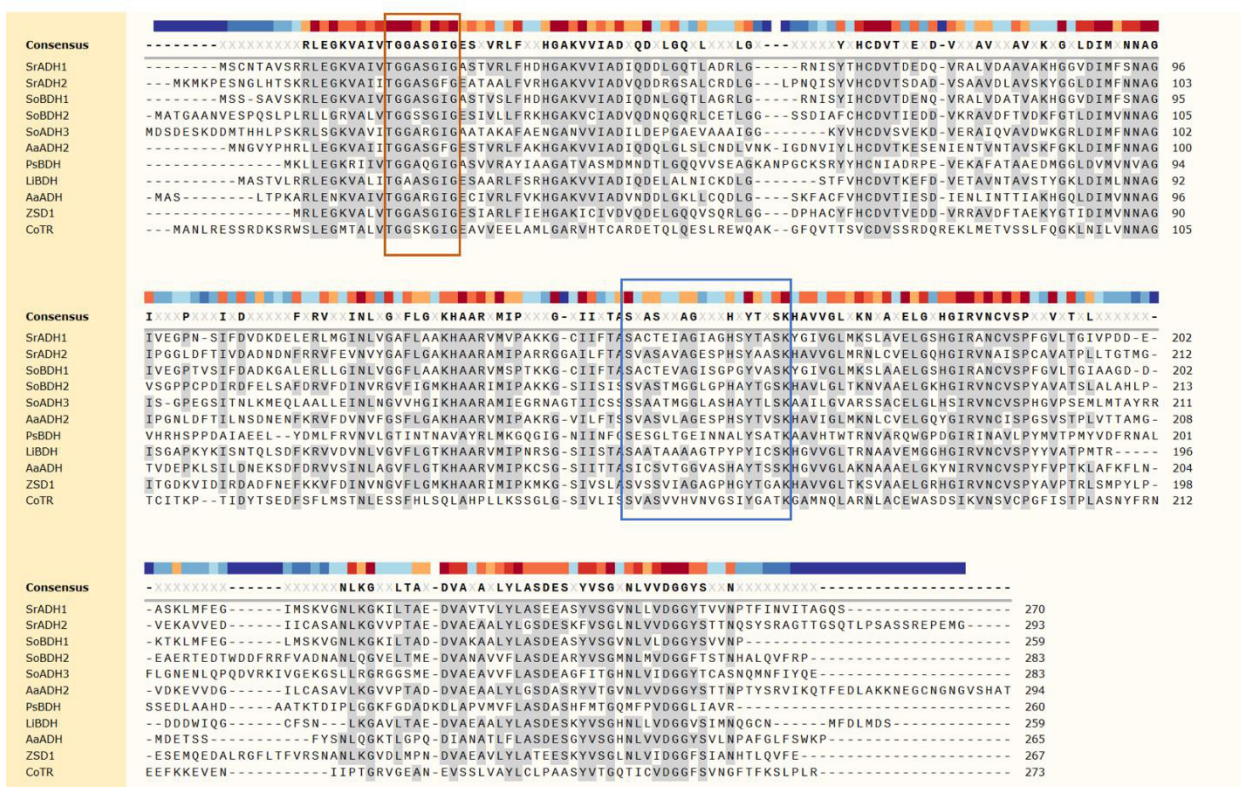
## **A Structural View on the Stereospecificity of Plant Borneol-Type Dehydrogenases**

Andrea M. Chánique<sup>+</sup>, Nicole Dimos<sup>+</sup>, Ivana Drienovská, Elia Calderini, Mónica P. Pantín, Carl P. O. Helmer, Michael Hofer, Volker Sieber, Loreto P. Parra, Bernhard Loll,\* and Robert Kourist\*

## Index

1. Supplementary Figures.....	3
Figure S1.....	3
Figure S2.....	4
Figure S3.....	5
Figure S4.....	6
Figure S5.....	7
Figure S6:.....	7
Figure S7.....	8
Figure S8.....	9
Figure S9.....	9
Figure S10.....	10
2. Supplementary tables .....	11
Table S1.....	11
Table S2.....	12
Table S3.....	13
Table S4.....	14
Table S5.....	15
3. Nucleotide and amino acid sequences of the enzymes used in this study .....	16
4. Complete name and accession numbers of SDR used in phylogenetic tree.....	19
5. Commands used for couple moves protocol .....	21
References .....	22

# 1. Supplementary Figures



**Figure S1.** Alignment of new amino acid sequences from *Salvia officinalis*, *Salvia rosmarinus* and enzymes previously described to catalyze the oxidation of borneol. The sequences were aligned using Mega X with parameters defined by default and visualized in SnapGene 5.1. Residues matching the consensus sequence are highlighted in grey. Degree of conservation is indicated on top, with blue representing the less conserved residues and red the most conserved ones. “Classical” SDR motif for cofactor binding (TGxxx[AG]xG) is highlighted in an orange square. The catalytic triad (S, Y, K), including the active site “classical” SDR motif (YxxxK) is highlighted in a blue square.



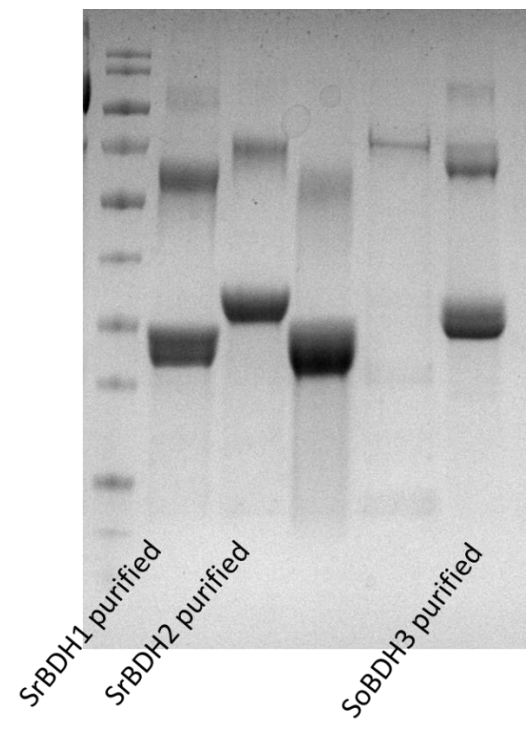
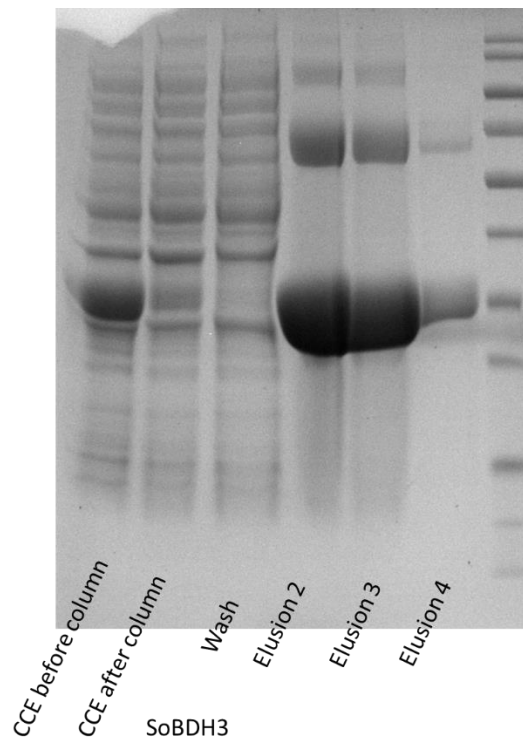
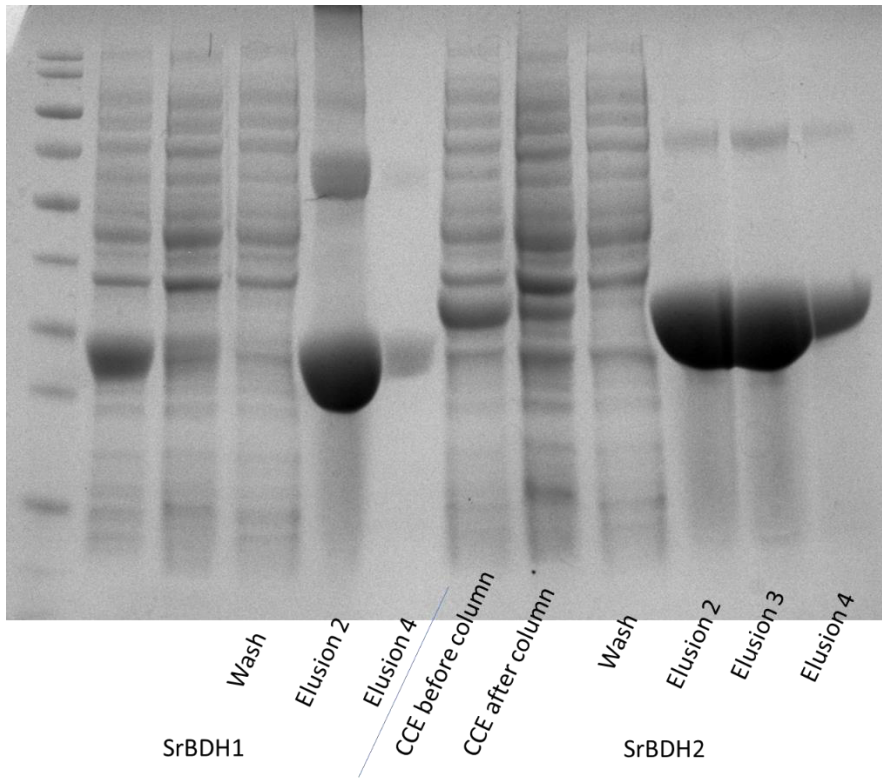
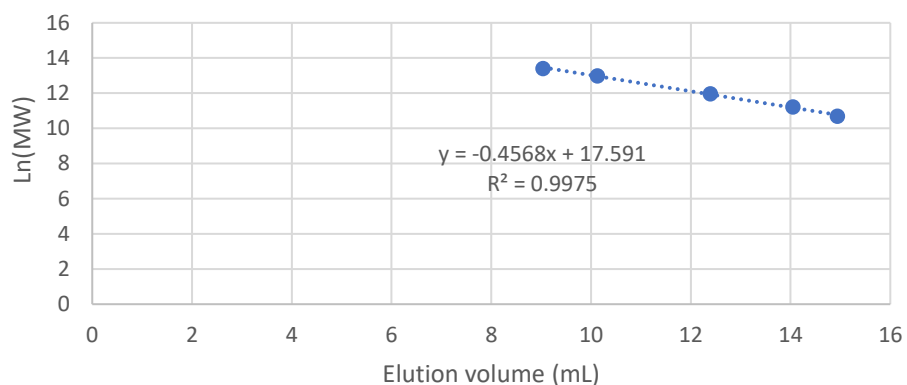


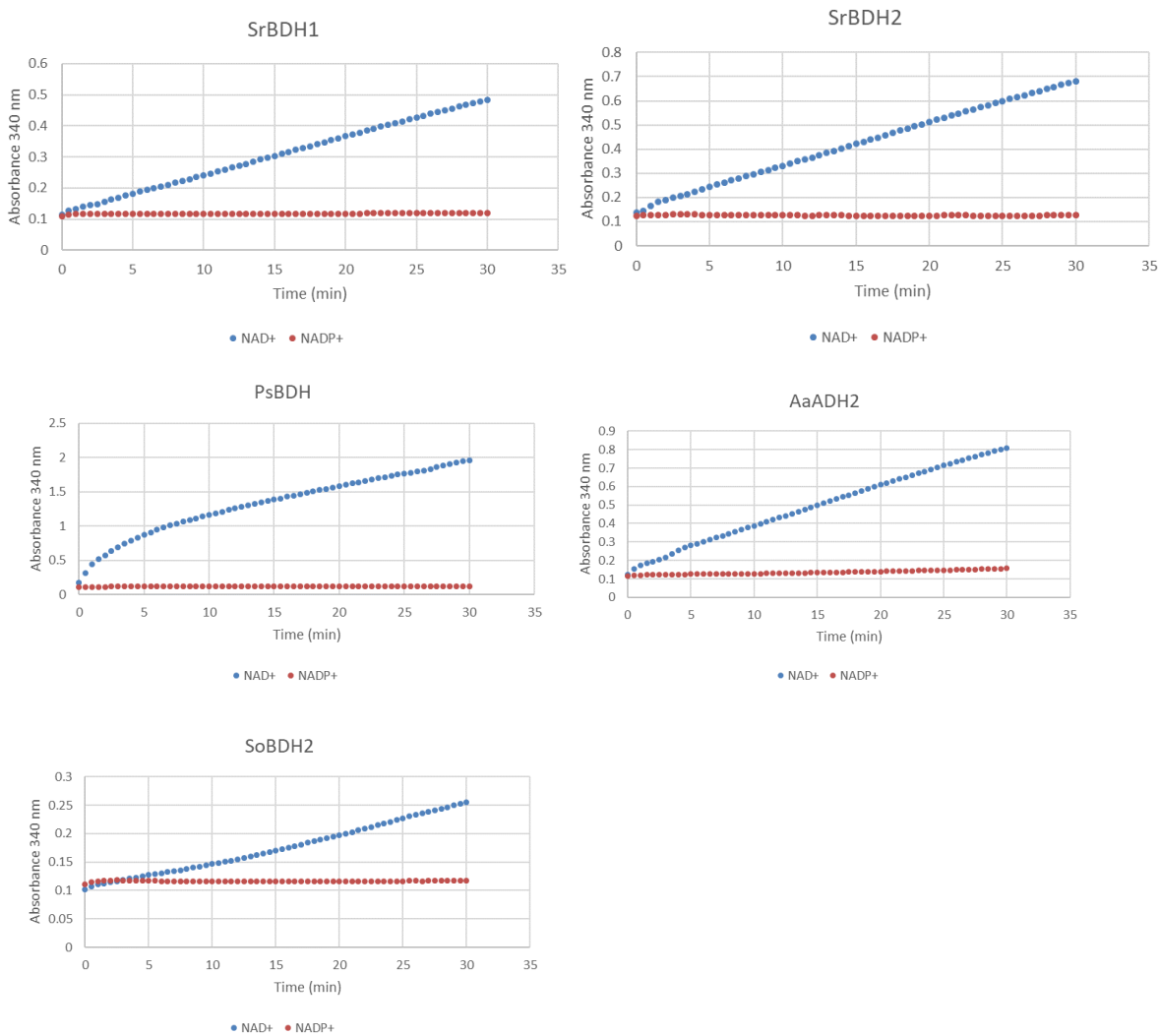
Figure S2. SDS gels for the purification of the three new BDH. CFE: cell free extract.

Calibration curve SEC



	Retention time (min)	Molecular weight according to calibration curve (Da)	Theoretical molecular weight of the monomer (Da)	Oligomeric state
<b>SrBDH1</b>	12.98242	115917.1	30247.53	3.8
<b>SrBDH2</b>	12.392845	151743.7	32395.52	4.7
<b>SrBDH3</b>	12.557154	140771.3	31812.02	4.4

**Figure S3.** Graph showing calibration curve for size exclusion chromatography and table with the oligomeric state of SrBDH1, SrBDH2 and SoBDH3.



**Figure S4.** Graphs showing absorbance at 340 nm for NAD<sup>+</sup> and NADP<sup>+</sup> for borneol dehydrogenase-like enzymes. 2  $\mu$ M of AaADH2, PsBDH, SoBDH2 and SrBDH1 or 20  $\mu$ M of SrBDH2, 2 mM NAD(P)<sup>+</sup>, 2 mM *exo-1a*, 1% DMSO, buffer Tris-HCl 100 mM pH 9.



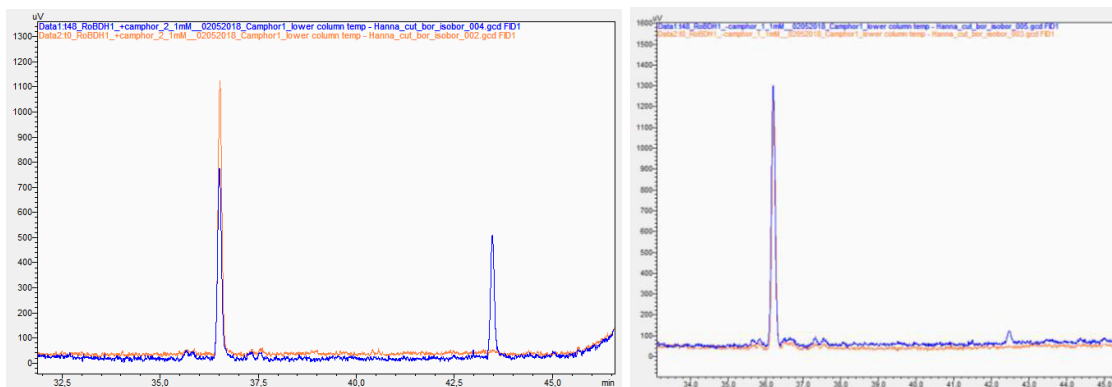


Figure S5. Chromatograms for conversion of pure enantiomers (+)-**1b** (left) and (-)-**1b** (right) catalyzed by SrBDH1. Orange for time 0 and blue for 48 hours.

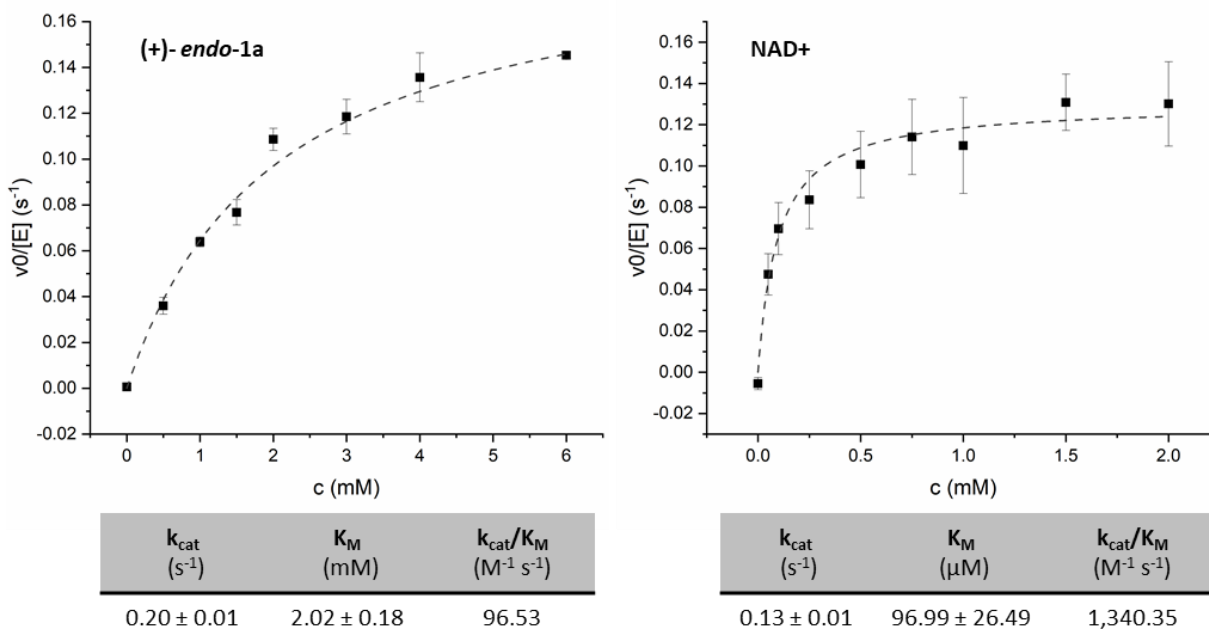
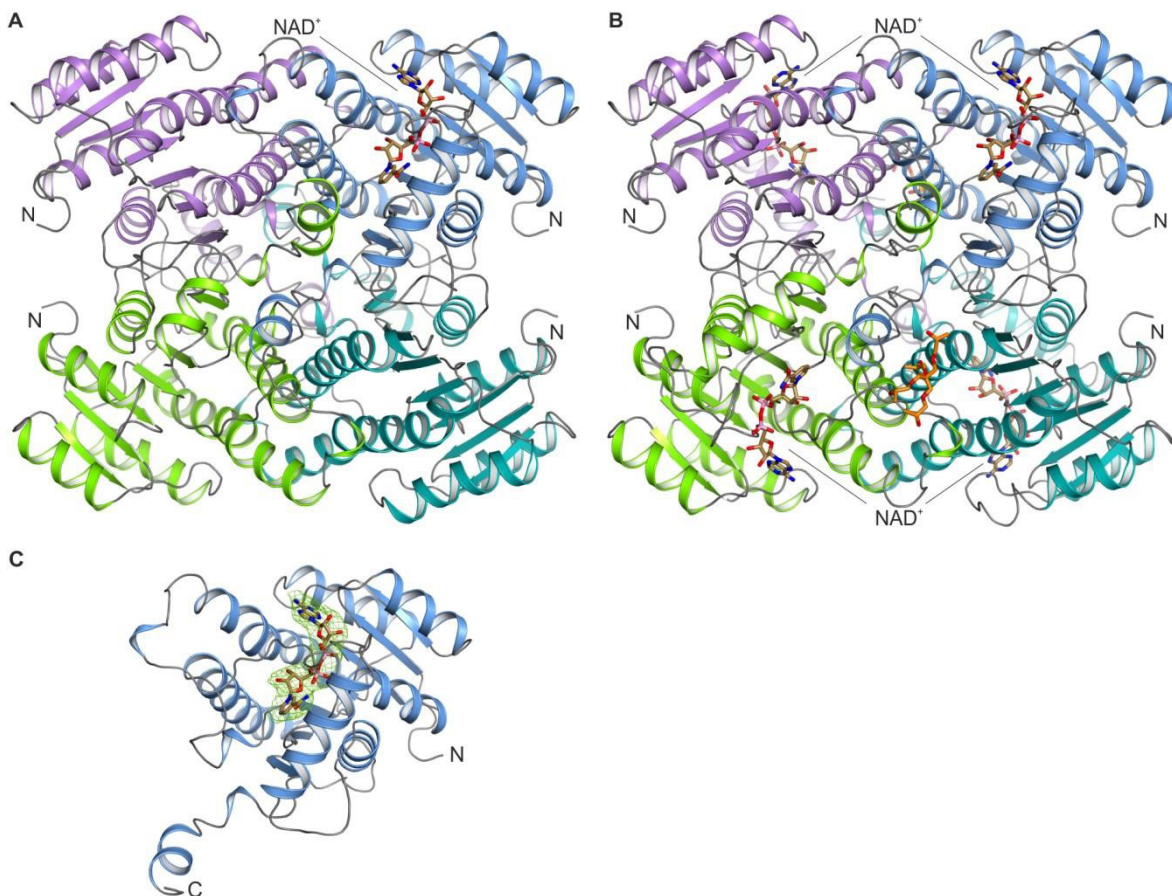
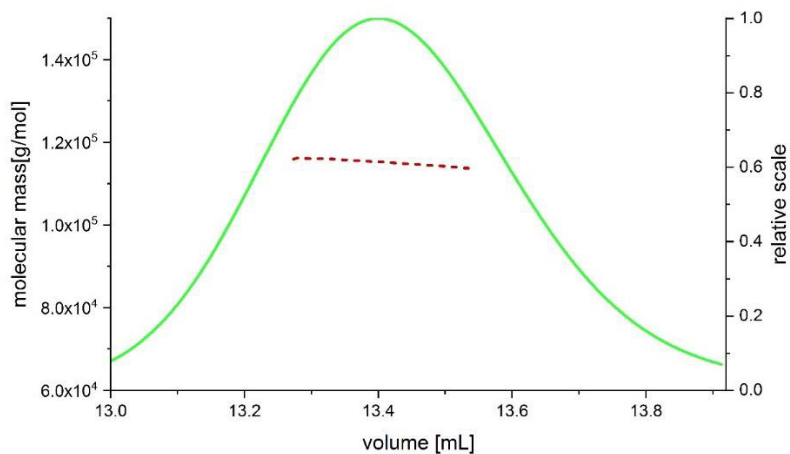


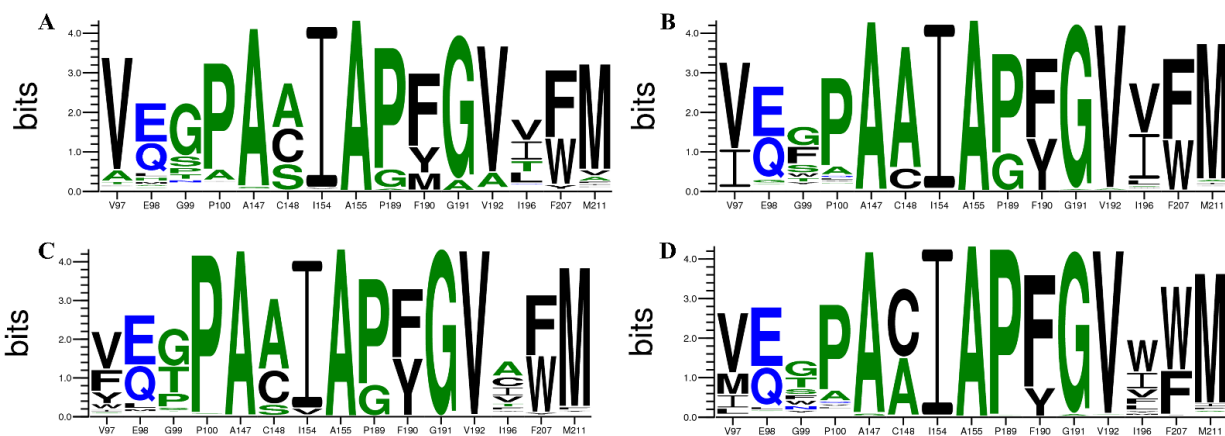
Figure S6: Kinetic characterization of SrBDH1 for (+)-**endo-1a** (left) and  $NAD^+$ . The lines represent the fit obtained using the Michaelis-Menten equation constructed using Origin 2019 software.



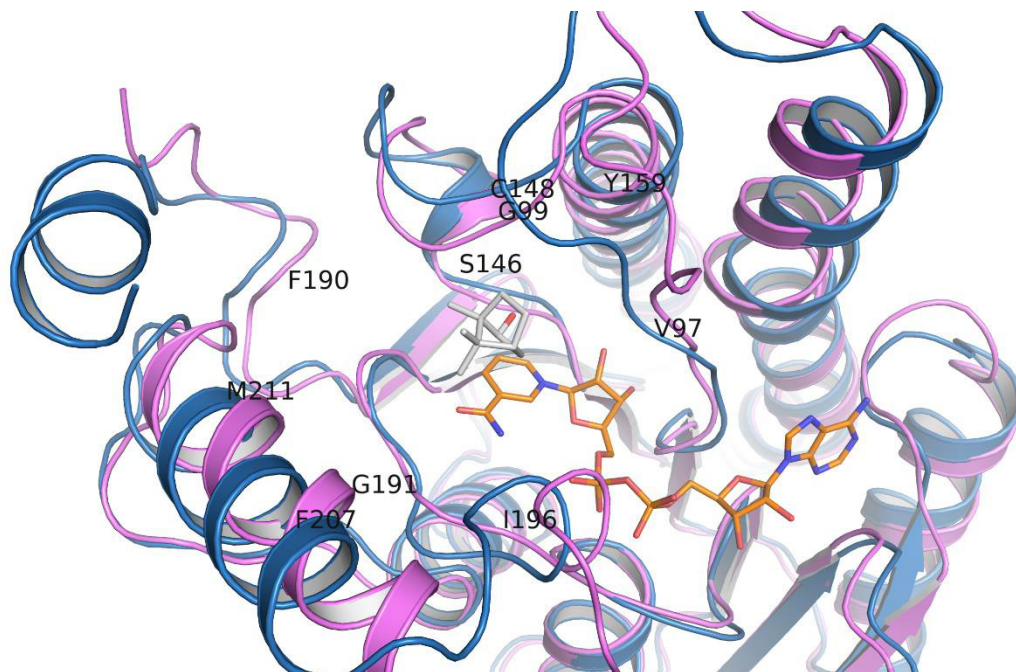
**Figure S7.** (A) Overall architecture of SrBDH1•NAD<sup>+</sup> / high salt forming a tetramer shown in cartoon representation. Only one NAD<sup>+</sup> molecule is bound to one monomer (chain C) depicted in stick representation. (B) Overall architecture of SrBDH1•NAD<sup>+</sup> / PO/OH. Identical view as in panel (A). NAD<sup>+</sup> is shown in brown stick representation. Quality of the electron density clearly indicates that the cofactor is not fully occupied. Two molecules of PO/OH are clearly defined in the electron density shown in orange stick representation. (C) Some view as in panel (A) but merely of the protein chain C in blue cartoon representation with the bound NAD<sup>+</sup>. A POLDER map<sup>[1]</sup> was calculated for the omitted NAD<sup>+</sup>. The electron density map is shown as green mesh contoured at 3  $\sigma$ .



**Figure S8.** SEC/MALS analysis of SrBDH1. Solid, green curve represents the refractive index trace and the red curve the molecular mass at the corresponding elution volumes.



**Figure S9.** Residues suggested by the coupled moves protocol implemented in the Rosetta framework by Ollikainen et al.<sup>[2]</sup> for a better fit in the active site of (+)-*endo-1a* (A), (+)-*exo-1a* (B), (-)-*endo-1a* (C) and (-)-*exo-1a* (D).



**Figure S10.** Secondary structure of the active site of SrBDH1 (blue) and AaADH2 (homology model) (violet). Positions mutated are labeled for SrBDH1. (+)-*endo*-**1a** showed in white and NAD<sup>+</sup> in orange.

## 2. Supplementary tables

**Table S1.** Specific activities measured for AaADH2: alcohol dehydrogenase from *A. annua*<sup>[3]</sup>; SoBDH2: borneol dehydrogenase from *S. officinalis* L.<sup>[4]</sup>; SrBDH1/2: borneol- like dehydrogenases from *S. rosmarinus*; SoBDH3: borneol-like dehydrogenase from *S. officinalis* L. PsBDH: borneol dehydrogenase from *Pseudomonas* sp. TCU-HL1.<sup>[5]</sup> The highest activity for each enzyme is highlighted in bold letters. Reactions with substrates displaying specific activities less than twice the blank are not shown. <sup>[a]</sup> n.c. = no conversion detected.

Substrate	Specific activity (mU/mg)					
	SrBDH1	SrBDH2	SoBDH3	AaADH2	SoBDH2	PsBDH
<i>endo-1a</i>	<b>30.22±0.46</b>	<b>9.98±0.46</b>	n.c.	<b>21.52±0.77</b>	<b>17.92±0.20</b>	<b>122.12±21.54</b>
<i>exo-1a</i>	23.66±0.36	5.05±1.71	0.30±0.05	53.34±1.29	12.74±0.56	114.78±5.17
<i>exo-2a</i>	4.80±0.04	1.54±0.13	n.c.	29.29±1.44	n.c.	68.59±3.58
<i>(+)-endo-2a</i>	3.92±0.36	5.89±1.04	0.27±0.04	22.50±0.81	n.d.	28.12±0.63
<i>endo-3a</i>	4.63±0.22	1.03±0.04	n.c.	14.48±0.53	7.05±1.03	29.54±1.86
<b>4a</b>	2.88±0.12	15.89±1.09	0.46±0.03	87.80±2.73	12.25±1.44	5.02±1.55
<b>5a</b>	3.33±0.07	n.c.	n.c.	29.10±2.87	6.69±0.78	28.49±3.12
<b>6a</b>	10.01±0.15	2.52±0.18	n.c.	16.76±0.73	18.03±0.046	22.81±1.46
<b>7a</b>	4.40±0.62	n.c.	0.23±0.02	20.54±0.47	38.79±2.94	17.51±0.83
<b>8a</b>	8.34±0.32	n.c.	0.21±0.11	14.75±2.11	3.76±0.25	10.69±0.66
<b>9a</b>	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
<b>10a</b>	n.c.	n.c.	n.c.	5.94±0.36	n.c.	n.c.
<b>11a</b>	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
<b>12a</b>	12.93±1.20	4.36±0.82	n.c.	8.70±5.17	2.26±0.10	42.98±5.8
<b>13a</b>	9.75±0.44	1.98±0.10	n.c.	62.73±2.74	6.33±0.13	6.15±0.12
<b>14a</b>	n.c.	n.c.	n.c.	9.96±1.97	n.c.	n.c.
<b>15a</b>	n.c.	n.c.	n.c.	n.c.	3.36±0.11	n.c.
<b>16a</b>	n.c.	n.c.	n.c.	4.77±0.34	4.72±0.54	n.c.
<b>17a</b>	3.27±0.14	-	0.20±0.01	-	6.45±0.49	3.24±0.18

**Table S2.** Conversions for (+)-**1b** and (-)-**1b** at 24 and 48 h with different concentrations of substrate.

(+) camphor	1 mM	2 mM	5 mM
24 h	22%	24%	51%
48 h	39%	45%	81%

(-) camphor	1 mM	2 mM	5 mM
24 h	7%	6%	6%
48 h	9%	10%	11%

Table S3. Crystallographic data

Data collection			
	SrBDH1•NAD <sup>+</sup> / high salt	SrBDH1•NAD <sup>+</sup> / PO/OH	SrBDH1 apo / high salt
Wavelength [Å]	0.9184	0.9184	1.0332
Space group	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2	<i>P</i> 6 <sub>5</sub>	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2
Unit cell parameters a, b, c [Å]	107.4, 107.4, 218.6	196.8, 196.8, 65.3	108.2, 108.2, 230.4
Resolution [Å] <sup>a</sup>	20.00 - 2.60 (2.76 – 2.60)	50.00 - 2.27 (2.41 - 2.27)	48.96-3.10 (3.29-3.10)
Reflections			
Total	473,592 (75,703)	561,781 (87,917)	663,719 (101,808)
Unique	40,188 (6,358)	66,509 (10,563)	25,633 (4,012)
Multiplicity	12.1 (12.0)	8.4 (8.3)	25.9 (25.4)
Completeness [%]	99.9 (99.7)	99.7 (98.4)	99.9 (99.6)
Mean I/σ(I)	14.33 (0.98)	9.88 (1.11)	16.49 (1.59)
R <sub>merge</sub> (I) <sup>b</sup>	0.145 (2.284)	0.205 (1.853)	0.194 (2.125)
R <sub>meas</sub> (I) <sup>c</sup>	0.152 (2.386)	0.218 (1.975)	0.198 (2.169)
CC <sub>1/2</sub> [%] <sup>d</sup>	99.9 (58.5)	99.7 (50.9)	99.9 (88.9)
Refinement			
Resolution [Å] <sup>a</sup>	20.00 - 2.59	50.00 - 2.27	50.00-3.10
Reflections			
Unique	40,006	66,484	25556
Test set [%]	5.0	3.1	5.0
R <sub>work</sub> <sup>e</sup>	0.210 (0.430)	0.197 (0.310)	0.255 (0.372)
R <sub>free</sub> <sup>f</sup>	0.254 (0.470)	0.235 (0.349)	0.305 (0.451)
Contents of A.U. <sup>g</sup>			
Non-H atoms	7437	8044	7293
Protein residues/atoms	1008 / 7363	1038/ 7621	997/ 7290
NAD <sup>+</sup> molecules	1	4	-
PO/OH molecules	-	2	-
Chloride ion	2	1	-
Water molecules	28	196	3
Mean B factors [Å <sup>2</sup> ]			
Wilson	71.9	47.9	82.7
Model atoms	81.0	51.8	94.8
Rmsd <sup>h</sup> from ideal geometry			
Bond lengths [Å]	0.006	0.007	0.004
Bond angles [°]	0.870	0.949	0.748
Model quality <sup>i</sup>			
Overall score	1.87	1.46	1.85
Clash score	8.0	4.1	9.73
Ramachandran favored [%]	97.6	96.9	95.12
Ramachandran outliers [%]	0.1	0.8	0.0
PDB ID	6ZZT	6ZYZ	6ZZO

<sup>a</sup> Values in parentheses refer to the highest resolution shells.

<sup>b</sup>  $R_{\text{merge}}(I) = \sum_h \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i I_{ih}$ , in which  $\langle I_h \rangle$  is the mean intensity of symmetry-equivalent reflections  $h$  and  $I_{ih}$  is the intensity of a particular observation of  $h$ .<sup>[6]</sup>

<sup>c</sup>  $R_{\text{meas}}(I) = \sum_h [N/(N-1)]^{1/2} \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i I_{ih}$ , in which  $\langle I_h \rangle$  is the mean intensity of symmetry-equivalent reflections  $h$ ,  $I_{ih}$  is the intensity of a particular observation of  $h$  and  $N$  is the number of redundant observations of reflection  $h$ .<sup>[6]</sup>

<sup>d</sup>  $CC_{1/2} = (\langle I^2 \rangle - \langle I \rangle^2) / (\langle I^2 \rangle - \langle I \rangle^2) + \sigma_e^2$ , in which  $\sigma_e^2$  is the mean error within a half-dataset.<sup>[6]</sup>

<sup>e</sup>  $R_{\text{work}} = \sum_h |F_o - F_c| / \sum F_o$  (working set, no  $\sigma$  cut-off applied).

<sup>f</sup>  $R_{\text{free}}$  is the same as  $R_{\text{work}}$ , but calculated on the test set of reflections excluded from refinement.

- <sup>g</sup> A.U. – asymmetric unit.  
<sup>h</sup> Rmsd – root-mean-square deviation  
<sup>i</sup> Calculated with MolProbity.<sup>[7]</sup>

**Table S4.** Residues in the active site for SrBDH1 and comparison with the equivalent residues in other members of the family of borneol dehydrogenases.

	99	146	154	155	159	190	191	211	Selective?	
									Borneol	Isoborneol
SrBDH1	G	S	I	A	Y	F	G	M	Yes	Yes
SoBDH1	G	S	I	S	Y	F	G	M	Yes	Yes
SrBDH2	G	S	E	S	Y	C	A	I	Yes	No
SoBDH2	P	S	L	G	Y	Y	A	V	Yes	No
AaADH2	N	S	E	S	Y	G	S	L	No	No
PsBDH	H	S	I	N	Y	Y	M	T	No	No
LiBDH	A	S	G	T	Y	Y	Y	F	No	Yes
AaBDH	E	S	V	A	Y	Y	F	-	No	No
ZsD1	D	S	A	G	Y	Y	A	V	?	?
CoTR	T	S	N	V	Y	G	F	-	?	?
SoBDH3	G	S	L	A	Y	H	G	G	-	-



Table S5. Primers for point mutation

Mutant	Replaced nucleotides	Primers used
G99N	g298a_g299a	5'-gatgctgttcggggtttcaacaataaccgcggtgctgaac-3'
		5'-gttcagcaacgcgggtattgtgaaaacccgaacagcatc-3'
G99H	g298c_g299a	5'-atgctgttcggggtttcaacaataaccgcggtgctga-3'
		5'-tcagcaacgcgggtattgtgaacacccgaacagcat-3'
G99D	g299a	5'-caaagatgctgttcgggtctcaacaataaccgcg-3'
		5'-cgcgggtattgtgaagacccgaacagcatctttg-3'
G99T	g298a_g299c	5'-atgctgttcggggtttcaacaataaccgcggtgctga-3'
		5'-tcagcaacgcgggtattgtgaaaacccgaacagcat-3'
F190Y	t572a_c573t	5'-ccggtcagaacaccatacgggctcacgcagtt-3'
		5'-aactgcgtgagcccgtatggtgttctgaccgg-3'
G191M	g574a_g575t_t576g	5'-cgatgccggtcagaacctgaacgggctcacgcag-3'
		5'-ctgcgtgagcccgttcatggttctgaccggcatcg-3'
G191S	g574a	5'-ccggtcagaacactgaacgggctcacg-3'
		5'-cgtgagcccgttcagtgttctgaccgg-3'
G191F	g574t_g575t	5'-gatgccggtcagaacaagaacgggctcacgcag-3'
		5'-ctgcgtgagcccgttcttctgaccggcatc-3'
M211V	a634g	5'-cccactttgctcacaataacctcaaacatcagcttgctc-3'
		5'-gagcaagctgatgtttgaaggattgtgagcaaaagtg-3'
M211L	a634t	5'-cccactttgctcaaaataacctcaaacatcagcttgctc-3'
		5'-gagcaagctgatgtttgaaggattgtgagcaaaagtg-3'
V97G	t293g	5'-ctgttcgggccttcaccaataaccgcggttc-3'
		5'-gcaacgcgggtattggtgaaggcccgaacag-3'
V97C	g292t_t293g	5'-gctgttcgggccttcacaataaccgcggtgctgaac-3'
		5'-gttcagcaacgcgggtattgtgaaggcccgaacagc-3'
V97P	g292c_t293c	5'-ctgttcgggccttcaggaataaccgcggtgctgaa-3'
		5'-ttcagcaacgcgggtattctgaaggcccgaacag-3'
Y159F	a479t	5'-ttgctcgcggtgaagctgtggccc-3'
		5'-cgggccacagcttcaccgcgagcaa-3'
S146A	a439g_g440c	5'-caatctcggtgcacgcggccggtgaaaatgatgc-3'
		5'-gcatcattttaccgcggccgctgcaccgagattg-3'
I196L	a589c	5'-cgtccggcacgaggccggtcagaac-3'
		5'-gttctgaccggcctcgtgccggaag-3'
V97F	g292t	5'-gttcgggccttcaaaaataaccgcggtgctga-3'
		5'-tcagcaacgcgggtattttgaaggcccgaac-3'
V97Y	g292t_t293a	5'-gctgttcgggccttcataataaccgcggtgctgaac-3'
		5'-gttcagcaacgcgggtattatgaaggcccgaacagc-3'
F207W	t623g_t624g	5'-ctttgctcataataacctcccacatcagctgctgcctc-3'
		5'-gaggcgagcaagctgatgtgggaaggtattatgagcaaag-3'
C148A	t445g_g446c	5'-cgcaatctcggtggccgctcgcggta-3'
		5'-taccgcgagcggccaccgagattgcg-3'
G99T+V98 C	g292t_t293g_g298a_g299 c	5'-gctgttcggggtttcaacaataaccgcggtgctgaaca-3'
		5'-tgttcagcaacgcgggtattgtgaacccgaacagc-3'

### 3. Nucleotide and amino acid sequences of the enzymes used in this study

#### a. Amino acid sequences

>AaADH2

MNGVYPHRLLEGKVAIITGGASGFGESTVRLFAKHGAKVVIADIQDQLGLSLCNDLVNKGIDNVIIYLHCDVTKES  
ENIENTVNTAVSKFGKLDIMFNAGIPGNLDFITLNSDNENFKRVFDVNVFGSFLGAKHAARVMIPAKRGVILFTS  
SVASVLAGEPHSYTVSKHAVIGLMKNLCVELGQYGRVNCISPGSVSTPLVTTAMGVDKEVVDGILCASAVLK  
GVVPTADDVAEAAALYLGSDASRYVTGVNVLVDGGYSTTNPTYSRVIKQTFEDLAKKNEGCNNGVSHAT

>PsBDH

MKLEGGKRIIVTGGAGGIGASVVRAIYAAGATVASMDMNDTLGQQVSEAGKANPGCKSRYYHCNIADRPEVE  
KAFATAAEDMGGLDVMVNVAGVHRHSPDAIAEELYDMLFRVNVLTINTNAVAYRLMKGGGIGNIIFGSES  
GLTGEINNALYSATKAAVHTWTRNVARQWGPDGIRINAVLPYVTPMYVDFRNALSSDLAAHDAATKTDIPL  
GGKFGDADKDLAPVMVFLASDASHFMTGQMFPVDGGLIAVR

>SoBDH1

MSSSAVSKRLEGKVAIVTGGASGIGASTVSLFHDHGAKVVIADIQDNLGQTLAAGRLGRNISYIHCVDVTDENQVRA  
LVDA TVAKHGGVDIMFNSAGIVEGPTVSIFDADK GALERLLGINLVGGFLAAKHAARVMSPKKGCIIFTASACT  
EVAGISPGYVASKYGIVGLMKSLAAELGSHGIRANCVSPFGVLTGIAAGDDKTKLMFEGLMKVGNLKKGKILT  
ADDVAKAALYLASDEASYVSGVNLVLDGGYSVVNP

>SoBDH2

MATGAANVESPQSLPLRLLGRVALVTGGSSGIGESIVLLFRKHGAKVCIADVQDNQGRQLCETLGGSSDIAFCHC  
DVTIEDDVKRAVDFTVDKFGTLDIMVNNAGVSGPPCPDIRDFELSAFDRVFDINVRGVFIGMKHAARIMIPAKKG  
SIISISSVASTMGGLGPHAYTGSKHAVLGLTKNVAELGKHGIRVNCVSPYAVATSLALAHLEAERTEDTWDDF  
RRFVADNANLQGVELTMEDVANAVVFLASDEARYVSGMNLMDVGGFTSTNHALQVFRP

>SrBDH1

MSCNTAVSRRELEGKVAIVTGGASGIGASTVRLFHDHGAKVVIADIQDDLQTLADRLGRNISYTHCDVTDEDQV  
RALVDAAVAKHGGVDIMFNSAGIVEGPNISIFDVKDELERLMGINLVGAFLAAKHAARVMVPAKKGCIIFTASA  
CTEIAAGIAGHSYTASKYGIVGLMKSLAVALGSHGIRANCVSPFGVLTGIVPDDEASKLMFEGIMSKVGNLKKGKILT  
AEDVAVTVLYLASEEASYVSGVNLVLDGGYTVVNPTFINVITAGQS

>SrBDH2

MKMKPESNGLHTSKRLEGKVAIITGGASGFGEATAALFVRHGAKVVIADVQDDRGSALCRDLGLPNQISYVHCD  
VTSADADVSAAVDLAVSKYGGLDIMFNAGIPGGLDFTIVDADNDNFRRVFEVNVYGAFLGAKHAARAMIPARR  
GGAILFTASVASAVAGESPHSYAASKHAVVGLMRNLCVELGQHGIRVNAISPCAVATPLLTGTMGVEKAVVEDI  
ICASANLKGVPPTAEDVAEAAALYLGSDSKFVSGLNLVVDGGYSTTNQSYSRAGTTGSQTLPSASSREPEMG

>SoBDH3

MDSDESKDDMTHHLPSKRLSGKVAIVTGGARGIGAATAKAFANGANVVIADILDEPGAEEAAAIGGKYVHCD  
VSVEKDVERAIQVAVDWKGRDLIMFNAGISGPEGSITNLKMEQLAALLEINLVVHGKHAARAMIEGRNAG  
TIICSSSSAATMGGLASHAYTSLKAAILGVARSSACELGLHSIRVNCVSPHGVPEMLMTAYRRFLGNENLQPQD  
VRKIVGEKGSLLRGRGSMEDVAEAVVFLASDEAGFITGHNLVIDGGYTCASNQMNFIYQE

b. Nucleotide sequence (codon optimized for *E. coli*)

>SoBDH1

ATGAGCAGCAGCGCGGTGAGCAAACGCCTGGAAGGCAAAGTGCGATTGTGACCGCGCGCGAGCGGCA  
TTGGCGCGAGCACCGTGAGCCTGTTTCATGATCATGGCGCGAAAGTGGTGATTGCGGATATTCAGGATAAC  
CTGGGCCAGACCCTGGCGGGCCGCTGGGCCGCAACATTAGCTATATTCATTGCGATGTGACCGATGAAAA  
CCAGGTGCGCGCGCTGGTGGATGCGACCGTGCGGAAACATGGCGGCGTGGATATTATGTTTAGCAACGCGG  
GCATTGTGGAAGGCCCGACCGTGAGCATTGTTGATGCGGATAAAAGGCGCGCTGGAACGCCTGCTGGGCATT  
AACCTGGTGGGCGGCTTCTGGCGGCGAAACATGCGGCGCGCGTGATGAGCCCGACAAAAAAGGCTGCAT  
TATTTTTACCGCGAGCGCGTGCACCGAAGTGGCGGGCATTAGCGGCCCGGGCTATGTGGCGAGCAAATATG  
GCATTGTGGGCTGATGAAAAGCCTGGCGGCGGAACTGGGCAGCCATGGCATTGCGCGAACTGCGTGAGC  
CCGTTTGGCGTGCTGACCGGCATTGCGGCGGGCGATGATAAAACCAAACCTGATGTTTGAAGGCCGTGATGAG  
CAAAGTGGCAACCTGAAAGGCAAAATTCGACCGCGGATGATGTGGCGAAAGCGCGCTGTATCTGGCG  
AGCGATGAAGCGAGCTATGTGAGCGGCGTGAACCTGGTGTCTGGATGGCGGCTATAGCGTGGTGAACCCGTG  
A

>SoBDH2

ATGGCTACAGGCGCTGCAAATGTTGAGTCACCACAGTCTCTCCCTTAAGATTATTAGGGAGAGTTGCTCTG  
GTCACCGGAGGTTTCGAGTGGCATTGGAGAGAGCATCGTGCTTCTGTTTCGTAAACATGGTGCAAAAGTTTGT  
ATAGCCGATGTTCAAGACAACCAAGGGCAACGTCTCTGCGAAACCCTAGGTGGCAGCTCAGACATCGCCTT  
TTGCCACTGCGATGTGACAATTGAAGATGATGTCAAGCGTGCAAGTGGACTTCACCGTGGACAAGTTCGGTA  
CCCTCGACATAATGGTGAACAACGCTGGGGTGTGGGCCACCCTGCCCCGATATCCGCGACTTTGAACTCT  
CCGCTTTCGACAGGGTCTTCGACATAAACGTGAGAGGGGTTTTTCATCGGAATGAAGCACGCGGCTCGCATA  
ATGATCCCGCCAAGAAAGGGTCGATAATATCAATCTCCAGCGTGGCGAGCACTATGGGCGGCTTAGGGCC  
TCACGCATACACGGGGTCCAAGCATGCTGTTCTGGGACTCACCAAGAACGTCGCGGCGGAGCTAGGGAAAC  
ACGGCATAACGCTGAACCTGCGTGTGCGCGTACGCGGTCGCGACTAGCTTGGCGCTGGCGCACTTGCCCCGAG  
GCGGAGAGGACGGAGGATACGTGGGATGATTTCCGTAGATTTGTGGCGGATAATGCAAACCTGACGGGAGT  
GGAATTGACTATGGAGGATGTGGCGAATGCGGTGGTTTTCTTGGCGAGTGATGAGGCAAGGTATGTAAGCG  
GCATGAATCTCATGGTTGATGGAGGCTTACATCTACAAATCATGCCCTCCAAGTATTTCTGCCCTGA

>SrBDH1

ATGAGCTGCAACACCGCGGTTAGCCGTCGTCTGGAGGGCAAGGTTGCGATTGTTACCGGTGGCGCGAGCGG  
TATTGGTGCAGACACCGTGCGTCTGTCCACGATCACGGCGCGAAAGTGGTTATCGCGGACATTCAGGACG  
ATCTGGGTCAAACCCTGGCGGATCGTCTGGGCCGTAACATCAGCTACACCCACTGCGACGTTACCGACGAG  
GATCAGGTTTCGTGCGCTGGTTGACGCGGCGGTTGCGAAGCATGGTGGCGTGGACATCATGTTACGCAACGC  
GGGTATTGTTGAAGGCCCGAACAGCATCTTTGACGTGGATAAAGGACGAGCTGGAACGTCTGATGGGTATTA  
ACCTGGTTGGTGCCTTCTGGCGGCGAAACATGCGGCGCGTGTGATGGTTCCGGCGAAGAAAGGTTGCATC  
ATTTTTACCGCGAGCGCGTGCACCGAGATTGCGGGTATTGCGGGCCACAGCTACACCGCGAGCAAGTATGG  
TATCGTTGGCCTGATGAAAAGCCTGGCGGTGGAACCTGGGTAGCCACGCGATTCTGCGCAACTGCGTGAGCC  
CGTTCCGGTGTCTGACCGGCATCGTGCCGACGATGAGGCGAGCAAGCTGATGTTTGAAGGTATTATGAGC  
AAAGTGGGTAACCTGAAGGGCAAAATCCTGACCGCGGAAGATGTGGCGGTTACCGTGCTGTACCTGGCGAG  
CGAGGAAGCGAGCTATGTTAGCGGTGTGAACCTGCTGGTTGACGGTGGCTATACCGTGGTTAACCCGACCT  
TTATCAACGTGATTACCGCGGGCCAAAGCTAA

>SrBDH2

ATGAAGATGAAACCGGAGAGCAACGGTCTGCACACCAGCAAGCGTCTGGAAGGCAAAGTTGCGATCATT  
CCGGTGGCGCGAGCGGTTTTGGTGAAGCGACCGCGGCGTGTGTTGTTTCGTCACGGTGCGAAGGTGGTTATC  
GCGGATGTGCAGGACGATCGTGGTAGCGCGCTGTGCCGTGACCTGGGTCTGCCGAACCAAATTAGCTACGT  
GCACTGCGATGTTACCAGCGATGCGGATGTGAGCGCGGCGGTTGATCTGGCTGTGAGCAAATATGGTGGCC  
TGGACATCATGTTCAACAACGCGGGCATTCCGGGTGGCCTGGACTTTACCATCGTTGACGCGGATAACGAC  
AACTTCCGTCGTGTGTTTGAAGTGAACGTTTACGGTGCCTTCTGGGCGCGAAACATGCGGCGCGTGCATG  
ATCCCGCGCGTCTGGTGGCGGATTCTGTTTACCGCGAGCGTTGCGAGCGCGGTTGCGGGCGAGAGCCC

GCACAGCTATGCGGCGAGCAAACACGCGGTGGTTGGCCTGATGCGTAACTGTGCGTTGAACTGGGTCAGC  
ACGGCATCCGTGTTAACGCGATTAGCCCGTGC GCGGTTGCGACCCCGCTGCTGACCGGTACCATGGGCGTG  
GAGAAGGCGGTGGTTGAAGATATCATTTGCGCGAGCGGAACCTGAAAGGTGTGGTTCCGACCGCGGAGG  
ATGTTGCGGAAGCGGCGCTGTACCTGGGTAGCGACGAGAGCAAGTTCGTGAGCGGCCTGAACCTGGTGGTT  
GACGGTGGCTACAGCACCACCAACCAGAGCTATAGCCGTGCGGGTACCACCGGCAGCCAGACCCTGCCGAG  
CGCGAGCAGCCGTGAGCCGAAATGGGTAA

>SoBDH3

ATGGACAGCGATGAAAGCAAGGACGATATGACCCACCACCTGCCGAGCAAGCGTCTGAGCGGCAAAGTGG  
CGGTTATCACCGGTGGCGCGCGTGGTATTGGTGC GCGGACCCGCGAAAGCGTTTGC GGAGAACGGCGCGAAC  
GTGGTTATCGCGGACATTCTGGATGAGCCGGGTGCGGAAGTGGCGGCGGCGATTGGTGGCAAGTACGTTCA  
CTGCGACGTGAGCGTTGAGAAAGATGTGGAACGTGCGATCCAGGTGGCGGTTGACTGGAAGGGTCGTCTGG  
ATATCATGTTTAAACAACGCGGGTATTAGCGGCCCGGAAGGTAGCATCACCAACCTGAAGATGGAGCAACTG  
GCGGCGCTGCTGAAAATCAACCTGAACGGCGTGGTTCACGGTATTAAGCATGCGGCGCGTGCATGATTGA  
GGGTCGTAACGCGGGTACCATCATTTGCAGCAGCAGCAGCGCGGCGACCATGGGTGGCCTGGCGAGCCATG  
CGTACACCCTGAGCAAAGCGGCGATTCTGGGCGTGGCGGTAGCAGCGCGTGC GAACTGGGTCTGCACAGC  
ATCCGTGTGAACTGCGTTAGCCCGCACGGCGTTCGAGCGAGATGCTGATGACCGCGTATCGTCGTTTCCTG  
GGTAACGAAAACCTGCAGCCGCAAGACGTGCGTAAGATTGTTGGCGAGAAAGGTAGCCTGCTGCGTGGTGC  
TGGTGGCAGCATGGAGGACGTGGCGGAAGCGGTGGTTTTCTGGCGAGCGATGAAGCGGGCTTTATTACCG  
GTCACAACCTGGTTATCGATGGTGGCTACACCTGCGCGAGCAACCAGATGAACTTTATCTATCAAGAGTAA

#### 4. Complete name and accession numbers of SDR used in phylogenetic tree

**LiBDH:** borneol dehydrogenase from *Lavandula intermedia* (AFV30207.1<sup>a</sup>)

**AaBDH:** borneol dehydrogenase from *Artemisia annua* (ANJ65052.1<sup>a</sup>)

**AaADH2:** alcohol dehydrogenase from *Artemisia annua* (ADK56099.1<sup>a</sup>)

**PsBDH:** borneol dehydrogenase from *Pseudomonas sp. TCU-HL1* (WP\_032492645.1<sup>a</sup>)

**SoBDH1:** borneol dehydrogenase 1 from *Salvia officinalis L.* (MT525100<sup>a</sup>)

**SoBDH2:** borneol dehydrogenase 2 from *Salvia officinalis L.* (MT525099<sup>a</sup>)

**SoBDH3:** alcohol dehydrogenase 3 from *Salvia officinalis L.*

**SrBDH1:** alcohol dehydrogenase 1 from *Rosmarinus officinalis* (MT857224<sup>a</sup>)

**SrBDH2:** alcohol dehydrogenase 2 from *Rosmarinus officinalis*

**PgADH:** short-chain alcohol dehydrogenase from *Panax ginseng* (ACL37155.1<sup>a</sup>)

**PaADH:** alcohol dehydrogenase from *Prunus armeniaca* (ABZ79222.1<sup>a</sup>)

**CmADH:** putative alcohol dehydrogenases from *Cucumis melo* (ABC02082.1<sup>a</sup>)

**AtADH:** putative dehydrogenase from *Arabidopsis thaliana* (AAM65725.1<sup>a</sup>)

**DsTR:** tropinone reductase-II from *Datura stramonium* (AAA33282.1<sup>a</sup>)

**CoTR:** Tropinone reductase from *Cochlearia officinalis*

**At2g29330:** Tropinone reductase from *Arabidopsis thaliana* (Q9ZW16<sup>b</sup>)

**At2g29150:** Tropinone reductase from *Arabidopsis thaliana* (Q9ZW03<sup>b</sup>)

**At2g29350:** Senescence-associated protein from *Arabidopsis thaliana* (Q9ZW18<sup>b</sup>)

**MMR:** Menthol dehydrogenase from *Mentha x piperita* (AAQ55960.1<sup>a</sup>)

**MNR:** neomenthol dehydrogenase from *Mentha x piperita* (AAQ55959.1<sup>a</sup>)

**SAD-C:** short-chain alcohol dehydrogenase from *Pisum sativum* (AAF04253.1<sup>a</sup>)

**DIHSDH:** 3-beta-hydroxysteroiddehydrogenase from *Digitalis lanata* (Q93Y47<sup>b</sup>)

**PeSDH:** Secoisolariciresinol dehydrogenase from *Podophyllum peltatum* (AAK38664.1<sup>a</sup>)

**FiSDH:** secoisolariciresinol dehydrogenase from *Forsythia x intermedia* (AAK38665.1<sup>a</sup>)

**HsCR:** carbonyl reductase from *Homo sapiens* (AAA52070.1<sup>a</sup>)

**SsHSDH:** 20-beta-hydroxysteroid dehydrogenase from *Sus scrofa* (AAA30980.1<sup>a</sup>)

**CgCR:** Carbonyl reductase from *Cricetulus griseus* (BAB07797.1<sup>a</sup>)

**MpIDH:** (-)-isopiperitenol dehydrogenase *Mentha x piperita* (AAU20370.1<sup>a</sup>)

**CrADH1:** Alcohol dehydrogenase 1 from *Catharanthus roseus* (5O98\_1<sup>c</sup>)

**saIR:** Salutaridine reductase from *Papaver somniferum* (3O26\_1<sup>c</sup>)

**ZSD1:** Short-chain dehydrogenase/reductase 1 from *Zingiber zerumbet* (BAK09296.1<sup>a</sup>)

**ZmSDP:** Sex determination protein tasselseed-2 from *Zea mays* (ACG37730.1<sup>a</sup>)

**CsADH:** Short chain alcohol dehydrogenase from *Citrus sinensis* (ADH82118.1<sup>a</sup>)

**StDR:** Short-chain dehydrogenase/reductase from *Solanum tuberosum* (AAT75153.1<sup>a</sup>)

**CbHCDH:** (S)-6 beta-hydroxycineole dehydrogenase from *Citrobacter braakii* (ACX31575.1<sup>a</sup>)

**RsGDH:** Galactitol dehydrogenase from *Rhodobacter sphaeroides* (ACM89305.1<sup>a</sup>)

**AaFDH:** Farnesol dehydrogenase 1 from *Aedes aegypti* (ADB03639.1<sup>a</sup>)

**AaDH:** Broad substrate reductase/dehydrogenase *Artemisia annua* (ACZ34296.1<sup>a</sup>).

<sup>a</sup>NCBI, <sup>b</sup>Uniprot, <sup>c</sup>PDB

## 5. Commands used for couple moves protocol

```
~/Rosetta/main/source/bin/coupled_moves.linuxgccrelease -s pdb_file -resfile res_file -  
database ~/Rosetta/main/database -mute protocols.backrub.BackrubMover -  
extra_res_fa params_file -ex1 -ex2 -extrachi_cutoff 0 -nstruct 99 -  
coupled_moves::mc_kt 0.6 -coupled_moves::ntrials 1000 -  
coupled_moves::initial_repack false -coupled_moves::ligand_mode true -  
coupled_moves::ligand_weight N
```

## References

- [1] D. Liebschner, P. V. Afonine, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, P. D. Adams, *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73*, 148–157.
- [2] N. Ollikainen, R. M. de Jong, T. Kortemme, *PLoS Comput. Biol.* **2015**, *11*, 1–22.
- [3] D. R. Polichuk, Y. Zhang, D. W. Reed, J. F. Schmidt, P. S. Covello, *Phytochemistry* **2010**, *71*, 1264–1269.
- [4] I. Drienovská, D. Kolanovi, A. Chánique, V. Sieber, M. Hofer, R. Kourist, *Phytochemistry* **2020**, *172*, DOI 10.1016/j.phytochem.2019.112227.
- [5] H. Tsang, J.-L. Huang, Y.-H. Lin, K. Huang, P. Lu, G. Lin, A. A. Khine, A. Hu, H.-P. Chen, *Appl. Environ. Microbiol.* **2016**, *82*, 6378–6385.
- [6] P. A. Karplus, K. Diederichs, *Science (80-. )*. **2012**, *336*, 1030–1033.
- [7] V. B. Chen, J. R. Wedell, R. K. Wenger, E. L. Ulrich, J. L. Markley, *J. Biomol. NMR* **2015**, *63*, 77–83.



## Publication II

Dimos, Nicole; Helmer, Carl P. O.; Chánique, Andrea M.; Wahl, Markus C.; Kourist, Robert; Hilal, Tarek; Loll, Bernhard (2022): CryoEM analysis of small plant biocatalysts at sub-2 Å resolution. *Acta cryst. D, Structural biology* 78 (Pt 1)  
DOI: <https://doi.org/10.1107/S205979832101216X>



# CryoEM analysis of small plant biocatalysts at sub-2 Å resolution

Nicole Dimos,<sup>a,‡</sup> Carl P. O. Helmer,<sup>a,‡</sup> Andrea M. Chánique,<sup>b,c</sup> Markus C. Wahl,<sup>a,d</sup> Robert Kourist,<sup>b,\*</sup> Tarek Hilal<sup>e,\*</sup> and Bernhard Loll<sup>a,f,\*</sup>

Received 21 July 2021

Accepted 16 November 2021

Edited by M. Rudolph, F. Hoffmann-La Roche Ltd, Switzerland

‡ These authors contributed equally.

**Keywords:** cryo-electron microscopy; camphor; terpenes; borneol dehydrogenases; high resolution; green chemistry; plant biocatalysts.

**EMDB references:** *Salvia officinalis* borneol dehydrogenase 2, EMD-12739; *Salvia rosmarinus* borneol dehydrogenase 1, EMD-12740

**PDB references:** *Salvia officinalis* borneol dehydrogenase 2, 7o6p; *Salvia rosmarinus* borneol dehydrogenase 1, 7o6q

**Supporting information:** this article has supporting information at journals.iucr.org/d

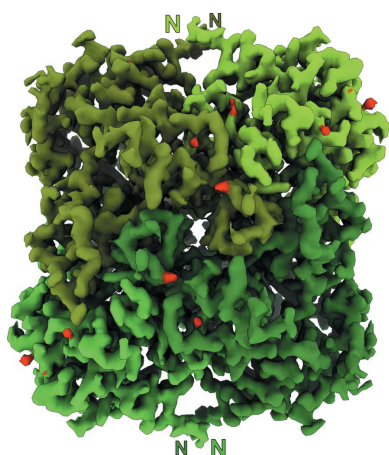
<sup>a</sup>Institute of Chemistry and Biochemistry, Department of Biology, Chemistry, Pharmacy, Laboratory of Structural Biochemistry, Free University of Berlin, Takustrasse 6, 14195 Berlin, Germany, <sup>b</sup>Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria, <sup>c</sup>Department of Chemical and Bioprocesses Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, 7810000 Santiago, Chile, <sup>d</sup>Macromolecular Crystallography, Helmholtz-Zentrum Berlin für Materialien und Energie, Albert-Einstein-Strasse 15, 12489 Berlin, Germany, <sup>e</sup>Institute of Chemistry and Biochemistry, Research Center of Electron Microscopy and Core Facility BioSupraMol, Free University of Berlin, Fabeckstrasse 36A, 14195 Berlin, Germany, and <sup>f</sup>moloX GmbH, Takustrasse 6, 14195 Berlin, Germany. \*Correspondence e-mail: kourist@tugraz.at, tarek.hilal@fu-berlin.de, loll@chemie.fu-berlin.de

Enzyme catalysis has emerged as a key technology for developing efficient, sustainable processes in the chemical, biotechnological and pharmaceutical industries. Plants provide large and diverse pools of biosynthetic enzymes that facilitate complex reactions, such as the formation of intricate terpene carbon skeletons, with exquisite specificity. High-resolution structural analysis of these enzymes is crucial in order to understand their mechanisms and modulate their properties by targeted engineering. Although cryo-electron microscopy (cryoEM) has revolutionized structural biology, its applicability to high-resolution structural analysis of comparatively small enzymes has so far been largely unexplored. Here, it is shown that cryoEM can reveal the structures of plant borneol dehydrogenases of ~120 kDa at or below 2 Å resolution, paving the way for the rapid development of new biocatalysts that can provide access to bioactive terpenes and terpenoids.

## 1. Introduction

Despite the stunning recent success of single-particle cryoEM in the structural analysis of many large molecular machines, comparatively small proteins remain a major challenge for this technique (Kühlbrandt, 2014; Lyumkis, 2019; Vinothkumar & Henderson, 2016). To date, the highest resolution achieved by cryoEM is 1.14 Å for the highly symmetric 480 kDa protein apoferritin (Yip *et al.*, 2020). Presently, the cryoEM structures of only four macromolecular complexes smaller than 120 kDa have been reported at a resolution better than 3.0 Å (Supplementary Table S1). Due to limited structural data and frequently insufficient understanding of the molecular basis of enzyme catalysis, protein engineering still mainly relies on combinatorial approaches for enzyme engineering such as random or saturation mutagenesis. An expansion of the scope of single-particle analysis towards the rapid elucidation of the structures of smaller proteins has tremendous potential to increase the rational element of protein engineering.

Enzyme catalysis offers an efficient and sustainable alternative to traditional chemical synthesis, as biocatalysts harbor excellent selectivity and work under mild reaction conditions. Today, enzymes are widely used in the chemical and pharmaceutical industries, and the application of biocatalysts to the manufacture of chemicals from renewable resources is a



rapidly growing field. However, due to limited structural data and, as a consequence, insufficient understanding of the molecular basis of enzyme catalysis, rational improvement of biotechnologically relevant enzymes has been severely hampered. Protein engineering relies mainly on directed evolution or semi-rational approaches. While often successful, these methods require the implementation of high-throughput screenings and a substantial effort in terms of laboratory work, leading to long time-to-market horizons. This situation could be alleviated by expanding the scope of cryoEM towards the elucidation of high-resolution structures of smaller proteins.

A particularly interesting application of enzyme catalysis is the synthesis and modification of bioactive terpenes and terpenoids. With more than 50 000 different structures, terpenes are a structurally and functionally extraordinarily diverse group of natural products (Oldfield & Lin, 2012). The outstanding selectivity of the enzymes involved in the formation of terpene carbon skeletons (Christianson, 2017), their primary functionalization (Bohlmann & Keeling, 2008) and their further derivatization (Rinkel *et al.*, 2019) could enable the formation of a myriad of new terpene derivatives with diverse, interesting properties for the food and pharmaceutical industries via environmentally friendly catalytic processes (Newman & Cragg, 2016; Oldfield & Lin, 2012). To this end, a detailed understanding of the molecular basis of the reaction mechanisms and selectivity of the enzymes is required.

Bornane-type monoterpenoids, such as borneol, isborneol and camphor, are found in essential oils from plants and are used in traditional medicine and cosmetics (Cheng *et al.*, 2013). Racemic borneol, isborneol and camphor are currently produced from  $\alpha$ -pinene, a side product of cellulose production. Essential oils from plants are often enriched in one of the enantiomers of these compounds, indicating the potential presence of highly stereoselective borneol dehydrogenases (BDHs). An enzymatic route towards pure enantiomers using enantioselective dehydrogenases would be highly desirable to avoid the labor-intensive and expensive extraction of pure enantiomers from plants.

BDHs belong to the family of short-chain dehydrogenase-reductases (SDRs; Chanique *et al.*, 2021). The members of this enzyme class have a TGXXX(AG)XG NAD<sup>+</sup>-binding motif and a YXXXK active-site motif (Ladenstein *et al.*, 2008; Kallberg *et al.*, 2002) and form dimers or tetramers. Some BDHs have a high twofold stereoselectivity in the conversion of chiral monoterpenoids (Chanique *et al.*, 2021; Croteau *et al.*, 1978; Drienovska *et al.*, 2020) by preferring one of the two substrate enantiomers and forming a stereocenter by asymmetric reduction of the diastereotopic keto group. The selective oxidation of (+)-borneol to (+)-camphor by a partially purified BDH from *Salvia officinalis* L. was first described by Croteau *et al.* (1978). Recently, the isolation and purification of two BDHs confirmed the high stereoselectivity of these enzymes (Figs. 1a, 1b and 1c; Drienovska *et al.*, 2020; Chanique *et al.*, 2021).

The understanding of terpene formation on a structural and mechanistic basis is important for the engineering of biosyn-

thetic pathways for the formation of new terpenoids (Kemper *et al.*, 2017). Unfortunately, the difficulty in producing enzymes from higher organisms in bacteria, and often their limited stability, make structure determination by classical crystallization very challenging. In order to obtain a structure under these circumstances, approaches such as truncation and homology modeling have been utilized, both of which have limited informational value for mechanistic studies and enzyme engineering. In the particular case of BDHs, only two crystal structures have been reported to date: those of the nonselective bacterial BDH from *Pseudomonas* sp. TCU-HL1 (*PsBDH*; PDB entry 6m5n; Khine *et al.*, 2020) and the enantioselective BDH from *Salvia rosmarinus* (*SrBDH1*; Chanique *et al.*, 2021). Although structural analysis of *SrBDH1* allowed us to identify a hydrophobic pocket that discriminates the monoterpenol isborneol, structures of additional BDHs, for example from *S. officinalis* (*SoBDH2*), are required to rationalize the selectivity of the enzymes towards (+)-borneol. Here, we report the determination of the structures of two stereoselective dehydrogenases, *SrBDH1* and *SoBDH2*, by single-particle cryoEM.

## 2. Experimental procedures

### 2.1. Cloning

The synthetic genes for the borneol-type dehydrogenases *SoBDH2* from *S. officinalis* (GenBank ID MT525099) and *SrBDH1* from *S. rosmarinus* (GenBank ID MT857224) were ordered from GenScript (USA), codon-optimized for *Escherichia coli* expression and cloned into the vector pET-15b in frame with an N-terminal His<sub>6</sub> tag (Chanique *et al.*, 2021; Drienovska *et al.*, 2020).

### 2.2. Expression and purification of *SoBDH2*

*E. coli* BL21-RIL cells (Stratagene) were transformed with a pET-15a vector containing *SoBDH2* fused to an N-terminal hexahistidine tag. Protein induction was carried out in auto-induction medium at 37°C for 7 h with subsequent cooling to 16°C (Studier, 2018). The cells were grown overnight and harvested by centrifugation (10 min at 7000 rev min<sup>-1</sup> at 4°C). The pellets were resuspended in 20 mM Tris-HCl pH 8.0, 500 mM NaCl (buffer A). The cells were lysed by homogenization at 4°C for 7 min after the addition of 0.5 mg l<sup>-1</sup> DNase and the lysate was cleared by centrifugation (30 min at 21 500 rev min<sup>-1</sup> at 4°C). All subsequent purification steps were performed at 4°C. A Ni<sup>2+</sup>-NTA column (1 ml column volume, Macherey Nagel) was equilibrated with buffer A, and *SoBDH2* was loaded onto the column and washed with 15 column volumes of buffer A. *SoBDH2* was eluted with buffer A supplemented with 300 mM imidazole. The protein was incubated with a threefold molar excess of NAD<sup>+</sup> (0.5 M in double-distilled H<sub>2</sub>O) for 10 min on ice prior to size-exclusion chromatography (SEC). SEC was performed using a HiLoad Superdex S200 16/60 column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 8.0, 125 mM NaCl. Pooled protein fractions were concentrated with Amicon Ultra-15 (Merck

KGaA) to  $11.2 \text{ mg ml}^{-1}$  as measured by the absorbance at 280 nm. *SrBDH1* was purified using a practically identical protocol (Chánique *et al.*, 2021).

### 2.3. Size-exclusion chromatography–multi-angle light scattering (SEC-MALS)

SEC-MALS experiments were performed at  $18^\circ\text{C}$ . *SoBDH2* was loaded onto a Superdex 200 Increase 10/300 column (GE Healthcare) coupled to a miniDAWN TREOS three-angle light-scattering detector (Wyatt Technology) in combination with a RefractoMax520 refractive-index detector. For calculation of the molecular mass, protein concentrations were determined from the differential refractive index with a specific refractive-index increment ( $dn/dc$ ) of  $0.185 \text{ ml}^{-1}$ . Data were analyzed using the *ASTRA* 6.1.4.25 software (Wyatt Technology).

### 2.4. Differential scanning fluorometry (DSF)

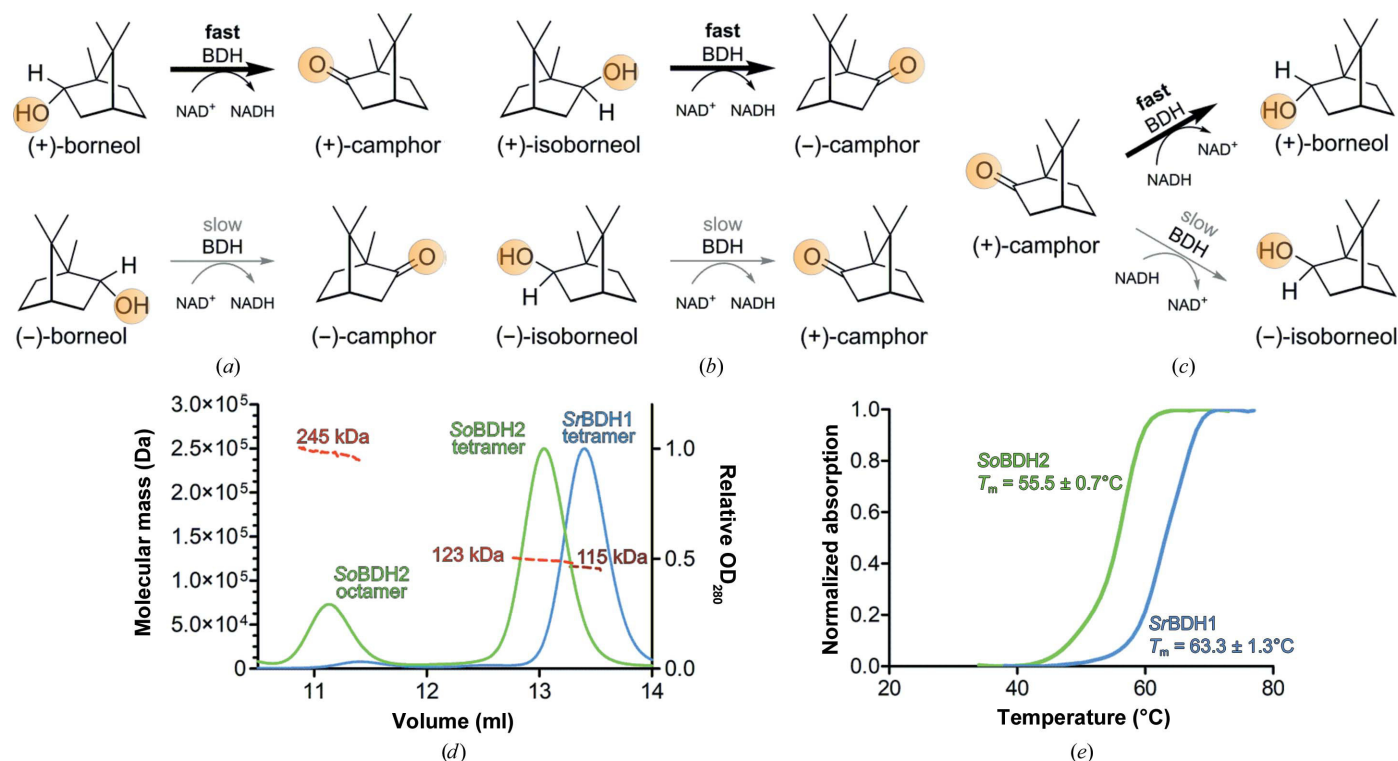
The melting temperatures of the proteins were measured using an Mx3005P qPCR system (Agilent) in 96-well plate format under the buffer condition  $20 \text{ mM}$  Tris–HCl pH 8.0,  $125 \text{ mM}$  NaCl as used for crystallization or cryoEM experiments. Each well contained  $10 \mu\text{l}$  buffer and  $10 \mu\text{l}$  protein ( $0.15 \mu\text{g} \mu\text{l}^{-1}$ ) with a final concentration of  $10\times$  SYPRO

Orange dye (Invitrogen). The program consisted of three steps: step 1 was a pre-incubation for 1 min at  $20^\circ\text{C}$  and steps 2 and 3 were cycles comprising a temperature increase of  $1^\circ\text{C}$  within 20 s. The temperature gradient proceeded from  $25$  to  $95^\circ\text{C}$  at  $1^\circ\text{C}$  per minute. Samples were measured in triplicate. The data were acquired using the *MxPro QPCR* software (Agilent) and analyzed using the *DSF Analysis* version 3.0.1 tool (<ftp://ftp.sgc.ox.ac.uk/pub/biophysics>) and *GraphPad Prism* 5.0.0.228 (Graph Pad Software). A *t*-test was performed with *GraphPad Prism* to validate the significance of the results.

### 2.5. Cryo-electron microscopy

Samples were diluted to  $1 \text{ mg ml}^{-1}$  and a total of  $3.8 \mu\text{l}$  was applied onto glow-discharged 300 mesh holey gold UltrAuFoil R1.2/1.3 grids (Quantifoil Micro Tools GmbH). Vitrification was conducted using a Vitrobot Mark IV (Thermo Fisher Scientific, Eindhoven, The Netherlands) set to  $10^\circ\text{C}$  and 100% humidity by plunging into liquid ethane after 4 s of blotting.

Data for *SoBDH2* were collected on an FEI Titan Krios G3i transmission electron microscope (Thermo Fisher Scientific, Eindhoven, The Netherlands) operated at 300 kV equipped with a Falcon 3EC at a nominal magnification of  $96\,000\times$ , corresponding to a calibrated pixel size of  $0.832 \text{ \AA}$ . Objective



**Figure 1**

Reaction schemes and enzyme characterization of *SoBDH2* and *SrBDH1*. The BDHs discussed here preferentially convert the (+)-enantiomers of borneol and isoborneol. *SrBDH1* is highly selective for both alcohols and catalyzes the reduction of (+)-camphor. *SoBDH2* is highly selective for borneol. *PsBDH* only shows a slight selectivity for both borneol and isoborneol and is capable of catalyzing the reduction of camphor (Khine *et al.*, 2020). (a, b) Reaction schemes of *SrBDH1* and *SoBDH2* in the enantiospecific oxidation of *rac*-borneol (a) and *rac*-isoborneol (b). (c) Reduction of (+)-camphor. (d) SEC-MALS analysis of *SrBDH1* (blue) and *SoBDH2* (green). For *SrBDH1* a single peak is observed consistent with a tetramer (theoretical molecular mass 120 kDa). The first peak in the chromatogram of *SoBDH2* corresponds to an octamer (theoretical molecular mass 258 kDa) and the second peak corresponds to a tetramer (theoretical molecular mass 129 kDa). The brown (*SrBDH1*) and red (*SoBDH2*) curves are refractive-index signals. (e) Differential scanning fluorometry reveals a significantly lower  $T_m$  for *SoBDH2* ( $55.5 \pm 0.7^\circ\text{C}$ ) compared with *SrBDH1* ( $63.3 \pm 1.3^\circ\text{C}$ ).

astigmatism and coma were corrected with *AutoCTF* (Thermo Fisher Scientific, Eindhoven, The Netherlands) under the final imaging conditions. To maximize beam coherence, a 50  $\mu\text{m}$  C2 aperture was chosen. Direct alignments were executed thoroughly and beam parallelism and condenser astigmatism were optimized using the ronchigram on a Volta phase plate (VPP), which was retracted during data acquisition. During imaging an electron flux of  $0.7\text{ e}^-$  per pixel per second on the detector was selected, corresponding to an exposure rate of  $1\text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$  on the sample. Images were taken at a nominal defocus of between  $-0.6$  and  $-1.6\text{ }\mu\text{m}$ , accumulating a total electron exposure of  $40\text{ e}^- \text{ \AA}^{-2}$  during a 40 s exposure, fractionated into 33 images. For automated data acquisition, *EPU* 2.8.1 (Thermo Fisher Scientific, Eindhoven, The Netherlands) was utilized with aberration-free image shift (AFIS) enabled, allowing  $6\text{ }\mu\text{m}$  image-beam-shift acquisition. The implemented ice filter was adjusted to exclusively image regions with the thinnest ice.

Data for *SrBDH1* were acquired on the same instrument with minor exceptions. The nominal magnification was increased to  $120\,000\times$ , yielding a pixel size of  $0.657\text{ \AA}$ . The electron flux was adjusted to  $0.6\text{ e}^-$  per pixel per second on the detector, resulting in a dose rate of  $1.3\text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$  on the sample. During an exposure time of 31 s, a total dose of  $40\text{ e}^- \text{ \AA}^{-2}$  was applied to the sample.

## 2.6. CryoEM image processing

Raw movies of the *SoBDH2* data set were aligned and dose-weighted with patch-motion correction implemented in *cryoSPARC* version 2.9 (Punjani *et al.*, 2017). Initial CTF estimation was achieved using *Patch CTF*. For initial particle picking, the *Blob Picker* was used with a particle diameter of  $120\text{--}160\text{ \AA}$ . Shiny class averages generated by reference-free 2D classification were selected as templates for template-based particle picking using a  $120\text{ \AA}$  circular mask. A total of  $1\,551\,724$  particle images were extracted with a box size of 224 pixels Fourier-cropped to 56 pixels ( $3.328\text{ \AA}$  per pixel) for initial analysis and subjected to 40 iterations of 2D classification. Shiny classes were selected for *ab initio* reconstruction imposing *D2* symmetry. Heterogeneous refinement with three classes did not guide further classification; therefore, particle images were re-extracted Fourier-cropped to a box size of 112 pixels ( $1.664\text{ \AA}$  per pixel). The best resolved structure after heterogeneous refinement was re-extracted with a box size of 256 pixels ( $0.832\text{ \AA}$  per pixel). Non-uniform (NU) refinement into a single class of  $290\,356$  particles yielded a reconstruction with  $2.32\text{ \AA}$  resolution. Global and local CTF correction did not improve the resolution; however, the reconstruction visually appeared to be better defined. In order to better account for anisotropic motion of the particles, local motion correction was applied followed by global CTF refinement, yielding a reconstruction after NU refinement at  $2.2\text{ \AA}$  resolution. Micrographs with estimated resolutions of worse than  $3.5\text{ \AA}$  were discarded, leaving  $254\,403$  particle images for another cycle of local motion correction followed by global CTF refinement and NU refinement. To account for the point

spread of the signal in the particle images, a box size of 384 pixels ( $320\text{ \AA}$ ) was used for re-extraction, giving a resolution after NU refinement of  $2.1\text{ \AA}$ . Another heterogeneous refinement run was conducted to isolate the final population of  $173\,781$  particle images, which was reconstructed after local motion correction by NU refinement to  $2.0\text{ \AA}$  resolution. In the later NU refinement runs, references were initially filtered to  $20\text{ \AA}$  to retain more structural information in the reference projections, which helped to stabilize refinement. We suspect that the similar appearance of BDH from perpendicular projections of the top view exacerbates the alignment which results in misaligned particles, thus limiting the resolution.

*SrBDH1* was refined similarly, with the exception that choosing the same final box size of 384 pixels resulted in smaller absolute dimensions of the box. From a total of  $1\,587$  micrographs  $1\,635\,690$  particle images were extracted, resulting in  $410\,573$  selected particle images after reference-free 2D classification. After iterative homogeneous and heterogeneous refinement cycles, a final subset of  $210\,505$  particle images were selected, yielding a reconstruction with  $1.88\text{ \AA}$  resolution after NU refinement.

## 2.7. Model building and refinement

An initial model of *SoBDH2* was obtained by automatic model building with *ARP/wARP ARPEM* (version 8.0; Chojnowski *et al.*, 2019) using the protein sequence as input and a sharpened Coulomb potential map. Sharpening was achieved by density modification with *phenix.resolve\_cryo\_em* with default settings using unfiltered, unmasked half-maps and the nominal resolution determined by gold-standard FSC. Sharpening of the *SrBDH1* reconstruction was conducted with *phenix.auto\_sharpen* using default settings starting with unfiltered, unmasked half-maps and the gold-standard resolution as the target resolution. Automatic model building comprised iterative refinement in *REFMAC5* (version 5.8.0258; Murshudov *et al.*, 2011). For comparison, the *phenix.map\_to\_model* procedure (Terwilliger *et al.*, 2018) as well as *Buccaneer* (Hoh *et al.*, 2020) as part of the *CCPEM* suite (Burnley *et al.*, 2017) were used for automated model building. The automated model-building programs were run with the standard settings, since they gave the best results. The obtained model was manually adjusted to the cryoEM density, supported by real-space refinement in *Coot* (version 0.8.9.1; Casañal *et al.*, 2020). The model was refined against the cryoEM map using the real-space refinement protocol in *Phenix* (version 1.19.1; Liebschner *et al.*, 2019; Afonine *et al.*, 2018). Water molecules were added in *Coot* and manually inspected, followed by an additional round of real-space refinement in *Phenix*. In the final stages of refinement, we fully released the restraints for secondary-structure elements, Ramachandran, noncrystallographic symmetry (NCS) and no corrections of energetically disfavored rotamer conformations. In final rounds of refinement, grouped atomic displacement factors were refined. The structures were evaluated with *EMRinger* (Barad *et al.*, 2015) and *MolProbity* (Williams *et al.*, 2018). Structure figures were prepared using *PyMOL*



**Table 1**  
CryoEM data-collection, refinement and validation statistics.

	<i>SrBDH1</i> (PDB entry 7o6q, EMDB entry EMD-12740)	<i>SoBDH2</i> (PDB entry 7o6p, EMDB entry EMD-12739)
Data collection and processing		
Microscope	FEI Titan Krios G3i	FEI Titan Krios G3i
Voltage (keV)	300	300
Camera	Falcon 3EC	Falcon 3EC
Magnification (nominal)	120000	96000
Pixel size at detector (Å per pixel)	0.657	0.832
Total electron exposure (e <sup>-</sup> Å <sup>-2</sup> )	40	40
Exposure rate (e <sup>-</sup> per pixel per second)	0.6	0.7
No. of frames collected during exposure	33	33
Defocus range (µm)	0.60–1.6	0.60–1.6
Automation software	<i>EPU</i> 2.8.1	<i>EPU</i> 2.8.1
No. of micrographs collected	1666	1439
No. of micrographs used	1587	1439
Total No. of extracted particles	1635690	1551724
No. of refined particles	410574	1061307
Final No. of particles	210505	173781
Point-group or helical symmetry parameters	<i>D</i> 2	<i>D</i> 2
Resolution (global) (Å)		
FSC 0.143 (unmasked/masked)	2.1/1.88	2.6/2.04
Resolution range (local) (Å)	1.66–30.00	1.84–9.96
Map-sharpening <i>B</i> factor (Å <sup>2</sup> )	–46	–43
Map-sharpening methods	Local <i>B</i> factor	Local <i>B</i> factor
Refinement package	<i>phenix.real_space_refine</i>	<i>phenix.real_space_refine</i>
Model composition		
Non-H atoms	7999	8276
Protein residues	977	1022
Water molecules	399	268
Model refinement		
Model-map scores		
CC (mask)	0.84	0.86
CC (volume)	0.85	0.83
Average FSC (unmasked/masked)	1.6/1.6	1.7/1.7
Average grouped <i>B</i> factors (Å <sup>2</sup> )		
Overall	18.4	34.9
Protein residues	18.3	36.2
Water	19.5	33.3
R.m.s.d. from ideal values		
Bond lengths (Å)	0.010	0.008
Bond angles (°)	0.708	0.582
Validation		
<i>MolProbity</i> score	1.5	1.7
<i>CaBLAM</i> outliers (%)	0.4	1.5
Clashscore	4.8	6.0
Poor rotamers (%)	1.3	0.9
<i>C</i> <sup>β</sup> deviations	0	0
<i>EMRinger</i> score	7.3	5.8
Ramachandran plot		
Favored (%)	96.8	98.2
Allowed (%)	3.2	1.8
Outliers (%)	0.0	0.0
Ramachandran plot Z-score (r.m.s.d.)		
Overall	–1.61 (0.23)	–0.54 (0.23)
Helix	–1.25 (0.19)	–0.01 (0.22)
Sheet	–0.75 (0.36)	0.38 (0.39)
Loop	–0.58 (0.30)	–0.96 (0.26)

(version 1.8; Schrödinger) and *UCSF Chimera* (Pettersen *et al.*, 2004). Secondary-structure elements were assigned with *DSSP* (Kabsch & Sander, 1983), and *ALSCRIPT* (Barton, 1993) was used for secondary-structure-based sequence alignments. The atomic models have been deposited in the Protein Data Bank (PDB) with the following accession codes: 7o6p for the 2.04 Å resolution structure of *SoBDH2* and 7o6q for the 1.88 Å resolution structure of *SrBDH1*. The cryoEM

maps have been deposited in the Electron Microscopy Data Bank as follows: *SoBDH2*, EMD-12739; *SrBDH1*, EMD-12740.

### 3. Results and discussion

#### 3.1. High-resolution cryoEM structure of *SoBDH2*

*SrBDH1* and *SoBDH2* exhibit 44% sequence identity and 60% sequence similarity. We produced *SoBDH2* with an N-terminal His<sub>6</sub> tag (theoretical molecular mass 32.2 kDa) in *E. coli* and prepared the protein at high purity. Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) revealed two distinct species (Fig. 1*d*) corresponding to an octameric and a tetrameric assembly.

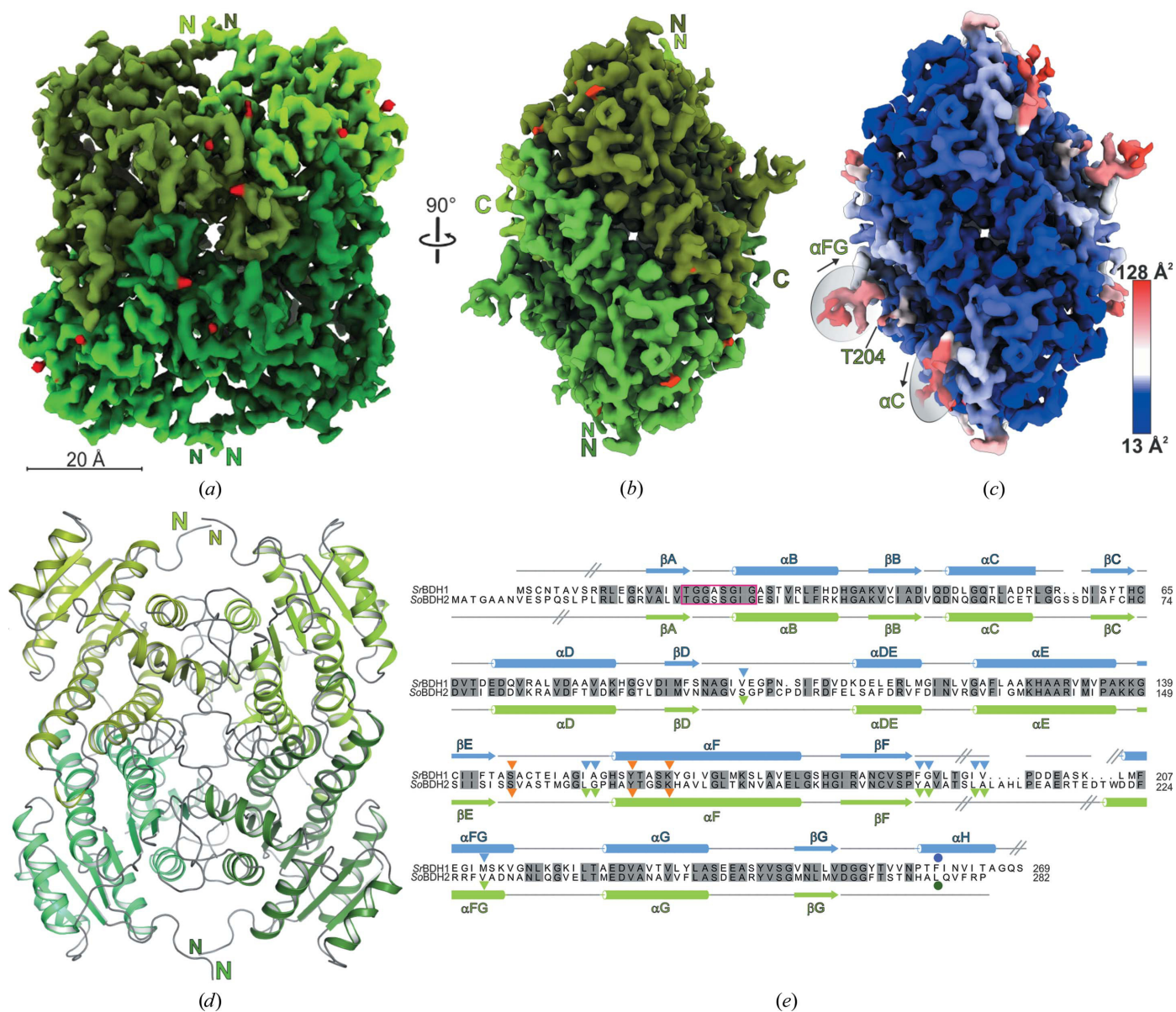
Encouraged by the possible occurrence of an octameric assembly, we considered cryoEM as powerful method to dissect structural heterogeneity, and prepared cryoEM grids. Imaging was conducted on a Titan Krios 300 kV TEM equipped with a Falcon 3EC detector operated in counting mode. We aligned the instrument thoroughly and aimed to maximize the beam coherence by choosing a 50 µm C2 aperture. To optimize the C2 intensity and stigmation, we used the ronchigram method on a Volta phase plate (VPP; Rodenburg & Macak, 2002). The VPP was only used for alignment and was retracted during data acquisition. A total of 1439 micrographs was acquired and subjected to motion correction and CTF estimation. From the 1 551 724 particle images that were initially picked, 173 781 particle images were selected by iterative 2D and 3D classification cycles for homogeneous 3D refinement (Supplementary Figs. S1 and S2). Although we had observed a fraction of octamers in solution (Fig. 1*d*), 3D refinement only yielded a tetrameric

structure (Supplementary Fig. S2); we also failed to detect octamers in negative-stain EM.

After the application of global and local CTF refinement, particle-based local motion correction and NU refinement within the *cryoSPARC* framework (Punjani *et al.*, 2017), a final gold-standard resolution of 2.04 Å was obtained. The obtained cryoEM density reflects the nominal resolution, as individual side chains could be unambiguously identified and built. Given

the high resolution of our cryoEM map (Figs. 2*a* and 2*b* and Table 1), we tested how the automated model-building programs *ARP/wARP* (Chojnowski *et al.*, 2019), *phenix.map\_to\_model* (Terwilliger *et al.*, 2018) and *Buccaneer* (Hoh *et al.*, 2020) would perform. The programs were run with the recommended standard settings and the results are summarized in Supplementary Table S2. All programs managed to fit

large portions of the protein sequence to the density (82–92%), with *ARP/wARP* outperforming the other two programs. We manually completed the initial *ARP/wARP* model. Spherical density regions clearly indicated water molecules, and well defined water molecules were automatically placed with *Coot* (Casañal *et al.*, 2020). The quality of the density allowed the modeling of 50 double conforma-



**Figure 2**  
CryoEM structure of *SoBDH2*. (a) Tetrameric assembly of *SoBDH2*. Density at a contour level of 1.1 is shown for each of the four protomers in different shades of green after sharpening with *Phenix*. The locations of very well defined water molecules are shown in red. (b) The same color-coding as in (a), but rotated by 90°. (c) Grouped *B* factor mapped onto the density. The color gradient is from blue to red corresponding to increasing *B* factors. Regions with high *B* factors are highlighted with gray ellipses and are labeled according to the assigned secondary structure. (d) *SoBDH2* structure in cartoon representation. The same view and color-coding as in (a) is used. (e) Structure-based sequence alignment of *SrBDH1* (GenBank ID MT857224) and *SoBDH2* (GenBank ID MT525099) as obtained by cryoEM. Secondary-structure elements are drawn above the alignment for *SrBDH1* and below the alignment for *SoBDH2*, with  $\alpha$ -helices depicted as cylinders and  $\beta$ -strands as arrows. Gray inclined lines indicate sections of the structures which could not be modeled since they were not resolved in the reconstruction. Orange triangles indicate the catalytic motif. Amino acids lining the putative active site of *SrBDH1*, based on its crystal structure (PDB entry 6zyz) with bound NAD<sup>+</sup>, are indicated by blue triangles and by a dark blue circle if derived from the C-terminal portion of another *SrBDH1* monomer within the tetramer. The dark green circle marks a residue derived from another protomer of *SoBDH2* that completes the active site. Gray-shaded amino acids are identical. The TGXXX(AG)XG NAD<sup>+</sup>-binding motif, between  $\beta$ A and  $\alpha$ B, is indicated with a magenta rectangle.

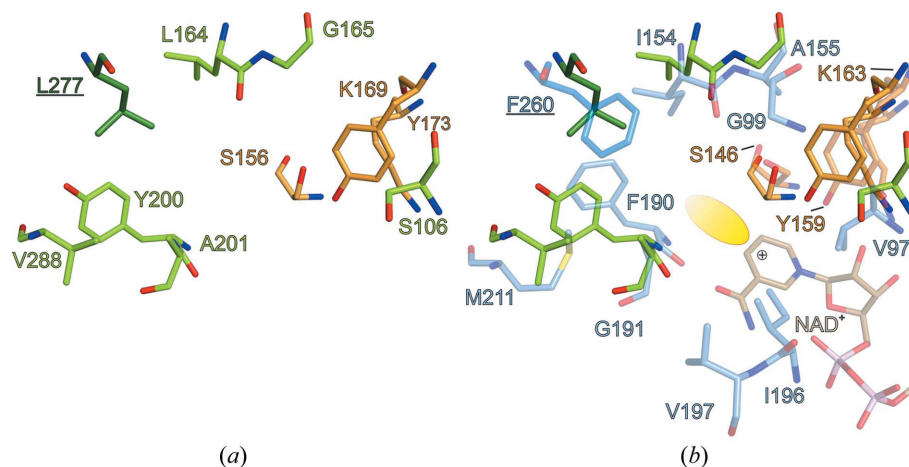


tions of amino-acid side chains and the localization of 268 water molecules. The final model exhibits an excellent fit to the density, with mask/volume correlation coefficients of 0.86/0.83 (Table 1).

As previously observed in the crystal structures of *Sr*BDH1 and *Ps*BDH, the *So*BDH2 homotetramer exhibits *D*<sub>2</sub> symmetry (Fig. 2 and Supplementary Fig. S1). The protomers adopt a Rossmann-like fold (Rossmann *et al.*, 1975) as required for binding of the NAD<sup>+</sup> cofactor (Supplementary Fig. S3). The 12 N-terminal residues and the preceding His<sub>6</sub> tag lack density (Fig. 2*e*). Very weak and fragmented density is observed for *So*BDH2 residues Gln52–Gly65 that fold into  $\alpha$ -helix  $\alpha$ C (Fig. 2*e*), reflected by elevated *B* factors (Fig. 2*c*). Moreover, the region from Ser205 to Glu218 is not resolved in the density and has not been modeled (Fig. 2*e*), which is in agreement with the observation that we could not observe any density for the NAD<sup>+</sup> cofactors in their binding pockets. The latter observation is in agreement with the apo-state crystal structure of *Sr*BDH1. However, the crystal structure of apo *Sr*BDH1 could only be obtained after co-crystallization with the substrate (+)-borneol, which led to the reduction of NAD<sup>+</sup> and the release of product and cofactor. Loss of the cofactor could not be prevented by adding a threefold molar excess of NAD<sup>+</sup> to *So*BDH2 before size-exclusion chromatography. While the loss of NAD<sup>+</sup> may have occurred during vitrification of the cryoEM sample, in the NAD<sup>+</sup>-bound crystal structures of *Sr*BDH1 the cofactor-binding site is stabilized by crystal contacts, suggesting that under the crystallization conditions the NAD<sup>+</sup>-binding site is artificially stabilized to prevent release of the cofactor.

### 3.2. Active site of *So*BDH2

Despite the absence of NAD<sup>+</sup>, the spatial arrangement of the catalytic Ser156, Lys169, Tyr173 motif (Fig. 2*e*) is maintained in *So*BDH2 compared with *Sr*BDH1–NAD<sup>+</sup> (PDB entry 6zyz; Chánique *et al.*, 2021; Fig. 3). The lysine residue, in concert with the positively charged nicotinamide, lowers the p*K*<sub>a</sub> value of the tyrosine, which acts as the catalytic acid/base. The serine residue is involved in stabilization and polarization of the carbonyl function of the substrate (Kavanagh *et al.*, 2008). As in *Sr*BDH1, the substrate-binding niche is very hydrophobic, but is decorated by different amino-acid residues. Moreover, in both enzymes the C-terminus of another protomer completes the active-site pocket (Fig. 3). Notably, the C-terminus of *So*BDH2 adopts a coiled-coil structure, in contrast to the C-terminal  $\alpha$ -helix  $\alpha$ H in *Sr*BDH1 (Supplementary Fig. S7*a*), but both Phe260 of *Sr*BDH1 and Leu277 of *So*BDH2 reside in the same position (Fig. 3*b*).



**Figure 3**

Active-site architecture of *So*BDH2. Residues of the catalytic motif are colored orange and residues of *So*BDH2 lining the active site are colored light green. The substrate-binding pocket is completed by Leu277 (underlined) from the other, neighboring protomer. (a) Substrate-binding pocket of *So*BDH2. Numbering refers to residues of *So*BDH2. (b) Superposition of the cryoEM structure of *So*BDH2 and the crystal structure of *Sr*BDH1–NAD<sup>+</sup> (PDB entry 6zyz; Chánique *et al.*, 2021). Residues of *Sr*BDH1 are drawn in light blue or marine for Phe260 (underlined) from the other, neighboring protomer. The numbering of amino acids refers to *Sr*BDH1. Residues in the equivalent positions to Ile196 and Val197 in *Sr*BDH1–NAD<sup>+</sup> are not resolved in the density of *So*BDH2 due to the absence of NAD<sup>+</sup>. The corresponding residues to the latter two residues in *So*BDH2 are Leu206 and Ala207, respectively. The yellow ellipse indicates the potential substrate-binding site.

Due to fold differences, the active-site architectures of plant BDHs and *Ps*BDH differ drastically (Supplementary Fig. S7*c*). In both *So*BDH2 and *Sr*BDH1 the single  $\alpha$ FG helix flanks the substrate-binding site, while the equivalent region in *Ps*BDH is divided into two discrete helices (Supplementary Fig. S7*c*):  $\alpha$ FG1 and  $\alpha$ FG2. Furthermore, the C-terminus of *Ps*BDH does not contribute to the substrate-binding site. The differences could be related to the natural functions of the enzymes. The bacterial enzyme, in contrast, participates in the degradation of monoterpenols. Development of stereoselectivity in a catabolic dehydrogenase would restrict the substrate scope as some potential substrates can no longer be converted. While catabolic enzymes generally have a broader substrate acceptance than their anabolic counterparts, in this particular case the development of enantiospecificity would preclude the oxidation of both enantiomers of borneol and isoborneol. As both the (+)- and (–)-enantiomers of these two terpenoids are constituents of the essential oils of many plants, it can be argued that the development of stereoselectivity does not provide an evolutionary advantage.

### 3.3. CryoEM structure of *Sr*BDH1

To explore the general applicability of cryoEM to the high-resolution structural analysis of small plant enzymes, we also subjected *Sr*BDH1 to cryoEM-based structure analysis. The *Sr*BDH1 preparation yielded a single peak in a SEC-MALS analysis, consistent with a tetramer in solution, in agreement with its crystal structure (Chánique *et al.*, 2021). As *Sr*BDH1 readily crystallized under various conditions, unlike *So*BDH2, we compared the thermal stabilities of the two proteins by differential scanning fluorometry. Interestingly, the readily

crystallizable *Sr*BDH1 is stabilized by approximately 8°C compared with *So*BDH2 (Fig. 1e).

Cryo-grid preparation for *Sr*BDH1 was performed as for *So*BDH2. To ensure that the resolution would not be limited by the sampling of the detector, we decided to increase the magnification during data acquisition. By picking 1 635 690 particle images from 1666 micrographs, we generated a data set of similar size to that for *So*BDH2. Following the same data-processing routine as used for *So*BDH2 yielded a final *Sr*BDH1 reconstruction at 1.88 Å resolution (Table 1, Supplementary Figs. S4 and S5). This, to the best of our knowledge, is the highest reported resolution of a sub-200 kDa

protein solved by single-particle cryoEM. Remarkably, the resolution of the cryoEM structure of apo *Sr*BDH1 is much higher compared with the best resolved crystal structure of *Sr*BDH1–NAD<sup>+</sup> (PDB entry 6zyz; Chánique *et al.*, 2021), with four bound NAD<sup>+</sup> molecules, at 2.27 Å resolution. As assumed, *Sr*BDH1 is arranged as a tetramer (Fig. 4 and Supplementary Fig. S4). During atomic modeling we followed the same refinement procedure as described for *So*BDH2 with the exception that we used the crystal structure of apo *Sr*BDH1 (PDB entry 6zz0; Chánique *et al.*, 2021) as the starting model. The crystal structure and cryoEM structure are practically identical (Supplementary Table S3). The density is

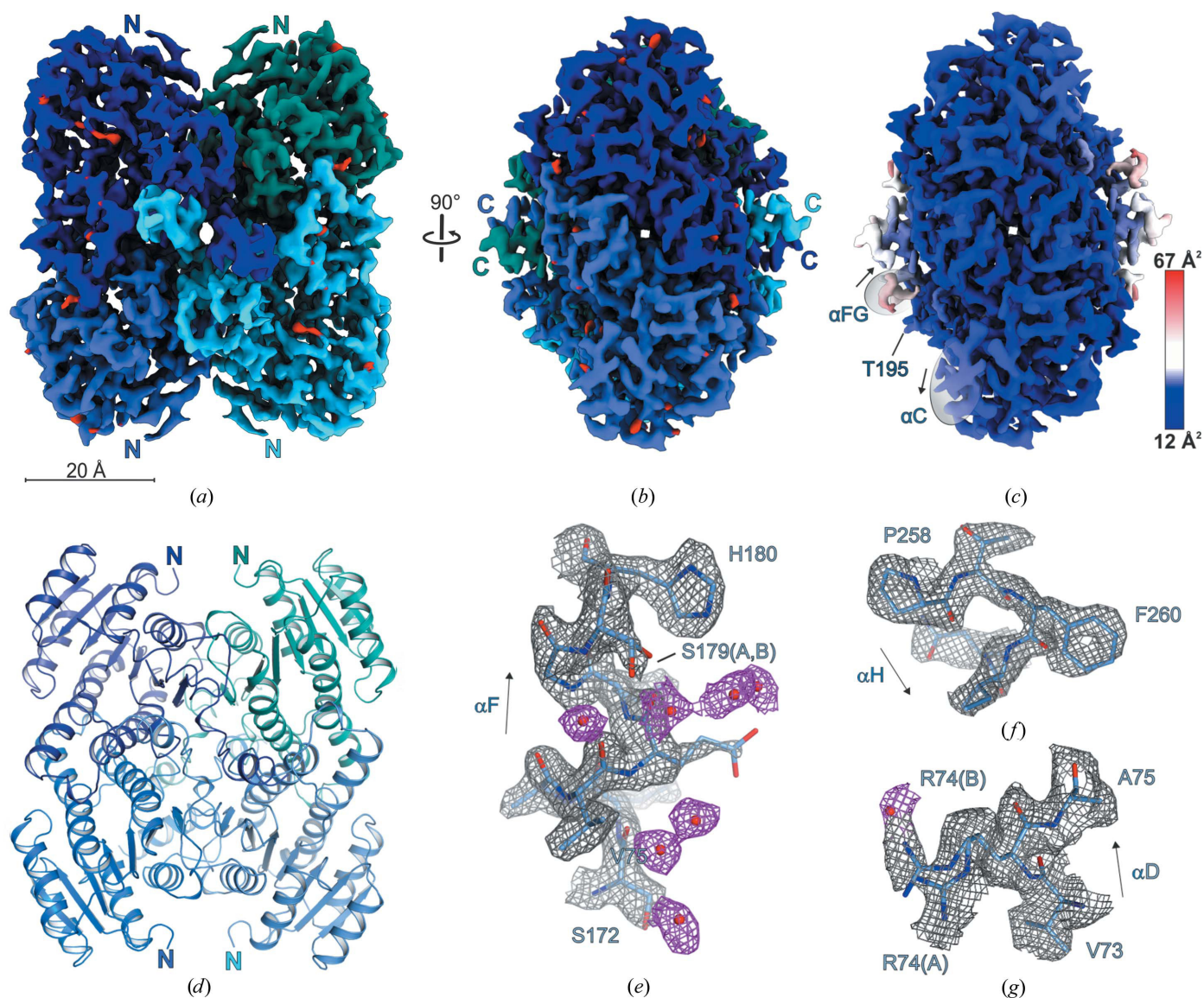


Figure 4

CryoEM structure of *Sr*BDH1. (a) Tetrameric assembly of *Sr*BDH1. Density-modified cryoEM reconstructions at a contour level of 9.5 are shown for each of the four protomers in different shades of blue. Locations of very well defined water molecules are shown in red. (b) The same color-coding as in (a) but rotated by 90°. (c) Grouped *B* factor mapped onto the density. The color gradient is from blue to red corresponding to increasing *B* factors. In contrast to *So*BDH2, the *B*-factor distribution is uniform. (d) The *Sr*BDH1 structure in cartoon representation. The same view and color-coding are used as in (a). (e) Enlargement of the C-terminal end of  $\alpha$ F with an alternate side-chain conformation of Ser179 and well defined water molecules, shown as red spheres. (f) C-terminal end of  $\alpha$ H with Phe260 with a characteristic hole in the density for the aromatic ring system. (g) Double conformation of Arg74. The water molecule is at a distance of 2.5 Å from the guanidinium function of Arg74 in side-chain conformation Arg74(B).

of outstanding quality, allowing the unambiguous assignment of amino-acid side chains in alternate conformations (Fig. 4 and Supplementary Fig. S6) and the placement of water molecules.

Almost the entire protein chain could be traced in the cryoEM map, which is reflected by an exceptional atom inclusion level at the moderate contour level of 0.3 for 97% of all backbone atoms and 93% of all non-H atoms. In addition to the first eight residues and the very C-terminal residue (Fig. 2e), the region from Leu193 to Leu205 is not defined in the density due to the missing NAD<sup>+</sup> cofactor, as in *SoBDH2* (Fig. 2e and Supplementary Fig. S7a). In comparison to the available crystal structures of *SrBDH1*, the total number of built residues is practically identical.

The *SrBDH1* model derived from the cryoEM map is virtually identical to the apo-state crystal structure (r.m.s.d. of 0.6 Å for 982 pairs of C<sup>α</sup> atoms; Supplementary Fig. S7b). At 1.88 Å resolution we could identify 399 water molecules, which uniformly cover the protein surface or are bound in cavities within the protein core. The ratio of water molecules to residues (0.4) is much lower compared with structures determined by X-ray crystallography, where one water molecule per residue is expected at a resolution of 2.0 Å (Carugo & Bordo, 1999). This discrepancy is explained by the absence of solvent channels in cryoEM structures and the missing local proximity of protein molecules. We observed 34 side chains with a double conformation, corresponding to about 3.5% of all residues. The observed ratio is perfectly in line with a detailed study reporting that 3% of residues present alternate side-chain conformations in protein crystal structures with a resolution between 1.0 and 2.0 Å (Miao & Cao, 2016).

Since the number of high-resolution cryoEM structures is limited, we wondered whether the Ramachandran *Z*-scores (Hoofst et al., 1997) of our structures (Table 1) would follow the distribution of Ramachandran *Z* ranges as observed for crystal structures in a similar resolution regime (Sobolev et al., 2020). The Ramachandran *Z*-scores of the *SrBDH1* and *SoBDH2* structures are in the expected region for crystal structures of similar resolution. Notably, we refined the models without Ramachandran restraints, demonstrating that the Ramachandran *Z*-score can also be a valuable measure for cryoEM densities.

#### 4. Summary

Structures of homomultimeric plant enzymes are under-represented in the fast-growing collection of protein structures analyzed by cryoEM. Here, we elucidated the cryoEM structures of two comparatively small plant BDHs to high resolution. Given the molecular mass of the tetrameric complex, here we report the highest resolution achieved by cryoEM so far (Supplementary Table S1), pushing the boundaries of this rapidly developing method.

The new *SoBDH2* structure we describe revealed details of the active-site architecture of the enzyme and allowed comparison to *SrBDH1*. To our surprise, we could not observe NAD<sup>+</sup> in the cryoEM structure of *SrBDH1*, although the

protein samples used for crystallization and cryoEM were identical. A possible explanation for this difference could be that in the crystal the cofactor-binding loop is stabilized by crystal contacts and thus may have trapped NAD<sup>+</sup>. Alternatively, vitrification of the sample for cryoEM may have led to the loss of the cofactor.

We attempted to find an explanation why *SrBDH1*, but not *SoBDH2*, could be crystallized. Firstly, *SrBDH1* has a considerably higher *T<sub>m</sub>* compared with *SoBDH2*, suggesting a higher fold stability that may be more amenable to crystallization. Furthermore, although the cryoEM structures superimpose with an r.m.s.d. of 1.3 Å for 952 pairs of C<sup>α</sup> atoms (Supplementary Fig. S7a), local structural differences might have hindered the crystallization of *SoBDH2*. The αC helix of *SoBDH2* (Gln52–Gly65) is weakly defined in the density and hence is much more flexible compared with that in *SrBDH1* (Figs. 2c, 2e and 3c). Furthermore, in the *SrBDH1* structure the αFG helix, upstream of the unresolved loop region, is stabilized by the C-terminal αH helix via hydrophobic contacts. In contrast, the C-terminus is shorter and is not folded in an α-helix in *SoBDH2* (Figs. 2c, 2e and Supplementary Fig. S7a). Lastly, we cannot rule out that the NAD<sup>+</sup> cofactor might stabilize *SrBDH1* to a larger extent, and its presence might support the crystallization process, which is not the case for *SoBDH2*.

Given the small size of our protein samples and the high particle density on the grids, sufficient data for high-resolution structure analysis could rapidly be acquired, reducing the use of valuable instrument time. Given the high resolution of our structures, model building was greatly facilitated by automated routines, in particular *ARP/wARP ARPEM* (Chojnowski et al., 2019) in combination with iterative refinement cycles in *REFMAC5* (Murshudov et al., 2011). Moreover, due to the small protein size, real-space refinement and validation was fast.

During the past two decades, X-ray crystallography has been the main structural biochemical method to support drug development. Our observation that high-resolution (≤2.0 Å) structures of rather small proteins can be elucidated by cryoEM in a short time emphasizes the important role that cryoEM has to play in future drug-development efforts, for example using high-throughput applications such as fragment-based screening. Apart from circumventing time-consuming crystallization screening and possible phasing problems, an additional considerable advantage of cryoEM in these and other endeavors is a much-reduced sample consumption compared with crystallography. Likewise, our findings show that cryoEM is already an attractive tool for the structural analysis of enzymes used in green industry.

The availability of high-resolution structural data on newly discovered enzymes is crucial for understanding the molecular basis of their catalytic properties. Furthermore, with this knowledge, characteristics such as stability and selectivity can be improved by rational protein engineering instead of the time-consuming random mutagenesis approaches (Jemli et al., 2016). Rational design will greatly facilitate the generation of tailor-made enzymes in relatively short time periods. CryoEM



is a valuable tool to achieve these goals, as it allows the fast and high-resolution structure determination of enzymes that prove difficult to crystallize.

## 5. Related literature

The following references are cited in the supporting information for this article: Cunha *et al.* (2021), Fan *et al.* (2019), Greber *et al.* (2021), Guntupalli *et al.* (2021), Herzik *et al.* (2017, 2019), Kern *et al.* (2021), Krissinel & Henrick (2004), Merk *et al.* (2016, 2020), Munir *et al.* (2021), Nakane *et al.* (2020) and Zhang *et al.* (2019).

## Acknowledgements

We acknowledge technical support from C. Langner in protein purification and support from Y. Huang with the SEC-MALS experiments. The authors would like to thank B. Kirmayer and B. Schade for assistance with cryoEM sample preparation and microscope operation. Open access funding enabled and organized by Projekt DEAL.

## Funding information

The authors acknowledge financial support from the German Federal Ministry of Education and Research (BMBF) for the project CbP – Camphor-based Polymers within the Bioeconomy International Program (grant No. 031B050B). RK and AC would also like to thank the Austrian Science Funds (FWF, P31001-B29) for financial support. CPOH is supported by the Hanns Seidel Foundation. We acknowledge access to electron microscopic equipment at the BioSupraMol core facility of Freie Universität Berlin, supported through grants from the Deutsche Forschungsgemeinschaft (HA 2549/15-2) and from the Deutsche Forschungsgemeinschaft and the state of Berlin for large equipment according to Art. 91b GG (INST 335/588-1 FUGG, INST 335/589-1 FUGG and INST 335/ 590-1 FUGG).

## References

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A. & Adams, P. D. (2018). *Acta Cryst.* **D74**, 531–544.

Barad, B. A., Echols, N., Wang, R. Y.-R., Cheng, Y., DiMaio, F., Adams, P. D. & Fraser, J. S. (2015). *Nat. Methods*, **12**, 943–946.

Barton, G. J. (1993). *Protein Eng. Des. Sel.* **6**, 37–40.

Bohlmann, J. & Keeling, C. I. (2008). *Plant J.* **54**, 656–669.

Burnley, T., Palmer, C. M. & Winn, M. (2017). *Acta Cryst.* **D73**, 469–477.

Carugo, O. & Bordo, D. (1999). *Acta Cryst.* **D55**, 479–483.

Casañal, A., Lohkamp, B. & Emsley, P. (2020). *Protein Sci.* **29**, 1069–1078.

Chánique, A., Dimos, N., Drienovská, I., Calderini, E., Pantín, M. P., Helmer, C. P. O., Hofer, M., Sieber, V., Parra, L. P., Loll, B. & Kourist, R. (2021). *ChemCatChem*, **13**, 2262–2277.

Cheng, C., Liu, X.-W., Du, F.-F., Li, M.-J., Xu, F., Wang, F.-Q., Liu, Y., Li, C. & Sun, Y. (2013). *Acta Pharmacol. Sin.* **34**, 1337–1348.

Chojnowski, G., Pereira, J. & Lamzin, V. S. (2019). *Acta Cryst.* **D75**, 753–763.

Christianson, D. W. (2017). *Chem. Rev.* **117**, 11570–11648.

Croteau, R., Hooper, C. L. & Felton, M. (1978). *Arch. Biochem. Biophys.* **188**, 182–193.

Cunha, E. S., Chen, X., Sanz-Gaitero, M., Mills, D. J. & Luecke, H. (2021). *Nat. Commun.* **12**, 230.

Drienovská, I., Kolanović, D., Chánique, A., Sieber, V., Hofer, M. & Kourist, R. (2020). *Phytochemistry*, **172**, 112227.

Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J. C., Zhao, L., Peng, H. L., Lei, J. & Wang, H. W. (2019). *Nat. Commun.* **10**, 2386.

Greber, B. J., Remis, J., Ali, S. & Nogales, E. (2021). *Biophys. J.* **120**, 677–686.

Guntupalli, S. R., Li, Z., Chang, L., Plapp, B. V. & Subramanian, R. (2021). *Biochemistry*, **60**, 663–677.

Herzik, M. A. Jr, Wu, M. & Lander, G. C. (2017). *Nat. Methods*, **14**, 1075–1078.

Herzik, M. A. Jr, Wu, M. & Lander, G. C. (2019). *Nat. Commun.* **10**, 1032.

Hoh, S. W., Burnley, T. & Cowtan, K. (2020). *Acta Cryst.* **D76**, 531–541.

Hooft, R. W. W., Sander, C. & Vriend, G. (1997). *Comput. Appl. Biosci.* **13**, 425–430.

Jemli, S., Ayadi-Zouari, D., Hlima, H. B. & Bejar, S. (2016). *Crit. Rev. Biotechnol.* **36**, 246–258.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kallberg, Y., Oppermann, U., Jörnvall, H. & Persson, B. (2002). *Eur. J. Biochem.* **269**, 4409–4417.

Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. (2008). *Cell. Mol. Life Sci.* **65**, 3895–3906.

Kemper, K., Hirte, M., Reinbold, M., Fuchs, M. & Brück, T. (2017). *Beilstein J. Org. Chem.* **13**, 845–854.

Kern, D. M., Sorum, B., Mali, S. S., Hoel, C. M., Sridharan, S., Remis, J. P., Toso, D. B., Kotecha, A., Bautista, D. M. & Brohawn, S. G. (2021). *Nat. Struct. Mol. Biol.* **28**, 573–582.

Khine, A. A., Chen, H.-P., Huang, K.-F. & Ko, T.-P. (2020). *Acta Cryst.* **F76**, 309–313.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.

Kühlbrandt, W. (2014). *Science*, **343**, 1443–1444.

Ladenstein, R., Winberg, J. O. & Benach, J. (2008). *Cell. Mol. Life Sci.* **65**, 3918–3935.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.

Lyumkis, D. (2019). *J. Biol. Chem.* **294**, 5181–5197.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J. L. S. & Subramaniam, S. (2016). *Cell*, **165**, 1698–1707.

Merk, A., Fukumura, T., Zhu, X., Darling, J. E., Grishammer, R., Ognjenovic, J. & Subramaniam, S. (2020). *IUCrJ*, **7**, 639–643.

Miao, Z. & Cao, Y. (2016). *Sci. Rep.* **6**, 37024.

Munir, A., Wilson, M. T., Hardwick, S. W., Chirgadze, D. Y., Worrall, J. A. R., Blundell, T. L. & Chaplin, A. K. (2021). *Structure*, **29**, 899–912.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., Grigoras, I. T., Malinauskaitė, L., Malinauskas, T., Miehl, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., Chirgadze, D. Y., Murshudov, G., Aricescu, A. R. & Scheres, S. H. W. (2020). *Nature*, **587**, 152–156.

Newman, D. J. & Cragg, G. M. (2016). *J. Nat. Prod.* **79**, 629–661.

Oldfield, E. & Lin, F.-Y. (2012). *Angew. Chem. Int. Ed.* **51**, 1124–1137.

- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. (2017). *Nat. Methods*, **14**, 290–296.
- Rinkel, J., Steiner, S. T. & Dickschat, J. S. (2019). *Angew. Chem. Int. Ed.* **58**, 9230–9233.
- Rodenburg, M. & Macak, E. B. (2002). *Microsc. Anal.* **90**, 5–7.
- Rossmann, M. G., Liljas, A., Brändén, C.-I. & Banaszak, L. J. (1975). *The Enzymes*, edited by P. D. Boyer, pp. 61–102. New York: Academic Press.
- Sobolev, O. V., Afonine, P. V., Moriarty, N. W., Hekkelman, M. L., Joosten, R. P., Perrakis, A. & Adams, P. D. (2020). *Structure*, **28**, 1249–1258.
- Studier, F. W. (2018). *Curr. Protoc. Mol. Biol.* **124**, e63.
- Terwilliger, T. C., Adams, P. D., Afonine, P. V. & Sobolev, O. V. (2018). *Nat. Methods*, **15**, 905–908.
- Vinothkumar, K. R. & Henderson, R. (2016). *Q. Rev. Biophys.* **49**, e13.
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, J. S. (2018). *Protein Sci.* **27**, 293–315.
- Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. (2020). *Nature*, **587**, 157–161.
- Zhang, K. M., Zhang, H. W., Li, S. S., Pintilie, G. D., Mou, T.-C., Gao, Y. Z., Zhang, Q. F., van den Bedem, H., Schmid, M. F., Au, S. W. N. & Chiu, W. (2019). *Proc. Natl Acad. Sci. USA*, **116**, 6800–6805.



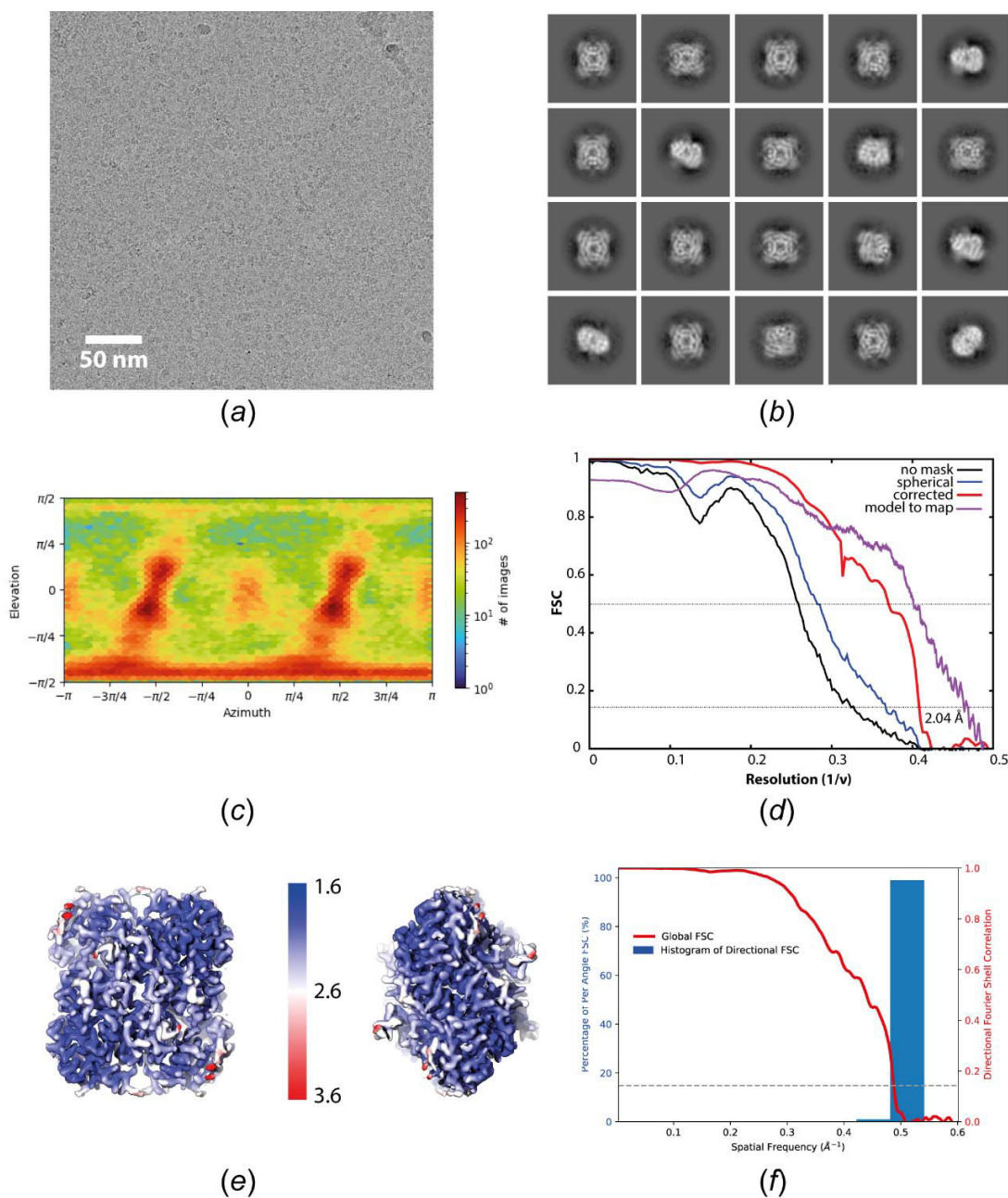
STRUCTURAL  
BIOLOGY

**Volume 78 (2022)**

**Supporting information for article:**

**CryoEM analysis of small plant biocatalysts at sub-2 Å resolution**

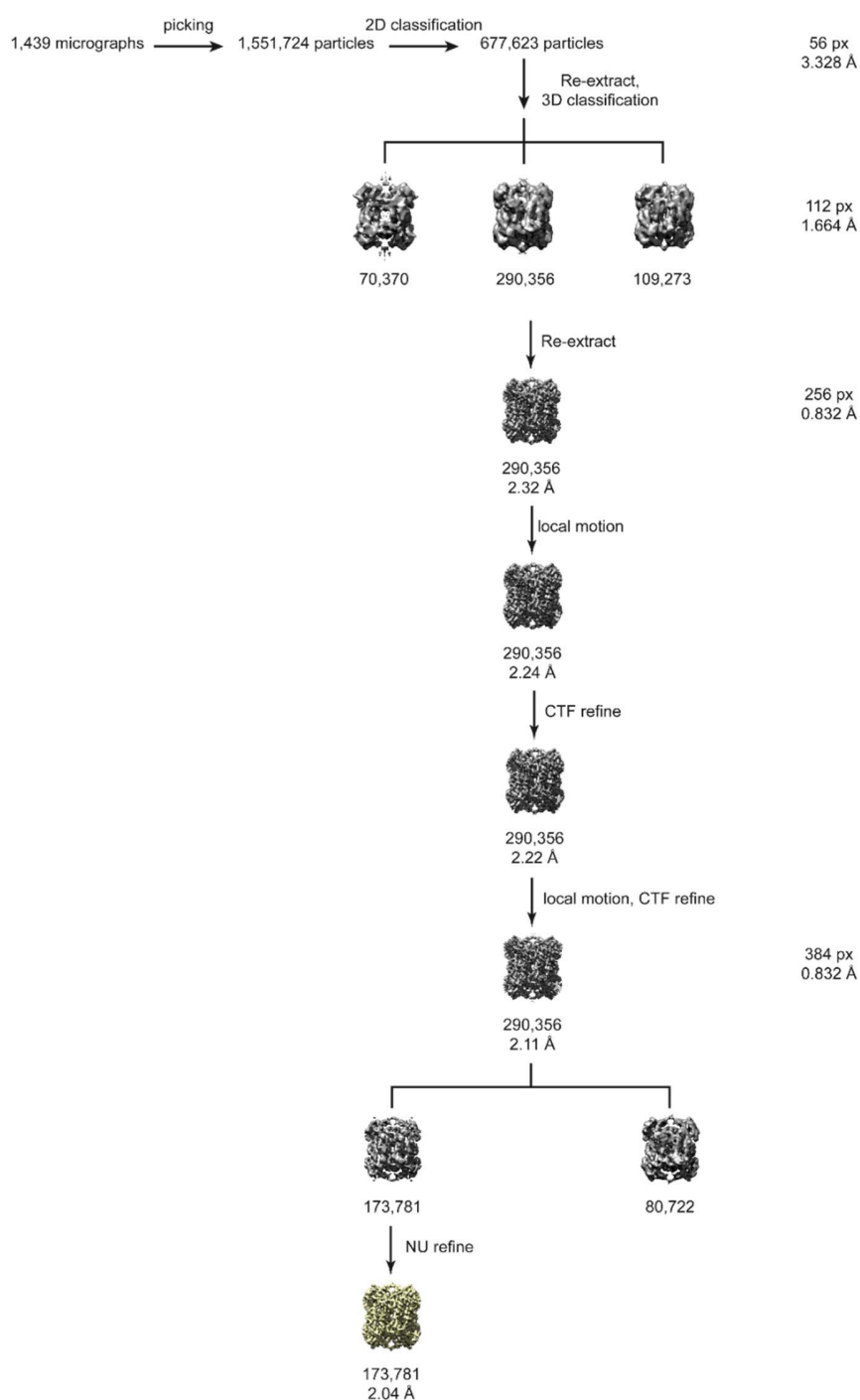
**Nicole Dimos, Carl P. O. Helmer, Andrea M. Chánique, Markus C. Wahl, Robert Kourist, Tarek Hilal and Bernhard Loll**



**Figure S1** CryoEM analysis of *SoBDH2*. (a) representative cryoEM micrograph, the scale bar indicates 50 nm spacing. (b) selected 2D class averages after reference-free 2D classification with cryoSPARC. Top and side views can be identified, excluding preferential orientation issues. A circular mask of 120 Å diameter was used during classification. (c) Viewing direction distribution as determined during non-uniform refinement with cryoSPARC. (d) Resolution estimates by fourier-shell correlation using either no mask (black line), a generous spherical mask (blue line) and after solvent correction by phase

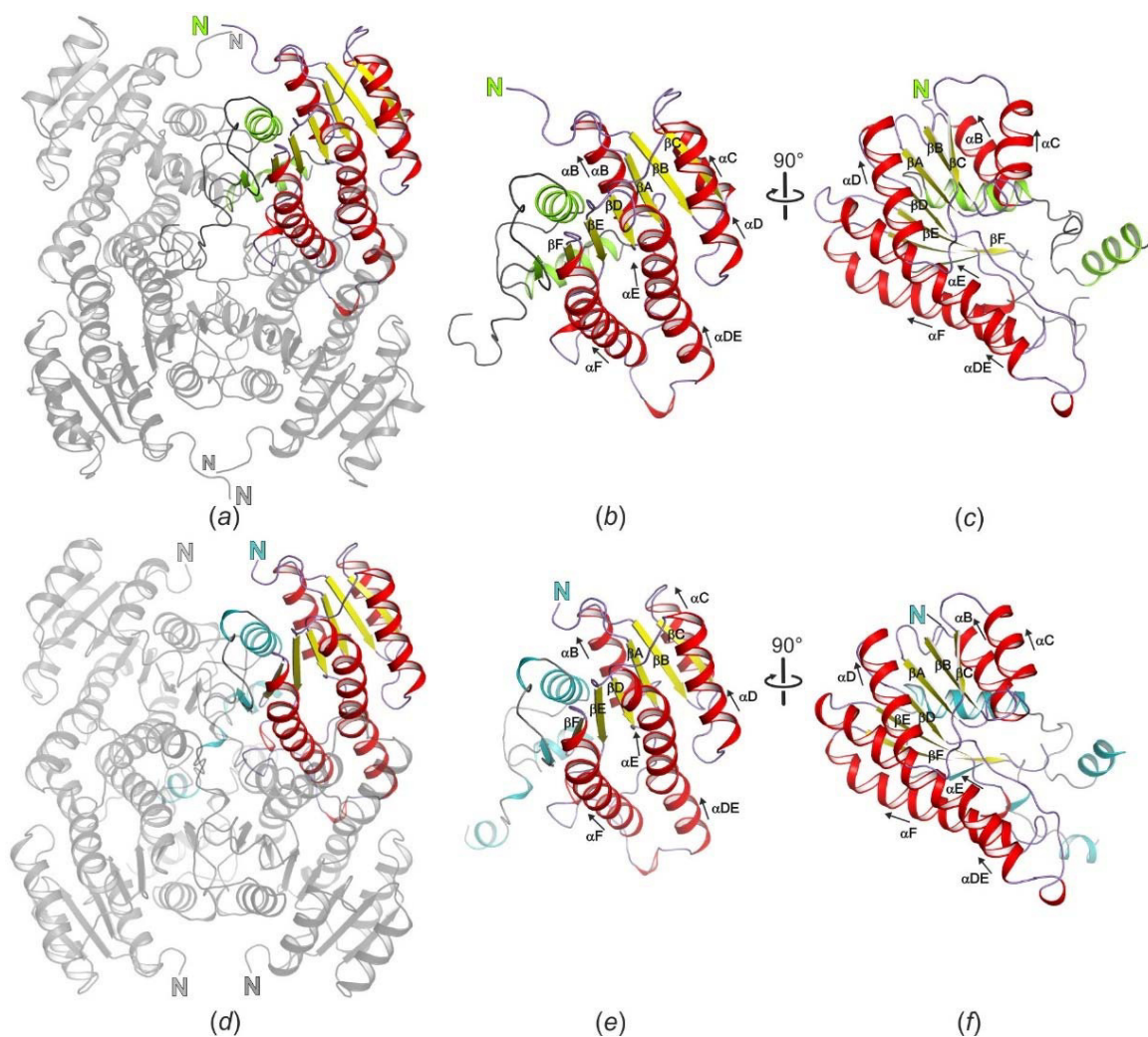
randomization (red line). Dashed lines represent FSC(0.5) and FSC(0.143) crossings. Model to map correlation as determined with PHENIX is colored purple. (e) Illustration of the local resolution estimation calculated with cryoSPARC for two different views of *SoBDH2* after rotation by 90°. Coloring of the cryoEM density reflects the local resolution ranging from 1.6 Å to 3.6 Å. A major fraction of the structure is resolved well beyond 2 Å, less resolved regions are mainly situated in the periphery of *SoBDH2*. (f) results from 3DFSC calculations show an overall good agreement of the directional FSC with the global FSC.



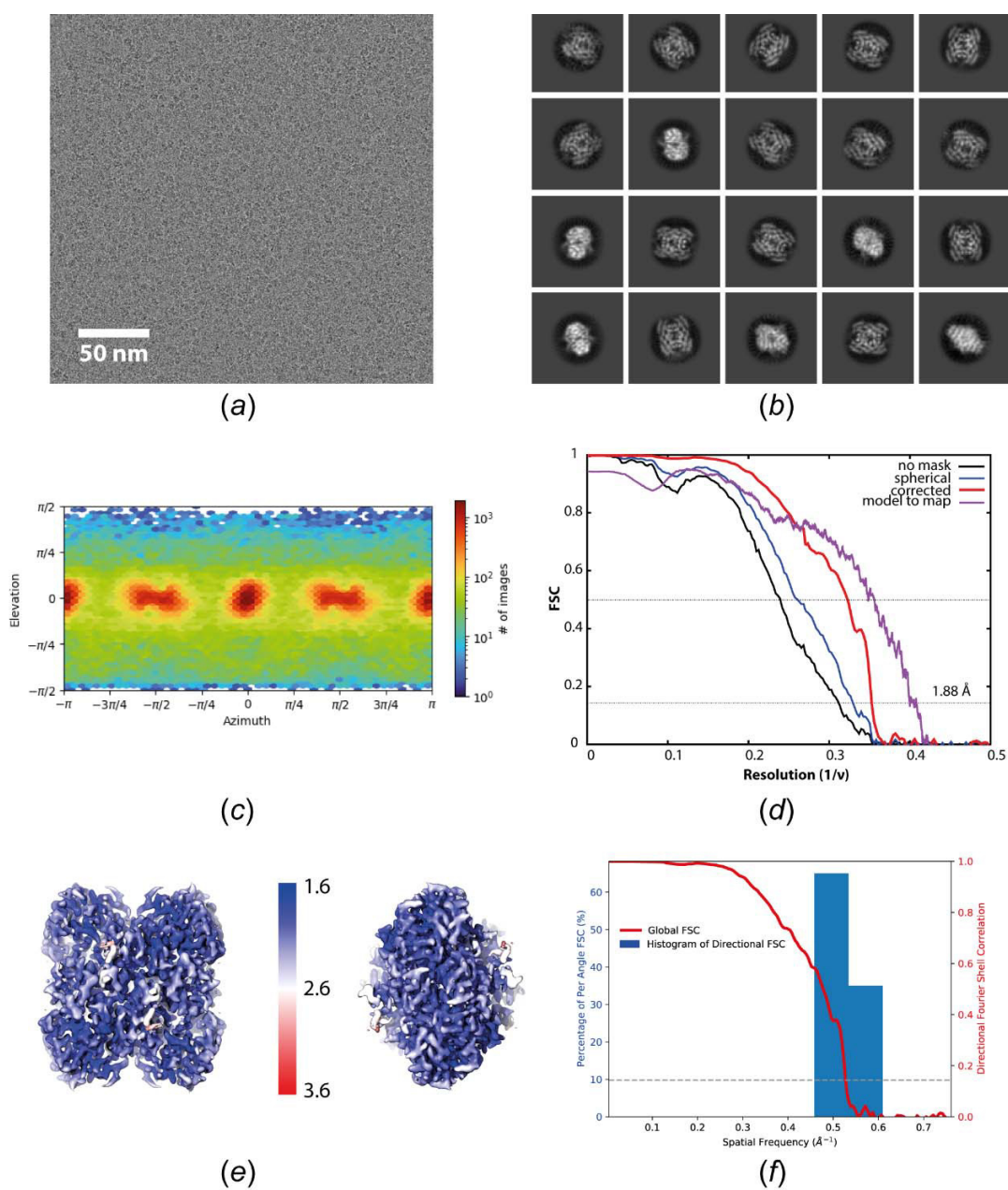


**Figure S2** Data processing workflow for the *SoBDH2* dataset. From initially selected 1,439 micrographs ~1.5 M particles were picked and subjected to reference-free 2D classification. ~678k particle images were re-extracted with a box-size of 224 px Fourier-cropped to 112 px giving a pixel-size of 1.664 Å. After 3D classification, a subset of 290,356 particle images was again re-extracted at full resolution with a larger box size of 256 px and homogeneously refined to 2.32 Å resolution. Particle

based local motion correction improved the resolution to 2.24 Å, which could be only marginally improved to 2.22 Å by CTF refinement. Another cycle of local motion correction was applied using a larger extraction box of 384 px to preserve high frequency information of the CTF. Following CTF refinement the resolution improved to 2.11 Å. By heterogeneous refinement, a final subset of 173,781 particle images was selected for homogeneous NU refinement, yielding the final reconstruction at 2.04 Å resolution.

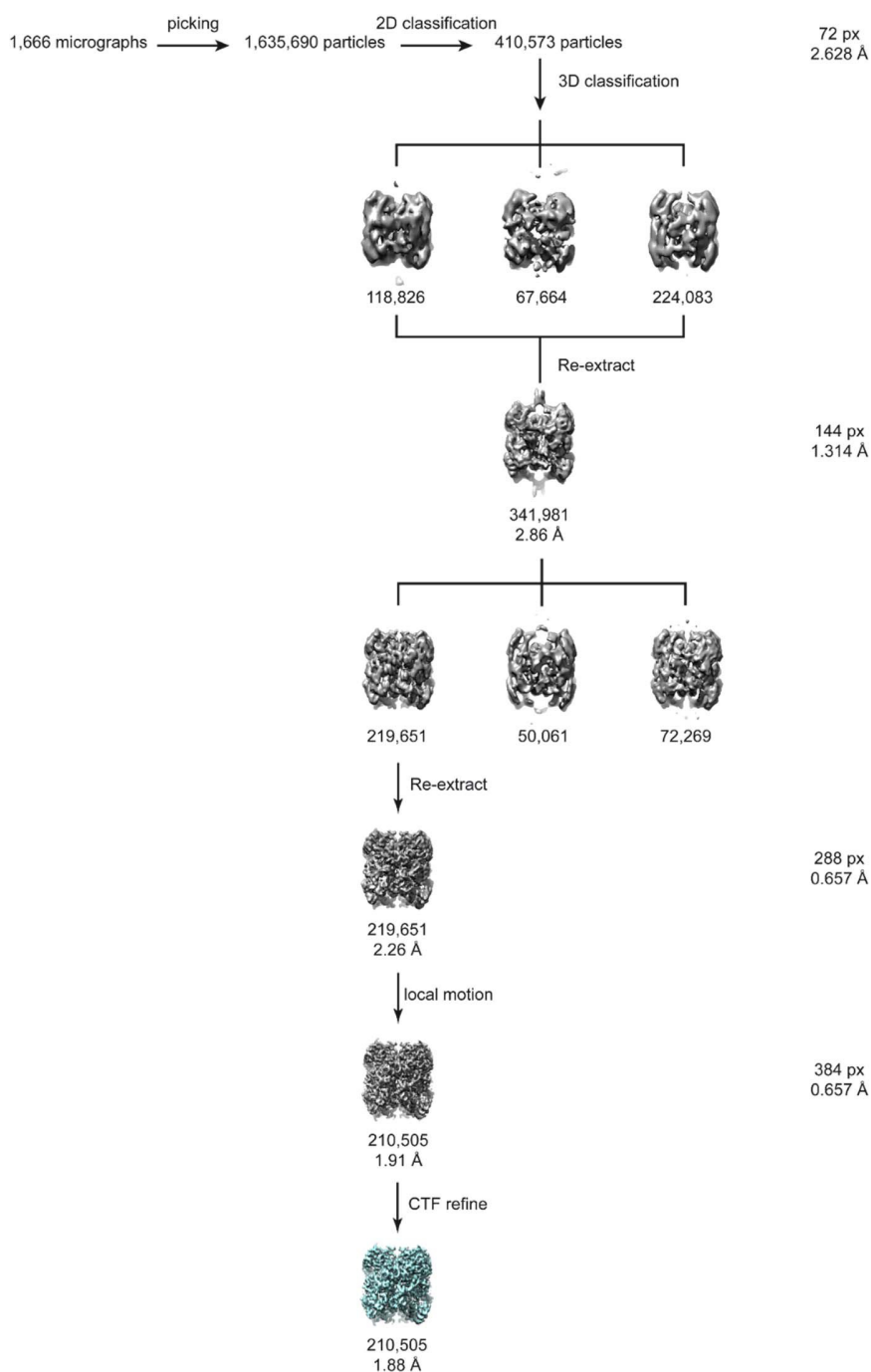


**Figure S3** The Rossmann fold in BDHs. (a) Same view as in Figure 1a. Tetrameric architecture of *So*BDH2 with three protomers in grey. In one monomer the Rossmann fold is highlighted with yellow  $\beta$ -strands, red  $\alpha$ -helices, and purple loop regions. The remaining structure is colored in green and loop regions in gray. (b) Identical view as in panel (a), zoom on one protomer. (c) view of (b) rotated by  $90^\circ$ . (d) Same view as in Figure 3a. Tetrameric architecture of *Sr*BDH1 with three protomers in grey. In one monomer the Rossmann fold is highlighted with yellow  $\beta$ -strands, red  $\alpha$ -helices, and purple loop regions. The remaining structure is colored in teal and loop regions in gray. (e) Identical view as in panel (a), zoom on one protomer. (f) view of (e) rotated by  $90^\circ$ .



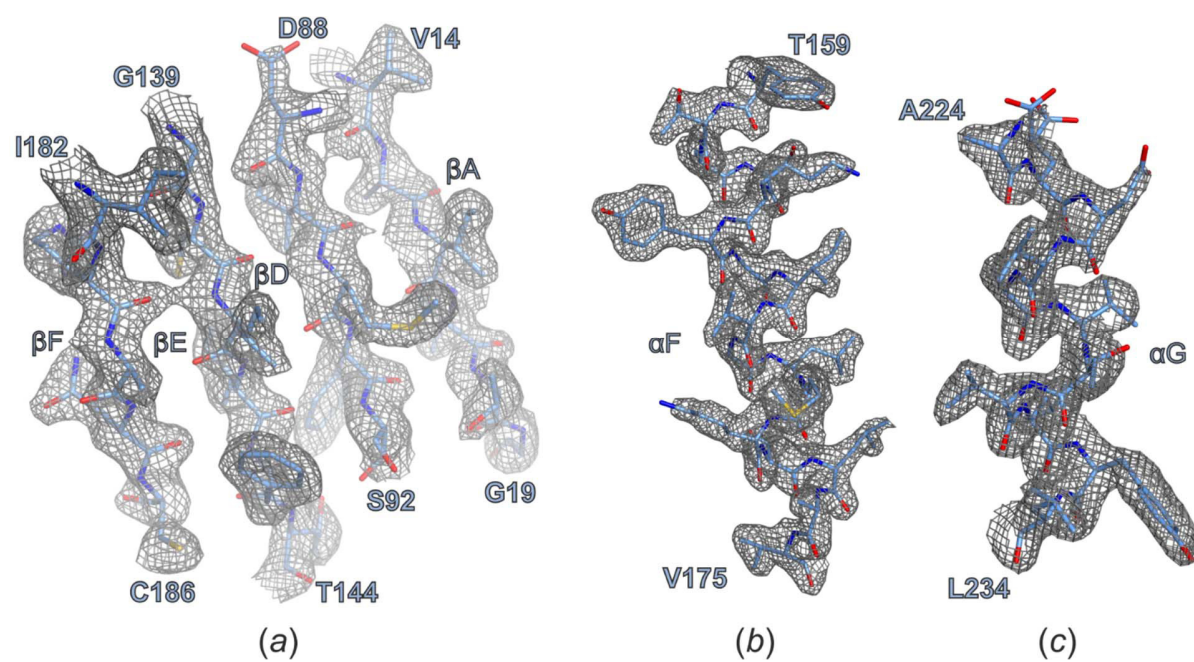
**Figure S4** CryoEM analysis of *SrBDH1*. (a) Representative cryoEM micrograph, the scale bar indicates 50 nm spacing. (b) selected 2D class averages after reference-free 2D classification with cryoSPARC. As for *SoBDH2*, top and side views can be identified. A circular mask of 100 Å diameter was used during classification. (c) Viewing direction distribution as determined during non-uniform refinement with cryoSPARC. (d) Resolution estimates by Fourier-shell correlation using either no mask (black line), a generous spherical mask (blue line) and after solvent correction by phase randomization (red line). (e) 3D surface and ribbon models of *SrBDH1* with a color scale for resolution from 1.6 Å (blue) to 3.6 Å (red). (f) Plot of Percentage of Per Angle FSC (%) and Histogram of Directional FSC vs Spatial Frequency ( $\text{\AA}^{-1}$ ), with a red line for Global FSC and a blue histogram for Directional FSC, and a secondary y-axis for Directional Fourier Shell Correlation.

Dashed lines represent FSC(0.5) and FSC(0.143) crossings. Model to map correlation as determined with PHENIX is colored purple. (e) Illustration of the local resolution estimation calculated with cryoSPARC for two different views of *Sr*BDH1 after rotation by 90°. Coloring of the cryoEM density reflects the local resolution ranging from 1.6 to 3.6 Å. The vast majority of the structure is resolved well beyond 2 Å. (f) 3DFSC calculations confirm that the global FSC falls in between the only two existing bins of directional FSCs. The directional resolution anisotropy did not result in obvious peculiarities of the reconstruction.



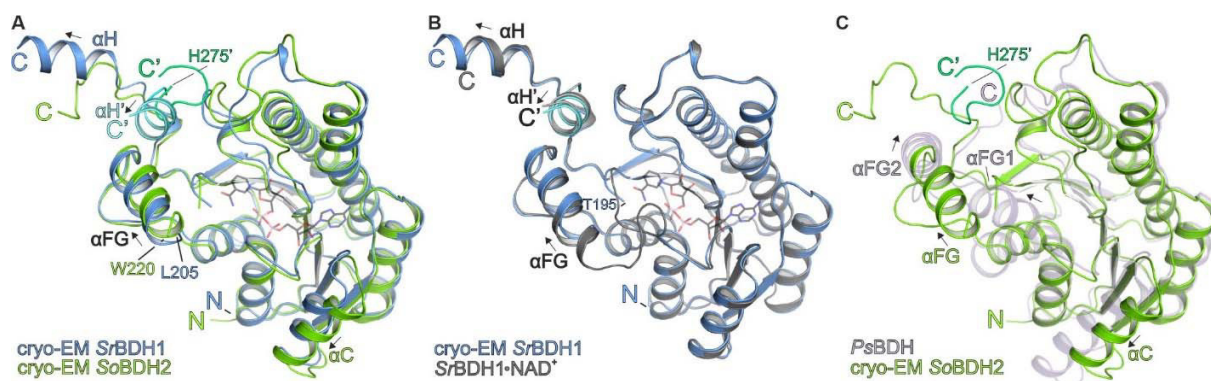
**Figure S5** Data analysis of the *SrBDH1* dataset. Using the *SoBDH2* structure as reference, ~1.6 M particles were automatically picked from 1,666 micrographs with cryoSPARC. Iterations of 2D classification were applied to select 410,573 particle images for heterogeneous 3D classification. A subset of 341,981 was re-extracted with a box-size of 288 px, fourier-cropped to 144 px and homogeneously refined to 2.86 Å resolution. After another heterogeneous refinement using 3 classes, 219,651 particles were re-extracted at full resolution (0.657 Å/pix) yielding a reconstruction of 2.16 Å. Local motion

correction was applied, after which 210,505 particles were re-extracted with a box size of 384 px and homogeneously refined to 1.91 Å resolution. CTF refinement followed by NU refinement generated the final reconstruction with 1.88 Å resolution (cyan).



**Figure S6** Examples of the high-quality electron volumes of *SrBDH1*. (a) Zoom on the central  $\beta$ -sheet. (b) and (c) examples of two  $\alpha$ -helices.





**Figure S7** Structural comparison of related BDH structures focusing on the substrate/cofactor binding site. Only one protomer of the tetrameric complexes is shown. The proteins are shown in cartoon representation. (a) Superposition of the cryoEM structures of *SrBDH1*, drawn in blue, as well as *SoBDH2*, drawn in green cartoon. The C-terminal  $\alpha\text{H}$  helix of another protomer completes the substrate binding site. The  $\text{NAD}^+$  molecule is drawn in black, obtained by a superposition with the crystal structure of *SrBDH1*• $\text{NAD}^+$  / PO/OH (PDB ID 6ZYZ) (Chánique *et al.*, 2021). Structural differences can be seen in particular for the C-terminus and the  $\alpha\text{C}$  helix. (b) Superposition of the cryoEM structure of *SrBDH1* drawn in blue and the crystal structure of *SrBDH1*• $\text{NAD}^+$  / PO/OH (PDB ID 6ZYZ (Chánique *et al.*, 2021)) drawn in gray. Binding of  $\text{NAD}^+$  leads to stabilization of the loop region upstream of helix  $\alpha\text{FG}$  and the helix itself. (c) Superposition of the cryoEM structure of *SoBDH2* and the crystal structure of *PsBDH* PDB ID 6M5N (Khine *et al.*, 2020)) shown in light purple. Major structural differences are observed for the C-terminal portion of the protein.

**Table S1** Overview of selected cryoEM structures with the highest achieved resolution.

The summary does not include larger multi-subunit complexes. Membrane proteins incorporated in nanodiscs are indicated with an asterisk.

enzyme	organism	assembly	resolution [Å]	total M <sub>r</sub> [kDa]	EMDB ID	reference
apoferritin	<i>H. sapiens</i>	24-mer	1.15	480	11668	(Yip <i>et al.</i> , 2020)
β <sub>3</sub> GABAA receptor*	<i>H. sapiens</i>	pentamer	1.7	200	11657	(Nakane <i>et al.</i> , 2020)
β-galactosidase	<i>E. coli</i>	tetramer	1.8	465	21995	(Merk <i>et al.</i> , 2020)
BDH1	<i>S. rosmarinus</i>	tetramer	1.88	120	12740	This study
urease	<i>H. pylori</i>	dodecamer	2.04	1100	11233	(Cunha <i>et al.</i> , 2021)
BDH2	<i>S. officinalis</i>	tetramer	2.04	129	12739	This study
ORF3a/apolipo*	SARS-CoV-2	tetramer	2.08	114	22898	(Kern <i>et al.</i> , 2021)
CDK-activating kinase	<i>H. sapiens</i>	dimer	2.51	119	12042	(Greber <i>et al.</i> , 2021)
aldolase	<i>O. cuniculus</i>	tetramer	2.6	150	8743	(Herzik <i>et al.</i> , 2017)
catalase- peroxidase	<i>M. tuberculosis</i>	dimer	2.68	161	11776	(Munir <i>et al.</i> , 2021)
alcohol dehydrogenase	<i>S. carlsbergensis</i>	tetramer	2.7	147	22807	(Guntupalli <i>et al.</i> , 2021)
methemoglobin	<i>H. sapiens</i>	tetramer	2.8	64	0407	(Herzik <i>et al.</i> , 2019)
lactate dehydrogenase	<i>G. gallus</i>	tetramer	2.8	144	8191	(Merk <i>et al.</i> , 2016)
alcohol dehydrogenase	<i>E. caballus</i>	dimer	2.9	82	0406	(Herzik <i>et al.</i> , 2019)
biotin-bound streptavidin	<i>S. avidinii</i>	tetramer	3.2	52	0689	(Fan <i>et al.</i> , 2019)
cytotoxin A	<i>H. pylori</i>	hexameric	3.2	530	0542	(Zhang <i>et al.</i> , 2019)

isocitrate dehydrogenase	<i>H. sapiens</i>	dimer	3.8	93	8193	(Merk <i>et al.</i> , 2016)
catalytic subunit protein kinase A	<i>M. musculus</i>	monomer	6.0	43	0409	(Herzik <i>et al.</i> , 2019)

---

**Table S2** Comparison of the performance of different automated model building programs.

ARP/wARP – ARPEM(Chojnowski *et al.*, 2019), phenix.map\_to\_model(Terwilliger *et al.*, 2018), as well as Buccaneer(Hoh *et al.*, 2020). Green numbers refer to the structure of *SoBDH2* and blue numbers to *SrBDH1*

	ARP/wARP	Phenix	CCPEM	Final model
	ARPEM	map_to_model	Buccaneer	
Total number of residues		1212 / 1160		
Residues built	1001 / 963	880 / 828	1159 / 972	1022 / 977
Residues sequenced	922 / 921	880 /	1047 / 955	1022 / 977
Completeness by residues built [%]	82.0 / 83.0	72.6 /	91.5 / 83.8	84.5 / 84.2

**Table S3** Structural comparison of the three different crystal structures of *Sr*BDH1.

*Sr*BDH1 apo (PDB ID 6ZZ0), *Sr*BDH1•NAD<sup>+</sup> / high salt (PDB ID 6ZZT) with one bound NAD<sup>+</sup> as well as *Sr*BDH1•NAD<sup>+</sup> / PO/OH (PDB ID 6ZYZ) with four bound NAD<sup>+</sup> molecules (Chánique *et al.*, 2021) as well as the two cryoEM structures of *Sr*BDH1 and *So*BDH2. R.m.s.d. for pairs of C $\alpha$ -atoms calculated with SSM (Krissinel & Henrick, 2004) as implemented in COOT (Casañal *et al.*, 2020).

	<i>Sr</i> BDH1 apo	<i>Sr</i> BDH1•NAD <sup>+</sup> / high salt	<i>Sr</i> BDH1•NAD <sup>+</sup> / PO/OH	cryoEM <i>Sr</i> BDH1	cryoEM <i>So</i> BDH2
<i>Sr</i> BDH1 apo					
<i>Sr</i> BDH1•NAD <sup>+</sup> / high salt	0.48				
<i>Sr</i> BDH1•NAD <sup>+</sup> / PO/OH	0.56	0.52			
cryoEM <i>Sr</i> BDH1	0.57	0.49	0.39		
cryoEM <i>So</i> BDH2	1.44	1.39	1.39	1.15	

## References

- Casañal, A., Lohkamp, B. & Emsley, P. (2020). *Protein Sci.* **29**, 1069-1078.
- Chánique, A., Dimos, N., Drienovska, I., Calderinia, E., Pantína, M. P., Helmer, C. P. O., Hilal, T., Hofer, M., Sieber, V., Parraf, L. P., Loll, B. & Kourist, R. (2021). *ChemCatChem* **13**, 2262-2277.
- Chojnowski, G., Pereira, J. & Lamzin, V. S. (2019). *Acta Crystallogr D Struct Biol* **75**, 753-763.
- Cunha, E. S., Chen, X., Sanz-Gaitero, M., Mills, D. J. & Luecke, H. (2021). *Nat Commun* **12**, 230.
- Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J. C., Zhao, L., Peng, H. L., Lei, J. & Wang, H. W. (2019). *Nat Commun* **10**, 2386.
- Greber, B. J., Remis, J., Ali, S. & Nogales, E. (2021). *Biophys J.*
- Guntupalli, S. R., Li, Z., Chang, L., Plapp, B. V. & Subramanian, R. (2021). *Biochemistry.*
- Herzik, M. A., Jr., Wu, M. & Lander, G. C. (2017). *Nat Methods* **14**, 1075-1078.
- Herzik, M. A., Jr., Wu, M. & Lander, G. C. (2019). *Nat Commun* **10**, 1032.
- Hoh, S. W., Burnley, T. & Cowtan, K. (2020). *Acta Crystallogr D Struct Biol* **76**, 531-541.
- Kern, D. M., Sorum, B., Hoel, C. M., Sridharan, S., Remis, J. P., Toso, D. B. & Brohawn, S. G. (2020). Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs, bioRxiv.
- Khine, A. A., Chen, H. P., Huang, K. F. & Ko, T. P. (2020). *Acta Crystallogr F Struct Biol Commun* **76**, 309-313.
- Krissinel, E. & Henrick, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268.
- Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J. L. S. & Subramaniam, S. (2016). *Cell* **165**, 1698-1707.
- Merk, A., Fukumura, T., Zhu, X., Darling, J. E., Grishammer, R., Ognjenovic, J. & Subramaniam, S. (2020). *IUCrJ* **7**, 639-643.
- Munir, A., Wilson, M. T., Hardwick, S. W., Chirgadze, D. Y., Worrall, J. A. R., Blundell, T. L. & Chaplin, A. K. (2021). *Structure.*
- Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehl, J., Uchanski, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., Chirgadze, D. Y., Murshudov, G., Aricescu, A. R. & Scheres, S. H. W. (2020). *Nature* **587**, 152-156.
- Terwilliger, T. C., Adams, P. D., Afonine, P. V. & Sobolev, O. V. (2018). *Nat Methods* **15**, 905-908.
- Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. (2020). *Nature* **587**, 157-161.
- Zhang, K. M., Zhang, H. W., Li, S. S., Pintilie, G. D., Mou, T. C., Gao, Y. Z., Zhang, Q. F., van den Bedeme, H., Schmid, M. F., Au, S. W. N. & Chiu, W. (2019). *P Natl Acad Sci USA* **116**, 6800-6805.

## Publication III

Ringel, Marion; Dimos, Nicole; Himpich, Stephanie; Haack, Martina; Huber, Claudia;  
Eisenreich, Wolfgang; Schenk, Gerhard; Loll, Bernhard; Brück, Thomas (2022):  
Biotechnological potential and initial characterization of two novel sesquiterpene  
synthases from Basidiomycota *Coniophora puteana* for heterologous production of  $\delta$ -  
cadinol. *Microb Cell Fact* 21 (1), S. 64  
DOI: <https://doi.org/10.1186/s12934-022-01791-8>





RESEARCH

Open Access



# Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of $\delta$ -cadinol

Marion Ringel<sup>1†</sup> , Nicole Dimos<sup>2†</sup> , Stephanie Himpich<sup>2</sup>, Martina Haack<sup>1</sup>, Claudia Huber<sup>3</sup> , Wolfgang Eisenreich<sup>3</sup> , Gerhard Schenk<sup>4</sup> , Bernhard Loll<sup>2\*</sup>  and Thomas Brück<sup>1\*</sup> 

## Abstract

**Background:** Terpene synthases are versatile catalysts in all domains of life, catalyzing the formation of an enormous variety of different terpenoid secondary metabolites. Due to their diverse bioactive properties, terpenoids are of great interest as innovative ingredients in pharmaceutical and cosmetic applications. Recent advances in genome sequencing have led to the discovery of numerous terpene synthases, in particular in Basidiomycota like the wood rotting fungus *Coniophora puteana*, which further enhances the scope for the manufacture of terpenes for industrial purposes.

**Results:** In this study we describe the identification of two novel (+)- $\delta$ -cadinol synthases from *C. puteana*, Copu5 and Copu9. The sesquiterpene (+)- $\delta$ -cadinol was previously shown to exhibit cytotoxic activity therefore having an application as possible, new, and sustainably sourced anti-tumor agent. In an *Escherichia coli* strain, optimized for sesquiterpene production, titers of 225 mg l<sup>-1</sup> and 395 mg l<sup>-1</sup>, respectively, could be achieved. Remarkably, both enzymes share the same product profile thereby representing the first two terpene synthases from Basidiomycota with identical product profiles. We solved the crystal structure of Copu9 in its closed conformation, for the first time providing molecular details of sesquiterpene synthase from Basidiomycota. Based on the Copu9 structure, we conducted structure-based mutagenesis of amino acid residues lining the active site, thereby altering the product profile. Interestingly, the mutagenesis study also revealed that despite the conserved product profiles of Copu5 and Copu9 different conformational changes may accompany the catalytic cycle of the two enzymes. This observation suggests that the involvement of tertiary structure elements in the reaction mechanism(s) employed by terpene synthases may be more complex than commonly expected.

\*Correspondence: loll@chemie.fu-berlin.de; brueck@tum.de

†Marion Ringel and Nicole Dimos contributed equally to this manuscript

<sup>1</sup> Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, Lichtenbergstr. 4, 85748 Garching, Germany

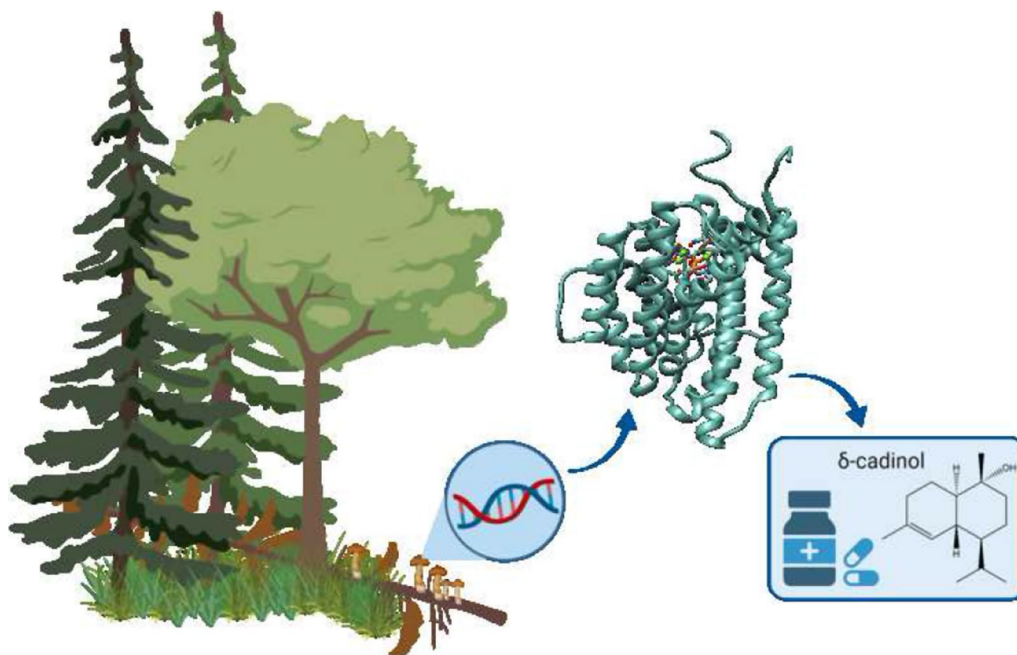
<sup>2</sup> Institute for Chemistry and Biochemistry, Structural Biochemistry Laboratory, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany  
Full list of author information is available at the end of the article



**Conclusion:** The presented product selectivity and titers of Copu5 and Copu9 may pave the way towards a sustainable, biotechnological production of the potentially new bioactive (+)- $\delta$ -cadinol. Furthermore, Copu5 and Copu9 may serve as model systems for further mechanistic studies of terpenoid catalysis.

**Keywords:**  $\delta$ -cadinol, Sesquiterpene, Basidiomycota, Terpene synthases, Active site architecture, Mutagenesis

### Graphical Abstract



### Introduction

The rapid emergence of new diseases (e.g., Covid-19) and the excess use of common drugs such as antibiotics continuously forces the scientific community towards the development of new, innovative drug leads. Over the past decades filamentous fungi have mainly been subject to research focused on the biosynthesis of major antibiotic agents, whereas the identification of bioactive terpenoids was largely based on the analysis of secondary metabolites of medicinal plants [1]. Little attention has been paid to filamentous fungi as potential source of new terpenoid-based bioactives. To date terpenoids represent the largest and structurally most diverse group of natural products encompassing over 80,000 characterized compounds [2]. Terpenoids (and in particular the sesquiterpenoid subfamily) are widely used in medicine and health care for their anti-insect, anti-inflammatory, anti-viral, anti-malarial, anti-microbial and anti-tumor activities [3, 4]. A prominent example for a clinically relevant sesquiterpenoid bioactive is artemisinin, which is

a first line treatment against malaria [5]. All sesquiterpenoids feature a complex  $C_{15}$  carbohydrate skeleton which is formed by the cyclization of the universal, aliphatic precursor farnesyl diphosphate (FPP), a reaction catalyzed by enzymes from the sesquiterpene synthase family. Sesquiterpene synthases typically belong to the class I terpene synthase (TPS) family, commonly exhibiting a  $\alpha$ ,  $\alpha\beta$  or  $\alpha\beta\gamma$  domain architecture for monofunctional enzymes, with the catalytic site located in their respective  $\alpha$ -domains. Bifunctional class I TPS also exist, exhibiting additional catalytic functions in either a second  $\alpha$ - or the  $\gamma$ -domain (displaying the class I—class I  $\alpha\alpha$  or class I—class II  $\alpha\beta\gamma$  domain architecture) [2]. All class I TPSs share highly conserved sequence motifs in their respective  $\alpha$ -domains, such as the aspartate-rich DDXXD (DD) dyad and the (N,D)D(L,I,V)X(S,T)XXXE (NSE) triad, with both being involved in complexing of three  $Mg^{2+}$  ions that are essential for catalysis [2]. Furthermore, class I TPSs share a WxxxxRY sequence motif that facilitates the closure of the active site via salt bridge formation

upon substrate binding [6, 7]. In this regard, the *Streptomyces*-derived class I di-TPS CotB2 [8] and the trichodiene synthase from *Fusarium sporotrichioides* [9], yielding the sesquiterpene trichodiene, belong to the best studied TPSs to date [6, 10]. Detailed computational studies on the mechanism employed by trichodiene synthase have highlighted the relevance of a bifacial active site architecture consisting of a highly polar region to promote binding of diphosphate (PP region) and a hydrophobic pocket lined with aromatic amino side chains to guide the propagation of carbocations [10, 11]. Upon the initial substrate binding within the PP region, which also involves the complexation of the diphosphate moiety by the tri-Mg<sup>2+</sup>-cluster, the active site is closed (induced fit) [7, 10]. In the first step of the chemical reaction, C–O bond cleavage and the abstraction of the PP moiety result in the formation of a farnesyl carbocation intermediate. Subsequently, the carbocation relocates to the hydrophobic pocket; the rate of this rearrangement, and therefore the maturing of the carbocation to the respective cyclized (sesqui-) terpene, is controlled by electrostatic interactions within the active site [11]. The structure of the mature terpene is ultimately defined by the amino acid residues lining the hydrophobic binding pocket [2, 11]. Therefore, detailed knowledge about catalytically essential residues within the active site of TPSs may guide the engineering of such enzymes towards targeted products (e.g., new natural bioactives).

This study reports the identification of two new sesqui-TPSs, Copu5 and Copu9, from the wood rotting fungus *Coniophora puteana* via a genome mining approach. Both enzymes can be classified as class I terpene synthases and catalyze the formation of (+)- $\delta$ -cadinol as main product from its direct precursor FPP. (+)- $\delta$ -cadinol, which is also known as “torreyol” or “pilgerol” [12], has been subject to numerous studies over the past decades including bioactivity tests [13, 14]; plant extracts containing cadinene-type sesquiterpenes (e.g.,  $\delta$ -cadinol) were shown to have anti-microbial, anti-fungal and anti-inflammatory properties [15, 16]. Furthermore, purified (+)- $\delta$ -cadinol exhibited cytotoxic activity against MCF7 cells with an IC<sub>50</sub> of 3.5 ± 0.58  $\mu$ g ml<sup>-1</sup> [17]. To date only two other terpene synthases predominantly forming  $\delta$ -cadinol have been identified, BvCS from *Boreostereum vibrans* [18] and GME3638 from *Lignosus rhinocerotis* [17] with sequence similarities of 41.4% and 58.3% compared to Copu9, respectively (Additional file 1: Fig. S3). Interestingly, exactly like *Coniophora puteana*, both organisms belong to the division of Basidiomycota. Analyses of the amino acid sequences and corresponding tertiary structure elements of Copu5 and Copu9, employing protein

crystallization and homology modelling techniques, revealed that both enzymes share an almost identical active site architecture. Structure-based mutagenesis was employed to probe the role of catalytically important residues, with a view to provide a platform for product-targeted engineering of TPSs.

## Results and discussion

### Identification of potential sesquiterpene synthases

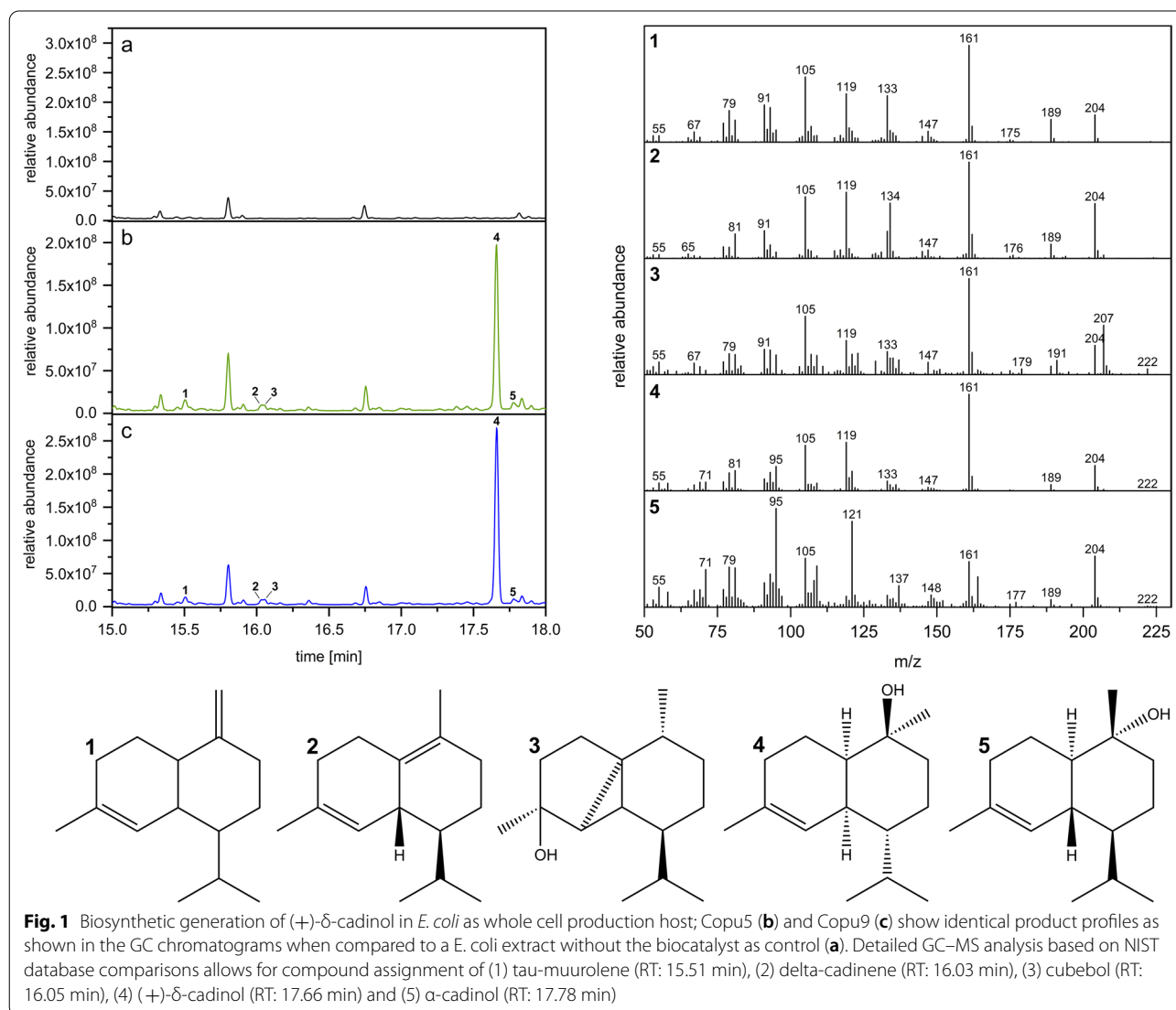
A previous study reported the identification and functional characterization of several putative TPSs within the genome of the wood rotting fungi *C. puteana*. Two of these candidates were indeed shown to be efficient and highly selective sesqui-TPSs producing cubebol and  $\beta$ -copaene, respectively [19]. However, the majority of these putative TPSs still await characterization. To gain further insight into the terpenom of *C. puteana*, the genome of *C. puteana* was probed for the presence of additional TPS-like sequences using a Basic Local Alignment Search Tool (BLAST) with the amino acid sequence of the recently identified cubebol synthase Copu3 [19] as reference. Six candidates were identified (Copu5: XP\_007765330, Copu6: XP\_007773189, Copu7: XP\_007767204, Copu9: XP\_007765560, Copu10: XP\_007766266.1 and Copu11: XP\_007767169.1), all of which contain the aspartate-rich DDXXD motif and the NSE triad, which provide ligands for the three essential Mg<sup>2+</sup> ions in the active site (Additional file 1: Fig. S1 and S2) [2, 7]. The class I TPS-specific WxxxxRY motif, a component of the induced fit mechanism, is also conserved [6]. Beyond these motifs the level of sequence conservation is considerably lower, with the six newly identified putative TPS sequences sharing only 24–37% sequence similarity with Copu3.

### Heterologous expression in *Escherichia coli* and characterization of produced sesquiterpenes

In order to biochemically and functionally characterize the novel TPSs from *C. puteana*, their respective open reading frames (ORFs) were codon-optimized for heterologous expression in *E. coli* and cloned into a single operon expression system as described previously [19]. The employed expression system, also includes the native *E. coli* non-mevalonate pathway (MEP) bottleneck enzymes 1-deoxy-D-xylulose-5-phosphate synthase (DXS; WP\_099145004.1) and isopentenyl-pyrophosphate isomerase (IDI; AAC32208.1) to enhance sesquiterpene production as previously described [19, 20]. In addition, the ORFs of the TPS candidates were also cloned into a two-plasmid diterpene production system, which includes a geranylgeranyl diphosphate synthase (crtE

from *Pantoea ananatis*; ADD79325.1) [21], to test for possible catalytic activity towards diterpene production. In preliminary experiments Copu6, Copu7, Copu10 and Copu11 did not show any catalytic activity when tested for sesquiterpene production using the single operon expression vector nor for diterpene production applying the two-plasmid system. Therefore, these four putative TPSs are considered as non-viable protein sequences. Basidiomycota like *C. puteana* are prone to alternative splicing events which especially occur in organisms under stress conditions [22, 23]. Therefore, it is likely that the in silico annotated TPS coding sequences encompass variations of the enzymes, that result in inactive enzyme variants. Active variants of the annotated genes might occur in vivo during splicing events caused by exogenous

stress conditions. In contrast, GC–MS analysis of Copu5- and Copu9-expressing *E. coli* extracts showed catalytic activity towards the production of five different sesquiterpenes, when co-expressed with the MEP bottleneck enzymes (Fig. 1). Both cell extracts show typical fragmentation patterns at 105, 119, 161 and 204  $m/z$  indicative of non-functionalized, cyclic sesquiterpenes. Moreover,  $m/z$  values at 105, 119, 161, 204 and 222 suggest that the generated sesquiterpenes are decorated with a single hydroxyl group. Interestingly, Copu5 and Copu9 appear to be very selective showing only one prominent GC–MS signal with a typical mass pattern representative of a mono-hydroxylated sesquiterpene (parent ion 222  $m/z$ ; RT: 17.09 min). Remarkably, the cyclisation products generated by both Copu5- and Copu9 display the same



fragmentation pattern, indicative of formation of identical sesquiterpenes (Fig. 1).

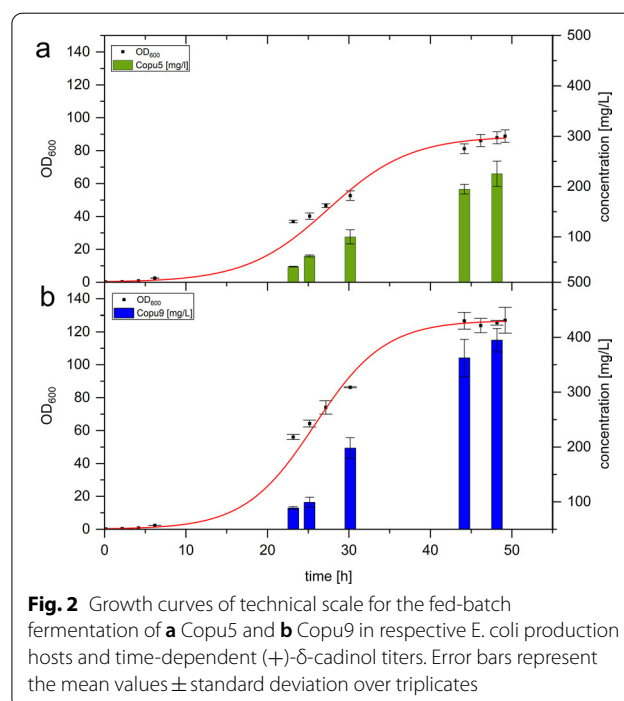
### Structure elucidation of generated sesquiterpenes

In order to structurally identify the sesquiterpenes generated by Copu5 and Copu9 a detailed comparison of their mass spectra with the National Institute of Standards and Technology (NIST) database was performed [24]. This evaluation of the *E. coli* extracts expressing Copu5 and Copu9 revealed (+)- $\delta$ -cadinol (RT: 17:66 min; parent ion mass 222 m/z, major daughter ions at 119, 161 and 204 m/z) as the main product (Fig. 1). The identity of this product was confirmed via NMR experiments (Standardized on solvent (CDCl<sub>3</sub>) peak: <sup>1</sup>H=7.26 ppm, <sup>13</sup>C=77.2 ppm <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  5.51 (dq, J=5.3, 1.6 Hz, 1H), 1.98 (m, J=13.9, 10.4, 6.2 Hz, 4H), 1.92–1.85 (m, 1H), 1.66 (s, 3H); 1.63–1.45 (m, 5H), 1.35–1.25 (m, 1H), 1.29 (s, 3H), 1.09 (qd, J=13.2, 4.2 Hz, 1H), 0.88 (d, J=6.9 Hz, 3H), 0.81 (d, J=6.9 Hz, 3H); <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)  $\delta$  134.36, 124.61, 72.55, 45.55, 44.08, 36.77, 35.31, 31.14, 27.97, 26.41, 23.66, 21.70, 21.52, 18.51, 15.31.) in conjunction with a comparison to reported NMR data (Additional file 1: Figs. S4–S11) [17]. In addition to the main cyclisation product (+)- $\delta$ -cadinol, the sesquiterpenes tau-muurolole (RT: 15.51 min), delta-cadinene (RT: 16.03 min), cubebol (RT: 16.05 min) and  $\alpha$ -cadinol (RT: 17.78 min) were putatively assigned as minor products in both extracts as indicated by comparison of GC–MS fingerprint spectra with NIST database references (Fig. 1). Based on these product profiles both Copu5 and Copu9 can be designated as new, highly selective (+)- $\delta$ -cadinol synthases. The product selectivity of (sesqui-) TPSs varies significantly within this versatile enzyme family ranging from single product formation (e.g., (+)- $\delta$ -cadinene synthase from *Gossypium arboreum*) to a product portfolio of over 50 different compounds (e.g.,  $\gamma$ -humulene synthase from *Abies grandis*) [19, 25, 26]. In contrast to the common function of an enzyme as an accelerator of a reaction rate, the catalytic challenge for TPSs rather lies in the control of the highly reactive carbocation intermediates alongside their reaction trajectory [27]. The product distribution in TPSs is guided by several factors such as: (i) the activation of the C–O bond by the pyrophosphate-Mg<sup>2+</sup>-cluster in the active site, (ii) electrostatic interactions that lead to the sequestration of the active site, (iii) the specific positioning of water molecules or acidic/basic residues, that facilitate site-specific hydroxylations or (de)protonations, and (iv) a specific active site architecture that pre-shapes the carbocation intermediate [11, 27]. For instance, for two fungal sesquiterpene synthases, Cop4 and Cop6 from

*Coprinus cinereus*, it was demonstrated that a smaller carbocation binding pocket lead to a more specific product profile as the carbocation intermediate is more restricted along its potential cyclization routes [28, 29]. At present, all functionally characterized sesquiterpene synthases from *C. puteana* show a highly specific product distribution, indicating that their active site architectures may be very effective in restricting the carbocation intermediates, thereby preventing undesired side reactions.

### Technical scale production of (+)- $\delta$ -cadinol

Mischko and co-workers demonstrated that the TPSs from *C. puteana* have both a high product selectivity as well as high product titers [19]. In order to investigate the performance of the newly identified (+)- $\delta$ -cadinol synthases Copu5 and Copu9 in an optimized *E. coli* production host, technical scale, fed-batch fermentation experiments were carried out using a 1.3 L parallel fermentation system as described previously [19]. *Escherichia coli* cultures co-expressing Copu5 and the respective MEP bottleneck enzymes reached stationary phase after 48 h with a final OD<sub>600</sub> of 88 and a (+)- $\delta$ -cadinol titer of 225 mg l<sup>-1</sup> (Fig. 2). Based on this data a Copu5-specific productivity of 4.7 mg l<sup>-1</sup> h<sup>-1</sup> was calculated. In contrast, Copu9-expressing cultures reached a final OD<sub>600</sub> of 126 and a (+)- $\delta$ -cadinol titer of 395 mg l<sup>-1</sup>, entering





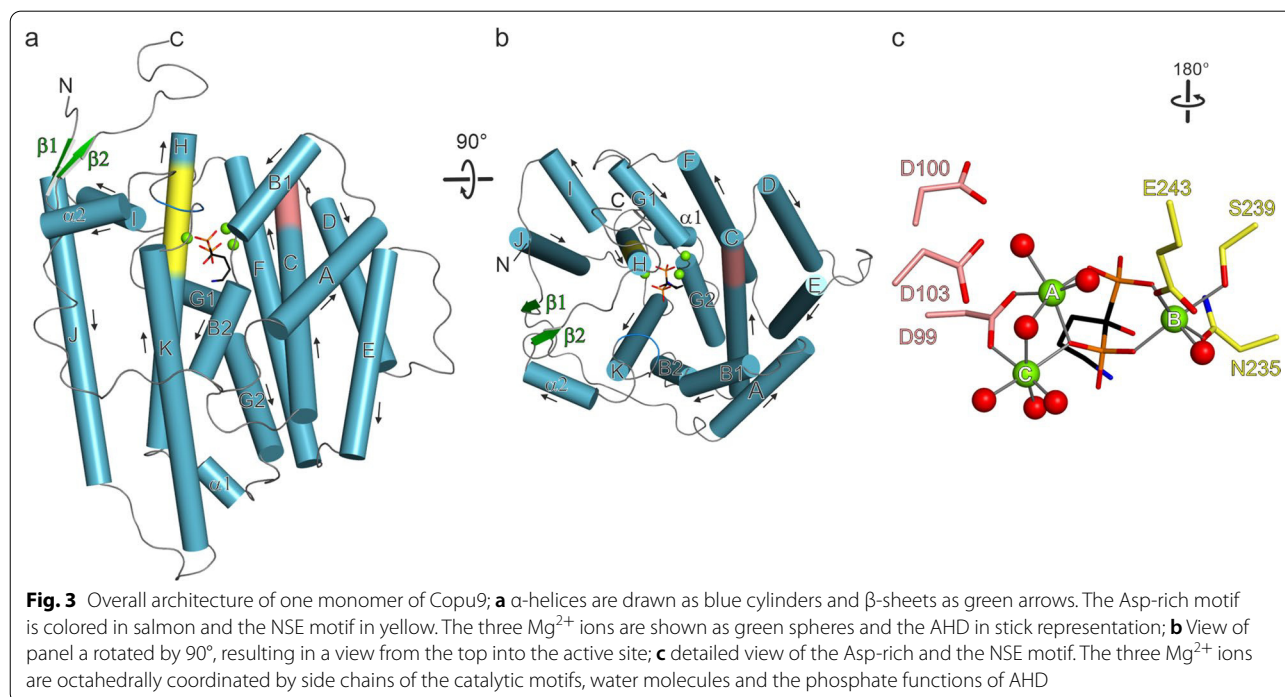
stationary phase after 48 h (Fig. 2). The calculated productivity of Copu9 was  $8.2 \text{ mg l}^{-1} \text{ h}^{-1}$ , respectively. Since equal fermentation parameters were maintained in the fermentation of Copu5- and Copu9-expressing strains, the only varying factor was the used TPS itself. The resulting different biomass ( $\text{OD}_{600}$ ) and  $\delta$ -cadinol accumulation is thus likely to be a result of a difference in metabolic burden. To date only a few attempts for the biotechnological production of  $\delta$ -cadinol by microbial hosts have been reported, all of them resulting only in minor yields (no larger than  $1 \text{ mg l}^{-1}$ ) [17, 18], and hence, despite its promising pharmaceutical properties,  $\delta$ -cadinol is mainly referred to as constituent of various plant extracts used in traditional medicine [15, 30, 31]. Therefore, our titers not only significantly exceed those of previous studies, but also, provide an opportunity to sustainably and scalably generate this highly valuable natural bioactive.

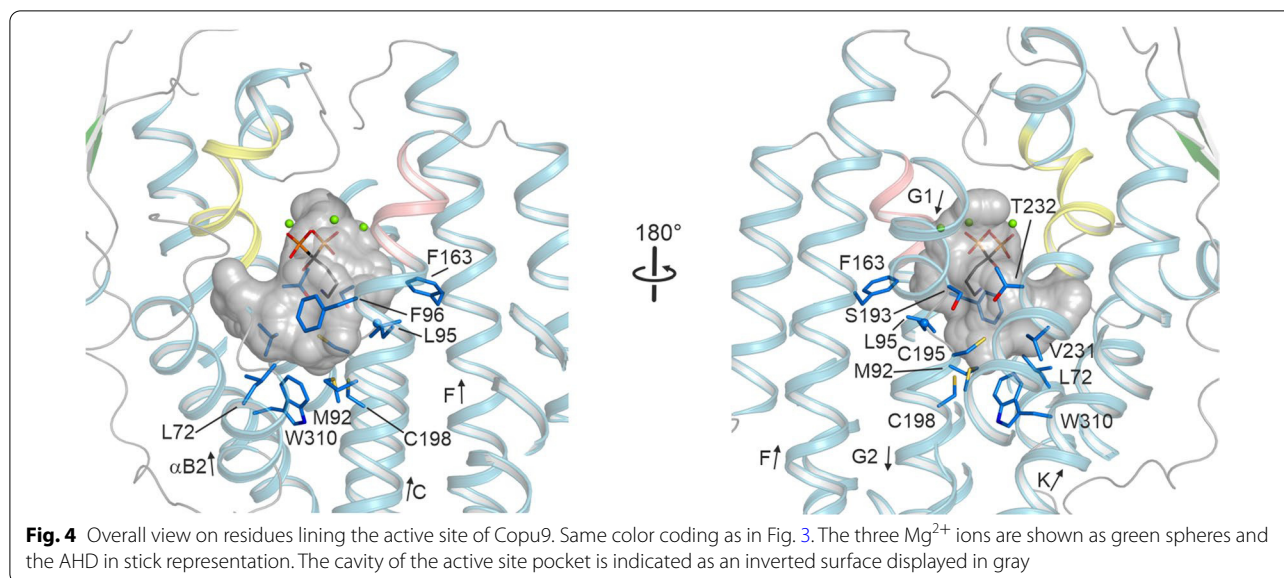
### Structural comparison of Copu5 and Copu9

Copu5 and Copu9 exhibit the same product profile (Fig. 1) but differ in their productivities (Fig. 2). The two enzymes share 52.7% sequence identity and a homology of 65.2% (Additional file 1: Figs. S2 and S3). In order to gain insight into residues, that promote the high selectivity of these enzymes, but also the enhanced productivity of Copu9, crystallization trials were carried out. Copu9 could not be crystallized in its open, resting state conformation, but co-crystallization with the non-hydrolysable FPP substrate

mimic (4-amino-1-hydroxybutylidene)bisphosphonic acid (alendronate, AHD) and  $\text{MgCl}_2$  resulted in crystals suitable for the collection of X-ray diffraction data. The obtained structure thus represents the closed, catalytically active Copu9 (Copu9· $\text{Mg}_3$ ·AHD) conformation. No crystals were obtained for Copu5. In order to further understand the differences of Copu9 and Copu5 a thermal shift assay was performed revealing a significantly lower melting temperature for Copu5 in its apo state as well as bound to AHD [ $27.6 \pm 1.0/33.5 \pm 0.8^\circ\text{C}$  compared to  $37.2 \pm 0.6/44.3 \pm 0.6^\circ\text{C}$  of Copu9 in pyrophosphate containing Copu5 buffer (Additional file 1 Fig. S15)]. This might be an explanation for the difficulties in purification of Copu5.

Crystals of Copu9· $\text{Mg}_3$ ·AHD diffracted to a resolution of  $1.83 \text{ \AA}$  (Additional file 1: Table S1). The enzyme forms a homodimer both *in crystallo* and in solution as observed in size exclusion chromatography. Inspection of the electron density clearly revealed bound AHD and the presence of three  $\text{Mg}^{2+}$  cations (Additional file 1: Fig. S13) in a closed conformation. Both poly-peptide chains are practically identical with a root mean square deviation (rmsd) of  $0.32 \text{ \AA}$  for 330 pairs of  $\text{C}\alpha$  atoms. The structure of Copu9 is complete except for its 13 N-terminal residues. While TPSs generally are helical bundle proteins lacking any  $\beta$ -strands, two short  $\beta$ -strands are present in Copu9, one at the N- and one at the C-terminal ends (Fig. 3a, b and Additional file 1: Fig. S12). These two  $\beta$ -strands (T17-L21 and R335-L339) form an antiparallel





$\beta$ -sheet, which might further stabilize the closed conformation. Copu9 shows the classical  ${}^{99}DDWLD^{103}$  (located on  $\alpha$ -helix C) and  ${}^{235}NSE^{243}$  (located on the opposing  $\alpha$ -helix H) motifs. The C-terminal  ${}^{317}WxxxxRY^{324}$  motif adopts a random coil conformation and folds onto the active site, reflecting the closed conformation of Copu9. Latter conformation of the  ${}^{317}WxxxxRY^{324}$  segment is identical as previously observed in CotB2 [6]. Therefore, both side chains of R223 and Y324 point towards the active site. R223 established a bidentate salt-bridge to Asp99 of the Asp-rich motif. R324 forms a hydrogen bond to one phosphate function of AHD. The pyrophosphate sensor [32] R188 is located on  $\alpha$ -helix G1 and establishes a bidentate salt-bridge to the second phosphate group of AHD.

The active site is mainly lined by hydrophobic residues: L72, M92, L95, F96, F163, S193, G194, C195, C198, V231, T232 and W310 (Fig. 4). Based on a DALI search [33], the closest structural homologue to Copu9 is Selinadiene synthase (SdS; PDB-ID 4OKM [7]) (Additional file 1: Table S2). The two structures superimpose with a rmsd of 1.45 Å for 296 pairs of C $\alpha$  atoms.

Since we could not obtain an experimental structure of Copu5, we predicted the structure by the ROBETTA server [34]. To validate the prediction, we initially predicted the structure of Copu9. The obtained model, in its open, inactive conformation superimposes very well, with the experimentally obtained structure of Copu9-Mg $^{2+}$ +3-AHD (Additional file 1: Fig. S16 and Table S3). The largest differences in the protein backbone are observed in the N- and C-terminal extensions of the protein (Additional file 1: Fig. S16). As anticipated,

the active site of the modeled structure is wider, due to the absence of the Mg $^{2+}$  ions as well as alendronate, since both  $\alpha$ -helices harboring the metal binding motifs are tilted away from the active side. A similar observation is made in the model of Copu5 that largely resembles the fold of Copu9 (Additional file 1: Fig. S16b, c and Table S3). The amino acid sequences in their hydrophilic PP binding pockets are highly conserved (DDXXD: Copu5:  ${}^{92}DDWSD^{96}$ , Copu9:  ${}^{99}DDWLD^{103}$ ; NSE: Copu5:  ${}^{227}NDVFSYNKE^{235}$ , Copu9:  ${}^{235}NDIFSYNKE^{243}$ ; WxxxxRY: Copu5:  ${}^{309}WSFETERY^{316}$ , Copu9:  ${}^{317}WSFDSHRY^{324}$ ) (Additional file 1: Fig. S2). Furthermore, the 12 residues involved in either pre-shaping the geometry or the propagation of the carbocations in the hydrophobic pocket of the active site are identical in the two enzymes (Additional file 1: Fig. S12 and Table S4) which is likely to be the cause of their identical catalytic activity. To the best of our knowledge Copu5 and Copu9 are the first reported TPSs from the same organism with the same product profile and an almost equivalent active site decoration. To this end, there are two residues located in the second shell of residues lining the active site and thus merely surrounding first shell residues, which are different between Copu9/Copu5: namely (F91/Y84), and (C198, V190), respectively. By contrast, both Copu5 and Copu9 only share four of the 12 relevant residues in the hydrophobic pocket with the previously reported cubebol synthase Copu3 [19] (Additional file 1: Fig. S12 and Table S4). The remaining eight amino acid side chains are thus likely to play an important role in guiding the product profiles of Copu3 and Copu5/Copu9.

### Structure-based mutagenesis targeting active site residues

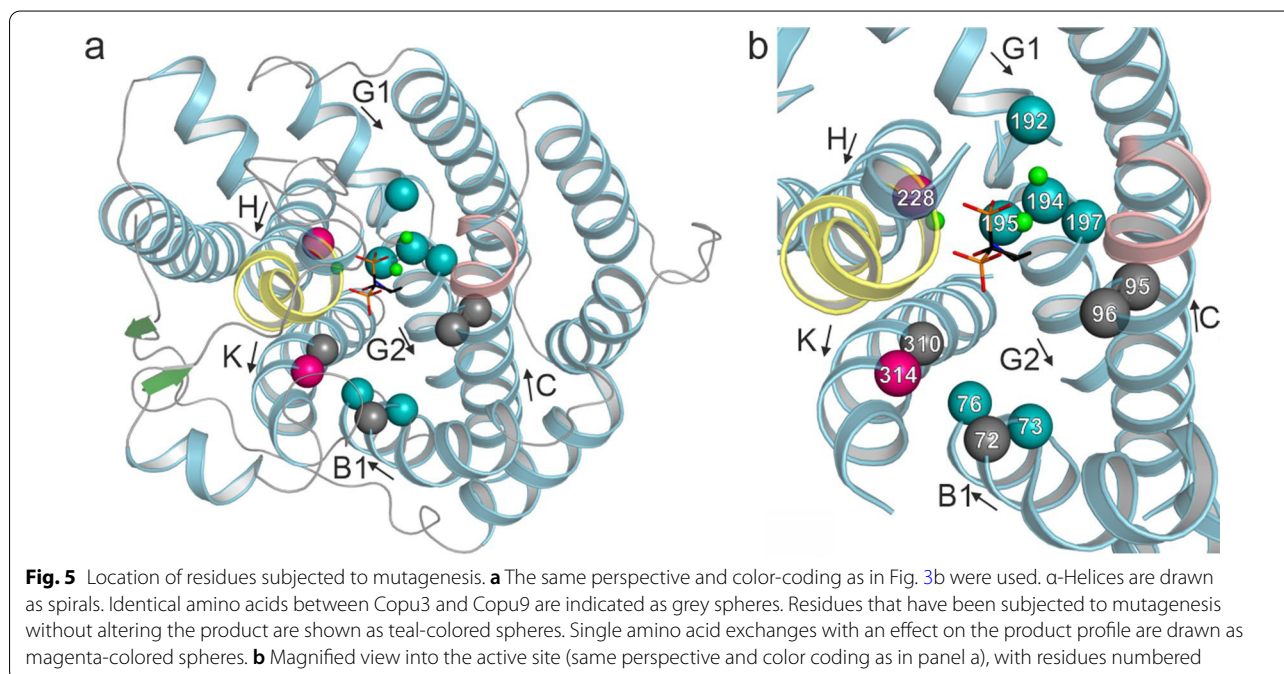
In order to evaluate their roles, each of them was iteratively changed in Copu5 and Copu9 to their counterpart present in Copu3. As the main product of Copu3 is cubebol [19], which is also produced in minor amounts by Copu5 and Copu9 (Fig. 1), the mutations introduced into Copu5 and Copu9 were anticipated to shift their product profile towards that of Copu3, i.e., generation of cubebol as the main cyclization product (Fig. 5).

The Copu5 variants T66C, C69V and T184N and the Copu9 variants T73C, C76V, T192N, G194A, C195V and P197C did not affect their product spectrum. In contrast, Copu5 variants G186A, C187V and S306N and Copu9 variants N228A and S314N showed increased synthesis of minor products (Fig. 1, Additional file 1: Fig. S17 and S18), while the P189C variant of Copu5 exhibited lower product formation indicating, that this residue is either essential for catalysis or interferes with the catalytically active, closed conformation. Notably, Copu5 variants C187V, N220A and S306N as well as Copu9 variants N228A and S314N showed the formation of an additional side product, germacrene D-4-ol (Additional file 1: Figs. S17–S19; RT: 16.85 min; identified by a comparison to the NIST database [24]). Interestingly, germacrene D-4-ol is also a side product of Copu3.<sup>12</sup>

In order to further evaluate the influence of the conducted point mutations on the synthases' catalytic properties in vitro kinetic experiments were performed. All kinetic parameters obtained from Copu9 and its variants

reaction with FPP are listed in Table 1 (Fig. 6, Additional file 1: Fig. S21). Copu9 WT and all variants show comparable binding affinity ( $K_m$ ) towards FPP. However, Copu9 variants C76V, N228A and S314N show a minor decrease in catalytic turnover ( $k_{cat}$ ) compared to Copu9 WT, while variants T73C, T192N, G194A, C195V and P197C either retain WT  $k_{cat}$  or show a slightly increased catalytic turnover compared to Copu9 WT (Table 1, Fig. 6). Copu9 WT and all variants also show comparable catalytic efficiency ( $k_{cat}/K_m$ ). However, variant C76V, which showed a minor decrease in catalytic turnover, exhibits slightly higher substrate specificity. Interestingly, the variants N228A and S314N, show a tendency towards decreased substrate specificity which possibly reflects the minor changes in their product spectrum. All kinetic parameters observed for Copu9 WT and its variants are within range of the respective kinetic constants of the kinetically well characterized fungal sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus* [29]. In contrast to Copu9, it was only possible to purify Copu5 using pyrophosphate containing buffers due to its significant in-vitro stabilizing effect. Hence, we were not able to determine the in-vitro kinetics of Copu5 and its variants.

However, none of the variants neither significantly changed the product profile towards another major product, such as cubebol, nor showed a drastic change in its respective catalytic properties. Copu5 appears to be more receptive to single amino acid changes whereas Copu9 largely compensates mutations and retains its



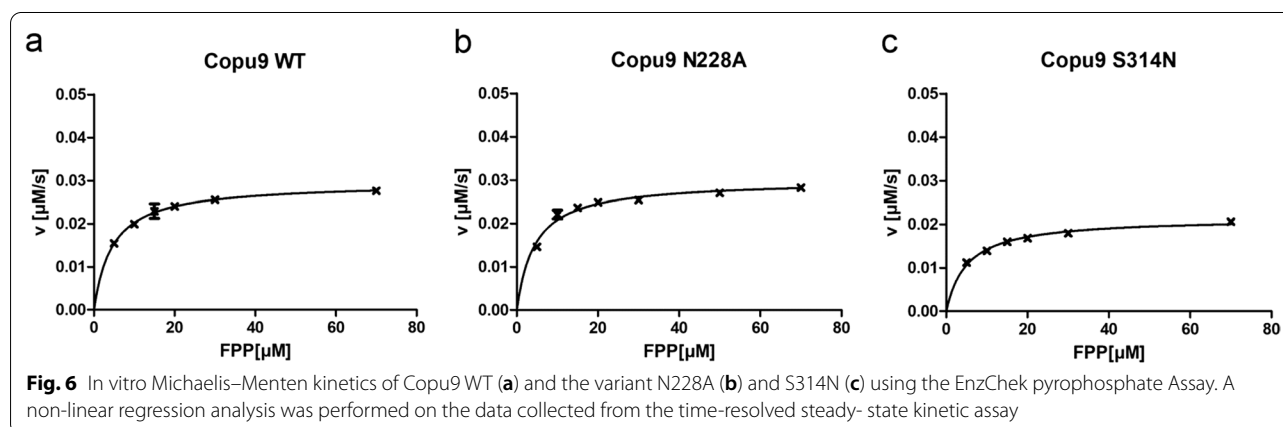


**Table 1** Steady-state kinetic parameters of Copu9 and its variants calculated from the EnzChek™ pyrophosphate assay

Copu9	$k_{cat}$ [ $s^{-1}$ ]	$K_m$ [ $\mu M$ ]	$k_{cat}/K_m$ ( $\times 10^3$ ) [ $s^{-1} M^{-1}$ ]
WT	$(3.24 \pm 0.03) \times 10^{-2}$	$4.59 \pm 0.17$	7.067
T73C	$(5.02 \pm 0.26) \times 10^{-2}$	$6.28 \pm 1.37$	7.995
C76V	$(2.91 \pm 0.07) \times 10^{-2}$	$3.14 \pm 0.45$	9.252
T192N	$(3.96 \pm 0.08) \times 10^{-2}$	$8.28 \pm 0.60$	4.782
G194A	$(3.60 \pm 0.05) \times 10^{-2}$	$7.00 \pm 0.37$	5.135
C195V	$(3.98 \pm 0.07) \times 10^{-2}$	$8.48 \pm 0.51$	4.692
P197C	$(3.81 \pm 0.10) \times 10^{-2}$	$6.82 \pm 0.77$	5.584
N228A	$(2.62 \pm 0.07) \times 10^{-2}$	$4.42 \pm 0.59$	5.940
S314N	$(1.92 \pm 0.05) \times 10^{-2}$	$5.10 \pm 0.54$	3.755

wild-type product profile. Recently, it was demonstrated that mutations in loop regions distinct from the active site may result in drastic catalytic differences in related synthases [28]. Considering the proposed cyclization mechanism required to form cadinene-type sesquiterpenes (Additional file 1: Fig. S20) every mutation carried out for Copu5 and Copu9 would be anticipated: (i) to have an effect (stabilizing or destabilizing) on the reaction of the carbocation intermediates as well as, (ii) to allow or restrict water molecule(s) to enter the active site [19, 26, 35]. In particular, residues C187, N220 and S306 in Copu5 and corresponding residues in Copu9 (N228 and S314) appear to have an impact on controlling the carbocation reaction trajectory of the germacryl cation, which is the common intermediate of the cadinene and germacrene cyclization trajectories (Additional file 1: Fig. S20) [19, 26, 29]. However, the impact of single amino acid substitutions in Copu5 and Copu9 is relatively small when compared to other TPSs (e.g., CotB2) [36–39]. In that context, residues lining the entrance of the active site (e.g., residues G186 and C187 in Copu5) affect product specificity.

The directed modulation of a product profile towards a defined product was previously demonstrated for (+)- $\delta$ -cadinene synthase from *G. arboreum* [26]. The main product of that synthase was altered towards the production of germacrene D-4-ol by specific mutations within the G1 and G2  $\alpha$ -helices in the hydrophobic pocket of the active site of this enzyme. These helices and the small linker region connecting them, were shown to be important for product specificity in class I TPSs [7, 26, 40]. Upon substrate binding this segment is subject to a conformational change triggered by a conserved effector triad, which was first identified in the C-terminal end of the G1  $\alpha$ -helix of SdS (residues Arg178, Asp181 and Gly182 [7]). This induced-fit mechanism triggers substrate ionization and therefore represents the starting point of the subsequent carbocation cyclization trajectory towards germacrene D-4-ol. By a superimposition of the SdS and Copu9 structures in conjunction with an inspection of a sequence alignment that also includes Copu5, this effector triad could also be identified in the C-terminal ends of the G1  $\alpha$ -helices of Copu5 (R182, D183 and S185) and Copu9 (R190, D191 and S193), respectively. However, instead of the commonly found glycine at the third position of this triad, both Copu5 and Copu9 employ a serine residue (S185 and S193, respectively). Single mutations in the vicinity of the effector triad of Copu9 (i.e., T192, G194, C195 and P197) have no significant effect on the catalytic properties nor product profile of that enzyme (vide supra). However, mutations of the corresponding glycine (G186) and cysteine (C187) residues in Copu5, both located in the linker region between the  $\alpha$ -helices G1 and G2, alter the product profile, with the G186A variant also promoting the formation of germacrene D-4-ol (Additional file 1: Figs. S15–S17). This glycine and cysteine residues are the closest residues within the hydrophobic pocket to the phosphate moiety of the bound substrate (see G194 and C195 in Fig. 5).



Therefore, the results presented herein point out: (i) that the serine residue in the effector triad may be important in promoting the production of (+)- $\delta$ -cadinol, but also (ii) that there are likely additional tertiary structural elements, that influence the product profile of sesqui-TPSs, possibly by affecting the interface between the hydrophilic and hydrophobic regions in the active site, and/or by interaction with the linker region between  $\alpha$ -helices G1 and G2. This is in agreement with studies exploring the influence of domain-domain interactions on the catalytic function of TPSs [41, 42]. It was shown that different bifunctional TPSs exhibit varying degrees of domain-domain interdependence regarding their catalytic activity. Upon separation of the  $\alpha$  and  $\beta$  domain in Ent-kaurene synthase from *Phaeosphaeria* sp. [42], the single domains still showed catalytic activity but catalytic activity was decreased by 30-fold compared to the wild-type full-length protein. Taxadiene synthase from *Taxus brevifolia* also showed severely compromised catalytic activity in its  $\alpha$ -domain when being separated from the  $\beta$ -domain [43]. In contrast, the separation of the same domains in abietadiene synthase from *Abies grandis* [41] led to a complete loss of function of the separate domains. Further studies on domain-domain chimeras of fusicoccadiene synthase from *Phomopsis amygdala* as well as ophiobolin F synthase from *Aspergillus clavatus* showed significantly altered cyclization fidelity and catalytic activity compared to the wild-type enzymes [43]. The architecture of multi-domain TPS might contribute to the overall stability of the proteins. Moreover, tertiary structure interactions play a significant role in shaping the active site for precise chemical control along the carbocation reaction trajectory. Especially for residues involved in conformational changes during catalysis or flexible tertiary structure elements this might play a bigger role than hitherto expected.

## Conclusion

The genomes of terrestrial and marine organisms bear an enormous and at present widely uncharacterized capacity of TPSs mediating the formation of natural bioactives [44, 45]. Similarly, although fungi are known for their broad variety of bioactive terpenoids, fungal TPSs have not yet been explored in depth for their potential for biotechnology applications [35]. In this study six novel putative TPSs from the Basidiomycota *C. puteana* were characterized. Only two of them, Copu5 and Copu9, were functional and identified as efficient and highly product selective (+)- $\delta$ -cadinol synthases (Fig. 1). Notably, these two synthases are the first TPSs that originate from the same organism and yet have virtually identical product profiles. The observed product selectivity of both synthases is likely to be a result of highly conserved

sequence motifs and a carefully and equally decorated active site. Both display excellent production and productivity for cadinol, exceeding currently available biosynthesis systems by far (Fig. 2).

The crystal structure of Copu9 in complex with three catalytically essential  $Mg^{2+}$  ions and the substrate mimic AHD was solved to a resolution of 1.83 Å (Fig. 3). This structure is the first of a class I TPS from Basidiomycota. Structure-informed mutations in the hydrophobic pockets of the active sites provided proof-of-concept, that single amino acid changes can alter the product profile of Copu5 and Copu9 from (+)- $\delta$ -cadinol to germacrene D-4-ol. However, despite having identical residues in the hydrophobic substrate binding pocket, Copu5 appears to be more flexible towards changing its product profile than Copu9. For Copu5 mutations targeting the linker between the  $\alpha$ -helices G1 and G2 affected the product spectrum (Additional file 1: Fig. S12). This structural segment has previously been shown to be important for the product specificity in class I TPSs [7, 26, 40]. To this end, the herein presented results in conjunction with the results previously shown for a (+)- $\delta$ -cadinene synthase [26] suggest that the helix-turn-helix motif may be a hotspot for product shifting mutations and thus serve as a guiding significance for the improvement of related class I terpene synthases. The question remains why Copu9, unlike other members of this class of enzymes, appears to be less affected by mutations in this region. This observation may suggest a structural robustness of Copu9 and may also be connected to its high product yield and specificity when expressed in *E. coli*. It is likely, that additional structural elements and associated conformational changes play an important role in modulating the catalytic performance and product profile of class I TPSs. In this respect, Copu5 and Copu9 present ideal model systems to enhance mechanistic insight into this important class of enzymes for applications in biotechnology and synthetic biology.

## Experimental section

### General

All media components and HPLC grade chemicals were purchased from Roth chemicals (Karlsruhe, Germany). Technical grade solvents were obtained from Westfalen AG (Münster, Germany).  $CDCl_3$  and Benzene- $d_6$  were purchased from Sigma-Aldrich (St. Louis, USA).

### Gene cloning, plasmid construction and culture condition

*Escherichia coli* strain DH5 $\alpha$  was used for all cloning steps and *E. coli* strain HMS174 (DE3) for terpene production. Genes encoding putative TPSs Copu5 (XP\_007765330), Copu6 (XP\_007773189), Copu7 (XP\_007767204), Copu9

(XP\_007765560), Copu10 (XP\_007766266.1) and Copu11 (XP\_007767169.1) from *C. puteana* were codon-optimized for *E. coli* using the GeneOptimizer™ software and subsequently synthesized by Eurofins Genomics GmbH (Ebersberg, Germany). For sesquiterpene production all genes were cloned into a pACYC-based vector system containing a single operon with selected bottleneck enzymes of the MEP pathway as previously described [19] and the respective TPS all set under the control of a *lac*-I-derived constitutive promoter [21]. For determination of potential catalytic activity towards diterpene production all genes were cloned into a pACYC-based expression vector system and co-transformed with a plasmid containing essential bottleneck enzymes for diterpene production as previously described [21]. All cloning experiments were performed according to standard protocols.

Cultures were grown in modified R-Media [46] ( $13.3 \text{ g l}^{-1} \text{ KH}_2\text{PO}_4$ ,  $4.0 \text{ g l}^{-1} (\text{NH}_4)_2\text{HPO}_4$ ,  $1.7 \text{ g l}^{-1}$  citric acid,  $5.0 \text{ g l}^{-1}$  yeast extract,  $4.88 \text{ ml l}^{-1} 1 \text{ M MgSO}_4$ ,  $2.45 \text{ ml l}^{-1} 0.1 \text{ M Fe(III) citrate}$ ,  $1.00 \text{ ml l}^{-1} 100 \times \text{Trace Element Solution}$  ( $5.0 \text{ g l}^{-1} \text{ EDTA}$ ,  $84 \text{ mg l}^{-1} \text{ ZnCl}_2$ ,  $13 \text{ mg l}^{-1} \text{ CuCl}_2 \cdot 2 \text{ H}_2\text{O}$ ,  $10 \text{ mg l}^{-1} \text{ CoCl}_2 \cdot 2 \text{ H}_2\text{O}$ ,  $10 \text{ mg l}^{-1} \text{ H}_3\text{BO}_3$ ,  $1.6 \text{ mg l}^{-1} \text{ MnCl}_2 \cdot 4 \text{ H}_2\text{O}$ ) and  $30 \text{ g l}^{-1}$  glycerol at  $30 \text{ }^\circ\text{C}$  and  $100 \text{ rpm}$  shaking. The appropriate antibiotics kanamycin ( $50 \text{ } \mu\text{g ml}^{-1}$ ), chloramphenicol ( $35 \text{ } \mu\text{g ml}^{-1}$ ) or ampicillin ( $100 \text{ } \mu\text{g ml}^{-1}$ ) were added as needed.

### Fermentation

Fermentation on a technical scale was performed using a DASGIP® 1.3 L parallel reactor system (Eppendorf AG, Germany) with a modified R-media as described above. An overnight preculture was used for the inoculation of the fermenters ( $\text{OD}_{600}=0.1$ ). Cultivation temperature was kept constant at  $30 \text{ }^\circ\text{C}$ . Initial stirring velocity and air flow were set to  $200 \text{ rpm}$  and  $0.2$  volumes of air per volumes of medium per minute (vvm), respectively. Dissolved oxygen was kept constant at  $30\%$  and maintained by a gradual increase in stirring velocity (max.  $1000 \text{ rpm}$ ), oxygen content (max.  $100\%$ ) and airflow (max.  $0.8$  vvm) during fermentation. A pH value of  $7.0$  was controlled by the addition of  $25\%$  aqueous ammonia. A pH-based feeding protocol was set as previously described [19, 21]. The feeding solution consisted of  $600 \text{ g l}^{-1}$  glycerol,  $5 \text{ g l}^{-1}$  yeast extract,  $35 \text{ g l}^{-1}$  collagen,  $20 \text{ g l}^{-1} \text{ MgSO}_4$ ,  $0.3 \text{ g l}^{-1}$  thiamine-HCl,  $5 \text{ ml l}^{-1} 1 \text{ M}$  ammonium iron(III) citrate,  $20 \text{ ml l}^{-1} 100 \times \text{trace element solution}$  ( $\text{pH}=7.0$ ) [19, 47].

### Terpene extraction

To extract terpenes during the screening process  $20 \text{ ml}$  of the *E. coli* culture broth were mixed with  $20 \text{ ml}$  of an extraction solution (ethanol, ethyl acetate and hexane;

$1:1:1$ ). The mixture was shaken for  $4 \text{ h}$  at room temperature and subsequently centrifuged down for  $5 \text{ min}$  at  $8000 \text{ rpm}$  to separate the organic phase. A sample from the organic phase was then analyzed via GC-MS.

The cultivation broth from either large-scale shaking flask experiments ( $1 \text{ l}$ ) or fermentation using a DASGIP®  $1.3 \text{ l}$  parallel reactor system (Eppendorf AG, Germany) was extracted by adding the same volume of ethanol. This mixture was shaken on a rotary shaker ( $80 \text{ rpm}$ ) at  $20 \text{ }^\circ\text{C}$  for  $12 \text{ h}$ . Subsequently  $\frac{1}{2}$  volume of ethyl acetate was added and shaken for  $3 \text{ h}$  ( $20 \text{ }^\circ\text{C}$ ,  $80 \text{ rpm}$ ) followed by a centrifugation step for  $15 \text{ min}$  at  $7000g$  to separate the supernatant from the cell debris. Afterwards the  $\frac{1}{2}$  volume of hexane was added to the supernatant and the extraction was carried out for  $3 \text{ h}$  ( $20 \text{ }^\circ\text{C}$ ,  $80 \text{ rpm}$ ). The organic phase was separated using a separation funnel and subsequently concentrated using a rotary evaporator.

### Terpene purification

The crude extract was evaporated until only the oily resin remained. The resin was dissolved in  $10 \text{ ml}$  of hexane. Subsequently, flash chromatography was carried out to separate the terpene fraction from fatty acid residues using the flash chromatography system PLC 2250 (Gilson, USA) equipped with a Luna  $10 \text{ } \mu\text{m}$  silica (2)  $100 \text{ A}$  column at a flow rate of  $10 \text{ ml min}^{-1}$ . Peaks were detected using an Evaporative Light Scattering Detector (ELSD) flushed with nitrogen gas and a diode array detector at  $40 \text{ }^\circ\text{C}$ . The following gradient was applied:  $100\%$  hexane (solvent A) for  $15 \text{ min}$ , followed by a rapid change (within  $3 \text{ s}$ ) to  $100\%$  EtOAc (solvent B), a  $15 \text{ min}$  wash with solvent B, a return to  $100\%$  solvent A within  $3 \text{ s}$  and a final wash with that solvent for further  $30 \text{ min}$ . Fractions of interest were reduced to approximately  $2 \text{ ml}$  using a rotary evaporator and mixed with acetonitrile. Subsequently, the residual hexane was evaporated until only acetonitrile (ACN) remained.

For further purification of the products, the samples were injected into an Ultimate 3000 UHPLC system (Thermo Scientific, USA) containing a binary pump, a diode array detector, an automated fraction collector, and a Jetstream b1.18 column oven. The purification of the respective sesquiterpenes was carried out on a NUCLEODUR® C18 HTec  $250/10 \text{ mm}$  and guard column holder  $8 \text{ mm}$  (Machery-Nagel GmbH & Co. KG, Germany) at  $30 \text{ }^\circ\text{C}$  and a flowrate of  $2.2 \text{ ml min}^{-1}$  using  $\text{H}_2\text{O}$  and ACN as solvents. The following gradient was applied:  $90\%$  ACN for  $0.5 \text{ min}$ , increased to  $100\%$  ACN within  $10 \text{ min}$  to remain for  $12 \text{ min}$ , decrease to  $90\%$  ACN within  $0.1 \text{ min}$  to remain for another  $10 \text{ min}$ . Fractions containing the sesquiterpene of interest were evaporated under low nitrogen flow to dryness and subsequently dissolved

in the solvent of interest (hexane for GC–MS analysis or  $\text{CDCl}_3$  for NMR analysis).

### Analytics

Analysis and quantification of terpenes was performed using a Trace GC–MS Ultra system with DSQII (Thermo Scientific, USA). The sample (1  $\mu\text{l}$ , 1/10 split) was injected by a TriPlus auto sampler onto a SGE BPX5 column (30 m, I.D 0.25 mm, film 0.25  $\mu\text{m}$ ) with an injector temperature of 280  $^\circ\text{C}$ . Helium was used as carrier gas with a flow rate of 0.8  $\text{ml min}^{-1}$ . Initial oven temperature was set to 50  $^\circ\text{C}$  for 2 min. The temperature was increased to 320  $^\circ\text{C}$  at a rate of 10  $^\circ\text{C/min}^{-1}$  and then held for 3 min. MS data were recorded at 70 eV (EI) in positive mode in a range between 50 and 650. GC-FID analysis was carried out accordingly. Quantification of sesquiterpenes was carried out by correlation of the FID peak area to a defined  $\alpha$ -humulene standard of known quantity as previously described [19].

Purified compounds for further NMR analysis were dissolved in  $\text{CDCl}_3$ .  $^{13}\text{C}$  NMR spectra were measured with a Bruker Avance-III 500 MHz spectrometer equipped with a cryo probe head (5 mm CPQNP,  $^1\text{H}/^{13}\text{C}/^{31}\text{P}/^{19}\text{F}/^{29}\text{Si}$ ; Z-gradient).  $^1\text{H}$  NMR spectra as well as 2D experiments (HSQC, HMBC, COSY, NOESY) were obtained on an Avance-I 500 MHz system with an inverse probehead (5 mm SEI;  $^1\text{H}/^{13}\text{C}$ ; Z-gradient). The temperature was set to 300 K. Resulting data were processed and analyzed by TOPSPIN 3.2 or MestreNova 11.0. Chemical shifts were given in ppm relative to  $\text{CDCl}_3$  ( $\delta=7.26$  ppm for  $^1\text{H}$  and  $\delta=77.16$  ppm for  $^{13}\text{C}$  spectra).

### Protein expression and purification for crystallization experiments

The codon optimized genes encoding Copu5 and Copu9 were fused to an N-terminal hexa-histidine-tag in a pET-M11 vector and transformed into *E. coli* BL21 RIL DE3. Overexpression was performed using auto-induction medium at 37  $^\circ\text{C}$  until an OD  $\sim 0.7$  was reached and subsequently cooled down to 18  $^\circ\text{C}$  [48]. Cells grew 48 h and were harvested by centrifugation (10 min, 6000 rpm at 4  $^\circ\text{C}$ ). For resuspension of the cell pellets, buffer A was used (Cocu9: 20 mM Tris/HCl pH 7.5, 500 mM NaCl, 5 mM  $\text{MgCl}_2$ ; Copu5: 100 mM Tris/HCl pH 7.5, 500 mM NaCl, 5 mM  $\text{MgCl}_2$ , 10% (w/v) glycerol, 10 mM sodium pyrophosphate, 1 mM DTT). Cells were lysed by homogenization at 4  $^\circ\text{C}$  and the lysate was cleared by centrifugation (1 h, 21,000 rpm at 4  $^\circ\text{C}$ ).  $\text{Ni}^{2+}$ -NTA beads (cv  $\sim 1$  ml; GE Healthcare) were equilibrated with buffer A. Copu9 was loaded on the column and washed with 10 cv of buffer A containing additional 30 mM imidazole. Copu9 was eluted with buffer A containing 250 mM imidazole. Size exclusion chromatography was performed

with a HighLoad Superdex S200 16/60 column (GE Healthcare), equilibrated with buffer B [(Cocu9: 20 mM Tris/HCl pH 7.5, 150 mM NaCl, 5 mM  $\text{MgCl}_2$ ; Copu5: 100 mM Tris/HCl pH 7.5, 250 mM NaCl, 5 mM  $\text{MgCl}_2$ , 10% (w/v) glycerol, 10 mM sodium pyrophosphate, 1 mM DTT)]. Pooled protein fractions were concentrated with an Amicon-Ultra 30,000 cell. Calibration runs were performed with the high molecular weight standard (GE Healthcare).

### Thermal shift assay

Melting temperatures of Copu9 variants and Copu5 were measured with the Mx3005P qPCR system (Agilent) in a 96-well plate format with and without alendronate. Each well contained 8  $\mu\text{l}$  SEC buffer (either of Copu9 or Copu5, as stated in the manuscript), 10  $\mu\text{l}$  protein (0.15  $\mu\text{g } \mu\text{l}^{-1}$ ) with  $1\times$  SYPRO Orange dye (Invitrogen) end concentration and either 2  $\mu\text{l}$  water or 2  $\mu\text{l}$  alendronate (0.6 mg/ml) dissolved in water. The program consisted of three steps: step 1 was a pre-incubation for 1 min at 20  $^\circ\text{C}$ , and steps 2 and 3 were cycles comprising the temperature increase of 1  $^\circ\text{C}$  within 20 s. The temperature gradient proceeded from 25 to 95  $^\circ\text{C}$  at 1  $^\circ\text{C}$  per minute. Samples were measured in triplicates. The data was acquired with MxPro QPCR software (Agilent, Germany) and analyzed with DSF Analysis v3.0.1 tool (<ftp://ftp.sgc.ox.ac.uk/pub/biophysics>) and Graphpad Prism 5.0.0.228 (Graph Pad Software Inc.). A t-test was performed with Graphpad Prism to validate the significance of the results.

### In-vitro kinetics

Time resolved kinetics were measured using the EnzChek<sup>TM</sup> pyrophosphate assay kit (Invitrogen<sup>TM</sup>, ThermoFisher Scientific) in a 96-well-plate with a reduced volume of 200  $\mu\text{l}$ . The assay was performed as stated in the manual using 0.8–1.2  $\mu\text{M}$  of the Copu9 variants and FPP (Sigma-Aldrich) concentrations ranging from 0 to 70  $\mu\text{M}$  dissolved in methanol. Everything except the substrate was mixed and preincubated at 25  $^\circ\text{C}$  for 10 min. The substrate was added shortly before measurement and mixed for 5 s prior the first measurement. The enzymatic reaction was performed at 25  $^\circ\text{C}$  and the absorption was measured at 360 nm in 30 s increments using a CM Spark plate reader (Tecan, Germany). The data was analyzed using Graphpad Prism 5.0.0.228 (Graph Pad Software Inc.).

### Crystallization

For co-crystallization experiments, Copu9 was concentrated to 28  $\text{mg ml}^{-1}$  as measured by the absorbance at 280 nm and incubated with a 10 -fold molar excess of AHD for 30 min on ice. Initial crystals were obtained by the sitting-drop vapor-diffusion method at 18  $^\circ\text{C}$  with a



reservoir solution composed of 15.0% (w/v) polyethylene glycol 3350, 100 mM Tris/HCl at pH 8.5 and 100 mM Mg formate. Initial, inter-grown crystals were used to prepare a seed stock. With a cat whisker, seeds were transferred to a freshly prepared crystallization drop and crystallization plates were subsequently stored at 4 °C. Prior to flash-freezing in liquid nitrogen, crystals were transferred to a cryo-protectant solution composed of the reservoir solution supplemented with 25% (v/v) ethylene glycol.

### Structure determination and refinement

Synchrotron diffraction data were collected at the beamline 14.1 of the MX Joint Berlin laboratory at BESSY (Berlin, Germany). Diffraction data were processed with XDS [49] (Additional file 1: Table S1). The structure was determined by molecular replacement using the coordinates of SVS\_A2 (PDB-ID: 6TJZ [50]) as search model using the PHASER software [51]. The structure was refined by maximum-likelihood restrained refinement in PHENIX Model building [52, 53] and water picking was performed with COOT [54]. Model quality was evaluated with MolProbity [55] and the JCSG validation server (JCSG Quality Control Check v3.1). Secondary structure elements were assigned with DSSP [56] and ALSCRIPT [57] was used for secondary structure based sequence alignments. Figures were prepared using PyMOL (Schrödinger, Inc.). The DALI server [33] was used to identify structures closely related to Copu9.

### Protein modelling and phylogenetic analysis

Sequence alignments were performed using Clustal Omega [58] by employing seeded guide trees and HMM profile techniques as previously described [19]. To predict the tertiary structure of Copu5 and Copu9 the ROBETTA server was used [34]. The primary sequence of both proteins was submitted, and the server was run with standard settings. All calculated models have been analyzed and verified by SAVES 6.0 (<https://saves.mbi.ucla.edu/>) which were subsequently analyzed within the UCSF Chimera environment [59, 60]. For further comparison of the sesquiterpene synthases sequences AliView [61] was used.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-022-01791-8>.

**Additional file 1.** Additional figures and tables.

### Author contributions

MR conceived the project, designed, and performed the experiments, analyzed the data, and prepared the manuscript. ND and BL performed protein

purification, crystallization and kinetic experiments, performed diffraction experiments, refined the structure, and contributed to manuscript preparation. SH performed protein purification, TSA measurements and kinetic experiments. MH supported the analytical work. CH and WE contributed to structural analysis with NMR expertise. GS contributed with scientific expertise in data analysis. TB supervised the work and contributed to manuscript preparation and finalization. MR and ND contributed equally to this manuscript. All authors read and approved the final manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL. MR, MH, and TB gratefully acknowledge funding by the Werner Siemens foundation for establishing the field of Synthetic Biotechnology at TUM. ND and BL acknowledge access to beamlines of the BESSY II storage ring (Berlin, Germany) via the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück-Centrum, the Leibniz-Institut für Molekulare Pharmakologie and Charité—Universitätsmedizin Berlin. CH and WE would like to acknowledge the support of the German ministry for Education and Research (BMBF; grant number 031A305A and 031B0823D). We are grateful to Gunther Stier for sharing the pET-M11 vector. We are grateful to Elvenstar's Pukipon for the kind donation of cat whiskers. The graphical abstract was created with BioRender.com. We acknowledge C. Freund for access to the plate reader.

### Availability of data and materials

All data generated or analyzed during this study are included in this article and the supporting information. Coordinates and structure factors have been deposited in the PDB (7OFL). Diffraction images have been deposited at [proteindiffraction.org \(https://doi.org/10.18430/m37OFL\)](https://doi.org/10.18430/m37OFL).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, Lichtenbergstr. 4, 85748 Garching, Germany. <sup>2</sup>Institute for Chemistry and Biochemistry, Structural Biochemistry Laboratory, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany. <sup>3</sup>Bavarian NMR Center - Structural Membrane Biochemistry, Department of Chemistry, Technical University of Munich, 85748 Garching, Germany. <sup>4</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Cooper Rd, Brisbane 4702, Australia.

Received: 14 September 2021 Accepted: 7 April 2022

Published online: 19 April 2022

### References

1. Demail AL, Martens E. Production of valuable compounds by molds and yeasts. *J Antibiot*. 2017;70:347–60. <https://doi.org/10.1038/ja.2016.121>.
2. Christianson DW. Structural and chemical biology of terpenoid cyclases. *Chem Rev*. 2017;117:11570–648. <https://doi.org/10.1021/acs.chemrev.7b00287>.
3. Jaeger R, Cuny E. Terpenoids with special pharmacological significance: a review. *Nat Prod Commun*. 2016;11:1934578X1601100946.
4. Joshee N, Dhekney SA, Parajuli P. Medicinal plants: from farm to pharmacy. Cham: Springer; 2019.
5. Meshnick SR. Artemisinin: mechanisms of action, resistance and toxicity. *Int J Parasitol*. 2002;32:1655–60. [https://doi.org/10.1016/S0020-7519\(02\)00194-7](https://doi.org/10.1016/S0020-7519(02)00194-7).

6. Driller R, Janke S, Fuchs M, Warner E, Mhashal AR, Major DT, et al. Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis. *Nat Commun*. 2018;9:3971. <https://doi.org/10.1038/s41467-018-06325-8>.
7. Baer P, Rabe P, Fischer K, Citron CA, Klapschinski TA, Groll M, Dickschat JS. Induced-fit mechanism in class I terpene cyclases. *Angew Chem Int Ed Engl*. 2014;53:7652–6. <https://doi.org/10.1002/anie.201403648>.
8. Janke R, Görner C, Hirte M, Brück T, Loll B. The first structure of a bacterial diterpene cyclase: CotB2. *Acta Crystallogr D Biol Crystallogr*. 2014;70:1528–37. <https://doi.org/10.1107/S1399004714005513>.
9. Rynkiewicz MJ, Cane DE, Christianson DW. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc Natl Acad Sci*. 2001;98:13543–8. <https://doi.org/10.1073/pnas.231313098>.
10. Dixit M, Weitman M, Gao J, Major DT. Chemical control in the battle against fidelity in promiscuous natural product biosynthesis: the case of trichodiene synthase. *ACS Catal*. 2017;7:812–8. <https://doi.org/10.1021/acscatal.6b02584>.
11. Major DT. Electrostatic control of chemistry in terpene cyclases. *ACS Catal*. 2017;7:5461–5. <https://doi.org/10.1021/acscatal.7b01328>.
12. Borg-Karlson A-K, Norin T, Talvitie A. Configurations and conformations of torreyol ( $\delta$ -cadinol),  $\alpha$ -cadinol, T-murolol and T-cadinol. *Tetrahedron*. 1981;37:425–30. [https://doi.org/10.1016/S0040-4020\(01\)92031-9](https://doi.org/10.1016/S0040-4020(01)92031-9).
13. Kundu A, Saha S, Walia S, Shakil NA, Kumar J, Annapurna K. Cadinene sesquiterpenes from *Eupatorium adenophorum* and their antifungal activity. *J Environ Sci Health B*. 2013;48:516–22. <https://doi.org/10.1080/03601234.2013.761921>.
14. Bande-Borujeni S, Zandi-Sohani N, Ramezani L. Chemical composition and bioactivity of essential oil from *Eucalyptus occidentalis* leaves against two stored product pests. *Int J Trop Insect Sci*. 2018;38:216–23. <https://doi.org/10.1017/S1742758418000085>.
15. Mulyaningsih S, Youns M, El-Readi MZ, Ashour ML, Nibret E, Sporer F, et al. Biological activity of the essential oil of *Kadsura longipedunculata* (Schisandraceae) and its major components. *J Pharm Pharmacol*. 2010;62:1037–44. <https://doi.org/10.1111/j.2042-7158.2010.01119.x>.
16. Chang H-T, Cheng Y-H, Wu C-L, Chang S-T, Chang T-T, Su Y-C. Antifungal activity of essential oil and its constituents from *Calocedrus macrolepis* var. *formosana* florin leaf against plant pathogenic fungi. *Bioresour Technol*. 2008;99:6266–70. <https://doi.org/10.1016/j.biortech.2007.12.005>.
17. Yap H-YY, Muria-Gonzalez MJ, Kong B-H, Stubbs KA, Tan C-S, Ng S-T, et al. Heterologous expression of cytotoxic sesquiterpenoids from the medicinal mushroom *Lignosus rhinocerotis* in yeast. *Microb Cell Fact*. 2017;16:103. <https://doi.org/10.1186/s12934-017-0713-x>.
18. Zhou H, Yang Y-L, Zeng J, Zhang L, Ding Z-H, Zeng Y. Identification and characterization of a  $\delta$ -cadinol synthase potentially involved in the formation of boreovibrins in *Boreostereum vibrans* of Basidiomycota. *Nat Prod Bioprospect*. 2016;6:167–71. <https://doi.org/10.1007/s13659-016-0096-4>.
19. Mischko W, Hirte M, Fuchs M, Mehlmer N, Brück TB. Identification of sesquiterpene synthases from the Basidiomycota *Coniophora puteana* for the efficient and highly selective  $\beta$ -copaene and cubebol production in *E. coli*. *Microb Cell Fact*. 2018;17:164. <https://doi.org/10.1186/s12934-018-1010-z>.
20. Kemper K, Hirte M, Reinbold M, Fuchs M, Brück T. Opportunities and challenges for the sustainable production of structurally complex diterpenoids in recombinant microbial systems. *Beilstein J Org Chem*. 2017;13:845–54. <https://doi.org/10.3762/bjoc.13.85>.
21. Hirte M, Mischko W, Kemper K, Röhrer S, Huber C, Fuchs M, et al. From microbial upcycling to biology-oriented synthesis: combining whole-cell production and chemo-enzymatic functionalization for sustainable taxanoid delivery. *Green Chem*. 2018;20:5374–84. <https://doi.org/10.1039/c8gc03126f>.
22. Sieber P, Voigt K, Kämmer P, Brunke S, Schuster S, Linde J. Comparative study on alternative splicing in human fungal pathogens suggests its involvement during host invasion. *Front Microbiol*. 2018;9:2313. <https://doi.org/10.3389/fmicb.2018.02313>.
23. Da Lage J-L, Binder M, Hua-Van A, Janeček S, Casane D. Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial  $\alpha$ -amylase gene to basidiomycetes. *BMC Evol Biol*. 2013;13:40. <https://doi.org/10.1186/1471-2148-13-40>.
24. Linstrom P. NIST chemistry WebBook, NIST standard reference database 69. Gaithersburg: National Institute of Standards and Technology; 1997.
25. Steele CL, Crock J, Bohlmann J, Croteau R. Sesquiterpene synthases from grand fir (*Abies grandis*). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. *J Biol Chem*. 1998;273:2078–89. <https://doi.org/10.1074/jbc.273.4.2078>.
26. Yoshikuni Y, Martin VJJ, Ferrin TE, Keasling JD. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chem Biol*. 2006;13:91–8. <https://doi.org/10.1016/j.chembiol.2005.10.016>.
27. Raz K, Levi S, Gupta PK, Major DT. Enzymatic control of product distribution in terpene synthases: insights from multiscale simulations. *Curr Opin Biotechnol*. 2020;65:248–58. <https://doi.org/10.1016/j.copbio.2020.06.002>.
28. López-Gallego F, Wawrzyn GT, Schmidt-Dannert C. Selectivity of fungal sesquiterpene synthases: role of the active site's H-1 alpha loop in catalysis. *Appl Environ Microbiol*. 2010;76:7723–33. <https://doi.org/10.1128/AEM.01811-10>.
29. Lopez-Gallego F, Agger SA, Abate-Pella D, Distefano MD, Schmidt-Dannert C. Sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus*: catalytic promiscuity and cyclization of farnesyl pyrophosphate geometric isomers. *ChemBioChem*. 2010;11:1093–106. <https://doi.org/10.1002/cbic.200900671>.
30. Ali NAA, Wurster M, Denkert A, Arnold N, Fadail I, Al-Didamony G, et al. Chemical composition, antimicrobial, antioxidant and cytotoxic activity of essential oils of *Plectranthus cylindraceus* and *Meriandra benghalensis* from Yemen. *Nat Prod Commun*. 2012;7:1099–102.
31. Moiteiro C, Esteves T, Ramalho L, Rojas R, Alvarez S, Zacchino S, Bragança H. Essential oil characterization of two Azorean *Cryptomeria japonica* populations and their biological evaluations. *Nat Prod Commun*. 2013;8:1785–90.
32. Dickschat JS. Bacterial terpene cyclases. *Nat Prod Rep*. 2016;33:87–110. <https://doi.org/10.1039/c5np00102a>.
33. Holm L. DALI and the persistence of protein shape. *Protein Sci*. 2020;29:128–40. <https://doi.org/10.1002/pro.3749>.
34. Hiranova N, Park H, Baek M, Anishchenko I, Dauparas J, Baker D. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat Commun*. 2021;12:1340. <https://doi.org/10.1038/s41467-021-21511-x>.
35. Quin MB, Flynn CM, Schmidt-Dannert C. Traversing the fungal terpenome. *Nat Prod Rep*. 2014;31:1449–73. <https://doi.org/10.1039/c4np00075g>.
36. Driller R, Garbe D, Mehlmer N, Fuchs M, Raz K, Major DT, et al. Current understanding and biotechnological application of the bacterial diterpene synthase CotB2. *Beilstein J Org Chem*. 2019;15:2355–68. <https://doi.org/10.3762/bjoc.15.228>.
37. Görner C, Häuslein I, Schrepfer P, Eisenreich W, Brück T. Targeted engineering of cyclooctat-9-en-7-ol synthase: a stereospecific access to two new non-natural fusicoccane-type diterpenes. *ChemCatChem*. 2013;5:3289–98. <https://doi.org/10.1002/cctc.201300285>.
38. Raz K, Driller R, Brück T, Loll B, Major DT. Understanding the role of active site residues in CotB2 catalysis using a cluster model. *Beilstein J Org Chem*. 2020;16:50–9. <https://doi.org/10.3762/bjoc.16.7>.
39. Schrepfer P, Buettner A, Goerner C, Hertel M, van Rijn J, Wallrapp F, et al. Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proc Natl Acad Sci USA*. 2016;113:E958–67. <https://doi.org/10.1073/pnas.1519680113>.
40. Karunanithi PS, Zerbe P. Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Front Plant Sci*. 2019;10:1166. <https://doi.org/10.3389/fpls.2019.01166>.
41. Peters RJ, Carter OA, Zhang Y, Matthews BW, Croteau RB. Bifunctional abi-etiadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. *Biochemistry*. 2003;42:2700–7. <https://doi.org/10.1021/bi020492n>.
42. Kawaide H, Sassa T, Kamiya Y. Functional analysis of the two interacting cyclase domains in ent-kaurene synthase from the fungus *Phaeosphaeria* sp. L487 and a comparison with cyclases from higher plants. *J Biol Chem*. 2000;275:2276–80. <https://doi.org/10.1074/jbc.275.4.2276>.
43. Pemberton TA, Chen M, Harris GG, Chou WKW, Duan L, Köksal M, et al. Exploring the influence of domain architecture on the catalytic function

- of diterpene synthases. *Biochemistry*. 2017;56:2010–23. <https://doi.org/10.1021/acs.biochem.7b00137>.
44. Jia Q, Li G, Köllner TG, Fu J, Chen X, Xiong W, et al. Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proc Natl Acad Sci USA*. 2016;113:12328–33. <https://doi.org/10.1073/pnas.1607973113>.
  45. Gross H, König GM. Terpenoids from marine organisms: unique structures and their pharmacological potential. *Phytochem Rev*. 2006;5:115–41. <https://doi.org/10.1007/s11101-005-5464-3>.
  46. Biggs BW, Lim CG, Sagliani K, Shankar S, Stephanopoulos G, de Mey M, Ajikumar PK. Overcoming heterologous protein interdependency to optimize P450-mediated taxol precursor synthesis in *Escherichia coli*. *Proc Natl Acad Sci USA*. 2016;113:3209–14. <https://doi.org/10.1073/pnas.1515826113>.
  47. Ringel M, Reinbold M, Hirte M, Haack M, Huber C, Eisenreich W, et al. Towards a sustainable generation of pseudopterosin-type bioactives. *Green Chem*. 2020;22:6033–46. <https://doi.org/10.1039/d0gc01697g>.
  48. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005;41:207–34. <https://doi.org/10.1016/j.pep.2005.01.016>.
  49. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010;66:125–32. <https://doi.org/10.1107/S0907444909047337>.
  50. Schriever K, Saenz-Mendez P, Rudraraju RS, Hendrikse NM, Hudson EP, Biundo A, et al. Engineering of ancestors as a tool to elucidate structure, mechanism, and specificity of extant terpene cyclase. *J Am Chem Soc*. 2021;143:3794–807. <https://doi.org/10.1021/jacs.0c10214>.
  51. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40:658–74. <https://doi.org/10.1107/S0021889807021206>.
  52. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66:213–21. <https://doi.org/10.1107/S0907444909052925>.
  53. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*. 2012;68:352–67. <https://doi.org/10.1107/S0907444912001308>.
  54. Casañal A, Lohkamp B, Emsley P. Current developments in coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci*. 2020;29:1069–78. <https://doi.org/10.1002/pro.3791>.
  55. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci*. 2018;27:293–315. <https://doi.org/10.1002/pro.3330>.
  56. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22:2577–637. <https://doi.org/10.1002/bip.360221211>.
  57. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng*. 1993;6:37–40. <https://doi.org/10.1093/protein/6.1.37>.
  58. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res*. 2013;41:W597–600. <https://doi.org/10.1093/nar/gkt376>.
  59. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–12. <https://doi.org/10.1002/jcc.20084>.
  60. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein structure modeling using modeller. *Curr Protoc Bioinformatics*. 2006. <https://doi.org/10.1002/0471250953.bi0506s15>.
  61. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30:3276–8. <https://doi.org/10.1093/bioinformatics/btu531>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



## Additional file 1

# Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of $\delta$ -cadinol

Marion Ringel<sup>1</sup> ‡, Nicole Dimos<sup>2</sup> ‡, Stephanie Himpich<sup>2</sup>, Martina Haack<sup>1</sup>, Claudia Huber<sup>3</sup>, Wolfgang Eisenreich<sup>3</sup>, Gerhard Schenk<sup>4</sup>, Bernhard Loll<sup>2\*</sup> and Thomas Brück<sup>1\*</sup>

<sup>1</sup>Werner Siemens Chair of Synthetic Biotechnology, Dept. of Chemistry, Technical University of Munich (TUM), Lichtenbergstr. 4, 85748 Garching, Germany

<sup>2</sup>Institute for Chemistry and Biochemistry, Structural Biochemistry Laboratory, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany

<sup>3</sup>Bavarian NMR Center—Structural Membrane Biochemistry, Department of Chemistry, Technische Universität München, 85748 Garching, Germany

<sup>4</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Cooper Rd, 4702 Brisbane, Australia



## Fig. S1. Amino acid sequences

### >Copu3 [1]

MSATPAPTEFILPNLFSVCPLTFGRSNPYDEVIPEARAWIAKYNPFVDSKRAEFVQGCNELLCSRVPYAGREEFR  
CCDFVNLLFVLDLSDDMGGADARSTCDSFIRVLNDPDPDTSLIAQMTREFRARVAERAKPGCLRRFIALCGTYVE  
AVCVEAELREQRVLDRSFILLRENSAVRCLLAEYALGLELPDAVFNDPAFQSVYFCAADMVCWSNDVYSYN  
MEQAKGHTGNNVVTVLMQEHGIDLQAAADRVEVFGQLMEHYTSGSRSLPTWGGKVDADAARFLEAAGQWV  
VGNLEWSFETPRYFGPDHDEVDRDTHRVLK

### >Copu5

MHLPEPFHLLPDFSSHCSYPLRLNKHCVAAAAASEDWLIRLAQLRSPRNGRKLKFMGLKAGYLTALCYPDCPRTE  
LRVVS DYMNFLFTLDDWSDEFAEAGVRLGECVMGMLYDPTVKTDKAAGRLARSFWLRMIRTAGPRVQHRFIVA  
FEDFFRAVEQQSRDRAGVMPDLESYIALRRDTSRCPVFLAEYAAGIELPDEVFEHPHIIQSMTEATNDLVTWSND  
VFSYNKEQALGDTNMITLLMAQHGLSLQGAVDFVGLCAASITRFESGRITLPSWGPVDCDVQKYVMGLQD  
WIAGSLHWSFETERYFGKRGKEVRQAGVVKLSPMKAPKV

### >Copu6

MLQLPSHFVLQDLCAISGRACELKVSPLQREAGALATKWFDSSIGVYDEIKFTKFTKFGKFDLFAALSPEADLRHLETC  
LMFFFWAFSTDDLDEGALQNRPDEVQAGHDVSNVIDHPEAPRPAYPYAAMLYDLLERFRETGTEGAYARFIRAF  
EDWSESQVQSQNRSEDRMPSIHEFILMRRATIGGAMVEAMIEYSLDIDLDFIFEHPTIIMSEATNDIMTWPND  
LCSFNKEQADGDYQNLVFIIMEERGVLQEGIDILDMLSQRVDDYLKASLPSFGPKVDYELARYLKALEHFTQGT  
VLWYYLSPRYFRTVDVSNRHNLVPLFPQSFH

### >Copu7

MPNRYLPDTMVTWPWKTRTNPHADEVEEKVCEWCNAFPMMTKAYKLLKPDHLLVSLFVLVILVDDCTDAEN  
ADVARKTADLVKDAFEHSDQPRPAGEGPIGEIVRRFWQFSIGVITANVQVAFLLKHFDEFLDSIVTQAGQRDEDVRLS  
VDKYLKLRDNDVGVMPFFPFLRTTSGPDLPEEIWDSAVIAEMTGHIIDMYIFDNDTISYGREYALGDTGHNIVTLLM  
QEHGIDVGSVAWATARHAAAQKAFKDGRLERLPSLGS HADAQVKEYLNGLGYWIRAYHVWSFKIERFYDGRGDE  
VKASRLVQLKPPASSVQSSAH

### >Copu9

MSPTATFTTTSSEENAPTKFILPDLVSDCTYPLLLNDNCEPVARASEQWLIAGARLQEPRTKFMGLLAGELTAACYP  
HADASHLRVCVDFMNWLFNMDDWLDDFDVDDTWGMRHCLGAFRDPVGFETDKLGLLMSKSFSSFRFRQDGG  
PGCTERFIHTMDLFFIAVAQQAGDRANGITPDLESYITVRRDTSRCPVFLAEYAAGIDLPDHVYHPTLAAMEEAT  
NDLVTWSNDIFSYNKEQVTDTHNMIPVLMRERGLDLQAVDFVGRCLCKGTIERFETERARLPSWGPPELDAQVQT  
YIEGLQNWIVGSLHWSFDSHRYFGKDGHAVKKHRIVKLLPKRVPQQA

### >Copu10

MSPSPTRFYLPDTMASWPWKTRTNPHADEVGGGDASSGHSLLSEFSRLIQLVHSFDHLLVSLFVLVILVDDCT  
DTENADVARTADLVNDAFEHSDQPRPAGEGPIGEIVRRFWQYAGIISPQVQVAFLLKHFGEFLESIVTQAGQRDED  
VRLNVDTYLKLRRDNDVGVMPFFPFLRPTSGPDLPEEIWESAVIAEMTGHIIDMYIFDNDTIFYGREYALGDTGHNIV  
TLLMQEYIDVGSVAWATARHAAAQKAFKDGRLERLPSHGAQADAQVKEYLNGLGYWIRAYHVWSFKIERFYDGR  
GDEVKVSRLVQLKPP

### >Copu11

MTSWPWKTKVNPYANAVRQRCAEWCSFPMMAKAYKLLDPNQLLVSLFSLVIVLVDDSTDVEDAAAAQQTAT  
LVRDALQHPHQARPFGEAPIGEIVRRFWQLAIQTTHVDVQSAFLTHFADFLESVVAQAGFRDHDSQLSIDAYLTIRR  
DTVGVMPFFPFLRPSADDRATENVWNSPIIAELTGYIVDMYIYDNDTISYAREHALGDIGHNIITLLMRDLNIDLGSA  
VSWAVMQHAVAQRAFIEGIARLPSWDADVDRQVRDYLDGLGHWAKAYHTWAFEVERYFGDRGKEVKATGLVR  
LRRS



copu3	-----
copu5	PKKV---
copu6	H-----
copu7	SVQSSAH
copu9	PQQA---
copu10	-----
copu11	-----



## Terpene structure elucidation: NMR Results

- Copu9:  $^1\text{H}$

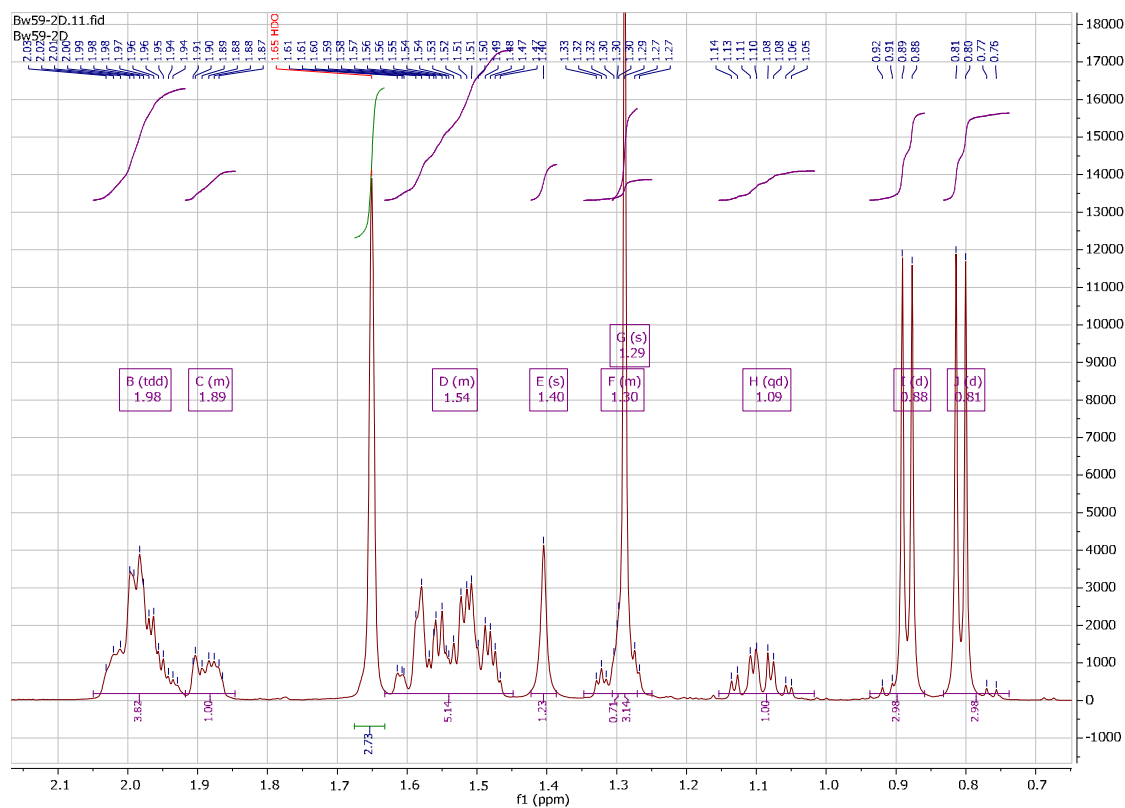


Fig. S4.  $^1\text{H}$  NMR spectrum of (+)- $\delta$ -cadinol produced by Copu9; Standardised on solvent ( $\text{CDCl}_3$ ) peak:  $^1\text{H} = 7.26$  ppm,  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  5.51 (dq,  $J = 5.3, 1.6$  Hz, 1H), 1.98 (m,  $J = 13.9, 10.4, 6.2$  Hz, 4H), 1.92 – 1.85 (m, 1H), 1.66 (s, 3H); 1.63 – 1.45 (m, 5H), 1.35 – 1.25 (m, 1H), 1.29 (s, 3H), 1.09 (qd,  $J = 13.2, 4.2$  Hz, 1H), 0.88 (d,  $J = 6.9$  Hz, 3H), 0.81 (d,  $J = 6.9$  Hz, 3H).

▪ Copu9:  $^{13}\text{C}$

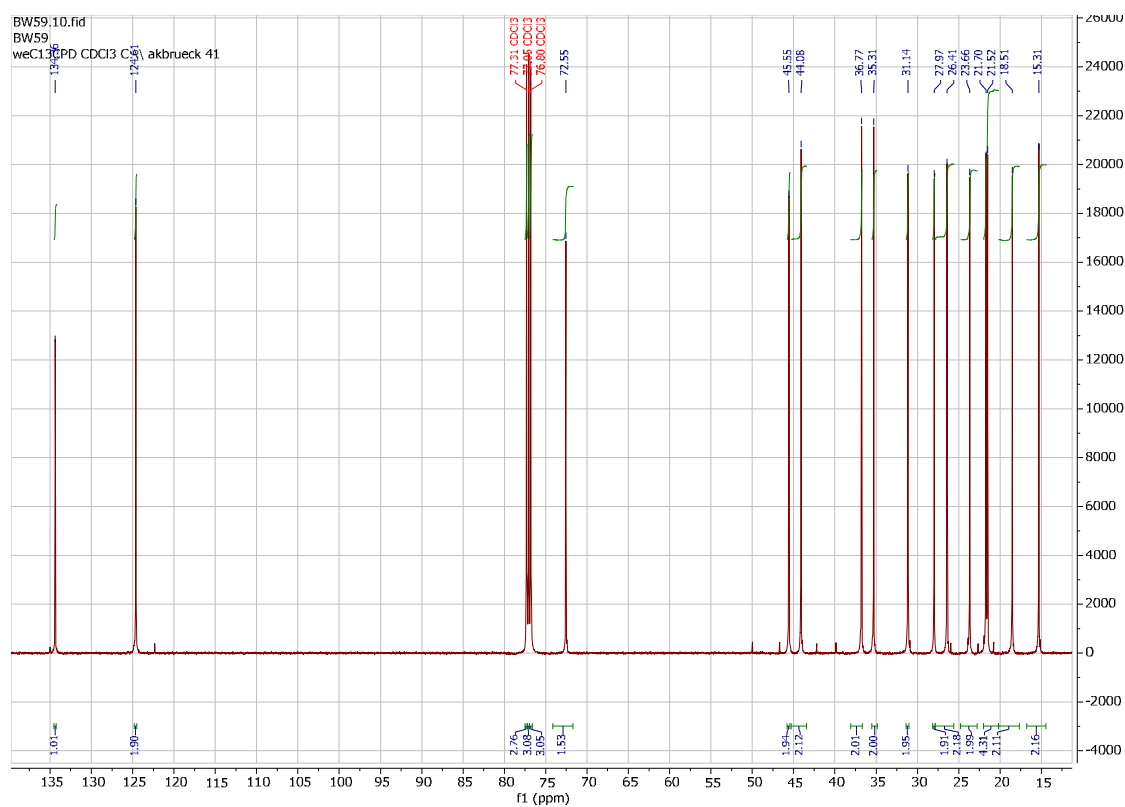


Fig. S5.  $^{13}\text{C}$  NMR spectrum of (+)- $\delta$ -cadinol produced by Copu9; Standardised on solvent ( $\text{CDCl}_3$ ) peak:  $^{13}\text{C} = 77.2$  ppm;  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  134.36, 124.61, 72.55, 45.55, 44.08, 36.77, 35.31, 31.14, 27.97, 26.41, 23.66, 21.70, 21.52, 18.51, 15.31.

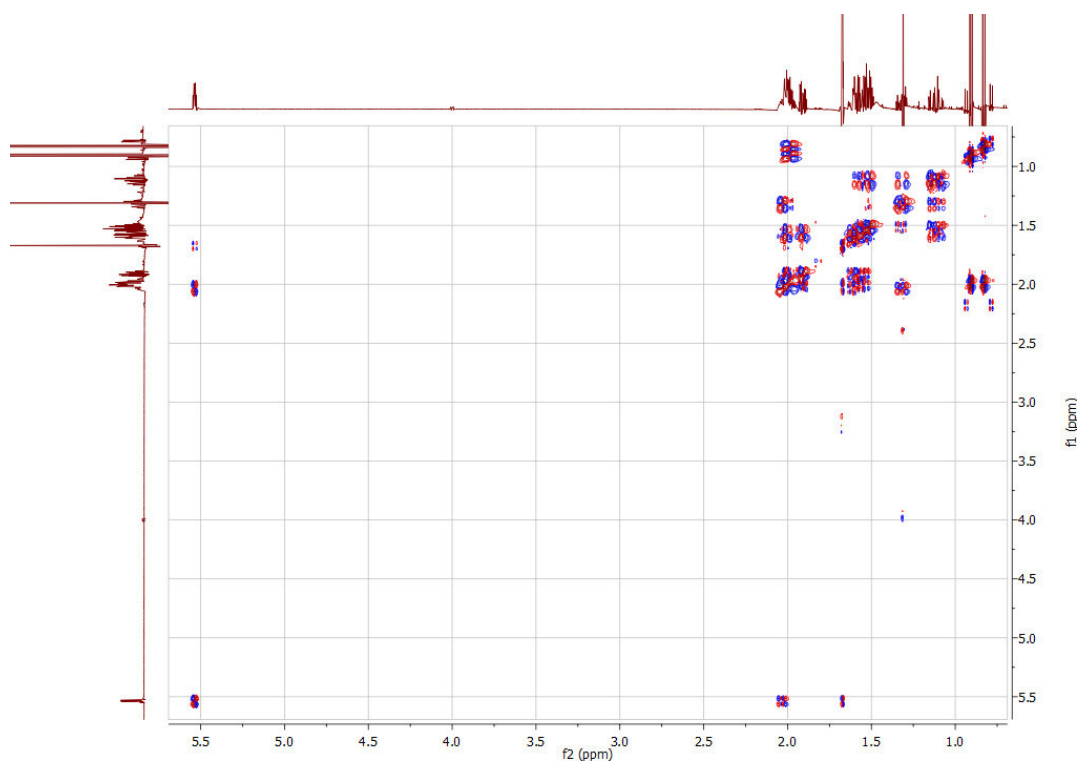


Fig. S6.  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of (+)- $\delta$ -cadinol produced by Copu9

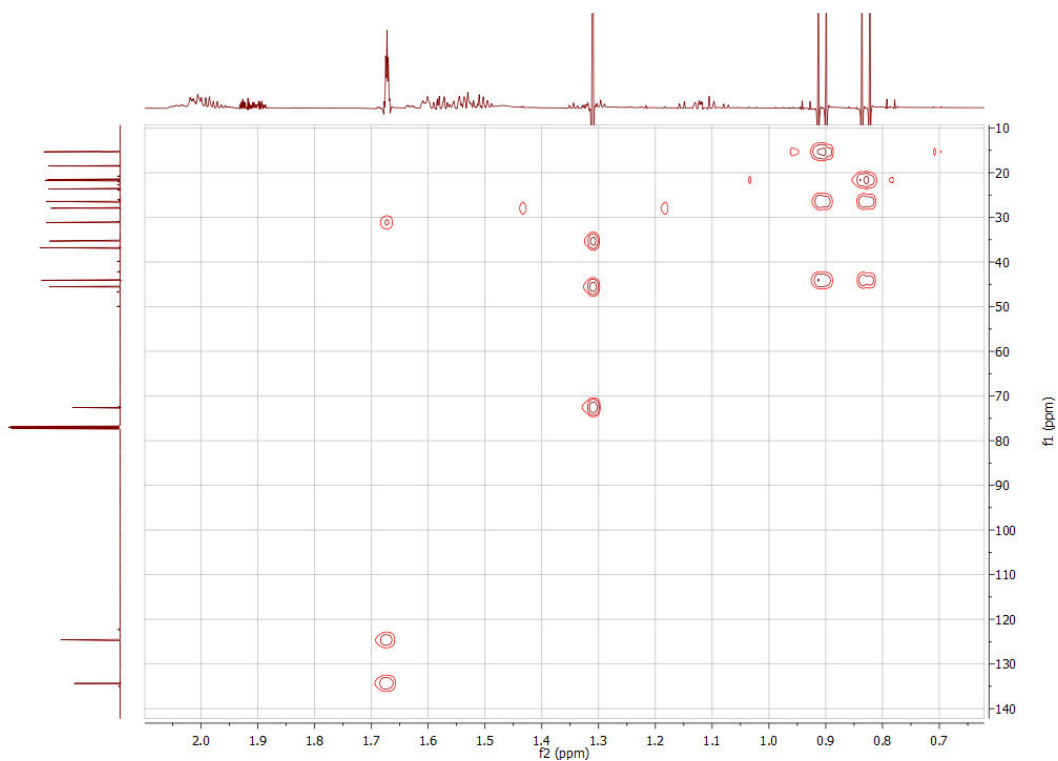


Fig. S7.  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectrum of (+)- $\delta$ -cadinol produced by Copu9

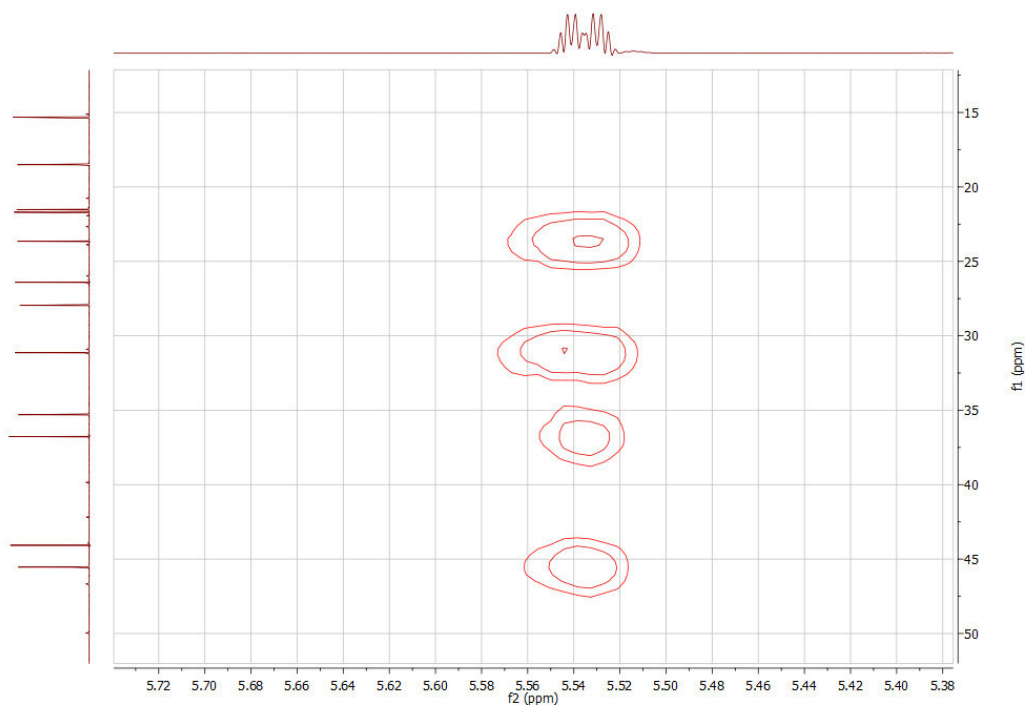


Fig. S8.  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectrum of (+)- $\delta$ -cadinol produced by Copu9

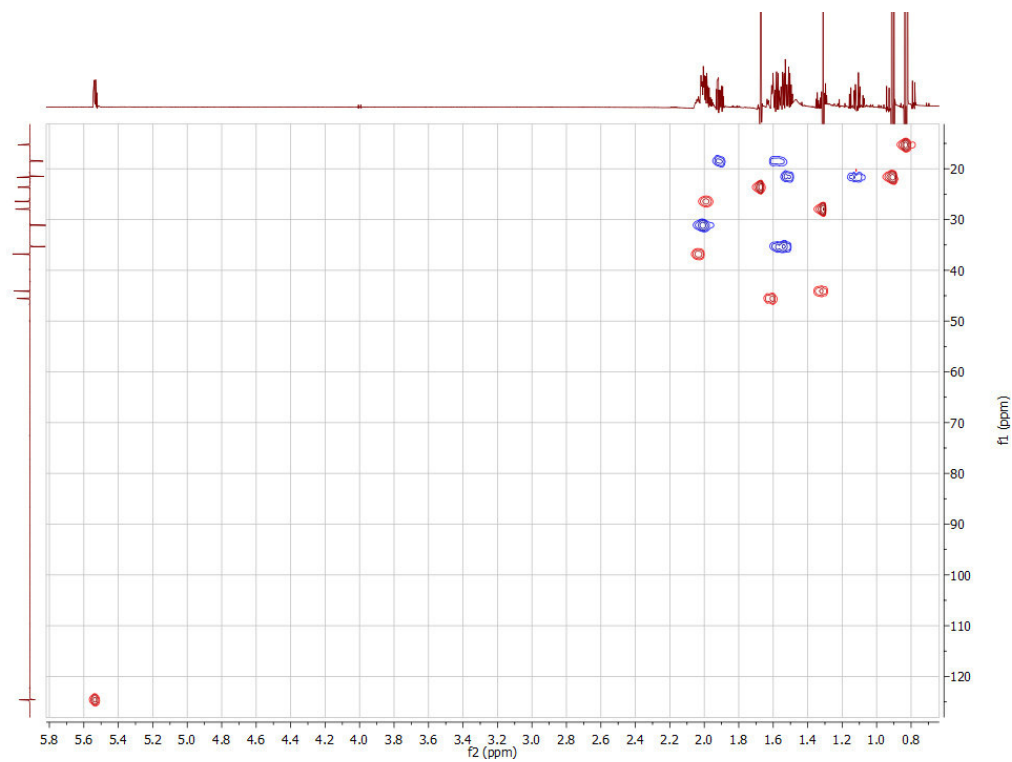


Fig. S9.  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of (+)- $\delta$ -cadinol produced by Copu9



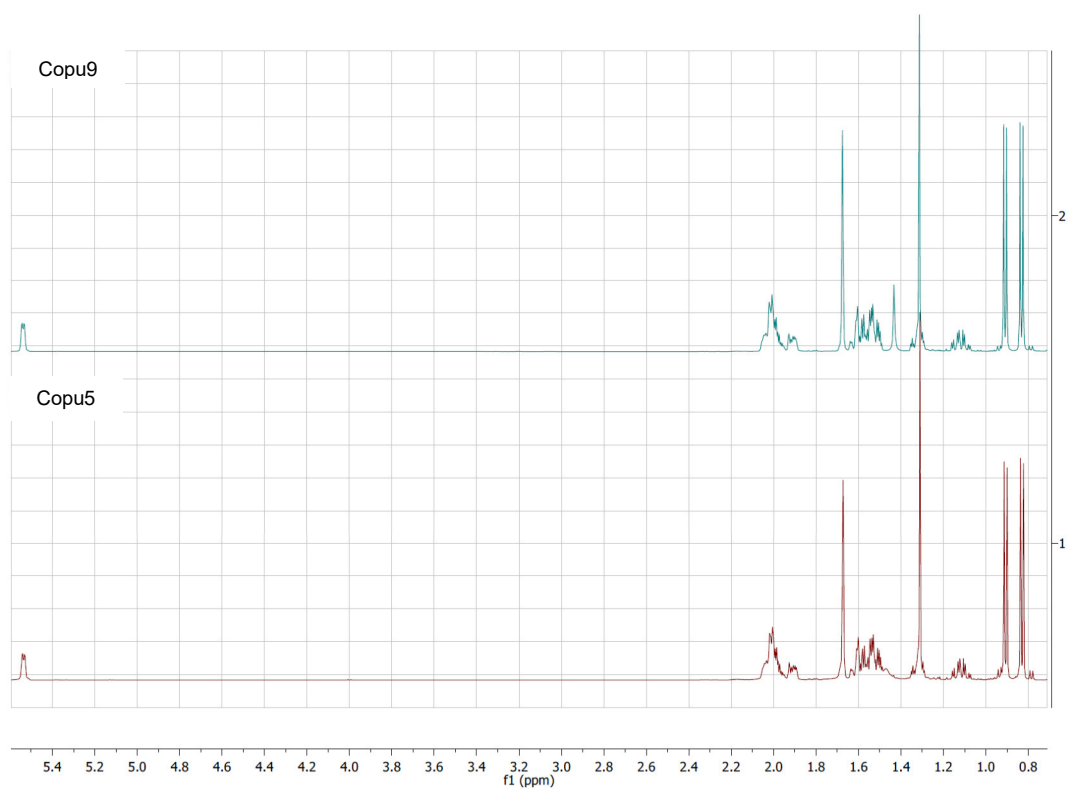


Fig. S10. Comparison of <sup>1</sup>H-NMR spectra of (+)- $\delta$ -cadinol produced by Copu9 and Copu5

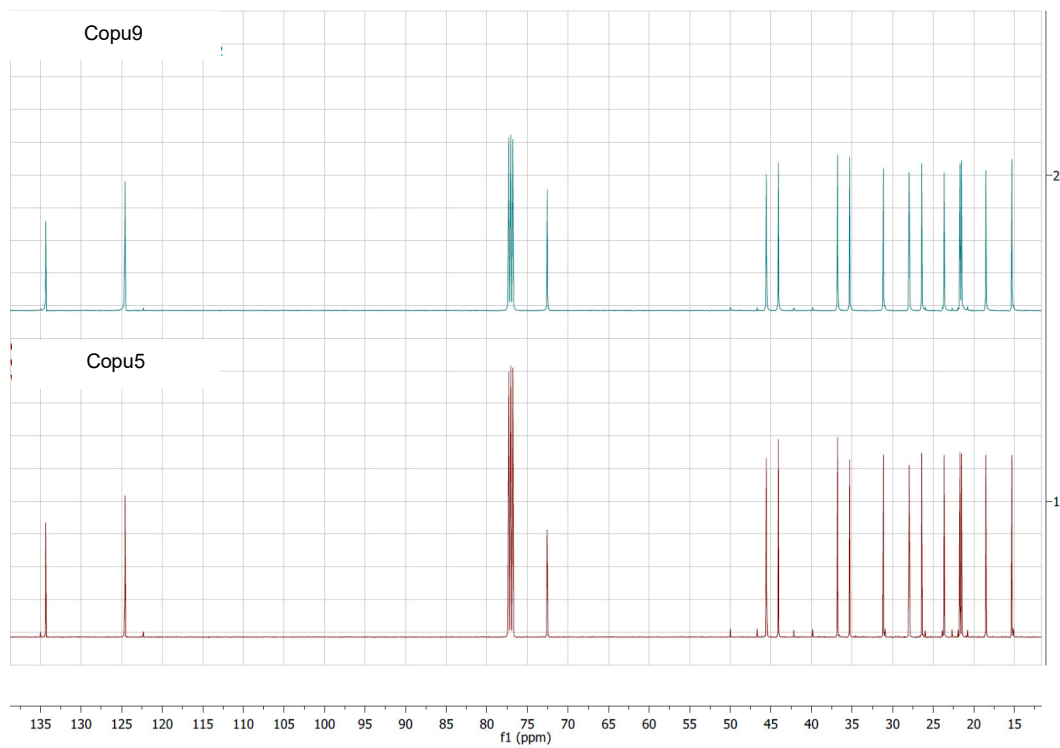


Fig. S11. Comparison of <sup>13</sup>C-NMR spectra of (+)- $\delta$ -cadinol produced by Copu9 and Copu5

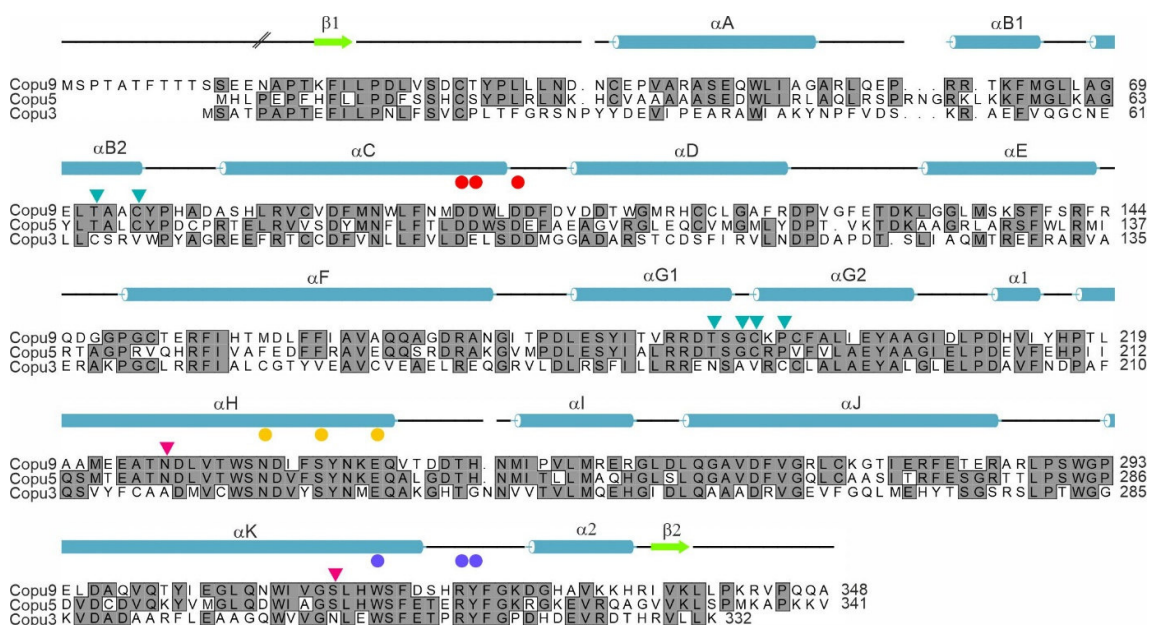


Fig. S12. Structure-based sequence alignment of Copu9, Copu5 and Copu3. On top of the primary sequence of Copu9 the secondary structure elements are drawn. Slashed lines at the N-terminus indicate the terminal residue, which is included in the crystal structure. The aspartate-rich motif (<sup>99</sup>DDWLD<sup>103</sup>) is highlighted with red spheres and the <sup>218</sup>NSE<sup>236</sup> motif with yellow spheres. The <sup>317</sup>WxxxxxRY<sup>324</sup> motif (with x as any amino acid) is indicated by blue spheres. Copu9 residues subjected to side directed mutagenesis are highlighted with triangles. Residues that have been subjected to mutagenesis, without altering the product, are shown as teal triangles, whereas single residue exchanges with an effect on the product profile are drawn as magenta triangles.

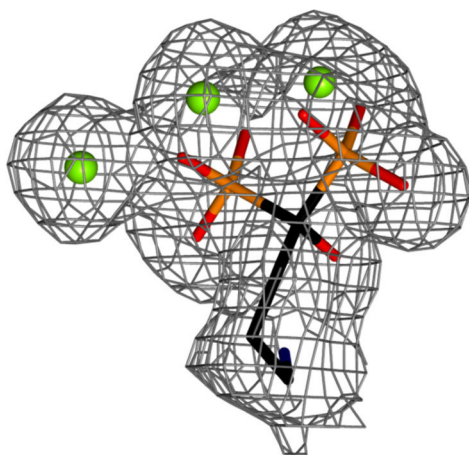


Fig. S13. Polder electron density map[2] shown as mesh at a  $\sigma$ -level of 2.0. AHD is presented as ball-stick-model with carbon atoms colored in black, oxygen in red, phosphorous in orange and nitrogen in light blue. Mg<sup>2+</sup> ions are shown as green spheres.

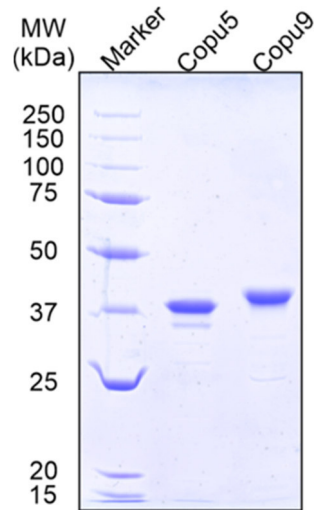


Fig. S 14: Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) of the purified Copu5 (theoretical mass including the His<sub>6</sub>-tag 41.9 kDa) and Copu9 (theoretical mass including the His<sub>6</sub>-tag 42.7 kDa) with Precision Plus unstained protein standard (Bio-Rad) on the left lane.

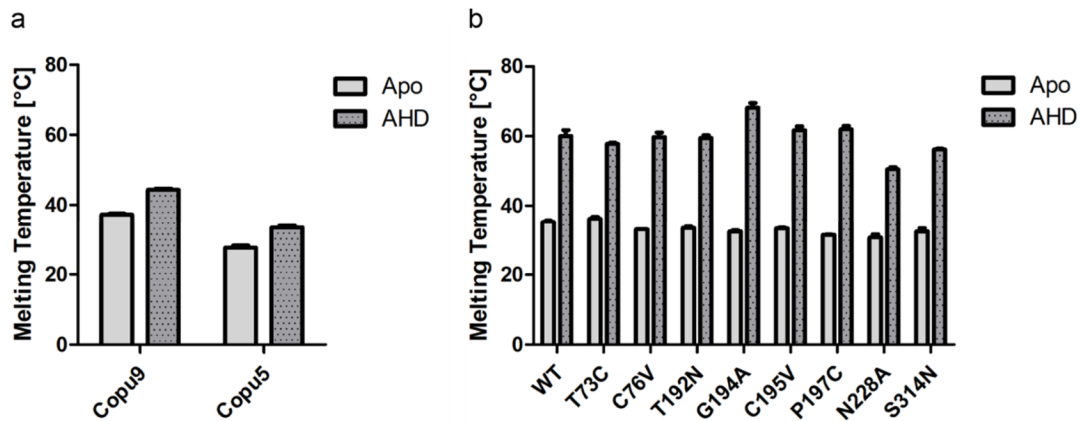


Fig. S15: Thermal shift analysis of Copu5 and Copu9 in pyrophosphate containing buffer with and without AHD (a) and Copu9 and its variants in Copu9 buffer with and without AHD

Table S1. Crystallographic data collection and model refinement statistics

<b>Dataset</b>	Copu9•Mg <sup>2+</sup> <sub>3</sub> •AHD
PDB entry	7OFL
<b>Data Collection</b>	
Wavelength [Å]	0.9184
Temperature [K]	100
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2
Unit Cell Parameters	
a, b, c [Å]	74.45 76.62 135.11
α, β, γ [°]	90.0 90.0 90.0
Resolution [Å] <sup>a</sup>	50.00 - 1.83 (1.94 - 1.83)
Reflections <sup>a</sup>	
Unique <sup>a</sup>	68,244 (10,301)
Completeness [%] <sup>a</sup>	98.6 (93.3)
Multiplicity <sup>a</sup>	7.3 (7.3)
Data quality <sup>a</sup>	
Intensity [I/σ(I)] <sup>a</sup>	8.73 (1.01)
R <sub>meas</sub> [%] <sup>a, b</sup>	52.5 (302.0)
CC <sub>1/2</sub> <sup>a, c</sup>	98.9 (44.0)
Wilson B value [Å <sup>2</sup> ]	35.7
<b>Refinement</b>	
Resolution [Å] <sup>a</sup>	50.00 - 1.83 (1.88 - 1.83)
Reflections <sup>a</sup>	
Number	67671
Test Set [%]	3.1
R <sub>work</sub> [%] <sup>a</sup>	18.3 (34.9)
R <sub>free</sub> [%] <sup>a</sup>	23.1 (36.5)
Asymmetric Unit	
Protein: Residues, Atoms	335 (A), 2,742 (A) 334 (B), 2,716 (B)
Ligands: Molecules	2 (AHD), 6 (Mg <sup>2+</sup> ), 14 (EDO),
Water molecules	697
Mean Temperature factors [Å <sup>2</sup> ] <sup>b</sup>	
All Atoms	29.3
Macromolecules	26.9 (A), 30.5 (B)
Ligands	25.2 (AHD), 21.4 (Mg <sup>2+</sup> ) 40.4 (EDO),
Water molecules	33.6
RMSD from Target Geometry <sup>d</sup>	
Bond Lengths [Å]	0.013
Bond Angles [°]	1.154
<b>Validation Statistics</b>	
Ramachandran Plot <sup>f</sup>	
Residues in Allowed Regions [%]	0.6
Residues in Favored Regions [%]	99.4
Ramachandran plot Z-score <sup>f</sup> (RMSD)	
whole	-1.20 (0.28)
helix	-0.86 (0.19)
sheet	-
loop	-0.04 (0.46)
MOLPROBITY Clashscore <sup>g</sup>	3.48
MOLPROBITY score <sup>f</sup>	1.16

<sup>a</sup> data for the highest resolution shell in parenthesis

<sup>b</sup>  $R_{meas}(I) = \sum_h [N/(N-1)]^{1/2} \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i I_{ih}$ , in which  $\langle I_h \rangle$  is the mean intensity of symmetry-equivalent reflections  $h$ ,  $I_{ih}$  is the intensity of a particular observation of  $h$  and  $N$  is the number of redundant observations of reflection  $h$ [3].

<sup>c</sup>  $CC_{1/2} = (\langle I^2 \rangle - \langle I \rangle^2) / (\langle I^2 \rangle - \langle I \rangle^2) + \sigma_e^2$ , in which  $\sigma_e^2$  is the mean error within a half-dataset[4].

<sup>d</sup> RMSD – root mean square deviation

<sup>e</sup> calculated with PHENIX [5]

<sup>f</sup> calculated with MOLPROBITY [6]

<sup>g</sup> Clashscore is the number of serious steric overlaps (> 0.4 ) per 1,000 atoms[6].

Table S 2: Results of a DALI [7] search with the coordinated of the structure of Copu9•Mg<sup>2+</sup>•AHD.

PDB ID	rmsd [Å]	sequence identity [%]	Z-score	protein	TPS family	Ligand in active site	resolution [Å]	Reference
4okz	1.7	23	39.1	selinadiene synthase	sesquiterpene	dihydrofarnesyl, pyrophosphate	1.90	[8]
4lz0	2.3	22	38.1	Epi-isozizaene synthase	sesquiterpene	pyrophosphate, benzyl triethyl ammonium cation	1.75	[9]
5dz2	1.9	24	36.8	germacradienol/geomisin synthase	sesquiterpene	alendronate	2.11	[10]
5nx7	2.0	20	36.0	pentalenene synthase	monoterpene	2-fluoroneryl diphosphate	1.51	[11]
6tjz	2.1	25	36.0	spiroviolene synthase	diterpene	none	2.40	[12]
5ivg	2.1	19	34.4	aristolochene synthase	sesquiterpene	farnesyl thiolodiphosphate	1.95	[13]
6oh6	2.3	16	34.1	labdane-related diterpene synthase	diterpene	pyrophosphate	2.07	[14]

Table S 3: Superposition of the experimentally obtained crystal structure of Copu9•Mg<sup>2+</sup>•AHD, with the models for Copu9 and Copu5, respectively. The root mean square deviation (rmsd) for aligned pairs of Ca-atoms is given in [Å]. The number of aligned residues is provided in parenthesis.

PDB ID	Copu9•Mg <sup>2+</sup> •AHD	Copu9 (model)	Copu5 (model)
Copu9•Mg <sup>2+</sup> •AHD	-	1.2 (317)	1.1 (319)
Copu9 (model)		-	0.8 (321)
Copu5 (model)			-

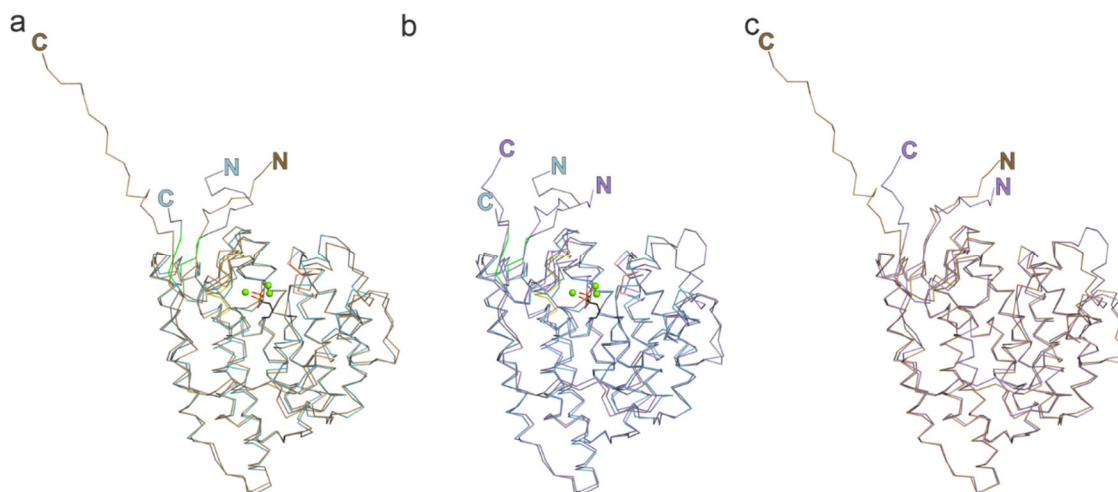
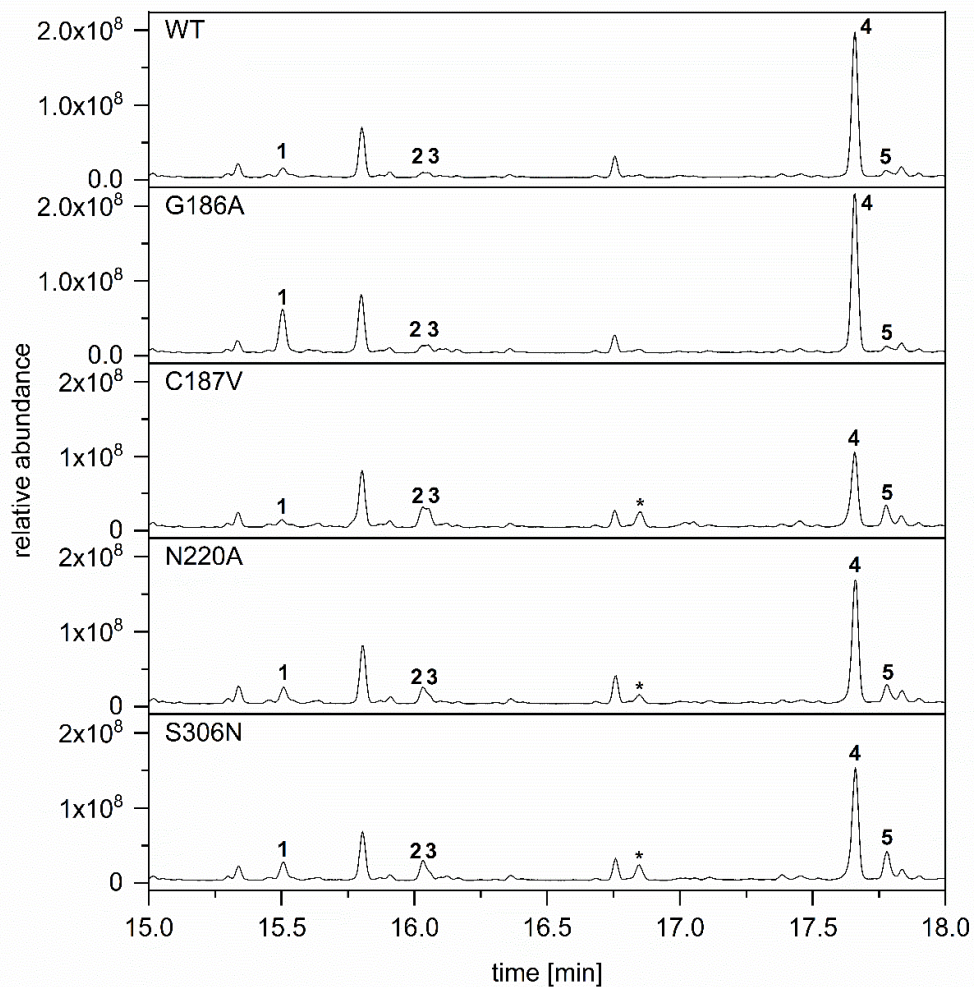


Fig. S 16: Comparison of the experimentally determined closed conformation of Copu9 with the modelled structures of Copu5 and Copu9. For clarity only one monomer is shown in ribbon representation. (a) Superposition **Copu9•Mg<sup>2+</sup><sub>3</sub>•AHD** with  $\alpha$ -helices are depicted in light blue and  $\beta$ -sheets in green. The Asp-rich motif is coloured in salmon and the NSE motif in yellow. The 3 Mg<sup>2+</sup> ions are indicated as green spheres and alendronate in stick representation. The modelled structure of Copu9 is shown in brown, representing the open, inactive conformation. (b) Superposition of **Copu9•Mg<sup>2+</sup><sub>3</sub>•AHD** and the modelled Cop5 shown in violet. (c) Superposition of the models of Copu5 (violet) and Copu9 (brown).

Table S4. Summary of active side residues in Copu3, Copu5 and Copu9. The listed residues are identical in Copu5 and Copu9. Only four of these residues are also identical in Copu3 (shown in grey). In the site directed mutagenesis study described the non-identical residues in Copu5 and Copu9 were replaced with the corresponding amino acids present in Copu3.

Copu3	Copu5	Copu9
L63	L65	L72
C64	T66C	T73C
V67	C69V	C76V
L86	L88	L95
F87	F89	F96
N182	T184N	T192N
A184	G186A	G194A
V185	C187V	C195V
C187	P189C	P197C
A218	N220A	N228A
W301	W302	W310
N305	S306N	S314N



*Fig. S17. GC-Chromatograms of Copu5 WT and variants G186A, C187V, N220A and S306N; Detailed GC-MS analysis based on NIST database comparison allows for compound assignment of (1) tau-murolene (RT: 15.51 min), (2) delta-cadinene (RT: 16.03 min), (3) cubebol (RT: 16.05 min), (4) (+)-δ-cadinol (RT: 17.66 min) and (5) α-cadinol (RT: 17.78 min); (\*) represents the formation of a new side product germacrene-D-4-ol (RT: 16.85 min); G186A shows enhanced production of tau-murolene; C187V shows enhanced production of δ-cadinene, cubebol and alpha-cadinol and formation of germacrene-D-4-ol; N220A shows a decrease in the formation of the main product, an increase in δ-cadinene and α-cadinol and the formation of germacrene-D-4-ol; S306N shows a decrease in the formation of the main product, an increase in δ-cadinene and α-cadinol and the formation of germacrene-D-4-ol.*



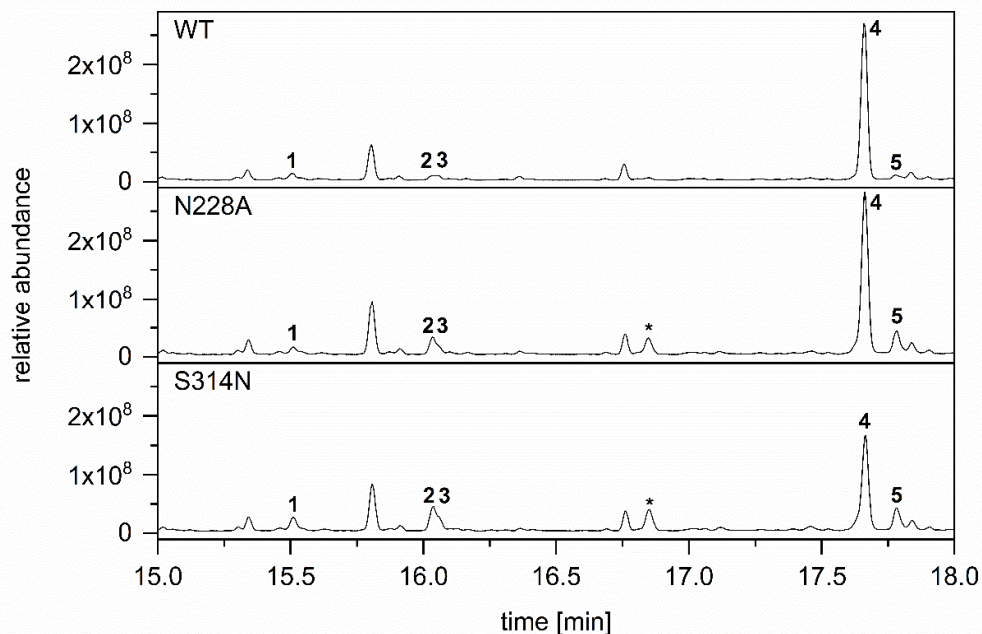


Fig. S18. GC-Chromatograms of Copu9 WT and variants N228A and S314N; Detailed GC-MS analysis based on NIST database comparison allows for compound assignment of (1) tau-murolene (RT: 15.51 min), (2)  $\delta$ -cadinene (RT: 16.03 min), (3) cubebol (RT: 16.05 min), (4) (+)- $\delta$ -cadinol (RT: 17.66 min) and (5)  $\alpha$ -cadinol (RT: 17.78 min); (\*) represents the formation of a new side product germacrene-D-4-ol (RT: 16.85 min); N228A shows an increase in  $\delta$ -cadinene and  $\alpha$ -cadinol and the formation of germacrene-D-4-ol; S314N shows a decrease in the formation of the main product, an increase in  $\delta$ -cadinene and  $\alpha$ -cadinol and the formation of germancrene-D-4-ol.

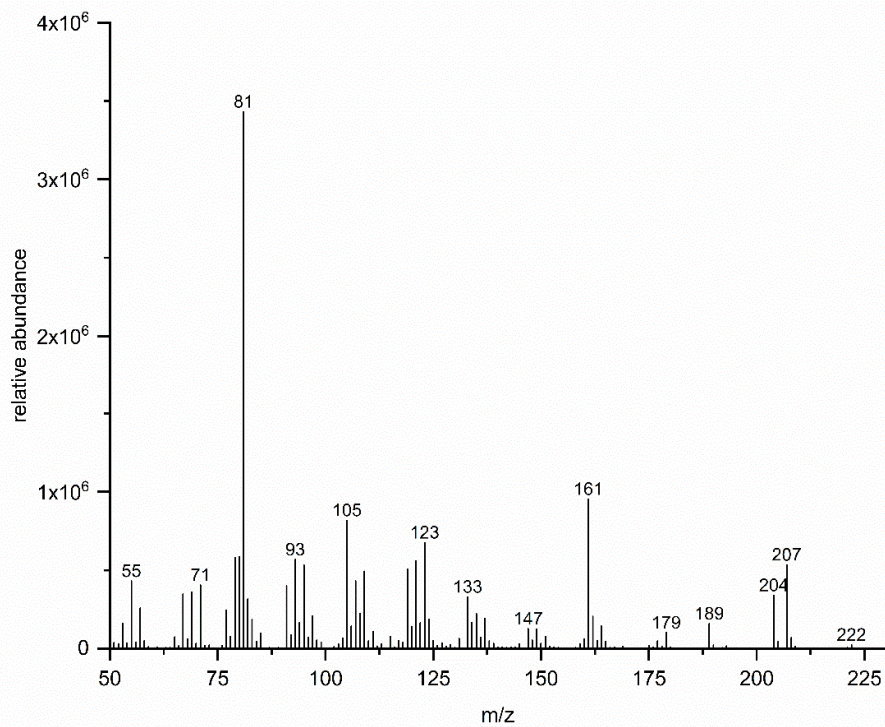


Fig. S19. GC-MS spectrum of germacrene-D-4-ol (RT: 16.85 min) as identified by NIST database[15]; assigned with (\*) in Fig. S 10 and Fig. S 11.



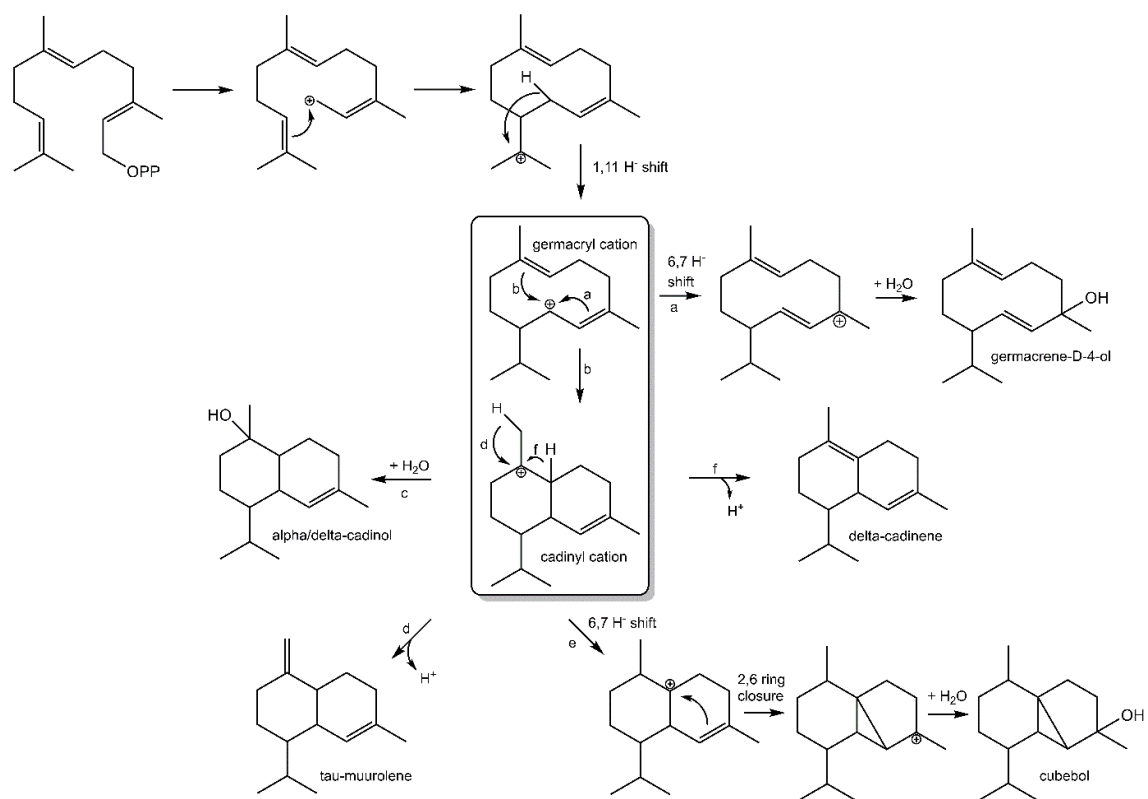


Fig. S20. Proposed cyclization pathway of all products of Copu5 and Copu9 leading the carbocation reaction trajectory via the key carbocations germacrlyl and cadinyl[1, 16, 17].

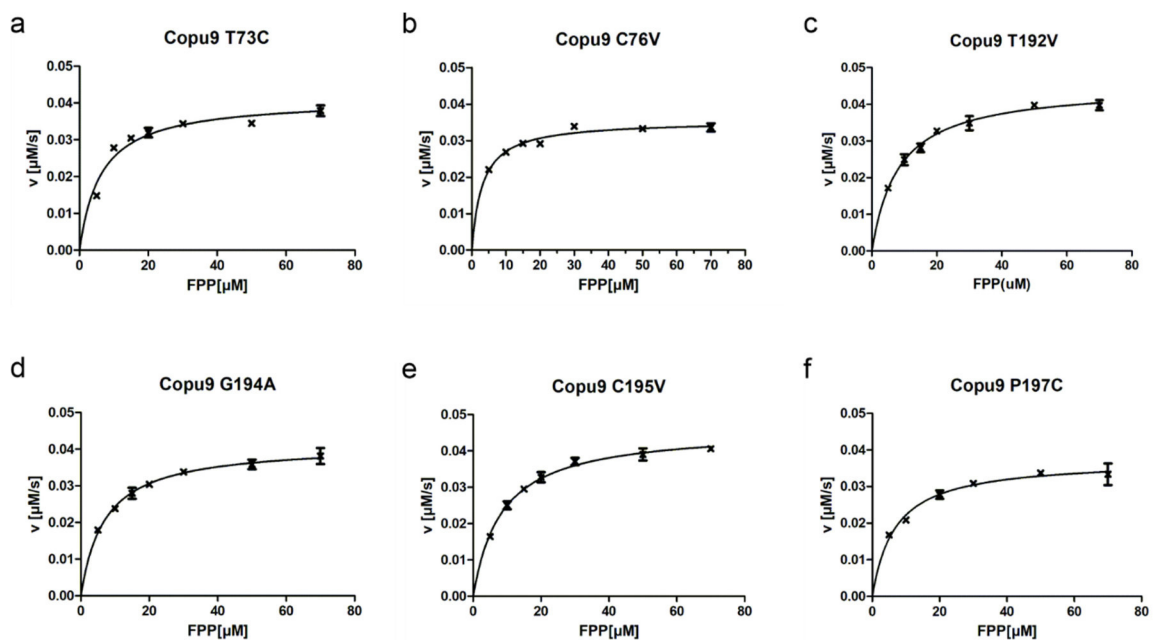


Fig S21: *In vitro* Michaelis-Menten kinetics of Copu9 variants T73C(a), C76V(b), T192N(c), G194A (d), C195V (e) and P197C (f) using the EnzChek™ Pyrophosphate Assay. A non-linear regression analysis was performed on the data collected from the time-resolved steady-state kinetic assay.

## References

1. Mischko W, Hirte M, Fuchs M, Mehlmer N, Brück TB. Identification of sesquiterpene synthases from the Basidiomycota *Coniophora puteana* for the efficient and highly selective  $\beta$ -copaene and cubebol production in *E. coli*. *Microb Cell Fact*. 2018;17:164. doi:10.1186/s12934-018-1010-z.
2. Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, Adams PD. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol*. 2017;73:148–57. doi:10.1107/s2059798316018210.
3. Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol*. 1997;4:269–75. doi:10.1038/nsb0497-269.
4. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science*. 2012;336:1030–3. doi:10.1126/science.1218231.
5. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66:213–21. doi:10.1107/S0907444909052925.
6. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*. 2018;27:293–315. doi:10.1002/pro.3330.
7. Holm L. DALI and the persistence of protein shape. *Protein Science*. 2020;29:128–40. doi:10.1002/pro.3749.
8. Baer P, Rabe P, Fischer K, Citron CA, Klapschinski TA, Groll M, Dickschat JS. Induced-fit mechanism in class I terpene cyclases. *Angew Chem Int Ed Engl*. 2014;53:7652–6. doi:10.1002/anie.201403648.
9. Li R, Chou WKW, Himmelberger JA, Litwin KM, Harris GG, Cane DE, Christianson DW. Reprogramming the chemodiversity of terpenoid cyclization by remolding the active site contour of epi-isozizaene synthase. *Biochemistry*. 2014;53:1155–68. doi:10.1021/bi401643u.
10. Harris GG, Lombardi PM, Pemberton TA, Matsui T, Weiss TM, Cole KE, et al. Structural Studies of Geosmin Synthase, a Bifunctional Sesquiterpene Synthase with  $\alpha\alpha$  Domain Architecture That Catalyzes a Unique Cyclization-Fragmentation Reaction Sequence. *Biochemistry*. 2015;54:7142–55. doi:10.1021/acs.biochem.5b01143.
11. Karuppiyah V, Ranaghan KE, Leferink NGH, Johannissen LO, Shanmugam M, Ní Cheallaigh A, et al. Structural Basis of Catalysis in the Bacterial Monoterpene Synthases Linalool Synthase and 1,8-Cineole Synthase. *ACS Catal*. 2017;7:6268–82. doi:10.1021/acscatal.7b01924.
12. Schriever K, Saenz-Mendez P, Rudraraju RS, Hendrikse NM, Hudson EP, Biundo A, et al. Engineering of Ancestors as a Tool to Elucidate Structure, Mechanism, and Specificity of Extant Terpene Cyclase. *J Am Chem Soc*. 2021;143:3794–807. doi:10.1021/jacs.0c10214.
13. Chen M, Chou WKW, Al-Lami N, Faraldos JA, Allemann RK, Cane DE, Christianson DW. Probing the Role of Active Site Water in the Sesquiterpene Cyclization Reaction Catalyzed by Aristolochene Synthase. *Biochemistry*. 2016;55:2864–74. doi:10.1021/acs.biochem.6b00343.
14. Centeno-Leija S, Tapia-Cabrera S, Guzmán-Trampe S, Esquivel B, Esturau-Esofet N, Tierrafría VH, et al. The structure of (E)-biformene synthase provides insights into the biosynthesis of bacterial bicyclic labdane-related diterpenoids. *J Struct Biol*. 2019;207:29–39. doi:10.1016/j.jsb.2019.04.010.
15. Linstrom P. NIST Chemistry WebBook, NIST Standard Reference Database 69: National Institute of Standards and Technology; 1997.

16. Yoshikuni Y, Martin VJJ, Ferrin TE, Keasling JD. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chemistry & Biology*. 2006;13:91–8. doi:10.1016/j.chembiol.2005.10.016.
17. Quin MB, Flynn CM, Schmidt-Dannert C. Traversing the fungal terpenome. *Nat Prod Rep*. 2014;31:1449–73. doi:10.1039/c4np00075g.

