

Stefan Etzl, BSc

Redesign of the catalytic activity of a flavin-dependent oxidoreductase

MASTER'S THESIS

to achieve the university degree of

Master of Science

Master's degree programme: Biochemistry and Molecular Biomedical Sciences

submitted to

Graz University of Technology

Supervisor

Univ.-Doz. Mag. Dr.rer.nat. Karl Gruber

Institute of Molecular Biosciences


University of Graz

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis dissertation.

31.03.2015

Date

Stefan 

Signature

Acknowledgments

I would like to thank Karl Gruber for attracting my interest in the field of structural biology in the first place, for allowing me to participate in his group, and for the excellent supervision in the course of this work. Special thanks go to Georg Steinkellner for providing assistance in bioinformatical issues and Altijana Hromic for always being a great help in questions concerning laboratory tasks. Also, I would like to express my gratitude to all the other "Strubi" members for their assistance in every situation, not to mention the good times apart from work.

Furthermore, my thanks go to Alexandra Binter and Michael Hetmann for providing me with recombinant cells for the expression of enzymes and to Kathrin Heckenbichler for providing me with compounds for the soaking experiments.

Finally, I would especially like to thank my parents for always supporting me in every aspect and thus giving me the freedom and opportunity to get to where I am now.

Contents

Abbreviations	ii
List of Figures	iii
List of Tables	iv
1 Introduction	1
1.1 Aims of this project	1
1.2 The enzyme: 12-oxophytodienoate reductase 3 from <i>Solanum lycopersicum</i>	1
1.3 Reductive C-C coupling	4
1.4 Enzyme design	5
2 Experimental procedures	5
2.1 Design process	5
2.1.1 Calculation of transition state models	6
2.1.2 Preparation of Rosetta input files	9
2.1.3 Initial design trials using the Rosetta match application	11
2.1.4 Protein design simply using the Rosetta enzyme design application	17
2.2 Protein expression and purification	24
2.3 Crystallization and structure determination	25
2.3.1 Crystallization of OPR3_5ring and OPR3_6ring	25
2.3.2 Crystallization of OPR3_Y190F	25
2.3.3 Soaking of OPR3_Y190F crystals with 4-bromobut-2-enal	25
2.3.4 Data processing and refinement	26
3 Results and discussion	27
3.1 Design process	27
3.1.1 Transition state structures	27
3.1.2 Initial design trials using the Rosetta match application	31
3.1.3 Protein design simply using the Rosetta enzyme design application	32
3.1.4 Summary	39
3.2 Protein expression and tests for catalytic activity	42
3.3 Crystallization	43
3.3.1 Crystal structure of OPR3_5ring	43
3.3.2 Crystal structure of OPR3_6ring	46
3.3.3 Crystal structure of OPR3_Y190F	49
4 Conclusions	55
References	56

Abbreviations

FMN	flavin mononucleotide
hal1	7-bromohept-2-enal
hal2	6-bromohex-2-enal
hal3	5-bromopent-2-enal
hal4	4-bromobut-2-enal
IPTG	isopropyl- β -D-1-thiogalactopyranoside
JA	jasmonic acid
NADH	nicotine adenine dinucleotide, reduced
NADPH	nicotine adenine dinucleotide phosphate, reduced
OPDA	12-oxo-phytodienoic acid
OPR3	12-oxophytodienoate reductase 3
PMSF	phenylmethanesulfonylfluoride

List of Figures

1	Reaction scheme for Old Yellow Enzymes	1
2	Jasmonic acid	2
3	Natural reaction of OPR3	2
4	Overall structure of OPR3	3
5	Active site of native OPR3	3
6	Halogenated substrates for reductive C-C coupling	4
7	Input structure for the calculation of a transition state, hal1	7
8	Input structure for the calculation of a transition state, hal2	7
9	Input structure for the calculation of a transition state, hal3	8
10	Input structure for the calculation of a transition state, hal4	8
11	Position of the substrate in the active site of OPR3 (hal2)	12
12	Potential energy profile for the reaction coordinate (hal1)	27
13	Potential energy profile for the reaction coordinate (hal2)	28
14	Potential energy profile for the reaction coordinate (hal3)	29
15	Potential energy surface (hal4)	30
16	Transition state models	31
17	Active site of OPR3_6ring	33
18	Active site of OPR3_5ring	34
19	Active site of OPR3_4ring_var1	35
20	Active site of OPR3_4ring_var2	36
21	Active site of OPR3_3ring_var1	37
22	Active site of OPR3_3ring_var2	37
23	Active site of OPR3_3ring_var3	38
24	Active site of OPR3_3ring_var4	38
25	Comparison: Crystal structure of OPR3_5ring vs design	45
26	Comparison: Crystal structure of OPR3_6ring vs design	48
27	Comparison: Substrate binding histidine residues of OPR3_6ring vs design	49
28	Electron density of a putative substrate molecule in OPR3_Y190F	52
29	Substrate fitted in OPR3_Y190F, 'reverse'	53
30	Substrate fitted in OPR3_Y190F	54
31	Substrate fitted in OPR3_Y190F, lower occupancy and different binding modes	54

List of Tables

1	Overview of the differences in the designed OPR3 variants	40
2	Experimental details for the crystal structure of OPR3_5ring	44
3	Experimental details for the crystal structure of OPR3_6ring	47
4	Experimental details for the crystal structure of OPR3_Y190F	50
5	Comparison of cell constants for OPR3_Y190F space groups	51

1 Introduction

1.1 Aims of this project

The aim of this project is to redesign the active site of a flavin-dependent oxidoreductase in a way that instead of a reduction of the olefinic substrate an intramolecular carbon coupling reaction is performed. Halogenated α, β -unsaturated aldehydes are thereby converted to cyclic compounds.

The transition state structures for the desired reactions are calculated for four different substrates and at least one enzyme variant for each substrate is designed. The enzyme design process is performed using bioinformatics tools, mainly the Rosetta software suite. The enzymes are expressed and crystallized. If x-ray crystal structures can be obtained they are compared with the computationally designed structures. Furthermore, an additional enzyme variant is crystallized, soaked with one of the substrates and the crystal structure is determined to get some insight into how the substrate binds in the active site.

1.2 The enzyme: 12-oxophytodienoate reductase 3 from *Solanum lycopersicum*

12-oxophytodienoate reductase 3 (OPR3, EC: 1.3.1.42) from *Solanum lycopersicum* (Tomato) belongs to a protein family of flavin mononucleotide (FMN) containing oxidoreductases known as 'Old Yellow Enzymes'. These enzymes generally catalyze the reduction of activated C=C bonds. Substrates of various members of the Old Yellow Enzyme family include various α, β -unsaturated compounds. The catalytic mechanism of Old Yellow Enzymes comprises the transfer of a hydride via the reduced cofactor (FMNH₂) to the C _{β} position of the substrate and the addition of a proton to C _{α} (oxidative half reaction)^[1]. For a complete catalytic cycle the oxidized FMN cofactor is reduced by the use of NAD(P)H (Fig. 1).

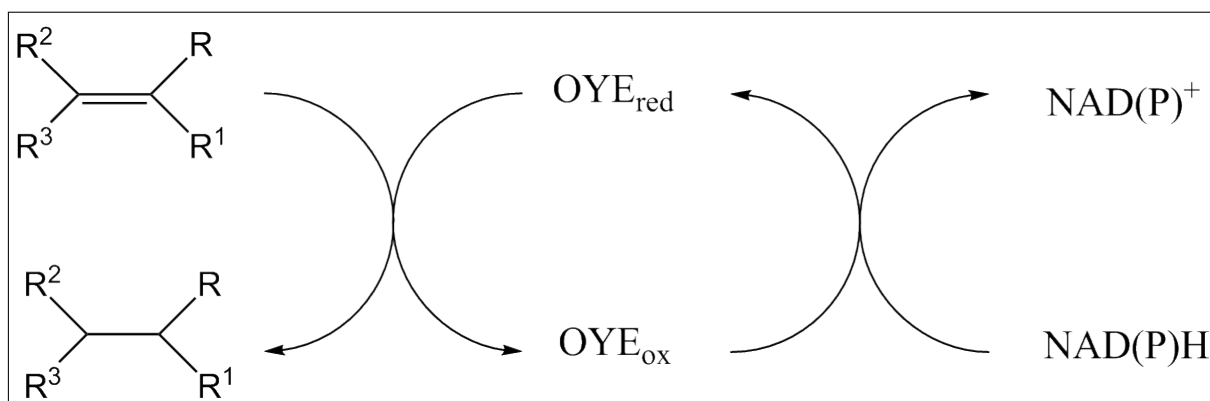


Figure 1 – General reaction scheme for Old Yellow Enzymes. A full catalytic cycle consists of two independent half reactions (ping pong mechanism). R = aldehyde, ketone, imide, nitro.

OPR3 consists of 396 amino acids on a single chain and contains one molecule of FMN per monomer as a cofactor. The enzyme is located in peroxisomes, and for the homolog from *Arabidopsis thaliana* it was shown that its natural role is in the biosynthesis pathway that leads to jasmonic acid (JA, Fig. 2)^[2;3]. JA is a signal molecule involved in stress response in plants^[4]. For the biosynthesis of JA a polyunsaturated fatty acid (linolenic acid) is oxygenated to yield a fatty acid hydroperoxide, which is converted to a cyclic compound that contains a cyclopentenone ring (12-oxo-phytodienoic acid, OPDA). OPR3 reduces the double bond in the ring, and in several cycles of β -oxidation the final product JA is formed^[5]. The preferred substrate of tomato OPR3 is the naturally occurring (9*S*,13*S*)-12-OPDA (Fig. 3), although it also accepts the other (9*R*,13*R*)-enantiomer^[3]. The fact that the enzyme also catalyzes the stereospecific reduction of various non-natural compounds like α , β -unsaturated aldehydes, ketones, nitroalkenes and N-substituted maleimides has shown a potential use of OPR3 in biocatalysis^[6].

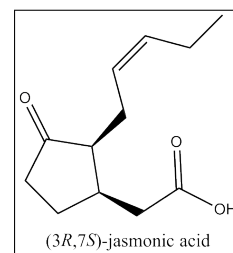


Figure 2 – Jasmonic acid.

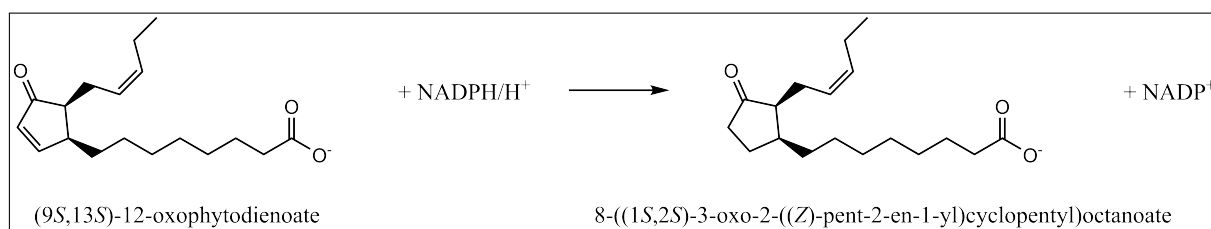


Figure 3 – A step in the biosynthetic pathway of jasmonic acid: OPR3 catalyzes the reduction of the 10,11-double bond in 12-oxophytodienoate.

The overall structure of OPR3 comprises an (α/β)₈-fold (TIM-barrel), which is typical for Old Yellow Enzymes (Fig. 4). Eight parallel beta strands form a barrel that is surrounded by alpha helices. The N-terminal side of the barrel is closed by a small beta hairpin. The FMN cofactor is located at the C-terminal side of the barrel and bound via hydrogen bonds to main chain and side chain atoms. The active site above the cofactor is mainly shaped by the loop regions between beta strands and alpha helices and is quite exposed to the solvent. Substrate binding involves hydrogen bonds to two histidine side chains (His-185 and His-188). A tyrosine residue (Tyr-190) acts as a proton donor for the substrate (Fig. 5).

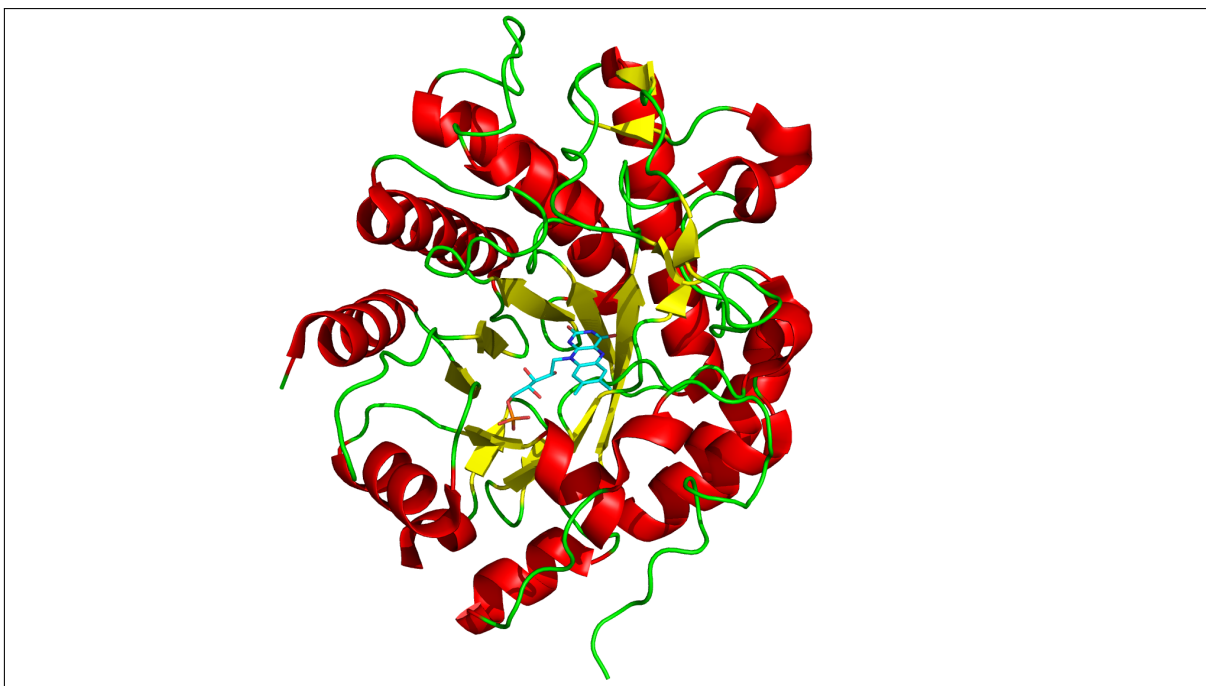


Figure 4 – Overall structure of OPR3. The central barrel is shaped by parallel beta sheets which are separated by the outer alpha helices. The FMN cofactor (cyan) sits at the C-terminal side of the barrel, the substrate binding pocket is made up by residues from the loop regions connecting alpha helices with beta strands (pdb: 3HGS).

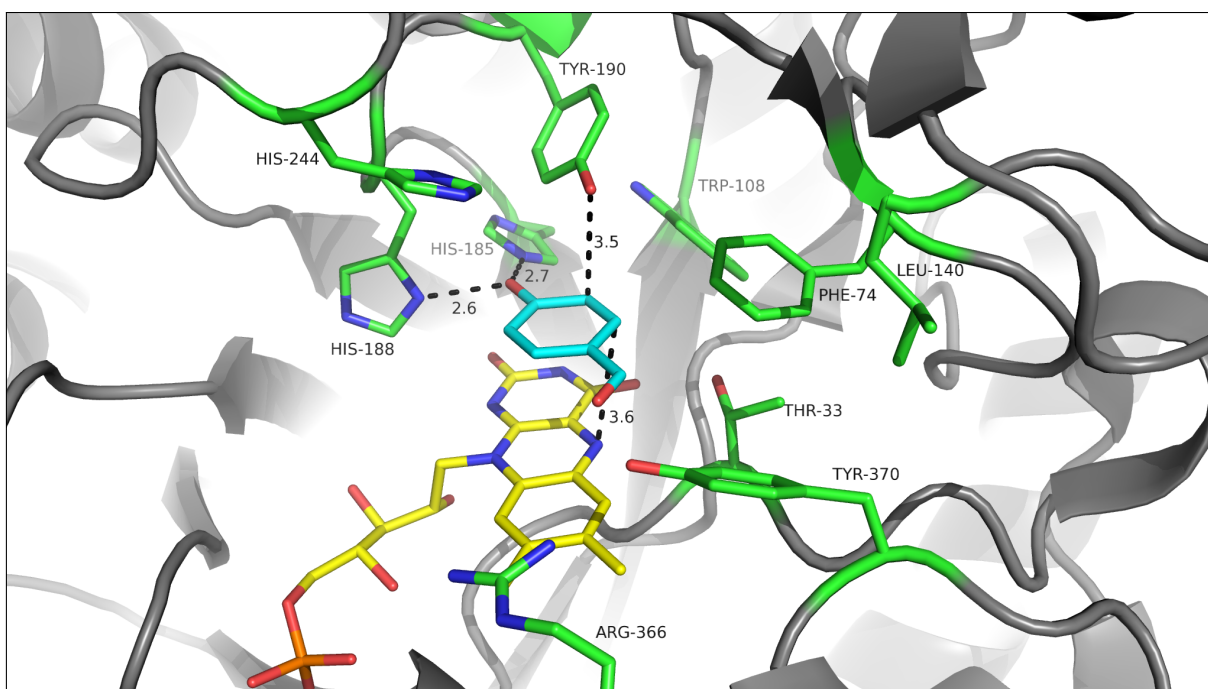


Figure 5 – Active site of OPR3 with *p*-hydroxybenzaldehyde (cyan) bound above the FMN cofactor (yellow). The side chains that shape the binding pocket are shown in green. The substrate is hydrogen bonded via two histidines, and a tyrosine residue acts as a proton donor in the oxidative half reaction (pdb: 3HGS).

1.3 Reductive C-C coupling

The reactions that are planned to be catalyzed by the designed enzymes are shown in Fig. 6. Transfer of a hydride from the reduced flavin to the β -carbon of the substrate can facilitate a nucleophilic attack of the α -carbon on the terminal carbon, with the bromine acting as a leaving group. For convenience the substrates were referred to as hal1 (7-bromohept-2-enal), hal2 (6-bromohex-2-enal), hal3 (5-bromopent-2-enal) and hal4 (4-bromobut-2-enal).

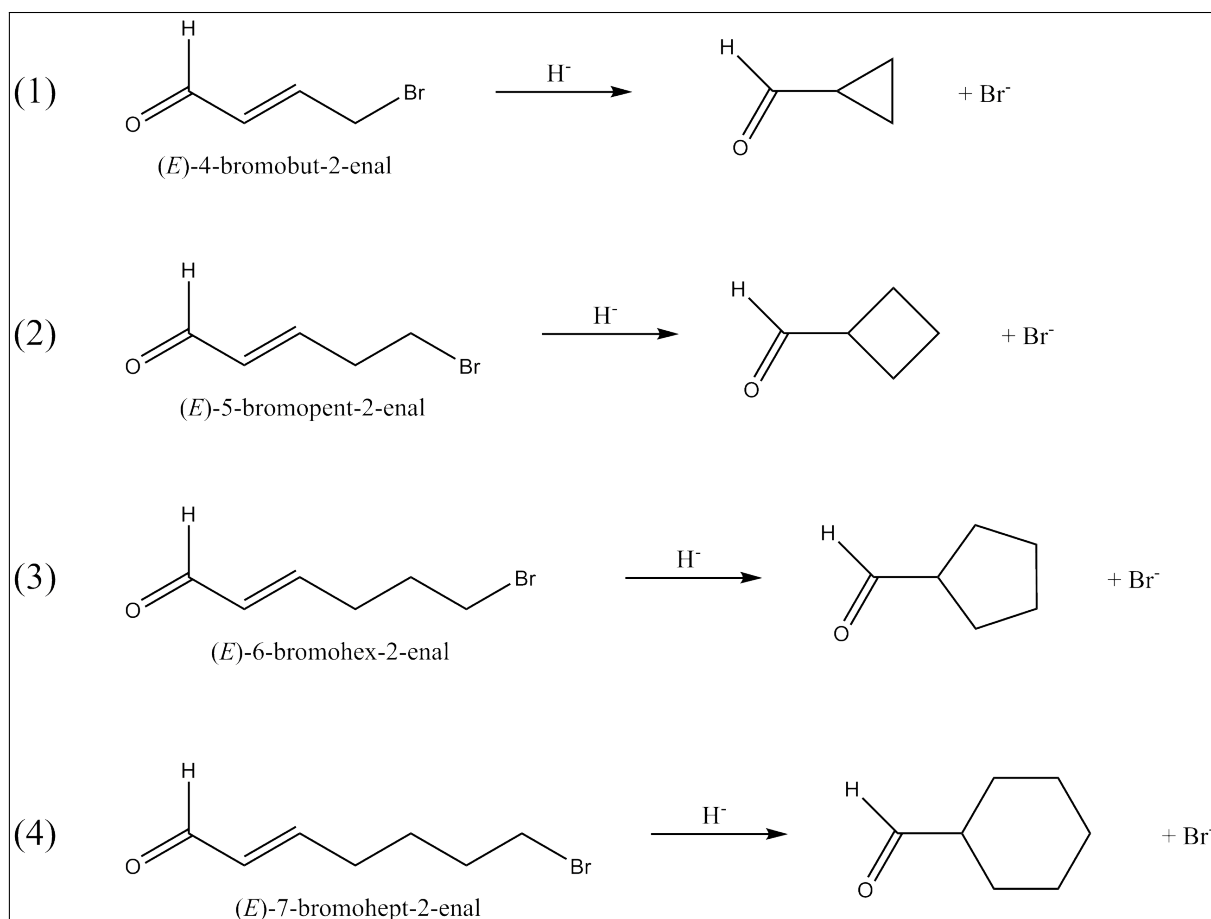


Figure 6 – The designed enzymes should facilitate the reaction of brominated α,β -unsaturated aldehydes to the respective cyclic carbonyl compounds. In the following part of this work the substrates will be referred to as hal4 (1), hal3 (2), hal2 (3), and hal1(4).

In mechanistic studies of Old Yellow Enzymes it was shown that the protonation of the substrate is significantly decreased when the main proton donor is eliminated^[1]. Since the C-C coupling reaction is in competition with a simple reduction, the suppression of proton transfer to the substrate was a basic requirement for new enzymes. Proof-of-concept that reductive C-C coupling using a modified OPR3 variant is possible could be shown in the group of Prof. Rolf Breinbauer at the Graz University of Technology. An enzyme, in which the main proton donor tyrosine (Tyr-190) was mutated to phenylalanine (OPR3_Y190F) converted 4-bromobut-2-enal to some extent to the desired cyclopropanecarbaldehyde (data not published).

1.4 Enzyme design

One of the ultimate goals in biocatalysis might be the ability to create any desired catalytic activity by altering existing enzymes or by de-novo-design of naturally non-existent enzymes. The function of a protein in general is based on its structure, which is dependent on the amino acid sequence. Due to the complexity of protein structures, it is still difficult to computationally model structures that are derived from artificial sequences. The more amino acids are altered, the more difficult is the prediction of how a designed protein structure might look like in reality. If the protein is an enzyme, its ability to catalyze a certain reaction is not only dependent on the overall structure, that provides a site where the reaction can occur, but also on the local geometry in the active site. In enzyme design, one has to find enzyme-substrate interactions that are crucial for catalysis, model an artificial active site that facilitates these interactions, and eventually provide a structure that stabilizes the active site and enables the binding of the substrate and the release of the product.

The practical approach for optimizing a catalytic activity for a certain reaction is, following the basic principle of catalysis, by determining the corresponding transition state models that are then stabilized via favorable interactions between substrate and protein. Artificial protein structures that provide these interactions are then calculated with computational tools.

2 Experimental procedures

2.1 Design process

The geometries of transition state structures were calculated using the ab-initio software Gaussian09^[7]. Together with a crystal structure of native OPR3 these structures were the input for the Rosetta design application. For the redesign of the catalytic activity the following requirements were taken into account:

- The need for a favorable binding mode of the substrate in order to allow transfer of a hydride from N5 of the flavin to the C_β position
- Stabilization of the negatively charged Br^- leaving group via hydrogen bonds and/or positively charged side chains in its vicinity
- Building an apolar and tightly packed environment around the C_α position in order to prevent protonation and therefore a simple reduction of the substrate.

To meet these requirements, certain amino acid residues were chosen to be exchanged for other residues that can provide the desired interactions. Given this input, the Rosetta design application can calculate a certain number of potential structures with favorable

mutations (no sterical clashes, tight packing, a maximum amount of stabilizing interactions, etc.). For technical reasons two different approaches were followed:

- Enzyme design using the Rosetta match application
- Enzyme design without using the Rosetta match application

The calculated design structures were visually investigated and ranked by chemical intuition. If not satisfactory, the designed structures were discarded and the gained experience was included in the setup for further design attempts until promising results could be obtained.

2.1.1 Calculation of transition state models

The geometries of transition state models were calculated using the software Gaussian09 via the graphical user interface GaussView^[8]. It was assumed that the crucial step in the reaction is rather the formation of a C-C bond and the leaving of the bromide than the transfer of a hydride. The initial model for the calculation was therefore a structure of the substrate with an additional hydrogen being bound to C_β (the hydride) and with the bromine still bound to the terminal carbon atom. The model was attributed an overall charge of -1. Using the 'scan' option in Gaussian09, the bond length between the two carbon atoms forming the new bond was increased in several increments of 0.15 - 0.20 angstroms in case of the hal1 substrate (Fig. 7), or decreased for hal2 and hal3 (Fig. 8 and Fig. 9). At each step the bond length was fixed and a geometry optimization was carried out. In case of the hal4 substrate a 2D scan, scanning both the newly formed C-C bond and the C-Br bond in increments, was set up (Fig. 10). The method used was B3LYP with the 6-311+G(d,p) basis set. A local maximum in the obtained energy profiles indicated the existence of a transition state structure. A structure very close to the local energy maximum was then used as input for an energy optimization towards a transition state (Gaussian09 job: opt=calcall,ts). The presence of a transition state was confirmed by the occurrence of a single negative (i.e. imaginary) vibrational frequency.

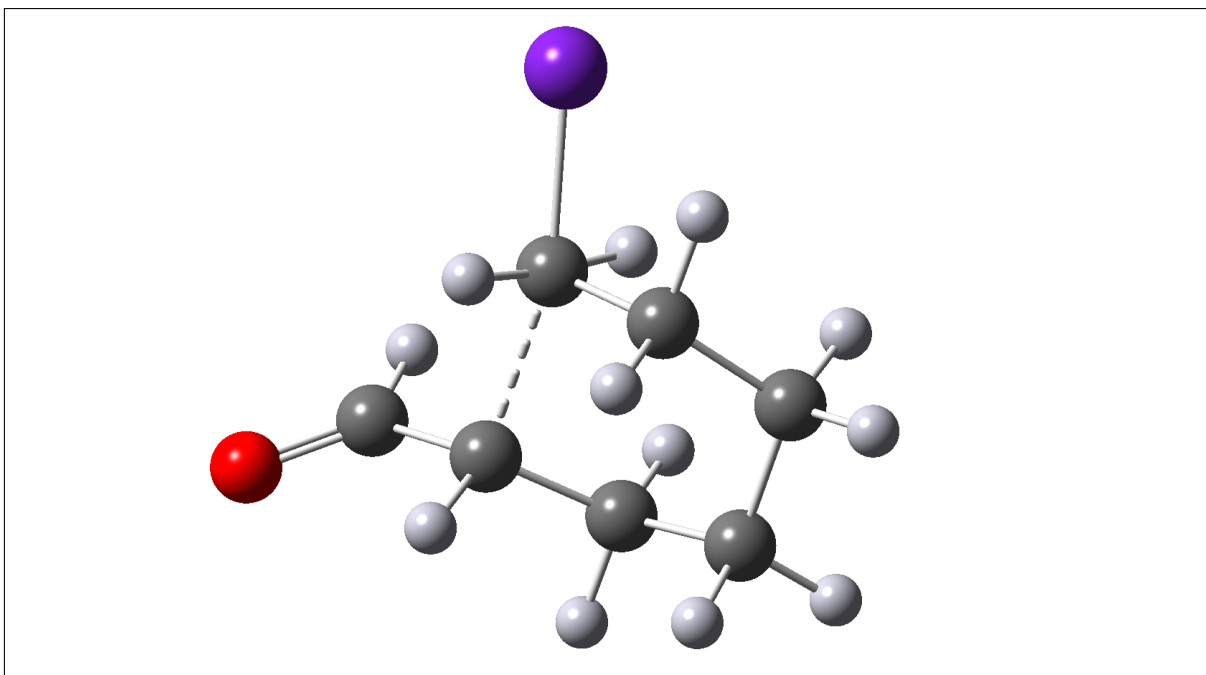


Figure 7 – The input structure for the calculation of a transition state between 7-bromohept-2-enal (hal1) and cyclohexanecarbaldehyde. The model is the substrate with an additional hydrogen atom bound to C3. The system was given an overall charge of -1. The C2-C7 bond length (dashed) was increased in seven steps of 0.15 angstroms and a geometry optimization was done after each step. (Red=oxygen, purple=bromine).

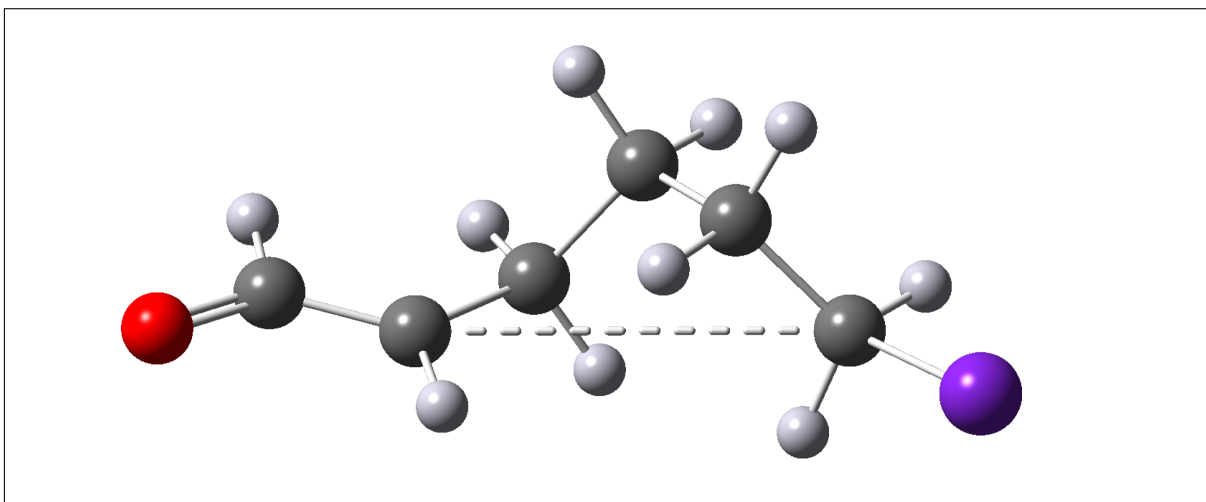


Figure 8 – The input structure for the calculation of a transition state between 6-bromohex-2-enal (hal2) and cyclopentanecarbaldehyde. The molecule was given an overall charge of -1. The C2-C6 bond length (dashed) was decreased in twelve steps of 0.2 angstroms. (Red=oxygen, purple=bromine).

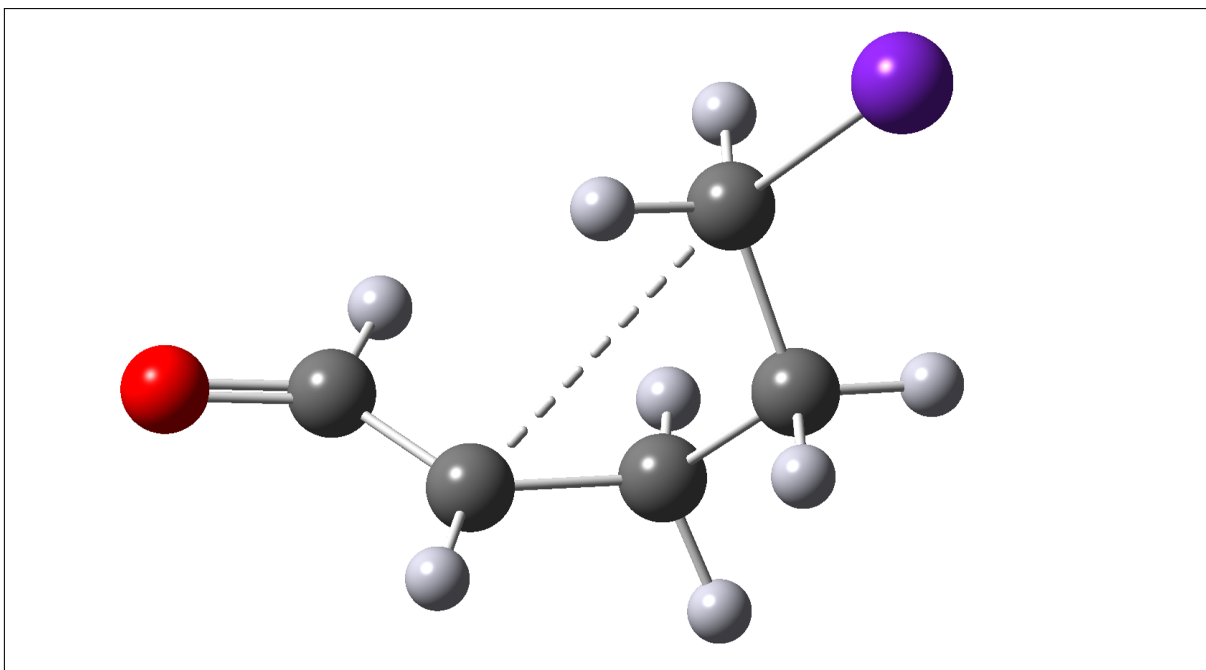


Figure 9 – The input structure for the calculation of a transition state for the reaction of 5-bromopent-2-enal (hal2) to cyclobutanecarbaldehyde. The molecule was given an overall charge of -1. The C2-C6 bond length (dashed) was decreased in twelve steps of 0.2 angstroms. (Red=oxygen, purple=bromine).

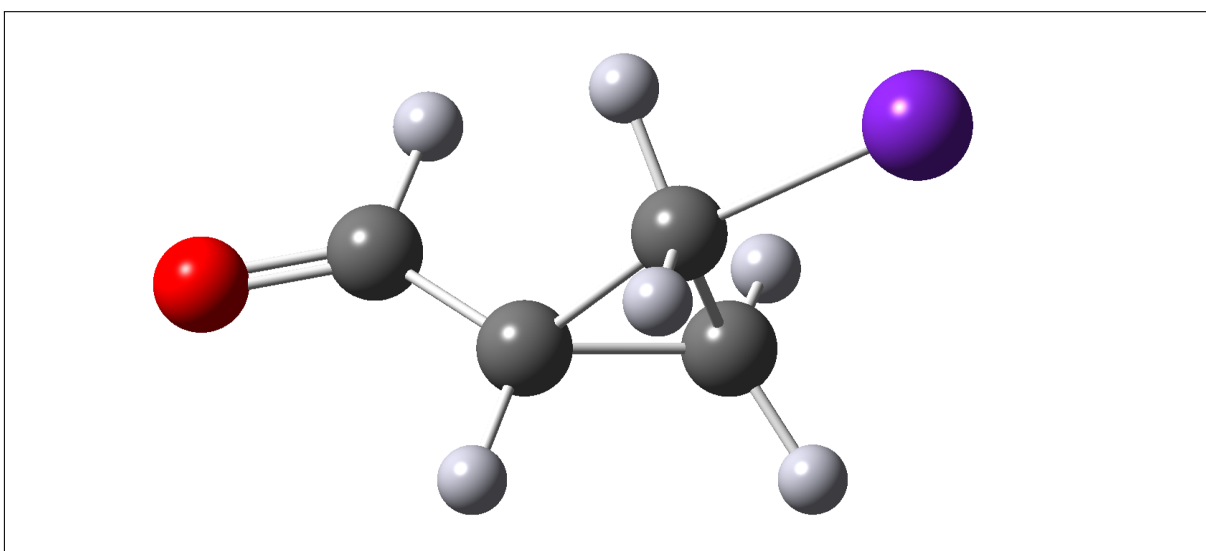


Figure 10 – The input structure for the calculation of a transition state between 4-bromobut-2-enal (hal4) to cyclopropanecarbaldehyde. For this compound the bond lengths between C2-C4 and C4-Br were increased in seven and four steps of 0.2 angstroms, respectively. The molecule was given an overall charge of -1. (Red=oxygen, purple=bromine).

2.1.2 Preparation of Rosetta input files

The starting model was a pdb file of a crystal structure of native OPR3 containing one chain with 376 residues (Asn-10 - Leu-385) and one molecule of FMN. In order to provide a proper input for the Rosetta programs the residue numbers were altered to make the first residue in the file (Asn-10) residue number 1 on chain 'A'. The FMN cofactor was transferred to a separate chain 'B'. Modifications were performed in PyMOL^[9]. In the pdb text file special amino acid three-letter-codes (e.g. HIP, HIE, CYD) were replaced with the standard three-letter-code (HIS, CYS) as Rosetta cannot handle this additional information.

In order to recognize non-protein compounds like FMN and the transition state molecules, Rosetta requires information on these compounds provided in parameter files (.params). To set up the file 'FMN_opr3.params' the FMN molecule was extracted from the initial OPR3 crystal structure pdb and saved in mol2 format. For this purpose the Schrödinger Maestro program was used^[10]. The mol2 file was then converted to a Rosetta params file using the script molfile_to_params.py provided in the Rosetta suite:

```
~/rosetta/rosetta_source/src/python/apps/public/molfile_to_params.py FMN.mol2 -name FMN --keep-names
```

The input file was 'FMN.mol2', the options used were:

```
-name FMN      name ligand 'FMN'  
--keep-names  important to keep atom names in the params file consistent with the  
              names in the scaffold pdb file.
```

The output written was a pdb file 'FMN.pdb' and the params file 'FMN.params'. The file 'FMN.params' was renamed to 'FMN_opr3.params'.

The same procedure was performed with the mol2 files of the transition state molecules. To assure that the script molfile_to_params.py was running properly, some minor changes had to be made in the Gaussian09 output mol2 files of the transition state molecules. The example below shows the modified mol2 file of the substrate hal2. The files for the other substrates were treated in analogy to this.

```
line |  
  1  | # hal2_scan1_9tsopt  
  ... |  
  8  | @<TRIPOS>MOLECULE  
  9  | Molecule Name  
 10  | 18 17  
 11  | SMALL
```

```

12 NO_CHARGES
13
14
15 @<TRIPOS>ATOM
16 1 C1      -1.2314  2.1033  0.1807  C
17 2 C2      -0.0065  1.3866 -0.4205  C
18 3 C3      -2.4454 -0.0870  0.4664  C
19 4 C4      -2.5281  1.3155 -0.0571  C
20 5 H5      -1.2953  3.1211 -0.2342  H
21 6 H6      -1.0742  2.2180  1.2617  H
22 7 H7      -0.1106  1.3195 -1.5087  H
23 8 H8       0.8765  2.0011 -0.2265  H
24 9 H9      -2.3910 -0.2304  1.5447  H
25 10 H10    -3.3619  1.8872  0.3886  H
26 11 H11    -2.7286  1.2823 -1.1376  H
27 12 C12    -2.7436 -1.2071 -0.3193  C
28 13 H13    -2.8936 -0.9587 -1.4044  H
29 14 O14    -2.8289 -2.4101  0.0200  O
30 15 H15     0.1561 -0.1221  1.2136  H
31 16 H16    -0.0416 -0.8655 -0.4267  H
32 17 C17     0.2211  0.0070  0.1443  C
33 18 Br18    2.5114 -0.3280  0.0115  Br
34 @<TRIPOS>BOND
35 1 1 2 1
36 2 1 4 1
37 3 1 5 1
38 4 1 6 1
39 5 2 7 1
40 6 2 8 1
41 7 2 17 1
42 8 3 4 1
43 9 3 9 1
44 10 3 12 ar
45 11 4 10 1
46 12 4 11 1
47 13 12 13 1
48 14 12 14 2
49 15 15 17 1
50 16 16 17 1
51 17 17 18 nc

```

Changes were made in line 10, where formerly '18 16' was substituted by '18 17'. 18 is the number of atoms in the file, 16 or 17 the number of bonds. The number of bonds had to be raised by 1, as the C-Br bond had not been present in the Gaussian mol2 output file. The additional bond was added in line 51: '17 17 18 nc' - bond number 17 between atom 17 (C17) and atom 18 (Br18) was assigned as 'nc' (non-covalent). Furthermore, in line 44 the original '10 3 12 Ar' was edited to '10 3 12 ar' as the molfile_to_params.py script only recognizes the lower case 'ar' (for aromatic bond order). All these changes did not affect atom coordinates and were therefore acceptable.

2.1.3 Initial design trials using the Rosetta match application

The first attempts to design a new active site were started using the Rosetta match application^[11;12]. The match application tries to find positions for catalytic side chains on a given protein backbone. The user has to specify the transition state molecule and the geometry of the interacting side chains with respect to that transition state (the theozyme). The theozyme geometry is described in constraint files (cstfiles), which are set up by the user. The match application then tries to find positions on a protein backbone (scaffold) that are suitable to introduce the desired theozyme geometry with respect to the given constraints. The output (called a 'match') is a pdb file of the scaffold containing the transition state molecule and the catalytic side chains. If the matcher cannot find a suitable position, there is no output. The match application cannot handle interactions between the transition state molecule and a ligand (or any other non-protein compound, like FMN in this case) that is already present in the scaffold pdb. This means that no direct interactions between the transition state molecule and FMN could be defined. To circumvent this problem, the interactions between the transition state molecule and the two histidine residues that are already present in native OPR3 (His-176 and His-179) were very tightly specified by setting the number of sampling steps to zero (last column in the CONSTRAINT section of the cstfile). This forces the match application to place the transition state molecule in a well defined orientation with respect to the two histidines as well as in the desired position relative to the isoalloxazine moiety of FMN.

The following part of this section shows an example of how the design process was performed in the case of the hal2-substrate. Analogous procedures were carried out for the substrates hal3 and hal4.

In order to set up a cstfile, the transition state molecule was manually positioned in the active site of the scaffold pdb (opr3_s.pdb) in the same position as a known inhibitor in a crystal structure (Fig. 11). In that way the first two blocks for the cstfile, specifying the interactions between the transition state molecule and His-176 and His-179, respectively, could be set up. For this purpose an in-house python script that automatically writes Rosetta constraint file blocks via manual selection of atoms in a PyMOL plugin was used. Also the block defining the interactions between FMN and the transition state molecule

was set up using this script. The other block, specifying an interaction between the bromine atom and an undefined hydrogen bond donor was set up manually. As the match application cannot handle non-protein compounds, the block that specifies the interactions between the transition state molecule and FMN was commented out during the matching process. For the later use of the design application this section was again uncommented.

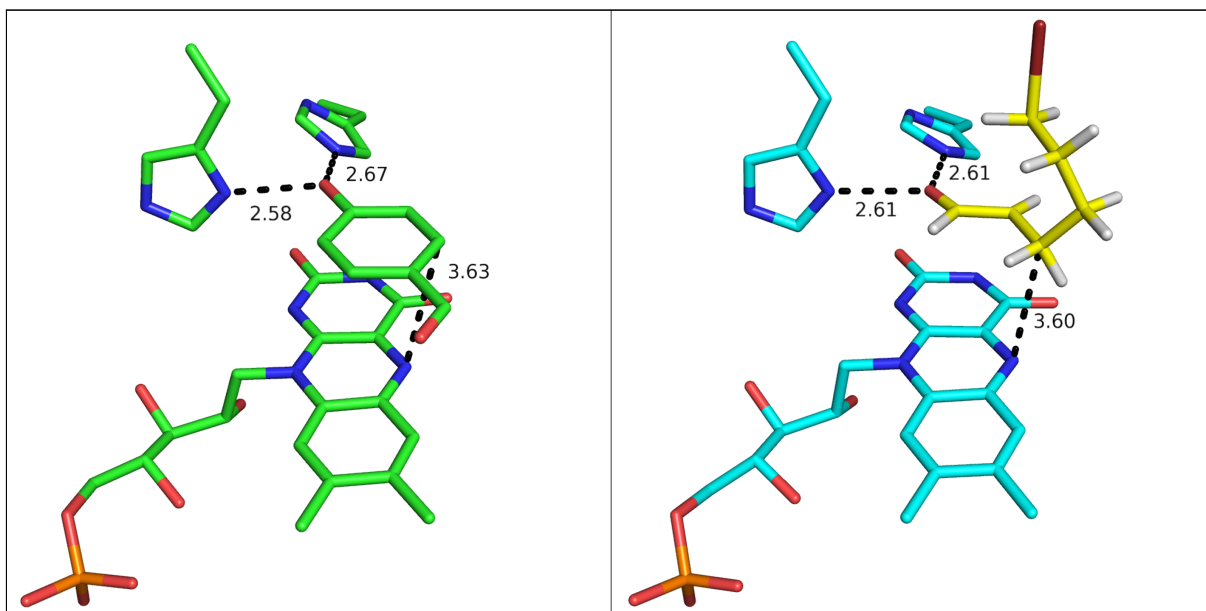


Figure 11 – Left: p-hydroxybenzaldehyde bound in the active site of native OPR3 (pdb code: 3HGS), right: manually positioned transition state model for hal2 in the scaffold pdb. The relevant parameters for the positioning of the transition state model in the active site were the distance to the two histidines that coordinate the carbonyl group, and the distance between the substrate's C_{β} and N5 of FMN.

Below the cstfile (opr3_hal2.cst) is shown:

```
#block 1 interaction H179 and TST
CST::BEGIN
    TEMPLATE::    ATOM_MAP: 1  atom_name:  014 C12 C3
    TEMPLATE::    ATOM_MAP: 1  residue3:  TST

    NATIVE
    TEMPLATE::    ATOM_MAP: 2  atom_type:  Ntrp
    TEMPLATE::    ATOM_MAP: 2  residue1:  H

    CONSTRAINT::  distanceAB:      2.61  0.2      100.   0.  0
    CONSTRAINT::   angle_A:        124.5  10.0     80.0  360.  0
    CONSTRAINT::   angle_B:        119.2  10.0     80.0  360.  0
    CONSTRAINT::   torsion_A:      179.5  10.0     80.0  360.  0
    CONSTRAINT::   torsion_AB:     92.5   10.0     0.0   360.  0
```

```
CONSTRAINT::  torsion_B:      -178.1  10.0          0.0  360.  0
CST::END
```

```
#block 2 interaction H176 and TST
```

```
CST::BEGIN
```

```
TEMPLATE::  ATOM_MAP: 1  atom_name:  014 C12 C3
```

```
TEMPLATE::  ATOM_MAP: 1  residue3:  TST
```

```
NATIVE
```

```
TEMPLATE::  ATOM_MAP: 2  atom_type:  Ntrp
```

```
TEMPLATE::  ATOM_MAP: 2  residue1:  H
```

```
CONSTRAINT::  distanceAB:      2.61  0.2          100.   0.  0
```

```
CONSTRAINT::   angle_A:      138.8  10.0          80.0  360.  0
```

```
CONSTRAINT::   angle_B:      107.6  10.0          80.0  360.  0
```

```
CONSTRAINT::  torsion_A:       6.4  10.0          80.0  360.  0
```

```
CONSTRAINT::  torsion_AB:      92.5  10.0           0.0  360.  0
```

```
CONSTRAINT::  torsion_B:     -178.1  10.0           0.0  360.  0
```

```
CST::END
```

```
#block 3 additional interaction with Br
```

```
CST::BEGIN
```

```
TEMPLATE::  ATOM_MAP: 1  atom_name:  Br18 C17 C2
```

```
TEMPLATE::  ATOM_MAP: 1  residue3:  TST
```

```
TEMPLATE::  ATOM_MAP: 2  atom_type:  Hpol
```

```
TEMPLATE::  ATOM_MAP: 2  residue1:  STKRHNQYW
```

```
CONSTRAINT::  distanceAB:      2.7  0.2          100.   0.  5
```

```
CONSTRAINT::   angle_A:      180.0  10.0          80.0  360.  5
```

```
CONSTRAINT::   angle_B:      180.0  10.0          80.0  360.  5
```

```
CONSTRAINT::  torsion_A:       0.0  180.0          80.0   10.  1
```

```
CONSTRAINT::  torsion_AB:      0.0  180.0          80.0   10.  1
```

```
CONSTRAINT::  torsion_B:     180.0  3.0           80.0  180.  3
```

```
CST::END
```

```
#block 4 interaction FMN and TST
```

```
#CST::BEGIN
```

```
#  TEMPLATE::  ATOM_MAP: 1  atom_name:  014 C12 C3
```

```
#  TEMPLATE::  ATOM_MAP: 1  residue3:  TST
```

```

#
# NATIVE
# TEMPLATE:: ATOM_MAP: 2 atom_name: N5 C4A C4
# TEMPLATE:: ATOM_MAP: 2 residue3: FMN
#
# CONSTRAINT:: distanceAB: 4.84 0.2 100. 0. 0
# CONSTRAINT:: angle_A: 50.3 10.0 80.0 360. 0
# CONSTRAINT:: angle_B: 43.1 10.0 80.0 360. 0
# CONSTRAINT:: torsion_A: 82.5 10.0 80.0 360. 0
# CONSTRAINT:: torsion_AB: 179.7 10.0 0.0 360. 0
# CONSTRAINT:: torsion_B: 98.0 10.0 0.0 360. 0
#CST::END

```

The constraints for the hydrogen bond donor (block 3) were derived from an idealized geometry between the donor and the bromine¹ atom. A distance of 2.7 ± 0.2 angstroms between H and Br was chosen, because a geometry analysis of the Cambridge Structural Database showed typical Br⁻ - donor distances in the range of 3.5 - 3.9 angstroms.

In order to define which sequence positions in the scaffold should be considered by the match application for introducing new residues, a posfile (opr3_hal2.pos) was set up:

```

N_CST 3
1: 179
2: 176
3: 64 99 101 104 140 181 235 236

```

The residue numbers for constraint block 3 were chosen by investigating the scaffold pdb and selecting favorable sequence positions that could function as a starting point for a newly introduced residue.

The match application was run using the following command line:

```

~/rosetta/rosetta_source/bin/match.linuxgccrelease -database ~/rosetta/r
osetta_database -extra_res_fa TST_hal2.params -extra_res_fa FMN_opr3.para
ms -match:lig_name TST -match:geometric_constraint_file opr3_hal2.cst -ma
tch:scaffold_active_site_residues_for_geomcsts opr3_hal2.pos -s opr3_s.pd
b -in:ignore_zero_occupancy false -ex1 -ex2 -ex1aro -ex2aro -use_input_sc

```

The run produced two match output files (UM_1_H179H176H104_opr3_s_opr3_hal2.1.pdb and UM_2_H179H176R104_opr3_s_opr3_hal2.1.pdb). These files were then passed on to

¹In design related contexts in this work, with 'bromine' the bromine/bromide moiety of the transition state model is meant

the Rosetta enzyme design application^[12]. It processes the output of the matcher in such a way that the introduced theozyme geometry is supported by the surrounding residues. The protein is redesigned (residues are mutated) and/or repacked in order to optimize the catalytic interactions defined in the cstfile and to minimize the structure. For this purpose the constraints block that specifies the interactions between FMN and the transition state molecule (block 4) was uncommented (other than the matcher the enzyme design application can deal with constraints between non-protein compounds). To make the FMN interaction recognizable for the enzyme design application the REMARKS section of the matching output pdb files had to be supplemented with an additional line:

```
REMARK 666 MATCH TEMPLATE X TST 0 MATCH MOTIV B FMN 1 4 1
```

A line like this is usually added automatically by the match application for every matched catalytic residue. Since the interaction between FMN and the transition state molecule was not part of the matching process this line had to be added manually. It tells the enzyme design application that in constraint block '4' in the cstfile, there is a specified interaction between catalytic residue 'FMN' on chain 'B' position '1' and the matched substrate 'TST' on chain 'X'.

Another required input and a very important aspect for controlling the behavior of the enzyme design application was the setup of a resfile. A resfile contains specific instructions on how to treat certain residues during the design process. Below the resfile 'opr3_hal2.res' is shown.

```
AUTO
USE_INPUT_SC

start

99  A  APOLAR

# FMN coordinating

24  A  NATRO
54  A  NATRO
55  A  NATRO
97  A  NATRO
228 A  NATRO
310 A  NATRO
312 A  NATRO
```

```
333 A NATRO
334 A NATRO
```

In this resfile residue number 99 was requested to be apolar to account for the need of a hydrophobic environment around the C α of the substrate. In order to avoid unnatural displacement of the cofactor, the residues that are directly involved in hydrogen bonding with FMN and some neighboring residues had to be prevented from being mutated or repacked by using the keyword 'NATRO'.

The enzyme design application was run with the following command line:

```
~/rosetta/rosetta_source/bin/enzdes.linuxgccrelease -database ~/rosetta/rosetta_database -extra_res_fa FMN_opr3.params -extra_res_fa TST_hal2.params -enzdes:cstfile opr3_hal2.cst -resfile opr3_hal2.res -in:file:list matchlist.txt -nstruct 25 -out:file:o scores.txt @design.flags
```

The file 'matchlist.txt' was a simple textfile containing the names of the two match output files in separate lines. By using this input flag, the enzyme design application calculates 25 structures for the first match output file (-nstruct 25), and then automatically moves on to the next file for the next 25 structures. The flag -out:file:o scores.txt induces the output of a text file containing scores for every catalytic constraint of each output design structure. The flag file 'design.flags' contained more general flags:

```
-linmem_ig 10
-use_input_sc
-ex1
-ex2
-enzdes::detect_design_interface
-enzdes::fix_catalytic_aa
-enzdes::cut1 6.0
-enzdes::cut2 8.0
-enzdes::cut3 10.0
-enzdes::cut4 12.0
-enzdes
  -cst_min
  -cst_opt
  -chi_min
  -cst_design
  -bb_min
  -design_min_cycles 3
```

```

-favor_native_res 0.8
-start_from_random_rb_conf
-enzdes::lig_packer_weight 1.5
-packing::soft_rep_design
-packing
  -ex1
  -ex2
  -ex1aro
  -ex2aro

```

From each of the two match output files 25 design structures were calculated, respectively. Based on the same principle, structures were also produced for the enzymes converting the substrates hal3 and hal4 (files not shown).

2.1.4 Protein design simply using the Rosetta enzyme design application

In this approach the matching process was skipped due to a lack of successful outputs (see 'Results and discussion' for details) and only the enzyme design application was used to produce designed structures. Instead of matching output files, pdb files of the native enzyme with manually positioned transition state molecules were used as input files. It was determined which sequence positions could be subjected to mutations and for which amino acids these positions might be exchanged. The necessary instructions were included in the resfiles. For each set of resfile instructions ten output structures were calculated. The designed structures were visually investigated (PyMOL) and promising modifications were kept, whereas for unsatisfactory mutations the resfile instructions regarding the sequence positions and types of possible amino acids were altered in subsequent design runs. Every run used the same starting pdb file. By a procedure like that, the design comprised a semi-manual, iterative process to find favorable changes that all together constitute a desired theozyme geometry.

The constraint files for this method only included interactions between His-176, His-179 and FMN with the transition state molecule, respectively.

Design of OPR3_6ring (hal1) The following part shows an example of how the design process was performed in the case of the hal1 substrate. The geometric constraint file (opr3_hal1.cst) used was:

```

#block 1 interaction H176
CST::BEGIN
  TEMPLATE::      ATOM_MAP: 1  atom_name:  018 C17 C5
  TEMPLATE::      ATOM_MAP: 1  residue3:  TST

```

```

NATIVE
TEMPLATE::  ATOM_MAP: 2  atom_type:  Ntrp ,
TEMPLATE::  ATOM_MAP: 2  residue1:  H

CONSTRAINT::  distanceAB:      2.71  0.2      100.    0  0
CONSTRAINT::  angle_A:        141.5  3.0      80.0  360.  0
CONSTRAINT::  angle_B:        107.1  3.0      80.0  360.  0
CONSTRAINT::  torsion_A:       -1.0  3.0      80.0  360.  0
CONSTRAINT::  torsion_AB:     -160.3  3.0      80.0  360.  0
CONSTRAINT::  torsion_B:     -171.9  3.0      80.0  360.  0
CST::END

#block 2 interaction H179
CST::BEGIN

TEMPLATE::  ATOM_MAP: 1  atom_name:  018 C17 C5
TEMPLATE::  ATOM_MAP: 1  residue3:  TST

NATIVE
TEMPLATE::  ATOM_MAP: 2  atom_type:  Ntrp ,
TEMPLATE::  ATOM_MAP: 2  residue1:  H

CONSTRAINT::  distanceAB:      2.61  0.2      100.    0  0
CONSTRAINT::  angle_A:        123.9  3.0      80.0  360.  0
CONSTRAINT::  angle_B:        117.7  3.0      80.0  360.  0
CONSTRAINT::  torsion_A:       171.3  3.0      80.0  360.  0
CONSTRAINT::  torsion_AB:       90.1  3.0      80.0  360.  0
CONSTRAINT::  torsion_B:     -175.8  3.0      80.0  360.  0
CST::END

#block 3 interaction FMN
CST::BEGIN

TEMPLATE::  ATOM_MAP: 1  atom_name:  C4 C3 C2
TEMPLATE::  ATOM_MAP: 1  residue3:  TST

NATIVE
TEMPLATE::  ATOM_MAP: 2  atom_name:  N5 C4A C4
TEMPLATE::  ATOM_MAP: 2  residue3:  FMN

CONSTRAINT::  distanceAB:      3.60  0.2      100.    0  0

```

```

CONSTRAINT::      angle_A:          125.4  3.0           80.0  360.  0
CONSTRAINT::      angle_B:           93.2  3.0           80.0  360.  0
CONSTRAINT::      torsion_A:         164.2  3.0           80.0  360.  0
CONSTRAINT::      torsion_AB:        -128.1  3.0           80.0  360.  0
CONSTRAINT::      torsion_B:          87.7  3.0           80.0  360.  0

```

CST:::END

In the input pdb file (opr3_hal1.pdb) the catalytic residues mentioned in the constraint files were again included in the REMARK section:

```

REMARK 666 MATCH TEMPLATE X TST 0 MATCH MOTIV A HIS 176 1 1
REMARK 666 MATCH TEMPLATE X TST 0 MATCH MOTIV A HIS 179 2 1
REMARK 666 MATCH TEMPLATE X TST 0 MATCH MOTIV B FMN 1 3 1

```

The desired mutations were implemented by altering the resfile instructions. With every subsequent design run, the previous instructions were commented out (except for the '#FMN coordinating' part, which was included in every run for obvious reasons). The resfile part for one run begins at '#opr3_hal1_n.DE' and ends at '#opr3_hal1_(n+1)_DE'. The following part shows the complete resfile 'opr3_hal1.res' to make the process more comprehensible.

```

AUTO
USE_INPUT_SC

start

#opr3_hal1_1.DE
#99  A  APOLAR
#235 A  PIKAA VLI
#236 A  PIKAA NH   #interaction with Br
#124 A  PIKAA A   #I124A required for Y140R
#140 A  PIKAA R   #interaction with Br
#104 A  PIKAA AS  #A104S for interaction with R140???

#opr3_hal1_2.DE
#99  A  APOLAR
#235 A  PIKAA VLI
#236 A  PIKAA NH
#124 A  PIKAA A
#140 A  PIKAA R
#104 A  PIKAA AS

```


#181 A PIKAA A #Y181 needs to be A for good ligand positioning

#opr3_hal1_3_DE

#99 A APOLAR

#235 A PIKAA VLI

#236 A PIKAA NH

#124 A PIKAA A

#140 A PIKAA R

#104 A PIKAA N #interaction with Br?

#181 A PIKAA A

#65 A PIKAA AVF #F too big? Maybe A,V?

#opr3_hal1_4_DE =same as op3_hal1_3_DE but without "irrelevant (?)" mutations

#99 A APOLAR

#235 A PIKAA VLI

#236 A PIKAA NH

#124 A PIKAA A

#140 A PIKAA R

#104 A PIKAA N

#181 A PIKAA A

#65 A PIKAA AVF

#these residues are not in the active site, but often mutated; this is to keep

#them native, but not force them into their native conformation

(maybe also Y361F ?)

#21 A PIKAA A #always mutated to D

#22 A PIKAA P #often mutated to S (?!)

#23 A PIKAA M #S,T

#274 A PIKAA R #Y

#313 A PIKAA Y #M,T

#336 A PIKAA F #A

#360 A PIKAA F #M,L,V

#-> residues are mostly in their native conformation, scores get slightly

#worse compared to opr3_hal1_3_DE

#opr3_hal1_5_DE

#99 A APOLAR

#235 A PIKAA VLI

#236 A PIKAA NH
#124 A PIKAA S
#104 A PIKAA N
#181 A PIKAA A
#65 A PIKAA AVF

#21 A PIKAA A
#22 A PIKAA P
#23 A PIKAA M
#271 A PIKAA T
#274 A PIKAA R
#313 A PIKAA Y
#336 A PIKAA F
#360 A PIKAA F

#opr3_hal1_6_DE

#99 A APOLAR
#235 A PIKAA VLI
#236 A PIKAA ST
#104 A PIKAA ST
#181 A PIKAA A
#65 A PIKAA AVF

#21 A PIKAA A
#22 A PIKAA P
#23 A PIKAA M
#271 A PIKAA T
#274 A PIKAA R
#313 A PIKAA Y
#336 A PIKAA F
#360 A NATRO

#opr3_hal1_7_DE

99 A APOLAR
235 A PIKAA VLI
236 A PIKAA ST
124 A PIKAA S
104 A PIKAA N
181 A PIKAA A
65 A PIKAA AVF

```
21  A  PIKAA A
22  A  PIKAA P
23  A  PIKAA M
271 A  PIKAA T
274 A  PIKAA R
313 A  PIKAA Y
336 A  PIKAA F
360 A  NATRO
```

```
#FMN coordinating
```

```
24  A  NATRO
54  A  NATRO
55  A  NATRO
97  A  NATRO
228 A  NATRO
310 A  NATRO
312 A  NATRO
333 A  NATRO
334 A  NATRO
1   B  NATRO
```

The design process was run via the following command line:

```
~/rosetta/rosetta_source/bin/enzdes.linuxgccrelease -database ~/rosetta/rosetta_database -extra_res_fa FMN_opr3.params -extra_res_fa TST_hal1.params -enzdes:cstfile opr3_hal1.cst -resfile opr3_hal1.res -s opr3_hal1.pdb -in:ignore_zero_occupancy false -nstruct 10 -out:file:o scores.txt @design.flags -out:suffix _n_DE
```

The file 'design.flags' was identical to the one used previously (see page 16). The flag '-out:suffix _n_DE' (with n being an integer) was used to add an identifier to the output file names of each run in order to keep track of which files were derived from which resfile instructions. As the transition state molecule appeared to be slightly misplaced in the output files, a new input scaffold pdb was generated (opr3_hal1_alt_lig_pos.pdb). The constraint file records were updated to fit with the new input structure and further design runs were performed using the resfile 'opr3_hal1_alt_lig_pos.res':

```
AUTO
USE_INPUT_SC
```

start

#opr3_hal1_alt_lig_pos_1_DE

#99 A APOLAR
#235 A PIKAA VLI
#236 A PIKAA ST
#124 A PIKAA S
#104 A PIKAA N
#181 A PIKAA A
#65 A PIKAA AVF

#opr3_hal1_alt_lig_pos_2_DE

#99 A APOLAR
#235 A PIKAA VLI
#236 A PIKAA N
#104 A PIKAA ST
#181 A PIKAA VLA
#65 A PIKAA AVFM

#opr3_hal1_alt_lig_pos_3_DE

99 A PIKAA M
235 A PIKAA L
236 A PIKAA N
104 A PIKAA S
181 A PIKAA A
65 A PIKAA M

#FMN coordinating

24 A NATRO
54 A NATRO
55 A NATRO
97 A NATRO
228 A NATRO
310 A NATRO
312 A NATRO
333 A NATRO
334 A NATRO
1 B NATRO

```

#other

21    A  PIKAA A
22    A  PIKAA P
23    A  PIKAA M
178   A  PIKAA A
271   A  PIKAA T
274   A  PIKAA R
313   A  PIKAA Y
332   A  PIKAA Y
336   A  PIKAA F
360   A  NATRO

```

The final sequence for the designed enzyme was derived from the output structures of 'opr3_hal1_alt_lig_pos_3_DE'.

For the design of enzymes converting the other substrates hal2, hal3 and hal4 procedures following the same principles were carried out (files not shown).

2.2 Protein expression and purification

The genes for eight designed OPR3 variants were obtained by site-directed mutagenesis of the wild type gene in a pET21a vector. For protein expression the plasmid was transformed into *Escherichia coli* BL21 and the enzymes were expressed and purified to be subjected to activity assays. All these procedures were done at the Graz University of Technology^[13]. In order to do crystallization experiments, some more protein was expressed as a part of this work. All designed enzymes except for the two variants of OPR3_4ring, which were reported to be heavily precipitating during purification, were expressed. Cells were grown at 37 °C in LB medium to OD₆₀₀ = 0.6. After induction with 0.2 mM IPTG (isopropyl- β -D-1-thiogalactopyranoside) cells were cultivated another 4 hours at 30 °C. Cells were harvested (3000 min⁻¹, 4 °C, 30 min) and the pellet was stored at -20 °C or directly used to extract protein. For protein purification the cells were suspended in 3 ml/mg lysis buffer (50 mM Na-phosphate, 150 mM NaCl, 20 mM imidazole, pH 7.5) and disrupted by sonication (10 min, pulsed). The cell extract was supplemented with 1 mM PMSF (phenylmethanesulfonylfluoride) as a protease inhibitor and debris was removed by centrifugation (16 000 min⁻¹, 30 min, 4 °C). The supernatant was supplemented with 1 μ M FMN and incubated o/n at 4 °C. The His-tagged protein was purified using a Ni-chromatography column (HisTrap HP 5 ml, GE Healthcare) equilibrated with 50 mM Na-phosphate, 150 mM NaCl, pH 7.5 on an ÄKTA-system. Elution was performed by a continuous gradient of imidazole from 0 - 300 mM. The protein was monitored via

absorption of the bound flavin at 465 nm. The fractions containing the protein were pooled and subjected to a size exclusion chromatography column (HiLoad 16/60 Superdex 200 pg, GE Healthcare) equilibrated with 50 mM Na-phosphate, 150 mM NaCl, pH 7.5. The protein was concentrated using centrifugal filters (Amicon Ultra, Merck Millipore).

2.3 Crystallization and structure determination

Crystallization experiments were set up with the OPR3_Y190F variant and all designed enzymes except for the two variants of OPR3_4ring. Commercial crystallization screens (Index HR2-144, Hampton Research; Morpheus MD1-46, Molecular Dimensions) were set up, and obtained crystals were further optimized. During the time of this study diffraction quality crystals could be obtained for three variants.

2.3.1 Crystallization of OPR3_5ring and OPR3_6ring

Initial crystals grew after three days at 20 °C in droplets consisting of equal parts of protein (4 mg/ml in 50 mM Na-phosphate pH 7.5, 150 mM NaCl) and reservoir solution (0.2 M NaCl, 0.1 M BIS-TRIS pH 6.5, 25% w/v polyethylene glycol 3,350). The method used was sitting drop vapor diffusion. The first crystals were used for microseeding (MicroSeed Beads MD2-14, Molecular Dimensions). Crystals were transferred into 50 μ l of reservoir solution and vortexed for 20 seconds to prepare a seeding stock. Optimization was done by varying dilutions of the seeding stock and testing different concentrations of NaCl solutions in the reservoirs in sitting drop vapor diffusion experiments. Good quality crystals were grown in droplets consisting of 2 μ l protein (4 mg/ml), 1.5 μ l seeding solution (diluted 50⁻¹ from the stock) and 0.5 μ l crystallization solution (0.2 M NaCl, 0.1 M BIS-TRIS pH 6.5, 25% w/v polyethylene glycol 3,350), with 0.6 M NaCl in the reservoir. Data were collected at the ESRF Grenoble, beamline ID29.

2.3.2 Crystallization of OPR3_Y190F

Small crystals were grown using the sitting drop vapor diffusion method in droplets containing 50 % protein (4 mg/ml in 50 mM Na-phosphate pH 7.5, 150 mM NaCl) and 50 % reservoir solution (0.2 M ammonium citrate tribasic pH 7.0, 20% w/v polyethylene glycol 3,350). From these crystals a microseeding stock was prepared (seed beads, 50 μ l) and diluted by 50⁻¹. Diffracting crystals were grown in 2 μ l droplets composed of 1 μ l protein (4 mg/ml), 0.5 μ l seeding solution and 0.5 μ l reservoir solution (sitting drop vapor diffusion). Data were collected at the ESRF Grenoble, beamline BM14.

2.3.3 Soaking of OPR3_Y190F crystals with 4-bromobut-2-enal

Since the flavin is in its oxidized state in aerobic conditions, without NAD(P)H no substrate conversion would take place and therefore the substrate compound ((*E*)-4-bromobut-2-enal)

was used in this experiment. For the soaking experiments 1 μ l of pure liquid substrate was pipetted to 10 μ l reservoir solution supplemented with 15 % glycerol v/v as a cryoprotectant. The substrate and reservoir solution turned out to be not miscible, however crystals were soaked in the reservoir phase for various timespans (1 min, 2 min, 5 min, 10 min, 20 min) before storing them in liquid nitrogen.

2.3.4 Data processing and refinement

Data were processed and scaled with XDS^[14] or iMosflm^[15], Pointless^[16;17] and Scala^[17]. The structures were solved by molecular replacement using Molrep^[18]. Refinement was done in alternating cycles of model building with Coot^[19] and Refmac^[20;21]. Further refinement was done using the PDB_REDO web server^[22]. For the generation of a library file for the restraints on the fitted substrate molecule Schrödinger Maestro and eLBOW^[23] from the Phenix suite were used.

3 Results and discussion

3.1 Design process

3.1.1 Transition state structures

Potential energy profiles were obtained by scanning the reaction coordinate of the various substrates. The presence of a local maximum is an indication for a transition state. In the graphs of hal1 and hal3 a maximum is visible (Fig. 12 and Fig. 14, respectively). In the case of hal2 there is small peak in the range of a supposed local maximum (Fig. 13). This is due to a change in the conformation of the molecule. For hal4 a more sophisticated 2D scan was performed, resulting in a potential energy surface (Fig. 15). Also in that case a small local maximum can be recognized. For calculation of the final transition state models, for each substrate an intermediate structure near the local maximum was chosen as an input model for a direct optimization towards a transition state (Fig. 16).

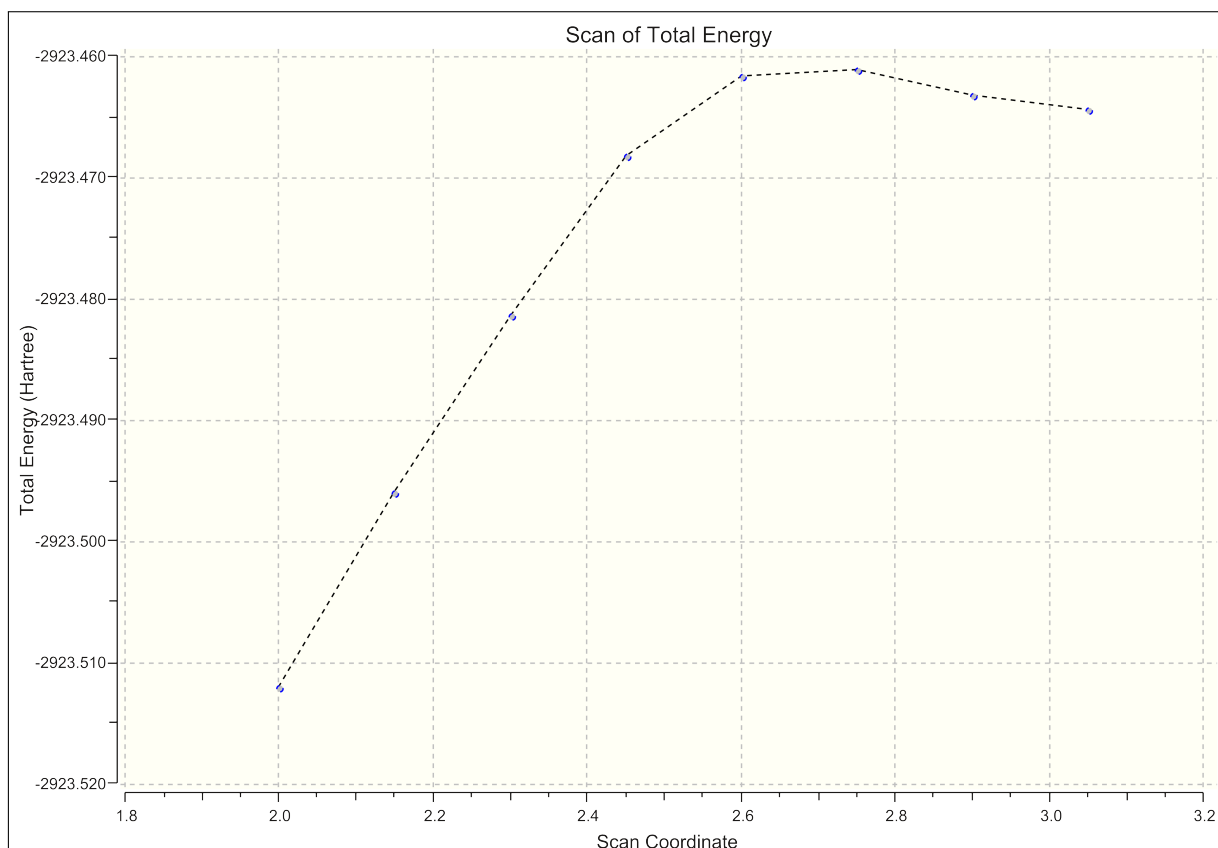


Figure 12 – Potential energy profile for the calculation of a transition state for hal1. The horizontal axis shows the C2-C7 bond length in Å. A local maximum is between 2.60 and 2.75 Å. The minimum at the left side is the energy of the initial model.

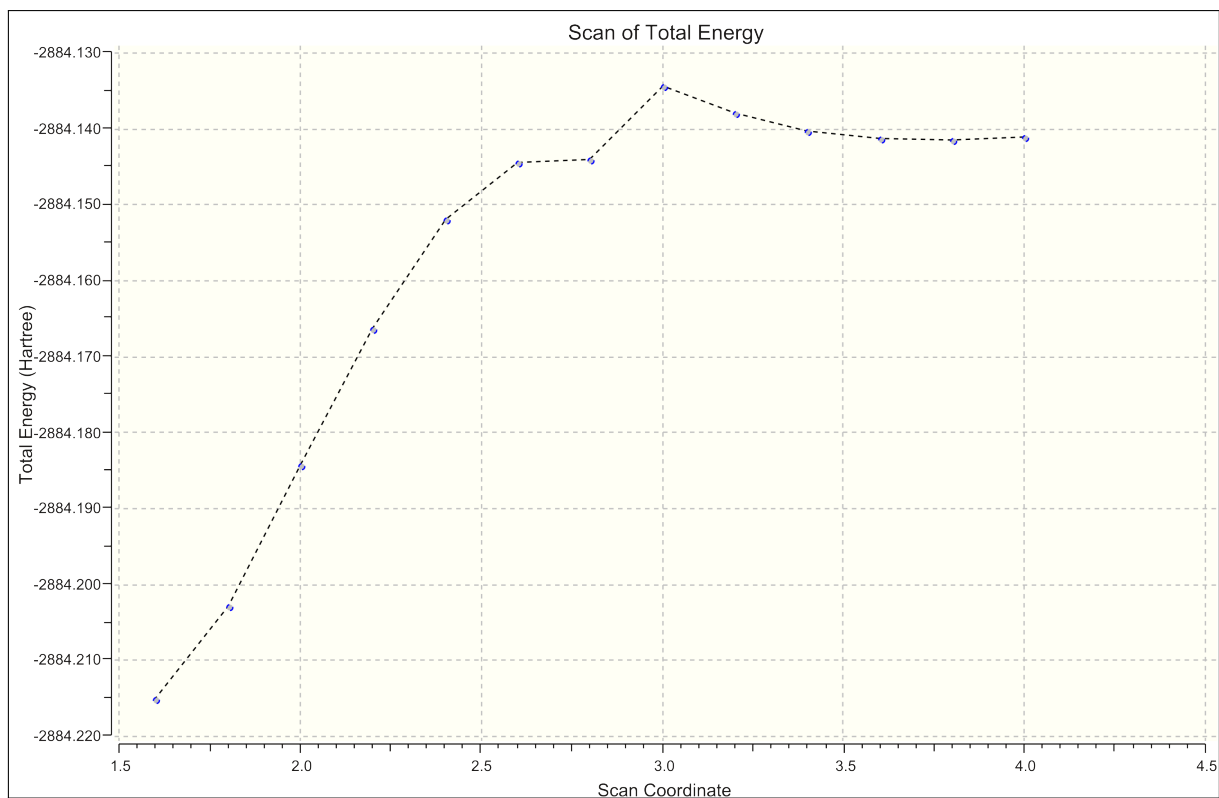


Figure 13 – Potential energy profile for the calculation of a transition state for hal2. The horizontal axis shows the C2-C6 bond length in Å. The minimum at the left side would be a state that is near the product - a cyclic compound and free bromide. The point at the very right side is the initial model, the path during the calculation went from right to left. The rise and sudden decrease of the energy at 3 Å is due to a conformational change. The input structure in that case was poorly modeled, a starting conformation analogous to hal1 or hal3 would have been the better choice.

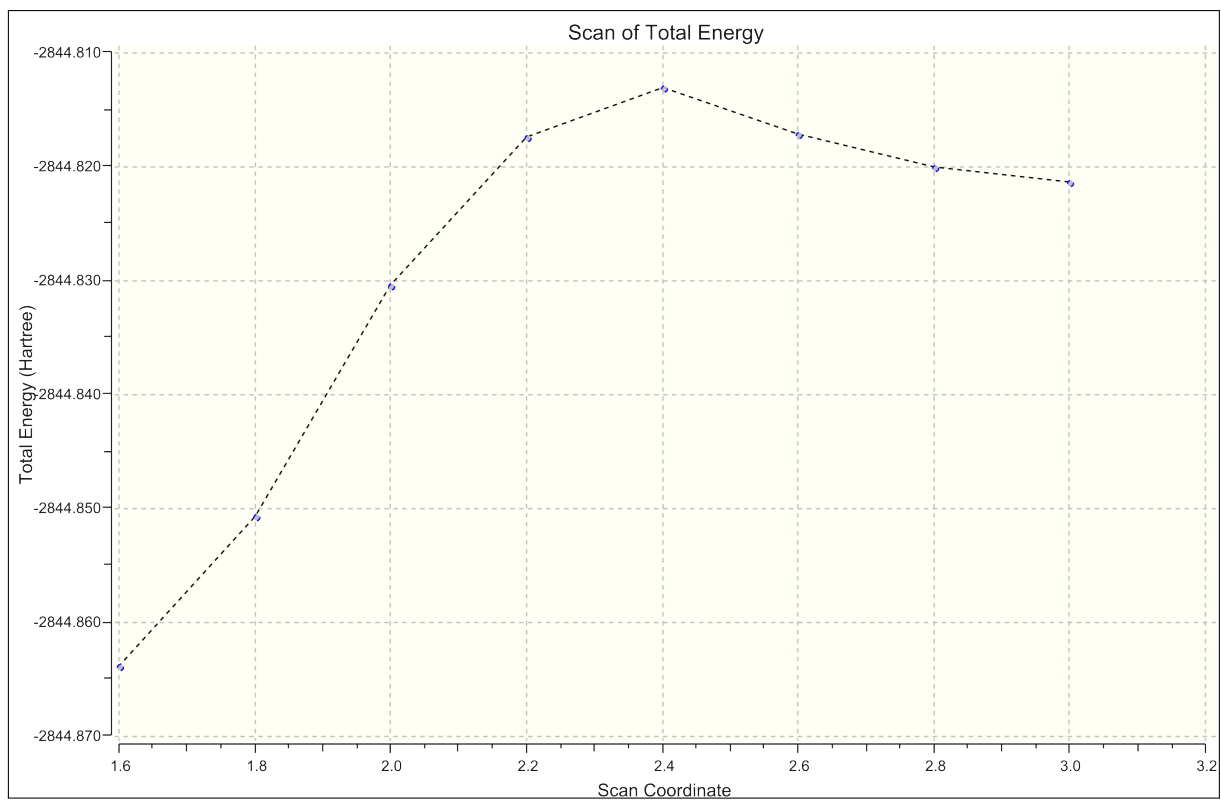


Figure 14 – Potential energy profile for the calculation of a transition state for hal3. The horizontal axis shows the C2-C5 bond length in Å. The minimum at the left side is a state that is near the product (C2-C5 bond length 1.6 Å). The path during the calculation went from right to left. A local maximum can be seen at 2.4 Å.

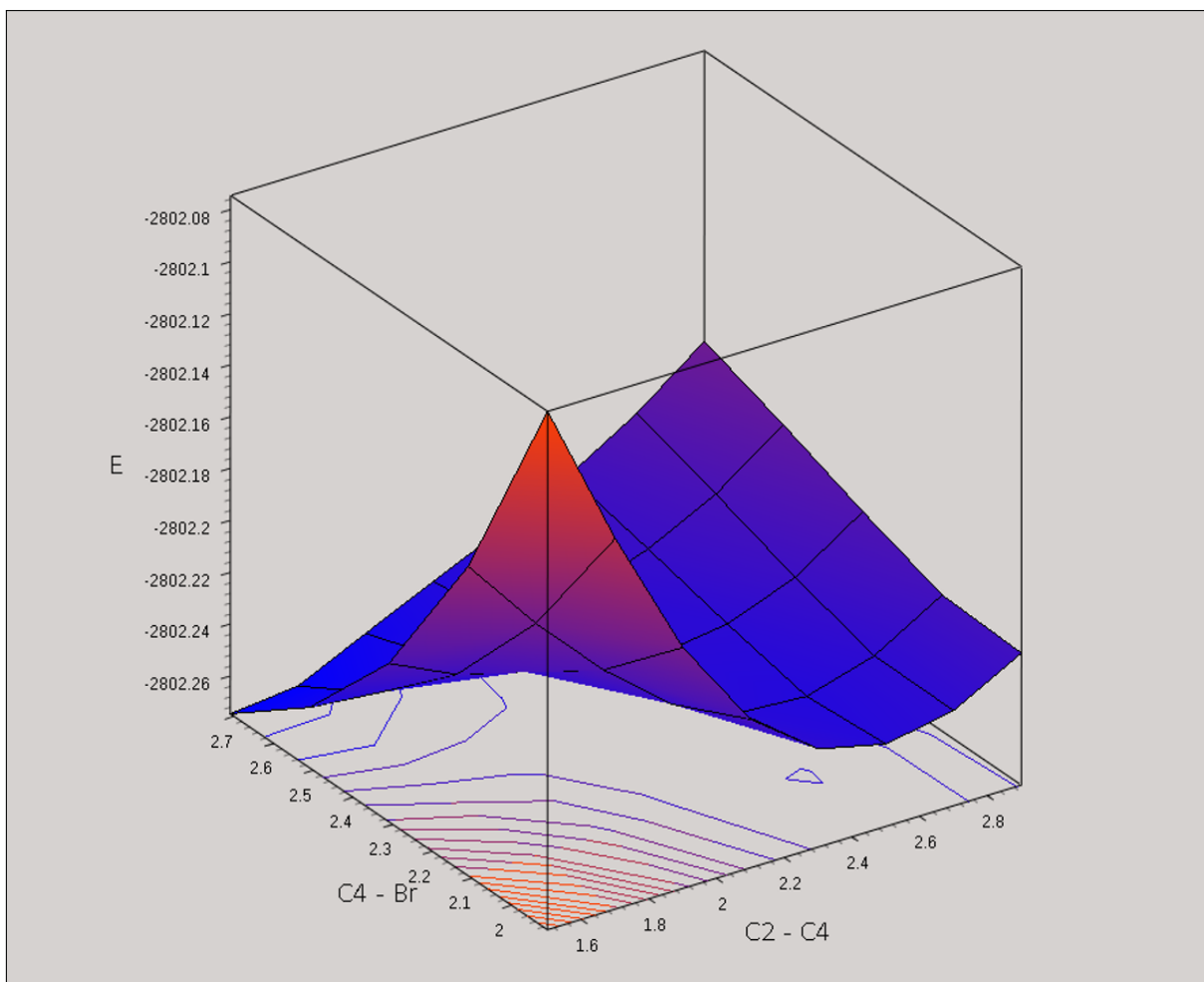


Figure 15 – Transition state search for hal4. With this more complex procedure a potential energy surface was obtained. The horizontal axes show distances in Å, the vertical axis the total energy in Hartree. The front corner (C2-C4 1.5 Å; C4-Br 1.9 Å) shows the high energy of the energetically unfavorable starting model where C4 has a valence of 5. In the left corner (C2-C4 1.5 Å; C4-Br 2.7 Å) there is a minimum that depicts the product and at C2-C4 2.5 Å; C4-Br 1.9 Å there is a local minimum which would be a state where the hydride is transferred to the substrate, but the bromine still bound. The small peak in the projection on the bottom plane between these two points indicates a local maximum, which refers to the desired transition state (C2-C4 2.5 Å; C4-Br 2.1 Å).

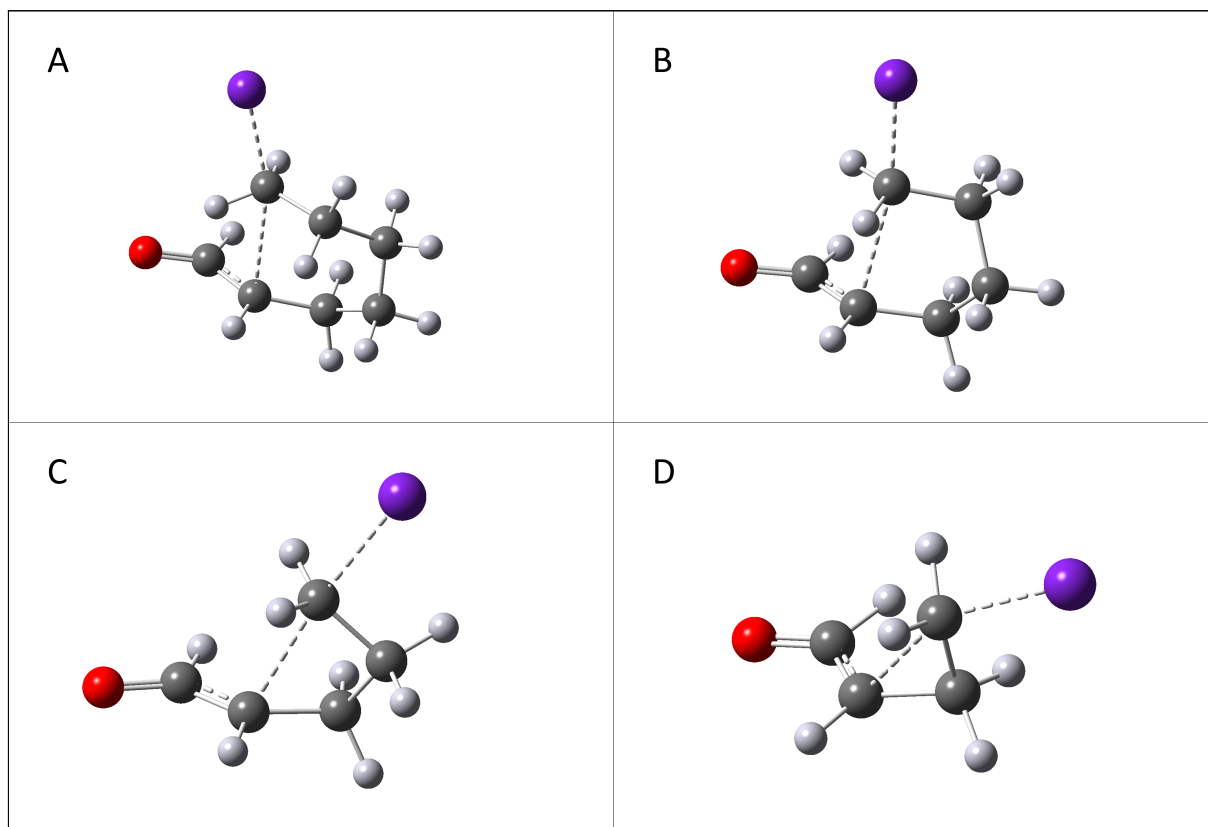


Figure 16 – The calculated transition state models. A: hal1 (distances C2-C7: 2.71 Å; C7-Br: 2.29 Å). B: hal2 (distances C2-C6: 2.69 Å; C6-Br: 2.32 Å). C: hal3 (distances C2-C5: 2.36 Å; C5-Br: 2.43 Å). D: hal4 (distances C2-C4: 2.21 Å; C4-Br: 2.37 Å). (Red=oxygen, purple=bromine).

3.1.2 Initial design trials using the Rosetta match application

In some of the first design attempts it was tried to find possible amino acid exchanges at various sequence positions with the match application. Besides the interactions of the two hydrogen bonding histidine side chains with the substrate's carbonyl oxygen, two additional hydrogen bond donors to interact with the bromine atom were requested. For a broad range of possible donors (Ser, Thr, Lys, Arg, His, Gln, Asn, Tyr, Trp) an optimized hydrogen bonding geometry was specified in the constraint files along with posfiles including many sequence positions all over the active site from where these residues could be built. The computational efforts were quite high, and even with loosening the geometric constraints for the interactions to very low levels (far from optimum geometry) no matching outputs could be obtained. The consequence was that in the next attempts only one interaction with the bromine atom was requested. This was tried for three substrates (hal2, hal3 and hal4), and by stepping back to a single interaction at least a few matches were found.

A problem with these designed structures turned up when it was not possible to use an important option for the enzyme design application: `-enzdes:final_repack_without_ligand`. This option causes the program to repack the structure after design without the catalytic

constraints in order to test if the mutated residues are not just in their current conformation because the constraints keep them there. For this option it seems that any non-protein compound is considered as a 'ligand', and thus the program could not distinguish between the transition state model and the FMN cofactor, which caused an error and the design run to fail. Without this option the predicted positions of catalytic residues, especially the long side chains of arginine and lysine, were not considered as very reliable, which was another reason why it was then tried to produce designed structures without the match application.

3.1.3 Protein design simply using the Rosetta enzyme design application

A different approach in comparison to using the match application to find suitable positions for amino acid exchanges was simply to skip the matching step and start right at the actual design process. The idea behind this method was that the active site geometry and the binding mode of the substrate was already very well defined. With the substrate binding to the two histidine side chains and right on top of the isoalloxazine moiety of FMN, there was only a limited number of possible sequence positions that could be subjected to mutations in order to further stabilize the binding of the transition state molecule. Including the experience gained from the design trials with the use of the match application, suitable sequence positions and promising mutations were tried out for amino acid exchanges, with the Rosetta enzyme design application as a tool to predict possible structures. The advantage was that there were no constraints on the catalytic residues that could force them into unnatural conformations (as the option `final_repack_without_ligand` was not available).

A total of eight different sequences were derived from the design outputs: Four variants for the conversion to the three-ring product (hal4), two variants for the four-ring product (hal3) and one variant for the five-ring (hal2) and six-ring product (hal1), respectively. All of the final design structures have two mutations in common: W99M (native position: W108M) appeared in nearly all output structures as a result of requesting Rosetta to make the residue at this position apolar to ensure a more hydrophobic surrounding of the substrate's C_α. For basically the same reason H235L (native: H244L) was introduced in all designed structures. Other mutations are specific for the various substrates.

OPR3_6ring (hal1) For this substrate one sequence was derived from the final design outputs. Besides Met-99 and Leu-235, four other mutations were introduced (Fig. 17). Phe-65 (native position: Phe-74) was eliminated to provide space for the substrate and changed to Met-65, which could build hydrophobic interactions and might help that the substrate molecule can only enter the active site in a bent conformation that brings C7 closer to C2. Tyr-181 (native Tyr-190) was changed in various design runs for either alanine, valine or leucine, from which most often alanine was selected by Rosetta. For interactions

with the bromine atom Leu-236 (native Leu-245) was mutated to a possible hydrogen bonding asparagine and Ala-104 (native Ala-113) was changed to serine. The predicted serine side chain conformations were not optimal for interaction with the substrate, but there was the idea that this serine residue might also help stabilizing the Asn-236 side chain via hydrogen bonding.

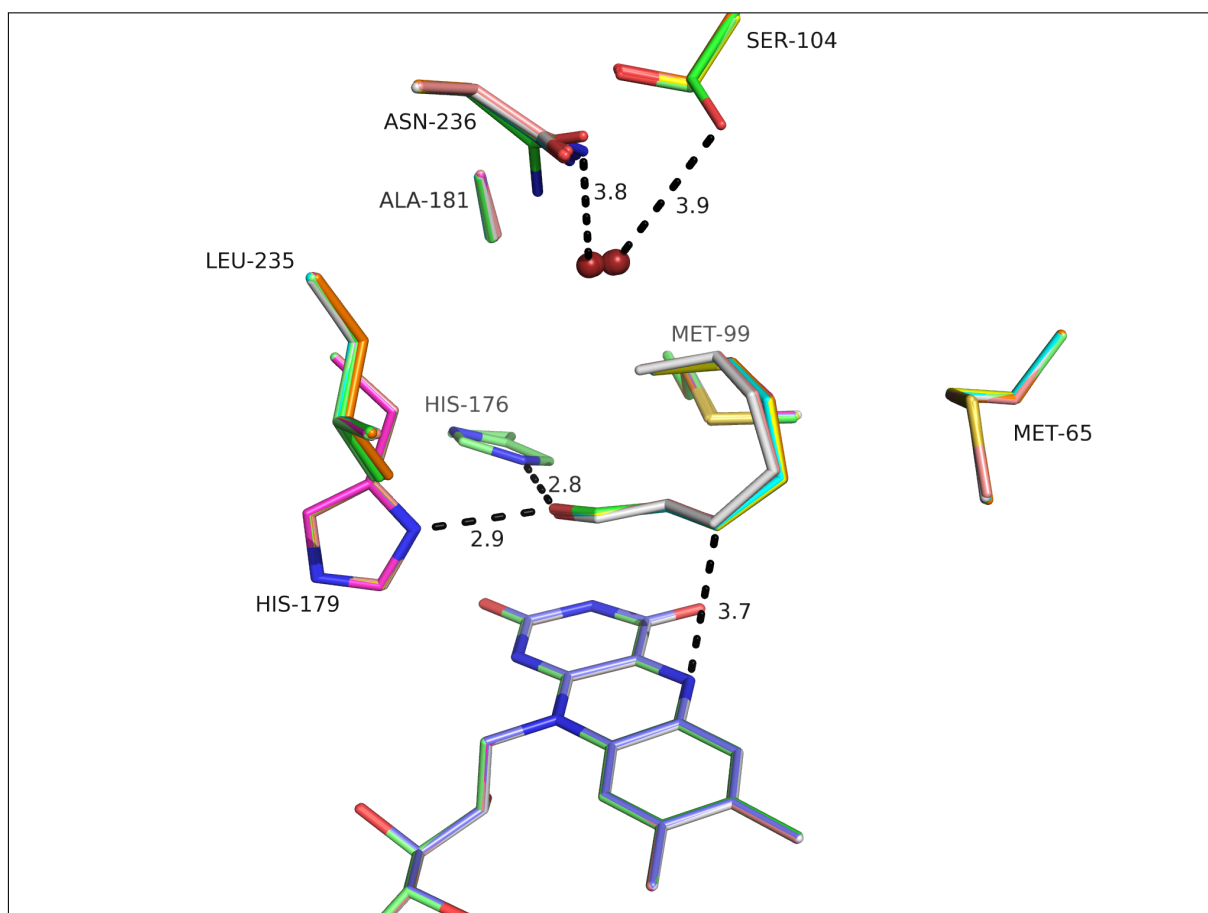


Figure 17 – Active site of the design outputs for OPR3_6ring (superposition). The transition state molecule for the conversion of 7-bromohept-2-enal is bound at the carbonyl group via His-176 and His-179. Leu-235 and Met-99 were introduced to make the active site pocket more hydrophobic. Asn-236 and Ser-104 interact with the bromine atom (red sphere, C-Br bond not shown). To provide space for the artificial substrate, Phe-65 was changed to methionine and Tyr-181 was mutated to alanine.

OPR3_5ring (hal2) The final version for this enzyme included six exchanged amino acids (Fig. 18). It only differs in one mutation from the variant OPR3_6ring. Due to the smaller substrate compared to OPR_6ring, Tyr-181 (native: Tyr-190) was changed to leucine (OPR3_6ring: alanine). At sequence position 99 (native: 108) the request for an apolar side chain led to introduction of either alanine or methionine. In order to get tighter packing in the relating area around the substrate, methionine was chosen.

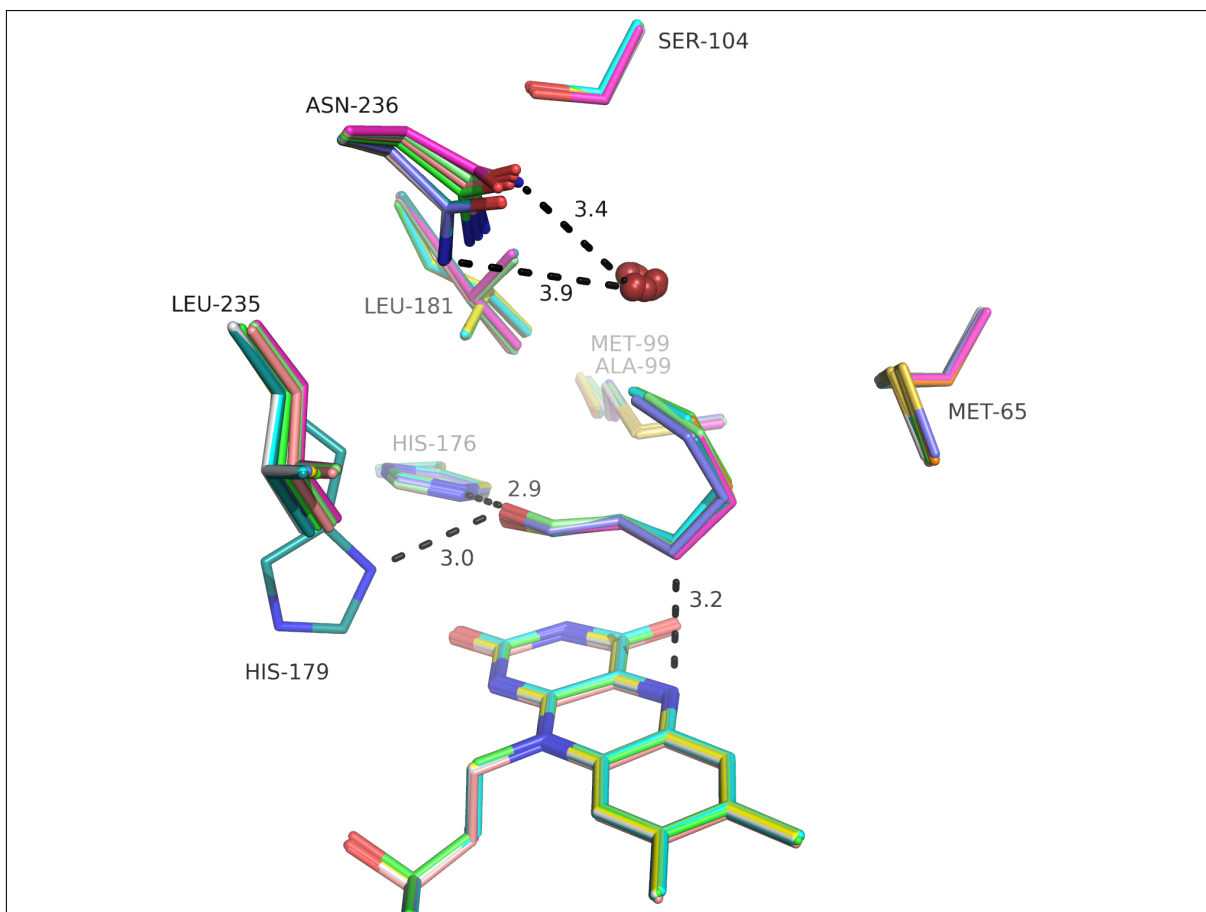


Figure 18 – Active site of the design outputs for OPR3_5ring (superposition). The transition state model is bound by the two native histidines. Asn-236 builds an interaction with the bromine atom (C-Br bond not shown). Despite Ser-104 did not appear to be in a good position to interact with the bromine atom, it was included in the final sequence, as a different side chain rotamer would be indeed in a promising position. At sequence position 99 in some structures the Rosetta program introduced alanine, in others methionine. For the final sequence of OPR3_5ring methionine was chosen.

OPR3_4ring (hal3) Two variants were selected for the hal3 substrate (OPR3_4ring.1-2). Each variant contained eight mutations, of which seven are the same in both enzymes (Fig. 19 and Fig 20). Tyr-181 (native Tyr-190) was changed to leucine, as phenylalanine would have been clashing with the transition state model, and smaller hydrophobic amino acids, like alanine or valine, would have left a too big gap. To interact with the bromine, Tyr-140 (native Tyr-149) was mutated to arginine. To help stabilizing the arginine side chain, Leu-236 (native Leu-245) was changed to asparagine and Ala-104 (native Ala-113) to serine. To provide enough space for the arginine residue, Ile-124 (native Ile-133) was mutated to a smaller valine. Phe-65 (native Phe-74) would have clashed with the transition state model and was therefore changed to alanine (variant 1) or, to provide an additional interaction with the substrate's bromine, to asparagine (variant 2). The other included mutations are the aforementioned Met-99 and Leu-235.

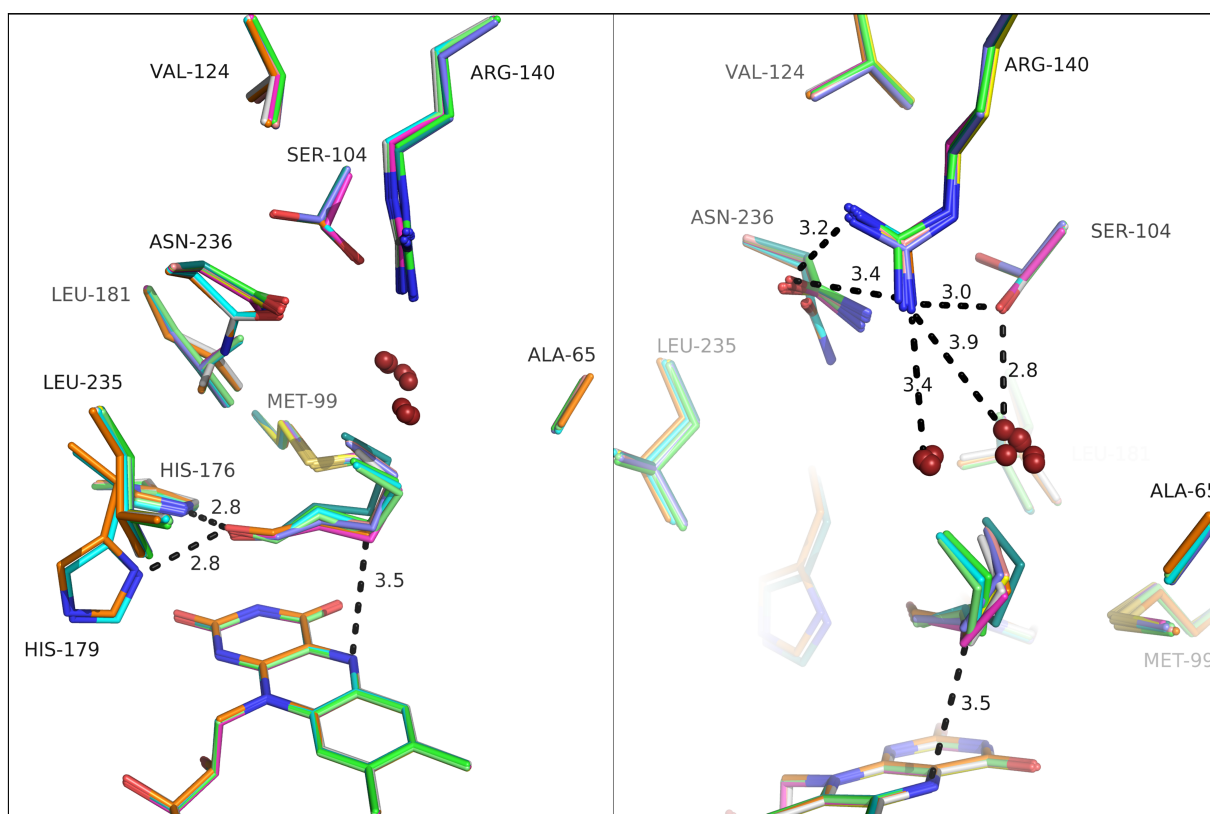


Figure 19 – Active site of the design outputs for OPR3_4ring_var1 (superposition). Leu-181 was chosen instead of the former proton donor tyrosine, Arg-140 was introduced to interact with the substrate’s bromine and Asn-236 and Ser-104 were planned to either stabilize the arginine side chain or to directly provide hydrogen bonds to the bromine (C-Br bond not shown). To provide enough space for Arg-104 the former Ile-124 was mutated to valine. Ala-104 was introduced instead of the native phenylalanine that would have occupied too much space in the active site.

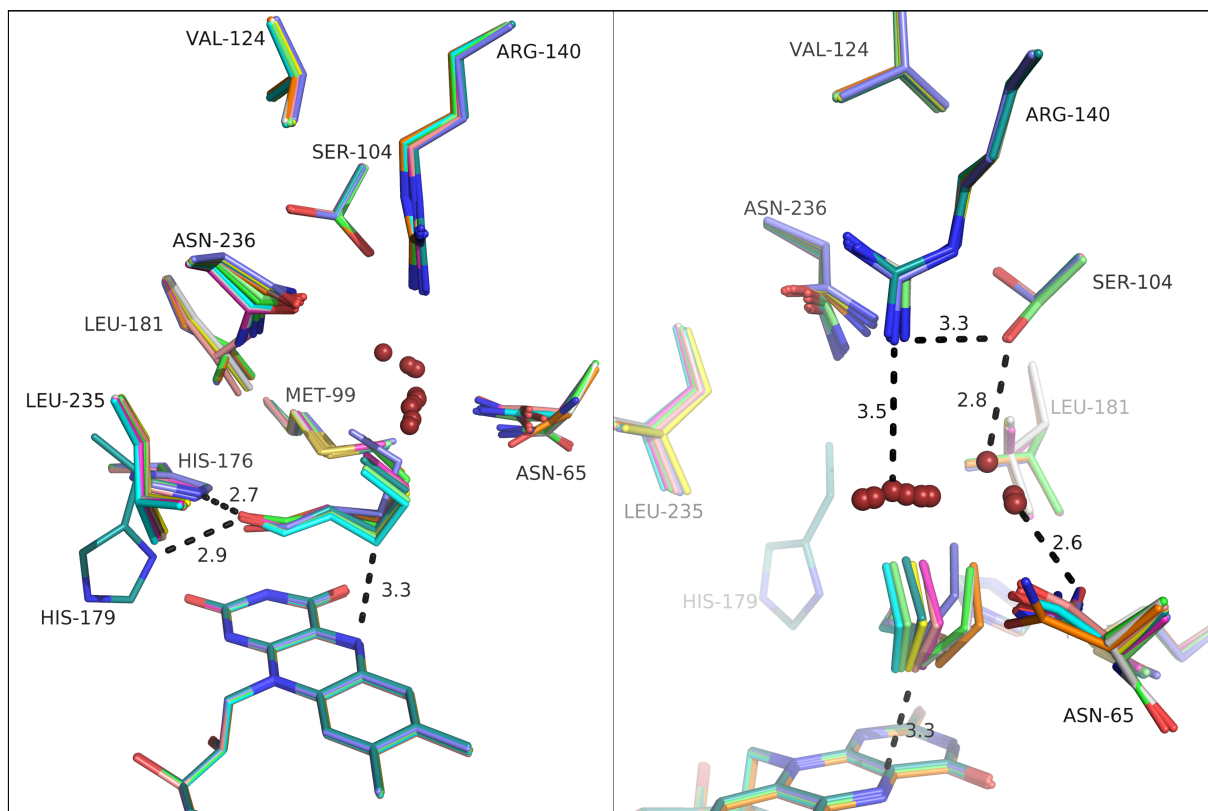


Figure 20 – Active site of the design outputs for OPR3_4ring_var2 (superposition). The only difference to OPR3_4ring variant 1 is Asn-65, which was chosen as a possible binding partner for the bromine/bromide leaving group (C-Br bond not shown).

OPR3_3ring (hal4) For the conversion of 4-bromobut-2-enal to cyclopropanecarbaldehyde a total of four slightly different variants were designed (OPR3_3ring_var1-4). All of them included six mutations (Fig. 21 - 24). To avoid direct protonation of the substrate upon hydride transfer, the Tyr-181 proton donor (native: Tyr-190) was exchanged for phenylalanine in all four designed enzymes. The native Phe-65 (native: Phe-74) would have clashed with the substrate. One possibility to introduce an interacting residue seemed to be an exchange for asparagine (variants 2 and 4). Another possibility was to introduce either alanine or valine to partially fill up the space that would have been occupied by tyrosine. In the vast majority of design output structures the Rosetta program chose alanine (present in variants 1 and 3). Leu-131 (native: Leu-140) was exchanged for either arginine (variants 1 and 2) or lysine (variants 3 and 4). The idea was to introduce a positively charged side chain in the vicinity of the substrate's bromine. In the calculated output structures this never worked well as Rosetta predicted the side chain always to be on the surface and not pointing towards the active site. However, the positively charged residues at this position were kept, as they might stabilize another mutated residue via hydrogen bonding: Tyr-361 (native: Tyr-370) was mutated to glutamine (variants 1-4). A glutamine side chain could build a hydrogen bond to the bromine. Especially for variants 3 and 4 an interaction between lysine and glutamine was not predicted by Rosetta, but nevertheless these mutations were tried to see how accurate Rosetta works for long side

chains and whether this lack of interaction was also true for the actual structures.

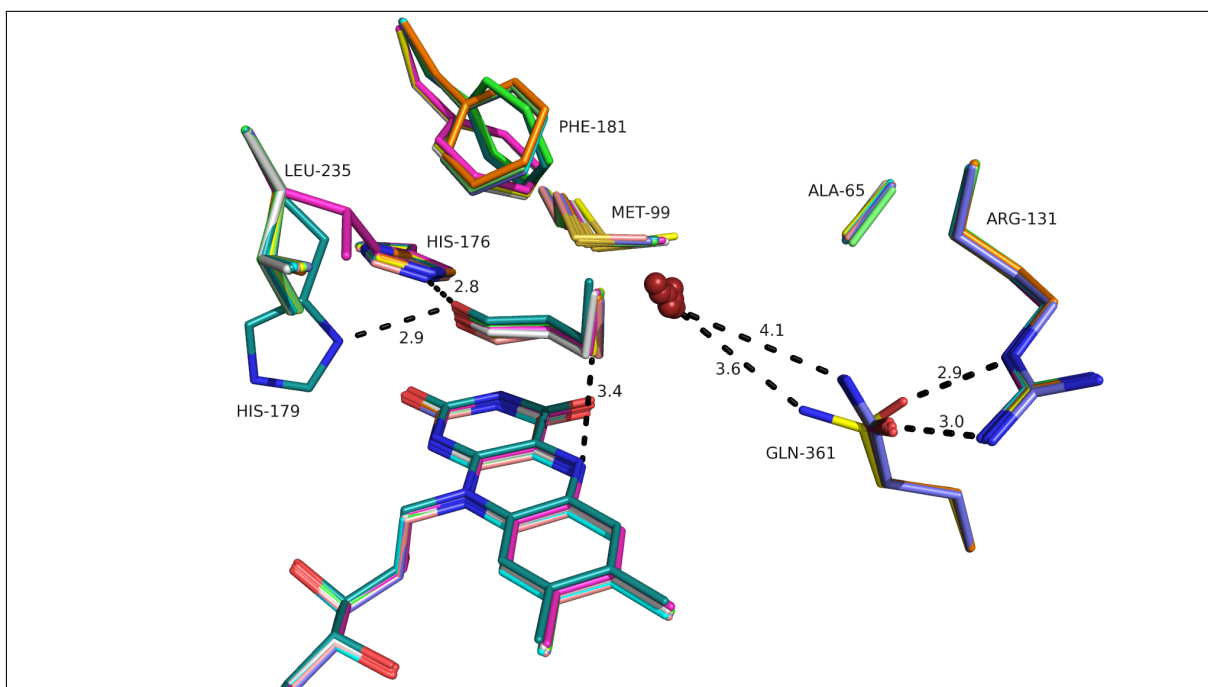


Figure 21 – Active site of the design outputs for OPR3_3ring_var1. A glutamine residue that is stabilized by an arginine side chain can interact with the bromine (C-Br bond not shown). The alanine side chain partially fills up a space that would be occupied by a tyrosine residue in the native structure. The image shows a superposition of ten output structures.

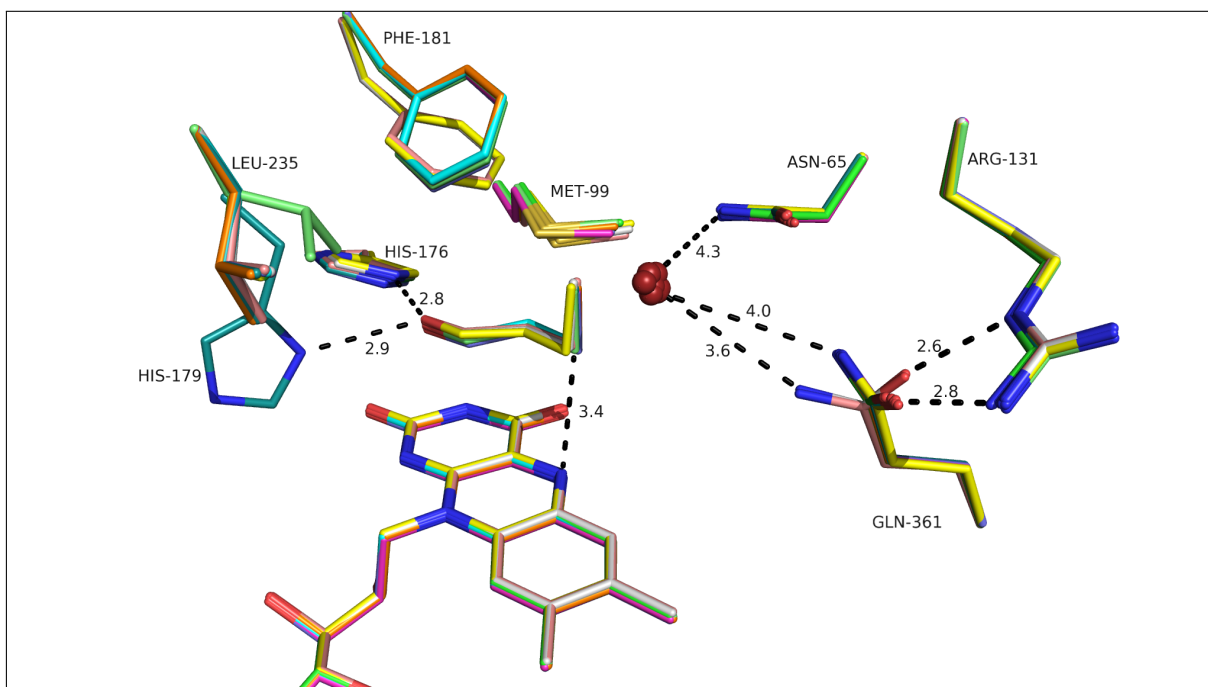


Figure 22 – Active site of the design outputs for OPR3_3ring_var2. Much like variant 1, an additional interaction between bromine and an asparagine side chain is introduced (C-Br bond not shown). The image shows a superposition of ten output structures.

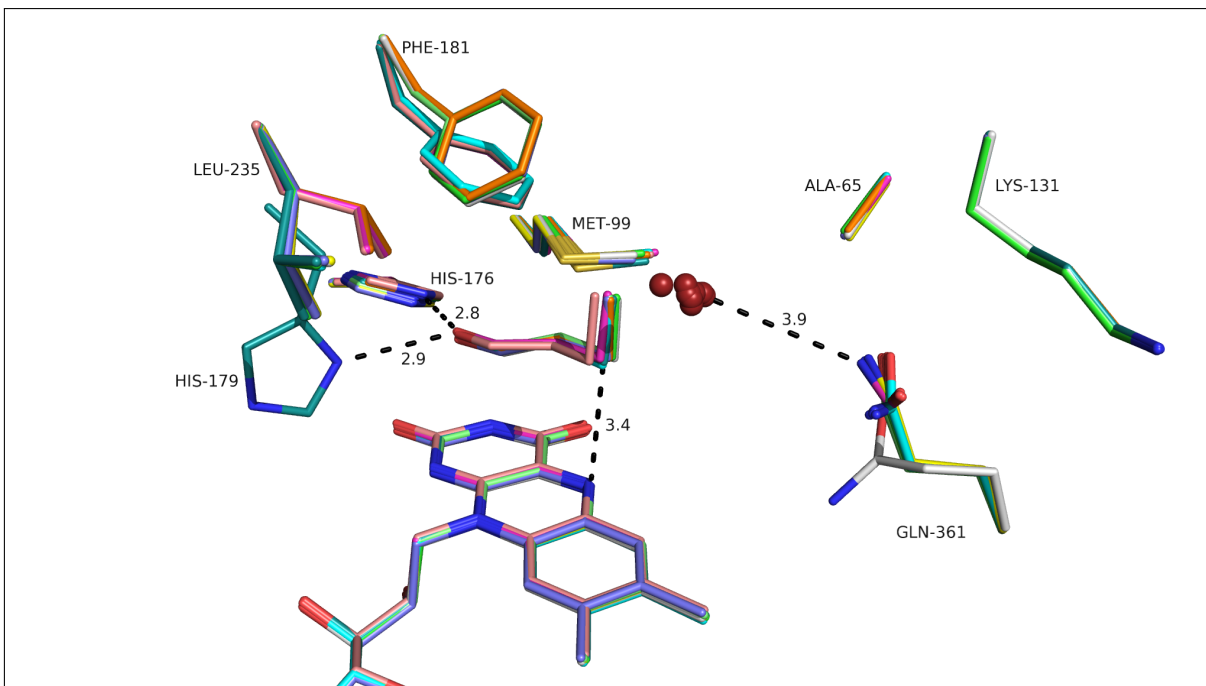


Figure 23 – Active site of the design outputs for OPR3_3ring_var3. Although the lysine side chain is predicted to be outside of the active site pocket, this variant was tried. If Rosetta’s prediction failed, there might be a chance that the lysine interacts with the substrate directly by stabilizing the negatively charged bromide leaving group (C-Br bond not shown). The image shows a superposition of ten output structures.

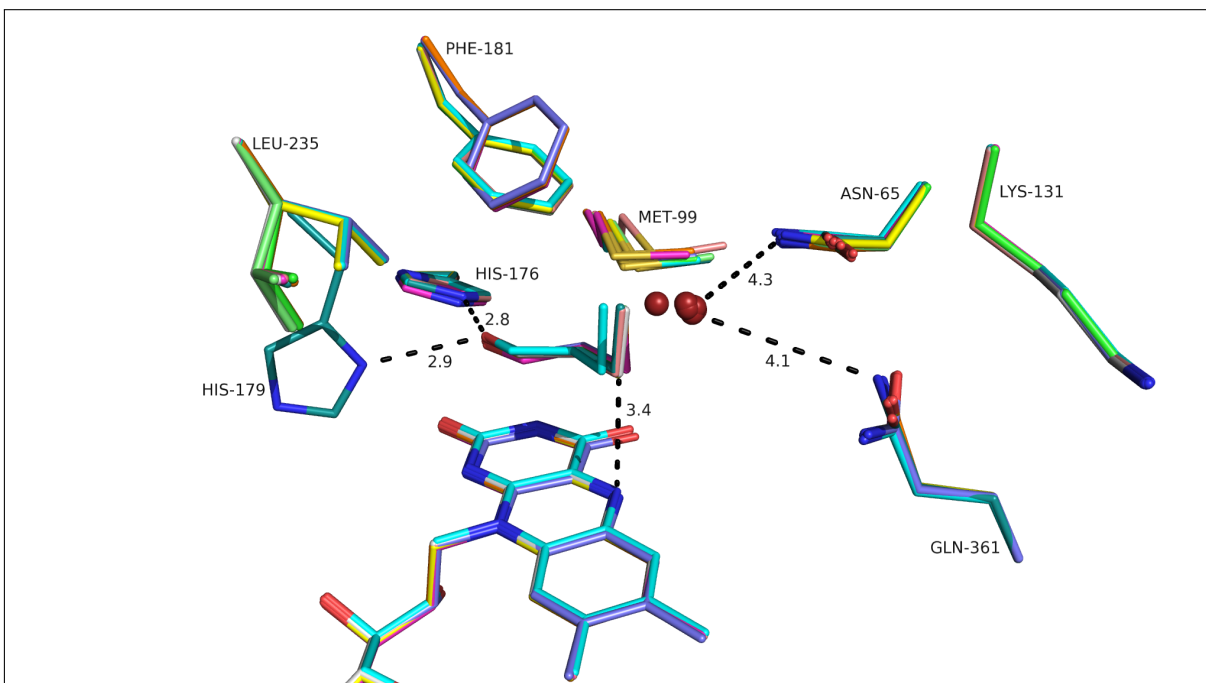


Figure 24 – Active site of the design outputs for OPR3_3ring_var4. This variant is like variant 3, but with an additional interaction between bromine and an asparagine side chain (C-Br bond not shown). The image shows a superposition of ten output structures.

3.1.4 Summary

The complete redesign of active sites for reactions that require a bound cofactor is in general a very delicate task, as any disruption of the binding sites might have a big influence on the desired reaction. On the one hand, the binding of the cofactor should not be affected by the design, on the other hand for a complete redesign a part of the protein near this cofactor is mutated at several sites. This issue makes the procedure quite a balancing act, and as every design process is different, there is no general solution provided in Rosetta and therefore the user has to think about how far the program is allowed to change the structure. However, one has to keep in mind that too rigid instructions could suppress the occurrence of optimal solutions or even force the program to create biased output.

A possible source for a systematic error in this design process was in the modeling of the design input structures regarding the distance between the substrate's C_α and N5 of the flavin cofactor. It was assumed that positioning the transition state molecules like an inhibitor in a known crystal structure is sufficiently accurate to obtain design structures in which actual hydride transfer can occur. However, since during the design process also repacking of the altered structures is done by Rosetta and therefore slight movements need to be allowed, modeling a distance between two atoms by less than tenths of angstroms seems to be beyond the program's capabilities, and therefore would not make sense.

The eight enzyme variants that were finally selected for expression and in-vitro testing for their desired catalytic activity were the result of a long process with many output structures, that all were slightly different. The sequences were derived after careful investigation from structures that looked the most promising regarding the interactions with the transition state models and their chemical plausibility. Some of the interactions between side chains and the transition state models did not appear to be in an optimal geometry. Taken into account that the whole design process was based on the static model of a single crystal structure, and being dependent on the accuracy of the underlying force fields that are used by the Rosetta applications, the reliability of the output had its limitations, and at some point the designed enzymes were given a try. Table 1 shows a summary of the mutations in the different variants.

Table 1 – Overview of the differences in the designed OPR3 variants.

Enzyme	Mutated amino acids:									
	F74	W108	A113	I133	L140	Y149	Y190	H244	L245	Y370
OPR3_6ring	M	M	S				A	L	N	
OPR3_5ring	M	M	S				L	L	N	
OPR3_4ring_var1	A	M	S	V		R	L	L	N	
OPR3_4ring_var2	N	M	S	V		R	L	L	N	
OPR3_3ring_var1	A	M			R		F	L		Q
OPR3_3ring_var2	N	M			R		F	L		Q
OPR3_3ring_var3	A	M			K		F	L		Q
OPR3_3ring_var4	N	M			K		F	L		Q

The complete sequences of the designed enzymes are shown in the following alignment:

```

OPR3 native      MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_6ring       MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_5ring       MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_4ring_var1  MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_4ring_var2  MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_3ring_var1  MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_3ring_var2  MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_3ring_var3  MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
opr3_3ring_var4  MASSAQDGNNPLFSPYKMGKFNLSHRVVLAPMTRCRALNNIPQAALGEYYEQRATAGGFL  60
*****

OPR3 native      ITEGTMISPTSAGFPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLWHVGRASHEVYQP  120
opr3_6ring       ITEGTMISPTSAGMPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRSSHEVYQP  120
opr3_5ring       ITEGTMISPTSAGMPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRSSHEVYQP  120
opr3_4ring_var1  ITEGTMISPTSAGAPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRSSHEVYQP  120
opr3_4ring_var2  ITEGTMISPTSAGNPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRSSHEVYQP  120
opr3_3ring_var1  ITEGTMISPTSAGAPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRASHEVYQP  120
opr3_3ring_var2  ITEGTMISPTSAGNPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRASHEVYQP  120
opr3_3ring_var3  ITEGTMISPTSAGAPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRASHEVYQP  120
opr3_3ring_var4  ITEGTMISPTSAGNPHVPGIFTKEQVREWKKIVDVVHAKGAVIFCQLMHVGRASHEVYQP  120
*****
*****
*****:*****

```

OPR3 native AGAAPISSTEKPISNRWRILMPDGTHGIYKPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_6ring AGAAPISSTEKPISNRWRILMPDGTHGIYKPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_5ring AGAAPISSTEKPISNRWRILMPDGTHGIYKPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_4ring_var1 AGAAPISSTEKPVSNRWRILMPDGTHGIRPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_4ring_var2 AGAAPISSTEKPVSNRWRILMPDGTHGIRPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_3ring_var1 AGAAPISSTEKPISNRWRIRMPDGTHGIYKPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_3ring_var2 AGAAPISSTEKPISNRWRIRMPDGTHGIYKPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_3ring_var3 AGAAPISSTEKPISNRWRIRMPDGTHGIYKPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180
 opr3_3ring_var4 AGAAPISSTEKPISNRWRIRMPDGTHGIYKPKPRAIGTYEISQVVEDYRRSALNAIEAGFD 180

*****:***** ***** *****

OPR3 native GIEIHGAHGYLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_6ring GIEIHGAHGALIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_5ring GIEIHGAHGLLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_4ring_var1 GIEIHGAHGLLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_4ring_var2 GIEIHGAHGLLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_3ring_var1 GIEIHGAHGFLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_3ring_var2 GIEIHGAHGFLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_3ring_var3 GIEIHGAHGFLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240
 opr3_3ring_var4 GIEIHGAHGFLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVSAIGADRVGVRVSP 240

***** *****

OPR3 native AIDHLDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_6ring AIDLNDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_5ring AIDLNDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_4ring_var1 AIDLNDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_4ring_var2 AIDLNDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_3ring_var1 AIDLLDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_3ring_var2 AIDLLDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_3ring_var3 AIDLLDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300
 opr3_3ring_var4 AIDLLDAMDSNPLSLGLAVVERLNKIQHSGSKLAYLHVTQPRYVAYGQTEAGRLGSEEE 300

*** *****

```

OPR3 native      EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_6ring      EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_5ring      EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_4ring_var1 EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_4ring_var2 EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_3ring_var1 EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_3ring_var2 EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_3ring_var3 EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
opr3_3ring_var4 EARLMRTLRLNAYQGTFICSGGYTRELGIEAVAQGDADLVSYGRLFISNPDLVMRIKLNAP  360
*****

```

```

OPR3 native      LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_6ring      LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_5ring      LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_4ring_var1 LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_4ring_var2 LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_3ring_var1 LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_3ring_var2 LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_3ring_var3 LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
opr3_3ring_var4 LNKYNRKTFTYTDQDPVVGTYDYPFLQGNGSNGPLSRL  396
*****

```

3.2 Protein expression and tests for catalytic activity

All eight designed enzymes were first expressed and purified at the Graz University of Technology. It was reported that the two variants of OPR3_4ring heavily precipitated during purification. As no further analyses were made the reason for this instability remains unknown, but it might be possible that the introduction of an arginine residue (Arg-149) which is positioned in a loop on the surface and pointing towards the center of the protein disturbs the proper folding process.

The remaining six variants were also expressed and purified in the course of this project. The proteins were stable in solution at the concentrations used, though multiple freeze-and-thaw cycles caused slight precipitation.

By the time this thesis was written, four variants of OPR3_3ring were tested at the Graz University of Technology for activity in converting 4-bromobut-2-enal to cyclopropanecarbaldehyde, but all of them failed in these first attempts. Further testing and optimization, as well as testing of the variants OPR3_5ring and OPR3_6ring for conversion of their intended substrates is subjected to future studies.

A very important consideration that could be crucial for successful conversion of the desired

substrates was intentionally left out in all these design attempts: It could not be assured that the binding of NAD(P)H and hence the regeneration of the oxidized flavin was still possible in the altered active sites. The fact that the overall catalytic cycle consists of two independent reactions was a problem that could not be dealt with in the Rosetta design process. Before expressing the designed enzymes, a possibility to obtain qualitative evidence that NAD(P)H can still bind could be by extensive docking studies. Due to lack of time, no such investigations were performed.

3.3 Crystallization

Crystallization experiments were set up with six variants, with two of them leading to well diffracting crystals that could be used for successful data collection and structure determination. Also a structure of OPR3_Y190F containing a ligand could be solved.

3.3.1 Crystal structure of OPR3_5ring

A crystal structure of OPR3_5ring was determined at 2.1 Å. The structure contains two molecules of OPR3_5ring: Chain A with residues number 10 - 283, 299 - 384 and chain B with residues number 10 - 283, 300 - 385. For nine residues at the N-termini, twelve residues at the C-termini and gaps of 15 (chain A) and 16 (chain B) residues no electron density could be interpreted. The structure showed the expected TIM-barrel fold and the sites of the FMN molecules could be easily determined in the difference electron density map. The final R-factor and R_{free} were 0.2017 and 0.2567 respectively. Table 2 shows an overview of the data and refinement parameters.

Both chains were superimposed with the predicted design outputs (Fig. 25). The protein backbone of the loops that set up the active site are only slightly different in comparison to the designed structures. The position of the FMN cofactor and the histidine side chains (His-185 and His-188) are almost equal. The side chains of Leu-255 and Asn-245 show the same conformation as in the design structures, but are slightly offset towards the active site pocket, since the whole loop region where these residues are located is slightly shifted to the inner of the protein. The hydroxyl group of Ser-113 is, other than predicted in the designs, oriented towards the substrate binding site. This might be positive, as the hydrogen bonding geometry with the bromide from the substrate is better. The side chains of Leu-190 and Met-74 are again almost at the same position as predicted by Rosetta. For Met-108 two alternate conformations were observed in both chains. Although both conformations are slightly different than expected, the space occupied by the side chains is still very close to that of the designed structures, and there might be no lack in providing a tight packing and exclusion of solvent. Overall, the active site architecture of the actual crystal structure is quite similar to that expected from the design runs, and the intended changes could be accomplished as planned.

Table 2 – Experimental details for the crystal structure of OPR3.5ring.

Data collection¹	
X-ray source	ESRF-ID29
Wavelength [Å]	0.91908
Temperature	100 K
Space group	$P 2_1$
Cell dimensions	
<i>a, b, c</i> [Å]	50.03, 89.84, 90.24
α, β, γ [°]	90, 98.85, 90
Resolution [Å]	63.29 - 2.10 (2.21 - 2.10)
Total number of reflections	170 295 (25 513)
Unique reflections	44 703 (6 482)
Multiplicity	3.8 (3.9)
Completeness [%]	97.1 (97.2)
$R_{p.i.m.}$	0.066 (0.207)
R_{merge}	0.118 (0.373)
$CC_{1/2}$	0.993 (0.914)
CC^*	0.998 (0.977)
Mean $I/\sigma(I)$	7.4 (3.1)
Refinement	
Resolution [Å]	89.17 - 2.10
R_{work}/R_{free}	0.2017/0.2567
No. of atoms	
Protein	5 665
Cofactor	62
Water	325
Other	2
Mean B-factors [Å ²]	
Protein	28.548
Cofactor	18.490
Water	29.509
Other	26.005
All atoms	28.496
rmsd bond lengths [Å]	0.015
rmsd bond angles [°]	1.652
Ramachandran outliers [%]	0.00

¹Values in parentheses are for the highest resolution shell

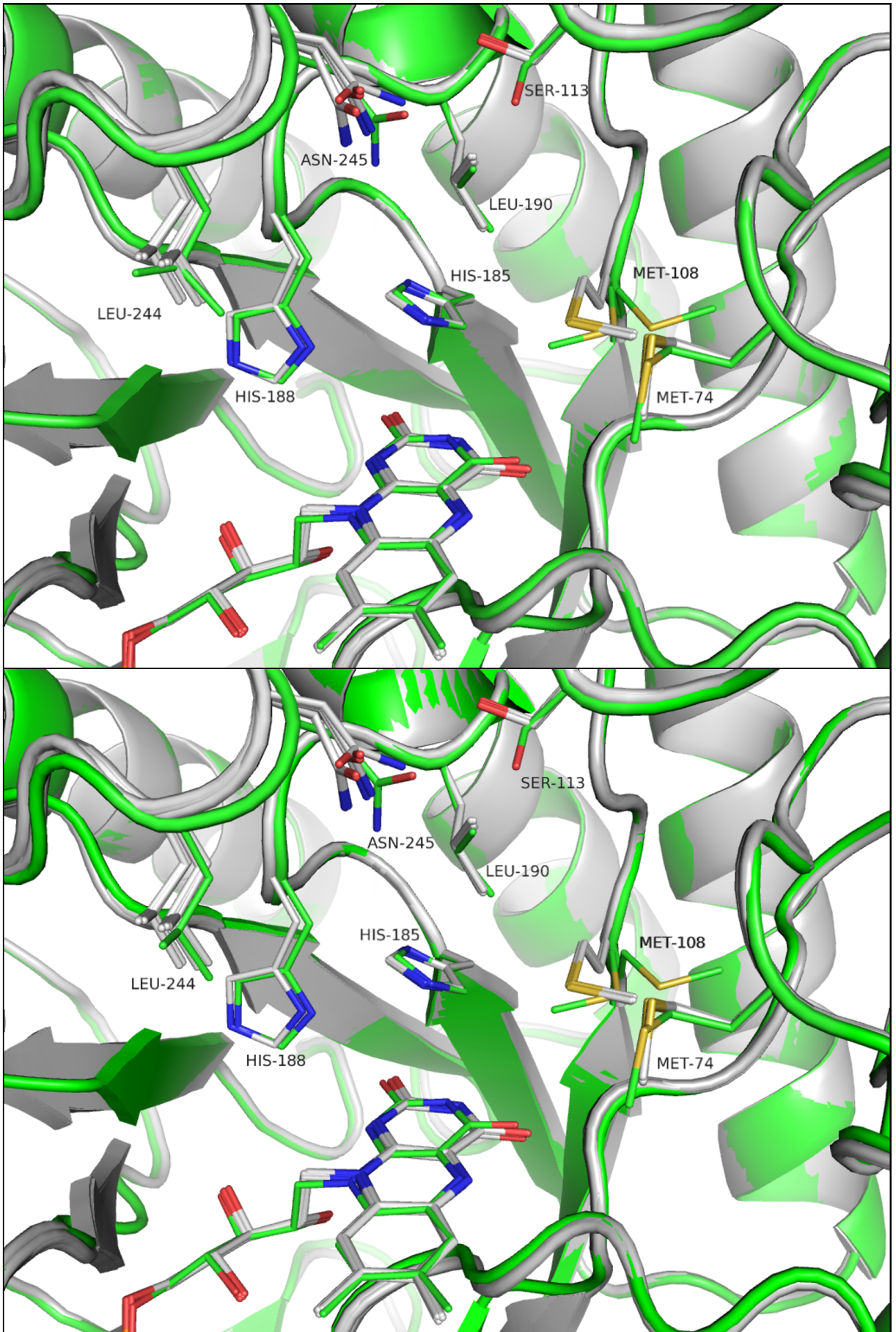


Figure 25 – (Legend on next page)

Figure 25 – (Previous page) Superposition of six design output structures that featured the final sequence in the active site (gray) with the two chains of the determined crystal structure of OPR3_5ring (green). Top: chain A, bottom: chain B. The FMN molecule and the two histidine side chains that coordinate the carbonyl group of the substrate are very close to the expected positions. The loop that contains residue numbers 244 and 245 is slightly offset compared to the designed structures, which brings the respective side chains a little further towards the inside of the active site pocket. Ser-113 is oriented towards the substrate binding site, which might help to interact with the substrate’s bromide leaving group. The two alternate conformations of Met-108 are not at the exact position as in the designed structures, but the requirement for occupation of the particular space with a hydrophobic residue is still fulfilled.

3.3.2 Crystal structure of OPR3_6ring

For this variant a structure at 2.1 Å could be solved. The structure contains two chains of OPR3_6ring. In each chain nine residues at the N-terminus, a gap between Pro-282 and Glu-299 and twelve residues at the C-terminus were not fitted due to a lack of electron density. The FMN cofactors were clearly visible in the difference electron density. The structure was refined to a final R-factor of 0.2459 and R_{free} of 0.2877. Details for data processing and refinement are shown in Table 3.

The two chains were superimposed with the predicted design structures (Fig. 26). The FMN cofactor is still at the same position, but other residues are more or less different than in the designs. The backbone in the loop regions around the active site pocket is slightly distorted in the experimentally determined structure. Leu-244 shows different rotamers in the crystal structure and probably there are alternate conformations, as the electron density in that area was not clearly defined. In contrast, it was obvious from the electron density that the two histidine side chains (His-185 and His-188) are misplaced (see also Fig. 27). The imidazole ring of His-188 is twisted in way that the nitrogen, acting as a hydrogen bond donor in the binding of substrates, points away from the supposed (and native) binding site. Also His-185 is slightly misplaced. The resulting distortion of the substrate binding site might lead to a much lowered affinity of substrate molecules to the enzyme or even a complete loss of activity. Interestingly, the only difference between OPR3_5ring, where the native geometry of the two histidines is still intact, and OPR3_6ring is the exchange of Leu-190 (OPR3_5ring) to Ala-190 (native: Tyr-190). It seems that the smaller alanine side chain causes enough flexibility in the region, that the adjacent His-185 and His-188 are displaced. Other mutated residues are slightly different than in the designed structures. Asn-245 is shifted a little towards the center of the active site pocket, but still in a position that would allow for the substrate to bind. The hydroxyl group of Ser-113 is not oriented towards the active site as desired, but the actual rotamer is in consistency with the majority of the predictions in the designed structures. Met-108 shows the same conformation, but is a bit further outside of the active site. Met-74 has another conformation than expected, but is still in the same place as in the designed structures. In summary, as the two histidine side chains that are positioning the substrate via hydrogen bonding are displaced, and therefore a fundamental requirement for catalysis is not fulfilled,

a successful conversion of the substrate to the desired product is unlikely. The other mutated residues, although being slightly different than in the predicted design structures, are still in positions where they would support the intended catalytic mechanism.

Table 3 – Experimental details for the crystal structure of OPR3_6ring.

Data collection¹	
X-ray source	ESRF-ID29
Wavelength [Å]	0.91908
Temperature	100 K
Space group	$P 2_1$
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> [Å]	49.85, 91.09, 90.27
α , β , γ [°]	90, 98.33, 90
Resolution [Å]	63.77 - 2.10 (2.21 - 2.10)
Total number of reflections	164 843 (24 858)
Unique reflections	44 793 (6 571)
Multiplicity	3.7 (3.8)
Completeness [%]	96.2 (97.1)
$R_{p.i.m.}$	0.091 (0.239)
R_{merge}	0.157 (0.413)
$CC_{1/2}$	0.987 (0.656)
CC^*	0.994 (0.890)
Mean $I/\sigma(I)$	5.6 (3.1)
Refinement	
Resolution [Å]	89.32 - 2.10
R_{work}/R_{free}	0.2459/0.2877
No. of atoms	
Protein	5 607
Cofactor	62
Water	201
Mean B-factors [Å ²]	
Protein	24.577
Cofactor	13.849
Water	22.182
All atoms	24.381
rmsd bond lengths [Å]	0.013
rmsd bond angles [°]	1.457
Ramachandran outliers [%]	0.00

¹Values in parentheses are for the highest resolution shell

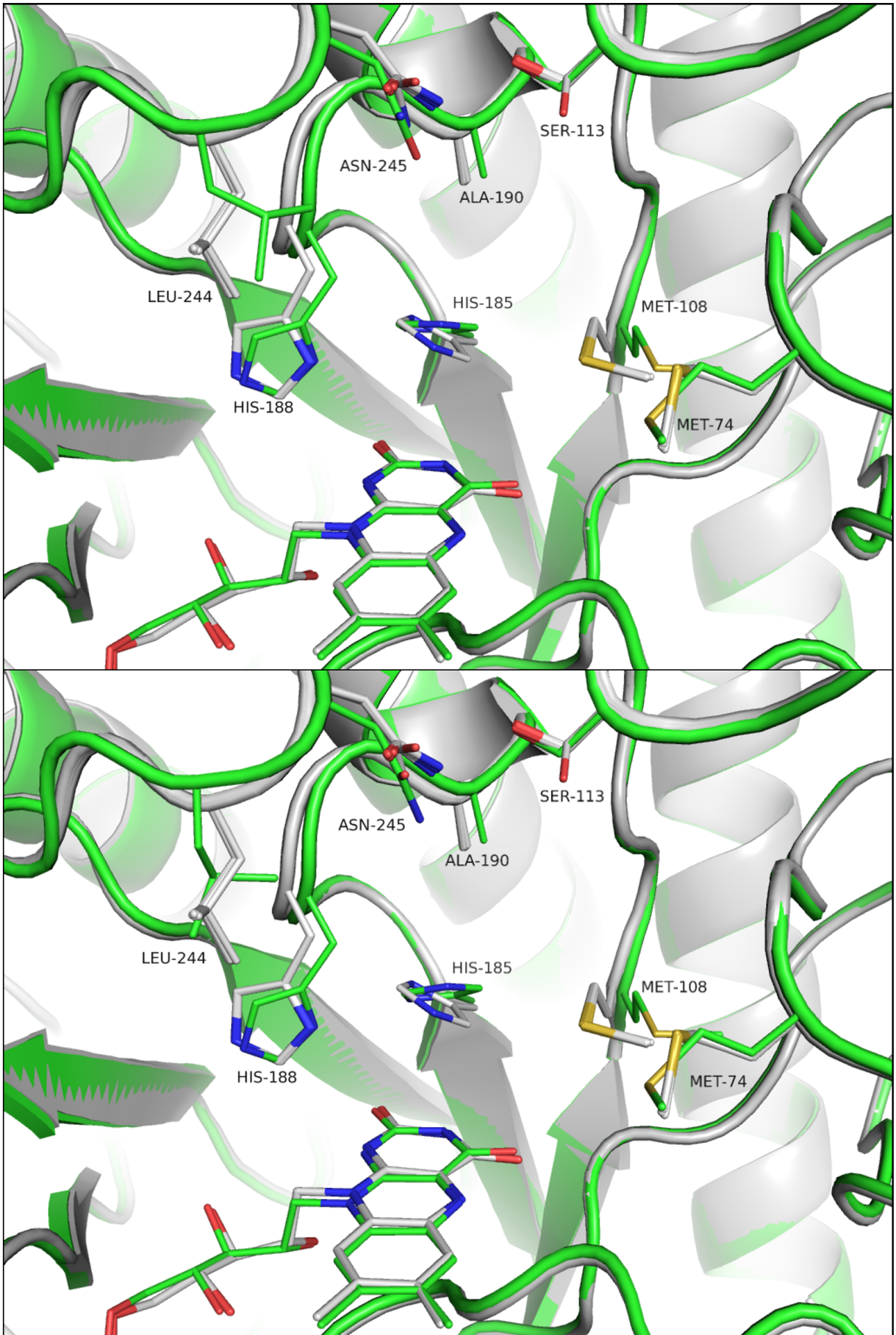


Figure 26 – (Legend on next page)

Figure 26 – (Previous page) Superposition of the predicted structures (grey) with the actual crystal structure (green) of OPR3_6ring. Top: chain A, bottom: chain B. The two residues that are crucial for binding the substrate in a position that allows for hydride transfer are displaced in comparison to the designed structures (see also Fig. 27). Other residues are also in slightly different positions, but their role in catalysis is not so fundamental and therefore their displacements would be still acceptable.

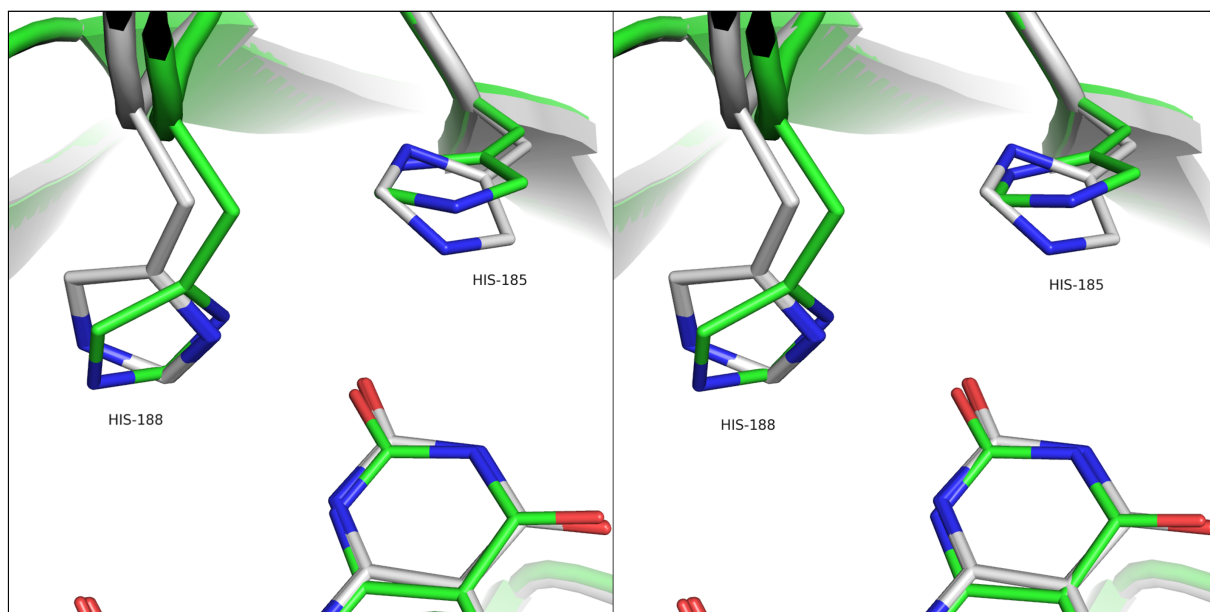


Figure 27 – Superposition of designed structures (gray) with the experimentally determined crystal structure (green). Left: chain A, right: chain B. The backbone in the region of the two substrate binding histidines is slightly distorted in the actual crystal structure, resulting in a displacement of the imidazole rings.

3.3.3 Crystal structure of OPR3_Y190F

A dataset to 2.00 angstroms could be collected from the crystal that was soaked with (*E*)-4-bromobut-2-enal for 1 min. It appeared that the crystals that were soaked for longer timespans were either of bad quality or destroyed by the soaking procedure. The asymmetric unit contained two molecules of OPR3_Y190F. In both chains nine residues at the N-termini, a gap between residues number 283 - 296, and twelve residues at the C-termini could not be fitted, as the electron density was too poorly defined. The position of the FMN cofactors and the mutated side chain (Tyr-190 to Phe) was clearly visible in the difference electron density. The structure was refined to an R-value and R_{free} of 0.1876 and 0.2253 respectively, until fitting of a potential ligand molecule was tried. Table 4 shows a summary of the data processing and refinement parameters.

Table 4 – Experimental details for the crystal structure of OPR3_Y190F.

Data collection¹	
X-ray source	ESRF-BM14
Wavelength [Å]	0.95373
Temperature	100 K
Space group	$P 2_1 2_1 2_1$
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> [Å]	89.27, 91.85, 97.44
α , β , γ [°]	90, 90, 90
Resolution [Å]	48.72 - 2.00 (2.05 - 2.00)
Total number of reflections	401 434 (26 893)
Unique reflections	54 700 (3 879)
Multiplicity	7.3 (6.9)
Completeness [%]	99.8 (97.7)
$R_{p.i.m.}$	0.099 (0.279)
R_{merge}	0.250 (0.687)
$CC_{1/2}$	0.989 (0.885)
CC^*	0.997 (0.969)
Mean $I/\sigma(I)$	7.5 (3.1)
Refinement	
Resolution [Å]	66.84 - 2.00
R_{work}/R_{free}	0.1876/0.2253
No. of atoms	
Protein	5 820
Cofactor	62
Water	413
Mean B-factors [Å ²]	
Protein	7.037
Cofactor	2.005
Water	13.046
All atoms	7.381
rmsd bond lengths [Å]	0.019
rmsd bond angles [°]	1.935
Ramachandran outliers [%]	0.00

¹Values in parentheses are for the highest resolution shell

The dataset for this structure was autoprocessed at the ESRF Grenoble using XDS. When the data were again processed manually using Mosflm with default settings, the best solution in indexing that was found was in space group $P 2_1$, and there was no solution for the cell in $P 2_1 2_1 2_1$ (see Table 5). The cell constants in $P 2_1$ were much like the same as in other crystals even from the same crystallization condition and other known crystals of OPR3. Merging and scaling this lower symmetry data led to higher R_{merge} values and scaling factors than with the data in $P 2_1 2_1 2_1$. In both datasets twinning was not detected and for each of them it was possible to solve the structure by molecular replacement. Both solutions looked plausible, but with better refinement results in the higher symmetry dataset ($P 2_1$: $R = 0.2511$, $R_{free} = 0.3078$; $P 2_1 2_1 2_1$: $R = 0.1876$, $R_{free} = 0.2253$). The question that arose was that how a solution in a supposedly wrong space group ($P 2_1 2_1 2_1$) could provide a better refinement result than the 'correct' one? An explanation for this was found when the original dataset was again processed using Mosflm, but with a lowered threshold for finding spots during indexing. The questionable cell in $P 2_1 2_1 2_1$ could be found and processing the data and solving the structure led to the same results as with the autoprocessed data. It seems that there were very weak reflections that were not included previously with using the default settings. This would be in agreement with the observation that two cell lengths were very similar and one was approximately double the length in $P 2_1 2_1 2_1$ compared to $P 2_1$. If every second reflection along one axis was weak and therefore excluded, this would lead to only a half of the cell length. For any other datasets of OPR3 crystals that were collected in the course of this work, only the $P 2_1$ symmetry with the usual cell constants was found. Although this is very infrequently observed, it is possible that the crystal that was used in this experiment did actually grow in a different symmetry than other crystals with the same crystallization condition or that the soaking with substrate led to a rearrangement in the crystal packing. However, the fact that it was possible to solve the structure in different space groups with plausible results remains a mystery that will be further investigated.

Table 5 – Comparison of the cell constants for the two observed space groups (lengths in angstroms; angles in degrees). Both of them gave solutions that looked plausible, but the one in $P 2_1 2_1 2_1$ led to better refinement results.

$P 2_1 2_1 2_1$	$P 2_1$
a = 89.27 $\alpha = 90$	a = 49.39 $\alpha = 90$
b = 91.85 $\beta = 90$	b = 91.93 $\beta = 98.92$
c = 97.44 $\gamma = 90$	c = 89.37 $\gamma = 90$

Fitting a putative substrate molecule In both chains there was positive electron density in the difference map above the isoalloxazine ring of FMN (Fig. 28). As the density could not be clearly interpreted at first, possible compounds that were known to be present in reasonable amounts in the crystallization batches were tried to fit the density. When none of them appeared to be suitable, the substrate compound used for soaking the crystals ((*E*)-4-bromobut-2-enal) was tried. The highest positive difference electron density was observed at the binding site that is coordinated by two histidines, which could have been evidence that the substrate binds in a 'reverse' orientation in the active site, i.e. that the bromine interacts with the two histidine side chains instead of the carbonyl group. If this was true, the fundamental assumption for the design process - that the substrate molecules bind with the carbonyl moiety at the site between the two histidines - would have been proven wrong. The substrate molecule was fitted into the density with the bromine at the aforementioned site to test this hypothesis. After refinement a highly negative difference electron density was observed around the bromine atom, which means that this binding mode is with high probability not present (Fig. 29). Still with the occupancy of the substrate molecule lowered to 0.5, negative electron density appeared (not shown).

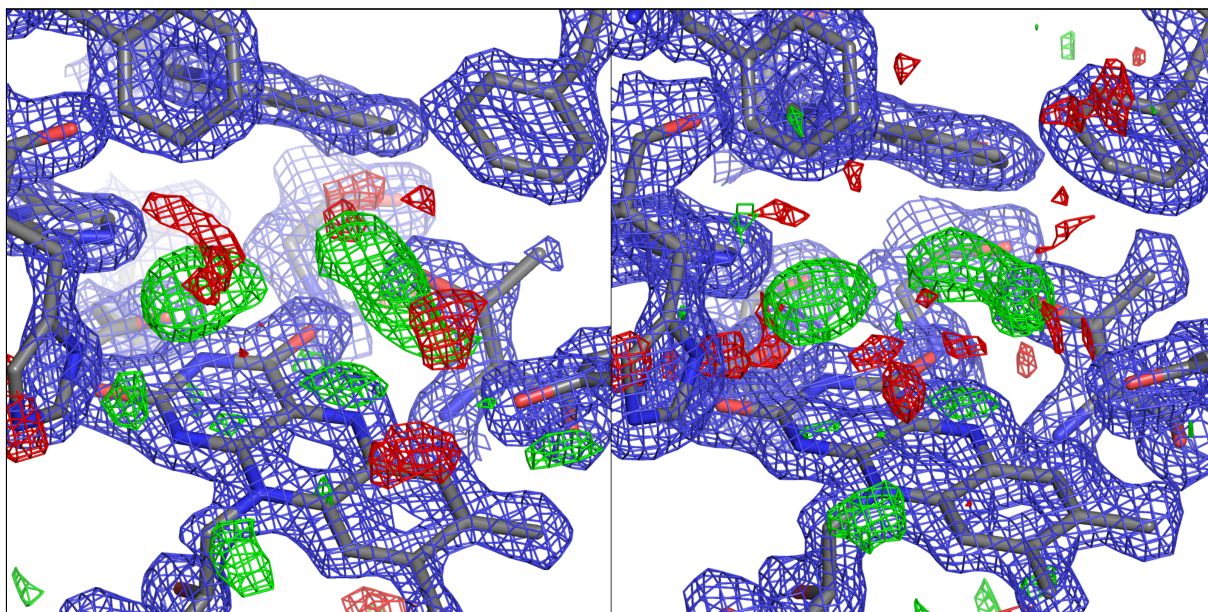


Figure 28 – Active site of OPR3_Y190F. Left: chain A, right: chain B. Two positive difference electron density peaks were located in the active site pocket of both chains. Electron density maps: 2Fo-Fc at 1.5 sigma (blue), Fo-Fc at 3.0 sigma (green = positive, red = negative).

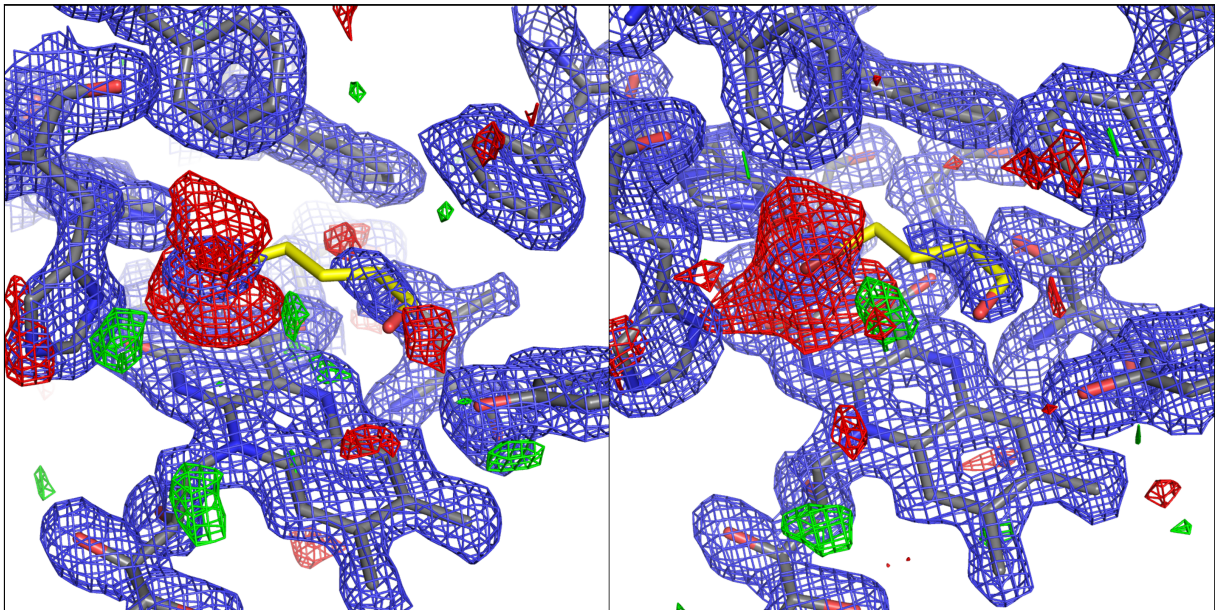


Figure 29 – Active site of OPR3_Y190F. Left: chain A, right: chain B. Refinement with the substrate molecule (yellow) oriented with the bromine atom at the binding site between the two histidine side chains resulted in a highly negative difference density. Electron density maps: 2Fo-Fc at 1.5 sigma (blue), Fo-Fc at 3.0 sigma (green = positive, red = negative).

It was then tried to fit the substrate molecule in the expected orientation into the electron density (i.e. the carbonyl group is coordinated by the two histidine side chains). After refinement, there was still negative electron density around the bromine atom, but to a lesser extent than in the 'reverse' orientation (Fig. 30). In a next step the model was refined with a lowered occupancy of the substrate molecules. In chain A a single molecule with an occupancy of 0.5 was fitted, whereas in chain B two molecules in different conformations that were supported by a slightly positive difference electron density were fitted with an occupancy of 0.5 respectively (Fig. 31). In chain A there was still a slightly positive difference density around the carbonyl moiety. Also next to C_{α} there was positive density, which was the reason to try a second conformation in chain B, where this density was even more distinct. In both chains there was only little negative electron density around the bromine atoms.

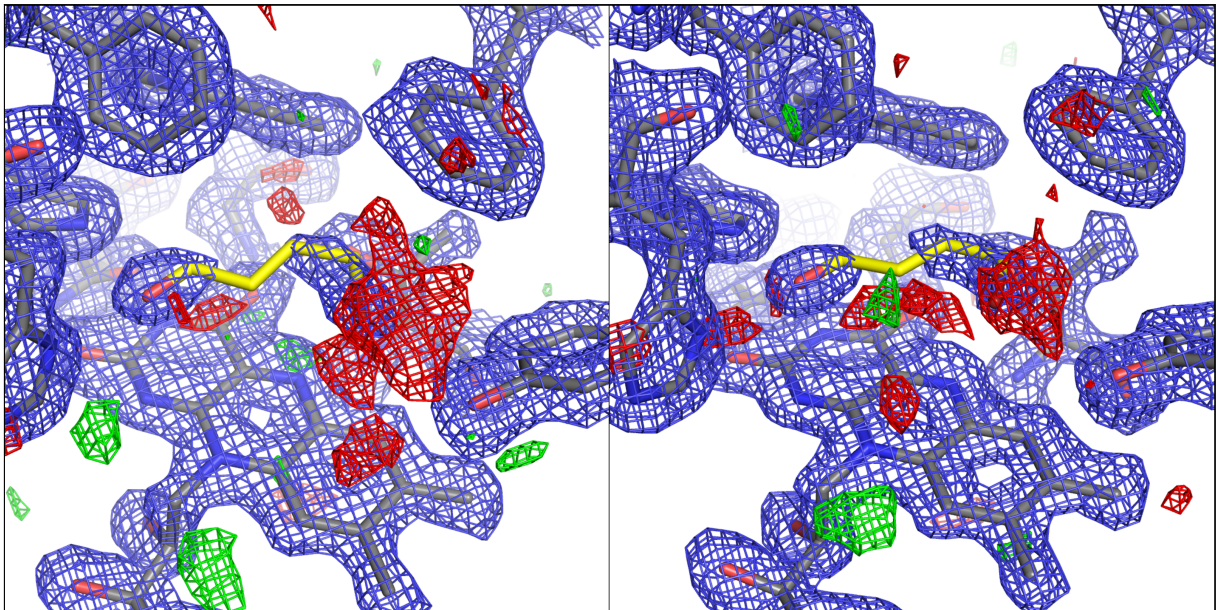


Figure 30 – Active site of OPR3_Y190F. Left: chain A, right: chain B. The soaked substrate compound (yellow) was modeled into the electron density in the expected orientation that would allow hydrogen bonding between the carbonyl group and the histidine side chains. There is still negative difference density around the bromine atom. Electron density maps: 2Fo-Fc at 1.5 sigma (blue), Fo-Fc at 3.0 sigma (green = positive, red = negative).

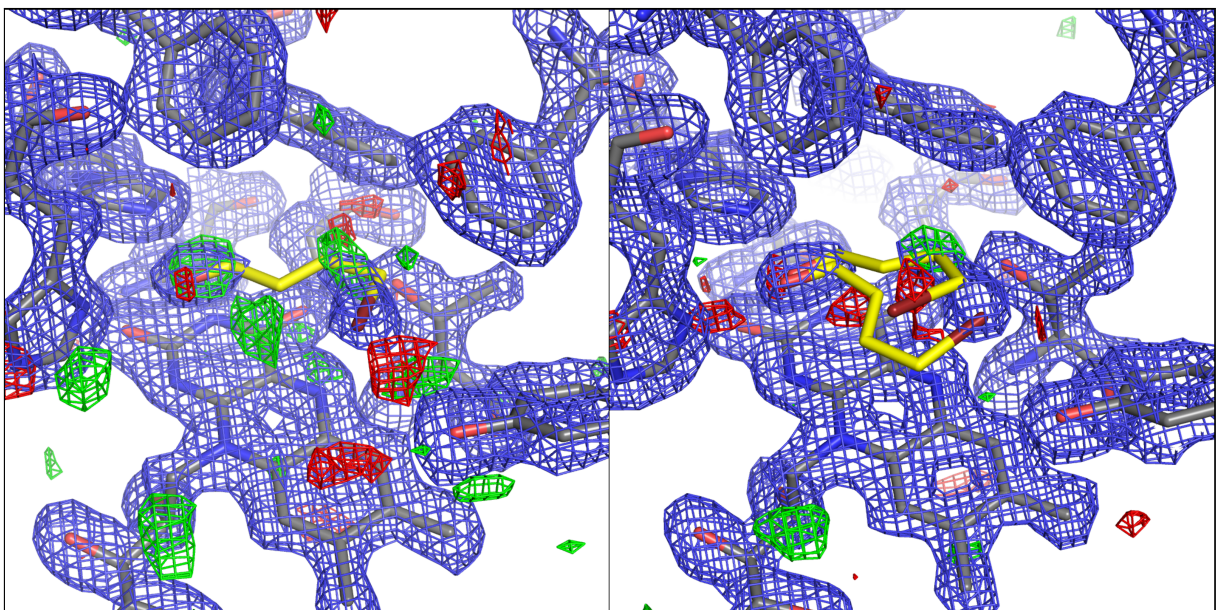


Figure 31 – Active site of OPR3_Y190F. Left: chain A, right: chain B. In chain A a single substrate molecule was fitted with an occupancy of 0.5. At the carbonyl group there is still slightly positive difference density. Also next to C_α (in the front) there is a small blob of positive density. This is why in chain B it was tried to model two molecules with an occupancy of 0.5. Electron density maps: 2Fo-Fc at 1.5 sigma (blue), Fo-Fc at 3.0 sigma (green = positive, red = negative).

In summary, there is not enough evidence to derive a particular binding mode of the substrate from the crystal structure, except that the substrate probably does not bind in the 'reverse' orientation in which the bromine is coordinated by the two histidine residues.

The electron density was too poorly defined to interpret it with enough certainty. This could be due to a too low occupation of the binding sites with substrate molecules in the crystal, the presence of a few slightly different binding modes or a partial decomposition of the substrate.

4 Conclusions

In the process that led to the final designed variants, the Rosetta applications were used as a tool to predict mutated protein structures. The possibilities with Rosetta would be many more, as the real potential of the Rosetta enzyme design package is in the de-novo-design of new catalytic sites. With the match application Rosetta provides a tool for the screening of several different protein scaffolds for ones that might support a strictly defined theozyme geometry. Once a fitting scaffold is found and the theozyme is introduced into the structure, the enzyme design application is used to alter the sequence in a way that the new catalytic site is supported by favorable interactions and therefore kept in the exact theozyme geometry.

As in this project only one target enzyme (OPR3) was used as a scaffold structure, the match application became almost unnecessary or was even a limitation, as the theozyme geometry constraints had to be loosened in order to get output structures. A possibility to still use the match application could have been by matching via the secondary match algorithm, which would accept any theozyme geometry within certain boundaries. In that case more outputs would have been generated, but still at the cost of accurate geometry. In the end the structures that were predicted with Rosetta were close to the real structures, but also little differences can have a big impact on the catalytic performance, as it can be supposed for the design of OPR3.6ring. Screening of several slightly different designs for the same activity might be a possibility to enhance the hit ratio for successful designs. The two crystal structures that were determined and compared with the theoretical designs give far too little evidence to provide a general statement on how the Rosetta design application performs in enzyme design. A limitation in this project was the presence of a cofactor that must not be displaced. Due to this, quite rigid instructions on the sequence positions that could be subjected to mutations had to be set up. In this way it could be possible that 'side' mutations, that are required to support the structure and the positioning of the catalytic side chains, were suppressed. Apart from that, it was also observed that Rosetta sometimes tends to 'overoptimize' a structure, i.e. when the design shells were set big enough, that single residues far from the active site were mutated. Nevertheless, considering the complexity of modeling such artificial protein structures, the Rosetta suite provides very powerful tools for approaching such an intricate task as enzyme design.

References

- [1] Kohli, R. M.; Massey, V. The Oxidative Half-reaction of Old Yellow Enzyme. *J. Biol. Chem.*, **273**(49):32763–32770, (1998).
- [2] Strassner, J.; Schaller, F.; Frick, U. B.; Howe, G. A.; Weiler, E. W.; Amrhein, N.; Macheroux, P.; Schaller, A. Characterization and cDNA-microarray expression analysis of 12-oxophytodienoate reductases reveals differential roles for octadecanoid biosynthesis in the local versus the systemic wound response. *The Plant Journal*, **32**:585–601, (2002).
- [3] Schaller, F.; Biesgen, C.; Müssig, C.; Altmann, T.; Weiler, E. W. 12-Oxophytodienoate reductase 3 (OPR3) is the isoenzyme involved in jasmonate biosynthesis. *Planta*, **210**:979–984, (2000).
- [4] Weiler, E. W. Octadecanoid-Mediated Signal Transduction in Higher Plants (review). *Naturwissenschaften*, **84**:340–349, (1997).
- [5] Vick, B. A.; Zimmerman, D. C. The biosynthesis of jasmonic acid: A physiological role for plant lipoxygenase. *Biochemical and Biophysical Research Communications*, **111**(2):470–477, (1983).
- [6] Hall, M.; Stueckler, C.; Ehammer, H.; Pointner, E.; Oberdorfer, G.; Gruber, K.; Hauer, B.; Stuermer, R.; Kroutil, W.; Macheroux, P.; Faber, K. Asymmetric Bioreduction of C=C Bonds using Enoate Reductases OPR1, OPR3 and YqjM: Enzyme-Based Stereocontrol. *Adv. Synth. Catal.*, **350**:411–418, (2008).
- [7] Gaussian 09, Revision A.02; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.
- [8] GaussView, Version 5.0.8, Dennington, R; Keith, T; Millam, J.; Semicem Inc., Shawnee Mission, KS, 2009.
- [9] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. August 2010.
- [10] Schrödinger, LLC. Maestro, version 9.4. New York, NY, 2013.

- [11] Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. New algorithms and an in silico benchmark for computational enzyme design. *Protein Science*, **15**(12):2785–2794, (2006).
- [12] Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS ONE*, **6**(5):e19230, (2011).
- [13] Hetmann, M. Bachelor’s thesis: Generation of Old Yellow Enzyme variants for the catalysis of reductive C-C coupling reactions, Graz University of Technology, 2015.
- [14] Kabsch, W. XDS. *Acta Crystallographica Section D - Biological Crystallography*, **66**:125–132, (2010).
- [15] Leslie, A.G.W.; Powell, H.R. Evolving Methods for Macromolecular Crystallography, 245, 41-51 (2007), ISBN 978-1-4020-6314-5.
- [16] Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D - Biological Crystallography*, **66**:282–292, (2011).
- [17] Evans, P. R. Scaling and assessment of data quality. *Acta Crystallographica Section D - Biological Crystallography*, **62**:72–82, (2006).
- [18] Vagin, A.; Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, **30**:1022–1025, (1997).
- [19] Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallographica Section D - Biological Crystallography*, **66**:486–501, (2010).
- [20] Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D - Biological Crystallography*, **53**(3):240–255, (1997).
- [21] Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D - Biological Crystallography*, **67**:355–367, (2011).
- [22] Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*, **1**:213–220, (2014).
- [23] Moriarty, N.W.; Grosse-Kunstleve, R.W.; Adams, P.D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallographica Section D - Biological Crystallography*, **65**:1074–1080, (2009).