Development and Screening of promiscuous aldolases

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Dissertation

submitted in fulfilment of the requirements for the
degree of *Doctor rerum naturalium*
(Dr. rer. nat.)

at the Department of Chemistry
of the Technischen Universität Darmstadt

Doctoral thesis
by Tobias Glesner

First assessor: Prof. Dr. Wolf-Dieter Fessner
Second assessor: Prof. Dr. Katja Schmitz

Darmstadt 2024

Glesner, Tobias: Entwicklung und Screening von promiskuitiven Aldolasen

Darmstadt, Technische Universität Darmstadt

Jahr der Veröffentlichung der Dissertation auf TUPrints: 2024

URN: urn:nbn:de:tuda-tuprints-284053

Tag der Einreichung: 15. Januar 2024

Tag der mündlichen Prüfung: 06. Mai 2024

Die vorliegende Dissertation wurde unter der Leitung von Prof. Dr. Wolf-Dieter Fessner am Clemen-Schöpf-Institut für organische Chemie und Biochemie der Technischen Universität Darmstadt im Zeitraum von August 2018 bis Dezember 2023 durchgeführt. Die praktischen Arbeiten im Labor waren am 31.12.2022 abgeschlossen.

Erklärung laut Promotionsordnung

§ 8 Abs. 1 lit c PromO
Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt und für die Durchführung des Promotionsverfahrens vorliegt.

§ 8 Abs. 1 lit d PromO
Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde und zu keinem früheren Zeitpunkt an einer in- oder ausländischen Hochschule eingereicht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§ 9 Abs. 1 PromO
Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§ 9 Abs. 2 PromO
Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, den 15. Januar 2024   _____

                                                                      Tobias Glesner

1 Inhaltsverzeichnis

Abbreviations

AIM	Auto induction media
APBS	Adaptive Poisson-Boltzmann Solver
Bl	Blind sample
CalB	Candida Antarctica Lipase B
CFE	Cell free extract
CD	Circular dichroism
D-ManNAc	N-acetyl-D-mannosamine
DHAP	Dihydroxyacetonephosphate
DHA	Dihydroxyacetone
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
DR5P	2-Deoxyribose-5-phosphate
DSF	Differential scanning fluorometry
DSC	Differential scanning calorimetry
*Ec*DERA	2-Deoxyribose-5-phosphate aldolase from *E. coli*
EcFSA	Fructose-6-phosphate aldolase from *E. coli*
EDTA	Ethylenediaminetetraacetic acid
EP	Empty plasmid BL21 background protein mixture
FruA	Fructose-biphosphate aldolase
FSA	Fructose-6-phosphate aldolase
GalA	2-Keto-3-deoxy-6-phospho-D-galactonate aldolase
GA3P	D-Glyceraldehyde-3-phosphate
GC	Gas chromatography
GndCl	Guanidinium chloride
GlcA	2-Keto-3-deoxy-6-phospho-D-gluconate aldolase
HA	Hydroxyacetone
HEWL	Hen egg white lysozyme
HPLC	High performance liquid chromatography
HTS	High through-put screenings
KdoA	2-Keto-3-deoxy-d-manno-octosonate aldolase
KDPGal	2-Keto-3-deoxy-6-phospho-D-galactonate
KDPGlc	2-Keto-3-deoxy-6-phospho-D-gluconate
KLD	Kinase, Ligase and DpnI enzymes
MTP	Multi titer plate
MS	Mass spectrometry
NeuAc	N-Acetylneuraminic acid lyase
NADH	Nicotinamide adenine dinucleotide
NMR	Nuclear Magnetic Resonance
NTA	Nitrilotriacetic acid
Pdb	Protein database

PCR	Polymerase chain reaction
PLP	Pyridoxal phosphate
Ppm	Part per million
PyrA	Pyruvate dependent aldolases
RP-18	Reversed phase with C18 side chains
Rpm	Rounds per minute
ScADH	Alcohol dehydrogenase from Saccharomyces cerevisiae
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SDM	Site directed mutagenesis
SEC	Size exclusion chromatography
SHMT	Serine hydroxymethyltransferase
TBE	Tris borate EDTA buffer
TEA	Triethanolamine
ThrA	L-Threonine aldolase
TLC	Thin layer chromatography
TRIS	Tris(hydroxymethyl)aminomethane
T_m	Protein melting point
UV	Ultraviolet

For amino acids the standard abbreviations were utilized

2   Introduction

The application of microbes and enzymes in synthetic chemistry describes the concept of biocatalysis[1-2]. The roots of applied biocatalysis dates back over 10000 years to ancient Mesopotamia, China and Japan. Biocatalysts have been used for fermentation processes or food preservation to produce bread, cheese, vinegar, alcoholic drinks and many more[3-4]. The biocatalysts were obtained from nature and were used as such without further modification. One example remaining till today is the classical baker yeast *Saccharomyces cerevisiae*, which was intensively studied for its synthetic purpose[5-9].

In the beginning of industrial biocatalysis, discovered enzymes have been used for different processes due to their excellent performance, reflecting the first wave of biocatalysts[1]. Famous examples are proteases in laundry detergents [10], glucose isomerase to yield the sweeter fructose[11] or penicillin G acylase for the production of antibiotics[12].

Nowadays, the chemical and pharmaceutical industries show a growing interest in the implementation of sustainable biocatalytic processes [13-18]. Water as solvent, reactions at room temperature, natural feed stocks for production of the biocatalysts, excellent reactivity and selectivity are just a few advantages compared to classic organic chemistry[19]. With increasing biochemical methods for enzyme modification and optimization, industrial demands can be met. In the beginning of synthetic biochemistry mostly ready-to-use enzymes were utilized [20]. Novel techniques like directed evolution, rational design, gene sequencing and synthesis enabled scientists to optimize enzymes for academic and industrial scale chemistry. The implemented modifications during the second wave of biocatalysts were structure-based modifications with minimal screening effort. Noteworthy examples are lipase-catalyzed resolution of chiral precursors for synthesis of diltiazem[21] or hydroxynitrile-lyase-catalyzed synthesis of intermediates for herbicides[22].

With the pioneering work of Pim Stemmer and Frances Arnold in the 90's on directed evolution another barrier was surpassed[23-26]. Now being able to implement random mutations with a specific rate enabled scientists to modify biocatalysts dependent on the required process properties rather than designing new processes based on the limitations of known biocatalysts. Multiple rounds of directed evolution produce enormously large libraries with thousands to millions of candidates to screen for novel or increased activity and stability. During that third wave of biocatalysts the focus lay on the development of novel screening methods to speed up the process of screening such enormously large libraries.

Today the transition into a fourth wave of biocatalysis has started[27]. With increasing number and through-put of screening technologies, it is possible to screen hundreds of thousands of candidates in an appropriate time scale enabling the screening of metagenomic enzyme sources. The diversity that nature put forth most likely will lift biocatalysis to the next level.

2.1 Aldolase Family

Aldolases are a specific group of lyases, which are able to catalyze the reversible formation of C-C bonds by aldol addition of one acceptor (electrophile) and one donor (nucleophile) molecule while creating up to two new stereocenters with high chiral induction[28].
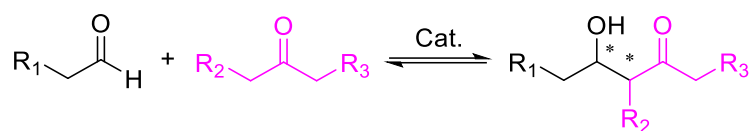
Scheme 1 Aldol addition of an electrophilic aldehyde as acceptor with a nucleophilic ketone donor.

While the acceptor molecule is an aldehyde the donor can be an aldehyde or ketone. In nature the acceptor species is usually activated by phosphorylation. When switching to non-phosphorylated acceptors the activity of the aldolase is usually significantly lower compared to the usage of phosphorylated acceptors. A well-studied example is 2-deoxyribose-5-phosphate aldolase from *Escherichia coli* (*Ec*DERA) having a relative activity orders of magnitude lower with D-glyceraldehyde compared to D-glyceraldehyde-3-phosphate (**3**) (GA3P)[29].

Aldol reactions are a corner stone for production of polyfunctional organic molecules like carbohydrates or iminocyclitols the product of the homo aldol condensation of two *n*-butanal (**53**) molecules, which is utilized for PVC softening in ton scale[30]. While aldolases work in water as solvent, have no need for elevated temperature and being highly stereo- and regioselective, traditional aldol chemistry utilizes organic solvents, elevated temperature and presenting lower stereo control compared to enzymatic results. Therefore, enzymatic aldol reactions promise an alternative tool for organic chemists for production of complex, enantiopure compounds for fine chemical production.

2.1.1 Classification of aldolases based on mechanism

The classification of aldolases is primarily based on the mechanism of the aldol reaction and the utilized donor species[31-32]. In Figure 1 the three possible reaction intermediates of aldolases are presented[33-35].

Class1 aldolase enamine intermediate   **Class2 aldolase metal-enolate intermediate**   **PLP dependent quinoid intermediate**
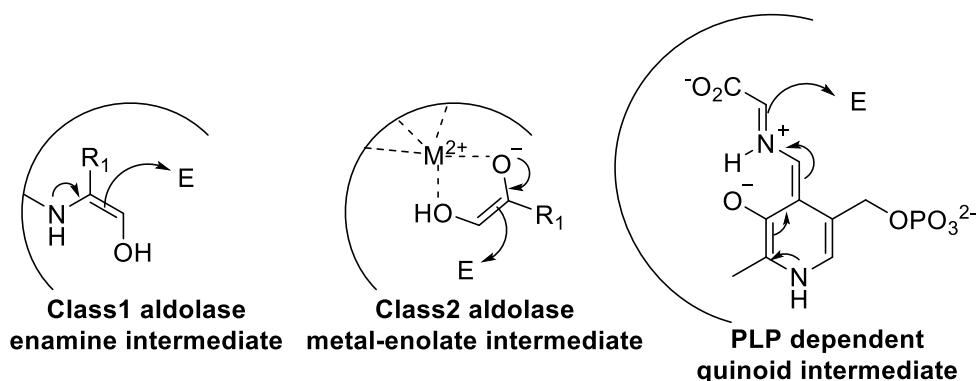
Figure 1 Characteristic reaction intermediates for Class 1, Class 2 and PLP dependent aldolases.

For class 1 aldolases an enaminic intermediate [33] is formed while for class 2 aldolases a metal-enolate species [34] was identified as intermediate. The last class is the pyridoxal phosphate (PLP) dependent aldolases forming a quinoid intermediate [35] with the cofactor PLP. Mechanistically class 1 aldolases bind the donor via Schiff base with an active lysin residue with its ε-amino group to form an enamine which attacks the aldol acceptor as a nucleophile. During that step up to two new stereocenters are formed with high chiral induction. The aldol product gets released after hydrolysis of the carbonyl function [33]. Class 1 aldolases were found in all groups of living organisms from prokaryotes to eukaryotes [36].
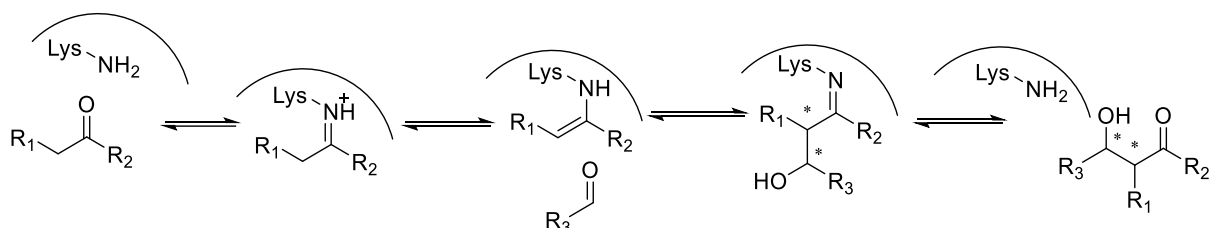
Scheme 2 Relevant reaction intermediates for class 1 aldolases.

A more detailed example is found in section 2.1.6 for the specific aldolase *Ec*DERA.

For class 2 aldolases the mechanism involves a divalent metal ion species which can either be Zn^{2+} for the majority of class 2 aldolases while Fe^{2+} or Co^{2+} are found with rarer occurrence [37]. The metal ion promotes the enolization of the donor substrate via Lewis acid complexation followed by deprotonation to form an enole. The aldol acceptor is attacked by the enamine species. Like for class 1 aldolases the same product with two newly formed stereocenters is produced with high chiral induction[38].
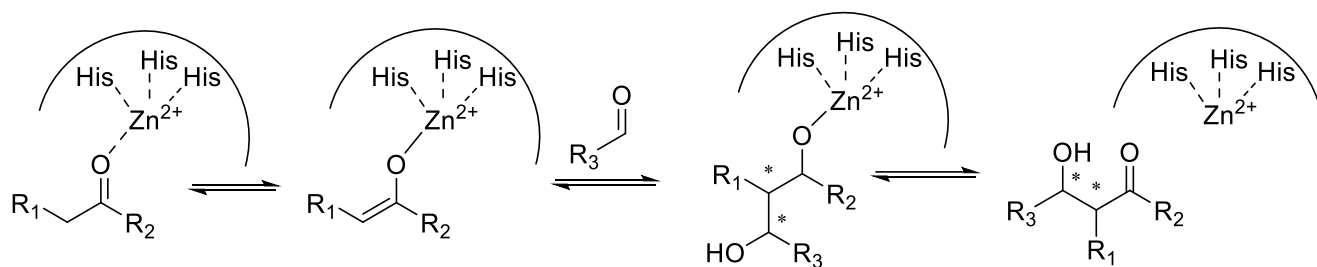
Scheme 3 Relevant reaction intermediates for class 2 aldolases.

After realizing that classification based solely on the catalytic mechanism represents only a small fraction of the complexity of aldolases, a more advanced classification was based on the utilized donor species of the aldolases in the corresponding natural reaction[32, 38-41]. Up to date five different aldolase families have been identified: Dihydroxyacetonephosphate-(DHAP), acetaldehyde-, glycine-, pyruvate- and hydroxyacetone (**32**) dependent aldolases. On that topic multiple excellent reviews are published[38-41].

2.1.2 Dihydroxyacetonephosphate (DHAP) dependent aldolases

The first extensively studied aldolase family was the DHAP dependent aldolases family. One of the most prominent and well-studied examples of aldolases is fructose-biphosphate aldolase FruA (EC 4.1.2.13). In nature it takes part in glycolysis and is responsible for the cleavage of fructose-1,6-bisphosphate into DHAP and GA3P (**3**)[37].
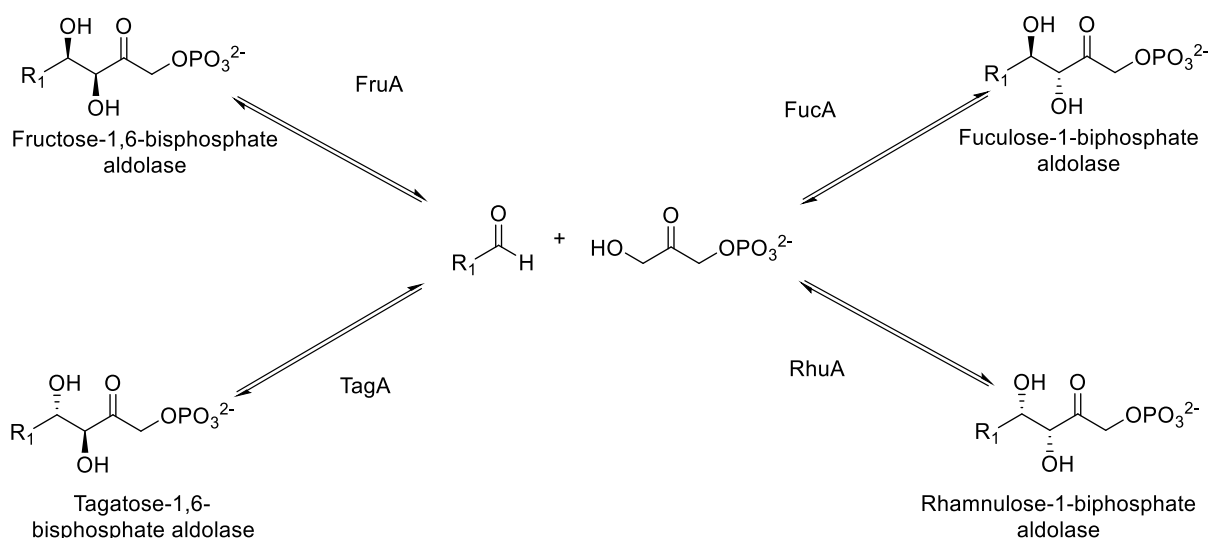
Scheme 4 Stereocomplementary set of DHAP aldolases: Fructose-1,6-biphosphate aldolase, Fuculose-1-phosphate aldolase, Tagatose-1,6-bisphosphate aldolase and Rhamnulose-1-phosphate aldolase for enzymatic aldol reaction of an aldehyde acceptor with dihydroxyacetone phosphate as donor.

Amazingly, nature evolved DHAP dependent aldolases as such that all possible diastereomeric outcomes are available from different enzyme sources. As presented in Scheme 4 fructose-

1,6-biphosphate aldolase produces (3S,4R)-orientation[42], fuculose-1,6-biphosphate aldolase (3R,4R)[43-44], L-rhamnulose-1-phosphate aldolase (3R,4S)[43] and tagatose-1,6-biphosphate aldolase builds mainly (3S,4S), but especially the latter enzyme produces more of a mixture and is therefore synthetically less valuable[28, 45].

DHAP dependent aldolases were utilized for many purposes like labeling experiments, synthesis of iminocyclitols or generation of complex polyhydroxylated monosaccharides for food industry and cosmetics or precursors for further chemical steps for production of fine chemicals[46-54]. In most applications the unphosphorylated product is the desired product. The removal of the phosphate group can be achieved by the use of alkaline or acidic phosphatase[46-47, 55].

In general, the utilization of DHAP dependent aldolases is limited by the disadvantage of the strict utilization of DHAP as donor, which is expensive and unstable under alkaline conditions[19, 56]. Enzymatic routes for the production of DHAP are favorably utilized due to chemical procedures resulting in a mixture DHAP and other phosphorylated substances that are challenging to separate during purification while direct enzymatic routes[43, 57] or routes towards precursors like D-glyceraldehyd-3-phosphate yield pure products[58]. Most common preparation methods start from cheap materials like dihydroxyacetone (DHA), glycerol or glucose. Therefore, the process is only limited by the enzyme of interest[52].

2.1.3 Glycine/alanine dependent aldolases

Glycine and alanine dependent aldolases utilize pyridoxal-5-phosphate (PLP) as cofactor [59-60] for the retro cleavage of threonine forming glycine and acetaldehyde (**17**)[61]. Famous examples are L-threonine aldolase (ThrA; EC 4.1.2.5) and serine hydroxymethyltransferase (SHMT; EC 2.1.2.1). ThrA aldolases are separated in two classes dependent on the cleavage of threonine enantiomers, L- and D-specific threonine aldolases[62-63]. Threonine aldolases were found in plants, vertebrates, bacteria and fungi[64-66].

L-ThrA can be utilized for the synthesis of β-hydroxy-α,ω-diamino acids, which are valuable precursors for the production of statin derivatives, protease inhibitors, antivirals, antibiotics and peptide mimetics[66-68].
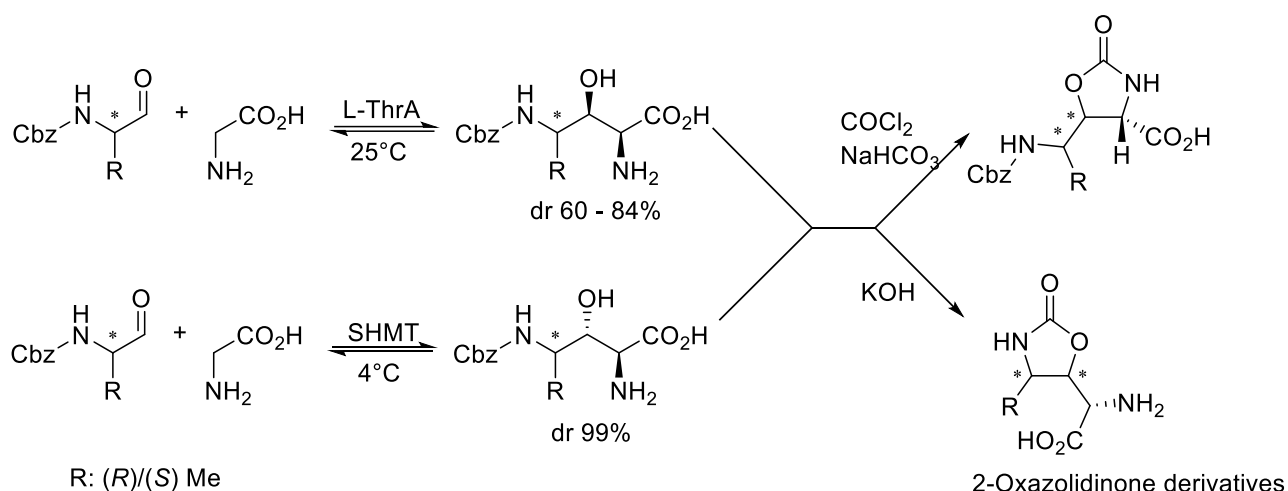
Scheme 5 Synthetic enzymatic route for production of 2-oxazolidinone derivatives by utilizing L-threonine aldolase ThrA or serine hydroxymethyltransferase.

The spectrum of acceptors for L-ThrA is considerably high allowing to convert many different aliphatic and aromatic aldehydes with high chiral induction at the α-carbon while achieving low chiral induction for the β-carbon [69-71].

One industrial relevant process utilizing L-ThrA from *Streptomyces avermitilis* MA-4680 as whole cells, is the synthesis of L-*threo*-3,4-dihydroxyphenylserine (DOPS) which is used for the treatment of Parkinson's disease. The process is designed as a whole-cell high-density bioreactor yielding productivities of 8 gl^{-1} in the aldol addition of glycine and 3,4-dihydroxybenzaldehyde to form L-*threo*-3,4-dihydroxyphenylserine under optimized conditions[72].

2.1.4 Pyruvate dependent aldolases (PyrA)

Most synthetically utilized PyrA are class 1 aldolases, while class 2 PyrA exist, both catalyzing the reversible aldol addition of pyruvate (**1**) to different aldehydes forming γ-hydroxy-α-oxoacids[42, 73-74]. PyrA is involved in different biological functions and nature presents a great diversity of PyrA variants[37]. This diversity enables the utilization of a variety of electrophiles combined with pyruvate (**1**) as donor[42, 75-76].
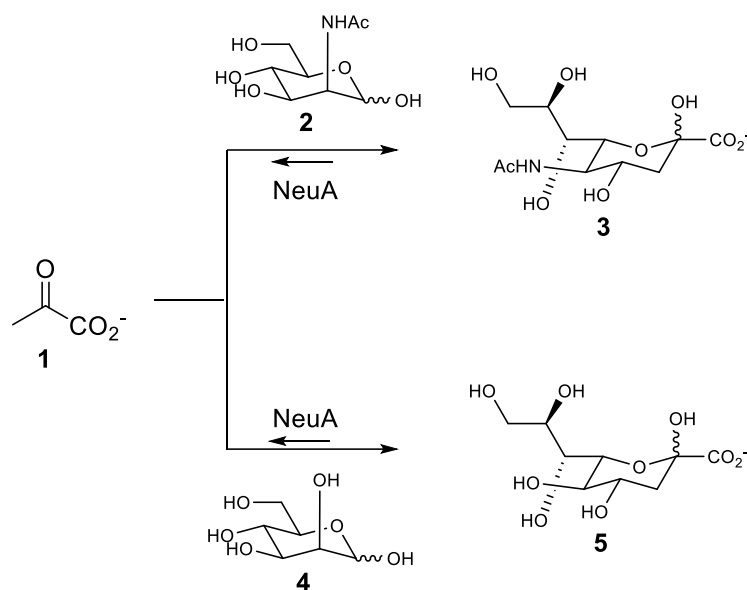
Scheme 6 Synthesis of sialic acid (**3**) (top) from *N*-acetyl-D-mannosamine (**2**) and deoxy-D-glycero-D-galacto-2-nonulosonic acid (**5**) (KDN bottom) from D-mannose (**4**) catalyzed by NeuA.

A famous example of the PyrA family is *N*-acetylneuraminic acid lyase or sialic acid aldolase (NeuAc; EC 4.1.3.3) which was utilized for the synthesis of sialic acids and corresponding analogs[75-78]. NeuA has the ability to utilize pyruvate (**1**) as donor in the aldol addition with different aldohexose sugars structurally related to *N*-acetyl-D-mannosamine (**2**) (D-ManNAc) and a few disaccharides[79-82]. One noteworthy example is the enzymatic synthesis of Neu5Ac as a precursor for the antiviral drug Zanamivir at multi-ton scale [83-84].

Noteworthy alternative PyrAs are the 2-keto-3-deoxy-d-manno-octosonate aldolase (KdoA, EC 4.1.2.23) with a preference for D-arabinose as acceptor, whereas 2-keto-3-deoxy-6-phospho-D-gluconate aldolase (KDPGlc/GlcA; EC 4.1.2.14) and 2-keto-3-deoxy-6-phospho-D-galactonate aldolase (KDPGal/GalA; EC 4.1.2.21) both show a preference for GA3P (**3**)[85]. Scheme 7 sums up the catalytic activity of the above mentioned PyrAs.
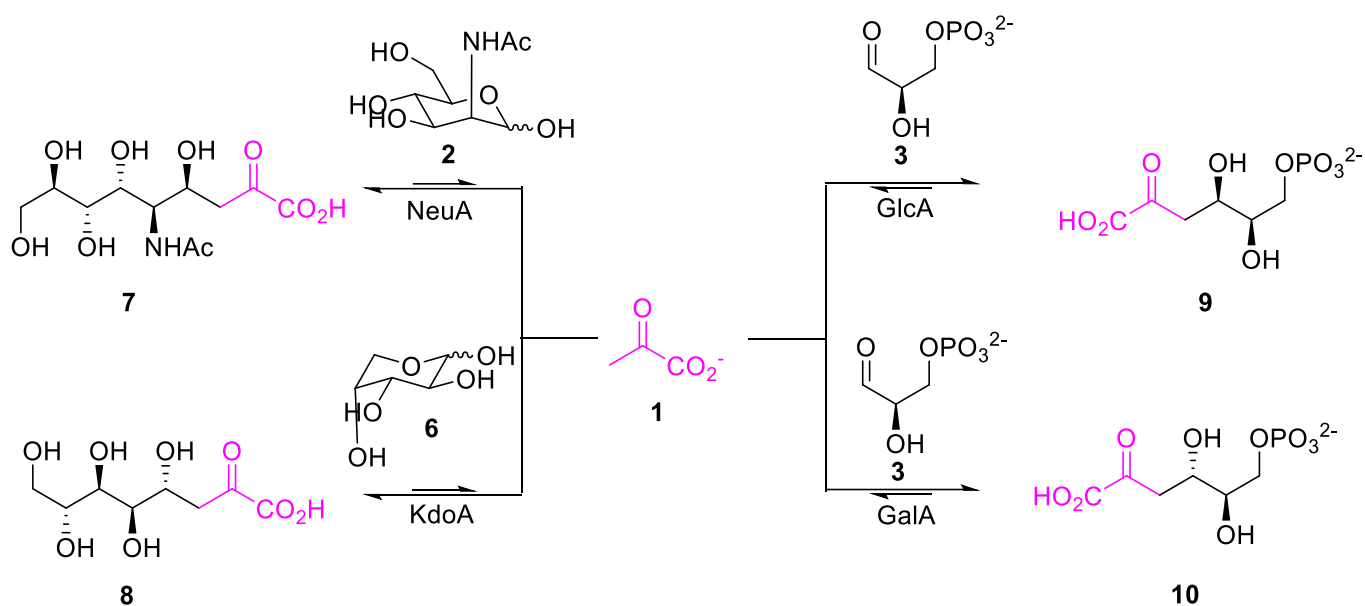
Scheme 7 Stereocomplementary sets of stereoselective PyrA: N-acetylneuraminic acid aldolase (NeuA; EC 4.1.3.3), 2-keto-3-deoxy-manno-octosonate aldolase (KdoA; EC 4.1.2.23), 2-keto-3-deoxy-6-phospho-D-gluconate aldolase (GlcA; EC 4.1.2.14) and 2-keto-3-deoxy-6-phospho-D-galactonate aldolase (GalA; EC 4.1.2.21)[85].

2.1.5 Dihydroxyacetone (DHA) dependent aldolase – Fructose-6-phosphate aldolase (FSA)

A promising alternative to DHAP dependent aldolases is the family of DHA (**13**) dependent aldolases. Fructose-6-phosphate aldolase from *E. coli (Ec*FSA) was discovered in 2001 after being wrongfully identified as a member of the transaldolase family[86], due to high structural similarity. *Ec*FSA is a class 1 aldolase with an active lysine residue in position 85, which was indicated by substitution with an arginine resulting in a loss of activity[86]. *Ec*FSA forms a decameric structure (dimer of a cyclic pentamer) which results in a remarkably stable quaternary structure for having a mesophilic origin[87-88]. With a melting point of 72.1 °C in 50 mM TEA*HCl buffer (pH 7.5) the enzyme can easily endure elevated temperatures during catalysis or heat shock purification. Each monomer possesses 220 amino acids and forms a TIM $(\alpha/\beta)_8$-barrel structure[88]. In Figure 2 the monomer with bound natural substrate fructose-6-phosphate (**12**) and decameric form of *Ec*FSA are presented with each monomer separately colored.
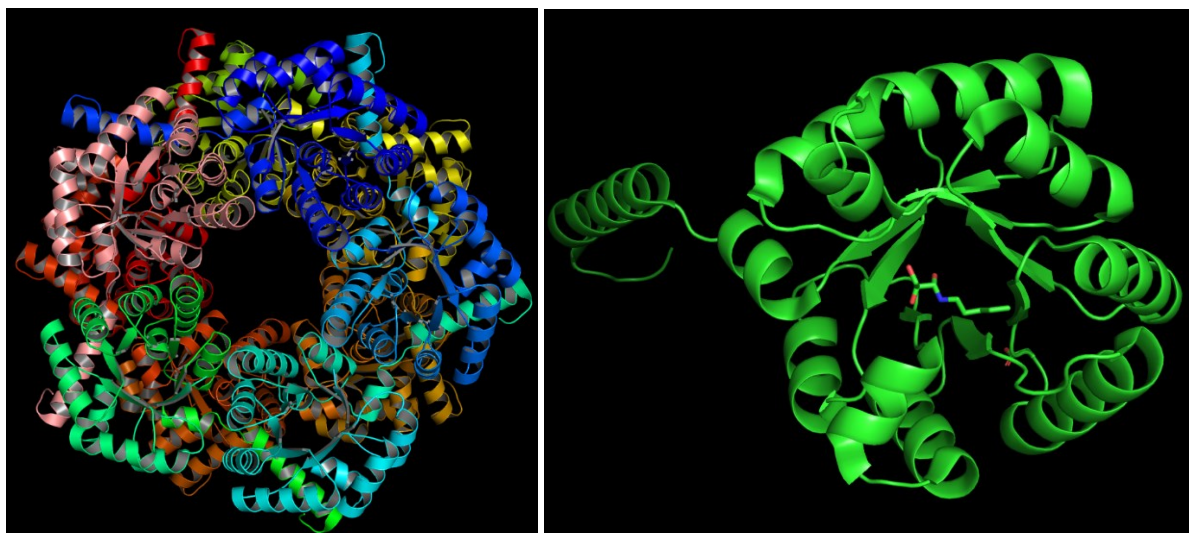
Figure 2 Crystal structure of *Ec*FSA (1L6W) presenting the decameric form with each monomer separately colored (left side). Additionally, a single monomer is presented with the bound natural substrate (right side).

The monomeric form of *Ec*FSA was never isolated cause of spontaneous arrangement into the decameric form. The active centers are found at the outer ring with a cone shaped geometry enabling bulky acceptors to take part in the reaction[89].

Usually, aldolases are named after the natural substrate which is consumed in metabolic pathways. But up to date no natural occurring reaction was identified[86]. The name given was based on the first identified substrate D-fructose-6-phosphate (**12**), which was cleaved into GA3P (**3**) and dihydroxyacetone (**13**). For the backwards reaction, the aldol addition, a strict (3*S*,4*R*) chirality is achieved[38-39, 86]. Scheme 8 presents the cleavage of D-fructose-6-phosphate (**13**) into GA3P (**3**) and DHA (**13**) catalyzed by *Ec*FSA.
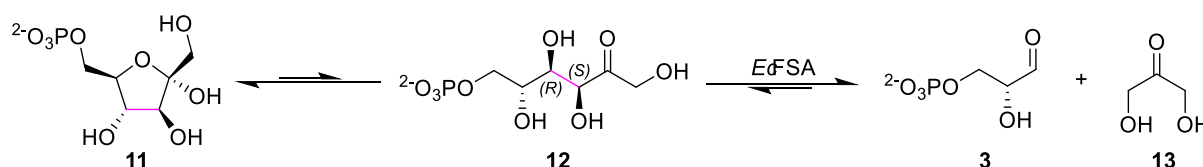
Scheme 8 Aldol cleavage of D-fructose-6-phosphate forming GA3P (**3**) and DHA (**13**) catalyzed by *Ec*FSA.

The optimal pH for catalysis is 8.5 while the enzyme can be utilized in a broader pH range of 5.5 – 11.0 [90]. Furthermore, cosolvents can be easily utilized at significant concentration without denaturing the enzyme under synthetically relevant conditions[87].

An astonishing feature of *Ec*FSA is the relaxed donor spectrum accepting hydroxyacetone (**32**), dihydroxyacetone (**13**), glycolaldehyde and 1-hydroxybutan-2-one (**33**) with high activity [86, 91-93]. The homo aldol addition of glycolaldehyde is especially fascinating due to the utilization of an aldehyde species as both the donor and acceptor, thereby forming D-threose[92]. Next to DERA, *Ec*FSA is unique for being able to utilize aldehydes instead of ketones as donor and acceptor at once, enabling the synthesis of rare deoxysugars or iminocyclitols. An important

example is the synthesis of D-fagomine, which has inhibitory activity against mammalian intestinal α-, β-glycosidase and α-, β-galactosidase and antihyperglycemic effects making it an interesting drug candidate for diabetes[91, 94]. The following Scheme 9 presents the synthetic pathway for the biocatalytic production of D-fagomine (**16**).
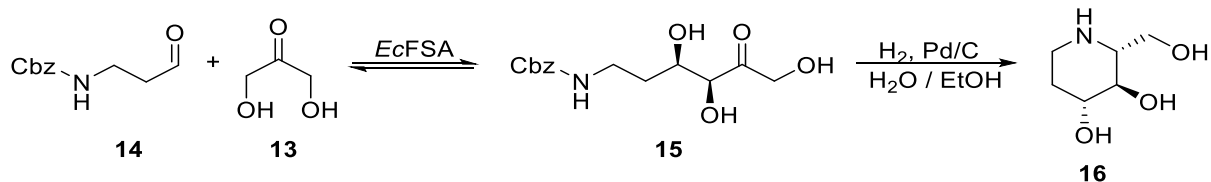
Scheme 9 Synthesis of D-fagomine (**16**) via *Ec*FSA.

Interestingly, hydroxyacetone (**32**) (HA) seems to be a more active substrate than DHA (**13**). Comparing the Kcat/kM values, *Ec*FSA shows 4 $s^{-1}mM^{-1}$ for DHA(**13**) while 145 $s^{-1}mM^{-1}$ for HA[95]. Considering that the utilization of non-phosphorylated substrates was already superior to DHAP aldolases, the relaxed donor spectrum makes DHA aldolases even more synthetically relevant. With the knowledge of the geometry of the active center, it was possible to further optimize *Ec*FSA to increase its substrate spectrum as discussed later in section **2.2.2**.

2.1.6 Acetaldehyde dependent aldolase – 2-Deoxyribose-5-phosphate aldolase

In nature, DERA catalyzes the reversible aldol cleavage of 2-deoxyribose-5-phosphate (**18**) (DR5P) into GA3P (**3**) and acetaldehyde (**17**) with a strict 3S selectivity. The reason for that strict orientation is explained by the catalytic mechanism in which the enamine can only attack from the si face.

In 1952 Racker identified DERA in cell extracts from *E. coli*[96]. Later in 1960, the equilibrium constant for the cleavage of DR5P (**18**) with DERA from *Lactobacillus plantarum* was determined at 4.2 x 10^3 M^{-1} favoring the formation of DR5P (**18**)[97]. The enzyme was identified in different metabolic processes like glycolysis, the Krebs cycle, the pentose-phosphate pathway and nucleotide catabolism[98] and in bacterial, archaeal and mammalian cells[96].
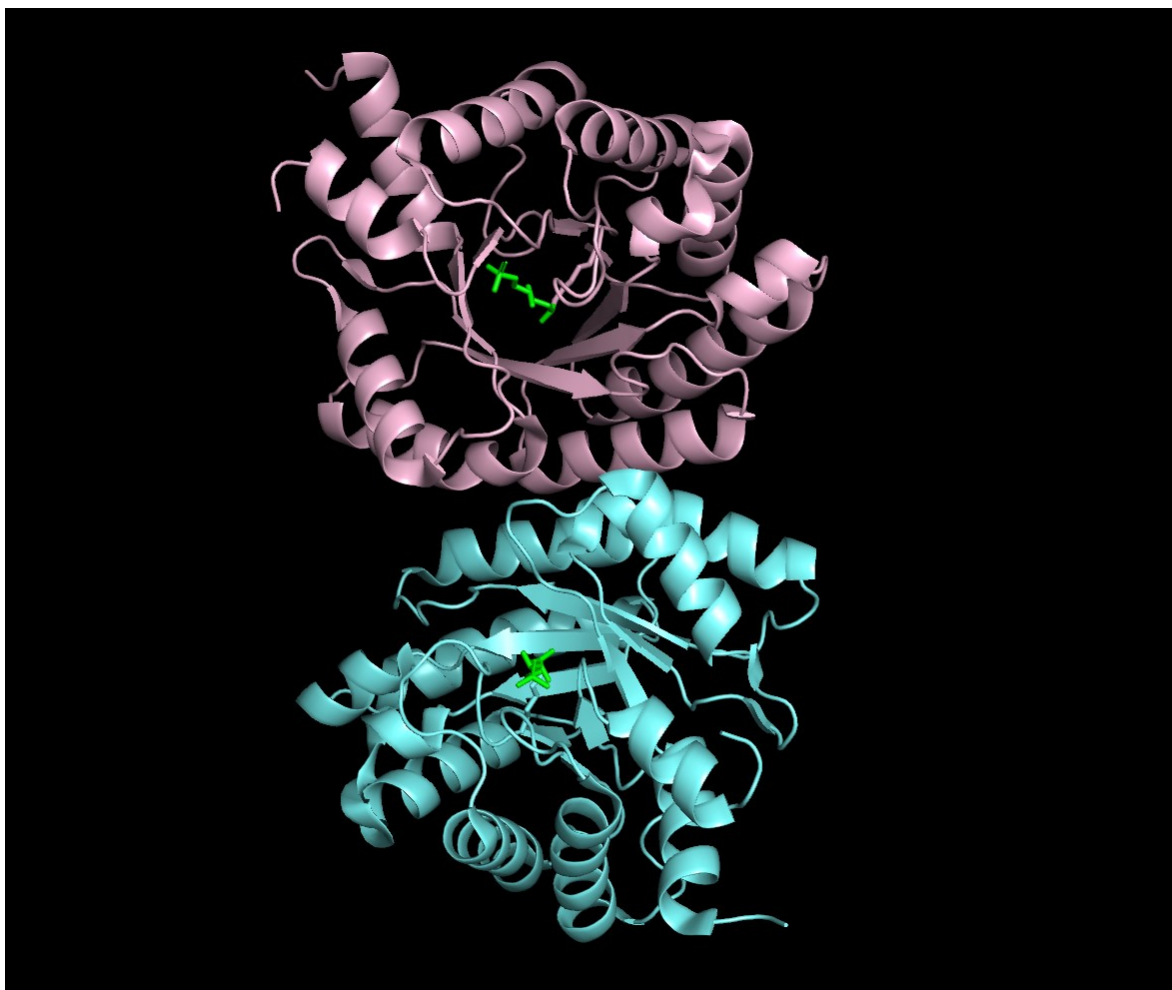
Figure 3 Crystal structure of 2-deoxyribose-5-phosphate aldolase (*Ec*DERA, PDB 1JCL) with the natural substrate 2-deoxyribose-5-phosphate bound at active Lys167.

DERA typically consists of two subunits forming a dimer with two active centers[99-103]. Related to *Ec*DERA, each monomer possesses 259 amino acids and forms a TIM $(\alpha/\beta)_8$-barrel structure[103]. Studies with the natural substrate DR5P (**18**) demonstrated that Lys167 is responsible for forming the Schiff base during catalysis[103].
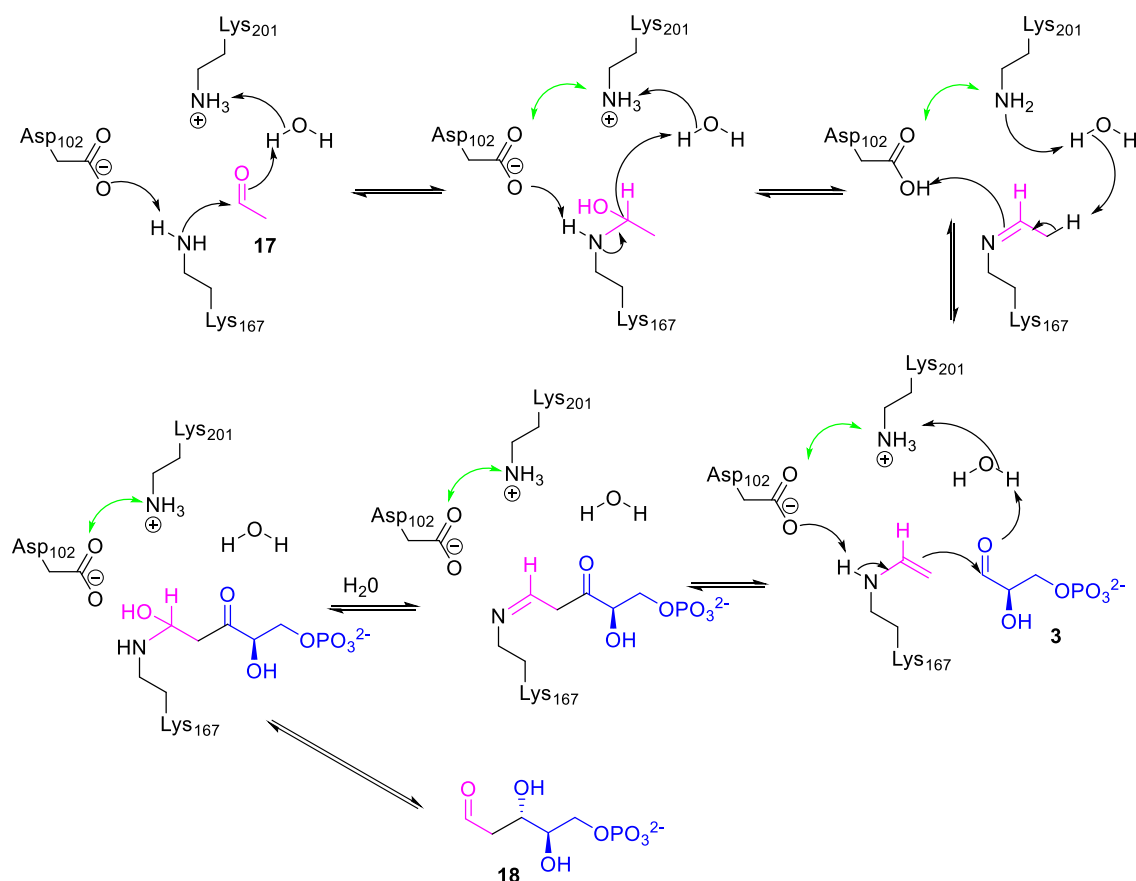
Scheme 10 Reaction mechanism for the aldol addition of acetaldehyde (**17**) as donor and GA3P (**3**) as acceptor with *Ec*DERA as catalyst forming DR5P (**18**)[104].

Interestingly, DERA variants show a tendency to get inhibited by high concentrations of the natural donor acetaldehyde (**17**), which was interpreted as a protective mechanism terminating the cleavage of DR5P (**18**) at too high concentrations of acetaldehyde (**17**). That synthetic downside was addressed by academic research with a search for novel DERAs with improved properties[105-106]. Also a random mutagenesis attempt on *Ec*DERA was conducted yielding multiple variants with increased activity and stability[107]. Later it was discovered that the product of the homo aldol addition of two acetaldehyde (**17**) molecules, which forms (*R*)-2-hydroxybutanal (**19**), is responsible for the irreversible inhibition of DERA variants due to high acetaldehyde (**17**) concentrations. The hydroxyl group of (*R*)-2-hydroxybutanal (**19**) can eliminate to yield crotonaldehyde (**20**), which can form a Schiff base with the active Lys167. The active site Cys47 can then attack the C_β-atom in a Michael type addition, forming an irreversible cross link in the active center (Scheme 11)[106].

24
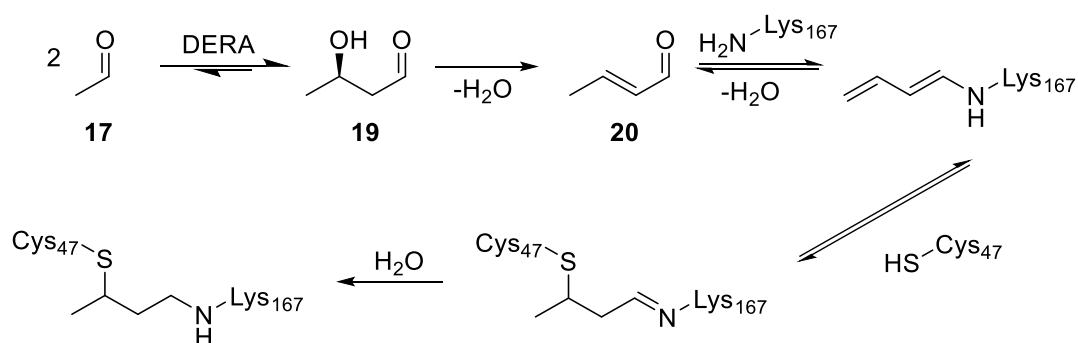
Scheme 11 Proposed mechanism for the inhibition of EcDERA by crotonaldehyde (**20**)[106].

DERA shows some fascinating features, being able to utilize aldehyde species as donor and acceptor at once enabling homo aldol reactions. Special attention got the sequential aldol addition of two acetaldehyde (**17**) molecules to chloroacetaldehyde as acceptor forming a "trimer" (3R,5S)-6-chloro-2,4,6-trideoxyhexose (**23**), which can cyclize. (**24**) can be used for the synthesis of statins or epithilones. DERA received great attention for the synthesis of chiral building blocks in the production of statins like atorvastatine (**25**), fluvastin (**26**), pravastatin (**27**), rosuvastatin (**28**) and for cancer drugs epothilone A-D based on (3R,5S)-6-chloro-2,4,6-trideoxyhexose (**23**) as a precursor. The most important statin is atorvastatin with over 100 000 000 prescriptions and patients per year[108].
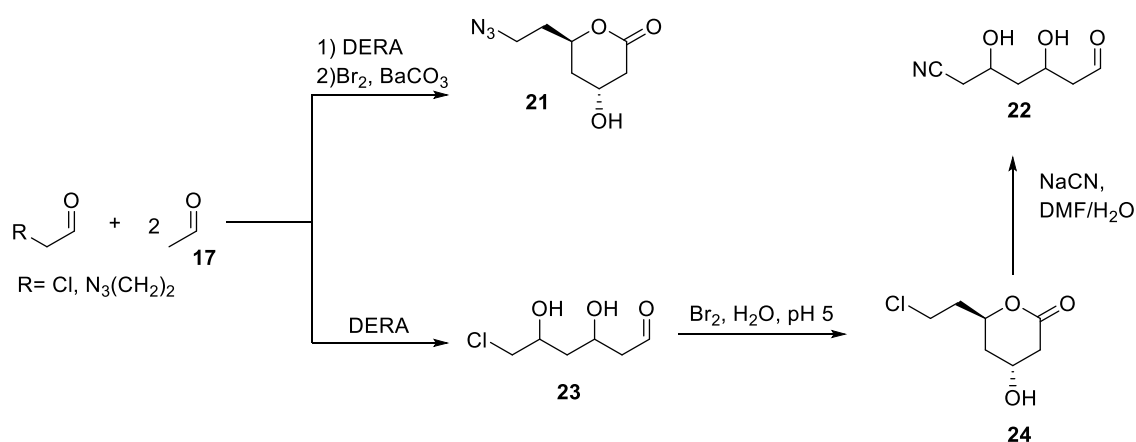
Scheme 12 Sequential aldol addition of two acetaldehyde (**17**) molecules as donor with chloroacetone (**66**) as acceptor for the production of the statin sidechain.

Considering that the lactone can be a substrate in its open form for an additional aldol addition, it may occur in rare cases when the enzyme possesses sufficient space in the active center for the conversion. The ability to form trimers was exploited in industrial scale chemistry and is still corner stone research for DERA. Scheme 13 presents a selection of statins and epothilones which are synthesized by utilizing DERA variants.
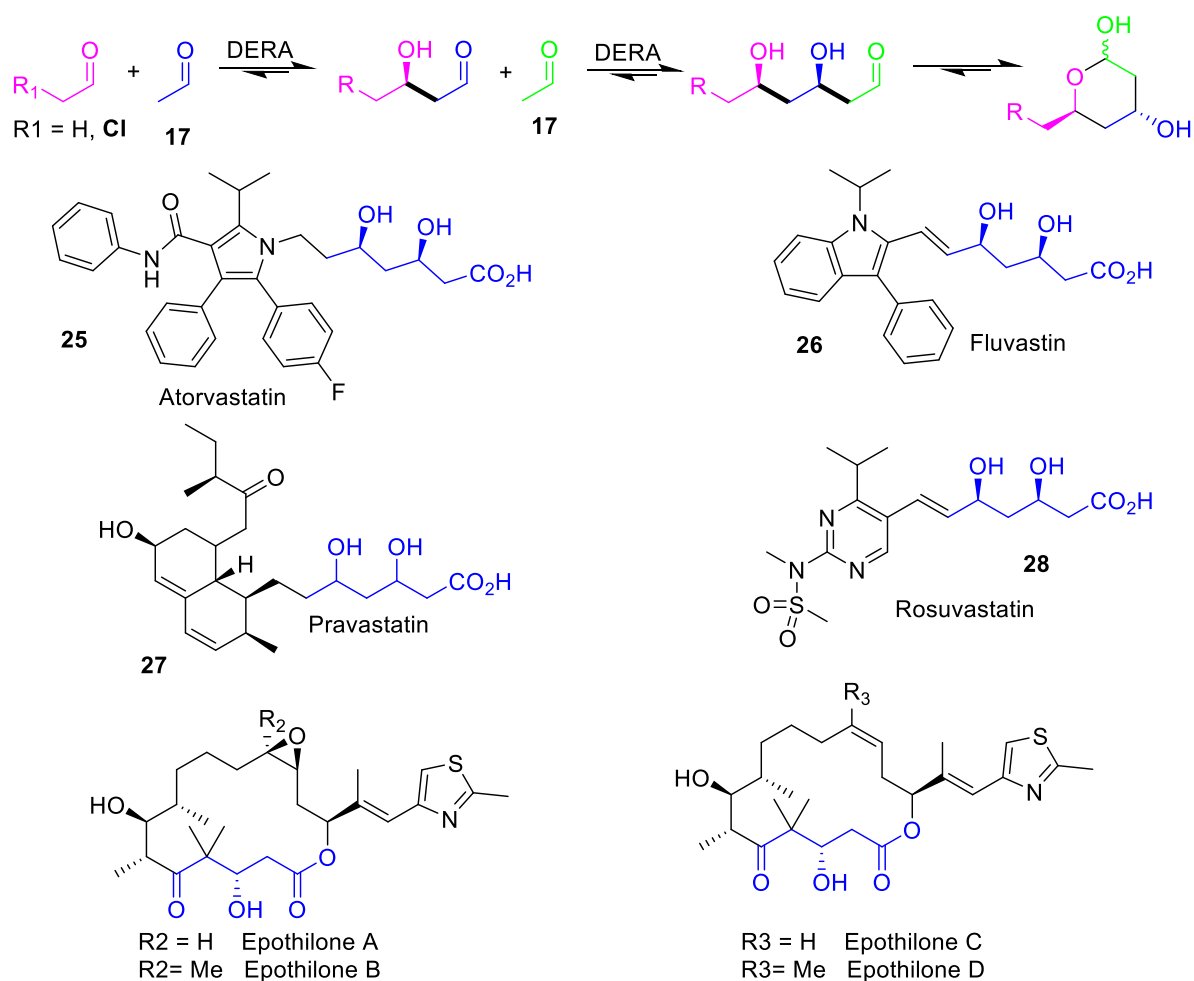
Scheme 13 Exemplary selection of statins and epothilones containing the trimer motif of acetaldehyde (**17**) catalyzed by DERA variants.

Another fascinating ability of *Ec*DERA is its relaxed donor spectrum being able to convert acetaldehyde (**17**), acetone (**50**), fluoroacetone (**77**) and propanal (47) as donor[29, 109]. In that selection, both aldehydes and ketones are included, which is unique among aldolases even though the activity is two to three digits lower when not using acetaldehyde (**17**) as donor[109]. In the natural reaction, DR5P (**18**) is utilized. When changing the acceptor part to D-glyceraldehyde instead of the phosphorylated GA3P (**3**) the same effect occurs and activity is two to three digits lower compared to the phosphorylated donor[29]. The exploration for novel donor activity was increasing over the last decade, including the search for novel DERAs from different organisms or manipulating wildtype DERA by directed mutagenesis experiments. One remarkably outstanding DERA variant is from the organism *Thermotoga maritima*[101], which was exploited for the synthesis of statine building blocks due to its higher activity and stability. Coming from a natural hot smoker habitat at sea bottom level, the variant was naturally evolved towards higher thermostability, while showing an increased donor spectrum compared to *Ec*DERA as discussed in chapter 2.

2.2 Protein engineering

In nature the ability to evolve, enables organisms to adapt to their natural habitat. While evolution is a time costly process over generations, it can be sped up in laboratory scale biochemistry. In the last 30 years tremendous progress concerning the ability to modify protein sequences was achieved.

The work of Francis Arnold during the 90's on directed evolution demonstrated a tool for scientists, which allows to introduce mutations on genetic level with a specific rate[26]. The outcome of a single mutations can lead to significant increase of activity and/or stability at once[107, 110-111]. Usually, directed evolution can yield enzymes with improved properties, while being hidden among thousands of variants to screen for the jackpot. Therefore, the success rate of the screening method to identify variants with improved properties is crucial.

Site directed mutagenesis (SDM) allows scientists to introduce specific mutations at planned positions[112-117]. SDM reflects one of the most important laboratory techniques for scientists in the field of biocatalysis. With the knowledge of the geometry and sequence of the active center from the crystal protein structure or computational calculations, scientists are able to modify the geometry and polarity of the catalytically relevant areas of the enzyme for increased substrate scope, activity or stability.

In this section the focus lays on examples of aldolases mutated with both techniques to demonstrate the possible benefits rather than explaining the techniques in depth. In the literature there are many examples of widening the substrate scope by mutagenesis experiments. In this section the most outstanding examples of mutagenic experiments for DERA and FSA will be presented.

2.2.1 Directed evolution
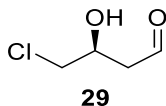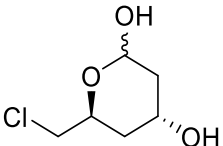
Directed evolution in vitro describes the optimization and modification of proteins and enzymes by mimicking natural selection in a speed up manner. By applying conditions in the polymerase chain reaction, which lead to errors during the process (error prone PCR), random mutations are implemented in the DNA sequence with specific rates. After screening the mutants for beneficial effects, the candidates are again subjected to error prone PCR producing multiple libraries based on the novel active mutants of the first round. The outcome after a few rounds of directed evolution results in thousands of candidates, which can be screened and analyzed for catalytic activity.

One outstanding example in the field of aldolases was produced with *Ec*DERA as a precursor enzyme by JENNEWEIN in 2006[107]. The focus of the project was on the acetaldehyde (**17**) tolerance of *Ec*DERA for increased substrate loading under industrial relevant conditions for the sequential aldol addition of two acetaldehyde (**17**) molecules with chloroacetaldehyde as acceptor forming (3R,5S)-6-chloro-2,4,6-trideoxyhexapyranoside which can be further oxidized to the lactone being an important precursor for statins.

For the screening of the resulting libraries, the cleavage of DR5P (**18**) was followed in micro titer plates. During the cleavage GA3P (**3**) is released which is reduced to D-glycerol-3-phosphate by using triose phosphate isomerase and glycerol phosphate dehydrogenase, following the oxidation of NADH to NAD+ at 340nm. This screening method allows high throughputs and the screening of thousands of variants per day, which was necessary due to the screening amount of approximately 10 000 variants obtained by error prone PCR.

The most active variants were rescreened in the sequential acetaldehyde (**17**) addition with chloroacetaldehyde as acceptor following the single addition and the formation of the trimer. Table 1 displays the results of that screening in comparison to *Ec*DERA.

Table 1 Results of the activity screening for selected mutants from error prone PCR experiment based on the sequential aldol addition of acetaldehyde (**17**) as donor and chloroacetaldehde as acceptor (Results taken from Ref. [107]).

Mutant	Amino acid change(s)	**29** [mmol/mg CFE]	**30** [mmol/mg CFE]
EcDERA	-	14.6	3.2
DERA[var1]	ΔY259	27.7	9.9
DERA[var2]	M185V	24.2	15.1
DERA[var3]	N80S, E127G, M185V, S258T, Y259T Additional C-terminal aa (KTQLSCTKW)	14.5	6.2
DERA[var4]	Y49F, M185T	15.3	4.2
nnnnnDERA[var5]	D84G, ΔY259	11.5	3.6
DERA[var6]	T19I, I166T	15.5	4.2
DERA[var7]	K13C	22.2	8.2
DERA[var8]	S93G, A174V	23.1	9.2
DERA[var9]	F200I	23.4	**44.2**
DERA[var10]	T19S	21.0	4.8
DERA[var11]	M185T	20.1	5.7
DERA[var12]	S239C	17.4	5.7

As demonstrated by the outcome of the screening, one mutant showed outstanding performance, possessing a 13fold increased activity based on the cyclized trimer. The best variant based on the formation of the trimer is DERA F200I. Figure 4 presents the active center of *Ec*DERA with the bound natural substrate DR5P (**18**) with the position F200I marked in red.

The purple marked amino acid residues reflect a selection of identified position for SDM project performed in this work.
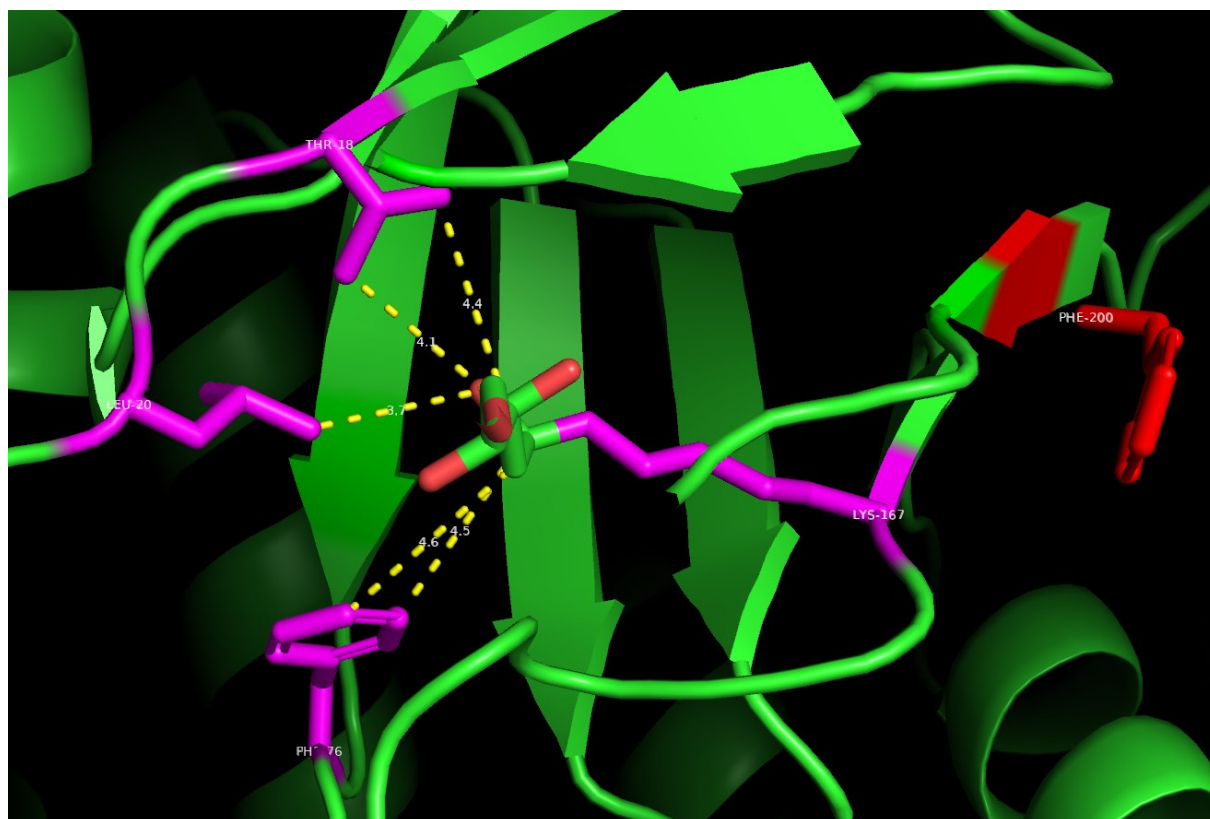
Figure 4 Active center of *Ec*DERA with the natural substrate bound with the active Lys167 residue showing the relation of the active center and position F200, which is not in close proximity of DR5P (**18**).

This example of directed evolution demonstrates the value of the techniques itself. The amino acid position of interest F200 is not in close proximity to the active center but oriented away from the active center. Therefore, the residue would rather not be considered in SDM projects since residues in the active center are more promising candidates due to possible space gained by substitution with smaller amino acids. With directed evolution, nonetheless, this highly active and more stable mutant was detected and characterized. *Ec*DERA F200I is utilized as precursor enzyme for an SDM project in Chapter 1 based on a rational design idea for increasing the space in the active site for increased donor spectrum.

2.2.2 Site directed mutagenesis
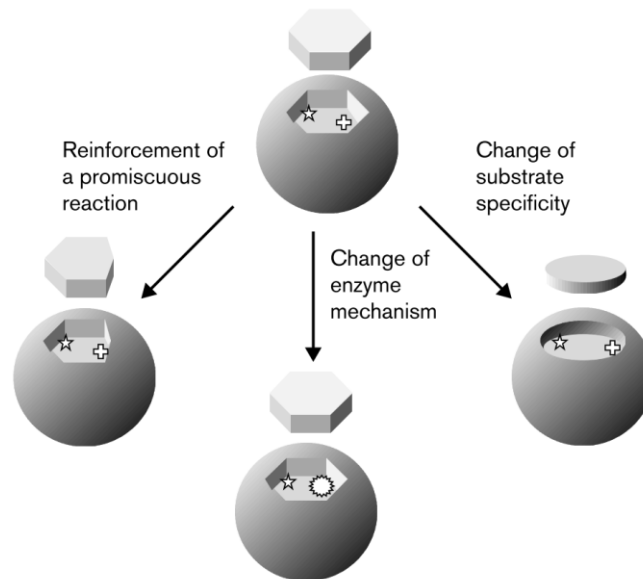
Scheme 14 Possible aims achievable by SDM including reinforcement of a promicuous reaction, change of the enzyme mechanism or change of substrate specificity (graphic taken from ref[118]).

Site directed mutagenesis has become a standard laboratory technique commonly utilized for introducing planned mutations for a rational redesign of protein structures. Usually, site directed mutagenesis is done via PCR with a set of primers containing each the mutation of interest in the primer design. After isolation of the plasmids and expression in *E. coli* cells, a screening for positive hits is conducted. After selection of the best variants, the introduced mutations to selected hits are determined by DNA sequencing.

The active center of an enzyme has usually multiple residues of interest, which are responsible for the orientation of the natural substrate and catalysis. By targeting those residues and exchanging them with amino acids with smaller residues or different polarity often results in an increased substrate spectrum for the enzyme of interest but can also lead to a total loss of activity.

Based on aldolases, and especially DERA and FSA, only *Ec*FSA was extensively studied in mutagenesis experiments with the aim of increased substrate scope. *Ec*FSA is remarkably stable due to its decameric quaternary structure while still being highly active with a relaxed donor spectrum. Those features make *Ec*FSA an attractive precursor for mutagenic experiments due to the crystal structure being available with excellent resolution[88].

The first example is from the group of Prof. Fessner[119]. In the experiment, residues requiring space in the active center were targeted with basic alanine screening. Due to a missing crystal structure with the bound natural substrate, the high structural similarity of transaldolase B was utilized for which a crystal structure with bound natural substrate was available. By analyzing the crystal structure of FSA with superimposed substrate, certain amino acid residues were identified for mutagenesis, namely L107, A129 and L163. Each targeted residue was separately mutated towards alanine or glycine (only for L129) and screened for novel activity.
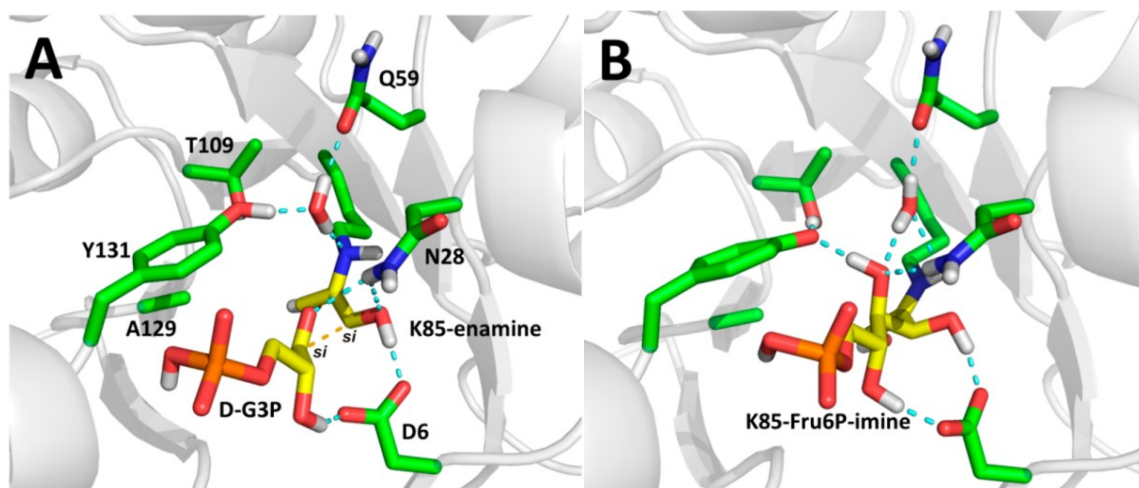
Figure 5 Molecular models of FSA-bound DHA-enanime/GA3P (A) and fructose-6-phosphate imine (B) intermediates. In both cases, the D6 residue is forming two hydrogen bonds with the C2 hydroxyl group of the GA3P (**3**) and the hydroxymethyl group of the DHA (**13**) (A), or the corresponding C3 and C5 hydroxyl group of fructose-6-phosphate (B) (graphic taken from ref [111]).

The FSA variants were screened for a selection of bulkier hydroxyacetone (**32**) analogs with 3-hydroxypropanal (**31**) as acceptor species. All mutants showed activity for the bulkier donors 1-hydroxybutan-2-one (**33**) and 1-hydroxy-3-methoxypropan-2-one (**42**) and especially A129G showed additionally activity for 1-hydroxypentan-2-one (**34**). By combining beneficial mutations, the spectrum of donors was significantly increased enabling multiple novel donors with carbon chain length of up to seven carbons instead of only 3 as in hydroxyacetone (**32**). One restriction that was not overcome is that no branching at position C3 is tolerated 1-hydroxy-3-methylbutan-2-one (**37**) but bulky structures at position C4 are accepted. Scheme 15 gives an overview of the reactions achieved.
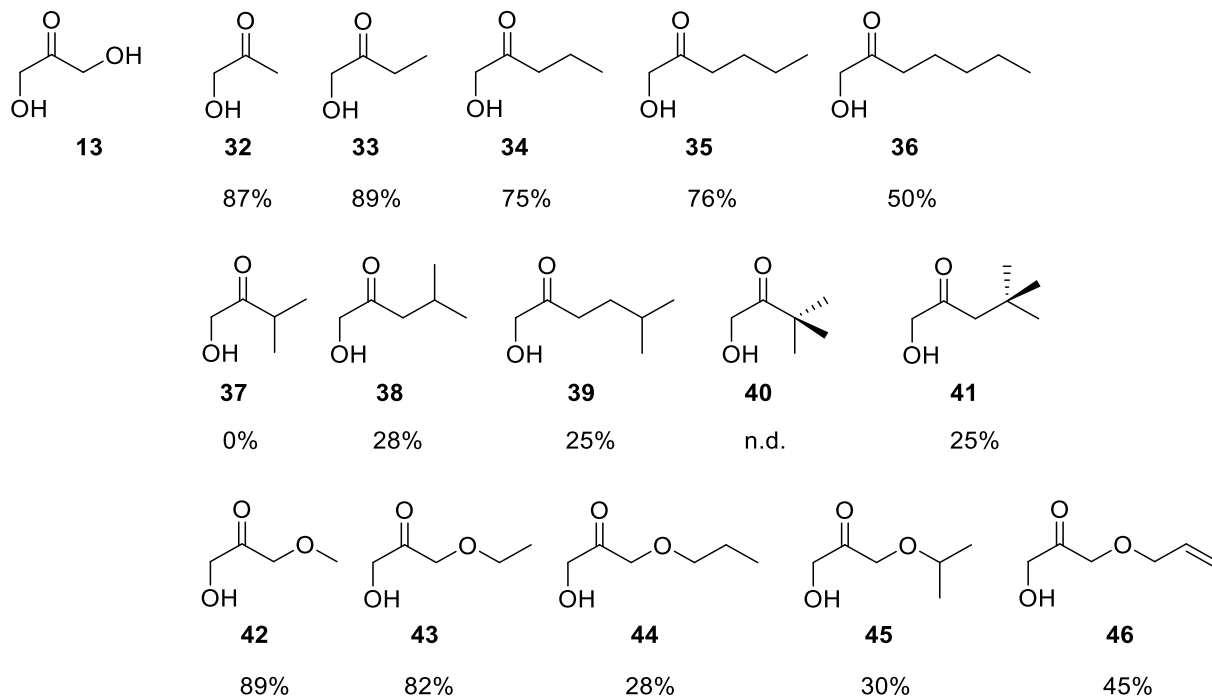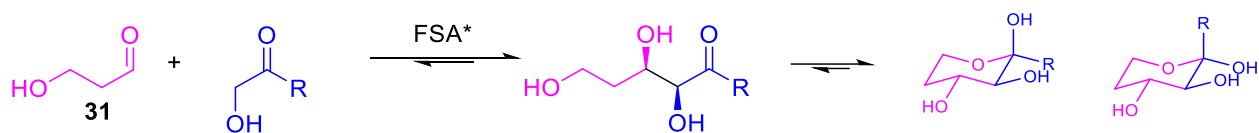
Scheme 15 Overview of novel donors accepted by *Ec*FSA modified by SDM experiments. The percentages below the presented donors corresponds to the isolated yield obtained in the cross-aldol reaction of 3-hydroxy-propanal (**31**) and the presented donors. No discrimination between α- and β-product was performed[119].

Considering that the *Ec*FSA is capable of utilizing DHA (**13**), HA and 1-hydroxy-3-methoxypropan-2-one (**42**) as donor with 3-hydroxypropanal (**31**) as acceptor, the outcome of the experiment is a great success enabling the usage of novel donors with a carbon length of up to seven carbons with the only noticed restriction that no α-branching is tolerated at the C3-position. In general, the high obtained yields are related to the nature of the acceptor species, enabling cyclization after successful aldol reaction and pushing the reaction equilibria towards the cyclized product side.

Another fascinating example utilizing *Ec*FSA as precursor for mutagenic experiments targeting the access for novel donors was performed by the FESSNER laboratory[110] and followed up by the CLAPÉS group[111, 120]. In the initial experiment of the FESSNER group, position D6 and T26 of *Ec*FSA were targeted in the active center for substitution with smaller amino acids. The newly produced FSA variants were tested in the homo aldol addition of propanal (**47**) and *n*-butanal (**53**). Additionally, it was tested if the homo aldol reaction is preferred when offering acetone (**50**) as an additional available donor. Figure 6 summarizes the results of the experiment.
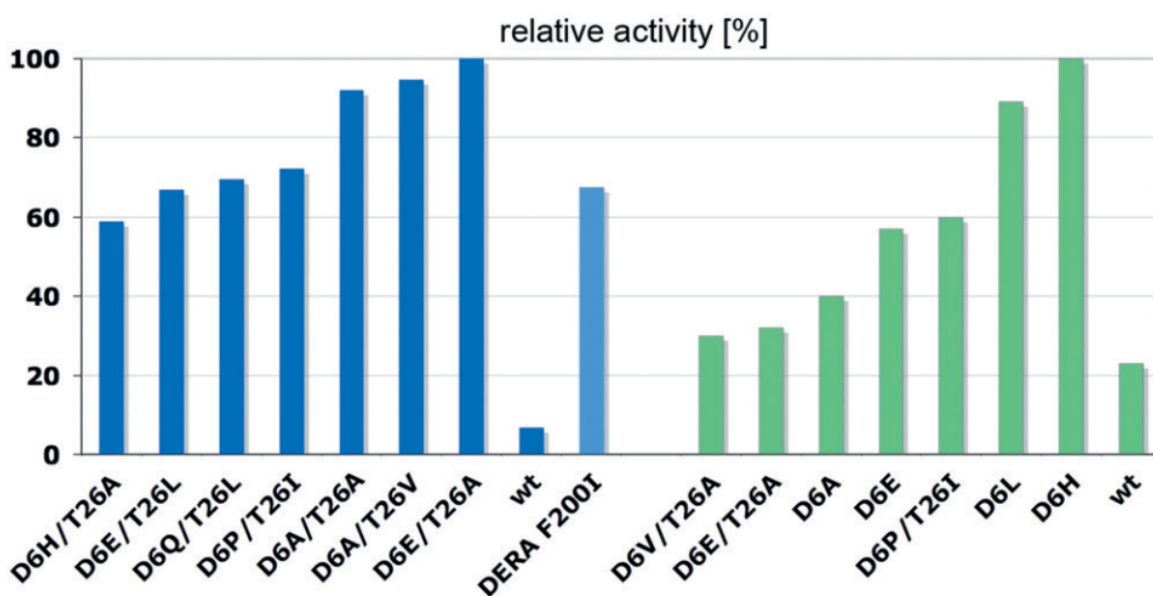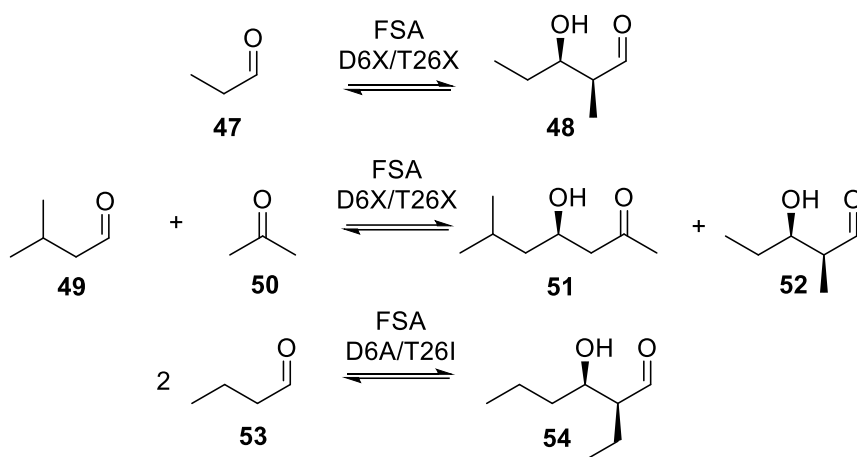
Figure 6 Blue bars: rate of formation of homo aldol product of propanal (**47**) catalyzed by FSA variants most active in primary screening. *Ec*FSA and *Ec*DERA included for reference. Green bars: rate of formation of cross aldol product of *i*-valeraldehyde (**49**) with acetone (**50**) using FSA variants (graphic taken from ref [119]).

For the homo aldol reaction of propanal (**47**), the most active variants contained two-point mutations at position D6 and T26. No single mutant led to the desired homo aldol product of propanal (**47**) or *n*-butanal (**53**). For the homo aldol reaction of *n*-butanal (**53**), the two FSA variants D6A/T26L/I were found to be active. Both homo aldol products of propanal (**47**) and *n*-butanal (**53**) are important Guerbet-type compounds for solvents/plasticizers[121]. It is important to notice, that the FSA variants were able to utilize aldehydes instead of ketones as donor species, which is only possible for the DHA (**13**) and acetaldehyde (**17**) dependent aldolase families.

The experiment of the CLAPÉS group continued on the results of the above-mentioned study. In the experiment, position D6 was targeted for SDM exchanging with A/L/N/Q/S/T/E/H. The following scheme presents the tested acceptors and the selection of donors to be tested with the generated FSA variants.
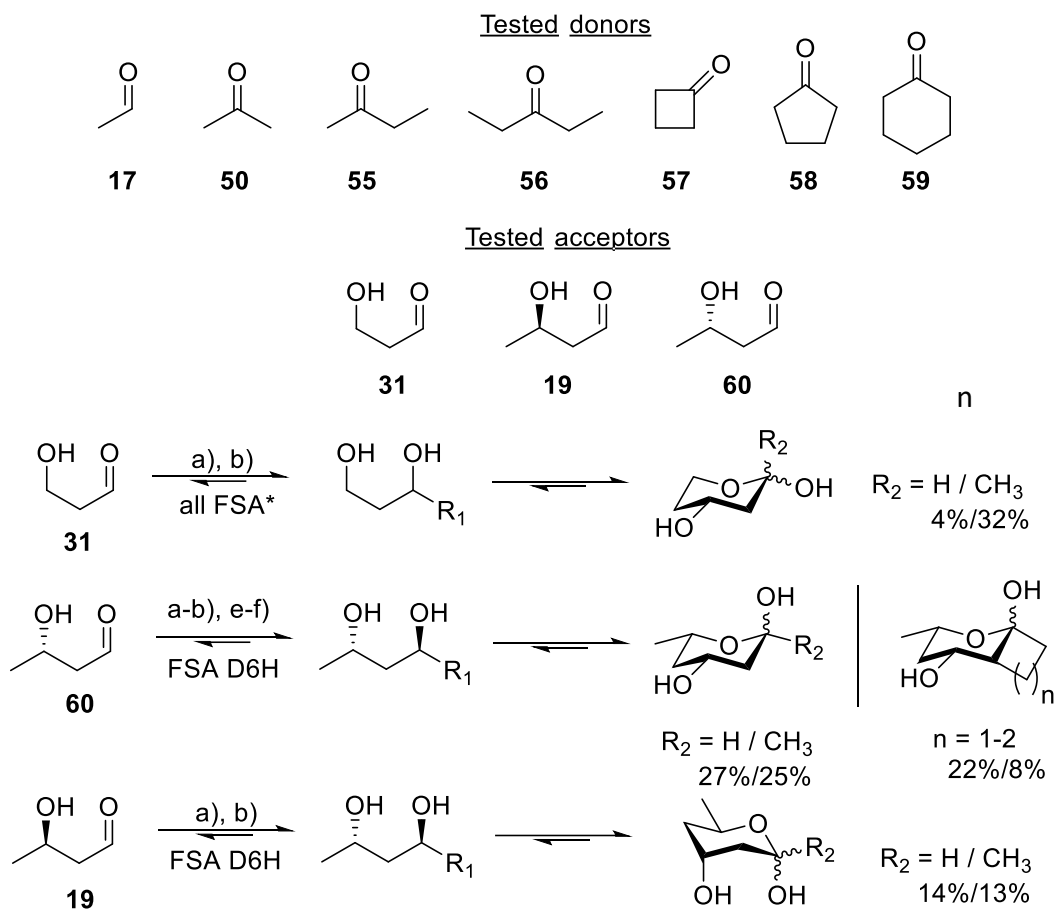
Scheme 16 Aldol additions catalyzed by FSA variants[111].

Preparative reactions were performed with 100 mM acceptor concentration and 20% v/v donor excess. The yields obtained with significant donor excess were found to be between 8 – 32%, dependent on the utilized donor/acceptor combination. The obtained yields in the cross-aldol reaction of 3-hydroxypropanal (31) and acetaldehyde (17) were found to be 8%, while the cross-aldol reaction of (R)-2-hydroxybutanal (19) yielded 27% of the desired product. When switching to (S)-2-hydroxybutanal (60) the reaction yielded approximately half the yield of the (R)-enantiomer and was determined to be 14%. This difference in obtained yields demonstrates the importance of acceptors for aldolase families.

In the standard aldol reaction of DHA (13) and D,L-GA3P (3), only the D6N variant showed similar activity (45.4 ± 0.1 U mg^{-1} vs. 46.0 0.1 U mg^{-1}) comparable to the wildtype. Since asparagine is isosteric to aspartate and has similar hydrogen bonding capabilities, this effect was expected by the authors. All other produced FSA variants showed approximately 100fold lower activity, demonstrating the importance of D6 or D6N in the wildtype for orientation of hydroxylated nucleophiles.

The D6H single amino acid exchange enabled the enzyme to accept two novel cyclic donors. Cyclobutanone (57) (22%) and cyclopentanone (58) (8%) were accepted as donors with (S)-2-hydroxybutanal (60) as acceptor, yielding the cyclized products after purification.

Interestingly, the linear donor butanone (**55**) was not accepted and seems all over more challenging for the FSA and DERA aldolases, since only one example is available in literature [39, 122].

The results of the above-mentioned studies inspired the work on mutagenic F200I *Ec*DERA library in this work.

3   Aim of work

Most of the results of this chapter were produced by M. Sc. Feodor Belov during the work on his master thesis "Directed Evolution of an aldolase from *E. coli* to increase its substrate promiscuity" under supervision of the author. Prior to this work, first-generation libraries targeting 6 positions in the active site of *Ec*DERA F200I were created by Dr. Magnusson. By overlapping and comparison of known DERA sequences from different bacterial sources, multiple variable amino acid residues were identified as target for mutagenesis. Figure 7 presents the enzyme crystal structure with the targeted amino acids highlighted.

Figure 7 Crystal structure of EcDERA (1JCL pdb) with the natural bound substrate DR5P (**18**) bound at the catalytic active Lys167. The targeted amino acid residues for SDM and the catalytic active Lys167 are highlighted in magenta.

Six variable amino acid residues were identified, which can possibly increase the space in the active site by substitution with smaller amino acids. The mutagenesis project was performed in accordance with the most common naturally occurring amino acid alterations, identified by 3DM[123], in the DERA aldolase family for the positions in question. Table 2 presents the identified positions and the planned mutations with their possible benefit on donor or substrate spectrum.

Table 2. Planned positions and amino acid alterations for the mutagenesis project in this work.

Amino acid position	Interaction with substrate molecule	Randomization	Activity with unnatural aldol donors
Threonine 18	Acceptor	Ala, Ser, Thr	+
Leucine 20	Donor, acceptor	Ala, Ser, Val, Leu	+
Valine 73	Donor	Ala, Ser, Leu, Val	−
Phenylalanine 76	Donor	Ala, Ser, Val, Phe	+
Isoleucine 139	Donor	Gly, Ser, Val, Ile	−
Serine 194	Phosphate group binding (acceptor)	Asp, Glu, Val	−

Only the mutagenesis of three of the six selected positions yielded a significant increase in substrate promiscuity for the obtained mutants: Thr18, Leu20 and Phe76. Yet, since the goal of the library creation was the widening of the substrate donor scope through the widening of the active site itself, it was determined that the first-generation libraries lacked glycine mutations on the randomized positions, whereas glycine constitutes the smallest possible amino acid residue.

Therefore, the first goal of this work was to create the missing five different mutants carrying glycine variants for positions Thr18, Leu20, Phe76. Those mutants were then characterized in the homo aldol addition of *n*-butanal (**53**), which had been utilized as test reaction before.

Further, by combining the mutations shown in Table 1 for Thr18, Leu20 and Phe76 and adding glycine to the randomization of those amino acid positions, a second-generation library was created for *Ec*DERA. The variants were tested in the homo aldol addition of *n*-butanal (**53**) and with a set of tested donors presented in Figure 8.

The next goal of this work was to screen this second-generation library for activity in the homo-aldol addition with *n*-butanal (**53**) for comparison with the first-generation library and cross-aldol additions with varying novel activity donors shown in Figure 8 and *n*-butanal (5**3**) as acceptor.
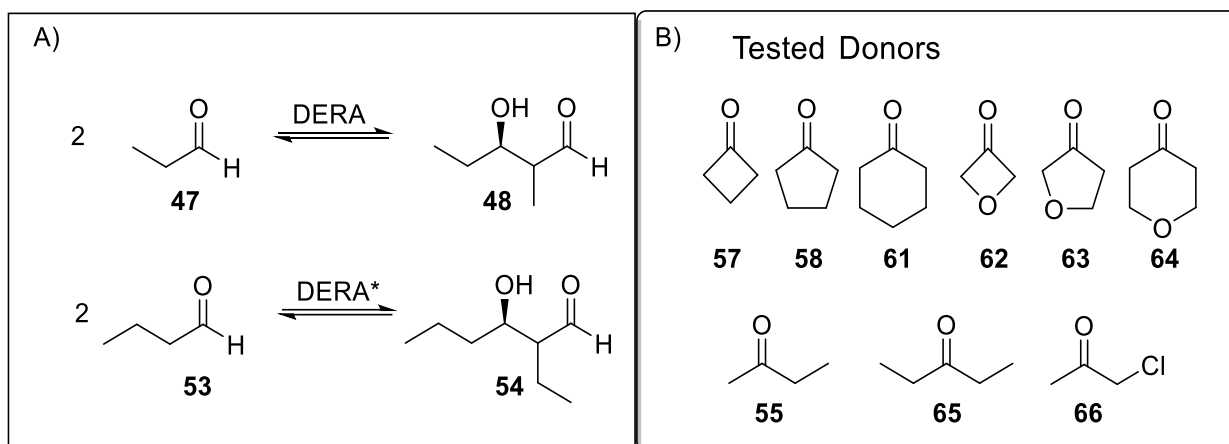
Figure 8: A. Reaction scheme of homo-aldol addition of propanal (**47**) and *n*-butanal (**53**). B. Tested novel donors for cross-aldol screening of second generation F200I *Ec*DERA library: Cyclobutanone (**57**), Cyclopentanone (**58**), cyclohexanone (**61**), 3-oxetanone (**62**), dihydrofuran-3(2H)-one (**63**), tetrahydro-4h-pyran-4-one (**64**), butanone (**55**), 3-pentanone (**65**), 1-chloroacetone (**66**).

Finally, mutants selected in the screening process had to be sequenced and characterized in terms of their turnover via reaction monitoring by HPLC analysis. Additionally, the most promising mutants were utilized in preparative scale reactions to allow product isolation and stereochemical analysis.

4   Results

4.1   Cloning of the first-generation supplementary glycine mutants

The design of the first-generation DERA-libraries was such that the grouped or single positions (T18/L20), (V73/F76), I139 and S194 were each mutated in a single experiment, resulting in four different libraries containing 96 variants each. The active center of *Ec*DERA is shown in Figure 7.

From the first generation it was known that mutations at position T18, L20 and (F76) can result in an increased activity for *n*-butanal (**53**) as novel donor. Meanwhile mutations at position V73 led to a loss of activity. The libraries of the position I139 and S238 showed no increased activity and were therefore not included in the second-generation experiments. I139 is in close proximity to the active lysine, which is responsible for the Schiff base formation in the catalytic cycle. S238 is located close to the phosphate binding site of the natural substrate DR5P (**18**).

Since the second-generation library was planned to include glycine-variants, it was decided to produce the following glycine-variants (T18G/L20V), (T18G/L20A), (T18G/L20G) and (F76G), since they were not included in the first-generation design of the enzyme library. The selection is based on the most promising DERA-variants of the first generation, namely (T18A/L20A), (T18A/L20V) and (F76A/S/V).

The (F76G)-variant was produced by implementing a point mutation at position F76, while for the generation of the other four variants, a different approach was followed. Both experiments use F200I as a precursor due to its reported higher stability and activity (acetaldehyde (**17**) trimerization)[107]. The *Ec*DERA gene cloned into the pET21a plasmid with ampicillin resistance as a selection marker, was already available from earlier experiments. The presence of the *lacI* repressor gene and its operator before the T7 promoter allowed for controlled expression of the gene of interest (deoC – *Ec*DERA) only in the presence of lactose, which would bind the repressor protein thus freeing the ribosome binding site (RBS) for the docking of the RNA polymerase for further expression. The T7 promoter itself ensured overexpression, being classified as a strong promoter. This will also be the *E. coli* vector of choice over the complete course of this work (see Figure 9). The plasmid carrying the F200I *Ec*DERA gene was successfully isolated from a cryogenic stock to be used in the further mutagenesis work.

For the verification of the purification process of the DERA-variants, SDS-PAGE was utilized. The obtained enzyme yields ranged between 200–500 mg per clone per 1 liter of expression culture.
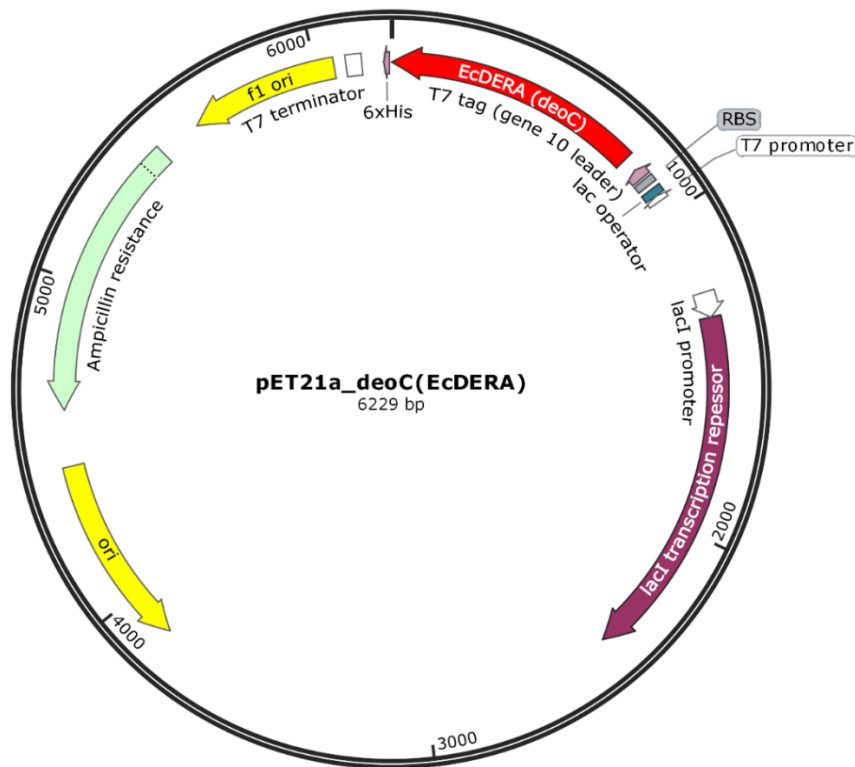
Figure 9: Plasmid map of the pET21a carrying the *Ec*DERA gene. This construct was used in all further mutagenesis and cloning work. The *Ec*DERA gene is marked in red, followed by the 6xHisTag and preceded by the T7 promoter.

For the position F76 a set of specific primers was designed following the recommendations of the "Q5 Site-Directed Mutagenesis Kit" by NEB, tailored for the mismatch replacement of the Phe76 TTC codon by a GGC glycine codon (primer: fwd. - DERA-F76G_f; rev. - DERA76-KYC_r). For the insertion of the mutation, polymerase chain reaction (PCR) was utilized. During the process the whole plasmid is amplified and due to the mutagenic primers, the final plasmid product contains the desired mutations. The success of the experiment was controlled via agarose gel electrophoresis. Since a circular plasmid is desired, the amplified plasmid is cyclized via KLD-Mix by NEB (Kinase-Ligase-DpnI enzyme mix). The transformation was carried out in *E. coli* DH5α cells (NEB) via heat shock.

After cell-growth on agar plates, randomly selected clones were cultivated, their plasmid isolated and sent for sequencing. The final cloning success was verified by the sequencing process (see Figure 10). As sequencing primer, T7-mod-fwd was utilized throughout all the mutagenesis experiments.

```
WT_sequence<       AACTCCGGTCGGCAATACCGCCGCTATCTGTATCTATCCTCGCTTTATCCCGATTGCTCG 170
F76G_mutant1<      AACTCCGGTCGGCAATACCGCCGCTATCTGTATCTATCCTCGCTTTATCCCGATTGCTCG 240
                   ************************************************************

WT_sequence<       CAAAACTCTGAAAGAGCAGGGCACCCCGGAAATCCGTATCGCTACGGTAACCAACTTCCC 230
F76G_mutant1<      CAAAACTCTGAAAGAGCAGGGCACCCCGGAAATCCGTATCGCTACGGTAACCAACGGCCC 300
                   *******************************************************  ***

WT_sequence<       ACACGGTAACGACGACATCGACATCGCGCTGGCAGAAACCCGTGCGGCAATCGCCTACGG 290
F76G_mutant1<      ACACGGTAACGACGACATCGACATCGCGCTGGCAGAAACCCGTGCGGCAATCGCCTACGG 360
                   ************************************************************
```

Figure 10: Plasmid map of the pET21a carrying the *Ec*DERA gene. This construct was used in all further mutagenesis and cloning work. The *Ec*DERA gene is marked in red, followed by the 6xHisTag and preceded by the T7 promoter.

The plasmid, carrying the mutant (F76G) was transformed into *E. coli* BL21 strain via electroporation. The BL21 strain was chosen because of deficiency of proteases in the cytoplasm and the outer membrane. A cryogenic stock of the mutant was prepared for further use.

For the preparation of the other four mutants, namely (T18A/L20G), (T18G/L20V), (T18G/L20A) and (T18G/L20G) a slightly different approach was utilized. For the mutation of position F76, mismatched primers were applied in the PCR. From experience in our group, mismatched primers work well with single point mutation, while for two mutations in close proximity at once, those primers can lead to low yields or unsuccessful PCRs. For the preparation of the (T18/L20)-variants it was decided to first delete the sequential part of the *Ec*DERA-gene coding for (T18/L20) (9 nucleotides). The next step is the insertion of the desired mutations via PCR. This convenient method results in an extra step but therefore yielded excellent results. The general procedures of both steps leading to the mutated clones are identical to those described in the cloning of the (F76G) mutant. The primers DERA18-20del_f (forward primer) and DERA18-20del_r (reverse primer) were used for the deletion of the original nine nucleotides. The design of the primers is listed in Table 3:

Table 3: Mutagenesis insertion primers for the cloning of the mutants T18A/L20G, T18G/L20V, T18G/L20A, T18G/L20G.

Mutant	Forward primer	Reverse primer
T18A L20G	DERA-L18/20G_f	DERA-T18A-L20G_r
T18G L20V	DERA-T18G-L20V_f	DERA-L18/20G_r
T18G L20A	DERA-T18G-L20A_f	DERA-L18/20G_r
T18G L20G	DERA-L18/20G_f	DERA-L18/20G_r

The cloning success was verified through sequencing and is shown in the alignment with the WT sequence in Figure 11:

```
T18GL20A_clone1<      CTGGGCACCGCGAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCAGGCC      179
T18GL20G_clone1<      CTGGGCACCGGCAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCAGGCC      179
T18AL20G_clone1<      CTGGCCACCGGCAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCAGGCC      179
WT_sequence<          CTGACCACCCTGAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCAGGCC      108
T18GL20V_clone2new<   CTGGGCACCGTGAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCAGGCC      178
                      ***  ****  *************************************************
```

Figure 11: Alignment of the successfully cloned T18A/L20G, T18G/L20V, T18G/L20A and T18G/L20G mutants with the *Ec*DERA gene WT sequence. The mutation site is marked with a red arrow (nucleotides 52-54 and 58-60 of the WT sequence). T18 (ACC) is replaced by A (GCC) or G (GGC), and L20 (CTG) is replaced by A (GCG), G (GGC) or V (GTG).

After sequencing the mutant plasmids were selected, transformed into *E. coli* BL21 expression strain and a cryogenic stock was prepared for further cultivation.

4.1.1 Expression and purification of the glycine supplementary mutants

After successful generation and verification of the newly generated DERA-variants, all variants were tested in the non-natural homo aldol test reaction of *n*-butanal (**53**). Therefore, it was started with the expression of all glycine-variants in auto induction media (AIM). As comparison, one of the best variants from the first generation (T18A/L20V) was used as a reference. The use of AIM media was found to be superior to lysogenic broth media with isopropyl-β-D-1-thiogalactopyranoside induction in terms of enzyme yield. The idea of AIM media is that glucose and lactose both are available but the cells first start consuming the glucose for their cell growth stage. When the glucose is fully consumed, the *lacI* repressor allows for the overexpression of the DERA-enzyme by consumption of lactose. The process of expression usually is initiated in the afternoon, starting with the inoculation of the AIM media with the desired *E. coli* cells. Afterwards, the expression media is placed in a shaker at 37°C for 18h. The next day, the cells can be centrifuged, lysed with lysozyme and DNase and proteins purified via the implemented His-tag. The desalted enzyme solution was lyophilized to obtain solid yields of purified enzyme. The appearance of the enzymes was a white fluffy foam.

All above mentioned steps were monitored via SDS PAGE. An exemplary picture of SDS PAGE is shown in Figure 12, monitoring the purification of variants (T18G/L20A) and (T18G/L20G):
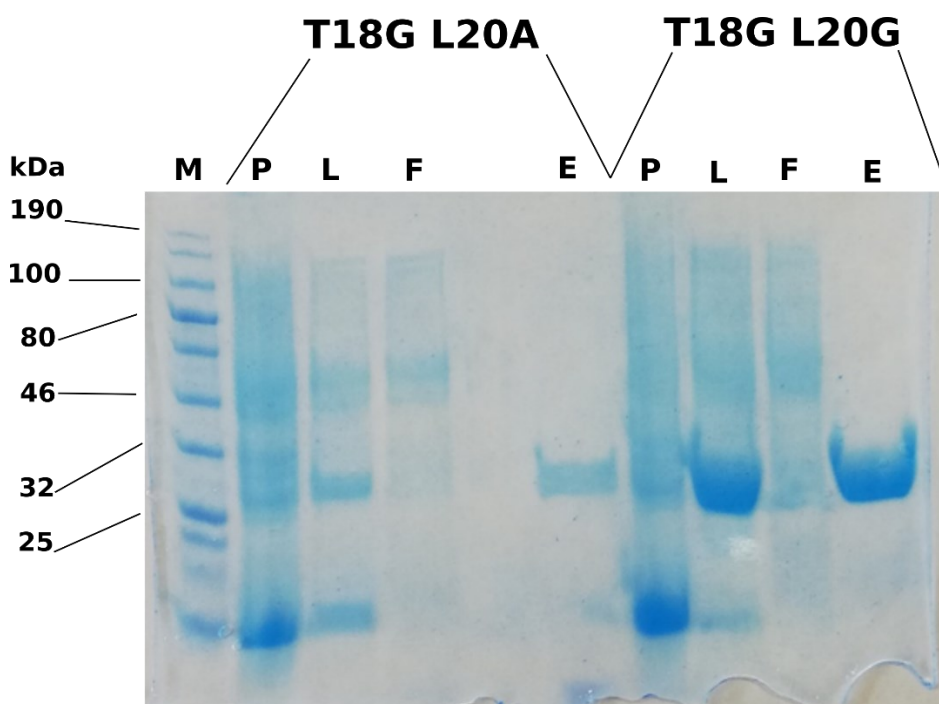
Figure 12: SDS PAGE of the purification process of the mutants T18G/L20A and T18G/L20G. Marker: Blue Prestained Protein Standard Broad Range (11-250 kDa; NEB). M – marker, P – cell debris pellet after lysis, L – cell lysate, F – HisTrap column flowthrough after lysate loading, E – HisTrap column elution.

All DERA variants were successfully generated, identified and purified, yielding between 200-500 mg enzyme per clone per one liter of expression media. Those results were satisfying and it was decided to proceed with the characterization of the DERA-variants in unnatural chemical reactions.

4.2 Characterization of first-generation supplementary glycine variants

The known donor spectra for *Ec*DERA are well studied and results in four naturally accepted donors with the natural acceptor GA3P (**3**), namely acetaldehyde (**17**), acetone (**50**), fluoroacetone (**77**) and propanal (47). While acetaldehyde (**17**) as the natural donor has the highest activity, the other three donors have a relative activity of around 1/100 the activity of acetaldehyde (**17**). Aldolases are more restrictive for the donor, but the change of acceptor from GA3P (**3**) to propanal (47) or *n*-butanal (**53**) is a major structural change that most likely results in a lower activity. Therefore, the use of both *n*-butanal (**53**) as donor and acceptor at once is a dramatical change compared to the naturally occurring reaction and will result in lower yields than with the natural pendant. But still *n*-butanal (**53**) is the next homologue after propanal (47), which is accepted by *Ec*DERA.

For initial analysis of positive or negative results in the enzyme reactions, thin layer chromatography (TLC) with anisaldehyde stain was utilized. Positive hits were then further analyzed via high performance liquid chromatography (HPLC). For the diode array UV-detection in HPLC, derivatization of the aldol product is necessary, which was achieved via

reaction with O-benzylhydroxylamin in methanol/pyridine. For the quantification by HPLC, a standard curve for *n*-butanal (**53**) was generated in the range of 0–200 mM. Here, the fit for the measurement points was to include the origin to get a clear approximation for peak area counts/mM of the measured compound without side parameters.
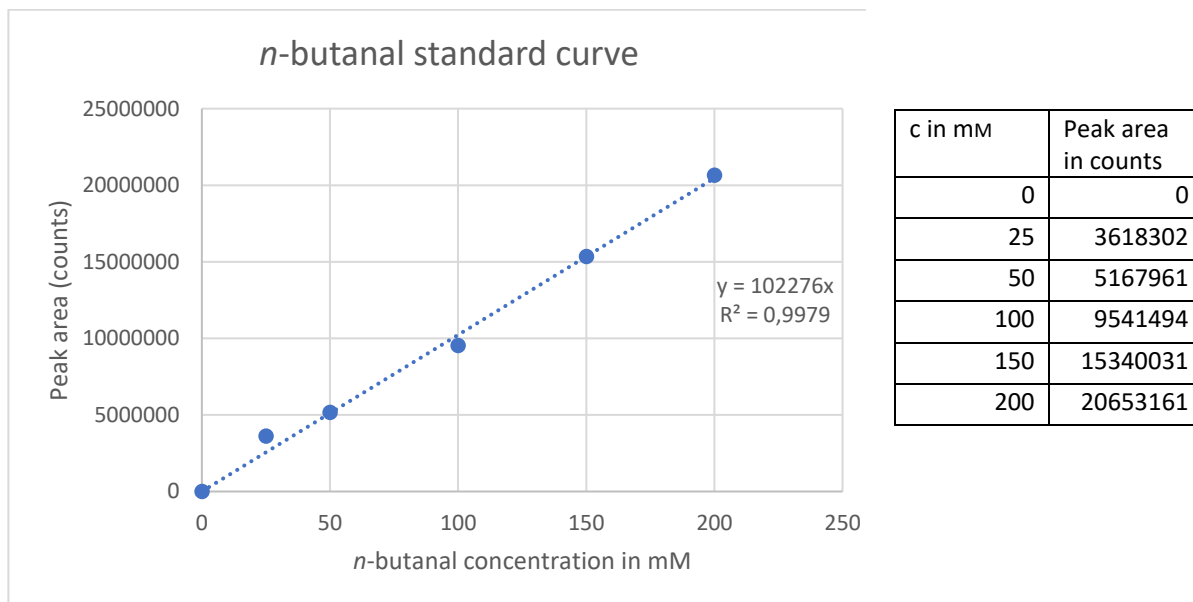
Figure 13 Standard curve derivated for *n*-butanal (**53**). Numerical data are listed in the table on the right.

c in mM	Peak area in counts
0	0
25	3618302
50	5167961
100	9541494
150	15340031
200	20653161

Control experiments with acetaldehyde (**17**) and *i*-butyraldehyd showed insignificant changes of the obtained peak area counts/mm value below 3% (data not shown), reflecting that the calibration curve can be used for a broad range of different aldehydes and ketones. On molecular view, with a methylene group and oxime-group in between the measurable phenyl ring and the aldehyde function, the phenyl ring is isolated enough to not experience deviating electronic effects. The slope directly indicates the necessary value, reflecting 102276 area counts/mM substance in the HPLC-sample.

4.2.1 Characterization of the homo-aldol addition activity with *n*-butanal (53)

For the characterization of the DERA-variants in the homo aldol reaction of *n*-butanal (**53**) the following conditions were utilized:

- *n*-Butanal (**53**) 100 mM
- TEA*HCl buffer, pH 7.5, 50 mM
- DMSO 20% v/v
- Purified DERA-variant 3.5 mg/mL

One factor that influences the resulting yields, is the high volatility of *n*-butanal (**53**) with a boiling point of 75°C. Therefore, the yields may be higher than the measured value due to evaporation of *n*-butanal (**53**) during the experiment. For minimizing that error, sealed HPLC-vials were used as reaction vessel.

The enzyme reactions were allowed to proceed overnight in a closed vessel at room temperature with constant shaking at 900 rpm. The next day, the reactions were analyzed via TLC followed by HPLC quantification. The results of these experiments are shown in Figure 14:
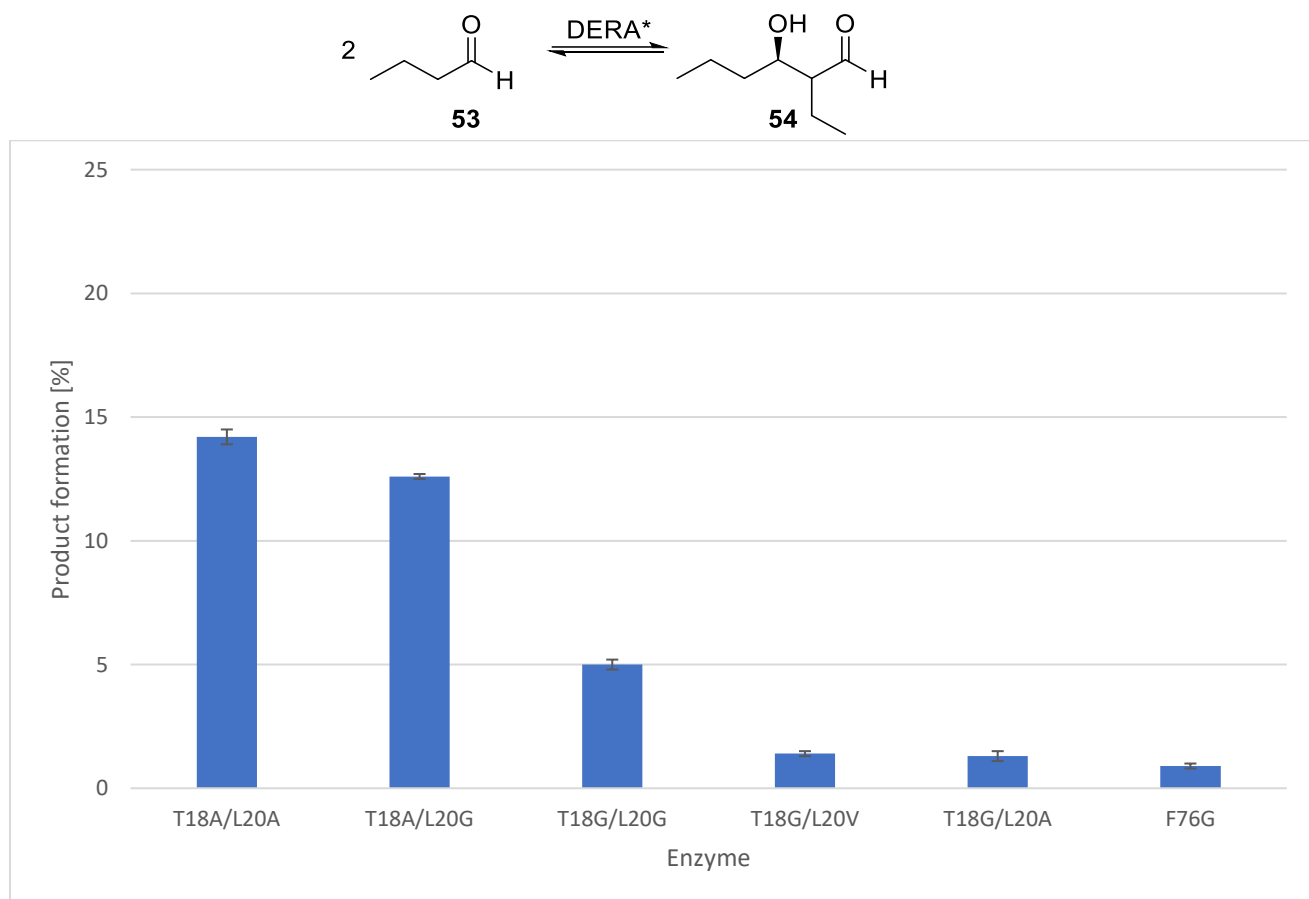
Figure 14 Results for the *n*-butanal (**53**) homo-aldol addition catalyzed by the supplementary first-generation mutants after 22 hours. Resulting educt and product concentrations, as well as conversion and product formation rates were calculated using the *n*-butanal (**53**) standard curve.

The experiments were performed in triplicates. The conversion factor of 102267 counts/mм was utilized for determination of the product yields. This conversion factor is utilized throughout the following results determined by HPLC in this chapter. The relative error of 1% up to 29% seems dramatic on first view, but the high error values were only obtained for samples with insufficient yields for further usage. Therefore, the error rate is dramatically higher for low conversion samples than for the samples with sufficient yields e. g. (T18A/L20G). DMSO was chosen as cosolvent due to its excellent solvating capabilities, and also from earlier studies [87] it was known that DMSO is one of the less harmful cosolvents on thermostability and chemostability of proteins in general.

From earlier studies it was known that the DERA-variants (T18A/L20A+V) both showed good results. With the natural substrate DR5P (**18**), T18 is responsible for orientation, through hydrogen bridges with a water molecule, of the hydroxyl groups at C1 and C3. Position L20 is not directly involved in the orientation of DR5P (**18**) but is in close proximity and is possibly blocking space for ketone acceptors and bulkier substrates than the natural substrates. The best variants from the first generation were those that bare a mutation at position T18 and a combination with L20, by increasing the space in those two areas surrounding L167 at which the binding of the donor occurs. One aspect that is quite fascinating is that by mutating T18 to aliphatic amino acids, the reaction with *n*-butanal (**53**) becomes possible but no candidates were found in the first generation having only a mutation at position L20, that could transform *n*-butanal (**53**). The mutations at F76 were less promising in theory since it is again not responsible in the orientation of DR5P (**18**) and the only gain by mutating it towards smaller amino acids is the increased space in the active center. But this feature alone was enough to increase activity in the *n*-butanal (**53**) homo aldol reaction even though by a factor of 10fold less than the T18/L20-variants from the first generation.

By comparing the (T18A/L20A), (T18G/L20G) and (T18A/L20G) mutant, one can say that the mutation of T18A has the most beneficial effect on activity in the *n*-butanal (**53**) homo aldol reaction. For T18G a loss of activity was observed for the DERA-variants (T18G/L20A+V) but interestingly a positive effect was observed when T18 and L20 were both mutated towards glycine. However, the resulting activity of T18G/L20G is still 3-fold lower than for the reference DERA-variant (T18A/L20A).

4.2.2 Characterization of the cross-aldol addition activity with acetone (50) and *n*-butanal (53)

After successfully testing *n*-butanal (**53**) as a novel donor for the generated glycine variants, the next step was to test for cross aldol activity with acetone (**50**). Of course, cross aldol reactions with *n*-butanal (**53**) as acceptor will always compete with the *n*-butanal (**53**) homo aldol reaction. Therefore, the enzyme will kinetically discriminate between *n*-butanal (**53**) and acetone (**50**) as a donor. Since acetone (**50**) is accepted by *Ec*DERA, while *n*-butanal (**53**) is not, there is a good chance that the desired cross aldol product will be formed more favorably than the homo aldol product. In previous experiments with the first-generation variants, the cross-aldol reaction of *n*-butanal (**53**) and acetone (**50**) was already achieved. Therefore, product formation was expected in this experiment. For both, acceptor and donor, a concentration ratio of 1:1 was used to see how the variants discriminate between *n*-butanal (**53**) and acetone (**50**) as donor, which reflects also the consumption preference of aldehydes to ketones. By using higher donor concentrations, in this case acetone (**50**), the cross-aldol product would be favored at a higher degree. Increasing donor concentrations is especially important for more challenging, unnatural donors like cyclopentanone (**58**) or cyclobutanone (**57**), which are to be tested subsequently to successfully utilizing acetone (**50**),

which is the simplest ketone. As a control, variant (T18A/L20A) was utilized, due to its good results from previous work. The following conditions were utilized during the experiment.

- *n*-Butanal (**53**) 100 mм
- Acetone (**50**) 100 mм
- TEA*HCl buffer, pH 7.5, 50 mм
- DMSO 20% v/v
- Purified DERA-variant 3.5 mg/mL

The use of DMSO as cosolvent was kept at 20% v/v for comparison reasons but is indeed unnecessary since acetone (**50**) is already a good cosolvent. Problematic is the fact that acetone (**50**) will also function as a donor with DERA-variants and therefore cannot be chosen as the standard cosolvent. Furthermore, acetone (**50**) has a more negative effect on thermostability than DMSO [87].

The reactions were analyzed after 24h at RT via TLC and HPLC. For determination of the resulting concentrations a conversion factor of 102276 area counts/mm was utilized. The results are listed in Figure 15.
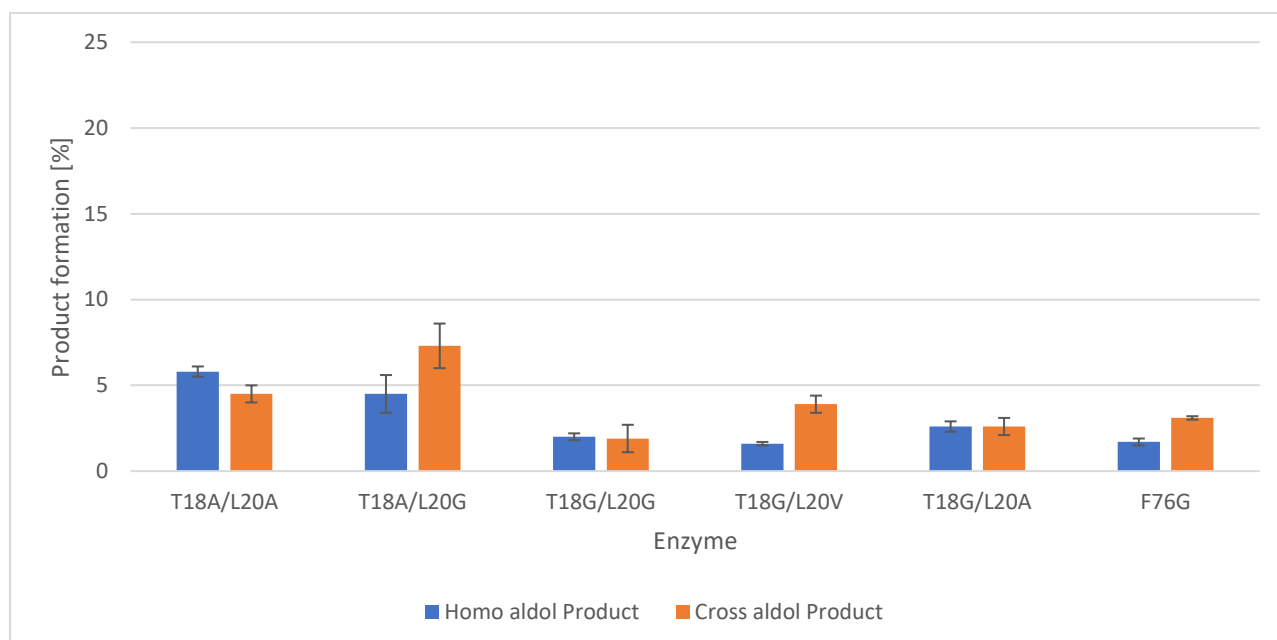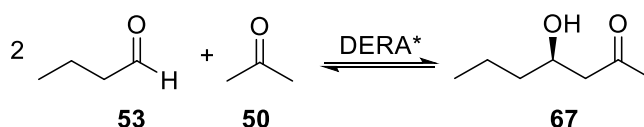
Figure 15 Results for the cross-aldol addition of acetone (**50**) and *n*-butanal (**53**) catalyzed by supplementary first-generation mutants after 24 hours. Resulting educt and product concentrations, as well as conversion and product formation rates were calculated using the *n*-butanal (**53**) standard curve.

As expected, all DERA-variants yielded both products in varying amounts. The wild type was not tested due to experience, yielding no product in the above tested cross and homo aldol reaction. Due to acetone (**50**) being a natural donor, while *n*-butanal (**53**) is not, the cross-aldol product was expected to be favored. Interestingly, this was not the case for all DERA-variants. (T18A/L20G), (T18G/L20G) and (F76G) favored the cross-aldol product, while (T18A/L20A), (T18G/L20A) and (T18G/L20V) had no preference for either product and both the cross-aldol and homo-aldol product were formed in equimolar amounts. Interestingly, the newly created DERA-variant (T18G/L20G) showed the highest overall product formation with 11.8%, while for (T18A/L20A) slightly lower total product formation of 10.3% was observed. Furthermore (T18A/L20G) favors cross-aldol reaction which is the desired reaction path, while the homo-aldol product is the unwanted byproduct. All other tested glycine DERA-variants yielded lower total product formation.

Additionally, the residual concentrations of *n*-butanal (**53**) and acetone (**50**) were determined. The results indicate a significant loss of acceptor and donor concentrations, plausibly due to evaporation during the reaction and preparation of analytical samples. As a result, the observed product quantities are lower compared to when relating the product formation to the overall acceptor concentration after the experiment. As an example, (T18A/L20A) would have ~ 20%/20% calculated for homo/cross aldol product respectively, which is 5fold higher. Noteworthy are some of the high error rates of up to 20%, which again result from low product formation rates. In principle, 1% product formation error is generally acceptable. But related to the acquired low yields during the experiment, the impact of those error rates is significant. However, the DERA-variants having those high error rates are anyway not synthetically useful because of their insufficient relative activity.

Reflecting on the observation that apparently all DERA-variants have enough space in the active center to utilize *n*-butanal (**53**) as a donor, one can conclude that mutations at positions T18, L20, F76 show an impact on the preference of aldehydes over ketones. All DERA- glycine -variants have at least equimolar ratio of homo to cross aldol product or even up to 1:2 in favor of the cross-aldol product, which is remarkable. Another possible explanation is that *n*-butanal (**53**) is already challenging concerning the space demanded in the active center. Acetone (**50**) is a natural donor, even though 100fold less active than acetaldehyde (**17**), but still can be favored due to its smaller size despite being a less reactive ketone.

The DERA-variants (T18A/L20A), (T18A/L20V), (T18A/L20G) and (T18G/L20G) were further analyzed with different ratios of acceptor to donor. Due to the time consuming HPLC-analysis, the following experiments were conducted only in unicates. The tested conditions and results are summarized in Figure 16–19.
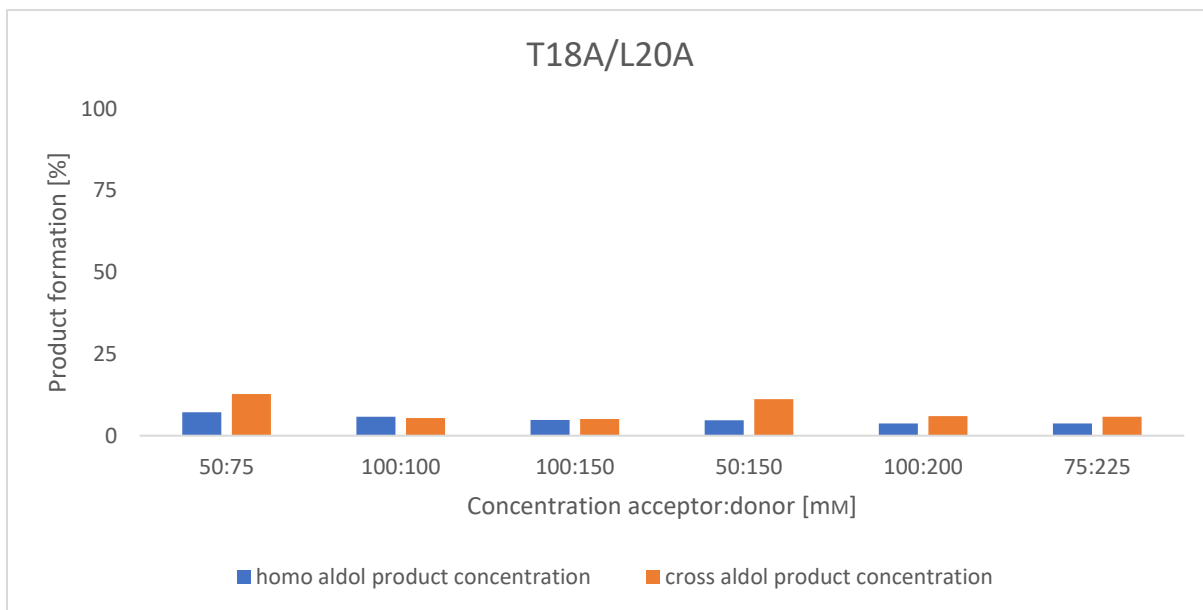
Figure 16 HPLC analysis of the cross-aldol reaction of acetone (**50**) and *n*-butanal (**53**) performed by variant T18A/L20A with varying donor/acceptor concentrations.
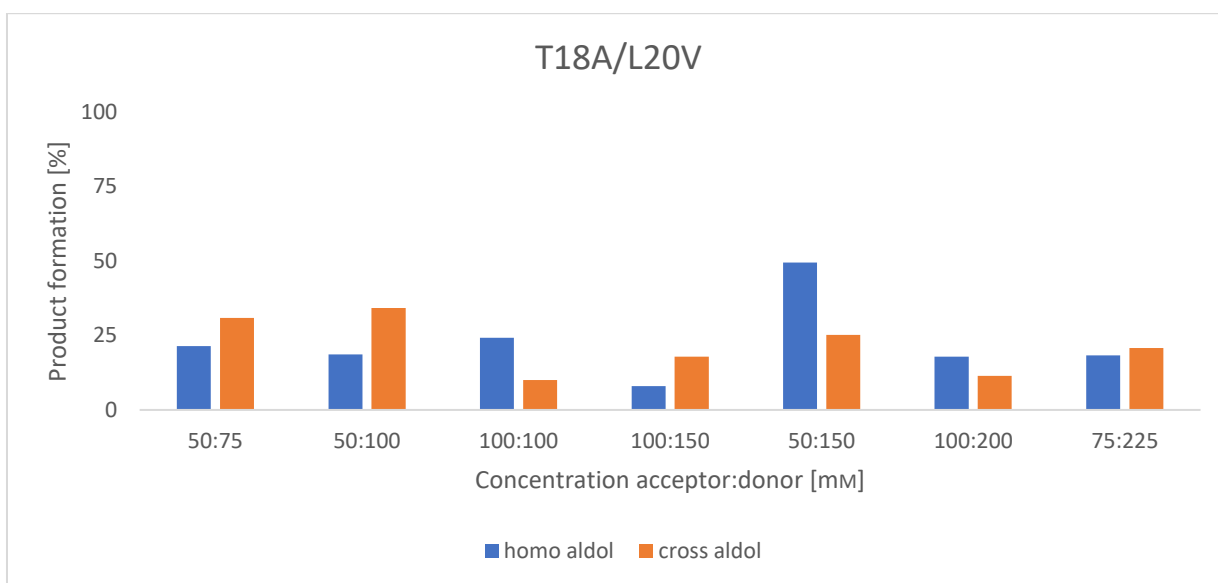
Figure 17 HPLC analysis of the cross-aldol reaction of acetone (**50**) and *n*-butanal (**53**) performed by mutant T18A/L20V with varying donor/acceptor concentrations.
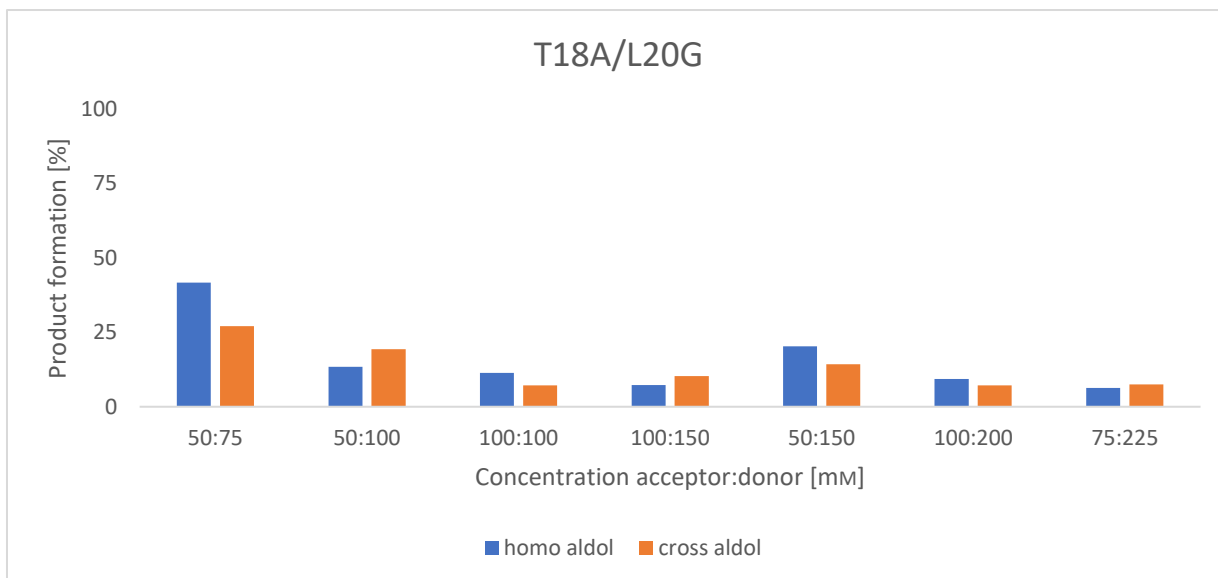
Figure 18 HPLC analysis of the cross-aldol reaction of acetone (**50**) and *n*-butanal (**53**) performed by mutant T18A/L20G with varying donor/acceptor concentrations.
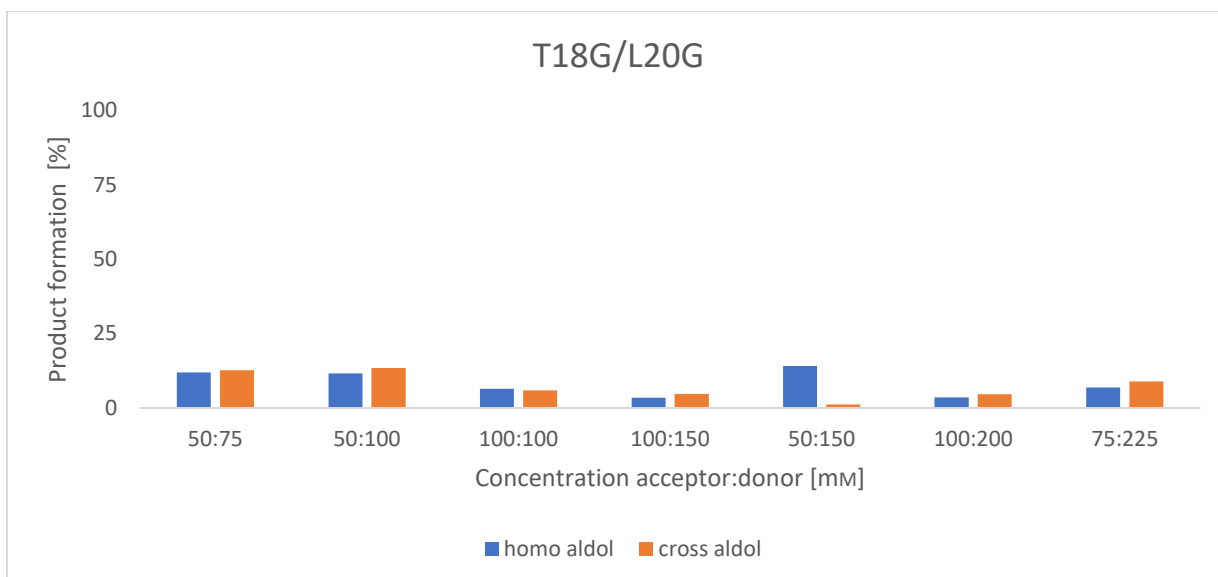
Figure 19 HPLC analysis of the cross-aldol reaction of acetone (**50**) and *n*-butanal (**53**) performed by mutant T18G/L20G with varying donor/acceptor concentrations.

The results indicate at first glance that the enzymes seem to be more productive with lower substrate loadings, while increasing the substrate loading leads to lower product formation. Considering that both substrates have a low boiling point, the true product formation will be higher as seen in the experiment before, when referenced to the unreacted substrate amounts.

The ratios of cross- to homo-aldol product demonstrate again that the glycine variants (T18A/L20G) and (T18G/L20G) both show a preference for the cross-aldol product. Interestingly, a small excess of donor (100:150) yields the highest ratios for (T18A/L20V), (T18A/L20G) and (T18G/L20G), while only the reference DERA-variant T18A/L20A shows the highest ratio for a 50:150 condition. Comparing the obtained ratios of cross to homo aldol product for both reference DERA-variants (T18A/L20A+V), one can say that (T18A/L20A) is more productive in the homo-aldol reaction.

The newly created DERA-variants (T18G/L20G) and (T18A/L20G) both show a clear preference for the cross-aldol product under equimolar conditions, resulting in a ratio of 1.8 and 1.3. Therefore, both mutations seem to lead to higher usage of ketones over aldehydes. Both reference DERA-variants showed a slight preference for the homo-aldol product with equimolar conditions.

One fact is noticeable, that (T18A/L20G) showed a switch in product formation ratios: 61% : 39% homo-aldol product to cross-aldol product versus 38% : 62% observed in the triplicate measurement. The shift may be explained by high standard errors of the HPLC measurements (mostly by sample preparation with derivatization) but then again questioning the general outcome of the conducted experiments. Due to the nature of a single measurement, of course the data could be an outlier or caused by a handling error.

The results of the experiments still demonstrate a clear tendency, which was desired for the planning of further experiments but due to relative inconsistencies should be interpreted with care. Since the glycine-variants performed well in the cross-aldol reaction with acetone (**50**), the next step clearly was to test for cyclic donors.

4.2.3 Characterization of the cross-aldol addition activity with cyclobutanone (57), cyclopentanone (58) and *n*-butanal (53)

From the literature it is known, that DERA in general has a more relaxed donor scope when compared to other aldolases (see Introduction). Cyclobutanone (**57**) and cyclopentanone (**58**) are known to function with different DERAs from manifold sources including reports from our group[39, 124-125]. Therefore, one aim of the created library was to utilize cyclopentanone (**58**) and cyclobutanone (**57**) as donors. The reactions were performed analogously to those of the cross-aldol reaction between *n*-butanal (**53**) and acetone (**50**):

- *n*-Butanal (**53**) 100 mм
- Donor (cyclopentanone (**58**) or cyclobutanone (**57**)) 100 mм
- TEA*HCl buffer, pH 7.5, 50 mм
- DMSO 20% v/v
- Purified DERA 3.5 mg/mL

As reference the DERA-variants (T18A/L20A+V) were chosen. Reactions were analyzed after 24h via TLC and for positive hits additionally by HPLC. From the first generation, it was known that cyclopentanone (**58**) is an accepted donor, while cyclobutanone (**57**) is not. Therefore, a positive result for cyclobutanone (**57**) would be an important achievement. Unfortunately, the glycine variants also showed no significant product formation (traces) to be used for synthetic purposes. The variants (T18A/L20V) and (T18G/L20G) both showed product formation for cyclobutanone (**57**) in trace amounts on TLC. Analyzing those samples by HPLC yielded non-detectable product quantities under the applied conditions.

The TLC analysis of cyclopentanone (**58**) samples however showed clear product formation for the variants (T18A/L20A), (T18A/L20V), (T18A/L20G) and (T18A/L20G). The results are shown in Figure 20–23:
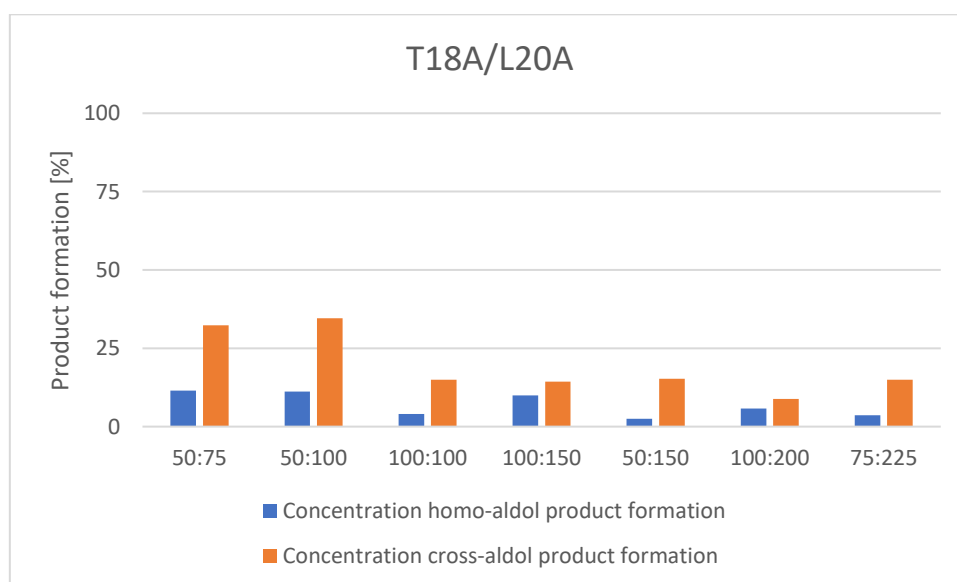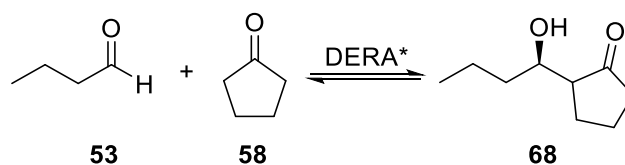
Figure 20 HPLC analysis of the cross-aldol reaction between cyclopentanone (**58**) and *n*-butanal (**53**) performed by mutant T18A/L20A with varying donor/acceptor concentrations.
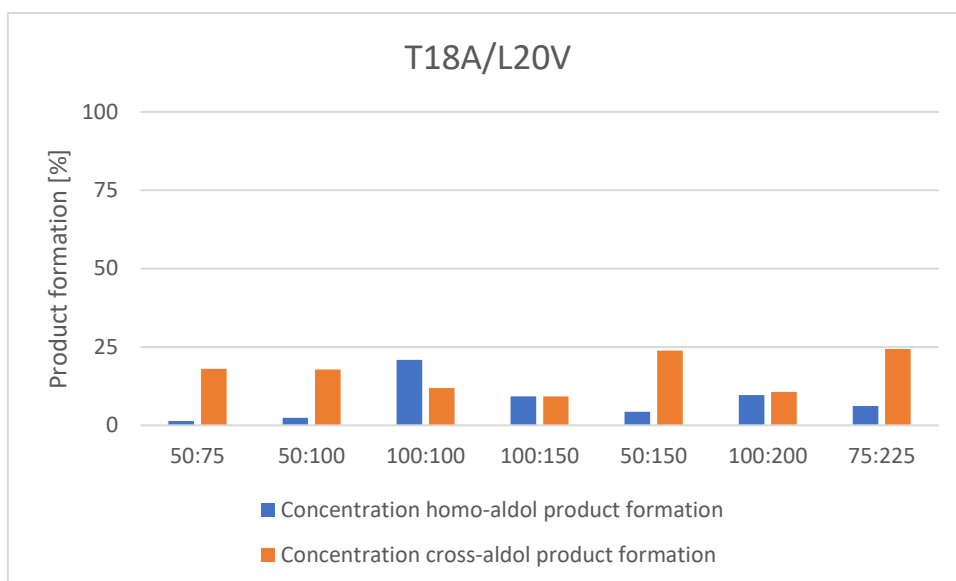
Figure 21 HPLC analysis of the cross-aldol reaction between cyclopentanone (**58**) and *n*-butanal (**53**) performed by mutant T18A/L20V with varying donor/acceptor concentrations.
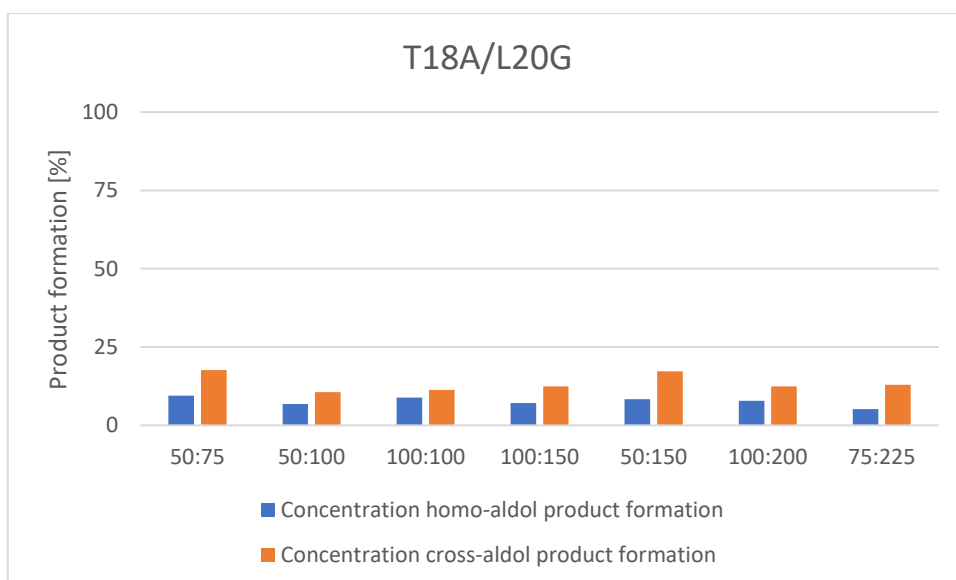
Figure 22 HPLC analysis of the cross-aldol reaction between cyclopentanone (**58**) and *n*-butanal (**53**) performed by mutant T18A/L20G with varying donor/acceptor concentrations.
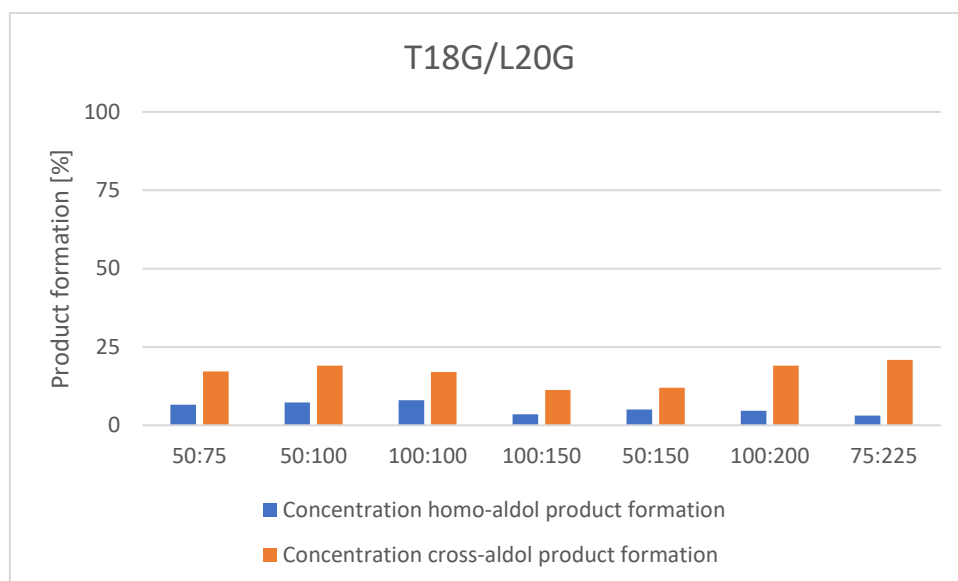
Figure 23 HPLC analysis of the cross-aldol reaction between cyclopentanone (**58**) and *n*-butanal (**53**) performed by mutant T18G/L20G with varying donor/acceptor concentrations.

One can conclude, that the reference DERA-variants (T18A/L20A+V) both performed superiorly under the tested conditions compared to DERA variants containing glycine mutations. The highest product formation for cross-aldol product was detected for (T18A/L20A) when utilizing 50 mM acceptor and 75/100mm donor concentrations. Amazingly, the same mutant displays the lowest cross-aldol activity when utilizing 100 mM acceptor to 200 mM donor concentration. Since the experiment was conducted in unicates, this effect can be the result of a single measurement error. (T18G/L20G) and (T18A/L20G) both yielded product formation between 10–21% for the cross-aldol product.

The DERA-variant (T18A/L20V) had been studied by the author previously during his master-thesis for cross aldol reaction of *n*-butanal (**53**) (100mM) with cyclopentanone (**58**) (150mM) as a donor, yielding 8% of isolated product[124]. Comparing the isolated yield with the product formation from the above-mentioned experiment, both experiments result in similar conversions in the cross-aldol reaction of *n*-butanal (**53**) with cyclopentanone (**58**) as a donor. Therefore, at least the measurement of that sample seems to fit well to known results.

Interestingly, only for one DERA-variant (T18A/L20V), a negative ratio of cross/homo-aldol was detected for equimolar conditions of 100mM acceptor to 100mM donor. All other tested DERA-variants used cyclopentanone (**58**) favorably in the cross-aldol reaction rather than producing the *n*-butanal (**53**) homo-aldol product even upon equimolar conditions. One can say that the glycine mutations at position T18 and L20, both increase the substrate scope towards bulkier donors. Disappointingly, the glycine-variants present lower overall activities under the tested conditions compared to (T18A/L20A+V). From the data no final interpretation is possible, whether the glycine mutations favor ketones over aldehydes.

As a summary, two DERA-variants were newly identified with increased substrate scope comparable to the variants selected from the first generation. (T18G/L20G) and (T18A/L20G) both are able to utilize *n*-butanal (**53**) and cyclopentanone (**58**) as unnatural donors.

Compared to the reference variants from the first generation, both glycine variants demonstrate lower activity (10-20% product formation) but still in synthetically acceptable parameters. When increasing the ratio of donor to acceptor, the detected product formation is increasing. Taking into account that with higher substrate loadings of volatile substances, the real product formations rank probably higher than the values reported due to evaporation of educts. Therefore, based on the obtained data a 2-3fold ratio of donor to acceptor seems the best option in general.

The DERA-variants (T18G/L20A), (T18G/L20V) and (F76G) all were able to catalyze the homo-aldol reaction n-butanal (**53**) but with insufficient activities for synthetic usage and were therefore not selected for further analysis.

4.2.4 Characterization of the thermostability of DERA-variants

As will be discussed in Chapter 3, nano differential scanning fluorimetry is a powerful tool for quick determination of protein thermostability[87]. With this method the intrinsic protein fluorescence is measured during thermal heating. By plotting the ratio of 350/330nm wavelength, the inflection point is reflecting the melting point (T_m) of a protein. At this temperature, the protein is half folded and half unfolded. Since thermostability is related to chemostability, the value gives direct indication how stable a protein is. The conditions of the T_m-measurement were following:

- TEA*HCl buffer, pH 7.5
- 20% DMSO
- 1.75 mg/mL enzyme

The measurement was conducted in ultra-thin capillaries with a volume of 10 µL. During the measurement, the capillary is heated with a temperature gradient of 1°C/min. The measurement was conducted in triplicates, even though the measurement is extremely precise. The results of the experiment are listed in Table 4:

Table 4 Thermostability assessment of first-generation mutants compared to the T18A/L20A mutant and the initial library *Ec*DERA origin clone F200I.

DERA-Variant	T_m in °C
F200I	62.4 ± 0.1
F200I (no DMSO)	63.1 +0.3
T18A/L20A	60.3 ± 0.1
T18A/L20G	55.1 ± 0.1
T18G/L20G	55.2 ± 0.5
T18G/L20A	57.3 ± 0.2
T18G/L20V	58.0 ± 0.2
F76G	57.3 ± 0.2

The data of the T_m measurement produced clear signals for all measured samples. The standard errors of that experiment account to less than 1%. Comparing the resulting T_m values of the first-generation DERA variants, it is clear that all point mutations lead to a loss of thermostability since the T_m values are lower than compared to the precursor DERA F200I. Especially the glycine variants have significantly lower T_m-values accounting up to 7°C compared to F200I. When mutating an amino acid residue towards glycine, which in fact has no side chain, will lead to extra space or different folding modes. One possible compensation for those resulting cavities, is placing water molecules in the cavity, which will of course alter the strength of the bonds surrounding that area. This usually leads to lower thermostability but of course can be vice versa. The measured T_m values all remain between 55-63°C with addition of 20% DMSO and all were therefore deemed useful under synthetic conditions.

4.3 Cloning of the second-generation library of *Ec*DERA

After successfully generating and characterizing the supplementary DERA-variants with glycine mutations, the next step was the combination of beneficial mutations by creating a second-generation library. The second generation should therefore combine mutations at position T18, L20 and F76.

For the generation of the second-generation library, the strategy was based on the method for generating the supplementary glycine variants. This called for first randomizing the position F76 towards a set of smaller amino acids glycine, alanine, valine and serine, followed by the deletion of 9 nucleotides coding for T18-L20. Finally, the randomized nucleotides coding for the desired amino acids needed to be inserted. The same rules for primer design apply from Section 3.1. The codons for the mutagenesis positions were optimized for serine, valine and alanine, while for glycine such an optimization was not possible and the same primers were taken as in Section 3.1.

All second-generation mutagenesis primers with their encoded amino acids and respective universal codons are listed in Table 5:

Table 5 Second-generation mutagenesis primers and their encoded amino acids.

Primer	Function	Encoded amino acids	Universal mutagenesis codon[a]
DERA76-KYC_f	Forward primer	Phe76 → Ala, Ser, Val, Phe	KYC
DERA-F76G_f	Forward primer	Phe76 → Gly	GGC
DERA76-KYC_r	Reverse primer	Non-mutagenic	-
DERA18-20ins_f	Forward primer	Leu20 →Ala, Ser, Val, Leu	KYG
DERA-L18/20G_f	Forward primer	Leu20 →Gly	GGC
DERA18-20ins_r	Reverse primer	Thr18 → Ala, Ser, Thr	DCC
DERA-L18/20G_r	Reverse primer	Thr18 → Gly	GGC

[a] K stands for nucleotides G/T, Y for C/T and D for A/G/T.

For the randomization of position F76 via PCR, the forward mutagenic primers DERA76-KYC_f and DERA-F76G_f were mixed in a ratio of 4:1 to yield a statistically equal ratio for the mutagenesis. The DERA76-KYC_f encodes for the amino acids alanine, serine, valine and phenylalanine, while DERA-F76G_f encodes for glycine only. After the PCR, the plasmids were amplified in DH5α cells and random clones (200 clones per library) were picked from the transformation plates, their plasmids were isolated and sent for sequencing. During sequencing one desired mutant (F76A) could not be identified and was therefore taken from an earlier experiment. For confirmation, this sample was also sequenced to prevent time

consuming mistakes. The alignment of the identified clones i.e., (F76A), (F76V), (F76S) and (F76G), is shown in Figure 24.

```
F76G_clone8_KYCG<    GCAAAACTCTGAAAGAGCAGGGCACCCCGGAAATCCGTATCGCTACGGTAACCAACGGCC    298
F76S_clone3_KYCG<    GCAAAACTCTGAAAGAGCAGGGCACCCCGGAAATCCGTATCGCTACGGTAACCAACTCCC    299
F76A_E6<             GCAAAACTCTGAAAGAGCAGGGCACCCCGGAAATCCGTATCGCTACGGTAACCAACGCCC    300
WT_sequence<         GCAAAACTCTGAAAGAGCAGGGCACCCCGGAAATCCGTATCGCTACGGTAACCAACTTCC    229
F76V_clone1_KYCG<    GCAAAACTCTGAAAGAGCAGGGCACCCCGGAAATCCGTATCGCTACGGTAACCAACGTCC    299
                     ********************************************************  **
```

Figure 24 Alignment of the successfully cloned (F76G), (F76S) and (F76V) mutants with the selected (F76A) mutant and the *Ec*DERA gene WT sequence. The mutation site is marked with a red arrow (nucleotides 226-228 of the WT sequence). All mutated codons A(GCC), V(GTC), S(TCC) and G(GGC) are shown.

The next step for generation of the second-generation library was the deletion of the nine nucleotides coding for positions T18-L20. For convenience, four different libraries were generated, based on mutants (F76A), (F76G), (F76S) and (F76V). The combination of each library containing 88 randomly selected clones, results in a total of 352 randomly selected clones overall. Since the libraries will be screened by TLC and positive hits by HPLC, the screening effort was expected to be time demanding.

The mutagenesis step itself, the deletion and insertion of nine nucleotides coding for T18-L20, were performed analogously to Section 3.1. The primers DERA18-20del_f and DERA18-20del_r could be used universally for the deletion of T18-L20 and were used in this mutagenesis step. The mutants (F76A/V/S/G) were mutated via PCR, their plasmids amplified in DH5α cells, clones randomly picked from the transformation plates, their plasmid isolated and sent for sequencing. The alignment of the selected clones from the deletion step is shown in Figure 25.

```
F76A_T18/L20Del12<   GGACCT----------GATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCA    97
WT_sequence<         GGACCTGACCCACCCTGAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCA    104
F76V_T18/L20Del11<   GG---------ACCTGAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCA    107
F76G_T18/L20Del14<   GG---------ACCTGAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCA    111
F76S_T18/L20Del13<   GG---------ACCTGAATGACGACGACACCGACGAGAAAGTGATCGCCCTGTGTCATCA    110
                     **              ***************************************
```

```
F76A_T18/L20Del12<   CGCCCCACACGGTAACGACGACATCGACATCGCGCTGGCAGAAACCCGTGCGGCAATCGC    277
WT_sequence<         CTTCCCACACGGTAACGACGACATCGACATCGCGCTGGCAGAAACCCGTGCGGCAATCGC    284
F76V_T18/L20Del11<   CGTCCCACACGGTAACGACGACATCGACATCGCGCTGGCAGAAACCCGTGCGGCAATCGC    287
F76G_T18/L20Del14<   CGGCCCACACGGTAACGACGACATCGACATCGCGCTGGCAGAAACCCGTGCGGCAATCGC    291
F76S_T18/L20Del13<   CTCCCCACACGGTAACGACGACATCGACATCGCGCTGGCAGAAACCCGTGCGGCAATCGC    290
                     *  ********************************************************
```

Figure 25 Alignment of the successful T18/L20 deletion clones of (F76G), (F76S), (F76A) and (F76V) mutants with the *Ec*DERA gene WT sequence. The mutation site is marked with a red arrow, the deleted 9 nucleotides are circled in red. The position of F76 is also shown below.

Due to a program bug and the presence of similar nucleotides on both ends of the deleted sequences, a group of five nucleotides (ACCTG) appears shifted for the clones (F76G), (F76S) and (F76V). Interestingly, for (F76A) the deletion step resulted in the deletion of 10 instead of 9 nucleotides. The deletion step was deemed successful, even though some steps had to be repeated due to unwanted point mutations in the gene frame or incomplete deletions in the deletion site. For the clone (F76A) it was assumed that the insertion step should not yield a frameshift and the deletion site should be regenerated with the desired nucleotides during the insertion step.

After finalization of the deletion step and characterization of the resulting clones, the next step was the insertion of randomized nucleotides for the positions T18-L20. Since the forward and reverse primers DERA18-20ins_f and DERA18-20ins_r do not include glycine as a possible outcome, the primers DERA-L18/20G_f and DERA-L18/20G_r were mixed 1:3 with the before mentioned primer pair to yield a statistically equal outcome of the possible amino acids. All mutagenesis steps were performed as before, resulting in the desired libraries in form of colonies of *E. coli* DH5α transformants on agar plates. For a confirmation of the successful mutagenesis, 10 randomly picked clones were cultivated, their plasmids isolated and sent for sequencing to check if any clone was dominant. The results showed that 8/10 clones had different mutations and therefore the insertion step was deemed successful.

The plates containing the four different libraries, based on (F76A/G/S/V), were scraped into 50 mL LB-media and cultured for 18h at 37° C. During that process all possible clones were cultured as a mixture. From this culture, the plasmids were isolated, resulting in a plasmid mix carrying all possible mutation outcomes. Afterwards, *E. coli* BL21 expression strains were transformed with those plasmids creating the final libraries. From the resulting plates 88 clones per library were randomly selected, cultured and stored in 96-well plates for further usage. In each 96-well plate, eight controls were implemented in column 12 row. A/B wells were used for the background protein control with cells carrying an empty BL21 plasmid, the negative reference variant F200I ("wildtype") was placed in C/D, the positive control mutant FSA D6A/T26I, which is able to catalyze *n*-butanal (**53**) homo-aldol reaction, from the work of Sebastian Junker[110] was placed in E/F, while G /H were used for chemical background. The layout of the 96-well plates is shown in Figure 26:

Library F76X	1	2	3	4	5	6	7	8	9	10	11	12
A												EP
B												EP
C												F200I
D	Randomly selected 88 clones											F200I
E												FSA D6A T26I
F												FSA D6A T26I
G												Blank
H												Blank

Figure 26 Exemplary layout of the 96-well plates used for the respective libraries. Column 12 was used as a control column, having each control as a duplicate. EP stands for empty plasmid, F200I – for the original starting clone of the library, FSA – a Fructose-6-phosphate aldolase variant and the Blank for the chemical background control.

For further usage, the plates were copied and stored with 20% v/v glycerol in each well at –80 °C. Glycerol was utilized to prevent damages during freezing/thawing cycles[126].

4.4 Primary screening of generated libraries for activity in homo- and cross-aldol additions utilizing *n*-butanal (53), clone selection and sequencing

After finalization of the second-generation libraries in form of 96-well MTPs, the next step was an initial screening of the libraries for positive hits in the *n*-butanal (53) homo aldol addition as well as cross-aldol addition of *n*-butanal (53) and acetone (50). It was decided to categorize the outcome of those screenings in categories of 0 for negative hits and positive hits from (+) – (+++). Depending on the number of hits, only the category (+++) would be further analyzed via HPLC and in case of fewer hits also the category (++) samples for broader coverage. Basically, the same procedure as in Section 3.2 was utilized in the screening following the same rules, with one exception being the biocatalyst itself. In Section 3.2 purified enzyme was utilized, but this procedure would be way too time consuming for the screening of the larger generated libraries. Therefore, it was decided to utilize whole cell catalysis instead. After culturing of the cells in AIM in MTP format, the cell suspension was centrifuged, the plate washed with 50 mM TEA*HCl buffer (pH 7.5) and after one round of centrifuging and removal of the washing buffer, the cells were suspended again in 200 µL TEA*HCl buffer (pH 7.5) and used to screen for positive hits. Alternatively, hypotonic extraction[87] or lysis extraction could be utilized but these methods are indeed more tedious and time consuming to perform.

The standard reaction conditions were utilized from Section 3.2 (50 mm TEA*HCl buffer pH 7.5, DMSO 20% v/v) with both acceptor and donor concentration at 100 mm. From the 200 µL of cell suspension, 150 µL were taken for the aldol reactions. The reactions were run overnight and analyzed the next day. The analysis was performed via TLC (1 µL/sample with anisaldehyde stain) and for positive hits via HPLC for quantification. For a first selection, the above mentioned categories from (0)–(+++) were utilized, which are of course rather subjective. As an example, the results of the (F76V) library are shown in Figure 27.
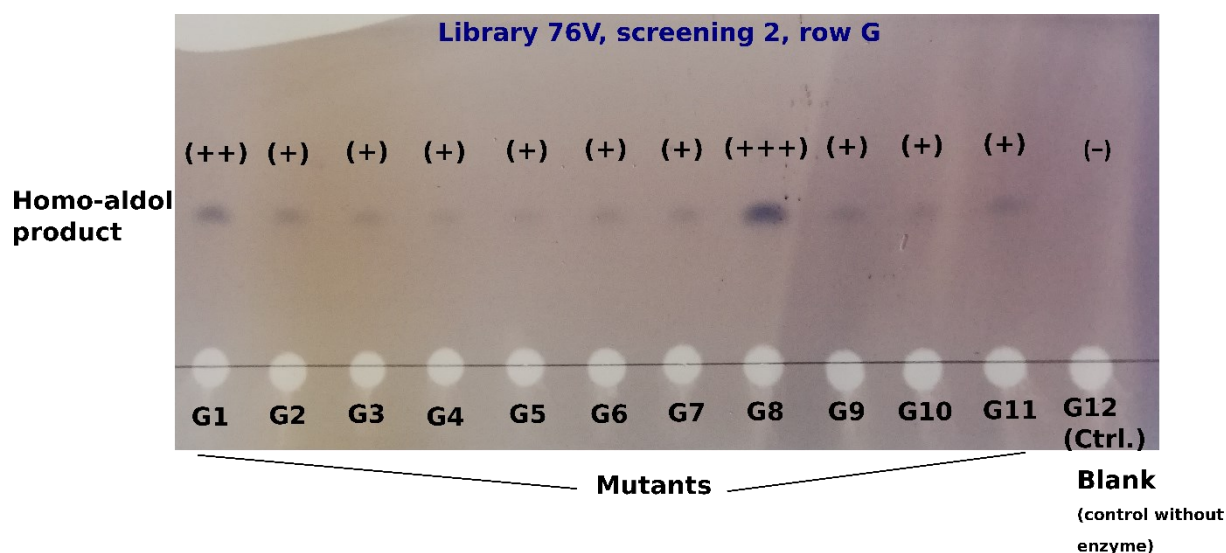
Figure 27 Exemplary TLC screening of the (F76V) library for the homo-aldol addition of *n*-butanal (**53**) depicting the product band intensity gradation from (+) as low product formation to (+++) for high product formation. The product bands were stained with 4-anisaldehyde. Eluent: cyclohexane:ethylacetate 2:1.

As mentioned earlier, M. Sc. Feodor Belov performed the screening and the decision making on categorizing the different variants, whereas borderline cases were decided accordingly to a four eyes principle. In general, categorizing TLC spots visually always involves a larger human variance and most likely will also differ from day to day. Thus, the qualitative results presented below are just a rough estimate. To minimize the error, the staining solution was always freshly prepared because impurities in the methanol solvent of the anisaldehyde stain, result in background staining of the reagent. Also, a defined volume of 1 µL was spotted for each reaction to produce more constant results. The reaction mixture itself was also produced from stocks containing all reagents except the whole cell, to prevent variance of the catalytic reaction conditions. In general, the screening via TLC will only lead to products being visible from volatile educts like *n*-butanal (**53**) or acetone (**50**), which will both evaporate during the drying of the TLC plates before the staining is performed.

The homo-aldol screening with *n*-butanal (**53**) was performed in duplicate to ensure sufficient data coverage for assessment. The product was easily detectable after staining with anisaldehyde due to educts being easily evaporated during heating of the TLC plate with a heat gun. Exemplary duplicate results for the (F76G) library are shown in Figure 28. The complete screening results are shown as a grid in the supplementary information.

Library F76G												
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	+	+	+	+	+++	+	+	-	-
B	-	-	+	-	+	+	+	++	++	-	-	-
C	-	+	+	+	+	+++	+++	++	++	+	+	+
D	-	-	+	-	+	+	+	+	-	-	-	+
E	+	+	+	++	+	++	+	+	++	++	+	+
F	-	-	-	-	+	-	-	-	-	+	+	+
G	-	-	-	-	-	++	+	+	+	-	+	-
H	++	+	+	+	+	+	++	++	+	++	++	-

	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	+	+	+	+	+++	+	++	+	-
B	+	+	+	+	+	+	+	+++	+++	+	+	-
C	+	+	++	++	+	+++	+++	+	+	+	+	+
D	-	-	-	+	++	++	+	+	+	+	+	+
E	+	+	+	++	+	+++	++	++	++	+++	+	++
F	-	-	-	-	-	+	+	++	-	++	+++	++
G	+	-	+	-	-	++	+	+	++	-	++	-
H	-	++	+	+	++	+	+	+	-	+	+	-

Figure 28 Exemplary results of the homo-aldol screening with *n*-butanal shown in duplicate for the library (F76G). Color code indicating quality of resulting hits from red for negative hits, pale green for traces, green for medium intensity product spots and dark green for strong hits.

From those screenings, all samples reaching the level (+++) at least once and once (+) or (++) were selected for further characterization. Multiple clones showed clear (+++)-intensities. The selected variants are summarized in Table 6:

Table 6 Mutant selection for the homo-aldol screening of the second-generation libraries. The clones are labeled by their respective positions on the library plates.

Library	Selected mutants
F76G	A8, B8, B9, C6, C7, E6, E10, F11
F76S	A2, A5, A7, B6, B11, C3, D10, E7, F8, F9, G5
F76V	A4, C1, D1, E5, E10, F4, G1, G8, H5
F76A	A8, C5, C8, D8, G11, H2

As a consequence, between 7% (F76A-library) and up to 13% (F76S) library of the selected clones were chosen for further analysis.

Next, the screening towards cross-aldol reaction with *n*-butanal (**53**) as acceptor and acetone (**50**) as donor was performed. Important to remember is that acetone (**50**) is a naturally

accepted donor of *Ec*DERA and *n*-butanal (**53**) that can act as both the donor and acceptor. Therefore, both the homo-aldol product of *n*-butanal (**53**) and the cross-aldol product with acetone (**50**) can be formed. For the screening, two different conditions were chosen. The first experiment was conducted with equimolar conditions at 100 mM each of donor and acceptor, while the second conditions contain 100 mM *n*-butanal (**53**) and 10% v/v acetone (**50**). Again positive hits were categorized between (0)–(+++) based on the cross aldol product staining on TLC with anisaldehyde as staining reagent, while provisionally ignoring the outcome of the homo aldol reaction. Both experiments were performed in unicates, and the results are presented in Figure 29.

Library F76G												
	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	+	-	+	-	+++	-	-
B	-	-	-	-	-	-	-	+++	+++	-	-	-
C	-	++	-	+++	-	+	+++	-	-	-	-	+
D	-	-	-	-	+	-	-	-	-	-	-	+
E	++	++	-	+	-	+++	++	+	+	+++	+	+++
F	-	-	-	-	-	-	-	+++	-	++	+++	+++
G	-	-	-	-	-	+	-	-	+	-	+++	-
H	+	+++	-	-	-	++	-	-	-	++	-	-

Library F76G with 10 % acetone												
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	+	+	+	+	++	+	+++	+	+
B	+	+	+	+	+	+	+	+++	+++	+	+	+
C	+	++	+	+++	+	++	+++	+	+	+	++	+
D	+	+	+	+	++	+	+	+	++	+	++	+
E	++	+	+	++	+	+++	+	+	+	+++	+	+++
F	+	+	+	+	+	+	+	+++	+	+++	+++	+++
G	+	+	+	+	+	+++	+	+	+++	+	+++	-
H	+	+++	+	+	+	+	+	+	+	++	+	-

Figure 29 Exemplary results of cross-aldol screening with acetone (**50**) as donor and *n*-butanal (**53**) as acceptor for the library (F76G). The upper table depicts the screening results under equimolar donor:acceptor concentrations, while the lower table shows the screening with 10 % (v/v) of acetone (**50**).

With equimolar conditions of 100 mM each donor and acceptor, over 13 DERA variants of the exemplary F76G-library produced spots were categorized as (+++) and 6 variants categorized as (++). The majority of DERA variants presented very faint spots (+) or none at all. By increasing the ratio of donor to acceptor the majority of negative hits now showed faint spots (+). DERA variant F10 improved from (++) to (+++) due to the higher ratio of donor to acceptor, while the other remained in the same category even though a few DERA variants lost activity

due to the higher acetone (**50**) concentration. All variants that were categorized as (+++) in at least one of both experiments were selected for sequencing and are presented in Table 7.

Table 7 Mutant selection for the cross-aldol screening of the second-generation libraries with acetone (**50**) as donor. The clones are labeled by their respective positions on the library plates. Cross-aldol selective clones not present in the previous homo-aldol selections are marked in bold and underlined.

Library	Selected mutants
F76G	**A10**, B8, B9, **C4**, C7, E6, E10, **F8**, **F10**, F11, **G6, G9, G11, H2**
F76S	A2, A7, **B3**, B6, B11, C3, **D2**, D10, E7, F8, G5, **G8, H10**
F76V	A4, **B9, B11,** C1, **C10, C11, D11**, E5, E10, F4, **G9**, H5
F76A	A8, **B3**, C8, **D10, F5**, G11, H2, **H11**

4.4.1 Sequencing of selected second-generation library hits

As a summary of the results from the *n*-butanal (**53**) homo aldol and *n*-butanal (**53**)/acetone cross aldol reaction screening, a total of 56 clones classified for the selection. All were separately cultured, their plasmids isolated and sent for sequencing. From that screening 23 unique clones were identified and prepared in MTP-format for further usage. The layout of that screening master plate is shown in Figure 30.

	1	2	3	4	5	6	7	8	9	10	11	12
A	**76G** **A8** **T18S** **L20G**	**76G** **B8** **T18** **L20**	**76G** **G6** **T18S** **L20**	**76S** **A2** **T18** **L20**	**76S** **A5** **T18** **L20V**	**76S** **A7** **T18S** **L20**	**76S** **F9** **T18G** **L20G**	**76S** **B2** **T18S** **L20G**	**76V** **A4** **T18** **L20**	**76V** **D1** **T18S** **L20G** **N114D**	**76V** **G1** **T18S** **L20G**	**76V** **B11** **T18A** **L20**
B	**76V** **C10** **T18S** **L20**	**76V** **C11** **T18G** **L20V**	**76V** **D6** **T18G** **L20**	**76V** **D11** **T18G** **L20A**	**76V** **F7** **Deletion** **mutant**	**76A** **A8** **T18** **L20**	**76A** **C5** **T18S** **L20G**	**76A** **A9** **T18** **L20A**	**76A** **B7** **T18** **L20G**	**76A** **D9** **T18G** **L20G**	**76A** **H11** **T18S** **L20**	**Blank**
C	76G A8 T18S L20G	76G B8 T18 L20	76G G6 T18S L20	76S A2 T18 L20	76S A5 T18 L20V	76S A7 T18S L20	76S F9 T18G L20G	76S B2 T18S L20G	76V A4 T18 L20	76V D1 T18S L20G N114D	76V G1 T18S L20G	76V B11 T18A L20
D	76V C10 T18S L20	76V C11 T18G L20V	76V D6 T18G L20	76V D11 T18G L20A	76V F7 Deletion mutant	76A A8 T18 L20	76A C5 T18S L20G	76A A9 T18 L20A	76A B7 T18 L20G	76A D9 T18G L20G	76A H11 T18S L20	Blank
E	76G A8 T18S L20G	76G B8 T18 L20	76G G6 T18S L20	76S A2 T18 L20	76S A5 T18 L20V	76S A7 T18S L20	76S F9 T18G L20G	76S B2 T18S L20G	76V A4 T18 L20	76V D1 T18S L20G N114D	76V G1 T18S L20G	76V B11 T18A L20
F	76V C10 T18S L20	76V C11 T18G L20V	76V D6 T18G L20	76V D11 T18G L20A	76V F7 Deletion mutant	76A A8 T18 L20	76A C5 T18S L20G	76A A9 T18 L20A	76A B7 T18 L20G	76A D9 T18G L20G	76A H11 T18S L20	Blank
G	EP	EP	EP	F200I	F200I	F200I	FSA D6A T26I	FSA D6A T26I	FSA D6A T26I	Blank	Blank	Blank
H	EP	EP	EP	F200I	F200I	F200I	FSA D6A T26I	FSA D6A T26I	FSA D6A T26I	Blank	Blank	Blank

Figure 30 Layout of the selection master plate after primary homo- and cross-aldol screenings. Rows G and H are used for controls. All selected clones are present in triplicates, always distributed in two continuous rows. Hence, one unicate is marked in orange. The EP slot stands for empty plasmid, F200I – for the original starting clone of the library, FSA – a Fructose-6-phosphate aldolase mutant and the Blank for the chemical background control.

As can be seen from the results, four selected clones possess no mutation at the positions T18 and L20. This can be explained by the primer design, which also encodes for the native amino acids to be able to get all possible mutation combinations. Interestingly, one of the selected mutants presented no insertion sequence in the T18-L20 region, but due to its activity during the experiment it was kept in the selection. With the selection finalized, the screening for bulkier ketone donors was the next part.

4.5 Screening of the selected variants for activity in cross-aldol reactions with novel donors

With the second-generation library finalized and the primary screening of the library in the test reactions, the next step was the screening for bulkier cyclic donors. From the literature it is know that cyclopentanone (**58**) can be a good substrate for DERA libraries[38-40]. The selection of 23 DERA-variants was tested in the cross-aldol reaction of *n*-butanal (**53**) as acceptor with the following cyclic donors: butanone (**55**), cyclobutanone (**57**), 3-oxetanone, cyclopentanone (**58**), dihydrofuran-3(2H)-one (**63**), cyclohexanone (**61**) and tetrahydro-4H-pyran-4-one (**64**).

Whole cell catalysis was again applied due to the excellent results in the previous screening of the DERA libraries. The reactions were performed as triplicates in MTP format in 200 µL of overnight culture resulting in a lower overall error rate due to homogenous handling of the samples. In all cases equimolar conditions were utilized amounting to 100 mM per acceptor and donor with 20% v/v DMSO and 50 mM TEA*HCl buffer (pH 7.5).

4.5.1 Screening of second-generation library with cyclobutanone (57) as the aldol donor

From previous experiments[124] it was known that cyclopentanone (**58**) can be used as donor for the first-generation library. Therefore, the expectation was that cyclobutanone (**57**) will possibly be accepted by some variants as donor, which was not the case for the first generation. An exemplary TLC-result of the cross-aldol reaction with cyclobutanone (**57**) as donor and *n*-butanal (**53**) as acceptor is presented in Figure 31.
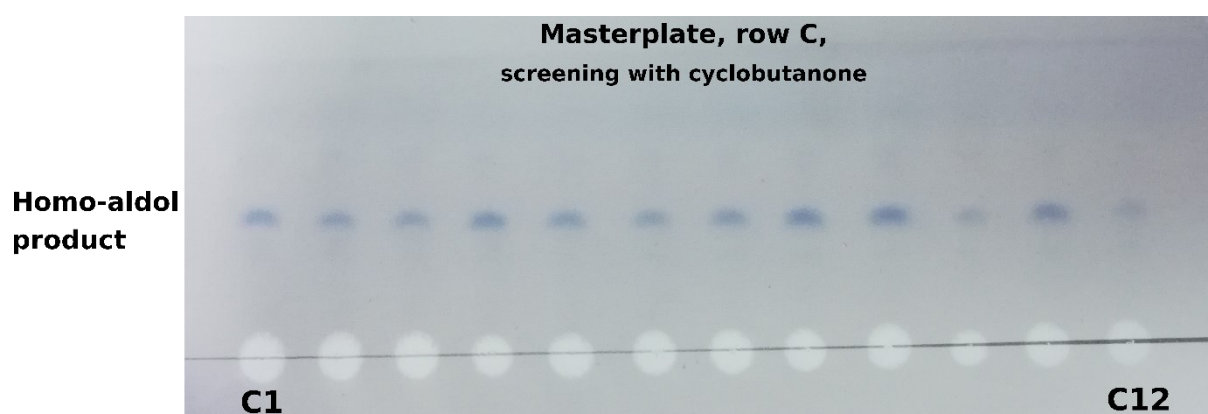
Figure 31 Screening results with cyclobutanone (**57**) as donor for the cross-aldol reaction. As can be seen, only the homo-aldol product is formed from *n*-butanal (**53**). Eluent: cyclohexane:ethylacetate 2:1.

Interestingly, cyclobutanone (**57**) was not utilized by the DERA-variants as donor for the cross-aldol reaction with *n*-butanal (**53**). The DERA-variants were active, at least in the homo aldol reaction, since the product is clearly detectable on TLC after staining with anisaldehyde, indicating that none of the variants is able to accept cyclobutanone (**57**) as donor.

4.5.2 Screening of second-generation library with cyclopentanone (58) as the aldol donor

Even though cyclobutanone (**57**) was tested unsuccessfully, cyclopentanone (**58**) was tested next with high expectations due to the experience with the first-generation library. The results of the screening for cyclopentanone (**58**) as donor with *n*-butanal (**53**) as acceptor are presented below.

A.

B.

	1	2	3	4	5	6	7	8	9	10	11	12
A	+++	+	+	+	+	+	+	++	+	+	+	+
B	+	+	+	+	+	++	+++	+	++	+++	++	+
C	+++	++	++	++	+	+	+	++	+	+	+	+
D	+	+	+	+	+	++	+++	+	++	+++	++	+
E	+++	++	+	++	+	+	++	++	+	+	+	+
F	+	+	+	+	+	++	+++	+	++	+++	++	+
G[a]	+ EP	+ EP	+ EP	+ F200I	+ F200I	+ F200I	++ FSA	++ FSA	+++ FSA	+ Blank	+ Blank	+ Blank
H[a]	+ EP	+ EP	+ EP	+ F200I	+ F200I	+ F200I	++ FSA	++ FSA	++ FSA	+ Blank	+ Blank	+ Blank

Figure 32 Screening results with cyclopentanone (**58**) as donor for the cross-aldol reaction. **A**: Exemplary TLC from the cyclopentanone (**58**) screening from row D of the master plate is shown. As can be seen, both the cross- (red to orange in color) and the homo-aldol products (blue) are formed. Eluent: cyclohexane:ethylacetate 2:1. **B**: Complete screening results of the master plate in the cross-aldol addition with cyclopentanone (**58**). Band intensity is depicted ranging from (+) for low to (+++) for high. [a] Controls.

The results from the experiment demonstrate that cyclopentanone (**58**) is accepted as a donor by various DERA variants. The DERA variants (T18S/L20G/F76G) and (T18S/L20G/F76A) showed excellent activity classifying as (+++). The DERA variants (T18S/L20G/F76S), (L20G/F76A) and F76A all classified as medium active, while all other samples showed insignificant activity with trace amounts of product only. Interestingly, the DERA variants

(T18S/L20G/F76G) and (T18S/L20G/F76A) showed increased activity compared to the positive control variants FSA (D6A/T26I), while the wildtype variant DERA F200I used as a reference control showed minor activity. Since the negative control with empty plasmid and blank all showed activity, it must be assumed that the reaction is also catalyzed to some degree by background proteins of the BL21 cell line or by buffer since the activity is also measured for the blanks without enzyme. Since those questionable "hits" only formed trace quantities of product, it was deemed acceptable for discrimination of true active hits. The identified positive hits were further analyzed via HPLC in Chapter 4.7.

4.5.3 Screening of second-generation library with 3-oxetanone (62) as the aldol donor

When cyclobutanone (**57**) was not accepted, the idea that 3-oxetanone (**62**) might be more reactive due to activation by the oxygen in β-position. The results of the screening are presented in Figure 33.
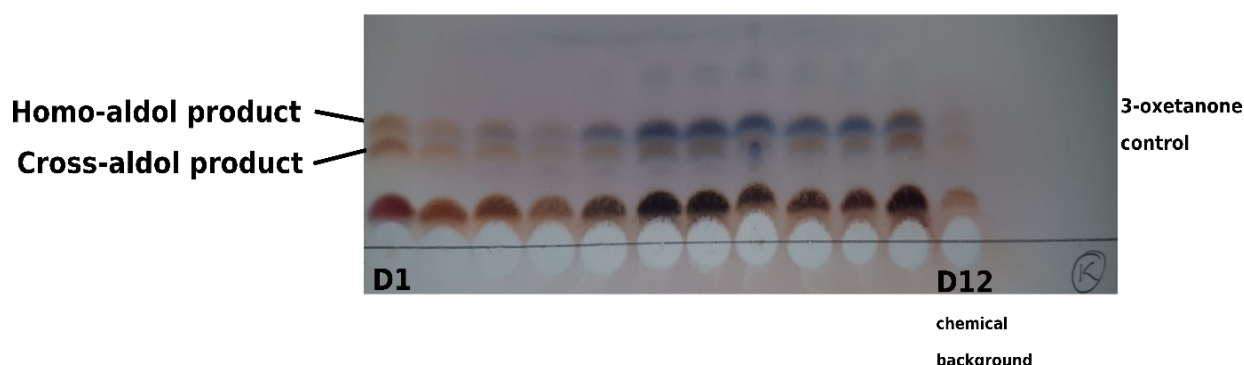
Figure 33 Screening results with 3-oxetanone (**62**) as donor for the cross-aldol reaction. Exemplary TLC from the 3-oxetanone (**62**) screening from row D of the master plate is shown. Eluent: cyclohexane:ethylacetate 2:1.

The first analysis via TLC showed newly formed spots that were identified initially as the product. Due to having constant faint spots in all control samples, it was assumed that this comes from chemical background. The DERA variants, which were showing an increased spot intensity for the cross-aldol product, were analyzed via HPLC. However, in the analysis no product formation could be detected. Most likely the product formation was too low to be detectable under the commonly implemented HPLC conditions. Therefore, 3-oxetanone (**62**) was deemed as an unacceptable donor for the DERA-variants, due to synthetically irrelevant yields.

4.5.4 Screening of second-generation library with dihydrofuran-3(2H)-one (63) as the aldol donor

After testing 3-oxetanone, the 5-ring analogue dihydrofuran-3(2H)-one (**63**) was tested for donor activity since cyclopentanone (**58**) clearly worked as a donor for some variants. Unfortunately, dihydrofuran-3(2H)-one (**63**) is such a reactive donor in the chemical aldol reaction that no enzyme catalyst is required for significant product formation, and TLC lanes were therefore indistinguishable. But again, the variants proved to be active by producing the homo aldol product in the known pattern. An exemplary TLC is shown in Figure 34.
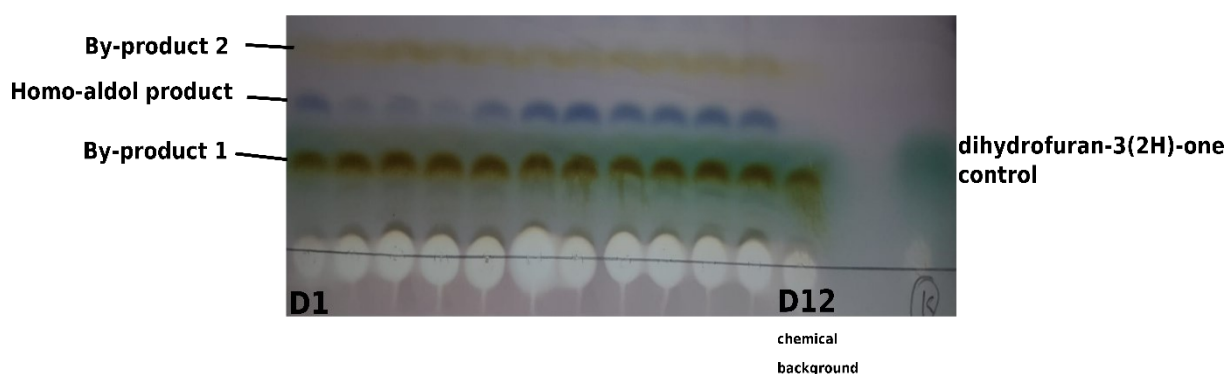
Figure 34 Screening results with dihydrofuran-3(2H)-one (**63**) as donor for the cross-aldol reaction (row D). As can be seen, the chemical background in well D12 is not distinguishable from the enzymatic reactions of the other 11 wells, except for not forming the homo-aldol product (deep blue). Hence, no statement about the selected mutants ability to utilize dihydrofuran-3(2H)-one (**63**) as an aldol donor in this reaction can be made. Eluent: cyclohexane:ethylacetate 2:1.

4.5.5 Screening of second-generation library with cyclohexanone (61) and tetrahydro-4H-pyran-4-one (64) as the aldol donors

Up to day, no DERA variant is known to utilize cyclohexanone (**61**) as a donor. In this work, both cyclohexanone (**61**) and the corresponding tetrahydro-4H-pyran-4-one (**64**) were tested as donors in the cross-aldol reaction with n-butanal (**53**). The low expectations indeed were met as none of the DERA variants were able to produce the cross-aldol product above questionable trace amounts. But again, the occurrence of traces may indicate that with multiple rounds of evolution, the enzyme could be better adjusted to such bulky donors.

4.5.6 Screening of second-generation library with butanone (55) as the aldol donor

Butanone was utilized as donor in the literature but seems challenging because it was only identified in one example with the natural acceptor GA3P (**3**)[122] in the retro aldol direction, meaning the cleavage of (5R,7S)-5,7,8-trihydroxyoctan-3-one.

Indeed, all variants showed no product formation for the cross-aldol reaction with butanone (**55**) as donor and *n*-butanal (**53**) as acceptor.

4.5.7 Summary of the cross-aldol screening results with novel aldol donors

In summary, the selected second-generation mutants showed new activity only in the reactions with cyclopentanone (**58**). The DERA variants (T18S/L20G/F76G) and (T18S/L20G/F76A) demonstrated excellent activity, both being second-generation variants. DERA variants (T18S/L20G/F76S), (L20G/F76A) showed medium activity comparable with first generation DERA-variant (F76A).

The distinguished clones showing those activities are mapped onto the master plate layout in Figure 35: activity with cyclopentanone (**58**) is highlighted in blue, with 3-oxetanone (**62**) in yellow and activity with both in green.

	1	2	3	4	5	6	7	8	9	10	11	12
A	76G A8 T18S L20G	76G B8 T18 L20	76G G6 T18S L20	76S A2 T18 L20	76S A5 T18 L20V	76S A7 T18S L20	76S F9 T18 G L20 G	76S B2 T18S L20G	76V A4 T18 L20	76V D1 T18S L20G N114D	76V G1 T18S L20G	76V B11 T18A L20
B	76V C10 T18S L20	76V C11 T18 G L20V	76V D6 T18 G L20	76V D11 T18 G L20A	76V F7 Deletion mutant	76A A8 T18 L20 G	76A C5 T18S L20 G	76A A9 T18 L20A	76A B7 T18 L20 G	76A D9 T18G L20G	76A H11 T18S L20	Blank

Figure 35 Layout of the master plate with the highlighted clones, which convert cyclopentanone (**58**) (blue), 3-oxetanone (**62**) (yellow) or both substrates (green) as their donor in the cross-aldol reaction with *n*-butanal (**53**).

4.6 Screening of the selected variants for activity in cross-aldol reactions with various donors and (*R,S*)-3-hydroxybutanal as an acceptor

It was decided that the DERA-library should also be screened for the cross-aldol reaction of 3-hydroxybutanal as an acceptor with cyclobutanone (**57**), cyclopentanone (**58**), 3-oxetanone (**62**) as possible donors. (*R*)-3-Hydroxybutanal (**19**) reflects the intermediate product of the homo aldol reaction of acetaldehyde (**17**), which is further converted by DERA to give the cyclic trimer. Therefore, it is an acceptor that is known to work well at least for acetaldehyde (**17**) as a donor. Another benefit is the β-hydroxyl group, which allows an intramolecular cyclization and therefore can possibly shift the equilibria towards the cyclized

product. For preventing the cleavage of the acceptor substrate, (S)-3-hydroxybutanal (**60**) was chosen. For a control, racemic and (*R*)-3-hydroxybutanal (**19**) were also screened.

For an access to the enantiomerically pure forms, a lipase such a CalB can be utilized, followed by deprotection of the aldehyde function via H+-functionalized resin. The reaction solution after deprotection is used as such without further purification for the enzyme reaction. The reaction scheme is shown in Scheme 17.
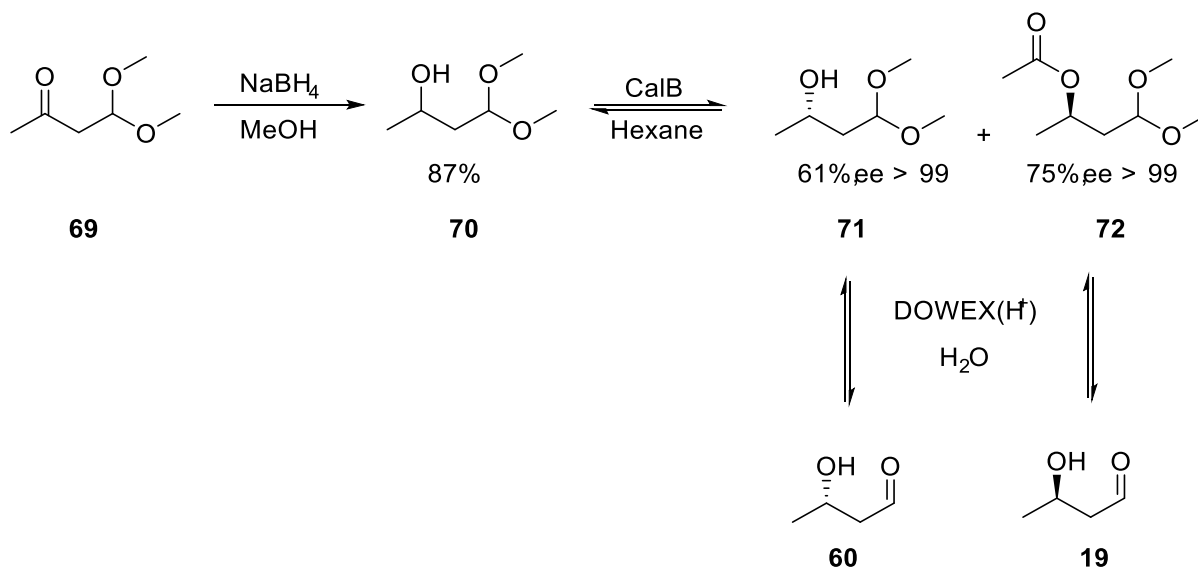
Scheme 17 Synthesis of (*R/S*)-3-hydroxybutanal.

For the synthesis of 3-hydroxybutanal the synthetic pathway of DURRWACHTER and WONG was utilized[127]. 3-Hydroxybutanal is accessible through reduction of 4,4-dimethoxybutan-2-one with sodium borohydride, which yields the dimethoxy-protected 3-hydroxybutanal as a racemate. By kinetic resolution performed with CalB and vinylacetate both enantiomers can be separated due to different reaction rates of both enantiomers. CalB will react exclusively with the (*R*)-enantiomer as long as it is available, enabling to stop the reaction before reaching low concentrations of the (*R*)-enantiomer. After isolation of the (*R*)-ester, the reaction is performed a second time until the (*R*)-enantiomer is fully consumed. By separation of the produced (*R*)-ester and (*S*)-ester from the remaining (*S*)-enantiomer, both enantiomers can be separated with high enantiomeric excess in good yield.

The reaction was followed via GC by using a chiral column yielding two peaks for both enantiomers. After 6h reaction time the conversion of the (*R*)-enantiomer was almost complete and the ester was isolated with 61% yield. The reaction was then repeated to completely convert the remaining (*R*)-enantiomer and stopped directly when the (*S*)-enantiomer started to be consumed, and the esters were separated from the remaining (*S*)-enantiomer. Both products were obtained with excellent enantiomeric excess of >99, based on the peaks obtained in the GC analysis.

The deprotection of the aldehyde function for liberating the (*R*)-enantiomer can be performed by utilizing H+-functionalized resin or inorganic acids. Due to the product possessing a β-

hydroxyl group, which can eliminate under acidic conditions, the milder version utilizing H$^+$-functionalized resin was found to be favorable by minimizing elimination effects.

The results of the screening showed no enzyme activity on TLC and also in control experiments with HPLC-MS towards the cross-aldol products of the corresponding donors. The most plausible reason for that is inhibition of DERA enzymes by crotonaldehyde (**20**), which can form by elimination of both enantiomers of 3-hydroxybutanal[106, 128].

4.7 Characterization of the selected second-generation variants

After the initial screening by TLC, successful enzyme reactions were subsequently analyzed via HPLC for independent verification and quantification. Whole cells after overexpression of the enzyme in AIM were utilized for the reactions. The cells were harvested by centrifugation and washed with 50 mM TEA*HCl buffer (pH 7.5). After centrifugation and resuspension of the cells in 50 mM TEA*HCl buffer (pH 7.5), the cell suspension was used as such for the characterization of the enzyme activity. The reactions were performed on a 200 μL scale.

4.7.1 Characterization of the selected second-generation variants for activity in homo-aldol addition with *n*-butanal (53)

The homo-aldol reaction of *n*-butanal (**53**) was used for initial screening of the DERA-library due to *n*-butanal (**53**) being the next bulkier aldehyde relative to propanal (**47**), which is a naturally accepted donor. Therefore, DERAs that can convert *n*-butanal (**53**) already show enhanced substrate scope compared to the wildtype *Ec*DERA.

The general conditions were kept constant to those of section 4.5, except for using whole cells instead of purified enzyme. Each measurement was conducted in triplicates. In Figure 36 the conversions are summarized.
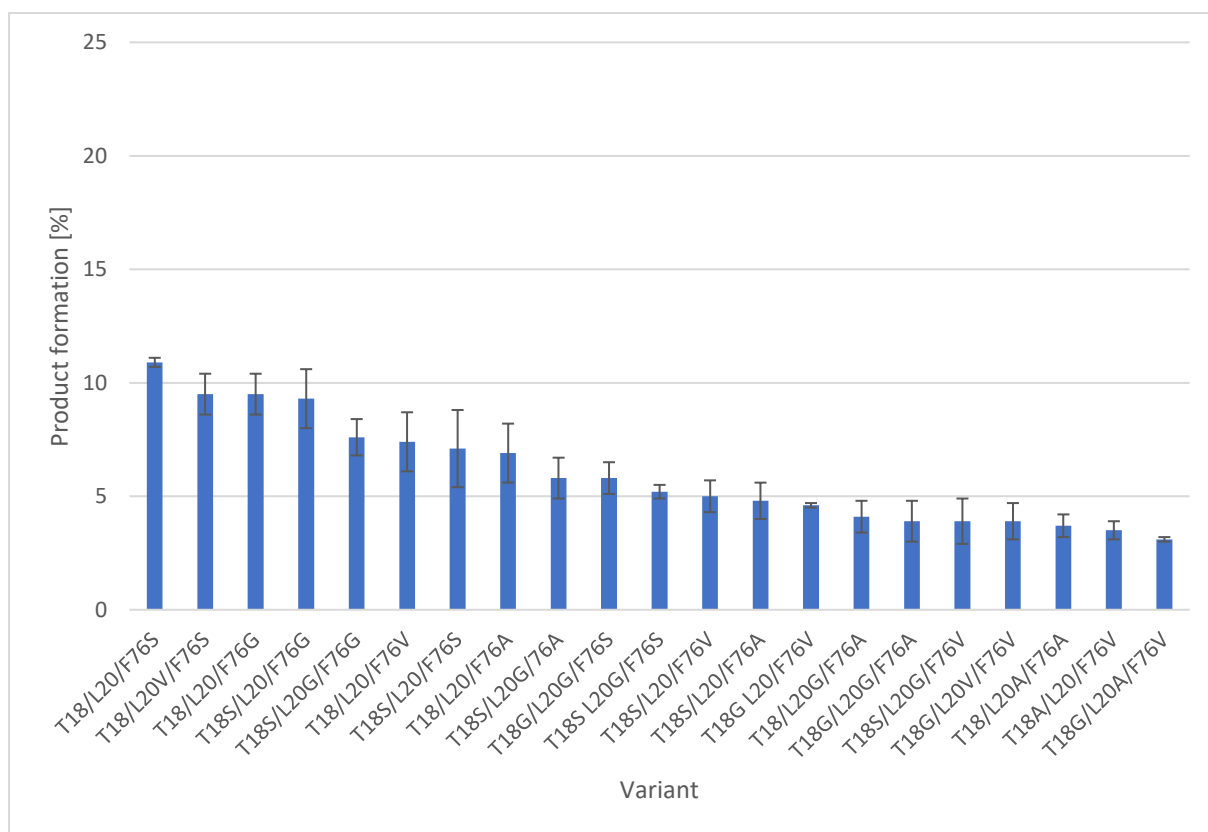
Figure 36 Results for the selected second-generation mutants (master plate) in *n*-butanal (**53**) homo-aldol addition after 22 hours. Resulting educt and product concentrations, as well the product formation rates were calculated using the *n*-butanal (**53**) standard curve (see Section **3.2**; 102276 counts/mM *n*-butanal (**53**)).

The results of the HPLC measurement of the enzyme reactions revealed that all variants were able to transform *n*-butanal (**53**) to the corresponding homo-aldol product in the range of 3-11% product formation. Due to the low product level the usually acceptable error rates seem high compared to the amount of product. The errors most likely result from the usage of whole cells instead of purified enzymes. The expression rates can be slightly different during each expression in MTP format due to different positions in the screening MTP. For example, samples that are placed at an edge or corner have different conditions than samples placed in the center of the plate, which is referred to the corner or edge effect[129].

The comparison with the first generation lacks same conditions since in my MSc Thesis I had measured the turnover with purified enzymes instead of whole cells, although certainly some comparison can be made. With purified enzymes, the DERA mutants F76A and F76S, both revealed four times higher product formation with purified enzyme. After that comparison, the best DERA mutants from the first generation were remeasured as whole cells in the homo-aldol reaction of *n*-butanal (**53**). The reactions were not performed as triplicates and were conducted just for comparison. The results are shown in Figure 37.
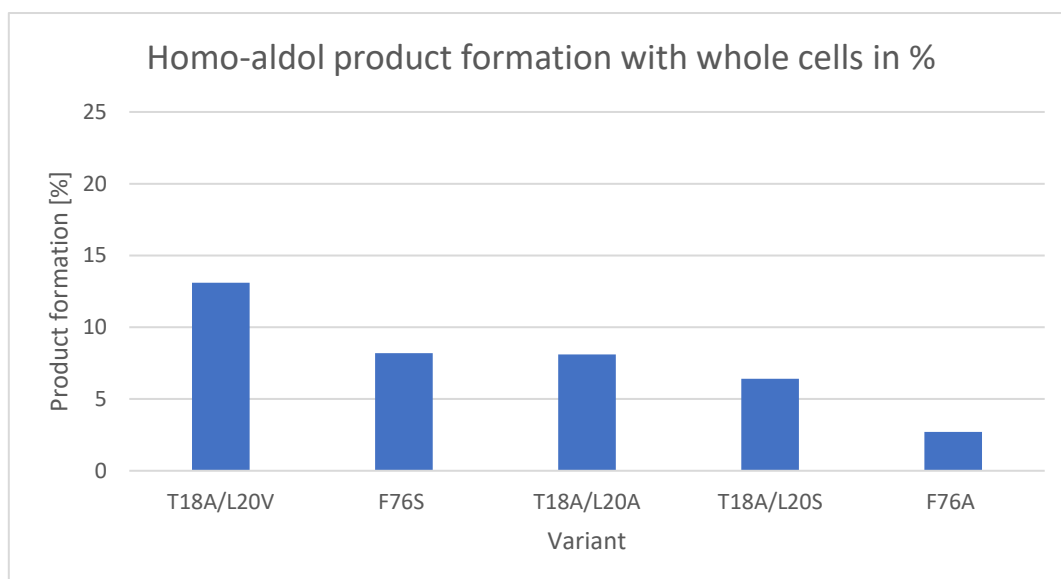
Figure 37 Selected clones of the first-generation library and their respective homo-aldol product formation with whole cell suspensions and purified enzymes used as catalysts.

The comparison of second-generation library DERA-variant F76S (11%) and first generation variant F76S (8%) both show comparable results while that does not account for DERA F76A yielding a range of 2.7-6.9%.

Interestingly, the variation of T18A/L20A/V in combination with a mutation at position F76 are not included in the selection. The expectation was clearly to find at least one DERA-mutant in the second-generation bearing those mutations and being able to transform *n*-butanal (**53**). At position T18 for the five best DERA-variants, the position is either conserved or presents a mutation towards serine, which was not found in the first generation. The best results were achieved with the position being conserved and a mutation for position F76 with smaller serine, alanine or valine residues. For position L20 all DERA mutants, the best and the worst, show the same combination of mutations. Mostly the position is either conserved or bears a mutation with glycine, valine or alanine. Alanine is only present as a mutation for the worst DERA-mutants, while valine and glycine are present in good as well as in bad mutants. In general, the turnovers of those reactions are very low.

4.7.2 Characterization of the selected second-generation variants for activity in cross-aldol reactions with novel donors and *n*-butanal (53) as the acceptor

From the screening for novel donors, three donors were identified for further HPLC analysis from the initial screening via TLC with anisaldehyde staining. 3-Oxetanone (**62**) and cyclopentanone (**58**) were accepted in the corresponding cross-aldol reaction with *n*-butanal (**53**) as acceptor. During the reaction the same reaction conditions as in section **4.5** were utilized for comparison.

First the cross-aldol reaction of cyclopentanone (**58**) with *n*-butanal (**53**) was repeated and analyzed via HPLC. The experiment was conducted in triplicates with the five most promising

DERA-variants from the second-generation library. The selected DERA-variants were (T18S/L20G/F76G), (T18S/L20G/F76A), (T18G/L20G/F76A), (T18S/F76A) and F76A. The results are presented in Figure 38.
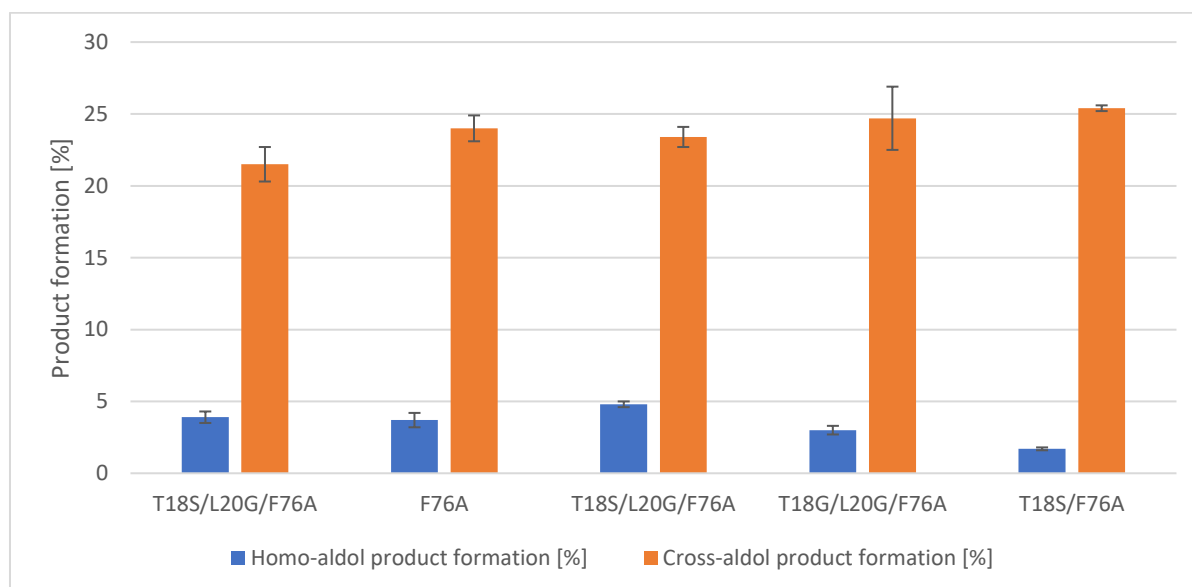
Figure 38 Measurement results for five selected second-generation mutants in the cross-aldol addition with cyclopentanone (**58**) after 22 hours. Resulting product concentrations, as well the product formation rates were calculated using the *n*-butanal (**53**) standard curve (102276 counts/mм *n*-butanal (**53**)).

The results present a clear preference of the DERA-variants towards the cross-aldol product of *n*-butanal (**53**) and cyclopentanone (**58**) rather than producing the homo-aldol product of *n*-butanal (**53**). The highest selectivity for the cross-aldol product is found for first generation DERA-variant F76A with a ratio of 15 to 1 in favor of the cross-aldol product while the lowest selectivity's are found for second generation DERA-variants T18S/L20G/F76G and T18S/L20G/F76A both having ratios around 5 to 1. On average, product formation for the cross-aldol product is slightly higher than those measured for the first-generation variants. The standard errors of this experiment range from below 1% up to 10% in the case of (T18G/L20G/F76A). Again, those errors can originate from pipetting issues and additionally from the usage of whole cells as enzyme source instead of purified enzyme. In general, the second-generation variants perform on average better than the first-generation variants, except for the (T18A/L20A) which yields up to 30% product formation for the cross-aldol product of *n*-butanal (**53**) with cyclopentanone (**58**) as donor. With the usage of purified enzyme instead of whole cells and an excess of donor concentration, those turnovers should yield even higher amounts of product.

The next step was the analysis of the newly formed spots on TLC from the cross-aldol reaction of *n*-butanal (**53**) and 3-oxetanone. Five DERA variants namely, (F76A), (T18S/F76A), (T18S/F76G), (T18S/F76S), (T18S/L20G/F76A), produced spots on TLC that exceeded the chemical background of the negative samples without enzyme or background proteins from BL21 cell strain. The samples of the corresponding reactions were analyzed via HPLC in

triplicates. No product signals could be identified during the measurement via HPLC. An additional repeat via HPLC-MS revealed that no cross-aldol product was formed.

As a summary of the results from both HPLC measurements, seven of the selected DERA-variants bear the mutation F76A. It seems that bulkier donor molecules are stabilized by that mutation. Another beneficial mutation seems to be T18S, which was not identified in the screening of the first-generation library. Both mutations were included in 7 of 10 ten DERA-variants considered for HPLC analysis.

5 Discussion

5.1 Generation of first-generation supplementary DERA-clones and second-generation DERA library, expression of clones and protein purification

The main part of this chapter was the generation of the second-generation library and the supplementary DERA variants bearing mutations with glycine at beneficial positions identified during the screening of the first-generation DERA library.

For the generation of supplementary DERA variants for the first-generation library no problems occurred. The mutations were implemented via SDM with mismatched primers only bearing the desired mutations. One of the downsides were low yields of colonies when transforming the plasmids into *E. coli* BL21 expression cell line. With all plates never reaching above 100 colonies per plate, it was assumed that the transformation efficiency was low. Most probably this is a result from mishandling after electroporation, adding the media somewhat too late since even a small delay can cause significant decrease in transformation efficiency[130]. An alternative factor could be an increased salt concentration after plasmid isolation also leading to lower transformation efficiencies.

For creating the second-generation DERA-library, mismatched primers were utilized for position F76 due to observations in our group, which highly recommended the usage of such primers. The usage of matched primers often yielded insufficient results.

For position T18 and L20, a different approach was tested due to its apparent convenience. In this case, the nine nucleotides coding for the area T18-L20 were first deleted on precursor plasmids bearing one of the desired mutations at position F76 namely F76A, F76V, F76G and F76S yielding all over four different libraries with a fixed mutation at position F76. Afterwards, the area coding for T18-L20 was inserted with randomized primers coding for the desired possible mutations at position T18 and L20 in an equal statistical outcome. The theoretical outcome in this experiment is exactly the same compared to using mismatched primers.

Most errors in mutagenesis experiments originate from the PCR itself and are identified when controlling the outcome by sequencing. While in all experiments genes of the correct size were produced, in some cases undesired point mutations were identified. Those cases had to be repeated while happening at the randomization of position F76 as well during the deletion and insertion process of position T18-L20. Therefore, the problem likely arose from the PCR condition instead of the primers themselves. The usage of "Q5 High-Fidelity Polymerase", which possesses a proof-reading mechanism, should lead to an error rate of 1×10^{-6} [131], which translates to one error per one million bases. The experimentally observed error rate was higher than expected and can only result from altering PCR-conditions by errors during pipetting. In the literature it was found that altering concentrations during PCR can lead to the observed error rates[132].

After all, the second-generation library was created successfully. Random sequencing of 10 clones proved sufficient randomization. Therefore, the second-generation library could be finalized and screened for novel or enhanced enzyme activity.

All DERA variants can be overexpressed in AIM with a protein yield of 200-500 mg from 500 mL of expression volume. Those yields are common experience in our group and were therefore satisfying. The losses during desalting via SEC and column purification via His6tag were controlled and deemed neglectable.

5.2 Primary-screening results

For the primary screening it was decided to utilize screening via TLC stained with anisaldehyde. Due to product similarity to the educts both being aldehydes all different photogenic or coupled reactions didn't work for screening since product and educt both yielding similar results in such screening methods. The screening via TLC with the test reaction being the homo aldol reaction of *n*-butanal (**53**), the next bulkier aldehyde after propanal (47), which is accepted by *Ec*DERA[29, 133], the screening will yield any DERA-variants with increased substrate scope. The supplier Merck estimated the detection limit of the utilized TLC-plates at 50-100 ng depending on the spotted compound [134]. From experience, undetectable samples or samples that show trace amounts of product formation are anyway synthetically irrelevant. During the initial screening, it is noteworthy that the room temperature exceeded 30° C which allows evaporation of volatile compounds like *n*-butanal (**53**) or acetone (**50**) in greater amounts than with standardized conditions. Therefore, each sample was carefully prepared in sealed HPLC-vials.

During the screening the homo aldol reaction of *n*-butanal (**53**) and the cross-aldol reaction between *n*-butanal (**53**) as acceptor and acetone (**50**) as donor was utilized as test reaction. The products were classified based on the spot intensity after applying 1μL sample on TLC plates and staining with freshly prepared anisaldehyde staining solution. The resulting spots were categorized between (0) - (+++) depending on the spot intensity of the product. Obviously, those screenings are very dependent on the person classifying the positive hits. Due to that fact, not only (+++) were collected but also samples that showed (++) - (+++) for receiving a broader selection of novel DERA-variants. All over, 56 DERA-variants were selected for sequencing and further screening. The screening was conducted in duplicates which yielded comparable results and same reaction patterns on the MTP. Therefore, the screening accuracy was deemed precise.

A clear expectation was the finding of mutations that showed the best results in the first-generation library, namely T18A/L20A and T18A/L20V. Those mutations were only found isolated in combination with mutations at position F76. Interestingly, the most common mutations for position T18 were T18S/T18G while for position L20 the mutations L20S/L20G, which were not identified in the first-generation screening. Of course, the mutations bearing a glycine were not considered in the design of the first-generation library. Nevertheless, the found combinations of mutations showed comparable product formations to the best DERA-variants from the first generation.

5.3 Comparison of homo-aldol activities

5.3.1 Data discussion

The comparison of the results from the second-generation DERA-library in this work with earlier studies with the first-generation DERA-library is hard to achieve. The HPLC measurement yielded overlapping peaks in earlier work due to different measurement method and the usage of an old RP-18 column. In this work, a new column with adjusted measurement method led to separated peaks which could be analyzed properly. Another factor, which should be neglectable, is the purification process of the enzymes. In this work SEC and His6tag column purification were utilized, while in earlier work the desalting was acquired by ultra-filtration cells.

The results produced during determination of the product formation via HPLC in this work yielded low standard errors of 1 – 13 % of the calculated homo aldol product formation value. With increasing amounts of product, the error usually is lower compared to samples with low amounts of product. Therefore, the measurement itself was deemed as precise and successful.

In nature, DERA is responsible for the cleavage of 2-deoxyribose-5-phosphate. Since the cleavage of DR5P (**18**) is reversible, the addition is possible but should be less favored[29, 97]. Furthermore, the utilization of unnatural acceptors and especially donors results in at least two digits loss of activity as demonstrated by WONG et al. [29]. Considering those effects, the mediocre to low conversions could be explained, since the enzyme is optimized for the natural reaction instead of being applicable to a broad selection of reactants. In fact, *Ec*DERA already has a relaxed donor spectrum compared to other aldolases. Instead of only utilizing acetaldehyde (**17**) as donor, it also accepts propanal (47), acetone (**50**) and fluoroacetone (**77**)[133]. Not only is *Ec*DERA able to utilize four different donors, but also is able to convert ketones and aldehydes, even though the natural reaction cleaves acetaldehyde (**17**) exclusively. Consequently, the achieved low turnovers seem normal considering the above-mentioned aspects of *Ec*DERA. A possible method for increasing the yield up to full conversion is the utilization of 3-hydroxyaldehydes or 3-aminoaldehydes as acceptors, allowing the aldol product to cyclize pushing the equilibria to the cyclized product form.

Comparing the newly produced DERA-variants from second generation with variants from the first generation, comparable variants were produced but indeed not exceeded the abilities from DERA-variants of the first generation. The combination of mutations from the two best DERA-variants from the first-generation library, namely T18A/L20A/V, were not identified in combination with mutations at position F76 in the second-generation library. Some triple mutants showed comparable results. T18G/L20G/F76A, T18S/L20G/F76G, T18S/L20G/F76A all proved to be able to catalyze the cross-aldol reaction of *n*-butanal (**53**) with cyclopentanone (**58**). Considering that *Ec*DERA is able to utilize acetaldehyde (**17**), acetone (**50**), fluroacetone and propanal (47) as donors, cyclopentanone (**58**) is a great achievement considering its bulkiness. Interestingly, cyclopentanone (**58**) with its bulky cyclic structure is accepted while butanone (**55**) with one carbon less is not. The reason for that remains elusive but

nonetheless, being able to utilize cyclopentanone (**58**) as a novel donor is still a great achievement even though cyclopentanone (**58**) was accepted already as donor by the first generation.

The other tested donors, all produced negative results or only trace amounts not judged as being synthetically useful.

Considering the aim of this work, combining beneficial mutations that lead to increase of substrate scope and/or activity in reactions with unnatural donors, the results are insufficient to classify a success. The substrate spectra of the second-generation were identically to the first-generation library. Also, the measured product formations were similar in the homo aldol reaction of *n*-butanal (**53**) and cross aldol reaction of *n*-butanal (**53**) as acceptor and cyclopentanone (**58**) as donor. In industry two rounds of evolution is just a starting point to optimize an enzyme towards unnatural reaction conditions. Sometimes reaching over 10 cycles of evolution. But considering the screening afford, TLC based screening is too time costly for such mutagenesis projects usually demanding photometric based detection methods of the reaction process. Therefore, a decision was made that will be explained in chapter 2 of this work.

5.3.2 Analysis of mutational influence on the homo-aldol activity of the created mutants

The active site of *Ec*DERA is optimized to fit perfectly for binding and cleavage of 2-deoxyribose 5-phosphate. Figure 39 presents the available space and the geometry of the active center.
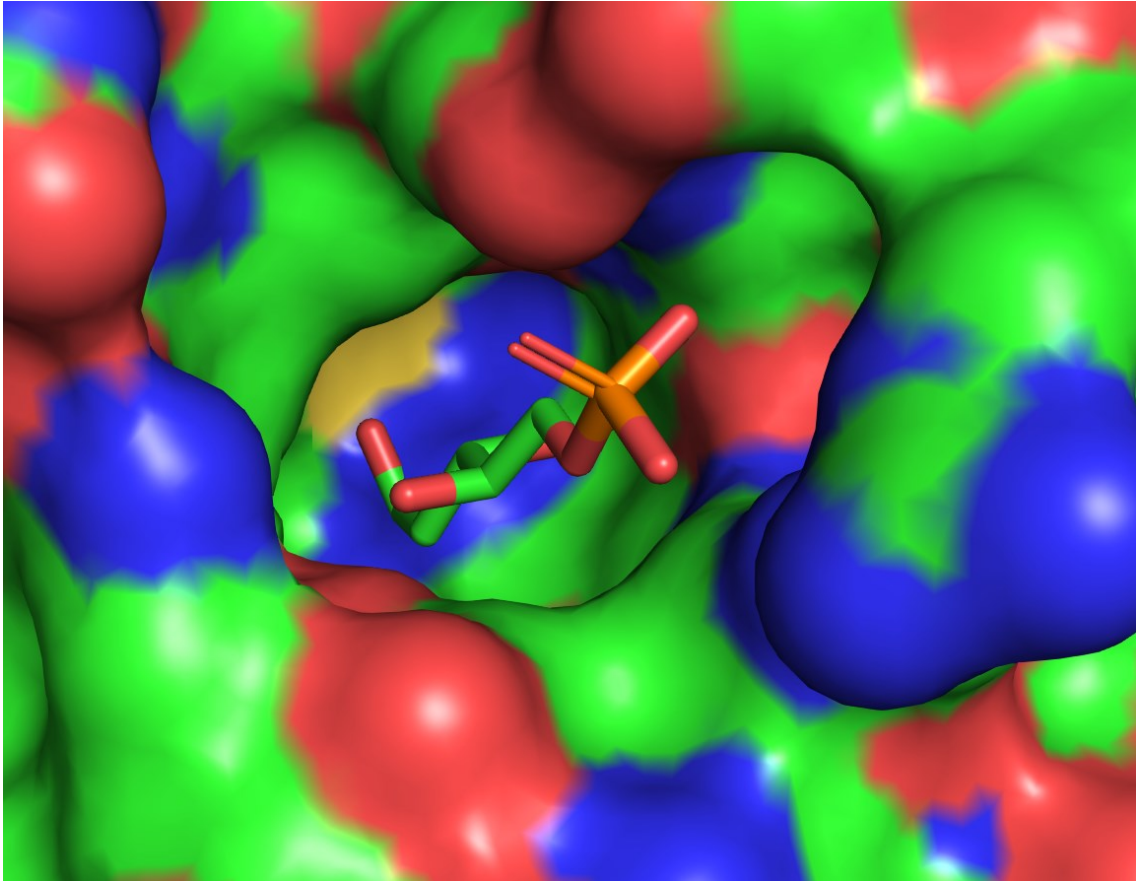
Figure 39 Active site geometry of *Ec*DERA. The blue color corresponds nitrogen atoms, red for oxygen atoms and green for carbon atoms.Picture created with Pymol on the basis of PDB-entry 1JCL.

Figure 40 demonstrates the optimized active center for the natural substrate in which the enzyme was crystallized. The active center is additionally shown in Figure 40, highlighting the positions of interest for mutagenesis, namely T18, L20 and F76. Also included is the natural substrate 2-deoxyribose 5-phosphate in the bound state as Schiff base with K167.

Figure 40 Active site configuration of the *Ec*DERA with the distances from the mutated amino acid positions (magenta) to the native substrate (green) measured in Å. Picture created with Pymol on the basis of PDB-entry 1JCL.

In Figure 40 the closest distances of the corresponding amino acid residue towards 2-deoxyribose 5-phosphate are presented for interpretation of the implemented mutations. The average length of a single bond between two carbon atoms with sp³ hybridization is ~1.5 Å, while the length of a single bond between a hydrogen and sp³ hybridized carbon is ~1.1 Å[135]. The displayed distances in Figure 26 result in a range of 3.7-5.2 Å to the closest non-hydrogen atoms of the highlighted residues T18, L20 and F76. Considering that those hydrogen-saturated atoms are practically in van der Waals contact to the substrate, the space left is barely enough to fit in another methyl-group, which is required for the binding of acetone (**50**) instead of acetaldehyde (**17**) as aldol donor. Important to notice is the excellent resolution of the crystal structure of 1.05 Å (pdb 1JCL)[103]. Furthermore, crystal structures always show a static picture of the state in which the protein has crystallized, while in solution the protein is in a dynamic situation being able to adjust on structural level towards different conditions.

Due to their positioning, F76 and L20 both stabilize the hydrophobic donor part while T18 can only stabilize the acceptor part of the natural substrate 2-deoxyribose-5-phosphate. The results from the first-generation DERA-library and the results presented in this work demonstrate that the positions T18 and L20 in combination possess the most promising potential for increased substrate scope and activity by mutagenesis while position F76, having the bulkiest side chain of all three residues, also proved to have great potential for increased

substrate scope by mutagenesis. The space gained by mutagenesis of position F76 towards the A/G/S/V variants is presented in Figure 41 based on the distances of the closest carbon towards 2-deoxyribose 5-phosphate.

Figure 41 Active site configuration of the *Ec*DERA with the distances from the mutated amino acid F76 to the native substrate (green) measured in Å. Picture created with Pymol on the basis of PDB-entry 1JCL.

The closest distance of F76 with its closest carbons of the phenyl ring towards 2-deoxyribose 5-phosphate is 4.6 Å. By changing to the smallest amino acid residue glycine an additional theoretical distance of 4.2 Å is obtained. It is questionable if the gained space will be free to occupy bulkier substrates, since glycine mutations can alter the folding pattern significantly[136]. Nevertheless, the second generation F76G DERA-library produced some of the best variants for the homo aldol addition of *n*-butanal (**53**), accounting for 3 out of 5 best variants selected for that purpose. Interestingly, the lowest turnover was detected for purified F76G of all first-generation glycine variants but when using whole cells during the screening of second-generation variants, F76G showed the second highest turnover of all selected DERA variants. The most plausible explanation seems to be protein denaturation during purification

of the DERA-variant, possibly resulting in a loss of activity. The usage of whole cells offers a more natural surrounding for the enzyme. Nevertheless, F76G proved to be a valuable DERA variant for synthesis.

The DERA variant F76S reflects the next bulkier residue in the selection, creating additional space of 2.9 Å compared to F76, leading surprisingly to an increased turnover of 11.0% vs. 9.5% for F76G. Serine is the only amino acid mutation with a polar side chain. The expectation in this case was to have increased substrate scope but decreased activity due to poorer orientation of the hydrophobic substrate *n*-butanal (**53**). Here, the serine mutation proved to be superior even though the difference is small. The gained distance shown in Figure 27 is the second highest of all four DERA variants, which is most likely the reason for the good turnovers.

Comparing F76A with F76S, it was expected that F76A should have higher turnovers than F76S due to the more hydrophobic side chain and higher distance of the side chain towards 2-deoxyribose 5-phosphate. These expectations were not met as DERA variant F76A showed the lowest conversion of all four variants. Nevertheless, compared to second generation of double or triple variants F76A still demonstrated high activity and creating additional space of 3.0 Å for bulkier substrates.

The last of the four DERA variants, F76V with the lowest distance gain of 1.9 Å, shows the third highest activity of the four DERA-variants. Due to the bulkier side chain of valine the gained space is accordingly smaller than for all other discussed DERA-variants.

For position T18 3 different mutations were identified to have beneficial effects, namely (T18A/G/S). All three mutations are presented in Figure 42 with displayed distances for residues T18A/G/S.
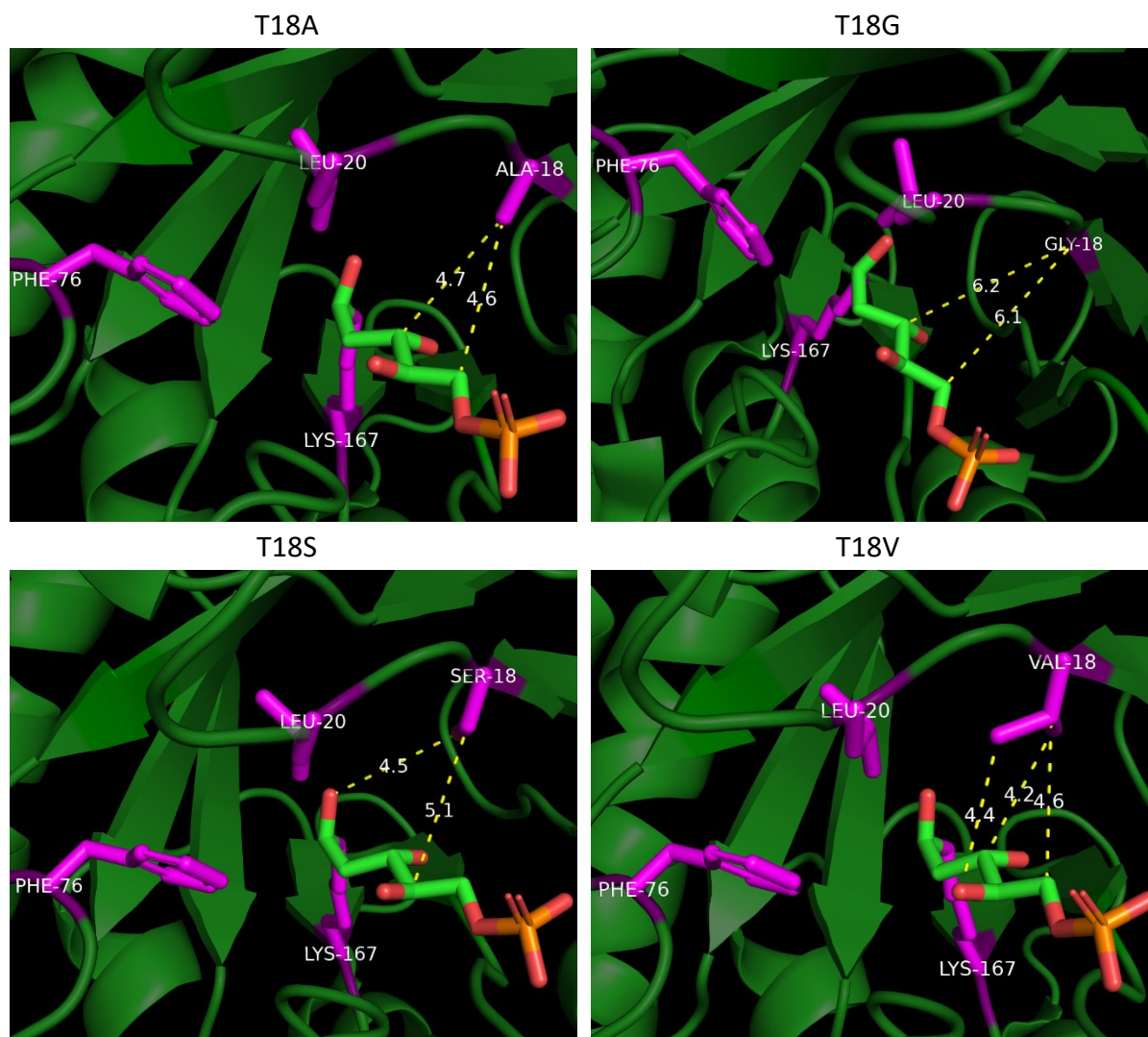
Figure 42 Active site configuration of the *Ec*DERA with the distances from the mutated amino acid T18 to the native substrate (green) measured in Å. Picture created with Pymol on the basis of PDB-entry 1JCL.

The closest proximities of T18 towards 2-deoxyribose 5-phosphate is 4.1-5.2 Å. Starting with F76G the gained distance is about 2-2.5 Å, which creates space for bulkier donors and acceptors during catalysis. Compared to T18A/S, the gained space for T18G is significantly larger. The same argument for position F76G with its missing side chain is of course a factor for any mutation with glycine, again possibly leading towards unpredictable structural effects. It is important to notice, that T18 was identified to stabilize the aldehyde function over a water molecule[103]. That stabilization is only possible for T18S, while T18A/G have a more apolar character. With 5 out of 21 selected DERA-variants bearing the T18G mutation, the mutation was represented in sufficient ratio to consider it a beneficial mutation even if all of the selected variants show only mediocre activity in the *n*-butanal (**53**) homo aldol reaction.

Comparison of T18S with T18 wild type, they have structural similarity with just a methyl group being lost. That will lead to a distance gain of 0.6-1.0 Å increasing space for bulkier donors and

acceptors while keeping the natural possibility to stabilize the donor part of 2-deoxyribose 5-phosphate. This is reflected in the number of selected DERA-variants bearing the T18S-mutation, which account to 8 out of 21 DERA variants.

Unexpectedly, T18A was found in only one of the selected DERA-variants, showing the second lowest turnover of all selected variants in the homo aldol reaction of *n*-butanal (**53**). During the screening of the first generation that very mutation proved to be a beneficial mutation in combination with mutations at position L20A/V. The distance gained for position T18A towards 2-deoxyribose 5-phosphate is 0.6 Å.

For position L20 mutations with alanine, valine and serine were identified. In the following Figure 43 all identified beneficial mutations are displayed for discussion.
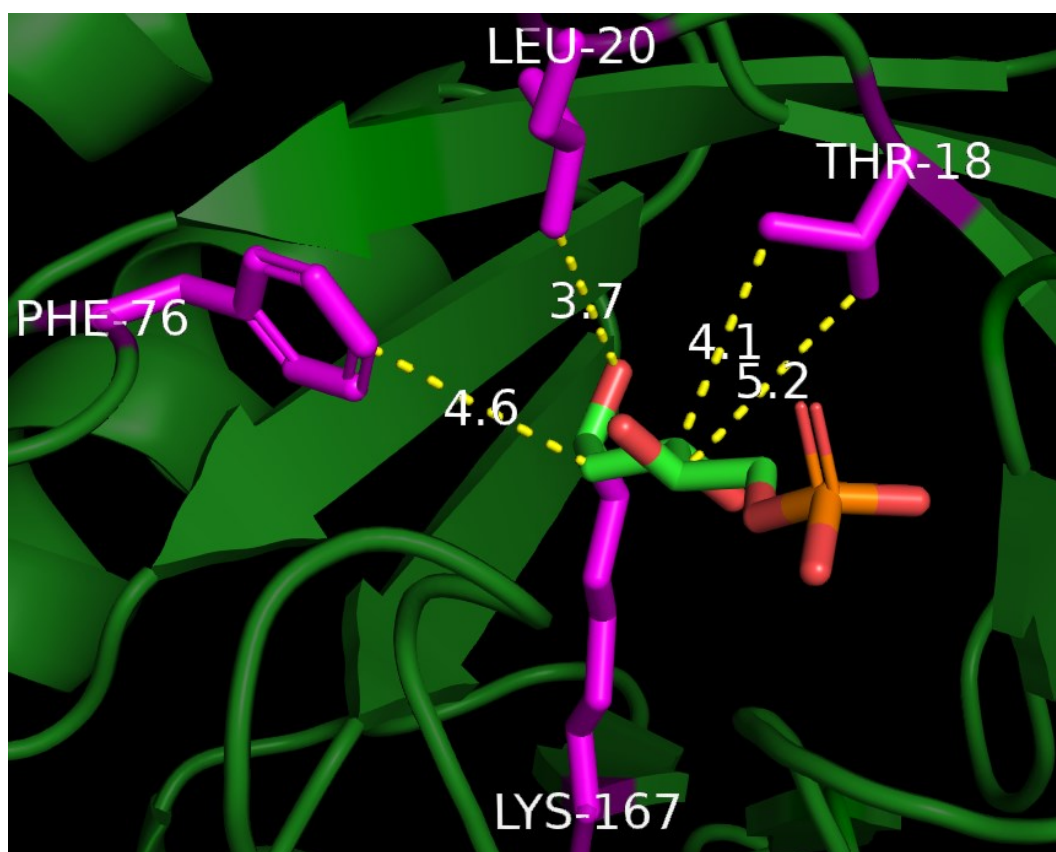
L20A

L20G

L20V
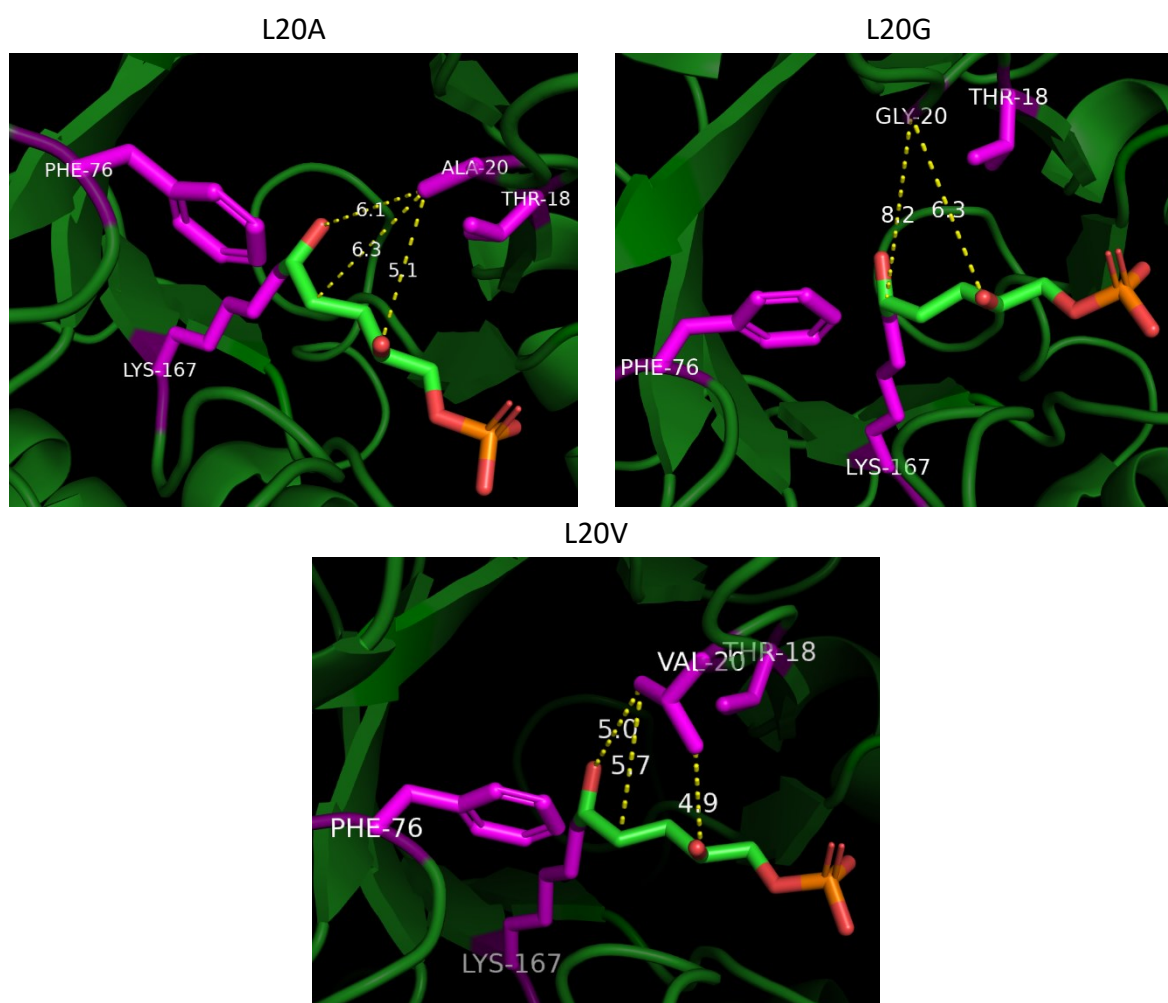
Figure 43 Active site configuration of the *Ec*DERA with the distances from the mutated amino acid L20 to the native substrate (green) measured in Å. Picture created with Pymol on the basis of PDB-entry 1JCL.

For (L20A/S/V) 0.9-2.1 Å of additional space is created, allowing bulkier donors and acceptors. Basically, the same arguments that are valid for position T18 and F76 also account for position L20. Again, the additional space gained allows the utilization of larger substrates. while the small amino acid residues allowed stabilizing effects of either said alkyl chain (L20A/V) or the

aldehyde group of the substrate (L20S). For L20G, additional 4 Å of space were freed, but the lack of an amino acid residue did not promote additional stabilization.

The variant (T18A/L20G) showed similar homo-aldol activity as (T18A/L20A). Most likely, the L20G mutation freed up enough space for the larger donor molecule *n*-butanal (**53**), while T18A mutation sufficiently stabilized the donor molecule in the active site. The combination of additional space and the ability to stabilize the substrates in the active center, resulted in enhanced activity in the *n*-butanal (**53**) homo aldol addition. On the other hand, the mutations (T18G/L20A/V) resulted in less additional space in the active center and no additional stabilization. This combination led to significantly lower yields in the homo aldol addition of *n*-butanal (**53**). The last mutant in line with this argumentation is (T18G/L20G), presenting lower conversions in the *n*-butanal (**53**) homo aldol addition without any stabilization effects to be expected from the exchanged replacements T18G or L20G, but freeing enough space to support catalytic activity of larger donors as compared to wild-type *Ec*DERA.

Additionally, free space can be filled with water molecules, that may further destabilize the binding and positioning of alkyl moieties in the donor. The best homo-aldol double mutants of the second-generation, (F76G/T18S) and (F76S/L20V) seem to present a good mix of the described properties: The F76G/S mutations freed space in the active site, and additionally F76S may stabilize the donor, while T18S and L20V may stabilize the acceptor and donor molecules, respectively.

5.4 Assessment of cross-aldol activities with novel donors

The triple mutant (T18S/L20G/F76G) was one of the selected five DERA variants for the cross-aldol reaction of *n*-butanal (**53**) with cyclopentanone (**58**) showing a product ratio of 5:1 in favor of the cross-aldol product. Those results could be interpreted towards a preference of ketones over aldehydes. But considering a 7% conversion that without cyclopentanone (**58**) against that of 4.0% with cyclopentanone (**58**) clearly is no real preference. The gained space is presented in Figure 44.
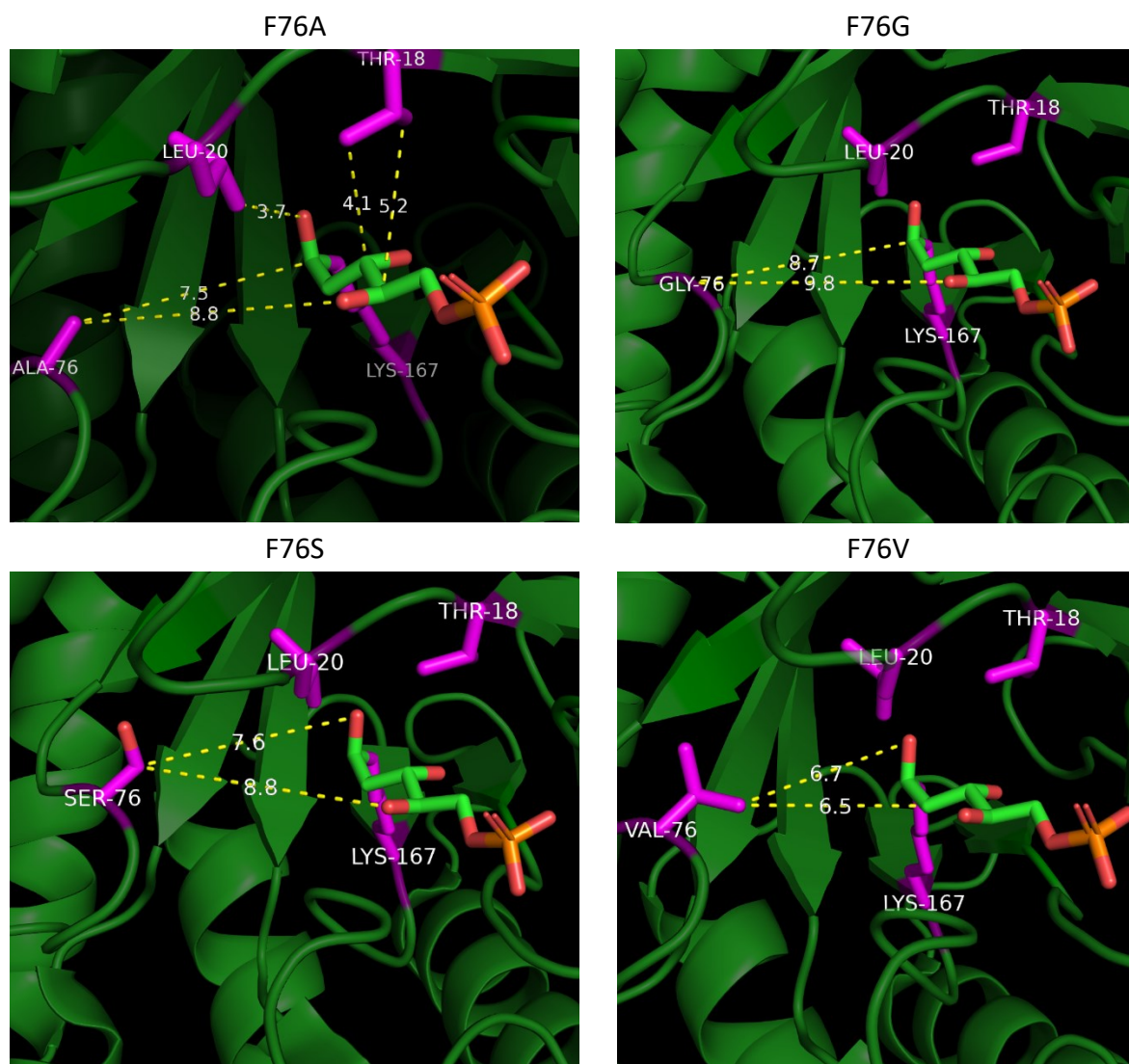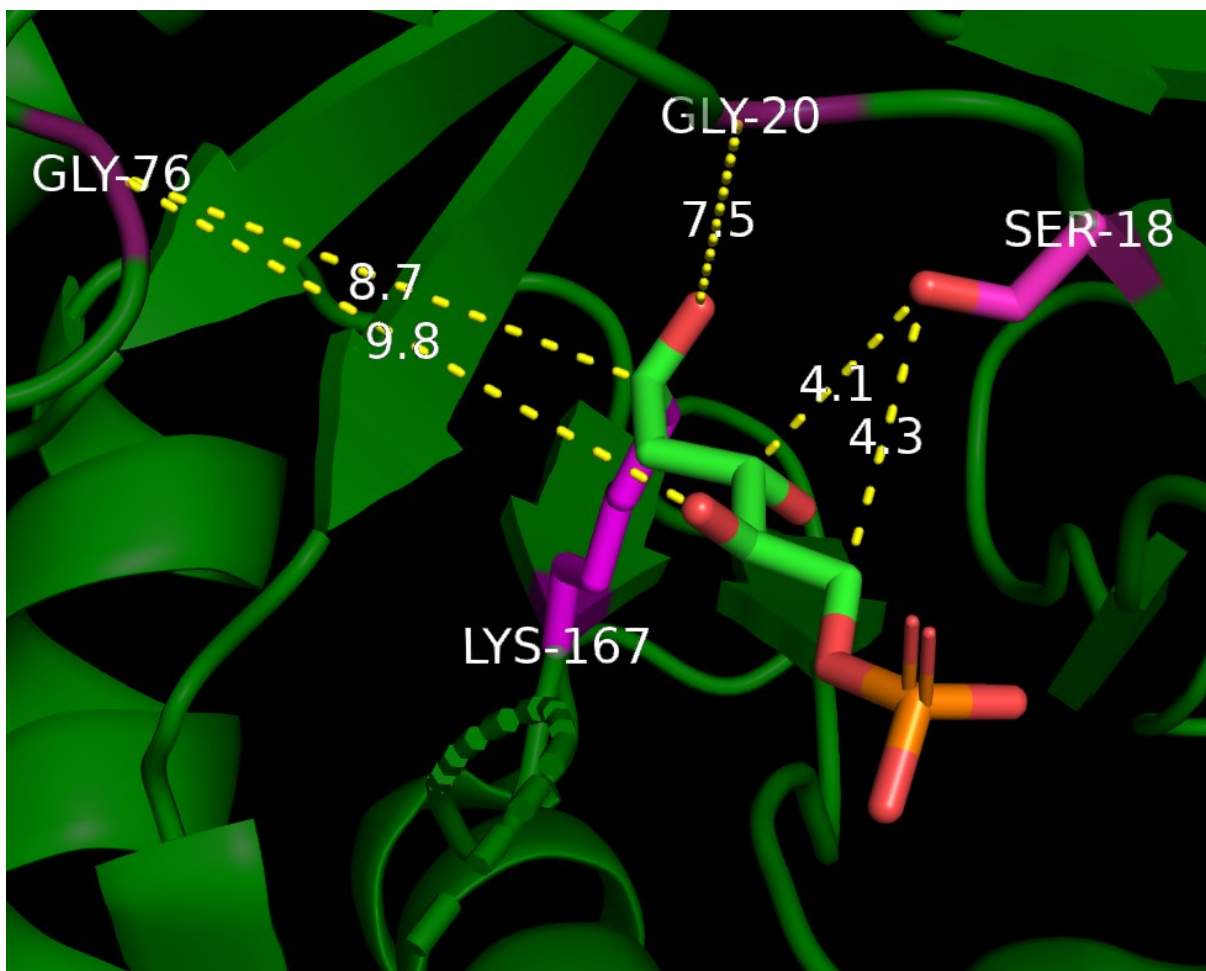
Figure 44 Active site configuration of the *Ec*DERA (T18S/L20G/F76G) with the distances from the mutated amino acids to the native substrate (green) measured in Å. The mutated amino acid residues T18S, L20, F76A and the active lysin residue are marked in magenta. Picture created with Pymol on the basis of PDB-entry 1JCL.

The mutation of F76A created space to accommodate the bulky ring structures of cyclopentanone (**58**), while the methyl residue of the alanine stabilized the alkyl ring in its position. T18S allowed for stabilization of the acceptor *n*-butanal (**53**) molecule. The alkyl residue of L20 had the property to also interact with the alkyl ring of the donor molecule, further stabilizing it.

Furthermore, one of the 5 best mutants for the cross-aldol reaction of *n*-butanal (**53**) and cyclopentanone (**58**) came from the F76G second generation library.

Interestingly, no F76V mutant was among the selected variants for cross aldol reaction of *n*-butanal (**53**) and cyclopentanone (**58**). The reason for that effect is most likely found in the space demanding side chain of valine, which may be incompatible with an effective binding of cyclopentanone (**58**) for catalysis.

1 Chapter 2 Metagenomic DERA and FSA library

2 Introduction

Given the limited success of the mutagenesis project presented in chapter 1, it was decided to implement a new approach by utilizing two metagenomic enzyme libraries that became available from a collaboration with Prozomix® UK. One enzyme library contained 200 DERA variants initially studied for acetaldehyde (**17**) trimerization[87], while the second library containing 100 FSA variants was created for the search of novel donor and acceptor activities. Some selected variants of the same metagenomic FSA library were additionally constructed to contain a D6Q mutation. That specific mutation at position D6 had produced excellent results in the literature (see Section 1.1.2)[110-111].

Metagenomics describes the strategy to analyze the genomes of microbial communities in specific environmental habitats in situ without cultivation. Metagenomics utilizes the advantage of direct DNA extraction from the environmental sample instead of first cultivating organisms in laboratory scale with a hit rate to identify 99% of the total microorganisms in a sample[137], indicating high precision of the technique itself. Direct DNA extraction from environmental samples was first introduced by PACE in 1985[138-139] but the term metagenomics was first coined in 1998 by HANDELSMANN[140]. It can either be evaluated on the sequences themselves or the corresponding function by suitable screening or selection. Metagenomics can be used to identify novel enzyme classes or extend the spectrum of known enzyme classes. In microbial communities the same enzyme types are expressed by different microbes, offering a diverse selection of the same enzymes from different organisms, which will therefore possess different properties such as thermostability, activity or substrate scope.

Metagenomic enzyme libraries offer a great diversity of active variants by utilizing environmental samples[87, 141-143]. Those samples can basically come from every habitat in which life exists[144]. One extreme condition can be high salt concentrations like the dead sea[145] or in general marine waters[146]. Some more extreme examples are salt saturated crystallizer ponds[147]. Acidic environments can also be the target for identifying extremophilic enzymes[148-150]. In general, environments outside of the pH range of 5–8.5 are considered as extreme habitats[151]. Additionally, other factors like extreme temperatures, presence of heavy metal pollution, pressure differences or organic solvents can have an impact on the evolution of the organisms living under those conditions.

The most prominent part of metagenomic enzyme mining takes place in environments located in areas with elevated temperatures. Organisms with an optimal growth temperature of above 45°C are considered thermophiles. With an optimal growth temperature of 65°C and higher, organisms are classified as hyperthermophiles[151]. Most prominent examples are extremophilic areas like hot smokers on sea ground, saline lakes or hot springs[152-154].

After taking the sample, the DNA is extracted and categorized based on sequence identity to known enzyme classes. Usually, after identification and amplification of the DNA, the carrying plasmid is expressed in *E. coli* cells under standard procedures. The enzymes are delivered as cell free extract by companies like Prozomix®. It is important to understand that not all

variants of a metagenomic enzyme library will be active due to different reasons. One reason is the misidentification of the corresponding enzyme class, some are inactive under the implemented conditions and some may have a substrate scope that is not tested in the study. Therefore, a reliable screening system is important to avoid false identification of inactive variants as active and vice versa.

3 Aim of work

After the limited success of the mutagenesis of *Ec*DERA F200I for creating space in the active center to unlock novel donor activity in Chapter 1, it was decided to utilize metagenomic enzyme sources. Two metagenomic enzyme libraries produced by Prozomix® were available from a finished collaborative project "Carbazymes". The first library contained 200 DERA orthologs while the second contained 100 FSA variants, plus the same number of FSA samples carrying a D6Q mutation.

The original purpose of those libraries was to find suitable candidates for sequential aldol addition of acetaldehyde (**17**) and propanal (**47**) for the DERA library, while the FSA library was meant for the search of novel donor and acceptor activity for FSA samples. With the knowledge available from those projects that some DERA enzymes were able to utilize cyclopentanone (**58**) as donor when using *i*-butyraldehyde (**71**) as acceptor, it was decided that the libraries should be rescreened for novel donor activity. The DERA library was screened for activity with *i*-butyraldehyde (**71**), while the FSA library was screened with *i*-valeraldehyde (**73**) instead of *i*-butyraldehyde (**71**) due to an unexpected loss of activity when using *i*-butyraldehyde (**71**). The work flow for screening both metagenomic libraries is presented in Scheme 1.
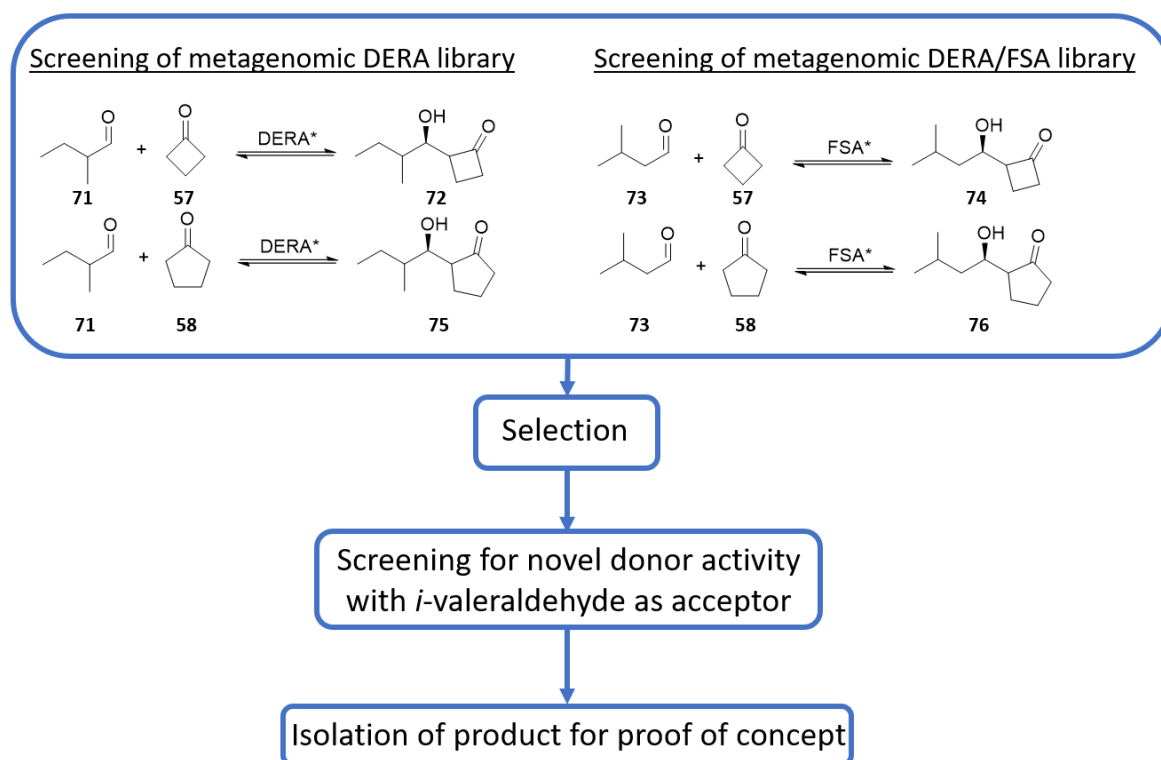
Scheme 1. Workflow for screening of metagenomic DERA and FSA libraries for identification of novel donor activity and synthesis of corresponding aldol products.

The aim of the initial screening was to identify active variants for cyclic ketones as donor. As shown in Scheme 2, different cyclic derivatives of cyclobutanone (**57**), cyclopentanone (**58**)

and cyclohexanone (**61**) were screened for activity with both libraries. Additionally, a set of acetone (**50**) derivatives and analogs were tested.

The selection was again based on the developed TLC spot intensity after staining with anisaldehyde staining solution. The spots on silica TLC sheets from the test reactions were classified from (0) – (+++). All samples with (++)/(+++) were selected and further screened with *i*-valeraldehyde (**73**) as donor and the set of donors presented in Scheme 2. All identified aldol products were subsequently synthesized using one of the corresponding most active enzymes and isolated on preparative scale to produce sufficient amounts for characterization by ^{1}H/^{13}C-NMR. The molecular mass of the corresponding aldol products was additionally determined in HPLC-MS during the screening to check for successful reactions before proceeding with preparative scale reactions.

Tested donors in aldol reaction with *i*-valeraldehyde as acceptor
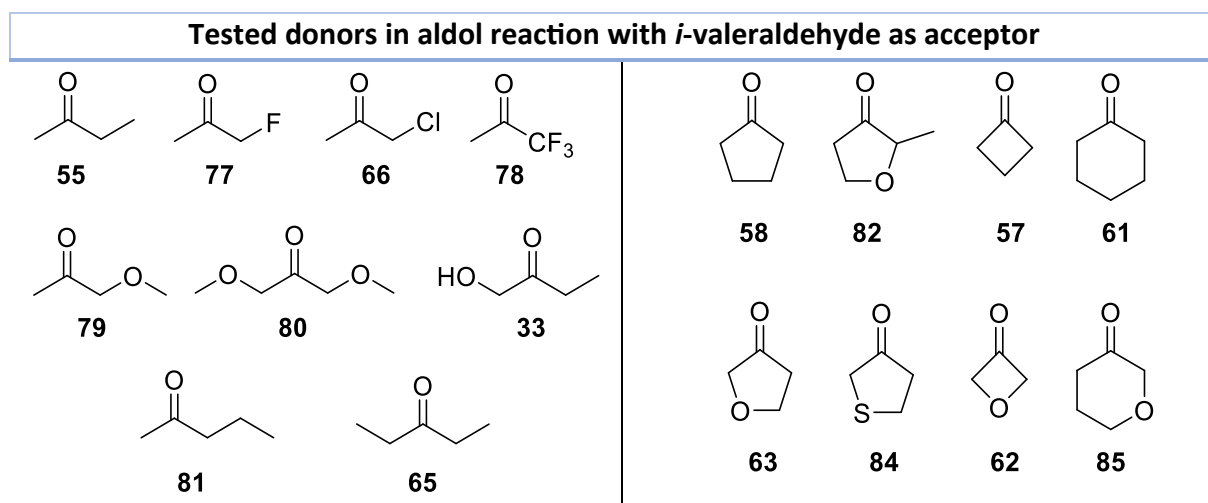
Scheme 2 Set of donors to be tested in enzymatic aldol addition with *i*-valeraldehyde (**73**), (*R*)/(*S*)-3-hydroxybutanal and 3-hydroxy-3-methylbutanal (**98**) as acceptors. The selection contains butanone (**55**), fluoroacetone (**77**), chloroacetone (**66**), trifluoroacetone (**78**), methoxyacetone (**79**), dimethoxyacetone (**80**), 1-hydroxybutan-2-one (**33**), pentan-2-one (**81**), pentan-3-one (**65**), cyclobutanone (**57**), 3-oxetanone (**62**), cyclopentanone (**58**), dihydrofuran-3(2H)-one (**63**), dihydrothiophen-3(2H)-one, cyclohexanone (**61**), dihydro-2H-pyran-3(4H)-one (**85**) and 2-methyldihydrofuran-3(2H)-one (**82**).

Additionally, after finalizing the screening and preparation of the corresponding aldol products, the accepted donors were tested for activity with 3-hydroxy-3-methylbutanal (**98**) and (*R*)/(*S*)-3-hydroxybutanal. The aim of this additional testing was to check if the aldol product will undergo an intramolecular cyclization after formation. The intramolecular cyclization should lead to higher conversion due to a shift of equilibria towards the cyclic aldol product.

4 Results

4.1 Screening of metagenomic DERA/FSA enzyme libraries for active variants with cyclic ketones

All measured values via HPLC-MS in this chapter are presented in form of bar diagrams. The exact values are presented in the attachments in table format.

4.1.1 Screening of metagenomic DERA library in cross aldol reaction of *i*-butyraldehyde (71) and cyclopentanone (58)

Prior to the systematic screening in this work, an initial screening of the metagenomic DERA library (Prozomix©) with another incentive. The goal was to find suitable candidates for sequential aldol additions of acetaldehyde (**17**) and propanal (**47**) on different acceptors. During that screening, activity was found for the donor cyclopentanone (**58**) with *i*-butyraldehyde (**71**) as acceptor for the metagenomic DERA library but no continuation had followed on those results.

Therefore, for the repeated screening of the metagenomic DERA-library (Prozomix©) similar reaction conditions were utilized first to reproduce the results from earlier screenings. The whole DERA-library was rescreened with the following conditions in a scale of 300 µl per reaction:

- 50 mM TEA*HCl buffer pH 7.5
- DMSO 10% v/v
- 100 mM *i*-Butyraldehyde (**71**)
- 150 mM Cyclopentanone (**58**)
- Enzyme cell free extract (CFE 8 mg/ml

From experience (see Chapter 3) choosing different buffer systems should not affect the reaction as long as the buffer is not deactivating the enzyme. Since FSA and DERA are classified as class 1 aldolases utilizing an active lysine residue for catalysis, most amine-based buffers will work without complications. A suitable alternative is standard phosphate buffer due to its non-extractable nature for larger scale reactions with purification of the product, while TEA*HCl can be extracted in small amounts when using a too high ratio of solvent to extractable solution.

While the cosolvent DMSO was not necessary for solubilizing the substrates, it was still added to reactions to keep comparability with earlier results. In general, it is advisable to use cosolvents to ensure solubility and keep evaporation effects limited. Also the destabilizing

effect of DMSO on enzymes, especially DERA and FSA, is rather mild with respect to lowering thermostability, and usually concentrations up to 20% v/v are very well tolerated[87].

Higher ratio of donor to acceptor, or better the excess of donor, usually increases the product formation rate. Considering that the donor needs to bind first with the active lysin to start catalysis, an excess of donor will force exactly that binding to faster rates. To keep the conditions comparable, it was decided not to increase the donor concentration initially to 300 – 500 mᴍ. The outcome with higher ratios of donor to acceptor will be presented later after the library screening was finished under the same conditions used in earlier screenings.

The enzyme samples in both libraries were provided as cell free extract (CFE) with the amount of enzyme varying around 30-70%. Therefore, it was decided to keep the enzyme load high. When working with purified enzyme, an enzyme load of 2 – 4 mg/ml is already considered relatively high, while the same amount as cell free extract can result in less than half of the desired amount. With 8 mg/ml CFE it was assured that those effects are minimalized and a sufficient catalytic activity is ensured.

The results of the screening were categorized from (0) – (+++). All samples obtaining at least (+++) were further analyzed via HPLC-MS. A total of nine variants showed sufficient activity for further analysis. The results of the HPLC-MS analysis of the selected hits are presented in Figure 1.

Figure 1. Results of the HPLC-MS measurement of the metagenomic DERA library in the cross-aldol reaction of *i*-butyraldehyde (**71**) and cyclopentanone (**58**) with 1 : 1.5 ratio of acceptor to donor in 50 mᴍ TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Based on the obtained results it was concluded, that the reactions are almost complete after two days, while the third day still yields more product but only to a very small extent. The obtained results were identical to earlier results, raising confidence in the screening method.

Compared to the results of the mutagenic DERA library (Chapter 1) based on *Ec*DERA, the metagenomic library yields superior results in the cross-aldol reaction of *i*-butyraldehyde (**71**) and cyclopentanone (**58**). Interestingly, while *i*-butyraldehyde (**71**) was initially utilized due to earlier results, it was found that the methyl group in α-position is already bulky enough to inhibit enzyme activity for most enzymes and, as presented later, will not yield any activity for all tested FSA enzymes of the metagenomic FSA library.

After the screening under the earlier conditions, it was decided to test an excess of donor up to 500 mM while keeping all other parameters constant to see if this will increase the conversion or if some enzymes will denature or show inactivity due to the higher concentrations of substrates. The results are presented in Figure 2.

Figure 2 Results of the HPLC-MS measurement of the metagenomic DERA-library in the cross-aldol reaction of *i*-butyraldehyde (**71**) and cyclopentanone (**58**) with 1 : 5 ratio of acceptor to donor in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

For most DERA samples a higher product formation was observed. Even though most DERA enzymes showed only slightly increased product formation, some orthologs like DERA180 and DERA125 showed significantly increased activity under the tested conditions with increased donor to acceptor ratio. The effect is most obvious for DERA180, doubling the product formation based on *i*-butyraldehyde's (**71**) initial concentration.

One of the best performing enzymes with 1:1.5 acceptor to donor ratio, DERA112, showed a complete loss of activity during the first day. DERA062 was known to perform well with acetone (**50**) as donor and was therefore a promising candidate for acetone (**50**) derivatives as donor but performed with one of the lowest product formations observed. Some enzyme samples, namely DERA067, DERA125, DERA143, showed a reduction of product formation

after 120 hours indicating that the reaction produced less product than available after the first two days. Therefore, it was decided to let the test reactions run for two days before analysis in future screenings for novel donor activity.

4.1.2 Screening of metagenomic DERA library in cross aldol reaction of *i*-butyraldehyde (71) and cyclobutanone (57)

The selected DERA enzymes were additionally screened for activity in the cross-aldol reaction of *i*-butyraldehyde (**71**) as acceptor and cyclobutanone (**57**) as donor. This specific reaction was desired for the mutagenic *Ec*DERA library in Chapter 1, but not achieved. Therefore, any positive hits were considered a success. It was decided to keep the acceptor to donor ratio of 1 to 5 for the screenings. The following reaction conditions were utilized in 300 µl scale reacting for 24h:

- 50 mᴍ TEA*HCl buffer pH 7.5
- DMSO 10% v/v
- 100 mᴍ *i*-Butyraldehyde (**71**)
- 500 mᴍ Cyclobutanone (**57**)
- Enzyme CFE 8 mg/ml
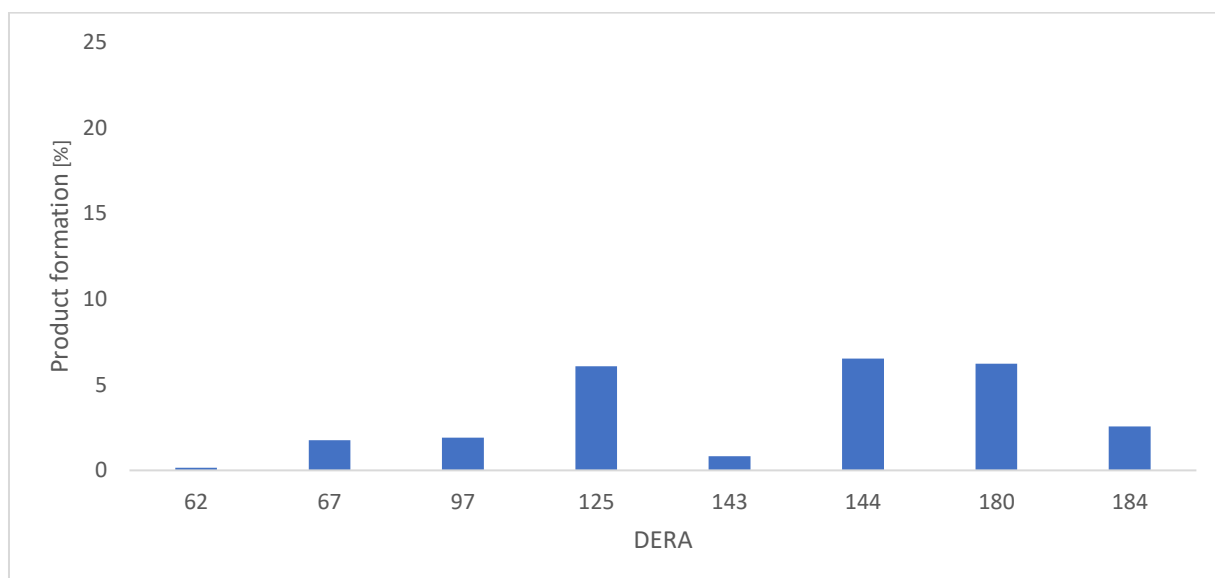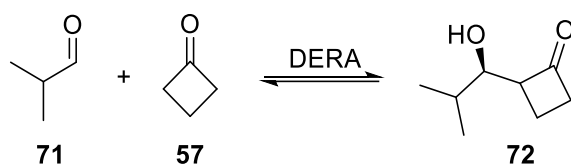
Figure 3 Results of the HPLC-MS measurement of the metagenomic DERA-library in the cross-aldol reaction of *i*-butyraldehyde (**71**) and cyclobutanone (**57**) with 1 : 5 ratio of acceptor to donor in 50 mᴍ TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Interestingly, most of the selected DERA samples showed product formation, even though in smaller quantities than with cyclopentanone (**58**) as donor. DERA126, which was already performing with low activity, showed no conversion by TLC and was therefore not further analyzed via HPLC-MS. While DERA097, DERA125, DERA144 and DERA180 all behaved similarly compared to cyclopentanone (**58**), DERA112 also showed no activity for cyclobutanone (**57**) as donor.

Comparing the results of the enzymes from metagenomic source with the mutagenic DERA library from chapter 1, one can already say that the successful DERA samples from metagenomic source show higher activity and a more relaxed donor scope. Still, one has to consider that the switch from *n*-butanal (**53**) towards *i*-butyraldehyde (**71**) changes the situation and steric effects can dominate the outcome. But nonetheless, the variants of the mutagenic *Ec*DERA library which showed conversion with cyclopentanone (**58**) had shown no conversion with cyclobutanone (**57**).

4.1.3 Screening of metagenomic FSA library in cross-aldol reaction of *i*-valeraldehyde (73) and cyclopentanone (58)/cyclobutanone (57).

Initially, it was intended to screen the metagenomic FSA library analogously to the screening of the metagenomic DERA library. During the screening with *i*-butyraldehyde (**71**), it was noticed that none of the tested FSA enzymes showed conversion. The reason for missing activity seemed to be the methyl group at α-position of *i*-butyraldehyde (**71**) since a switch of acceptor to *i*-valeraldehyde (**73**) yielded excellent results. The following conditions were utilized for the screening with cyclopentanone (**58**) as donor and *i*-valeraldehyde (**73**) as acceptor. The results of the screening are presented in Figure 4.

- 50 mM TEA*HCl buffer pH 7.5
- DMSO 10% v/v
- 100 mM *i*-Valeraldehyde (**71**)
- 500 mM Cyclopentanone (**58**)
- Enzyme CFE 8 mg/ml

Figure 4 Results of the HPLC-MS measurement of the metagenomic FSA-library in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**) with 1 : 5 ratio of acceptor to donor in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Due to the high rate of active variants, the hits were again classified in categories of (+) – (+++) and all samples with at least (+++) were analyzed via HPLC-MS. The screening yielded 12 variants with product formation above 10%. Especially FSA042, FSA049, FSA057, FSA061, FSA071, FSA074, FSA075 and FSA087 showed excellent results, deemed useful for larger scale reactions. Interestingly, the FSA samples FSA002-FSA028 showed no conversion in all conducted experiments. The reason for that can either be false identification as FSA orthologs since the classification is based on sequence identity. The next possible reason can be that those orthologs are highly specialized on the natural donor, which is commonly referred to DHA (**13**) even though the natural reaction is still unknown for FSA. Another possible reason is that the conditions implemented for the screening lie outside of the active range of those variants. It is important to notice that the metagenomic libraries of Prozomix are solely based on sequence identity and were not analyzed for activity in the natural reaction.

From the literature it is known that some FSA orthologs can utilize cyclopentanone (**58**) as donor and therefore the results were pleasing enough to continue with the screening for activity with cyclobutanone (**57**) as donor.

The same conditions were implemented as in the screening discussed above. Figure 5 presents the results of the screening with cyclobutanone (**57**) as donor.
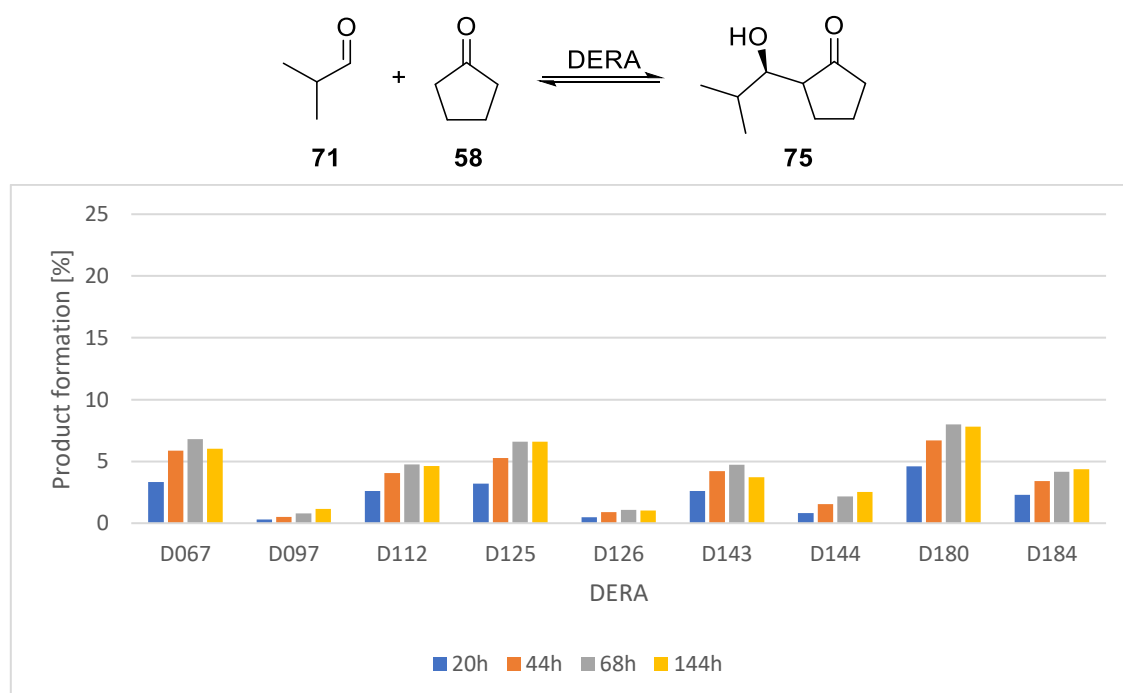
Figure 5 Figure 6 Results of the HPLC-MS measurement of the metagenomic FSA-library in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclobutanone (**57**) with 1 : 5 ratio of acceptor to donor in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

From the results, one can clearly state that cyclobutanone (**57**) seems to be a rather well-accepted donor for most FSA orthologs of the metagenomic FSA library. The average observed conversion with cyclobutanone (**57**) is higher than for cyclopentanone (**58**). The metagenomic library contains samples that can utilize both, cyclobutanone (**57**) and cyclopentanone (**58**), as donor with significantly higher conversions than those obtained from the mutagenic DERA library. Interestingly, the most active enzymes for cyclopentanone (**58**) like FSA049 performed still as one of the most active variants for cyclobutanone (**57**) but others like FSA057 showed higher levels of product formation with cyclobutanone, indicating that a broad selection of active variants should be considered for further screenings towards yet unknown donor activities to increase chances of success.

Both experiments were repeated with a random selection of active variants to check for standard errors of the screening method. The results of the rescreening for activity with cyclobutanone (**57**) and cyclopentanone (**58**) in triplicates are presented in Figure 7.
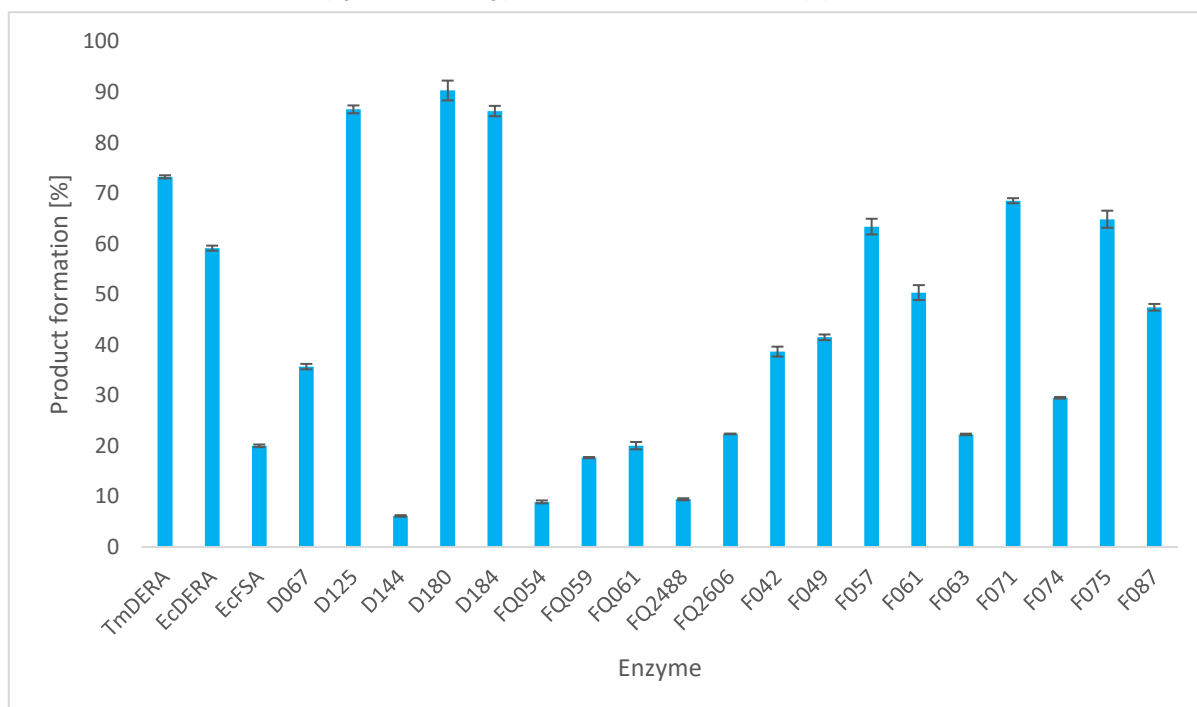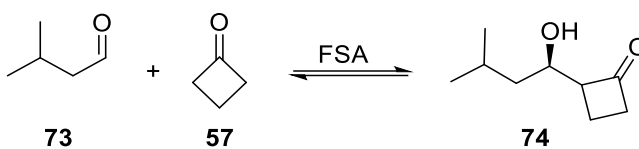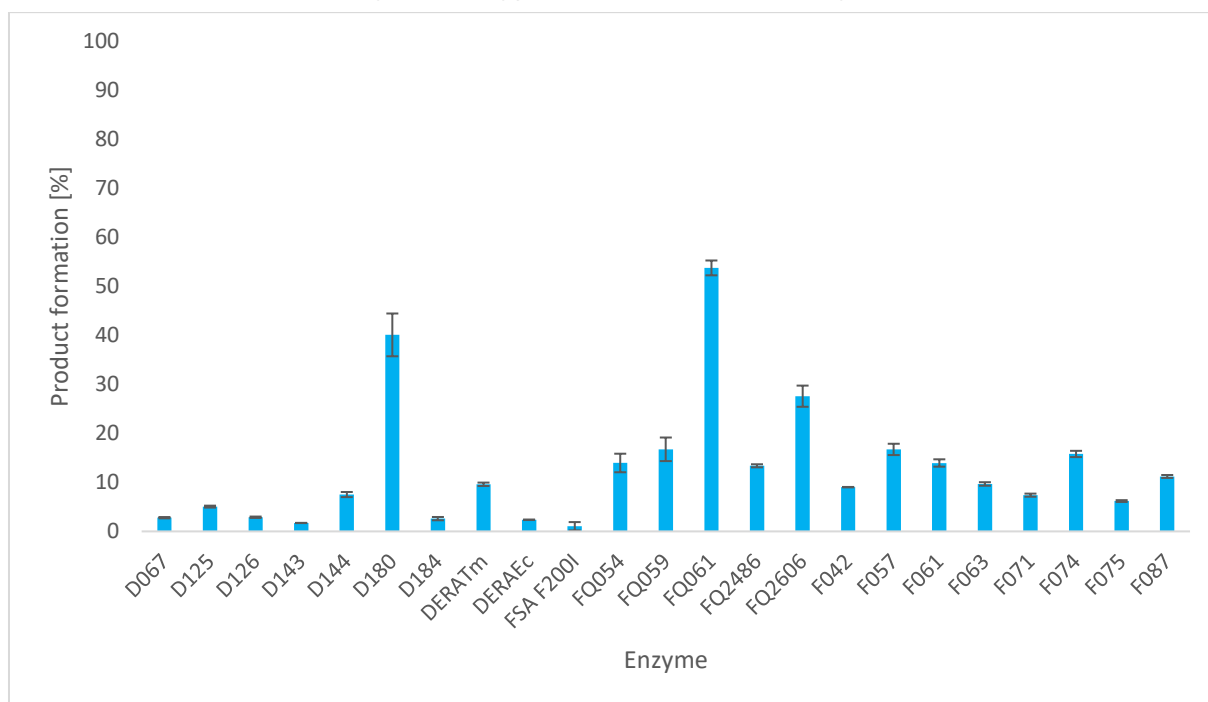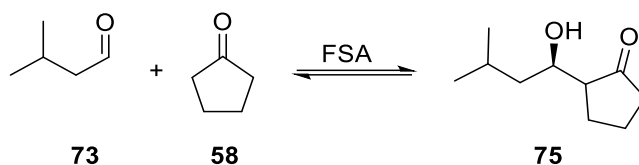
Figure 7 Results of the rescreening of active variants in cross aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**) (top) and cyclobutanone (**57**) (bottom) in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO.

The variations in both experiments concerning the product formation likely stems from reaction sample handling and sample preparation for HPLC-MS. For the measurement with HPLC-MS the sample needs to be derivatized, which was achieved under basic conditions with O-benzylhydroxylamine hydrochloride during which evaporation effects and sample handling can produce such error rates. However, low standard deviation, high reproducibility and in general the selected enzymes showed reproducible product formation useful for synthetic preparation of the desired products.

Compared to the EcDERA library, the diversity of the two metagenomic enzyme panels allowed to conduct reactions that were impossible with the mutagenic library. The questionably high yields with DERA125 and DERA180 in the screening runs could be confirmed in the preparative experiments making them excellent candidates for the synthesis of the product in sufficient yields. It is important to notice that cyclopentanone (**58**) and cyclobutanone (**57**) were claimed before as acceptable donors under certain criteria[39]. All following screenings will be performed with the same reaction conditions successfully applied in the screening for activity with cyclopentanone (**58**) and cyclobutanone (**57**).

4.1.4 Screening of metagenomic FSA library (D6Q mutations) in cross-aldol reaction of *i*-valeraldehyde (73) and cyclopentanone (58).

To achieve a broader selection of variants, it was decided to screen the same FSA library with D6Q mutations, which was also available from the earlier collaborative project. The selection of active variants from that library was solely based on activity in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**).

With the same conditions utilized in Section **3.1.1**, the metagenomic FSA library containing the D6Q mutation was screened for activity in the cross-aldol reaction of *i*-valeraldehyde (**73**) as acceptor and cyclopentanone (**58**) as donor. For comparison, DERA125, DERA126 and DERA180 were rescreened in that experiment. The following Figure 8 presents the results of the screening. Following conditions were utilized:

- 50 mM TEA*HCl buffer pH 7.5
- DMSO 10% v/v
- 100 mM *i*-valeraldehyde (**73**)
- 500 mM Cyclopentanone (**58**)
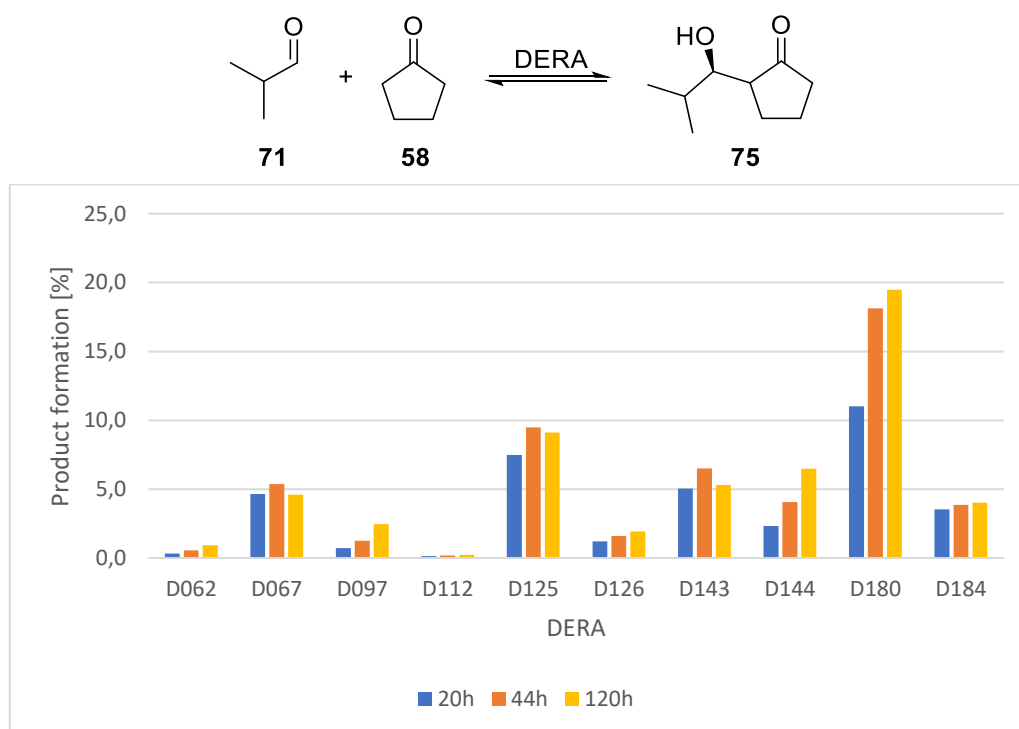- Enzyme CFE 8 mg/ml
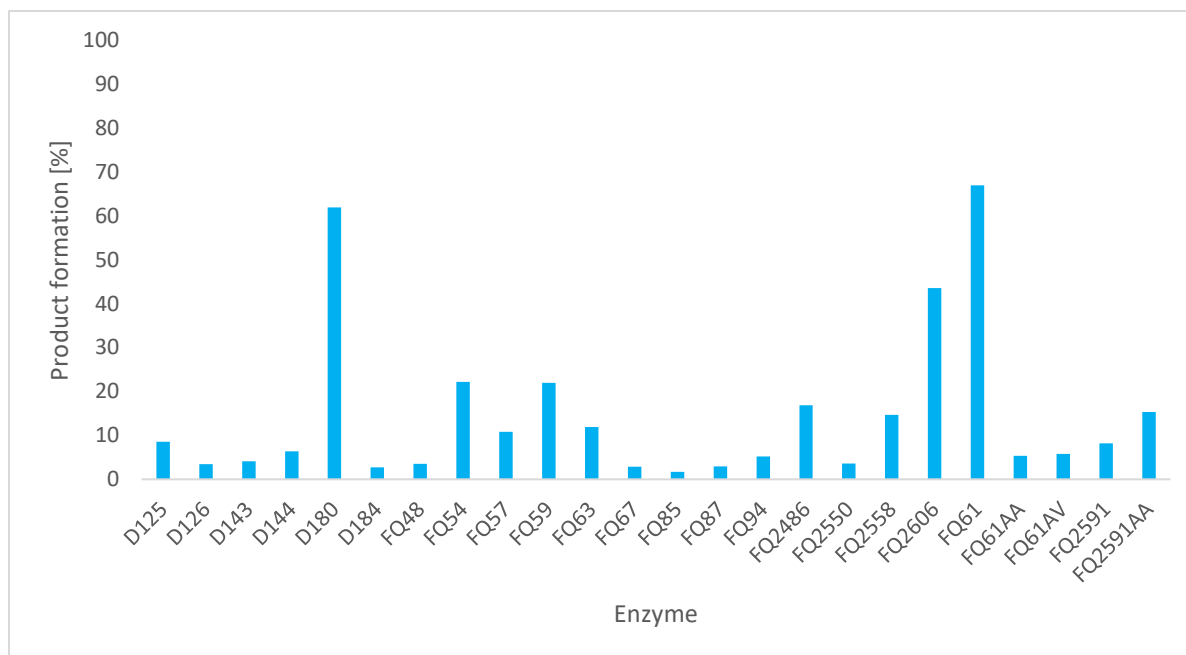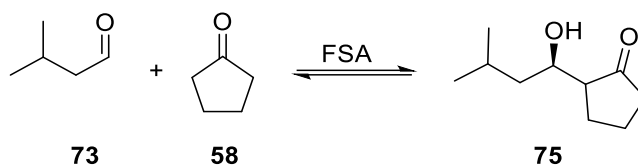
Figure 8 Results of the HPLC-MS measurement of the metagenomic FSA-library (D6Q mutation) in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**) with 1 : 5 ratio of acceptor to donor in 50 mм TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Interestingly, DERA180 showed a threefold higher product formation compared to the results produced with *i*-butyraldehyde (**71**) as acceptor demonstrating the disadvantage of *i*-butyraldehyde (**71**) for the screening. FSA FQ061 showed comparable product formation to DERA180. Interestingly, the introduced mutation D6Q resulted in increased activity for some variants like FSA061, which presented product formation values of ~50% in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**) (Figure 7), while with D6Q mutation ~70% product formation was achieved. Noteworthy, the D6Q mutation apparently can also have a negative impact, for variants FSA049, which had lost all its activity for the same reaction.

4.1.5 Preparative synthesis of 2-((1-hydroxy-3-methylbutyl)-cyclopentan-1-one (75)

For confirmation of the screening results by TLC and HPLC-MS, additionally an NMR spectrum was required for confirmation of the identity of the desired aldol product. Initially, the same conditions were applied for larger scale synthesis as for the screening test reactions, including the use of 10% DMSO as cosolvent.
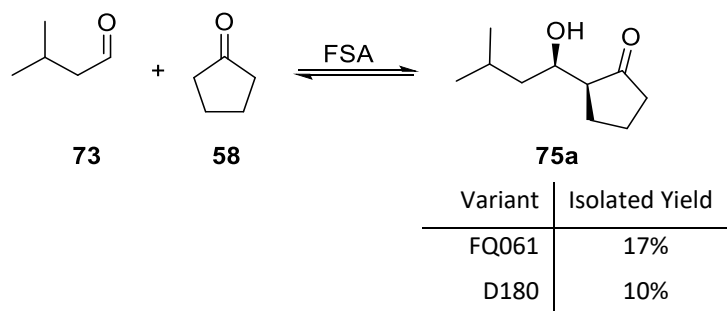
Scheme 3 Preparative aldol addition of *i*-valeraldehyde (**73**) (100 mM) as acceptor with cyclopentanone (**58**) (500 mM) as donor in 50 mM TEA*HCl buffer (pH 7.5) and 10% v/v DMSO as cosolvent and without any cosolvents.

Variant	Isolated Yield
FQ061	17%
D180	10%

After extraction of the product from the aqueous solution, it was necessary to wash the organic layer with water at least 6 times for the removal of DMSO. Unpleasantly, the product was re-extracted in significant amounts yielding significantly lower yields than when the same experiment was conducted without DMSO as cosolvent. The experiment with DMSO as cosolvent yielded for the variant D180 around 3% purified yield, while without DMSO for the same variant an isolated yield of 10% was obtained. The ^{1}H-NMR and ^{13}C-NMR of the purified product is presented in Figure 9+10.
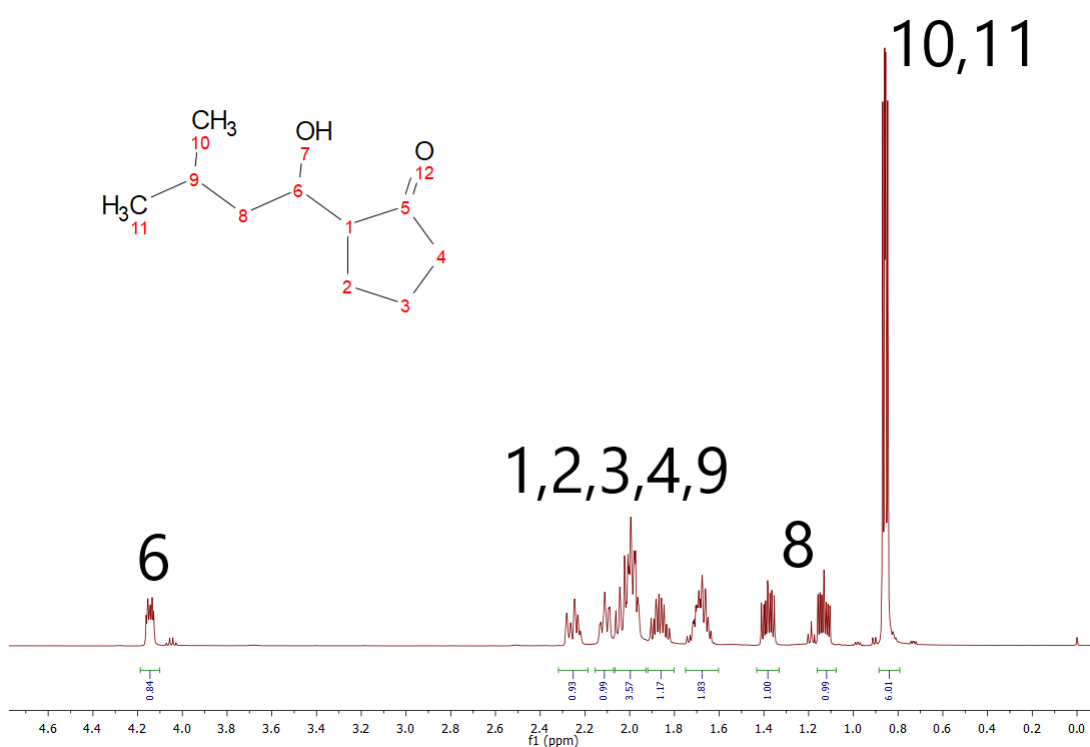
Figure 9 ^{1}H NMR of 2-(*R*)-1-hydroxy-3-methylbutyl-cyclopentan-1-one (**75**).
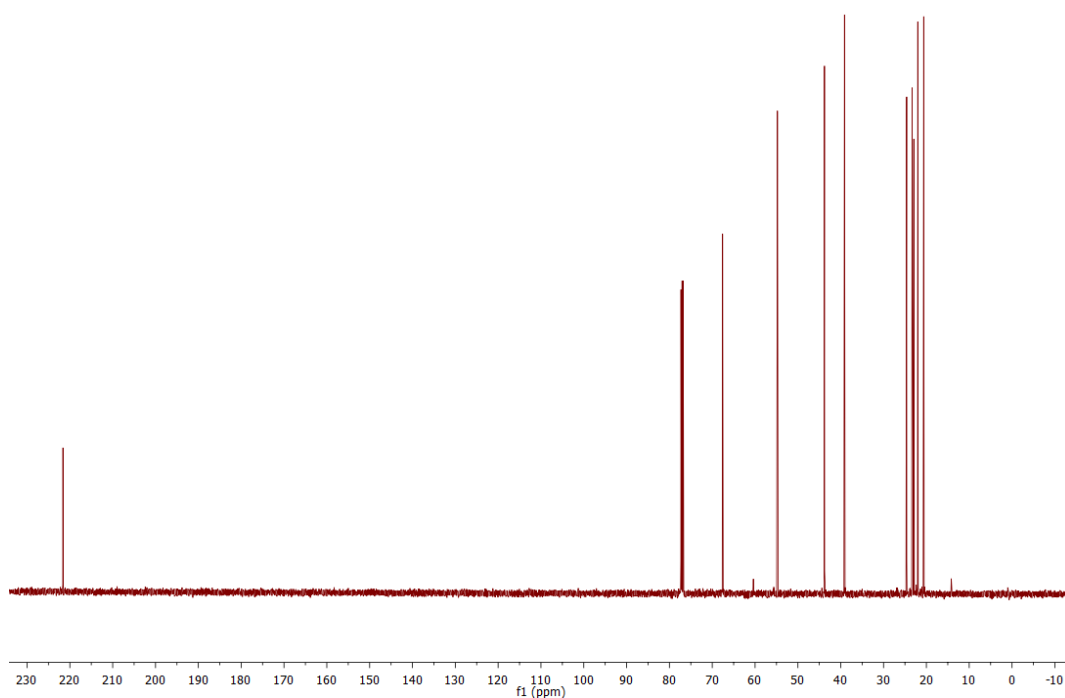
Figure 10 ^{13}C NMR (bottom) of 2-(*R*)-1-hydroxy-3-methylbutyl-cyclopentan-1-one (**75**).

Comparing the results with ^{1}H NMR spectra[120] from the literature, the isolated substance could be clearly identified as (*S*)-2-((*R*)-1-hydroxy-3-methylbutyl)cyclopentane-1-one (**75a**) based on the chemical shift of signal 6 (Figure 9). Minor amounts of (*R*)-2-((*R*)-1-hydroxy-3-methylbutyl)cyclopentane-1-one can be identified comparing the integrals of the signals in the area 3.8 and 4.2 ppm. The evaluation of the signals indicates a diastereomeric mixture of 93:7, which can be seen in Figure 11.
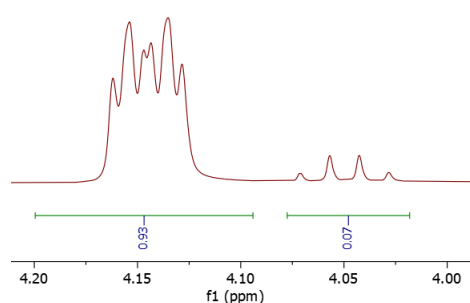
Figure 11 Signals of 1H NMR of aldol product (**75**) in the area of 4.0-4,2 ppm.

The general workup for the enzyme reactions starts with precipitation of the enzyme after two days of reaction time. This can be achieved by using methanol, acetone (**50**) or any other volatile solvent that can be mixed with water in at least three to one ratio. By adding the triple amount of an organic phase compared to reaction volume most enzymes will precipitate in less than 30 minutes. Centrifugation is the easiest method to remove the precipitated enzyme

since filtration can take hours to complete because the precipitated enzyme can clog the filter. After removal of the solvent by rotary evaporation, the water layer can be extracted with ethyl acetate. After removal of the ethyl acetate, the crude product is filtered through silica or when having a product mixture separated by column chromatography.

Some of the products prepared in this Chapter, and especially the product of cyclopentanone (**58**) and *i*-valeraldehyde (**73**), tend to eliminate the β-hydroxyl group (an NMR spectrum of the elimination product, which was separated during purification, can be found in the supporting information). This was especially the case for elevated temperatures, concentration effects and during NMR-sample preparation when using deuterated chloroform, which may contain minor amounts of hydrochloric acid.

4.1.6 Preparative synthesis of 2-(1-hydroxy-3-methylbutyl)cyclobutan-1-one (74)

Additionally, a preparative synthesis of the aldol product of *i*-valeraldehyde (**73**) as acceptor and cyclobutanone (**57**) as donor, namely 2-(1-hydroxy-3-methylbutyl)cyclobutan-1-one **(74)** was required. For that purpose, DERA180 was chosen because of the good activity in the initial screening, and since it was available in greater quantity.

Variant	Isolated Yield
D180	19%
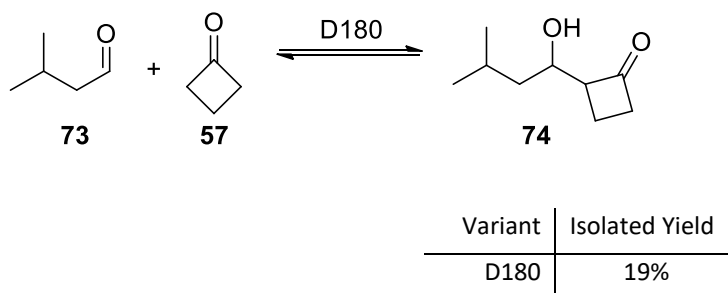
For the preparation of the 2-(1-hydroxy-3-methylbutyl)cyclobutan-1-one (**74**), no DMSO was utilized as cosolvent in consideration of the losses to be expected during removal of the DMSO by washing of the organic phase with water.

The aldol product was obtained with a yield of 19%. Compared to the product of *i*-valeraldehyde (**73**) and cyclopentanone (**58**), no elimination effects troubled the purification of the aldol product. The following Figures 12+13 present the ^{1}H NMR and ^{13}C NMR spectra, respectively.
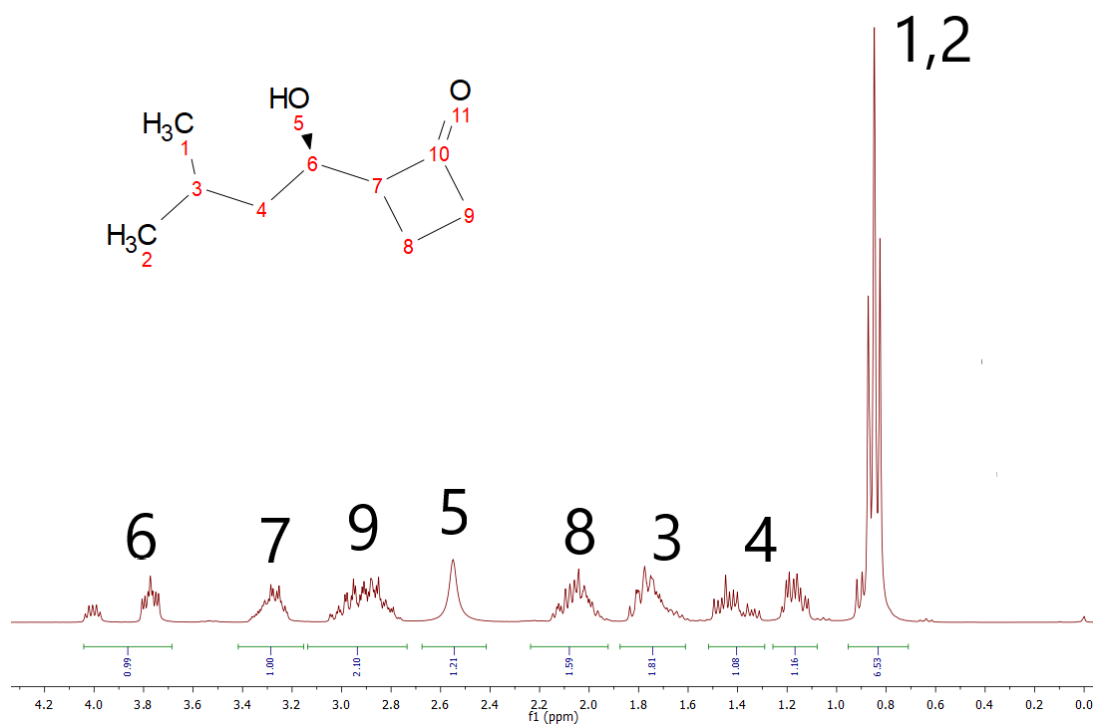
Figure 12 ^{1}H NMR of 2-(1-hydroxy-3-methylbutyl)cyclobutan-1-one (**74**).
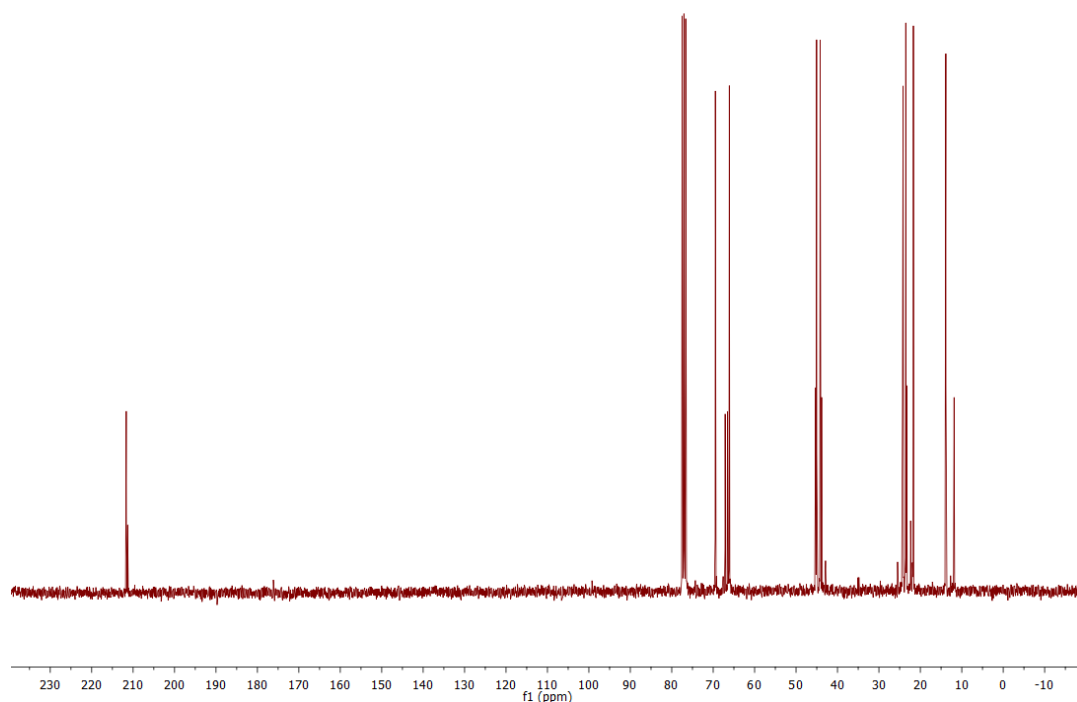
Figure 13 ^{13}C NMR of 2-(1-hydroxy-3-methylbutyl)cyclobutan-1-one (**74**).

The compound could be identified as 2-(1-hydroxy-3-methylbutyl)cyclobutan-1-one. Both 1H and 13C NMR spectra, but especially the latter, indicated a diastereomeric mixture of (1*R*, 2*S*)-2-(1-hydroxy-3-methyl-butyl)cyclobutan-1-one and (1*R*, 2*R*)-2-(1-hydroxy-3-methyl-butyl)cyclobutan-1-one. Comparing the obtained 1H NMR spectrum with results from the

literature[120], the main product could be identified as (*R*)-2-((*R*)-1-hydroxy-3-methylbutyl)cyclobutan-1-one (**87**) while (S)-2-((R)-1-hydroxy-3-methylbutyl)cyclobutan-1-one (**86**) was produced in smaller amounts.
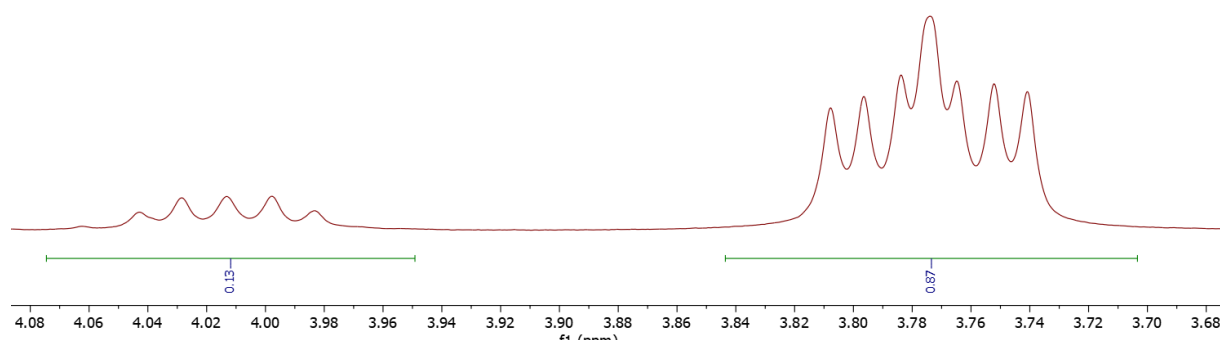
Figure 14 Signals of 1H NMR of aldol product (75) in the area of 3.7-4,1 ppm.

Figure 15 shows both possible diastereomers of the aldol product. It is important to notice that the hydroxyl group of the discussed aldol product always shows (*R*)-orientation. This can be explained mechanistically by the enamine nucleophilic attack of the *si* face of the electrophile[97].

86

(*S*)-2-((*R*)-1-hydroxy-3-methylbutyl)cyclobutan-1-one

87
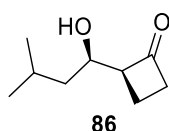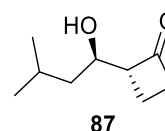
(*R*)-2-((*R*)-1-hydroxy-3-methylbutyl)cyclobutan-1-one

Figure 15 Both possible aldol products of the aldol reaction of *i*-valeraldehyde (**73**) and cyclobutanone (**57**) (S)-2-((R)-1-hydroxy-3-methylbutyl)cyclobutan-1-one on (**86**) the left and (R)-2-((R)-1-hydroxy-3-methylbutyl)-cyclobutan-1-one (**87**) on the right side.

The occurrence of a diastereomeric mixture does not necessarily result from the enzymatic reaction itself but results from subsequent of keto-enol-tautomerism, which can occur during workup at non-neutral pH. Nonetheless, the product identity could be confirmed by NMR analysis and by HPLC-MS during the screening. With the results of the initial screening now in hand, the next step was the screening for unknown donor activity in Section **4.2.**

4.2 Screening of the selected enzymes for novel donor activity with *i*-valeraldehyde (73) as acceptor

One important factor that is worth mentioning is that the selection of enzymes could not be based on results of all Screenings in section **4.1** and had to be decided before having all results available due to disposition issues with the British partner Prozomix after the Brexit. Therefore, some interesting variants could not be received in greater quantities than on a 20 – 200 mg CFE scale and were therefore only used on an analytical scale.

After the initial screening for activity with cyclic ketones, the selection of most promising enzymes was tested for novel unknown donor activity. The selection from the initial screening contained the DERA orthologs DERA019, DERA029, DERA062, DERA067, DERA090, DERA125, DERA126, DERA135, DERA143, DERA144, DERA147, DERA180, DERA190, DERA192. From the metagenomic FSA library FSA049, FSA054, FSA055, FSA063, FSA071, FSA075 were selected and also the FSA orthologs of the metagenomic FSA library with D6Q mutation FQ054, FQ059, FQ061, FQ2486, FQ 2588 and FQ2606 were additionally selected for a broader and more diverse selection. All over 26 active variants were selected and as control *Ec*DERA (F200I), DERA from *Thermotoga maritima* (*Tm*DERA), the empty plasmid BL21 background protein mixture (EP) and a blind sample without enzyme (Bl) were chosen to identify background activity not related to the DERA and FSA catalysts.

The selected donors to be tested were butanone (**55**), fluoroacetone (**77**), chloroacetone (**66**), trifluoroacetone, methoxyacetone (**79**), dimethoxyacetone (**80**), 1-hydroxybutan-2-one (**33**), 2-pentanone (**81**), 3-pentanone (**65**), cyclopentanone (**58**), dihydrofuran-3(2H)-one (**63**), 2-methyldihydrofuran-3(2H)-one (**82**), dihydrothiophen-3(2H)-one (**84**), cyclobutanone (**57**), 3-oxetanone (62), cyclohexanone (**61**), dihydro-2H-pyran-3(4H)-one **(85)**.

In the literature, there is only one example of butanone (**55**) reported as donor for *Ec*FSA variant D6H[122]. In the example, butanone (**55**) was used as aldol donor with the natural acceptor GA3P (**3**).

For DERA there is only one example of butanone (**55**) reported as donor[39]. In the example, butanone (**55**) was cleaved in a retro-aldol fashion as donor with the natural acceptor D-glyceraldehyde 3-phosphate, while no example is available for any other acceptors or in the synthesis direction to produce the aldol product.

Additionally, 3-pentanone (**65**), cyclohexanone (**61**) and 2-hexanone were tested. The cyclic oxo/thio-variants of cyclobutanone (**57**), cyclopentanone (**58**) and cyclohexanone (**61**) had never appeared in the literature in context with aldolases and were therefore selected as possible new donors.

Apparently fluoroacetone (**77**) was never tested as possible donor for FSA, while being identified to work with wildtype *Ec*DERA. The next bulkier derivative chloroacetone (**66**) would be an unknown activity for both enzyme classes and was therefore tested.

All screening samples were produced in unicates only because of the goal of identifying suitable candidates for preparative scale, which is perfectly achievable with single samples.

4.2.1 Screening of selected DERA and FSA enzymes for activity with fluoroacetone (77)

While fluoroacetone (**77**) is accepted by *Ec*DERA wild type, no examples were published for any FSA variants. Therefore, we were interested to see, if any FSA type can accept fluoroacetone (**77**) as donor at all and how this compares to DERA variants. In principle three different products are possible for the cross-aldol reaction to *i*-valeraldehyde (**73**) as acceptor with fluoroacetone (**77**) as donor. The first possible product is the linear product with the fluorine atom at C1 whereas the second and third products are the branched product with fluorine at C3 in two different stereo configurations. The results of the TLC screening and the possible reaction products are presented in Figure 16.

Figure 16 Reaction scheme of cross aldol addition of *i*-valeraldehyde (**73**) (black) as acceptor and fluoroacetone (**77**) (blue) as donor with the newly formed bond (pink) coloured and the obtained yields in larger scale (top). Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of *i*-valeraldehyde (**73**) and fluoroacetone (**77**) based on TLC analysis (2:1 cyclohexane/ethylacetate) with anisaldehyde staining (bottom).

No product formation was detected for the negative control samples using the protein background of BL21 (EP) and the chemical background (Blind). While *Ec*DERA and *Tm*DERA both were active in the cross-aldol addition of fluoroacetone (**77**) to *i*-valeraldehyde (**73**), *Ec*FSA showed only minimal product formation on TLC. *Tm*DERA showed the formation of two different product spots, while *Ec*DERA produced solely one product staining as a blackish spot on TLC. In general, most DERA variants i.e. DERA029, DERA126, DERA135, DERA143, EcDERA, DERA144 showed high selectivity for the product that resulted in a blackish spot on TLC after staining with anisaldehyde, while the other DERA enzymes showed slight preference for one of both products. All FSA samples showed high selectivity for the other product presenting a brownish spot on TLC.

For the most active variants, samples for HPLC-MS analysis were prepared. The linear and branched product overlapped during HPLC-MS analysis and could not be separated with optimized conditions. Figure 17 presents the results of the HPLC-MS analysis.
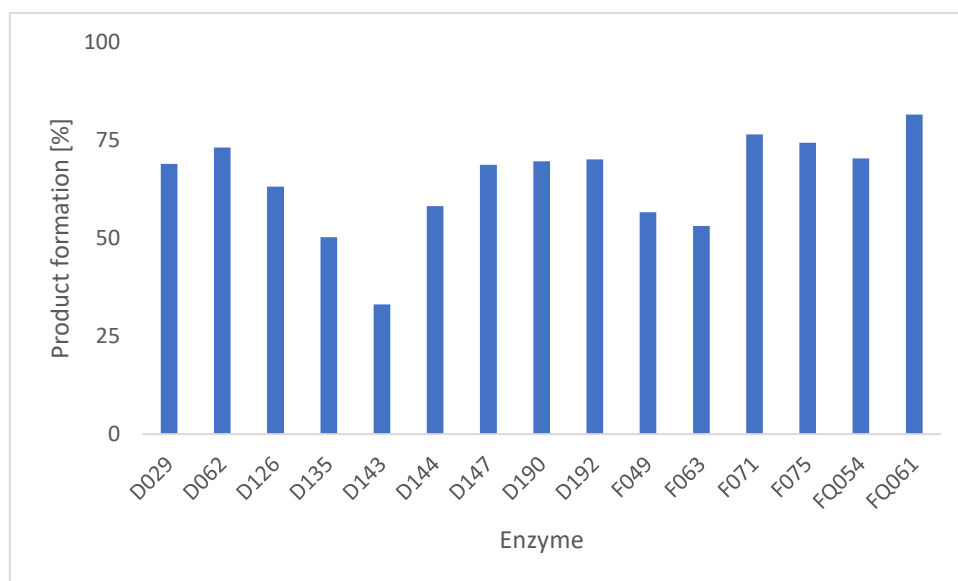
Figure 17 Results of the screening of most active variants in cross aldol reaction of *i*-valeraldehyde (**73**) (100 mM) and fluoroacetone (**77**) (500 mM) in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO.

For identification of the product spots, FQ061 and DERA029 were used in a preparative scale reaction with 200 mL reaction volume each. These enzymes were chosen because both produced mostly one of the two product spots making separation of the products more convenient.

The upper spot from FQ061 sample contained mainly one of branched products and no linear product could be isolated during purification. The product was obtained with excellent yield of 42%. It is important to notice that *Ec*DERA shows a two-digit lower activity for fluoroacetone (**77**) when utilizing the natural donor GA3P (**3**). Therefore, such high yields came unexpected when changing to an aliphatic acceptor.

Figure 18 presents the obtained ^{1}H/^{13}C NMR spectra of the main product fraction produced with the FQ061 variant (brownish spot on TLC) corresponding to the formation of the branched aldol product.
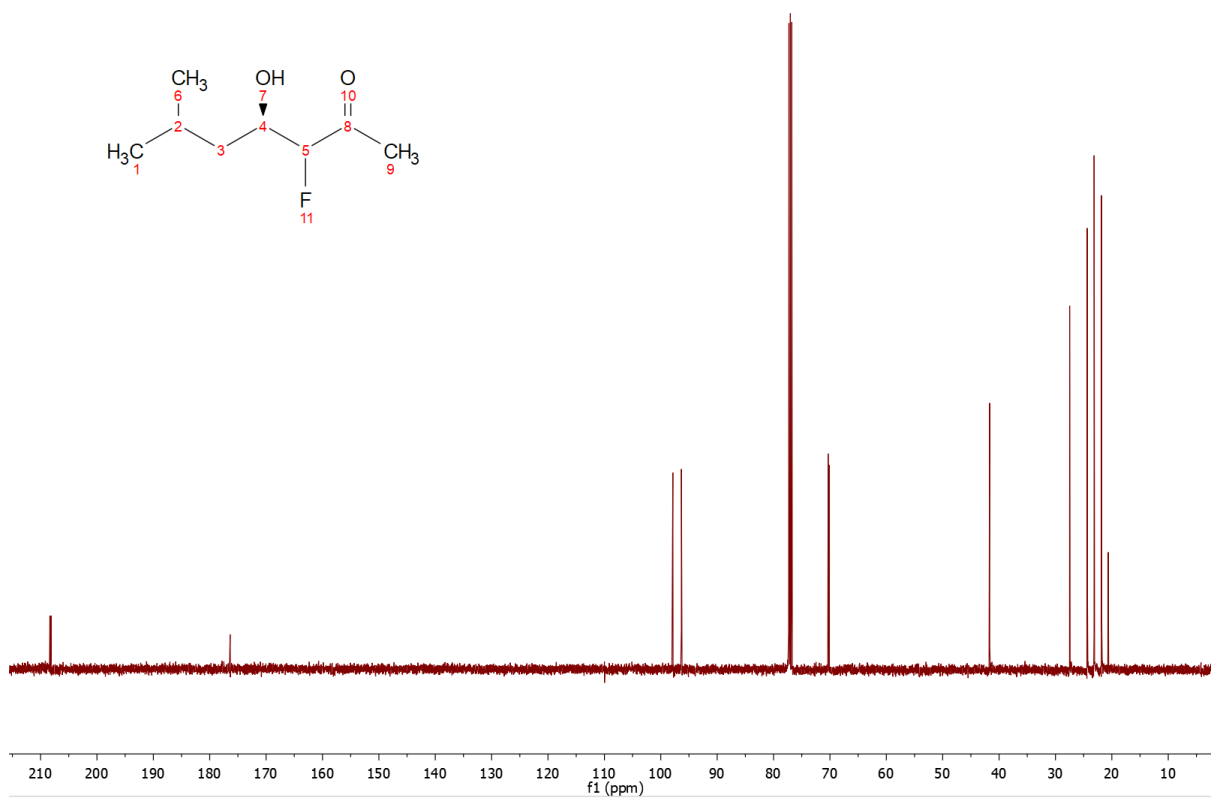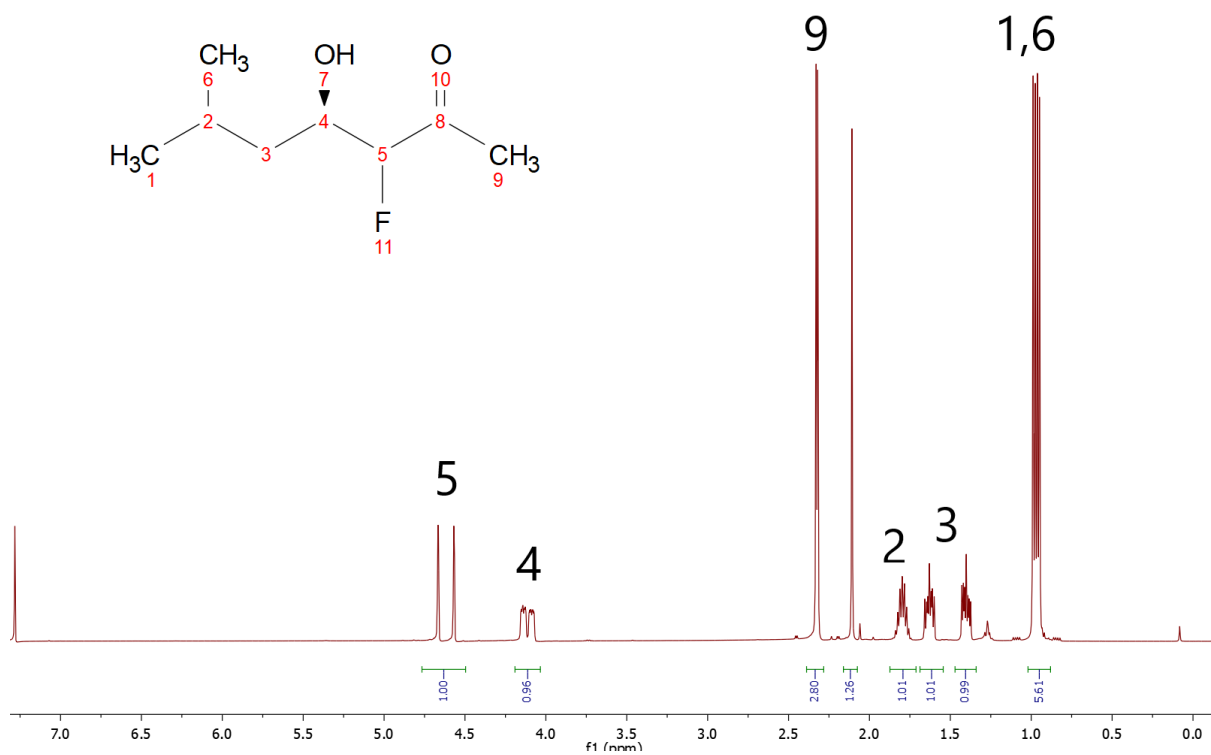
Figure 18 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of the branched aldol product (**89**).

The ^{1}H NMR and ^{13}C spectrum indicate a racemic mixture based on signals for position 4+5 with a chemical shift of 4.0-4.75 confirmed by the signals in 13C spectrum at 70 ppm and 95-100 ppm.

For the second product spot (blackish spot on TLC), DERA029 was utilized in the preparative scale experiment. The product was obtained with an excellent yield of 48%. The ^{1}H/^{13}C NMR-spectra of the linear aldol product produced with DERA029 are presented in Figure 19.
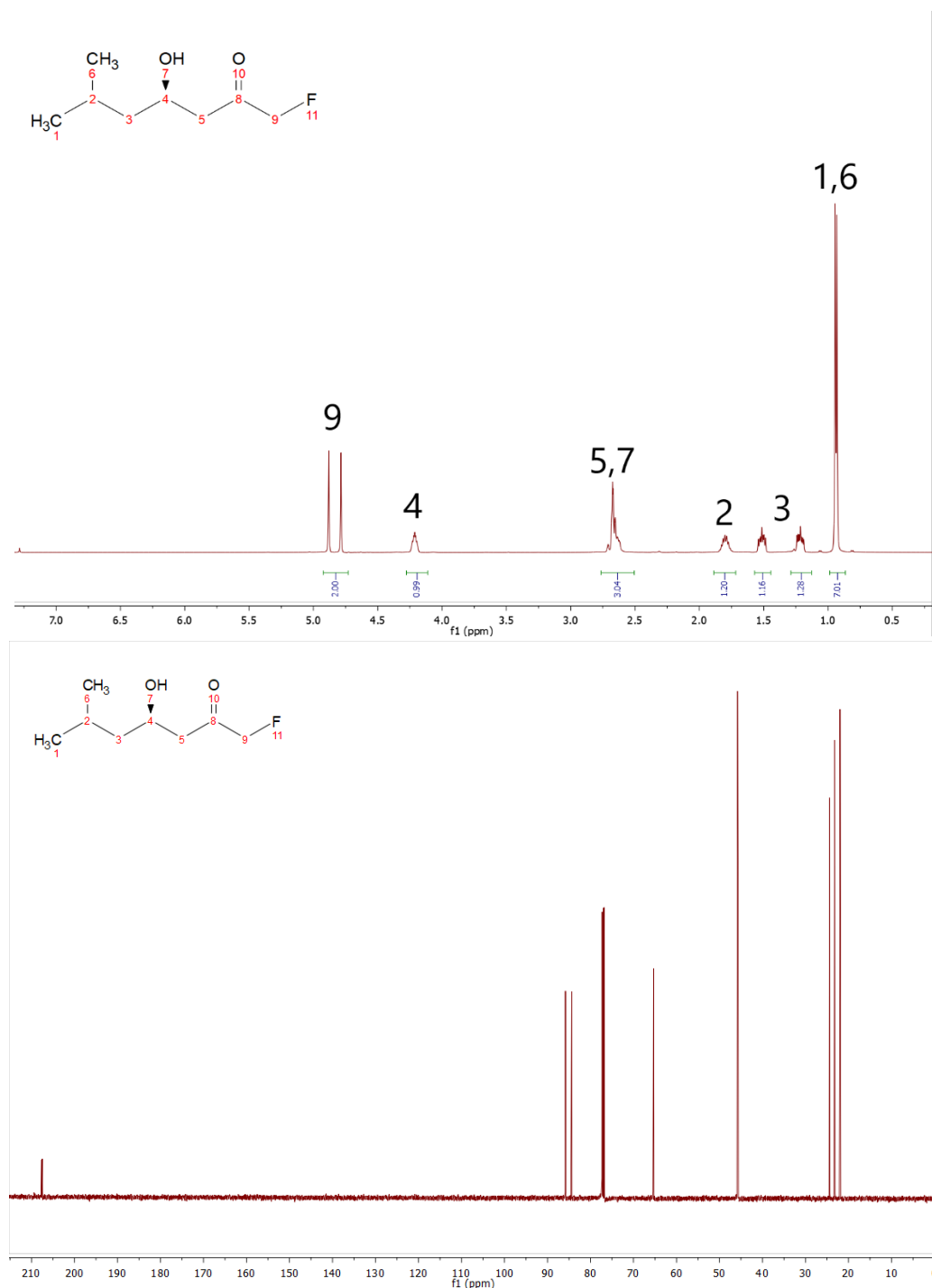
Figure 19 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of the linear aldol product (**88**) of i-valeraldehyde (**73**) as acceptor and fluoroacetone (**77**) as donor.

This result is impressive due to both products being individually accessible by catalysts from complementary aldolase families with good selectivity, demonstrating the synthetic value of native aldolases from different sources. While all FSA enzymes clearly favored the branched product, the situation for DERA samples was different. Some DERA enzymes, e.g. DERA019, DERA090 and *Tm*DERA, produced the branched product with little amounts of the linear product. Some even produced a mixture with no clear preference, such as DERA125, DERA147 and DERA190 which showed only minor discrimination for one of the products. The enzymes DERA029, DERA062, DERA126, DERA135, DERA143 and wildtype *Ec*DERA (F200I) showed a clear selectivity towards the linear product with almost no branched product detectable on TLC and HPLC-MS.

During the screening of fluoroacetone (**77**) as an accepted donor, further screening with additional acceptors deemed unnecessary due to fluoroacetone (**77**) being accepted as a donor and should therefore be applicable with many different acceptors.

4.2.2 Screening of selected DERA and FSA enzymes for activity with chloroacetone (66)

In light of the positive results for fluoroacetone (**77**) as donor, it was decided to test if some of the selected variants can also accept the related chloroacetone (**66**) as donor. In the literature up to now, there was no example of DERA or FSA enzymes accepting chloroacetone (**66**) as donor. Therefore, any product formation would be a great starting point for further protein engineering to optimize the enzymes for chloroacetone (**66**) as donor. Figure 20 presents the results of the TLC screening of the cross-aldol reaction of *i*-valeraldehyde (**73**) as acceptor and chloroacetone (**66**) as donor.
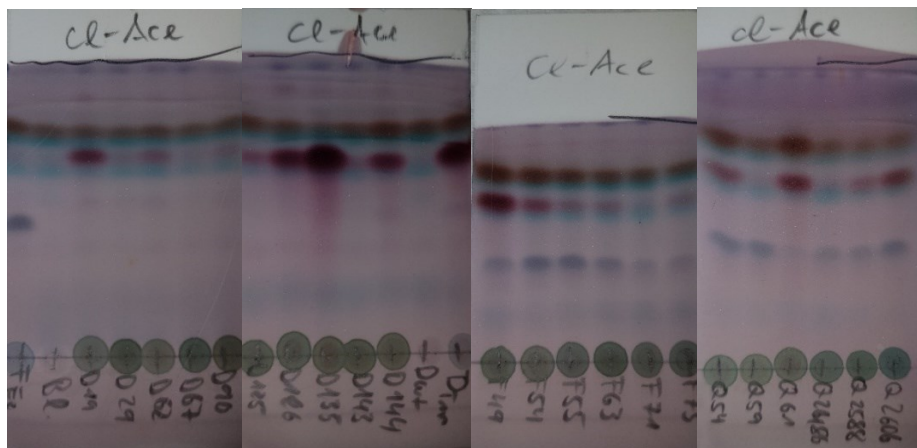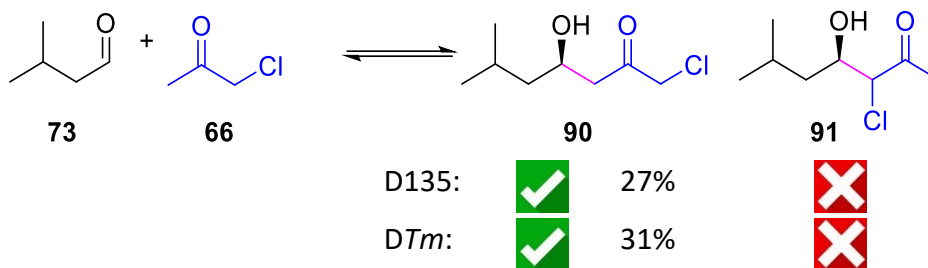
Figure 20 Reaction scheme of cross aldol addition of *i*-valeraldehyde (**73**) (black) as acceptor and chloroacetone (**66**) (blue) as donor with the newly formed bond (pink) colored and the obtained yields in large scale (top). Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of *i*-valeraldehyde (**73**) and chloroacetone (**66**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

From the results, it is clearly visible that most positive hits are produced by DERA variants e.g., DERA135 shows the highest conversion followed by *Tm*DERA, while most FSA variants were unable to form the product. From all reactions with positive hits, accounting to 30% of all variants, samples were prepared for HPLC-MS analysis. However, the results were inconclusive since peak overlapping resulted in non-meaningful data even with differently modified methods for HPLC-MS measurements.

For most FSA enzymes an additional new, blueish spot on TLC was stained by anisaldehyde. Yet, in the HPLC-MS analysis no new product was detected when compared to the sample produced using DERA enzymes. The most plausible explanation is that spontaneous hydrolysis of chloroacetone (**66**) to hydroxyacetone (**32**) occurred, which was then used as an excellent donor, producing the cross-aldol product of *i*-valeraldehyde (**73**) and hydroxyacetone (**32**).

It was decided to produce the aldol product utilizing chloroacetone (**66**) as donor by using D135 and *Tm*DERA. Both enzymes had performed outstandingly in other reactions with different donors. Due to this, it was decided to utilize these catalysts for comparison in preparative scale reaction.

Both enzymes produced the linear product with high rates and similar isolated yield for DERA variant D135 a yield of 27% was obtained, while 31% for *Tm*DERA. Figure 21 presents the ^{1}H and ^{13}C NMR spectra.
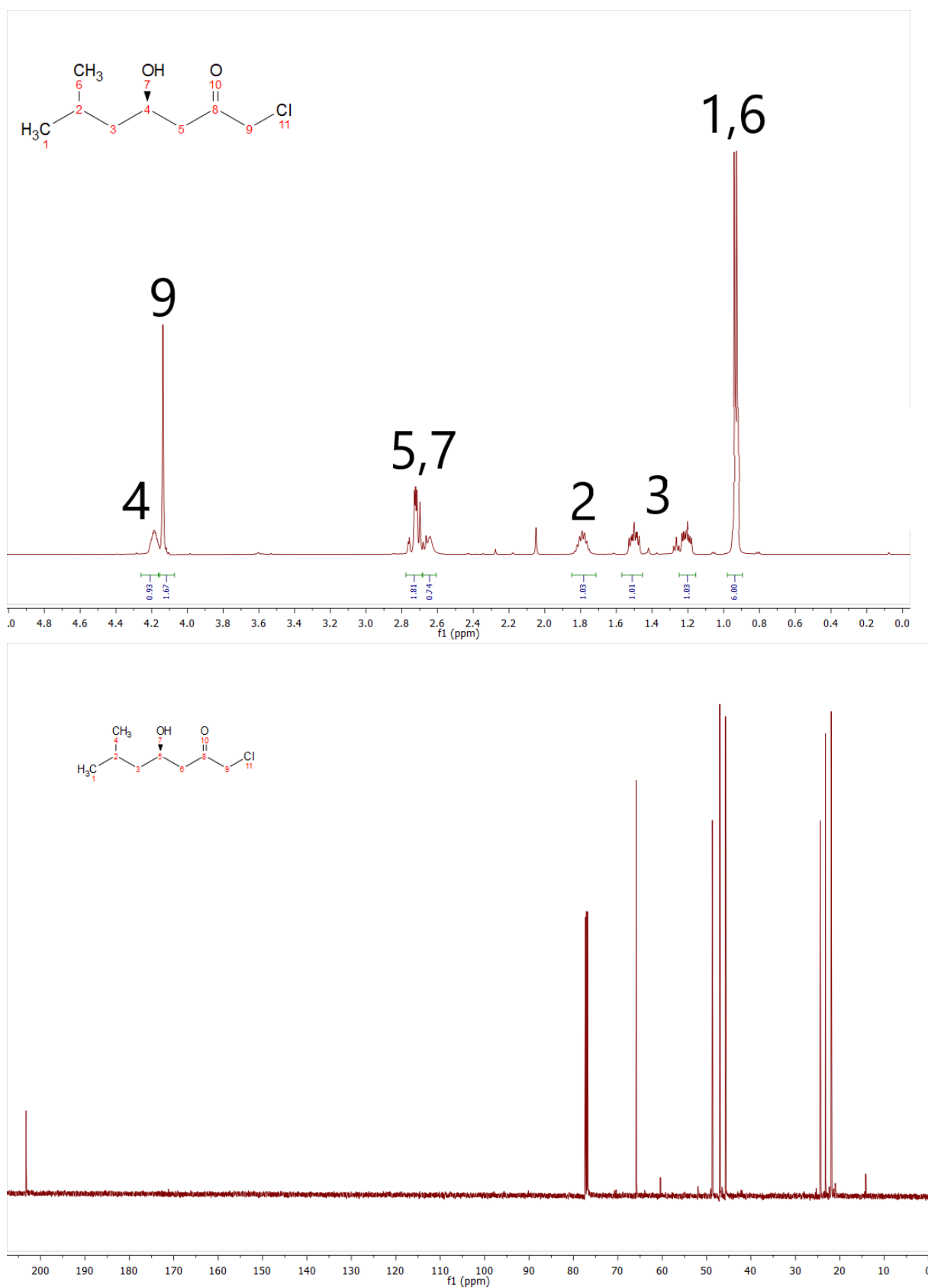
Figure 21 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of the linear aldol product (**90**).

While the same product should be available through classic organic chemistry, the outcome of the reaction is in some cases with different acceptors not as selective as with aldolases[155-

[156]. Additionally, reaction and purification conditions can reach critical points resulting in partial elimination of the valuable β-hydroxylgroup of the product resulting in unwanted side products[157]. Comparing the possible yields of the reaction, aldolases lack in that sense. But since the enzyme or the organocatalyst is in most cases the cost relevant factor, aldolases are easily and cost efficient prepared from frozen cell stocks.

4.2.3 Screening of selected DERA and FSA enzymes for activity with butanone (55)

For butanone (**55**) there is only one example in the literature related to its potential reactivity as donor for DERA enzymes. In this special case GA3P (**3**) was the electrophile compensating as the natural acceptor for the lack of activity for butanone (**55**) as a weakly accepted donor[39]. The natural donor GA3P (**3**) has an attractive feature, namely the β-hydroxyl group which allows the product to form the cyclized form by an intramolecular nucleophilic attack of the β-hydroxyl group on the ketone group. It would be a great step forward, using a non-natural, unphosphorylated and aliphatic acceptor with butanone (**55**) as donor such as *i*-valeraldehyde (**73**). Figure 22 presents the results of the TLC screening and the results from preparative scale reactions.
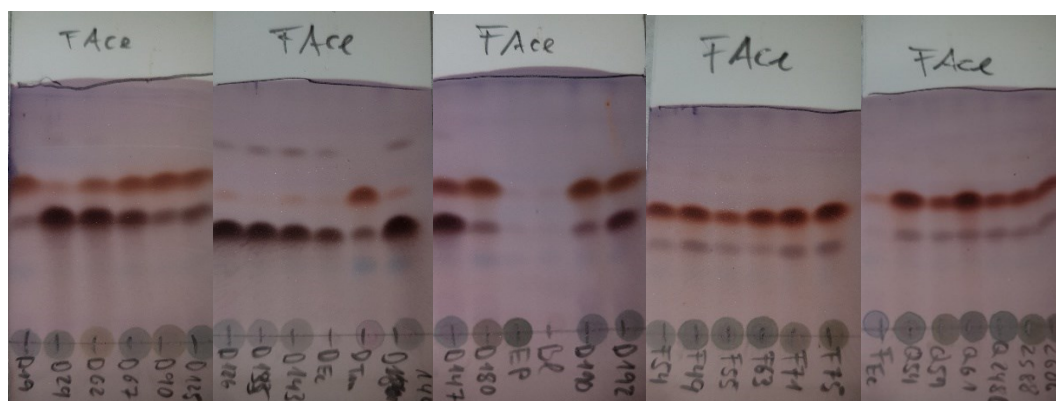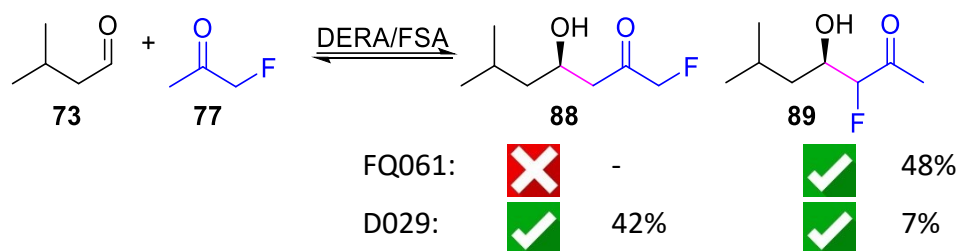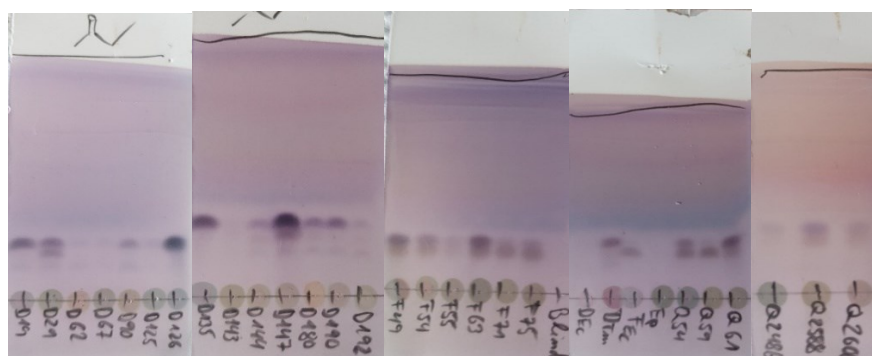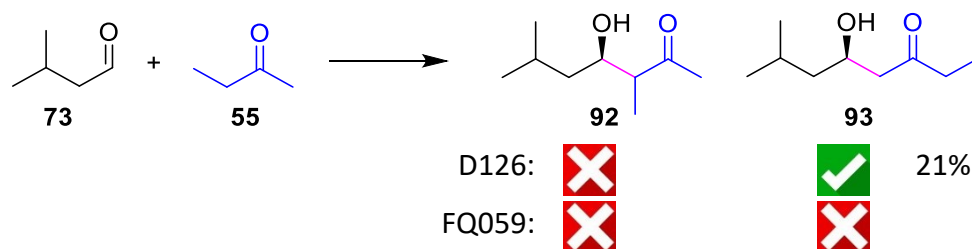

Figure 22 Reaction scheme of cross aldol addition of *i*-valeraldehyde (**73**) (black) as acceptor and butanone (**55**) (blue) as donor with the newly formed bond (pink) colored and the obtained yields in big scale (top). Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of *i*-valeraldehyde (**73**) and butanone (**55**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

Amazingly, around half of the tested enzymes were able to convert butanone (**55**) as donor, mostly producing two product spots. Important to notice, all educts were evaporated from the TLC sheet by careful heating with a heat gun before the anisaldehyde staining. Therefore,

by this procedure all visible spots should correspond to the desired products even though most samples present spot intensities related to traces of the product only.

The enzyme variants DERA126, DERA135, DERA147, FSA063 and FQ061 were selected for HPLC-MS analysis. The results of the HPLC-MS analysis are presented in Figure 23.

| Variant	Product formation [%]
DERA126	16.8
DERA135	9.8
DERA147	24.0
FSA063	6.4
FQ061	6.6
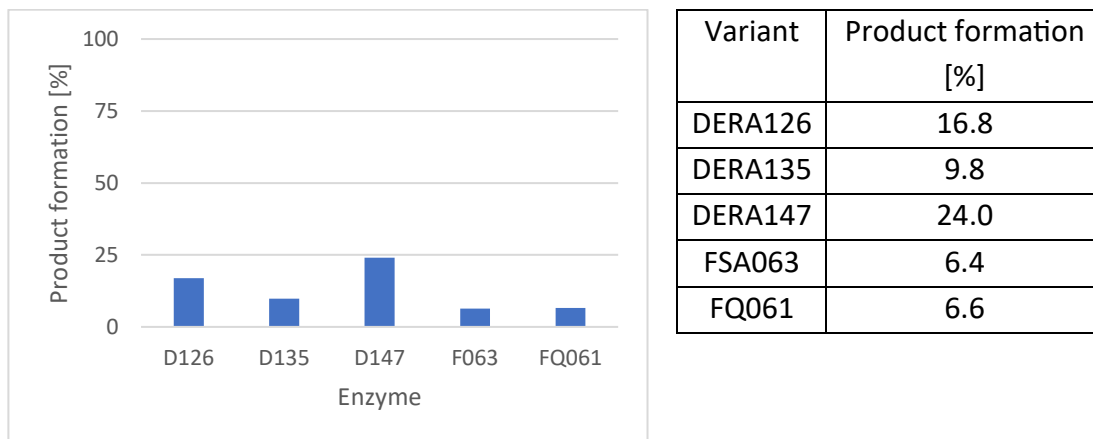
Figure 23 Results of the HPLC-MS screening of most active variants in the cross-aldol reaction of *i*-valeraldehyde (**73**) (100 mM) and butanone (**55**) (500 mM) in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO.

It was decided to proceed with larger scale reactions using variant DERA126 for analysis of the upper spot, while utilizing enzyme FQ059 for analysis of the bottom spot. In principle three different products are possible from the cross-aldol addition of butanone (**55**) to *i*-valeraldehyde (**73**).

The initial interpretation of the results from TLC screening was that one spot should correspond to the linear product, while the other spot would correspond to the branched product, which can occur as two diastereomers.

From the reaction with enzyme FQ059 no product could be isolated in preparative scale reaction. The most plausible reason is that the product concentration was too low even though a spot on TLC was visible. From experience extremely low amounts of aldols can produce clearly visible spots on TLC after staining.

Meanwhile, the reaction utilizing DERA126 gave good results. The isolated yield in large scale with DERA126 was determined to be 21%. Considering that the only result in the literature with butanone (**55**) as donor was with the usage of the natural acceptor, being able to switch the acceptor to a non-phosphorylated, aliphatic acceptor is a tremendous achievement. The obtained yield of 21% is considered a good yield in this case. Figures 24 and 25 present the ^{1}H and ^{13}C NMR spectra of the linear aldol product of *i*-valeraldehyde (**73**) as acceptor and butanone (**55**) as donor.
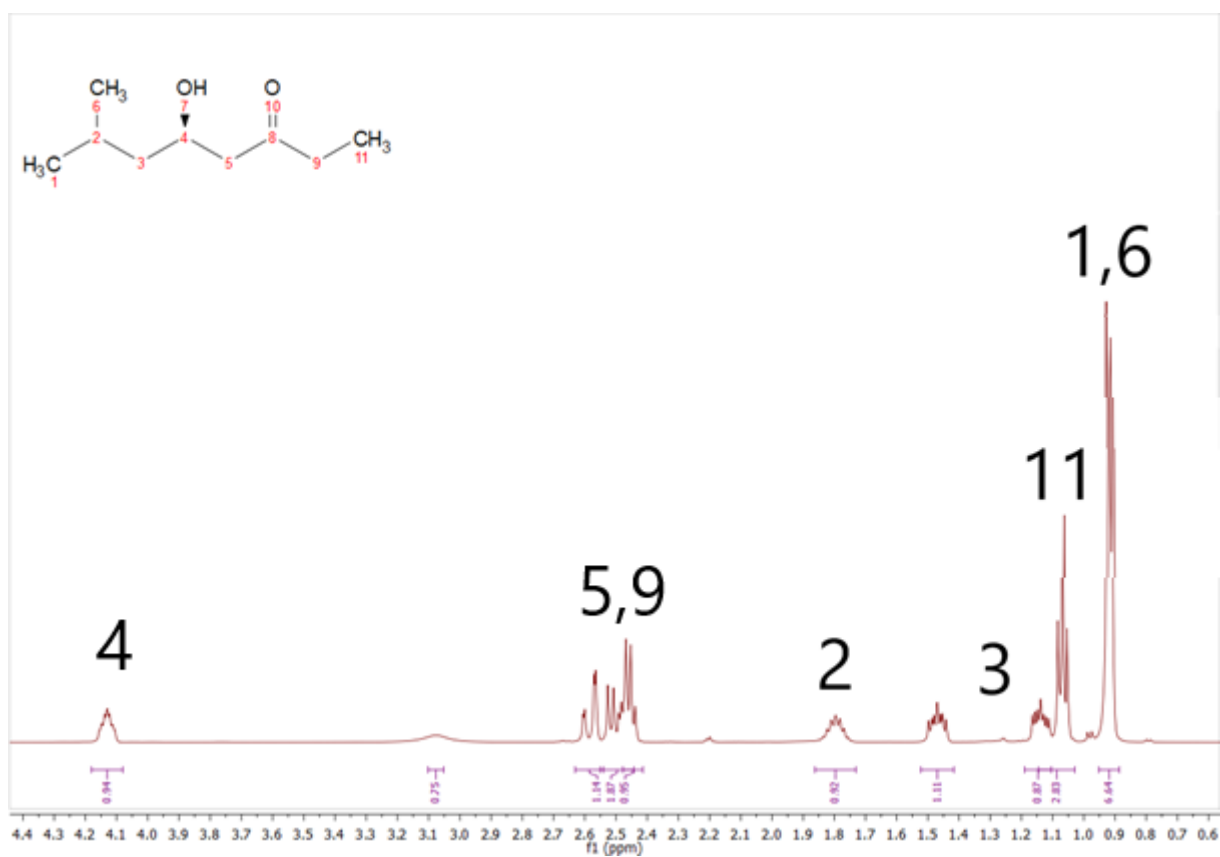
Figure 24 ^{1}H-NMR of (R)-5-hydroxy-7-methyloctan-3-one (**93**).
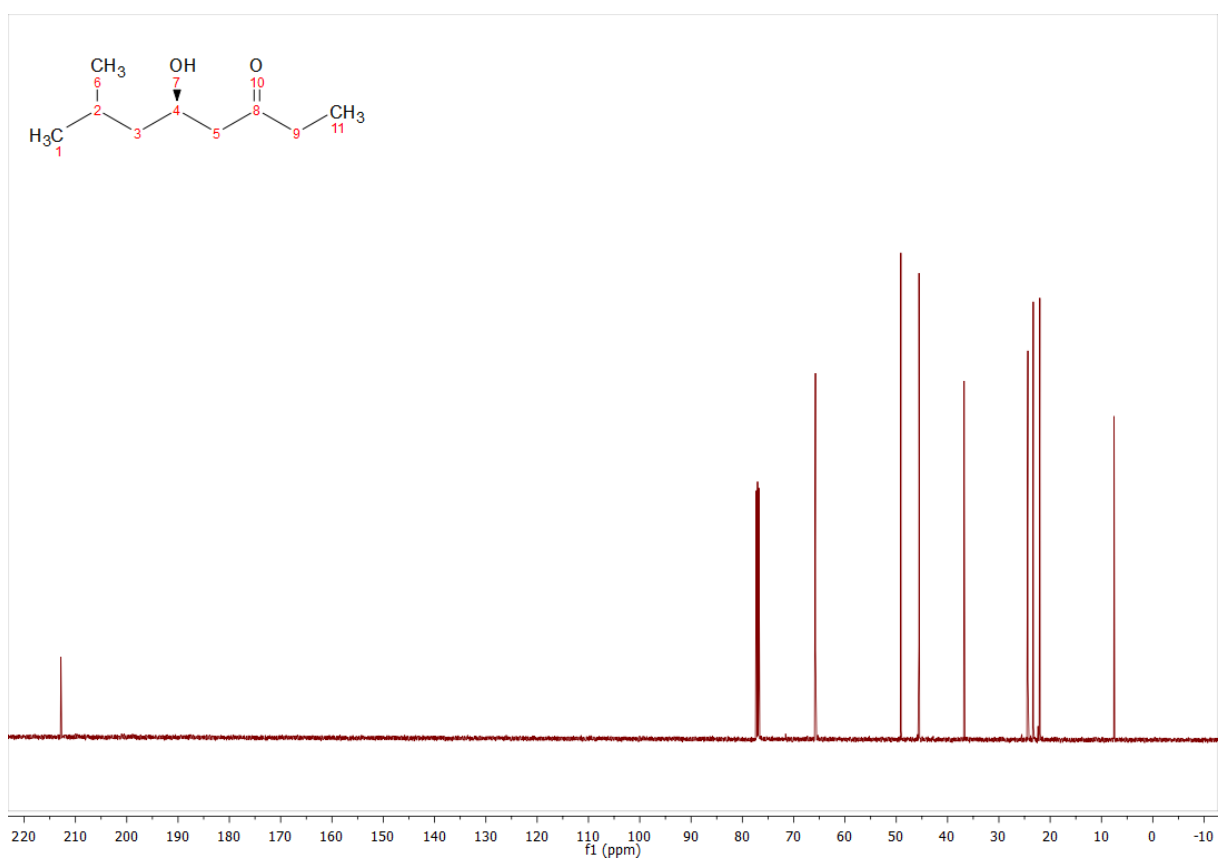
Figure 25 ^{13}C-NMR of (R)-5-hydroxy-7-methyloctan-3-one (**93**).

Obviously, the upper spot on TLC corresponds to the linear aldol product while the bottom spot most likely corresponds to the branched aldol product. In all samples, only spot intensities were identified corresponding to traces of the product, indicating that all variants could not properly orientate butanone (**55**) in the active center for producing the branched aldol product. The branched product can only be considered a side product in this experiment. But nonetheless, protein engineering may possibly enable suitable enzymes to form the branched product.

4.2.4 Screening of selected DERA and FSA enzymes for activity with methoxyacetone (79)

With hydroxyacetone (**32**) being a donor for wildtype FSA, there was hope that some of the FSA variants will be able to accept even methoxyacetone (**79**) as donor. For the DERA variants hits were less likely to be expected. The results of the screening of the aldol addition of *i*-valeraldeyhde as acceptor and methoxyacetone (**79**) as donor with the standard screening conditions and the results from preparative scale reaction are presented in Figure 26.
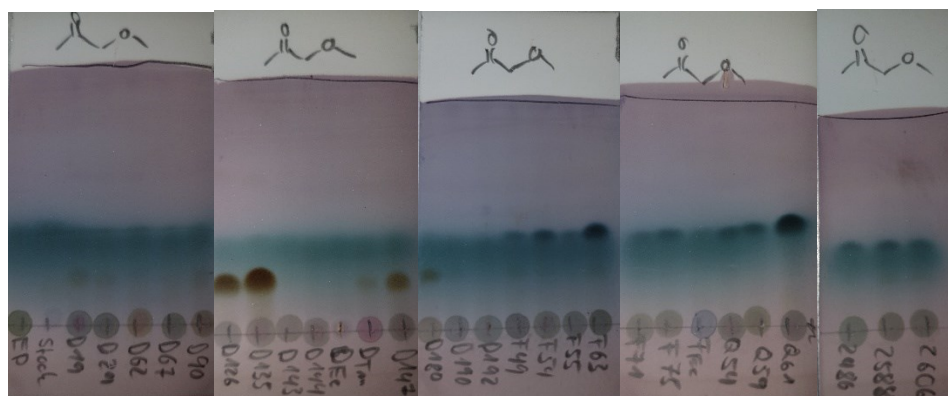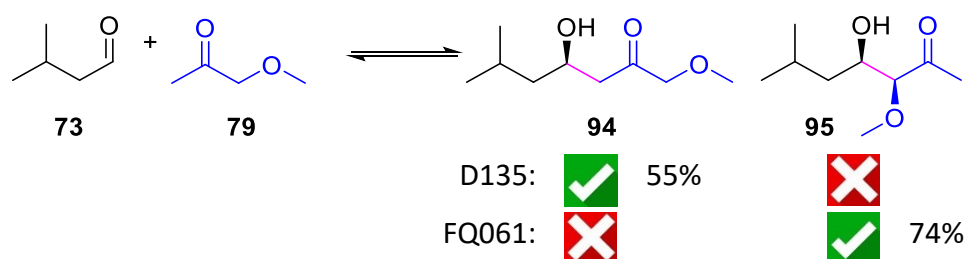
Figure 26 Reaction scheme of cross aldol addition of *i*-valeraldehyde (**73**) (black) as acceptor and methoxyacetone (**79**) (blue) as donor with the newly formed bond (pink) colored and the obtained yields in large scale (top). Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of *i*-valeraldehyde (**73**) and methoxyacetone (**79**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

Only 5 of the 26 tested DERA and FSA enzymes gave synthetically relevant product spots in the TLC screening of the cross-aldol reaction of *i*-valeraldehyde (**73**) and methoxyacetone (**79**). Interestingly, two different products were formed with high selectivity dependent on the

utilized aldolase family. While active FSA variants gave a deep blue spot, all positive DERA hits gave a brownish spot after staining with anisaldehyde. Important to mention, the color intensity from the staining with anisaldehyde is no clear indication for the total amount of a substance since concentration effects can play a dominant role, also concerning the resulting staining color. For both aldolase types, one enzyme was selected for preparative scale reaction. Enzyme FQ061 was clearly the FSA enzyme from the selection with the highest conversion for the blueish product spot, while D135 was the DERA variant with highest conversion for synthesizing the brownish product spot. Most other enzymes, except DERA126 and FSA063, became synthetically irrelevant from experience based on the intensity of the product spot.

The reactions of the most active hits DERA126, DERA135, DERA147, FSA054, FSA063 and FQ061 were analyzed via HPLC-MS. The results of the HPLC-MS analysis are presented in Figure 27.

Variant	Product formation [%]
D126	29
D135	62
D147	6
F054	9
F063	29
FQ061	58
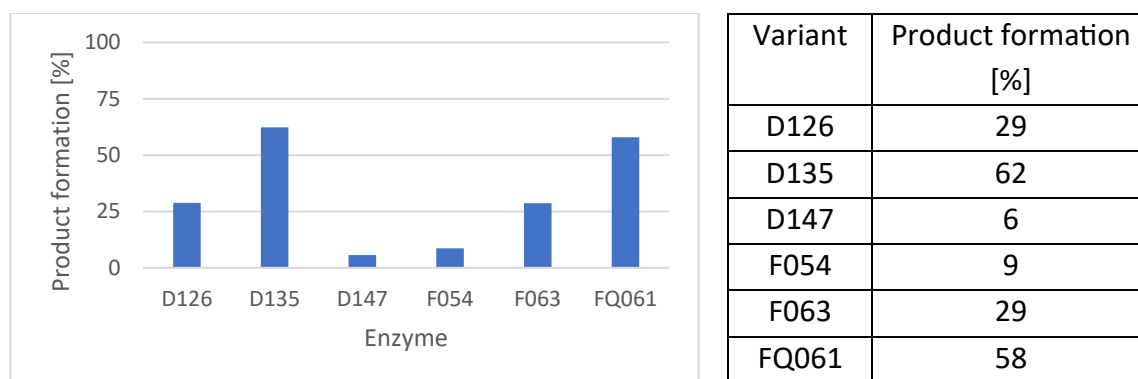
Figure 27 Results of the screening of most active variants in cross aldol reaction of *i*-valeraldehyde (**73**) (100 mM) and methoxyacetone (**79**) (500 mM) in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO evaluated with HPLC/MS.

The most active variants DERA126, DERA135, FSA063 and FQ061 generated product in the range of 29% up to 62%, while DERA147 and FSA054 gave values below 10%. All samples with values above 10% were deemed as synthetical useful in preparative scale reaction.

The result of the product identification indicates that DERA variant D135 produced solely the linear product, while FSA variant FQ061 exclusively produced the branched product. Figure 28 presents the ^{1}H and ^{13}C NMR spectra of the branched product produced by variant FQ061.
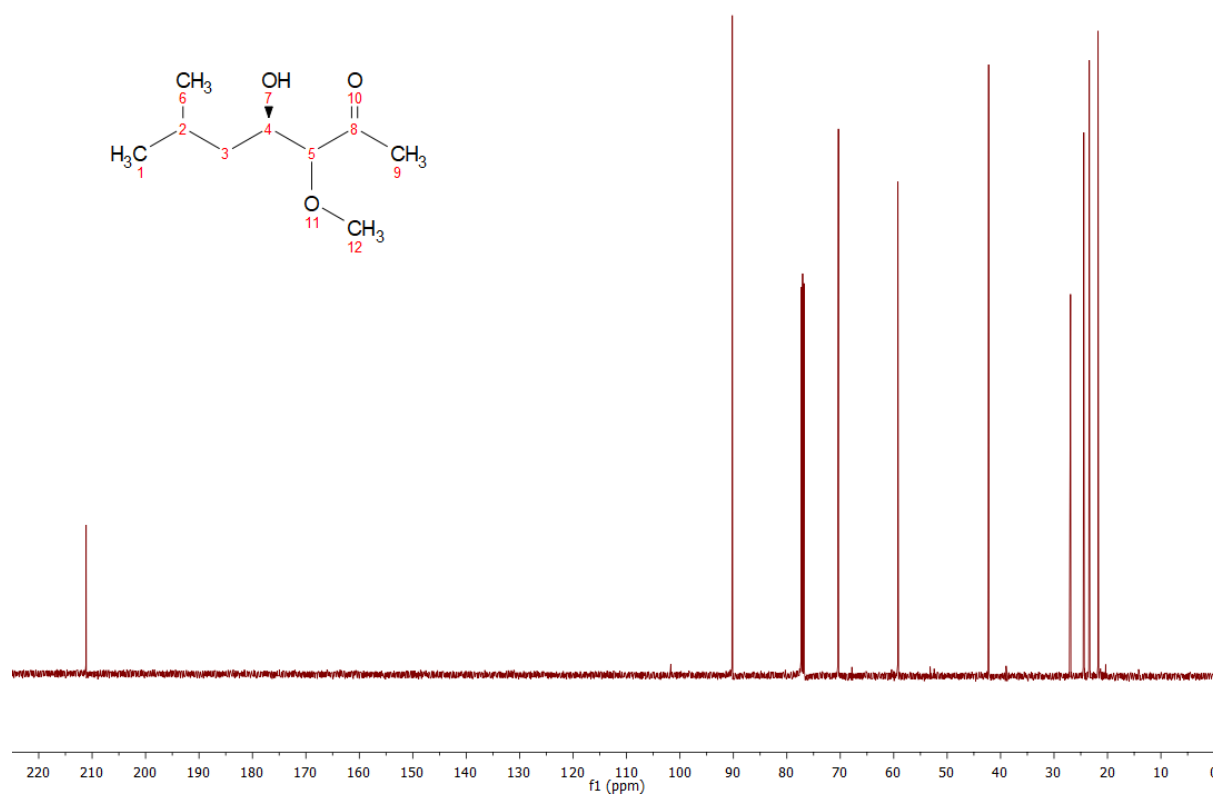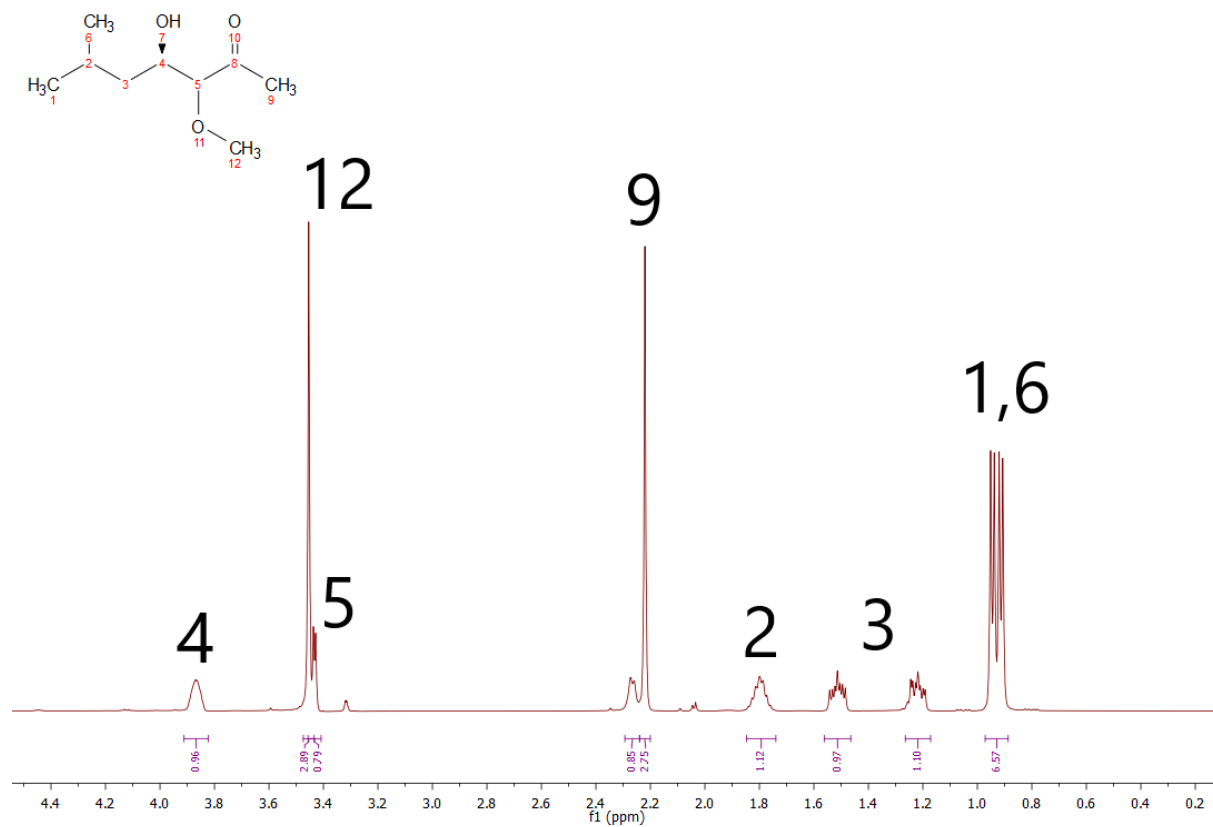
Figure 28 ^{1}H-NMR (top) and ^{13}C-NMR (bottom) spectra of the branched aldol product (**95**) (branched aldol product).

The following spectra in figure 29 were measured for the product sample produced by D135 indicating the linear product.
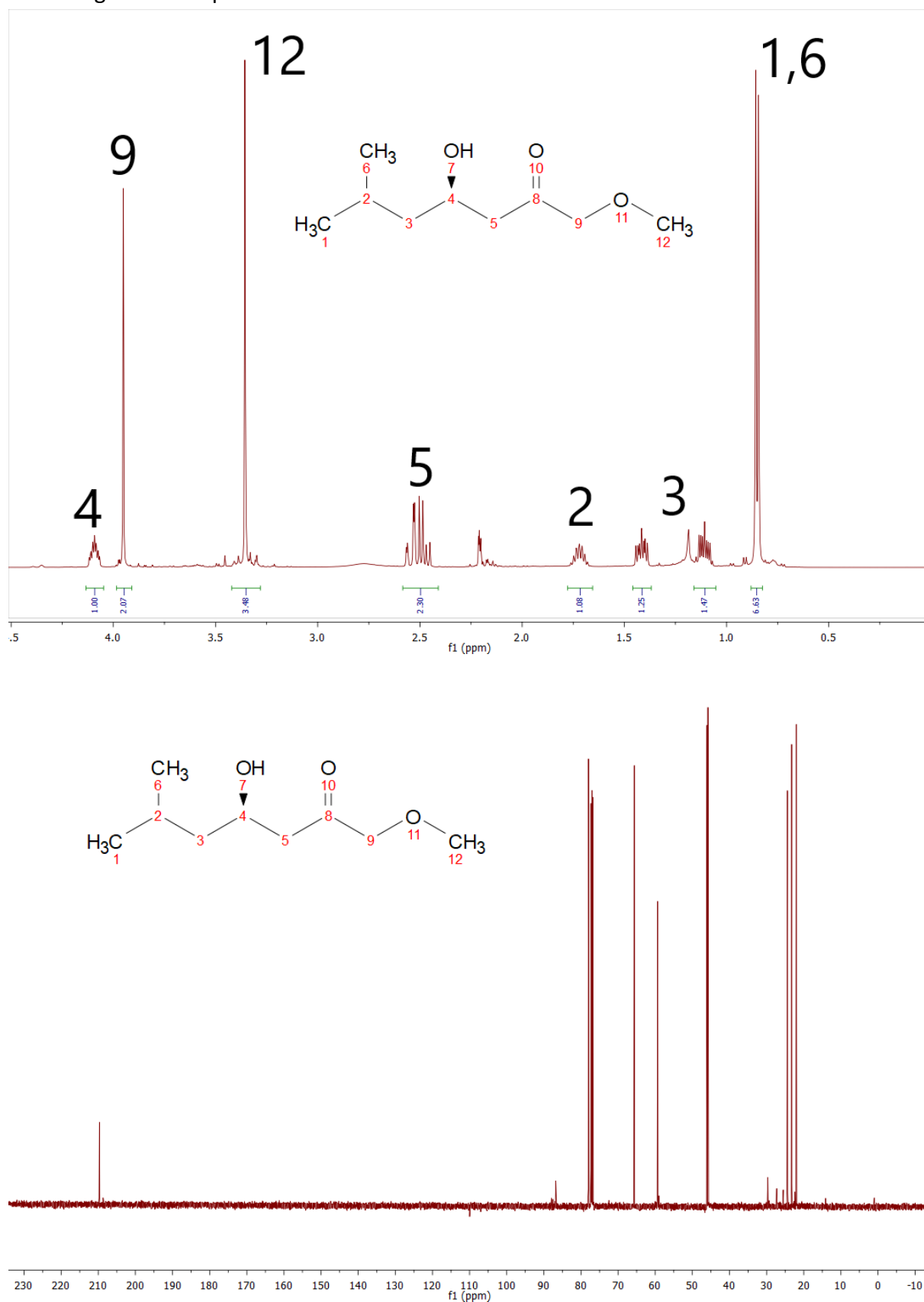
Figure 29 ^{1}H-NMR (top) and ^{13}C-NMR (bottom) spectra of linear aldol product (**94**).

The result that both products, the linear and the branched aldol product, become available through different complementary enzyme families is fascinating. Additionally, DERA135 and FQ061 are both very active in that reaction and furthermore performed very well, also with different reactions rendering them as universal aldol catalysts for most of the donor scope identified in this work. This is especially the case for DERA variant D135.

4.2.5 Screening of selected DERA and FSA enzymes for activity with 3-pentanone (65) and trifluoroacetone (78) – the borderline cases

Due to fluoroacetone (**77**) being a naturally accepted donor by *Ec*DERA, we were interested in higher fluorinated acetone (**50**) derivatives such as, 1,1-difluoroacetone and 1,1,1-trifluoroacetone as donors. While 1,1,1-trifluoracetone is commercially available, difluoroacetone was not commercially available and is additionally challenging to synthesize. Therefore, it was decided to only test for activity with trifluoroacetone. The results with trifluoroacetone indicated only very faint amounts of a newly formed product on TLC. With only traces produced by the enzyme, chances for optimization of enhanced activity by directed evolution or SDM seems possible. Sadly, no product could be isolated in preparative scale nor detected by HPLC-MS analysis and therefore could not be confirmed beyond TLC analysis.

Similar results were observed for the donors 3-pentanone (**65**)/2-pentanone (**81**), the next bulkier donors next to butanone (**55**). Again, no product formation could be supported by HPLC-MS or by test reactions via TLC analysis.

4.2.6 Screening of selected DERA and FSA enzymes for activity with cyclohexanone (61), 1,3-dimethoxyacetone (80), 2-pentanone (81), 1-hydroxybutanone, 3-oxetanone, dihydrofuran-3(2H)-one (63), tetrahydro-4H-pyran-4-one (64)

By using *i*-valeraldehyde (**73**) as acceptor the screening of the tested selection of possible donors namely cyclohexanone (**61**), dimethoxyacetone (**80**), 2-pentanone (**81**), 1-hydroxybutanone (**33**), 3-oxetanone (**62**), dihydrofuran-3(2H)-one (**63**), tetrahydro-4H-pyran-4-one (**64**), cyclohexanone (**61**), yielded all negative results.

The three potential donors 3-oxetanone, dihydrofuran-3(2H)-one (**63**) and tetrahydro-4H-pyran-4-one (**64**) were never tested in the literature before as possible donors by none of the two aldolase families. A significant problem observed with those donors was the high chemical background reactivity. For all three compounds no enzyme activity was detectable while a constant chemical background arose in each sample with different velocity. Hence, these ketones were deemed as not accepted.

While cyclobutanone (**57**) and cyclopentanone (**58**) both are reported donors for both aldolase families, literature lacks examples of cyclohexanone (**61**) products. Indeed, similar results were observed during the screening for activity with cyclohexanone (**61**), which was not accepted by any sample of the two metagenomic enzyme libraries.

With 1,3-dimethoxyacetone (**80**) another borderline case for aldolase activity of DERA and FSA was found. While methoxyacetone (**79**) was accepted by some enzymes of both aldolase families, 1,3-dimethoxyacetone (**80**) was not acceptable by any enzyme, presenting a clear limit of donor activity.

4.3 Screening of the selected enzymes for novel donor activity with 3-hydroxy-3-methylbutanal (98) as acceptor

After finalizing the results with *i*-valeraldehyde (**73**) as acceptor, it was decided to further evaluate the active enzymes with the selection of positive donors with an alternative acceptor substrate.

One way to increase conversion is the implementation of a sequential reaction, in which the aldol product takes part reversibly, thereby pushing the equilibrium state towards the aldol product.
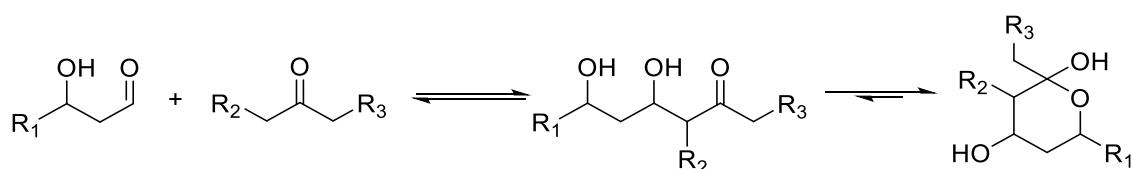
Scheme 3 Schematic presentation of intramolecular cyclization reaction of aldol products influencing the equilibrium of the aldol reaction.

For that purpose, a β-hydroxyl group is instrumental to cyclize the product to the ring-closed hemiacetal form, as in case of the natural acceptor GA3P (**3**). For that purpose, 3-hydroxybutanal (**98**) was prepared following the reaction pathway presented in Scheme 4. Initially it was planned to use 3-hydroxybutanal in both its enantiomerically pure forms, accessible by kinetic resolution catalyzed by CalB lipase from *candida antarctica*[158-159]. Scheme 4 shows the reaction pathway towards (*R*)/(*S*)-3-hydroxybutanal and the donors to be tested.
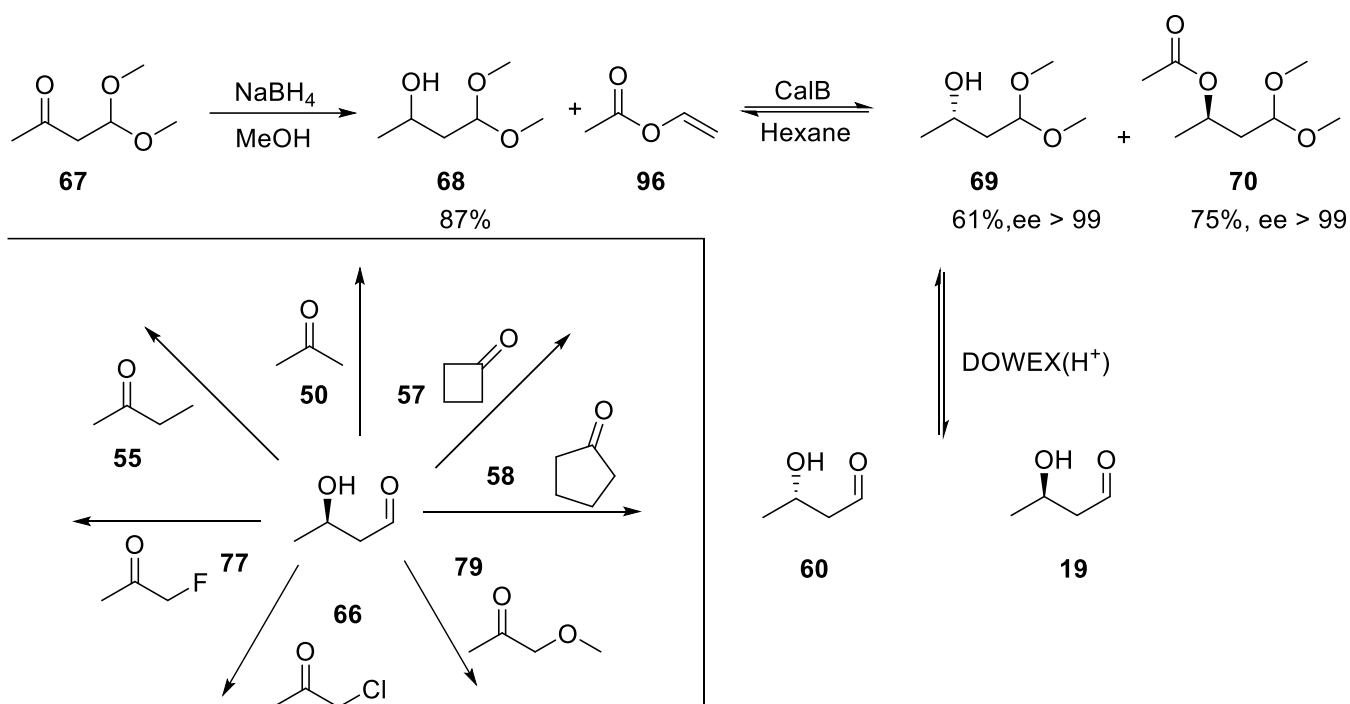
Scheme 4 Workflow for the screening of metagenomic FSA and DERA libraries in the cross-aldol reaction of (R)/(S)-3-hydroxybutanal with the accepted donors from section 3.2 including preparation of (R)/(S)-3-hydroxybutanal.

For the synthesis of 3-hydroxybutanal the proposed synthetic pathway of DURRWACHTER and WONG was utilized[127]. 3-Hydroxybutanal is accessible through reduction of 4,4-dimethoxybutan-2-one (**67**) with sodium borohydride, which will yield the dimethoxy-protected 2-hydroxybutanal as a racemate. By kinetic resolution performed with CalB and vinyl acetate, both enantiomers can be separated due to significantly different reaction rates for both enantiomers. CalB will react almost exclusively with the (R)-enantiomer as long as it is available, enabling to stop the reaction before consumptions of the (S)-enantiomer starts. After isolation of the (R)-ester, the reaction is continued until the (R)-enantiomer is completely consumed. By separation of the ester from the remaining (S)-enantiomer, both enantiomers are obtained with high enantiomeric excess.

The reaction was followed by GC with a chiral column, which yield two peaks corresponding to the two possible enantiomers. After 6h reaction time, the conversion of the (R)-enantiomer was almost complete and the ester was isolated. The reaction was then repeated to completely convert the remaining (R)-enantiomer and was stopped directly after signs appeared that the (S)-enantiomer started to be consumed, when the (S)-enantiomer was isolated. Both products were obtained with excellent enantiomeric excess of >99 based on the peaks obtained in the GC analysis.

The deprotection of the aldehyde function and the removal of the acetyl ester of the (R)-enantiomer can be achieved by utilizing H+-functionalized resin or inorganic acids. Because the product possesses a β-hydroxyl group which can eliminate under acidic conditions, the milder method utilizing H+-functionalized resin was favored to minimize the risk of elimination.

Because 3-hydroxybutanal is the reaction product of the homo aldol addition of two ethanal molecules, it can be cleaved by DERA catalysis to produce two ethanal molecules in a retro aldol fashion. When ethanal is becoming available, DERA will start to utilize ethanal as its natural donor for addition of the dimer aldehyde resulting in the formation of the ethanal trimer. The latter can be present in an open and a closed hemiacetal form, which will shift the reaction equilibrium towards the undesired but more stable intramolecular cyclization product. This unwanted side reaction predominated in the DERA catalyzed reactions only. All FSA enzymes were unable to accept 3-hydroxybutanal and presented all negative results in all tested reactions with the same donor selection from section **3.2**. Apparently, the cleavage will always be more active compared to the reaction with unnatural donors, even though a cyclization can occur to influence the equilibrium of the reaction.

Therefore, we were interested if the problem of DERA cleaving the acceptor will be less critical for 3-hydroxy-3-methylbutanal (**98**). In principle its hydroxyl group would act to cyclize the product while also avoiding chirality at position C3, which would reduce the number of possible diastereomers. This would also raise the chance for successful purification of the product since diastereomers can be hard to separate. The desired product in protected form could be made by using a methylmagenesium Grignard reaction with 4,4-dimethoxybutan-2-one (**67**) (Scheme 5).
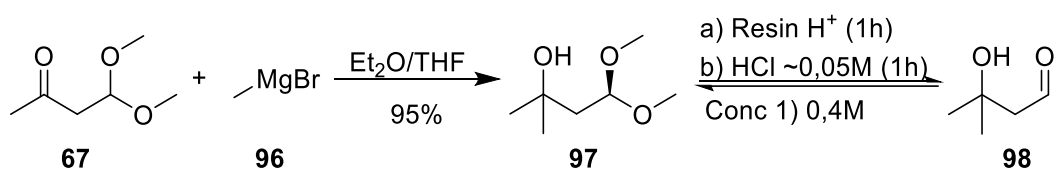
Scheme 5 Preparation of 3-hydroxy-3-methylbutanal (**98**).

4,4-Dimethoxy-2-methylbutan-2-ol (**97**) was obtained with a yield of 95% after filtration through silica and removal of solvent. The deprotection of the aldehyde function was tested with different resins containing H$^+$ functionalization. The resin yielding the best results possesses sulfonate groups which can be activated by simple washing with HCl and additional washing with distilled water. For drying the resin, an additional washing step with diethyl ether was performed yielding a dry resin with constant weight for more convenient and constant usage.

By adding the same mass of resin compared to the educt, good results were obtained after one hour reaction time at room temperature. After separation of the resin, the solution showed a close to neutral pH based on pH paper analysis. It was decided to use it as such directly after deprotection for the enzyme reactions.

When using HCl instead of resin or elevated temperatures, the number of side products increased, clearly detectable as the solution became cloudy. During deprotection polymerization can occur because elimination of the hydroxyl group is an issue. Therefore, the mild deprotection conditions achieved with H$^+$-resin at room temperature was the best available option and was therefore implemented. During the experiments with 3-hydroxy-3-

methylbutanal (**98**) no analysis was conducted by HPLC-MS. The main aim during this series of test reactions was to identify suitable enzyme candidates for larger scale reactions, which is more easily achievable by TLC analysis.

4.3.1 Screening of selected DERA and FSA enzymes for activity with cyclopentanone (58) as donor

Cyclopentanone (**58**) and cyclobutanone (**57**) were the donors for which the metagenomic enzyme libraries of Prozomix® had already been positively screened for. The expectation was therefore that most selected DERA enzymes should also show conversion with 3-hydroxy-3-methylbutanal (**98**). For the FSA enzymes it was less clear if the new acceptor 3-hydroxyl-3-methylbutanal (**98**) will be accepted since it is even more bulky than 3-hydroxybutanal which was not accepted with unnatural donors by all selected FSA enzymes. The results demonstrated that indeed none of the FSA enzymes was able to convert 3-hydroxy-3-methylbutanal (**98**) with any tested conditions or donor and therefore will not be presented here.

Figure 30 presents the results of the TLC screening for activity with cyclopentanone (**58**) as donor and 3-hydroxy-3-methylbutanal (**98**) as acceptor.

DERA135:   4%

Figure 30 Reaction scheme of the cross-aldol addition of 3-hydroxy-3-methylbutanal (**98**) (black) as acceptor and cyclopentanone (**58**) (blue) as donor with the newly formed bond (pink) colored and the obtained yields in larger scale (top). Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of 3-hydroxy-3-methylbutanal (**98**) and cyclopentanone (**58**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

From the results of the TLC-screening, 3-hydroxy-3-methybutanal is clearly less well accepted as an electrophile compared to *i*-valeraldehyde (**73**), judged by the low number of positive hits

showing a deep product spot intensity. Still two enzymes, DERA135 and DERA143, presented clear product spots that deemed the enzymes useful for larger scale reaction.

The only two hits useful for synthetic purposes seemed to be DERA135 and DERA143 of which DERA135 was selected for a preparative scale reaction. The synthesis yielded 4% aldol product after purification with a total mass of 85 mg. The ^{1}H and ^{13}C NMR spectra indicated a mixture of the desired product and the corresponding elimination product (see attachment). Due to the low yield and limited amounts of both the most promising enzymes as CFE further studies had to be discontinued.

4.3.2 Screening of selected DERA and FSA enzymes for activity with cyclobutanone (57) as donor

Regarding the results with cyclopentenone (in Chapter **3.3.1**), the expectations for active hits were limited to find only a few good hits. The results of the screening with cyclobutanone (**57**) as donor and 3-hydroxy-3-methylbutanal (**98**) are illustrated in Figure 31.

DERA180: 16% Isolated yield
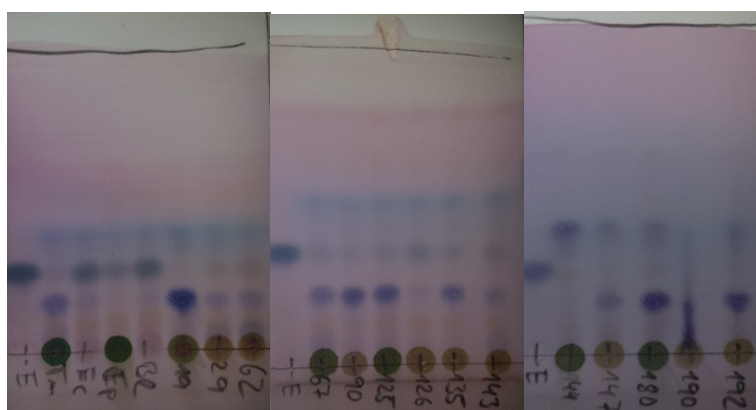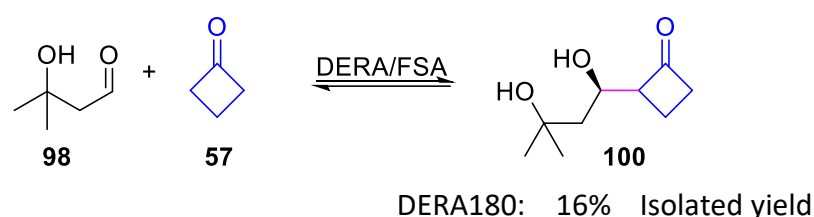
Figure 31 Reaction scheme of the cross aldol addition of 3-hydroxy-3-methylbutanal (**98**) (black) as acceptor and cyclobutanone (**57**) (blue) as donor with the newly formed bond (pink) colored and the obtained yields in larger scale (top). Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of 3-hydroxy-3-methylbutanal (**98**) and cyclopentanone (**58**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

Interestingly, cyclobutanone (**57**) was found to be a well-accepted donor by most DERA variants with 3-hydroxy-3-methylbutanal (**98**) as the electrophile. Based on the baseline spot, it seemed that a side reaction took place that is even catalyzed by the protein samples from the empty plasmid. The corresponding TLC lane showed only traces of residual 3-hydroxy-3-

methylbutanal (**98**), which indicates substrate loss to an unspecified consumption process. In general, the aldol product stained less well and was hard to detect, but since most samples show full conversion of the acceptor with varying degree of product spots, there is clearly a side reaction taking place consuming the acceptor molecules. DERA180 was available in greater quantities and was therefore selected for a preparative scale reaction. The synthetsis yielded 16% of the purified product. The ^{1}H and ^{13}C NMR spectra of the aldol product of 3-hydroxy-3-methybutanal and cyclobutanone (**57**) are presented in Figure 32+33.
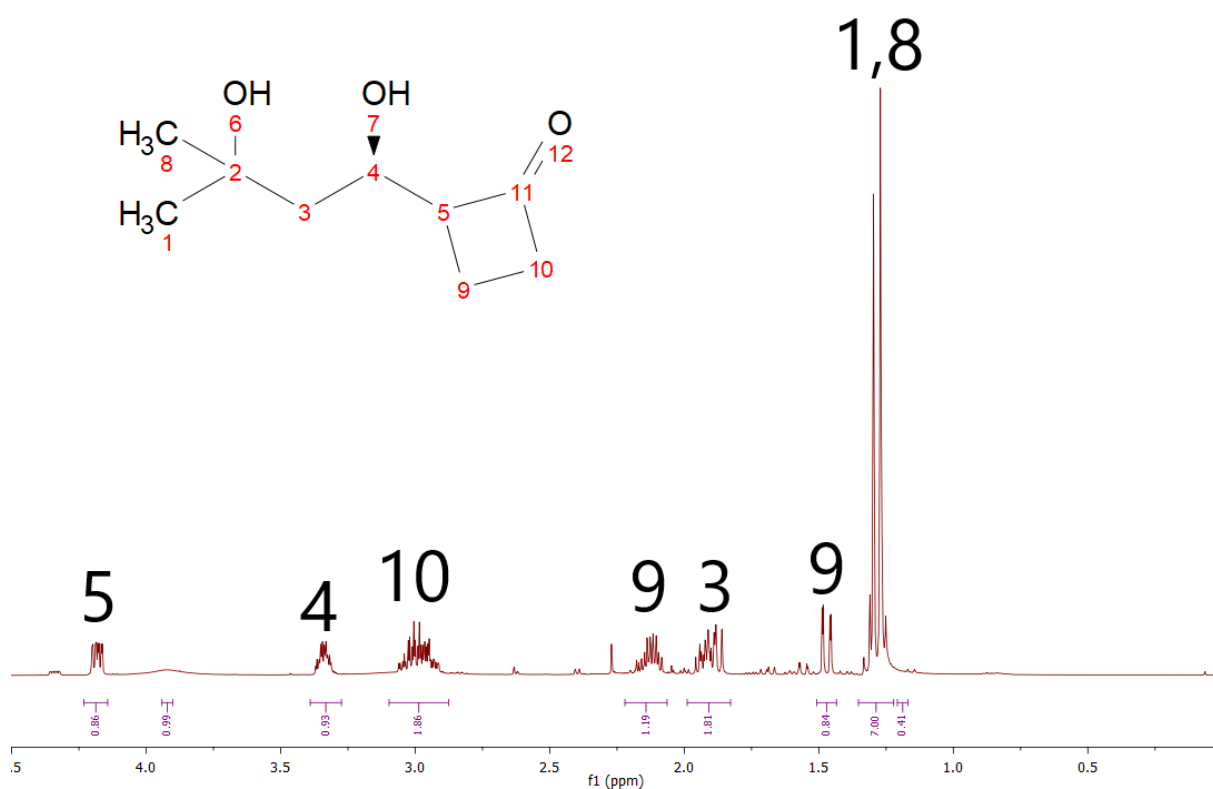
Figure 32 ^{1}H-NMR of 2-((*R*)-1,3-dihydroxy-3-methylbutyl)cyclobutan-1-one (**100**).
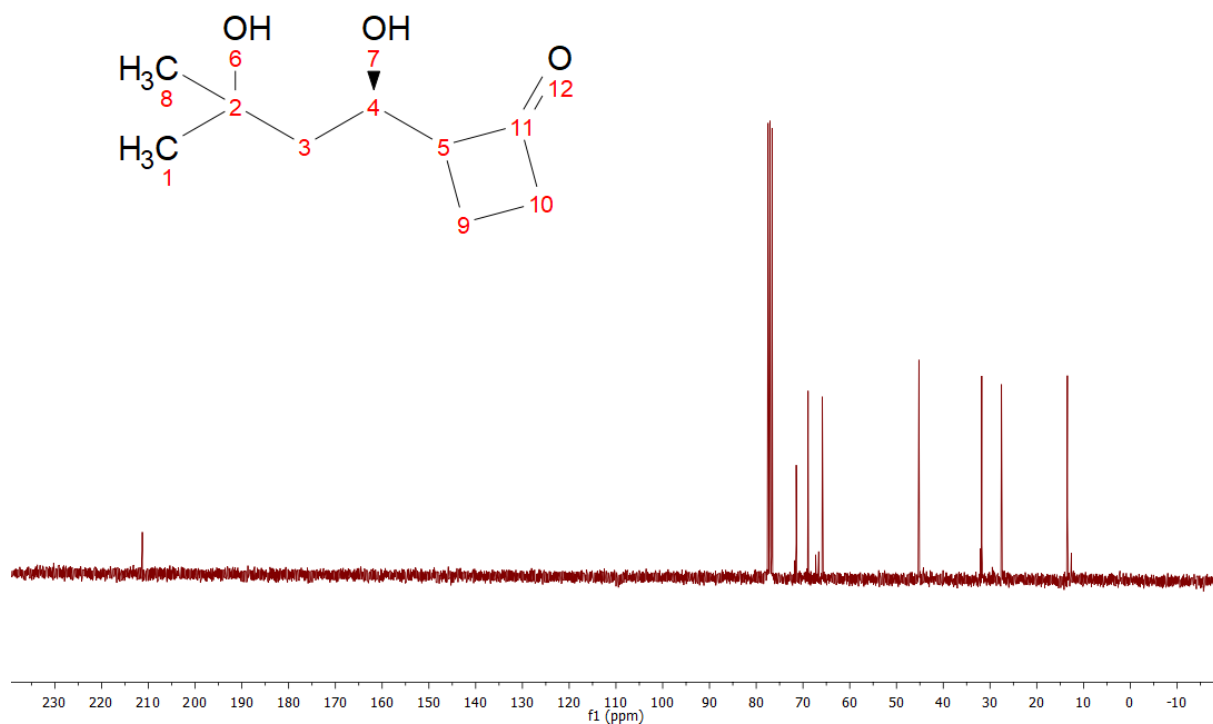
Figure 33 ^{13}C-NMR of 2-((R)-1,3-dihydroxy-3-methylbutyl)cyclobutan-1-one.

All signals could be interpreted for the desired product. Minor traces were found, which can possibly be interpreted for the closed isomer.

4.3.3 Screening of selected DERA and FSA enzymes for activity with butanone (55) as donor

Results of the TLC screening of the selected DERA variants with butanone (**55**) as donor and 3-hydroxy-3-methylbutanal (**98**) as acceptor are presented in Figure 34.
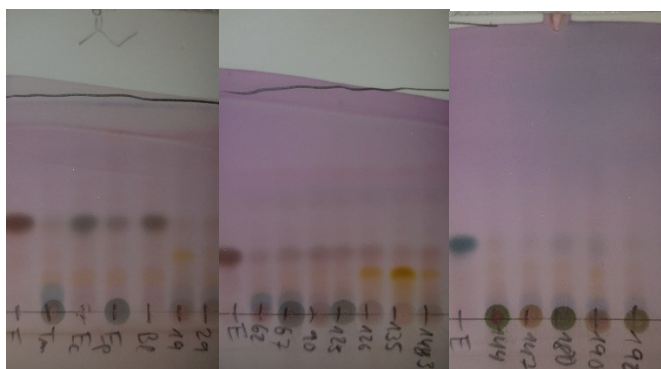
Figure 34 Reaction scheme of the cross-aldol addition of 3-hydroxy-3-methylbutanal (**98**) (black) as acceptor and butanone (**55**) (blue) as donor with the newly formed bond (pink) colored and the obtained yields in larger scale (top). Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of 3-hydroxy-3-methylbutanal (**98**) and butanone (**55**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

The results indicated product formation for only three of the samples, namely DERA126, DERA135 and DERA143. Only the staining of the product spot for DERA135 indicated synthetically sufficient product formation, whereas the DERA samples DERA126 and DERA143 showed only faintly stained product spots. Therefore, the apparently universal DERA enzyme DERA135 was again clearly identified as the most active variant and used for a preparative scale reaction. Figure 35 presents the ^{1}H and ^{13}C NMR spectra obtained for the isolated product produced with DERA enzyme DERA135.
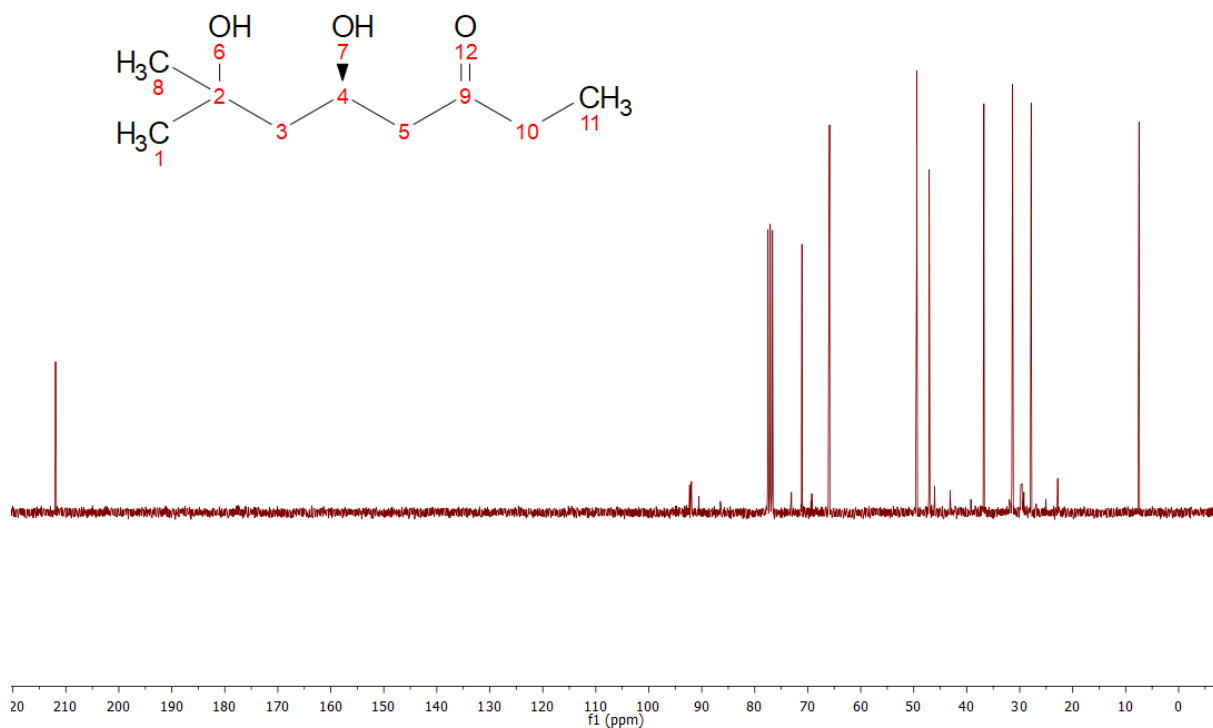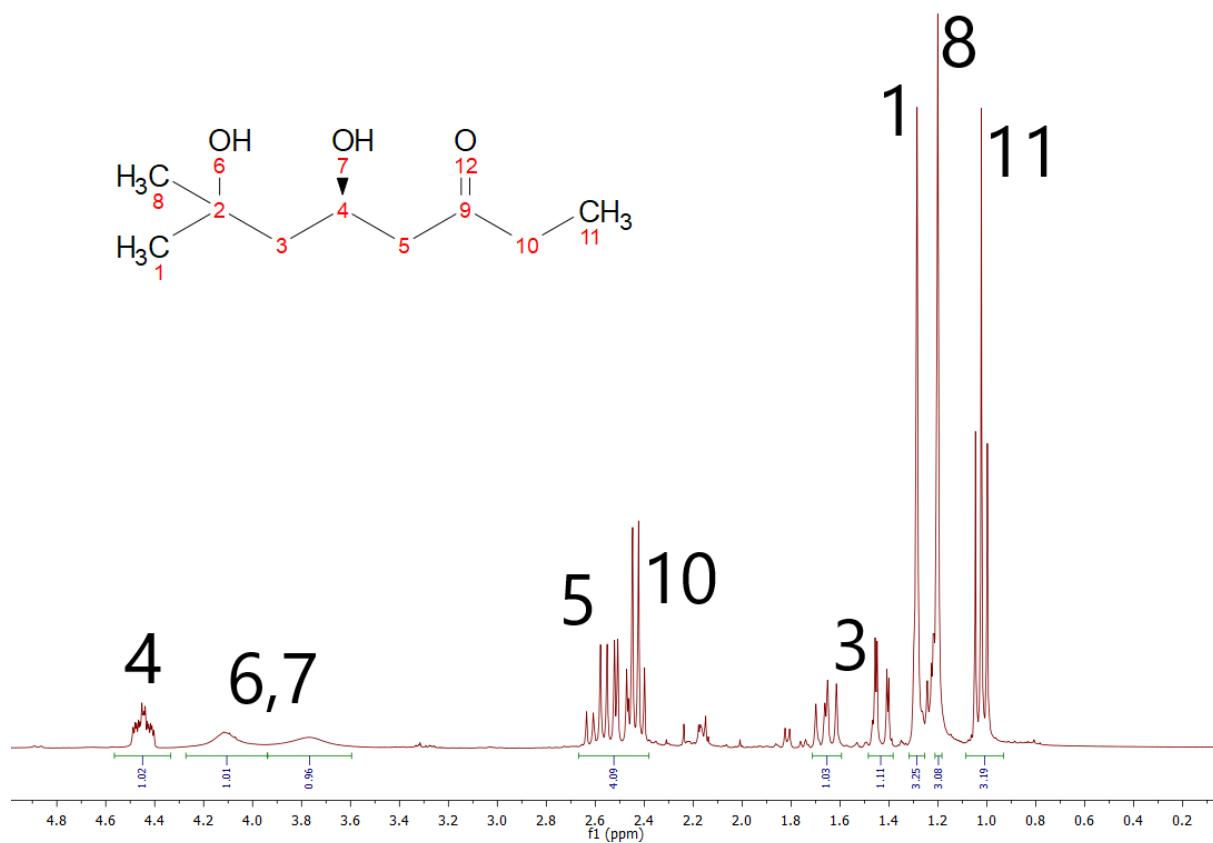
Figure 35 1H-NMR (top) and 13C-NMR (bottom) of (*R*)-5,7-dihydroxy-7-methyloctan-3-one (**101**).

The product was obtained in good purity with 5% yield after purification via column chromatography on silica gel. The product could be identified as the linear product.

In the closed form, the hydroxyl group and methyl group block each other making the cyclized form unfavorable compared to the linear form. The cyclized form is presented in Figure 36.
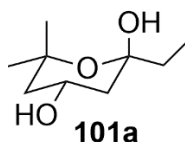
Figure 36 Cyclized aldol product (101a)

4.3.4 Screening of selected DERA and FSA enzymes for activity with chloroacetone (66) as donor

For the reaction with chloroacetone (**66**) as donor, the results indicated the formation of a novel product spot. The results of the TLC screening are presented in Figure 37.
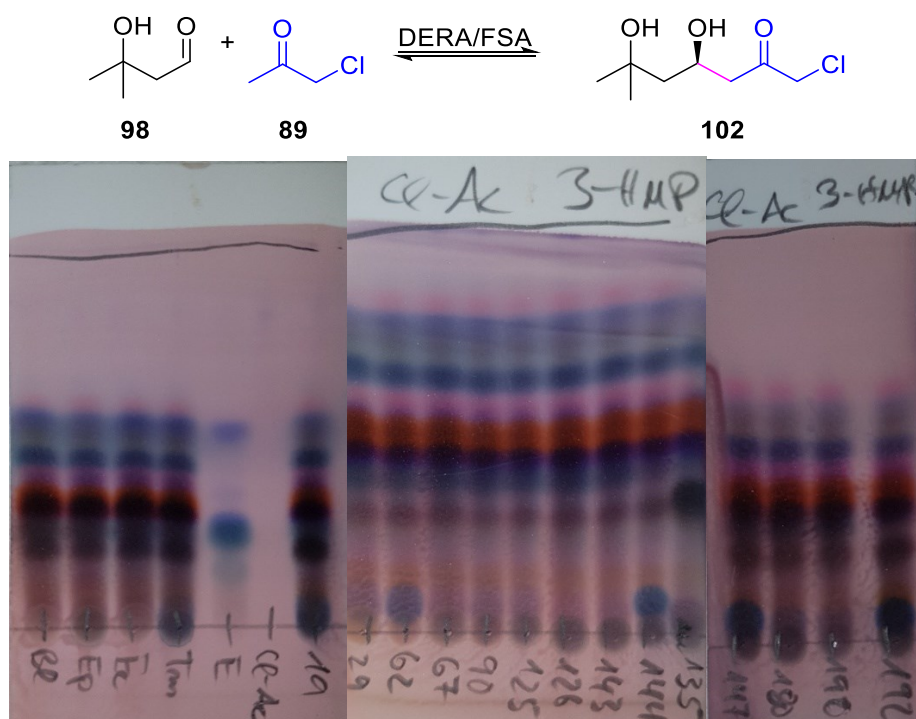
Figure 37 Screening results of the metagenomic DERA/FSA libraries in the cross aldol addition of 3-hydroxy-3-methylbutanal (**98**) and chloroacetone (**66**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

The samples DERA62, DERA144, DERA147 and DERA192 were known from earlier projects to be highly active for the natural donor acetaldehyde (**17**). The newly formed hits close to baseline corresponded to the acetaldehyde (**17**) trimer and were therefore not further

followed up in preparative scale. The acetaldehyde (**17**) trimer can be formed after cleavage of 3-hydroxy-3-methylbutanal (**98**) forming 3-hydroxybutanal, which can than react again with another acetaldehyde (**17**) from the cleavage reaction.

4.3.5 Screening of selected DERA and FSA enzymes for activity with methoxyacetone (79) as donor

Methoxyacetone (**79**) was a challenging donor for DERA enzymes since only three DERA samples showed product formation in the cross-aldol reaction of *i*-valeraldehyde (**73**) as acceptor and methoxyacetone (**79**) as donor. DERA135 was the only DERA enzyme which had presented synthetically relevant results. Therefore, the expectations were limited for the cross-aldol reaction of 3-hydroxy-3-methylbutanal (**98**) as acceptor with methoxyacetone (**79**) as donor. The results of the TLC screening are presented in Figure 38.
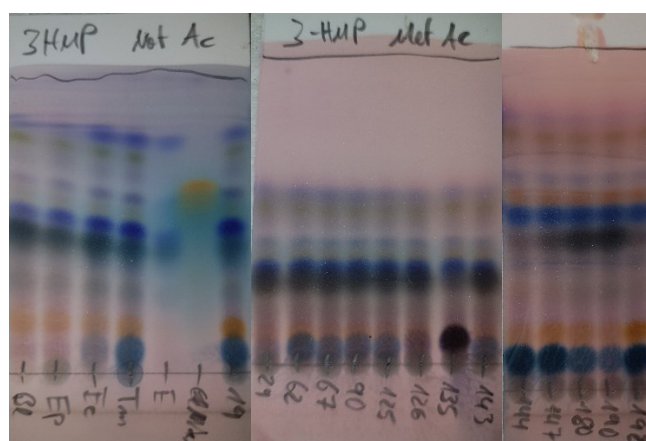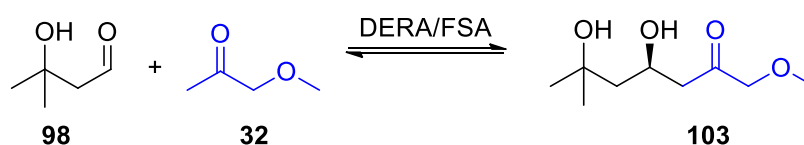
Figure 38 Screening results of the metagenomic DERA/FSA libraries in the cross-aldol addition of 3-hydroxy-3-methylbutanal (**98**) and methoxyacetone (**79**) based on TLC analysis (2:1 cyclohexan/ethylacetate) with anisaldehyde staining (bottom).

Interestingly, almost all samples showed a new product spot above the baseline. For DERA135 the intensity of the spot color was the most intense of all samples. During HPLC-MS analysis, the mass of the expected aldol product could not be identified but the mass of the acetaldehyde (**17**) trimer was found instead. Hence, the DERA variant proved to have the highest activity in the cleavage of 3-hydroxy-3-methylbutanal, followed by a trimerization of acetaldehyde (**17**). Therefore, the cross-aldol reaction of 3-hydroxy-3-methylbutanal (**98**) with methoxyacetone (**79**) was deemed unsuccessful, and further studies were discontinued.

5 Discussion

In the literature there are many published reviews on studies of novel donors and acceptors for aldolases. The latest review focusing on aldolases is from 2022[39], which summarizes the known donors and acceptors that are converted by different aldolase families. The Figure 39 summarizes the known donors for FSA- and DERA variants.
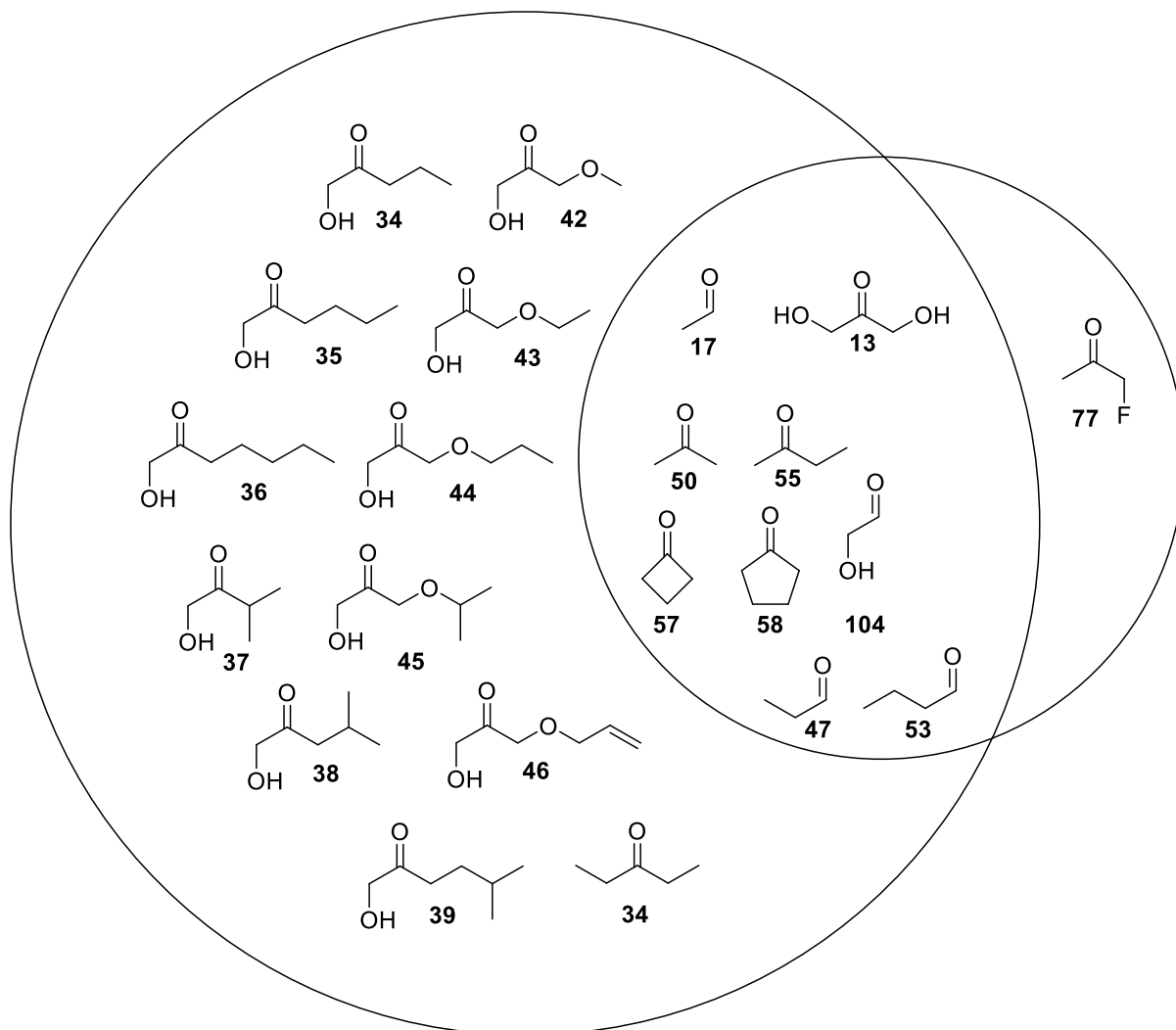
Figure 39 Shared donor scope for DHA and acetaldehyde dependent aldolases from the latest published aldolase review (graphic taken from ref [39]).

The main goal of the present work was the search for novel unknown donor activity for DERA and FSA enzymes. Figure 39 demonstrates that the two enzyme classes, DHA and acetaldehyde dependent aldolases, can both utilize a variety of different donors and for some donors both enzymes present a shared donor scope. It is important to understand that most of the results in the literature were not obtained by using wildtype enzymes but with engineered variants produced by directed evolution or SDM. Another factor is the utilized acceptor substrate. With the natural acceptor GA3P (**3**), butanone (**55**) was cleaved from the corresponding aldol product by DERA, while there is no example with an alternative

electrophile[39], or in direction of aldol synthesis. Butanone (**55**) was accepted by FSA variant D6H to produce the corresponding aldol product with the natural acceptor GA3P (**3**).

Most experiments make use of acceptor molecules that enable the aldol product to cyclize, pushing the equilibria of the aldol reaction in favor of the product. In the present work, the metagenomic and mutagenic libraries were screened for activity with aliphatic acceptor substrates. With the natural acceptor GA3P (**3**) possessing a hydroxyl group at α-position and a phosphorylated OH group in β-position, the switch to aliphatic, non-hydroxylated acceptors is a challenging task for DERA and FSA enzymes.

Nonetheless, both metagenomic DERA+FSA libraries unraveled enzymes able to convert aliphatic acceptors like *n*-butanal (**53**), *i*-valeraldehyde (**73**), 3-hydroxyl-3-methybutanal and *i*-butyraldehyde (**71**) for DERA only. That alone was already synthetically interesting, but being able to convert novel donors with aliphatic acceptors seemed unlikely after evaluation of the mutagenic DERA library in Chapter 2. The results proved different and the results of the screening for unknown activity with metagenomic enzyme sources were fruitful.

During the study multiple new donors were identified for both aldolase families. Methoxyacetone (**79**) was accepted by enzymes from both libraries enabling the synthesis of the branched product and the linear product dependent on the enzyme family with excellent yields in preparative scale reaction. While this activity was only observed with *i*-valeraldehyde (**73**) as acceptor, DERA126 and DERA135 both showed excellent conversion with methoxyacetone (**79**) as donor. The enzymes DERA135 and FQ061 were utilized in preparative scale for the proof of the corresponding product since two different spots were formed by both enzyme classes. The DERA enzymes produced solely the linear product while FSA enzymes produced the branched product with high selectivity. The experiments in preparative scale yielded 55% of the linear product for DERA135 and 74% of the branched product for FSA enzyme FQ061. Interestingly, both enzymes showed high activity during the experiment series mostly yielding products with all tested donors that showed activity with at least one variant. Nonetheless, all tested reactions with 3-hydroxy-3-methylbutanal (**98**) were negative for the FSA.

The acceptance of butanone (**55**) as donor for almost all selected DERA and FSA variants with *i*-valeraldehyde (**73**) as acceptor is another great achievement of this study. Some of the variants, namely DERA019/126/135/180/190 produced the linear product with high selectivity, while only FQ059 showed good selectivity for the branched product even though only in trace amount. Most other variants produced a mix of both products, while DERA062/067/125/192 only produced trace amounts of the linear product. Those results were unexpected due to only examples being available in literature with butanone (**55**) as donor utilizing the natural electrophile GA3P (**3**). This study presents the first results of DERA and FSA accepting butanone (**55**) as donor with aliphatic acceptors in combination.

Another fascinating result is the use of chloroacetone (**66**) as donor. Up to date, there were no examples of chloroacetone (**66**) being an acceptable donor for any DERA or FSA enzymes. One limitation with chloroacetone (**66**) as donor is the sole production of the linear product, which is of course in line with expectations since the chloro substituent is quite bulky. Different from chloroacetone (**66**), the experiments with fluoroacetone (**77**) revealed some enzymes

with excellent selectivity for either the linear or the branched product. Because chloro atoms are larger than fluoro atoms the reason most likely originate from steric hindrance in the active center. With chloro and fluoroacetone (**77**) being acceptable donors, potential application of even bulkier derivatives as donors seems possible and may lead to new pathways for the sustainable production of potential drug candidates.

Regarding the current limitations of novel donors, clearly with 3-pentanone (**65**), the next higher homolog of butanone (**55**), product formation occurred only in traces. Interestingly, 3-pentanone (**65**) is already known to be reactive with a modified *Ec*FSA variant, which was mutated at the D6 position [111, 120]. The aim of the mutagenesis was to target amino acid residues in the active center of *Ec*FSA for substitution having amino acids with smaller side chains. The study involved a library with D6X mutations, from which a novel D6T *Ec*FSA variant was enabled to convert 3-pentanone (**65**). Accordingly, there is a good chance that most of the metagenomic FSA enzymes, will accept 3-pentanone (**65**) after rational redesign of the inactive center.

In total three novel donor activities were detected for DERA, namely chloroacetone (**66**), methoxyacetone (**79**) and butanone (**55**). For FSA enzymes the novel donors fluoroacetone (**77**), chloroacetone (**66**) and methoxyacetone (**79**) were identified. By comparison of the FSA and DERA metagenomic panels of prozomix, interesting differences were identified concerning the stereoselectivity of the aldol reaction, when linear and branched products are possible. That was especially the case for fluoroacetone (**77**), butanone (**55**) and methoxyacetone (**79**). Many enzymes showed excellent activity for one of both products.

Many studies have successfully tried to use acceptors, which allow spontaneous cyclization of the aldol product to achieve an equilibrium shift towards the product side. Therefore, this approach was also followed with 3-hydroxybutanal and 3-hydroxy-3-methylbutanal (**98**) as acceptors.

Because 3-hydroxybutanal is chiral both enantiomers were synthesized, using a kinetic resolution via CalB to separate the enantiomers. The deprotection of the aldehyde function worked best with H$^+$-resin at 0.4 M concentrations, yielding the concentration desired for the stock of 3-hydroxybutanal. After removal of the resin by filtration after 1 h, no neutralization of the reaction solution was necessary. No further purification was conducted to prevent elimination and polymerization.

The main problem with 3-hydroxybutanal as acceptor was the cleavage of the acceptor substrate by DERA in a retro aldol fashion yielding two equivalents of acetaldehyde (**17**), while FSA variants did not convert 3-hydroxybutanal under the chosen conditions, yielding only negative results. The cleavage product acetaldehyde (**17**) can react with 3-hydroxybutanal forming a formal aldol trimer of acetaldehyde (**17**). The trimer can cyclize to a stable ring hemiacetal isomer, thus pushing the equilibrium of the aldol reaction to the product side for practical full conversion. Since all reaction steps are in equilibrium, the reaction with the thermodynamically most stable product becomes preferred, and with cyclization of the product there was no chance for the desired complementary aldol products with non-natural donors to be formed.

To minimize that problem, it was decided to use 3-hydroxy-3-methylbutanal (**98**) as a bulkier analog to 3-hydroxybutanal. The synthesis was easily achieved by a Grignard reaction of methylmagenesium bromide and 4,4-dimethoxybutan-2-one in THF. Compared to 3-hydroxybutanal, no kinetic resolution was necessary since the β-carbon is not chiral and therefore does not lead to diastereomer formation upon aldol addition, which can be challenging in some cases. The product of the Grignard reaction, 4,4-dimethoxybutan-2-ol, could be deprotected analogously to the above described deprotection of 3-hydroxybutanal.

In principle, the β-hydroxyl group of 3-hydroxybutanal can support a cyclization of the aldol product, but was actually not confidentially observed during the experiments. Additionally, no cleavage of the educt was observed, enabling the variants to perform the desired aldol addition instead of producing the acetaldehyde (**17**) trimer. Interestingly, no FSA variant was able to utilize 3-hydroxy-3-methylbutanal (**98**) as acceptor, in line with the negative results for 3-hydroxybutanal.

In general, the DERA enzymes showed fewer positive hits with 3-hydroxy-3-methylbutanal (**98**) than with *i*-valeraldehyde (**73**) as acceptor. Both metagenomic aldolase panels were screened for activity with *i*-valeraldehyde (**73**) as acceptor and cyclobutanone (**57**)/cyclopentanone (**58**) as donor. Retrospectively, not only one acceptor should have been used for the screening. Instead, a variation of acceptor molecules with different properties would have been useful but increasing the screening effort. Since TLC screening with HPLC-MS confirmation was utilized for detection of positive hits, the screening is a very tedious and costly process given the 400+ possible variants to be screened. All attempts to create a fluorometric assay yielded negative results due to product and educt similarity, making assays only possible in special cases with coupled enzyme reactions (not shown in present work)[107, 160].

During the screening with 3-hydroxy-3-methylbutanal (**98**) as acceptor, a total of three donors showed activity towards the desired aldol product, namely cyclobutanone (**57**), cylopentanone (**58**), butanone (**55**).

Interestingly, the product of the aldol reaction with cyclopentanone (**58**) and 3-hydroxy-3-methylbutanal (**98**) yielded a mixture of diastereomers in a 1:1 ratio. There can be different reasons for a diastereomeric mixture to be formed. Either the enzyme has no face selectivity for the nucleophilic α-CH position, which is not absent in the natural reaction, or the primarily formed aldol enantiomer becomes equilibrated at the α-carbon due to keto-enol tautomery, either during purification or less likely already during reaction at higher pH. The effect should not be dramatic during the aldol reaction since the reaction solution is buffered at pH 7.5 not strongly supporting keto-enol tautomery.

The main goal of the present work was to identify novel donor activity. The results demonstrate that Nature's diversity is a seemingly endless source for diverse biocatalysts. The mutagenic *Ec*DERA library from Chapter 2 was based on a rational design approach, which seemed plausible but yielded insufficient results, whereas the metagenomic diversity from a handful of offshore dirt yielded different active variants. The next step could be the crystallization of the most active variants to obtain a crystal structure for either improvements by rational design or even by directed evolution for increasing the diversity of the donor and

acceptor scope. The rational design is usually aimed at the active center to create space, change polarity or adjust substrate channels. This method often misses second layer amino acid residues which can have a beneficial effect on activity and stability[107].

1 Chapter 3 Hofmeister series

2 Introduction

Life on earth arose under extreme conditions. A variety of extremophilic environmental factors stimulated rather than disturbed a steady evolution. Conditions included, for example, extreme temperatures, high pressure or high salt concentrations. The development of life has therefore adapted to these conditions from the beginning.

The interplay of salts or ions on biological macromolecules has troubled scientific investigations for over 130 years. In 1888 Franz Hofmeister, then working in the German part of the Faculty of Medicine at Charles University in Prague, was researching salt effects on proteins in aqueous solution. In his manuscript "Zur Lehre von der Wirkung der Salze"[161-162], which appeared as a part of a series of publications, he described for the first time the effects of specific ion effects on biological and organic macromolecules. His series of measurements of different ions on the salting-out behavior of aqueous protein solutions resulted in the so-called "Hofmeister series", which is still a relevant and topic of research until today. The Hofmeister series is also known as "lyotropic series".

In Hofmeister's experiments, egg foam from beaten chicken eggs was utilized as study sample, which was left to stand overnight beforehand. The separated liquid from the foam was collected and used as such as a protein solution sample. By mixing the sample with concentrated sodium and chloride salts and measuring the time at which the solution turned cloudy, he was able to classify the various anions and cations according to their "salting out" behavior. For him at that time, many new theories probably played a role in his considerations. Arrhenius had just published the theory of ionic dissociation only four years before Hofmeisters article "Zur Lehre von der Wirkung der Salze"[163]. Accordingly, he had a well-educated understanding of his research topic at the time. His research was mainly focused on protein chemistry, but he also compared his results on proteins with results produced with gelatin, colloidal iron oxide or sodium oleate.

Hofmeister initially explained his results with a "water absorbing effect" of varying degree, which he wanted to link directly to the salting-out behavior of the respective salts. The terms "kosmotrope" and "chaotrope" developed from his theory. The terms were chosen because of the questionable theory that strongly hydrated ions such as fluorides and sulfates can orient water molecules in coordination shells around the anion and thus forming solvent layers which orient and influence each other. Accordingly, the chaotropic salts were unable to form such solvent shells and brought about the opposite, a disorder in the orientation of the solvating water molecules. This terminology is still in use until today and used colloquially. For example, denaturants are also known as chaotropes. Figure 1 shows Hofmeister's series from 1888 immortalized in front of Charles University in Prague.

Figure 1 Commemorative plaque to Franz Hofmeister and his research in front of the Charles University in Prague[164].

Referring to the Hofmeister series chloride has historically been classified as the borderline ion between chaotropic and kosmotropic ions. The kosmotropic ions are found to the left of chloride, while the chaotropic ions are found to the right side of chloride.

In this Chapter, Hofmeister effects are discussed with relation to the thermostability of proteins. By using "nano differential scanning fluorometry" (nanoDSF), the Hofmeister effects of sodium and chloride salts on the thermostability of three enzymes, namely lysozyme from *Gallus gallus*, alcohol dehydrogenase from *Saccharomyces cerevisiae* and fructose-6-phosphate aldolase from *E. coli*, which differ significantly in their structural properties, were determined based on the change of melting temperature T_m induced by salt additives.

To identify or screen enzymes as potential candidates for biotechnological processes, it is important to establish a reproducible high-throughput screening method. Thousands of proteins are often screened for a wide variety of properties in just one day. It is necessary to test the enzymes for various factors like thermostability, chemostability and substrate scope. NanoDSF equipment, which was used during this work, is suitable for the screening for thermostability of proteins in a reliable and high-throughput manner.

2.1 Thermal and chemical stability of proteins

The mechanisms of protein unfolding, a switch between stable forms or the inhibition of enzyme activity can be crucial for treating diseases[165]. The main physical forces accounting to the thermostability of proteins were summarized by KAUZMANN in 1959[166]. The interplay of hydrophobic effects, electrostatic interactions, hydrogen bonds and covalent cross links define the quaternary structure of proteins. Environmental conditions such as temperature, pressure, pH, ionic strength and concentrations of ligands, stabilizers and denaturants can have a significant impact on the stability of proteins[167].

The search for thermostable enzymes and methods to optimize known enzymes for improved thermostability, forms a large body of work on protein stability[168-170]. A main source for thermostable enzymes are extremophilic bacteria that optimally grow at 85°C or higher, which are generating enzymes able to endure elevated temperatures without denaturation[171]. Through natural evolution, for example the DERA variant from *Thermotoga maritima* (*Tm*DERA) got optimized by nature towards higher thermal stability, which reflects the energy necessary to break its native structure. Interestingly, this stabilizing evolution usually leads to a decline in room temperature activity[172]. By comparison of intramolecular interactions within the structures of *Ec*DERA and *Tm*DERA, it is clearly visible that the thermophilic variant from *Thermotoga maritima* possesses more interactions, packing the structure in a more rigid manner compared to *Ec*DERA. Figure 2 presents both enzyme structures and the attractive forces defining the quaternary structure of the corresponding enzyme.
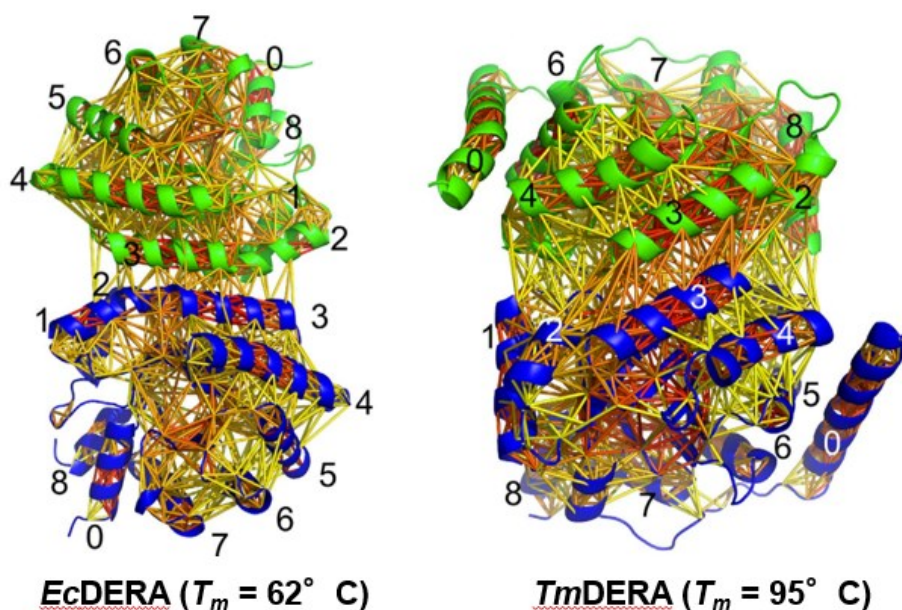
Figure 2 Structural rigidity of mesophilic *Ec*DERA and thermophilic *Tm*DERA of amino acid residues in the protein structure. The strength of the interactions is color coded from yellow (weak) to red (strong) (picture taken from ref [172]).

The higher degree and strength of interactions is especially the case for the contact area of both monomers forming the dimeric structure. Due to a higher degree of polar and hydrophobic interactions, the energy necessary for breaking those bonds, or in other words for denaturing the proteins' quaternary structure, is increased with each interaction.

It is important to notice, that *Thermotoga maritima* for example has an optimal growth temperature of 85 – 90°C reflecting its natural habitat at sea ground level living near hot smokers [173]. In contrast, *Escherichia coli* the optimal growth temperature is 37°C[174].

It is necessary to differentiate between thermodynamic and kinetic stability of proteins. The unfolding of the protein structure is a measure of its thermodynamic stability and reflected by the melting temperature T_m, the change of free energy necessary for the unfolding ΔG_{unfold} and the unfolding equilibrium constant K_{unfold}. The most prominent methods for measuring thermodynamic stability are differential scanning calorimetry (DSC) or differential scanning fluorometry (DSF), circular dichroism (CD) or the adaption of DSF and DSC to follow intrinsic tryptophan or tyrosin fluorescence or the change of free energy at the nano scale [175]. All methods for determination of protein melting temperatures will be discussed in section **2.2**.

Kinetic stability refers to the length of time for a protein to remain active before losing activity due to irreversible unfolding[175]. For an example of an experimental procedure, a protein sample solution gets incubated under certain conditions including specific elevated temperatures, pH and chemical gradients. Samples are then taken at different time points for testing the residual activity of the protein sample in standardized test reactions. A commonly used value is the half-life of the activity T_{50}, the point at which the activity of the enzyme is halved. Additionally, the optimal working temperature T_{opt} is also commonly used for enzyme characterization.

Interestingly, thermostability or the ability to avoid irreversible thermal unfolding seems related to the ability to avoid irreversible denaturation under chemical stress[171]. Hence, it is interesting to measure not only the protein's thermal stability but also the chemical stability with denaturants like guanidinium hydrochloride, urea, lithium perchlorate or sodium dodecyl sulfate[176-178].

2.2 Enzyme and protein stability measurement

Industrial processes demand the biocatalyst to be sufficiently stable under the implemented process conditions[179-180]. The most prominent factors that have an impact on protein stability in industrial processes are high substrate loads, non-neutral pH, co-solvent addition and elevated temperatures. A common trend to overcome insufficient biocatalyst stability against such factors is the search for novel thermostable enzymes from extremophilic environments such as hot smokers on sea ground[173, 181], deserts[182] and many more reflecting an immense diversity of enzymes from different microbial or bacterial origins. While harvesting those enzymes, one factor that gives direct indication about their thermostability is the melting point of the protein.

The definition of the melting point of a protein (T_m) is the state at which half of the protein molecules are unfolded, while the other half is still in the native state. Another informative value which can be automatically detected, is the onset of unfolding T_{on}. This value reflects the temperature at which the protein starts to unfold. Another factor which can be obtained with a modified heating cycle yields the point after which the protein is unable to refold in its native state. This value is described as the point of irreversible unfolding[183].

In the following part three different thermal shift assay methods will be explained for determination of the melting point of protein samples. The selection contains differential scanning calorimetry (DSC), differential scanning fluorimetry (DSF) and the further development nanoDSF. An alternative method for the determination of the melting points of protein samples is circular dichroism (CD), which will not be explained in detail. In comparison, the necessary amount of protein sample makes it unfeasible, compared to the other methods, for high throughput screenings.

It is important to notice that buffers are commonly used in such experiments, potentially having a (de)stabilizing effect on protein stability. All methods follow the same concept of heating the protein sample at a specific concentration along a defined temperature gradient. A commonly found value for the temperature gradient in the literature is 1 °C/min, while the protein concentration is bound to the limits of detection of each method. A clear dependence of T_m of a protein is found for the heating rate from many examples in literature[184-187].

The results from one of these studies[187] demonstrate that the heating rate and the concentration of the protein sample can have a direct impact on the observed melting point of the sample. The differences in T_m were approximately less than 10% of the total measured T_m but nonetheless significant. Therefore, it is important to keep protein concentration, buffer concentration, pH and heating rate constant for thermal shift assays to maintain comparability of the results.

2.2.1 Differential scanning fluorometry (DSF)

DSF has developed to one of the main techniques for rapid determination of the melting point of a purified protein sample in biological laboratories[188] after its initial presentation of SEMISOTNOV et al. in 1991[189-190].

DSF experiments are performed in adjusted PCR devices, where it is possible to follow fluorescence of the sample during heating cycles[191]. The sample is prepared before the measurement to contain environment-sensitive fluorescence dyes such as Sypro orange®, Proteostat® or, as initially reported in 1991, 1-anilino-naphthalene-8-sulfonate, which are all able to bind to hydrophobic parts of proteins. The property of the fluorescence dye results in a shift of fluorescence intensity upon binding to the protein's hydrophobic parts during unfolding. When bound to a hydrophobic environment, the dye is protected against the excited state quenching effect of water, resulting in a change of fluorescence[192].

On the contrary, in the native state of the protein such binding is not possible because hydrophobic amino acid residues are usually deeply buried in the hydrophobic protein core.

The technique yields reproducible results when sample preparation is done carefully and using fresh stock solutions especially for the fluorescence dye. It is important to mention that additives like the fluorescence dye can have a minor effect on the protein stability itself[193]. For the thermal shift assay, it is important that the target protein unfolds hydrophobic parts for the binding of the fluorescence dye. Therefore, not all protein structures can be similarly well analyzed by DSF.

2.2.2 Nano Differential scanning fluorometry nanoDSF

A further development of DSF (nanoDSF) operates by monitoring the intrinsic fluorescence of tryptophan and tyrosine residues during the unfolding of the protein sample without the need for any dye additives. A typical nanoDSF device is the *Prometheus* system of the NanoTemper company[194]. The supplier claims that for the detection of the melting temperatures based on tryptophan fluorescence only a single tryptophan residue in the protein structure is required to give sufficient signal intensity for the measurement. For detection of tyrosine fluorescence at least 5 – 10 residues are required in the protein structure. Of course, all detectable residues need to experience a shift from hydrophobic to hydrophilic environment to yield a detectable signal change.

A typical result of a melting point measurement is shown in Figure 3, produced with the *Prometheus* system with lysozyme as protein sample in TEA*HCl buffer (100 mM, pH 7.5).

Run	T_m
1	72.5
2	72.4
3	72.4

Sample:
- 100 mm TEA*HCl pH 7.5
- 0.5 mg/ml lysozyme
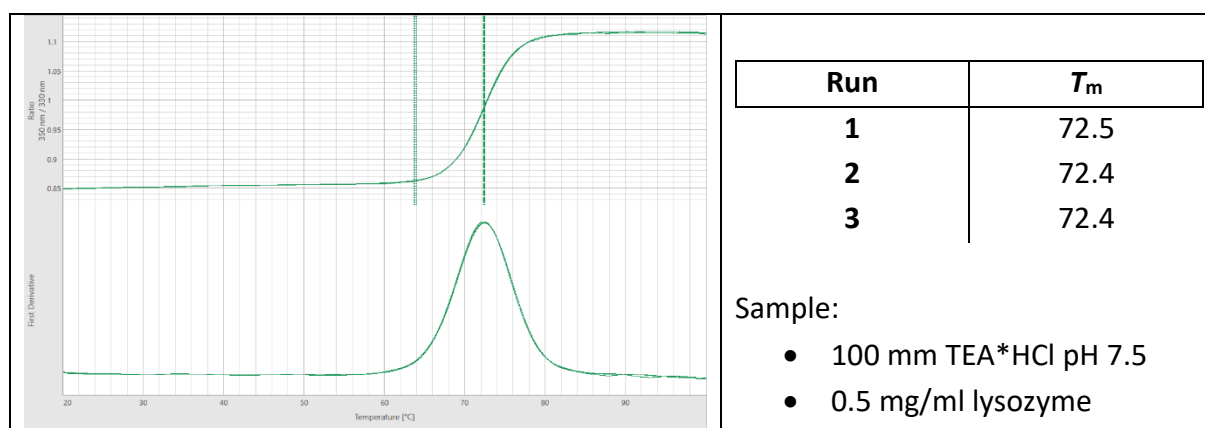
Figure 3 T_m measurement of lysozyme in 100 mm TEA*HCl buffer (pH 7.5) with 0.5 mg/ml lysozyme from hen egg white (*Gallus gallus*) in triplicates. The sigmoidal curve (top) reflects the experimental signal measurement. The parabola (bottom) is the first derivative of the sigmoidal growth curve yielding the T_m of the sample as the maximum.

Most hydrophobic amino acid residues form a hydrophobic core in the native state of a protein, which is surrounded by a more hydrophilic layer of amino acids interacting with water molecules. During unfolding, the hydrophobic amino acids experience a shift of environment towards the hydrophilic surrounding. For provoking the unfolding, usually a temperature gradient is applied. This leads to a shift of the intrinsic fluorescence of tryptophans and tyrosins from 330 to 350 nm. By detecting the change of fluorescence at 330 and 350 nm and

calculating the ratio of both signals (350 nm/330 nm), this yields a sigmoidal growth curve with the turning point reflecting the melting temperature of the protein sample. By calculating the turning point by way of the peak of the first derivative, the melting point of the sample is acquired. It is also possible to plot a single wavelength against the temperature but this will limit the sensitivity of the measurement.

A typical nanoDSF sample only needs water and protein to determine the melting point of the protein. Usually, the sample contains additional buffer to give sharper signals and also to mimic the conditions of interest. After sample preparation, the sample is filled into ultra-thin quartz capillaries which can hold 10 μL of sample solution. For detection of melting points of thermophilic proteins, the sample capillary can be sealed with sealing paste and heated to a maximum of 110°C. In the literature, there are examples of metagenomic enzyme sources showing such high melting points for some variants in simple buffer solutions[87].

Amazingly, nanoDSF can directly determine the melting points of an unpurified protein sample which contains cell free extract without additives[195]. Cell free extract, for example from BL21 expression cell lines, contains a broad background of proteins. This background is neglectable as long as the target protein is sufficiently overexpressed. The necessary degree of overexpression depends on the available tryptophan residues in the protein structure contributing to the fluorescence signal[195].

Through nanoDSF measurements, four important protein parameters can be obtained. The first is the melting point itself. The onset of the melting event, indicating the conditions and temperature at which the unfolding of the sample starts, is obtained during a standard measurement at once with the melting point. This value gives valuable information on stable storage conditions which do not induce denaturation. By using a temperature gradient for the cooling after the experiment is finished, one can identify if the folding is reversible. The last important information is the point of irreversible unfolding, which can be obtained with a modified heating cycle[183].

A protein can yield more than one peak, if domains unfold separately as it is the case for yeast hexokinase B, when no stabilizing substrates or additives are present[196].

Especially the high throughput manner of the *Prometheus* is important to notice. A single run yields the melting points of up to 48 samples in a single parallel run. With a typical temperature gradient from 20°C – 90°C and the maximum possible heating rate of 7°C/min, the melting points of 48 samples can be obtained in 10 minutes. Considering that the device can be handled by robotic arms systems, the automatic determination of thousands of melting temperatures can be accomplished in one day. With current trends like metagenomic enzyme mining, tools like nanoDSF are required to quickly get important information on protein thermostability.

A prominent example from the literature utilized nanoDSF equipment in the search for thermostable and methanol stable lipases for the production of biofuels[197]. In the study, the wild type lipase LipT6 was compared to a triple mutant LipT6 H86Y/A269T/R374W termed LipT6$_M$. The authors were able to demonstrate that the triple mutant had considerably higher melting temperatures in the presence of methanol under all tested conditions with high

precision. The highest obtained error was less than 1%, reflecting the high precision of the *Prometheus* system with purified enzymes.

Another study demonstrated that even protein mixtures from a cell free extract with sufficient overexpression can be analyzed[195]. The protein of interest, the cyclohexanone (**61**) monooxygenase from *Acinetobacter calcoaceticus,* possesses (pdb 6A37) 31 tryptophanes with a total amino acid acount of 552[198]. Therefore, the melting event is easily detectable considering that one tryptophan is already enough for precise detection. In the study, the protein was easily detectable at a slight overexpression rate of 20% or more, while at 10% overexpression the signal could not be analyzed with confidence since protein background can produce equivalent signals.

2.2.3 Differential scanning calorimetry

DSC is a label-free technique allowing to measure melting temperatures of protein samples with high reliability based on the heat capacity of states and the excess heat associated with transitions [199]. The pioneering work on DSC measurements for determination of protein thermostability was performed in the 1960's and 1970's[200-204]. Those early studies were performed with proteins that were commercially available and known from the literature to be appropriate to study the thermodynamic driving forces related to protein stability, folding and binding interactions[199].

DSC measurements are nowadays commonly used for determination of thermostability of proteins with drastic improvements concerning the necessary amount of sample and automatization of the measuring process[205]. During the measurement, the sample temperature is increased in a calorimeter while detecting the heat necessary for the temperature increase. The sample preparation and amount of protein necessary for most devices is comparable to other techniques. Results produced with DSC are highly reproducible. An exemplary measurement of lysozyme in glycerol is presented in Figure 4.
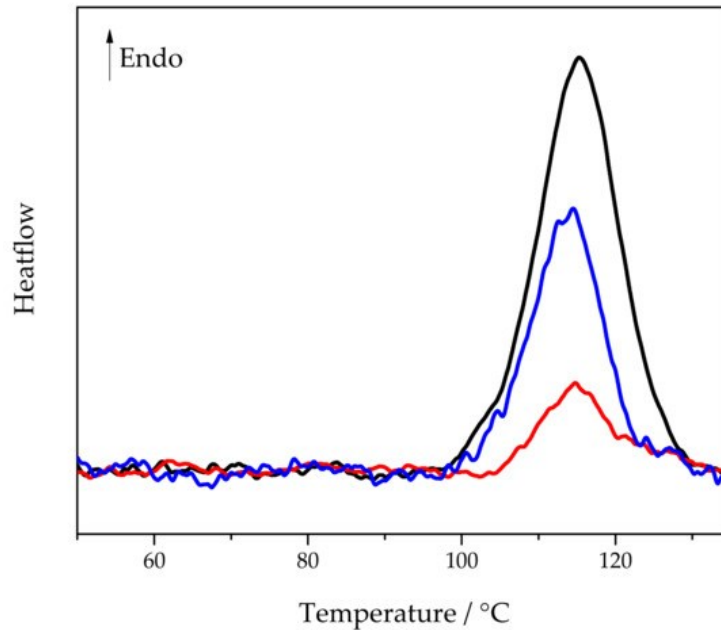
Figure 4 Exemplary DSC thermogram of lysozyme in glycerol. Black line reflects initial measurement, blue line produced in remeasurement after annealing at 70°C for 600s while for red lines annealing time is 25s (graphic taken from ref [206]).

The peaks of the melting curves from DSC measurement reflect directly the melting temperature of the protein sample.

The further development of the state of the art of the technology nanoDSC enables measurements in a desired 96-well plate format with 500 µL sample volume and minimal detectable amount of 2 µg protein, which enables HTS of large mutagenic or metagenomic enzyme sources [207].

2.3 Effect of additives on protein stability

In biochemical industrial processes, an enzyme catalyst needs to be robust to endure certain stress factors such as elevated temperature, high salt freight, high substrate loads or presence of cosolvents. In nature, a local environment can offer equivalently harsh conditions. Hence, natural adaption developed measures for preventing lethal damage to organisms by implementing protective mechanisms. One way is the biosynthesis of so-called compatible solutes that interact with biological structures like proteins to stabilize them against stress factors[208-211].

Small molecules with low molecular weight, which can improve the stability of proteins in an aqueous solution are typically polyols[212-213], buffers[214-215], most amino acids[216], inorganic salts [217-219] and specific polymers like gelatine[220].

2.3.1 Stabilizing effect of polyols

A common observation in biological chemistry is that polyols have a stabilizing effect on proteins in aqueous solution[221]. The underlaying effect, namely the concept of preferential exclusion, describes the solvophobic effect of polyols on the protein backbone[222], meaning that the denatured form is less well hydrated and is therefore energetically unfavored leading to a force on the protein to stay in its native form.

Interestingly, the stabilizing effect increases with the size of the polyol[222]. The chymotrypsin inhibitor 2 was simulated in different aqueous solutions containing different polyols, namely glycerol, xylitol, sorbitol, trehalose and sucrose, and the results were compared with the MD simulation using solutions containing ethanol and glycol. The results of the simulation demonstrated that only a few hydrogen bonds between the protein and a polyol are prevailing and that the number of bonds is similar for different polyols. Additionally, an increase of rigidity of the first hydration shell surrounding the proteins surface was observed with increasing size of the polyol. The competition with solvating water molecules, the number of which is lower in the native state of a protein than in its unfolded state, results in an energetically unfavorable situation, which does not allow the protein to unfold.

2.3.2 Stabilizing effect of buffers

In 1966 GOOD published an initial list of 14 different buffer systems, which were later extended with additional eight buffer substances[223-224] that can be used for biological experiments in different pH ranges, focusing mainly on the range of pH 6-8[214].

Buffers were long interpreted to stabilize protein structures mainly by pH regulating effects rather than direct interactions. It was later shown that protein formulations, especially at higher protein concentrations, are mainly buffered by the protein itself rather than by the buffer[225]. Direct binding of the buffer species to the protein's surface and the influence of the solute on the hydration shell of the protein were identified as the origin for enhanced protein stability[226].

2.3.3 Hofmeister effects

In 1888, Franz Hofmeister published his famous study "Zur Lehre der Wirkung der Salze" [161], which translates to „About the science of the effects of salts"[162]. One part of his study utilized separated liquid from beaten hen egg white as model protein solution. Hofmeister was interested in the ability of ions to salt-out proteins from aqueous solution. It is important to set his experiments in scientific context, since for example ionic dissociation was just proposed by Arrhenius a few years earlier[163]. Hence, Hofmeister followed on the idea of ionic dissociation and was interested in the specific effects of each ion separately. Therefore, he

used a series of sodium and of corresponding chloride salts and compared the effects of various anions and cations separately.

After salt addition, the time was measured until the solution turned cloudy, which Hofmeister interpreted as protein precipitation. The famous Hofmeister series was created based on those acquired time values for protein precipitation. Anions on the left side of the series were considered to have "salting out" ability, while anions on the right side have "salting in" ability. Chloride was considered to be the borderline case between salting in and salting out behavior. Based on the Hofmeister series, the terms kosmotropes and chaotropes were proposed. The terms were created based on early interpretations of the Hofmeister effects that anions can either bring order or chaos to the ordering of the water network. Kosmotropes would therefore build hydration shells surrounding the ion influencing each other, while chaotropes do not possess that ability and will decrease the order of the water network. The terms are commonly utilized in the literature up to date with an adjusted meaning. For example, the term chaotropes is commonly used for denaturant agents like guanidinium hydrochloride, lithium perchlorate or urea.

The theory of water ordering effects of kosmotropes was only disproved by OMTA et al. in 2003[227], by measuring the orientational correlation time of water molecules in magnesium perchlorate, sodium perchlorate and sodium sulfate solutions utilizing of femtosecond pump probe spectroscopy. In the experiment, no effects of salt ions could be detected that would influence water molecules apart from the first solvation shell of an ion. Considering that Hofmeister's experiment was published in 1888 and the first plausible explanation was falsified just in 2003, a precise mechanistic explanation is still needed. The main reason for such a late explanation or scientific proof is the difficulty imminent in the observation of water itself since most spectroscopic methods are incapable of observing such fast events like water reorientation, on a femtosecond scale.

The best correlation for Hofmeister effects up to date is found in the Jones-Dole-equation[228], which relates salt/solvent effects to the viscosity of the salt solution (η) relative to the viscosity of water (η_0).

$$\frac{\eta}{\eta_0} = 1 + Ac^{0.5} + Bc + Dc^2$$

Ac⁻	SO₄²⁻	F⁻	Cl⁻	Br⁻	NO₃⁻	ClO₄⁻	I⁻
0,250	0,206	0,127	-0,005	-0,033	-0,045	-0,061	-0,073

The first term $Ac^{0.5}$ describes electrostatic effects (Debye-Hückel) and is relevant at concentrations below 0.1 M. The second term Bc describes ion-solvent interactions, where B is specific for each ion. The B-value gives the best correlation to Hofmeister effects with chaotropic ions having a negative B-viscosity value, while kosmotropic ions possess a positive

B-value. Many B-viscosity values are available in the literature[229-232]. The last term Dc^2 is only necessary at very high salt concentrations for describing ion-ion interactions.

One explanation attempt for Hofmeister effects describes the concept of preferential interactions. The idea that the solute is favorably interacting with the protein instead of water leads to an unfolding force to create additional binding sites for the solutes inducing the unfolding of the native state with high salt concentrations. Preferential binding or exclusion can be measured by equilibrium dialysis or vapor pressure osmometry[233-234], directly reflecting the transfer energy necessary for converting a protein from its native to an unfolded state.

3   Aim of work

Considering current geopolitical trends and the impact of the energy crisis on chemical industry[235], the implementation of green processes gains in importance[197, 236]. For the identification of industrially relevant biocatalysts, extensive screening of thousands of candidates is performed for identification of suitable candidates for biocatalytic processes[237]. The candidates are screened for activity in the desired reactions and additionally for their thermo- and chemostability. Thermostable enzymes especially were found to be promising candidates for further development to optimize the enzyme's tolerance to crucial process conditions such as high concentrations of cosolvent or elevated temperatures.

Prior to this work, our group was pioneering studies into the effects of cosolvent on the thermostability using a collection of enzymes[87]. The determination of the melting points of the enzymes was performed with nanoDSF equipment (*Prometheus*) from Nanotemper Technologies GmbH.

In this work, specific ion effects of sodium and chloride salts were investigated concerning their impact on the thermostability of hen egg white lysozyme (HEWL), alcohol dehydrogenase from *Saccharomyces cerevisiae* (ADHY), fructose-6-phosphate aldolase from *Escherichia coli* (*Ec*FSA) in aqueous solution added high concentrations of sodium salts of various mono-, bi- and trivalent anions and measured the time until the protein solution showed first signs of precipitated protein. His results on the so-called Hofmeister series[238] should be compared with modern T_m measurement technology in this work.

For this purpose, nanoDSF equipment was utilized following the intrinsic fluorescence of tryptophanes during thermal unfolding. The melting points of the enzymes should be determined in defined sodium or chloride salt solutions up to 1 M concentration. Equivalent studies on lysozyme performed with DSC were available in the literature, and lysozyme was therefore chosen as a prime candidate for comparison. Lysozyme possesses a monomeric structure with a positively charged surface at pH 8[239-240]. ADHY, the second enzyme of our selection possesses a dimeric structure with a slightly negative surface charge[241]. *Ec*FSA as the last enzyme of the selection possesses a decameric structure (dimer of pentamer) with a clearly negatively charged surface[88] and was already used in previous studies of our group[87, 110, 119, 122].

4 Results

4.1 T_m measurement of HEWL, ADHY and FSA

ADHY (pdb code 5ENV) is a dimeric enzyme, containing 347 amino acids per monomer. *Ec*FSA (pdb 1L6W) posseses a decameric structure with each monomer containing 220 amino acids. HEWL (pdb 1DPX) is a monomeric enzyme containing 129 amino acids. The monomeric form indicates that no hydrophobic areas are available at the outer sphere of the protein indicating mostly hydrophilic amino acids at the proteins surface. Therefore, HEWL is an ideal enzyme for testing thermostability on a small globular protein. Lysozyme is special concerning the surface charge of the crystal structure when applying adaptive Poisson-Boltzmann Solver (APBS)[240, 242]. The result of the calculation with APBS indicates that the surface net charge of Lysozyme is of a positively charged nature, while both other enzymes ADHY and *Ec*FSA possess a more negatively charged surface. Figure 5 presents the graphical demonstration of the APBS plugin for pymol[240, 242].
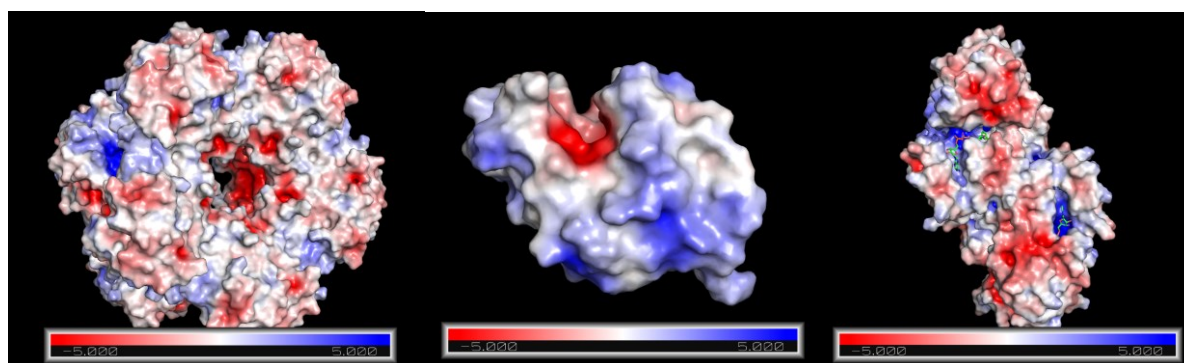
Figure 5 Surface potential of EcFSA (left), HEWL (middle) and ADHY (right) calculated with the APBS plugin for PyMoL indicating a positive surface net charge for HEWL, while for EcFSA and ADHY a negative surface potential is obtained[240, 242].

As a starting point, it was necessary to measure all three enzymes, namely HEWL, ADHY and *Ec*FSA, without any additives except 50 mM TEA*HCl (pH 7.5) buffer. In a previous study from our group, the measurement was performed in 50 mM TEA*HCl buffer at pH 7.5 with *Ec*FSA[87]. The same conditions were utilized for the T_m measurement of the above-mentioned enzymes and for all following experiments for comparison.

For the determination of the melting points via nanoDSF, the sample is filled in provided ultrathin quartz capillaries. The capillaries were sealed with the provided sealing paste and placed in the Prometheus device. As the temperature gradient, the literature standard of 1° C/min was chosen. The applied temperature ranged from 20° C up to 110 °C. In Figure 6 the result of the measurement for HEWL is presented as an example with 7 repeats to evaluate the experimental error of the measurement. The measurement was performed in TEA*HCl buffer (50 mM, pH7.5).

Run	T_m
1	73.0
2	73.3
3	73.0
4	73.1
5	73.2
6	73.0
7	73.1
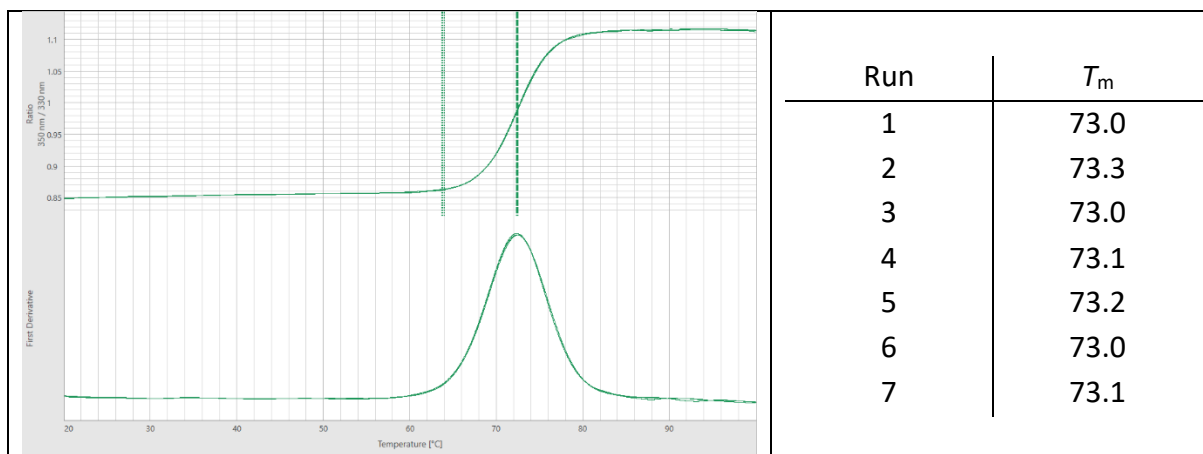
Figure 6 Exemplary melting point measurement for HEWL of 7 different samples in 50 mM TEA*HCl buffer (pH 7.5)

The average error of the measurement was small with all three tested enzymes. Additionally, the obtained average value of 73.1±0.1 °C for HEWL is very well comparable with the literature values[243].

With a standard deviation of ±0.1°C of the total T_m value, the measurements for additive screening will be performed in unicates instead of triplicates to increase throughput and lower the total screening costs due to capillaries being one time use.

4.2 Hofmeister effects on protein thermostability

4.2.1 Anionic Hofmeister series

In general, the effects of anions on protein structures are more prominent than the effects of cations. Therefore, the effect of Na^+ cations as standard counterion should have minimal effect on the measurement of the tested enzymes as can be seen in section **4.2.2**.

The effects of sodium salts on the thermostability of HEWL are especially well studied in the literature[243-245]. Lysozyme was therefore utilized as a reference enzyme, for which the salt effects on thermostability had been published for comparison, but were determined by using nanoDSC instead[243]. It is important to notice that absolute T_m values can differ slightly but trends obtained with different methods should be very well comparable. Table1–3 shows the results of the thermostability measurement within a $0 - 1$ M concentration range using the following sodium salts: NaF, NaCl, NaBr, NaI, Na_2SO_4, $Na_2S_2O_3$, NaOAc, $NaNO_3$, $NaClO_4$. The measurements were performed under standard conditions in 50 mM TEA*HCl buffer adjusted to pH 7.5. The salt solutions were prepared as 2 M stock solutions for generating a dilution series, so that all the necessary final salt concentrations were reached. In general, the enzyme stock solution contained twice the desired amount of enzyme and concentration of buffer as to be utilized in the actual measurement.

Table 1 Resulting ΔT_m measured by nanoDSF in TEA*HCl buffer (50mм, pH 7.5) for *Ec*FSA. The colour code corresponds to the total value of ΔT_m (ΔT_m>10°C, ΔT_m>5°C, ΔT_m>0°C, ΔT_m>–5°C, ΔT_m > −10°C, ΔT_m<−10°C).

EcFSA									

Conc.	ΔT_m [°C]								
[mм]	Na_2SO_4	Na-Ac	$Na_2S_2O_3$	NaF	NaCl	$NaNO_3$	NaBr	NaI	$NaClO_4$
1000	6,2	9,2	9,3	7,2	5,7	1,9	1,9	-5,4	-7,3
500	7,6	6,8	8,3	6,9	4,6	2,1	2,6	-1	-2,5
250	6,4	5	6,5	6,2	3,8	2,2	2,7	1,1	0
125	5,2	3,4	5,5	4,1	2,9	2	2,2	1,5	1
62,5	4,2	2,2	4,7	3	2,1	1,6	1,5	1,2	1,3

Table 2 Resulting ΔT_m measured by nanoDSF in TEA*HCl buffer (50mм, pH 7.5) for ADHY. The colour code corresponds to the total value of ΔT_m ($\Delta T_m{>}10°C$, $\Delta T_m{>}5°C$, $\Delta T_m{>}0°C$, $\Delta T_m{>}{-}5°C$, $\Delta T_m > -10°C$, $\Delta T_m{<}{-}10°C$).

ADHY

Salt conc. [mм]	ΔT_m [°C]								
	Na$_2$SO$_4$	Na-Ac	Na$_2$S$_2$O$_3$	NaF	NaCl	NaNO$_3$	NaBr	NaI	NaClO$_4$
1000	20,1	1,5	3,8	0,2	-4,1	-9,1	-9,4	-15,3	-19,1
500	0,5	-1,5	-1,6	-0,6	-4,4	-7,5	-6,9	-10,3	-13,6
250	-3,2	-2,1	-2,9	-1,2	-3,5	-5,6	-4,9	-6,8	-8,6
125	-3,5	-1,7	-3,2	-1,7	-2,4	-3,8	-3,4	-4,1	-5,5
62,5	-2,5	-1,5	-2,5	-1,5	-1,9	-2,4	-2,1	-2,7	-3,1

Table 3 Resulting ΔT_m measured by nanoDSF in TEA*HCl buffer (50mM, pH 7.5) for HEWL. The colour code corresponds to the total value of ΔT_m ($\Delta T_m > 10°C$, $\Delta T_m > 5°C$, $\Delta T_m > 0°C$, $\Delta T_m > -5°C$, $\Delta T_m > -10°C$, $\Delta T_m < -10°C$).

HEWL

Salt conc. [mM]	ΔT_m [°C]								
	Na$_2$SO$_4$	Na-Ac	Na$_2$S$_2$O$_3$	NaF	NaCl	NaNO$_3$	NaBr	NaI	NaClO$_4$
1000	7,1	3,6	0,3	2,4	1,6	-1,6	-2	-6,8	-11,1
500	4,2	1,5	-3,5	2,3	-0,3	-2	-2	-4,6	-8,1
250	1,5	0,4	-4,7	1,8	-0,7	-1,7	-1,9	-3,5	-6
125	0,1	0	-4,7	0,7	-0,7	-1,2	-1,5	-2,5	-4,2
62,5	-0,3	-0,1	-4,2	0,3	-0,5	-1	-1,1	-1,7	-2,8

When comparing the data for HEWL with chloride, iodide, sulfate and perchlorate to available data from the literature produced with nanoDSC[243], the results are very well comparable and show only small differences in absolute values, while the trends are exactly the same. The conclusion from the comparison is that the nanoDSF method is obviously well suited under the implemented conditions, for gathering reliable data with very limited sample consumption and little experimental effort.

For an interpretation of the observed effects, a useful differentiation of the anions is found in the classification based on level of hydration, which results from different charge densities of the anions[228, 246-247]. COLLINS demonstrated with aqueous gel filtration chromatography[247] on Sephadex G-10 columns with cross-linked dextran gel containing a hydrophobic surface that strongly hydrated anions flow through the column without any delay, while less well hydrated

anions apparently interact with the hydrophobic surface, resulting in a delay and an increase of the retention time. This simple method demonstrates that strongly hydrated anions are unlikely to bind to hydrophobic areas of a protein, whereas less well hydrated anions are able to dissociate water molecules from the hydrate shell, enabling an interaction with the hydrophobic surface. This finding was supported by another study[248] on anion binding using a non-charged 600-residue elastin-like polypeptide (VPGVG)$_{120}$. For the latter study, molecular dynamic simulations were applied that demonstrated the areas for anion-peptide interaction. The authors could show that weakly hydrated anions like SCN$^-$ have significantly more interaction sites and longer interaction times than strongly hydrated anions like sulfate. Furthermore, the authors confirmed the results of the MD simulation with an ^{1}H-NMR study on the backbone protons of the peptide.

The general Hofmeister trend from the series obtained in 1888[161-162] was confirmed by the measurement of thermostability for all three tested enzymes. Using sodium salts as constant additives the observed effects are clearly based on the chaotropic or kosmotropic effect of the anion.

$$SO_4^{2-} < Acetate^- < F^- \quad < Cl^- < \quad S_2O_3^{2-} < NO3^- < Br^- < I^- < ClO_4^- \qquad \text{HEWL}$$

$$SO_4^{2-} < S_2O_3^{2-} < Acetate^- < F^- \quad < Cl^- < \quad NO3^- < Br^- > I^- < ClO_4^- \qquad \text{ADHY}$$

$$S_2O_3^{2-} < Acetate^- < F^- < SO_4^{2-} \quad < Cl^- < \quad NO3^- = Br^- < I^- < ClO_4^- \qquad Ec\text{FSA}$$

$$F^- = SO_4^{2-} < Acetate^- \quad < Cl^- < \quad NO3^- < Br^- < I^- < ClO_4^- \qquad \text{Hofmeister}$$

By comparing the resulting Hofmeister trends at 1 M sodium salt concentration for each enzyme, it can be concluded that the destabilizing weakly hydrated anions show a constant ordering independent of the protein's structure, while the stabilizing and well hydrated anions can switch position relative to the original Hofmeister series.

Sulfate is a well-known protein stabilizer, commonly used with ammonium as a counterion for protein precipitation[249]. The main attractive features of ammonium sulfate are its very high-water solubility and low costs for high purity material. Purifying proteins is nowadays usually performed by applying a tag to the protein (such as a His$_6$tag) to aid in the purification itself. With ammonium sulfate precipitation such a tag is not required, and therefore this method is often used as initial purification step for native small globular proteins. [250-251] With all three enzymes, sulfate shows a clear tendency to stabilize at high concentrations. The highest increase of T_m overall was measured for ADHY at 1 M concentration of sulfate, increasing the T_m by 20.1 °C, while HEWL experiences an increase of T_m by 7.1 °C. Only for EcFSA a unique behavior is observed. For HEWL and ADHY sulfate is clearly the strongest stabilizer, while for EcFSA sulfate is observed as the fourth strongest stabilizer increasing the T_m by 6.2 °C. Interestingly, for EcFSA at 1 M sulfate concentrations the stabilization is less strong than with 0.5 M concentration. The same effect was observed for ammonium sulfate.

Thiosulfate shows an interesting behavior when compared to sulfate. At concentrations up to 0.5 M the stabilization effect was directly comparable to sulfate for EcFSA and ADHY. Surprisingly, HEWL experiences a strong destabilization at low concentrations of thiosulfate with a reverse trend to stabilize at higher concentrations. The reason for such a behavior so far remains unclear. At 1 M concentration of thiosulfate, EcFSA is stabilized by 9.3 °C, ADHY by 3.8 °C and HEWL by 0.3 °C. The main difference between sulfate and thiosulfate is the charge distribution of thiosulfate. Thiosulfate possesses a more localized charge at the sulfur compared to sulfate in which both charges are delocalized. Since the effect is only observed with HEWL, which has a positive surface charge, one can speculate that attractive electrostatic interactions of thiosulfate with the proteins surface induce the destabilizing effect. An alternative interpretation could originate from a conformational destabilization of HEWL, for which low salt concentration might induce change from closed to a more flexible open confirmation.

Acetate is a commonly available anion in nature. Therefore, a destabilization was not expected since nature most likely would adapt to common intracellular solutes. With all three enzymes, acetate was found as the second-best stabilizer with exception of the behavior of thiosulfate with ADHY. EcFSA experienced a stabilization under all tested conditions, demonstrating the highest stabilization of 9.2°C at 1 M concentrations, whereas HEWL and ADHY experienced a minor destabilization at low acetate concentrations. At 1 M concentration, HEWL is stabilized by 3.6°C and ADHY by 1.5°C. Comparing the results for acetate, apparently with increasing complexity of the quaternary structure of the enzymes the stabilization effect is increasing.
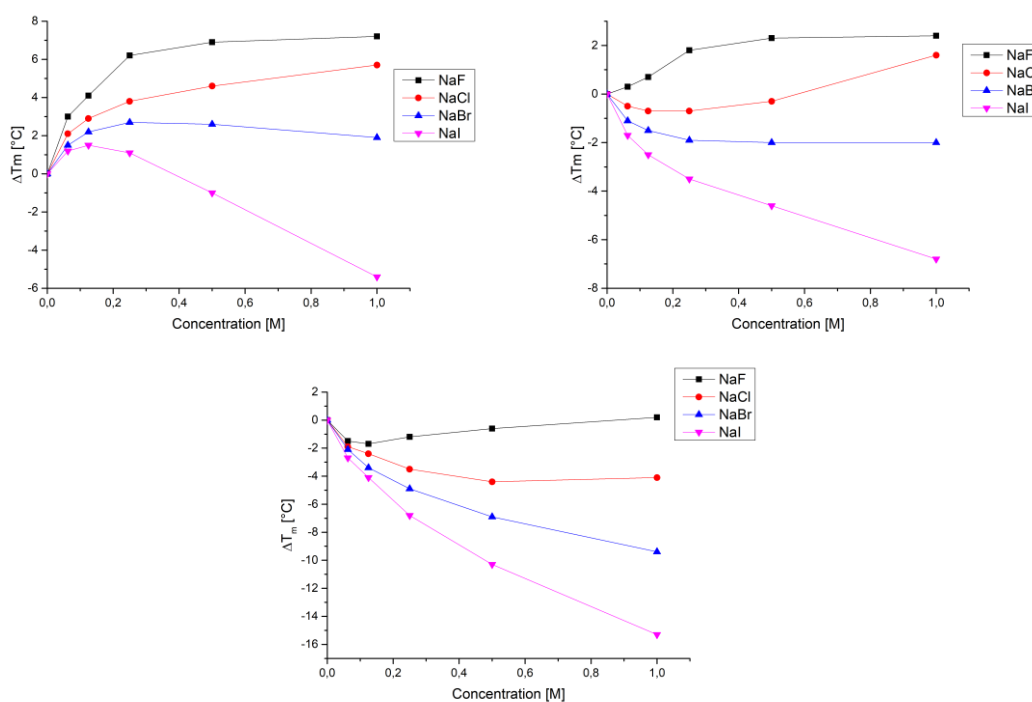
Figure 7 Resulting ΔT_m measured by nanoDSF in TEA*HCl buffer (50mM, pH 7.5) for EcFSA (top left), ADHY (bottom) and HEWL (top right) for the halide series

For all selected enzymes the classical Hofmeister trend was reproduced for the halogen ions (F⁻ > Cl⁻ > Br⁻ > I⁻). As the halogen ions differ in size but retain the same charge, this causes differences in charge density and therefore to different states of hydration of the ion[252].

Fluoride is the anion in the halogenides series with the highest ratio of charge to volume, resulting in an exclusion from direct interactions with protein surfaces due to strong hydration[248]. At 1 M concentrations, fluoride possesses a stabilizing effect for all three tested enzymes. The significantly stronger stabilization of 7.2 °C for EcFSA could be the result of the higher occurrence of hydrophobic interactions in the decameric structure of EcFSA compared to the increase in T_m for monomeric HEWL by 2.4 °C and dimeric ADHY by only 0.2 °C. Since high salt concentrations favor hydrophobic interactions, the higher stabilization could be explained by hydrophobic effects[253-255]. For the dimeric ADHY and monomeric HEWL the effect is clearly less dominant. With low concentration of fluoride only ADHY experiences some destabilization, while EcFSA and HEWL are stabilized under those conditions.

Chlorides usually position in the middle of the Hofmeister series. Chlorides are known to almost not disrupt hydrogen bonds in bulk water[256-257], reflecting that it is a weakly hydrated anion. This is also the reason why chloride is generally accepted as the borderline between the kosmotropic and chaotropic part of the series.[258] Interestingly, the resulting effect can either be destabilizing or stabilizing dependent on the nature of the native structure of the enzyme. With low concentrations, chloride has a destabilizing effect on ADHY and HEWL, while for EcFSA a clear stabilization is observed. Increasing the chloride concentration up to 1 M, EcFSA is stabilized by 5.7 °C, while HEWL is stabilized by 1.6 °C. Interestingly, for ADHY a decrease of T_m by −4.1 °C was detected.

For HEWL and ADHY, bromide is a destabilizing anion, inducing under all tested conditions a reduction of stability. At 1 M concentrations of bromide, HEWL is destabilized by −2.0 °C, while ADHY is strongly destabilized by −9.4 °C. Surprisingly, EcFSA experiences a stabilizing effect in concentrations below 200 mM, which then starts to destabilize with increasing concentrations, resulting in a stabilization of 1.9 °C at 1 M concentration of bromide. These results indicate that with higher complexity of the quaternary structure, a strengthening of internal protein interactions is induced with high salt concentrations. The stronger binding of substructures leads to an increase in thermostability.

Iodide was a strong destabilizer at 1 M concentrations for all selected enzymes, showing the second strongest destabilization overall. Noteworthy is that bromide and iodide both showed stabilizing effects with low concentrations only for decameric EcFSA. At 1 M concentrations, iodide destabilizes EcFSA by −5.4 °C, ADHY by −15.3 °C and HEWL by −6.8 °C. In the halide series, iodide is the counterpart of fluoride, having the lowest charge density and state of hydration. Therefore, the highest destabilization effect in the halide series was to be expected for iodide.

Nitrate usually induces destabilizing effects comparable to bromides. The reason for that is the dissipated charge distribution in nitrates. Both nitrates and bromides have a similar charge density and state of hydration[259]. At 1 M concentrations, nitrate always showed slightly higher stabilization than bromides, while at lower concentrations the situation shifted dependent on the selected protein. At concentrations below 200 mM, only ADHY is destabilized, while EcFSA

and HEWL experience an increase of the T_m. At 1 M concentrations the resulting (de)stabilization is well comparable to the results of bromide. ADHY is destabilized by –9.1 °C, *Ec*FSA is stabilized by 1.9 °C while HEWL is destabilized by –1.6 °C.

The destabilizing effect of perchlorate is well known for lithium perchlorate, which is commonly used as a denaturant agent.[260] Only for *Ec*FSA a stabilizing effect was observed at concentrations below 200 mM. One possible explanation is again the higher occurrence of protein/protein interactions in the decameric structure of *Ec*FSA. At 1 M concentration a strong destabilizing effect was observed for all three enzymes. *Ec*FSA was destabilized by –7.3°C, while HEWL showed a reduction of T_m by –11.1°C. The highest destabilization for all three enzymes was measured for ADHY at 1 M concentration of perchlorate, reducing the T_m by a record –19.1°C. Since ADHY possesses a T_m of 48.6°C in the absence of salt additives in 50 mM TEA buffer, the decrease of T_m is so dramatic, that the protein remains barely stable at room temperature. Interestingly, at very low concentrations *Ec*FSA showed some stabilizing effect, which becomes surpassed at increasing concentrations of perchlorate. For ADHY and HEWL a destabilizing effect was detected even at low concentrations.

Comparing the selected enzymes, it is apparent that *Ec*FSA demonstrates a unique tolerance towards most conditions. The reason for that is most likely the decameric structure, possessing a T_m of 81.7°C in 50 mM TEA*HCl buffer (pH 7.5) without salt additives, which is remarkably high for a protein of mesophilic origin. Because of greater prevalence of beneficial internal protein/protein interactions, which are favored at high salt concentrations, *Ec*FSA even gets stabilized with conditions that destabilize HEWL or ADHY. Nonetheless, the general Hofmeister ordering of the anions is retained with only small inconsistencies. Results indicate that there is no universal effect of anions on protein stability, since bromide for example can either be stabilizing or destabilizing at high concentrations. Indeed, the extreme cases of strong denaturants like perchlorate or iodide always destabilize at high concentrations, but the shift of T_m is dependent on the nature of the protein. The same accounts for the stabilizing sulfate, showing a shift of T_m in the range of 6.2 – 20.1°C being positive with all selected enzymes at 1 M concentrations of sulfate.

For low salt concentrations a different behavior was expected for HEWL and ADHY due to the opposite surface net charge. Results indicate however, that this seems to have minimal effect on the ordering of the anions regarding their ability to (de)stabilize the protein structure. More relevant seems the particular stabilization through quaternary structure elements. By increasing protein/protein interactions, the loss of hydrating water molecules is compensated by the protein. Molecular dynamics simulations indicate that with low salt concentrations favorable electrostatic interactions between positively charged residues or polarizable areas and weakly hydrated anions dominate the resulting (de)stabilization[248, 261]. The interactions can be repulsive or attractive dependent on the charge density of the anion and the surface nature of the protein. The positively charged side chains of the amino acids arginine, lysine and histidine are sites for possible electrostatic interactions for anions. Furthermore, the amide unit of the protein backbone with its α-carbon is a favorable interaction site for weakly hydrated anions, while small, strongly hydrated anions, like fluoride, are almost fully excluded from such kind of interactions.[248]

The most likely explanation in summary is that with low salt concentrations the anions interact with charged side chains of surface amino acid residues. With increasing concentrations, the weakly hydrated anions interact with the protein. By interacting with the protein, the energy for hydrating the hydrophobic core of the protein gets lowered, which results in a destabilization. At high concentrations the well hydrated anions like sulfate or acetate compete for water molecules with the protein, resulting in an increase of the energy required for hydrating the hydrophobic core of the protein and resulting in a stabilizing effect[243].

4.2.2 Cationic Hofmeister series

The effects of cations are generally less prominent than the effects of anions on the thermostability of proteins. Nevertheless, it was important to quantify the effects of cations separately. Chloride as counterion is known to have minimal effect on protein structures, making it the perfect standard counterion to evaluate the effects of the cation. The results of the melting point measurement with varying concentrations of chloride salts, namely sodium-, potassium-, lithium-, caesium-, ammonium-, magnesium-, calcium and guanidinium chloride, are presented in Table 4–6. The selection of cations is based on Hofmeister selection used in his experiment with the exception of the guanidinium ion, which was used because of the strong destabilizing effect.

Table 4 Resulting ΔT_m measured by nanoDSF in TEA*HCl buffer (50mM, pH 7.5) for *Ec*FSA. The colour code corresponds to the total value of ΔT_m ($\Delta T_m > 10°C$, $\Delta T_m > 5°C$, $\Delta T_m > 0°C$, $\Delta T_m > -5°C$, $\Delta T_m > -10°C$, $\Delta T_m < -10°C$).

Conc.	ΔT_m [°C]							
[mM]	GndCl	NH$_4$Cl	NaCl	KCl	LiCl	MgCl$_2$	CaCl$_2$	CsCl
1000	-3,3	0,5	5,7	5,6	0,4	-8,4	-12,7	4
500	0,5	1,8	4,6	4,5	2,2	-3	-6,8	4,5
250	1,6	1,9	3,8	3,5	2,3	-0,2	-2,9	2,5
125	1,8	1,6	2,9	2,8	1,6	1,1	-0,8	1,9
62,5	1,3	1	2,1	1,9	1	1,4	0,2	1
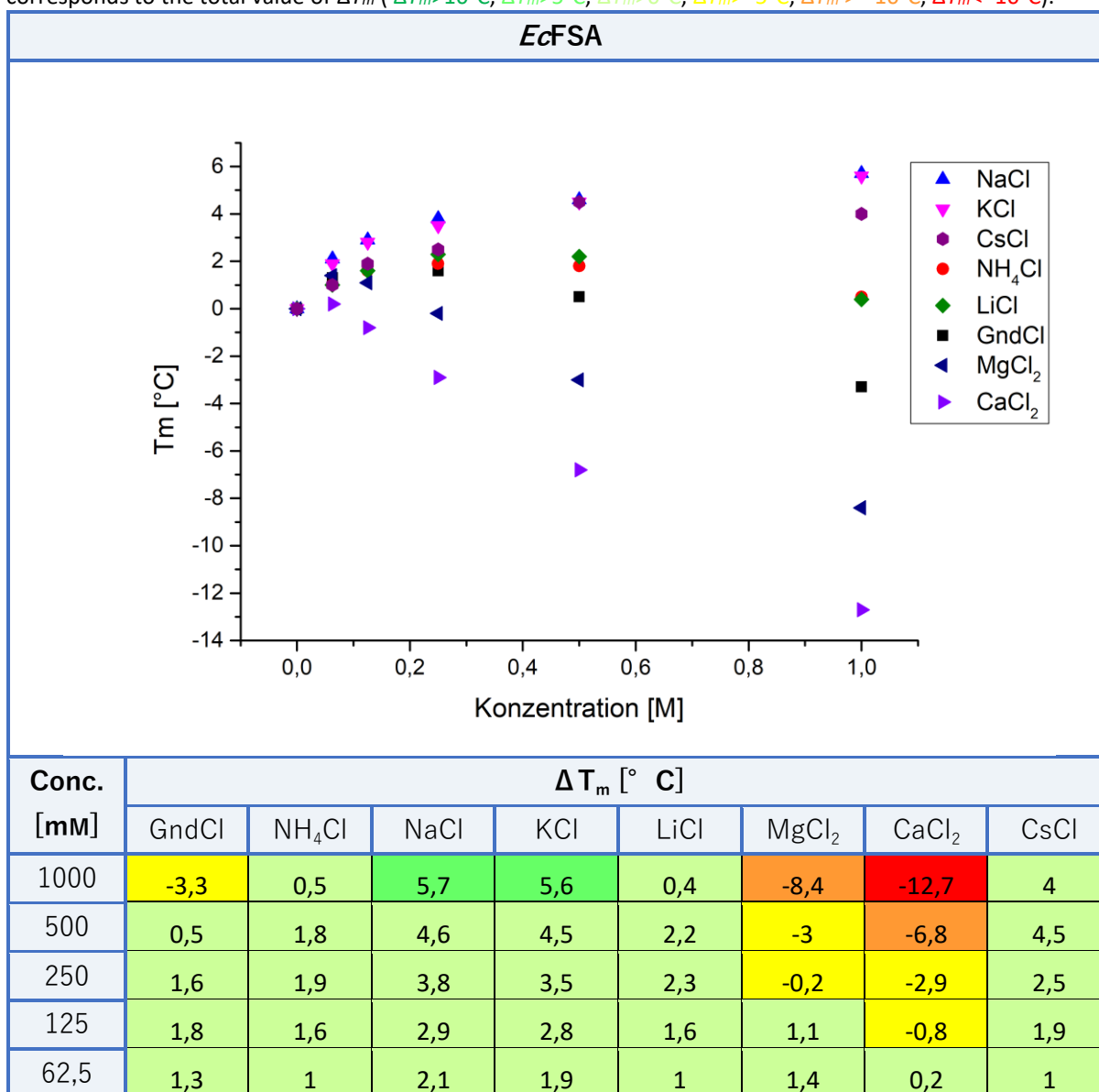
Table 5 Resulting ΔT_m measured by nanoDSF in TEA*HCl buffer (50mM, pH 7.5) for ADHY. The colour code corresponds to the total value of ΔT_m ($\Delta T_m{>}10°C$, $\Delta T_m{>}5°C$, $\Delta T_m{>}0°C$, $\Delta T_m{>}{-}5°C$, $\Delta T_m > -10°C$, $\Delta T_m{<}{-}10°C$).

Salt conc. [mM]	ΔT_m [°C]							
	GndCl	NH$_4$Cl	NaCl	KCl	LiCl	MgCl$_2$	CaCl$_2$	CsCl
1000	den.	-7,9	-4,5	-5	-5,8	-10,2	den.	-5,5
500	-10,4	-5,9	-3,3	-3,4	-4,2	-8,3	-12,3	-3,8
250	-7,6	-4,4	-2,5	-2,7	-3,2	-7	-10,1	-3
125	-5,4	-3	-1,6	-2	-1,8	-5,9	-7,8	-2
62,5	-3	-2	-0,9	-0,9	-1,3	-4,6	-6,5	-1,1

Table 6 Resulting ΔT_m measured by nanoDSF in TEA*HCl buffer (50mм, pH 7.5) for HEWL. The colour code corresponds to the total value of ΔT_m (ΔT_m>10°C, ΔT_m>5°C, ΔT_m>0°C, ΔT_m>−5°C, ΔT_m > −10°C, ΔT_m<−10°C).

HEWL

Salt conc. [mм]	ΔT_m [°C]							
	GndCl	NH$_4$Cl	NaCl	KCl	LiCl	MgCl$_2$	CaCl$_2$	CsCl
1000	-8.8	1	1.1	0.6	0.1	1.5	-3.4	-0.1
500	-4.5	0.1	-0.3	-0.5	-0.3	1.4	-1.5	-1.2
250	-2.5	-0.2	-0.7	-0.8	-0.3	1.6	-0.3	-1.2
125	-1.4	-0.2	-0.7	-0.7	-0.3	1.6	0.4	-1.1
62,5	-0.8	-0.3	-0.6	-0.6	-0.4	1.4	0.7	-0.9
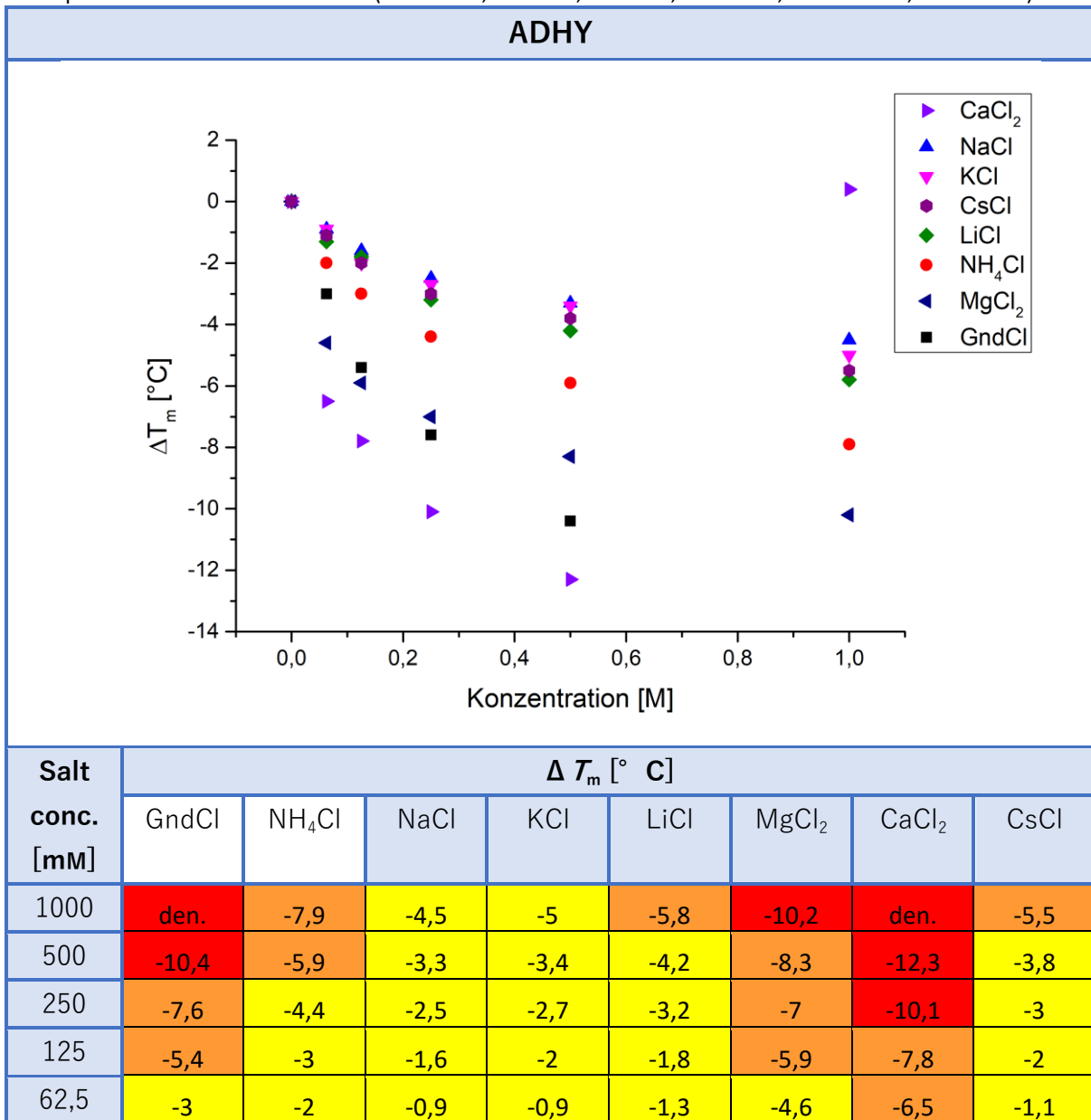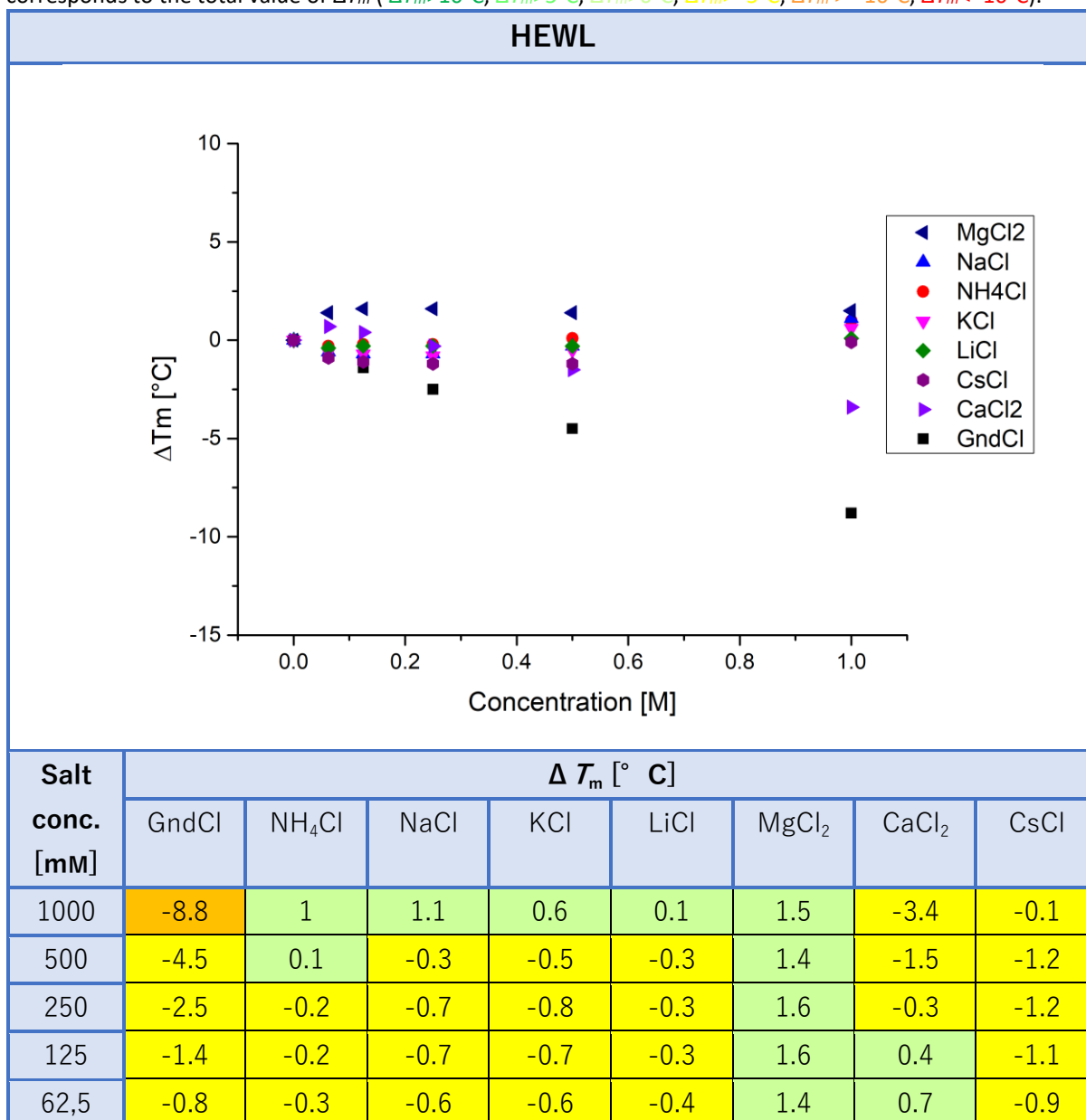
The interactions of cations and proteins are usually less prominent compared to effects of anions[261]. Interestingly, the simple categorization that strongly hydrated ions stabilize the native fold, while weakly hydrated ions favor unfolding, failed for cations. The weakly hydrated cations like ammonium indeed showed stabilizing effects whereas strongly hydrated cations such as magnesium and calcium showed destabilizing effects. It is important to not only to consider the binding of ions and water molecules, but to consider direct interactions of ions with the protein surface[261].

$$Gnd^+ > Ca^{2+} > Cs^+ \quad > Li^+ > \quad K^+ > NH_4^+ > Na^+ > Mg^{2+} \qquad \text{HEWL}$$

$$Ca^{2+} > Gnd^+ > Mg^{2+} > NH_4^+ \quad > Li^+ > \quad Cs^+ > K^+ > Na^+ \qquad \text{ADHY}$$

$$Ca^{2+} > Mg^{2+} > Gnd^+ \quad > Li^+ > \quad NH_4^+ > Cs^+ > K^+ > Na^+ \qquad \textit{Ec}\text{FSA}$$

$$Gnd^+ > Ca^{2+} > Mg^{2+} \quad > Li^+ > \quad Na^+ > K^+ > NH_4^+ \qquad \text{Hofmeister}^{[161\text{-}162]}$$

The general Hofmeister trend was reproduced in that sense, that divalent cations and the guanidium ion have the strongest destabilizing effect on all tested enzymes, while the monovalent cations possess a less dramatic effect. The cations in general showed less strong (de)stabilizing effect as compared to the anions. The only exception is found for the magnesium cation inducing a stabilizing effect only for HEWL.

Guanidinium chloride GndCl is a well-known denaturing agent, usually applied with significantly higher concentrations(4 M) than the highest measured concentration in this study [262-263]. Even at 1 M concentrations, the ADHY sample was already denatured at room temperature before the measurement. For *Ec*FSA the third strongest destabilization was detected for GndCl, destabilizing the enzyme by −3.3 °C. HEWL experienced the strongest destabilization by GndCl, lowering the T_m by −8.8 °C. GndCl performed as one of the best destabilizers for all three tested enzymes, in strong competition with both the tested divalent anions Mg^{2+} and Ca^{2+}. At low concentrations, guanidinium chloride already induced destabilizing effects on all three enzymes. One possible explanation is the breaking of salt bridges of the protein structure by interacting with accessible amino acid residues of the protein structure. With increasing concentrations, a strong destabilizing effect was detected as expected of a commonly utilized denaturant agent[262].

The calcium cation also induced such strong destabilizing effects on ADHY that at 1 M concentrations the sample already denatured at room temperature. For *Ec*FSA a shift of T_m by −12.7 °C was detected, demonstrating the highest negative value for *Ec*FSA, while HEWL was destabilized by −3.4 °C under the same conditions.

The magnesium cation showed unexpected behavior for HEWL, where a stabilizing effect was detected with all tested concentrations, while for ADHY and *Ec*FSA a strong destabilizing effect was detected at 1 M concentration, lowering the T_m by −8.4 °C for *Ec*FSA and −10.2 °C for ADHY. The reason for that effect is most likely found in the monomeric form of HEWL possessing significantly less hydrophobic parts for interactions with the cation than the dimeric ADHY and decameric *Ec*FSA.

For all monovalent cations, only a minor impact on thermostability could be detected, where the highest (de)stabilization was detected for ADHY with a maximum ΔT_m of −7.9° C for ammonium. In general, ADHY was destabilized under all tested conditions, and in higher degrees compared to both other tested enzymes. This result is interesting, since *Ec*FSA was stabilized by all monovalent chloride salts under any tested conditions. Meanwhile, HEWL was

only destabilized significantly by guanidine chloride and was even stabilized by the divalent cation magnesium ion and, at low concentrations only, by calcium ion.

Comparing the monovalent cations Na⁺, K⁺ and Cs⁺, all showed similar behavior with minor changes in total (de)stabilization. Figure 8 shows the trends obtained for each tested enzyme with the above-mentioned monovalent cations.
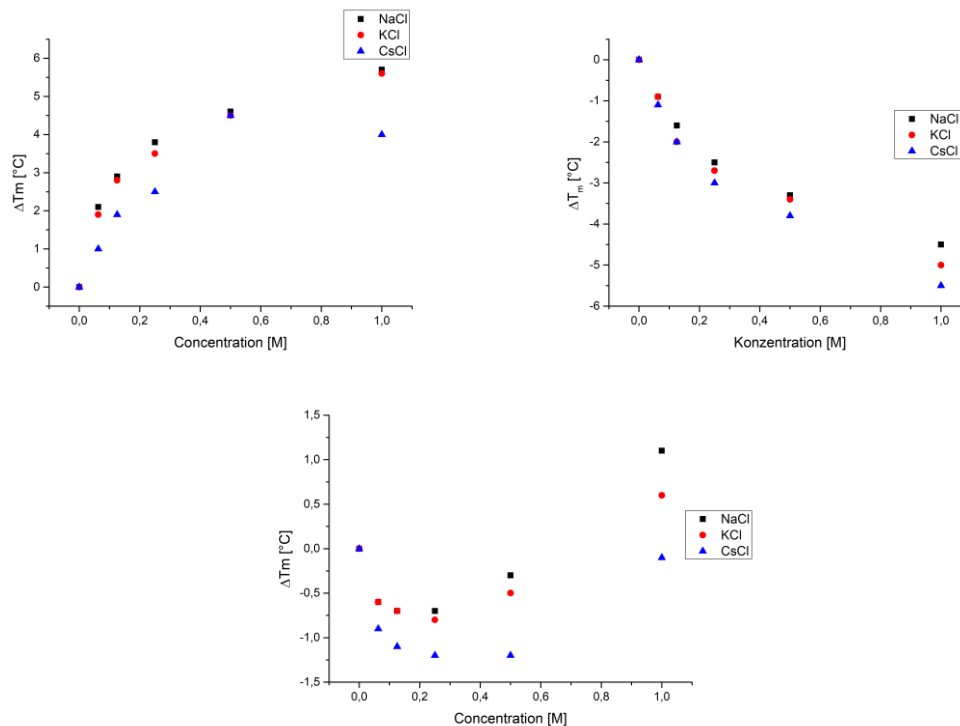
Figure 8 Comparison of the resulting (de)stabilization of *Ec*FSA (top left), ADHY (top right) and HEWL (bottom) with sodium chloride, potassium chloride and cesium chloride as additive with 50 mM TEA*HCl buffer (pH 7.5) based on the melting point of the sample.

It is important to notice that the effects of cations are determined in the presence of chloride cations, which can have a stronger effect on thermostability than the cations itself. For HEWL no clear tendency was detected since at low concentrations a minor destabilization was detected, which turned into a stabilizing effect at 1 M concentrations. For *Ec*FSA a clearer tendency was observed. All tested concentrations of the monovalent cations Na⁺, K⁺ and Cs⁺ increased the T_m of *Ec*FSA. Sodium chloride showed the highest stabilization at 1 M concentrations, increasing the T_m by 5.7 °C, followed by potassium chloride increasing the T_m by 5.6 °C, and cesium chloride by 4.0 °C.

Interestingly, for ADHY a different trend was obtained with all tested conditions, which lowered the T_m under all tested conditions. Sodium chloride lowered the T_m by −4.5 °C at 1 M concentration, while potassium chloride lowered the T_m by −5.0 °C and cesium chloride by −5.5 °C.

A plausible explanation for the strong destabilization could be found in the dimeric state of ADHY. The addition of salts can lead to protein/ion interactions which can possibly lead to a separation of the dimer into monomeric form which most likely is rather unstable and the

presence of cations and anions may favor the unfolding process. For *Ec*FSA those interactions can be possibly compensated by the strong stabilization resulting from the decameric state of the enzyme. For monomeric HEWL, hydrophobic interactions are less favorable due to mostly polar surface that hinder ion/protein interactions. Additionally, the positive surface potential should result in attraction of anions.

Lithium chloride was found in the center of the resulting series for all tested proteins. For the enzyme HEWL a maximum destabilization of −0.3 °C was detected, showing minimal effects on the protein structure. For ADHY a clear dependency on lithium chloride concentration was detected, rapidly decreasing the T_m with increasing concentrations. At 1 M concentrations, a destabilization of −5.8 °C was determined.

For *Ec*FSA, the opposite trend was obtained. With increasing lithium chloride concentration, the protein experienced a stabilization up to 500 mM, which is lost at 1 M concentration. The most plausible explanation for the different trends obtained for all enzymes is the favorable occurrence of hydrophobic protein/ion interactions. The hydrophobic interactions yielding the dimeric state of ADHY most likely get disrupted by the presence of lithium chloride favoring the monomeric form, which is unstable and therefore enters into the unfolding process.

Interestingly, the same trend as for lithium chloride was obtained for ammonium chloride. Up to 1 M concentrations a minor stabilization is found for HEWL of up to 1.0 °C. For *Ec*FSA a stabilization effect of 1.9 °C was found at 250 mM concentrations, which is reduced at 1 M concentrations to 0.5 °C. A negative trend with increasing concentrations was found for ADHY, constantly decreasing the T_m of the sample with increasing concentrations of ammonium chloride.

4.2.3 Influence of glucose, trehalose, sorbitol and saccharose on thermostability of *Ec*FSA

The impact of sugars on protein stability has been a subject of considerable research interest in the field of biochemistry. Sugars, including monosaccharides and disaccharides, have been shown to exert diverse effects on the stability of proteins under various conditions.

Proteins, as indispensable macromolecules in living entities, demand stability for the preservation of biological functionality. Sugars intricately participate in modulating protein stability by engaging with the protein surface, influencing water structure, and interacting with other solute molecules. This comprehension proves pivotal for diverse applications, spanning biotechnology, pharmaceuticals, and the food industry.

In this work, the effect on thermostability on *Ec*FSA of glucose, trehalose, sorbitol and saccharose was measured with nanoDSF. Figure 9 presents the results of the nanoDSF measurement.
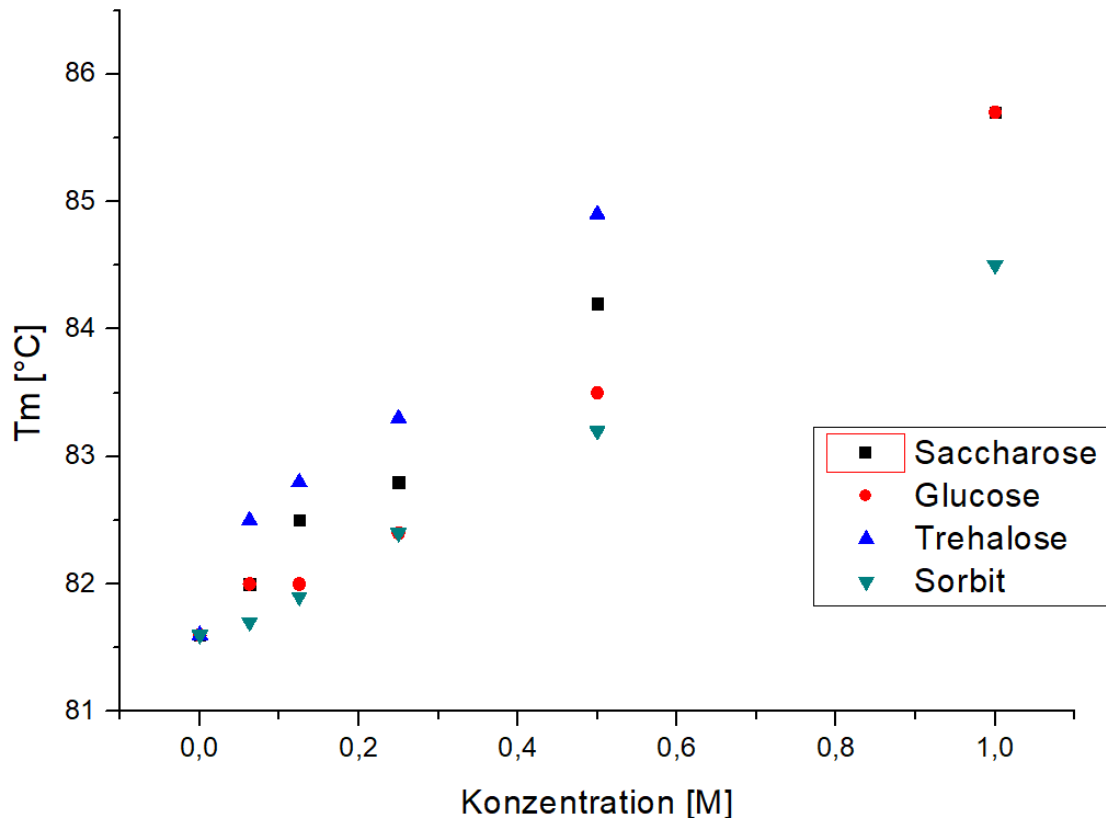
Figure 9 Thermostability measurement of *Ec*FSA with varying concentrations of glucose, trehalose, sorbitol and saccharose in 50 mᴍ TEA*HCl buffer at pH 7.5.

The results of the measurement demonstrate a clear stabilizing effect for all samples. One aspect of sugar-protein interactions involves the ability of sugars to act as osmolytes, contributing to the stabilization of protein structures under osmotic stress or environmental fluctuations. Additionally, sugars can influence protein stability by altering the hydration shell around proteins, leading to changes in conformational dynamics and unfolding pathways.

Sugars are hydrophilic molecules with hydroxyl groups able to form hydrogen bonds. By interacting with the protein by hydrogen bonding, sugars can influence the surrounding water of the hydrate shell of the protein. This altered water structure may provide a more favorable environment for maintaining protein hydration and stability.

Furthermore, sugar molecules can replace water molecules surrounding the protein, creating a more ordered and less dynamic hydration shell. This replacement can reduce the entropy loss associated with protein folding and contribute to the overall stability of the protein. By forming a protective layer around the protein, sugars mitigate potential disruptive interactions, thereby stabilizing the native conformation of the protein.

Sugars can additionally influence the mechanism of unfolding by increasing the necessary energy for the unfolding and thereby making the native folded state more favorable.

It is important to notice, that the mechanism of sugar-protein interaction is highly related to the protein's properties, the sugar and the environmental conditions. The interplay between these factors contributes to the overall (de)stabilizing effect of the sugar on proteins, making them interesting parameters for biotechnological and pharmaceutical applications.

We were interested if hydroxyl groups on small molecules also show a stabilizing effect on *Ec*FSA. Figure 10 shows the results of the measurement with glycerol, 1,2-propylenegylcol, 1,3-propyleneglycol and ethylene glycol.

Figure 10 Thermostability measurement of *Ec*FSA with varying concentrations of glycerol, 1,2-propyleneglycol, 1,3-propyleneglycol and ethylene glycol in 50 mM TEA*HCl buffer at pH 7.5.

Interestingly, a stabilizing effect with higher concentrations was only observed for glycerol. 1,2-propylenegylcol, 1,3-propylenglycol and ethyleneglycol all presented destabilizing effects in high concentrations, while1,3-propylenglycol and ethyleneglycol showed a stabilizing effect in up to 15% V/V.

Glycerol is well known for its stabilizing effect, utilized for freeze-thaw cycles with cell cultures. Also, glycerol is commonly used for liquid protein solutions for increasing the shelf life of the protein preparation.

5 Experimental

5.1 Methods and Materials

All chemicals were bought from Sigma Aldrich, if not mentioned otherwise. All enzymatic reactions were carried out at room temperature with a pH of 7.5. Distillation with the tag "removed in vacuo" refers to usage of a rotary evaporator with a membrane pump and a water bath of 40 °C.
All tools and media used for working with DNA and microorganisms were autoclaved before use. For biomolecular experiments standard procedures were used.

5.2 Analytic

5.2.1 Thin layer chromatography

For analytic thin layer chromatography for reaction control and screening of enzyme reactions, silica gel alumina plates from *Macherey-Nagel* were utilized. Detection of the corresponding product and educt spots was performed after staining with anisaldehyde staining solution.

Anisaldehyde staining solution:
- 200 ml EtOH
- 7.5 ml H_2SO_4 (conc.)
- 2.2 ml AcOH (conc.)
- 5.5 ml anisaldehyde

The TLC plates were dipped into the staining solution and dried via heat gun.

5.2.2 HPLC-MS

The HPLC measurement was performed on a HPLC-MS from *Shimadzu* with a SPD-M20A UV-diaodes array detector. The measurements were analyzed with a wavelength of λ=215 nm with a flow rate of 0.4 ml/min. During the measurements, XbridgeTM C18 column (pore size 3.5 µm, length: 150 mm, diameter: 3.0mm) bought from *Waters* (Milford, USA) was utilized and the measurement was performed at 40°C.

Following solvent systems were used during the measurement:

- Solvent A: dest. Water + 0.1% formic acid
- Solvent B: Acetonitrile + 0.1% formic acid

Samples for HPLC-MS were analyzed with the following method:

0 min	→	30 min	10% B	→	90% B
30 min	→	31 min	90% B	→	95% B
31 min	→	34 min	95% B	→	
34 min	→	36 min	95% B	→	10% B
36 min	→	40 min	10% B	→	

5.2.3 Sample preparation for HPLC MS

For detection of aldehydes and ketones (≤100 mM, 10 µl sample volume) via HPLC-MS, compounds were derivatized with the following derivatization solution (50 µl): 210 mg O-Benzylhydroxylamine hydrochloride, 6.6 ml pyridine, 3.0 ml MeOH, 400 µl dest. H_2O. The mixture was kept at rt for 30 min before diluting to 1 ml total volume with MeOH, followed by centrifugation and finally the sample was transferred to HPLC-vials for the measurement.

5.2.4 Protein melting temperature determination

The Prometheus NT.48 instrument (NanoTemper Technologies) was used to determine the melting temperatures of protein samples. The capillaries were filled with 10 µL sample and placed on the sample holder. A temperature gradient of 1 °C min^{-1} from 25 to 100 °C was applied and the intrinsic protein fluorescence at 330 and 350 nm was recorded.

5.3 Materials

If not mentioned otherwise, all chemicals were purchased from different manufacturers (TCI, Acros Organics, Sigma-Aldrich-Merck, Grüssing, Iris Biotech, Maybridge) and were utilized without further purification.

5.3.1 Mutagenesis Project:

Enzymes

Enzyme	Manufacturer
Lysozyme	Ovoblast Eierprodukte
Alcohol dehydrogenase	Sigma-Aldrich
DNase 1	New England Biolabs (NEB)
Q5 High Fidelity DNA Polymerase	NEB
KLD enzyme mix	NEB

Composition of relevant buffers and solutions

All buffers were -if necessary- adjusted with 1 M NaOH or 1 M HCl to obtain the necessary pH.

Buffer	Components
Washing buffer pH 7.5	9.4 ml 1 M monobasic (KH_2PO_4), 40.6 ml 1 M dibasic(K_2HPO_4) →filled up with dest. H_2O to 1 l
SDS-PAGE sample buffer (4x) pH 6.8	2 ml 1 M TRIS pH 6.8, 0.8 g SDS, 5 ml Glycerol, 1 ml 0.5 M EDTA, Bromphenolblue (2 % in ethanol)
Standard lysozyme buffer	50 mg lysozyme, 2 µl DNase I →filled up with 50 ml washing buffer pH 7.5
SDS-PAGE gel running buffer	25 mM (3.1 g) TRIS, 192 mM (14,41 g) glycine, 0.1 % (1 g) SDS →filled up with dest. H_2O to 1 l
Ni-NTA Wash buffer	20 mM (2.76 g) NaH_2PO_4, 0.05 % NaN_3 →filled up with dest. H_2O to 1 l, pH adjusted to 7.4
Ni-NTA Wash buffer with NaCl	20 mM (2.76 g) NaH_2PO_4, 0.5 M (29.29 g) NaCl, 0.05 % NaN_3 →filled up with dest. H_2O to 1 l, pH adjusted to 7.4 →If needed, 30 mM imidazole added
Ni-NTA Elution buffer	20 mM (2.76 g) NaH_2PO_4, 0.5 M (29.29 g) NaCl, 50 mM (14.61 g) EDTA, 0.05 % NaN_3 →filled up with dest. H_2O to 1 l, pH adjusted to 7.5

TEA buffer, 1M, pH 7.5	132.1 ml triethanolamine →fill up to 1 l with dest. H_2O, pH adjusted to 7.5
10x Q5 DNA Polymerase buffer	Obtained from NEB
2x KLD Reaction buffer	Obtained from NEB
0.5x TBE buffer	5.4 g TRIS and 2.25g boronic acid →filled with 900 ml dest. H_2O →2 ml of 0.5 M Na_2EDTA (pH 8.0) added →filled up to 1 l with dest. H_2O
6x Purple Loading Dye	Obtained from NEB
Blue Prestained Protein Standard Broad Range (11-250 kDa)	Obtained from NEB
1 KB Plus DNA Ladder	Obtained from NEB

Media

Medium	Compounds
LB	10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl →filled up to 800 ml dest. H_2O →the pH was adjusted to 7.5 with NaOH. →filled up to 1 l with dest. H_2O and autoclaved.
AIM	10 g peptone, 5 g yeast extract, 2.68 g NH_4Cl, 0.71 g Na_2SO_4, 5 g glycerol (85%), 0.5 g glucose, 2 g lactose →filled up with 950 ml dest. H_2O and autoclaved. →1 ml $MgSO_4$ (1 M), 40 ml autoclaved 1 M K_2HPO_4 and 10 ml autoclaved 1 M KH_2PO_4 were added to cooled media (room temperature)
SOC	Obtained from NEB

Microorganisms:

Strain	Application
E. coli BL21	Expression of LbDERA-mutants
E. coli DH5α (NEB 5α Competent High Efficiency E. coli)	Amplification of plasmid DNA/cloning products

Kits:

Kit	Manufacturer
GenElute HP Plasmid Miniprep Kit	Sigma-Aldrich
KLD Enzyme Mix Kit	NEB

Primers:

Primer	Type	Usage	Sequence
DERA18-20del_f	fwd	Deletion AA positions 18-20	5´-AATGACGACGACACCGACGAGAAAGTGATCGC-3´
DERA18-20del_r	rev	Deletion AA positions 18-20	5´-CAGGTCCATCAATTTCAGTGCACGCAGGC-3´
DERA76-KYC_f	fwd	Randomization of position 76	5´-GTATCGCTACGGTAACCAACKYCCCACACGGTAACG-3´
DERA F76G_f	fwd	Mutation F76G	5´-GTATCGCTACGGTAACCAACGGCCCACACGGTAACG-3´
DERA76-KYC_r	rev	Randomization of position 76	5´-GGATTTCCGGGGTGCCCTGCTCTTTCAGAGTTTTGCG-3´
DERA18-20ins_f	fwd	Randomized insertion in position 20	5´-CCKYGAATGACGACGACACCGACGAGAAAGTGATCGC-3´
DERAL18/20G_f	fwd	Insertion of glycine in position 20	5´-CCGGCAATGACGACGACACCGACGAGAAAGTGATCGC-3´
DERA18-20ins_r	rev	Randomized	5´-TGGHCAGGTCCATCAATTTCAGTGCACGCAGGC-3´

		insertion in position 18	
DERAL18/ 20G_r	rev	Insertion of glycine in position 18	5´-TGCCCAGGTCCATCAATTTCAGTGCACGCAGGC-3´
DERAT18 AL20G_r	rev	Insertion of alanine in position 18	5´-TGGCCAGGTCCATCAATTTCAGTGCACGCAGGC-3´
DERAT18 GL20A_f	fwd	Insertion of alanine in position 20	5´-CCGCGAATGACGACGACACCGACGAGAAAGTGATCGC-3´
DERAT18 GL20V_f	fwd	Insertion of valine in position 20	5´-CCGTGAATGACGACGACACCGACGAGAAAGTGATCGC-3´
T7-modfwd	fwd	Sequencing of genetic constructs on vector pET21a	5´-CCCGCAAATTAATACGACTCAC-3´

5.4 Methods

Mutagenesis: Polymerase chain reaction (PCR)

The primer design was executed according to the recommendations of the "Q5 Site-Directed Mutagenesis Kit" by NEB.
For nucleotide substitutions, forward primers with at least ten overhang nucleotides at the 3´-end counting from the mutation site were designed, placing the mismatch codon in the middle of the primer.
The reverse primers were tailored to start at the 5´-end of the forward primer. For nucleotide deletions, primers complementary to the plasmid sequence were created, ending adjacent to the 5´-end of the fragment to be deleted. For nucleotide insertions, the primers had a similar plasmid complementary sequence, but carried a part of the split insertion sequence at their respective 5´-end. For the reverse primers, this were four nucleotides, while the for forward primers, the overhang was five nucleotides long, amounting to the 9 nucleotides long insertion sequence.

PCR mix composition

Component	Volume	Concentration in reaction mix
Dest. H₂O	14.25 µL	-
5x Q5 Reaction buffer (NEB)	5 µL	1x
dNTPs (2 mM)	2.5 µL	0.1 mM
3´-primer (10 µM)	1.25 µL	0.5 µM
5´-primer (10 µM)	1.25 µL	0.5 µM
Template DNA	0.5 µL	-
Q5 High Fidelity DNA Polymerase (NEB, 2000 units/ml)	0.25 µL	0.5 units
Reaction volume	25 µL	

The PCRs were performed according to the time program shown in Table. Due to the melting temperatures of the primers, it was decided to perform a two-step PCR, discarding the primer annealing slot in the time program.

PCR program

Phase	Temperature [°C]	Duration [s]	Cycles
Initial denaturation	98	30	x1
Denaturation	98	10	X25
Annealing and Elongation	72	150 (20s/kB)	
Final Elongation	72	5	X1
Hold	15	-	-

The success of the PCR was verified via agarose gelelectrophoresis.

Agarose gelelectrophoresis:

For agarose gel electrophoresis, gels were prepared via mixing approximately 50 ml of 0.8% (w/v) agarose solution in 0.5x TBE buffer with 10 µl ethidium bromide and pouring this mix into the gel chamber after preheating it to 100 °C first. After solidifying, the gel was covered with 0.5x TBE buffer. The DNA marker and samples were loaded onto the gel's wells. The samples had to be prestained with 6x Purple Loading Dye (NEB).

After loading, the gel was run at 100 V for 30-40 min. Then, a picture of it was taken in the UV light.

The gel was discarded.

KLD reactions for PCR product ligation

The obtained PCR products had to be ligated to a circular plasmid prior to transformation. This was achieved with the KLD (Kinase-Ligase-DpnI) mix from NEB. While the ligase performed a blunt-end ligation of the PCR product, the original template was digested by the DpnI enzyme, preserving only the recombinant PCR products. The reaction was performed with the following protocol:

- 2.5 µl 2x KLD reaction buffer (NEB) were mixed with 1.5 µl dest. H2O.
- 0.5 µl PCR mix were added.
- 0.5 µl KLD enzyme mix (NEB) were added.
- The reaction was left at room temperature for 30 min. For transformation, the whole reaction mix had to be used.

Transformation of ligated PCR products/ plasmids into *E. coli* Heat shock protocol for DH5α

The freshly ligated PCR products were transformed into the DH5α cell line for DNA amplification (to be followed by plasmid preparation and sequencing). It was performed with NEB 5α Competent High Efficiency *E. coli* according to the following protocol:

1. Competent cells (50 µl aliquots) were thawed on ice for 10 min.
2. The KLD reaction mix (5 µl) was added to the cell suspension. The cells were then incubated for 30 min on ice.
3. The cell aliquots were placed into a prewarmed heating block for 30 s at 42 °C and placed back onto ice immediately after.
4. The cells were chilled on ice for 5 min.
5. 950 µl of SOC medium was added to the cells. This suspension was incubated at 37°C and 900 rpm for 1 hour.
6. Variable volumes were plated on selective LB-agar plates. The plates were incubated overnight at 37 °C.

Electroporation protocol for BL21

The electroporation protocol as used to transform amplified recombinant plasmid DNA (carrying finalized verified mutants) into the BL21 cell line. The following protocol was used:

1. 50 µl electrocompetent *E. coli* BL21 cells were incubated with 1 µl of plasmid DNA for 5 minutes on ice.
2. Electroporation cuvettes were precooled on ice.
3. The cell-DNA-mix was transferred into the electroporation cuvettes.
4. The cuvette was quickly wiped and placed into the electroporator.
5. The cells were pulsed one time at 1.8 kV for 4.5-6 ms.
6. 950 µl SOC medium was added immediately after pulsing. The resulting suspension was put into a 2 mL microcentrifuge tube.
7. The suspension was incubated at 37 °C and 900 rpm for 1 hour.
8. Variable volumes were plated on selective LB-agar plates. The plates were incubated overnight at 37 °C.

Cultivation of overnight cultures for plasmid preparation

All overnight cultures were inoculated as 1.6 ml of LB medium and supplemented with antibiotics if necessary. Cultures were grown overnight (18 h) at 37 °C and 900 rpm.

Cryostocks

To preserve the plasmids made during this work, cryostocks of those were made. For this purpose, 800 µl of overnight cultures were mixed with 200 µl glycerol (resulting 20% v/v glycerol stocks) in a special tube and frozen at −80 °C.

Plasmid preparation and sequencing

Plasmid preparation was performed with the GenElute HP Plasmid Miniprep Kit (Sigma-Aldrich) according to the manufacturer's manual.

For sequencing, 10 µl of the isolated plasmid elution were mixed with 4 µl of a 5 µM primer stock solution and sent to GATC Biotech AG for sequencing. The sequences were obtained in the ".fasta" and ".ab1" formats and evaluated with SnapGene.

Cell cultivation for protein overexpression

The expression cultures were grown in AIM in 2 l flasks. 500 mL of AIM were taken per flask. 500 µl of ampicillin (1:1000) and 500 µl of an antifoam additive were added. The cultures were inoculated by taking some cell material from the master plates via a pipette tip and leaving the tip in the culture. The cultures were first incubated at 37 °C for 5-6 hours and afterwards at 30 °C for 18 hours under constant shaking at 220 rpm.

After a 24-hour cultivation the cells were harvested via centrifugation at 7000xg for 5 min at room temperature. The pellets were resuspended in 5 ml of utilized buffer (either 1:10 Wash buffer for lyophilization, or Ni-NTA Wash buffer for affinity chromatography), unified if needed and frozen at –20 °C prior to lysis.

Cell lysis

For cell lysis, the frozen cell suspension was thawed in a water bath at 37°C. 1-2 mg/ml lysozyme and 2 µl/ml DNase I solution were added to the suspension. The suspension remained in the water bath for 1 hour under constant stirring. The lysozyme had to be solved before addition to the cell suspension to avoid clumping.

Protein purification

Protein purification was performed using the "HisTrap FF crude" system from GE Healthcare. The purification steps can be described as followed (1 CV – 1 column volume – 5 ml):

1. Addition of 0.5 M NaCl and 300 mM imidazole to the crude cell extracts in Ni-NTA Wash buffer
2. Equilibration of the HisTrap column with 10 column volumes (CV) Ni-NTA Wash buffer with NaCl
3. Loading of the crude cell extracts onto the column; preservation of the flowthrough
4. Washing of the column with 10 CV of Ni-NTA Wash buffer with NaCl and imidazole
5. Washing of the column with 10 CV of Ni-NTA Wash buffer with NaCl
6. Elution of the protein with 5 CV of Ni-NTA Elution buffer (50 mM EDTA)
7. Washing with 10 CV of Ni-NTA Wash buffer
8. Column regeneration with 1 CV of 100 mM $NiSO_4$-solution
9. Washing with 5 CV Ni-NTA Wash buffer with NaCl

Protein desalting: Size exclusion chromatography (SEC)

Prior to lyophilization and storage the eluted proteins had to be desalted. For this purpose, size exclusion chromatography (SEC) was used. The CentriPure P100 columns from *emp Biotech* were utilized for that in accordance with the protocol below:

1. The column was emptied from the storage buffer
2. The column was equilibrated with 10 CV (1 CV – 1 column volume – 10 ml) of 5 mM TAE buffer.
3. 1 CV of the protein sample was loaded onto the column.
4. The elution was performed with 14 ml of 5 mM TAE buffer. The first and last 2 ml of the elution flowthrough were discarded. The other 10 ml were collected as the desalted elution fraction.

5. The column was washed with 10 CV of Ni-NTA Wash buffer with NaCl and left with some of it on top of the gel compartment for storage

Lyophilization

For long-term storage, the purified and desalted proteins were lyophilized. For that, the desalted elution fractions were first frozen with liquid nitrogen under constant rotation and then connected to a lyophilizer and pressurized to vacuum to withdraw the liquid.
The lyophilized proteins were stored at −20°C.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

To analyze the lysis success, SDS-PAGE was utilized. For this purpose, 10%-acrylamide/bisacrylamide gels were prepared.

Component composition for two 10%-acrylamide/bisacrylamide gels

Component	Quanitity
760 mM Tris pH 7.4	1 mL
1 M Serine pH 7.4	1 mL
1 M Glycine pH 7.4	1 mL
Asparagine	132 mg
ddH₂O	3 mL
Acrylamide 30 %	2.75 mL
Bis-acrylamide 2 %	1.16 mL
10x APS 10%	60 µL
TEMED	10 µL

The components were stirred for 1 min and the resulting solution was poured into the gel chambers. The gel combs were inserted and the gels were given approximately 1 hour to polymerize at room temperature. If the gels needed to be stored, they were wrapped in wet tissues and stored at 4 °C.
The electrophoresis samples for the electrophoresis were prepared with 2x SDS-PAGE sample buffer. 15 µl of the protein sample were mixed with 15 µl of the 2x SDS-PAGE sample buffer, dithiothreitol (DTT) was added to a final concentration of 0.1 M from a 2 M stock (1.5 µl). Those samples were incubated at 95 °C for 10 min to ensure protein denaturation. Afterwards, 15 µl of the prepared samples were loaded onto the gels.
Gel electrophoresis was performed with the Mini Protean II (Bio-Rad) system at 200 V until the bromophenol band would run out of the gel (approximately 45 min). 5 µl of the Blue Prestained Protein Standard, Broad Range (NEB) were used as a protein marker.

The protein bands were stained with a 0.1% solution of Coomassie Brilliant Blue. The gels with the staining solution were heated for 30 s in the microwave, then left to shake gently for 15 min at room temperature. This cycle was repeated. To destain the gels, the staining solution was discarded and the gels were put into water. The water was heated, until the surplus Coomassie would be dissolved in it.

Reactions in MTP-format

Reactions in the 96-well plate format were performed with whole cell suspensions. The cells were grown in 1 ml AIM medium per well in 96-well deep-well plates (2 ml, Sarstedt) at 37 °C and 900 rpm overnight (approx. 24 h). Afterwards the cells were harvested through centrifugation at 4000 xg for 20 minutes. The medium was discarded and the cell pellets were resuspended in 200 µl 50 mM TEA buffer. The plates were centrifuged again at 4000 xg for 20 minutes, the buffer was discarded. Cell pellets were resuspended in 200 µl 50 mM TEA buffer again, to be used as the cell suspension in the reactions. The analyzed reactants were dissolved in DMSO. The concentrations were chosen following the guideline of 20 % (v/v) of DMSO in the reaction mix.

Reactions were prepared to have an end volume of 200 µl per well. The composition was chosen as follows:

- 10 µl 1 M TEA buffer
- 40 µl reactants in DMSO (20 % v/v)
- 150 µl of cell suspension in 50 mM TAE buffer

For homo-aldol reactions, 40 µl of 0.5 M *n*-butanal (**53**) solution in DMSO was used, amounting to 100 mM reactants per well. For crossed aldol additions, 20 µl of 1 M *n*-butanal (**53**) solution in DMSO and 20 µl of a 1 M solution of the other reactant (enolic donor) in DMSO were used, amounting to 100 mM of each reactant per well.

The prepared 96-well reaction plates were left shaking for 22 hours at room temperature and 900 rpm.

The screening of the reaction success was performed via thin layer chromatography.

Reactions with purified enzymes (chapter 1)

In parts of this work, similar reactions were performed with purified enzyme variants. For this purpose, enzyme solutions with a concentration of 3.5 mg/ml protein were prepared. Reactions were prepared to have an end volume of 200 µl per well. The composition was chosen as follows:

- 10 µl 1 M TEA*HCl buffer
- 40 µl reactants in DMSO (20 % v/v)
- 150 µl of enzyme solution in 50 mM TAE buffer (500 µg enzyme per reaction)

The DMSO concentration was kept at 20% (v/v), while the reactant concentrations varied. Those reactions were incubated in HPLC vials at room temperature and 900 rpm for 22 hours and analyzed via thin layer chromatography and HPLC.

General Workup of enzyme reactions (chapter 2):

Purification:

For precipitating the enzyme, 150ml methanol were added to the reaction solution and the mixture was stirred for 30 minutes at room temperature. The mixture was centrifuged and the solid precipitate was isolated from the mixture. For removal of the methanol, the mixture was applied to the rotary evaporator (30°C water bath). The mixture was extracted with ethyl acetate and dried with magnesium sulfate. After removal of the magnesium sulfate by filtration, the ethyl acetate was removed in vacuo. The obtained raw product was additionally purified via column chromatography on silica gel (cyclohexane/ethyl acetate 4:1 → 1:1).

HPLC samples preparation and analysis

For HPLC analysis, the reactions had to undergo certain procedures. 10 µl samples were drawn from the reaction mixtures (96-well or purified enzyme formats) and mixed with 50 µl O-benzoylhydroxylamine hydrochloride in pyridine. Those samples were then incubated for 1 h at 70 °C to obtain the reactant and product derivatives detectable in the HPLC. The derivatized samples were supplemented with 940 µl HPLC-grade methanol to precipitate the enzyme and centrifuged at 13000 xg to remove this precipitate. Afterwards the whole sample was carefully (without disturbing the pellet) transferred into fresh HPLC vials and loaded onto the HPLC to be analyzed with the corresponding program.

For quantification purposes, standard solutions (100 mM) of all reactants were prepared in 50 mM TEA buffer and 20 % (v/v) DMSO, allowing the calculation of reaction turnovers.

Protein expression of Fructose-6-phosphate aldolase

*Ec*FSA is prepared as previously reported with slight modifications.[110] For expression of *Ec*FSA *E. coli* BL21 (DE3) was transformed with the corresponding plasmid. Cells were grown in auto induction media [264] (4 l) containing ampilicin (100 mg·l⁻¹) for 22 h at 37 °C. Cells were harvested by centrifugation (2254 x g, 30 min) and suspended in 200 ml TEA buffer (50 mM, pH 7.5) containing lysozyme (1600 kU). Cell suspension was frozen at –20 °C. After thawing up to room temperature DNAse (800 U) and DTT (2 mM) were added. After incubation for 1 h cellular debris was removed by centrifugation (2540 x g for 30 min). The clear supernatant was purified by heat-shock treatment (70 °C, 30 min) followed by centrifugation (16000 x g for 10 min). The supernatant was concentrated by ultrafiltration to 5 ml and diluted to 250 ml with TEA buffer (2 mM, pH 7.5). After two cycles of diluting and concentrating the resulting solution was lyophilized. The lyophilized powder was stored at –20 °C until usage.

Effects of salts on protein stability

*Ec*FSA and HEWL were dissolved in twice the desired concentration of buffer. *Ec*FSA and HEWL were dissolved in 100 mM TEA-Buffer (pH=7.5). The final concentrations used in this study were 2 mg/mL for *Ec*FSA for comparison with older studies [87] and 0.5 mg/ml HEWL in TEA buffer (50 mM, pH 7.5).

Dilution series of 2 M sodium salt stock solutions with deionized water yields 1, 0.5, 0.25, 0.125, 0.0625 M concentrated salt solutions. Mixing the diluted salt solutions in equal parts with the stock enzyme solutions yields the desired concentrations used in this study.

Deprotection of dimethyl acetal

Before performing the deprotection it was necessary to wash the resin with 2N HCl, dest. water and THF.

- 100 mg dimethyl acetal protected aldehyde
- 90 mg Resin (SO₄ functionalization)
- Dilute to desired concentration with water (Not above 0.4 M)
- 1h reaction time

After filtration of the resin, the solution was applied to the enzyme reactions without further neutralization steps due to sufficient buffer concentration in the enzyme reaction.

Preparation of 2-((R)-1-hydroxy-3-methylbutyl)cyclopentan-1-one (75)

Reaction conditions:

- 200 ml total reaction volume
- 10.0 ml TEA*HCl buffer (1 м, pH 7.5) → 50 мм
- 2.194 ml dest. *i*-Valeraldehyde (**73**) → 100 мм
- 7.087 ml Cyclopentanone (**58**) → 500 мм
- 1.6 g Enzyme CFE D180/ FQ061
- 180.72 ml Dest. water
- Reaction time 2 days

Yield:

D180:

- Purified: 340 mg, 1.98 mmol, 10%

FQ061

- Purified: 711 mg, 4.2 mmol, 18%

Purification:

Standard purification

Elimination effects of β-hydroxyl group during purification with temperatures above 30°C.

Analytic:
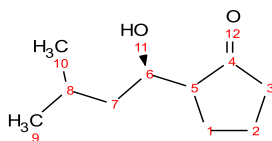
TLC R_f: 0.67 (4:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (300 MHz, Chloroform-*d*) δ 3.68 (tt, *J* = 8.3, 2.6 Hz, 1H), 3.50 (s, 1H), 2.41 – 2.21 (m, 1H), 2.20 – 1.77 (m, 4H), 1.77 – 1.58 (m, 1H), 1.53 – 1.30 (m, 2H), 1.06 (ddd, *J* = 13.8, 10.0, 2.6 Hz, 1H), 0.84 (dd, *J* = 8.5, 6.6 Hz, 6H).

^{13}C NMR (126 MHz, CDCl$_3$) δ 223.90, 70.33, 54.42, 44.56, 38.52, 26.82, 23.95, 23.82, 21.37, 20.51.
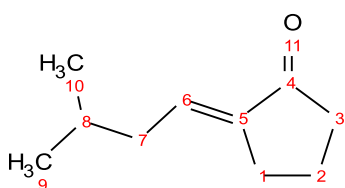
^{1}H NMR (300 MHz, Chloroform-*d*) δ 6.51 (tt, *J* = 7.7, 2.7 Hz, 1H), 2.51 (ddt, *J* = 10.0, 5.8, 1.4 Hz, 2H), 2.26 (t, *J* = 7.8 Hz, 2H), 1.97 (ddt, *J* = 8.1, 6.8, 1.5 Hz, 2H), 1.86 (p, *J* = 7.5 Hz, 2H), 1.71 (dt, *J* = 13.3, 6.7 Hz, 1H), 0.86 (d, *J* = 6.6 Hz, 7H).

^{13}C NMR (75 MHz, Chloroform-*d*) δ 207.09, 137.87, 135.17, 38.80, 38.63, 28.25, 26.88, 22.50, 19.79.

Preparation of 2-((R)-1-hydroxy-3-methylbutyl)cyclobutan-1-one (74)
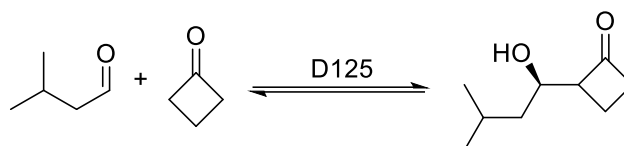
Reaction conditions:

- 200 ml total reaction volume
- 10.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 2.194 ml dest. *i*-Valeraldehyde (**73**) → 100 mM
- 5.866 ml Cyclobutanone (**57**) → 500 mM
- 181.94 ml Dest. water
- 1.6 g Enzyme CFE D125
- Reaction time 2 days

Yield:

D125:

- Purified: 581 mg, 3.72 mmol, 19%

Purification:

Standard purification

Analytic:
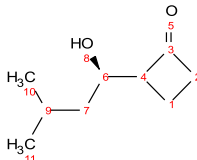
TLC R_f: 0.54 (4:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (300 MHz, Chloroform-*d*) δ 3.77 (ddd, *J* = 9.6, 7.2, 3.4 Hz, 1H, H6), 3.32 – 3.15 (m, 1H, H4), 3.07 – 2.73 (m, 2H, H2), 2.13 – 1.96 (m, 2H, H1), 1.77 (dtd, *J* = 11.5, 9.5, 7.4 Hz, 2H), 1.45 (ddd, *J* = 14.3, 9.6, 4.9 Hz, 1H), 1.16 (ddd, *J* = 13.9, 9.2, 3.4 Hz, 1H, H7), 0.85 (dd, *J* = 7.6, 6.5 Hz, 6H, H10 + H11).

^{13}C NMR (75 MHz, CDCl₃) δ 211.59, 69.56, 66.12, 45.06, 44.22, 24.19, 23.54, 21.70, 13.88.

Preparation of (4*R*)-3-fluoro-4-hydroxy-6-methylheptan-2-one (89)

Reaction conditions:

- 50 ml total reaction volume
- 2.5 ml TEA*HCl buffer (1 м, pH 7.5) → 50 mм
- 0.548 ml dest. *i*-Valeraldehyde (**73**) → 100 mм
- 1.804 ml Fluoroacetone (**77**) → 500 mм
- 46.15 ml Dest. water
- 400mg Enzyme CFE FQ061
- Reaction time 2 days

Yield:

D29:

- Purified linear product: 53 mg, 0.327 mmol, 7%
- Purified branched product: 342 mg, 2.109 mmol, 42%

Purification:

Standard purification

Analytic:
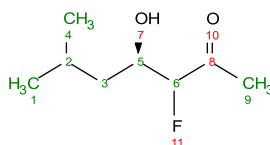
TLC R_f: 0.59 (4:1 Cyclohexane/Ethyl acetate)

^{1}H-NMR: (500 MHz, Chloroform-d) δ 4.62 (ddd, J = 48.8 Hz, 1H, H6), 4.22 – 4.02 (m, 1H, H5), 2.32 (d, J = 5.0 Hz, 3H, H9), 1.80 (dt, J = 13.6, 6.7 Hz, 1H, H2), 1.69 – 1.58 (m, 1H, H3), 1.40 (ddd, J = 13.7, 8.5, 4.7 Hz, 1H, H3), 0.97 (dd, J = 13.2, 6.6 Hz, 6H, H1+H4).

^{13}C NMR: (126 MHz, Chloroform-*d*) δ 208.1, 97.8, 96.4 70.2, 41.7, 27.4, 23.1, 21.8.

Preparation of (*R*)-1-fluoro-4-hydroxy-6-methylheptan-2-one (88)

Reaction conditions:

- 50 ml total reaction volume
- 2.5 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 0.548 ml dest. *i*-Valeraldehyde (**73**) → 100 mM
- 1.804 ml Fluoroacetone (**77**) → 500 mM
- 46.15 ml Dest. water
- 400mg Enzyme CFE D29
- Reaction time 2 days

Yield:

FQ061:

- Purified linear product: 390 mg, 2.404 mmol, 48%

Purification:

Standard purification

Analytic:
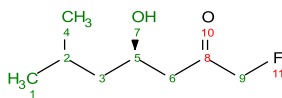
TLC R_f: 0.47 (4:1 Cyclohexane/Ethyl acetate)

^{1}H NMR: (500 MHz, Chloroform-*d*) δ 4.83 (ddd, *J* = 47.6 Hz, 2H, H9), 4.25 – 4.11 (m, 1H, H5), 2.67 (pt, J = 10.3, 5.3 Hz, 3H, H6+H7), 1.87 – 1.72 (m, 1H, H2), 1.51 (s, 1H, H3), 1.21 (ddd, *J* = 13.4, 8.6, 4.4 Hz, 1H, H3), 0.94 (d, *J* = 6.6 Hz, 6H, H1+H4).

^{13}C NMR: (126 MHz, CDCl3) δ 207.65, 85.83, 84.36, 65.32, 45.81, 24.40, 23.22, 21.97.

Preparation of (*R*)-1-chloro-4-hydroxy-6-methylheptan-2-one (90)
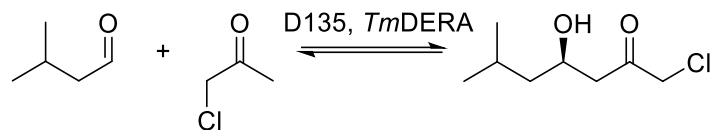
Reaction conditions:

- 100 ml total reaction volume
- 5.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 1.097 ml dest. *i*-Valeraldehyde (**73**) → 100 mM
- 4.023 ml Chloroacetone (**66**) → 500 mM
- 89.88 ml Dest. Water
- 800mg Enzyme CFE D135/*Tm*DERA
- Reaction time 2 days

Yield:

D135:   *Tm*DERA

- Purified: 471 mg, 2.636 mmol, 26%   • Purified: 550 mg, 3.078 mmol, 31%

Purification:

Standard purification

Analytic:
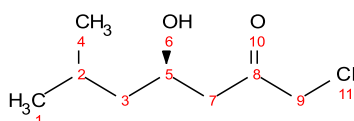
TLC R$_f$: 0.59 (4:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (500 MHz, Chloroform-d) δ 4.19 (ddd, J = 8.2, 3.9 Hz, 1H, H5), 4.14 (s, 2H, H9), 2.84 − 2.61 (m, 2H, H7), 1.88 − 1.74 (m, 1H, H2), 1.50 (ddd, J = 14.2, 8.9, 5.5 Hz, 1H, H3), 1.21 (ddd, J = 13.3, 8.6, 4.4 Hz, 1H, H3), 0.93 (d, J = 6.6 Hz, 6H, H1+H4).

^{13}C NMR (126 MHz, CDCl$_3$) δ 203.24, 65.85, 48.72, 47.04, 45.73, 24.41, 23.23, 21.96.

Preparation of (*R*)-1-fluoro-4-hydroxy-6-methylheptan-2-one (93)
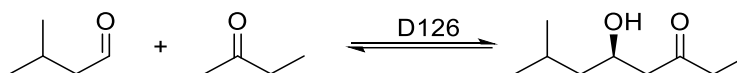
Reaction conditions:

- 100 ml total reaction volume
- 5.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 1.097 ml dest. *i*-Valeraldehyde (**73**) → 100 mM
- 3.608 ml Butanone (**55**) → 500 mM
- 90.30
- 800mg Enzyme CFE D126
- Reaction time 2 days

Yield:

- 319 mg, 2.01 mmol, 20%

Purification:

Standard purification

Analytic:
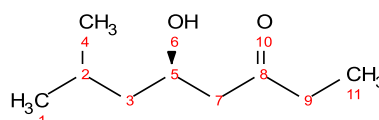
TLC R$_f$: 0.29 (2:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (500 MHz, Chloroform-*d*) δ 4.13 (ddd, *J* = 9.1, 4.5, 2.8 Hz, 1H, H5), 3.08 (s, 1H, H6), 2.58 (dd, *J* = 17.4, 2.9 Hz, 1H, H7), 2.55 – 2.44 (m, 2H, H9), 2.45 (d, *J* = 7.3 Hz, 1H, H7), 1.79 (ddt, *J* = 15.3, 13.1, 6.7 Hz, 1H, H2), 1.47 (ddd, *J* = 14.1, 9.0, 5.5 Hz, 1H, H3), 1.15 (dt, *J* = 9.5, 4.5 Hz, 1H, H3), 1.07 (t, *J* = 7.3 Hz, 3H, H11), 0.92 (dd, *J* = 6.5, 1.5 Hz, 6H, H1+H4).

^{13}C NMR (126 MHz, CDCl$_3$) δ 212.83, 65.75, 49.10, 45.57, 36.77, 24.35, 23.29, 22.01, 7.54.

Preparation of (4*R*)-4-hydroxy-3-methoxy-6-methylheptan-2-one (95)
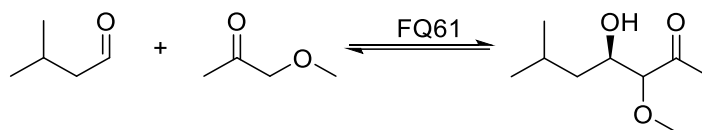
Reaction conditions:

- 100 ml total reaction volume
- 5.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 1.097 ml dest. *i*-Valeraldehyde (**73**) → 100 mM
- 4.604 ml methoxyacetone (**79**) → 500 mM
- 89.30 ml Dest. water
- 800mg Enzyme CFE FQ061
- Reaction time 2 days

Yield:

D135:

- Purified: 1085 mg, 7.37 mmol, 74%

Purification:

Standard purification

Analytic:
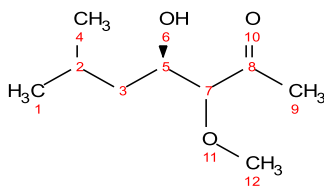
TLC R$_f$: 0.22 (4:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (500 MHz, Chloroform-*d*) δ 3.87 (ddd, *J* = 10.3, 3.6 Hz, 1H, H5), 3.45 (s, 3H, H12), 3.43 (d, *J* = 4.2 Hz, 1H, H7), 2.22 (s, 3H, H9), 1.86 – 1.74 (m, 1H, H2), 1.51 (ddd, *J* = 14.5, 9.6, 5.2 Hz, 1H, H3), 1.22 (ddd, *J* = 13.7, 8.8, 3.7 Hz, 1H, H3), 0.94 (d, *J* = 6.7 Hz, 3H, H1), 0.91 (d, *J* = 6.6 Hz, 3H, H6).

^{13}C NMR (126 MHz, CDCl$_3$) δ 211.12, 90.20, 70.35, 59.21, 42.24, 26.91, 24.42, 23.40, 21.75.

Preparation of (*R*)-4-hydroxy-1-methoxy-6-methyheptan-2-one (94)
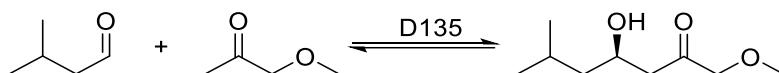
Reaction conditions:

- 100 ml total reaction volume
- 5.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 1.097 ml dest. *i*-Valeraldehyde (**73**) → 100 mM
- 4.604 ml methoxyacetone (**79**) → 500 mM
- 89.30 ml Dest. water
- 800mg Enzyme CFE D135
- Reaction time 2 days

Yield:

D135:

- Purified: 1085 mg, 4.91 mmol, 49%

Purification:

Standard purification

Analytic:
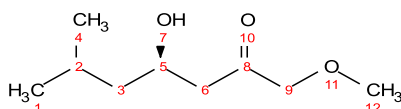
TLC R_f: 0.46 (4:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (500 MHz, Chloroform-*d*) δ 4.09 (ddd, *J* = 12.1, 8.8, 3.7 Hz, 1H, H5), 3.95 (s, 3H, H9), 3.36 (s, 3H, H12), 2.59 – 2.43 (m, 2H, H6), 1.72 (dp, *J* = 13.3, 6.6 Hz, 1H, H2), 1.42 (ddd, *J* = 14.0, 8.9, 5.6 Hz, 1H, H3), 1.11 (ddd, *J* = 13.3, 8.5, 4.5 Hz, 1H, H3), 0.86 (s, 3H, H1), 0.84 (s, 3H, H4).

^{13}C NMR (126 MHz, CDCl$_3$) δ 209.69, 77.99, 59.34, 46.05, 45.80, 24.39, 23.24, 22.01.

Preparation of 5,5-Diemthyl-4-hydroxyhexan-2-one
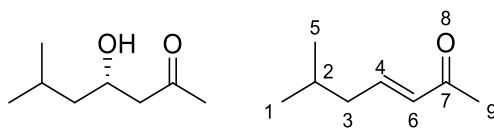
Synthesis following LUESCH *et al.*

D-Proline (1.020 g, 8.865 mmol) was added to a mixture of acetone (**50**) (50 ml) and DMSO (200 ml) at room temperature and the mixture was stirred for 1 h. *i*-valeraldehyde (**73**) (2.545 g, 29.55 mmol) was added and the resulting mixture was stirred for 5 days. The reaction was followed via thin layer chromatography. NH_4Cl (aq.) was added to quench the reaction. After removal of acetone (**50**) the solution was extracted with ethyl acetate and the extract was diluted with ethyl acetate and washed 7 times with small portions of water to remove most of the DMSO. The resulting organic layer was dried via $MgSO_4$ and the solvent was removed in vacuo. The residue was purified via column chromatography with a gradient of 10:1 → 2:1 (CH:EE) to give the elimination product of compound (0.668g, 6.4 mmol, 22%).

^{1}H NMR (300 MHz, Chloroform-*d*) δ 6.71 (dt, *J* = 15.3, 7.4 Hz, 1H, H-4), 5.99 (d, *J* = 15.8 Hz, 1H, H-6), 2.18 (s, 3H, H-9), 2.05 (t, *J* = 7.3 Hz, 2H, H-3), 1.71 (dt, *J* = 13.4, 6.7 Hz, 1H, H-2), 0.88 (s, 3H, H-1), 0.85 (s, 3H, H-5).

Preparation of 2-((R)-1,3-dihydroxy-3-methylbutyl)cyclopentan-1-one (99)
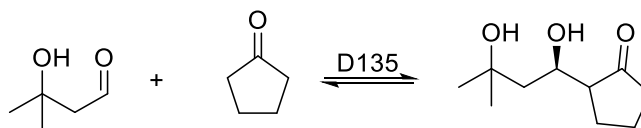
Reaction conditions:

- 100 ml total reaction volume
- 5.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 33.33 ml 0.4 M 3-Hydroxy-3-methylbutanal (**98**) (aq.) → 100 mM
- 4.426 ml Cyclopentanone (**58**) → 500 mM
- 57.24 ml Dest. water
- 800mg Enzyme CFE D135
- Reaction time 2 days

Yield:

D135:

- Purified: 85 mg, 0.46 mmol, 5%
  As a mixture of aldol product and the eliminated form indicated by ^{13}C-NMR and HPLC/MS.

Purification:

Standard purification

Analytic:

TLC R_f: 0.18 (2:1 Cyclohexane/Ethyl acetate)

^{13}C NMR (126 MHz, CDCl$_3$) δ 223.42, 220.88, 144.16, 133.93, 71.47, 70.80, 70.73, 67.66, 55.06, 54.08, 46.63, 45.78, 45.42, 39.10, 38.48, 31.87, 31.29, 29.51, 27.71, 27.68, 26.41, 23.33, 20.60, 20.46.

Preparation of 2-((R)-1,3-dihydroxy-3-methybutyl)cyclobutan-1-one (100)
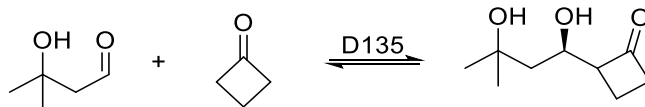
Reaction conditions:

- 100 ml total reaction volume
- 5.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 33.33 ml 0.4 M 3-Hydroxy-3-methylbutanal (**98**) (aq.) → 100 mM
- 3.670 ml Cyclobutanone (**57**) → 500 mM
- 58.00 ml Dest. water
- 800mg Enzyme CFE D180
- Reaction time 2 days

Yield:

D135:

- Purified: 180 mg, 1.05 mmol, 11%

Purification:

Standard purification

Analytic:
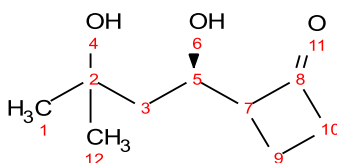
TLC R_f: 0.17 (2:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (300 MHz, Chloroform-*d*) δ 4.19 (ddd, *J* = 11.0, 6.3, 2.1 Hz, 1H, H5), 3.91 (s, 1H, H4/H9), 3.43 – 3.27 (m, 1H, H7), 3.00 (dtd, *J* = 9.8, 7.9, 2.2 Hz, 2H, H10), 2.23 – 2.03 (m, 1H, H9), 2.02 – 1.82 (m, 2H, H3), 1.48 (dd, *J* = 14.5, 2.1 Hz, 1H, H9), 1.30 (s, 3H, H1/H12), 1.28 (s, 4H, H1/H12).

^{13}C NMR (75 MHz, CDCl₃) δ 211.30, 71.44, 68.90, 65.85, 45.29, 45.20, 31.76, 27.54, 13.44.

Preparation of (*R*)-5,7-dihydroxy-7-methyloctan-3-one (101)
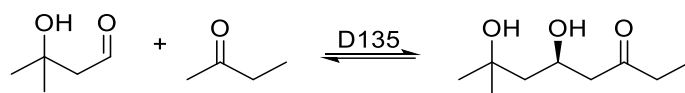
Reaction conditions:

- 100 ml total reaction volume
- 5.0 ml TEA*HCl buffer (1 M, pH 7.5) → 50 mM
- 33.33 ml 0.4 M 3-Hydroxy-3-methylbutanal (**98**) (aq.) → 100 mM
- 4.480 ml Cyclobutanone (**57**) → 500 mM
- 57.19 Dest. Water
- 800mg Enzyme CFE D135
- Reaction time 2 days

Yield:

D135:

- Purified: 103 mg, 0.59 mmol, 6%

Purification:

Standard purification

Analytic:
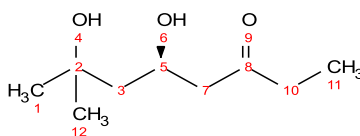
TLC R_f: 0.16 (2:1 Cyclohexane/Ethyl acetate)

^{1}H NMR (300 MHz, Chloroform-*d*) δ 4.45 (dddd, *J* = 10.7, 8.1, 4.0, 2.4 Hz, 1H, H5), 2.67 – 2.37 (m, 8H, H6+H10), 1.71 – 1.59 (m, 1H, H3), 1.43 (dd, *J* = 14.4, 2.4 Hz, 1H, H3), 1.29 (s, 7H, H1), 1.21 (s, 2H, H12), 1.02 (t, *J* = 7.3 Hz, 5H, H11).

^{13}C NMR (75 MHz, CDCl$_3$) δ 211.95, 71.09, 65.89, 49.44, 47.08, 36.75, 31.35, 27.83, 7.48.

Preparation of 4,4-dimethoxybutan-2-ol (68)
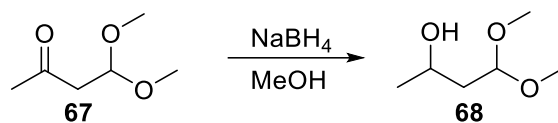
The reaction was conducted in preheated glassware. A solution of sodium borohydride (5.68 g, 150 mmol, 1.5 eq) in methanol (150ml) was prepared. To this solution 4,4-dimethoxybutan-2-one (13.2 g, 100 mmol, 1 eq) was slowly. After cooling the solution to 0°C, the reaction was quenched with aq. $NaHCO_3$ (2.7 g, 1 eq) and stirred for 30 minutes.

The product was extracted with diethyl ether. After water removal with $MgSO_4$, the solvent was removed in vacuo. The product (12.8 g, 95 mmol) as a colorless oil.

^{1}H NMR (300 MHz, CDCl₃) δ 4.50 (t, J = 5.6 Hz, 1H), 3.86 (dqd, J = 7.6, 6.3, 4.6 Hz, 1H), 3.28 (d, J = 3.2 Hz, 6H), 1.66 (ddd, J = 5.7, 5.1, 2.5 Hz, 2H), 1.11 (d, J = 6.3 Hz, 3H).

^{13}C NMR (75 MHz, CDCl₃) δ 103.18, 64.06, 53.08, 52.64, 40.86, 23.30.

Preparation of 4,4-dimethoxy-2-methylbutan-2-ol (97)
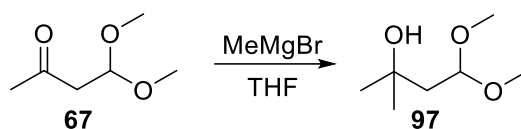
The reaction was conducted in preheated glassware. A solution of 4,4-dimethoxybutan-2-one (13.2 g, 100 mmol, 1 eq) in THF (200ml) was prepared. To this solution methylmagnesium bromide (3M in ethyl ether, 38.0 ml, 110 mmol, 1.1 eq) was slowly added under argon atmosphere. The solution was stirred for 30 minutes at room temperature. The solution was quenched with 50 ml of acetone. After removal of the solvent a white solid was obtained. After addition of water (200 ml), the product was extracted three times with ethyl acetate (75 ml). After water removal with MgSO$_4$, the solvent was removed in vacuo.

The product (13.5 g, 91 mmol) was obtained as a colorless oil.

^{1}H NMR (300 MHz, CDCl$_3$) δ 4.67 (t, J = 6.0 Hz, 1H), 3.37 (s, 6H), 1.81 (d, J = 6.0 Hz, 2H), 1.24 (s, 6H).

^{13}C NMR (75 MHz, DMSO) δ 97.95, 64.27, 48.09, 39.13, 24.84.

[1]   U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185-194.

[2]   M. D. Truppo, *ACS Medicinal Chemistry Letters* **2017**, *8*, 476-480.

[3]   J. B. Pyser, S. Chakrabarty, E. O. Romero, A. R. H. Narayan, *ACS Central Science* **2021**, *7*, 1105-1116.

[4]   U. T. Bornscheuer, K. Buchholz, *Engineering in Life Sciences* **2005**, *5*, 309-323.

[5]   R. Csuk, B. I. Glaenzer, *Chemical Reviews* **1991**, *91*, 49-97.

[6]   S. Servi, *Synthesis* **1990**, *01*, 1-25.

[7]   C. Fuganti, *Pure and Applied Chemistry* **1990**, *62*, 1449-1452.

[8]   O. P. Ward, C. S. Young, *Enzyme and Microbial Technology* **1990**, *12*, 482-493.

[9]   R. d. S. Pereira, *Critical Reviews in Biotechnology* **1998**, *18*, 25-64.

[10]   D. A. Estell, T. P. Graycar, J. A. Wells, *Journal of Biological Chemistry* **1985**, *260*, 6518-6521.

[11]   V. J. Jensen, S. Rugh, in *Methods in Enzymology, Vol. 136*, Academic Press, **1987**, pp. 356-370.

[12]   A. Bruggink, E. C. Roos, E. de Vroom, *Organic Process Research & Development* **1998**, *2*, 128-133.

[13]   A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, *Nature* **2001**, *409*, 258-268.

[14]   A. R. Alcántara, P. Domínguez de María, J. A. Littlechild, M. Schürmann, R. A. Sheldon, R. Wohlgemuth, *ChemSusChem* **2022**, *15*, e202102709.

[15]   J. B. Zimmerman, P. T. Anastas, H. C. Erythropel, W. Leitner, **2020**, *367*, 397-400.

[16]   R. A. Sheldon, D. Brady, **2022**, *15*, e202102628.

[17]   R. D. Lewis, S. P. France, C. A. Martinez, *ACS Catalysis* **2023**, *13*, 5571-5577.

[18]   S. P. France, R. D. Lewis, C. A. Martinez, *JACS Au* **2023**, *3*, 715-735.

[19]   K. Faber, W. D. Fessner, N. J. Turner, *Science of Synthesis: Biocatalysis in Organic Synthesis Vol. 1*, Georg Thieme Verlag KG, Stuttgart, **2015**.

[20]   H. E. Schoemaker, D. Mink, M. G. Wubbolts, *Science* **2003**, *299*, 1694-1697.

[21]   H. Matsumae, M. Furui, T. Shibatani, *Journal of Fermentation and Bioengineering* **1993**, *75*, 93-98.

[22] H. Griengl, H. Schwab, M. Fechter, *Trends in Biotechnology* **2000**, *18*, 252-256.

[23] W. P. Stemmer, *National Acad Sciences* **1994**, *91*, 10747-10751.

[24] H. Zhao, L. Giver, Z. Shao, J. A. Affholter, F. H. Arnold, *Nature Biotechnology* **1998**, *16*, 258-261.

[25] A. Crameri, E. A. Whitehorn, E. Tate, W. P. C. Stemmer, *Nature Biotechnology* **1996**, *14*, 315-319.

[26] F. H. Arnold, *Accounts of Chemical Research* **1998**, *31*, 125-131.

[27] U. T. Bornscheuer, *Philosophical Transactions of the Royal Society A* **2018**, *376*, 20170063.

[28] T. D. Machajewski, C.-H. Wong, *Chemie International Edition* **2000**, *39*, 1352-1375.

[29] C. F. Barbas, III, Y. F. Wang, C. H. Wong, *Journal of the American Chemical Society* **1990**, *112*, 2013-2014.

[30] B. Elvers, S. Hawklins, G. Schulz, *Vol. A*, 5th ed. (Eds.: B. Elvers, S. Hawklins, G. Schulz), Wiley - VCH Verlag GmbH & Co., Hoboken, NJ, **1991**.

[31] B. L. Horecker, O. Tsolas, C. Y. Lai, in *The Enzymes, Vol. 7* (Ed.: P. D. Boyer), *Academic Press*, **1972**, pp. 213-258.

[32] W.-D. Fessner, C. Walter, in *Bioorganic Chemistry: Models and Applications* (Ed.: F. P. Schmidtchen), Springer Berlin Heidelberg, Berlin, Heidelberg, **1997**, pp. 97-194.

[33] T. Gefflaut, C. Blonski, J. Perie, M. Willson, *Progress in biophysics and molecular biology* **1995**, *63*, 301-340.

[34] W.-D. Fessner, A. Schneider, H. Held, G. Sinerius, C. Walter, M. Hixon, J. V. Schloss, *Angewandte Chemie International Edition* **1996**, *35*, 2219-2221.

[35] M. L. di Salvo, S. G. Remesh, M. Vivoli, M. S. Ghatge, A. Paiardini, S. D'Aguanno, M. K. Safo, R. Contestabile, *The FEBS journal* **2014**, *281*, 129-145.

[36] J. Rutter W, *Fed. Proc.* **1964**, *23*, 1248-1257.

[37] A. K. Samland, G. A. Sprenger, *Applied Microbiology and Biotechnology* **2006**, *71*, 253-264.

[38] P. Clapés, W.-D. Fessner, G. A. Sprenger, A. K. Samland, *Current Opinion in Chemical Biology* **2010**, *14*, 154-167.

[39] V. Hélaine, C. Gastaldi, M. Lemaire, P. Clapés, C. Guérard-Hélaine, *ACS Catalysis* **2022**, *12*, 733-761.

[40]   P. Clapés, in *Green Biocatalysis*, **2016**, pp. 267-306.

[41]   P. Clapés, X. Garrabou, *Advanced Synthesis & Catalysis* **2011**, *353*, 2263-2283.

[42]   W.-D. Fessner, *Current Opinion in Chemical Biology* **1998**, *2*, 85-97.

[43]   W.-D. Fessner, G. Sinerius, A. Schneider, M. Dreyer, G. E. Schulz, J. Badia, J. Aguilar, *Angewandte Chemie International Edition* **1991**, *30*, 555-558.

[44]   A. Ozaki, E. J. Toone, C. H. Von der Osten, A. J. Sinskey, G. M. Whitesides, *Journal of the American Chemical Society* **1990**, *112*, 4970-4971.

[45]   O. Eyrisch, G. Sinerius, W.-D. Fessner, *Carbohydrate Research* **1993**, *238*, 287-306.

[46]   M. D. Bednarski, E. S. Simon, N. Bischofberger, W.-D. Fessner, M.-J. Kim, W. J. Lees, T. Saito, H. J. Waldmann, G. M. Whitesides, *Journal of the American Chemical Society* **1989**, *111*, 627-635.

[47]   G. Guanti, L. Banfi, M. T. Zannetti, *Tetrahedron Letters* **2000**, *41*, 3181-3185.

[48]   L. Azéma, F. Bringaud, C. Blonski, J. Périé, *Bioorganic & Medicinal Chemistry* **2000**, *8*, 717-722.

[49]   G. Guanti, M. T. Zannetti, L. Banfi, R. Riva, *Advanced Synthesis & Catalysis* **2001**, *343*, 682-691.

[50]   M. Mitchell, L. Qaio, C.-H. Wong, *Advanced Synthesis & Catalysis* **2001**, *343*, 596-599.

[51]   M. Schuster, W.-F. He, S. Blechert, *Tetrahedron Letters* **2001**, *42*, 2289-2291.

[52]   A. Li, L. Cai, Z. Chen, M. Wang, N. Wang, H. Nakanishi, X. D. Gao, Z. Li, *Carbohydrate research* **2017**, *452*, 108-115.

[53]   E. Busto, *ChemCatChem* **2016**, *8*, 2589-2598.

[54]   M. Bilal, H. M. N. Iqbal, H. Hu, W. Wang, X. Zhang, *Critical Reviews in Food Science and Nutrition* **2018**, *58*, 2768-2778.

[55]   L. Espelt, T. Parella, J. Bujons, C. Solans, J. Joglar, A. Delgado, P. Clapés, *Chemistry–A European Journal* **2003**, *9*, 4887-4899.

[56]   M. Schümperli, R. Pellaux, S. Panke, *Appl Microbiol Biotechnol* **2007**, *75*, 33-45.

[57]   C. J. Hartley, N. G. French, J. A. Scoble, C. C. Williams, Q. I. Churches, A. R. Frazer, M. C. Taylor, G. Coia, G. Simpson, N. J. Turner, C. Scott, *PLOS ONE* **2017**, *12*, e0184183.

[58] J. Hettwer, H. Oldenburg, E. Flaschel, *Journal of Molecular Catalysis B: Enzymatic* **2002**, *19-20*, 215-222.

[59] R. A. John, *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1995**, *1248*, 81-96.

[60] R. Percudani, A. Peracchi, *EMBO reports* **2003**, *4*, 850-854.

[61] J. Steinreiber, K. Fesko, C. Reisinger, M. Schürmann, F. van Assema, M. Wolberg, D. Mink, H. Griengl, *Tetrahedron* **2007**, *63*, 918-926.

[62] J.-Q. Liu, T. Dairi, N. Itoh, M. Kataoka, S. Shimizu, H. Yamada, *Journal of Molecular Catalysis B: Enzymatic* **2000**, *10*, 107-115.

[63] M. Wada, M. Sakamoto, M. Kataoka, J.-Q. Liu, H. Yamada, S. Shimizu, *Bioscience, Biotechnology, and Biochemistry* **1998**, *62*, 1586-1588.

[64] G. Jander, S. R. Norris, V. Joshi, M. Fraga, A. Rugg, S. Yu, L. Li, R. L. Last, *The Plant Journal* **2004**, *39*, 465-475.

[65] A. J. Edgar, *BMC Genomics* **2005**, *6*, 32.

[66] J. Q. Liu, M. Odani, T. Yasuoka, T. Dairi, N. Itoh, M. Kataoka, S. Shimizu, H. Yamada, *Applied Microbiology and Biotechnology* **2000**, *54*, 44-51.

[67] S. Hajra, A. Karmakar, T. Maji, A. K. Medda, *Tetrahedron* **2006**, *62*, 8959-8965.

[68] K. C. Nicolaou, C. N. C. Boddy, S. Bräse, N. Winssinger, *Angewandte Chemie International Edition* **1999**, *38*, 2096-2152.

[69] T. Kimura, V. P. Vassilev, G.-J. Shen, C.-H. Wong, *Journal of the American Chemical Society* **1997**, *119*, 11734-11742.

[70] T. Miura, T. Kajimoto, *Chirality* **2001**, *13*, 577-580.

[71] V. P. Vassilev, T. Uchiyama, T. Kajimoto, C.-H. Wong, *Tetrahedron Letters* **1995**, *36*, 4081-4084.

[72] S.-H. Baik, H. Yoshioka, *Biotechnology Letters* **2009**, *31*, 443-448.

[73] W. Wang, S. Y. K. Seah, *Biochemistry* **2005**, *44*, 9447-9455.

[74] H. Ling, G. Wang, Y. Tian, G. Liu, H. Tan, *Biochemical and Biophysical Research Communications* **2007**, *361*, 196-201.

[75] W.-D. Fessner, in *Asymmetric Organic Synthesis with Enzymes*, **2008**, pp. 275-318.

[76] W.-D. Fessner, N. He, D. Yi, P. Unruh, M. Knorst, in *Cascade Biocatalysis*, **2014**, pp. 361-392.

[77] A. Bolt, A. Berry, A. Nelson, *Archives of Biochemistry and Biophysics* **2008**, *474*, 318-330.

[78] G. J. Williams, T. Woodhall, A. Nelson, A. Berry, *Protein Engineering, Design and Selection* **2005**, *18*, 239-246.

[79] M. J. Kim, W. J. Hennen, H. M. Sweers, *Journal of the American Chemical Society* **1988**, *110*, 6481-6486.

[80] W. Fitz, J.-R. Schwark, C.-H. Wong, *The Journal of Organic Chemistry* **1995**, *60*, 3663-3670.

[81] C. H. Lin, T. Sugai, R. L. Halcomb, Y. Ichikawa, Wong, *Journal of the American Chemical Society* **1992**, *114*, 10138-10145.

[82] H. Yu, X. Chen, *Organic Letters* **2006**, *8*, 2393-2396.

[83] S. Blayer, J. M. Woodley, M. D. Lilly, M. J. Dawson, *Biotechnology* **1996**, *12*, 758-763.

[84] A. J. Humphrey, C. Fremann, P. Critchley, Y. Malykh, R. Schauer, T. D. H. Bugg, *Bioorganic & Medicinal Chemistry* **2002**, *10*, 3175-3185.

[85] R. Mahrwald, *Modern methods in stereoselective aldol reactions*, John Wiley & Sons, **2013**.

[86] M. Schürmann, G. A. Sprenger, *Journal of Biological Chemistry* **2001**, *276*, 11055-11061.

[87] A. O. Magnusson, A. Szekrenyi, H.-J. Joosten, J. Finnigan, S. Charnock, W.-D. Fessner, *The FEBS Journal* **2019**, *286*, 184-204.

[88] S. Thorell, M. Schürmann, G. A. Sprenger, G. Schneider, *Journal of Molecular Biology* **2002**, *319*, 161-171.

[89] F. Acke Dissertation, Technische Universität (Darmstadt) AK Fessner, **2020**.

[90] A. K. Samland, M. Rale, G. A. Sprenger, W.-D. Fessner, *ChemBioChem* **2011**, *12*, 1454-1474.

[91] J. A. Castillo, J. Calveras, J. Casas, M. Mitjans, M. P. Vinardell, T. Parella, T. Inoue, G. A. Sprenger, J. Joglar, P. Clapés, *Organic Letters* **2006**, *8*, 6067-6070.

[92] X. Garrabou, J. A. Castillo, C. Guérard-Hélaine, T. Parella, J. Joglar, M. Lemaire, P. Clapés, *Angewandte Chemie International Edition* **2009**, *48*, 5521-5525.

[93] A. L. Concia, C. Lozano, J. A. Castillo, T. Parella, J. Joglar, P. Clapés, *Chemistry–A European Journal* **2009**, *15*, 3808-3816.

[94] G. Masdeu, L. M. Vázquez, J. López-Santín, G. Caminal, S. Kralj, D. Makovec, G. Álvaro, M. Guillén, *PLOS ONE* **2021**, *16*, e0250513.

[95] J. A. Castillo, C. Guérard-Hélaine, M. Gutiérrez, X. Garrabou, M. Sancelme, M. Schürmann, T. Inoue, V. Hélaine, F. Charmantray, T. Gefflaut, L. Hecquet, J. Joglar, P. Clapés, G. A. Sprenger, M. Lemaire, *Advanced Synthesis* **2010**, *352*, 1039-1046.

[96] E. Racker, *Journal of Biological Chemistry* **1952**, *196*, 347-365.

[97] W. E. Pricer, B. L. Horecker, *Journal of Biological Chemistry* **1960**, *235*, 1292-1298.

[98] M. G. Tozzi, M. Camici, L. Mascia, F. Sgarrella, P. L. Ipata, *The FEBS journal* **2006**, *273*, 1089-1101.

[99] P. Hoffee, P. Snyder, C. Sushak, P. Jargiello, *Archives of Biochemistry and Biophysics* **1974**, *164*, 736-742.

[100] T.-P. Cao, J.-S. Kim, M.-H. Woo, J. M. Choi, Y. Jun, K. H. Lee, S. H. Lee, *Journal of Microbiology* **2016**, *54*, 311-321.

[101] H. Sakuraba, K. Yoneda, K. Yoshihara, K. Satoh, R. Kawakami, Y. Uto, H. Tsuge, K. Takahashi, H. Hori, T. Ohshima, *Applied and environmental microbiology* **2007**, *73*, 7427-7434.

[102] X.-C. Jiao, J. Pan, G.-C. Xu, X.-D. Kong, Q. Chen, Z.-J. Zhang, J.-H. Xu, *Catalysis Science & Technology* **2015**, *5*, 4048-4054.

[103] A. Heine, G. DeSantis, J. G. Luz, M. Mitchell, C.-H. Wong, I. A. Wilson, *Sience* **2001**, *294*, 369-374.

[104] M. Schulte, D. Petrović, P. Neudecker, R. Hartmann, J. Pietruszka, S. Willbold, D. Willbold, V. Panwalkar, *ACS Catalysis* **2018**, *8*, 3971-3984.

[105] H. Fei, C.-c. Zheng, X.-y. Liu, Q. Li, *Process Biochemistry* **2017**, *63*, 55-59.

[106] M. Dick, R. Hartmann, O. H. Weiergräber, C. Bisterfeld, T. Classen, M. Schwarten, P. Neudecker, D. Willbold, J. Pietruszka, *Chemical Science* **2016**, *7*, 4492-4502.

[107] S. Jennewein, M. Schürmann, M. Wolberg, I. Hilker, R. Luiten, M. Wubbolts, D. Mink, *Biotechnology Journal: Healthcare Nutrition Technology* **2006**, *1*, 537-548.

[108] https://clincalc.com/DrugStats/Drugs/Atorvastatin.

[109] L. Chen, D. P. Dumas, Wong, *Journal of the American Chemical Society* **1992**, *114*, 741-748.

[110] S. Junker, R. Roldan, H. J. Joosten, P. Clapes, W. D. Fessner, *Angewandte Chemie (International ed. in English)* **2018**, *57*, 10153-10157.

[111] R. Roldán, K. Hernandez, J. Joglar, J. Bujons, T. Parella, I. Sánchez-Moreno, V. Hélaine, M. Lemaire, C. Guérard-Hélaine, W.-D. Fessner, P. Clapés, *ACS Catalysis* **2018**, *8*, 8804-8809.

[112] G. D. Haki, S. K. Rakshit, *Bioresource Technology* **2003**, *89*, 17-34.

[113] T. Narancic, R. Davis, J. Nikodinovic-Runic, K. E. O' Connor, *Biotechnology Letters* **2015**, *37*, 943-954.

[114] J. M. Woodley, *Curr Opin Chem Biol* **2013**, *17*, 310-316.

[115] M. T. Reetz, J. D. Carballeira, *Nature Protocols* **2007**, *2*, 891-903.

[116] M. Reetz, **2022**, *23*, e202200049.

[117] G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, **2020**, *59*, 13204-13231.

[118] F. Cedrone, A. Ménez, E. Quéméneur, *Current Opinion in Structural Biology* **2000**, *10*, 405-410.

[119] D. Güclü, A. Szekrenyi, X. Garrabou, M. Kickstein, S. Junker, P. Clapés, W.-D. Fessner, *ACS Catalysis* **2016**, *6*, 1848-1852.

[120] R. Roldán, K. Hernández, J. Joglar, J. Bujons, T. Parella, W.-D. Fessner, P. Clapés, *Advanced Synthesis* **2019**, *361*, 2673-2687.

[121] B. Elvers, *Ullmann's encyclopedia of industrial chemistry, Vol. 17*, Verlag Chemie Hoboken, NJ, **1991**.

[122] R. Roldán, I. Sanchez-Moreno, T. Scheidt, V. Hélaine, M. Lemaire, T. Parella, P. Clapés, W.-D. Fessner, C. Guérard-Hélaine, *Chemistry–A European Journal* **2017**, *23*, 5005-5009.

[123] https://3dm.bio-prodict.com/.

[124] T. Glesner, *Masterthesis TU Darmstadt AK Fessner* **2017**.

[125] D. Chambre, C. Guérard-Hélaine, E. Darii, A. Mariage, J.-L. Petit, M. Salanoubat, V. de Berardinis, M. Lemaire, V. Hélaine, *Chemical Communications* **2019**, *55*, 7498-7501.

[126] S. Fadanka, S. Minette, N. Mowoh, *protocols.io* **2022**.

[127] J. R. Durrwachter, Wong, *Journal of the American Chemical Society* **1988**, *53*, 4175-4181.

[128] J. Bramski, M. Dick, J. Pietruszka, T. Classen, *Journal of Biotechnology* **2017**, *258*, 56-58.

[129] B. K. Lundholt, K. M. Scudder, L. Pagliaro, *SLAS Discovery* **2003**, *8*, 566-570.

[130] N. E. Biolabs.

[131] N. E. Biolabs.

[132] K. A. Eckert, T. A. Kunkel, *Nucleic acids research* **1990**, *18*, 3739-3744.

[133] L. Chen, D. P. Dumas, C. H. Wong, *Journal of the American Chemical Society* **1992**, *114*, 741-748.

[134] M. Sigma-Aldrich.

[135] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, R. Taylor, *Journal of the Chemical Society, Perkin Transactions 2* **1987**, S1-S19.

[136] K. A. Scott, D. O. V. Alonso, S. Sato, A. R. Fersht, V. Daggett, *Proceedings of the National Academy of Sciences* **2007**, *104*, 2661-2666.

[137] L. d. F. Alves, C. A. Westmann, G. L. Lovate, G. M. V. de Siqueira, T. C. Borelli, M.-E. Guazzaroni, *International Journal of Genomics* **2018**, *2018*, 2312987.

[138] R. Pace N, *ASM News* **1985**, *51*, 4-12.

[139] C. Simon, R. Daniel, *Applied and environmental microbiology* **2011**, *77*, 1153-1161.

[140] F. A. Prayogo, A. Budiharjo, H. P. Kusumaningrum, W. Wijanarka, A. Suprihadi, N. Nurhayati, *Journal of Genetic Engineering and Biotechnology* **2020**, *18*, 39.

[141] E. E. Kempa, J. L. Galman, F. Parmeggiani, J. R. Marshall, J. Malassis, C. Q. Fontenelle, J.-B. Vendeville, B. Linclau, S. J. Charnock, S. L. Flitsch, N. J. Turner, P. E. Barran, *JACS Au* **2021**, *1*, 508-516.

[142] M. Valancauskaite, G. W. Black, S. J. Charnock, *Access Microbiology* **2020**, *2*.

[143] J. R. Marshall, P. Yao, S. L. Montgomery, J. D. Finnigan, T. W. Thorpe, R. B. Palmer, J. Mangas-Sanchez, R. A. M. Duncan, R. S. Heath, K. M. Graham, D. J. Cook, S. J. Charnock, N. J. Turner, *Nature Chemistry* **2021**, *13*, 140-148.

[144] A. Lewin, A. Wentzel, S. Valla, *Current Opinion in Biotechnology* **2013**, *24*, 516-525.

[145] P. Narasingarao, S. Podell, J. A. Ugalde, C. Brochier-Armanet, J. B. Emerson, J. J. Brocks, K. B. Heidelberg, J. F. Banfield, E. E. Allen, *The ISME Journal* **2012**, *6*, 81-93.

[146] J. C. Venter, K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R.

Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H. Rogers, H. O. Smith, *Science* **2004**, *304*, 66-74.

[147] L. Pašić, B. Rodriguez-Mueller, A.-B. Martin-Cuadrado, A. Mira, F. Rohwer, F. Rodriguez-Valera, *BMC Genomics* **2009**, *10*, 570.

[148] C. L. Hemme, Y. Deng, T. J. Gentry, M. W. Fields, L. Wu, S. Barua, K. Barry, S. G. Tringe, D. B. Watson, Z. He, T. C. Hazen, J. M. Tiedje, E. M. Rubin, J. Zhou, *The ISME Journal* **2010**, *4*, 660-672.

[149] D. S. Jones, H. L. Albrecht, K. S. Dawson, I. Schaperdoth, K. H. Freeman, Y. Pi, A. Pearson, J. L. Macalady, *The ISME Journal* **2012**, *6*, 158-170.

[150] B. J. Baker, L. R. Comolli, G. J. Dick, L. J. Hauser, D. Hyatt, B. D. Dill, M. L. Land, N. C. VerBerkmoes, R. L. Hettich, J. F. Banfield, *Proceedings of the National Academy of Sciences* **2010**, *107*, 8806-8811.

[151] N. M. Mesbah, J. Wiegel, *Annals of the New York Academy of Sciences* **2008**, *1125*, 44-57.

[152] C. S. Chan, K.-G. Chan, Y.-L. Tay, Y.-H. Chua, K. M. Goh, *Front. Microbiol.* **2015**, *6*.

[153] W. Xie, F. Wang, L. Guo, Z. Chen, S. M. Sievert, J. Meng, G. Huang, Y. Li, Q. Yan, S. Wu, X. Wang, S. Chen, G. He, X. Xiao, A. Xu, *The ISME Journal* **2011**, *5*, 414-426.

[154] H. K. Kotlar, A. Lewin, J. Johansen, M. Throne-Holst, T. Haverkamp, S. Markussen, A. Winnberg, P. Ringrose, T. Aakvik, E. Ryeng, K. Jakobsen, F. Drabløs, S. Valla, *Environmental microbiology reports* **2011**, *3*, 674-681.

[155] G. Guillena, M. del Carmen Hita, C. Nájera, *Tetrahedron: Asymmetry* **2007**, *18*, 1272-1277.

[156] X.-Y. Xu, Y.-Z. Wang, L.-Z. Gong, *Organic Letters* **2007**, *9*, 4247-4249.

[157] K. Sato, M. Kuriyama, R. Shimazawa, T. Morimoto, K. Kakiuchi, R. Shirai, *Tetrahedron Letters* **2008**, *49*, 2402-2406.

[158] K. Zhang, Z. Pan, Z. Diao, S. Liang, S. Han, S. Zheng, Y. Lin, *Enzyme and Microbial Technology* **2018**, *110*, 8-13.

[159] A. Ghanem, *Tetrahedron* **2007**, *63*, 1721-1754.

[160] M. Haridas, C. Bisterfeld, L. M. Chen, S. R. Marsden, F. Tonin, R. Médici, A. Iribarren, E. Lewkowicz, P.-L. Hagedoorn, U. Hanefeld, E. Abdelraheem, *Catalysts* **2020**, *10*, 883.

[161] F. Hofmeister, *Archiv für experimentelle Pathologie und Pharmakologie* **1888**, *24*, 247-260.

[162] W. Kunz, J. Henle, B. W. Ninham, *Current Opinion in Colloid & Interface Science* **2004**, *9*, 19-37.

[163] S. J. Z. P. C. Arrhenius, **1887**, *1*, 631.

[164] https://upload.wikimedia.org/wikipedia/commons/thumb/3/3e/Deska_Franz_Hofmeister.jpg/1920px-Deska_Franz_Hofmeister.jpg.

[165] K. Brown, J. A. Mastrianni, *Journal of geriatric psychiatry and neurology* **2010**, *23*, 277-298.

[166] W. Kauzmann, in *Advances in Protein Chemistry, Vol. 14* (Eds.: C. B. Anfinsen, M. L. Anson, K. Bailey, J. T. Edsall), *Academic Press* **1959**, pp. 1-63.

[167] T. Alber, *Prediction of protein structure and the principles of protein conformation* **1989**, 161-192.

[168] R. S. Singh, R. R. Singhania, A. Pandey, C. Larroche, *Biomass, biofuels, biochemicals: advances in enzyme technology* **2019**, pp. 469-495.

[169] O. López-López, M. E. Cerdán, M. I. González-Siso, *Life* **2013**, *3*, 308-320.

[170] O. López-López, M. E Cerdan, M. J. C. P. I Gonzalez Siso, P. *Science* **2014**, *15*, 445-455.

[171] D. A. Cowan, *Comparative Biochemistry and Physiology Part A: Physiology* **1997**, *118*, 429-438.

[172] M. Dick, O. H. Weiergräber, T. Classen, C. Bisterfeld, J. Bramski, H. Gohlke, J. Pietruszka, *Scientific Reports* **2016**, *6*, 17908.

[173] R. Huber, T. A. Langworthy, H. König, M. Thomm, C. R. Woese, U. B. Sleytr, K. O. Stetter, *Archives of Microbiology* **1986**, *144*, 324-333.

[174] M. Ferrer, T. N. Chernikova, M. M. Yakimov, P. N. Golyshin, K. N. Timmis, *Nature Biotechnology* **2003**, *21*, 1267-1267.

[175] K. M. Polizzi, A. S. Bommarius, J. M. Broering, J. F. Chaparro-Riggers, *Current Opinion in Chemical Biology* **2007**, *11*, 220-225.

[176] B. J. Spix, M. Veurink, *Chemical and Thermal Unfolding,* https://www.mstechno.co.jp/techno/docs/applications/prometheus/Application%20Note%20NT-PR-009%20-%20Chemical%20and%20thermal%20unfolding.pdf.

[177] L. Hamborg, E. W. Horsted, K. E. Johansson, M. Willemoës, K. Lindorff-Larsen, K. Teilum, *Analytical Biochemistry* **2020**, *605*, 113863.

[178] F. Bleffert, V. Misetic, M. Jerabek, K.-E. Jaeger, F. Kovacic, www.nanotempert-technologies.com.

[179] Y. Cao, X. Li, J. Ge, *Trends in Biotechnology* **2021**, *39*, 1173-1183.

[180] J. M. Woodley, *Applied Microbiology and Biotechnology* **2019**, *103*, 4733-4739.

[181] H. Sakuraba, K. Yoneda, K. Yoshihara, K. Satoh, R. Kawakami, Y. Uto, H. Tsuge, K. Takahashi, H. Hori, T. Ohshima, *Applied and environmental microbiology* **2007**, *73*, 7427-7434.

[182] T. Aanniz, M. Ouadghiri, M. Melloul, J. Swings, E. Elfahime, J. Ibijbijen, M. Ismaili, *Brazilian Journal of Microbiology* **2015**, *46*, 443-453.

[183] H. L. Svilenov, T. Menzen, K. Richter, G. Winter, *Molecular Pharmaceutics* **2020**, *17*, 2638-2647.

[184] B. Zelent, K. A. Sharp, J. M. Vanderkooi, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2010**, *1804*, 1508-1515.

[185] K. Sasahara, H. Naiki, Y. Goto, *Journal of Molecular Biology* **2005**, *352*, 700-711.

[186] D. S. Goldberg, S. M. Bishop, A. U. Shah, *Journal of pharmaceutical Sciences* **2011**, *100 4*, 1306-1315.

[187] I. C. Guillermo Senisterra, and Masoud Vedadi, *Assay and drug development* **2012**, *10*, 128-136.

[188] T. Wu, J. Yu, Z. Gale-Day, A. Woo, A. Suresh, M. Hornsby, J. E. Gestwicki, Student Thesis **2020**.

[189] G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas', R. I. Gilmanshin, *Biopolymers: Original Research on Biomolecules* **1991**, *31*, 119-128.

[190] K. Gao, R. Oerlemans, M. R. Groves, *Biophysical Reviews* **2020**, *12*, 85-104.

[191] K. Huynh, C. L. Partch, *Current protocols in protein science* **2015**, *79*.

[192] J. Maillard, K. Klehs, C. Rumble, E. Vauthey, M. Heilemann, A. Fürstenberg, *Chemical Science* **2021**, *12*, 1352-1362.

[193] A. O. Magnusson, A. Szekrenyi, H.-J. Joosten, J. Finnigan, S. Charnock, W.-D. Fessner, *The FEBS Journal* **2019**, *286*, 184-204.

[194] nanoTemper Technologies www.nanotemper-technologies.de

[195] S. Wedde, C. Kleusch, D. Bakonyi, H. Groger, *Chembiochem : a European journal of chemical biology* **2017**, *18*, 2399-2403.

[196] F. Catanzano, A. Gambuti, G. Graziano, G. Barone, *The Journal of Biochemistry* **1997**, *121*, 568-577.

[197] S. Gihaz, D. Weiser, A. Dror, P. Sátorhelyi, M. Jerabek-Willemsen, L. Poppe, A. Fishman, ChemSusChem **2016**, *9*, 3161-3170.

[198] Y. Zhang, Y.-Q. Wu, N. Xu, Q. Zhao, H.-L. Yu, J.-H. Xu, *ACS Sustainable Chemistry & Engineering* **2019**, *7*, 7218-7226.

[199] C. M. Johnson, *Archives of Biochemistry and Biophysics* **2013**, *531*, 100-109.

[200] W. M. Jackson, J. F. Brandts, *Biochemistry* **1970**, *9*, 2294-2301.

[201] S. J. Gill, K. Beck, *Review of Scientific Instruments* **1965**, *36*, 274-276.

[202] P. L. Privalov, N. N. Khechinashvili, B. P. Atanasov, *Biopolymers: Original Research on Biomolecules* **1971**, *10*, 1865-1890.

[203] J. M. Sturtevant, *Annual review of biophysics and bioengineering* **1974**, *3*, 35-51.

[204] R. Biltonen, A. T. Schwartz, I. Wadso, *Biochemistry* **1971**, *10*, 3417-3423.

[205] A. Balmori, R. Sandu, D. Gheorghe, A. Botea-Petcu, A. Precupas, S. Tanasescu, D. Sánchez-García, S. Borrós, *Frontiers in Bioengineering and Biotechnology* **2021**, *9*, 650281.

[206] A. Fatkhutdinova, T. Mukhametzyanov, C. Schick, *International Journal of Molecular Sciences* **2022**, *23*, 2773.

[207] T. Instruments.

[208] J. F. Imhoff, *FEMS Microbiology Reviews* **1986**, *2*, 57-66.

[209] H. Santos, M. S. Da Costa, *Environmental Microbiology* **2002**, *4*, 501-509.

[210] M. F. Roberts, *Saline Systems* **2005**, *1*, 5.

[211] G. J. Gregory, E. F. Boyd, *Computational and Structural Biotechnology Journal* **2021**, *19*, 1014-1027.

[212] A. Taravati, M. Shokrzadeh, A. Ebadi, P. Valipour, A. T. M. Hassan, F. J. W. A. S. J. Farrokhi, World Appl Sci. J. **2007**, *2*, 353-362.

[213] S. Ajito, H. Iwase, S.-i. Takata, M. Hirai, *The Journal of Physical Chemistry B* **2018**, *122*, 8685-8697.

[214] N. E. Good, G. D. Winget, W. Winter, T. N. Connolly, S. Izawa, *ACS Publications* **1966**, *5*, 467-477.

[215] N. A. Kim, I. B. An, D. G. Lim, J. Y. Lim, S. Y. Lee, W. S. Shim, N.-G. Kang, S. H. Jeong, *Biological and Pharmaceutical Bulletin* **2014**, *37*, 808-816.

[216] N. Gheibi, A. A. Saboury, K. Haghbeen, A. A. Moosavi-Movahedi, *Journal of Biosciences* **2006**, *31*, 355-362.

[217] K. Kar, B. Alex, N. Kishore, *The Journal of Chemical Thermodynamics* **2002**, *34*, 319-336.

[218] R. L. Baldwin, *Biophysical Journal* **1996**, *71*, 2056-2063.

[219] X. Tadeo, B. López-Méndez, D. Castaño, T. Trigueros, O. Millet, *Biophysical Journal* **2009**, *97*, 2595-2603.

[220] M. Vrkljan, T. M. Foster, M. E. Powers, J. Henkin, W. R. Porter, H. Staack, J. F. Carpenter, M. C. Manning, *Pharmaceutical Research* **1994**, *11*, 1004-1008.

[221] J. K. Yadav, V. Prakash, *Journal of Biosciences* **2009**, *34*, 377-387.

[222] F.-F. Liu, L. Ji, L. Zhang, X.-Y. Dong, Y. Sun, *The Journal of Chemical Physics* **2010**, *132*, 225103.

[223] N. E. Good, S. Izawa, in *Methods in Enzymology, Vol. 24*, Academic Press, **1972**, pp. 53-68.

[224] W. J. Ferguson, K. I. Braunschweiger, W. R. Braunschweiger, J. R. Smith, J. J. McCormick, C. C. Wasmann, N. P. Jarvis, D. H. Bell, N. E. Good, *Analytical Biochemistry* **1980**, *104*, 300-310.

[225] Y. R. Gokarn, E. Kras, C. Nodgaard, V. Dharmavaram, R. M. Fesinmeyer, H. Hultgen, S. Brych, R. L. Remmele Jr, D. N. Brems, S. Hershenson, *Journal of Pharmaceutical Sciences* **2008**, *97*, 3051-3066.

[226] T. J. Zbacnik, R. E. Holcomb, D. S. Katayama, B. M. Murphy, R. W. Payne, R. C. Coccaro, G. J. Evans, J. E. Matsuura, C. S. Henry, M. C. Manning, *Journal of Pharmaceutical Sciences* **2017**, *106*, 713-733.

[227] A. W. Omta, M. F. Kropman, S. Woutersen, H. J. Bakker, *Science* **2003**, *301*, 347-349.

[228] K. D. Collins, G. W. Neilson, J. E. Enderby, *Biophysical Chemistry* **2007**, *128*, 95-104.

[229] G. J. S. Krestov, H. Energetics, New York, **1991**.

[230] M. Kaminsky, *Discussions of the Faraday Society* **1957**, *24*, 171-179.

[231] J. B. Robinson Jr, J. M. Strottmann, S. Stellwagen, *National Acad Sciences* **1981**, *78*, 2287-2291.

[232] H. D. B. Jenkins, Y. J. C. r. Marcus, Environmental Science **1995**, *95*, 2695-2724.

[233] J. C. Lee, S. N. J. B. Timasheff, *Archives of Biochemistry and Biophysics* **1974**, *13*, 257-265.

[234] E. S. Courtenay, M. W. Capp, C. F. Anderson, M. T. Record, *Biochemistry* **2000**, *39*, 4455-4471.

[235] A. Vaughan, *New Scientist* **2022**, *253*, 18-21.

[236] J. L. Ramos, B. Pakuts, P. Godoy, A. García-Franco, E. Duque, *Microbial Biotechnology* **2022**, *15*, 1026-1030.

[237] Y. V. Sheludko, W.-D. Fessner, *Current Opinion in Structural Biology* **2020**, *63*, 123-133.

[238] F. Hofmeister, *Archiv für experimentelle Pathologie und Pharmakologie* **1888**, *25*, 1-30.

[239] M. S. Weiss, G. J. Palm, C. Hilgenfeld, *Acta Crystallographica Section D* **2000**, *56*, 952-958.

[240] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, N. A. Baker, Protein Science **2018**, *27*, 112-128.

[241] B. V. Plapp, H. A. Charlier, S. Ramaswamy, *Archives of Biochemistry and Biophysics* **2016**, *591*, 35-42.

[242] https://pymol.org/.

[243] J. W. Bye, R. J. Falconer, *Protein Science* **2013**, *22*, 1563-1570.

[244] M. Ciolkowski, B. Pałecz, D. Appelhans, B. Voit, B. Klajnert, M. Bryszewska, *Colloids and Surfaces B: Biointerfaces* **2012**, *95*, 103-108.

[245] A. Blumlein, J. J. McManus, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013**, *1834*, 2064-2070.

[246] K. D. Collins, M. W. Washabaugh, *Quarterly Reviews of Biophysics* **1985**, *18*, 323-422.

[247] K. D. Collins, *Proceedings of the National Academy of Sciences* **1995**, *92*, 5553-5557.

[248] K. B. Rembert, J. Paterová, J. Heyda, C. Hilty, P. Jungwirth, P. S. Cremer, *Journal of the American Chemical Society* **2012**, *134*, 10039-10046.

[249] R. R. Burgess, in *Methods in Enzymology, Vol. 463* (Eds.: R. R. Burgess, M. P. Deutscher), Academic Press, **2009**, pp. 331-342.

[250] **2016**, *84*, A.3F.1-A.3F.9.

[251] K. C. Duong-Ly, S. B. Gabelli, in *Methods in Enzymology, Vol. 541* (Ed.: J. Lorsch), Academic Press, **2014**, pp. 85-94.

[252] A. A. Zavitsas, *Current Opinion in Colloid & Interface Science* **2016**, *23*, 72-81.

[253] T. Ghosh, A. Kalra, S. Garde, *The Journal of Physical Chemistry B* **2005**, *109*, 642-651.

[254] T. Fujita, H. Watanabe, S. Tanaka, *Chemical Physics Letters* **2007**, *434*, 42-48.

[255] H. Hamada, T. Arakawa, K. Shiraki, *Current Pharmaceutical Biotechnology* **2009**, *10*, 400-407.

[256] R. Mancinelli, A. Botti, F. Bruni, M. A. Ricci, A. K. Soper, *The Journal of Physical Chemistry B* **2007**, *111*, 13570-13577.

[257] S. J. Stuart, B. J. Berne, *The Journal of Physical Chemistry* **1996**, *100*, 11934-11943.

[258] Y. Cho, Y. Zhang, T. Christensen, L. B. Sagle, A. Chilkoti, P. S. Cremer, *The Journal of Physical Chemistry B* **2008**, *112*, 13765-13771.

[259] P. A. Bergstroem, J. Lindgren, O. Kristiansson, *The Journal of Physical Chemistry* **1991**, *95*, 8575-8580.

[260] M. Kugimiya, C. C. Bigelow, *Can J Biochem* **1973**, *51*, 581-585.

[261] H. I. Okur, J. Hladílková, K. B. Rembert, Y. Cho, J. Heyda, J. Dzubiella, P. S. Cremer, P. Jungwirth, *The Journal of Physical Chemistry B* **2017**, *121*, 1997-2014.

[262] C. Camilloni, A. Guerini Rocco, I. Eberini, E. Gianazza, R. A. Broglia, G. Tiana, *Biophysical Journal* **2008**, *94*, 4654-4661.

[263] H. Meuzelaar, M. R. Panman, S. J. A. C. Woutersen, *Angewandte Chemie* **2015**, *127*, 15470-15474.

[264] F. W. Studier, *Protein Expression and Purification* **2005**, *41*, 207-234.

Supplementary data

Chapter 1 Supplementary data

4.2 Characterization of first-generation supplementary glycine variants

4.2.1 Characterization of the homo-aldol addition activity with *n*-butanal (53)

Table 4 Results for the *n*-butanal (**53**) homo-aldol addition catalyzed by the supplementary first-generation mutants after 22 hours. Resulting educt and product concentrations, as well as conversion and product formation rates were calculated using the n-butanal (53) standard curve.

Mutant	Average peak value for *n*-butanal (counts)	Final concentration of *n*-butanal in mᴍ	*n*-butanal conversion in %	Average peak value for aldol product (counts)	Final concentration of aldol product in mᴍ	Product formation in %
T18A L20A	3342268	32.7	67.3 ± 16.1	724330	7.1	14.2 ± 0.3
T18A L20G	2838006	27.8	72.3 ± 2.6	641871	6.3	12.6 ± 0.1
T18G L20G	2539931	24.8	75.2 ± 11.0	255854	2.5	5.0 ± 0.2
T18G L20V	1651210	16.1	83.9 ± 7.4	69296	0.7	1.4 ± 0.1
T18G L20A	2069022	20.2	79.8 ± 3.5	67817	0.7	1.3 ± 0.2
F76G	4336700	42.4	57.6 ± 6.8	47453	0.5	0.9 ± 0.1

4.2.2 Characterization of the cross-aldol addition activity with acetone (50) and *n*-butanal (53)

Mutant	Final concentration *n*-butanal in mM	Concentration homo-aldol product in mM	Homo-aldol product formation in %	Final concentration acetone in mM	Concentration cross-aldol product in mM	Cross-aldol product formation in %
T18A L20A	10.4	2.9	5.8 ± 0.3	36.7	4.5	4.5 ± 0.5
T18A L20G	6.1	2.3	4.5 ± 1.1	41.6	7.3	7.3 ± 1.3
T18G L20A	9.1	1.0	2.0 ± 0.2	35.6	1.9	1.9 ± 0.8
T18G L20G	6.5	0.8	1.6 ± 0.1	37.1	3.9	3.9 ± 0.5
T18G L20V	10.4	1.3	2.6 ± 0.3	39.1	2.6	2.6 ± 0.5
F76G	9.5	0.9	1.7 ± 0.2	39.4	3.1	3.1 ± 0.1

Table 6 HPLC analysis of the cross-aldol reaction of acetone (**50**) and *n*-butanal (**53**) performed by mutants T18A/L20A, T18A/L20V, T18A/L20G and T18G/L20G with varying donor/acceptor concentrations.

Concentration *n*-butanal (acceptor) in mM	Concentration acetone (donor) in mM	Concentration homo-aldol product in mM	Homo-aldol product formation in %	Concentration cross-aldol product in mM	Cross-aldol product formation in %	Ratio cross-aldol to homo-aldol
T18A/L20A						
50	75	7.2	29.0	12.8	25.6	1,8
100	100	5.8	11.6	5.4	5.4	0,9
100	150	4.8	9.5	5.1	5.1	1,1
50	150	4.7	19.0	11.2	22.3	2,3
100	200	3.7	7.5	6.0	6.0	1,6
75	225	3.7	10.0	5.8	7.8	1,6
T18A/L20V						
50	75	5.4	21.4	15.5	30.9	2,9
50	100	4.6	18.6	17.1	34.2	3,7
100	100	12.1	24.2	10.0	10.0	0,8
100	150	4.0	8.0	17.9	17.9	4,5
50	150	12.4	49.5	12.6	25.2	1
100	200	9.0	17.9	11.4	11.4	1,3
75	225	6.8	18.3	15.6	20.8	2,3
T18A/L20G						

50	75	10.4	41.7	13.5	27.1	1.3
50	100	3.4	13.4	9.7	19.3	2.9
100	100	5.7	11.4	7.2	7.2	1.3
100	150	3.6	7.3	10.3	10.3	2.8
50	150	5.1	20.3	7.2	14.3	1.4
100	200	4.7	9.3	7.2	7.2	1.5
75	225	2.3	6.3	5.6	7.5	2.4
T18G/L20G						
50	75	3.0	11.9	6.4	12.7	2.1
50	100	2.9	11.6	6.7	13.4	2.3
100	100	3.2	6.4	5.9	5.9	1.8
100	150	1.7	3.4	4.7	4.7	2.7
50	150	3.5	14.1	5.8	1.2	1.6
100	200	1.8	3.5	4.6	4.6	2.6
75	225	2.6	6.9	6.6	8.9	2.6

4.2.3 Characterization of the cross-aldol addition activity with cyclobutanone (57)/cyclopentanone (58) and *n*-butanal (53)

Concentration *n*-butanal (acceptor) in mM	Concentration cyclopentanone (donor) in mM	Concentration homo-aldol product in mM	Homo-aldol product formation in %	Concentration cross-aldol product in mM	Cross-aldol product formation in %	Ratio cross- to homo aldol product formation
T18A/L20A						
50	75	2.9	11.5	16.2	32.3	2.8
50	100	2.8	11.2	17.3	34.6	3.1
100	100	2.0	4.0	15.0	15.0	3.7
100	150	5.0	10.0	14.4	14.4	1.4
50	150	0.6	2.5	7.7	15.3	6.2
100	200	2.9	5.8	8.8	8.8	1.5
75	225	1.3	3.6	11.3	15.0	4.2
T18A/L20V						
50	75	0.4	1.4	9.0	18.0	12.8
50	100	0.6	2.4	8.9	17.8	7.6
100	100	10.5	20.9	11.9	11.9	0.6
100	150	4.6	9.2	9.2	9.2	1
50	150	1.1	4.3	11.9	23.8	5.6

100	200	4.8	9.6	10.7	10.7	1.1
75	225	2.3	6.2	18.2	24.3	3.9
T18A/L20G						
50	75	2.4	9.5	9.5	17.7	1.9
50	100	1.7	6.8	6.8	10.6	1.6
100	100	4.5	8.9	8.9	11.3	1.3
100	150	3.6	7.1	7.1	12.4	1.7
50	150	2.1	8.4	8.4	17.3	2.1
100	200	4.0	7.9	7.9	12.4	1.6
75	225	2.0	5.2	5.2	13.0	2.5
T18G/L20G						
50	75	1.7	6.6	6.6	17.2	2.6
50	100	1.8	7.3	7.3	19.0	2.6
100	100	4.0	8.0	8.0	17.0	2.1
100	150	1.7	3.5	3.5	11.3	3.3
50	150	1.3	5.0	5.0	12.0	2.4
100	200	2.3	4.6	4.6	19.0	4.2
75	225	1.2	3.1	3.1	20.9	6.7

4.4 Homo-aldol screening results of the second-generation libraries

A.

Library F76G												
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	+	+	+	+	+++	+	+	-	-
B	-	-	+	-	+	+	+	++	++	-	-	-
C	-	+	+	+	+	+++	+++	++	++	+	+	+
D	-	-	+	-	+	+	+	+	-	-		+
E	+	+	+	++	+	++	+	+	++	++	+	+
F	-	-	-	-	+	-	-	-	-	+	+	+
G	-	-	-	-	-	++	+	+	+	-	+	-
H	++	+	+	+	+	+	++	++	+	++	++	-
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	+	+	+	+	+++	+	++	+	+/-
B	+	+	+	+	+	+	+	+++	+++	+	+	+/-

	1	2	3	4	5	6	7	8	9	10	11	12
C	+	+	++	++	+	+++	+++	+	+	+	+	+
D	-	-	-	+	++	++	+	+	+	+	+	+
E	+	+	+	++	+	+++	++	++	++	+++	+	++
F	-	-	-	-	-	+	+	++	-	++	+++	++
G	+	-	+	-	-	++	+	+	++	-	++	-
H	-	++	+	+	++	+	+	+	-	+	+	-

B.

Library F76S												
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+++	++	++	+++	++	+++	+	++	+	+	+
B	+	++	+	+	+	++	+	++	+	+	+++	+
C	+	-	+++	+	+	+	-	-	-	-	-	+
D	-	++	-	-	+	-	-	+	+	+++	+	+
E	+	+	+	-	+	-	+++	+	+	-	+	++
F	-	-	-	+	+	+	+	+++	++	+	-	+
G	-	-	+	-	+++	-	+	++	-	++	-	-
H	-	-	+	+	+	+	++	+	+	++	++	-
	1	2	3	4	5	6	7	8	9	10	11	12
A	++	+++	+	+	+++	+	++	-	+	+	+	-
B	+	++	++	+	+	+++	+	++	+	+	+++	+
C	-	-	+++	+	+	++	-	+	+	+	+	+
D	-	++	-	+	+	-	+	+	+	+++	+	+
E	+	+	+	+	++	-	+++	+	++	+	+	+++
F	+	+	+	+	+	+	+	+++	+++	+	-	+
G	+	-	+	++	+++	+	+	++	+	++	-	-
H	+	+	+	++	++	+	+	+	-	++	-	-

C.

Library F76V												
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	++	+	+	+	+	+	+	+	-
B	+	+	+	+	+	+	+	+	+	+	+	-
C	+++	+	+	+	+	+	+	++	-	++	++	+
D	++	+	+	+	+	++	++	+	+	+	+	+
E	-	+	+	+	+++	+	+	+	+	++	+	++
F	+	+	+	+++	+	+	++	+	+	+	+	++
G	+++	+	+	+	+	+	+	+++	+	+	+	-
H	+	+	+	+	++	+	-	-	+	-	+	-

	1	2	3	4	5	6	7	8	9	10	11	12
A	++	+	+	+++	++	++	+	+	+	+	++	+/-
B	+	+	+	+	++	+	-	+	+	+	+	-
C	+++	+	+	+	+	+	+	++	+	++	+	+
D	+++	+	+	+	+	++	++	+	+	+	+	+
E	+	+	++	+	+++	+	+	+	+	+++	++	++
F	+	+	+	+++	+	+	++	+	+	+	+	++
G	++	+	+	+	+	+	+	+++	+	+	+	-
H	+	-	-	+	+++	+	+	+	+	+	++	-

D.

Library F76A												
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	-	-	+	+	-	+++	++	-	+	-
B	+	+	++	+	+	+	++	++	+	+	+	-
C	++	++	+	+	+++	+	+	++	+	-	-	+
D	+	+	+		+	-	+	++	++	+	+	+
E	-	++	+	+	-	-	-	-	+	+	-	+++
F	+	-	-	+	++	++	-	-	+	-	-	+
G	+	+	+	+	+	+	+	+	+	+	+++	-
H	++	+++	-	++	-	+	-	+	-	+	-	-

	1	2	3	4	5	6	7	8	9	10	11	12
A	++	++	-	-	++	++	-	+++	++	-	++	-
B	+	++	++	+	+	-	++	+	+	-	+	-
C	+	-	-	+	++	++	-	+++	+	+	-	-
D	+	+	+	-	+	-	++	+++	++	++	++	+
E	-	+	+	+	-	-	-	-	++	++	+	++
F	++	-	-	+	+	++	-	-	-	-	+	++

Results of homo-aldol screening with *cy*anal shown in duplicate for: A – library F76G, B – library F76S, C – library F76V, D – library F76A.

4.5 Cross-aldol screening results of the second-generation libraries with acetone (50) as donor

A.

Library F76G												
	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	+	-	+	-	+++	-	-
B	-	-	-	-	-	-	-	+++	+++	-	-	-
C	-	++	-	+++	-	+	+++	-	-	-	-	+
D	-	-	-	-	+	-	-	-	-	-	-	+
E	++	++	-	+	-	+++	++	+	+	+++	+	+++
F	-	-	-	-	-	-	-	+++	-	++	+++	+++
G	-	-	-	-	-	+	-	-	+	-	+++	-
H	+	+++	-	-	-	++	-	-	-	++	-	-
10 % acetone	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	+	+	+	+	++	+	+++	+	+
B	+	+	+	+	+	+	+	+++	+++	+	+	+
C	+	++	+	+++	+	++	+++	+	+	+	++	+
D	+	+	+	+	++	+	+	+	++	+	++	+
E	++	+	+	++	+	+++	+	+	+	+++	+	+++
F	+	+	+	+	+	+	+	+++	+	+++	+++	+++
G	+	+	+	+	+	+++	+	+	+++	+	+++	-
H	+	+++	+	+	+	+	+	+	+	++	+	-

B.

Library F76S												
	1	2	3	4	5	6	7	8	9	10	11	12
A	-	+++	-	-	-	-	+++	-	+	+	-	-
B	-	-	++	-	-	+++	-	-	-	-	+++	-
C	-	-	++	-	-	-	-	-	-	-	-	+++

	1	2	3	4	5	6	7	8	9	10	11	12
D	-	++	-	-	-	-	-	-	-	+++	-	+++
E	-	-	-	-	-	-	+	-	-	-	-	+
F	-	-	-	-	-	-	+++	+	-	-	-	+
G	-	-	+	-	+++	-	+	+++	+	-	-	-
H	-	-	-	-	-	-	-	-	-	+++	-	-

10 % acetone	1	2	3	4	5	6	7	8	9	10	11	12
A	-	++	-	-	-	-	+	-	-	-	-	-
B	+	+	+++	-	-	+++	-	+	+	+	+++	-
C	-	-	+++	-	-	+	-	+	+	+	+	+++
D	-	+++	-	-	-	-	-	-	-	+++	-	+++
E	-	++	+	+	+	+	+++	-	+	+	+	+++
F	-	-	-	-	-	-	-	+++	+	-	-	+++
G	-	-	+	+	+++	-	+	+++	+	+	-	-
H	-	-	-	+	+	-	-	-	-	++	-	-

C.

Library F76V												
	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	++	-	-	-	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	-
C	+	-	-	-	-	-	-	-	-	-	-	+
D	-	-	-	-	-	-	-	-	-	-	-	+
E	-	-	-	-	++	-	-	-	-	++	-	+++
F	-	-	-	+	-	-	-	-	-	-	-	+++
G	-	-	-	-	-	-	-	-	+++	-	-	-
H	-	-	-	+	-	-	-	-	-	-	-	-

10 % acetone	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	+	+++	+	+	+	-	+	++	+	-
B	+	+	+	+	+	+	+	+	+++	++	+++	++
C	+++	+	+	+	+	+	+	+	++	+++	+++	++
D	++	++	+	+	+	+	+	+	+	+	+++	++
E	+	+	+	+	+++	+	+	+	+	+++	++	+++
F	+	+	+	+++	+	+	+	+	+	+	+	+++
G	+	+	+	+	+	+	+	+	+++	+	+	+/-
H	+	+	+	+	+++	+	+	+	+	+	+	-

226

D.

Library F76A												
	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	+	-	+++	-	-	-	-
B	-	-	+	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	-	-	+++	-	-	-	+
D	-	-	-	-	-	-	-	++	+	+++	+	+
E	-	-	-	-	-	-	-	-	-	+	-	+++
F	-	-	-	-	+++	-	-	-	-	-	-	+++
G	-	-	-	-	-	-	-	-	-	-	+++	-
H	+	++	-	+	-	++	-	-	-	-	+++	-

10 % acetone												
	1	2	3	4	5	6	7	8	9	10	11	12
A	+	+	-	-	+	+	-	+++	-	-	-	-
B	+	+	+++	+	+	-	+	+	+	+	+	+
C	-	-	-	-	+	-	-	+	-	-	-	++
D	-	-	-	-	-	-	-	++	+	++	+	++
E	-	-	-	-	-	-	-	-	-	-	-	+++
F	-	-	-	-	++	-	-	-	-	-	-	++
G	-	-	-	-	-	-	-	-	-	-	-	-
H	+	+++	-	-	-	+	-	-	-	-	++	+

Results of cross-aldol screening with acetone (**50**) as donor and *n*-butanal as acceptor shown as follows: A – library F76G, B – library F76S, C – library F76V, D – library F76A. The upper table depicts the screening results under equimolar donor:acceptor concentrations, while the lower table shows the screening with 10 % (v/v) of acetone (**50**).

4.7 Characterization of the selected second-generation variants

4.7.1 Characterization of the selected second-generation variants for activity in homo-aldol addition with *n*-butanal (53)

Table 12 Results for the selected second-generation mutants (master plate) in *n*-butanal (**53**) homo-aldol addition after 22 hours. Resulting educt and product concentrations, as well the product formation rates were calculated using the n-butanal (**53**) standard curve (see Section 3.2; 102276 counts/mM *n*-butanal (**53**)).

Mutant	Final concentration of homo-aldol product in mM	Homo-aldol product formation in %
T18/L20/F76S	5.5	10.9 ± 0.2
T18/L20V/F76S	4.8	9.5 ± 0.9
T18/L20/F76G	4.8	9.5 ± 0.9
T18S/L20/F76G	4.6	9.3 ± 1.3
T18S/L20G/F76G	3.8	7.6 ± 0.8
T18/L20/F76V	3.7	7.4 ± 0.3
T18S/L20/F76S	3.5	7.1 ± 1.7
T18/L20/F76A	3.5	6.9 ± 1.3
T18S/L20G/76A	2.9	5.8 ± 0.9
T18G/L20G/F76S	2.9	5.8 ± 0.7
T18S L20G/F76S	2.6	5.2 ± 0.3
T18S/L20/F76V	2.5	5.0 ± 0.7
T18S/L20/F76A	2.4	4.8 ± 0.8
T18G L20/F76V	2.3	4.6 ± 0.1
T18/L20G/F76A	2.1	4.1 ± 0.7
T18G/L20G/F76A	2.0	3.9 ± 0.9
T18S/L20G/F76V	1.9	3.9 ± 1.0
T18G/L20V/F76V	1.9	3.9 ± 0.8
T18/L20A/F76A	1.8	3.7 ± 0.5
T18A/L20/F76V	1.8	3.5 ± 0.4
T18G/L20A/F76V	1.5	3.1 ± 0.1

Table 13 Selected clones of the first-generation library and their respective homo-aldol product formation with whole cell suspensions and purified enzymes used as catalysts.

Mutant	Homo-aldol product formation with whole cells in %
T18A L20V	13.11
F76S	8.2
T18A L20A	8.1
T18A L20S	6.4
F76A	2.7

4.7.2 Characterization of the selected second-generation variants for activity in cross-aldol reactions with novel donors and *n*-butanal (53) as the acceptor

DERA-Variant	Final concentration of homo-aldol product in mM	Homo-aldol product formation in %	Final concentration of cross-aldol product in mM	Cross-aldol product formation in %
T18S/L20G/F76G	2.0	3.9 ± 0.4	21.5	21.5 ± 1.2
76A	1.9	3.7 ± 0.5	24.0	24.0 ± 0.9
T18S/L20G/F76A	2.4	4.8 ± 0.2	23.4	23.4 ± 0.7
T18G/L20G/F76A	1.5	3.0 ± 0.3	24.7	24.7 ± 2.2
T18S/F76A	0.9	1.7 ± 0.1	25.4	25.4 ± 0.2

5.5 NMR-assessment of the preparative scale reaction of *n*-butanal (53) and cyclopentanone (58)
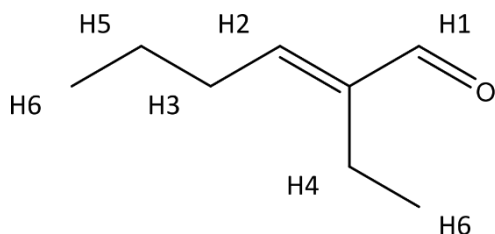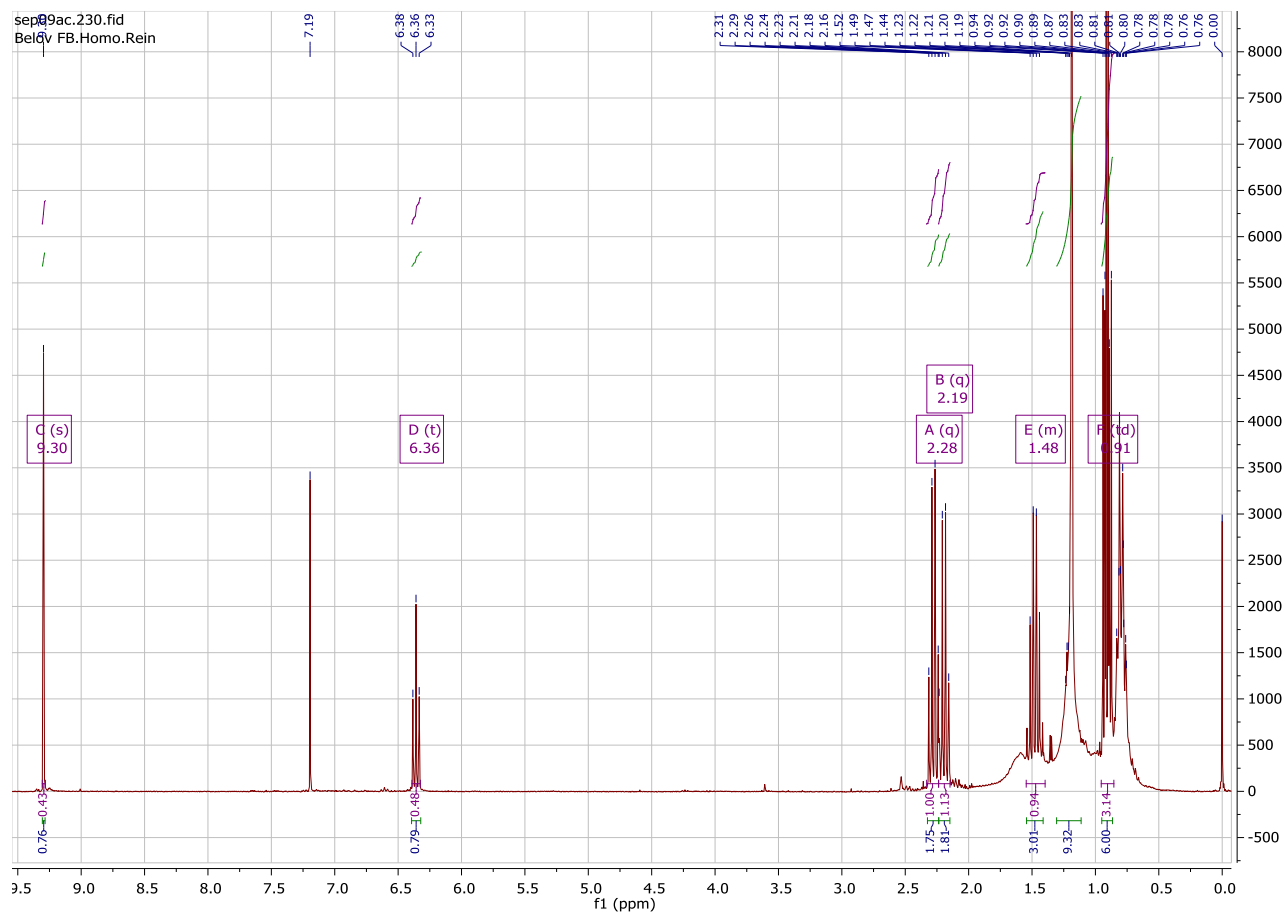
^{1}H NMR (300 MHz, Chloroform-*d*) δ 9.30 (s, 1H, H1), 6.36 (t, *J* = 7.5 Hz, 1H, H2), 2.28 (q, *J* = 7.4 Hz, 2H, H3), 2.19 (q, *J* = 7.5 Hz, 2H, H4), 1.55 – 1.40 (m, 2H, H5), 0.91 (td, *J* = 7.5, 4.8 Hz, 6H, H6).

Cross-aldol product
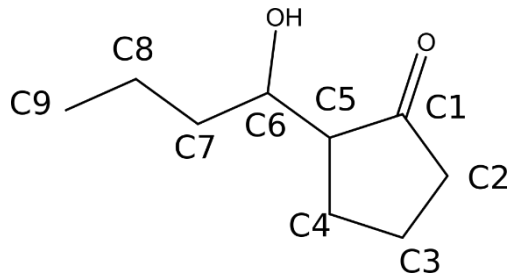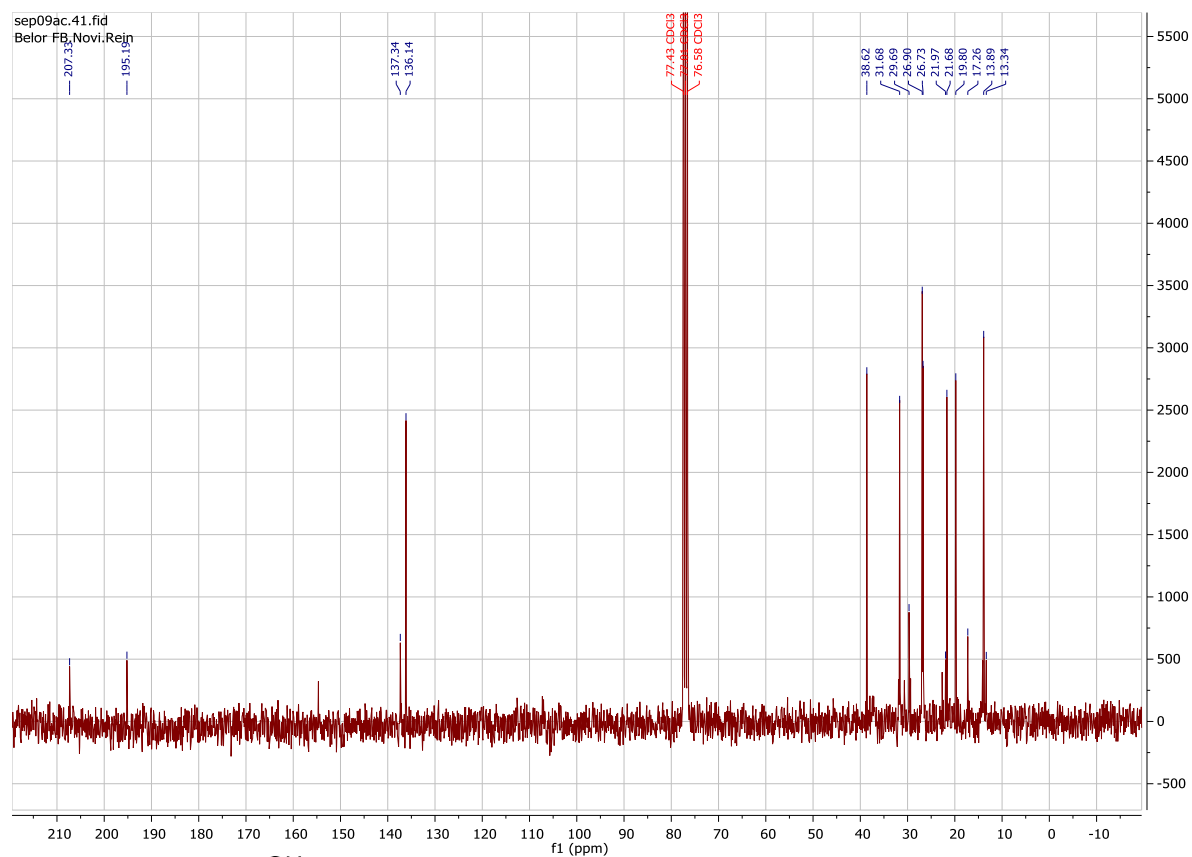
¹³C NMR (75 MHz, CDCl₃) δ 207.33 C1, 195.19, 137.34 C6, 136.14 C5, 77.43, 77.01, 76.58, 38.62 C2, 31.68 C7, 29.69, 26.90, 26.73, 21.97, 21.68 C3, 19.80, 17.26 C4, 13.89 C8, 13.34 C9

C (tt)
6.48

B (q)
4.05

H9

OH

H7

H8

H6 H5 H4

H3

H2

H1

O
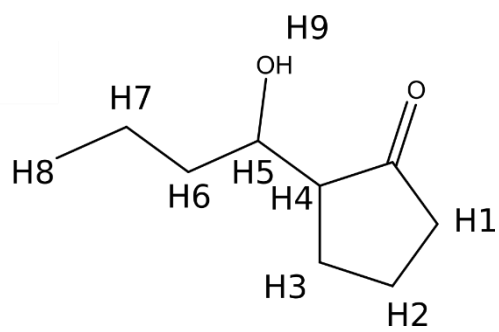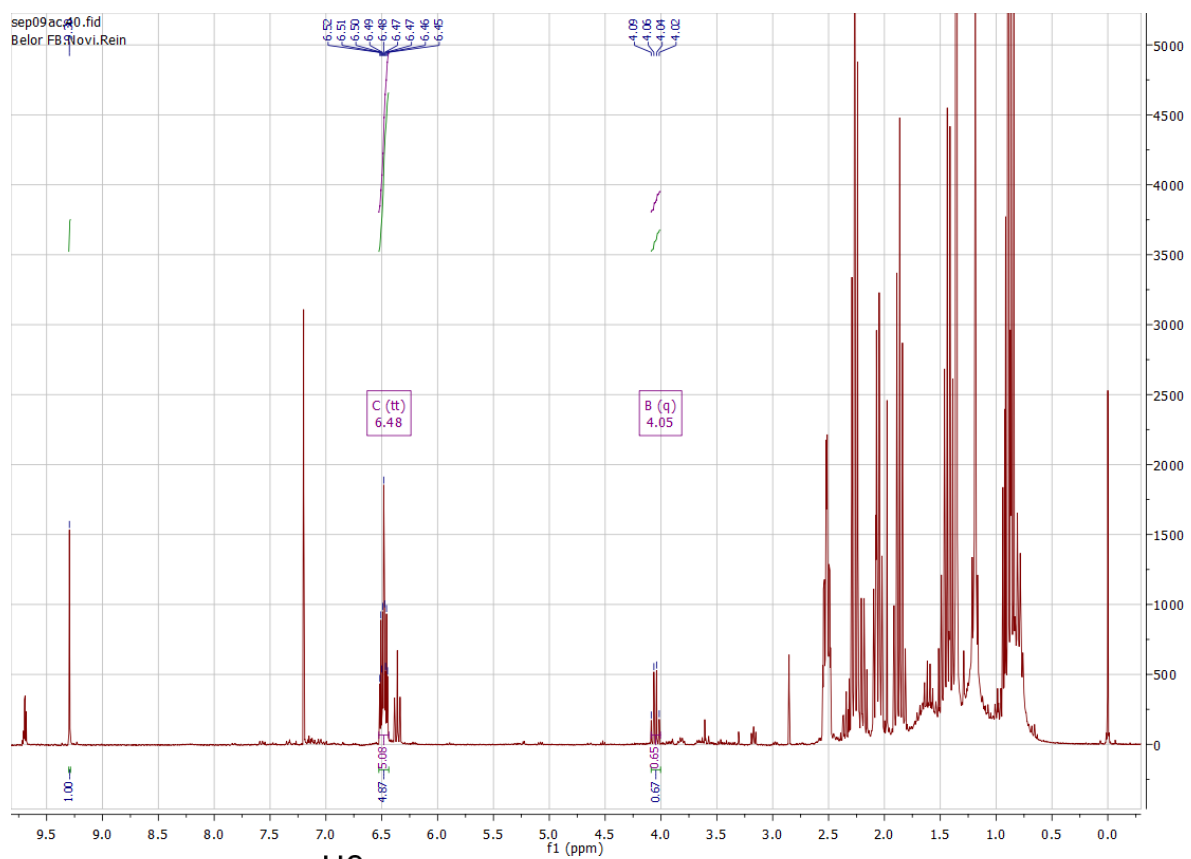
^{1}H NMR (300 MHz, Chloroform-*d*) δ 6.48 (tt, *J* = 7.6, 2.7 Hz, 6H, H1-H3), 4.05 (q, *J* = 7.1 Hz, 1H, H5).

Chapter 2 Supplementary data

4.1.1 Screening of metagenomic DERA library in cross aldol reaction of *i*-butyraldehyde (71) and cyclopentanone (58)

Table 1 Results of the HPLC-MS measurement of the metagenomic DERA library in the cross-aldol reaction of *i*-butyraldehyde (**71**) and cyclopentanone (**58**) with 1 : 1.5 ratio of acceptor to donor in 50 mм TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Enzyme	Product formation [%]			
	20h	44h	68h	144h
D067	3,3	5,9	6,8	6,0
D097	0,3	0,5	0,8	1,2
D112	2,6	4,1	4,8	4,6
D125	3,2	5,3	6,6	6,6
D126	0,5	0,9	1,1	1,0
D143	2,6	4,2	4,7	3,7
D144	0,8	1,5	2,2	2,5
D180	4,6	6,7	8,0	7,8
D184	2,3	3,4	4,2	4,4

Table 2 Results of the HPLC-MS measurement of the metagenomic DERA library in the cross-aldol reaction of *i*-butyraldehyde (**71**) and cyclopentanone (**58**) with 1 : 1.5 ratio of acceptor to donor in 50 mм TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Enzyme	Product formation [%]		
	20h	44h	120h
D062	0,3	0,6	0,9
D067	4,7	5,4	4,6
D097	0,7	1,3	2,5
D112	0,1	0,2	0,2
D125	7,5	9,5	9,1
D126	1,2	1,6	1,9
D143	5,0	6,5	5,3
D144	2,3	4,1	6,5
D180	11,0	18,1	19,5
D184	3,5	3,9	4,0
blind		0,0	0,0

4.1.2 Screening of metagenomic DERA library in cross aldol reaction of *i*-butyraldehyde (71) and cyclobutanone (57)

Table 3Results of the HPLC-MS measurement of the metagenomic DERA-library in the cross-aldol reaction of *i*-butyraldehyde (**71**) and cyclobutanone (**57**) with 1 : 5 ratio of acceptor to donor in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Enzyme	Product Formation [%]
D062	0,2
D067	1,8
D097	1,9
D125	6,1
D143	0,8
D144	6,5
D180	6,2
D184	2,6

4.1.3 Screening of metagenomic FSA library in cross-aldol reaction of *i*-valeraldehyde (73) and cyclopentanone (58)/cyclobutanone (57).

Table 4 Results of the HPLC-MS measurement of the metagenomic FSA-library in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**) with 1 : 5 ratio of acceptor to donor in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Enzyme	18h Product formation [%]
F001	11,9
F029	7,7
F036	3,2
F037	9,7
F042	20,1
F043	14,1
F046	10,1
F047	4,5
F049	46,5
F051	6,4
F057	31,7
F058	3,2
F061	28,5
F063	11,6
F071	15,8
F074	18,1
F075	15,2
F087	20,0
F094	2,8

Table 5 Results of the HPLC-MS measurement of the metagenomic FSA-library in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclobutanone (**57**) with 1 : 5 ratio of acceptor to donor in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Probe	18h Product formation [%]
F001	20,4
F029	17,1
F030	11,6
F032	26,5
F037	40,1
F039	4,0
F040	12,0
F041	5,5
F042	51,3
F043	36,6
F045	21,5
F047	9,2
F049	24,9
F051	10,0
F054	16,1
F055	12,4
F056	5,8
F057	60,2
F058	15,8
F059	15,1
F060	9,0
F061	37,3
F064	12,4
F071	36,0
F074	17,9
F075	45,3
F077	7,8
F078	10,8
F080	9,5
F081	24,1
F082	24,3
F083	22,9
F084	7,3
F085	14,6
F086	34,8
F087	43,8
F090	34,0
F091	9,7
F093	9,3
F094	5,0

F095	27,9
D125	96,9
D144	3,7
D180	117,8

Table 6 Results of the rescreening of active variants in cross aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**) in 50 mм TEA*HCl buffer (pH 7.5) with 10% v/v DMSO.

Enzyme	Mittelwert	Standard deviation
D067	2,8	0,1
D125	5,1	0,2
D126	2,9	0,1
D143	1,7	0,0
D144	7,5	0,5
D180	40,1	4,4
D184	2,6	0,3
DERATm	9,6	0,3
DERAEc	2,4	0,1
FSA F200I	1,0	0,9
FQ054	14,0	1,9
FQ059	16,8	2,4
FQ061	53,8	1,5
FQ2486	13,4	0,3
FQ2606	27,6	2,2
F042	9,0	0,0
F057	16,7	1,2
F061	14,0	0,8
F063	9,7	0,4
F071	7,4	0,3
F074	15,8	0,6
F075	6,2	0,2
F087	11,2	0,3

Table 7 Results of the rescreening of active variants in cross aldol reaction of *i*-valeraldehyde (**73**) and cyclobutanone (**57**) in 50 mM TEA*HCl buffer (pH 7.5) with 10% v/v DMSO.

Enzyme	Product formation [%]	Standardabweichung
*Tm*DERA	73,2	0,3
*Ec*DERA	59,1	0,5
*Ec*FSA	20,0	0,3
D067	35,7	0,5
D125	86,6	0,8
D144	6,1	0,1
D180	90,3	2,0
D184	86,2	1,0
FQ054	8,9	0,3
FQ059	17,7	0,1
FQ061	20,0	0,7
FQ2488	9,4	0,2
FQ2606	22,4	0,1
F042	38,6	1,0
F049	41,5	0,6
F057	63,4	1,6
F061	50,3	1,5
F063	22,3	0,2
F071	68,5	0,5
F074	29,5	0,2
F075	64,8	1,7
F087	47,4	0,7

4.1.4   Screening of metagenomic FSA library (D6Q mutations) in cross-aldol reaction of *i*-valeraldehyde (73) and cyclopentanone (58).

Table 7 Results of the HPLC-MS measurement of the metagenomic FSA-library (D6Q mutation) in the cross-aldol reaction of *i*-valeraldehyde (**73**) and cyclopentanone (**58**) with 1 : 5 ratio of acceptor to donor in 50 mᴍ TEA*HCl buffer (pH 7.5) with 10% v/v DMSO cosolvent.

Enzyme	18h Product formation [%]
D125	8,5
D126	3,5
D143	4,1
D144	6,4
D180	62,0
D184	2,7
FQ048	3,5
FQ054	22,2
FQ057	10,8
FQ059	22,0
FQ063	11,9
FQ067	2,8
FQ085	1,7
FQ087	2,9
FQ094	5,2
FQ2486	16,9
FQ2550	3,6
FQ2558	14,7
FQ2606	43,6
FQ061	67,0
FQ061AA	5,3
FQ061AV	5,8
FQ2591	8,2
FQ2591AA	15,3

4.2.1 Screening of selected DERA and FSA enzymes for activity with fluoroacetone (77)

Table 8 Screening results of the metagenomic DERA/FSA libraries in cross aldol addition of *i*-valeraldehyde (**73**) and fluoroacetone (**77**) based on TLC analysis (2:1 cyclohexane/ethylacetate) with anisaldehyde staining (bottom).

Enzyme	Product formation [%]
D29	68,9
D62	73,1
D126	63,1
D135	50,2
D143	33,0
D144	58,2
D147	68,7
D190	69,6
D192	70,1
F49	56,6
F63	53,1
F71	76,4
F75	74,3
Q54	70,3
Q61	81,5

4,4-Dimethoxybutan-2-ol (68) (1H + 13C NMR)
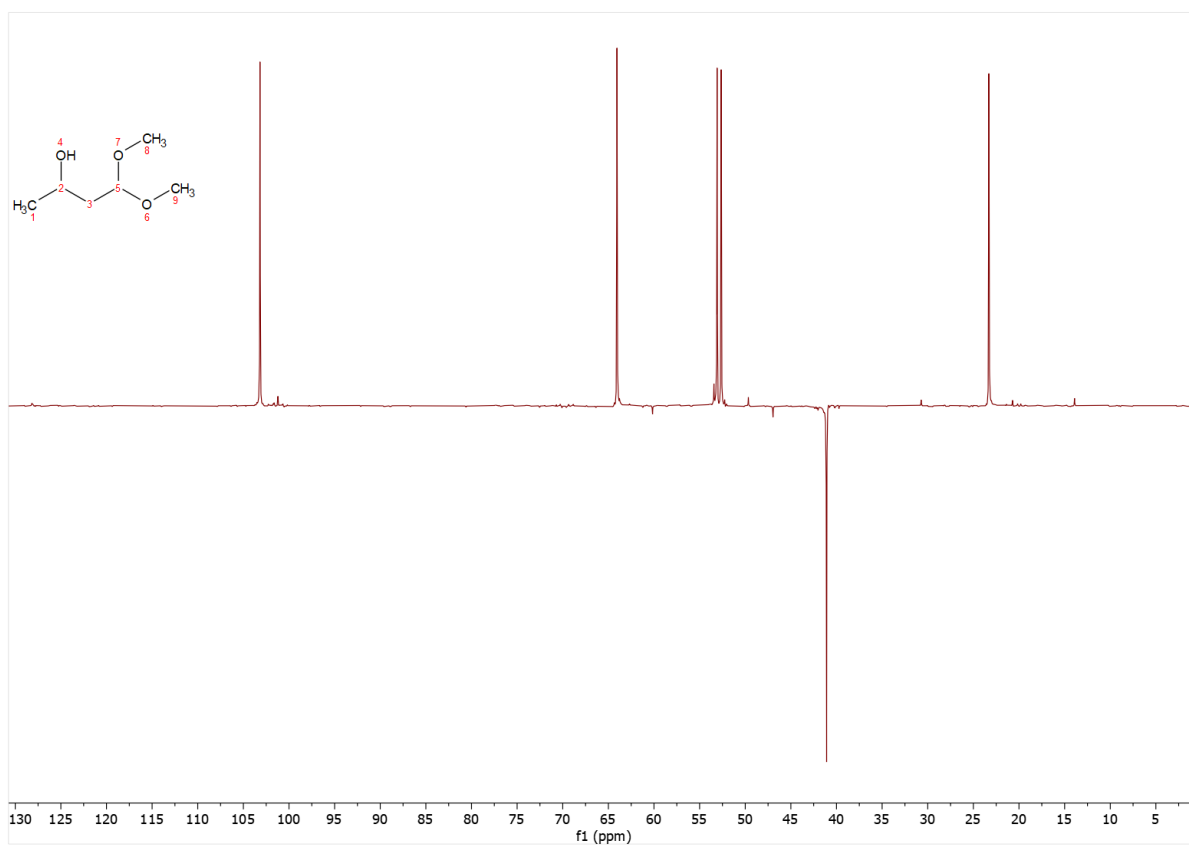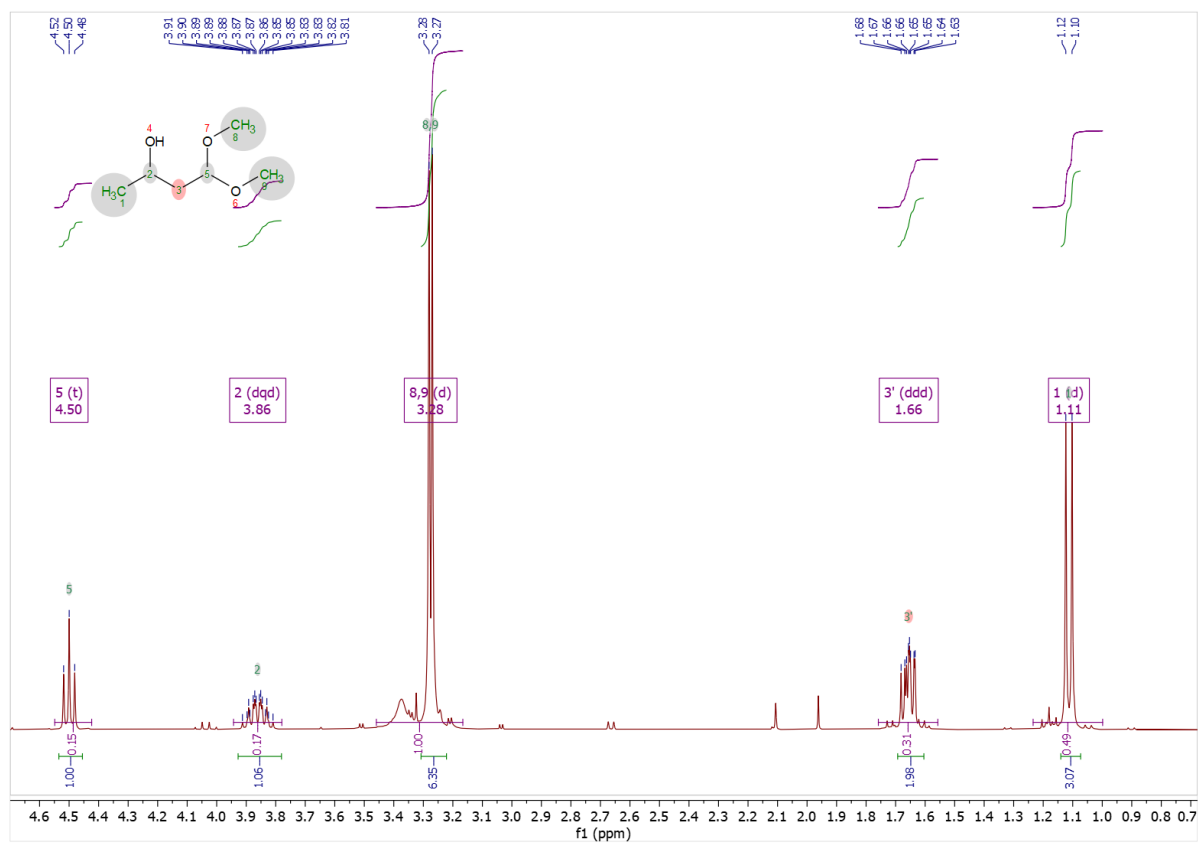
4,4-Dimethoxy-2-methylbutan-2-ol (97) (1H + 13C NMR)
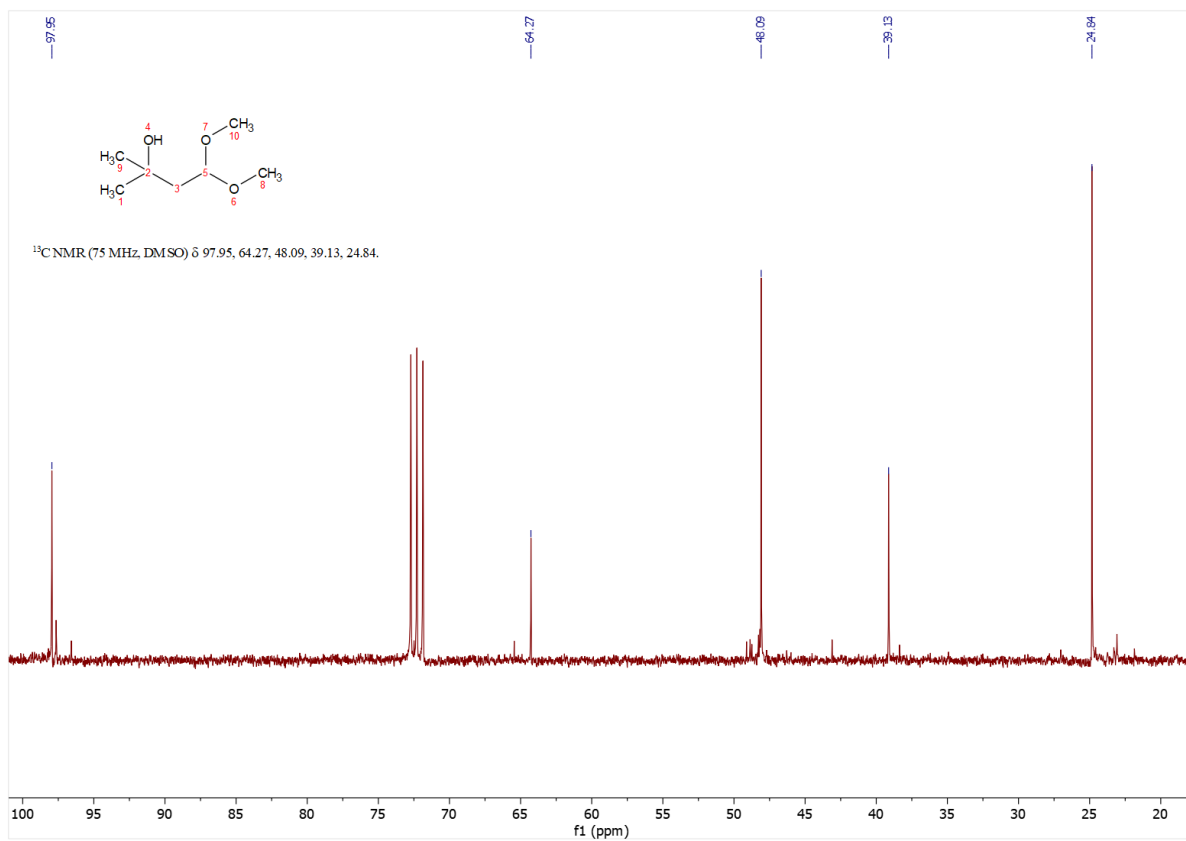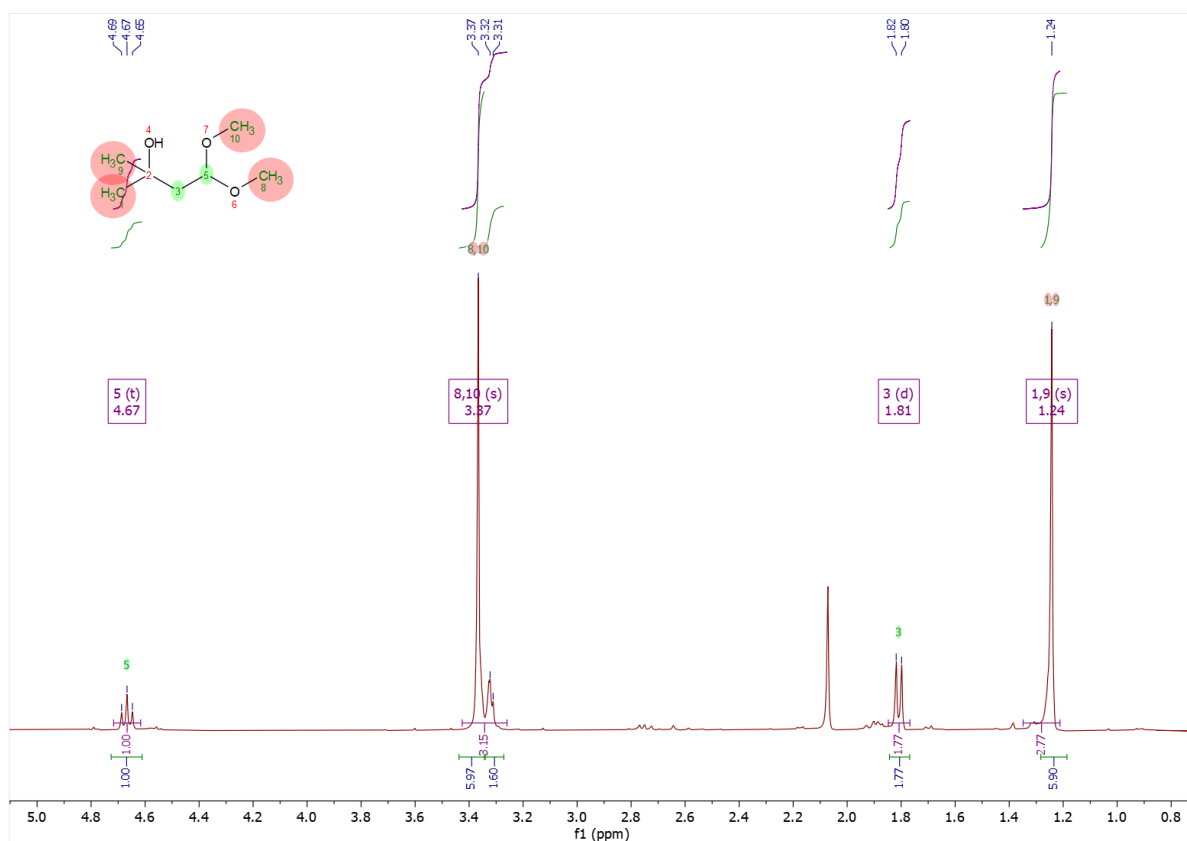
¹³C NMR (75 MHz, DMSO) δ 97.95, 64.27, 48.09, 39.13, 24.84.

4.3.1 Screening of selected DERA and FSA enzymes for activity with cyclopentanone (58) as donor

Diastereomeric mixture
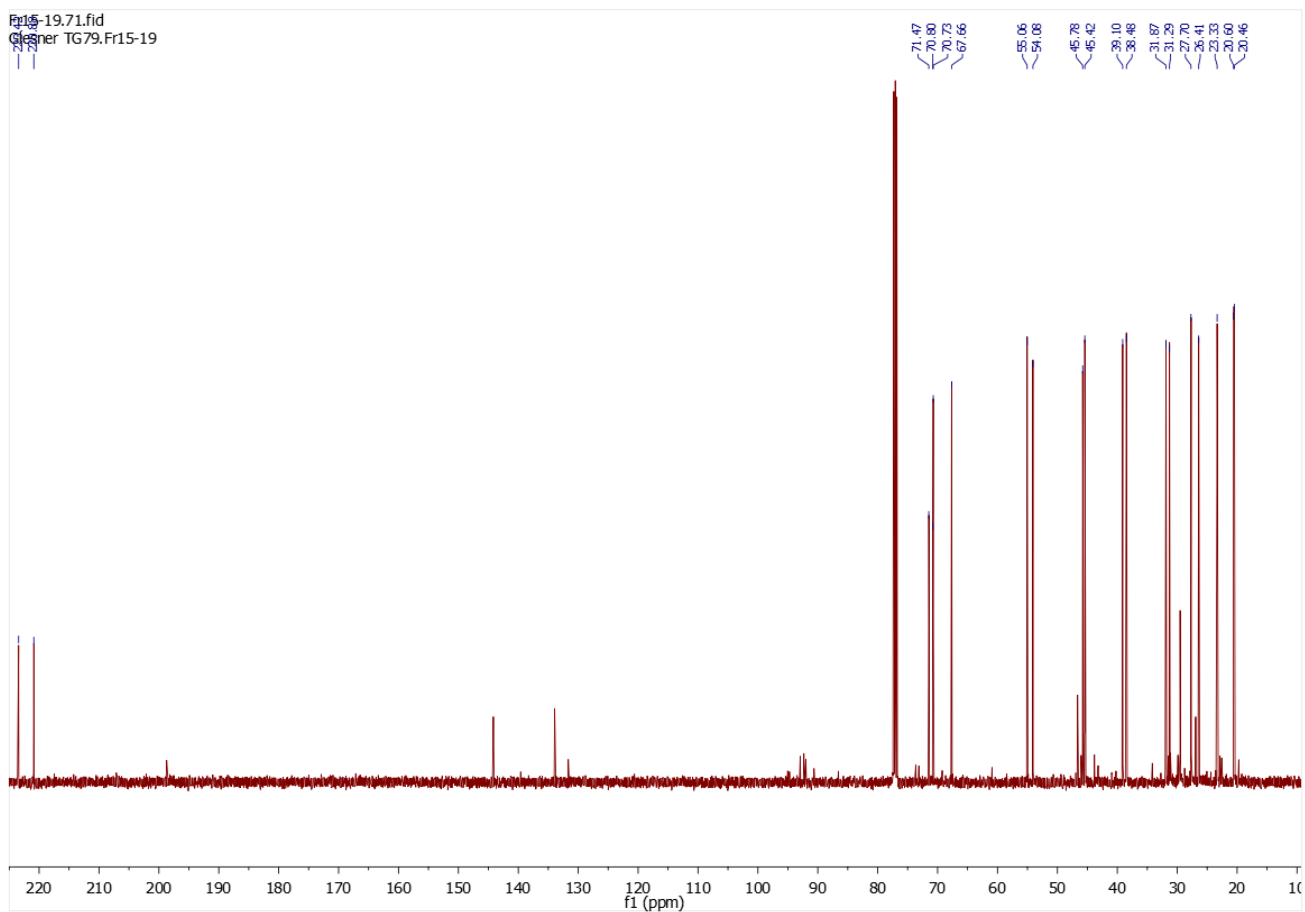
Abstract Deutsch

Aldolasen sind Enzyme, die eine entscheidende Rolle im Stoffwechsel von Kohlenhydraten spielen. Diese Enzyme gehören zu der Klasse der Lyasen und zeigen insbesondere in Aldolreaktion katalytische Aktivität, bei welcher Aldehyde oder Ketone mit einem Elektrophil reagieren und so eine Aldolbindung bilden.

Trotz der beeindruckenden Errungenschaften in der Entwicklung von Aldolasen, wurden diese bisher nur sehr spezifisch eingesetzt, wie z.B. für die Synthese von Building Blocks für die industrielle Produktion von Statinen für den Fall von DERA. Die starke Restriktion bei Aldolasen bezüglich der akzeptierten Donoren für die Aldolreaktion schränkte die breitflächige Verwendung jedoch ein.

Durch die geometrische Analyse des aktiven Zentrums auf Basis vorhandener Kristallstrukturen ist es möglich die katalytische Aktivität von Aldolasen mittels Mutagenese zu erweitern. Hierbei ist das Ziel den vorhandenen Platz in dem aktiven Zentrum zu erweitern und damit das Substratspektrum zu erweitern. In dieser Arbeit wurde als Ausgangsaldolase *Ec*DERA verwendet, welche schon eine Einzelpunktmutation (F200I) auf der Oberfläche trägt. Durch Analyse des aktiven Zentrums auf Basis vorhandener Kristallstrukturen wurden Enzymbibliotheken generiert, welche bis zu drei weitere Mutationen tragen. Die eingeführten Mutationen befinden sich alle im aktiven Zentrum mit dem Ziel den vorhandenen Platz zu erweitern durch Einführung von „platzsparenden" Mutationen mit kleineren Aminosäuren.

*Ec*DERA akzeptiert als Wildtyp Acetaldehyd, Aceton, Propanal und Fluoroaceton als Donor und zeigt damit ein entspanntes Donorspektrum im Vergleich zu anderen erforschten Aldolasen. Durch die Einführung der Mutationen im aktiven Zentrum sollten zyklische Ketone wie Cyclobutanon, Cyclopentanon und Cyclohexanon akzeptiert werden. Zusätzlich wurden weitere Ketone wie Chloraceton, Butanon, 1-Hydroxybutanon und weitere auf Aktivität getestet. Die Hypothese, dass zyklische Ketone akzeptiert werden können bestätigte sich. Einige der modifizierten Enzyme der mutagenen *Ec*DERA-Bibliothek waren in der Lage Cyclopentanon als Donor umzusetzen, jedoch zeigte keiner der Enzymvarianten Aktivität für Cyclobutanon und Cyclohexanon.

Aufgrund des limitierten Erfolgs der mutagenen *Ec*DERA-Bibliothek und der Verfügbarkeit der metagenomischen Enzymbibliotheken DERA und FSA unseres Partners Prozomix®, wurden diese verwendet um aktive Enzyme für die Auswahl der getesteten Donoren zu finden. Durch die Verwendung metagenomischer Enzymbibliotheken konnte das aus der Literatur bekannte Donorenspektrum für DERA erweitert werden. Als neue bisher unbekannte Aktivitäten für DERA-Enzyme konnte *i*-Valeraldehyd als Akzeptor mit Butanon, Methoxyaceton und Chloraceton umgesetzt werden. Bemerkenswert ist hierbei die Verwendung eines aliphatischen Akzeptors anstatt des natürlichen Akzeptors D-Glyceraldehyde-3-phosphat. Die in der Literatur präsentierten Beispiele sind hauptsächlich mit D-Glyceraldehyde-3-phosphat generiert worden, wobei D-Glyceraldehyde-3-phosphat als natürlicher Akzeptor die verminderte Aktivität unnatürlicher Donoren kompensiert. Somit stellen die präsentierten

Ergebnisse einen großen Schritt zur universellen, synthetischen Verwendung von Aldolasen für die Produktion von Feinchemikalien dar.

Das dritte Kapitel der vorliegenden Arbeit beschäftigt sich mit der Wirkung von anorganischen Salzen auf die Stabilität von Enzymen. Im Jahr 1888 wurde von Franz Hofmeister die Publikation "Zur Lehre der Wirkung der Salze" seine Ergebnisse publiziert, welche sich mit dem Aussalzverhalten von proteinhaltigen Lösungen unter Einfluss von anorganischen Salzen beschäftigt hat. Für sein Experiment verwendete Hofmeister separierte Flüssigkeit von geschlagenen Eiern, welche sich über Nacht absetzte. Nach Zugabe konzentrierter Salzlösungen beobachtete er die unterschiedlichen Zeitspannen die sich ergaben, bis erste Zeichen von ausgefallenem Protein sichtbar waren in Form einer weißlichen Trübung. Seine Ergebnisse resultierten in der bis heute relevanten „Lyotropen Serie" oder auch umgangssprachlich Hofmeister-Serie.

Für die vorliegende Arbeit wurde die Beeinflussung der Thermostabilität durch anorganische Salze und weiterer Verbindungen von drei Beispielenzymen analysiert. Die verwendeten Enzyme umfassen Lysozym von *Gallus gallus*, Alkoholdehydrogenase von *Sachharomyces cerevisiae* und FSA von *Escherichia coli*. Für die Bestimmung der Thermostabilität wurde nanoDSF Equipment eingesetzt für die Bestimmung des Schmelzpunkts der Enzyme in wässriger Lösung. Die Methodik erlaubt es den Schmelzpunkt von insgesamt 48 wässrigen Proteinproben basierend auf der intrinsischen Fluoreszenz von Tryptophan in einem Zeitrahmen unterhalb einer Stunde abhängig von verwendeten Temperaturgradienten zu bestimmen. Der verwendete Literaturstandard liegt bei 1°C/Minute. Mit dem verwendeten Gerät „Prometheus" von NanoTemper kann die Probe bis zu 115°C erhitzt werden. Die Methodik ist somit geeignet für Hochdurchsatz-Screenings von (hyper)thermophilen Enzymbibliotheken.

Die Ergebnisse der vorliegenden Arbeit bestätigen den Verlauf der Hofmeister-Serie. Mit hohen Salzkonzentrationen konnte der Trend der Hofmeister Serie für alle getesteten Enzyme beobachtet werden. Besonders der destabilisierende Anteil der Serie zeigte einheitliches Verhalten zu den Ergebnissen von Hofmeisters Experiment.

Abstract English

Aldolases are enzymes that play a crucial role in the metabolism of carbohydrates. These enzymes belong to the class of lyases and exhibit catalytic activity, especially in the Aldol reaction, where aldehydes or ketones react with an electrophile to form an aldol bond.

Despite impressive achievements in the development of aldolases, their application has been highly specific, such as in the synthesis of building blocks for statin synthesis, as seen in the case of DERA. The significant restriction regarding accepted donors for the aldol reaction in aldolases has limited their widespread use in industrial applications.

Geometric analysis of the active center based on existing crystal structures allows for the expansion of aldolase catalytic activity through mutagenesis. The goal is to enlarge the existing space in the active center. In this study, *Ec*DERA enzyme was used as the starting aldolase, which already carries a single-point mutation (F200I) on the surface. Enzyme libraries with up to three additional point mutations in the active center were generated based on the analysis of existing crystal structures. The introduced mutations, all located in the active center, aim to expand the available space by introducing "space-saving" mutations with smaller amino acids.

Wild-type *Ec*DERA accepts acetaldehyde, acetone, propanal, and fluoroacetone as donors, demonstrating a relaxed donor spectrum compared to other known aldolases. The introduction of mutations in the active center aims to enable the acceptance of cyclic ketones such as cyclobutanone, cyclopentanone, and cyclohexanone. Additionally, other ketones like chloroacetone, butanone, 1-hydroxybutanone, and many others were tested for activity. The hypothesis that cyclic ketones could be accepted was confirmed. Some enzymes from the mutagenic *Ec*DERA library were able to convert cyclopentanone, but none of the enzyme variants showed activity for cyclobutanone and cyclohexanone.

Due to the limited success of the mutagenic *Ec*DERA library and the availability of metagenomic enzyme libraries DERA and FSA from our partner Prozomix®, these were used to find active enzymes for the selection of donors. The use of metagenomic enzyme libraries expanded the known donor spectrum for DERA by three donors. New, previously unknown activities for DERA enzymes included *i*-valeraldehyde as an acceptor with butanone, methoxyacetone, and chloroacetone. Notably, an aliphatic acceptor was used, unlike the examples presented in the literature, mainly generated with D-glyceraldehyde-3-phosphate, where the natural acceptor compensates for the lower activity with unnatural donors. Thus, the results represent a significant step towards the universal application of aldolases for the production of fine chemicals.

The third chapter of this thesis deals with the effects of inorganic salts on enzyme stability. In 1888, Franz Hofmeister published the results of his study "Zur Lehre der Wirkung der Salze" on the salting-out behavior of protein-containing solutions under the influence of inorganic salts. For his experiment, Hofmeister used separated liquid from beaten eggs, which settled overnight. After adding concentrated salt solutions, he observed different time spans until the

first signs of precipitated protein appeared in the form of a whitish cloudiness. His results led to the still-relevant "Lyotropic Series" or colloquially known as the Hofmeister series.

For this study, the influence of inorganic salts and other compounds on the thermostability of three example enzymes was analyzed. The enzymes used include lysozyme from *Gallus gallus*, alcohol dehydrogenase from *Saccharomyces cerevisiae*, and FSA from *Escherichia coli*. NanoDSF equipment was used to determine the melting point of enzymes in aqueous solution for assessing thermostability. The methodology allows determining the melting point of a total of 48 aqueous protein samples based on intrinsic tryptophane fluorescence within an hour, depending on the temperature gradients used. The literature standard for the temperature gradient is 1°C/minute, and the Prometheus device from NanoTemper can heat the sample up to 115°C, making the methodology suitable for high-throughput screenings of (hyper)thermophilic enzyme libraries.

The results of this study confirm the Hofmeister series from 1888. With high salt concentrations, the trend of the Hofmeister series was observed for all tested enzymes. Particularly, the destabilizing portion of the series exhibited consistent behavior with Hofmeister's results.