# Chemoenzymatic Synthesis of Carbohydrates and Derivatives with Engineered D-Fructose-6-Phosphate Aldolase

Anna Szekrényi

**ADVERTIMENT**. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA**. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING**. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

# Chemoenzymatic Synthesis of Carbohydrates and Derivatives with Engineered D-Fructose-6-Phosphate Aldolase

Instituto de Química Avanzada de Cataluña (IQAC)
Consejo Superior de Investigaciones Científicas (CSIC)

Facultad de Farmacia, Universidad de Barcelona
Departamento de Bioquímica y Biología Molecular, Programa de Biotecnología

**Anna Szekrényi**
**2014**

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS (CSIC)
INSTITUTO DE QUÍMICA AVANZADA DE CATALUÑA (IQAC)

UNIVERSIDAD DE BARCELONA
Facultad de Farmacia

Departamento de Bioquímica y Biología Molecular
Programa de Doctorado de Biotecnología

# Chemoenzymatic Synthesis of Carbohydrates and Derivatives with Engineered D-Fructose-6-Phosphate Aldolase

Memoria presentada por Anna Szekrényi para optar al Título de Doctor
por la Universidad de Barcelona

| | |
|---|---|
| Dirigida por: | Tutor: |
| Dr. Pere Clapés Saborit | Dra. Josefa Badía Palacín |
| | |
| Codirectores: | |
| Dr. Jordi Bujons Vilàs | Dr. Jesus Joglar Tamargo |

Doctoranda:
Anna Szekrényi

**Anna Szekrényi, 2014**

**“I don’t know anything,**
**but I do know that everything is interesting**
**if you go into it deeply enough.”**

**Richard Feynman**
**Nobel Laureate in Physics (1965)**

# Acknowledgments

*When writing this thesis I had to think through those four years that I've worked in the Biotransformation and Bioactive Molecules Group, and I realized how grateful I am to all those people with whom I was surrounded. I would like to thank all of them this way for their help and kindness.*

*First I would like to express my sincere gratitude to my supervisor Dr. Pere Clapés for his continuous support and guidance throughout my thesis, to give me the opportunity to form part of his group, and for helping me discover the discipline of biocatalysis. Furthermore I greatly appreciate the help I received during these years from Dr. Jesús Joglar, Dr. Jordi Bujons and Dr. Josep Lluís Torres.*

*I would like to thank all my colleagues in the lab for sharing these four years with me, with all its great moments, diners, "raves" and the little things that made the CSIC a great place to work at. The first must be Mariana who showed me the lab and kept me company in the beginning. I am really grateful for her endless kindness and patience. I thank Xavi for being my mentor during a couple of years, to Karel whose pieces of advice showed his infinite knowledge and for always being there, whenever I needed to be cheered up a little, to Sara, for filling all our lives with energy and with her love of volley. I would like to thank Gerard and Roman their help and advice with my doubts in biology. I would like to acknowledge all my colleagues from the labs 209, 207 and 206. Esther and Carles for being the motors of the group, continuously organizing something to do. To Anna S., Raquel, Alda, Susana, Sara R., Chris, Aris, Anna C., Jara, Marisa, Jordi P. Dani and Bruno, and all the people who came and went by like Jenni, Pedro and Isa. I really appreciate the friendship of Liva and Maria, who were always there for me. I will miss all of you and I will keep in my memory all that time we spent together.*

*I would like to thank to Professor Fessner for giving me the opportunity to work with his group and in his laboratories in Darmstadt, to all his group for making me feel home, and to Deniz for being so welcoming, enthusiastic and for our chitchats. To Professor Pericàs and Dr. Sonia Sayalero for helping me with my work in Tarragona.*

## Acknowledgements

*I would like to thank my family and friends who brought me little pieces of home during their numerous visits to Barcelona, and to Guillaume for those great weekends and his support during those years.*

*And finally I would like to show my most sincere appreciation to my family, to Kati, Apu és Anyu for everything. It is impossible to express how grateful I am for your love and for all you have done for me.*

*I really would like to thank all of you for your support and encouragement! GRACIAS A TODOS!*

# Summary

Biocatalysis gained importance in the latter half of the 20th century with high potential for application in pharmaceutical, food or cosmetics industries, and is finding increasing acceptance among the organic chemists, since with enzymes molecular complexity can be rapidly built without the need of protection and deprotection steps, with extremely high stereoselectivity. A particularly important class of enzymes is the aldolases, which catalyze the C-C bond formation between a nucleophilic donor and an electrophilic acceptor substrate. These biocatalysts have been widely exploited in the asymmetric synthesis of complex polyfunctional carbohydrates. A recently discovered group of aldolases the DHA dependent aldolases have been widely used in the synthesis of carbohydrates and iminocyclytols, due to their unique ability to accept non-phosphorylated donors.

Unfortunately many aldolases show narrow substrate scope, or insufficient stability under operating conditions. With the development of enzyme engineering these problems can now be addressed. In this thesis enzyme engineering plays a central role, since structure guided redesign by site-directed and site-saturation mutagenesis was used for the modification of the D-fructose-6-phosphate aldolase from *E. coli* (FSA). These strategies have been exploited in engineering the tolerance of FSA for a wide range of donor and acceptor substrates.

In **Chapter 3.1.** a library of FSA mutants was developed by the combination of site-saturation and site-directed mutagenesis to increase the catalytic efficiency of FSA towards glycolaldehyde as donor substrate in cross-aldol reactions. The A129 residue was targeted for modification and it lead to a mutant with 20-fold increased activity in the self-aldol reactions. This mutation in combination with others in the active site led to the variant FSA L107Y/A129G/A165G with highly improved acceptor tolerance. The library of mutants showed unmatched selectivity towards glycolaldehyde as donor substrate in the competence experiments with equimolar mixtures of glycolaldehyde and dihydroxyacetone or hydroxyacetone. To test the best variants' synthetic applicability the enzymes were also assayed in cross-aldol reactions with *N*-Cbz-aminoaldehydes, and aldol-adducts were isolated with 80-98% yields.

In **Chapter 3.2.** the FSA mutants with different substitutions in the Ala 129 position were used in the synthesis of rare sugars from achiral prebiotic substrates. The enzyme variants catalyze the consecutive double addition of glycolaldehyde to the selected acceptor aldehydes, furnishing different rare aldose sugars and their derivatives. Furthermore, the stereochemical course of the first glycolaldehyde addition was also altered by the mutagenesis of the catalyst, thus allowing the one-pot synthesis of L-glucose and derivatives from successive glycolaldehyde additions on the electrophile substrate by the combination of FSA mutants. This method represents the first stereoselective synthesis of rare sugars in one pot.

In **Chapter 3.3.** another novel utility of the genetically modified FSA is shown. We have engineered the enzyme to accept an unprecedented variety of 1-hydroxy-2-alkanones and the related ether components as donor substrates. The novel mutants, particularly, the variant FSA L107A/L163A was found to convert linear and branched-chain structures up to seven carbon atoms with high stereoselctivity, thus opening new avenues towards the synthesis of novel product families.

In **Chapter 3.4.** a method for the coupling of organo- and biocatalytic aldol reactions is shown. It was demonstrated that proline catalyzed self-aldol reaction of propionaldehyde can be coupled with FSA catalyzed addition of HA and DHA on the 3-hydroxy-2-methylpentanal, proving once again the versatility of this aldolase to tolerate the highly non-polar aldehyde and catalyze the addition of HA and DHA to the pentanal with high stereoselectivity.

By the structure guided redesign of FSA the efficient synthesis of several novel carbohydrates was carried out, or new methods for their synthesis was discovered.

Supplementary material including NMR spectra can be found in the CD attached.

Collaborations with other members of the group rendered the following publications, not included in this thesis.

Mifsud M, Szekrenyi A, Joglar J, Clapés P. **IN SITU aldehyde generation for aldol addition reactions catalyzed by D-fructose-6-phosphate aldolase**. *Journal of Molecular Catalysis B: Enzymatic*. 2012, **84**, 102.

Guérard-Hélaine, C., Debacker, M., Clapés, P., Szekrenyi, A., Hélaine, V., Lemaire, M.; **Efficient biocatalytic processes for highly valuable terminally phosphorylated C5 to C9 D-ketoses**. *Green Chemistry*. 2014, **16**, 1109.

# Index

# Frequently used abbreviations

ACN: Acetonytrile
Cbz: Benzyloxycarbonyl
DAB: 1,4-Dideoxy-1,4-imino-D-arabinitol
DERA: 2-Deoxy-D-ribose-5-phosphate aldolase
D-F6P: D-Fructose-6-phosphate
DHA: Dihydroxyacetone
DHAP: Dihydroxyacetone phosphate
DMF: *N*,*N*-Dimethylformamide
DNJ: 1-Deoxynojirimycin
DTT: Dithiotheritol
*E. coli*: Escherichia coli
ep-PCR: Error prone polymerase chain reaction
FSA: D-Fructose-6-phosphate aldolase
FucA: L-Fuculose-1-phosphate aldolase
GA: Glyceraldehyde
GO: Glycolaldehyde
G3P: D,L-Glyceraldehyde-3-phosphate
HA: Hydroxyacetone
HB: Hydroxybutanone
HPLC: High performance liquid chromatography
IBX: 2-Iodoxybenzoic acid
LAB: 1,4-Dideoxy-1,4-imino-L-arabinitol
NMR: Nuclear magnetic resonance
PCR: Polymerase chain reaction
PDB: Protein Data Bank
RhuA: L-Rhamnulose-1-phosphate aldolase
Tal: Transaldolase
TFA: Trifluoroacetic acid
TLC: Thin layer chromatography

# 1. Introduction

## 1.1. Enzyme catalysis in organic synthesis

Enzymes are nature's catalysts that build unmatched complexity from simple substrates in environmentally benign conditions. These unique catalysts accomplish a myriad of transformations in living organisms in highly regio- and stereoselective manner. Due to their natural abundance – currently there are more than 5000 entries in the International Union of Biochemistry (IUB) database [1] and this number is continuously growing thanks to proteomic research – and wide range of catalyzed reactions, enzymes can be the adequate catalysts for innumerable transformations [2-7].

Biocatalysis is the application of enzymes in synthetic chemistry for purposes that they were not evolved. The pioneering discovery of Buchner, who in 1897 reported the successful fermentation of sugar by cell free yeast extract overthrew the long established paradigm of *vis vitalis* [8], and opened the door for modern biocatalysis. Since then enzyme catalysis have been applied in food technology, production of textile, and, last but not least in the pharmaceutical industry [2, 5, 9-12]. The main advantage of enzymes is that they offer effective transformations under very mild conditions. These ecologically and economically acceptable practices with which enzymes operate (water as solvent, temperatures between 20-40 °C, pH optimum between 5-9) make biocatalytic transformations perfect candidates for industrial processes where the costs and the "greenness" of the technology are highly considered [13].

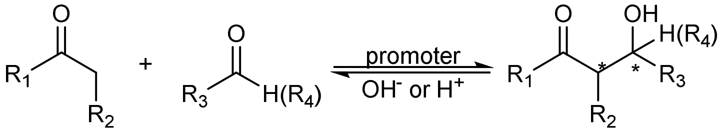
From the 1970s, one of the long standing limitations to generalizing biocatalysis in synthetic organic chemistry has vanished thanks to the development of recombinant DNA methodology. As a consequence, an enzyme can be overexpressed in a host organism, e. g. in *E. coli*, and therefore can be obtained in sufficient quantities for practical applications. Furthermore, protein and reaction engineering have been employed to improve the applicability of enzymes in industrial processes [5, 14, 15].

Due to the superb selectivity and efficiency offered by enzyme catalysis, and since the conditions of these transformations are generally compatible with most functional groups, enzymes offer advantages for the synthesis of different classes of polyfunctional compounds such as carbohydrates [16, 17]. The synthesis of carbohydrates is an area where enzyme catalysis has shown to be especially useful and exploited. The two main enzyme groups for the asymmetric synthesis of these compounds are the transketolases and aldolases [18, 19]. This product class is of fundamental importance since sugars play a significant role in several biological processes.

The main focus of this thesis is the application of protein engineering for the synthesis of different carbohydrates.

## 1.2. Aldol reactions

Asymmetric carbon-carbon coupling is one of the most important and challenging transformations in synthetic organic chemistry. Among the methods for C-C bond formation, aldol addition is a cornerstone transformation from the chemists' toolbox for the simultaneous formation of up to two new stereogenic centers and for the installation of functionality. Aldol addition was discovered in 1864 independently by Charles Wurtz and Alexander Borodin (i.e. the famous Russian compositor) [20, 21]. They described the most simple traditional aldol reaction that runs under thermodynamic control. It is a direct base or acid mediated addition that occurs between a nucleophile (enolate or ketone) and an electrophile (aldehyde) (Scheme 1.1.) [22, 23]. This transformation is usually reversible, and side reactions or dehydration can occur while it normally lacks stereoselectivity, which are serious limitations for the method's practical utility.



**Scheme 1.1.** The direct aldol reaction.

The practicality of these transformations was improved starting from the 1960s by application of directed, stepwise aldol additions with preformed enolates. Zimmermann proposed the chair transition state model for the kinetic selectivity (Scheme 1.2.A.) [24]. The model explained why the Z- or E-enolates preferentially generate *syn*- or *anti*-aldol products, respectively (Scheme 1.2.A.). The directed aldol reactions were initially performed with lithium or magnesium enolates, but later these were substituted by other metals or silicone enolates, e.g. Mukaiyama reaction [25, 26].

A different strategy to induce stereoselectivity is by chiral auxiliaries (Figure 1.1.). These are normally sterically hindered molecules that guide the stereochemical outcome of the reactions when used in stoichiometric amounts.

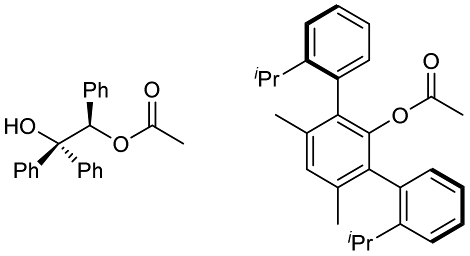
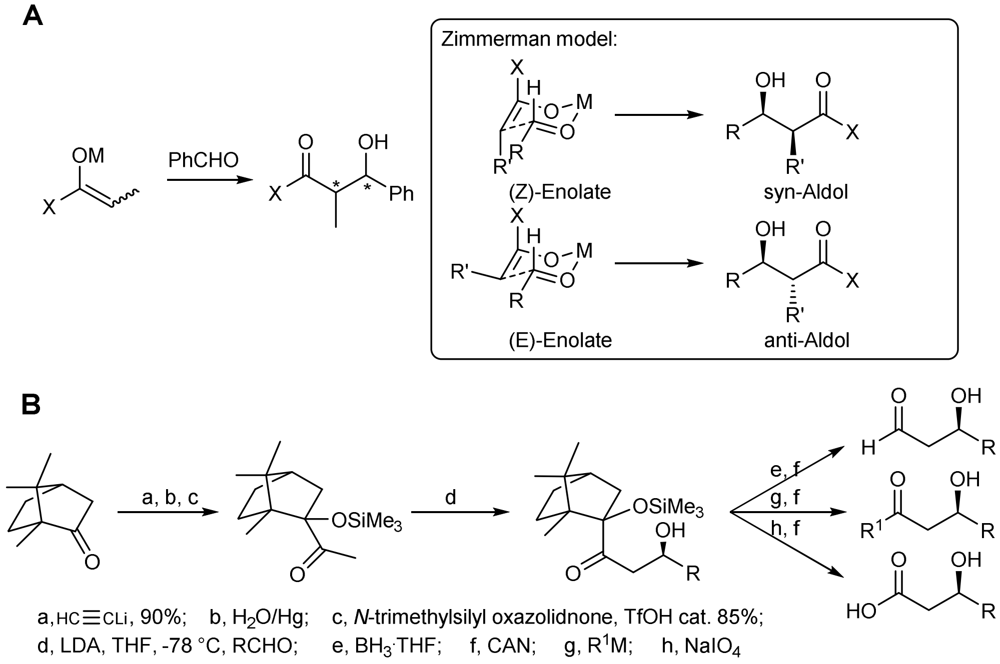


**Figure 1.1.** Sterically crowded chirality inducers display high diastereoselectivity in lithium-mediated aldol reactions [23].

With these chirality inducers, high enantiomeric excess can be achieved [23, 27]. The aldol reactions with chiral auxiliaries (Scheme 1.2.B.) are less atom-efficient compared to other methods, however, they can be more straightforward to prepare than some catalysts and they are efficient in fine tuning regio- and stereoselectivity.



a, $HC{\equiv}CLi$, 90%; b, $H_2O$/Hg; c, *N*-trimethylsilyl oxazolidnone, TfOH cat. 85%;
d, LDA, THF, -78 °C, RCHO; e, $BH_3{\cdot}THF$; f, CAN; g, $R^1M$; h, $NaIO_4$

**Scheme 1.2.** The directed aldol reactions. Governing the stereoselectivity of the reactions: **A**: Z- or E- enolates generate preferentially *syn*- and *anti*-adducts, respectively. **B**: Lithium mediated aldol reaction with camphor for achieving high level of diastereoselectivity [28].

## 1.2.1. Catalytic methods

The direct catalytic aldol reactions are alternatives with higher atom economy for the procedures involving stoichiometric amounts of auxiliaries and they facilitate the reactions in a controlled and selective manner. There are two main areas of research of catalytic aldol reactions. The first field includes chemical processes such as metal catalyzed aldol reactions, and organocatalysis. While the other field gathers the biological processes, which include enzyme catalysis and catalytic antibodies [29].

A high degree of complementarity exists between biochemical and chemical processes, since biocatalysis is generally performed in aqueous solutions with substrates with unprotected functional groups, whereas the metal catalyzed aldol reactions work well in various organic solvents with a wide range of less polar substrates.

Not long after the discovery of the Mukaiyama aldol reaction, its catalytic version was also developed [30], initiating a wave of chiral-Lewis acid catalyzed processes [31]. These metal-enolate based methods resemble the Class II aldolases (Figure 1.2.) [31, 32]. The coordination of the Lewis acid to an acceptor aldehyde creates a chiral environment. This activated aldehyde is then attacked by an enolate from a less hindered face. The majority of these complexes include tin, titanium, boron or silicone as heteroatom, although currently other metals are also being investigated [29].

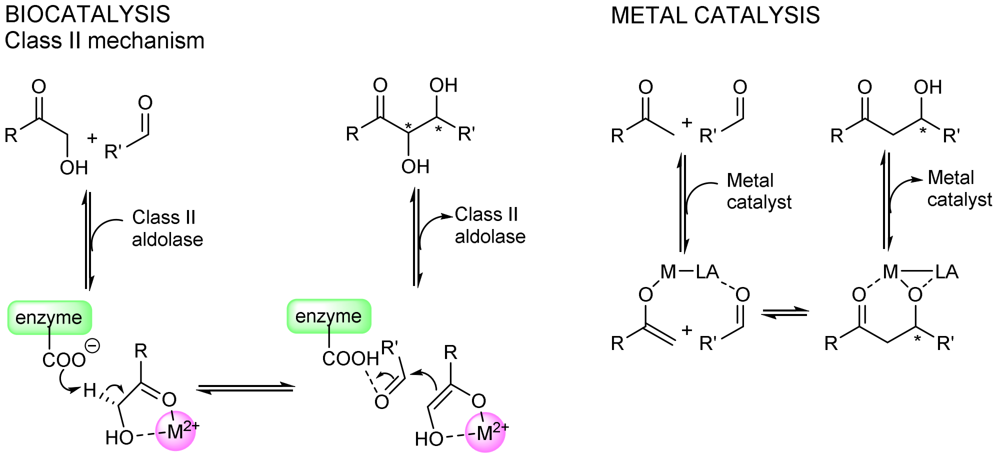


**Figure 1.2.** Comparison of the mechanisms of Class II aldolases and metal-catalysis.

The most dynamically growing field of catalytic aldol reactions of the last decade is organocatalysis. Small amino acids such as proline, which is the most widespread organocatalyst for these transformations, are functioning with enamine mechanism similarly to Class I aldolases (Figure 1.3), for this reason, and for its high stereoselectivity proline has sometimes been praised as the "simplest enzyme" [33]. The first report of its catalytic function dates back to 1971 when Hajos, Parrish, Eder, Sauer and Wiechert used L-proline in an intramolecular cyclization to perform the ring closing of steroids [34]. The revival of proline catalysis took place in 2000 when List and co-workers described the first intermolecular proline-catalyzed direct aldol reaction [35]. Even though the reaction showed only low yields and modest enantioselectivity, the simplicity of the process surpassed these restrictions. After the fast development of the organocatalytic reactions, proline can now be considered a general catalyst for aldol reactions thanks to its universal character in catalysis and its natural abundance [36, 37]. The mechanism of the intermolecular proline catalyzed aldol reaction has been studied thoroughly and shown to be basically the same as the one of Class I aldolases [32, 33], starting with enamine formation, followed by the addition of the acceptor governed by the carboxylic group of the proline, and finalized by the hydrolysis of the iminium ion (Figure 1.3).

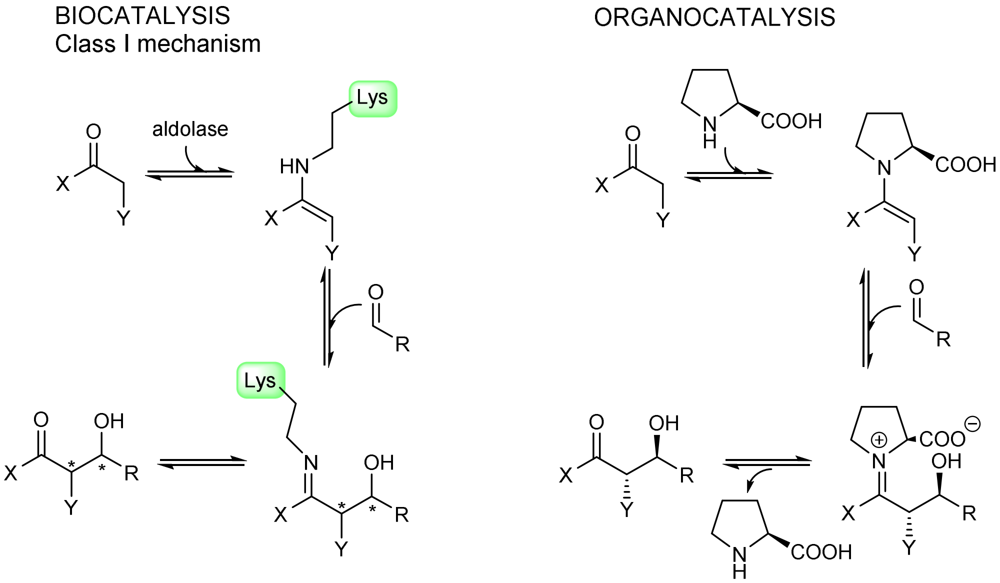


**Figure 1.3.** Comparison of the mechanisms of Class I aldolases and proline-catalysis.

In the first reports, the reaction was restricted to acetone which was also used as solvent, so when extended to other ketones – for practical reasons – only simple substrates, such as cyclohexanone, were used [38]. In these reactions the *anti*-aldol adducts were favored (Figure 1.4.), as well as in the reactions with hydroxyacetone [39] and a protected dihydroxyacetone [40, 41]. With these later additions, 2,2 dimethyl-1,3-diox-5-one protected sugar derivatives were generated.

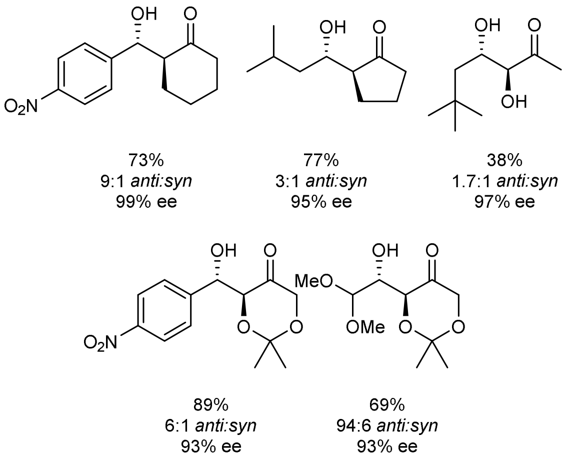


**Figure 1.4.** Products with ketone donors in the proline-catalyzed aldol reactions with 10-30 mol% L-proline. This catalysist has *anti*-selectivity.

Aldehydes can also be used as donors in the organocatalytic aldol reactions. The first self aldol reaction of acetaldehyde was reported by Barbas *et al*. [42]. Later Northup and MacMillan reported the remarkable cross-aldol reaction of aldehydes using slow addition of the donor [43]. The cross aldol products were obtained in excellent yield and enantioselectivity. Later MacMillan *et al.* expanded the group of the utilized aldehydes to dimerization of protected glycolaldehyde derivatives, to produce aldose sugars [43].

Despite the advantages of proline catalysis, the search for new organocatalysts has been also progressing. Although when developing new catalysts, one should consider that they ought to be more reactive and selective, and bear increased lipophilicity to gain advantage over the low cost and wide scope of proline. Generally these new catalysts are proline derivatives or analogs with increased hydrophobicity to improve the catalyst's solubility in organic solvents, while the other trend is to prepare catalysts with bulky groups to enhance stereoselectivity [32, 44].

Another disadvantage of organocatalytic reactions is the lack of stereocomplementary catalysts. The majority of catalysts remain *anti*-selective, and the opposite diastereomers are difficult to prepare. However, a couple of catalyst were developed that display *syn*-selectivity [45, 46].

In addition to chemical catalysis of the aldol reaction, a wide range of biocatalytic methods exist. The most significant classes of biological processes are antibody and enzyme catalysis.

In the mid-1980s the first catalytic antibodies were discovered by Lerner and Schultz, and at once they were considered alternatives to enzymes [47]. The last two decades have witnessed an intense exploration of catalytic antibodies. These catalysts are designed by synthetic chemists to complement the substrate scope of aldolases [48, 49]. Reactions are realized through a method called reactive immunization, where a β-diketone reacts with the active site lysine of a catalytic antibody [29, 48]. This method was also used for the functionalization of the commercially available aldolase antibody of 38C2, which exhibits an extremely wide substrate scope. The complementary features of the Catalytic antibodies with the aldolases are demonstrated by the tolerance of a wide range of ketones as donor substrates [29]. Small aliphatic ketones are very well accepted by antibodies unlike with natural aldolases, while again unlike aldolases, antibodies do not tolerate polyhydroxylated aldehydes such as glyceraldehyde. This tolerance towards hydrophobic substrates comes from the generally hydrophobic nature of the binding pocket, where the lysine residue is located. Catalytic antibodies operate with a similar mechanism like Class I aldolases (Figure 1.3.). However great the expectations were in the period when the first studies were published, the generalization of these protein-based catalysts was never achieved. This is likely due to their generally low activity, due to the lack of activating effects in the optimal binding of the substrates.

### 1.2.2. Enzyme catalysis

Aldolases are a class of lyases that catalyze the reversible addition of a nucleophile donor compound (usually a ketone) to an electrophilic acceptor (an aldehyde) in a reversible and stereocontrolled fashion. In living organisms aldolases catalyze the formation or cleavage of C-C bonds of carbohydrates, keto-acids and some amino acids [50]. Up to now, more than 30 aldolases are known, and most – if not all –

organisms contain them, so possibly more of this class of enzymes are waiting to be discovered [51].

One of the main characteristics of aldolases is that, in most instances they can predictably control the stereoselectivity of the reactions. These enzymes bind their respective donors with high specificity, and normally do not accept a broad structural variation but just extremely close isosteric analogues at much lower rates. Interestingly enough, they are flexible with the structure of the acceptors [50, 51], making them useful for chemical synthesis.

The hallmark of the biocatalytic aldol additions is their ability to produce new and diverse structure types by an individual variation of the electrophilic and nucleophilic components. This generation of diversity can be accomplished by a sequential combination of independent aldol additions catalyzed by different aldolases in which the aldol product of one reaction becomes the substrate for the next. Furthermore, the existence of stereocomplementary enzymes with similar substrate tolerance offers the potential of stereodivergent product generation from a common pair of synthetic precursors.

The aldol addition catalyzed by aldolases was discovered in 1936 by a team of German researches, Meyerhof, Lohmann and Schuster, at the Institut für Physiologie am·Kaiser Wilhelm-Institut für Medizinische Forschung in Heidelberg when using an extract from rabbit muscle what they called “*ein Kohlenstoff-verknüpfendes Ferment*” [52, 53]. Later in 1954, Albert Lehninger published a paper in the Journal of American Chemical Society about “Enzymatic Preparation of Two Ketohexose 1-Phosphates” using D-fructose 1,6-diphosphate(FDP). He considered the method “*simple and capable of yielding fairly large quantities of ketose phosphates which are otherwise rather difficult to prepare by purely chemical procedures*.” [54].

### 1.2.2.1. Mechanism

Based on their way of action, aldolases are divided into two classes [50, 55, 56]. The common ground in their mechanism is the abstraction of the α-proton of the donor substrate, generating a carbon nucleophile in the active site of the enzyme (Scheme 1.5.). The way of the generation of nucleophile species is the point of distinction of the two classes [17].

For the reactions with class I aldolases no cofactor is necessary. The donor substrate is activated through a Schiff base with a highly conserved lysine residue. This class is present mainly in higher eukaryotes: animals and plants. The reaction starts with the formation of an imine between the donor substrate and the ε-amino group of the lysine, followed by the subsequent the abstraction of the α-proton, and then formation of the enamine tautomer that attacks the acceptor aldehyde strictly from the suited face in a retention of configuration fashion. The enzyme-bound imine is then restored and hydrolyzed, releasing the product (Figure 1.5.A.) [29, 56].

Class II aldolases, on the other hand, only exist in prokaryotes or lower eukaryotes. These enzymes possess a metal co-factor usually a $Zn^{2+}$ that is bound to the active site coordinated by histidine residues. The metal ion acts as a Lewis acid, facilitating the deprotonation of the donor that performs a nucleophilic attack on the aldehyde to afford the aldol product (Figure 1.5.B.) [55].

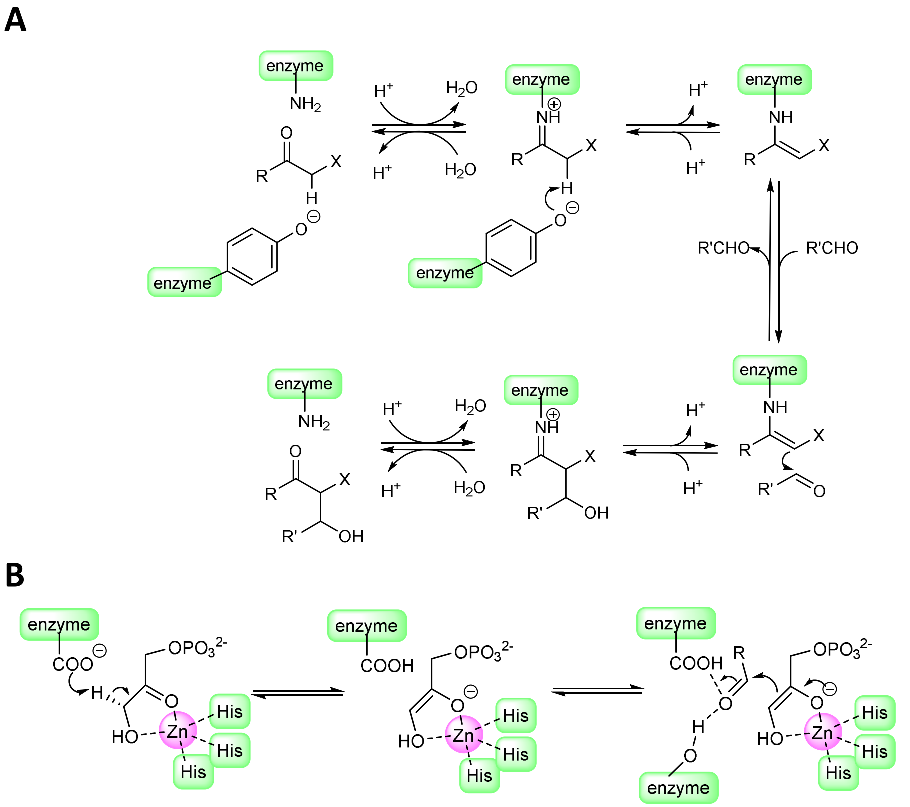


**Figure 1.5.** Schematic mechanism of aldolases. **A**: Class I aldolases activate the donor substrate via imine-enamine formation. B: Class II aldolases require a metal ion which acts as a Lewis acid in the activation of the nucleophile.

### 1.2.2.2. Donors

An alternative classification of aldolases is according to their donor specificity. As mentioned before, these lyases are highly restrictive with their donor substrate, and generally tolerate only small isosteric modifications, if any, in the accepted nucleophile. This specificity provides a structural basis for their categorization: dihydroxyacetone phosphate (DHAP), dihydroxyacetone (DHA), pyruvate and phosphoenolpyruvate, acetaldehyde and glycine can serve as donor substrates (Figure 1.6.) [51].

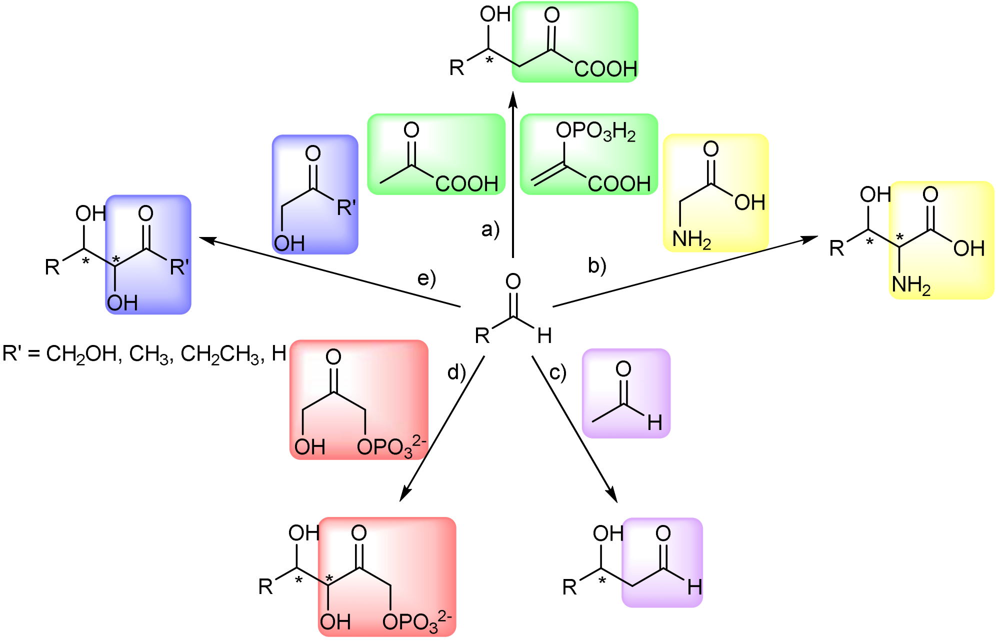


**Figure 1.6.** Classification of aldolases by their donor substrate into five main groups. The different aldolases generate one or two stereocenters.

(a) Pyruvate and phospoenolpyruvate dependent aldolases
(b) Glycine dependent aldolases
(c) Acetaldehyde dependent aldolases
(d) Dihydroxyacetone phosphate (DHAP) dependent aldolases
(e) Dihydroxyacetone (DHA) (and analogs)dependent aldolases

A class of aldolase that accepts unphosphorylated analogues of DHAP constitutes an exception to this strict donor specificity of aldolases. This DHA-utilizing aldolases such as D-fructose-6-phosphate aldolase, accepts a wide range of donor analogues of DHA (see section 3.3.).

Interestingly, a broad variety of aldehydes is tolerated as acceptors for all classes of aldolases. Partially for this reason, aldolases are considered as extremely powerful catalysts in the synthesis of highly functionalized products with high catalytic efficiency under mild conditions and with uncompromised stereochemical fidelity.

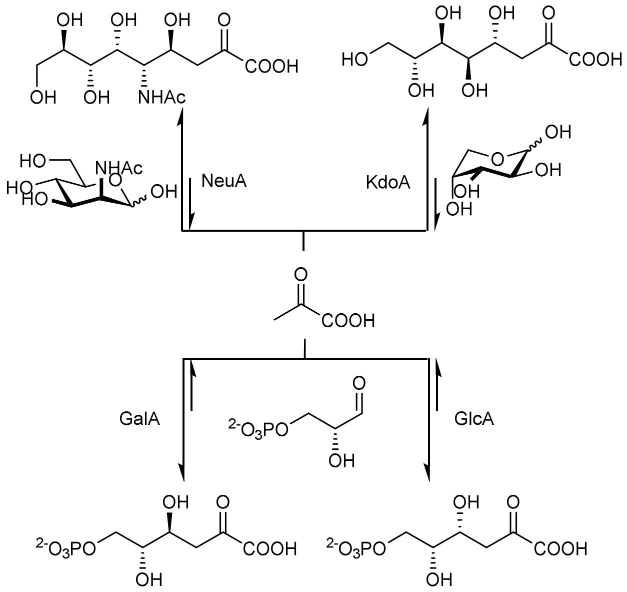
### *a. Pyruvate dependent aldolases*

Pyruvate dependent aldolases belong to a large family of enzymes that have the function of degrading sialic acids, or generally hexoses and pentoses. Due to this metabolic function the reaction equilibrium is mostly shifted towards the retro-aldol direction, thus for their application in synthesis an excess of pyruvate has to be used.

The most studied aldolases of this group are:

- *N*-Acetylneuraminic acid aldolase (NeuA): generally referred to as sialic acid aldolase that catalyzes the synthesis of sialic acid from pyruvate and *N*-acetyl-mannosamine,
- 2-Keto-3-deoxy-*manno*-octosonate aldolase (KdoA) uses arabinose as acceptor,
- 2-Keto-3-deoxy-6-phospho-D-gluconate aldolase (GlcA) and 2-keto-3-deoxy-6-phospho-D-galactonate aldolase (GalA) (also known as KDPG aldolases) are structurally related and have a preference for smaller acceptors,

These enzymes assemble a stereocomplementary set of pyruvate aldolases (Scheme 1.3.).



**Scheme 1.3.** Stereocomplementary set of pyruvate dependent aldolases [17].

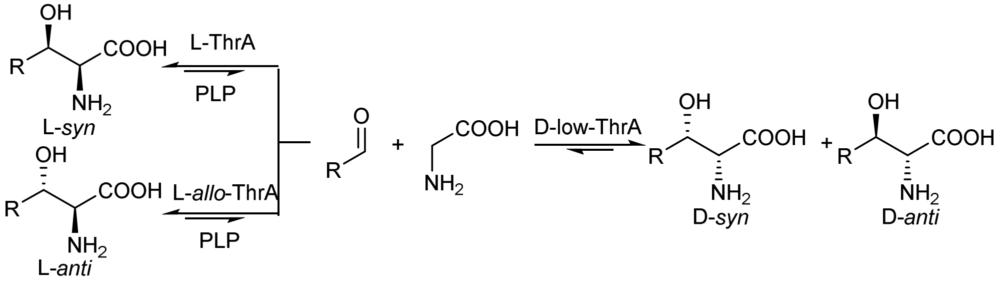
NeuA is the most intensively investigated member of the pyruvate dependent aldolases. It has very broad acceptor tolerance towards sugar derivatives [57-59], and besides pyruvate, fluoropyruvate is also accepted as nucleophile [60]. NeuA has been widely exploited in the synthesis of sialic acid containing structures. These molecules are of pharmaceutical interest because of their potential anti-bacterial, anti-virus and tumor metathesis properties [61-63].

Another well studied pyruvate dependent aldolase is KdoA. This enzyme’s preparative application has mainly focused on the synthesis of the analogs of its natural product. The aldolases GlcA and GalA are mainly involved in the synthesis of the precursors of enantiomers of an antibiotic Nikkomycin K [64].

### *b. Glycine dependent aldolases*

Glycine-dependent aldolases catalyze the reversible aldol reaction of glycine with an aldehyde to form β-hydroxy-α-amino acids. There are two groups in this class: serine hydroxymethyl-transferase (SHMT) and the group of threonine aldolases (ThrA). Both catalysts require the cofactor pyridoxal-5’-phosphate (PLP) for their action, which does not need to be regenerated. These aldolases create two new stereocenters with the addition of glycine to an acceptor, thus formally four stereocomplementary products can be obtained with the threonine aldolases.

These enzymes are highly selective for the configuration on the α-carbon of threonine, so L-and D-ThrAs can be distinguished. Nevertheless, the selectivity on the β-position is generally decreased. For L-ThrA the different selectivities can be divided into L-, L-*allo* and L-*low*-ThrA. The former two preferably produces the *syn*- and *anti*-conformers, respectively, while the later has no selectivity. From D-ThrAs so far only aldolases with D-*low* specificity have been found in nature (Scheme 1.4.). Genes encoding L-ThrA are widespread in nature. They exist in fungi, bacteria, plants, etc., in contrast to only a few D-ThrA genes reported up to date [65].

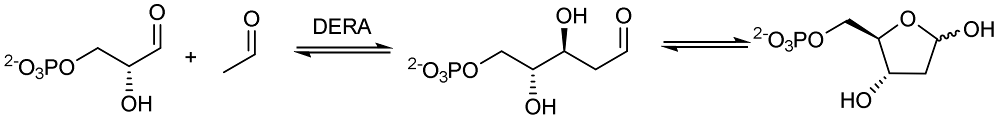


**Scheme 1.4.** Stereocomplementary set of glycine dependent aldolases.

Glycine dependent aldolases are exploited in the synthesis of β-hydroxy-α-amino acids [1, 66-68], component of several natural products or pharmaceutically interesting compounds [68-75]. In general, reactions proceed with high stereocontrol at the $C_\alpha$, whereas usually a mixture of *syn*- and *anti*-products is obtained. This is because of the low stereoselectivity in the β-carbon of the formed products, leading to the thermodynamic product that is a mixture of the *syn*- and *anti*-isomers. Diastereoselectivity can be enhanced with reactions ran at low temperatures or with shorter reaction times trapping the kinetic product [67, 72, 76].

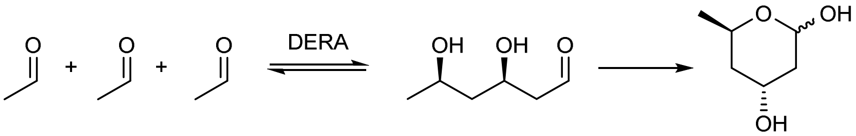
### *c. Acetaldehyde-dependent aldolases*

The group of acetaldehyde dependent aldolases has only one known member, 2-deoxy-D-ribose-5-phosphate aldolase (DERA). In vivo, DERA catalyzes the synthesis of D-2-deoxyribose 5-phosphate from acetaldehyde and glyceraldehyde 3-phosphate (Scheme 1.5.). DERA catalyzes aldol additions between two aldehydes, what is unprecedented among aldolases, with the exception of the recently found activity of D-fructose-6-phospate aldolase see Section 3.1. Another interesting trait of this aldolase is that its donor specificity is not as strict as of other aldolases. Besides acetaldehyde, propanal, acetone and fluoroacetone are also accepted but at much lower rates [77].



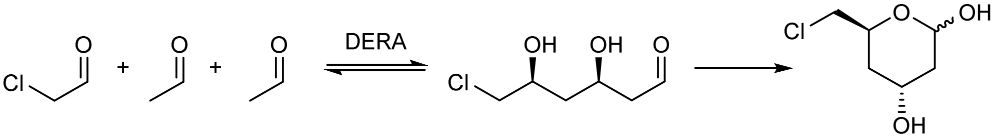
**Scheme 1.5.** The natural reaction of DERA: formation of D-2-deoxyribose 5-phosphate from glyceraldehyde 3-phosphate and acetaldehyde.

Since the product of the aldol addition of DERA is an aldehyde, a second, sequential aldol reaction can be performed by this aldolase, furnishing deoxysugars, which constitute an important class of sugars analogues and can be used as chiral building blocks as well. In each addition, strict control of the streochemical outcome was performed by the enzyme (Scheme 1.6.). Gijsen and Wong showed that starting from three achiral aldehydes (3*R*, 5*R*)-2,4,6-trideoxy hexoses can be synthesized [78].



**Scheme 1.6.** DERA catalyzed trimerization of acetaldehyde.

DERA can be used in the synthesis of important pharmaceutical targets. For example a new strategy for the synthesis of novel pyranose synthons for epothilones is the enzymatic addition of acetaldehyde to α-and β-hydroxyaldehydes catalyzed by DERA [79]. Another use of DERA in an industrialized process is the synthesis of an intermediate to atrovastin (a cholesterol lowering drug) by the double addition of acetaldehyde on chloroacetaldehyde (Scheme 1.7.) [80]. The enzyme's tolerance to chloroacetaldehyde (an inhibitor of DERA) was improved by directed evolution (see Section 1.3.) [81].



**Scheme 1.7.** Application of DERA in the synthesis of the precursor of atorvastatin from two sequential acetaldehyde additions on chloroacetaldehyde.

### *d. Dihydroxyacetone phosphate(DHAP) dependent aldolases*

Dihydroxyacetone phosphate (DHAP) dependent aldolases catalyze the reversible addition of DHAP to a large variety of aldehyde acceptors. In the aldol reactions with DHAP two new stereocenters are created in concert with the C-C bond formation, therefore formally four stereoisomers can be obtained. Nature has evolved the four DHAP-dependent aldolases to produce each of these four stereoisomers with excellent selectivity (Figure 1.7.). Up to now, three of the known four DHAP-dependent aldolases have found a broad synthetic applicability.

- D-Fructose 1,6-bisphosphate aldolase (FruA) is an ubiquitous glycolytic enzyme that catalyzes the reversible addition of DHAP to D-glyceraldehyde 3-phosphate to produce D-fructose 1,6-bisphosphate. FruA is the most prominent member of the family of DHAP-dependent aldolases, and RAMA the class I aldolase isolated from rabbit muscle is probably the most widely studied aldolase by virtue of its commercial availability.

- L-Rhamnulose-1-phosphate aldolase (RhuA) and fuculose-1-phosphate aldolase (FucA) catalyze the addition of L-lactaldehyde to DHAP creating the L-*threo*- and L-*erythro* conformations, respectively.

- D-Tagatose 1,6-bisphosphate aldolase (TagA) catalyzes the aldol cleavage of D-tagatose 1,6-bisphosphate to D-glyceraldehyde 3-phosphate and DHAP. TagA is the only member of this family which lacks of steroselectivity for most catalyzed aldol reactions [82], consequently this is a DHAP-dependent aldolase with modest synthetic applicability.

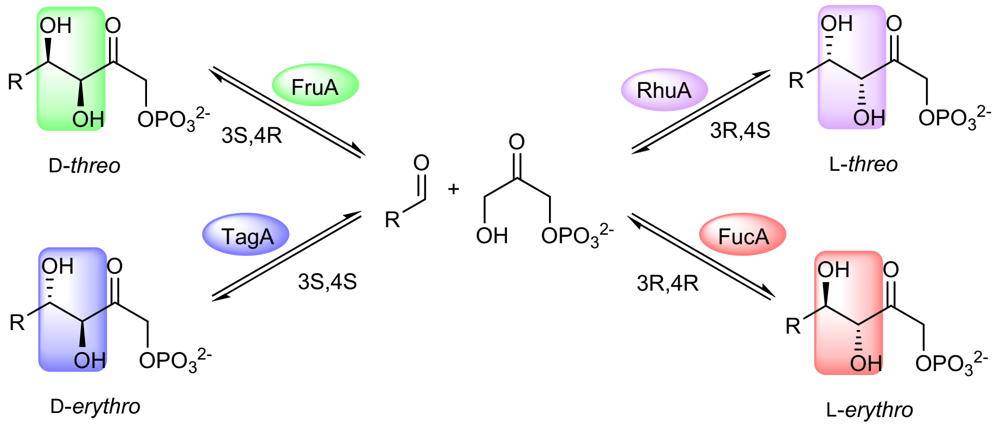


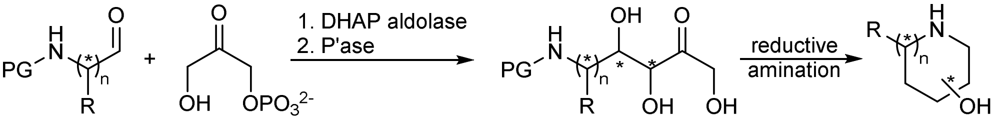
**Figure 1.7.** The stereocomplementary set of the four DHAP-dependent aldolases.

Each aldolase accept a wide range of acceptors including aliphatic and aromatic aldehydes, hydroxy-, thio- and aminoaldehydes with full stereoselectivity [83-85] at the C3 of the adduct (i.e., the configuration generated by the nucleophile) based on mechanistic requirements [86]. The stereocenter at C4 (i.e., the one from the aldehyde acceptor) is also highly conserved generally, however, in some instances, the selectivity depend on the aldehyde component, due to the incorrect binding of the electrophile, obtaining epimeric mixtures or reverse configuration as compared to that with the natural acceptor substrate [87].

The synthetic utility of the DHAP-dependent aldolases is broad. They were used in the synthesis of $^{13}$C labeled sugars, deoxy sugars, fluoro sugars, iminocyclitols etc. [17, 51, 88-90]. Bidirectional chain elongation was achieved as well by FruA and RhuA using dialdehydes, to produce disaccharide mimics [91]. The dialdehydes were obtained by ozonolysis of substituted α,ω-dienes. Thiosugars can be prepared from 2- and 3-thiolated aldehydes similarly to hydroxyaldehydes [92], these products are structural variations to carbohydrates, therefore they possess potential biological activity, for instance as glycosidase inhibitors [93]. As an example of industrial application of FruA the synthesis of a macrolactone (+)-aspicillin can be mentioned [94].

An important application of DHAP dependent aldolases is in the synthesis of iminosugars [95]. Iminocyclitols have potential value as inhibitors of glycosidases and glycosyltransferases [96], therefore the straightforward synthesis of these compounds with both functional and configurational diversity is of paramount importance. The synthetic strategy consists of the enzyme-catalyzed addition of DHAP to an aminoaldehyde equivalent followed by a reductive amination (Scheme

1.8.) [97-99]. Using different enzymes, diastereomeric piperidine, pyrrolidine, pyrrolizidine, indolizidine, quinolizidine derived iminocyclitols were prepared [86, 97, 100-102]. Because free aminoaldehydes are unstable substrates, the acceptors can be either *N*-protected aminoaldehydes, or azido aldehydes [86, 103].



**Scheme 1.8.** Chemo-enzymatic synthesis of iminosugars. PG: protective group.

Despite their synthetic utility, reactions catalyzed by DHAP-dependent aldolases suffer from the strict donor specificity. Only a few isosteric replacements (e.g. methylene phosphonate) are tolerated [104]. Although a great deal of investigation has been dedicated in the last decade to the substitute the phosphorylated donor efficient methods for the synthesis of DHAP are still essential. DHAP is chemically unstable and easily decomposes into inorganic phosphate and methyl glyoxal. Several methods exist for the synthesis of DHAP. It can be prepared chemically, enzymatically or by the combination of the two techniques [105]. Moreover, the phosphate group in the product is not always desirable, so in most cases the 1-phosphate group has to be removed from the product. This is most generally realized by the use of acidic or alkaline phosphatase in an additional step [86, 106].

Recently, an aldolase-kinase fusion protein was reported [107]. It is built up by a homodimeric dihydroxyacetone kinase (DHAK) from *Citrobacter freundii* and two units of FruA from *Staphylococcus carnosus*. The fusion protein retains both kinase and aldolase activity and catalyzes the phosphorylation of DHA, and its subsequent aldol addition with ~20-fold increase in the reaction rate compared to the parent enzymes.

Arsenate and borate (and occasionally vanadate) esters mimicking DHAP have been used to mimic the phosphorylated substrates. The arsenate esters are generated in situ from alcohols and inorganic arsenate salts, and have been used successfully with DHAP dependent aldolases [88, 108, 109]. However, due to the toxicity of the these salts and the low rates which characterize these reactions the utility of this method can be questioned. As an alternative DHA-borate esters can be used. Borates have lower toxicity, and RhuA was described to accept DHA in the presence of borate buffer with different aminoaldehydes, to afford L-iminocyclitols with improved yields. This effect might be explained by the product trapping of borate esters [110, 111].

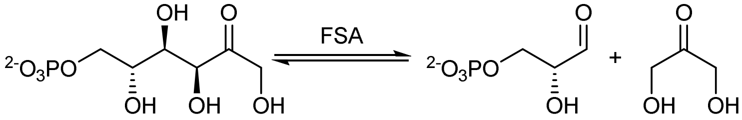
### *e. Dihydroxyacetone (DHA) utilizing aldolases*

The group of DHA dependent aldolases has to date only one natural member, the D-fructose-6-phosphate aldolase (FSA) from *Escherichia coli* and a genetically modified variant of transaldolase B also from *E. coli*, TalB F178Y.

These enzymes have the advantage over DHAP dependent aldolases that they directly use unphosphorylated donor substrates, greatly simplifying their application by avoiding the additional steps of handling the phosphate groups. FSA can be considered as an alternative of FruA since it catalyzes reactions with a similarly broad range of acceptors and with excellent steroselectivity to (3*S*,4*R*)-configuration.

Since this thesis is exclusively focusing on the engineering FSA and its utilization in asymmetric synthesis, its structure and characteristics will be discussed more in detail.

FSA has the unique ability of cleaving D-fructose-6-phoshate (D-Fru6P) to DHA and D-glyceraldehye-3-phosphate (D-G3P) (Scheme 1.9.) [112]. Its role in the *E. coli* metabolism is still uncertain, and this activity has not been found in any other organism so far. Possibly the *fsa* gene is an evolutionary remain of a catabolic pathway to rare sugars [113]. Two isozymes are reported from *E. coli* FSA A and FSA B. They are structurally very similar, but due to practical reasons FSA A is studied more deeply [112, 114].



**Scheme 1.9.** The putative natural reaction of FSA: Reversible formation of D-fructose 6-phosphate from D-glyceraldehyde 3-phosphate and dihydroxyacetone (DHA).

FSA belongs to the aldolase enzyme class although it does not show any sequence similarity to any DHAP dependent aldolases. Nevertheless, it shows a striking resemblance to transaldolases in sequence and structure. Indeed, the first time the two isozymes were mentioned as transaldolase-like proteins. The catalytic machinery of FSA is very similar to that of classical transaldolases, with subtle differences that might be responsible for the variation in its activity [115].

The three dimensional structure of FSA A was determined by crystallization [116] (PDB ID: 1L6W). It shows a decameric quaternary structure, built up from two identical rings of tightly packed and slightly rotated (18°) pentamers, more simply it can be referred to as a “yo-yo” shape (Figure 1.8.). The active sites in the rings face each other. All subunits contain 220 amino acid residues (23 kDa) [112, 116] that fold into an α/β barrel. The C-terminal helix runs across the neighboring subunit partially shielding its active site. This interaction can be responsible for the tight packing, and consequently the high temperature resistance of FSA, as it shows remarkably high thermo stability ($t_{1/2}$ of 16 h at 75 °C) [112]. This feature facilitates the purification procedure since most *E. coli* proteins are denaturated at this temperature. FSA exhibits a wide pH range (pH 6-12), with optimal activity at pH 8.5 [112]. It also tolerates organic cosolvents for example DMF, assisting to its synthetic utility.

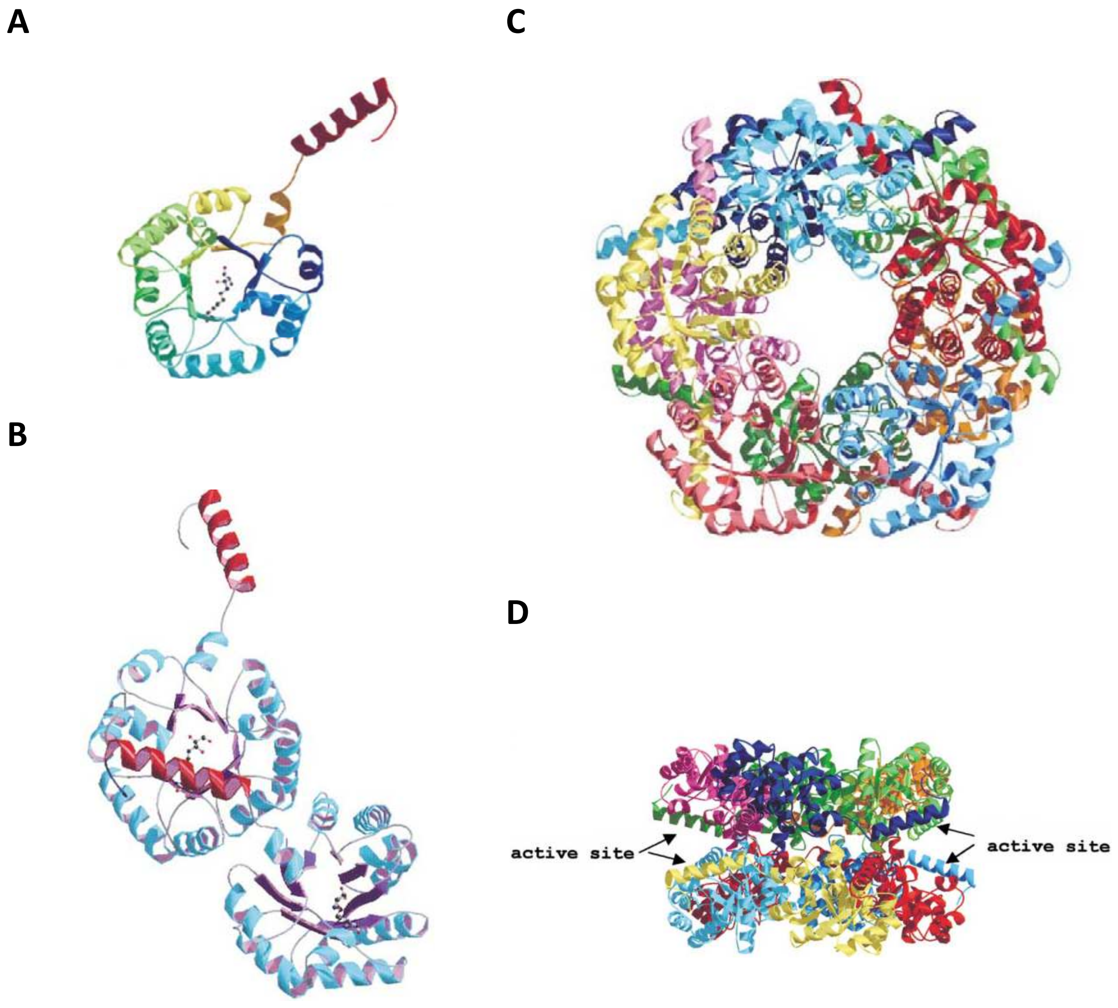


**Figure 1.8.** Crystal structure of FSA [116]. **A:** Schematic view of the FSA monomer. The chain is colored gradually from dark blue at the N terminal to red at the C terminal. The catalytic lysine is shown as ball-and-sticks. **B:** Two interacting subunits. The C-terminal helix is partially covering the active site of the adjacent subunit. **C:** The quaternary structure of FSA from the top view. Each subunit is shown in a different color. **D:** Side-view of the decamer. The entrance to the active sites are indicated with arrows.

Mechanistically FSA is a Class I aldolase with a catalytic lysine (i.e., K85) located at the center of the active site on the $\beta_4$ strand of the structure (Figure 1.9.). This lysine is surrounded by several amino acid residues that are apparently involved in the catalytic mechanism of transaldolases [117-119], thus possibly also important in the mechanism of FSA. These are D6, N28 and T109, which are conserved residues, whereas the Q59 and A129 are replacing a glutamate and a serine, of the transaldolases, respectively. The residues R134, S166 and K168 are suggested to be involved in the binding of the phosphate group of the acceptor, since the similarly positioned amino acids constitute the putative phosphate-binding site of TalB.

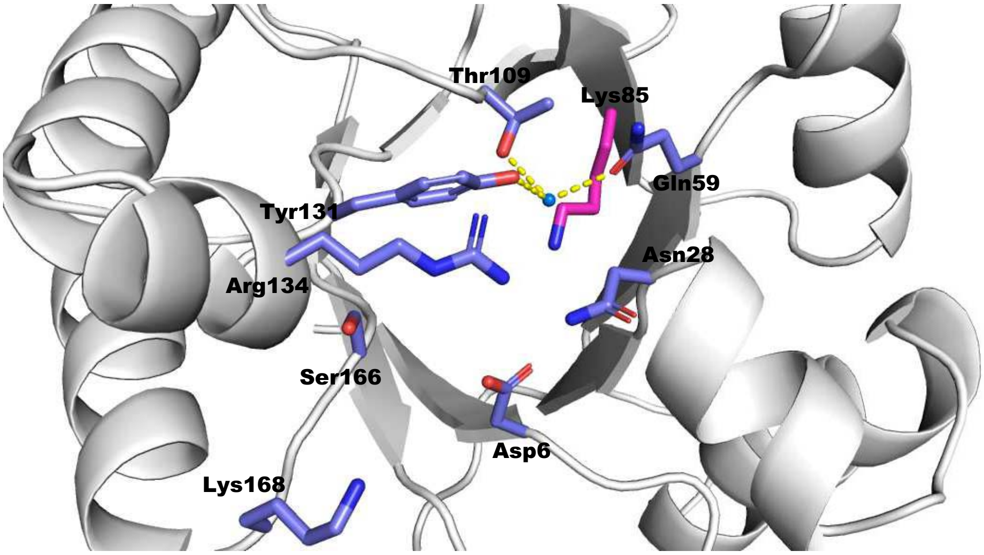


**Figure 1.9.** The active site of FSA with the catalytically important residues. The catalytic lysine is shown in magenta and the rest of the residues important for catalysis are shown in light blue: Asp6, Asn28, Gln59, Thr109, Tyr131, Arg134, Ser166 and Lys168.

The aspartate in equal position to the D6 residue of FSA could be considered to be the acid base residue that is responsible for the protonation of the acceptor substrate in transaldolases. The suggestion from Thorell *et al.* endows the aspartate in position 6 of FSA with a similar role. However, recent results require the review of this theory [120]. The most recent mutagenic studies, which are carried out on FSA in our group in collaboration with several other European research groups, also support the theory that proposes that this residue is only responsible for the stabilization of the substrate by hydrogen bonding.

The catalytic water molecule is bound to the active site residues Q59, T109 and Y131 in the proximity of the O1 of the donor. It was proposed that this water molecule participates in the proton transfer to and from the substrate during the formation of the Schiff-base intermediate (Figure 1.9.). This was recently proposed as a result of a mechanistic investigation of a transaldolase from *Thermoplasma acidophilum* [120].

As stated above, the active site of FSA and TalB from *E. coli* are very similar. There is variation in only a few amino acids (Table 1.1.). The changes in these residues are responsible for the different catalytic profile of the enzymes. As proof, a mutagenic study can be mentioned where the transaldolase was converted to an aldolase by replacing the phenylalanine by a tyrosine in position 178 of TalB (i.e., the residue in the equivalent position of FSA) [115]. However, in a study where the authors intended to confer transaldolase activity to FSA, the A129 residue of FSA *wild-type* was replaced by serine. The FSA A129S variant had a greatly improved affinity towards dihydroxyacetone [121], albeit no transaldolase activity was detected with this variant.

**Table 1.1.** Non-conserved residues in the active site of TalB and FSA.

| **FSA** | Phe57 | Gln59 | Leu107 | Ala129 | Tyr131 | Leu163 | Lys168 |
|---|---|---|---|---|---|---|---|
| **TalB** | Ser94 | Glu96 | Asn154 | Ser176 | Phe178 | Met223 | Arg228 |

Only the non-conserved residues are shown.

For most aldolases the donor substrate is invariable or very small variations in the nucleophiles are excepted, as they only show flexibility with the acceptor substrates [55]. However, FSA has a broad substrate spectrum including at least four different donor compounds [122-125]. The “natural donor” dihydroxyacetone (DHA) can be substituted by hydroxyacetone (HA), 1-hydroxybutan-2-one (HB) and glycolaldehyde (GO) with full stereo and regiospecificity (Figure 1.10.) [113, 123]. It is important to note that the presence of the hydroxymethylene unit in all possible donors is most probably mandatory for the Schiff base formation and the nucleophile generation. According to a comprehensive study non-hydroxylated acetone and butanone were found to be unreactive with wild-type FSA [123], however a recent report states otherwise [126].

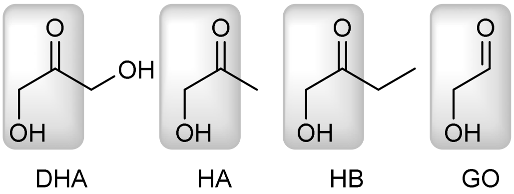
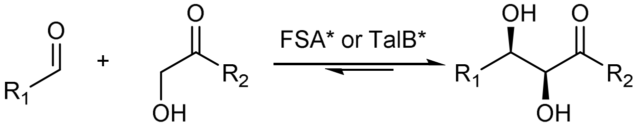


**Figure 1.10.** The tolerated four donor substrates of FSA with the unmodifiable hydroxymethylene unit highlighted.

It was discovered in our group that an unusual donor substrate, an aldehyde, i.e. glycolaldehyde can act as a nucleophile in aldol additions catalyzed by FSA. This novel activity lead to the synthesis of several aldose-type sugars, by the self- and cross-aldol addition of GO in which this substrate will act invariably as donor component [127]. Owing to these products are themselves aldehydes, they can be used as acceptors for further aldol additions with FSA or other aldolases as it will be demonstrated in Section 1.4.2. Investigations in this direction are described in Section 3.2..

The two known DHA dependent aldolases FSA and TalB F178Y have different substrate preferences due to the different hydrophobicity of their active site. FSA has a highly hydrophobic environment (i.e. Leu107, Ala129 and Leu163, Ala165), which facilitates the binding of hydrophobic substrates such as HA and HB while the variant TalB F178Y provides H-bonding with DHA, so it has a strong preference for this nucleophile. Site-directed mutagenesis was employed on both of the aldolases, where a residue in equivalent position was changed *vice versa.* FSA A129S was designed to resemble the donor binding site of TalB with a serine in this position, while TalB S176A/F178Y was created on the contrary to resemble the binding site of FSA [123]. These variants indeed show similar preferences to TalB and FSA, respectively. The donor preferences of each variant are shown in Table 1.2 The variant with increased activity towards unphosphorylated acceptors TalB F178Y/R181E has also been tested with different donor substrates, and was shown to improve the TalB's the tolerance towards GO.

**Table 1.2.** Comparison of the donor preferences of members of the DHA-dependent aldolase family [128, 129].



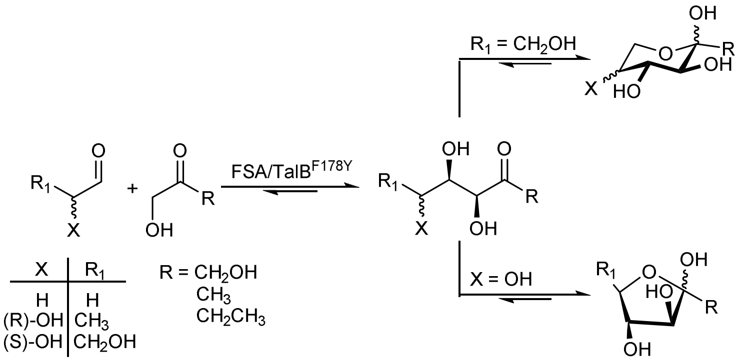
| **Donor compounds ($R_2$)** | **FSA *wild-type*** | **FSA A129S** | **TalB F178Y** | **TalB S176A/F178Y** | **TalB F178Y/R181E** |
|---|---|---|---|---|---|
| $CH_2OH$ (DHA) | + | +++ | +++ | ++ | ++ |
| H (GO) | ++ | + | + | n.d. | ++ |
| $CH_3$ (HA) | +++ | ++ | + | ++ | + |
| $CH_2CH_3$ | +++ | n.d. | - | - | n.d. |

FSA* or TalB*: Different FSA or TalB variants; Donor quality is indicated as +++: excellent, ++: medium, +: poor, -: no activity, n.d.: not determined

In this thesis, mutagenic studies are presented where we intend to improve the FSA activity towards different donor substrates. Our work on the engineering of FSA to accept different novel donor variations is presented in Chapter 3.3., and the mutagenetic studies for better tolerance towards GO is presented in Chapter 3.1.

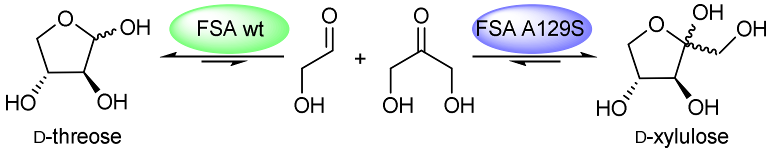
One of the most studied applications of this enzyme family is their use as catalyst in the preparation of iminocylitols and deoxysugars. FSA was shown to accept several *N*-Cbz-protected and azido-substituted aldehydes in combination with DHA, HA, HB and GO to produce iminocyclitols [125, 127, 130-132]. In a recent work some of the limitation of FSA to accept α-substituted acceptors was investigated, and the variant FSA A165G was designed, which significantly improved the yields with these acceptors. Combining the A165G mutation with the earlier described A129S gave an unexpected synergy and showed between 5- to 900-fold higher activity towards *N*-Cbz-aminoaldehydes compared to the wild-type enzyme with both HA and DHA as donors [132]. Several iminosugars were also synthesized with potential glycosidase inhibitory activity such as D-fagomine and *N*-alkylated derivatives [131], 1-deoxymannojirimycin (DNJ) [125] and 1,4-dideoxy-1,4-imino-D-arabinitol (DAB) [130].

In a recent study FSA was also demonstrated to be a good catalyst in the synthesis of several sugar derivatives with various level of deoxygenation [123]. All the investigated aldehydes (8) except D-lactaldehyde, which is a strong inhibitor of the enzyme, were found to be substrates for FSA and TalB F178Y. Out of the entire set of 24 combinations, 22 products were isolated in good to excellent yields (Scheme 1.10.).



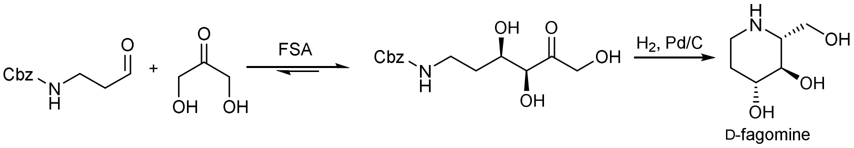
**Scheme 1.10.** Aldol addition of various donors to aliphatic and hydroxyaldehydes catalyzed by FSA and TalB F178Y.

Owing to the high affinity of FSA A129S towards DHA, this variant was also applied in synthesis of polyhydroxylated compounds, in which cases the substituted variant showed improved activity compared to the wild-type [121]. For example D-xylulose and D-fructose were synthesized with this variant, which were before unavailable with the wild-type enzyme (Scheme 1.11.).



**Scheme 1.11.** Scheme of the different donor preference of FSA wt and FSA A129S, in the synthesis of D-threose and D-xylulose, respectively.

The industrial applicability of this enzyme was demonstrated in the synthesis of the iminosugar D-fagomine (Scheme 1.12.) [131], that may be used as a dietary ingredient or functional food component to reduce the health risks associated with an excessive intake of fast-digestible carbohydrates [133].



**Scheme 1.12.** Application of FSA to the synthesis of the naturally occurring iminocyclitol D-fagomine.

## 1.3. Protein engineering of aldolases

Notwithstanding all the advantages of enzyme catalysis, i.e., high stereospecificity, mild reaction conditions, avoiding the need of cumbersome protection and deprotection schemes, it has its own limitations in synthesis. The aldolase specificity leads to narrow range of accepted substrates, although this limitation is also found in organocatalysis, with the difficulty to alter the stereochemical outcome of a given enzymatic transformation. Furthermore, the enzyme activity and stability in the presence of organic solvents might mean restrictions for the use of these catalysts. These limitations have prevented the aldolases to become general catalysts in industrial processes. Fortunately, starting from the last decade, enzyme engineering has been used to deal with and to overcome most of these problems [14, 51, 134] (Figure 1.11.).

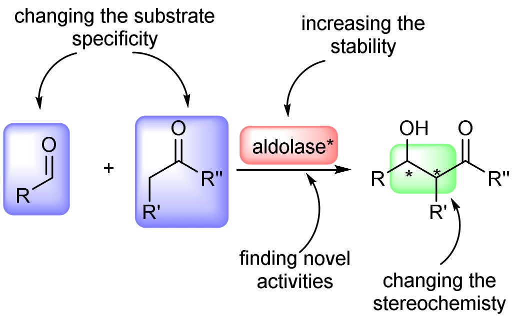


**Figure 1.11.** The main foci of enzyme engineering on aldolases.

Widening the synthetic applicability of aldolases, in particular FSA, by enzyme engineering is the main focus of this thesis, as well as, it is a main field of study of our research group.

There are two main paths for enzyme engineering (Figure 1.12.). First, structure based redesign of the active site of an enzyme for selected purposes. In this case, chosen residues are changed based on knowledge of the structure and mechanism of the aldolase. The advantage of this approach is that the number of produced variants is much lower, so the screening process is faster. The limitation is that a precise knowledge of the structure and mechanism, which is always a great challenge, is of paramount importance.

Second, directed evolution techniques, based on the generation of randomly modified enzymes tested against the desired properties or substrates. This approach does not require the knowledge of the enzyme structure. There are two main approaches used in directed evolution to generate genetic diversity: i) DNA shuffling, i.e. random recombination of related sequences [135] and ii) error prone PCR (ep-PCR) i.e. introducing random changes in single protein sequences [136, 137]. In these techniques, hundreds or even thousands of variants are tested in each round, which is tedious and time consuming. The bottleneck of directed evolution is the development of reliable high-throughput assay methods [81, 138, 139].

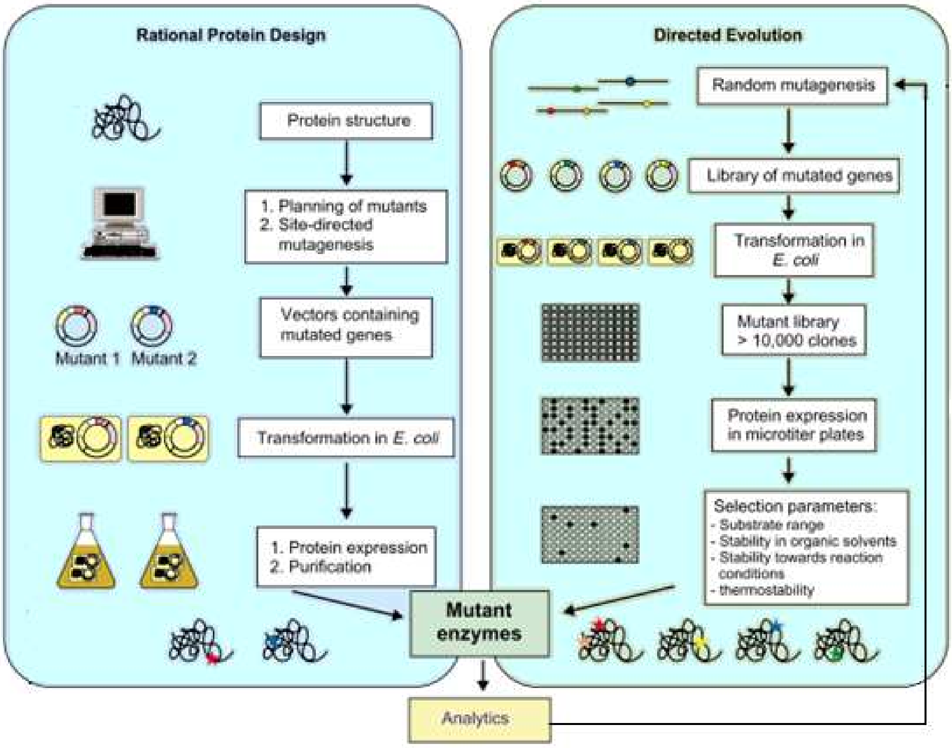


**Figure 1.12.** Rational protein design vs. directed evolution [140].

The ultimate goals of protein engineering are i) to find novel activities, ii) to change the stereochemical course of a reaction or the enantioselectivity of the enzyme, iii) to increase an enzyme's stability, or iv) to alter its substrate selectivity. In this chapter some examples will be presented to all of these approaches.

## 1.3.1. Changing the substrate specificity of aldolases

Since wild-type enzymes only tolerate a limited range of substrates, enzyme engineering is often used as a way to broaden the substrate scope of aldolases.

Many aldolases, e.g. DHAP-dependent aldolases, transaldolases, DERA, have high selectivity towards phosphorylated substrates that can be of interest, but in most instances are not the final goals of synthesis. Therefore, aldolases are often altered to accept the non-phosphorylated substrates.

Sprenger and coworkers engineered the TalB F178Y variant with aldolase activity to increase its affinity towards non-phosphorylated acceptors (e.g. D-glyceraldehyde, D-GA). They performed site saturation mutagenesis on the three residues of the putative phosphate-binding site of the enzyme and screened for the formation of D-fructose with a color assay. The best variant (TalB F178Y/R181E) has more than two-fold increased affinity towards D-GA as compared with the single variant. This increase was obtained through an inversion of the electrostatic charge in the phosphate-binding site [129]. In a similar approach, DERA was also engineered to accept glyceraldehyde instead of glyceraldehyde phosphate. When the three dimensional structure of DERA became available [141], DeSantis *et al.* designed a changed electrostatic environment of the active site – i.e. in five positions glutamic or aspartic acid was introduced – and the variant S238D exhibited a 2,5-fold improvement in $k_{cat}/K_m$ [142] towards the non-phosphorylated substrate.

Enzymes can also be modified to tolerate non-natural electrophiles to improve their versatility in synthesis. Many examples can be mentioned how protein engineering was used to overcome the structural limitations of the acceptor substrates. In our group, FucA from *E. coli* was modified by site directed mutagenesis to catalyze the addition of DHAP to *N*-Cbz-aminoaldehyde acceptors including (*R*)- and (*S*)-*N*-Cbz-prolinal. Among the variants that were produced, F131A was shown to be the most effective, as it accepted (*R*)-*N*-Cbz-prolinal while the wild-type enzyme gave no activity with this substrate. This variant also possesses a 4-25-fold increased reactivity towards other bulky *N*-Cbz-aminoaldehyde acceptors. [143]. FSA was modified in the same direction, as described in Chapter 1.2.2.2..

Woodhall *et al.* attempted the synthesis of sialic acid mimics by rationally designed aldolases. They applied site saturation mutagenesis on three residues of the *N*-acetylneuraminic acid aldolase, and screened the variants against *N,N*-di-n-propyl amide substituent, which substitution means a great change in the size of the acceptor, compared to the natural substrate *N*-acetyl-D-mannosamine, as shown in Figure 1.13. The E192N variant showed 50-fold increased activity compared to the *wild-type*, and also accepted various tertiary amides [144].

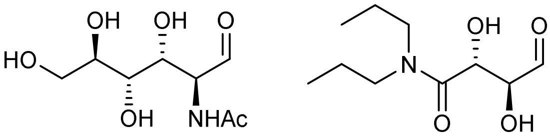


**Figure 1.13**. Altered substrate specificity of a sialic acid aldolase by site saturation mutagenesis.

Attempts to change the selectivity of the nucleophilic substrate of aldolases, e.g. DHAP-dependent aldolases, are an even more challenging field since these enzymes hardly accept small isosteric modifications of their natural donor substrtates.

Probably the synthetically most exploited group of aldolases is the family of DHAP-dependent ones, even though using DHAP in synthesis has certain disadvantages, as mentioned before. This problem was addressed in our group by site directed mutagenesis on the donor binding pocket of RhuA. Redesigning its phosphate binding site lead to the variant N29D that was found to has 3,6-fold higher affinity towards DHA compared to the wild-type [145].

## 1.3.2. Changing the stereochemical course of the reactions

Reactions catalyzed by aldolases are normally showing high stereoselectivity even though in some instances the stereoisomers of the natural products are needed. Two different strategies exist for the preparation of stereoisomeric products (Figure 1.14.). One is to engineer aldolases that can accept stereoisomers of the natural substrates, furnishing diastereomers of the natural products (Figure 1.14.A.). The other approach consists of engineering aldolases to catalyze reactions with modified stereochemical course where the formed carbon-carbon bond is stereochemically altered (Figure 1.14.B.) [14].

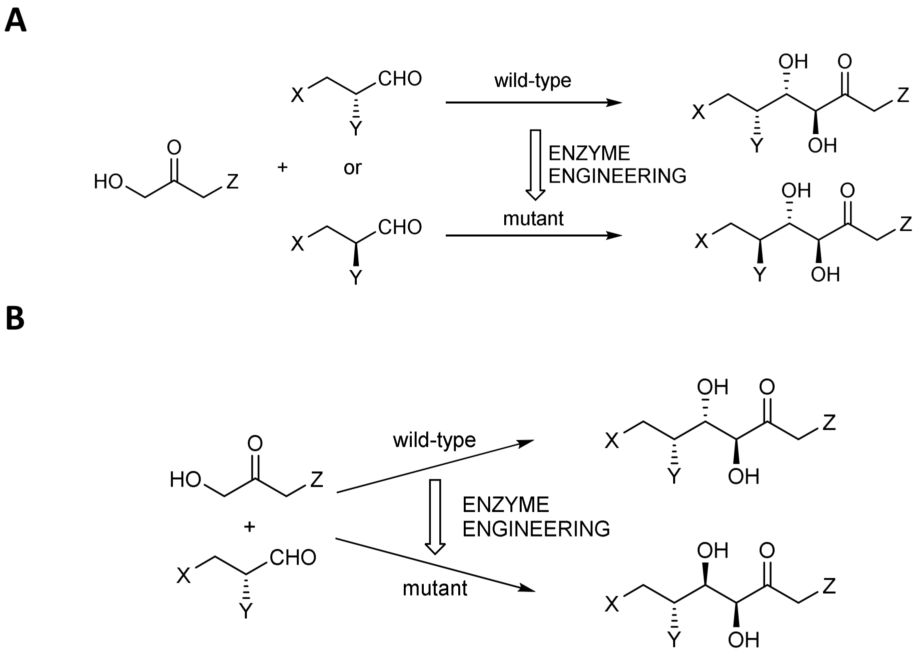


**Figure 1.14.** Strategies of enzyme engineering for the formation of stereoisomeric products. **A:** Modifying the substrate specificity of aldolases. **B:** Modifying the stereochemical course of a reaction.

Several examples exist for both approaches. An example of the first approach was demonstrated by the engineering of D-2-keto-3-deoxy-6-phosphogluconate aldolase, which was used as a template for ep-PCR, and DNA shuffling. Colonies were screened for their ability to accept D- and L-glyceraldehyde leading to the synthesis of D- and L-sugars [146].

As demonstration of the second approach, protein engineering was performed on different classes of aldolases to change the stereochemical course of the reaction. As illustration, the already mentioned NeuA E192N could be shown (Figure 1.15.). This variant exhibits poor stereochemical control of the aldol reaction. It catalyses the cleavage of both diastereomeric products, so this variant was used as starting point for ep-PCR [138]. Two variants were selected which displayed greater stereoselectivity than the template, each possessing 4 and 6 mutations, respectively. Then, site-directed mutagenesis was used to find the residues that play dominant role in the stereochemical control. Finally, two variants were identified to be *R* and *S* selective (Figure 1.15.).

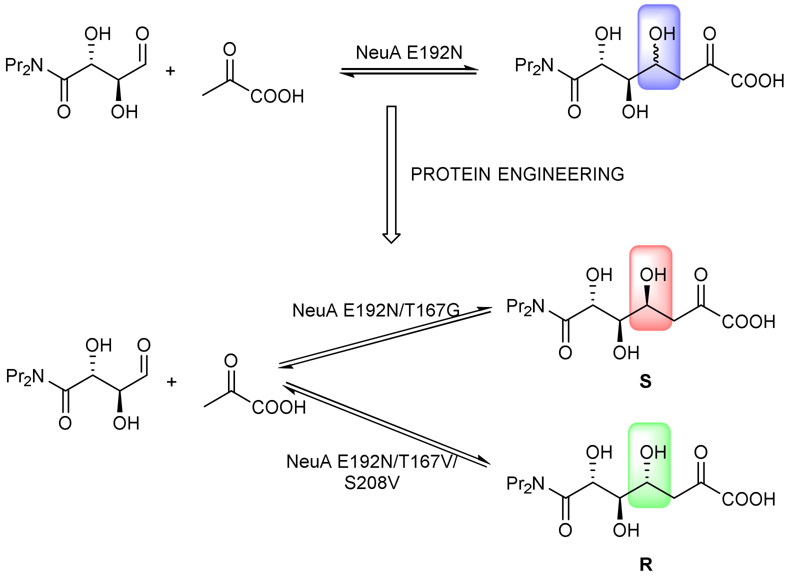
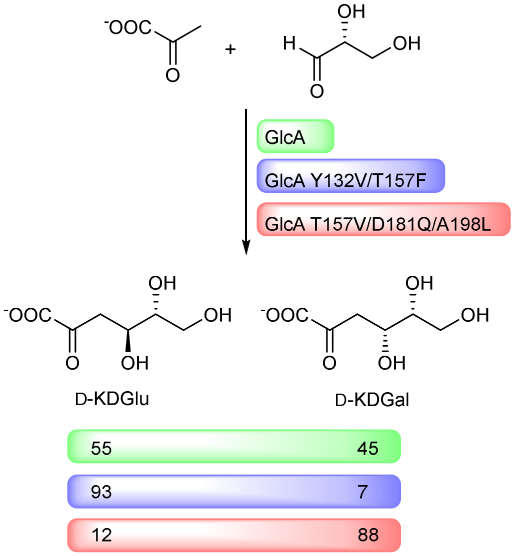


**Figure 1.19.** Enzyme engineering of a pair of complementary aldolases which catalyze aldol reactions with different stereochemical courses.

Stereoselectivity can be also achieved by simpler, structure-guided redesign. 2-Keto-3-deoxygluconate aldolase (GlcA) from *Sulfolobussolfataricus* is a rare example of an aldolase that has no stereocontrol when catalyzing its natural reaction [147, 148] (Scheme 1.14.). This problem was first indirectly addressed by substrate engineering (when a structurally rigid acetonid was used instead of glyceraldehyde) [149]. The study of the crystal structure of the enzyme, revealed that one residue T167 plays a crucial role in the hydrogen bonding in the epimeric $C_4$-OH of both diastereomers, so changes of this amino acid were studied by site saturation mutagenesis. Two single variants were found for the synthesis of both diastereomers, and by further mutagenesis a double and a triple mutant were designed for the stereoselective synthesis of D-KDGlu and D-KDGal, respectively [150] (Scheme 1.14.).



**Scheme 1.14.** Aldol addition of pyruvate and D-glyceraldehyde to afford D-KDGlu and D-KDGal with the wild-type and variants of GlcA aldolase, respectively.

## 1.3.3. Increasing the enzymes' stability

Enzyme inactivation under synthesis conditions is also a great problem when working with aldolases. In the laboratories of DSM, DERA was engineered for industrial applications [81]. They were interested in the production of a statin drug intermediate, 6-chloro-2,4,6-deoxyhexapyranoside, for its synthesis chloro-acetaldehyde was used as acceptor, however under high concentrations of this substrate the enzyme swiftly inactivates. To overcome this problem, error-prone PCR was used to generate genetic diversity. The clones were tested with a 2 min treatment of chloroacetaldehyde followed by measuring the residual aldol activity in the natural reaction. After several rounds of iteration, and the combination of beneficial mutations by site-directed mutagenesis a variant was selected with a 10-fold increased productivity.

In another example, a thermostable FruA variant was also produced by DNA shuffling [151]. Around190-fold improvement in thermostability was observed after the fourth round of screening. The best variant was also tried for its compatibility with organic solvents, and it was found to be especially resistant to irreversible inactivation by organic solvents, so the overall stability of the enzyme was increased.

## 1.3.4. Finding novel activities

The search for enzymes with novel activities is an important field of research since novel “aldolases” might act as complementary catalysts to the already existing ones.

By a single amino acid substitution a PLP-dependent alanine racemase from *Geobacillus stearothermophilus* was converted into a threonine aldolase [152]. It was hypothesized that the H166 in the active site could act as a key residue in the proton abstraction, analogously to threonine aldolases. However, in this enzyme, the histidine does not have a direct connection to PLP, but it forms a hydrogen bond with the residue Y265 that was substituted by an alanine. The engineered enzyme, Y265A, exhibits a $2{,}3 \times 10^5$-fold increase in the retroaldol reaction with β-phenylserine while a huge drop in the alanine racemase activity was observed.

TalB was converted into fructose-6 phosphate aldolase by site saturation mutagenesis [115] by one single amino acid replacement (as described in Section 1.2.2.2.).

As shown in the examples above, protein engineering can be applied to broaden the scope of enzyme catalyzed transformations, and increase the utility of aldolases in organic synthesis.

## 1.4. Cascade reactions in the synthesis of carbohydrates

Carbohydrates are an important class of natural products. They are involved in several processes of life such as glucolysis or glucogenesis [153]. In addition to their role as energy source they also play an important role in several biological signaling and recognition processes, such as immune response, inflammatory reactions, cancer metathesis and viral infections [154, 155]. Carbohydrates are frequent building blocks for the synthesis of drugs. Since carbohydrates are highly functionalized molecules with several stereocenters, their *de novo* synthesis is a challenging issue although, an arsenal of methods have been developed to facilitate their production. Unfortunately most of the approaches are lengthy and require sophisticated protecting-group strategies [156]. However, the aldol reaction is particularly useful in the synthesis of carbohydrates since they render directly β- or α/β-dihydroxycarbonyl compounds [23, 157].

The retrosynthetic analysis of carbohydrate frameworks show that the precursors can be either small aldehydes (to produces aldose) or a small ketone can be combined with an aldehyde to allow the direct construction of ketoses (Figure 1.16.) [158].

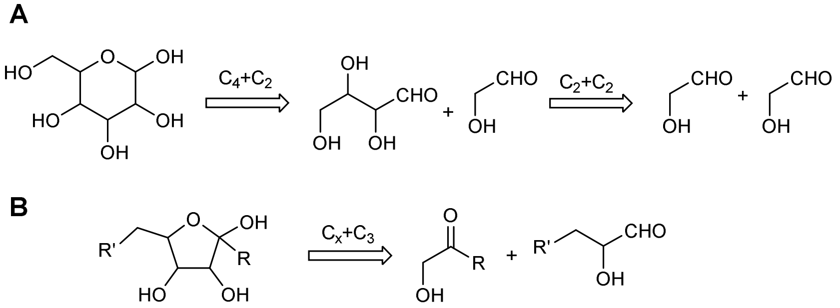


**Figure 1.16.** Retrosynthetic analysis of the main carbohydrate frameworks [158].

Cascade reactions (also referred to as tandem or domino reactions) are defined as processes of two or more bond-forming reactions under identical conditions, in which the subsequent transformation takes place at the functionalities obtained in the former transformation. They allow the efficient synthesis of complex molecules from simple substrates in an ecologically and economically favorable way. These reactions have the benefits of atom economy (the conversion efficiency of a chemical process) as well as they are more time, labor, resource efficient and they generate less waste [159, 160], in this sense, these transformations would be ideal for the construction of carbohydrate frameworks, to avoid additional protection and deprotection steps.

## 1.4.1. Synthesis of aldoses

Aldohexose structures can be built up by two strategies (Figure 1.17.). The aldoses can be synthesized either by the sequential aldol reaction of three aldehydes [161] ($C_2$+$C_2$+$C_2$) or alternatively by a biomimetic method which consists of an initial synthesis of a ketose and its modification by auxiliary enzymes [162] ($C_x$+$C_3$). Both strategies can lead to the synthesis of rare and unnatural sugars which have high potential for applications in pharmaceutical, cosmetic and food industries [163-165].

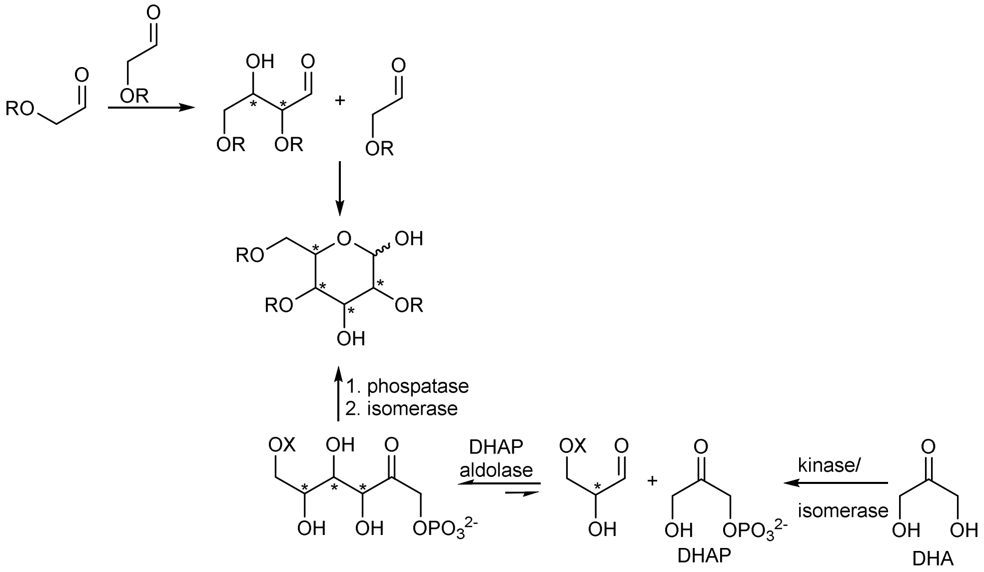
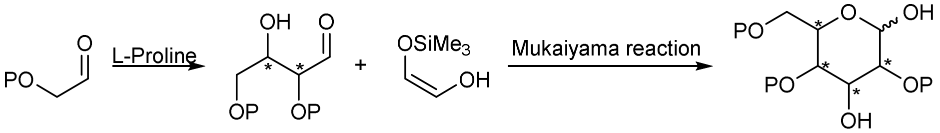


**Figure 1.17.** Two competing strategies for the synthesis of aldohexoses.

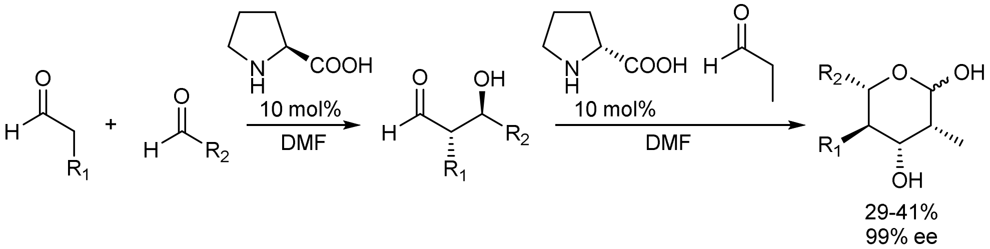
The first strategy would be the most simple, since it only requires the catalysis of two sequential aldol reactions, such as the one accomplished by Wong and his coworkers for the synthesis of trideoxy hexoses (see section 1.2.2.2.) [78]. They isolated the hexoses from the double aldol addition of acetaldehyde on different acceptor substrates.

This approach was also intended by MacMillan who envisaged an organocatalytic two-step carbohydrate synthesis using protected α-oxyaldehydes. On paper, this is a viable strategy, yet its first successful application required the combination of two conceptually different aldol technologies in sequential two-step reactions [166, 167]. The initial dimerization of α-oxyaldehydes lead to an erythrose derivative and was catalyzed by L-proline, followed by a tandem Mukaiyama aldol addition (Scheme 1.15.). Differentially protected glucose, allose and mannose stereoisomers were accessed by simply changing the solvent and the Lewis acid.



**Scheme 1.15.** MacMillan's approach for the two-step synthesis of protected aldohexoses. P: protecting group

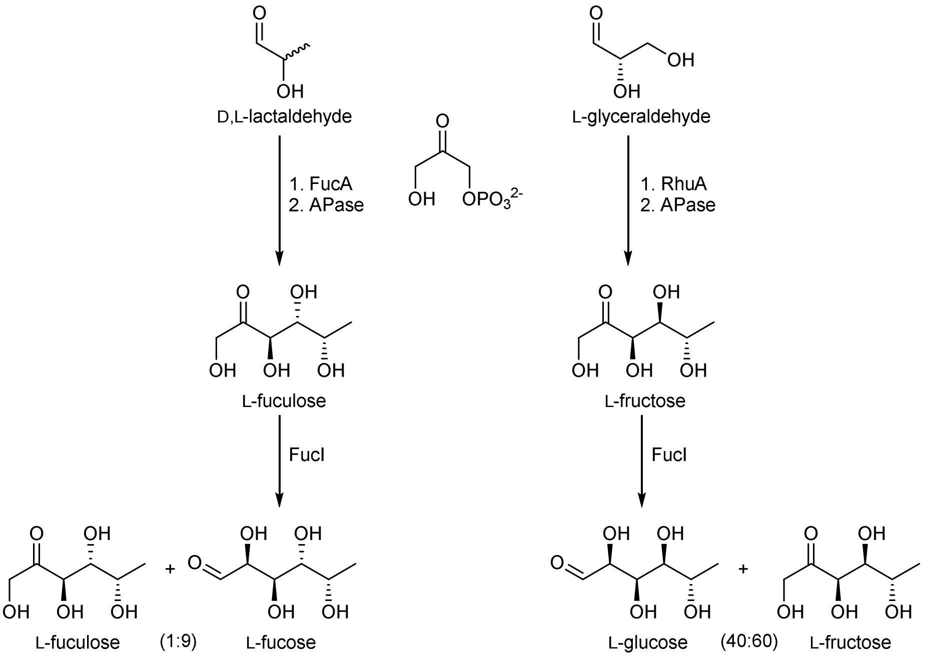
In the method of MacMillan, two subsequent organocatalytic reactions were not observed, while Cordova could develop a two-step protocol under complete amino acid control (Scheme 1.16.) [168, 169]. The authors tested the trimerization of aldehydes, but unfortunately the one-pot procedure only provided trace amount of hexoses. It was observed that yields can be enhanced by the isolation of the product of the first addition, so consequently the hexose synthesis was performed in two vessels.



**Scheme 1.16.** Cordova's hexose synthesis by the trimerization of aldehydes in two steps [169].

It is important to mention that the yields of these procedures for the synthesis of aldohexoses are comparable or higher than most conventional multistep sugar synthesis while the above proposed protocols are far more simple [161].

The other approach is a biomimetic strategy which involves first the synthesis of the donor, DHAP by one of the several methods [17]. Then the synthesis of a ketose by aldol reaction is carried out, followed by dephosphorylation and isomerization. Wong *et al*. described the synthesis of several aldohexoses using the combination of FruA and glucose isomerase [88, 170], and in a similar approach used the same methodology combining different DHAP-dependent aldolases and isomerases for the preparation of several aldohexoses (Scheme 1.17.). For example L-glucose was synthesized from L-glyceraldehyde and DHAP with RhuA, a Pase and fucose isomerase (FucI) [171], and L-fucose was synthesized from D,L-lactaldehyde, DHAP with FucA, a Pase and FucI [172]. Although the bottleneck of the application of this strategy is the limited substrate specificity of the isomerases, moreover the isomerization is reversible and ketoses are generally more stable than aldoses, so the reaction only yields to considerable amount of aldose sugar if it has a stable aldopyranose form [89].

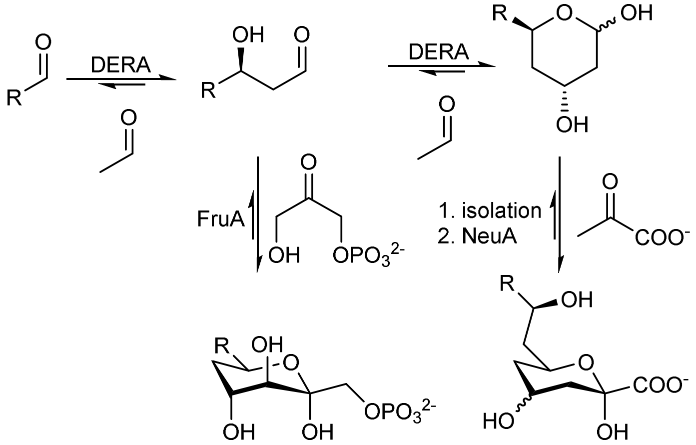


**Scheme 1.17.** Enzymatic synthesis of L-glucose and L-fucose by aldolization and isomerization [162].

## 1.4.2. Synthesis of ketoses

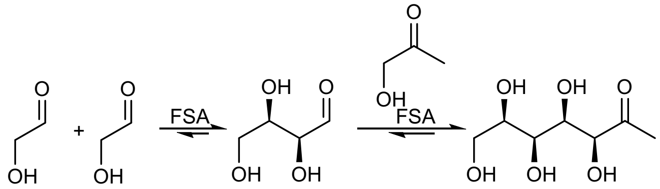
Ketose carbohydrates can be generated with a combination of aldolases in a cascade fashion. The first step of such a cascade has to be the enzymatic generation of the acceptor substrate, followed by the addition of a ketone.

DERA was combined with the DHAP-dependent aldolase FruA and with the pyruvate-dependent aldolase NeuA in tandem reactions to produce ketoses (Scheme 1.18.) [173]. In the combination of DERA and FruA only low amount of DERA was applied so the first reaction stopped at first acetaldehyde addition. In this way a variety of 7-substituted 5-deoxyheptuloses were synthesized in one-pot reaction. The formation of a stable cyclic hemiketal is the driving force for these sequential aldol reactions. Sequential four-substrate aldol reaction was performed combining DERA and NeuA, although it was not possible to perform this combination in one-pot reaction. In this manner, several sialic acid derivatives were synthesized in good yields.



**Scheme 1.18.** Cascade reactions involving three aldolases: DERA, FruA and NeuAc producing ketoses.

It was mentioned earlier that glycolaldehyde can act as a nucleophile in aldol reactions with FSA [127]. Since these products are themselves aldehydes, they can be used as acceptors for a next aldol addition. This was demonstrated by the synthesis of 1-deoxy-D-ido-heptulose, what is the product of the FSA catalyzed dimerization of GO and a second, *in situ* aldol addition of HA to D-threose (Scheme 1.19.). It is interesting to note that no reaction was observed when the donor was changed to DHA from HA.



Scheme 1.19. Cascade self- and cross-aldol reaction catalyzed by FSA.

In the present thesis a couple of approaches have been developed which are presented for the the synthesis of aldose and ketose carbohydrates by coupling aldose reactions in a cascade fashion. First thesynthesis of rare aldose sugars and derivatives by the enzymatic double addition of glycolaldehyde can be mentioned (Chapter 3.2.), and in Chapter 3.4. ogano- and biocatalytical aldol reactions were coupled to prepare deoxysugars.

# 2. Objectives

## 2.1. Main Objectives

The work presented in this thesis is part of a research project which is aiming the preparation of new polyhydroxylated products through new chemo-enzymatic strategies, which enables the control of the constitutional diversity and the structure of the final products.

This work is a contribution to the study of D-fructose-6-phosphate aldolase (FSA) as catalyst in asymmetric carbon-carbon bond formation, the main objective of this thesis was the study of different cascade cross- or self-aldol reactions with the main focus on the key enzymatic step, or steps. The substrates in the focus of this work were mainly small aldehydes and ketones from which polyhydroxylated compounds can be constructed in a building block fashion. The aldol additions were catalyzed by genetically engineered FSA, leading to the investigation of novel reaction types.

## 2.2. Specific Objectives

I. Design of an FSA variant which has high selectivity for glycolaldehyde in cross aldol reactions, and the synthesis of iminosugars from different aminoaldehydes and glycolaldehyde.

   - Design of the library of FSA A129X to improve the enzyme's activity with glycolaldehyde.
   - Combination of active site mutations to improve the variants' selectivity towards glycolaldehyde in competition experiments.
   - Synthesis, purification and characterization of iminosugars.

II. Development of biocatalytic $C_X$+$C_2$+$C_2$ one pot carbohydrate synthesis mediated by the novel FSA variants by two consecutive aldol reactions.

- Synthesis of rare aldose carbohydrates and derivatives by the enzymatic double addition of glycolaldehyde on different aldehyde acceptors.
- Engineering FSA to modify the stereochemical course of the reaction, to allow the one-pot synthesis of L-glucose from two-successive glycolaldehyde additions.
- Purification and characterization of the synthesized aldose carbohydrates.

III. Application of structure-based engineering of FSA to accept an unprecedented variety of novel donor substrates with hydroxymethylketone moiety.

- Design of an active site mutant which converts a wide variety of 1-hydroxy-2-alkanones and related ether components as novel donor substrates.
- Synthesis, purification and characterization of novel deoxysugar products.

IV. Coupling of organo- and biocatalytic aldol reactions in a cascade fashion. The products obtained of the proline catalyzed reactions are tested as acceptors in cross-aldol reactions with FSA or other aldolases, with dihydroxyacetone (DHA), hydroxyacetone (HA).

- Design of active site FSA mutants which accept the organocatalitically produced aldehydes.
- Synthesis, purification and characterization of novel deoxysugar products.

# 3. Results and Discussion

## 3.1. Engineering the donor selectivity of FSA for the asymmetric cross-aldol addition of glycolaldehyde

Aldol additions are important transformations in the toolbox of the synthetic chemists since they are key reactions for the formation of chiral polyhydroxylated molecules as described in the introduction [18, 37, 121, 168, 174]. The aldol reactions where the nucleophilic compounds are also aldehydes are exceptionally valuable, since the products – which themselves are aldehydes – can be used in further reaction, to create complex polyol structures from simple, achiral building blocks [43, 78, 158].

Glycolaldehyde and its derivatives are ideal building blocks for the assembly of aldohexoses by 2 consecutive aldol additions (C2+C2+C2) [158, 161, 175]. However the stereoselective addition of GO with the C-C bond forming methodologies is challenging. It was recently reported by our group that FSA can utilize GO as nucleophile substrate both in homo- and cross-aldol reactions [122], even though this report showed several successful examples for reactions, the addition of GO catalyzed by FSA can be hampered when dealing with weak acceptor substrates.

Our goal became the engineering FSA to have a strong activity and selectivity towards GO, together with a broad tolerance towards structurally diverse acceptors.

### 3.1.1. Engineering of FSA for increased activity with glycolaldehyde

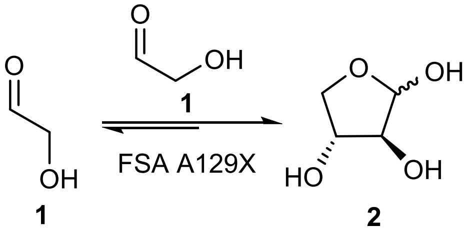
FSA has the unique property that it catalyzes the reaction with at least three structurally different nucleophiles, DHA, HA and GO. Its affinity for the donor substrates follows the order HA>GO>>DHA. To change the catalytic preferences of the fructose-6-phosphate aldolase, the residue A129 was targeted by by Schürmann and Sprenger [6], and the variant A129S showed a remarkable increase in the $k_{cat}/K_M$ value for DHA.

#### *a. Generation and screening the FSA A129X mutant library*

Thus the X-ray structure of FSA and from previous results [121], we hypothesized that the A129 position is the clue for the selectivity and activity towards the donor.

So to verify if any FSA variants with high selectivity towards GO exists, and to prove our hypothesis, this residue was targeted by site saturation mutagenesis.

The chosen screening reaction was the self-aldol reaction (Scheme 3.1.1.) of GO to furnish D-threose (**2**). Our assumption was that any increase in the reactivity would be due to the variants' increased activity towards GO as nucleophile, since the A129 residue is situated in the donor binding site of the enzyme.



**Scheme 3.1.1.** Screening reaction: homo-aldol addition of glycolaldehyde to furnish D-threose catalyzed by FSA wild-type and mutants.

The site saturation mutagenesis was carried out using the QuikChange® site saturation mutagenesis protocol with oligonucleotides bearing NNK (N: G/C/T/A; K: T/G) degeneracy, which covers 32 codons and all 20 proteinogenic amino acids with 95% probability in the case of randomly picking at least 94 colonies. In the screening 96 clones were grown, and expressed for testing in 96-well plates (for procedures see Experimental part). To determine whether the protein expression level can be considered even, the expression of the crude extract of 18 randomly picked clones was checked via SDS-PAGE.

In the screening reactions clones were considered as positives if they furnished >50% of D-threose after 3 hours under identical conditions. The reactions were conducted using 100 μL of lysed cell-free extracts. The reaction mixture of 500 μL final volume contained 100 mM glycolaldehyde. The analysis of the reactions was performed by TLC for "fast-checking" the results, and more accurately by calculating the conversion to D-threose by measuring the peak areas with HPLC (see Experimental section and Electronic Supplementary Material). Of the screened clones 6 showed consistently high activity which were sequenced. Three different mutants were found, with high activity towards GO, namely the alanine was changed for a glycine (two clones), for a threonine (three clones) and a valine (one clone).

For more precise analysis and for the comparison of the three obtained mutants FSA A129G, FSA A129T and FSA A129V, were expressed and purified with the earlier described methods. The purified enzymes were tested under identical conditions (i.e., 150 mM GO in reaction mixtures of 500 μL, catalyzed by 0.5 mg protein in each case) in the synthesis of D-threose (**2**) (Scheme 3.1.1.). The wild-type FSA was also included in the characterization of the mutants, to have a standard for the comparison of the FSA variants and to have a clearer picture of their catalytic abilities. The results of the screening are shown in Figure 3.1.1.. All variants showed increased affinity towards GO in comparison with the wild-type, and among them the most active mutant was FSA A129G with a ~20-fold increase of activity, while the other variants – FSA A129T and FSA A129V – also showed higher reaction rates with a ~10-fold increment in each case.

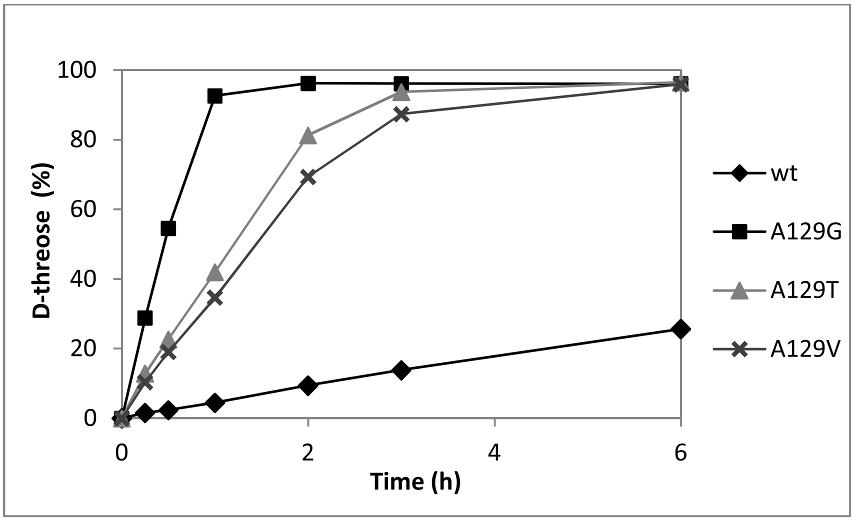


**Figure 3.1.1.** Comparison of the obtained FSA variants and wild-type FSA in the self-aldol reaction of GO. Percentage of D-threose formation is displayed in the function of time. The reactions were catalyzed by 0.5 mg protein. The conversions were calculated with HPLC from the peak areas with an external standard method.

To rationalize our experimental observations computational studies were carried out in our group by Dr. Jordi Bujons with the package Schrödinger Suite 2013.

The activity enhancement relative to FSA wild-type that was observed for mutants FSA A129T and FSA A129V is due to the T129 and V129 exposing a hydrophobic surface to the donor site, similarly to FSA wild type. The models of the corresponding hemiaminals (Figure 3.1.2.) show that the hemiaminal hydroxyl group does not establish any direct interaction with the mutated residues. This could have the effect of favoring the unprotonated state of the essential lysine (K85) by decreasing the p$K_a$ of its ε-amino group, [176] which is required for the formation of the initial hemiaminal with the donor aldehyde.

The activity of the A129G variant is more difficult to rationalize based on the models available (Figures 3.1.2.), since this mutant does not directly introduce any additional interaction with the substrates but instead generates additional empty space in the active site cavity. It could be speculated that this empty space could be either filled with water molecules which could interact with the substrates modifying their reactivity, or that it provides larger flexibility to the protein, allowing some residues to adopt different conformations and modulate the enzymatic reaction.

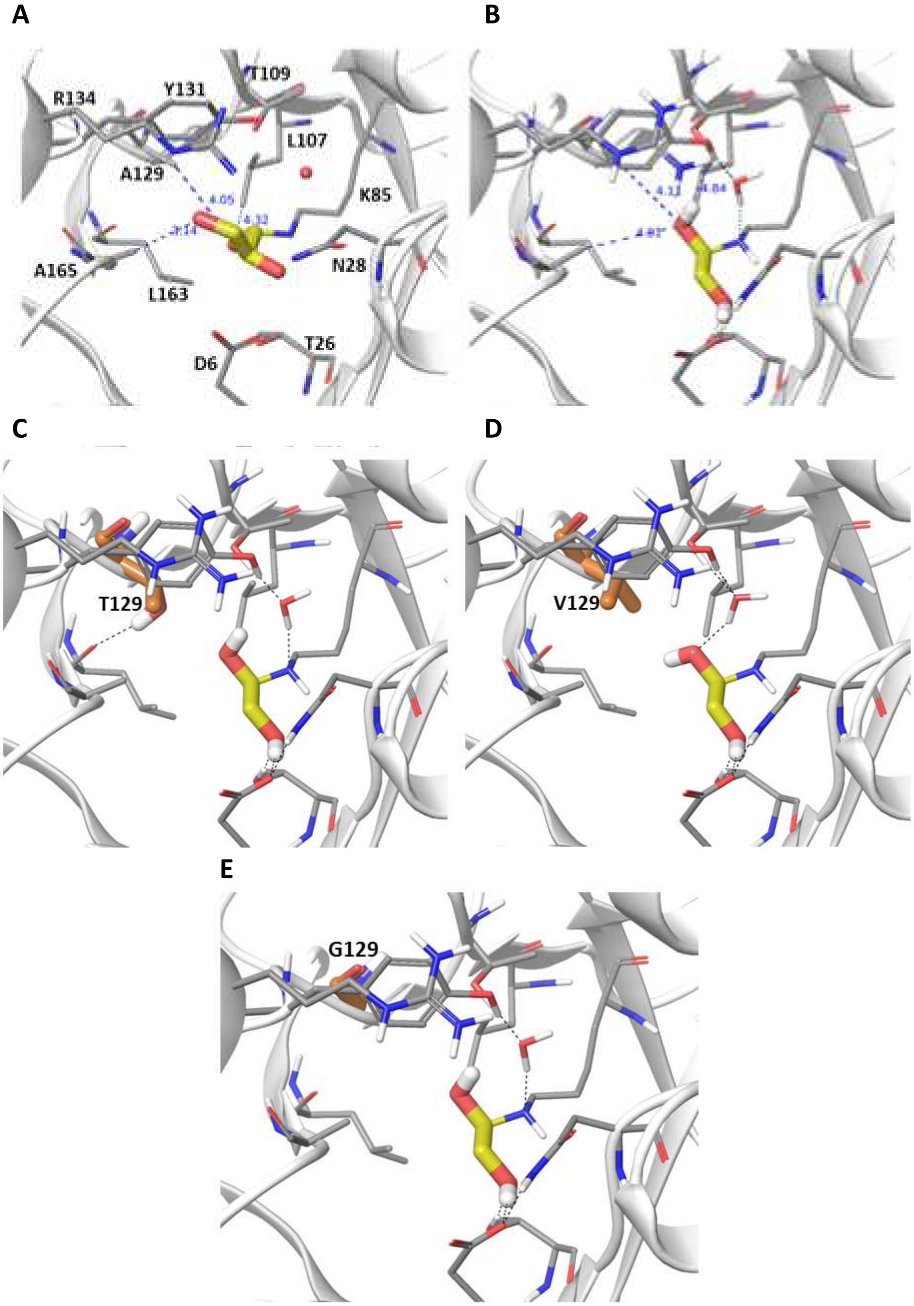
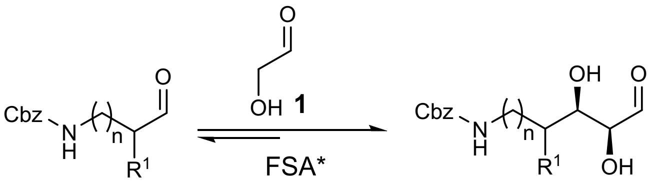


**Figure 3.1.2.** (**A**) Crystal structure (PDB 1L6W [177]) of the FSA active site showing the covalent complex between Lys85 and the hemiaminal of glyceraldehyde. (**B**, **C**, **D**, **E**) Minimum energy structure of the active site of FSA wt (B), FSA A129T (**C**), FSA A129V (**D**) and FSA A129G (**E**) with the Lys85-bound hemiaminal of glycolaldehyde. Hemiaminals (thick yellow), mutated residues (thick orange), near protein residues (thin gray), and a catalytically important water molecule (thin sticks) are shown.

### b. Testing the obtained FSA variants in aldol reactions with N-Cbz-protected amino aldehydes and glycolaldehyde

*N*-Cbz-amino aldehydes are widely used as electrophile substrates in the synthesis of iminocyclitols in our research group. They were proved to have excellent acceptor properties in reactions catalyzed by FSA [122, 130-132]. For this reason, we have decided to use them as model substrates to investigate the applicability of the library of mutants in cross-aldol reaction with GO.

To this end, the activity of the successful FSA variants from the saturation mutagenesis, i.e., FSA A129G, A129T and A129V were assayed as catalysts in the cross aldol addition reactions of GO to four selected amino aldehydes (Scheme 3.1.2.). Besides these variants, the wild-type and two previously constructed mutants, i.e., FSA A129S and FSA A129S/A165G [132] were also included in the tests as controls. FSA A129S/A165G has been proved to expand the catalytic efficiency of FSA in reactions both with HA and DHA, and FSA A129S which changed the donor preference of the aldolase [132]. Thus these enzymes were thought to be relevant examples for the comparison.



| n = 0 | $R_1$ |
|---|---|
| **3** | H |
| (*S*)-**4** | $CH_3$ |
| (*R*)-**4** | $CH_3$ |
| n = 1 | |
| **5** | H |

| n = 0 | $R_1$ |
|---|---|
| **6** | H |
| **7** | $CH_3$ |
| **8** | $CH_3$ |
| n = 1 | |
| **9** | H |

**Scheme 3.1.2.** Cross aldol reactions of GO and selected *N*-Cbz-amino aldehydes with FSA wild-type and selected mutants.

The mutants were screened under identical conditions (i.e., 100 mM GO and 80 mM acceptor aldehyde, using 2 mg $mL^{-1}$ of the FSA catalyst) with an excess of the donor substrate. The concentration of the adduct was measured by HPLC after six hours reaction time (Table 3.1.1.). These are highly unfavorable conditions for the cross-aldol additions, because the D-threose formation (i.e. homo-aldol addition of glycolaldehyde) is favored at high concentrations of GO and short reaction times [122], so for the identification of the best mutants, we used these reaction conditions. It was reasoned that if a mutant can furnish high conversion under these adverse conditions, its performance can be further improved by optimization.

**Table 3.1.1.** FSA wild-type and mutants-catalyzed aldol addition of GO to selected acceptor aldehydes. Percentage of aldol adduct formation after 6 h.

| *N*-Cbz-amino aldehyde | **3** | (*S*)-**4** | (*R*)-**4** | **5** |
|---|---|---|---|---|
| FSA catalyst | Aldol adduct (%) | | | |
| wild-type | 3 | nd | nd | 38 |
| A129S | 12 | nd | nd | 41 |
| A129G | **67** | 2 | nd | **82** |
| A129T | 44 | 2 | nd | **79** |
| A129V | 46 | 1 | nd | **88** |
| A129S/A165G | 42 | 3 | 28 | 39 |

Product formation was measured by HPLC using an external standard method. Percentage calculated with respect to the limiting acceptor aldehyde substrate. Results are the average of three measurements. Highest conversions are shown in bold.

The results revealed that the catalysts FSA wild-type and, FSA A129S furnished low yields with the *N*-Cbz-glycinal (**3**) and only moderate yields with *N*-Cbz-3-aminopropanal (**5**), whilst they did not give any product with the (*S*)- and (*R*)-Cbz-alaninal ((*S*)-**4** and (*R*)-**4**, respectively). FSA A129G gave the best results for *N*-Cbz-glycinal, but A129T and A129V gave also remarkably increased conversions. The three novel A129 variants, i.e., FSA A129G, FSA A129T and FSA A129V, provided excellent conversions for the reaction of *N*-Cbz-3-aminopropanal (**5**) with glycolaldehyde. While the (*S*)-and (*R*)-Cbz-alaninal were not accepted. The variant FSA A129S/A165G tolerated most of the amino aldehydes, except (*S*)-Cbz-alaninal ((*S*)-**4**), although only with moderate yields.

These results verified our earlier observations that the best mutant with the highest activity towards GO is A129G. This variant gave excellent results with the non-branched acceptors, but unfortunately did not tolerate the sterically more demanding aldehydes such as (*S*)-**4** and (*R*)-**4**.

### c. *Improving acceptor tolerance by adding further substitutions in the active site of the obtained FSA variants*

In a study of our research group, the acceptor binding site of the FSA was modified to improve the enzymes catalytic performance towards *N*-Cbz amino aldehydes, and the residue A165 was altered. Substitution of glycine for alanine allows more space and flexibility to allocate the aldehydes especially those with branching on the C-α. The result revealed that both enantiomers of N-Cbz-alaninals (which were not substrates for the FSA *wt*) were tolerated as acceptors by the mutant [132]. In this study the double mutant FSA A129S/A165G was also designed, and it catalyzed the additions in the additions of DHA and HA to *N*-Cbz-amino aldehydes remarkably high activity.

As our results above demonstrate (Table 3.1.1.) that (*S*)-**4** and (*R*)-**4** are not tolerated by the single mutants, only by FSA A129S/A165G, so to accommodate the C-α branching of these substrates additional mutations are necessary in the acceptor binding site. We thought that a broader acceptor tolerance could be achieved by combining the substitutions that are favorable for GO as donor (e. g. A129G and A129T) and the one for branched electrophiles (A165G). In this manner two double mutants FSA A129G/A165G and FSA A129T/A165G were constructed by site-directed mutagenesis. The modifications were verified by DNA sequencing and the protein was expressed and purified with the previously described methods (see Experimental Section).

These enzymes were tried in similar screening reactions as described above. Their catalytic performance was evaluated in the formation of D-threose (Scheme 3.1.1., Figure 3.1.3.) and also in the reaction of selected *N*-Cbz-protected amino aldehydes and glycolaldehyde (Sheme 3.1.2.).

The activity of the double mutant FSA A129G/A165G towards the formation of D-threose (i.e., homo-aldol addition of **1**) considerably diminished. This loss of activity can be explained by the effect of introducing two glycine residues, which can result in abolishing all interactions between residues and the substrates during the different steps of activation and catalysis. On the other hand the variant FSA A129T/A165G exhibited a slight increase in the activity compared to A129G and A129T in the homo-aldol addition of **1** (Figure 3.1.2.).

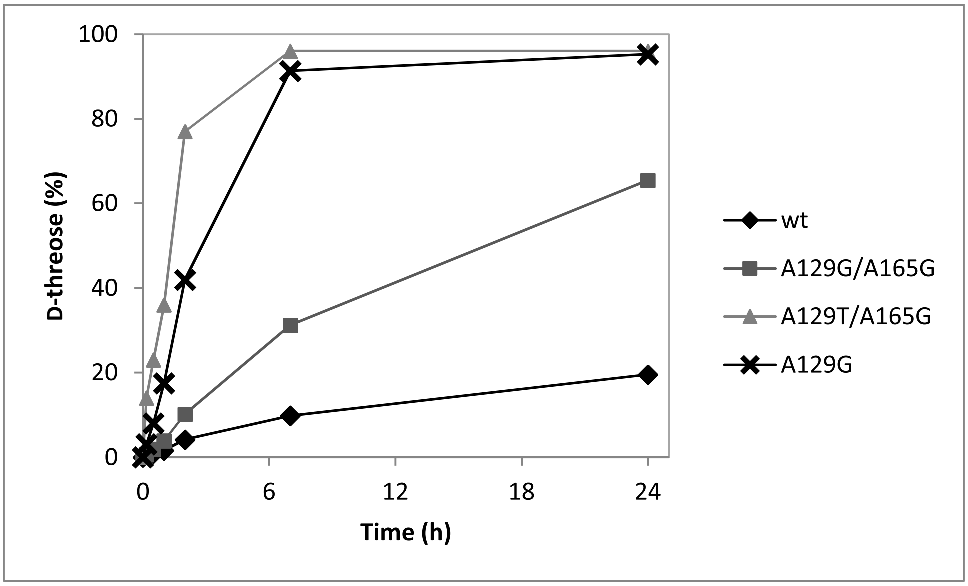


**Figure 3.1.3.** Comparison of the obtained FSA variants and wild-type FSA in the self-aldol reaction of GO. Percentage of D-threose formation is displayed in the function of time. The reactions were catalyzed by 0.5 mg protein. The conversions were calculated with HPLC from the peak areas with an external standard method.

In the reactions with the *N*-Cbz-protected amino aldehydes (Table 3.1.2.), it was observed that the yield with the variant FSA A129G/A165G decreased more than 2-fold in the formation of **3** and by 4-fold in the formation of **5**, compared to both the variants A129G and FSA A129T/A165G. The other designed double mutant FSA A129T/A165G for the non-branched aldehydes (**3** and **5**) furnished similar conversions to FSA A129G, and compared to the results with FSA A129T the yield to **6** increased almost 2-fold. As it was expected from the mutants with Gly substitution in the A165 position [132] both variants FSA A129G/A165G and FSA A129T/A165G tolerated the branched amino aldehydes (*R*)-**4** and (*S*)-**4** as acceptors, however the A129G/A165G mutant gave very low conversions, and even FSA 129T/A165G hardly improved the yields of the cross-aldol reactions, compared to FSA A129S/A165G (Table 3.1.2.).

**Table 3.1.2.** FSA wild-type and mutant-catalyzed aldol addition of GO (**1**) to selected acceptor aldehydes. Percentage of aldol adduct formation after 6 h.

| *N*-Cbz-amino aldehyde | **3** | (*S*)-**4** | (*R*)-**4** | **5** |
|---|---|---|---|---|
| FSA catalyst | % aldol adduct | | | |
| wild-type | 3 | nd | nd | 38 |
| A129G/A165G | 29 | 6 | 8 | 21 |
| A129T/A165G | 67 | 6 | 31 | 75 |

Product formation measured by HPLC using an external standard method; percentage calculated with respect to the limiting acceptor aldehyde substrate; **1** (100 mM), acceptor aldehyde (80 mM), FSA catalyst (2 mg $mL^{-1}$ of protein). Results are the average of three measurements. nd: no product detected.

To further extend the FSA's activity in the cross-aldol reactions, especially with C-α branched aldehydes, we decided to add a further mutation on the platform of A129G/A165G to reduce the "nakedness" (lack of interactions) of the donor binding site. We found the L107 position as a possible target for these changes. It was hypothesized that changing the leucine to a tyrosine might restore some polar interactions. So the mutant FSA L107Y/A129G/A165G was constructed by QuickChange® site directed mutagenesis. The modifications were verified by DNA sequencing and mass spectroscopy of the purified enzyme. The protein was expressed, purified and characterized following the previously described protocols (see Experimental Section).

The resulted variant showed extremely high activity in the homo-aldol addition of GO (furnished almost full conversion to D-threose in two hours) and it also gave high yields in the cross aldol reaction with the different amino aldehydes (88% to **6** and 79% to **9**). (Figure 3.1.5., Table3.1.3.) This FSA variant furnished the highest conversion with the *N*-Cbz-(*R*)-alaninal (72%), and it was the only enzyme that catalyzed the product formation with the *N*-Cbz-(*S*)-alaninal (32%).

According to the computational models a similar structure of the hemiaminal of GO can be observed in the wild-type and in FSA L107Y/A129G/A165G (Figure Figure 3.1.4.), but in the later case the hemiaminal hydroxyl group acts as hydrogen bond donor to the phenol group of Y107, which at the same time is hydrogen-bonded to the backbone carbonyl of L163. This interaction should increase the basicity of the hydroxyl group of the hemiaminal, facilitating its protonation and dehydration to yield the corresponding Schiff base.

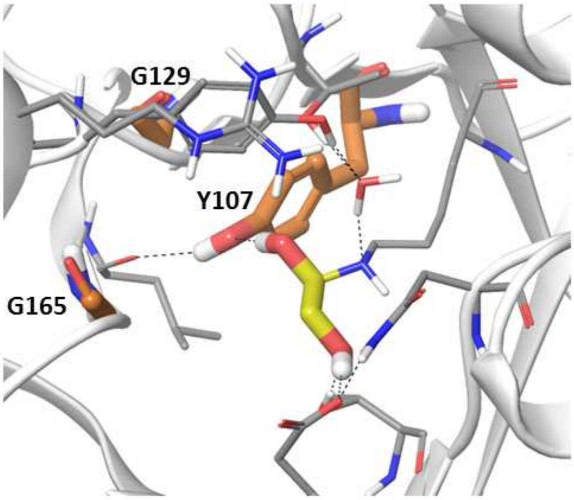


**Figure 3.1.4.** Minimum energy structure of the active site of FSA L107Y/A129G/A165G, with the Lys85-bound hemiaminal of glycolaldehyde. Hemiaminal (thick yellow), mutated residues (thick orange), near protein residues (thin gray), and a catalytically important water molecule (thin sticks) are shown.

To investigate the effect of each changed residue in the triple mutant FSA L107Y/A129G/A165G, and their possible synergy, the single and double mutants FSA L107Y and FSA L107Y/A129G were also designed by site directed mutagenesis. And the enzymes were expressed and purified, as described before. In the formation of threose FSA L107Y showed low activity. Nevertheless when the methyl group of the alanine was removed (FSA L107Y/A129G) the conversion to threose increased considerably (Figure 3.1.5.). The activity of the double mutant is slightly higher than of the single mutant A129G, while it is slightly lower than the activity of the triple mutant FSA L107Y/A129G/A165G, by synergistic effects, of the three substitutions similar to the ones observed in the case of the double mutant FSA A129S/A165G [132].

To further examine the effect of the polar Tyr residue, a phenylalanine was introduced in the position 107. We thought that the difference between the activity of the L107F/A129G and L107Y/A129G can be due to the intensified polar effect of the phenolic group of the tyrosine. And looking at the reaction conversions in the formation of D-threose (i.e. homo-aldol addition of **1**) it seems that the lack of the polar environment created by the tyrosine is detrimental for the catalysis (Table 3.1.3.).

After the confirmation of the positive effect of the polar tyrosine residue, we sought other potential amino acids that can have similar effects. We discovered that histidine can be an alternative, so both the double mutant FSA L107H/A129G and FSA L107H/A129G/A165G were designed and expressed with the previously described methods, and indeed both of them showed similar, but slightly lower conversions than their counterparts with tyrosine, in the self-aldol reaction of glycolaldehyde (**1**) (Figure 3.1.5.).

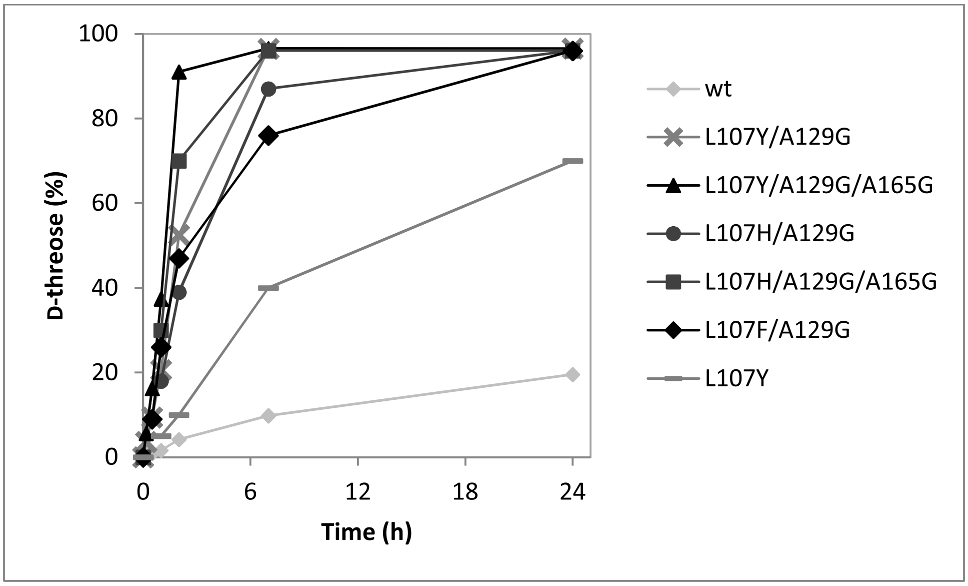


**Figure 3.1.5.** Comparison of the obtained FSA variants and wild-type FSA in the self-aldol reaction of GO. Percentage of D-threose formation is displayed in the function of time. The reactions were catalyzed by 0,5 mg protein. The conversions were calculated with HPLC from the peak areas with an external standard method.

All designed mutants were also tested in the cross aldol reactions of glycolaldehyde and the set of chosen *N*-Cbz-protected amino aldehydes. The mutants where the polar residues were introduced together with the A129G substitution – the variants FSA L107Y/A129G, FSA L107H/A129G and FSA L107H/A129G/A165G – act with similar activity as the FSA L107Y/A129G/A165G triple mutant in the reaction of *N*-Cbz-glycinal (**3**) and GO, however with **5** the conversions with the 3 variants (FSA L107Y/A129G, FSA L107H/A129G and FSA L107H/A129G/A165G) were much lower in comparison with FSA L107Y/A129G/A165G. All tested *N*-Cbz-amino aldehyde acceptors (**3**-**5**) furnished lower conversions with the single mutant L107Y and the L107F/A129G.

For the reaction of the C-α branched acceptor aldehydes the presence of the A165G mutation seems to be indispensable for the catalysis. FSA L107H/A129G/A165G showed some activity with (*R*)-**4** and (*S*)-**4**, furnishing both cases around half of the conversion than FSA L107Y/A129G/A165G. These results discouraged further investigations with FSA variant bearing the combination L107H/A129G.

**Table 3.1.3.** FSA wild-type and mutant-catalyzed aldol addition of GO (**1**) to selected acceptor aldehydes. Percentage of aldol adduct formation after 6 h.

| *N*-Cbz-amino aldehyde | **3** | (*S*)-**4** | (*R*)-**4** | **5** |
|---|---|---|---|---|
| FSA catalyst | % aldol adduct | | | |
| wild-type | 3 | nd | nd | 38 |
| L107Y/A129G/A165G | **88** | **32** | **73** | **79**[a)] |
| L107Y | 40 | nd | 1 | 40 |
| L107Y/A129G | 85 | 4 | 2 | 57 |
| L107F/A129G | 32 | nd | nd | 31 |
| L107H/A129G | 83 | 3 | nd | 57 |
| L107H/A129G/A165G | 88 | 14 | 31 | 41 |

Product formation measured by HPLC using an external standard method; percentage calculated with respect to the limiting acceptor aldehyde substrate; **1** (100 mM), acceptor aldehyde (80 mM), FSA catalyst (2 mg mL$^{-1}$ of protein). Results are the average of three measurements. a) The percentage of product corresponded to the area of the formed two diastereoisomers observed in the HPLC chromatogram during the reaction monitoring. nd: no product detected.

To explain the behavior of these variants the computational models of the double and single mutants L107Y/A129G and L107Y were created and they were compared to the model of the variant L107Y/A129G/A165G. In the comparison a gradation of activities was observed. First, as already mentioned, the absence of the A165G mutation causes a very low or null capacity to transform substrates (*R*)-**4** and (*S*)-**4** (Table 3.1.3.). In addition, while L107Y/A129G gave similar conversions to L107Y/A129G/A165G in the aldol addition of **1** to **3** and to **5**, mutant L107Y was less effective (Table 3.1.3.). This was also observed for the homo aldol addition of **1**, where the activities followed the order triple mutant>double mutant>>single mutant. The model of the hemiaminal of **1** bound to the L107Y variant (Figure 3.1.6.A) shows that the methyl group from A129 is oriented towards the side chain of Y107, “pushing” it away from the backbone carbonyl of residue L163. This prevents the formation of the hydrogen bond interaction between the Y107 phenol group and L163, which was detected for the triple mutant (L107Y/A129G/A165G, Figure 3.1.4.) and which is also present in the double mutant (L107Y/A129G, Figure 3.1.6.A).

The models show a weaker and altered interaction with the hydroxyl group of the hemiaminal, with the L107F/A129G mutant (Figure 3.1.6.C), which shows an even lower activity than FSA L107Y. On the contrary, the models for the mutants that contain a His residue in position 107 (L107H/A129G and L107H/A129G/A165G) show different hydrogen bond interaction patterns with the hemiaminals (Figures 3.1.6.D-E). Hence, their relatively high activity could be more related to the acid-base properties of the histidine residue, which could have an influence on the protonation/deprotonation equilibria of the hemiaminal intermediates.

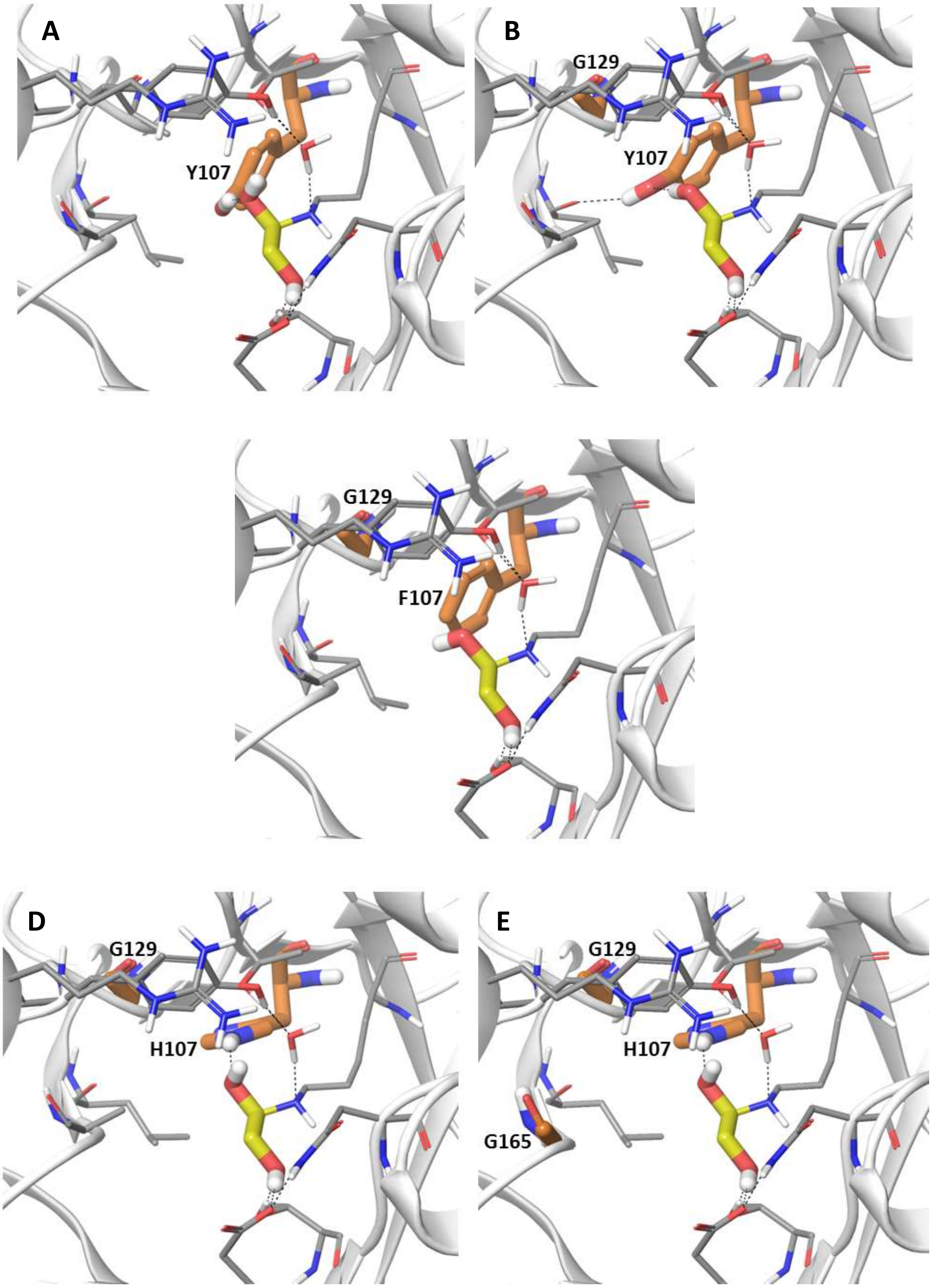


**Figure 3.1.6.** Minimum energy structures of the active site of mutants (**A**) L107Y, (**B**) L107Y/A129G, (**C**) L107F/A129G, (**D**) L107H/A129G, and (**E**) L107H/A129G/A165G, with the Lys85-bound hemiaminal of glycolaldehide. Hemiaminals (thick yellow), mutated residues (thick orange), near protein residues (thin gray), and a catalytically important water molecule (thin sticks) are shown.

### 3.1.2. Kinetic parameters of the newly designed FSA mutants

The steady state kinetic of the self aldol reaction of glycolaldehyde was determined by the group of Professor Lemaire from the Université Blaise Pascal, Clermont-Ferrand, France.

Measuring the steady state kinetic parameters of the mutants was necessary to get a better comprehension of the catalytic features of the newly designed variants, particularly the ones bearing the A129G and the L107Y mutations. The steady state kinetics was measured with glycolaldehyde both as donor and acceptor. The apparent $V^{app}_{max}$ and $K^{app}_{M}$ were calculated for the enzymatic homo aldol addition of glycolaldehyde using the methodology reported by Neuhaus [178].

From the apparent catalytic parameters calculated for GO as donor ($V^{app}_{max}/K^{app}_{M/donor}$) the best FSA variant was the triple mutant FSA L107Y/A129G/A165G (Table 3.1.4.) with more than 70-fold increment in the efficiency compared to the wild-type. A great elevation was also measured in the efficiency of the double mutant FSA L107Y/A129G, with a ~50-fold increase. The FSA A129G mutant also increased the efficacy of the catalysis by ~5-fold. The FSA variant that provided the smallest change compared to *w*t was the single mutant L107Y.

The catalytic efficiency for GO as acceptor was the highest again in the case of FSA L107Y/A129G/A165G with a more than 1000-fold increase over the wild-type. FSA L107Y/A129G also showed a more than 700-fold raise in the efficacy, while the single mutants acted very similarly both with a ~15-fold increment.

Examining the results of Table 3.1.4. it appears that the change in position A129 for a glycine is the variation that dominates the catalytic properties, since all the mutants with highly increased efficiency had this change. It is also apparent that the L107Y combined with A129G enhances the efficacy in the formation of threose and increases the enzymes efficiency towards glycolaldehyde both as donor and as acceptor, thus a synergistic effect can be observed.

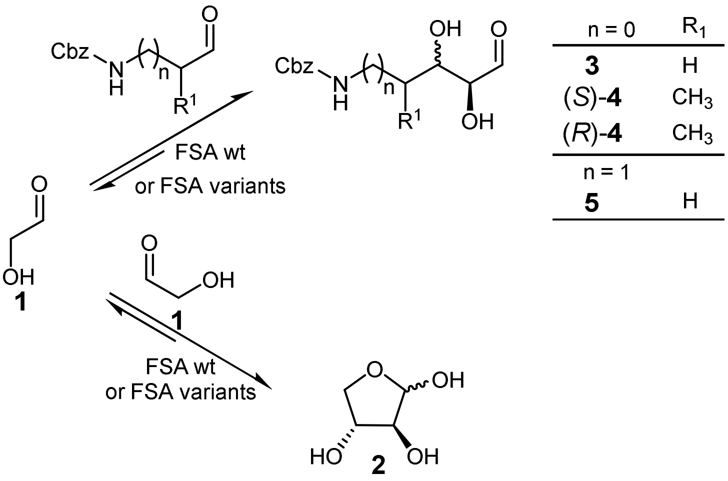
**Table 3.1.4.** Steady state kinetic parameters for glycolaldehyde in the homo aldol addition reaction catalyzed by D-fructose-6-phosphate aldolase wild-type and variants.

| | **Wild-Type**$^{b)}$ | **A129S/ A165G** | **A129G** | **L107Y/ A129G/ A165G** | **L107Y** | **L107Y/ A129G** |
|---|---|---|---|---|---|---|
| $\mathbf{K^{app}_{M/donor}}$ | 0.197±0.045 | 5.8±0.9 | 3±1 | 0.11±0.03 | 0.12±0.02 | 0.094±0.008 |
| $\mathbf{K^{app}_{M/acceptor}}$ | 62.8±5.5 | 340±90 | 287±3 | 2.5±0.5 | 4.6±0.8 | 2.1±0.3 |
| $\mathbf{V^{app}_{max}}$ | 0.22±0.02 | 7.2±1.5 | 15.6±0.1 | 8.8±0.4 | 0.240±0.003 | 5.3±0.3 |
| $\mathbf{V^{app}_{max}/K^{app}_{M/donor}\ /10^{-6}}$ | 1117 | 1241 | 5200 | 80000 | 2000 | 56383 |
| $\mathbf{V^{app}_{max}/K^{app}_{M/acceptor}\ /10^{-6}}$ | 3.5 | 21.2 | 54.4 | 3520 | 52.2 | 2524 |

a) Km (mM); $V_{max}$ (U mg$^{-1}$ protein); U=10$^{-6}$ mol min$^{-1}$. For further details, see experimental section. b) Results from Garrabou et al.[122]

Initial reaction rates ($v_0$, μmol min$^{-1}$ mg$^{-1}$ of protein) were also measured for the cross-aldol reaction of glycolaldehyde to the *N*-Cbz-amino aldehydes. The purpose was to gain a better insight into new enzymes ability to catalyze these reactions comparing the extent of the competition between the cross and the homo aldol addition reactions. The experiments were conducted under identical conditions using 10 mM glycolaldehyde, to not reach the saturation concentration of GO and 80 mM of the amino aldehyde. The enzyme amount was adjusted to ensure linear dependence of the product formation versus time (0-10 min). To calculate the initial velocities the reactions were only monitored up to 10% conversion to both the cross- and self-aldol products (Scheme 3.1.3., Table 3.1.5.).

It is important to notice, that due to the low $K_M$ values of glycolaldehyde as acceptor, as a rule, results of the cross-aldol reactions can be improved in the concentration of GO is kept at a low level (e.g. at <10 mM). This way saturation concentration is reached for glycolaldehyde as donor, but not as acceptor.



**Scheme 3.1.3.** Scheme of the self- and cross-aldol reactions studied for the determination of the initial reaction rates with the newly designed FSA mutants.

The initial velocity of the cross- aldol reactions of the examined FSA variants were in a good agreement with the yields obtained. As expected, the highest activities were found with FSA A129G and FSA L107Y/A129G/A165G for the self- and cross-aldol reactions as well. In most cases the amino aldehydes were the preferred acceptors over glycolaldehyde, especially in the case of the FSA variants baring the A129G mutation. The ratio between the initial velocities in the case of the non-branched acceptors was at least 3-fold higher with both FSA A129G and FSA L107Y/A129G/A165G. In the case of the (R)-4, which can be considered as poor acceptor substrate, the ratio $v_{0\ cross}/v_{0\ self}$ was around 1 for the triple mutant, and the threose formation exceeded the cross-aldol reaction with FSA A129S/A165G. The (S)-4 can be considered as the most difficult acceptor substrate from the tested ones. With this electrophile, the self-aldol reaction exceeded the cross-aldol even in the case of FSA L107Y/A129G/A165G.

The $K_M$ of glycolaldehyde both as donor and acceptor are the lowest in the case of the FSA mutants bearing the tyrosine residue in the 107 position (Table 3.1.4.), but this low $K_M$ value for GO as acceptor did not seem to have an effect on the cross aldol reaction (Table 3.1.5.).

**Table 3.1.5.** Initial rate of cross-aldol reaction ($v_{0cross}$ μmol min$^{-1}$ mg$^{-1}$ of protein) and the ratio of cross/self aldol reaction. The FSA variants are compared with wt and FSA A129S/A165G in the cross-aldol addition reactions of **1** to *N*-Cbz-aminoaldehydes and the self-aldol reaction.

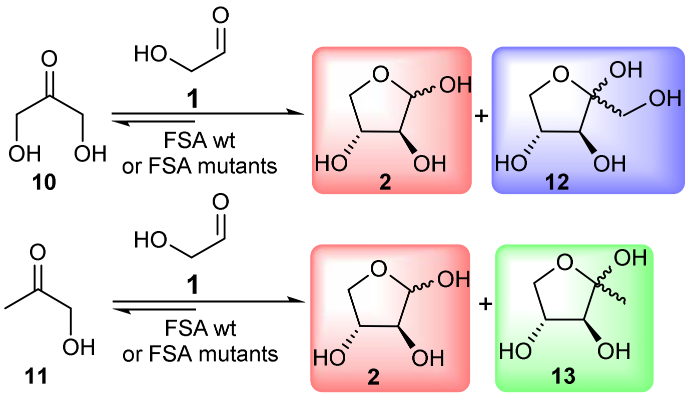
| | $v_{0\ cross}$ [a), b)], ($v_{0\ cross}/v_{0\ self}$) (yield%) [c)] | | | |
|---|---|---|---|---|
| **Aldehyde** | **Wild-Type** | **A129S/ A165G** | **A129G** | **L107Y/ A129G/ A165G** |
| **3** | nr | 0.39,<br>(1.4), (42) | 0.95,<br>(5.1), (67) | 3.1,<br>(3.3), (88) |
| **(*S*)-4** | nr | nd | nr | 0.26,<br>(0.7), (32) |
| **(*R*)-4** | nr | 0.15,<br>(0.5), (28) | nr | 1.2,<br>(1.1), (73) |
| **5** | 0.55,<br>(6.7), (38) | 0.41,<br>(1.8), (39) | 2.4,<br>(5.7), (82) | 2.1,<br>(3.6), (79)[d)] |

a) Different amounts of enzyme were used for obtaining linearity between 0% and 10% of aldol adduct. For details, see experimental section. b) Values are the mean of three-four determinations. c) Values from Tables 3.1.1 and 3.1.3. d) The $v_{0cross}$ corresponded to the formation of the two diasteroisomers. nd: not determined; nr: no reaction.

### 3.1.3. The newly designed mutants selectivity towards the donor substrates

As a next step we endeavored to investigate the designed enzymes' selectivity towards GO in competition experiments with two of the hydroxyketone donors, hydroxyacetone (**11**) and dihydroxyacetone (**10**). The selectivity was examined, by determining which is the preferential donor substrate i.e., GO (**1**), HA (**11**) or DHA (**10**) of the enzymes. Hence the initial reaction rates ($v_0$) of the formation of D-threose (**2**) versus D-xylulose (**12**) and the formation of D-threose (**2**) versus 1-deoxy-D-xylulose (**13**) were measured under identical conditions (Scheme 3.1.4.).

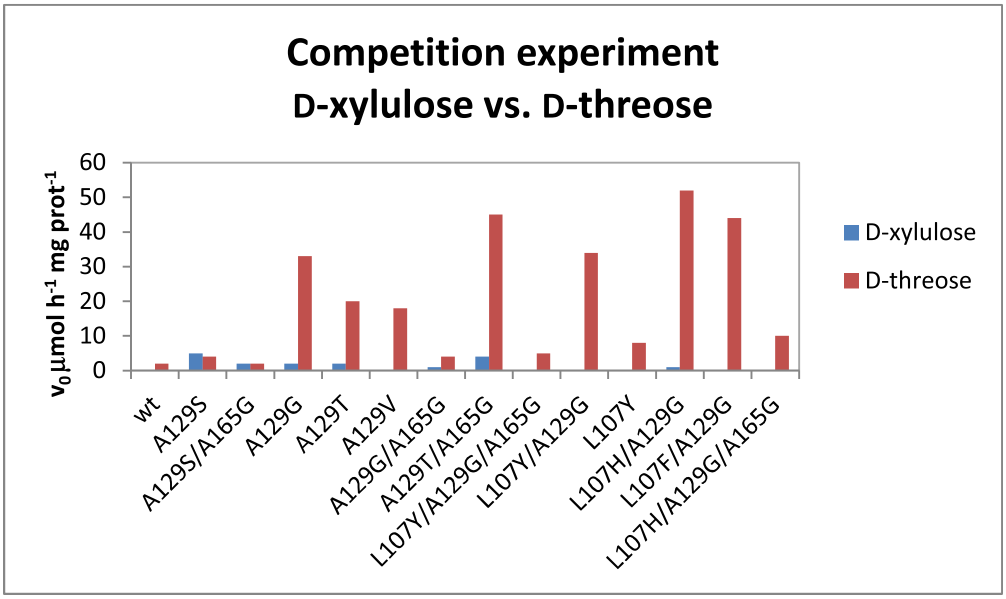
Two set of separate experiments were carried out in 500 µL volume with mixtures of **10**+**1** (100 mM each) and of **11**+**1** (also 100 mM each), catalyzed by 0.5 mg enzyme. To ensure linear dependence of the formation versus time samples were taken in the beginning of the reaction at 0, 20 min, 30 min and 1 hour, and the initial reaction rates to each product were calculated from the product concentrations measured with the HPLC.



**Scheme 3.1.4.** Donor competence reactions between glycolaldehyde, dihydroxyacetone (**10**) and hydroxyacetone (**11**) leading to the formation of D-threose (**2**), D-xylulose (**12**) and 1-deoxy-D-xylulose (**13**).

Under these conditions most of the FSA variants showed a preference for D-threose formation (Figure 3.1.4.) consistent with the fact that $K_M$ value for **1** as donor for the wild type is ~10-fold lower than that of **10** and **11** [127, 132].

**A**



**B**

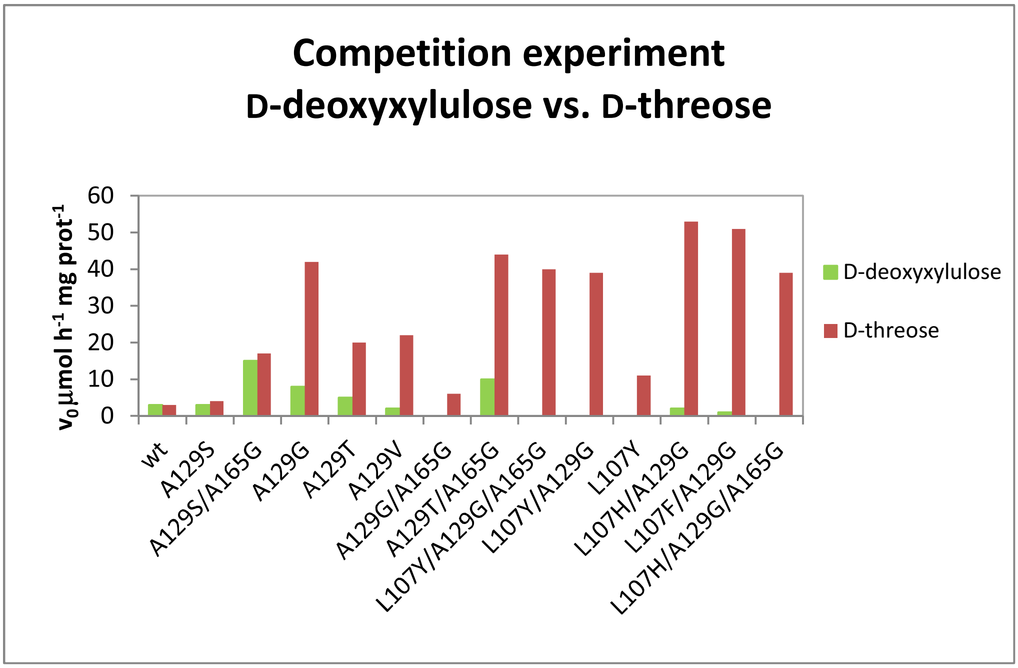


**Figure 3.1.7.** Donor selectivity of FSA wild-type and variants. Donor substrate competition experiments between **1** and **10** (A) and **1** and **11** (B).

The mutants bearing with L107Y substitution gave exclusively D-threose and the L107H mutants also showed a strong preference for the homo-aldol addition of **1**. The donor DHA was not accepted with any variants with L107Y and L107H mutation. As expected from previous results FSA A129G/A165G and FSA L107Y show little activity towards all donors.

Computational models of the D-F6P bound to the active site show steric impediment for the terminal hydroxymethyl group by the side chain of Y107. This group on the on wild-type FSA would occupy the deeper part of the active site cavity. The same applies to the rest of L107 mutants which also show a low activity for DHA.

Unexpectedly low activity for the formation of D-threose was observed for both triple mutants FSA L107Y/A129G/A165G and FSA L107H/A129G/A165G in the reaction of **10**+**1**. This low activity was due to inhibition of the variants by dihydroxyacetone with an $IC_{50}$ = 6,7 mM and 78,6 mM, respectively (Figure 3.1.5.).

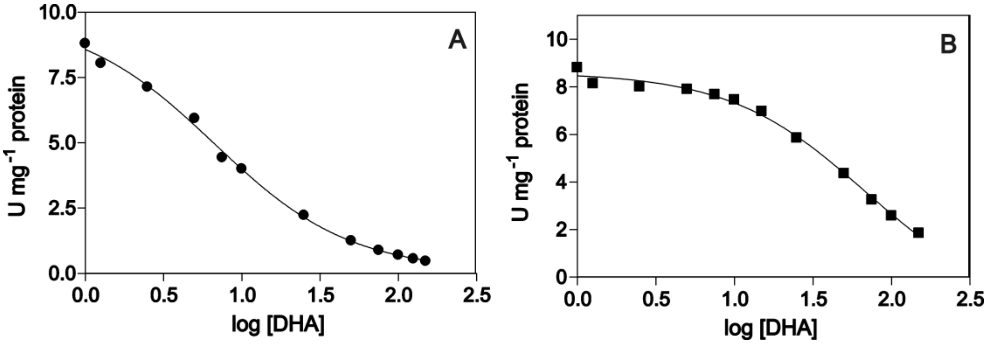


**Figure 3.1.8.** Inhibition of FSA L107Y/A129G/A165G (A) and FSA L107H/A129G/A165G (B) by dihydroxyacetone (DHA). Units of FSA L107Y/A129G/A165G (A) and L107H/A129G/A165G (B) vs. the logarithm of DHA concentration. The reactions solutions of GO (0.05 mmol, 100 mM) was added to triethanolamine buffer (400 µL) and DHA (12 different concentrations: 0, 1, 2.5, 5, 7.5, 10, 25, 50, 75, 100, 125 and 150 mM) was added to the mixture. The reactions were started by adding the respective FSA mutant. Samples (10 µL) were withdrawn at 0, 2, 5, 10 20 min reaction time, to ensure a linear dependence of the product formation versus time at the beginning of the reaction, and followed the sample derivatization protocol prior to HPLC analysis; U of activity: the amount of D-threose formation per minute per mg of protein at 25 ºC.

*In situ* NMR reaction monitoring was carried out on the additions **10**+**1** (Figure 3.1.9.) and **11**+**1** (Figure 3.1.10.)under optimized conditions of excess of **10** and **11** donors. With these conditions FSA variants A129S and A129S/A165G gave D-xylulose (**12**) and 1-deoxy-D-xylulose (**13**), respectively as ascertained in the NMR experiments. In contrast, the triple mutant FSA L107Y/A129G/A165G gave exclusively D-threose even in the presence of an excess of **10** and **11**.

Hence a set of mutants was designed with complementary donor preference compared to the FSA variants described before. These catalysts are valuable new entries for the cross-aldol additions of glycolaldehyde, expanding the biocatalytic possibilities towards the preparative applications of aldose like carbohydrate compounds.

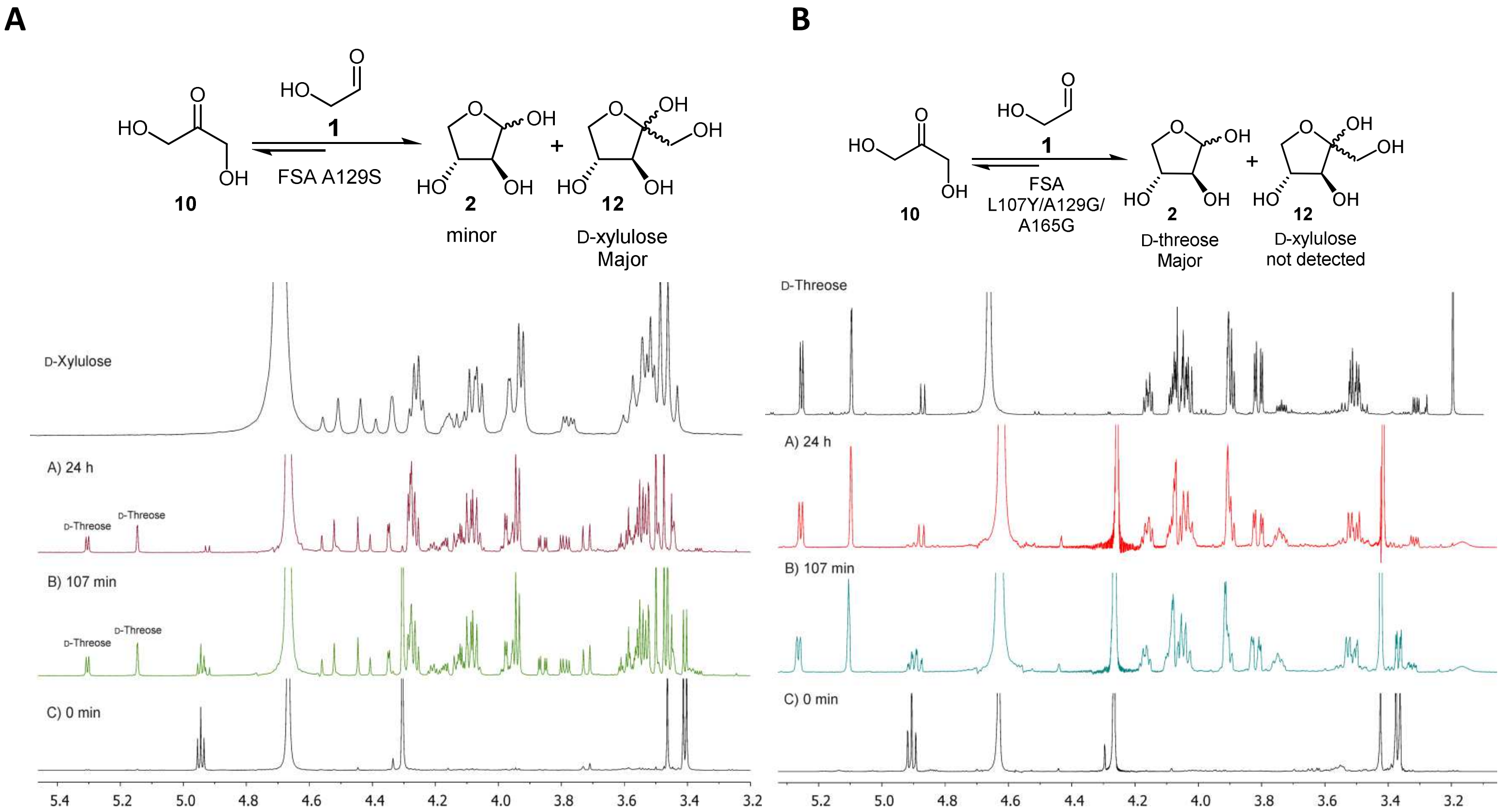


**Figure 3.1.9.** Selectivity of FSA A129S (**A**) and FSA L107Y/A129G/A165G (**B**) towards donor substrates. In situ NMR donor competence experiments between glycolaldehyde (**1**) and dihydroxyacetone (**10**).

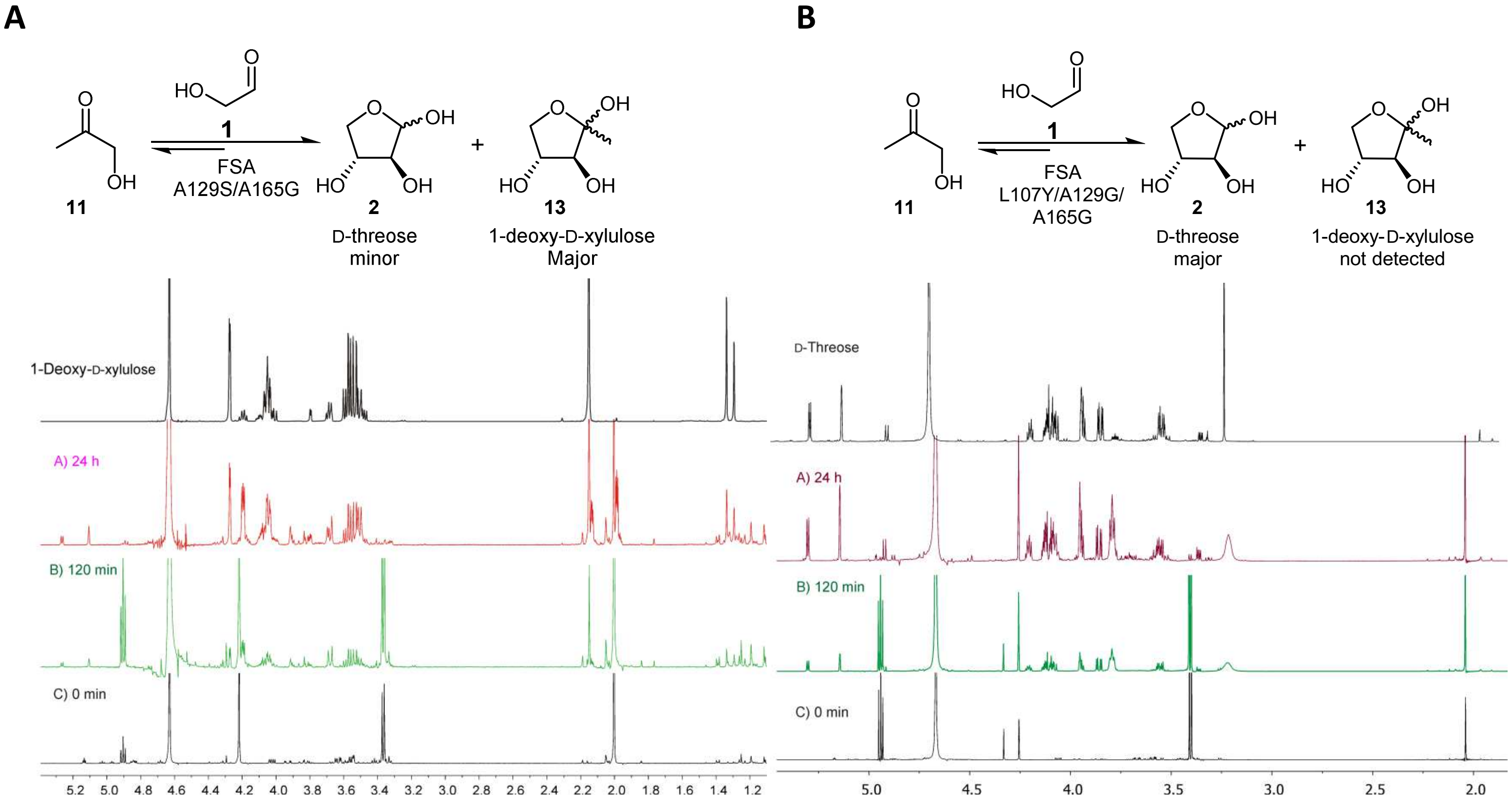


**Figure 3.1.10.** Selectivity of FSA A129S/A165G (**A**) and FSA L107Y/A129G/A165G (**B**) towards donor substrates. In situ NMR donor competence experiments between glycolaldehyde (**1**) and hydroxyacetone (**11**).

## 3.1.4. Synthetic examples

To prove the synthetic value of the newly designed mutants the preparative scale reactions of the selected *N*-Cbz-aminoaldehydes with glycolaldehyde were carried out in close collaboration with Anna Soler, a PhD student of our research group. The structure and the stereochemistry of the isolated products were also examined.

The reactions were conducted with portion wise addition of GO and were scaled up to 10 mL of total volume. The concentration of the aldehyde was 80 mM, and 24 mg $mL^1$ enzyme was added to each reaction. The conversions to the aldol adducts were high (between 80-98%) in all cases.

The stereochemical outcome of the reactions was established by high field NMR was determined from the hemiaminals **14a-b** and **17a-b** and from the synthesized iminocyclitols **15** and **16** (Scheme 3.1.5.).

Some of the obtained aldol adducts were treated with reductive amination to unequivocally determine the relative configuration of each stereocenter by NMR spectroscopy (Figure 3.1.11.). To achieve the iminocyclitols the reductive amination with $H_2$ (50 psi) treatment in the presence of Pd/C was realized. In these reactions the hydrogenation takes place by the opposite side than the hydroxyl of the C4, independently of the stereochemistry of the other substituents [86, 97, 100, 179, 180], therefore in the majority of cases a syn configuration can be observed between the centers C2 and C4. Also in the reductive amination the stereochemistry of the centers C3, C4 and C5 is conserved, no epimerization occurs [179, 180].

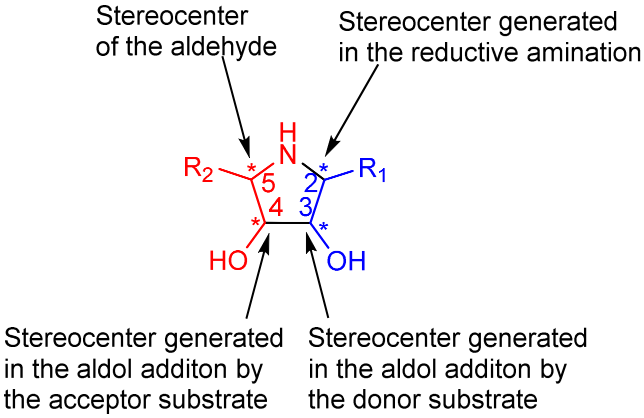
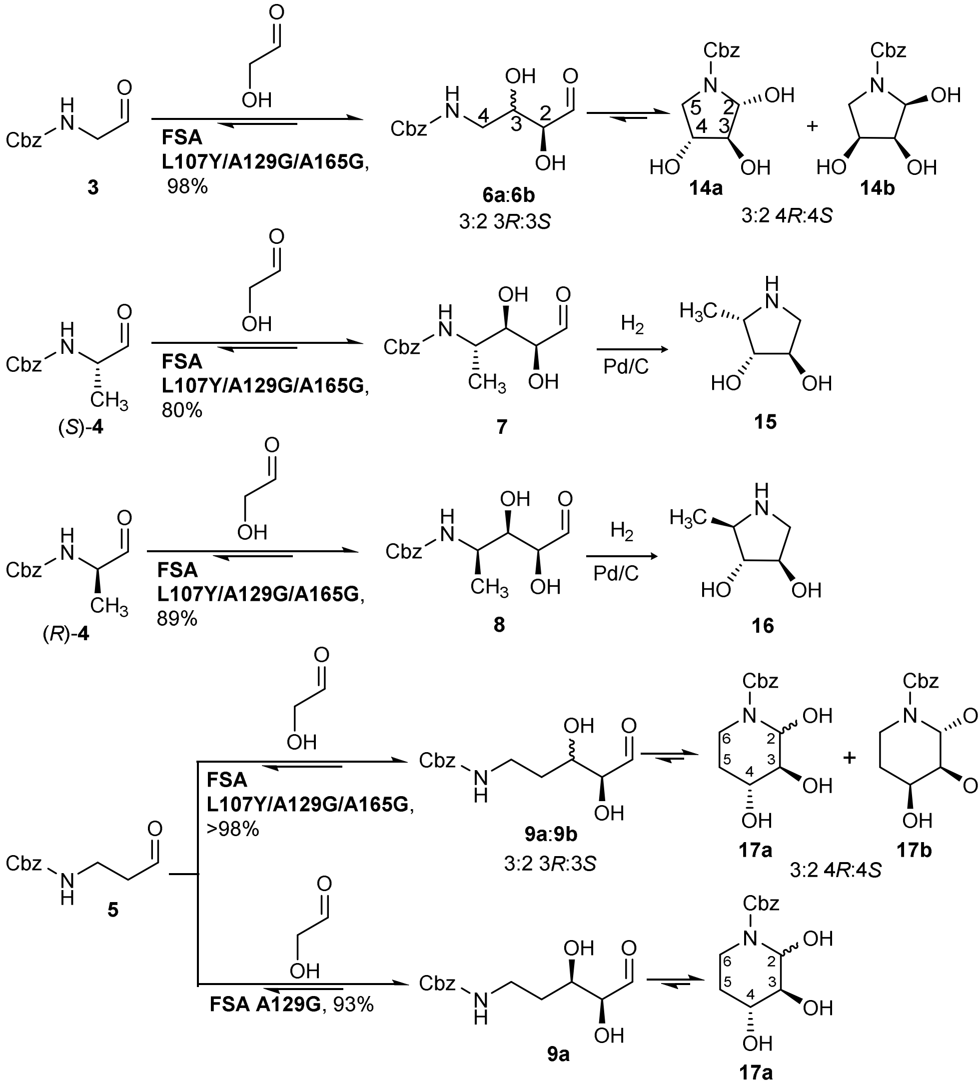


**Figure 3.1.11.** General scheme of the iminiocyclitoles obtained in reductive amination reactions with $H_2$ (50 psi), Pd/C. The configuration of the stereocenters in the pirrolidine derivatives. $R_1$: $CH_2OH$, $CH_3$, H (from the donor substrate DHA, HA and GO respectively); $R_2$: C-α substituent of the *N*-Cbz-aminoaldehydes.

According to the NMR spectra, the cross-aldol additions of glycolaldehyde to (*S*)- and (*R*)-**4** catalyzed by FSA L107Y/A129G/A165G were fully stereoselective. The aldol reaction with **5** was also fully stereoselective with FSA A129G, whereas a 3:2 3R:3S **9a**:**9b** (or 4R:4S in the cyclic hemiaminal **17a**:**17b**) mixture of aldol adduct was obtained with FSA L107Y/A129G/A165G catalyst. On the other hand, acceptor **3** gave a mixture of ~3:2 3R:3S **6a**:**6b** (or 4R:4S **14a**:**14b** in the cyclic heminaminal) aldol adducts using either of the A129G, A129G/A165G, L107Y/A129G/A165G, A129S/A165G, A165G and A129T FSA variants (for spectra see Supporting Info).

The loss of facial differentiation of the electrophilic carbonyl was not observed before, when dihydroxyacetone or hydroxyacetone were used as donors with any of the tried FSA variants.



**Scheme 3.1.5.** FSA variants L107Y/A129G/A165G and A129G catalyzed the cross aldol addition of glycolaldehyde to *N*-Cbz-aminoaldehyde acceptors. Structure of the hemiaminals **14a-b** and **17a-b** and the iminocyclitols **15** and **16** synthesized.

The models of the hemiaminal intermediates from compounds **6a** and **6b** (Figure 3.1.12A-B), indicate that both pairs of epimeric adducts (3*R* and 3*S*, respectively) can fit in the active site of FSA L107Y/A129G/A165G, with the (3*R*)-OH group conformation stabilized by hydrogen bond interactions with residues D6 and N28 and that of the (3*S*)-OH group with N28. Formation of both epimers requires that the acceptor aldehyde approaches the enamine presenting either its *si*- or *re*-face, to yield the (3*R*) or (3*S*)-adduct respectively, as exemplified for aldehyde 3 in Figure 4. The models of Figure 4 show that in a pre-reactive situation the carbonyl oxygen can accept a hydrogen-bond either from the essential water molecule (Figure 3.1.13A) or from the side-chain amide of residue N28 (Figure 3.1.13B), interactions that would increase the electrophilicity of the carbonyl carbon.

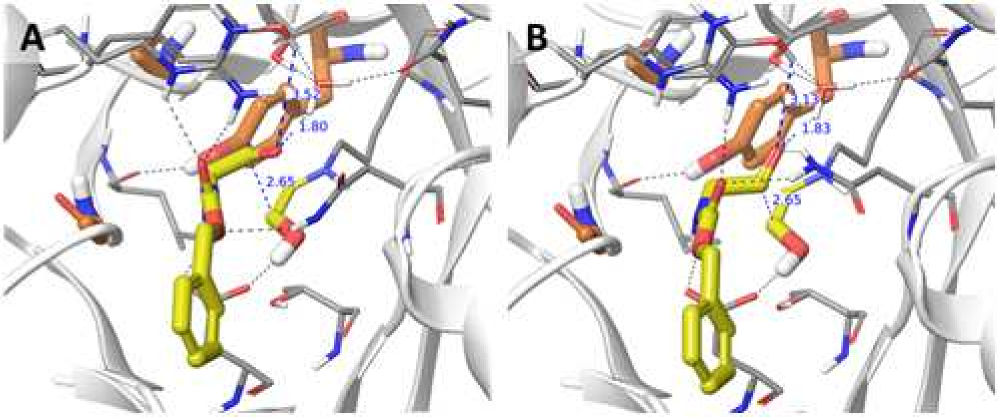


**Figure 3.1.12.** Modeled active site structure of FSA L107Y/A129G/A165G in complex with the hemiaminal intermediates of compounds (**A**) 6a, (**B**) 6b, The hemiaminals (yellow) and the mutated residues (orange) are shown as thick sticks, while the essential water molecule (hydrogen bonded to K85, T109, Y131 and Q59) and the other residues (grey) are shown as thin sticks. The figure shows the point of view from the entrance to the enzyme active site.

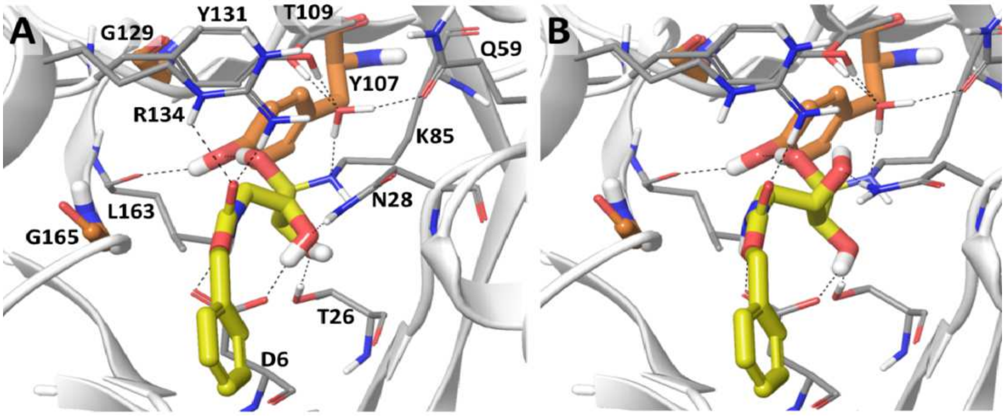


**Figure 3.1.13.** Modeled active site structure of FSA L107A/A129G/A165G in complex with the glycolaldehyde enamine and aldehyde **3** in the pre-reactive orientations presenting its (**A**) *si*- or (**B**) *re*-face to provide adducts **6a** or **6b**, respectively. Distances (Å) between the reactive carbon atoms, and the carbonyl oxygen atom of **3** and the protons from the essential water molecule and the phenol from residue Y131 are shown. Both structures were optimized using QM/MM methods (see Experimental Section).

## 3.1.5. Conclusions

A library of FSA mutants with extremely high activity and selectivity towards glycolaldehyde was designed with a combined program of structure-guided site saturation and site directed mutagenesis.

It was shown that the Ala 129 is the key position in modifying the donor selectivity of the D-fructose-6-phosphate aldolase. The substitution of alanine in position 129 by a serine showed an increment in the activity for DHA and its substitution by glycine, threonine or valine increased FSA's activity towards glycolaldehyde. When combining it with other mutations (for example L107Y) this selectivity can be further enhanced and the acceptor tolerance can also be improved by an additional mutation in the acceptor binding site (A165G).

When testing the generated mutants in donor competition experiments FSA L107Y/A129G, FSA L107Y/A129G/A165G, FSA L107H/A129G and FSA L107H/A129G/A165G gave exclusively D-threose (from the homo-aldol addition of GO), whilst not accepting HA or DHA as nucleophile. *In situ* NMR experiments with an excess of HA and DHA also showed that the mutant FSA L107Y/A129G/A165G did not furnish D-xylulose or D-deoxyxylulose in detectable amount, the only product was D-threose. Thus these variants show unusual selectivity for glycolaldehyde as donor.

The enzymes were tried in cross-aldol reactions with glycolaldehyde and *N*-Cbz-aminoaldehydes, and furnished the products with high yields. However, unexpectedly the stereochemistry of the $C_3$ position of the aldol addition products depended on the substrates and the FSA mutant. This was unprecedented with FSA in the case of other donors. These novel FSA variants combined with the previously described enzymes constitute a set of highly complementary enzymes with selectivity towards HA, DHA or GO.

It was also demonstrated that minimal changes in the active site of FSA can provoke great alternations in the selectivity towards donors and electrophiles. These mutants are being further investigated in various lines of research in our group, including in cross-aldol reactions with other amino aldehydes as well as in enzymatic cascade reactions coupled with benzaldehyde lyase.

## 3.2. Synthesis of rare aldoses with consecutive biocatalytic aldol reactions

Carbohydrates are the most abundant biomolecules that are present in all living organisms. Their role as energy supply for various vital processes is well established, but they also encompass important function in signaling and recognition processes [154, 181]. Carbohydrates are often conjugated with other biomolecules to form oligosaccharides, glycoproteins and glycolipids, or they are incorporated in metabolites [182, 183]. Their importance in biological processes fuels the increasing need for methods for an efficient synthesis of analogues, e.g. rare sugars, in the search of different probes or potential drugs.

Rare sugars are monosaccharides including their derivatives that exist in nature in extremely small amounts [184]. Among all possible aldoses – hexoses and pentoses – only six of them are present in significant amounts in nature, i.e. D-glucose, D-galactose, D-mannose, D-fructose, D-xylose, D-ribose and L-arabinose. Despite their low natural abundance, rare sugars have various known biological functions as well as a high potential for application in pharmaceutical, food or cosmetics industries [163, 165, 185]. For this reason, methodologies that allow the expedient access to these molecules are of paramount importance.

The first synthesis of complex carbohydrates was possibly completed by the formose reaction, one of the theories for the origin of life. It is a hypothesis explaining the abiotical formation of pentose and hexose carbohydrates from small achiral organic molecules such as formaldehyde and glycolaldehyde in the prebiotic world [186-188]. Studies show that by this reaction, a wide variety of carbohydrates can be constructed in the presence of bases or minerals [189, 190].

Nowadays an arsenal of more sophisticated stereoselective methods exist for the synthesis of carbohydrates and analogues including rare sugars, however, they are generally tedious procedures, proceeding by a successive and cumbersome protection and deprotection steps, which are costly and time consuming [156]. During the last 25 years, enzyme catalysis became an alternative approach for the production of rare sugars. Izumori and co-workers have extensively studied converting enzymes such as D-tagatose 3-epimerase from *Pseudomonas cichorii* for the mass production of aldoses e.g. D-psichose from D-fructose [191]. However, more efficient catalytic methods may be used as alternatives. One of them could be the consecutive connection of small building blocks by carbon-carbon bond forming reactions with the concomitant hydroxyl functionalization and installation of chiral positions by aldol additions [161]. This approach was exploited by Northup and MacMillan. They described the synthesis of protected hexoses by two sequential aldol additions of α-oxyaldehydes on a synthetic glycolaldehyde equivalent by combining a direct organocatalytic aldolization with a metal-catalyzed Mukaiyama aldol reaction (Scheme 3.2.1. A) [166, 167]. With this procedure, protected sugars and derivatives, i.e. including carbon, nitrogen and sulfur substitutions at C2 were obtained.

The biocatalytic method uses DHAP-dependent aldolases to first generate ketose carbohydrates, which are then converted to aldoses by dephoshorylation and isomerization [88]. The shortcoming of this method is the need for a number of assisting enzymes acting in concert to convert specific substrates and further transform the aldol adduct into an aldose sugar (Scheme 3.2.1. B).

MacMillan's procedure provides conveniently protected sugar building blocks, whereas the straightforward synthesis of unprotected carbohydrates from simple unmodified materials such as formaldehyde and glycolaldehyde remains highly valuable. Such procedure would mimic the conditions in the early stages of the prebiotic world. In this regard, suitably engineered highly evolved aldolases could present a promising potential towards a fully direct stereoselective catalytic carbohydrate preparation. The ideal aldolase catalyzed pentose or hexose carbohydrate synthesis should involve the successive connection of e.g. formaldehyde, or glycolaldehyde in direct C1+C2+C2, C1+C2+C3 or C2+C2+C2 type aldol additions (Figure 3.2.1. C).

An excellent candidate for our pursuit is the ᴅ-fructose-6-phosphate aldolase (FSA) with unprecedented nucleophile tolerance, accepting for example glycolaldehyde (**1**) as nucleophilic substrate. Homo aldol addition of **1** using FSA wild-type affords exclusively ᴅ-threose, as a result of one single glycolaldehyde addition, however the control on the number of GO additions prevents the one pot preparation of more complex polyoxygenated structures. In the preceding section of this thesis the substrate selectivity, towards glycolaldehyde and different electrophilic substrates has been tailored by a structure-guided approach combining saturation and site-directed mutagenesis of the active site (Chapter 3.1). This has provided a toolbox of FSA mutants with novel synthetic applicability. In the present chapter, it was attempted to combine the knowledge gained on the FSA activity towards GO with further structure guided engineering of the enzyme to achieve carbohydrate synthesis by the trimerization of GO, as well as its double addition onto other acceptor aldehydes. Moreover, modification of its stereoselectivity for the one pot synthesis of different aldose sugars was also investigated.

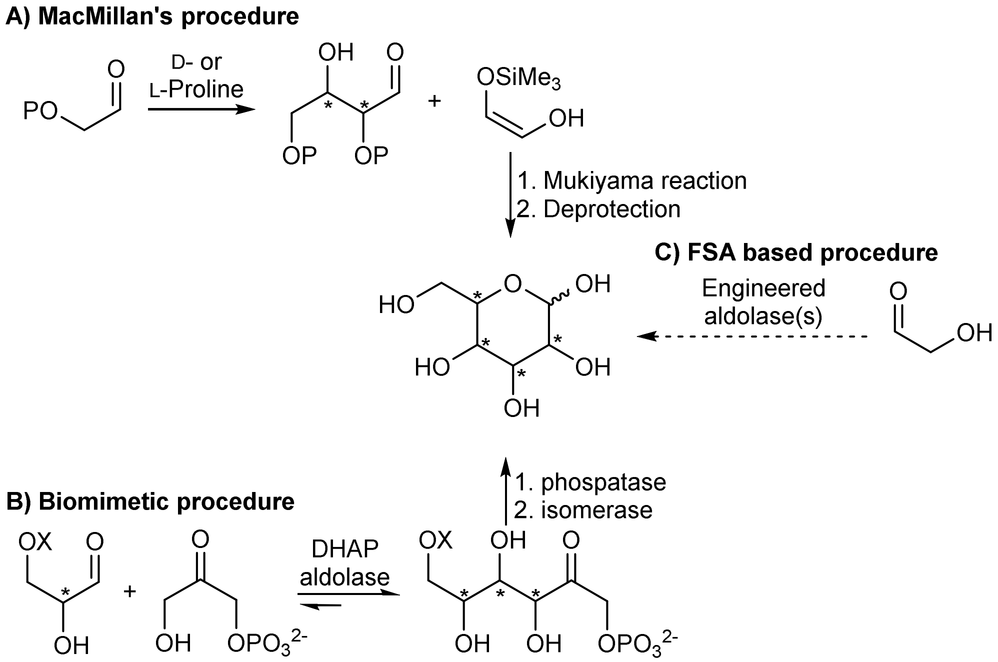
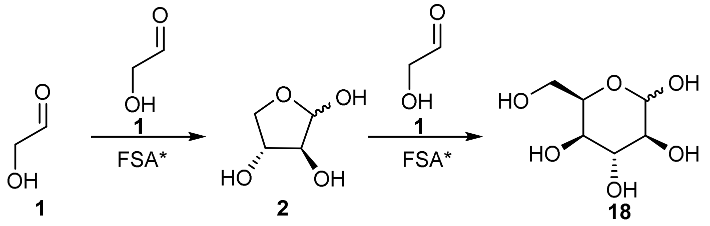


**Figure 3.2.1.** Synthesis of aldohexoses from small achiral starting materials through a chemical or biocatalytic routes. **A)** Combination of organocatalysis and metal assisted Mukaiyama aldol addition; **B)** Biomimetic procedure using a combination of enzymes to first furnish a ketose carbohydrate which is then converted to an authentic aldose by phosphatases and isomerases; **C)** Novel prebiotic approach based on D-fructose-6-phosphate aldolase from *E. coli* (FSA)

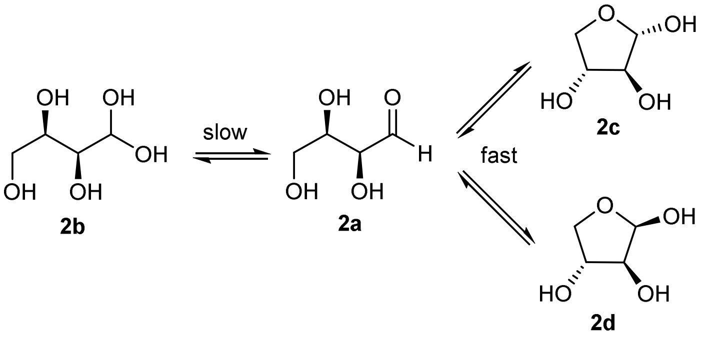
## 3.2.1. Formation of aldoses with consecutive aldol reactions

The results obtained in the reactions of glycolaldehyde with the library of mutants of FSA (Chapter 3.1) encouraged us to investigate a second consecutive addition of GO to prepare aldopentoses and aldohexoses. While evaluating the recently uncovered set of FSA mutants we observed that some of them catalyze the trimerization of GO (i.e. C2+C2+C2 connection), furnishing the non-natural carbohydrate D-idose (**18**) (Scheme 3.2.1.).



**Scheme 3.2.1.** Trimerization of glycolaldehyde to form D-idose (**18**) catalyzed by the FSA mutants with enhanced activity and selectivity towards GO (**1**). The first addition furnishing D-threose is catalyzed by the wild-type, however for the second aldol reaction engineered FSA is necessary (FSA*).

The FSA wild-type catalyzed trimerization of GO is precluded by two factors. First, the limited donor quality of GO, since some non-cyclic aldehydes (e.g. D-glyceraldehyde) did not react with GO. The second factor is that the furanose ring of D-threose limits its availability as acceptor since the amount of the free aldehyde is low in the aqueous solution. Hexoses, pentoses and tetroses are generally poor acceptors for aldolases, since the open chain form is rarely observed in aqueous solutions. They preferentially form cyclic hemiacetals, or undergo hydration on the carbonyl function, hampering their reactivity [192, 193]. NMR analysis of D-(–)-threose (**2**) in $D_2O$ revealed that two cyclic products, **2c** and **2d**, are in fast equilibrium with the open form **2a** and a slow equilibrium with the corresponding hydrate **2b** (Scheme 3.2.2.) [194]. As observed from the $^1$H NMR spectrum only a small amount of aldehyde exists in the aqueous solution, causing threose to be considered as a poor acceptor substrate.



**Scheme 3.2.2.**Equilibrium of D-(–)-threose by NMR analysis [194].

Moreover, it appears that for the catalysis of these reactions high $k_{cat}/K_M$ values are required for the donor. For instance, HA possesses such donor qualities $k_{cat}/K_M$ =

145 $s^{-1}$ m $M^{-1}$ [121], and it is able to furnish 1-deoxy-D-ido-hept-2-ulose from the aldol addition of HA to D-threose. On the other hand, GO has poorer donor qualities ($k_{cat}/K_M$ = 84 $s^{-1}$ m $M^{-1}$ [121]) and its addition to D-threose was not achieved with FSA wild-type catalysis. Both of these factors are important, therefore the formation of pentose or hexose sugars with FSA requires the increased tolerance towards polyhydroxylated acceptor aldehydes combined with the improvement of the donor reactivity of GO the as it was performed in the previous chapter. Hence, the mutants produced there could possibly be used as catalyst for the double addition of GO in the synthesis of rare aldoses.

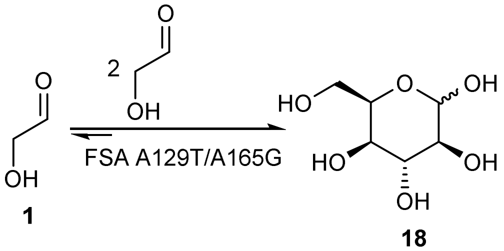
### *a. One-pot formation of D-Idose*

To overcome these hardships in the way of the trimerization of GO, two strategies were applied to engineer FSA. The substitutions that improved the reactivity of the GO as donors, described in Chapter 3.1, were combined with structure-guided engineering of the aldolase, mainly in the acceptor site.

When testing the single mutants with the substitutions in position 129, i.e. A129G, A129T, A129V; Chapter 3.1., we observed the formation of a hexose, nevertheless only in a small amount. Furthermore, in the screening of the double and triple mutants, i.e. FSA A129G/A165G, FSA A129T/A165G, FSA L107Y/A129G and FSA L107Y/A129G/A165G, the formation of a hexose was observed, particularly with FSA A129T/A165G,which produced a considerable amount (>40%) of hexose in 24 hours. This observation proves our hypothesis that the combination of mutations in the donor and acceptor binding site are required for the trimerization of glycolaldehyde. These FSA variants were tested together to find the best combination for the D-idose formation.

The assay reactions were performed on a 500 µL scale, with 150 mM GO solutions using 1 mg protein in each case (see Experimental part). Samples were withdrawn after 72 hours and the concentration of each substrate was measured by HPLC by an external standard method after derivatization (see Experimental section). The wild-type FSA was also included in the screening as a reference (Figure 3.2.2.). The single mutants, FSA A129G, FSA A129T and FSA A129V as well as the wild-type FSA did not catalyze the double addition of glycolaldehyde, or only with very low conversion. As expected from earlier results FSA A129G/A165G, FSA L107Y/A129G did not catalyze the trimerization either, and surprisingly neither did FSA

L107Y/A129G/A165G although it bares mutations in the donor as well as the acceptor site of the enzyme. FSA A129T/A165G was the best variant, which furnished over 70% D-idose (**18**) after 3 days. This is an extremely remarkable and unprecedented result because the formation of D-idose (**18**) by the trimerization of GO (Scheme 3.2.3.) implies the concomitant functionalization and creation of four stereogenic centers accomplished in a one-pot two-step reaction.



**Scheme 3.2.3.** Synthesis of D-idose (**18**) by the trimerization of GO (**1**) catalyzed by the genetically engineered FSA variants.

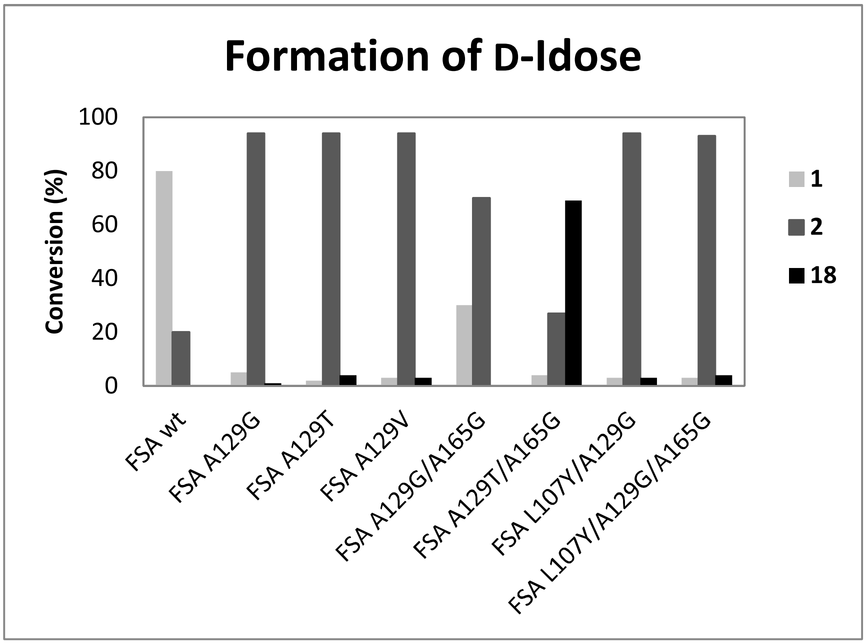


**Figure 3.2.2.** Testing the FSA variants in the trimerization of GO, the bars represent the conversions after 72 hours observed by HPLC analysis of the screening. The reactions were conducted under identical conditions with 2 mg mL$^{-1}$ enzyme amount.

To prove unequivocally the formation of a hexose and to analyze the stereochemistry of the aldol reactions the product was isolated and analyzed by NMR. The preparative scale reaction was straightforward (i.e. 150 mM glycolaldehyde was incubated with 3 mg mL$^{-1}$ FSA A129T/A165G at room temperature for 96 hours) such long incubation time was necessary to overcome the limitation of the poor acceptor quality of D-threose. Under scale up conditions the enzyme catalyzed the stereoselective trimerization of glycolaldehyde (i.e. C2+C2+C2 connection), furnishing the non-natural monosaccharide D-idose (**6**) with 85% conversion (Figure 3.2.3.). The purification of carbohydrates are always cumbersome, especially to separate idose from threose, therefore under non-optimized conditions a remarkable 48% isolated yield was achieved after flash column chromatography.

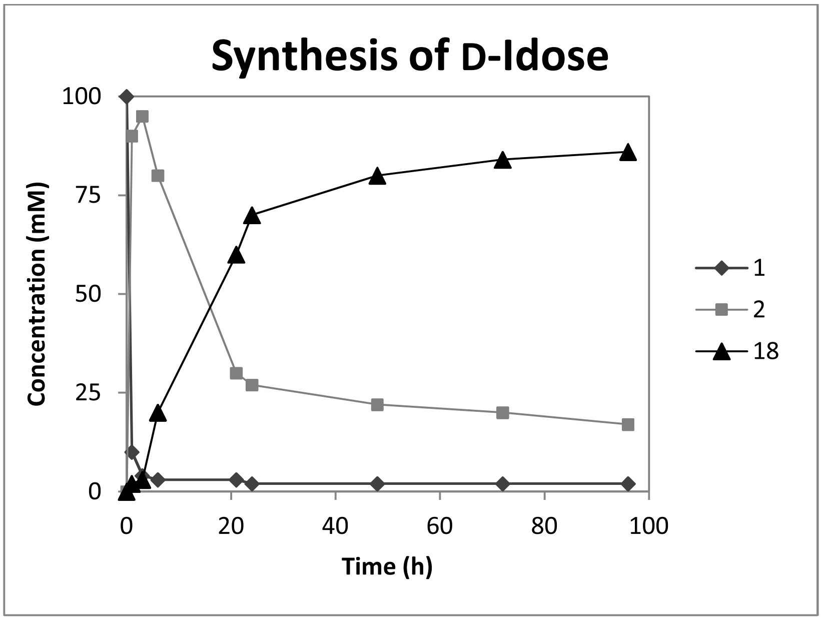


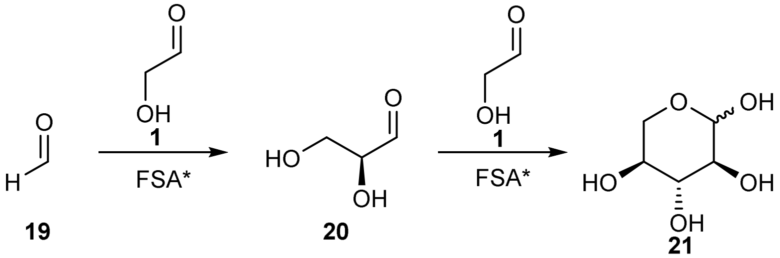
**Figure 3.2.3.** Synthesis of D-idose by the scale up of the double aldol addition of glycolaldehyde by FSA A129T/A165G catalysis. The concentrations of the substrates are shown in the function of time. The concentration of D-threose and D-idose are normalized.

Our method represents the first synthesis of D-idose in one pot with full control over the four stereogenic centers using one single enzyme, and the first enzymatic trimerization of GO directly furnishing a carbohydrate.

### *b. Formation of L-Xylose*

Inspired by the unprecedented ability of FSA A129T/A165G, we started seeking new synthetic challenges. Our next goal was the successive C1+C2+C2 connection by the double addition of glycolaldehyde to formaldehyde to first render L-glyceraldehyde (**20**), with an ensuing *in situ* GO addition to furnish the rare aldose L-xylose (**21**) (Scheme 3.2.4.).

First we tested the FSA variants (FSA wt, FSA A129G, FSA A129T, FSA A129V, FSA A129G/A165G, FSA A129T/A165G, FSA L107Y/A129G, FSA L107Y/A129G/A165G) in the screening reaction, in the formation of L-xylose (Scheme 3.2.4., Figure 3.2.4.). The assays were performed with portion-wise addition of GO in 20 steps over 10 hours, to avoid the formation of threose (**2**) and idose (**18**). Samples were withdrawn after 24 hours and the concentration of each substrate was measured by HPLC after derivatization (see Experimental section). FSA A129T/A165G was the best variant to catalyze the one pot formation of L-xylose (**21**) in full conversion already in 24 hours. Interestingly, the double addition of glycolaldehyde to formaldehyde was also achieved by the single mutants FSA A129G, FSA A129T and FSA A129V and with FSA L107Y/A129G. These mutants already exhibit high activity towards GO, thus for accommodating L-glyceraldehyde no substitutions are necessary in the acceptor binding site. This theory was justified by FSA A129G/A165G and the variant FSA L107Y/A129G/A165G, since both of these mutants exhibited diminished activity compared to the variants without substitution in the acceptor site, FSA A129G and FSA L107Y/A129G, respectively.



**Scheme 3.2.4.** Double addition of glycolaldehyde on formaldehyde to form L-xylose (**21**) catalyzed by the FSA mutants with enhanced activity and selectivity towards GO (**1**). For the double addition reaction engineered FSA is necessary (FSA*).

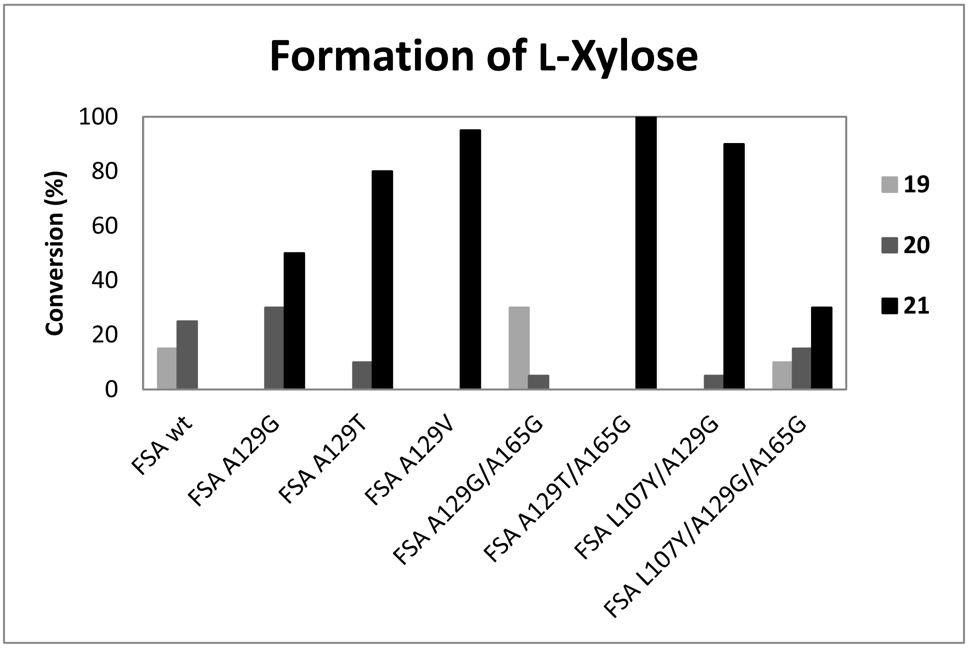


**Figure 3.2.4.** Testing the FSA variants in the synthesis of L-xylose by the double addition of GO on formaldehyde. [1]Only the formation of L-glyceraldehyde (**20**) and L-xylose (**21**) are shown, GO, threose and idose concentrations are not calculated. [2] Due to the volatility of formaldehyde in cases of catalysts with lower activity, the amount of substrates and products do not add up to 100%.

L-Xylose was synthesized by syringe-pump mediated addition of two equivalents of GO onto one equivalent of formaldehyde catalyzed by FSA A129T/A165G. Full conversion and stereoselectivity (2*S*, 3*R*, 4*S*, 5*R*) was observed in 24 hours (see Experimental section). The product was purified by flash column chromatography on silica to afford 62% isolated yield of L-xylose.

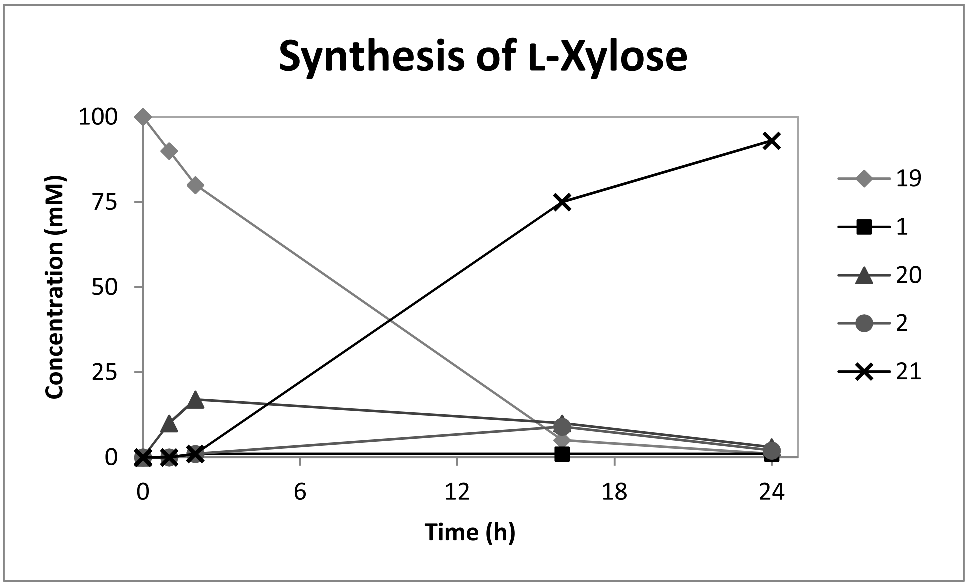
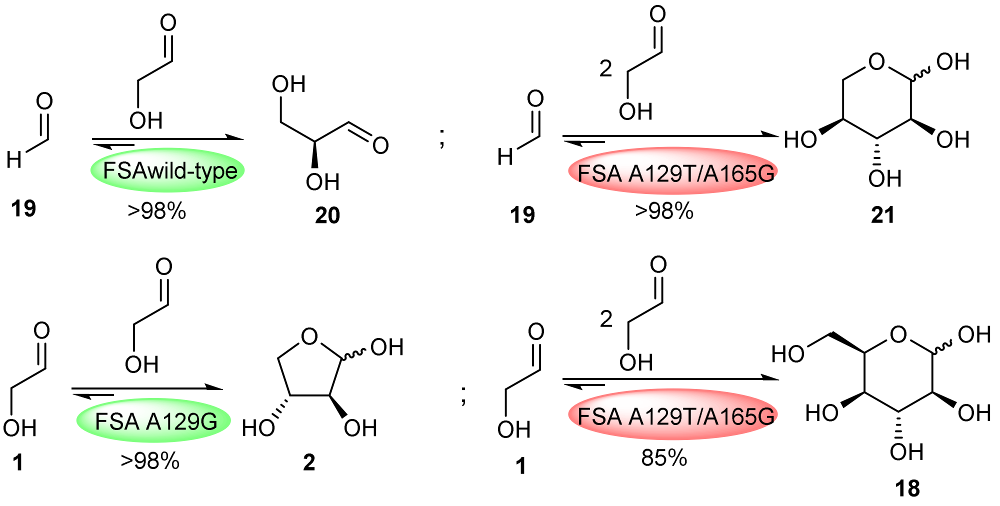


**Figure 3.2.5.** Synthesis of L-xylose (**21**) by syringe pump mediated double addition of glycolaldehyde to formaldehyde by 2 mg mL$^{-1}$ FSA A129T/A165G catalysis. The concentrations of the substrates were calculated from the HPLC peak areas by an external standard method and are shown in the function of time.

The synthesis of this product proves our initial hypothesis that by engineering the aldolase both pentose and hexose carbohydrates can be synthesized with full stereoselectivity.

### *c. Synthesis of a series of three- to six-carbon aldose sugars*

For the synthesis of the aforementioned rare sugars, the selection of the FSA variant plays a crucial role in the control on the number of glycolaldehyde additions. In this manner, the choice of FSA wild-type, FSA A129G or FSA A129T/A165G mutants, allowed the preparation of a complete, stereochemically consistent series of three- to six- carbon aldose sugars from the prebiotic substrates such as formaldehyde and glycolaldehyde. In the case of L-glyceraldehyde (**20**), and D-threose (**2**) only one GO addition takes place, catalyzed by FSA wt and FSA A129G, respectively, while FSA A129T/A165G catalyzes the double addition to furnish L-xylose (**21**) and D-idose (**18**) (Scheme 3.2.5.). It can be stated that we have developed a set of aldolases with which reactions mimicking the prebiotic carbohydrate formation can be carried out with full stereoselectivity.



**Scheme 3.2.5.** Biocatalytic approach mimicking prebiotic reactions for the synthesis of the stereochemically consistent series of three- to six- carbon aldose sugars namely L-glyceraldehyde (**21**) [127], L-xylose (**21**), D-threose (**2**), and D-idose (**18**) catalyzed by FSA wild-type, FSA A129G or FSA A129T/A165G mutants.

### 3.2.2. Synthesis of D-idose derivatives

Next synthetic challenge was the use of glycolaldehyde derivatives and simple achiral aldehydes as acceptors for the one-pot two-step double addition of glycolaldehyde, i.e. C2+C2+C2 or C3+C2+C2 connection, to generate 6-deoxy-D-idose and 6-*O* substituted D-idose derivatives.

A library of small aldehydes with wide range of hydrophobicity and functionalities were tried for the double aldol additions. Acetaldehyde (**22a**), methoxy- (**22b**), benzyloxy- (**22c**), azido- (**22f**), chloro- (**22g**), phenyl- (**25**), phenoxy- (**22e**), benzylthio- (**22d**), and dimethoxy- (**24**) acetaldehyde as well as glyoxal (**28**), methyl-glyoxal (**29**), glyoxilic acid (**26**) and ethyl-glyoxylate (**27**) and electrophiles with 3 carbon chain length i.e., propionaldehyde (**23a**), 3-phenylpropionaldehyde (**23b**) and 3-hydroxypropionaldehyde (**23c**) were tested as acceptors with glycolaldehyde to achieve the double additions (Figure 3.2.6.).

These acceptor substrates were screened with all the mutants that previously showed a high activity towards GO.

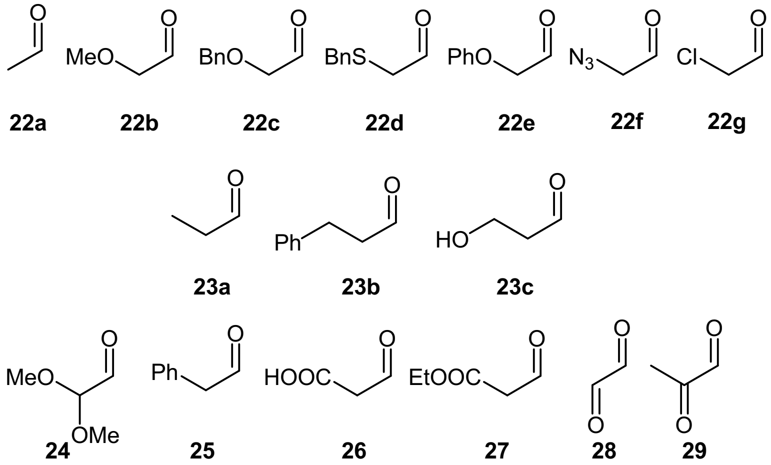


**Figure 3.2.6.** The library of small aldehydes to be tested in the synthesis of D-idose derivatives.

For the accommodation of the bulkier aldehydes we anticipated that further modification in the acceptor binding site would be necessary. Therefore, we decided to test the effect of additional mutations, to expand the scope of accepted electrophiles for the enzyme. To this end, the substitution of the glycine in position 166 for serine (Figure 3.2.7.) was also introduced. This is the position directly neighboring the A165, which residue when substituted by glycine showed a high impact on the acceptor tolerance of the enzyme (Chapter 3.1. and [132]). Hence, the following FSA variants were designed and expressed, FSA A129G/S166G, FSA A129T/S166G and FSA A129T/A165G/S166G.

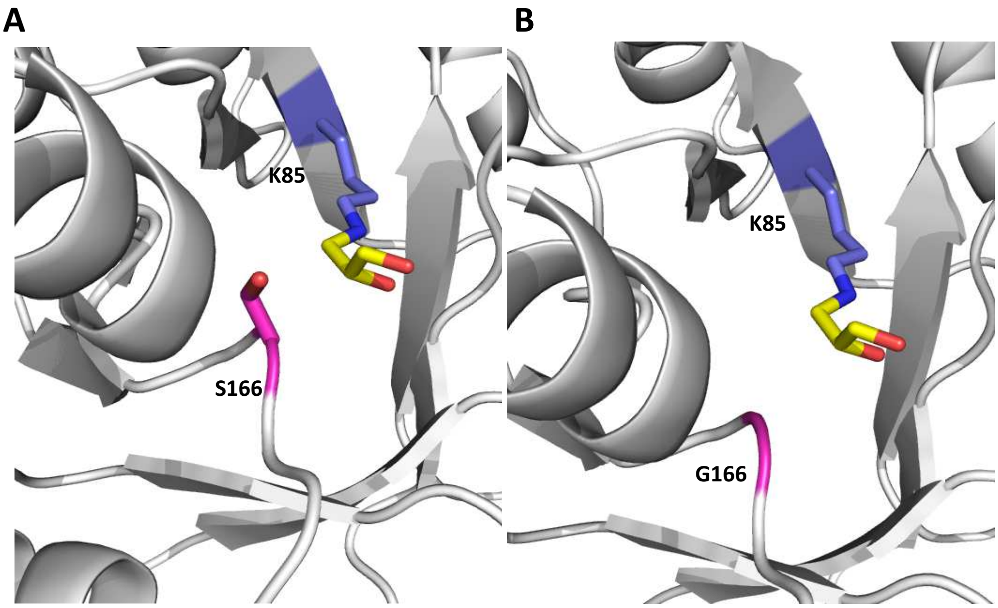


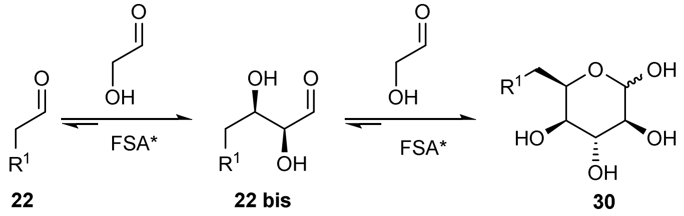
**Figure 3.2.7.** Figure of the active site of FSA. **A)** FSA wt and **B)** FSA S166G the changed residue is shown in magenta.

### *a. One-pot C2+C2+C2 Connection*

The first screening was conducted with the electrophiles containing two carbon atoms to accomplish C2+C2+C2 type additions (Scheme 3.2.6.).

In the screenings glycolaldehyde was added portion-wise (in 20 steps over 10 hours), to minimize the formation of D-threose and D-idose. Samples were withdrawn after 24 hours and the concentration of each compound was measured by HPLC (for details see Experimental section).

From the tested potential acceptors most aldehydes (**22 a**-**g**) furnished moderate to high conversion in the double addition reactions (Table 3.2.1), however dimethoxy acetaldehyde (**24**) gave very low conversion likely due to the C-$\alpha$ substitution. Glyoxylic acid (**26**) and ethyl-glyoxylate (**27**) no aldol adduct formation was detected as well as using glyoxal (**28**) and methyl-glyoxal (**29**). Probably, as observed with DHAP-dependent aldolases, methyl glyoxal is an inhibitor of the enzyme [87]. However, some product appears to be detected in the aldol addition of DHAP to glyoxal using fructose-1,6-bisphosphate aldolases from different sources, although no consistent proof of the product structure was reported [195].



$R^1$ = H (**a**), OMe (**b**), OBn (**c**), SBn (**d**), OPh (**e**), $N_3$ (**f**), Cl (**g**)

**Scheme 3.2.6.** Screening reaction small aldehydes in C2+C2+C2 type double addition of GO.

**Table 3.2.1.** Results of the screening of small aldehydes in C2+C2+C2 type double addition of glycolaldehyde.

| | **H** | | **OMe** | | **OBn** | | **SBn** | | **OPh** | | **$N_3$** | | **Cl** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **22a bis[1]** | **30a** | **22b bis[1]** | **30b** | **22c bis[1]** | **30c** | **22d bis[1]** | **30d** | **22e bis[1]** | **30e** | **22f bis[1]** | **30f** | **22g bis[1]** | **30g** |
| **wild-type** | 2 | - | 6 | - | 5 | - | 62 | - | 36 | - | 58 | - | 100 | - |
| **A129G** | 4 | 8 | 30 | 4 | 50 | 25 | 35 | 51 | 77 | 8 | 50 | 8 | 87 | 5 |
| **A129T** | - | 19 | 27 | 35 | 25 | 55 | 11 | 51 | 52 | 35 | 13 | 31 | 96 | - |
| **A129V** | - | 16 | 26 | 32 | 40 | 40 | 49 | 40 | 62 | 24 | 13 | 31 | 96 | - |
| **A129G/A165G** | 7 | - | 10 | - | 5 | - | 34 | - | 19 | - | 19 | - | 79 | - |
| **A129G/S166G** | 6 | 14 | 25 | 4 | 55 | 15 | 47 | 38 | 78 | 4 | 52 | 9 | 96 | - |
| **A129T/A165G** | - | 17 | 5 | 79 | 10 | 80 | 41 | 18 | 57 | 22 | 7 | 38 | 41 | 7 |
| **A129T/S166G** | - | 16 | 16 | 36 | 20 | 60 | 17 | 77 | 39 | 48 | 5 | 47 | 41 | 48 |
| **A129T/A165G/S166G** | - | 23 | 6 | 79 | 3 | 90 | 45 | 18 | 49 | 35 | 5 | 38 | 38 | 17 |
| **L107Y/A129G** | - | 14 | 42 | 11 | 40 | 40 | 65 | 17 | 67 | 20 | 33 | 10 | 64 | 5 |
| **L107Y/A129G/A165G** | - | - | 23 | 6 | 35 | 25 | 38 | 9 | 66 | 7 | 30 | 6 | 62 | 1 |

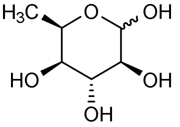
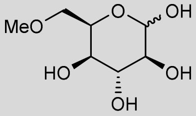
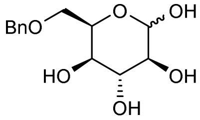
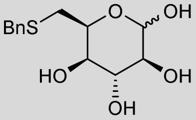
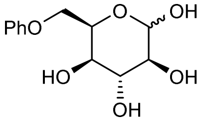
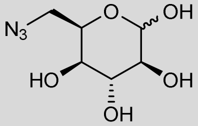
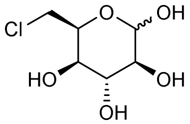
Conversions were measured to the products of the first (**22 bis**) and second (**30**) GO addition by HPLC. The reactions were catalyzed by the FSA mutants with high activity towards GO, and were carried out with 100 mM acceptor concentration and portion-wise addition of GO catalyzed by 1 mg $mL^{-1}$ FSA variant. [1]Not isolated products, concentration was only determined from the chromatogram.

The best FSA variants for the glycolaldehyde double additions were in all cases the enzymes combining the A129T substitution with mutations in the acceptor site either A165 or S166, e.g. FSA A129T/A165G, FSA A129T/S166G or the combination of all three mutations FSA A129T/A165G/S166G. To accommodate aromatic, azido- and chloro-groups FSA A129T/S166G mutant was used, while the smallest aldehydes, methoxy-acetaldehyde and acetaldehyde gave the highest conversions with FSA A129T/A165G variant.

For the smaller aldehydes such as **22a** and **22f** the single mutants FSA A129T and FSA A129V showed similar conversions to the double and triple mutants, similarly to the case of formaldehyde (**19**). As before, only low amount of the double addition product was observed with the variants FSA A129G/A165G, FSA L107Y/A129G and FSA L107Y/A129G/A165G.

The preparative scale reactions were conducted with syringe pump mediated addition, of two equivalent of GO onto one equivalent of the acceptor substrate. The additions lasted between 10 and 25 hours, and then the reactions were at room temperature for additional 12-48 hours. Owing to the low boiling point of acetaldehyde (20.2 °C), both the donor and acetaldehyde (**22a**) were added slowly by a syringe pump to the aqueous buffer, enzyme mixture, to avoid its loss by evaporation. The conversions to **30b**-**g** were between 65 and 95% however, the isolated yields were between 39-64% after flash column chromatography. These results are remarkable, considering the difficult task of carbohydrate purification, nevertheless by optimizing the isolation methodology probably higher yields may be achieved, however, this was out the scope of this study. The isolated yield of **30a** was rather low (7%), due to the need for an extra purification step with activated charcoal column to separate the D-threose from the product **30a** (Table 3.2.2.). NMR spectroscopy showed that the reactions were fully stereoselective and seven D-idose derivatives were synthesized by the C2+C2+C2 type connections.

**Table 3.2.2.** The obtained D-idose derivatives with C2+C2+C2 type additions.

| | Mutant | Conversion | Isolated yield |
|---|---|---|---|
|  **30a** | A129T/A165G | 70% | 7% |
|  **30b** | A129T/A165G | 69% | 39% |
|  **30c** | A129T/A165G/S166G | 94% | 41% |
|  **30d** | A129T/S166G | 66% | 39% |
|  **30e** | A129T/S166G | 79% | 44% |
|  **30f** | A129T/S166G | 82% | 64% |
|  **30g** | A129T/S166G | 82% | 57% |

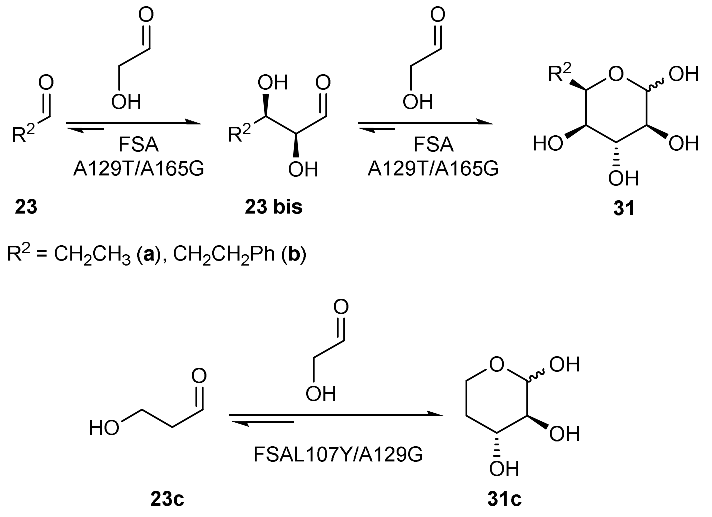
The FSA mutant indicated next to each product is the variant used in the scale up reactions. The stereochemistry of the products was analyzed by NMR. The conversions were calculated from the HPLC chromatograms, at the final reaction time.

### *b. One-pot $C_3$+$C_2$+$C_2$ connection*

Simple aldehydes with 3 carbons were also assayed for the double addition of glycolaldehyde, i.e. C3+C2+C2 connection, to generate high carbon content aldoses (Scheme 3.2.7.).

The screenings were performed as explained before (see Experimental section).

The tests showed that using FSA A129T/S166G the synthesis of 6,7-dideoxy-D-ido-heptose (**31a**) and 6,7-dideoxy-7-phenyl-D-ido-heptose (**31b**) can be accomplished (Table 3.2.3.) with 47-48% conversion. 3-hydroxypropionaldehyde (**23c**) only gave one single glycolaldehyde addition most likely because of the stability of the 4,5,6-trideoxy-D-idose (**31c**) hemiacetal that was formed after the first aldol addition. The most active mutant for this latter reaction was FSAL107Y/A129G developed in a previous work (Chapter 3.1.), (Table 3.2.3.).



**Scheme 3.2.7.** Screening reaction of small aldehydes to glycolaldehyde for the C3+C2+C2 type aldol additions.

The best mutants to catalyze the double glycolaldehyde addition to **23a** and **23b** were again the FSA variants that combine the substitutions A129T and some variations in the acceptor-binding site such as A165G or S166G or their combination. However, the single variant FSA A129T also showed remarkable activity. The mutants with A129G substitutions (FSA A129G, FSA A129G/A165G, FSA L107Y/A129G and FSA L107Y/A129G/A165G) gave only moderated activity for the double additions with these substrates.

**Table 3.2.3.** Results of screening of small aldehydes for the C3+C2+C2 type double addition of glycolaldehyde.

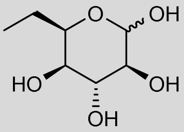
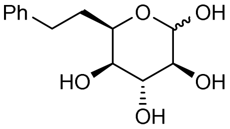
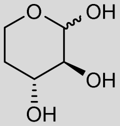
| | **$CH_2CH_3$** | | **$CH_2CH_2Ph$** | | **$CH_2CH_2OH$** |
|---|---|---|---|---|---|
| | **23a bis[1]** | **31a** | **23b bis[1]** | **31b** | **31c** |
| **wild-type** | 5 | - | 30 | - | 13 |
| **A129G** | - | 10 | 47 | 18 | 76 |
| **A129T** | - | 25 | 32 | 39 | 80 |
| **A129V** | - | 30 | 48 | 21 | 82 |
| **A129G/A165G** | - | - | 16 | - | 15 |
| **A129G/S166G** | - | 5 | 47 | 18 | 72 |
| **A129T/A165G** | - | 40 | 22 | 48 | 80 |
| **A129T/S166G** | - | 47 | 22 | 48 | 80 |
| **A129T/A165G/S166G** | - | 44 | 22 | 48 | 80 |
| **L107Y/A129G** | - | 10 | 42 | 22 | 91 |
| **L107Y/A129G/A165G** | - | 15 | 35 | 2 | 78 |

Conversions were measured to the products of the first (**23 bis**) and second (**31**) GO addition by HPLC. The reactions were catalyzed by the FSA mutants with high activity towards GO, and were carried out with 100 mM acceptor concentration and portion-wise addition of GO catalyzed by 1 mg $mL^{-1}$ FSA variant. [1]Not isolated products, concentration was only determined from the chromatogram.

The preparative scale reaction to furnish the heptoses was carried out by syringe pump mediated additions of glycolaldehyde (for 2 and 13 hours). In the case of **23a**, both the donor and acceptor were added slowly to the aqueous buffer solution of the enzyme to avoid its evaporation, since in the screening reactions the amount of propionaldehyde decreased over the time with the tested reaction conditions like before with acetaldehyde. The conversions were between 47 and 75%, and the isolated yields were 17 and 47% to **31a** and **31b** respectively, furnishing two new D-idose derivatives (Table 3.2.4.).

The aldol addition of GO to with 3-hydroxypropanal (**23c**) was catalyzed by the mutants bearing the A129T or A129V mutation, including double and triple mutations. Interestingly the double variant FSA L107Y/A129G furnished almost full conversion in 5 hours (Table 3.2.4.), being the most active mutant. In this case, the substitution L107Y was necessary since the A129G variant was not as efficient and the double variant A129G/A165G gave only meager conversions.

**Table 3.2.4.** Synthesized products and their yields and the mutants used in the reaction of different aldehydes by the (C3+C2+C2) additions of glycolaldehyde.

| | Mutant | Conversion | Isolated yield |
|---|---|---|---|
|  **31a** | A129T/S166G | 47% | 17% |
|  **31b** | A129T/S166G | 75% | 47% |
|  **31c** | L107Y/A129G | 91% | 38% |

The FSA mutant indicated next to each product is the variant used in the scale up reactions. The stereochemistry of the products were analyzed by NMR. The conversions were calculated from the HPLC chromatograms, at the final reaction time.

## 3.2.3. Modification of the stereochemical outcome of FSA catalysis

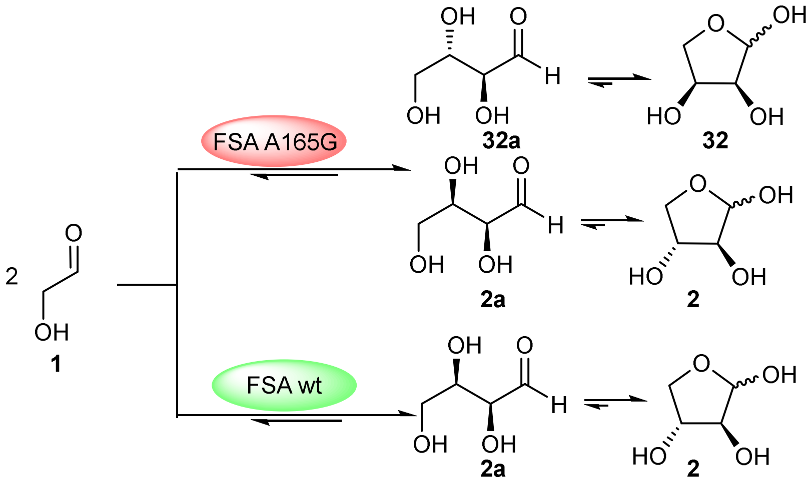
The stereochemistry of the products formed in an enzyme catalyzed aldol addition is, in most instances, controlled by the aldolase, with usually very high level of stereocontrol. This is one of the main advantages of biocatalysis. FSA wild-type proved to have very strict control over the stereochemistry of the formed products, with all described donor substrates [112, 121, 122, 130]. The additions of glycolaldehyde as nucleophilic component with FSA wild-type have a conserved stereochemistry of 2*S*, 3*R*, however, as seen in Chapter 3.1. the stereochemical outcome of the aldol reaction catalyzed by some FSA variants with selected acceptors can produce mixtures, depending on the electrophile.

Thus, our next objective was to take advantage of the compromised stereoselectivity of the FSA variants and to produce mutants that will provide diastereomeric products that are stereochemically complementary to those obtained with the wild-type or variants providing the D-threo, (2*S*, 3*R*) stereochemistry.

Studying the mechanism of FSA, it was supposed that the stereochemistry of the donor is difficult to change, but the stereocenter originating from the acceptor is more easily altered, similarly to other aldolases. Indeed, the stereochemical approach of the aldehyde to the enzyme-donor complex is often compromised when non-natural substrates are used as reactants with some aldolases [97, 196]. Thus we decided to design some FSA variants that produce adducts with changed configuration on the C3 (i.e. the stereocenter formed by the addition of the aldehyde acceptor).

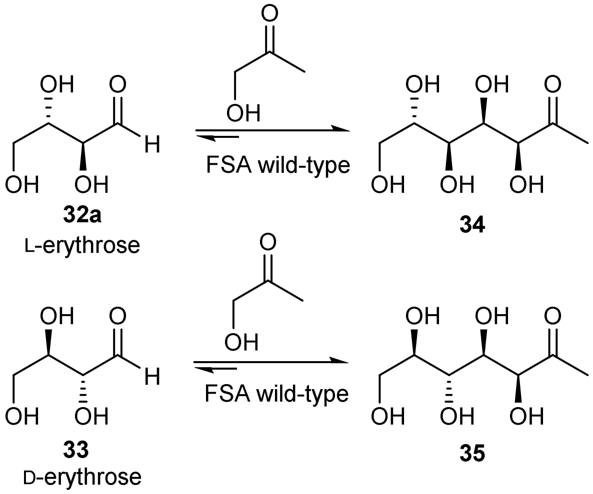
### a. *Engineering FSA for a modified outcome stereochemistry of C3 of the products*

Our next goal was to engineer FSA to modify the sterochemical outcome of the self-aldol reaction of glycolaldehyde. In the course of our investigations, we discovered that a previously described variant, FSA A165G, gave a mixture of D-threose (**2a**) and L-erythrose (**32a**) during the homo-aldol addition of GO (Scheme 3.2.8.). Thus, we envisaged that a second GO addition to L-erythrose (**32a**) could take place by the catalysis of an FSA variant furnishing L-glucose.



**Scheme 3.2.8.** Catalysis of the self-aldol reaction of GO by FSA wild-type furnishing uniquely threose, and FSA A165G providing a mixture of diastereomers (results confirmed by HPLC).

We observed that FSA A165G catalyzed the self-aldol addition of GO with an 8:2 ratio of D-threose (**2**) : L-erythrose (**32**). These results were confirmed by both HPLC, and by a follow up reaction on the erythrose-threose mixture. The follow up biocatalytic reaction consisted of the aldol addition of HA, a good donor for FSA, to the erythrose-threose mixture catalyzed by wild-type FSA (Scheme 3.2.9.). The reaction was followed by HPLC, and the resulting new peaks were compared with those from the same aldol addition reaction but using commercially available D- and L-erytrose independently. Comparison of the retention times between the products of the reaction on the erythrose-threose mixture with those that are products from the commercial source proved preliminarily that the two tetroses of the mixture are likely D-threose (**2**) and L-erythrose (**32**).

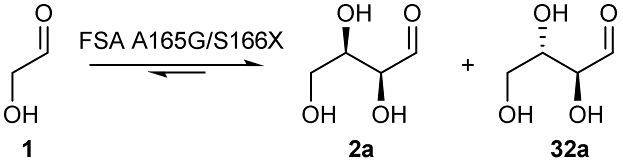


**Scheme 3.2.9.** Aldol addition of D- and L-erythrose with HA, to determine which of the potentially produced tetroses is actually present in the reaction mixture of the self-aldol reaction of GO catalyzed by FSA A165G.

As next step, we sought for an FSA variant with high activity towards the formation of L-erythrose. We envisioned that this enzyme could be constructed by protein engineering, and specifically, by modification of the residue immediately after A165, the serine in position 166. To this end, the library FSA A165G/S166X was created by site saturation mutagenesis.

The mutagenesis was carried out using the QuickChange® site directed mutagenesis protocol on FSA A165G as template, utilizing oligonucleotides with NNS (N: G/C/T/A; S: G/C) degeneracy, which covers 32 codons and all 20 proteinogenic amino acids with 95% probability when randomly picking at least 94 colonies. In the screening 137 clones were grown, and expressed for testing in 96-

well plates (for procedures see Experimental part). The screening reaction was again the self-aldol reaction of GO measuring the formation of L-erythrose (Scheme 3.2.10.). Clones were considered positives if they furnished more L-erythrose than D-threose after 24 hours under identical conditions. Four different variants were found, namely the serine was changed for glycine, asparagine, aspartate and proline: FSA A165G/S166G, FSA A165G/S166N, FSA A165G/S166D and FSA A165G/S166P.



**Scheme 3.2.10.** The screening reaction of the FSA A165G/S166X library.

For more precise analysis and comparison of the variants, they were expressed and purified (see Experimental section). The purified enzymes were assayed with the screening reaction under identical conditions (Table 3.2.5.). The best mutant A165G/S166P improved the L-erythrose formation to 90%, however the activity of this variant was low (0.46 mM $h^{-1}$ $mg^{-1}$protein).

**Table 3.2.5.** Screening the positive mutants of the saturation mutagenesis, comparing the ratio of the formed D-threose (**2**) and L-erythrose (**32**).

| | **Ratio 2a:32a** | **Conversion (total)** |
|---|---|---|
| **A165G** | 4:1 | |
| **A165G/S166G** | 1:1 | |
| **A165G/S166N** | 1:1 | |
| **A165G/S166D** | 1:1 | |
| **A165G/S166P** | 1:9 | |

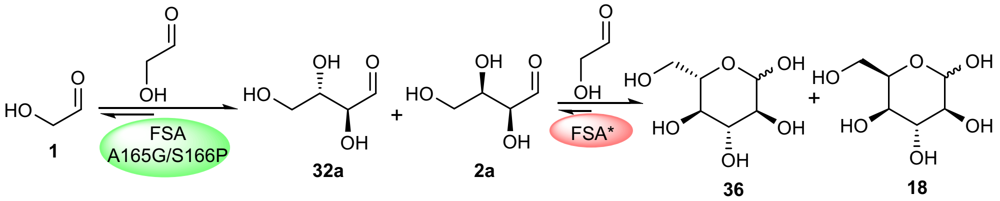
To further improve the L-erythrose ratio in the reaction, we envisioned one more mutation to the FSA A165G/S166P platform. In position 164 the alanine was substituted by a glycine. This residue is pointing away from the active site, however, we hypothesized that the removal of the methyl group would increase the flexibility of the chain that became more rigid due to the substitution of the proline for a serine. Hence, the triple mutant FSA A164G/A165G/S166P was designed by site directed mutagenesis, and tried in the self-aldol reaction of glycolaldehyde. It did not furnish higher amount of L-erythrose than the original FSA A165G/S166P: i.e., after 72 hours FSA A165G/S166P furnished 5.17 mM threose and 54.13 mM erythrose, whereas the FSA A164G/A165G/S166P gave 3.32 mM threose and 32.90 mM erythrose, and it did not improve the activity either: 63% erythrose formation with the double mutant, in contrast to 49% furnished by the FSA A164G/A165G/S166P variant.

Another attempt to improve the activity of FSA A165G/S166P in the formation of erythrose was tried combining the mutations A165G and S166P with the previously mentioned substitutions L107Y/A129G. The latter variant enhanced the activity in the reactions with glycolaldehyde as nucleophile. The library FSA L107Y/A129G/A165G/S166X (X: NNS (N: G/C/T/A; S: G/C) degeneracy) was constructed by site saturation mutagenesis, and screened for the formation of L-erythrose, like before.

According to the screening, the saturation mutagenesis did not provide any positive clones. This result is consistent with a previous observation, that the FSA variants that combine the A165G mutations with a substitution in the A129 position, i.e. FSA A129G/A165G, FSA A129T/A165G, FSA A129S/A165G or FSA L107Y/A129G/A165G , do not show any stereochemical promiscuity in the self aldol reaction of GO (Chapter 3.1).

### *b. One-pot two step biocatalytic synthesis of L-Glucose*

Once the production of L-erythrose was accomplished in good yield, a second aldol addition of GO must be performed with full *syn* stereochemistry (2*S*,3*R*,4*S*,5*S*) to produce L-glucose (Scheme 3.2.11.). FSA A165G/S166P was not able to catalyze another GO addition, and thus FSA variants with good capacity in hexose synthesis were tested as catalyst using the L-erythrose generated *in situ* by FSA A165G/S166P (i.e. as 1:9 mixture of D-threose : L-erythrose) as acceptor (Scheme 3.2.11., Table 3.2.6).



**Scheme 3.2.11.** Screening reaction for the synthesis of L-glucose (**36**) from a mixture of L-erythrose and D-threose.

**Table 3.2.6.** Testing the FSA mutants in the synthesis of L-glucose (**36**).

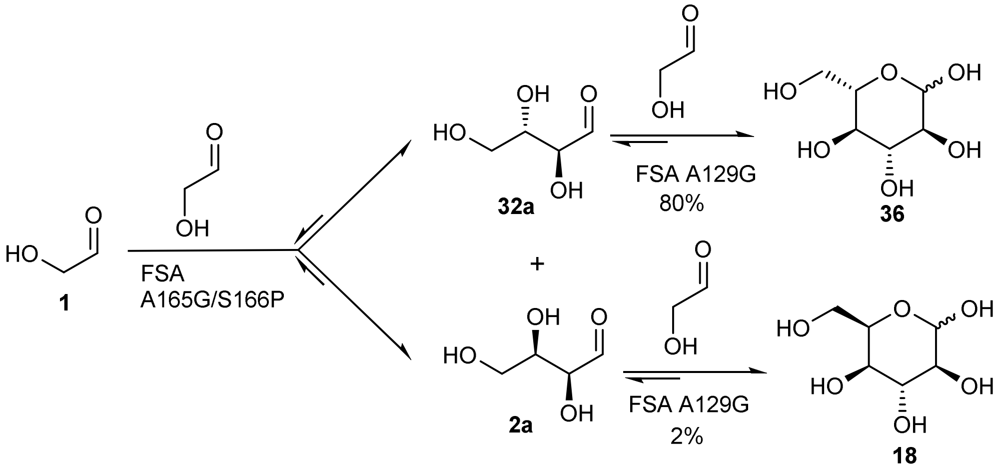
| | **2a** | **32a** | **18** | **36** |
|---|---|---|---|---|
| **wild-type** | 21 | 64 | - | - |
| **A129G** | 15 | 48 | 2 | 26 |
| **A129T** | 20 | 56 | 1 | 16 |
| **A129V** | 20 | 58 | 1 | 13 |
| **A129G/S166G** | 18 | 50 | 2 | 20 |
| **A129T/A165G** | 19 | 64 | 2 | 7 |
| **A129T/S166G** | 18 | 59 | 4 | 11 |
| **A129T/A165G/S166G** | 14 | 61 | 8 | 7 |
| **L107Y/A129G** | 14 | 57 | 5 | 15 |
| **L107Y/A129G/A165G** | 20 | 60 | 2 | 10 |

Conversions are calculated from peak areas using HPLC chromatograms after 24 hour reaction time, using 2 mg mL$^{-1}$ enzyme.

From the screening FSA A129G appears to tolerate **32a** more efficiently than any other variant. Since as observed in the earlier screening (Figure 3.2.3.) this variant barely accepts **2** as electrophile, thus with this enzyme the formation of D-idose was minimized (2%). Surprisingly the mutants furnishing D-idose in high conversion, i.e., FSA A129T/A165G, FSA A129T/S166G and FSA A129T/A165G/S166G did not produce a high amount from either of the hexoses, possibly due to the low amount of free GO or due to some inhibition effect. However, the variant FSA A129G/S166G furnished similarly high amount of **36** to A129G, which can be an effect originating from the substitution of glycine for alanine in position 129.

To ascertain unequivocally the stereochemistry of the formed product, preparative scale synthesis was carried out in one-pot two-step reaction of GO using a tandem of biocatalysts (Scheme 3.2.12.).

The first step of the reaction catalyzed by FSA A165G/S166P furnished 76% L-erythrose and 7% D-threose in 7 days, while some GO remained in the mixture. To this solution, FSA A129G was added which catalyzed the addition of GO to erythrose to furnish L-glucose. On Figure 3.2.8. the slow diminution of D-threose can be observed after 169 hours, as the result of the retro-aldol reaction of **2**, and the GO formed by this reaction is then probably used in the formation of L-glucose what is accumulating over long reaction times.



**Scheme 3.2.12.** Synthesis of L-glucose by double addition of GO catalyzed by the combination FSA A165G/S166P and FSA A129G.

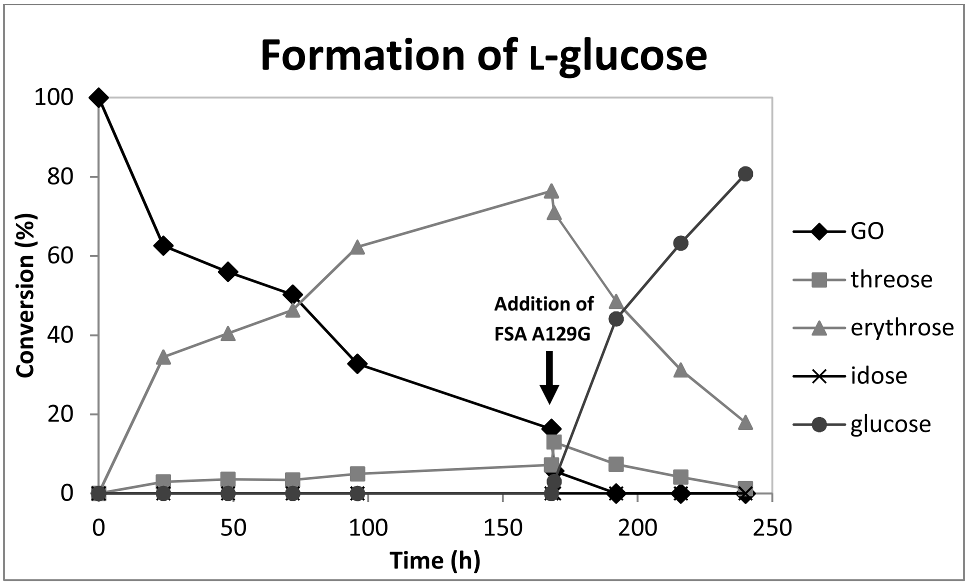


**Figure 3.2.8.** Changes in the substrate concentration during the course of the reactions. The first part of the reaction (up to 168 hours) is catalyzed by FSA A165G/S166P and at this point FSA A129G is added to the mixture (marked with an arrow).

In this reaction the variant FSA A129G selectivity assists the L-glucose formation, since no D-idose generation is observed during the course of the reaction.

To simplify the procedure we have tried to use the two enzymes together in one-pot, hoping, that L-glucose will accumulate, since it is the thermodynamically favored product. However, its formation was not observed, and with this combination of FSA A165G/S166P and FSA A129G the only in a small amount of D-idose (<10%) was formed. This is due to the very low reactivity of the variant FSA A165G/S166P in the formation of erythrose, compared to the extremely high activity of the variant FSA A129G in furnishing threose.

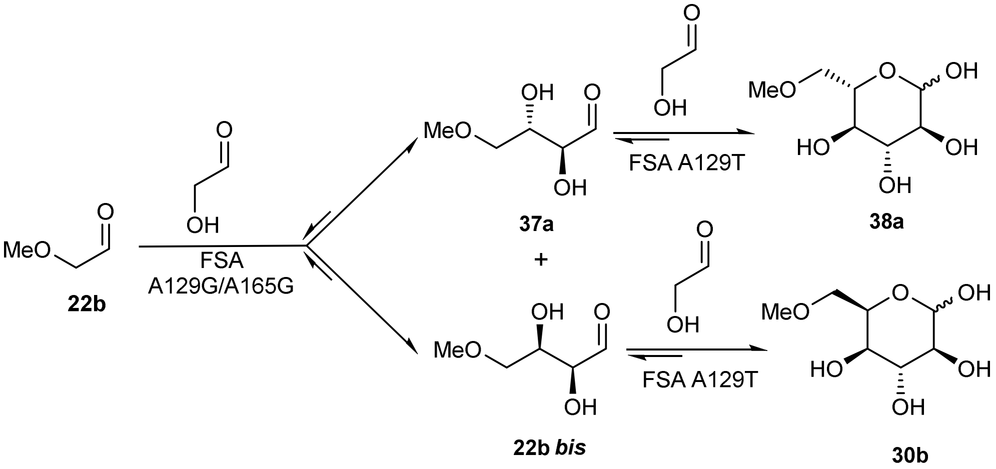
### *c. Synthesis of L-Glucose derivatives*

To furnish other L-glucose derivatives with L-erythro configuration in the first GO addition the earlier mentioned acceptor substrates (**22**-**29**)(Figure 3.2.6.) were tried with the variants of FSA obtained thus far, and interestingly some mutants showed promiscuity in certain cases (Table 3.2.7. and Table 3.2.8.)

With none of the tested aldehydes (**22**-**29**) the preferential formation of L-erythrose derivative was observed with the mutants of the FSA A165G/S166X library.

We found that in some cases the earlier described variants, furnished a mixture of D-threo and L-erythro products after the first GO addition. These mutants, bare the substitution of glycine for alanine at position 165. E.g. FSA A129T/A165G and FSA A129G/A165G were found to catalyze the synthesis of two diastereomers in the additions of different aldehydes and GO.

Methoxyacetaldehyde (**22b**) and glycolaldehyde with FSA A129G/A165G produced mixture of the two diastereomers (**22b bis** and **37a**) in a ~1:2 ratio (Scheme 3.2.13., Table 3.2.7.). Upon the addition of a second FSA variant to the reaction, a mixture of **38a** and **30b** were formed in all cases. Among the tested variants, FSA A129T furnished the best ratio between **38a** and **30b**, giving 6-methyl-L-glucose (**38a**) with 38% conversion, and only 23% idose derivative.



**Scheme 3.2.13.** Synthesis of 6-methoxy-L-glucose.

**Table 3.2.7.** Screening the mutants in the synthesis of 6-methoxy-L-glucose

| | 22b | 37a | 30b | 38a |
|---|---|---|---|---|
| **1st step** | | | | |
| **A129/** | 30 | 3 | | |
| **A129G/A165G** | **10** | **22** | | |
| **A129T/A165G** | 25 | 2 | | |
| **A165G** | 25 | 17 | | |
| **A165G/S166D** | 23 | 18 | | |
| **A165G/S166G** | 21 | 19 | | |
| **A165G/S166N** | 22 | 18 | | |
| **A165G/S166P** | 2 | 4 | | |
| **L107Y/A129G/A165G** | 23 | 13 | | |
| **2ndstep[b)]** | | | | |
| **Wild-type** | 49 | 41 | 3 | 7 |
| **A129G** | 43 | 22 | 9 | 26 |
| **A129T** | **30** | **10** | **23** | **38** |
| **A129V** | 35 | 18 | 17 | 30 |
| **A129G/A165G** | 42 | 20 | 10 | 28 |
| **A129T/A165G** | 25 | 18 | 27 | 30 |
| **A129T/S166G** | 24 | 17 | 28 | 31 |
| **A129T/A165G/S166G** | 18 | 23 | 34 | 25 |
| **L107Y/A129G** | 44 | 21 | 8 | 27 |
| **L107Y/A129G/A165G** | 52 | 42 | 0 | 6 |

a) Percentage of product formed at 24 h b) Percentage calculated from the decrease of the educts.

During the screening of **22g** we have found that some FSA variants furnish 6-chloro-L-glucose (**38b**). Among these variants FSA A129T/A165G preferentially catalyzes the synthesis of **38b** in one pot (Scheme 3.2.14.) with 12% conversion after 24 hours. In this case the same enzyme catalyses both the first and the second GO addition.

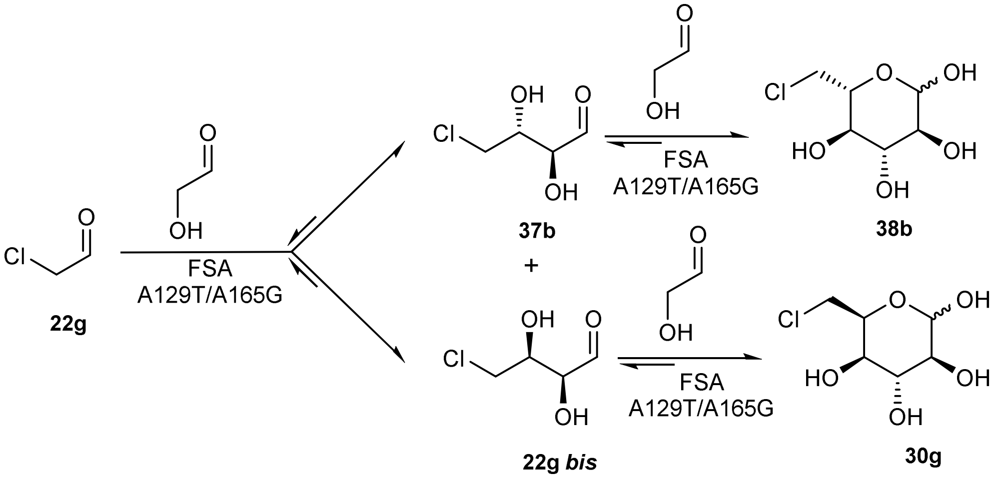


**Table 3.2.14.** Synthesis of 6-chloro-L-glucose.

**Table 3.2.8.** Screening the FSA mutants in the synthesis of 6-chloro-L-glucose

| | 22g bis | 37b | 30g | 38b |
|---|---|---|---|---|
| **Wild-type** | 100 | - | - | - |
| **A129G** | 87 | 4 | 5 | 3 |
| **A129T** | 96 | 4 | - | - |
| **A129V** | 85 | 6 | 6 | 3 |
| **A129G/A165G** | 79 | 21 | - | - |
| **A129G/S166G** | 96 | 4 | - | - |
| **A129T/A165G** | 41 | 40 | 7 | 12 |
| **A129T/S166G** | 41 | 4 | 48 | 7 |
| **A129T/A165G/S166G** | 38 | 39 | 17 | 6 |
| **A165G/S166G** | 56 | 44 | - | - |
| **A165G/S166P** | 46 | 53 | - | - |
| **L107Y/A129G** | 64 | 25 | 5 | 6 |
| **L107Y/A129G/A165G** | 62 | 36 | 1 | 2 |

Both reactions were scaled up to prove the stereochemistry of the obtained product carried out on a preparative scale. To produce methoxy-L-glucose the mixture of two FSA variants were used, while for chloro-L-glucose only one variant was applied and the products furnished with 68-60% conversion and isolated with 51-14% yield after flash column chromatography.

### 3.2.4. Conclusions

We have developed a procedure for the preparation of a number of aldose carbohydrate derivatives from simple achiral building blocks, such as formaldehyde, glycolaldehyde and other small aldehydes or glycolaldehyde derivatives by biocatalytic tandem reactions using engineered FSA. This method mimics the ancient prebiotic processes such as the formose reaction, but using highly stereoselective evolved biocatalyst with complete control over the stereochemical outcome of the reactions.

The aldopentose, aldohexoses and aldoheptoses synthesized in this work were prepared by consecutive aldol additions of two glycolaldehyde molecules to the chosen acceptor substrate, and the sugars were biocatalytically obtained from achiral substrates, avoiding intermediate isomerization or phosphorylation steps. Such aldolase activities are unprecedented in nature and represent a biocatalytic alternative for the construction of artificial metabolic pathways.

Interestingly, variants of FSA with minimal modifications in the active site were found sufficient enough to efficiently catalyze each reaction, which points out that a relatively small number of FSA mutants suitably cover for a survey of acceptor substrates in the synthesis of aldohexose derivatives.

As a proof of concept, we demonstrated that the stereochemical outcome of FSA catalyzed reactions may be also engineered by modification of the aldolase by a structure guided approach. Site saturation mutagenesis was performed on FSA and the mutant furnished the two stereoisomers of the self-aldol reaction of glycolaldehyde: L-erythrose and D-threose. Later we achieved the preparation of L-glucose by successive *anti* and *syn* aldol additions by a tandem of FSA mutants. Using variants of FSA developed in this work the L-glucose derivatives such as 6-metoxy-L-glucose and 6-chloro-L-glucose were also obtained in one pot.

This method is a uniquely straightforward way for the preparation of aldose sugars in one pot, with one aldolase, and without the need of protection and deprotection steps, or different auxiliary enzymes and isomerization.

# 3.3. Engineering FSA for unprecedented donor substrate flexibility

Aldolases generally provide very strict donor specificity originated from the mechanistic requirement of creating the carbanion nucleophile precisely shielded from the bulk water media and adequately enantiotopically oriented. Even small structural variations in the nucleophile structure, such as bioisosteric phosphonate mimicking DHAP donor of the DHAP-dependent aldolases, resulted in a decrease of activity of up to several orders of magnitude [87, 104]. This indicates the strong influence of steric and electronic factors to stabilize and orient the activated nucleophile in the catalytic active site of the aldolase. Donors of larger size than the native substrate can clash with catalytic residues, and when changing or adding a functional group in the donor structure, the nucleophilicity may be strongly altered thus the substrate could evolve to an unproductive binding mode.

FSA proved to have a broad substrate spectrum including at least four different donor compounds, i.e. DHA, HA, HB and GO, which is rather unprecedented as compared to the single nucleophilic substrate of the DHAP-dependent aldolases. It is important to note that the presence of the hydroxymethylene unit in all donors is, to the best of our knowledge, mandatory for the enamine nucleophile generation. Non-hydroxylated acetone and butanone are unreactive with wild-type FSA [123] as reported up to date. However, a communication stated that FSA may also accept acetone as donor (Biotrans, 2011) [126].

Successful attempts to modify the nucleophile preference of FSA by enzyme engineering were reported, e.g. the mutant FSA A129S has completely changed the order of donor reactivity as compared to the preferences of FSA wild-type, i.e. wild-type HA>GO>>DHA and FSA A129S variant DHA>HA>GO [6, 121]. As of today, no studies have been published about structure-guided engineering of the donor binding pocket of FSA to obtain variants with broader donor selectivities, e.g. nucleophiles with increased size and with additional functional groups.

In collaboration with the group of Professor Wolf-Dieter Fessner from Department of Organic Chemistry and Biochemistry, Technische Univesität Darmstadt (TU Darmstadt), Germany, we investigated the utility of structure-guided protein engineering to design variants of FSA wild-type that offer an extended tolerance towards 1-hydroxy-2-alkanones and 1-hydroxy-3-alkoxy ketones baring long linear, branched and functionalized moieties as donor substrates. This will lead to new biocatalysts with tremendous possibilities for the synthesis of innovative structures resulting in the discovery of novel lead compounds.

## 3.3.1. Selection of novel donor substrates and model aldol-reaction

In this collaborative study we intended to synthesize donor substrates homologous to the hydroxyketone nucleophiles (i.e., HA and DHA) to test in model aldol reactions using redesigned FSA variants as catalyst. The synthesis of a consistent collection of donors was conducted by Deniz Güclü, PhD student from the German team. Secondments of the author of this thesis to the Darmstadt laboratory were done in the frame of a COST Action to work there in this collaborative project

### *a. Selection and synthesis of new nucleophiles*

First, the synthesis of hydroxyacetone analogs (i.e. 1-hydroxy-2-alkanones) were attempted since FSA wild-type tolerates this donor with the highest activity [121, 125]. To this end, new 1-hydroxy-2-alkanones **39b**-**d** (Figure 3.3.1.) containing chain extension from five to seven carbon atoms were synthesized by the selective oxidation of commercially available diols (Scheme 3.3.1.) [197]. Hydroxybutanone (**39a**) which is known to be a good donor of FSA wild-type [123] was also included in this study.

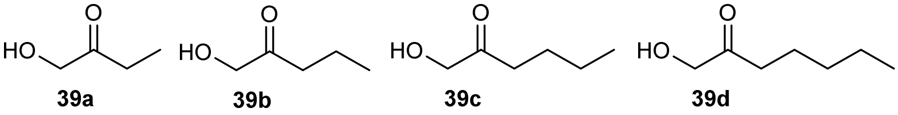
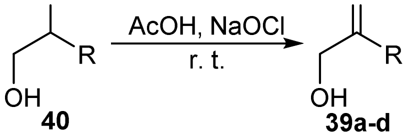


**Figure 3.3.1.** 1-Hydroxy-2-alkanones to be tested in aldol-reactions.



**Scheme 3.3.1.** Synthesis of 1-hydroxy-2-alkanones by the selective oxidation of diols [197].

Isomeric donors were also designed in which the alkyl part was branched demanding more steric space to the enzyme. These nucleophilic substrates can either have isopropyl (**39e**), isobutyl (**39f**), isopentyl (**39g**), tert-butyl (**39h**) and neopentyl (**39i**) substitutions (Figure 3.3.2.).

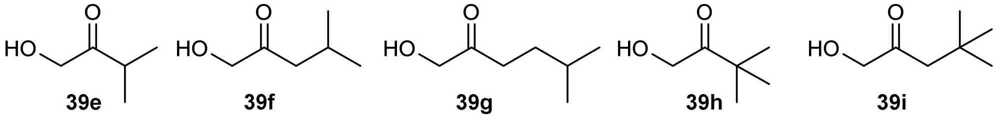
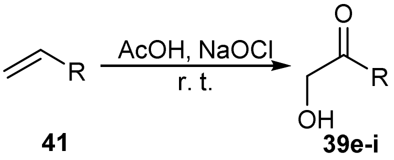


**Figure 3.3.2.** Branched 1-hydroxy-2-alkanones considered for screening with the engineered FSA variants with increased substrate space.

The branched 1-hydroxy-2-alkanones (**39e**-**i**) were synthesized by the ketohydroxylation of commercially available olefins (Scheme 3.3.2.) [198].



**Scheme 3.3.2.** Synthesis of branched 1-hydroxy-2-alkanones by the keto-hydroxylation of olefins [198].

To fully investigate the catalytic flexibility of FSA towards different donors, novel nucleophiles that incorporate changes in the chemical functionality were also designed. The utilization of these 1-hydroxy ketones in enzyme catalyzed aldol addition would grant the access to different and novel classes of chiral compounds achieved by FSA catalysis. These donor substrates bear other functionalities exceeding the ones that are indispensable for the enzyme catalysis (i.e. the hydroxymethylene moiety). Thus, a set of ketones that are ether derivatives of DHA (i.e. 1-hydroxy-3-alkoxy ketones) (Figure 3.3.3.) were synthesized to be tested in aldol reactions. An increase in the activity of these nucleophiles can be anticipated compared to 1-hydroxy-2-alkanones of a similar size due to the electron-withdrawing effect of the ether oxygen and also to the extra polar contacts provided by the additional oxygen atom in the active site of the enzyme.

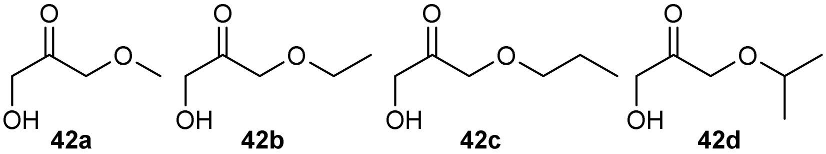
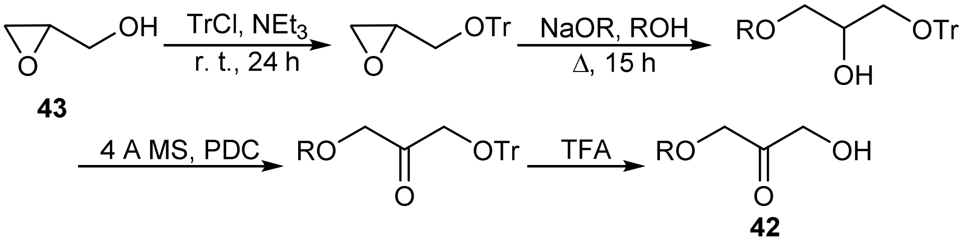


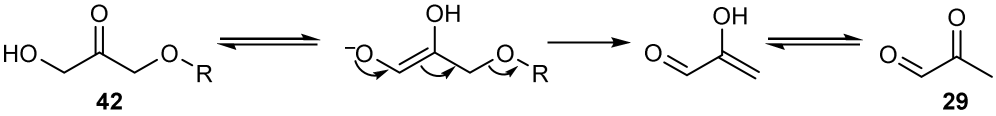
**Figure 3.3.3.** 1-Hydroxy-3-alkoxy ketones to be tested with the engineered FSA variants as possible donor substrates.

The library of ketones **42a**-**d** was synthesized with the general procedure of the modification of epoxides to synthesize 1-hydroxy-3-alkoxy ketones (Scheme 3.3.3.).



**Scheme 3.3.3.** Synthesis of 1-hydroxy-3-alkoxy ketones.

It was discovered, that these ethers can undergo decomposition under acidic conditions (Scheme 3.3.4.). Hence, their stability was tested at different pH (range pH 1-6) using compound **42a** as a model substrate. Initially, the solution of **42a** was at pH 1 due to the trityl deprotection with TFA in the last step of its synthesis (Scheme 3.3.3.). HPLC analysis revealed that the concentration of **42a** was 10-fold lower than expected, and it was observed that it decomposes to methylglyoxal (**29**). This was determined by comparing the new peaks with standard commercial sample of methylglyoxal (Figure 3.3.4.). It was observed that the substrates with shorter chain were the ones that decompose faster.



**Scheme3.3.4.** The possible decomposition of the ether substrates.

To overcome this problem, these 1-hydroxy-3-alkoxy ketones were synthesized again, with a special care taken removing the TFA from the reaction mixture. When performing the aldol reaction the solution containing the compound was neutralized to pH 5, and then to reach the pH necessary for the aldolase catalyzed reaction it was mixed with aqueous buffer solution of pH 8. By this methodology the decomposition of **42a**-**d** could be minimized.

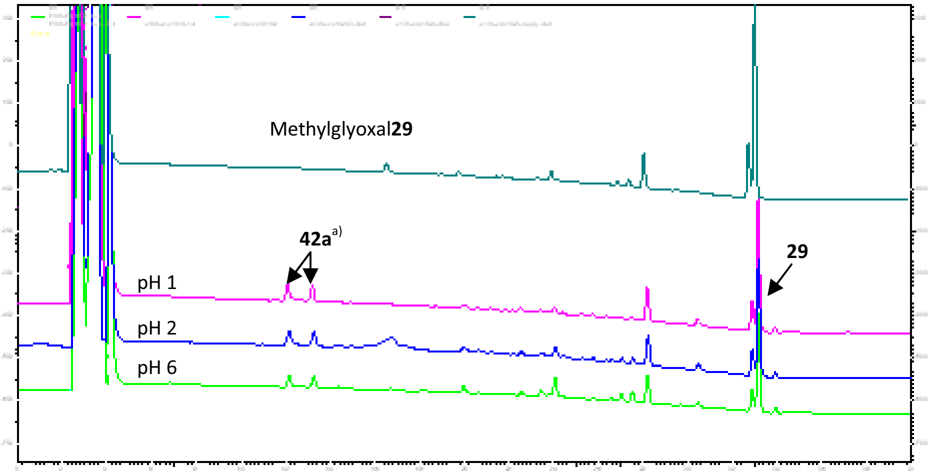


**Figure 3.3.4.** Chromatograms of the solutions of **39a** at different pH, compared to commercially available methyl-glyoxal (**29**). a) The two peaks represent the cis- and trans-isomer of the oxime generated by the addition of the substrate and the derivatization reagent (*O*-benzylhydroxylamine). b) The pH of the solution was adjusted to different values (pH 1, pH 2, and pH 6).

Unsaturated ether derivatives were also targeted to elucidate if these structures have advantageous or detrimental effect on the catalysis with FSA variants, and more importantly to synthesize products which can be further modified by follow-up chemistry to other sugar analogues.

The three derivatives (Figure 3.3.5.) were produced following the general protocol described, in the synthesis of 1-hydroxy-3-alkoxy ketones (Scheme 3.3.3.).

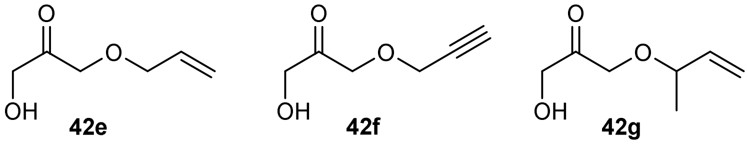


**Figure 3.3.5.** Unsaturated ether derivatives tested with the FSA mutants.

Finally, other potential donor substrates were also synthesized or were purchased from commercial sources which cannot be classified in the categories mentioned above (Figure 3.3.6.). These donors are either functionalized, or bear bulky substituents.

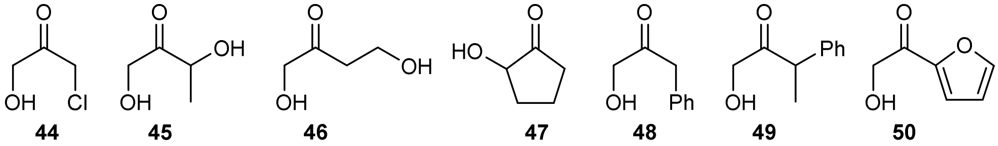


**Figure 3.3.6.** Possible donor substrates in aldol addition reactions with the novel FSA mutants.

## 3.3.2. Rational engineering of FSA for accepting chain-extended and functionalized nucleophiles

FSA was modified by rational protein engineering using site directed mutagenesis to modify the residues that prevent the synthesized sterically more demanding donor substrates to be tolerated by the enzyme. The mutations were introduced in the active site of the enzyme and are intended to remove the space restrictions required to accommodate the donor substrates with increasing size and functionality.

From the crystal structure of the enzyme, three residues were the obvious choices for modification to increase the size of the donor binding site. The target residues were Leu107, Ala129 and Leu163 (Figure 3.3.6.), that are in contact with the modified moiety of the 1-hydroxy ketone donors. To enlarge the active site the two leucines were substituted by alanines. This amino acid was chosen since it has a hydrophobic character and it can safely be assumed that the optimal folding of the protein will be maintained. The Ala129 was substituted by glycine, the only choice to provide extra space and the variant FSA A129S, that was earlier described to provide a rate acceleration effect with some donors, was also included in the study.

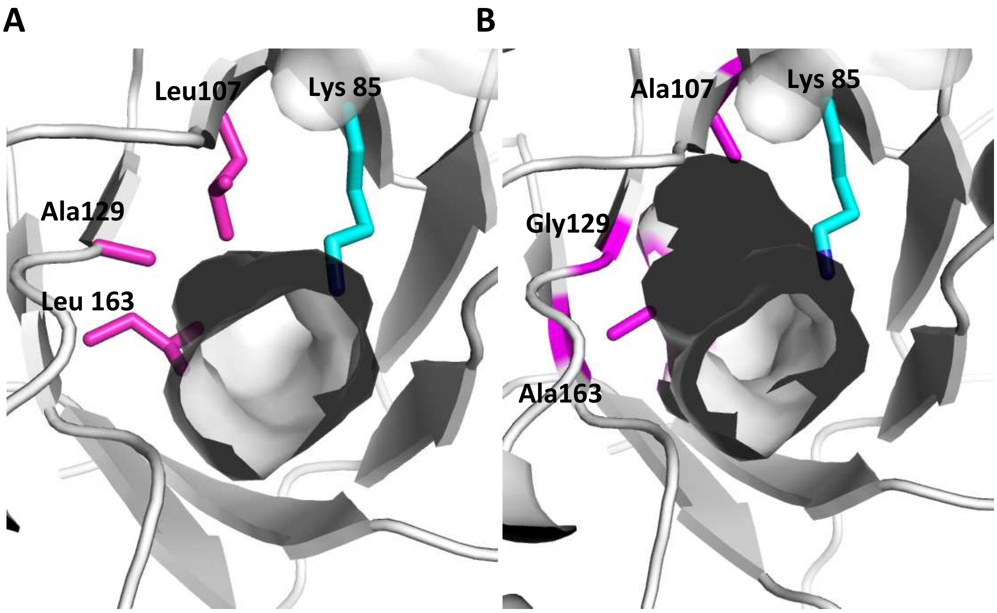


**Figure 3.3.6.** Active site of FSA showing the binding pocket that can accommodate the aliphatic side chain of the novel donor substrates. Catalytic lysine shown in cyan and the mutated residues shown in magenta **A:** FSA wild-type. **B:** the increased cavity due to substitutions in position 107 alanine for leucine, in position 129 glycine for alanine and in position 163 alanine for leucine.

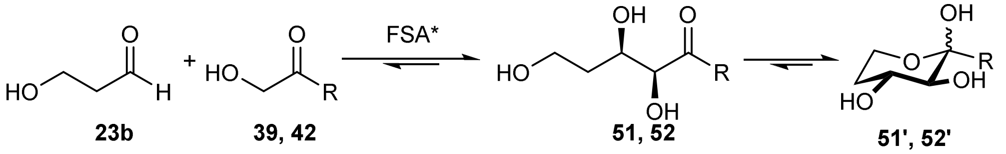
### *a. First round of mutagenesis*

In the first round, the following mutants were constructed: FSA L107A, FSA A129G, FSA A129S and FSA L163A. The enzymes were expressed and purified according to previously established protocols (see Experimental section).

The nucleophiles **39a**-**i**, **42a**-**g**, and **44**-**50** were tested in aldol reactions catalyzed by FSA wild-type and variants in our research group of IQAC-CSIC in Barcelona.

3-Hydroxypropanal was selected as the acceptor substrate. This electrophile has already shown to have high activity in FSA catalyzed reactions with all described ketone donors of the enzyme [123]. In a previous study conducted by Professor Fessner's team, among a small library of 8 structurally varied aldehydes, 3-hydroxypropanal was found the most active with the three tested donor substrates of FSA. This electrophile is an ideal acceptor substrate for aldol additions because 6-membered ring structures are formed by intramolecular hemiketal formation, consequently shifting the reaction equilibrium towards the adduct formation

(Scheme 3.3.5.). Moreover, the stereochemistry of the six-membered rings can be easily and unequivocally determined by NMR.



**Scheme 3.3.5.** The screening reaction with the homologous donor substrates, for the evaluation of the engineered FSA variants (FSA*) and ring formation by intramolecular hemiketal generation of the aldol adducts.

The screening reactions were conducted under identical conditions (Scheme 3.3.5.) with an excess of the acceptor substrate to ensure the shift of equilibrium to the product formation. An FSA variant (2 mg mL$^{-1}$) catalyzed the additions. The formation of the aldol product was monitored at 2 and 24 h reaction times by HPLC (for protocol see Experimental section).

**Table 3.3.1.** Results of the screening reaction with the enzymes baring mutations in the donor binding site with novel donor substrate.

| | **39a** | | **39b** | | **39c** | | **369** | | **39e** | | **36f** | | **39g** | | **39h** | | **39i** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| time (h) | 2 | 24 | 2 | 24 | 2 | 24 | 2 | 2 | 24 | 2 | 24 | 24 | 2 | 24 | 2 | 24 | 24 | 48 | | | | | | |
| **FSA wt** | 65 | 95 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | | | | | | |
| **FSA L107A** | 47 | 94 | nr | 13 | nr | 7 | nr | nr | nr | nr | nr | nr | nr | 13 | nr | nr | nr | nr | | | | | | |
| **FSA A129G** | 54 | 94 | 3 | 47 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | | | | | | |
| **FSA A129S** | 53 | 93 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | | | | | | |
| **FSA L163A** | 95 | | 95 | | 62 | 96 | 37 | 88 | nr | nr | nr | nr | nr | 4 | nr | nr | nr | nr | | | | | | |
| | | | **42a** | | **42b** | | **42c** | | | | | | **42d** | | | | | | **42e** | | **42f** | | **42g** | |
| time (h) | | | 2 | 24 | 2 | 24 | 2 | 24 | | | | | 2 | 24 | | | | | 2 | 24 | 2 | 24 | 2 | 24 |
| **FSA wt** | | | 19 | 95 | nr | nr | nr | nr | | | | | nr | nr | | | | | nr | nr | nr | nr | nr | nr |
| **FSA L107A** | | | 20 | 94 | 5 | 19 | nr | 13 | | | | | nr | nr | | | | | 6 | 10 | 5 | 8 | nr | nr |
| **FSA A129G** | | | 23 | 95 | nr | nr | nr | nr | | | | | nr | nr | | | | | nr | nr | nr | nr | nr | nr |
| **FSA A129S** | | | 12 | 55 | nr | nr | nr | nr | | | | | nr | nr | | | | | nr | nr | nr | nr | nr | nr |
| **FSA L163A** | | | 81 | 100 | 78 | 99 | 20 | 88 | | | | | nr | 4 | | | | | 85 | 95 | 84 | 96 | nr | nr |

Percentage of aldol adduct formation of different donor substrates (100 mM) to hydroxypropionaldehyde (140 mM) at 2 and 24 h. 2 mg mL$^{-1}$ of protein was used. Percentage calculated with respect to the limiting donor substrate. n.r.: no reaction.

The FSA L163A variant catalyzed the reaction with all tried 1-hydroxy-2-alkanones with linear chain, furnishing full conversion, in 2 hours with **36a** and **36b**, and it was the only variant that tolerated **39c** and **39d** (Table 3.3.1.). Nucleophile **36a** was fully converted after 2 h using FSA L163A, whereas the rest of the variants needed 24 h. The homolog, **39b**, was not tolerated by FSA A129S, therefore for this donor increased space at the active site is needed. Nucleophiles **39b** and **39c** were only converted by FSA L107A and FSA L163A variants.

As expected, the branched donors **39e**-**i** and **42d** were not tolerated by these variants (Table 3.3.1.), since to accommodate these substrates more space is necessary. Thus further combination of mutations was envisaged.

The 1-hydroxy-3-alkoxy ketone derivatives (**42a**-**d**) were tolerated with a similar pattern to that of 1-hydroxy-2-alkanone donors (Table 3.3.1.). In this case, FSA L107A and FSA L163A were the most active variants, uniquely tolerating ethers with the long alkyl chain, **42b** and **42c**. Interestingly **42a** was tolerated by the wild-type aldolase, while its alkanone analog 3**9b** was not. It is likely that the presence of the ether as an electron withdrawing group increased its reactivity and improved the binding efficiency by establishing H-bonds.

The FSA variant L163A was also the best mutant for the unsaturated ether derivatives **42e**-**f**. It furnished high conversions for **42e**-**f** after 2 h (84% and 85%, respectively) and full conversion after 24 h. None of these variants was able to accept **42g** as nucleophile (Table 3.3.1.).

The donor substrates **44**-**50** did not furnish any products with the set of FSA variants with single mutations.

### *b. Second round of mutagenesis*

In the second round of mutagenesis, the mutations described above were combined to create double mutants seeking beneficial substitutions and synergistic effects. All the variants were designed on the FSA L163A substitution. Hence, FSA L107A/L163A, FSA A129G/L163A and FSA A129S/L163A were constructed, expressed and purified. FSA L107A/A129G was also designed as a control. This variant is potentially advantageous, since these substitutions also provide increased size for the substrate binding pocket. The new FSA variants thus constructed were tested against the selected donor substrates (Table 3.3.2.). Among them, the ones baring the mutation L163A gave high conversions (>90%) with i) linear1-hydroxy-2-alkanones **39a**-**d**, ii) 1-hydroxy-3-alkoxy ketones **42a**-**c**, iii) unsaturated ether derivatives **42 e**-**f** iv) **42g**, and good conversion (up to 66%) was achieved with **39f**, **39i** and **42d**. Most interestingly, the variant containing the double leucine to alanine replacement FSA L107A/L163A, excelled by achieving full conversion after short reaction times (i.e., 2 hours), even with ketones with the longest chains **39c**-**d**, **42b**-**c** and with the branched ketone donor **39g**.

The variants incorporating the substitution A129G, i.e. A129G/L163A and L107A/A129G displayed lower activity, even though this mutation alone showed improved capacity to tolerate enlarged substrates compared to the wild-type. From the two combinations baring A129G mutation FSA A129G/L163A was the more successful, possibly because of the beneficial effect of L163A mutation. The double mutant FSA L107A/A129G increased the conversions to some substrates, but this increment was not significant (Table 3.3.2.), and with substrates **39a**-**b** and **42a** the combination was disadvantageous, as the conversions decreased compared to the corresponding single variants.

The donor substrates **44**-**50** did not furnish any products with the double mutants.

**Table 3.3.2.** Results of the screening reaction with the second round of enzymes baring combination of mutations in the donor binding site with sterically demanding donors.

| | **39a** | | **39b** | | **39c** | | **39d** | | **39e** | | **39f** | | **39g** | | **39h** | | **39i** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| time (h) | 2 | 24 | 2 | 24 | 2 | 24 | 2 | 2 | 24 | 2 | 24 | 24 | 2 | 24 | 2 | 24 | 24 | 48 | | | | | | |
| **FSA L107A/A129G** | 10 | 50 | 3 | 15 | 4 | 20 | nr | 5 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | | | | | | |
| **FSA L107A/L163A** | 81 | 95 | 97 | | 95 | | 95 | | nr | nr | nr | 52 | 74 | 100 | nr | nr | | | | | | | | |
| **FSA A129G/L163A** | 64 | 100 | 48 | 91 | 13 | 74 | 3 | 29 | nr | nr | nr | 19 | nr | 19 | nr | nr | | | | | | | | |
| **FSA A129S/L163A** | 96 | | 69 | 94 | 38 | 96 | 26 | 92 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | | | | | | |
| | | | **42a** | | **42b** | | **42c** | | | | | | **42d** | | | | | | **42e** | | **42f** | | **42g** | |
| time (h) | | | 2 | 24 | 2 | 24 | 2 | 24 | | | | | 2 | 24 | | | | | 2 | 24 | 2 | 24 | 2 | 24 |
| **FSA L107A/A129G** | | | 18 | 86 | 70 | 94 | 12 | 60 | | | | | nr | nr | | | | | 70 | 88 | 78 | 96 | nr | nr |
| **FSA L107A/L163A** | | | 71 | 99 | 99 | 99 | 99 | 99 | | | | | 12 | 66 | | | | | 82 | 96 | 80 | 96 | nr | nr |
| **FSA A129G/L163A** | | | 57 | 93 | 30 | 71 | 10 | 46 | | | | | nr | nr | | | | | 70 | 96 | 69 | 96 | nr | nr |
| **FSA A129S/L163A** | | | 95 | | 40 | 75 | 25 | 60 | | | | | nr | nr | | | | | 68 | 94 | 70 | 96 | nr | nr |

Percentage of aldol adduct formation of different donor substrates (100 mM) to hydroxypropionaldehyde (140 mM) at 2 and 24 h. 2 mg mL$^{-1}$ of protein was used. Percentage calculated with respect to the limiting donor substrate. n.r.: no reaction.

### *c. Third round of mutagenesis*

In the third stage of mutagenesis all tried substitutions were combined in two triple mutants, thus FSA L107A/A129G/L163A and FSA L107A/A129S/L163A were designed. To compare their activity to the ones described earlier, these FSA variants were tested against the selected substrates (**36a**-**i** and **39a**-**g**) (Table 3.3.3.). The triple mutant FSA L107A/A129S/L163A showed indistinguishable activity as compared with the double mutant L107A/L163A. Both variants achieved full conversion with **39a**-**d**, **39g**, **42a**-**c** and **42e**-**f** substrates, while with the branched donors **39f** and **42d** lower conversions were obtained (23% and 7% with FSA L107A/A129S/L163A vs. 52% and 66% with FSA L107A/L163A, respectively) (Table 3.3.3.).

The enzyme baring the A129G substitution, FSA L107A/A129G/L163A showed decreased activity compared to the original FSA L107A/L163A. This loss of activity indicates that further increment in the catalytic cavity is not necessary for accommodating the donor nucleophiles with increased side chains. Although the mutation A129G increased the FSA's activity as compared to the wild-type with **39a**, **39b** and **42a**, the combination of this mutation with either L107A or L163A substitutions were counterproductive. This can be possibly explained by the low rate of interaction between the enzyme and substrate due to the volume of the cavity of the active site. This means that enzyme-substrate interactions were lost between the catalytic residues and the substrates during the different steps of catalysis, similarly to the mutant FSA A129G/A165G from Section 3.1.2. Another explanation could be that by replacing the alanine by glycine placed next to the already increased void by the replacement of L163 and L107 is creating higher flexibility on this site, which may be detrimental for the effective catalysis.

The FSA variant with the double leucine replacement, FSA L107A/L163A was the most efficient with all branched ketones, even superior to triple mutants FSA L107A/A129G/L163A and FSA L107A/A129S/L163A.

The tests revealed that none of the enzymes showed tolerance to the nucleophiles **39e** and **39h**, which have the branching directly adjoining the carbonyl group. This phenomenon is most likely to be a steric effect, which causes major clashes with some catalytically essential residues of the active site. Nucleophiles with branched substitutions where the most bulky part was separated from the carbonyl group, by at least one methylene moiety, were tolerated by the enzymes that had both the substitutions in the L107 and L163 positions. It seems that the further was the branching from the catalytically essential parts of the nucleophile, the better it was tolerated by the enzyme variants (e. g. the isopropyl **39e** versus the isopentyl **39g** substituted donors). Nevertheless the reaction rates of the branched hydroxyalkanones are significantly reduced compared to the corresponding linear chain isomers (Table 3.3.2. and Table 3.3.3.).

Unfortunately, as expected no conversion was observed with any of the donors **44**-**50**, and by any of the tested enzymes, so probably further mutagenic study is necessary, which will be later carried out in our research group.

**Table 3.3.3.** Results of the screening reaction with the third round of enzymes baring combination of mutations in the donor binding site with sterically demanding donors.

| | **39a** | | **39b** | | **39c** | | **39d** | | **39e** | | **39f** | | **39g** | | **39h** | | **39i** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| time (h) | 2 | 24 | 2 | 24 | 2 | 24 | 2 | 2 | 24 | 2 | 24 | 24 | 2 | 24 | 2 | 24 | 24 | 48 |
| **FSA L107A/A129G/L163A** | 5 | 14 | 15 | 67 | 20 | 72 | 27 | 69 | nr | nr | nr | 23 | 12 | 63 | nr | nr | | |
| **FSA L107A/A129S/L163A** | 87 | 97 | 97 | | 97 | | 91 | 96 | nr | nr | nr | 23 | 56 | 100 | nr | nr | | |

| | **42a** | | **42b** | | **42c** | | **42d** | | **42e** | | **42f** | | **42g** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| time (h) | 2 | 24 | 2 | 24 | 2 | 24 | 2 | 24 | 2 | 24 | 2 | 24 | 2 | 24 |
| **FSA L107A/A129G/L163A** | 26 | 54 | 3 | 9 | 9 | 20 | nr | nr | 10 | 20 | 8 | 15 | nr | nr |
| **FSA L107A/A129S/L163A** | 95 | | 99 | 99 | 98 | 99 | 15 | 7 | 80 | 95 | 81 | 96 | nr | nr |

Percentage of aldol adduct formation of different donor substrates (100 mM) to hydroxypropionaldehyde (140 mM) at 2 and 24 h. 2 mg mL$^{-1}$ of protein was used. Percentage calculated with respect to the limiting donor substrate. n.r.: no reaction.

### 3.3.3. Kinetic measurements of the cross aldol reactions

To efficiently compare the catalytic activity of the FSA variants towards the different donor substrates the initial reaction rates ($v_0$, µmol min$^{-1}$ mg$^{-1}$ of protein) were measured.

Thus, aldol reactions (Scheme 3.3.1.) were carried out on analytical scale under identical conditions, using 150 mM of 3-hydroxypropanal and 100 mM solution of the selected nucleophile. The most productive enzyme of each mutagenic round (i.e., FSA L163A, FSA L107A/L163A and FSA L107A/A129S/L163A including FSA wild-type with **39a** and **42a**) was evaluated. The enzyme amount was adjusted to ensure linear dependence of the product formation versus time (0-20 min). To calculate the initial velocities the reactions were only monitored up to 10% conversion of the aldol products.

The initial rates ($v_0$) of the examined FSA variants were consistent with the yields achieved. The anticipated difference of the activities between the 1-hydroxy-3-alkoxy ketone and 1-hydroxy-2-alkanones was only obvious in case of the **39b** vs. **42a**, where the activity of each FSA variant was higher with the 1-hydroxy-3-alkoxyketone (**42a**) as hypothesized before. The rest of the structurally analogous alkanones and ethers showed similar reaction rates with the same mutants (Table 3.3.4.). As expected, the highest activities were found with the double mutant FSA L107A/L163A with of most of the substrates, and the $v_0$ decreased in the order of FSA L107A/A129S/L163A>FSA L163A. The only exception to this rule was **39a**, which showed the highest activity with FSA L163A. 1-Hydroxy-2-butanone (**39a**) has the shortest alkyl chain from the tested analogs, and this would probably result in a loss of binding contacts with the FSA variants with the increased steric space (e.g. FSA L107A/L163A, FSA L107A/A129S/L163A).

The best substrate for all analyzed FSA variants was **42a**, which possibly have the most effective contact with the binding site of the enzyme and it also have higher carbonyl electrophilicity than the equivalent alkanone due to the presence of an electron withdrawing group.

**Table 3.3.4.** Table of initial reaction rates of the homologous donor substrates: 1-hydroxy-2-alkanones and 1-hydroxy-3-alkoxy-ketones with 3-hydroxypropanal, catalyzed by the most effective FSA variants from each round of mutagenesis.

| | $v_0$ (mmol min$^{-1}$ mg$^{-1}$of protein$^{a)}$) | | | | |
|---|---|---|---|---|---|
| | **39a** | **39b** | **39c** | **39d** | **39g** |
| **FSA wt** | 0.31 ± 0.02 | nr | nr | nr | nr |
| **FSA L163A** | 0.98 ± 0.04 | 0.71 | 0.11 | 0.06 | nr |
| **FSA L107A/L163A** | 0.48 | 1.33 | 1.25 | 0.94 ± 0.10 | 0.32 ± 0.01 |
| **FSA L107A/A129S/L163A** | 0.47 | 0.73 | 0.61 ± 0.08 | 0.60 | 0.30 ± 0.01 |

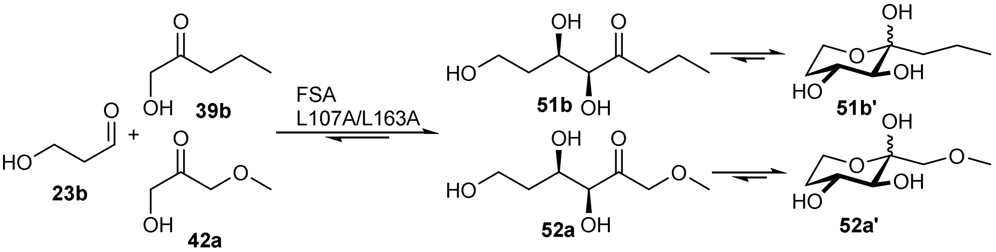
| | $v_0$ (mmol min$^{-1}$ mg$^{-1}$of protein$^{a)}$) | | | |
|---|---|---|---|---|
| | **42a** | **42b** | **42c** | **42d** |
| **FSA wt** | 0.12 ± 0.02 | nr | nr | nr |
| **FSA L163A** | 0.76 ± 0.05 | 0.17 ± 0.02 | 0.18 ± 0.01 | nr |
| **FSA L107A/L163A** | 1.84 ± 0.20 | 1.28 ± 0.13 | 0.54 ± 0.08 | 0.08 |
| **FSA L107A/A129S/L163A** | 1.08 ± 0.19 | 0.58 ± 0.07 | 0.62 ± 0.08 | 0.08 |

Different amounts of enzyme were used for obtaining linearity between 0% and 10% of aldol adduct. For details, see experimental section. a) amount of protein was determined with the Bradford method [199]. b) Values are the mean of three-four determinations. nr: no reaction. The homologous substrates are grouped together, and the structurally similar ones are shown in the same column.

## 3.3.4. Competition experiments

To unequivocally assess the difference of reaction rates between the 1-hydroxy-2-pentanone **39b** and 1-hydroxy-3-methoxy ketone **42a** under the exact same reaction conditions, direct competition experiment monitored by HPLC and *in situ* NMR was performed (Scheme 3.3.6.).

The relative kinetics of **36b** and **39a** were determined for the aldol addition to 3-hydroxypropanal using FSA L107A/L163A as catalyst (Scheme 3.3.6.). Hence, the initial reaction rates ($v_0$) of the formation of **51b** versus **52a** were measured.



**Scheme 3.3.6.** Competition experiments of the aliphatic and ether type donors. Aldol reactions with **39b** and **42a** and 3-hydroxy propionaldehyde catalyzed by the FSA variant L107A/L163A.

The concentration of the donor substrates was kept lower (i.e. 75 mM) as compared to the reactions performed before, to avoid high concentration of organic material, which might inhibit the enzyme (for the rest of experimental conditions see the Experimental Section). To ensure linear dependence of the product formation versus time samples were taken at the beginning of the reaction (i.e., 0 min, 5 min, 15 min, 30 min and 2 hours). Analysis of the reaction after 24 hours was conducted to evaluate the conversion. Initial reaction rates to each product were determined from the product concentrations measured with the HPLC using an external standard method.

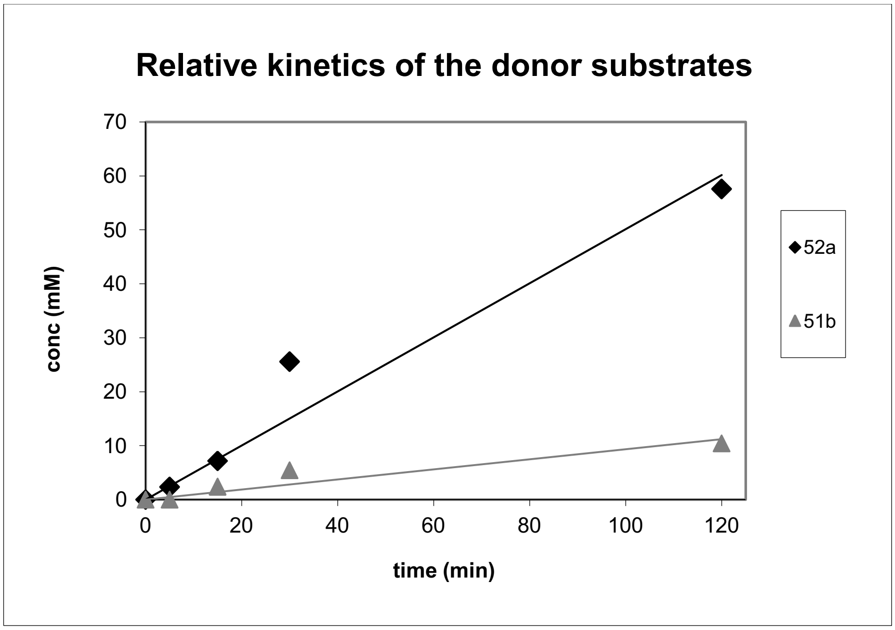


**Figure 3.3.7.** Comparison of the relative kinetics of **39b** and **42a** in aldol reactions to 3-hydroxypropanal with FSA L107A/L163A. Concentration of the adducts were measured by HPLC by an external standard method.

As expected the 1-hydroxy-3-methoxy ketone **42a** was more active in the aldol reaction than the 1-hydroxy-2-pentanone **39b** with around 5-fold increased activity i.e. 0.5 mmol min$^{-1}$ mg$^{-1}$ of protein versus 0.09 mmol min$^{-1}$ mg$^{-1}$ of protein, respectively (Figure 3.3.7.). The conversions after 24 hours were 89% vs. 17% to the aldol adducts **52a** and **51b**, respectively, confirming the $v_0$ results. Under identical conditions the competition reaction was also monitored by NMR. A similar result was observed as by the HPLC, i.e., **52a** showing higher reaction rate than **51b** (Figure 3.3.8.).

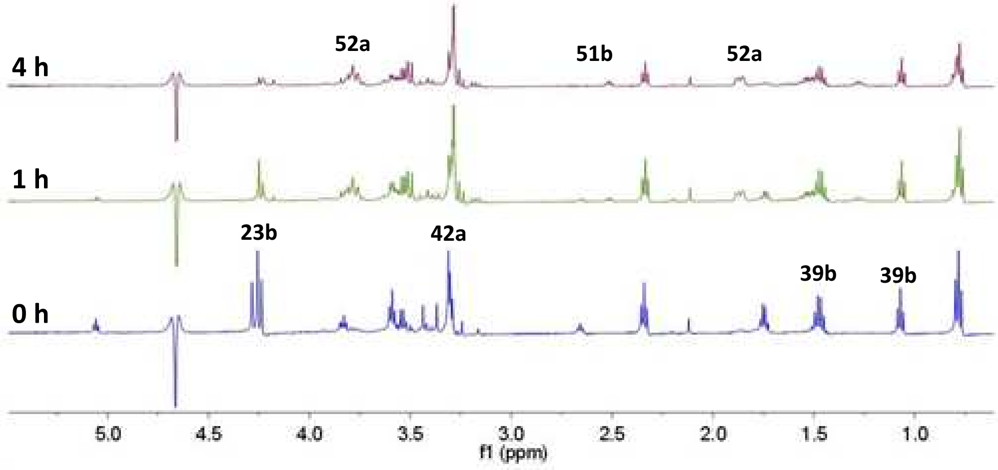


**Figure 3.3.8.** *In situ* reaction monitoring of the competition reaction of **39b** and **42a** and 3-hydroxypropanal with FSA L107A/L163A.

Hence it can be stated that **42a** have a higher reactivity in the aldol reactions with FSA mutants than **39b**, even though it can be considered homologue of the best donor of the aldolase, i.e. hydroxyacetone. The reason for the higher activity is probably the increased electrophilicity of the ether and the additional polar contacts in the donor binding site of the enzyme as stated before.

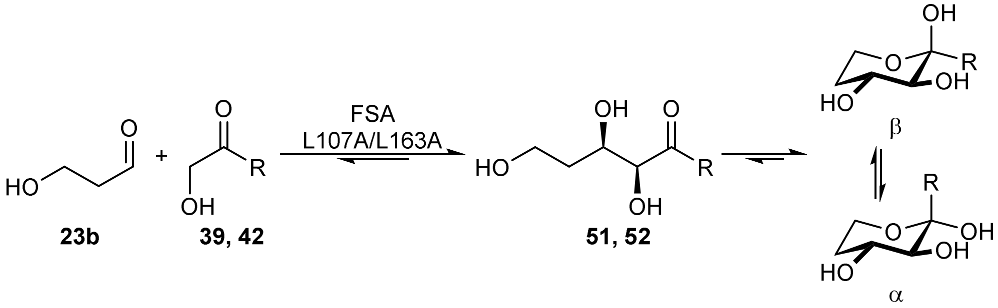
### 3.3.5. Preparative scale reactions

It was proved that direct biocatalytic generation of novel deoxysugars can be achieved by the newly designed FSA mutant L107A/L163A. These products were to the best of our knowledge, unknown from natural sources and as demonstrated in this work accessible by biocatalytic routes for the first time with the novel FSA variants. The isolation of these products was necessary to unequivocally analyze the stereochemical outcome of the enzymatic aldol reactions.

The preparative scale aldol additions were performed by Deniz Güclü from the TU Darmstadt. The reactions were scaled up to 10 mL of total volume, containing the respective donor in 150 mM concentration and 3-hydroxy propionaldehyde at 100 mM concentration. 1.5 mg $mL^{-1}$ enzyme was added to each reaction. 11 aldol adducts were produced with isolated yields between 25-89% (Scheme 3.3.7.).

Reactions with all tested donor substrates were scaled up, except for **39a**, which was already characterized [123], and **39h**, **39e** and **42g** since none of the variants of FSA showed activity with these substrates. The highest isolated yields were achieved with **42a** and **42b** with 89% and 82% respectively. Slightly lower yields were accomplished with the 1-hydroxy-2-alkanones **39b**, **39c** and **39d** with 75%, 76% and 50% respectively. The *O*-allyl substituted ether **42e** gave a moderate yield of 45%, and all products from the branched nucleophiles, **39f**, **39g**, **39i** and **42d** furnished low yields, between 25-30% (Table 3.3.5.).

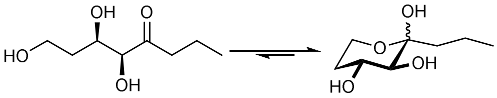
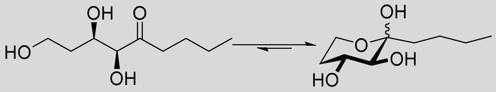
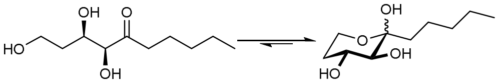
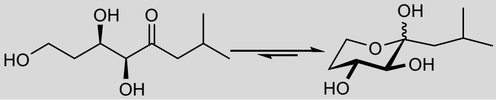
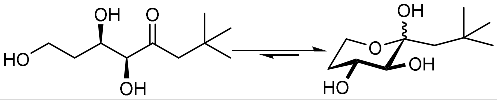
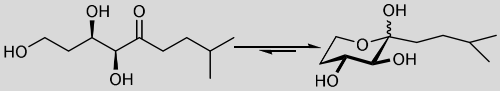
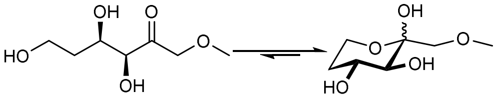
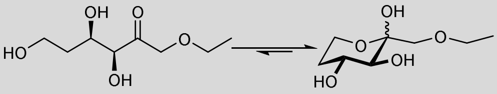
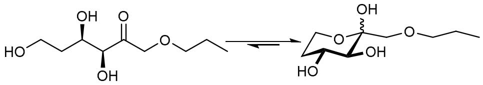
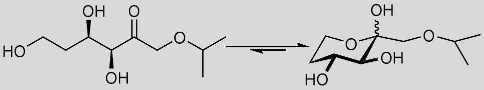
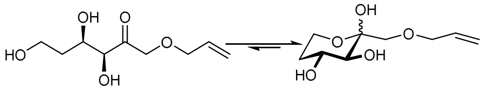
Structural analysis by NMR spectroscopy confirmed that in aqueous solution the products exist in equilibrium between the open chain and the hemiketal 6-membered ring structures, which are in turn, in equilibrium between the α/β-anomeric forms (Scheme 3.3.7.). In both the 1-hydroxy-2-alkanones and 1-hydroxy-3-alkoxy ketone series, the amount of α-anomer strongly decreases with increasing bulkiness of the variable constituent, while the ratio of the open chain isomer increases (Table 3.3.5.).



**Scheme 3.3.7.** Synthesis of aldol products using FSA L107A/L163A as catalyst for addition of novel donors to 3-hydroxypropanal. The products participate in a constitutional equilibration with the open-chain isomer and it reversibly cyclize to furnish 6-membered ring of the α/β-anomeric forms.

FSA wild-type was shown to have extremely high stereoselectivity, with many of the donor and acceptor variations [122, 123, 131, 194]. In this study full stereoselectivity was observed also with all donors using FSA L107A/L163A as catalyst. This is highly remarkable since the mutagenesis of the donor binding site that increases the cavity, i.e. the possibility of different binding modes, could therefore decrease the stereoselectivity. Most interestingly, the NMR spectroscopy revealed that the addition reactions with FSA L107A/L163A also furnished exclusively the D-*threo* (3*S*,4*R*) configuration with all tested nucleophiles being the stereochemistry exclusively controlled by the enzyme.

**Table 3.3.5.** Isolated yield and product ratio of the aldol adducts produced by the preparative scale reactions of the novel donors catalyzed by the newly designed FSA L107A/L163A.

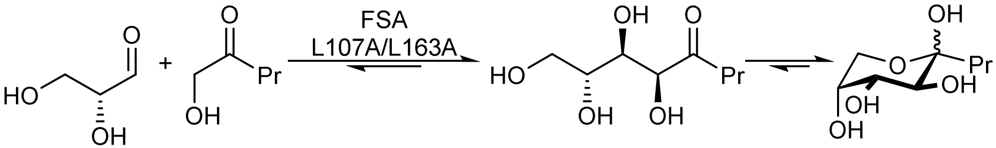
| | Isolated yield | Product ratio OC/α/β |
|---|---|---|
|  | 75% | 67/5/34 |
|  | 76% | 54/5/41 |
|  | 50% | 62/4/34 |
|  | 28% | 64/2/34 |
|  | 25% | 66/2/32 |
|  | 25% | 54/5/41 |
|  | 89% | -/1/99 |
|  | 82% | -/-/100 |
|  | 28% | -/-/100 |
|  | 30% | -/3/97 |
|  | 45% | -/1/99 |

Products synthesized by the aldol addition of the novel donor substrates (150 mM) and 3-hydroxypropanal (100 mM) catalyzed by 1.5 mg $mL^{-1}$ FSA L107A/L163A. The products are shown in equilibrium of the open chained and the 6-membered hemiketal ring form. OC: open chain form, α: ratio of the α-anomer, β: ratio of the β-anomer.

## 3.3.6. Testing other acceptor substrates

To further demonstrate the scope and utility of the novel FSA L107A/L163A variant in aldol addition reactions, other acceptor substrates, such as D-glyceraldehyde and propanal were tested. Simple aldehyde substrates render carbohydrates with higher level of deoxygenation, while D-glyceraldehyde produces deoxy analogs of D-fructose.

The first choice for testing the diverse acceptors was D-glyceraldehyde. This substrate is the most obvious choice, since it is the unphosphorylated derivative of the putative natural substrate of FSA, D-glyceraldehyde 3-phosphate. The aldol addition of **39b** to **53** catalyzed by FSA L107A/L163A (Scheme3.3.8.) was monitored by NMR spectroscopy, which revealed that the product was obtained as one single diastereomer.

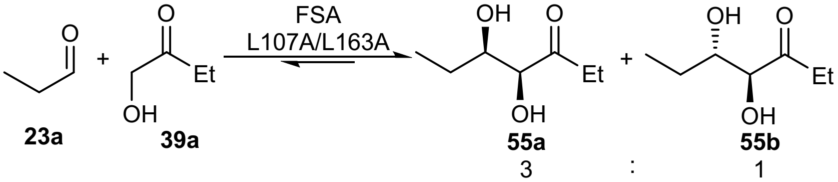


**53** **39b** **54** **54'**

**Scheme 3.3.8.** Synthesis of aldol products using FSA L107A/L163A as catalyst for addition of **39b** to D-glyceraldehyde. The products participate in a constitutional equilibration with the open-chain isomer and it reversibly cyclize to furnish 6-membered ring of the α/β-anomeric forms.

The other more generic acceptor choice was propanal. This non-hydroxylated acceptor substrate was chosen to prove the new enzyme’s ability to effectively catalyze aldol reactions with simple electrophiles unable to perform any asymmetric induction.

The addition of **39a** to propanal catalyzed by FSA L107A/L163A furnished the aldol adduct in high yield (> 95%), however NMR spectroscopy revealed the presence of both D-*threo* : L-*erythro* configuration in a 3:1 ratio (Scheme 3.3.9.). The addition was monitored by HPLC in the early stages of the reaction, and it confirmed that the formation of both diastereomers was of kinetic origin, and not of final configurational equilibration.



**Scheme 3.3.9.** Aldol reaction of propanal and **36a** catalyzed by FSA L107A/L163A produces a mixture of the two diastereomers.

It is remarkable that the same reaction catalyzed by wild-type FSA uniquely furnished the D-*threo* configuration [123]. This change in the diastereoselectivity appeared to be related with the lack of effective selective orientation of the acceptor in the binding site. It is likely that more space in the donor binding site facilitates change in the orientation of the alkyl chain of the propanal. As a consequence, this would change the relative direction of the enantiotopic face of **23a**.

### 3.3.7. Conclusions

In this study we present an unprecedented expansion of the nucleophilic substrate tolerance of FSA. This expansion was accomplished by structure-guided rational protein engineering. Different 1-hydroxy-2-alkanones, 1-hydroxy-3-alkoxyketones, and other 1-hydroxy ketone analogs with branched, functionalized (e.g. unsaturated) structures were tested as donor components.

Mutants were designed to improve the steric space of the FSA active site in the donor binding space. Several active site mutants were tested with substitutions in the donor binding site, which included reducing the bulky leucine residues in position 107 and 163, and the testing of the previously advantageous variations A129G and A129S. The most beneficial substitutions were the L107A and L163A, especially their combination the double mutant L107A/L163A showed increased activity.

The synthetic capacity of these enzymes was tested, and the variants were found to tolerate several hydroxyalkanones, hydroxy-alkoxy ketones, and their branched and functionalized derivatives. The analysis of the obtained products confirmed a high level of stereoselectivity of the enzymatic additions, since all products were D-*threo* stereoisomers. These deoxysugars are from now on accessible by biocatalytic reaction routes.

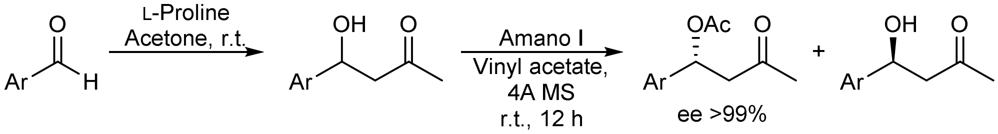
By this study we show that the designed FSA variants open new avenues towards the synthesis of novel deoxysugar product families, and the power of protein engineering in changing the substrate scope of aldolases, enzymes which generally have very strict donor specificity.

## 3.4. Sequential organo- and biocatalytic aldol reactions in the synthesis of deoxysugars

As mentioned in the introduction and in the previous chapter, cascade reactions, a term that encompasses “domino”, “tandem”, and “sequential” reactions including sequences in which further reagents are added at various points (“one-pot” transformations), have undeniable benefits such as increased atom-economy, as well as economies of time, labor, resource management, and waste generation. Examples of catalytic cascade reactions are scarce in the literature but few remarkable examples can be found in the field of organocatalysis [200, 201], biocatalysis [202, 203] and combining organometallic and organocatalysis [204-206]. However, the combination of organocatalysts and enzymes are rare.

Organocatalysis refers to a form of catalysis that use single amino acids and their derivatives for the rate acceleration of chemical reactions. The most commonly used organocatalyst is proline, what is sometimes regarded as the simplest enzymes [33]. Hence, biocatalyst could be considered a kind of highly evolved organocatalysts, and, therefore, a potential complementarity may exist between them, which could be exploited in the field of catalytic aldol additions. Both catalysts have general applicability, although they share particular restrictions, which can be overcome by their beneficial tandem application. Consequently, the sequential application of organo- and biocatalysts may allow the synthesis of complex compounds, and molecular scaffolds with greater chemical diversity than using both technologies independently.

However, only a few examples are reported for their combination in cascade reactions [207-209]. One example is the coupling of organocatalytic aldol addition and enzyme catalyzed kinetic resolutions to furnish enantiomerically pure β-hydroxyketones [207]. The acetone was removed after organocatalytic reaction was finished to be compatible with the ensuing enzyme catalysis (Scheme 3.4.1.). The two systems were efficiently combined sequentially to afford enantiomerically pure aldol adducts in higher conversion than standard resolution procedures.



**Scheme 3.4.1.** Sequential direct catalytic aldol reactions and lipase-catalyzed kinetic resolution [207].

Gröger *et*. *al*. showed the combination of asymmetric organo- and biocatalytic reactions which lead to a sequential construction of both stereogenic centers of

1,3-diols [209] (Figure 3.4.1.). This modular chemo-enzymatic synthetic concept allows the efficient access to all four stereoisomers in enantiomerically pure form.

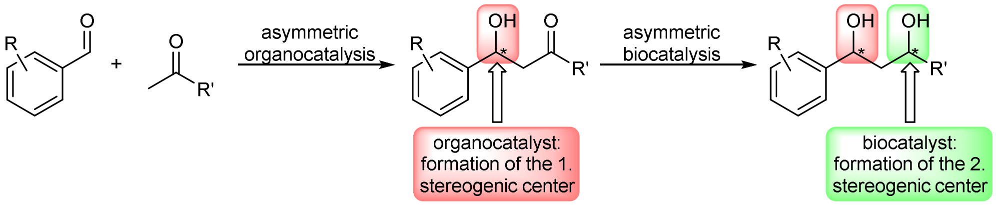
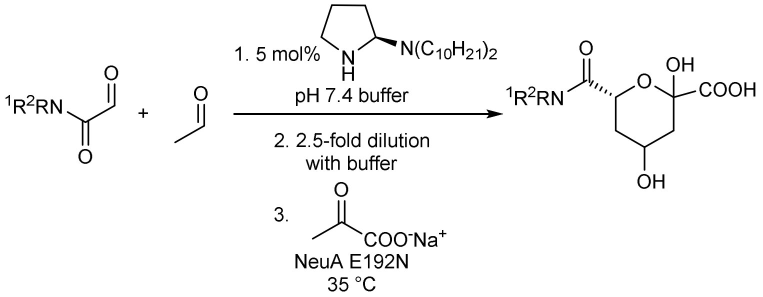


**Figure 3.4.1.** Synthesis of 1,3-diols by combination of organo- and biocatalysis. The first stereogenic center is formed by an organocatalytic aldol reaction, while the second center is formed by enzymatic reduction, with alcohol dehydrogenases. All four isomers were generated by this method.

However in these examples the organocatalytic carbon-carbon bond formation is followed by the enzymatic modification of the functional group earlier generated (i.e. reduction of carbonyl, or esterification).

Up to now, only one report was published about the combination of organo- and biocatalytic aldol additions, where the enzymatic step stereoselectively expanded the structure of the adduct formed by organocatalysis [210].This report describes the specific combination of organo- and enzymatic catalysis in three component reactions leading to heterocyclic products. A system was developed where both carbon-carbon bond formation steps could be carried out in buffered aqueous medium (Scheme 3.4.2.) [210]. The overall stereoselectivity of the process was poor (78:22 cis/trans ratio and ee<5%) due to the lack of enantioselectivity of the organocatalytic step and the low diastereoselectivity of the aldolase catalyzed reaction. Nonetheless, this study demonstrated the possibility of combining organo- and enzymatic aldol reactions in catalytic cascades.



**Scheme 3.4.2.** The aldol reactions catalyzed the combination of by organocatalyst and enzyme catalyst. NeuA: Neuraminic acid aldolase.

## 3.4.1. Cascade organo- and biocatalytic reactions to the synthesis of deoxysugars

The goal of this study was to combine organo- and biocatalytic aldol additions for the synthesis of deoxysugars. Deoxysugars play a crucial role as recognition elements in bioactive molecules [211-213] and can usually be found as conjugates with natural products along with other carbohydrates in polysaccharide structures [214, 215]. The role that they play in many physiological processes is partially attributed to the enhanced hydrophobicity displayed with respect to the fully oxygenated analogues [214, 216-218].

This project was carried out in collaboration with the group of Prof. Miquel A. Pericàs from the Institut Català d'Investigación Química (ICIQ) from Tarragona, who originally prepared the chiral beta-hydroxyaldehydes using polystyrene-supported proline as catalyst in water [219].

When planning the coupled reaction it was taken into consideration that organocatalysis often requires the protection of hydroxyl functions [40, 41, 166]. Thus, it was chosen as the first organocatalytic aldol addition reaction the one that use the lower functionalized starting materials. The second step of the cascade was to be performed by biocatalysis, in which the more functionalized substrates are easily handled and used without the need of protecting the functional groups generated in the first reaction (Figure 3.4.2.).

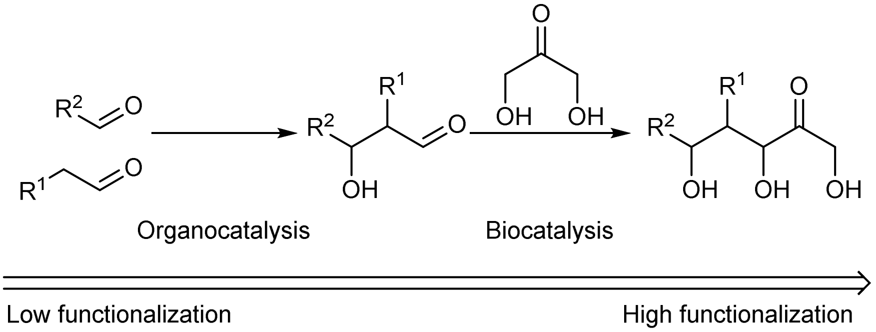


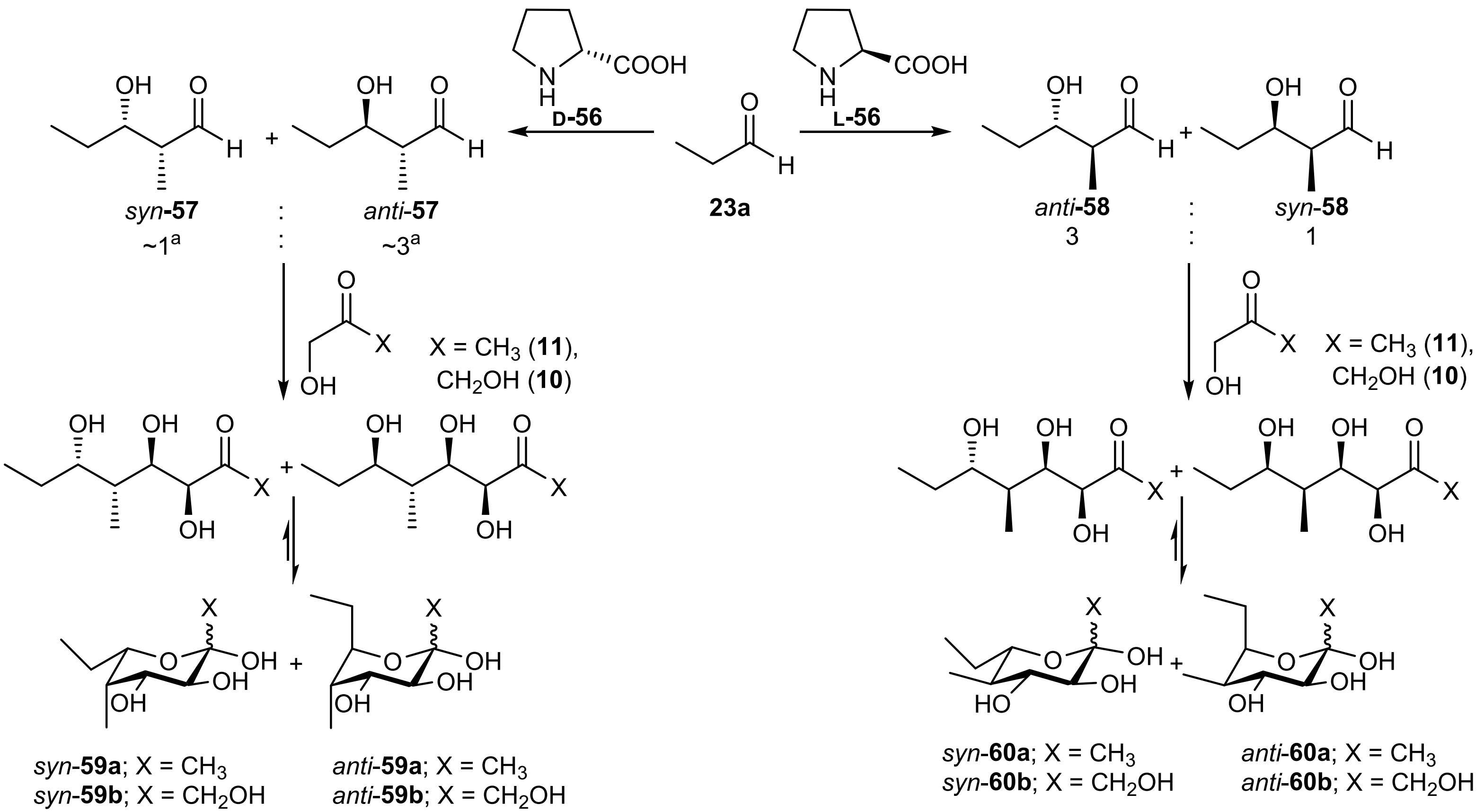
**Figure 3.4.2.** The strategy of coupling organo- and biocatalysis by starting from simple small aldehydes to achieve highly functionalized substrates.

To combine both catalytic methodologies we searched for an organocatalytic reaction between two aldehydes. Thus, the aldehyde formed can be used as an acceptor for a second aldol reaction that takes place by enzyme catalysis (Figure 3.4.2.). To this end, the proline catalyzed dimerization of propionaldehyde was chosen as the model organocatalytic reaction for its high diastereo- (3:1 *anti*:*syn* ratio) and enantioselectivity (99% ee), its simplicity and easy access [43]. The reaction renders (2*S*, 3*S*)-3-hydroxy-2-methylpentanal with a stereochemistry that depends on the organocatalyst L-proline. The same reaction catalyzed by D-proline was also tested to obtain the corresponding enantiomer, i.e. (2*R*, 3*R*)-3-hydroxy-2-methylpentanal (Scheme 3.4.3.). The organocatalytic preparation of (2*S*, 3*S*)-3-hydroxy-2-methylpentanal was the first report on direct and enantioselective cross-aldol reaction of aldehydes catalyzed by L-proline reported by MacMillan and Northup [43].

For the enzymatic aldol reactions we selected FSA catalyst, which proved to be one of the most versatile aldolase, with broad nucleophile and electrophile tolerance [123, 194]. This aldolase was shown to be the perfect candidate to access deoxysugars with different levels of deoxygenation, and with excellent stereoselectivities [123]. The planned reactions are depicted in Scheme 3.4.3. It may be anticipated that after the FSA catalyzed addition of DHA or HA to the aldehydes **57** and **58**, the hemiketals will be formed and the cyclization will shift the reaction equilibrium towards synthesis.

The synthesis of the pentanal derivatives **57** and **58** was first carried out following the described methodologies [43], i.e. 2 M solution of propionaldehyde in DMF with 10% mol of D- or L-proline, to access both enantiomers of the aldol adducts **57** and **58**. Our line of thinking was to first test the suitability of the pentanal enantiomers **57** and **58** in the enzymatic reactions before trying with the supported proline catalyst developed by the ICIQ team.

Because the activity of the biocatalyst was incompatible with the organocatalytic reaction conditions, the enzymatic aldol addition was conducted in a different reactor. Thus, when the homo aldol addition of propionaldehyde was completed, the insoluble proline was removed by filtration and the filtrate added to an aqueous buffer solution containing the donor substrate for enzymatic aldol reaction. The amount of aqueous buffer was adjusted to obtain a final DMF concentration of 18% v/v, which was then fully compatible with the FSA activity. The final donor and acceptor concentrations were 150 mM and 100 mM, respectively (for protocol see Experimental Section).



**Scheme 3.4.3.** The coupled organo- and biocatalytic aldol reactions. The first step was catalyzed by D- and L-proline and the second addition of HA or DHA was enzyme catalyzed with FSA. a) Theoretical ratio of the diastereomers according to [43].

FSA wild-type and variants were tested as potential catalyst for the aldol addition of DHA and HA to pentanals **57** and **58**. Reactions were monitored by HPLC following the appearance of a new peak. As judged by the HPLC analysis, FSA wild-type furnished moderate to low conversion of both **57** and **58** using HA as donor: 44% conversion to **59a** and 21% to **60a**. The results were especially disappointing using DHA as donor: 1 % and 2% conversion was achieved to **59b** and **60b**, respectively (Figure 3.4.5.).

A site directed mutagenesis program was envisaged to increase the FSA tolerance for **57** and **58** acceptors. The mutations were applied in the supposed acceptor binding site of the aldolase, in particular on the residues: Arg 134, Ala 165 [220],Ser 166, Lys 168, Thr 185, and the alanine in position 129 which was described as positive modification for the tolerance of DHA [121](Figure 3.4.3.).

The following substitutions were made that are not yet described in the literature (Figure 3.4.3.):

- **R134E:** This residue was already modified to an alanine in an earlier study, but this substitution did not provide the expected results [220]. The rationale behind this substitution was to change the ionic character to make more favorable the hydrophobic and non-ionic interactions.
- **S166G:** The reason to modify this residue was similar to the change of alanine in position 165, i.e. to improve the steric space of the enzyme for the acceptor. This serine also forms sterical hindrance, thus removing the side chain can help the binding of more bulky acceptor substrates.
- **K168E:** This lysine is right at the entrance of the active site. Its substitution by glutamic acid will change the polarity and could prevent the amino group of lysine to trap the aldehydes as a shift base or even the subsequent formation of a reactive enamine.
- **T185S:** This substitution was intended to remove a methyl group and increase the steric space for the acceptor with suppressing the possibility of hydrogen bond of the $CH_2OH$ side chain of the serine. Moreover, the crystallographic structure of FSA shows three methyl groups in small distance from each other. Thus removing one or more of them would facilitate bulky or branched aldehydes to be accepted.

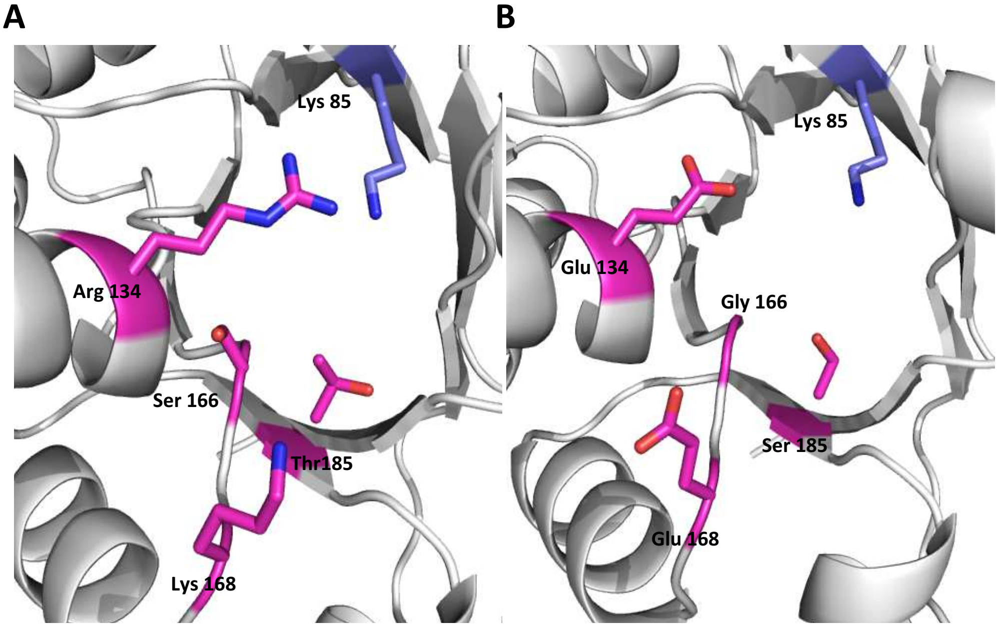


**Figure 3.4.3.** Active site of FSA showing the residues of the acceptor binding site. Catalytic lysine shown in blue and the mutated residues shown in magenta **A:** FSA wild-type. **B**: The increased cavity due to substitutions in position 134 glutamate for arganine, in position 166 glycine for serine, in position 168 glutamate for lysine and in position 185 serine for threonine.

Hence, the following mutants and combination of mutations were initially tried with aldehydes **57** and **58**: FSA A129S [121], FSA A129S/A165G [220], FSA A165G [220], FSA S166G, FSA K168E, FSA T185S, FSA A165G/T185S, and the triple mutant prepared by Mariana Gutierrez PhD, with the objective to increase the space in the acceptor binding site of the enzyme for the allocation of substituents in the α-position of Cbz-aminoaldehydes, FSA A129S/A165G/S166G. Models of the imine form of the adduct of HA and 3-hydroxy-2-methylpentanal in the active site of FSA were constructed and indicate that the substrate can fit better in the FSA A129S/A165G/S166G active site than in the FSA wild-type (Figure 3.4.4.).

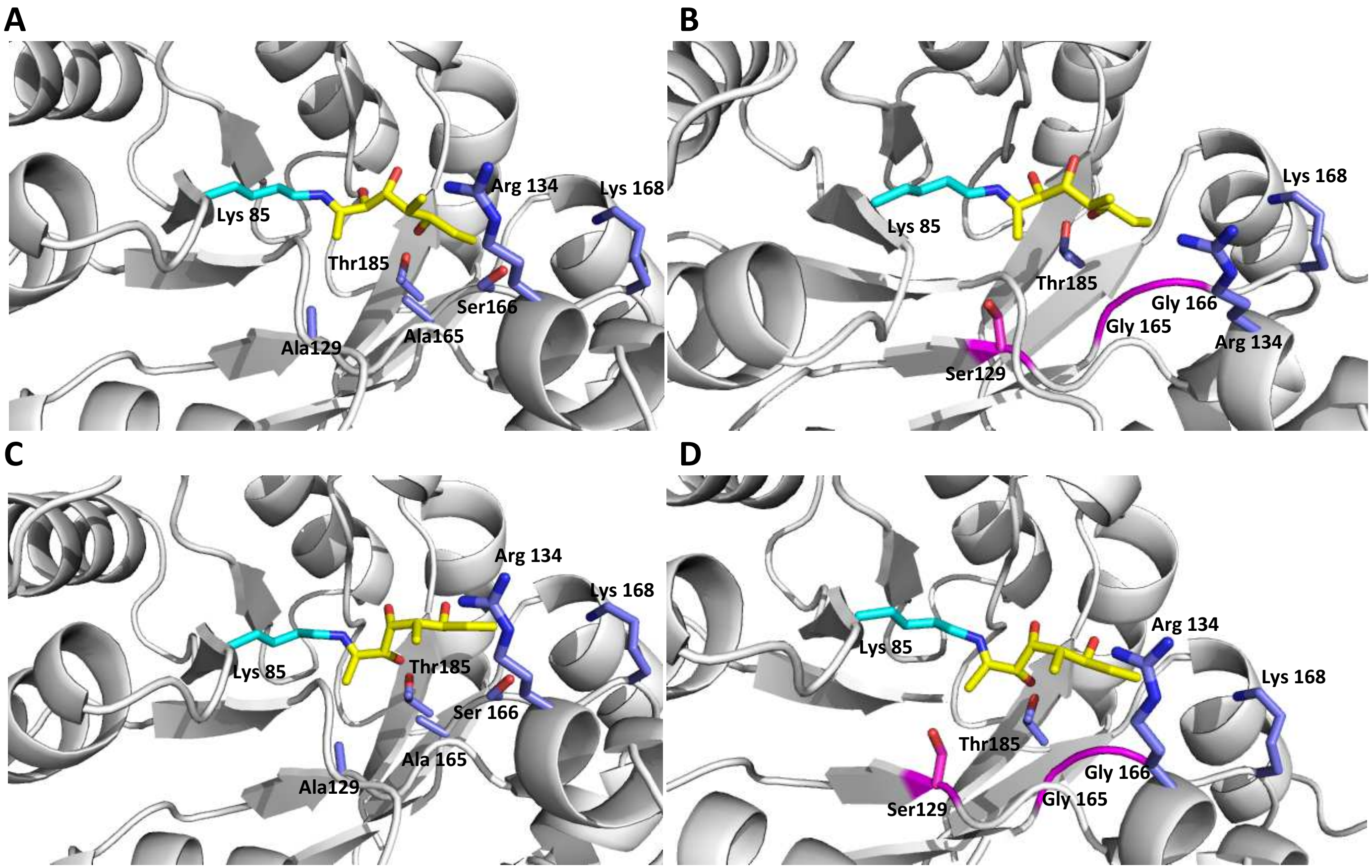


**Figure 3.4.4.** The model of the imine form of the adduct of HA and 3-hydroxy-2-methylpentanal in the active site of FSA. The residues of the acceptor site are shown in light blue. A) Imine intermediate of **60a** in the active site of FSA wt. B) Imine intermediate of **60a** in the active site of FSA A129S/A165G/S166G.The mutated residues are shown in magenta. C) Imine intermediate of **59a** in the active site of FSA wt. D) Imine intermediate of **59a** in the active site of FSA A129S/A165G/S166G. The mutated residues are shown in magenta.

Both **57** and **58** were tolerated by some FSA variants (Figure 3.4.5.). The screening of the enzymatic reactions showed that the additions with HA as donor substrate furnished higher conversions (up to 60%), and among the acceptors, the aldehyde synthesized by D-proline (**57**) was better substrate. Compared to the wild-type the conversions increased by 15% with FSA A129G/A165G/S166G. For the reactions with HA, the substitution of the A129 was not mandatory, since good results, i.e. 58% and 31% were achieved with FSA A165G and FSA S166G for **59a** and with **60a**, respectively. Using DHA as donor, the variants that bared the A129S mutation, i. e. FSA A129S/A165G and FSA A129S/A165G/S166G, were the only ones that furnished aldol adduct although in moderate yields (16% and 34%, respectively). As stated before, the A129S substitution is essential for the tolerance of DHA. However, modifications in the acceptor binding site were also necessary, since the single mutant FSA A129S hardly produced any adduct (6% **59b** and 2% **60b**). These results show that the most beneficial FSA mutations for the catalysis of this reaction were the ones that open up space for the binding of the acceptor substrate and for all tested reactions the FSA variant that appeared to fulfill this requirement was FSA A129S/A165G/S166G.

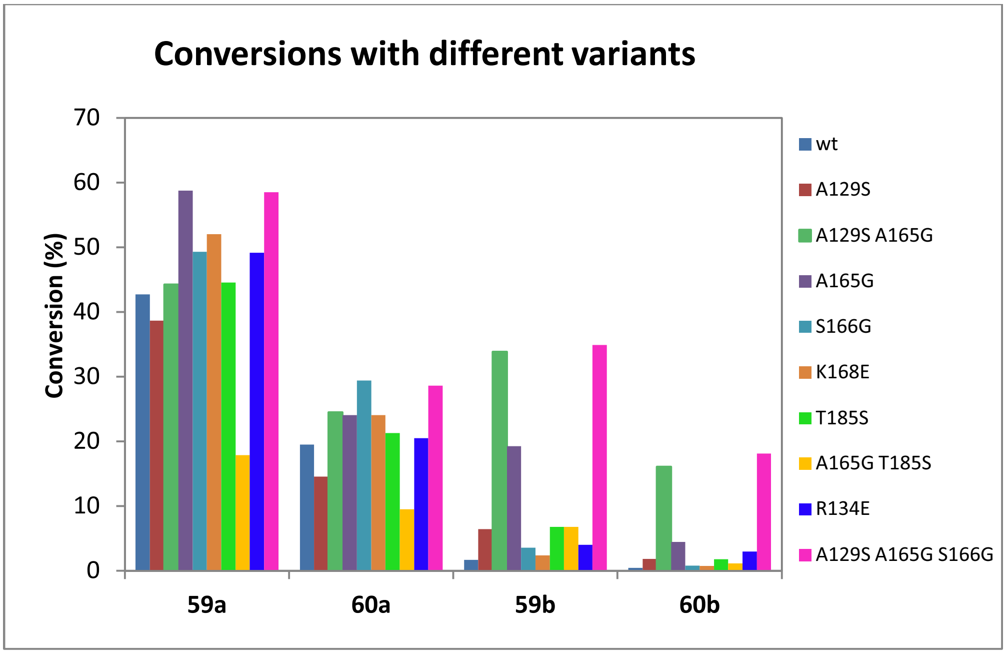
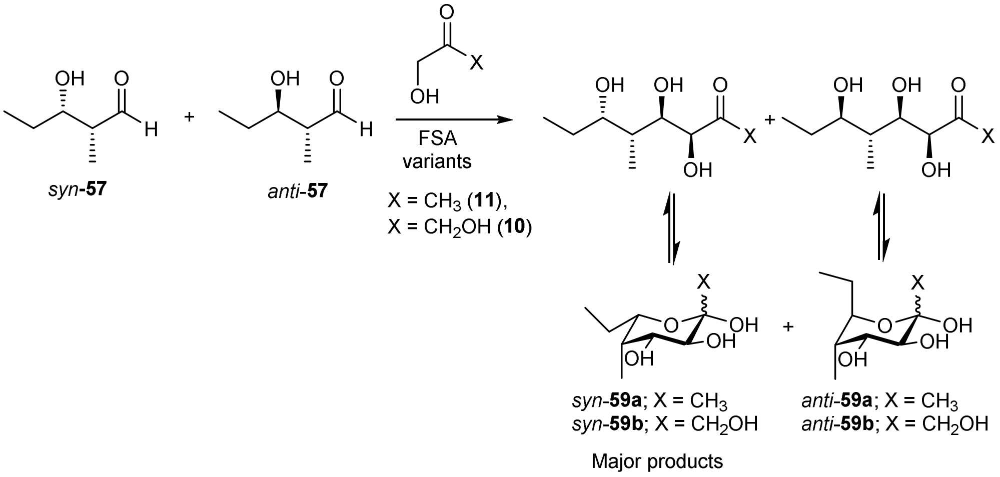


**Figure 3.4.5.** Testing the organocatalytically prepared aldehydes in the enzyme catalyzed reactions, with FSA wild-type and mutants, using HA and DHA as donor substrates.

To fully characterize the products and assess unequivocally the level of stereoselectivity of carboligations, the aldol addition reaction of HA and DHA to **57** and **58** were scaled up using FSA A129S/A165G/S166G as catalyst. Formation of side products were noticed in the HPLC chromatogram, which were not detected during the screening, probably due to unknown side reactions of the aldehydes. It was also hypothesized that the enzyme was able to catalyze the retro-aldol reaction of **57** and **58** generating propionaldehyde, which might furnish aldol adduct with DHA and HA because of FSA catalysis. Propionaldehyde was reported to be a good acceptor for FSA [123]. This experimental observation of the different aldolase behavior between the analytical and preparative scale reactions remains puzzling.

With the idea that the aldehydes were most probably the source of the side reactions, several different conditions were tried to minimize the side-reactions, which included changing the buffer and the concentrations, but none of them had any significant positive effect. Finally, the slow addition of the aldehyde using a syringe pump was applied, to keep the concentration of **57** and **58** as low as possible minimizing the formation of byproducts. Hence, the reactions were carried out in 30 h with the slow addition of the aldehydes in 24 h, with 3 mg $mL^{-1}$ enzyme added.

Addition of HA and DHA to **57** gave moderate conversions, and when purified, four products were isolated from the reactions of both donors (i. e., *syn*-**59a***, anti*-**59a**, and *syn*-**59b** and *anti*-**59b**,) (Scheme 3.4.4., Table 3.4.1.). These products are formed from the addition to the *anti*-**57** and *syn*-**57** stereoisomers of 3-hydroxy-2-pentanal. The characterization of the aldol adducts by NMR revealed, that *syn*-**59a** and syn-**59b** were the major products (Scheme 3.4.4., Table 3.4.1.), which is the result of the addition of HA or DHA to the *syn*-**57**. It was mentioned earlier that the aldol reaction catalyzed by proline provides a 3:1 *anti*:*syn* ratio. This indicates that the *syn*-**57** stereoisomer is better substrate for FSA than the corresponding major one, i.e. *anti*-**57**.



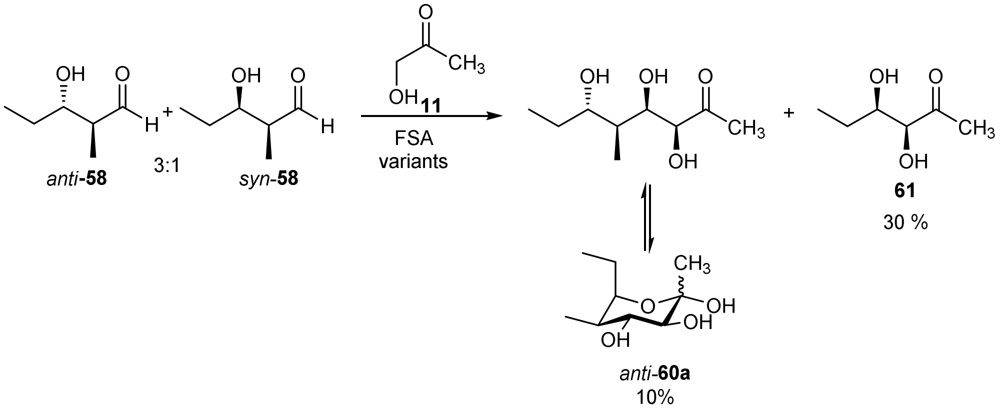
**Scheme 3.4.4.** Different products of the aldol additions of **57** and HA or DHA. *anti*-**5öa**: X=$CH_3$; *anti*-**59b**: X=$CH_2OH$;*syn*-**59a**: X=$CH_3$; *syn*-**59b**: X=$CH_2OH$.

**Table 3.4.1**. Results of the enzymatic aldol additions of **57** and HA and DHA.

| | **Conversion** | **Isolated yield** | **mg** |
|---|---|---|---|
| *anti*-**59a** | 54% | 23% | 39 |
| *syn*-**59a** | | | 72 |
| *anti*-**59b** | 34% | 11% | 26 |
| *syn*-**59b** | | | 40 |

Reactions were conducted in aqueous buffer with 150 mM donor concentration and 3 mg $mL^{-1}$ FSA A129S/A165G/S166G enzyme variant and the syringe pump mediated addition of the aldehyde to reach a final concentration of 100 mM and 18 v/v% DMF. Conversions were calculated by HPLC after application the derivatization protocol.

The aldol reaction of HA and DHA to **58**, furnished less than 10% of **60a** and ~30% of the adduct from HA and propionaldehyde (**61**) (Scheme 3.4.5). Isolation of the products gave 4 mg of *anti*-**60a** and 39 mg of the aldol adduct of HA with pentanal as ascertained by NMR. This is very surprising since in the starting mixture the concentration of propionaldehyde was low, according to the HPLC analysis. Hence, propionaldehyde can only be supplied through retro-aldol addition of **58**. Therefore, we sought for explanations how can the retro-aldol reaction be performed on **58**.



**Scheme 3.4.5.** Aldol reaction between **58** and HA furnishing a mixture of *anti*-**60a** and **61**.

Several hypotheses were made that lead to devise some additional experiments. First, we incubated aldehydes **57** and **58** with 3 mg mL$^{-1}$ of FSA wild-type, FSA S166G, FSA A129S/A165G/S166G. In a similar experiment, the FSA K85M, a variant with the modified catalytic lysine and therefore without aldolase activity was assayed as well. These variants were tried in order to show if there was any non-specific protein catalysis of the retro-aldol reaction of **57** and **58** and to ascertain if the enzyme can perform the retro-aldol reaction. No change was observed in the concentration of the aldehyde in any case, and neither any product was formed. Hence, no retro-aldol activity appeared to occur with any of the enzymes, and even the non-specific protein catalysis could be ruled out. Thermodynamic limitations on the retro-aldol reaction were discarded because aldehydes **57** and **58** cannot form cyclic hemiketal products. The same assay was tried in the aldol reaction direction, i.e. homo aldol addition of propionaldehyde, to investigate if 3-hydroxy-2-methylpentanal is formed using the above mentioned enzymes, but no reaction took place in this case either.

Another possibility was that a small amount of proline remained dissolved in the DMF, which, under the conditions of the biocatalytic reaction, could certainly catalyze the retro-aldol reaction. Furthermore, since HA is well tolerated as donor substrate by proline catalysis, this could favor the aldol reaction of HA to propanal. Nevertheless, these are hypotheses that must be demonstrated experimentally. In this sense, current experiments in our lab are conducted to clarify the selectivity of the FSA towards the aldehydes **57** and **58**, which include some control experiments such as the ones described above.

It is also noteworthy that among the **57** and **58** pentanal derivatives, only **57** was substrate of FSA. This result was at first glance surprising considering the stability of the final hemiketals formed, regarding their relative axial-equatorial substitutions. Indeed the predicted *anti*-**59** have two substituents in axial position, and is formally less stable than the adducts from the major *anti*-**58** aldehyde with all-equatorial substitutions (*anti*-**60**). However, comparing the structure of aldehydes **57** and **58** with L- and D-lactaldehyde may shed some light about their reactivity. Aldehyde **57** shares the same orientation of the α-methyl group with L-lactaldehyde, which is known as a good substrate for FSA. On the other hand, D-lactaldehyde, a known inhibitor of FSA, and **58** have the α-methyl group with the same orientation. These similarities could explain the difference in the aldehydes' reactivity (Figure 3.4.6.).

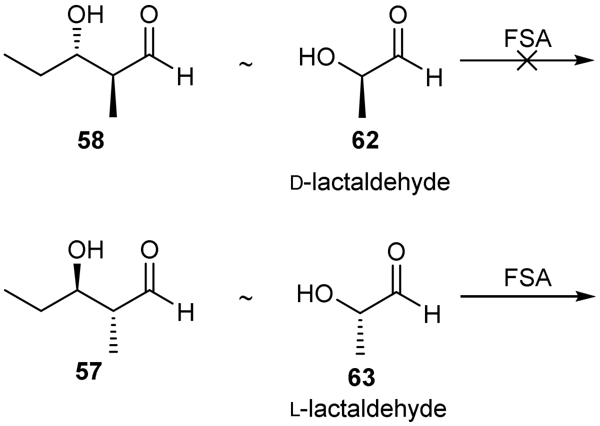


**Figure 3.4.6.** The possible reasons for the low reactivity of FSA with **58**. The similarity between the organocatalytically prepared aldehydes and lactaldehyde may be the clue for the lower tolerance for **58**, since the D-lactaldehyde which has the same configuration in the C-α position does not furnish product in the FSA catalyzed aldol reactions, while L-lactaldehyde (similar configuration to **57**) is a known substrate of the enzyme.

## 3.4.2. Conclusions

We have developed a method for the coupling of organo- and biocatalytic aldol reactions in a cascade fashion. Even though both reactions are widespread for the stereoselective formation of carbon-carbon bonds, their combination have not been extensively studied. This is likely due to that they are considered complementary catalytic methods, e.g. organocatalysis operates in organic solvents and with high catalyst loadings, which are generally unfavorable conditions for aldolases, while biocatalysis can be employed in water and with unprotected substrates. Only a few examples exist for organocatalytic reactions under these conditions [221-223].

As a proof of concept, we showed that proline catalyzed self-aldol reaction of propionaldehyde can be coupled with FSA catalyzed addition of HA and DHA on the 3-hydroxy-2-methylpentanal, proving once again the versatility of this aldolase to tolerate the highly non-polar aldehyde **57** and catalyze its additions with high stereoselectivity.

We have tested several FSA variants with substitutions in the acceptor binding site, and have registered in most cases slightly increased reactivity in the reactions with HA as donor substrate. The best variant was the triple mutant FSA A129S/A165G/S166G designed in our group. However with DHA no reaction took place with the FSA variants without the A129S substitution, consequently only two mutants provided product formation FSA A129S/A165G and the triple mutant FSA A129S/A165G/S166G, and again in this case also the triple mutant was the most effective catalyst. Hence this FSA variant was used in the preparative scale reactions.

Unfortunately the product with L-proline (**58**) was tolerated very poorly by the enzyme variants, especially under the scale-up conditions, where principally product with propanal was observed. This difference between the outcome of the smaller and larger scale reactions is puzzling. But the difference of reactivity between the two organocatalytic products can be explained by their similarity to the lactaldehydes, from which D-lactaldehyde is a known inhibitor of FSA while L-lactaldehyde is accepted well.

# 4. Conclusions

In this thesis mutagenic studies were carried out on the D-fructose-6-phosphate aldolase for the synthesis of different carbohydrates in the positions indicated on Figure 4.1.

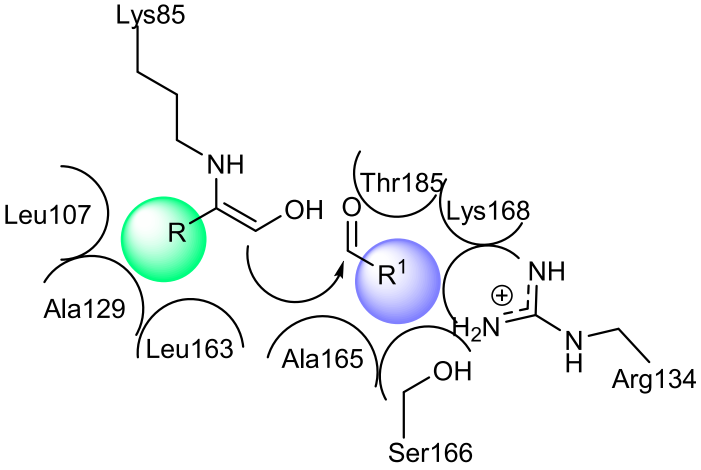


**Figure 4.1.** Schematic representation of the active site of FSA indicating the residues which were modified in this study.

The following general conclusions correspond to the four objectives previously described in **Chapter 2**.

I. Combination of a structure-guided program of saturation, site directed mutagenesis has led us to find a set of FSA mutants highly active for the cross-aldol additions of glycolaldehyde.

   - The library FSA A129X furnished variants with high glycolaldehyde selectivity, and combined with other mutations in the active site improved the acceptor tolerance. Thus, a toolbox of new FSA mutants was build up, expanding the biocatalytic possibilities for the synthesis of aldoses.

   - Donor competence experiments with equimolar mixtures of glycolaldehyde and dihydroxyacetone and glycolaldehyde and hydroxyacetone revealed that the designed FSA mutants are highly selective for glycolaldehyde as donor, since FSA L107Y, FSA L107Y/A129G/A165G, FSA L107Y/A129G and FSA

L107H/A129G/A165G catalyzed exclusively the homoaldol addition of GO furnishing D-threose.

- The reaction of glycolaldehyde with selected *N*-Cbz-aminoaldehydes catalyzed by the developed FSA mutants, furnished between 80-98% aldol adducts. However, the stereochemical outcome of the reactions depended, on both the substrates and the FSA mutant used.

II. A procedure was developed for the preparation of a number of aldose carbohydrate derivatives from simple formaldehyde, glycolaldehyde and other simple aldehydes or glycolaldehyde derivatives by biocatalytic one-pot two step tandem reactions using engineered FSA mimicking prebiotic processes. The developed method was the first direct aldohexose preparation with the biocatalytic $C_X$+$C_2$+$C_2$ one pot carbohydrate synthesis.

- The process might mimic the ancient prebiotic process, but using highly stereoselective evolved biocatalyst with complete control on the stereochemical outcome of the reactions. FSA mutants with minimal modifications were found to catalyze the double aldol reactions with the selected aldehydes.

- We demonstrated that the stereochemical outcome of FSA catalyzed reactions may be also altered. Engineering the enzyme led us to a variant that preferably furnishes the L-erythro configuration in FSA catalyzed reaction. By coupling the mutants of FSA we achieved the preparation of L-glucose by successive *anti*- and *syn*- aldol additions by tandem aldol additions of coupling FSA mutants. Using mutants developed in this work the L-glucose derivatives such as 6-metoxy-L-glucose and 6-chloro-L-glucose were also obtained in one pot using a single FSA mutant, or by the combination of FSA variants.

- Several rare aldose sugars and derivatives (D-idose or L-glucose derivatives) were synthesized with full control on the stereochemical outcome of the reactions, by one of the most straightforward procedures in carbohydrate synthesis.

III. We devised an unprecedented, wide expansion of the nucleophilic substrate tolerance of FSA by structure-guided protein engineering.

- Several active site substitutions were tested which catalyze the addition of a wide variety of 1-hydroxy-2-alkanones and related ether components as novel donor substrates. The double-active-site FSA L107A/L163A mutant was found to convert linear and branched structures of up to seven carbon atoms.

- Eleven deoxysugar products were synthesized, retaining consistently high D-*threo* diastereoselectivity, and high yield.

- These FSA variants open new avenues towards the synthesis of novel product families that were up to now inaccessible by biocatalysis.

IV. As proof of concept organo- and biocatalytic aldol additions were coupled in a cascade fashion. The products obtained by the proline catalyzed reactions were tested as acceptors in cross-aldol reactions with FSA wt. and FSA variants, with dihydroxyacetone, hydroxyacetone.

- Several FSA variants were tested with substitutions in the acceptor binding site, and the triple mutant FSA A129S/A165G/S166G was the most effective catalyst for the reactions of both 3-hydroxy-2-methylpentanals and both donors.

- No aldol reactions took place with the substrates obtained by L-proline catalysis, due to the configuration of the α-substituent, but two products were synthesized, purified and characterized from the aldehyde furnished by D-proline.

# 5. Experimental Section

## 5.1. General equipment

MAGNETIC STIRRER: MR 2002, Heidolph
VIBRAX SHAKER: KS130 basic, IKA
ULTRASOUND BATH: Selecta
pH METER: GLP21, Crison
ANALYTICAL HPLC: LaChrom Elite with L-2130 pump and L-2200 auto sampler and UV detector L-2400, VWR-Hitachi
SEMIPREPARATIVE HPLC: with UV-Vis L4000 detector, LC4000 injector and CF4000 flow controller, Waters
NMR spectrometer: Varian 500 MHz from the IQAC-CSIC Spectroscopy service
NMR spectrometer: Varian 400 MHz from the IQAC-CSIC Spectroscopy service
NMR spectrometer: Brucker 500 MHz from the UAB Spectroscopy service
NANODROP: 2000C UV-Vis spectrophotometer, Thermo
SPECTROPHOTOMETER: Cary 500 Bio UV-Vis
LYOPHILIZER: Christ Alpha 2-4
POLARIMETER: 341, Perkin-Elmer
THERMOCYCLER: MJ Mini, Bio-Rad
THERMOCYCLER: personal, Eppendorf
INCUBATOR: Infors AG CH-4103, Bottmingen
INCUBATOR: Minishaker, VWR
ROTARY EVAPORATOR:
CENTRIFUGES: Eppendorf 5810R, Eppendorf 5427R

## 5.2. Chemical and other reactives

Phosphonoglucose isomerase from baker's yeast (*S. cerevisiae*), glucose-6-phosphate dehydrogenase from baker's yeast (*S. cerevisiae*), D,L-glyceraldehyde 3-phosphate (G3P) and nicotinamide adenine dinucleotide phosphate (NADP$^+$) were from Sigma-Aldrich. Antibiotics, acrylamide-bisacrylamide, buffer components and glycolaldehyde dimer were from Sigma-Aldrich. Culture media components for bacteria were from Pronadisa. Milli-Q grade water was used for analytical and preparative HPLC and for preparations of buffers and other assay solutions obtained from an Arium® Pro Ultrapure Water Purification System (SartoriusStedim Biotech). All the other solvents used were of analytical grade.

# 5.3. Biological material

## 5.3.1. Microorganisms

The bacterial strains and their most relevant characteristics are shown in Table 5.1

Two types of bacterial strains were used in this work: for the molecular biology experiments *E. coli* Nova Blue and XL-1 Blue cells were used, and for protein expression *E. coli* M15 and Nova Blue were used.

**Table 5.1.** Bacterial strains and their most relevant characteristics used in this study

| **Strains** | **Relevant genotype** | **Ref./Origin** |
|---|---|---|
| *E. coli* M15 [pREP4] | *na*l$^{s}$, *str*$^{s}$, *rif*$^{s}$, *thi*$^{-}$, *lac*$^{-}$, *ara*$^{+}$, *gal*$^{+}$, *mt*l$^{-}$, F$^{-}$, *rec*A$^{+}$, *Uvr*$^{+}$, *Lon*$^{+}$ Km$^{R}$, lac F'[proA$^{+}$B$^{+}$lacl$^{a}$ZΔM15:Tn10] (Amp$^{R}$), (Kan$^{R}$). | Qiagen |
| *E. coli* XL-1 Blue | endA1, gyrA96(nal$^{R}$) thi-1 recA1 relA1, supE44lac F'[ ::Tn10 proAB$^{+}$ lacI$^{q}$ Δ(lacZ)M15] thi hsdR17(r$_{K}^{-}$ m$_{K}^{+}$) (Tet$^{R}$) | Stratagene |
| *E. coli* Nova Blue | *end*A1, *hsd*R17(rB$^{+}$, mB$^{+}$), *sup*E44, *thi-1*, *rec*A1, *gyr*A96, *rel*A1 *lac* F´ [ *pro*A$^{+}$B$^{+}$, *lac*I$^{q}$ ZΔM15::Tn10] (*Tet*$^{R}$) | Novagen |

## 5.3.2. Vectors

The plasmid pQE*fsa* containing the gene for expression of FSA was constructed in our lab using routine procedures of molecular biology.

The plasmid vectors used in this study and their construction and most important characteristics are listed in Table 5.2.

**Table 5.2.** Plasmid vectors used in this study and their construction.

| Plasmid | Relevant characteristics | Reference/Origin |
|---|---|---|
| pQE40 | $P_{T5}$, $Amp^r$, $ColE1_{ori}$ | Quiagen |
| pQE*fsa* | *fsa* gene (660 bp) cloned in pQE40 (*BseRI/HindIII*), the C-terminal His-tag was deleted | [127] |
| pQE*fsaL107A* | *fsaL107A* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL107A/A129G* | *fsaL107A/A129G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL107A/L163A* | *fsaL107A/L163A* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL107A/A129G/L163A* | *fsaL107A/A129G/L163A*gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL107A/A129S/L163A* | *fsaL107A/A129S/L163A* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL107F/A129G* | *fsaL107F/A129G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL107H/A129G* | *fsaL107H/A129G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL107H/A129G/A165G* | *fsaL107H/A129G/A165G* gene (660bp) cloned in pQE40 | This study |
| pQE*fsaL107Y* | *fsaL107Y* gene (660bp) cloned in pQE40 | This study |
| pQE*fsaL107Y/A129G* | *fsaL107Y/A129G* gene (660bp) cloned in pQE40 | This study |
| pQE*fsaL107Y/A129G/A165G* | *fsaL107Y/A129G/A165G* gene (660bp) cloned in pQE40 | This study |
| pQE*fsaA129G* | *fsaA129G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA129G/L163A* | *fsaA129G/L163A*gene (660bp) cloned in pQE40 | This study |
| pQE*fsaA129G/A165G* | *fsaA129G/A165G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA129G/S166G* | *fsaA129G/S166G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA129S* | *fsaA129S* gene (660 bp) cloned in pQE40 | [121] |
| pQE*fsaA129S/L163A* | *fsaA129S/L163A* gene (660bp) cloned in pQE40 | This study |

| | | |
|---|---|---|
| pQE*fsaA129S/A165G* | *fsaA129S/A165G* gene (660 bp) cloned in pQE40 | [132] |
| pQE*fsaA129S/A165G/S166G* | *fsaA129S/A165G/S166G* gene (660 bp) cloned in pQE40 | Thesis Mariana Gutierrez 2011 |
| pQE*fsaA129T* | *fsaA129T* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA129T/A165G* | *fsaA129T/A165G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA129T/S166G* | *fsaA129T/S166G* gene (660bp) cloned in pQE40 | This study |
| pQE*fsaA129T/A165G/S166G* | *fsaA129T/A165G/S166G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA129V* | *fsaA129V* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaR134E* | *fsaR134E* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaL163A* | *fsaL163A* gene (660bp) cloned in pQE40 | This study |
| pQE*fsaA164G/A165G/S166P* | *fsaA164G/A165G/S166P* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA165G* | *fsaA165G* gene (660 bp) cloned in pQE40 | [132] |
| pQE*fsaA165G/S166G* | *fsaA165G/S166G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA165G/S166N* | *fsaA165G/S166N* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA165G/S166D* | *fsaA165G/S166D* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA165G/S166P* | *fsaA165G/S166P* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaA165G/T185S* | *fsaA165G/T185S* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaS166G* | *fsaS166G* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaK168E* | *fsaK168E* gene (660 bp) cloned in pQE40 | This study |
| pQE*fsaT185S* | *fsaT185S* gene (660 bp) cloned in pQE40 | This study |

## 5.3.3. Oligonucleotides

The oligonucleotides were specifically designed for this study. They are listed in Table 5.3., with the mutated codons highlighted in bold and italics, and silent mutations shown in italics. Some of the codons were designed following the recommendations of the QuickChange® Site-Directed Mutagenesis protocol from Stratagene, and some others were designed with a partially overlapping primer design [224].

All the oligonucleotides were produced by the commercial provider Eurofins MWG-Operon with the purity of >70% and they were purified with the system HPSF (High Purity Salt Free).

**Table 5.3.** Sequences of oligonucleotides used in this study in the site directed and site saturation mutagenesis.

| Oligonucleotides | Oligonucleotide sequences (5´→ 3´) | Ref./Origin |
|---|---|---|
| FSAA129X up | TATGTT***NNK***CCTTACGTTAATCGTATTGATGC | This study |
| FSAA129X down | ACGTAAGG***MNN***AACATATTCCGCACCTGC | This study |
| FSAA165G up | GAAAGTGCTGGCA***GGG***AGTTTCAAAACC | [132] |
| FSAA165G down | GGTTTTGAAACT***CCC***TGCCAGCACTTTC | [132] |
| FSAL107F up | CGACG***TTT***GGAACCGCGGTATATGGAGCAGC | This study |
| FSAL107F down | GGTTCC***AAA***C GTCGGTATCC CTTCCCTTCC | This study |
| FSAL107H up | CGACG***CAT***GGAACCGCAGTATATGGC | This study |
| FSAL107H down | GCGGTTCC***ATG***CGTCGGAATCCCTTCC | This study |
| FSAL107Y up | CGACG***TAT***GGAACCGCAGTATATGGC | This study |
| FSAL107Y down | GCGGTTCC***ATA***CGTCGGAATCCCTTCC | This study |
| FSAA165G/S166X up | GCAGGA***NNS***TTCAAAACCCCGCGTCAGGC | This study |
| FSAA165G/S166X down | GTTTTGAA***SNN***TCCTGCCAGCACTTTC | This study |
| FSAS166G up | GGCAGCG***GGT***TTCAAAAC*T*CC*A*CGTCAGGCGC | This study |
| FSAS166G down | GAA***ACC***CGCTGCCAGCACTTTCGCCTGCG | This study |
| FSAA165G/S166G up | GAAAGTGCTGGCA***GGGGGT***TTCAAAACC | Thesis Mariana Gutierrez 2011 |

| | | |
|---|---|---|
| FSAA165G/S166G down | GGTTTTGAA***ACCCCC***TGCCAGCACTTTC | Thesis Mariana Gutierrez 2011 |
| FSAA164G/A165G/S166P up | GCTG***GGAGGACCG***TTCAAAACCCCGCGTC | This study |
| FSAA164G/A165G/S166P down | GAA***CGGTCCTCC***CAGCACTTTCGCCTGCGGC | This study |
| FSAK168E up | GAGTTTC***GAA***ACCCCGCGTCAGGCACTGGAC | This study |
| FSAK168E down | GGT***TTC***GAAACTCGCTGCCAGCACTTTC | This study |
| FSAL107A up | CGACG***GCA***GGAAC*A*GCGGTATATGGCG | This study |
| FSAL107A down | ACCGCTGTTCCTG***CCG***TCGGAATCCCTTCC | This study |
| FSAL163A up | GAAAGT***GGC***GGCAGCGAGTTTCAAAAC | This study |
| FSAL163A down | GAAACTCGCTGCC***GCC***ACTTTCGCCTGC | This study |
| FSAR134E up | CGTTAAT***GAG***ATTGATGCTCAGGGCGGTAGCGG | This study |
| FSAR134E down | GCATCAAT***CTC***ATTAACGTAAGGCGCAACATATTCCG | This study |
| FSAT185S up | GTGAATCAATT***TCT***CTGCCACTGGATGTG | This study |
| FSAT185S down | GCAG***AGA***AATTGATTCACATCCTGCCAGTAAGC | This study |

## 5.4. Culture media and solutions

All the culture media and solutions used in this work were prepared with ultrapure distilled water and were sterilized by autoclaving during 30 minutes at 121 ºC. The thermolabile (for example antibiotics), and concentrated microelement solutions were sterilized by filtration with 0.22 μm filters.

### LB medium (Luria-Bertani)

LB medium is a rich medium that is used extensively in recombinant DNA work to culture *E. coli*.

Its composition:

| | |
|---|---|
| Triptone | 10 g/L |
| Yeast extract | 5 g/L |
| Nacl | 10 g/L |

To use this medium in solid form (in plates), 1.5% agar was added to it after sterilization.

In most cases LB medium was complemented with antibiotics necessary for maintaining the plasmid selection, which is contained in the interior of the cells. Antibiotics composition of the medium was:

LB-Amp: 100 μg/mL ampicillin concentration was added to LB medium

LB-Kan: 25 μg/mL kanamycin concentration was added to LB medium

## SOB medium

This salt rich medium is used to obtain competent cells.

Composition:

| | |
|---|---|
| Triptone | 2% |
| Yeast extract | 0.5% |
| Nacl | 10 mM |
| KCl | 2.5 mM |
| $MgCl_2$ | 10 mM |
| $MgSO_4$ | 10 mM |

## SOC medium

This medium has the same composition as SOB, but glucose is also added as carbon source. It is used for growing the recently transformed bacteria, during the incubation

SOC Medium is a rich media used primarily in the recovery step of *Escherichia coli* competent cell transformations. The use of SOC maximizes the transformation efficiency of competent cells, if used for growing the recently transformed bacteria during the incubation, before plating the culture.

Composition:
SOB media + 20 mM glucose

**TFB solution**

For the preparation of competent cells this solution is used.

Composition:
30 mM potassium acetate
10 mM $CaCl_2$
50 mM $MnCl_2$
100 mM RbCl
15% glycerol

pH was adjusted to 5.8 with 1M acetic acid and the solution was filter-sterilized.

## 5.5. Molecular Biology techniques

### 5.5.1. Isolation and purification of plasmid DNA (Miniprep)

A plasmid preparation is a method used to extract and purify plasmid DNA on a small scale from E. coli cultures. It is based on alkaline lysis according to the method of Bimboim and Doly [225]. All methods are based on three steps: growth of the bacterial culture, harvesting and lysis of the bacteria and purification of the plasmid DNA. In this study most often a commercial kit from Roche – Applied Sciences was used, following the protocol provided by the manufacturer.

The kit relies on alkaline lysis to release plasmid DNA from bacteria. RNase removes all RNA in the lysate. After cellular debris and (entrapped) genomic DNA are removed by centrifugation, the remaining supernatant is mixed with a chatotropic salt and applied to the glass fiber fleece in a High Pure Spin Filter Tube. Under the buffer conditions used in the procedure, the plasmid binds to the glass fiber fleece, while contaminating substances (salts, proteins, and other cellular contaminants) do not. Brief wash-and-spin steps readily remove these contaminants. Once purified, the plasmid can be easily eluted in a small volume of water.

The procedure (Figure 5.1.) is the following: 10 mL overnight culture ($OD_{600}$=1,5) was centrifuged for 2 minutes at 6000 x g, the supernatant was discarded and the bacterial pellet was resuspended with 250 μL Suspension buffer (with RNase) and the mixture was incubated for 5 minutes with 250 μL Lysis buffer. And it was mixed with 350 μL Binding buffer, and incubated on ice for 5 minutes, centrifuged for 10 minutes at 13000 x g. The supernatant was filtered with the high pure filter tube and washed twice with wash buffers I and II. And finally eluted from the filter with water and stored at -20 ºC.

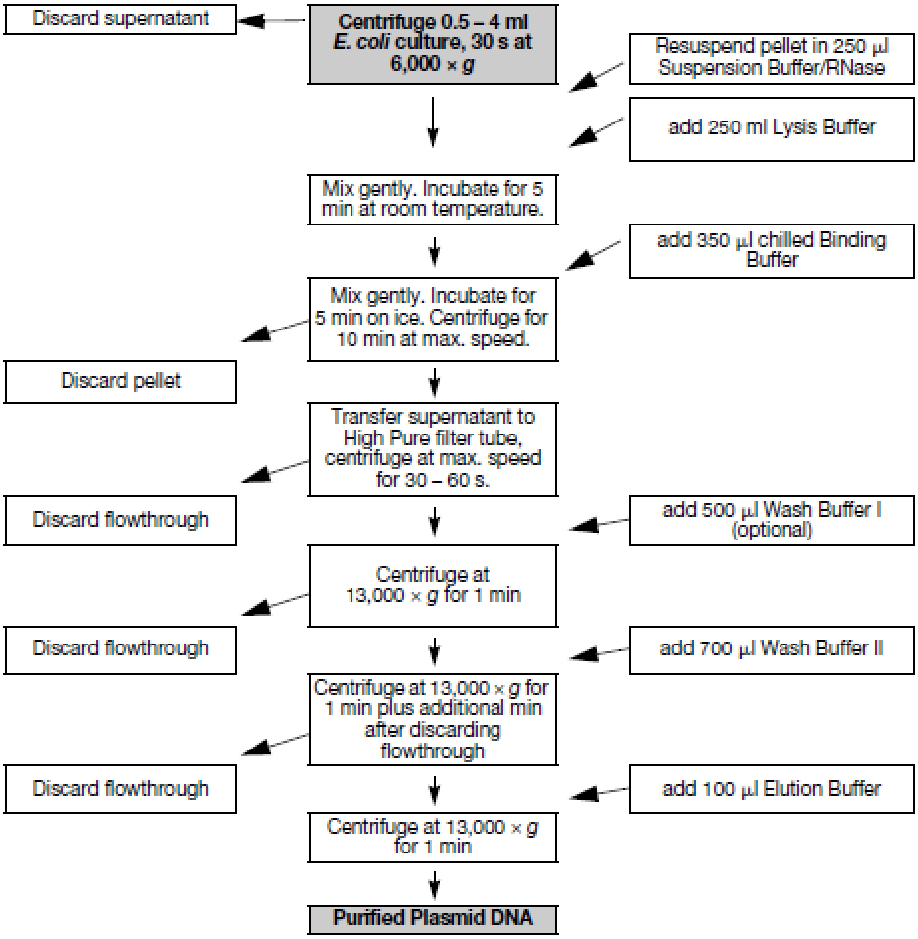


**Figure 5.1.** Experimental overview of the plasmid isolation protocol (Roche – Applied Sciences).

## 5.5.2. Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was used for the separation and visualization of the plasmids or PCR products. The electrophoresis takes place on a horizontally submerged agarose gel (analytical grade, Roche – Applied sciences) with the adequate concentration for the DNA size. The gels were run using the Wide Mini-Sub Cell GT System from Bio-Rad. The gel was prepared with dissolving agarose (usually 1%) in TAE 1X buffer (Tris-base 40 mM, acetic acid 20 mM, EDTA 1 mM, pH set to 8.1), and 7 µL SYBR-Safe DNA stain (Invitrogen) was added before casting. The samples were prepared with 1:5 volume equivalent of DNA loading dye 6X

(Fermentas), and were loaded into the gel. To facilitate the identification of DNA by the size a DNA ladder (GeneRuler DNA ladder mix, Fermentas) was also applied. For the separation a constant 5-10 V/cm Voltage is applied, and the gel is run until the front reaches until approx. 50% of the gel. When the electrophoresis is finished, the DNA is visualized by fluorescence under ultraviolet light (302 nm) in a transilluminator (TFX-20.M, Ecogen).

### 5.5.3. Quantification of nucleic acids

To determine the concentration of DNA from a solution with a spectrophotometric method, first described by Sambrook *et al.* [226]. The measurements were realized with a NanoDrop ND 2000, Thermo. The absorbancy values are converted to concentration, with respect to a standard value of $A_{260}$=1 what is generally accepted as 50 μg/mL double-stranded DNA. With each measurement the absorbency at 260 nm is compared to other absorbency at 280 nm and 230 nm to establish the purity of each sample. The ratio $A_{260}/A_{280}$ if lower than 1,8 shows contamination of protein in the sample, and the ratio $A_{260}/A_{230}$ if higher than 2 indicates RNA contamination of the sample.

### 5.5.4. Plasmid transformation: Transformation in *E. coli* with chemical methods

To transform *E. coli* cells, RbCl treatment was used (modification of Hanahan's method [227]), with this method high transformation efficiency can be obtained (up to 5 X $10^8$ transformants for each μg of used DNA).

#### a. *Preparation of competent cells*

To obtain the competent cells a saturated overnight culture ($OD_{600}$=1,5) of the chosen strain was inoculated into a mixture of 200 mL SOB medium with 12 mL of 5 M NaCl. It was incubated at 30 °C with shaking at 200 rpm, to growth up to 0,5 $OD_{600}$ (around 3-4 h). Then the growth was stopped with placing the culture on ice, and centrifugation for 10 minutes at 12000 x g. After discarding the supernatant, the bacterial pellet was resuspended in 20 mL TFB solution and it was incubated on ice for 15 minutes. Then 700 μL DMSO was added to the culture and left on ice for 10 minutes. Then with gentle shaking further 700 μL DMSO was added to the culture and it was incubated on ice for 5 minutes. Finally the suspension was divided to 500 μL aliquots in pre-cooled plastic tubes, and frozen with dry ice/ethanol bath and stored until use at -80 °C.

### *b. Transformation with heat shock*

For plasmid transformation 10-50 ng DNA was added to 100 μL competent cells, mixing it gently with circular motion then incubated on ice for 5 minutes. Then the heat shock was applied either for 5 minutes at 37 °C or for 30 seconds at 42 °C followed by incubation of 2 minutes on ice again. Finally 800 μL SOC was added and incubated at 37 °C, 250 rpm for 30 to 60 minutes, and plated on LB agar plate, and finally incubated overnight at 37 °C.

## 5.5.5. Site directed and saturation mutagenesis with specific oligonucleotides

To obtain point mutations on the fsa gene the Stratagene, QuickChange® Site directed mutagenesis system was used. The basic procedure utilizes a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest and two synthetic oligonucleotide primers containing the desired mutation (Figure 5.2.). The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by *PfuTurbo* DNA polymerase. Incorporation of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with *Dpn* I. The *Dpn* I endonuclease is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. The nicked vector DNA containing the desired mutations is then transformed into *E. coli* competent cells. The small amount of starting DNA template required to perform this method, the high fidelity of the *PfuTurbo* DNA polymerase, and the low number of thermal cycles all contribute to the high mutation efficiency and decreased potential for generating random mutations during the reaction.

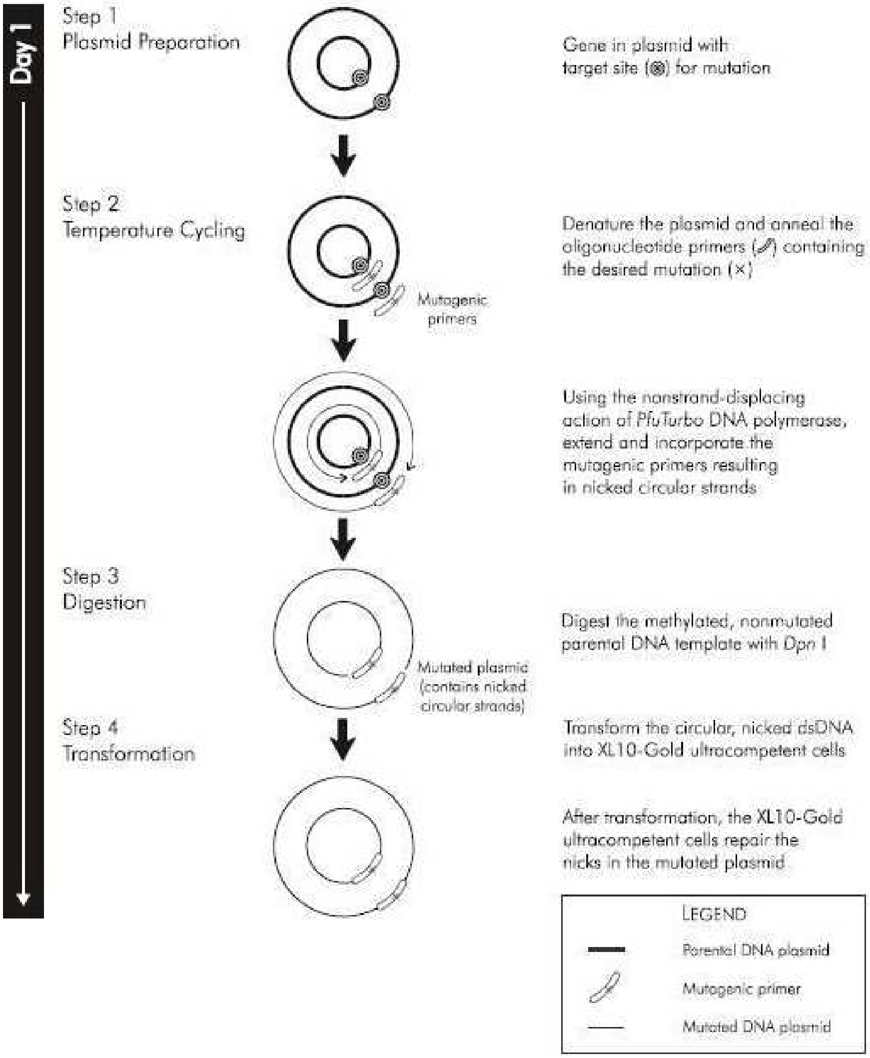


**Figure 5.2.** Overview of the QuickChange® site directed mutagenesis method (Stratagene).

### *a. Mutagenic primer design*

One of the most influential factors in the outcome of the mutagenesis is the design of the oligonucleotides. The design of the mutagenic primers were performed taking into account the Primer Design Guidelines of Stratagene and the recommendations from the publication of Zheng et al. [224]. The primers designed for this study are listed in Table 5.3. In this study mostly partially overlapping primer design was used. The most important considerations for the design are the following: (i) The length of the primer should be between 20-45 base pairs; (ii) $T_m \geq 78$ °C; (iii) Minimal GC content is 40%; (iv) At least one G or C should be placed at each end of the terminus; (v) At least eight non-overlapping bases should be introduced at the 3'-terminus of primer; (vi) the targeted mutation(s) should be included into both primers. The designed oligonucleotides were analyzed with the freeware Gene Runner (Heliogenetics, http://www.generunner.net/) to examine the most important parameters ($T_m$, %GC, structural artifacts).

### *b. Mutant strand synthesis reaction*

To obtain a vector with the desired mutation a couple of parallel reactions were ran simultaneously, with different DNA template concentration, in the range between 5-50 ng of DNA for each 50 μL of reaction mixture while always keeping the concentration of the oligonucleotides in excess. The general constitution of the reaction mixture and the reaction conditions (including the reaction temperatures used in each cycle) are shown in Table 5.3 and Table 5.4, respectively.

**Table 5.4.** General constitution of the reaction mixture used for the site directed mutagenesis, based on the composition of the commercial kit QuickChange® Site directed mutagenesis (Stratagene)

| | |
|---|---|
| Pfu DNA polymerase Buffer (10 X) | 5 μL |
| dNTP mix 25 mM | 1 μL |
| primer 5'3' | 125 ng |
| primer 3'5' | 125 ng |
| DNA template | 5-50 ng |
| Sterile ultrapure water | to 50 μL |
| Pfu DNA polymerase | 2,5 U |

**Table 5.5.** General sequence and conditions of the temperature cycles for the site directed mutagenesis, based on the composition of the commercial kit QuickChange® Site directed mutagenesis (Stratagene)

| Segment | | Cycles | Temperature | Time |
|---|---|---|---|---|
| Initial denaturalization | | 1 | 96 ºC | 1 minute |
| Amplification | Denaturalization | 16[1] | 96 ºC | 30 seconds |
| | Annealing | | 55 ºC | 1 minute |
| | Polymerization | | 68 ºC | 1 minute[2] |
| Final extension | | 1 | 68 ºC | 30 minutes |

[1] The number of cycles depends on the type of the desired mutation (point mutation/deletition/insertions).

[2] The length of the extension depends on the size of the fragment to amplify.

To verify the amplification agarose gel electrophoresis was used.

### c. *DpnI digestion of the amplification product*

To eliminate the parental plasmids (non-mutated plasmid) 1 μL *Dpn*I restriction enzyme (10 U/μL, Fermentas) was added to the reaction mixture, and was incubated at 37 ºC for 4 hours.

### d. *Transformation of the competent cells*

*E. coli* NovaBlue competent cells were transformed with 1 μL DNA solution treated with according to the method described in Chapter 5.5.4. In all cases at least 20 colonies/plate grew. From these 3-10 colonies were selected and their plasmid was extracted. To verify the presence of the mutations the DNA was analyzed by DNA sequensation by the DNA Sequensation Service of CRAG-CSIC.

## 5.6. Analysis of the DNA sequences

The analyses of the DNA sequences were performed using various computer programs. To visualize and analyze the obtained chromatograms the software Chromas (Technelysium Pty Ltd) was used. The sequences with high homology were identified using *Basic Local Alignment Search Tool* (BLAST) [228]. To translate the DNA sequence, and find the open reading frames pDRAW32 (AcaClone) was used. The protein sequence Blast searches were realized on line using the database and web page of the SIB Bioinformatics Resource Portal (Swiss Institute of Bioinformatics), from the Proteomic Service ExPASy (http://www.expasy.org/). The alignment of the protein sequences were carried out using the program ClustalW, offered online also by the ExPASy service.

## 5.6.1. Computational Methods

The structures of the models shown in this thesis were designed by Dr. Jordi Bujons. Protein complexes were modeled with the package Schrödinger Suite 2013, [229] through its graphical interface Maestro [230]. The program MacroModel [231] with its default force field OPLS 2005, [232] a modified version of the OPLS-AA force field, and GB/SA water solvation conditions [233] were used for all energy calculations. Coordinates of decameric *E. coli* fructose-6-phosphate aldolase [177] were obtained from the Protein Data Bank [234] at Brookhaven National Laboratory (PDB code 1L6W). Since each FSA monomer has a C-terminal helix (residues 200-215) which runs across the active site of the neighboring monomer, in order to simplify the system and reduce the computation time, all calculations were performed with a minimal functional unit that included one protein chain (A) with residues 1-199 and a second chain (B) that included residues 199-220. In addition, the C-terminal end of chain A and the N-terminal end of chain B were modified with a N-methyl amide and an acetyl capping groups, respectively. This minimal functional unit was mutated in silico to achieve the structures of the variants under study. Protein structures were prepared using the Protein Preparation Wizard [235, 236] included in Maestro to remove solvent molecules and ions, adding hydrogens, setting protonation states [237] and minimizing the energy using the OPLS 2005 force field. The structures of the different adducts bound into the active center of FSA enzymes were modeled starting from the coordinates of a glyceraldehyde hemiaminal molecule bound to the essential Lys-85 residue, which is already present in the FSA crystal structure. The structures were then minimized, first applying constraints to the protein (force constant = 100 kJ $Å^{-2}$ $mol^{-1}$) to avoid large changes on its structure, and afterwards allowing free movement of the whole system until reaching a gradient < 0.01 kcal $mol^{-1}$ $Å^{-1}$. In order to find low energy conformers of the different protein-bound hemiaminal adducts, a conformational search was carried out using the mixed MCMM/LMCS method[238] as implemented in the program MacroModel, [231] During this search, the dihedrals of the hemiaminal moiety were varied, while all residues of the protein with atoms within 6 Å were kept constrained with an harmonic potential (force constant = 100 kJ $Å^{-2}$ $mol^{-1}$), and the rest of residues were kept frozen. All the confomers detected by this search were then minimized allowing free movement of the whole system, and the resulting structures were ranked according to their potential energies. To further analyze the conformational space of the complexes, short stochastic dynamics runs were carried out on the lowest energy conformer determined from each complex. For that purpose, the MacroModel Dynamics application was used. The following conditions were applied: equilibration time = 50.0 ps, simulation time = 500 ps, time step = 2 fs, temperature = 350 K, implicit GB/SA water solvation, an harmonic potential (force constant = 200 kJ $Å^{-2}mol^{-1}$) set to constrain all residues which atoms were at>6 Å of the hemiaminal moiety, bond lengths to hydrogens constrained with SHAKE [239] . 100 conformations from each run were sampled and re-minimized without constraints and the resulting confomers where ranked according to their potential energy. All the structures shown in the paper correspond to the lowest energy minima determined using this protocol, with the exception of the active site structures of FSA L107A/A129G/A165G in complex with the glycolaldehyde enamine and aldehyde 3 in its si-/re-face pre-reactive orientations (Figure 4),for which the QM/MM program Q Site was used [240]. In that later case, the structures were built from the optimized hemiaminal moieties shown in Figure 3A,B by removing the bonds between C2-C3 and manually modifying the structures to build the corresponding aldehyde 3 and protein-bound enamine. These structures were optimized, first, by using Macro Model with its standard settings and imposing an arbitrary distance constraint between atoms C2 of the enamine and C1 of 3 (d = 2.4 ± 0.2 Å, force constant = 500 kJ $Å^{-2}mol^{-1}$) to avoid that the two moieties move too far apart. The resulting structures were then used as input for Q site. The QM/MM boundary was defined by placing hydrogen caps between atoms Cα-Cβ of residues N28 and Y131, and between atoms Cγ-Cδ of residue K85, and also including in the QM region the K85-bound enamine, the aldehyde 3 and the essential water molecule, totaling 65 QM atoms. This QM region was optimized at the DFT B3LYP/lacvp* level of theory, setting a distance constraint (d= 2.65 Å) between the two reactive carbon atoms, while the MM region was also optimized without constraints with the OPLS 2005 force field.

# 5.7. Microbiology techniques for the study of proteins

## 5.7.1. Protein Expression from *E. coli* cultures

For the expression of FSA *wt* and mutants the following conditions were used: The plasmids pQE*fsa* and variants (Table 5.2.) were transformed into *E. coli* M-15 [pREP-4] (QIAGEN) and plated on LB-agar-Amp/Kan and grown overnight at 37 ºC. From the grown colonies 3-5 were picked and expression tests were performed. From the colony which had the highest level of expression a preinocculum culture was grown in 100 mL LB-Amp/Kan at 37 ºC and 250 rpm during 12-18 hours until $OD_{600}$=3-4. 25 mL was transferred from the preinocculum to 2 L Erlenmeyers containing 1 L LB-Amp/Kan, and it was incubated at 37 ºC with 250 rpm agitation up to 0,6-0,8 $OD_{600}$. For protein expression, temperature was lowered to 30 ºC (to avoid the formation of inclusion bodies) and 1 mM final concentration of IPTG was added. After additional incubation of 12 h cells were harvested by centrifugation at 8000 rpm 4 ºC for 30 minutes. The discarding the supernatant the pellet was stored at -20 ºC until the extraction of the proteins.

## 5.7.2. Protein Expression from *E. coli* cultures for testing the library of mutants obtained by site saturation mutagenesis

The degenerated oligonucleotides contained NNK degeneracy (N, A/C/G/T; K, G/T), thus 32 codons and all the 20 proteinogenic amino acid residues were obtained. The library was transformed into Nova Blue competent cells (EMD Millipore). For 17 randomly picked clones the expression level in crude extract was checked by SDS-PAGE. Single colonies (96) were randomly picked and arrayed in 96-deepwell plates (volume 1 mL; Sigma-Aldrich). Each well contained YT broth (600 µL) with ampicillin (60 µg $mL^{-1}$). The plate was sealed with a gas-permeable adhesive film and incubated at 37 ºC with overnight shaking at 1000 rpm. These cultures were used to inoculate expression cultures (1:6) containing YT broth (700 µL) with ampicillin (70 µg $mL^{-1}$). Cells were grown up to an optical density of 0.8 at 600 nm. For protein expression, isopropyl-*ß*-D-thiogalactoside (IPTG) (1 mM final concentration) was added. These expression cultures were grown overnight at 37 ºC shaking at 1000 rpm. The cells were collected by centrifugation at 1000 × g for 30 min and the supernatants were discarded. Cell pellets were resuspended in triethanolamine buffer (100 µL, 50 mM, pH 8) containing lysozyme from chicken egg (1 mg $mL^{-1}$) (Sigma-Aldrich). After 2 h of incubation at 37 °C cell debris was collected by centrifugation at 1000 × *g* for 30 min, and the cell free extracts were transferred to a new 96 deep well plate for the assay.

For the design of the degenerated codons the Degenerate Codon Optimizer was used (Manfred T. Reetz Group).

### 5.7.3. Extraction of proteins from *E. coli* cells

The soluble proteins were extracted using a 0.75 kW 40K Cell Disrupter (Constant Systems) "French Press". For the lysis the cells were resuspended in 50 mL starting buffer (50 mM GlyGly, pH 8.5 containing dithiothreithol (DTT) (1 mM)) and lysed using the "French Press" at 1.65 KBar pressure. The lysate was cleared by centrifugation and the cellular debris was removed 8228 × g for 30 min at 4 °C. The clear supernatant was saved for further purification.

### 5.7.4. Purification of the recombinant proteins

The soluble fraction of the lysate was treated with heat-shock (70 °C during 30 min). Right after finishing this step the protein solution was directly placed on ice, and centrifuged at 8228 × g for 30 min at 4 °C. The precipitate was discarded and the supernatant was dialyzed against 5 L GlyGly buffer (5 mM, pH 8.5 containing DTT (0.1 mM)) and finally frozen and lyophilized. The proteins were quantified and analyzed by SDS-PAGE.

### 5.7.5. Purification of the recombinant proteins

In all cases the protein concentrations were calculated with the Bradford method [199]. To measure the protein concentration 0,5-1 mg sample was dissolved in 1 mL aqua milliQ. From this solution 1-10 μL was added to milliQ water to 100 μL volume, and to this mixture 900 μL Bradford reagent (Bio-Rad) was added and left for 5 minutes, the absorbency was measured at 595 nm. The measured absorbencies were interpolated from the calibration curve. For this reason a dilution series of bovine serum albumin (BSA) was used, with 6 different concentrations (0, 5, 10, 15, 25 and 50 μg/mL) from a mother solution of 0,25 mg/mL.

### 5.7.6. Protein electrophoresis (SDS-PAGE)

The protein electrophoresis was carried out with denaturated proteins in the presence of sodium dodecyl sulphate (SDS) according to the method described by Laemmli [241]. During this study 12 % acrylamide/bis-acrylamide separator gel conditions were used. In Table 5.5 the composition of gels used for the separation of proteins with a molecular weight between 10 and 100 KDa was used.

For the assembly and the development of the SDS-PAGE the Mini Protean II (Bio-Rad) system was used for the gels 7 cm x 8 cm. For the preparation of the samples a loading buffer was used with the constitution 0,5 M Tris/HCl pH 6,8, 20% glycerol, 10% SDS 0,05% 2-mercaptoethanol and 0,005% bromophenol blue, and it was mixed with the protein solutions in 3:1 proportion. Once the mixture is homogenized, a heat shock was used for 5 minutes at 95 $^{o}$C. The samples (10 μL) were then loaded into the wells of the gel, along with the ladder – unstained protein weight marker (Fermentas), which permits the identification of the corresponding protein bands between 10 and 200 KDa. For the electrophoresis an electrophoresis buffer was used (Tris 3%, glycine 14,4%, SDS 1%, pH 8,8), and the applied electric field is variable, between 90 V starting current, followed by 120 V, once the samples have arrived in the separator part. The electrophoresis is finished once the blue dye of the bromophenol blue arrives to the bottom of the gel.

**Table 5.6.** Composition of the gels of SDS-PAGE used in this work.

| | Final concentrations | |
|---|---|---|
| | **SEPARATOR GEL** | **STACKING GEL** |
| Tris/HCl | 375 mM (pH 8,8) | 125 mM (pH 6,8) |
| Acrylamide/Bis-acrylamide | 12% / 0,3 % | 4% / 0,1% |
| SDS | 0,1% | 0,1% |
| Ammonium persulfate | 0,1% | 0,1% |
| TEMED | 0,02% (v/v) | 0,02% (v/v) |

### a. *Visualization of the proteins: Staining with Coomasie Brillant Blue*

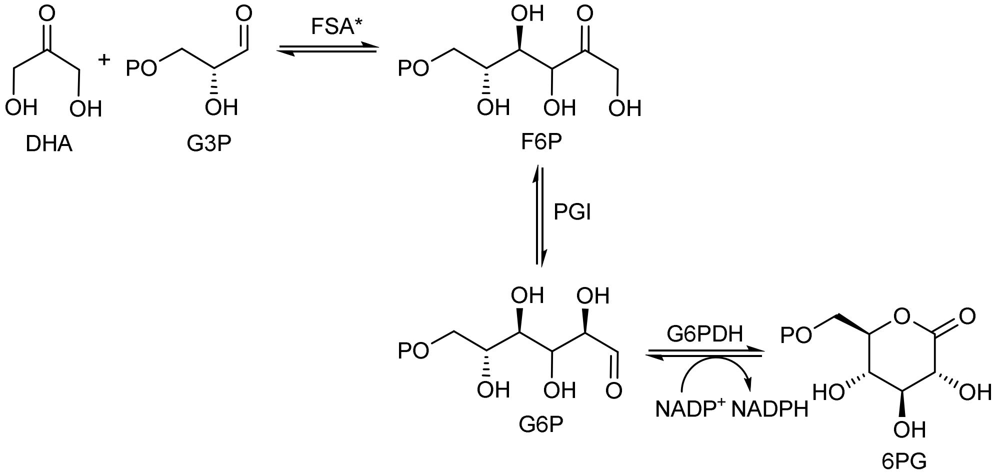
For staining the SDS-PAGE and the visualization of the proteins coomasie brilliant blue dye was used, which binds to the aromatic and basic residues of the protein, permitting the visualization and coloration of the protein with a detection limit of 30-100 ng, depending on the protocol [242]. The staining took place in 3 steps (i) Fixation: once the electrophoresis is finished, the gel was removed from the cast and incubated during 2 minutes in a solution of 45% methanol and 9% acetic acid, (ii) Staining: The prestaining solution was discarded and the gel was submerged in dying solution (40% methanol, 7% acetic acid and 0,25% Coomasie Brillant Blue (Bio-Rad)) and warmed in a microwave oven for 30 seconds, (iii) Destaining: the staining solution was removed and the gel was left in the destaining solution (5% MeOH, 7,5% acetic acid) and warmed in the microwave oven for 30 seconds, and left for a couple of hours with slight mixing. The solution was changed until the coloration was saturated. Finally the gels were scanned and analyzed.

## 5.7.7. Activity assay of the enzyme

To determine the enzymatic activity of the fructose-6 phosphate aldolase coupled spectrophotometric assays were used. In all assays a Cary 500 Bio UV-Visible (Varian) were used, where the enzyme activity was monitored during 5-10 minutes.

To determine the activity of fructose-6-phosphate aldolase the procedure of Schürmann and Sprenger (2001) [112] was used. In the assay the formation of D-fructose 6-phosphate from dihydroxyacetone (DHA) and D,L-glyceraldehyde 3-phosphate (D,L-GAP) was monitored with a coupled at 30 ºC using phosphonoglucose isomerase (PGI) and glucose-6-phosphate dehydrogenase

(G6PDH). The reduction of NADP$^+$ was monitored at 340 nm for 5 minutes in the formation of 1 mmol of NADPH was set equivalent to 1 mmol of D-fructose 6-phosphate formed.



**Scheme 5.1.** The scheme of the coupled enzymatic reactions for the measurement of the activity of D-fructose-6-phosphate aldolase.

The composition of the assay is described in Table 5.7.. The reaction is initiated with the addition of D,L-GAP, once all the other components of the reaction are well mixed. The enzyme sample is prepared from purified and lyophilized protein dissolved in the reaction buffer to have 5-10 μg final protein amount in the assay reaction. The final volume in the assay reaction is 1 mL.

**Table 5.7.** Composition of the reaction mixtures used in the enzymatic assay for the quantification of the transaldolase activity.

| Component | Volume (μL) | Final Concentration |
|---|---|---|
| DHA 5M | 60 | 50 mM |
| NADP$^+$ 50 mM | 10 | 0,5 mM |
| PGI | 1 | 3-6 U/mL |
| G6PDH | 1 | 3-6 U/mL |
| Buffer[1] 50 mM | 868 | ~50 mM |
| FSA*[2] | | 1-100 μg |
| D,L-GAP 280 M | 10 | 2,8 mM |

[1] Glycylglycine 50 mM, DTT 1 mM, pH 8,5
[2] FSA, and FSA mutants

One Unit is defined to as the enzyme amount that will convert 1 μmol D,L-glyceraldehyde phosphate to D-fructose 6-phosphate (D-F6P) (equivalent to the formation of 1 μmol NADPH) per minute at 30 ºC.

**Table 5.8.** Specific aldolase activity of the FSA wild-type and mutants used in this work.

| Enzyme | Specific activity U mg$^1$ protein (Bradford) |
|---|---|
| FSA | 0.22±0.01 |
| FSA A129G | 0.36±0.1 |
| FSA A129T | 0.54±0.04 |
| FSA A129V | 0.05±0.004 |
| FSA A129G/A165G | 0.09±0.02 |
| FSA A129G/S166G | 0.11±0.03 |
| FSA A129T/A165G | 0.26±0.07 |
| FSA A129T/S166G | 0.15±0.03 |
| FSA A129T/A165G/S166G | 0.18±0.006 |
| FSA A165G/S166G | 0.11±0.01 |
| FSA A165G/S166D | 0.011±0.002 |
| FSA A165G/S166N | 0.012±0.004 |
| FSA A165G/S166P | 0.04±0.01 |
| FSA A164G/A165G/S166P | 0.012±0.003 |
| FSA L107Y/A129G | 0.03±0.01 |
| FSA L107Y/A129G/A165G | 0.05±0.005 |
| FSA T185S | 0.85±0.15 |
| FSA A165G/T185S | 0.79±0.10 |
| FSA A129S | 9.4±1.3 |
| FSA A165G | 0.9±0.2 |
| FSA L107A | 0.52±0.03 |
| FSA L163A | 0.41±0.10 |
| FSA L107A/129G | 0.36±0.04 |
| FSA L107A/L163A | 0.50±0.01 |
| FSA A129G/L163A | 0.42±0.03 |
| FSA A129S/L163A | 0.57±0.04 |
| FSA L107A/A129G/L163A | 0.13±0.01 |
| FSA L107A/A129S/L163A | 0.56±0.04 |
| FSA R134E | 0.23±0.08 |
| FSA S166G | 0.88±0.10 |
| FSA A129S/A165G/S166G | 0.22±0.05 |
| FSA A129S/A165G | 2.5±0.9 |

| | |
|---|---|
| FSA L107Y | 0.11±0.01 |
| FSA L107H/A129G | 0.03±0.01 |
| FSA L107F/A129G | 0.02±0.006 |
| FSA L107H/A129G/A165G | 0.04±0.004 |

## 5.7.8. Electrospray ionization mass spectrometry of proteins

The analysis of the obtained mutant proteins was carried out in the Mass Spectroscopy Service of the CID-CSIC. Sample preparation was done in the following way: each protein (1 mg of lyophilized powder) was resuspended in water (1 mL) and formic acid (150 µL) and acetonitrile (50 µL) was added. Samples (10 µL) were analyzed by UPLC-ESI-MS using an Acquity UPLC$^{TM}$ BEH300 C18 column (1.7 µm, 2.1 × 100 mm), and an ESI-TOF mass spectrometer (LCT Premier Waters, Milford, MA, USA) equipped with a 4 GHz time-to-digital converter (TDC) with a dual ESI source (LockSpray). The second sprayer provided the lock mass calibration with leucine enkephalin (m/z 556.2771). The ESI-TOF was operated in the W-optics mode, thus providing a mass resolution of at least 10 000 full-width at half maximum (FWHM). The acquisition time per spectrum was set to 0.2 s, and the mass range was from 200 to 1800 Da. Data were acquired using a cone voltage of 50 V, capillary voltage of 3000 V, desolvation temperature of 350 ºC, and source temperature of 100 ºC. The desolvation gas flow was set at 400 L h$^{-1}$ and the cone gas flow was set at 30 L h$^{-1}$. The solvent system used for the elution was: solvent (A): formic acid 0.1% (v/v) in acetonitrile and solvent (B) aqueous formic acid 0.1% (v/v), gradient elution 0% B for 5 minutes, from 0% to 70% B in 14 min, from 70% to 100% B in 2 min, flow rate 0.3 mL min$^{-1}$. MassLynx 4.1 (Waters, Milford, MA, USA) was used for data acquisition and processing. MagTran software kindly provided by Dr. Zhongqi Zhang (Amgen, Inc.; Thousand Oaks, CA, USA) was used for molecular weight deconvolution from ESI-MS spectra of proteins.

**Table 5.9.** Mass spectroscopy of the FSA mutants used in this work.

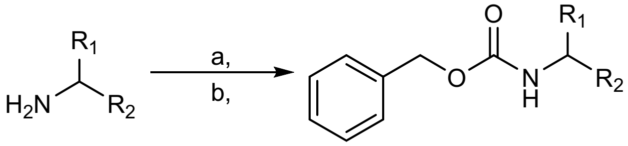
| Enzyme | Calculated | Observed | Error abs (Da) |
|---|---|---|---|
| FSA A129G | 22982.03 | 22982.0 | 0.03 |
| FSA A129T | 23026.08 | 23025.0 | 1.08 |
| FSA A129V | 23024.11 | 23024.0 | 0.11 |
| FSA A129G/S166G | 22952.00 | 22953.0 | 1.00 |

| | | | |
|---|---|---|---|
| FSA A129T/A165G | 23012.05 | 23012.0 | 0.05 |
| FSA A129T/S166G | 22996.06 | 22996.0 | 0.06 |
| FSA A129T/A165G/S166G | 22982.03 | 22982.0 | 0.03 |
| FSA A165G/S166G | 22952.00 | 22952.0 | 0.00 |
| FSA A165G/S166D | 23009.03 | 23009.0 | 0.03 |
| FSA A165G/S166N | 23009.05 | 23009.0 | 0.05 |
| FSA A165G/S166P | 22992.07 | 22992.0 | 0.07 |
| FSA L107Y/A129G | 23032.04 | 23033.0 | 0.96 |
| FSA L107Y/A129G/A165G | 23018.02 | 23018.0 | 0.02 |
| FSA L107A | 22953.98 | 22954.0 | 0.02 |
| FSA L163A | 22953.98 | 22954.0 | 0.02 |
| FSA L107A/129G | 22939.95 | 22940.0 | 0.05 |
| FSA L107A/L163A | 22911.90 | 22912.0 | 0.10 |
| FSA A129G/L163A | 22939.95 | 22940.0 | 0.05 |
| FSA A129S/L163A | 22969.97 | 22969.0 | 0.97 |
| FSA L107A/A129G/L163A | 22897.87 | 22898.0 | 0.13 |
| FSA L107A/A129S/L163A | 22927.90 | 22928.0 | 0.10 |
| FSA L107F/A129G | 23016.04 | 23016.0 | 0.04 |
| FSA L107H/A129G/A165G | 22992.49 | 22993.0 | 0.51 |

## 5.8. Chemical and analytical techniques

### 5.8.1. Preparation of *N*-Cbz-aminoaldehydes

The N-Cbz aminoaldehydes were synthesized from the corresponding amino alcohols using the procedures worked out in our group [86]. These were first protected as Cbz derivatives followed by oxidation of the alcohol function using 2-iodoxybenzoic acid (IBX) (Scheme 5.2.) with a yield between 90-95%. In most instances, a simple workup followed by a rough crystallization or precipitation provided the aminoaldehydes in a form pure enough to be used in the enzymatic aldol reactions.



a, Cbz-OSu, dioxane/$H_2O$
b, Oxidation with IBX
$R_1$: H, $CH_3$
$R_2$: CHO, $CH_2CHO$

**Scheme 5.2.** General scheme for the preparation of N-Cbz-aminoaldehydes used in this work.

### *a. Synthesis of N-benzyloxycarbonyl (Cbz) protected α-aminoalcoholes*

A solution of Cbz-OSu (11-56 mmol) in dioxane/water 4:1 (10-50 mL) was added to a solution of the α-aminoalcohol (13-65 mmol) in dioxane/water 4:1 (50-150 mL). After stirring for 12 h, the mixture was evaporated to dryness under reduced pressure. The residue was dissolved with ethyl acetate (150 mL) and washed successively with citric acid 5% (w/v) (3 x 50-75 mL), NaHCO3 10% (w/v) (3 x 50-75 mL), and brine (3 x 50-75 mL). After drying over $Na_2SO_4$, the organic layer was evaporated under reduced pressure. NMR spectrum and HPLC analysis of the residue showed only one product that coincides with the previously described ones [86].

### *b. Synthesis of 2-Iodoxybenzoic acid*

2-Iodoxybenzoic acid was prepared as described in the literature [243]. 2-Iodobenzoic acid (160 mmol) was dissolved in 1,5 L distilled water and to this was added all at once to a solution of Oxone® (2 $KHSO_5 \cdot K_2SO_4 \cdot KHSO_4$, 450 mmol). The reaction mixture was warmed to 70-73 °C over 20 min and mechanically stirred at this temperature for 3 h. The aspect of the mixture varies consistently during the reaction. The initial thick slurry coating the walls of the flask eventually becomes a finely dispersed, easy to stir suspension of a small amount of solid that sedimented easily upon stopping the stirring. The suspension was then cooled to 4 °C and left at this temperature for 12 h. The mixture was filtered through a medium porosity sintered-glass funnel, and the solid was repeatedly rinsed with water (4 x 120 mL) and acetone (4 x 120 mL). The white, crystalline solid was left to dry at reduced pressure and stored at 4 °C protected from light (79-81% yield).

### *c. Synthesis N-Cbz protected α-aminoaldehydes (3-5)*

The oxidation of the N-Cbz-aminoalcoholes to aldehydes was realized with IBX. The N-Cbz-aminoalcoholes (9-30 mmol) were dissolved in 30-50 mL ethyl acetate, to this solution IBX (24-48 mmol) was added and was left under reflux for 5-11 hours. The reaction was monitored by HPLC until no alcohol was detected. At this point, the reaction mixture was diluted with water (30-60 mL) and the mixture was extracted with ethyl acetate (3 x 75-100 mL). The organic layers were pooled, washed with $NaHCO_3$ 5% (w/w) (3 x 100 mL) and brine (3 x 100 mL), dried over $Na_2SO_4$, and evaporated under reduced pressure to give the title compounds.

## 5.8.2. Preparation of aromatic aldehydes

### *a. Synthesis of phenoxy acetaldehyde (22e)*

The oxidation of the phenoxyethanol to aldehyde was realized with IBX. The phenoxyalcohole (1 g, 903 μL, 7,24 mmol) was dissolved in 120 mL ethyl acetate, to this solution IBX (4,05 g, 14,48 mmol) was added and was left under reflux for 6 hours. The reaction was monitored by TLC until no alcohol was detected. At this point, the reaction mixture was diluted with water (30 mL) and the mixture was extracted with ethyl acetate (3 x 75-100 mL). The organic layers were pooled, washed with $NaHCO_3$ 5% (w/w) (3 x 100 mL) and brine (3 x 100 mL), dried over $Na_2SO_4$, and evaporated under reduced pressure to give the title compound (98% yield).

### *b. Synthesis of benzylthio acetaldehyde (22d)*

S-Benzylthioglycolic acid (5 g, 27 mmol) was dissolved in 50 mL methanol, to this solution thionyl chloride (9,8 mL, 16 g, 135 mmol) was added and left stirring for 24 hours. The reaction was monitored by TLC. Once the reaction was finished the solution was neutralized with saturated $NaHCO_3$ solution. The methanol was evaporated under reduced pressure, and extracted with ethyl acetate (3 x 75-100 mL) The organic layers were pooled, washed with $NaHCO_3$ 5% (w/w) (3 x 100 mL) and brine (3 x 100 mL), dried over $Na_2SO_4$, and evaporated under reduced pressure to give the S-Benzylthioglycolic acid-methyl-esther (98% yield).

S-Benzylthioglycolic acid-methyl-esther (5,22 g, 26,6 mmol) was dissolved in 150 mL anhydrous ethanol, and left stirring cold. To the solution $CaCl_2$ (5,9 mg, 53,2 mmol) was added with 50 mL dried THF. After 15 minutes $NaBH_4$ (4 g, 106,4 mmol) was added with 30 mL THF. The reaction was monitored by TLC. Once all the esther was transformed the mixture was neutralized with saturated $NH_4Cl$ solution. The reaction mixture was left to cool, and 400 mL saturated $NH_4Cl$ solution was slowly added to it. It was left for 30 minutes stirring. The THF and EtOH were removed by extraction with ethylacetate (25 mL x 4) dried over $Na_2SO_4$, and evaporated under reduced pressure affording the S-benzylthio-ethanol (92% yield).

The oxidation of the S-benzylthio-ethanol to aldehyde was realized with IBX. The S-benzylthio-ethanol (0,5 g, 2,97 mmol) was dissolved in 60 mL ethyl acetate, to this solution IBX (1,68 g, 6 mmol) was added and was left under reflux for 6 hours. The reaction was monitored by TLC until no alcohol was detected. At this point, the reaction mixture was diluted with water (40 mL) and the mixture was extracted with ethyl acetate (3 x 75-100 mL). The organic layers were pooled, washed with $NaHCO_3$ 5% (w/w) (3 x 100 mL) and brine (3 x 100 mL), dried over $Na_2SO_4$, and evaporated under reduced pressure to give the title compound (92% yield).

### 5.8.3. Preparation of 3-Hydroxypropionaldehyde (23b)

To a solution of 1.2 mL of $H_2SO_4$ (43 mequiv.) and 440 mL of $H_2O$ was added 30 mL (25.2 g, 449 mmol)of freshly distilled acrolein. The solution was allowed to stand in the dark in a sealed bottle at r.t. for 3 days, after which it was stored in the refrigerator (0-4°C). NMR-data compared with propionaldehyde showed a concentration of 1.0 M for the desired aldehyde [88].

## 5.8.4. Preparation of azidoacetaldehyde (22f)

A solution of sodium azide (4,56 g, 70,1 mmol) and (*E*)-1,4-dibrombut-2-ene (5,0 g, 23,3 mmol) in DMSO (70 mL) was stirred at room temperature for 3 days. The mixture was extracted with $Et_2O$ (200 mL × 3). The ether phase was washed with water (50 mL × 3) and brine (50 mL), then dried with $Na_2SO_4$. The solution was filtered and concentrated to 15-20 mL at 20 °C. To this solution MeOH (30 mL) and $CH_2Cl_2$ (120 mL) were added. The mixture was treated with ozone at 78 °C until blue color persisted and reduced with dimethyl sulfide (7,2 g, 116.5 mmol). The crude azidoacetaldehyde was purified by flash column chromatography (silica, 15:1 $CH_2Cl_2$/MeOH) to afford product (2,97 g, 75% in two steps) [125].

## 5.8.5. Preparation of the donor substrates: Ketohydroxylation of Olefins and Protected Olefins (39e-i)

The olefin was added to a solution of acetone, water and acetic acid. A solution of $KMnO_4$ in acetone and water was added dropwise, and the resulting mixture was stirred at r.t. until complete disappearance on TLC. After complete conversion, the crude mixture was filtered through a pad of Celite and washed several times with acetone. The filtrate was concentrated, and diluted with $Et_2O$ and washed with saturated aqueous $NaHCO_3$ until pH = 7. The organic layer was then washed with brine, filtered and dried over $MgSO_4$. The crude product was purified either by distillation or by column chromatography [198].

## 5.8.6. Preparation of the donor substrates: Selective oxidation of diols (39a-d)

The diol was dissolved in 3mL of glacial acetic acid and stirred magnetically. Aqueous 1.86 M NaOCl was added drop wise at room temperature over 15 minutes initiated a rapid exothermic reaction. Stirring was continued for one hour at room temperature. The solution was extracted with dichloromethane (3 x), washed with aqueous $NaHCO_3$ until pH = 7 and dried over $MgSO_4$. After evaporation of the solvent the product could be used after further purification [197].

## 5.8.7. Preparation of the donor substrates: Synthesis of ethers (42)

### *a. General procedure for the synthesis of epoxides*

A mixture of 40% aqueous NaOH, epichlorohydrin and TBABr was vigorously stirred at 0 °C for 30 minutes and allowed to reach room temperature. To this mixture the corresponding alcohol was gradually added for 30 minutes at r.t. The progress of the reaction was monitored by TLC. After completion of the reaction the mixture was quenched with water and extracted with $Et_2O$ (3 x). The organic phase was dried over $MgSO_4$, filtered and evaporated to give the respective epoxides without further purification [244].

### *b. General procedure for the synthesis of ethers*

Trityl chloride and $NEt_3$ were added to a stirred solution of glycidol or any prepared epoxide in anhydrous dichloromethane under argon atmosphere at 0 °C. Then the mixture was warmed to r.t. and the reaction was monitored by TLC. After complete conversion of the starting material, the reaction was quenched with satured aqueous $Na_2CO_3$. The aqueous layer was extracted with dichloromethane and the combined organic extracts were dried, evaporated, and purified by column chromatography to afford the corresponding epoxides. The protected epoxide was diluted in anhydrous alcohol using the corresponding sodium salt as a nucleophile. The suspension was refluxed for 15 hours and quenched with $H_2O$. The organic layer was dried with $MgSO_4$, filtered and evaporated. The crude product was purified by column chromatography. A suspension of Alcohol, PDC (or PCC), and 4 Å molecular sieves in anhydrous dichloromethane was vigorously stirred at r.t. When no alcohol remained the reaction mixture was worked up by addition of three volumes of diethyl ether. After decantation, the solution was filtered through a glass filter and afterwards through a pad of Celite. The combined filtrates were concentrated, and purified by silica-filtration. The oxidized compound was stirred in TFA until reaction was complete, quenched with H2O, extracted with $Et_2O$ and the water phases evaporated to dryness. Crude product was purified by column chromatography and pH was set neutral [245].

## 5.8.8. Preparation of (2*R*,3*R*)-3-hydroxy-2-methylpentanal (59) and (2*S*,3*S*)-3-hydroxy-2-methylpentanal (59)

The organocatalytic aldehydes used in the coupled reactions were prepared according to the procedure described by MacMillan *et al.* [43]. A suspension of freshly distilled propionaldehyde (3,61 mL, 50 mmol) and L- or D- proline (5 mmol) in 25 mL of DMF was stirred for 24 hours at 4 °C. Then the mixture was either worked up or after centrifugation of the proline used directly in the enzymatic aldol reactions. Work up: The resulting solution was diluted with diethyl ether and washed successively with water and brine. The aqueous layers were extracted with dichloromethane and dried over $MgSO_4$, and concentrated under reduced pressure.

## 5.8.9. HPLC analysis of the substrates and products by HPLC (High-performance liquid chromatography)

The aldol reactions and the synthesis of the different substrates were generally followed by HPLC. The analyses were performed on an X-Bridge$^{TM}$ C18, 5 µm, 4.6 ×250 mm reverse phase column from Waters (Milford, USA). Generally samples (30 µL) were taken at different reaction times and they were injected and eluted with the following conditions: solvent system (A) aqueous trifluoroacetic acid (TFA) 0.1% (*v/v*) and (B): TFA 0.095% (*v/v*) in $CH_3CN/H_2O$ 4:1, (in some cases (B) was changed to TFA 0.095% (*v/v*) in $CH_3CN$), flow rate 1 mL min$^{-1}$, detection at 215 nm, column temperature 30 °C. The concentration of products or substrates was quantified from the peak areas using and external standard methodology. Conversions to the aldol adducts were calculated from the concentration of the limiting starting substrate. In most cases gradient elution systems were used, with different conditions, which are described in each chapter.

### *a. Sample derivatization protocol*

For the visualization of the substrates or the aldol adducts that don't have any chromophore, sample derivatization was applied. The samples (10 µl) were mixed with a solution of *O*-benzylhydroxylamine hydrochloride (50 µl 21.1 mg mL$^{-1}$; 0.14 mmol mL$^{-1}$) in pyridine:methanol:water 33:15:2. After incubation at 50 °C for 60 min, samples were diluted in methanol (940 µl) and after centrifugation analyzed by HPLC chromatography.

### *b. Other samples*

The other samples were directly analyzed with HPLC. In general samples were taken (25-50 µl) and dissolved in methanol. After centrifugation they were directly injected to the HPLC.

## 5.8.10. Analysis of the products by NMR (Nuclear magnetic resonance)

Routine, $^1$H (400-500 MHz) and $^{13}$C (101 MHz) NMR spectra of compounds were recorded with a Varian Mercury-400 and Varian Anova-500 spectrometers, respectively. High-field $^1$H and $^{13}$C NMR analyses were carried out by using a Bruker Avance 500 spectrometer equipped with a high-sensitive CryoProbe for $D_2O$ and $CD_3OD$ solutions (studies performed by Dr. Teodor Parella, NMR service, Autonomous University of Barcelona). Full characterization of the described compounds was performed using typical gradient-enhanced 2D experiments: COSY, NOESY and HSQC, recorded under routine conditions. When possible, NOE data was obtained from selective 1D NOESY experiments by using a single pulsed-field-gradient echo as a selective excitation method and a mixing time of 500 ms.

## 5.8.11. Determination of the specific optical rotations ($\alpha_s$)

Specific optical rotations ($[\alpha]_D^{22}$) of the products of the aldol reactions were measured with a Perkin Elmer Model 341 (Überlingen, Germany) polarimeter (Na lamp, 589nm). The products (1,0 -2,0 mg) were dissolved in 1 mL methanol and the samples were measured at room temperature with a 0,1 dm cell with the polarized light.

## 5.9. Aldol additions

### 5.9.1. Screening of the library FSA A129X

The 96 clones were tested at analytical scale using the homo aldol addition of glycolaldehyde as screening reaction. The formation of D-threose was monitored by TLC and HPLC (see Supporting Info). The reactions (500 µL total volume) were performed in a 96-deepwell plate (1 mL) stirred with a vortex mixer (VIBRAX VXR basic, Ika) at 1200 rpm and 25 °C. A freshly prepared glycolaldehyde solution (45 µL, 0.05 mmol, from a stock solution of 1 M) was added to a triethanolamine buffer (355 µL, 50 mM, pH 8). The reactions were started by adding cell-free extract (100 µL) to the plate. Samples were withdrawn at 1 h, 3 h, 4 h and 24 h reaction time. (The chromatograms are presented in the Electronic supplementary material.)

### 5.9.2. Screening of the newly designed library of FSA mutants with modification in the Ala 129 position

The reactions (500 µL total volume) were performed using 96-deepwell plates stirred with a vortex mixer at 1200 rpm, at room temperature. Freshly prepared glycolaldehyde solution (68 µL, 0.075 mmol, from a stock solution of 1.1 M) was added to triethanolamine buffer 50 mM, pH 8 (332 µL). The reactions were started by adding the enzyme (100 µL, 0.5 mg protein, from a stock solution of 5 mg protein $mL^{-1}$). Samples were withdrawn (10 µL) at 0 min, 15 min, 30 min, 1 h, 2 h, 3 h, 6 h and 24 hour reaction times and followed the sample derivatization protocol. The reactions were carried out with FSA wild-type, FSA A129G, FSA A129T, FSA A129V, FSA A129G/A165G, FSA A129G/S166G, FSA A129T/A165G, FSA L107Y, FSA L107Y/A129G, FSA L107Y/A129G/A165G, FSA L107H/A129G, FSA L107H/A129G/A165G and FSA L107F/A129G to compare their affinity towards glycolaldehyde.

### 5.9.3. Competition reactions between HA, DHA and GO as donor substrates catalyzed by newly designed library of FSA mutants with modification in the Ala 129 position

A set of reactions were carried out with FSA wild-type and mutants to compare their efficacy towards the dihydroxyacetone (**14**) or hydroxyacetone (**15**) and glycolaldehyde (**1**) as donor substrates. The reactions (500 µL total volume) were performed using 96-deepwell plates stirred with the vortex mixer at 1200 rpm, at 25 °C. Freshly prepared solutions of glycolaldehyde and **14** or **15** (45 µL and 50 µL, respectively 0.05 mmol from a stock solutions of 1.1 M and 1 M, respectively) were added to triethanolamine buffer (50 mM, pH 8) (355 µL). The reactions were started by adding the enzyme in solution (50 µL, 0.25 mg protein, from a stock solution of 5 mg protein $mL^{-1}$) for the units used for each particular mutant see Table 5.8.). Samples (10 µL) were withdrawn at 0, 20 min, 30 min and 1 h reaction times, to ensure a linear dependence of the product formation versus time at the beginning of the reaction, and followed the sample derivatization protocol prior to HPLC analysis. The reactions were carried out with FSA wild-type, FSA A129S, FSA A129S/A165G, FSA A129G, FSA A129T, FSA A129V, FSA A129G/A165G, FSA A129T/A165G, FSA L107Y, FSA L107Y/A129G, FSA L107Y/A129G/A165G, FSA L107H/A129G, FSA L107H/A129G/A165G and FSA L107F/A129G

### 5.9.4. *In situ* NMR competition experiments between HA, DHA and GO as donor substrates catalyzed by FSA A129S, FSA A129S/A165G and FSA L107Y/A129G/A165G

The experiments were conducted on a Varian Anova-500 spectrometer. All components were prepared in $D_2O$. Solutions of **10** or **11** and glycolaldehyde (1 mL, 100 mM final concentration) were prepared. From this mixture, a sample (700 µL) was withdrawn and transferred into an NMR tube. The initial spectra were taken and the FSA mutant (1.4 mg protein, for Units see Table 5.8.) was added. The progress of the reaction was monitored by $^1H$ NMR (500 MHz) recording the spectra each 15 or 30 min for 3 to 24 h of reaction time.

## 5.9.5. Inhibition of FSA L107Y/A129G/A165G and FSA L107H/A129G/A165G by dihydroxyacetone (10)

A set of reactions were conducted with FSA L107Y/A129G/A165G and FSA L107H/A129G/A165G to measure their inhibition by dihydroxyacetone (**10**). The reactions (500 μL total volume) were performed using 96-deepwell plates stirred with a vortex mixer at 1200 rpm, at room temperature. Freshly prepared solution of glycolaldehyde (45 μL, 0.05 mmol, from a stock solution of 1.1 M) was added to triethanolamine buffer 50 mM, pH 8 (454.5 to 380 μL) and **10** (0.5, 1.25, 2.5, 3.75, 5, 12.5, 25, 37.5, 50, 62.5, 75 μL from a stock solution of 1 M, giving 12 different concentrations were: 0, 1, 2.5, 5, 7.5, 10, 25, 50, 75, 100, 125 and 150 mM) was added to the mixture. The reactions were started by adding the FSA triple mutants (100 μL, 0.5 mg protein, from stock solutions of 5 mg protein mL$^{-1}$). Samples (10 μL) were withdrawn at 0, 2, 5, 10 20 min reaction time, to ensure a linear dependence of the product formation versus time at the beginning of the reaction, and followed the sample derivatization protocol prior to HPLC analysis; U of activity: the amount of D-threose formation per minute per mg of protein in triethanolamine buffer 50 mM, pH 8 at 25ºC. The graphical representation of the activity vs. the logarithm of concentration of dihydroxyacetone gave the $IC_{50}$ value.

## 5.9.6. Initial reaction rates of the self- and cross aldol addition of glycolaldehyde

The most productive enzymes were chosen for a second set of reactions to calculate the initial rates of the cross-aldol ($v_{o\ cross}$, μmol min$^{-1}$ mg$^{-1}$) and auto-aldol addition ($v_{o\ auto}$, μmol min$^{-1}$ mg$^{-1}$). Reactions were carried out on the analytical scale as follows. In a 96-deepwell plate (1 mL) the aldehyde acceptor (0.024 mmol, 80 mM) was dissolved in dimethylformamide (60 μL). Then to this solution freshly prepared glycolaldehyde solution (10 μL, 0.0143 mmol mL$^{-1}$ in triethanolamine buffer 50 mM, pH 8) was added. The reaction was initiated by the addition of the corresponding FSA mutant (between 30 μL and 100 μL of protein from stock solutions between 5 to 15 mg protein mL$^{-1}$ to ensure a linear dependence of the product formation versus time at the beginning of the reaction; reaction conversions <10%) in triethanolamine buffer 50 mM, pH 8 (30 μL) and stirred with the vortex mixer at 1200 rpm and 25 °C. At different reaction times 0-10 min in all cases, samples (10 μL) were withdrawn for measuring the concentration of cross-aldol products and mixed with methanol (990 μL) centrifuged to remove the precipitated protein, and analyzed directly by HPLC, and at the same time samples (10 μL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Concentrations of the aldol products were calculated with an external standard method from peak areas.

### 5.9.7. Kinetic parameters for glycolaldehyde as donor and acceptor

The steady state kinetic parameters were determined using a previous published methodology [122]. Reaction: formation of D-threose and measurement of GA consumption with glycerol dehydrogenase from *Cellulomonas sp.* ($GDH_{cell}$). To a solution of GA (at concentration from 0.4 mM to 200 mM) in 100 mM TEA buffer pH 7.5 at 25 °C, FSA (3 mg to 7.5 mg powder, for Units see Table 5.8.) was added. The final volume was 1 mL. Aliquots were withdrawn at different times and the amount of GA was determined using the enzyme $GDH_{cell}$ (10 U) in presence of NADH (0.7 mM). The oxidation of NADH was monitored as indicated above. One mmol of NADH oxidized was equivalent to 1 mmol of D-threose formed.

### 5.9.8. Enzymatic aldol reactions on analytical scale with glycolaldehyde as donor and *N*-Cbz-aminoaldehydes as acceptor by the newly designed library of FSA mutants with modification in the Ala 129 position

A preliminary set of reactions was conducted to test the obtained mutant proteins with selected *N*-Cbz-aminoaldehydes. Reactions (300 μL total volume) were performed using 96-deepwell plates stirred with vortex mixer at 1200 rpm, at room temperature. The aldehyde (24 μmol, 80 mM) was dissolved in DMF (60 μL) and then triethanolamine buffer (50 mM, pH 8) was added (120 μL). To this solution, freshly prepared glycolaldehyde solution was added (2.7 μL, 3 μmol, from stock solution in plain water 1.1 M). The reactions were started by adding the corresponding FSA mutant (120 μL, 0.6 mg protein, for Units see Table 5.8., from a stock solution of 5 mg protein $mL^{-1}$) dissolved in triethanolamine buffer 50 mM, pH 8 (120 μL), followed by a portion wise addition of glycolaldehyde stock solution (9 additions of 2.7 µL containing 3 µmol glycolaldehyde, total amount of glycolaldehyde added: 30 µmol) during the following time intervals: 30 min, 1 h, 1.5 h, 2 h, 2.5 h, 3 h, 4 h, 5 h, and 6 h. At 0, 1, 6 and 24 h samples were withdrawn (25 µL) and dissolved in 475 µL MeOH, and subsequently analyzed by HPLC.

## 5.9.9. Scale up cross-aldol addition reaction of glycolaldehyde to *N*-Cbz-aminoaldehydes

General procedure: Reactions (10 mL total volume) were conducted using screw capped conical-bottom polypropylene tubes (50 mL) stirred with the vortex mixer (200 rpm) at 25 ºC. The aldehyde (0.8 mmol) was dissolved in dimethylformamide (2 mL). To this solution triethanolamine buffer 50 mM, pH 7.5 (7 mL) was added and, finally, freshly prepared glycolaldehyde solution (0.1 mL, 0.1 mmol, 1 M solution) was added. The reactions were started by adding the FSA A129G or FSA L107Y/A129G/A165G, depending on the experiment, followed by a portion wise addition of glycolaldehyde solution (9 additions of 0.1 mL containing 0.1 mmol of glycolaldehyde, total amount of glycolaldehyde added: 1 mmol) during the following time intervals: 30 min, 1 h, 1.5 h, 2 h, 3.5 h, 4.5 h, 5.5 h, 6.5 h, and 8 h. Reactions were monitored by HPLC analysis: gradient elution 10 to 70% B in 30 min. When the reaction was finished, MeOH (10 mL) was added to stop the reaction and precipitate the enzyme. The excess of MeOH was then removed in a rotary evaporator and the suspension was filtered through a nylon filter (0.45 µm). The filtrate was loaded onto a semi-preparative column (Waters X-Terra Prep MS C-18, 10 µm, 19x250mm). Elution conditions for each compound are given below. The flow rate was 10 mL min$^{-1}$ and the products were detected at 215 nm. The fractions were analyzed under isocratic conditions of 35.6% ACN during 15 min. Products were structurally characterized by NMR. A copy of the recorded spectra with the peak assignments is depicted in the Electronic Supporting Information.

### a. *(2R,3S,4R)-N-benzyloxycarbonyl-2,3,4-trihydroxypyrrolidine (6a) and (2S,3S,4S)-N-benzyloxycarbonyl-2,3,4-trihydroxypyrrolidine (6b)*

The title compounds were obtained following the procedure described above. *N*-Cbz-Glycinal (**3**) (143.0 mg, 0.74 mmol) donor and FSA L107Y/A129G/A165G (38.5 mg protein, for Units see Table 5.8.) were used. The maximum conversion 98% was achieved after 22 h. Semi-preparative HPLC elution conditions: isocratic 20% of ACN. Pure fractions were pooled and lyophilized, obtaining 131 mg (0.52 mmol) of a 3:2 mixture of **6a**:**6b** product (98.1% of purity **6a** + **6b** by HPLC, 70 % isolated yield **6a** + **6b**). Major (**6a**): $^1$H NMR (400 MHz, $CD_3OD$) $\delta$ 7.34 (m, 5H), 5.29 (dd, *J* = 5.0, 1.5 Hz, 1H), 5.14 (m, 2H), 4.21 (q, *J* = 7.0 Hz, 1H), 3.80 (m, 1H), 3.74 (m, 1H), 3.12 (m, 1H). $^{13}$C NMR (101 MHz, $CD_3OD$) $\delta$ 155.44 (rotamer 155.38), 136.64, 127.75, 79.71 (rotamer 79.43), 76.39 (rotamer 75.84), 72.63 (rotamer, 72.03), 66.77, 49.08. Minor (**6b**) $^1$H NMR (400 MHz, $CD_3OD$) $\delta$ 7.34 (m, 5H), 5.21 (m, 1H), 5.14 (m, 2H), 4.06 (m, 1H), 3.96 (d, *J* = 9.7 Hz, 1H), 3.74 (m, 1H), 3.47 (m, 1H). $^{13}$C NMR (101 MHz, $cd_3od$) $\delta$ 155.74 (rotamer 155.60), 136.54, 127.75, 86.85 (rotamer 86.35), 80.39 (rotamer 79.70), 74.11 (rotamer 73.39), 66.67, 52.37 (rotamer 52.19). HRMS-ESI: *m*/*z* calcd for $C_{12}H_{15}NO_5Na$ 276.0848 [M+Na$^+$]; found: 276.0858.

The test with other FSA mutants was conducted under identical conditions but using a reaction volume of 1 mL: *N*-Cbz-Glycinal (100 mM), glycolaldehyde (initial 10 mM) added portionwise (10 $\mu$L of stock solution in triethanolamine 50 mM, pH 8 (1 M)) during the following time intervals: 30 min, 1 h, 1.5 h, 2 h, 2.5 h, 3 h, 3.75 h, 4.5 h, 5.25 h, 6 h and 6.75 h. FSA mutants (1 mg, for Units see Table 5.8.) FSA A129G/A165G, FSA A129G, FSA A129S/A165G, FSA A165G. and FSA A129T (6.4 mL total volume, for Units see Table 5.8., 19 mg enzyme) were used. Work up and purification as described above. $^1$H NMR spectra of the aldol adducts obtained are depicted in Supporting information.

### b. *(2S,3R,4R)-2-methylpyrrolidine-3,4-diol (15a) and (2S,3S,4R)-2-methylpyrrolidine-3,4-diol (15b) 4:1*

The title compounds were obtained following the procedure described above. (*S*)-*N*-Cbz-Alaninal ((*S*)-**4**) (166.0 mg, 0.80 mmol) and FSA L107Y/A129G/A165G (**B6**) (38.5 mg, for Units see Table 5.8.). The maximum conversion 80 % was achieved after 49 h. Semi preparative HPLC conditions: isocratic elution at 20% of ACN. The pure fractions were pooled and lyophilized, obtaining 150.4 mg (0.56 mmol) of product (97.1% of purity, 56 % isolated yield). The aldol adduct was dissolved in $H_2O$:EtOH 4:1 (120 mL) and treated with $H_2$ (50 psi) in the presence of Pd/C (250 mg) at room temperature during 24h. After removal of the catalyst by filtration through a 0.45µm nylon filter, the filtrate was lyophilized obtaining a solid material (65.9 mg), which was characterized without further purification. Two diasteroisomers were identified by NMR in 4:1 proportion corresponding to the title compounds. Major (**11a**). $^1$H NMR (400 MHz, $CD_3OD$) δ 4.17 (dt, *J* = 5.1, 1.4 Hz, 1H), 3.87 (dd, *J* = 3.2, 1.4 Hz, 1H), 3.53 (dd, *J* = 6.8, 3.2 Hz, 1H), 3.45 (dd, *J* = 12.5, 5.1 Hz, 1H), 2.89 (dd, *J* = 12.5, 1.6 Hz, 1H), 1.29 (d, *J* = 6.8 Hz, 3H). $^{13}$C NMR (101 MHz, $CD_3OD$) δ 77.3, 76.1, 56.8, 51.4, 10.9. Minor (**11b**): $^1$H NMR (400 MHz, $CD_3OD$) δ 4.05 (m, 1H), 3.92 (q, *J* = 6.0, 5.5 Hz, 2H), 3.59 (m, 1H), 3.18 (dd, *J* = 11.1, 5.7 Hz, 1H), 1.13 (d, *J* = 5.6 Hz, 3H). $^{13}$C NMR (101 MHz, $CD_3OD$) δ 76.9, 73.4, 54.5, 50.6, 13.6. HRMS-ESI: *m/z* calcd for $C_5H_{12}NO_2$ 118.0868 [M+H$^+$]; found: 118.0864.

### *c. (2R,3R,4R)-2-methylpyrrolidine-3,4-diol (16)*

The title compound was obtained following the procedure described above. (*R*)-*N*-Cbz-Alaninal ((*R*)-**4**) (166.0 mg, 0.80 mmol) and FSA L107Y/A129G/A165G (38.5 mg, for Units see Table 5.8.). The maximum conversion 89 % was achieved after 49 h. Semi preparative HPLC conditions: gradient elution from 0 to22.5% of ACN in 90 min; second run under isocratic elution at 20% of ACN. The pure fractions were pooled and lyophilized, obtaining 79 mg (0.30 mmol) of product (99.9% of purity, 37.0% isolated yield). The aldol adduct was dissolved in $H_2O$:EtOH 4:1 (75 mL) and treated with $H_2$ (50 psi) in the presence of Pd/C (157 mg) at room temperature during 24h. After removal of the catalyst by filtration through a 0.45 µm nylon filter, the filtrate was lyophilized obtaining the title compound as a solid material (34.0 mg), which was characterized without further purification. $^1$H NMR (400 MHz, $CD_3OD$) δ 4.08 (dt, *J* = 5.3, 2.7 Hz, 1H), 3.68 (m, 1H), 3.21 (dd, *J* = 12.2, 5.1 Hz, 1H), 3.10 (qd, *J* = 6.9, 4.5 Hz, 1H), 2.99 (dd, *J* = 12.2, 2.5 Hz, 1H), 1.34 (d, *J* = 6.9 Hz, 4H). $^{13}$C NMR (101 MHz, $CD_3OD$) δ 82.7, 77.1, 61.2, 51.3, 16.6. HRMS-ESI: *m/z* calcd for $C_5H_{12}NO_2$ 118.0868 [M+H$^+$]; found: 118.0864.

### *d. (2R,3S,4R)-N-benzyloxycarbonyl-2,3,4-trihydroxypiperidine (17a)*

The title compounds were obtained following the procedure described above. *N*-Cbz-3-aminopropanal (**5**) (172.0 mg, 0.83 mmol) and FSA A129G (21 mg protein, for Units see Table 5.8.) were used. The maximum conversion 93% was achieved after 44 h. Semi preparative HPLC conditions: isocratic elution at 20% ACN. The pure fractions were pooled and lyophilized, obtaining 83 mg (0.31 mmol) of **13a** (98.0 % of purity, 35 % isolated yield). NMR data matched those described in a previous work using the same procedure but using FSA wild-type [122]. A copy of the recorded spectra with the peak assignments is depicted in Supporting Info. HRMS-ESI: *m*/*z* calcd for $C_{13}H_{17}NO_5Na$ 290.1004 [M+Na$^+$]; found: 290.1010.

### *e. (2*R*,3*S*,4*R*)-*N*-benzyloxycarbonyl-2,3,4-trihydroxypiperidine-1-carboxylate (17a) and (2*R*,3*S*,4*S*)-*N*-benzyloxycarbonyl-2,3,4-trihydroxypiperidine-1-carboxylate (17b)*

The scale up was carried out following the procedure described above with the following variations: reaction volume (5 mL), the aldehyde (0.4 mmol) dissolved in 1 mL dimethylformamide, triethanolamine buffer (3.55 mL, 50 mM, pH 8), glycolaldehyde solution (45 µL, 0.05 mmol, 1.1 M) was added. The reaction was started by adding FSA L107Y/A129G/A165G (10 mg protein, for Units see Table 5.8.). During the reaction additional 0.05 mmol glycolaldehyde solution (45 µL, 1.1 M) was added at 0.5 h, 1 h, 1.5 h, 2 h, 2.5 h, 3 h, 4 h 6 h and 8 h reaction times. The progress of the reaction was monitored by HPLC analysis. Two new peaks appeared in the chromatogram, one at the retention time of **17a**. When the peak of the acceptor aldehyde disappeared (24 h) the reaction was stopped by adding MeOH (5 mL) and the work up was conducted as described above. Semi-preparative HPLC conditions: solvent system (A) plain water and (B) $CH_3CN/H_2O$ (4:1), gradient elution from 0-50% B in 30 min. Fraction analysis on HPLC: isocratic elution 33% of solvent B during 15 min. Pure fractions were lyophilized to give **17a** and **17b** diastereomers (17 and 20 mg, respectively). NMR spectra of **17a** matched those described above and in a previous work (see a copy in Supporting Info.). For the **13b** (see Supporting Info. Fig. S17): $^1H$ NMR (400 MHz, $CD_3OD$) δ 7.32 (m, 5H), 5.68 (dd, *J* = 3.0, 1.1 Hz, 1H), 5.10 (m, 2H), 3.95 (m, 1H), 3.78 (s, 1H), 3.13 (m, 1H), 1.82 (qd, *J* = 12.7, 4.9 Hz, 1H), 1.57 (m, 1H). $^{13}C$ NMR (101 MHz, $CD_3OD$) δ 128.0, 127.6, 127.4, 79.1, 70.0, 66.9, 65.7, 37.1, 26.9. HRMS-ESI: *m*/*z* calcd for $C_{13}H_{17}NO_5Na$ 290.1004 [M+Na$^+$]; found: 290.1008.

## 5.9.10. Screening the libraries: FSA A165G/S166X and FSA L107Y/A129G/A165G/S166X

The 137 and 96 clones were tested at analytical scale using the homo-aldol addition of glycolaldehyde as screening reaction. The formation of D-threose and L-erythrose was monitored by and HPLC. The reactions (500 µL total volume) were performed in a 96-deepwell plate (1 mL) stirred with a vortex mixer (VIBRAX VXR basic, Ika) at 1200 rpm and 25 ºC. A freshly prepared glycolaldehyde solution (100 mM, 0.05 mmol from a stock solution of 1M) was added to triethanolamine buffer (350 µl, 50 mM, pH 8). The reactions were started by adding cell-free extract (100 µL) to the plate. Samples were withdrawn at 24 h reaction time, and analyzed by HPLC (See Supporting Info.). Clones were considered positive with higher conversion to L-erythrose than to D-threose.

### 5.9.11. Screening of the FSA mutants with high activity towards glycolaldehyde with selected aldehydes (22-29)

A preliminary set of reactions was conducted to test the obtained mutant proteins (FSA wild-type, FSA A129G, FSA A129T, FSA A129V, FSA A129G/A165G, A129G/S166G, FSA A129T/A165G, FSA A129T/S166G, FSA A129T/A165G/S166G, FSA L107Y/A129G, FSA L107Y/A129G/A165G, FSA A165G, FSA A165G/S166D, FSA A165G/S166G, FSA A165G/S166N, FSA A165G/S166P) with selected aldehydes. Reactions (500 µL total volume) were performed using 96-deepwell plates stirred with vortex mixer at 1200 rpm, at room temperature. To each aldehyde (0.05 mmol, 100 mM) triethanolamine buffer (50 mM, pH 8) was added (290 µL). To this solution freshly prepared glycolaldehyde solution was added (9 µL, 10 µmol from a stock solution of 1.1 M). The reactions were started by adding the corresponding FSA variant (100 µL, 1 mg protein, (for units see Table 5.8.) from a sock solution of 5 mg protein mL$^{-1}$) dissolved in triethanolamine buffer solution (50 mM, pH 8), followed by a portion wise addition of glycolaldehyde stock solution (14 additions of 4.5 or 9 µL glycolaldehyde stock solution containing 5 or 10 µmol glycolaldehyde, total amount of glycolaldehyde added: 100 µmol) during the following time intervals: 30 min, 1 h, 1.5 h, 2 h, 2.5 h, 3 h, 3.5 h, 4 h, 5 h, 6 h, 7 h, 8 h, 9 h,10 h. Samples (10 µL) were withdrawn at 0, 3, 24 hour reaction times and followed the sample derivatization protocol prior to HPLC analysis.

## 5.9.12. Scale up of double additions of glycolaldehyde to selected aldehydes

### *a. D-idose (18)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the glycolaldehyde (1.5 mmol) triethanolamine buffer (50 mM, pH 8, 10 mL) was added. The reaction was started by FSA A129T/A165G (30 mg for Units see Table S2). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 20 to 60% B in 16 min. After 96 h, the reaction was stopped with MeOH (10 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 5.02. The MeOH was removed with rotary evaporator and the filtrate was lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 7:3 yielded 43 mg (48%). $[\alpha]_D^{20}$= + 22.8 (c= 0.69 in MeOH) (lit. [246] $[\alpha]_D^{25}$= + 7.7 (c = 0.3 in $H_2O$). ESI-TOF: calculated for $[M+Na]^+$ $C_6H_{12}O_6Na$: 203.0532, found 203.0527.

**Table 5.10.** $^1$H and $^{13}$C NMR shifts (ppm) of **18**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-furanose | 5.20 (103.0) | 4.10 (82.1) | 4.16 (76.16) | 4.13 (82.70) | 3.97 (72.28) | 3.70 (64.0) | 3.60 |
| β-furanose | 5.41 (96.8) | 4.14 (77.41) | 4.29 (76.35) | 4.18 (79.1) | 3.81 (71.42) | 3.70 (63.9) | 3.60 |
| α-pyranose | 4.97 (94.4) | 3.44 (74.2) | 3.72 (73.2) | 3.84 (71.4) | 4.13 (74.3) | 3.82 (59.8) | 3.75 |
| β-pyranose | 5.05 (93.7) | 3.66 (71.1) | 4.05 (71.0) | 3.64 (69.2) | 3.99 (76.1) | 3.84 (62.6) | 3.77 |

### *b. L-xylose (21)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the aldehyde (1 mmol) triethanolamine buffer (50 mM, pH 8, 10 mL) was added and, finally, FSA A129T/A165G (20 mg for Units see Table S2). The reaction was started by the fed-batch addition of glycolaldehyde solution (2 mmol, 400mM, 5 mL) with a syringe pump for 12 h (0.42 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90% B in 30 min. After 24 h, the reaction was stopped with MeOH (15 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.95. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 4:1 yielded 93 mg (62%). $[\alpha]_D^{20} = -21.9$ (c= 1.45 in MeOH) (lit. [247] $[\alpha]_D^{20} = -18.5$ (in $H_2O$). ESI-TOF: calculated for $[M+Na]^+$ $C_5H_{10}O_5Na$: 173.0426, found 173.0427.

**Table 5.11.** $^1$H and $^{13}$C NMR shifts (ppm) of **21**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_{5'}$ |
|---|---|---|---|---|---|---|
| α-pyranose | 4.41 (96.5) | 3.06 (73.9) | 3.27 (75.7) | 3.46 (69.1) | 3.77 (65.1) | 3.16 |
| β-pyranose | 5.03 (92.1) | 3.36 (71.4) | 3.47 (72.7) | 3.44 (69.3) | 3.52 (60.8) | 3.52 |

### *c. 6-deoxy-D-idose (30a)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To triethanolamine buffer (50 mM, pH 8, 10 mL) FSA A129T/A165G/S166G (20 mg for Units see Table 5.8) was added. The reaction was started by the fed-batch addition of a mixture of glycolaldehyde and acetaldehyde solution (2 mmol, 400mM, and 1 mmol, 200 mM, respectively in 10 mL water) with a syringe pump for 25 h (0.25 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90%B in 30 min. After 72 h, the reaction was stopped with MeOH (20 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.74. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 85:15, followed by purification on activated charcoal eluted with gradient $H_2O$ – MeOH 1:0 to 8:2 yielded 11 mg (7%) $[\alpha]_D^{20}$= + 6.3 (c= 0.63 in MeOH) (lit. [246] $[\alpha]_D^{25}$= + 9.0 (c = 1.50 in $H_2O$). ESI-TOF: calculated for $[M+Na]^+$ $C_6H_{12}O_5Na$: 187.0582, found 187.0578.

**Table 5.12.** $^1H$ and $^{13}C$ NMR shifts (ppm) of **30a**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ |
|---|---|---|---|---|---|---|
| α-furanose | 5.06<br>(101.9) | 3.95<br>(80.7) | 3.99<br>(74.7) | 3.81<br>(85.7) | 3.91<br>(69.3) | 1.08<br>(18.2) |
| β-furanose | 5.31<br>(95.8) | 3.97<br>(76.1) | 4.09<br>(75.3) | 3.86<br>(82.1) | 3.92<br>(66.9) | 1.06<br>(18.0) |
| α-pyranose | 4.73<br>(92.1) | 3.17<br>(73.7) | 3.54<br>(71.4) | 3.55<br>(71.7) | 3.50<br>(69.6) | 1.10<br>(12.5) |
| β-pyranose | 4.89<br>(92.2) | 3.50<br>(69.5) | 3.94<br>(69.7) | 3.32<br>(70.0) | 3.94<br>(70.2) | 1.11<br>(15.5) |

### *d. 6-O-methyl-D-idose (30b)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the aldehyde (1 mmol) triethanolamine buffer (50 mM, pH 8, 10 mL) was added and, finally, FSA A129T/A165G (20 mg for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (2 mmol, 400mM, 5 mL) with a syringe pump for 12 h (0.42 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90% B in 30 min. After 24 h, the reaction was stopped with MeOH (15 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.87. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 4:1 yielded 76 mg (39%). $[\alpha]_D^{20}$= + 10.3 (c= 0.85 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_7H_{14}O_6Na$: 217.0688, found 217.0690.

**Table 5.13.** $^1H$ and $^{13}C$ NMR shifts (ppm) of **30b**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-furanose | 5.06<br>(101.6) | 3.96<br>(80.8) | 4.00<br>(76.0) | 3.97<br>(81.3) | 3.87<br>(68.2) | | |
| β-furanose | 5.27<br>(95.4) | 4.00<br>(77.8) | 4.01<br>(76.0) | 4.02<br>(77.8) | 4.03<br>(74.7) | | |
| α-pyranose | 4.83<br>(93.3) | 3.28<br>(72.5) | 3.55<br>(69.5) | 3.64<br>(70.3) | 4.14<br>(70.7) | 3.62<br>(69.3) | 3.51 |
| β-pyranose | 4.91<br>(92.2) | 3.54<br>(69.5) | 3.92<br>(69.5) | 3.96<br>(72.8) | 3.96<br>(72.8) | 3.58<br>(71.9) | 3.52 |

### *e. 6-O-benzyl-D-idose (30c)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. The aldehyde (1 mmol) was dissolved in dimethylformamide (2 mL) To this solution triethanolamine buffer (50 mM, pH 8, 8 mL) was added and, finally, FSA A129T/A165G/S166G (20 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (2 mmol, 400mM, 5 mL) with a syringe pump for 12 h (0.42 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 30 to 100% B (TFA (0.095% (*v/v*)) in $CH_3CN$) in 30 min. After 48 h, the reaction was stopped with MeOH (15 mL) to precipitate the enzyme, and the pH of the mixture was adjusted to 4.97. The excess MeOH was removed with rotary evaporator and the mixture was filtered through a nylon filter (0.45 µm). The filtrate was loaded onto a semi-preparative column (Waters X-Terra Prep MS C-18, 10 µm 19 × 250 mm) Elution conditions: solvent system (A) milliQ grade water and (B) $CH_3CN/H_2O$ (4:1), gradient elution 0-60% B in 30 min. The flow rate was 10 mL $min^{-1}$ and the products were detected at 215 nm. The fractions were analyzed under isocratic conditions of 30 % B during 15 min. Pure fractions were pooled and yielded 124 mg (46%). $[\alpha]_D^{20}$= + 14.1 (c= 0.92 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_{13}H_{18}O_6Na$: 293.1001, found 293.0993.

**Table 5.14.** $^1H$ and $^{13}C$ NMR shifts (ppm) of **30c**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-furanose | 5.07<br>(102.7) | 4.00<br>(81.5) | 4.15<br>(76.2) | 4.15<br>(81.7) | 4.05<br>(69.3) | 3.57<br>(71.1) | 3.57 |
| β-furanose | 5.34<br>(96.1) | 3.97<br>(76.9) | 4.15<br>(76.4) | 4.02<br>(76.1) | 4.08<br>(69.4) | 3.58<br>(71.3) | 3.58 |
| α-pyranose | 5.02<br>(94.8) | 3.49<br>(70.8) | 3.81<br>(70.5) | 3.68<br>(69.8) | 4.36<br>(67.9) | 3.71<br>(69.4) | 3.71 |
| β-pyranose | 4.97<br>(92.5) | 3.57<br>(70.8) | 4.01<br>(69.6) | 3.54<br>(68.6) | 4.09<br>(72.9) | 3.73<br>(69.4) | 3.73 |

### *f. 6-deoxy-6-thiobenzyl-D-idose (30d)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. The aldehyde (1 mmol) was dissolved in dimethylformamide (2 mL) To this solution triethanolamine buffer (50 mM, pH 8, 8 mL) was added and, finally, FSA A129T /S166G (30 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (2 mmol, 400mM, 5 mL) with a syringe pump for 12 h (0.42 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 30 to 100% B (TFA (0.095% (*v/v*)) in $CH_3CN$) in 30 min. After 52 h, the reaction was stopped with MeOH (15 mL) to precipitate the enzyme, and the pH of the mixture was adjusted to 4.88. The excess MeOH was removed with rotary evaporator and the mixture was filtered through a nylon filter (0.45 µm) and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 9:1 yielded 111.8 mg (39%). $[\alpha]_D^{20}$= + 15.3 (c= 0.77 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_{13}H_{18}O_5SNa$: 309.0773, found 309.0771.

**Table 5.15.** $^1$H and $^{13}$C NMR shifts (ppm) of **30d**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-furanose | 5.06<br>(102.6) | 3.96<br>(81.6) | 4.00<br>(76.3) | 4.14<br>(82.5) | 3.90<br>(69.6) | 3.75<br>(36.0) | 3.75 |
| β-furanose | 5.32<br>(96.0) | 3.96<br>(77.0) | 4.17<br>(78.7) | 4.15<br>(76.4) | 3.90<br>(69.6) | 3.75<br>(35.9) | 3.75 |
| α-pyranose | 4.91<br>(92.6) | 3.47<br>(70.3) | 3.67<br>(69.7) | 3.76<br>(70.2) | 4.22<br>(67.8) | 3.75<br>(35.7) | 3.75 |
| β-pyranose | 4.95<br>(94.6) | 3.52<br>(70.6) | 4.00<br>(69.8) | 3.57<br>(68.6) | 3.90<br>(73.5) | 3.75<br>(35.6) | 3.75 |

### *g. 6-O-phenyl-D-idose (30e)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. The aldehyde (1 mmol) was dissolved in dimethylformamide (2 mL) To this solution triethanolamine buffer (50 mM, pH 8, 8 mL) was added and, finally, FSA A129T /S166G (30 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (2 mmol, 400mM, 5 mL) with a syringe pump for 12 h (0.42 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 30 to 100% B (TFA (0.095% (*v/v*)) in $CH_3CN$) in 30 min. After 52 h, the reaction was stopped with MeOH (15 mL) to precipitate the enzyme, and the pH of the mixture was adjusted to 4.93. The excess MeOH was removed with rotary evaporator and the mixture was filtered through a nylon filter (0.45 µm) and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 9:1 yielded 111.2 mg (43%). $[\alpha]_D^{20}$= + 30.0 (c= 0.80 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_{12}H_{16}O_6Na$: 279.0845, found 279.0845.

**Table 5.16.** $^1$H and $^{13}$C NMR shifts (ppm) of **30e**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-furanose | 5.10<br>(102.7) | 3.99<br>(81.6) | 4.20<br>(76.3) | 4.28<br>(81.3) | 4.23<br>(69.1) | 4.05<br>(68.7) | |
| β-furanose | 5.37<br>(96.1) | 4.00<br>(77.0) | 4.07<br>(76.3) | 4.28<br>(77.8) | 4.16<br>(68.9) | 4.09<br>(69.0) | |
| α-pyranose | 5.02<br>(94.9) | 3.51<br>(70.6) | 3.85<br>(70.45) | 3.74<br>(69.5) | 4.49<br>(67.4) | 4.17<br>(66.5) | 4.17 |
| β-pyranose | 5.00<br>(92.6) | 3.57<br>(70.8) | 4.02<br>(69.7) | 3.62<br>(68.3) | 4.21<br>(72.5) | 4.17<br>(67.0) | 4.17 |

### *h. 6-deoxy-6-azido-D-idose (30f)*

Reaction (50 mL starting volume) was conducted in a screw capped Erlenmeyer flask (100 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the aldehyde (5 mmol) triethanolamine buffer (50 mM, pH 8, 50 mL) was added and, finally, FSA A129T /S166G (150 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (10 mmol, 500mM, 10 mL) with a syringe pump for 17 h (0.59 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90% B in 30 min. After 27 h, the reaction was stopped with MeOH (75 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.82. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 4:1 yielded 640 mg (62%). $[\alpha]_D^{20}$= + 7.9 (c= 0.79 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_6H_{11}N_3O_5Na$: 228.0596, found 228.0605.

**Table 5.17.** $^1$H and $^{13}$C NMR shifts (ppm) of **30f**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-furanose | 5.07<br>(101.7) | 3.96<br>(80.6) | 4.03<br>(74.5) | 3.99<br>(82.0) | 3.98<br>(69.7) | 3.37<br>(53.1) | 3.28 |
| β-furanose | 5.28<br>(95.5) | 4.03<br>(75.9) | 4.15<br>(74.7) | 4.03<br>(70.2) | 3.86<br>(68.8) | 3.40<br>(53.1) | 3.29 |
| α-pyranose | 4.83<br>(93.1) | 3.29<br>(74.7) | 3.55<br>(71.6) | 3.65<br>(70.2) | 4.14<br>(71.4) | 3.52<br>(50.8) | 3.34 |
| β-pyranose | 4.95<br>(92.4) | 3.54<br>(69.5) | 3.93<br>(69.5) | 3.96<br>(73.2) | 3.48<br>(68.1) | 3.54<br>(48.2) | 3.34 |

### *i. 6-deoxy-6-chloro-D-idose (30g)*

Reaction (25 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the aldehyde (2.5 mmol) triethanolamine buffer (50 mM, pH 8, 25 mL) was added and, finally, FSA A129T /S166G (75 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (5 mmol, 500 mM, 10 mL) with a syringe pump for 10 h (1 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90% B in 30 min. After 52 h, the reaction was stopped with MeOH (25 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 5.0. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 75:25 yielded 54 mg (11%) $[\alpha]_D^{20}$= + 7.9 (c= 0.76 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_6H_{11}O_5ClNa$: 221.0193, found 221.0197.

**Table 5.18.** $^1H$ and $^{13}C$ NMR shifts (ppm) of **30g**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-furanose | 5.06<br>(101.7) | 3.97<br>(80.6) | 4.07<br>(74.4) | 4.08<br>(81.7) | 4.02<br>(75.9) | 3.64<br>(46.0) | 3.64 |
| β-furanose | 5.28<br>(95.5) | 4.02<br>(75.8) | 4.17<br>(74.7) | 4.14<br>(78.0) | 3.89<br>(69.5) | 3.57<br>(45.7) | 3.50 |
| α-pyranose | 4.78<br>(92.7) | 3.27<br>(72.5) | 3.54<br>(71.5) | 3.71<br>(70.3) | 4.12<br>(72.6) | 3.68<br>(41.0) | 3.68 |
| β-pyranose | 4.94<br>(92.6) | 3.53<br>(69.6) | 3.94<br>(69.5) | 3.60<br>(67.8) | 3.97<br>(74.7) | 3.63<br>(42.8) | 3.63 |

### *j. 6,7-dideoxy-D-ido-heptose (31a)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To triethanolamine buffer (50 mM, pH 8, 10 mL) FSA A129T /S166G (20 mg, for Units see Table 5.8) was added. The reaction was started by the fed-batch addition of a mixture of glycolaldehyde and propionaldehyde solution (1 mmol, 400mM, and 0.5

mmol, 200 mM, respectively in 5 mL water) with a syringe pump for 13 h (0.38 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90%B in 30 min. After 48 h, the reaction was stopped with MeOH (15 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.69. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 9:1, yielded 28 mg (31%). $[\alpha]_D^{20}$= + 12.5 (c= 0.66 in MeOH) (lit. [248] $[\alpha]_D^{20}$= + 11.0 (c = 2.50 in $H_2O$). ESI-TOF: calculated for $[M+Na]^+$ $C_7H_{14}O_5Na$: 201.0739, found 201.0744.

**Table 5.19.** $^1H$ and $^{13}C$ NMR shifts (ppm) of **31a**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ | $H_7(C_7)$ |
|---|---|---|---|---|---|---|---|---|
| α-furanose | 5.06 (102.6) | 3.95 (81.0) | 3.97 (76.2) | 3.95 (82.4) | 3.76 (71.8) | 1.62 (25.9) | 1.62 | 0.96 (8.9) |
| β-furanose | 5.34 (96.0) | 3.93 (77.1) | 4.11 (76.6) | 3.98 (80.2) | 3.67 (71.1) | 1.57 (25.6) | 1.48 | 0.96 (9.0) |
| α-pyranose | 4.83 (94.0) | 3.40 (71.4) | 3.70 (70.9) | 3.54 (70.9) | 3.94 (70.9) | 1.62 (21.2) | 1.62 | 0.96 (9.0) |
| β-pyranose | 4.90 (92.5) | 3.51 (70.8) | 3.98 (69.9) | 3.37 (69.3) | 3.69 (75.3) | 1.70 (23.1) | 1.60 | 0.96 (9.6) |

### ***k. 6,7-dideoxy-7-phenyl-D-ido-heptose (31b)***

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. The aldehyde (1 mmol) was dissolved in dimethylformamide (2 mL) To this solution triethanolamine buffer (50 mM, pH 8, 8 mL) was added and, finally, FSA A129T /S166G (30 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (2 mmol, 400mM, 5 mL) with a syringe pump for 12 h (0.42 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 30 to 100% B (TFA (0.095% (*v/v*)) in $CH_3CN$) in 30 min. After 48 h, the reaction was stopped with MeOH (15 mL) to precipitate the enzyme, and the pH of the mixture was adjusted to 4.93. The excess MeOH was removed with rotary evaporator and the mixture was filtered through a nylon filter (0.45 µm) and lyophilized. Purification on silica gel eluted with gradient

$CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 4:1 yielded 118.9 mg (47%). $[\alpha]_D^{20}$= + 16.7 (c= 0.99 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_{13}H_{18}O_5Na$: 277.1052, found 277.1042.

**Table 5.20.** $^1H$ and $^{13}C$ NMR shifts (ppm) of **31b**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ | $H_7(C_7)$ | $H_{7'}$ |
|---|---|---|---|---|---|---|---|---|---|
| α-furanose | 5.08 (102.6) | 3.96 (81.5) | 3.98 (76.2) | 3.97 (84.3) | 3.87 (69.9) | 1.79 (34.7) | 1.79 | 2.77 (31.5) | 2.65 |
| β-furanose | 5.36 (96.1) | 3.95 (77.2) | 4.12 (76.5) | 4.00 (80.6) | 3.79 (69.6) | 1.85 (25.1) | 1.73 | 2.77 (31.4) | 2.65 |
| α-pyranose | 4.89 (92.6) | 3.52 (70.8) | 3.98 (70.0) | 3.33 (70.0) | 3.77 (79.9) | 2.05 (32.3) | 1.76 | 2.77 (31.3) | 2.65 |
| β-pyranose | 4.91 (94.1) | 3.41 (71.5) | 3.71 (71.0) | 3.53 (71.0) | 4.07 (68.7) | 1.93 (30.4) | 1.81 | 2.77 (31.3) | 2.65 |

### *l. 4,5,6-trideoxy-D-idose (31c)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the aldehyde (0.5 mmol) triethanolamine buffer (50 mM, pH 8, 10 mL) was added and, finally, FSA L107Y/A129G (20 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (0.5 mmol, 500 mM, 0.5 mL) with a syringe pump for 2 h (0.254 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90% B in 30 min. After 5 h, the reaction was stopped with MeOH (10 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.89. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 4:1 yielded 29 mg (43%). $[\alpha]_D^{20}$= + 8.9 (c= 1.21 in H2O). ESI-TOF: calculated for $[M+Na]^+$ $C_5H_{10}O_4Na$: 157.0477, found 157.0481.

**Table 5.21.** $^1$H and $^{13}$C NMR shifts (ppm) of **31c**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_{4'}$ | $H_5(C_5)$ | $H_{5'}$ |
|---|---|---|---|---|---|---|---|
| α-pyranose | 4.34<br>(97.5) | 3.03<br>(76.6) | 3.52<br>(71.1) | 1.86<br>(33.1) | 1.54 | 3.84<br>(61.0) | 3.44 |
| β-pyranose | 5.05<br>(93.2) | 3.27<br>(73.8) | 3.82<br>(67.4) | 1.86<br>(32.6) | 1.54 | 3.91<br>(57.5) | 3.54 |

### *m. L-glucose (36)*

Reaction (10 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the glycolaldehyde (1.5 mmol) triethanolamine buffer (50 mM, pH 8, 10 mL) was added. The reaction was started by FSA A165G/S166P (40 mg, for units see Table 5.8). After 7 days of reaction time FSA A129G (20 mg for Units see Table S2) was added to the mixture. Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 20 to 60%B in 16 min. After additional 3 days, the reaction was stopped with MeOH (10 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.91. The MeOH was removed with rotary evaporator and the filtrate was lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 7:3 yielded 36 mg (38%). $[\alpha]_D^{20}$= – 39.7 (c= 0.78 in MeOH) (*ent*-$[\alpha]_D^{20}$= + 44.3 (c= 1.07 in MeOH)) (lit. [249] $[\alpha]_D^{21}$= – 10.8 (c = 1.41 in $H_2O$)) ((lit. [250] $[\alpha]_D^{25}$= – 107.4 (c= 1.05 in $H_2O$)). ESI-TOF: calculated for $[M+Na]^+$ $C_6H_{12}O_6Na$: 203.0532, found 203.0533.

**Table 5.22.** $^1$H and $^{13}$C NMR shifts (ppm) of **36**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-pyranose | 4.49<br>(95.8) | 3.09<br>(74.0) | 3.32<br>(75.8) | 3.25<br>(69.5) | 3.29<br>(65.9) | 3.74<br>(60.7) | 3.59 |
| β-pyranose | 5.07<br>(92.0) | 3.38<br>(71.3) | 3.56<br>(72.6) | 3.27<br>(69.6) | 3.65<br>(71.3) | 3.65<br>(60.5) | 3.67 |

### *n. 6-O-methyl-L-glucose (38a)*

Reaction (50 mL starting volume) was conducted in a screw capped Erlenmeyer flask (100 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the aldehyde (5 mmol) triethanolamine buffer (50 mM, pH 8, 50 mL) was added and, finally, FSA A129G/A165G (200mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (10 mmol, 400mM, 10 mL) with a syringe pump for 18 h (0.55 mL/h flow rate). After 22 h of reaction time FSA A129T (150 mg) was added to the mixture. Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90%B in 30 min. After additional 6.5 h, the reaction was stopped with MeOH (75 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.96. The MeOH was removed with rotary evaporator and the filtrate was lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 4:1 yielded 49 mg (51%). $[\alpha]_D^{20}= -32.6$ (c= 1.28 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_7H_{14}O_6Na$: 217.0688, found 217.0683.

### *o. 6-deoxy-6-chloro-L-glucose (38b)*

Reaction (25 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the aldehyde (2.5 mmol) triethanolamine buffer (50 mM, pH 8, 25 mL) was added and, finally, FSA A129T /A165G (75 mg, for Units see Table 5.8). The reaction was started by the fed-batch addition of glycolaldehyde solution (5 mmol, 500 mM, 10 mL) with a syringe pump for 10 h (1 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 30 to 100% B in 30 min. After 52 h, the reaction was stopped with MeOH (25 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4.82. The MeOH was removed with rotary evaporated and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 75:25 yielded 68 mg (14%). $[\alpha]_D^{20}= -73.6$ (c= 1.80 in MeOH). ESI-TOF: calculated for $[M+Na]^+$ $C_6H_{11}O_5ClNa$: 221.0193, found 221.0194.

**Table 5.23.** $^{1}$H and $^{13}$C NMR shifts (ppm) of **38b**.

| Comp | $H_1(C_1)$ | $H_2(C_2)$ | $H_3(C_3)$ | $H_4(C_4)$ | $H_5(C_5)$ | $H_6(C_6)$ | $H_{6'}$ |
|---|---|---|---|---|---|---|---|
| α-pyranose | 4.54<br>(96.0) | 3.11<br>(74.0) | 3.35<br>(75.3) | 3.36<br>(70.0) | 3.52<br>(74.4) | 3.77<br>(43.8) | 3.67 |
| β-pyranose | 5.08<br>(92.2) | 3.40<br>(71.3) | 3.58<br>(72.4) | 3.37<br>(70.1) | 3.91<br>(70.2) | 3.73<br>(44.3) | 3.73 |

### 5.9.13. Enzymatic aldol reactions with the new donors and the newly designed library of FSA mutants modified in the donor positions

A preliminary set of reactions was conducted to test the obtained mutant proteins (FSA wild-type, FSA A129G, FSA A129S, FSA L107A, FSA L163A, FSA L107A/A129G, FSA L107A/L163A, FSA A129G/L163A, FSA A129S/L163A, FSA L107A/A129G/L163A, FSA L107A/A129S/L163A) with selected nucleophiles. Reactions (200 µL total volume) were performed using 96-deepwell plates stirred with vortex mixer at 1200 rpm, at room temperature. To each donor (0.02 mmol, 100 mM) triethanolamine buffer (50 mM, pH 8) was added (29 µL). To this solution 3-hydroxypropanal solution was added (85 µL, 150 mM from a stock solution of 350 mM). The reactions were started by adding the corresponding FSA variant (50 µL, 2 mg protein, for units see table before from a sock solution of 5 mg protein mL$^{-1}$) dissolved in triethanolamine buffer solution (50 mM, pH 8). Samples (10 µL) were withdrawn at 0, 2, 24 hour reaction times and followed the sample derivatization protocol prior to HPLC analysis.

### 5.9.14. Initial reaction rates of the enzymatic aldol reactions with the new donors with the best mutants

The most productive enzymes were chosen for a second set of reactions to calculate the initial rates of the aldol reaction ($v_0$, µmol min$^{-1}$ mg$^{-1}$). Reactions were carried out on analytical scale as follows. In a 96-deepwell plate (1 mL). To the donor substrates (0.03 mmol, 100 mM) triethanolamine buffer (50 mM, pH 8) was added (75 µL). To this mixture 3-hydroxypropanal solution was added (113 µL, 150 mM from a stock solution of 350 mM). The reactions were initiated by the addition of the corresponding FSA mutant (between 30 µL and 100 µL of protein from stock solutions between 3 to 5 mg protein mL$^{-1}$ to ensure a linear dependence of the product formation versus time at the beginning of the reaction; reaction conversions <10%) in triethanolamine buffer 50 mM, pH 8 (100 µL) and stirred with the vortex mixer at 1200 rpm and 25 ºC. At different reaction times 0-20 min in all cases, samples (10 µL) were withdrawn for measuring the concentration of the aldol products and followed the sample derivatization protocol prior to HPLC analysis. Concentrations of the aldol products were calculated with an external standard method from peak areas.

### 5.9.15. *In situ* NMR competition experiments between 2, 5 and 3-hydroxypropionaldehyde catalyzed by FSA L107A/L163A

The experiments were conducted on a Varian Anova-500 spectrometer. All components were prepared in $D_2O$. Solution of **2**, **5** (0.075 mM, 75-75 mM final concentrations) were prepared in bicarbonate buffer (50 mM, pH 8, 870 µL). To this mixture 3-hydroxypropanal solution was added (113 µL, 150 mM from a stock solution of 750 mM). This mixture was transferred into an NMR tube. The initial spectra were taken and the FSA L107A/L163A mutant (1 mg protein, for the activity see Table S2) was added. The progress of the reaction was monitored by 1H NMR (500 MHz) recording the spectra each 15 for 6 h of reaction time.

### 5.9.16. Competition experiments between 2, 5 and 3-hydroxypropionaldehyde catalyzed by FSA L107A/L163A

A competition reactions were carried out with FSA L107A/L163A to compare the reactivity of **2** and **5** with and 3-hydroxypropanal in the aldol reactions. The reactions (500 µL total volume) were performed using 96-deepwell plates stirred with the vortex mixer at 1200 rpm, at 25ºC. Solution of **2** and **5** (0.075 and 0.075 mM, 75 and 75 mM final concentrations, respectively) was prepared in triethanolamine buffer (50 mM, pH 8, 287 µL). To this mixture 3-hydroxypropanal solution was added (113 µL, 150 mM from a stock solution of 750 mM.) The reactions were started by adding the variant FSA L107A/L163A in solution (100 µL, 0.5 mg protein, from a stock solution of 5 mg protein mL$^{-1}$) for the activity see Table S2. Samples (10 µL) were withdrawn at 0, 5 min, 15 min, 30 min, 2 h and 24 h reaction times, to ensure a linear dependence of the product formation versus time at the beginning of the reaction, and followed the sample derivatization protocol prior to HPLC analysis.

## 5.9.17. Scale up cross-aldol addition reaction of different donors to 3-hydroxypropionaldehyde

General Procedure: Lyophilized fructose 6-phosphate aldolase powder (15 mg, for Units see Table 5.8) was added to a solution (10 mL total reaction volume) containing the respective donor (150 mM) and acceptor (100 mM) components in glycyl-glycine buffer (50 mM, pH 8.5) and the resulting mixture was stirred at r.t. The reaction was monitored at regular intervals by TLC (CM 5:1) and worked up after 24-48 h depending on the rate of consumption of the aldehyde. The individual reaction mixtures were worked up by lyophilization of the crude reaction mixture followed by silica gel column chromatography of the residue using chloroform / methanol (15:1 – 5:1) as the eluent to provide pure product [123].

### *a. D-threo-oct-4-ulose (51a)*

According to General Procedure: Colorless liquid; 75% yield; product ratio: α-anomer (5%):open chain (67%): β-anomer (34%). Open chain: $^1$H NMR (500 MHz, $D_2O$) δ 4.31 – 4.25 (m, 1H), 3.94 – 3.80 (m, 2H), 3.77 – 3.71 (m, 1H), 2.63 (t, J = 7.3 Hz, 1H), 1.91 – 1.54 (m, 6H), 0.90 (td, J = 7.4, 3.9 Hz, 5H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 215.16, 79.24, 68.43, 58.25, 40.37, 35.11, 16.51, 12.85 ppm. β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 3.78 – 3.69 (m, 1H), 1.97 (ddt, *J* = 12.9, 4.9, 1.9 Hz, 1H), 1.91 – 1.54 (m, 6H), 1.46 – 1.33 (m, 1H), 0.90 (td, *J* = 7.4, 3.9 Hz, 5H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 99.45, 74.82, 68.23, 58.52, 39.80, 32.83, 15.64, 13.55 ppm.

### *b. D-threo-non-5-ulose (51b)*

According to General Procedure: Colorless liquid; 76% yield; Product ratio: α-anomer (5%), open chain (54%), β-anomer (41%). Open chain: $^1$H NMR (500 MHz, $D_2O$) δ 4.38 – 4.31 (m, 1H), 3.75 – 3.68 (m, 1H), 2.72 (t, *J* = 7.4 Hz, 1H), 2.04 (ddd, *J* = 6.8, 5.0, 2.2 Hz, 1H), 1.95 – 1.88 (m, 1H), 1.89 – 1.59 (m, 5H), 1.47 – 1.34 (m, 6H), 1.00 – 0.92 (m, 5H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 215.41, 79.30, 68.60, 58.37, 38.27, 35.22, 25.12, 21.68, 13.11 ppm. β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 3.98 – 3.90 (m, 2H), 3.81 (td, *J* = 6.5, 2.5 Hz, 1H), 3.37 (d, *J* = 9.4 Hz, 1H), 2.04 (ddd, *J* = 6.8, 5.0, 2.2 Hz, 1H), 1.95 – 1.88 (m, 1H), 1.89 – 1.59 (m, 5H), 1.47 – 1.34 (m, 6H), 1.00 – 0.92 (m, 5H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 99.60, 74.98, 68.36, 58.61, 37.34, 32.93, 24.39, 22.35, 13.29 ppm.

### *c. D-threo-dec-6-ulose (51c)*

According to General Procedure: Colorless liquid; 50% yield; product ratio: α-anomer (4%): open chain (62%), β-anomer (34%). Open chain: $^1$H NMR (500 MHz, $D_2O$) δ 4.39 – 4.28 (m, 2H), 3.81 (td, *J* = 6.5, 2.3 Hz, 2H), 2.71 (t, *J* = 7.4 Hz, 2H), 1.97 – 1.64 (m, 10H), 1.64 – 1.25 (m, 19H), 1.01 – 0.90 (m, 10H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 215.30, 79.31, 68.64, 58.37, 48.96, 38.52, 35.23, 30.65, 22.67, 13.26 ppm. β-anomer:$^1$H NMR (500 MHz, $D_2O$) δ 4.01 – 3.88 (m, 10H), 3.72 (ddd, *J* = 11.2, 7.0, 4.7 Hz, 6H), 3.37 (d, *J* = 9.4 Hz, 4H), 2.09 – 1.98 (m, 5H), 1.97 – 1.64 (m, 27H), 1.64 – 1.25 (m, 50H), 1.01 – 0.90 (m, 26H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 99.59, 74.99, 68.38, 58.60, 37.63, 32.95, 31.39, 21.90 (d, *J* = 10.2 Hz), 13.36 ppm.

### *d. 2-methyl-D-threo-oct-4-ulose (51f)*

According to General Procedure: Colorless liquid; 28% yield; product ratio: α-anomer (2%), open chain (64%), β-anomer (34%). Open chain: $^1$H NMR (500 MHz, $D_2O$) δ 4.39 – 4.28 (m, 2H), 3.81 (td, *J* = 6.5, 2.3 Hz, 2H), 2.71 (t, *J* = 7.4 Hz, 2H), 1.97 – 1.64 (m, 10H), 1.64 – 1.25 (m, 19H), 1.01 – 0.90 (m, 10H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 215.30, 79.31, 68.64, 58.37, 48.96, 38.52, 35.23, 30.65, 22.67, 13.26 ppm. β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 4.01 – 3.88 (m, 10H), 3.72 (ddd, *J* = 11.2, 7.0, 4.7 Hz, 6H), 3.37 (d, *J* = 9.4 Hz, 4H), 2.09 – 1.98 (m, 5H), 1.97 – 1.64 (m, 27H), 1.64 – 1.25 (m, 50H), 1.01 – 0.90 (m, 26H) ppm. $^{13}$C NMR (126 MHz, $D_2O$) δ 99.59, 74.99, 68.38, 58.60, 37.63, 32.95, 31.39, 21.90 (d, *J* = 10.2 Hz), 13.36 ppm.

### *e. 2,2-dimethyl-D-threo-oct-4-ulose (51i)*

According to General Procedure: Colorless liquid; 25% yield. Product ratio: α-anomer (2%), open chain (66%), β-anomer (%). Open chain: $^1$H NMR (500 MHz, $D_2O$) δ 4.39 – 4.28 (m, 2H), 3.81 (td, *J* = 6.5, 2.3 Hz, 2H), 2.71 (t, *J* = 7.4 Hz, 2H), 1.97 – 1.64 (m, 10H), 1.64 – 1.25 (m, 19H), 1.01 – 0.90 (m, 10H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 215.30, 79.31, 68.64, 58.37, 48.96, 38.52, 35.23, 30.65, 22.67, 13.26 ppm. β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 4.01 – 3.88 (m, 10H), 3.72 (ddd, *J* = 11.2, 7.0, 4.7 Hz, 6H), 3.37 (d, *J* = 9.4 Hz, 4H), 2.09 – 1.98 (m, 5H), 1.97 – 1.64 (m, 27H), 1.64 – 1.25 (m, 50H), 1.01 – 0.90 (m, 26H) ppm. $^{13}$C NMR (126 MHz, $D_2O$) δ 99.59, 74.99, 68.38, 58.60, 37.63, 32.95, 31.39, 21.90 (d, *J* = 10.2 Hz), 13.36 ppm.

### *f. 2-methyl-D-threo-non-5-ulose (51g)*

According to General Procedure: Colorless liquid; 25% yield;product ratio: α-anomer (5%), open chain (54%), β-anomer (41%). Open chain: $^1$H NMR (500 MHz, $D_2O$) δ 4.31 – 4.23 (m, 2H), 4.08 – 3.78 (m, 7H), 3.77 – 3.33 (m, 9H), 2.68 – 2.61 (m, 2H), 1.88 – 1.40 (m, 20H), 1.26 (ddd, *J* = 10.3, 8.5, 6.7 Hz, 6H), 0.88 (dt, *J* = 8.1, 4.1 Hz, 27H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 215.56, 79.33, 68.63, 58.37, 36.65, 35.24, 31.89, 27.13, 21.61 (d, *J* = 8.9 Hz) ppm. β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 4.79 (s, 37H), 3.89 – 3.78 (m, 10H), 3.63 (ddd, *J* = 11.7, 5.1, 1.2 Hz, 6H), 3.29 (d, *J* = 9.4 Hz, 4H), 2.01 – 1.73 (m, 18H), 1.73 – 1.40 (m, 29H), 1.26 (ddd, *J* = 10.3, 8.5, 6.7 Hz, 11H), 0.88 (dt, *J* = 8.1, 4.1 Hz, 53H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 99.72, 75.00, 68.38, 58.62, 35.53, 32.96, 31.16, 27.72, 21.89 (d, *J* = 5.0 Hz) ppm.

### *g. 5-deoxy-1-O-methyl-D-fructose (52a)*

According to General Procedure: Colorless liquid; 89% yield; product ratio: α-anomer (1%), β-anomer (99%). β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 3.99 (td, *J* = 11.6, 5.9 Hz, 4H), 3.88 – 3.65 (m, 5H), 3.58 – 3.42 (m, 11H), 2.07 (dd, *J* = 7.9, 5.1 Hz, 2H), 1.72 (qd, *J* = 12.9, 5.2 Hz, 2H), 1.26 (t, *J* = 7.1 Hz, 1H) ppm. $^{13}$C NMR (126 MHz, $D_2O$) δ 97.78, 73.97, 72.64, 68.19, 58.97, 58.79, 32.85 ppm.

### *h. 5-deoxy-1-O-ethyl-D-fructose (52b)*

According to GP5: Colorless liquid; 82% yield; product ratio: α-anomer (<1%), β-anomer (99%).β-anomer:$^1$H NMR (500 MHz, $D_2O$) δ 4.04 – 3.85 (m, 9H), 3.82 – 3.62 (m, 19H), 3.51 (dd, *J* = 9.9, 1.4 Hz, 7H), 2.10 – 2.00 (m, 4H), 1.72 (ddd, *J* = 24.6, 13.0, 5.3 Hz, 4H), 1.26 (t, *J* = 7.1 Hz, 12H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 97.87, 72.65, 71.74, 68.23, 67.40, 58.73, 32.80, 14.08 ppm.

### *i. 5-deoxy-1-O-propyl-D-fructose (52c)*

According to GP5: Colorless liquid; 28% yield; product ratio: α-anomer (<1%), β-anomer (99%). β-anomer: $^1$H NMR (500 MHz, D2O) δ 4.79 (s, 6H), 4.42 – 3.91 (m, 3H), 16.59 – 3.66 (m, 16H), 3.70 (d, *J* = 10.4 Hz, 1H), 3.67 – 3.53 (m, 1H), 3.50 (dd, *J* = 9.9, 8.5 Hz, 2H), 2.10 – 1.97 (m, 1H), 1.82 – 1.51 (m, 4H), 0.96 (t, *J* = 7.4 Hz, 4H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 97.94 (s), 73.62, 72.66, 71.90, 68.29, 58.72, 32.81, 21.92, 9.75 ppm.

### *j. 5-deoxy-1-O-isopropyl-D-fructose (52d)*

According to GP5: Colorless liquid; 30% yield; product ratio: α-anomer (3%), β-anomer (97%). β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 4.11 – 3.92 (m, 2H), 3.87 – 3.74 (m, 2H), 3.69 (d, *J* = 10.3 Hz, 1H), 3.63 (s, 1H), 3.52 (dd, *J* = 9.9, 4.4 Hz, 1H), 2.05 (ddd, *J* = 6.9, 5.1, 2.5 Hz, 1H), 1.72 (ddd, *J* = 24.5, 13.0, 5.3 Hz, 1H), 1.37 – 1.22 (m, 7H) ppm.$^{13}$C NMR (126 MHz, $D_2O$) δ 98.08, 73.57, 72.73, 69.76, 68.44, 58.81, 32.88, 21.13 (dd, *J* = 22.3, 10.3 Hz) ppm.

### *k. 1-O-allyl-5-deoxy-D-fructose (52e)*

According to GP5: Colorless liquid; 45% yield; product ratio: α-anomer (1%), β-anomer (99%). β-anomer: $^1$H NMR (500 MHz, $D_2O$) δ 6.09 – 5.97 (m, 1H), 5.48 – 5.31 (m, 2H), 4.79 (s, 3H), 4.16 (dd, *J* = 5.9, 1.2 Hz, 2H), 4.04 – 3.92 (m, 2H), 3.83 – 3.70 (m, 2H), 3.52 (dd, *J* = 9.9, 2.0 Hz, 2H), 2.10 – 2.01 (m, 1H), 1.79 – 1.66 (m, 1H) ppm. $^{13}$C NMR (126 MHz, $D_2O$) δ 133.98, 118.49, 98.01, 72.78, 72.51, 71.55, 68.34, 58.87, 32.93 ppm.

## 5.9.18. Enzymatic aldol reactions with (2*S*,3*S*)-3-hydroxy-2-methylpentanal (58) and (2*R*,3*R*)-3-hydroxy-2-methylpentanal (57) and hydroxyacetone and dihydroxyacetone

Reaction (25 mL starting volume) was conducted in a screw capped conical bottom polypropylene tube (50 mL) stirred with a vortex mixer (750 rpm) at 25 °C. To the HA or DHA (2,5 mmol) triethanolamine buffer (50 mM, pH 8, 25 mL) was added and, finally, FSA A129S/A165G/S166G (75 mg for units see Table S2). The reaction was started by the fed-batch addition of (2*S*,3*S*)-3-hydroxy-2-methylpentanal solution in DMF (3 mmol, 800 mM, 8 mL) with a syringe pump for 18 h (0,4 mL/h flow rate). Reaction was monitored by HPLC analysis: Samples (10 µL) were withdrawn and followed the sample derivatization protocol prior to HPLC analysis. Gradient elution was applied: 10 to 90% B in 30 min. After 24 h, the reaction was stopped with MeOH (25 mL) to precipitate the enzyme, and the mixture was filtered through activated charcoal. The pH of the filtrate was adjusted to 4,94. The MeOH was removed with rotary evaporator and lyophilized. Purification on silica gel eluted with gradient $CHCl_3$:EtOAc (1:1)-MeOH from 1:0 to 85:15 yielded 109 mg (**59a**)(23%), and 66 mg (**59b**)(11%).

# 6. References

# 6. References

1. Steinreiber, J., et al., *Synthesis of γ-halogenated and long-chain β-hydroxy-α-amino acids and 2-amino-1,3-diols using threonine aldolases.* Tetrahedron, 2007. **63**(34): p. 8088-8093.
2. Gröger, H.,Y. Asano, *Introduction – Principles and Historical Landmarks of Enzyme Catalysis in Organic Synthesis*, in *Enzyme Catalysis in Organic Synthesis*. 2012, Wiley-VCH Verlag GmbH & Co. KGaA. p. 1-42.
3. Walsh, C., *Enabling the chemistry of life.* Nature, 2001. **409**(6817): p. 226-231.
4. Koeller, K.M.,C.-H. Wong, *Enzymes for chemical synthesis.* Nature, 2001. **409**(6817): p. 232-240.
5. Bornscheuer, U.T., et al., *Engineering the third wave of biocatalysis.* Nature, 2012. **485**(7397): p. 185-194.
6. Wendorff, M., et al., *Catalytic Asymmetric Synthesis: Section 2.2*, in *Asymmetric Synthesis with Chemical and Biological Methods*. 2007, Wiley-VCH Verlag GmbH & Co. KGaA. p. 298-413.
7. Reetz, M.T., *Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future.* J. Am. Chem. Soc., 2013. **135**(34): p. 12480-12496.
8. Buchner, E., Ber. D. Chem. Ges., 1897. **30**: p. 117-124.
9. Wenda, S., et al., *Industrial biotechnology-the future of green chemistry?* Green Chem., 2011. **13**(11): p. 3007-3047.
10. Burton, S.G., D.A. Cowan,J.M. Woodley, *The search for the ideal biocatalyst.* Nat. Biotech., 2002. **20**(1): p. 37-45.
11. Vasic-Racki, D., *History of Industrial Biotransformations – Dreams and Realities*, in *Industrial Biotransformations*. 2006, Wiley-VCH Verlag GmbH & Co. KGaA. p. 1-36.
12. Schoemaker, H.E., D. Mink,M.G. Wubbolts, *Dispelling the Myths--Biocatalysis in Industrial Synthesis.* Science, 2003. **299**(5613): p. 1694-1697.
13. Anastas, P.,N. Eghbali, *Green Chemistry: Principles and Practice.* Chem. Soc. Rev., 2010. **39**(1): p. 301-312.
14. Bolt, A., A. Berry,A. Nelson, *Directed evolution of aldolases for exploitation in synthetic organic chemistry.* Arch. Biochem. Biophys., 2008. **474**(2): p. 318-330.
15. Jaeger, K.-E.,T. Eggert, *Enantioselective biocatalysis optimized by directed evolution.* Curr. Opin. Biotechnol., 2004. **15**(4): p. 305-313.
16. Buchholz, K.,J. Seibel, *Industrial carbohydrate biotransformations.* Carbohydr. Res., 2008. **343**(12): p. 1966-1979.
17. Clapes, P.,W.D. Fessner, *Stereoselective Reactions of Carbonyl and Imino Groups*, in *Science of Synthesis. Stereoselective Synthesis 2.* , G.A. Molander, Editor. 2011, Georg Thieme Verlag KG: Stuttgart (Germany). p. 677.
18. Fessner, W.-D.,V. Helaine, *Biocatalytic synthesis of hydroxylated natural products using aldolases and related enzymes.* Curr. Opin. Biotechnol., 2001. **12**(6): p. 574-586.

19. Illanes, A., *Introduction*, in *Enzyme Biocatalysis*, A. Illanes, Editor. 2008, Springer Netherlands. p. 1-56.
20. Wurtz, C.A., *Sur un aldéhyde-alcool.* C.R. Acad. Sci., 1872. **74**: p. 1361.
21. Borodin, A., *Ueber einen neuen abkdmmling des valerals.* Ber. Dtsch. Chem. Ges., 1873. **6**(

): p. 982-985.
22. Heathcock, C.H., *1.5 - The Aldol Reaction: Acid and General Base Catalysis*, in *Comprehensive Organic Synthesis*, M.T. Editor-in-Chief: Barry and F. Ian, Editors. 1991, Pergamon: Oxford. p. 133-179.
23. Palomo, C., M. Oiarbide,J.M. García, *The Aldol Addition Reaction: An Old Transformation at Constant Rebirth.* Chem. Eur. J., 2002. **8**(1): p. 36-44.
24. Zimmerman, H.E.,M.D. Traxler, *The Stereochemistry of the Ivanov and Reformatsky Reactions. I.* J. Am. Chem. Soc., 1957. **79**(8): p. 1920-1923.
25. Mukaiyama, T., K. Banno,K. Narasaka, *New cross-aldol reactions. Reactions of silyl enol ethers with carbonyl compounds activated by titanium tetrachloride.* J. Am. Chem. Soc., 1974. **96**(24): p. 7503-7509.
26. Mukaiyama, T.,J.-I. Matsuo, *Boron and Silicon Enolates in Crossed Aldol Reaction*, in *Modern Aldol Reactions*. 2008, Wiley-VCH Verlag GmbH. p. 127-160.
27. Braun, M., *Fundamentals and Transition-state Models. Aldol Additions of Group 1 and 2 Enolates*, in *Modern Aldol Reactions*. 2008, Wiley-VCH Verlag GmbH. p. 1-61.
28. Palomo, C., et al., *Design and Evaluation of a Practical Camphor-Based Methyl Ketone Enolate for Highly Stereoselective "Acetate" Aldol Reactions.* Angew. Chem. Int. Ed., 1998. **37**(1-2): p. 180-182.
29. Machajewski, T.D.,C.-H. Wong, *The Catalytic Asymmetric Aldol Reaction.* Angew. Chem. Int. Ed., 2000. **39**(8): p. 1352-1375.
30. Kobayashi, S., Y. Fujishita,T. Mukaiyama, *The Efficient Catalytic Asymmetric Aldol-type Reaction.* Chem. Lett., 1990. **19**(8): p. 1455-1458.
31. Shibasaki, M., S. Matsunaga,N. Kumagai, *Direct Catalytic Asymmetric Aldol Reaction Using Chiral Metal Complexes*, in *Modern Aldol Reactions*. 2008, Wiley-VCH Verlag GmbH. p. 197-227.
32. Trost, B.M.,C.S. Brindle, *The direct catalytic asymmetric aldol reaction.* Chem. Soc. Rev., 2010. **39**(5): p. 1600-1632.
33. Movassaghi, M.,E.N. Jacobsen, *Chemistry. The simplest "enzyme".* Science, 2002. **298**(5600): p. 1904-5.
34. Hajos, Z.G.,D.R. Parrish. 1971, DE 2 102 623: Germany.
35. List, B., R.A. Lerner,C.F. Barbas, *Proline-Catalyzed Direct Asymmetric Aldol Reactions.* J. Am. Chem. Soc., 2000. **122**(10): p. 2395-2396.
36. Markert, M.,R. Mahrwald, *Total Syntheses of Carbohydrates: Organocatalyzed Aldol Additions of Dihydroxyacetone.* Chem. Eur. J., 2008. **14**(1): p. 40-48.

37. Enders, D.,A.A. Narine, *Lessons from nature: Biomimetic organocatalytic carbon-carbon bond formations.* J. Org. Chem., 2008. **73**(20): p. 7857-7870.
38. List, B., P. Pojarliev,C. Castello, *Proline-Catalyzed Asymmetric Aldol Reactions between Ketones and α-Unsubstituted Aldehydes.* Org. Lett., 2001. **3**(4): p. 573-575.
39. Notz, W.,B. List, *Catalytic Asymmetric Synthesis of anti-1,2-Diols.* J. Am. Chem. Soc., 2000. **122**(30): p. 7386-7387.
40. Enders, D.,C. Grondal, *Direct Organocatalytic De Novo Synthesis of Carbohydrates.* Angew. Chem. Int. Ed., 2005. **44**(8): p. 1210-1212.
41. Grondal, C.,D. Enders, *A Direct Organocatalytic Entry to Selectively Protected Aldopentoses and Derivatives.* Adv. Synth. Catal., 2007. **349**(4-5): p. 694-702.
42. Córdova, A., W. Notz,C.F. Barbas, *Proline-Catalyzed One-Step Asymmetric Synthesis of 5-Hydroxy-(2E)-hexenal from Acetaldehyde.* J. Org. Chem., 2001. **67**(1): p. 301-303.
43. Northrup, A.B.,D.W.C. MacMillan, *The First Direct and Enantioselective Cross-Aldol Reaction of Aldehydes.* J. Am. Chem. Soc., 2002. **124**(24): p. 6798-6799.
44. Notz, W., F. Tanaka,C.F. Barbas, *Enamine-Based Organocatalysis with Proline and Diamines: The Development of Direct Catalytic Asymmetric Aldol, Mannich, Michael, and Diels−Alder Reactions.* Accounts of Chemical Research, 2004. **37**(8): p. 580-591.
45. Luo, S., et al., *A Simple Primary−Tertiary Diamine−Brønsted Acid Catalyst for Asymmetric Direct Aldol Reactions of Linear Aliphatic Ketones.* J. Am. Chem. Soc., 2007. **129**(11): p. 3074-3075.
46. Utsumi, N., et al., *Mimicking Aldolases through Organocatalysis: syn-Selective Aldol Reactions with Protected Dihydroxyacetone.* Org. Lett., 2007. **9**(17): p. 3445-3448.
47. Lerner, R., S. Benkovic,P. Schultz, *At the crossroads of chemistry and immunology: catalytic antibodies.* Science, 1991. **252**(5006): p. 659-667.
48. Wagner, J., R.A. Lerner,C.F. Barbas, 3rd, *Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes.* Science, 1995. **270**(5243): p. 1797-800.
49. Tanaka, F.,C.F. Barbas, *Antibody-catalyzed Aldol Reactions*, in *Modern Aldol Reactions*. 2008, Wiley-VCH Verlag GmbH. p. 273-310.
50. Samland, A.,G. Sprenger, *Microbial aldolases as C–C bonding enzymes—unknown treasures and new developments.* Appl Microbiol Biotechnol, 2006. **71**(3): p. 253-264.
51. Dean, S.M., W.A. Greenberg,C.-H. Wong, *Recent Advances in Aldolase-Catalyzed Asymmetric Synthesis.* Adv. Synth. Catal., 2007. **349**(8-9): p. 1308-1320.
52. Meyerhof, O., K. Lohmann,P. Schuster, *Über die Aldolase, ein Kohlenstoff-verknüpfendes Ferment. I. Mitteilung: Aldolkondensation von*

*Dioxyacetonphosphorsaure mit Acetaldehyd.* Biochem. Z., 1936. **286**: p. 301-318.

53. Meyerhof, O., K. Lohmann,P. Schuster, *Über die Aldolase, ein Kohlenstoff-verknüpfendes Ferment. II. Mitteilung: Aldolkondensation von Dioxyacetonphosphorsaure mit Acetaldehyd.* Biochem. Z., 1936. **286**: p. 319-335.
54. Lehninger, A.L.,J. Sice, *Enzymatic Preparation of Two Ketohexose 1-Phosphates.* J. Am. Chem. Soc., 1955. **77**(20): p. 5343-5345.
55. Fessner, W.-D., *Enzyme-catalyzed Aldol Additions*, in *Modern Aldol Reactions*. 2008, Wiley-VCH Verlag GmbH. p. 201-272.
56. Gefflaut, T., et al., *Class I aldolases: Substrate specificity, mechanism, inhibitors and structural aspects.* Prog. Biophys. Mol. Biol., 1995. **63**(3): p. 301-340.
57. Kragl, U., et al., *New synthetic applications of sialic acid aldolase, a useful catalyst for KDO synthesis. Relation between substrate conformation and enzyme stereoselectivity.* J. Chem. Soc., Perkin Trans. 1, 1994(1): p. 119-124.
58. Halcomb, R.L., W. Fitz,C.-H. Wong, *Enzymatic synthesis of 7-deoxy-N-acetylneuraminic acid and 7-O-methyl-N-acetylneuraminic acid.* Tetrahedron: Asymmetry, 1994. **5**(12): p. 2437-2442.
59. Huang, S., H. Yu,X. Chen, *Disaccharides as Sialic Acid Aldolase Substrates: Synthesis of Disaccharides Containing a Sialic Acid at the Reducing End.* Angew. Chem. Int. Ed., 2007. **46**(13): p. 2249-2253.
60. Chokhawala, H.A., et al., *Enzymatic Synthesis of Fluorinated Mechanistic Probes for Sialidases and Sialyltransferases.* J. Am. Chem. Soc., 2007. **129**(35): p. 10630-10631.
61. Chen, X.,A. Varki, *Advances in the Biology and Chemistry of Sialic Acids.* ACS Chem. Biol., 2009. **5**(2): p. 163-176.
62. Sugai, T., et al., *Improved Enzymatic Procedure for a Preparative-Scale Synthesis of Sialic Acid and KDN.* Bull. Chem. Soc. Jpn., 1995. **68**(12): p. 3581-3589.
63. Honda, T., et al., *Synthesis and anti-Influenza virus activity of 7-O-Alkylated derivatives related to zanamivir.* Bioorg. Med. Chem. Lett., 2002. **12**(15): p. 1925-1928.
64. Henderson, D.P., et al., *Stereospecific Preparation of the N-Terminal Amino Acid Moiety of Nikkomycins KX and KZ via a Multiple Enzyme Synthesis.* J. Org. Chem., 1997. **62**(23): p. 7910-7911.
65. Dückers, N., et al., *Threonine aldolases—screening, properties and applications in the synthesis of non-proteinogenic β-hydroxy-α-amino acids.* Appl Microbiol Biotechnol, 2010. **88**(2): p. 409-424.
66. Steinreiber, J., et al., *Threonine aldolases—an emerging tool for organic synthesis.* Tetrahedron, 2007. **63**(4): p. 918-926.

67. Gutierrez, M.L., et al., *Serine Hydroxymethyl Transferase from Streptococcus thermophilus and L-Threonine Aldolase from Escherichia coli as Stereocomplementary Biocatalysts for the Synthesis of β-Hydroxy-α,ω-diamino Acid Derivatives.* Chem. Eur. J., 2008. **14**(15): p. 4647-4656.
68. Vassilev, V.P., et al., *An efficient chemo-enzymatic synthesis of α-amino-β-hydroxy-γ-butyrolactone.* Tetrahedron Lett., 1995. **36**(28): p. 5063-5064.
69. Liu, J.-Q., et al., *Diversity of microbial threonine aldolases and their application.* J. Mol. Catal. B: Enzym., 2000. **10**(1–3): p. 107-115.
70. Kimura, T., et al., *Enzymatic Synthesis of β-Hydroxy-α-amino Acids Based on Recombinant d- and l-Threonine Aldolases.* J. Am. Chem. Soc., 1997. **119**(49): p. 11734-11742.
71. Saeed, A.,D.W. Young, *Synthesis of L-β-hydroxyaminoacids using serine hydroxymethyltransferase.* Tetrahedron, 1992. **48**(12): p. 2507-2514.
72. Shibata, K., et al., *Kinetic and thermodynamic control of L-Threonine aldolase catalyzed reaction and its application to the synthesis of mycestericin D.* Tetrahedron Lett., 1996. **37**(16): p. 2791-2794.
73. Tanaka, T., et al., *Design and Synthesis of Peptide Mimetics of GDP-Fucose: Targeting Inhibitors of Fucosyltransferases.* Synlett, 2004. **2004**(EFirst): p. 0243-0246.
74. Liu, J.Q., et al., *A new route to l-threo-3-[4-(methylthio)phenylserine], a key intermediate for the synthesis of antibiotics: recombinant low-specificity d-threonine aldolase-catalyzed stereospecific resolution.* Appl Microbiol Biotechnol, 1999. **51**(5): p. 586-591.
75. Baik, S.-H.,H. Yoshioka, *Enhanced synthesis of L-threo-3,4-dihydroxyphenylserine by high-density whole-cell biocatalyst of recombinant L-threonine aldolase from Streptomyces avermitilis.* Biotechnol. Lett., 2009. **31**(3): p. 443-448.
76. Fesko, K., et al., *Four types of threonine aldolases: similarities and differences in kinetics/thermodynamics.* J. Mol. Catal. B: Enzym., 2008. **52-3**: p. 19-26.
77. Barbas, C.F., Y.F. Wang,C.H. Wong, *Deoxyribose-5-phosphate aldolase as a synthetic catalyst.* J. Am. Chem. Soc., 1990. **112**(5): p. 2013-2014.
78. Gijsen, H.J.M.,C.-H. Wong, *Unprecedented Asymmetric Aldol Reactions with Three Aldehyde Substrates Catalyzed by 2-Deoxyribose-5-phosphate Aldolase.* J. Am. Chem. Soc., 1994. **116**(18): p. 8422-8423.
79. Liu, J.,C.-H. Wong, *Aldolase-Catalyzed Asymmetric Synthesis of Novel Pyranose Synthons as a New Entry to Heterocycles and Epothilones.* Angew. Chem. Int. Ed., 2002. **41**(8): p. 1404-1407.
80. Greenberg, W.A., et al., *Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates.* Proc. Nat. Acad. Sci. USA, 2004. **101**(16): p. 5788-5793.
81. Jennewein, S., et al., *Directed evolution of an industrial biocatalyst: 2-deoxy-D-ribose 5-phosphate aldolase.* Biotechnol. J., 2006. **1**(5): p. 537-548.

82. Hall, D.R., et al., *Structure of Tagatose-1,6-bisphosphate Aldolase: INSIGHT INTO CHIRAL DISCRIMINATION, MECHANISM, AND SPECIFICITY OF CLASS II ALDOLASES.* J. Biol. Chem., 2002. **277**(24): p. 22018-22024.
83. Maliakel, B.P.,W. Schmid, *Chemo-enzymatic synthesis of natural products: synthesis of sphydrofuran.* Tetrahedron Lett., 1992. **33**(23): p. 3297-3300.
84. Zannetti, M.T., et al., *Fructose 1,6-Bisphosphate Aldolase from Staphylococcus carnosus: Overexpression, Structure Prediction, Stereoselectivity, and Application in the Synthesis of Bicyclic Sugars.* Chem. Eur. J., 1999. **5**(6): p. 1882-1890.
85. Fessner, W.-D., et al., *Diastereoselective Enzymatic Aldol Additions: L-Rhamnulose and L-Fuculose 1-Phosphate Aldolases from E. coli.* Angew. Chem. Int. Ed. Engl., 1991. **30**(5): p. 555-558.
86. Espelt, L., et al., *Stereoselective aldol additions catalyzed by dihydroxyacetone phosphate dependent aldolases in emulsion systems: preparation and structural characterization of linear and cyclic aminopolyols from aminoaldehydes.* Chem. Eur. J., 2003. **9**(20): p. 4887-4899.
87. Fessner, W.-D.,C. Walter, *Enzymatic C-C bond formation in asymmetric synthesis.* Top. Curr. Chem., 1997. **184**(Bioorganic Chemistry): p. 97-194.
88. Durrwachter, J.R., et al., *Enzymic aldol condensation/isomerization as a route to unusual sugar derivatives.* J. Am. Chem. Soc., 1986. **108**(24): p. 7812-7818.
89. Fessner, W.-D., et al., *Short Enzymatic Synthesis of L-Fucose Analogs.* Eur. J. Org. Chem., 2000. **2000**(1): p. 125-132.
90. Li, Z., et al., *Enzymatic synthesis of d-sorbose and d-psicose with aldolase RhaD: Effect of acceptor configuration on enzyme stereoselectivity.* Bioorg. Med. Chem. Lett., 2011. **21**(23): p. 7081-7084.
91. Eyrisch, O.,W.-D. Fessner, *Disaccharide Mimetics by Enzymatic Tandem Aldol Additions.* Angew. Chem. Int. Ed. Engl., 1995. **34**(15): p. 1639-1641.
92. Chou, W.-C., et al., *A New Route to Deoxythio sugars Based on Aldolases.* J. Am. Chem. Soc., 1994. **116**(14): p. 6191-6194.
93. Witczak, Z.,J. Culhane, *Thiosugars: new perspectives regarding availability and potential biochemical and medicinal applications.* Appl Microbiol Biotechnol, 2005. **69**(3): p. 237-244.
94. Chênevert, R., M. Lavoie,M. Dasser, *Use of aldolases in the synthesis of non-carbohydrate natural products. Stereoselective synthesis of aspicilin C-3—C-9 fragment.* Can. J. Chem., 1997. **75**(1): p. 68-73.
95. Clapés, P., G.A. Sprenger,J. Joglar, *Novel Strategies in Aldolase-Catalyzed Synthesis of Iminosugars*, in *Modern Biocatalysis*. 2009, Wiley-VCH Verlag GmbH & Co. KGaA. p. 299-311.
96. Winchester, B.G., *Iminosugars: from botanical curiosities to licensed drugs.* Tetrahedron: Asymmetry, 2009. **20**(6–8): p. 645-651.

97. Calveras, J., et al., *Dihydroxyacetone phosphate aldolase-catalyzed synthesis of structurally diverse polyhydroxylated pyrrolidine derivatives and evaluation of their glycosidase inhibitory properties.* Chem. Eur. J., 2009. **15**: p. 7310-7328.

98. Mitchell, M.L., L.V. Lee,C.-H. Wong, *Synthesis and evaluation of six-membered GDP-iminocyclitol.* Tetrahedron Lett., 2002. **43**(32): p. 5691-5693.

99. Fesko, K.,M. Gruber-Khadjawi, *Biocatalytic Methods for C C Bond Formation.* ChemCatChem, 2013. **5**(6): p. 1248-1272.

100. Espelt, L., et al., *Aldol additions of dihydroxyacetone phosphate to N-Cbz-amino aldehydes catalyzed by L-fuculose-1-phosphate aldolase in emulsion systems: Inversion of stereoselectivity as a function of the acceptor aldehyde.* Chem. Eur. J., 2005. **11**(5): p. 1392-1401.

101. Calveras, J., et al., *Chemoenzymatic Synthesis and Inhibitory Activities of Hyacinthacines A1 and A2 Stereoisomers.* Adv. Synth. Catal., 2007. **349**(10): p. 1661-1666.

102. Concia, A.L., et al., *Chemo-enzymatic synthesis and glycosidase inhibitory properties of DAB and LAB derivatives.* Org. Biomol. Chem., 2013. **11**(12): p. 2005-2021.

103. Von der Osten, C.H., et al., *Use of a recombinant bacterial fructose-1,6-diphosphate aldolase in aldol reactions: preparative syntheses of 1-deoxynojirimycin, 1-deoxymannojirimycin, 1,4-dideoxy-1,4-imino-D-arabinitol, and fagomine.* J. Am. Chem. Soc., 1989. **111**(11): p. 3924-3927.

104. Fessner, W.-D.,G. Sinerius, *Synthesis of Dihydroxyacetone Phosphate (and Isosteric Analogues) by Enzymatic Oxidation; Sugars from Glycerol.* Angew. Chem. Int. Ed. Engl., 1994. **33**(2): p. 209-212.

105. Schümperli, M., R. Pellaux,S. Panke, *Chemical and enzymatic routes to dihydroxyacetone phosphate.* Appl Microbiol Biotechnol, 2007. **75**(1): p. 33-45.

106. Guanti, G., L. Banfi,M.T. Zannetti, *Phosphonic derivatives of carbohydrates: chemoenzymatic synthesis.* Tetrahedron Lett., 2000. **41**(17): p. 3181-3185.

107. Iturrate, L., et al., *Substrate channelling in an engineered bifunctional aldolase/kinase enzyme confers catalytic advantage for C-C bond formation.* Chem. Commun., 2009(13): p. 1721-1723.

108. Schoevaart, R., F. Van Rantwijk,R.A. Sheldon, *Facile Enzymatic Aldol Reactions with Dihydroxyacetone in the Presence of Arsenate.* J. Org. Chem., 2001. **66**(13): p. 4559-4562.

109. Drueckhammer, D.G., et al., *Reversible and in situ formation of organic arsenates and vanadates as organic phosphate mimics in enzymatic reactions: mechanistic investigation of aldol reactions and synthetic applications.* J. Org. Chem., 1989. **54**(1): p. 70-77.

110. Sugiyama, M., et al., *Borate as a Phosphate Ester Mimic in Aldolase-Catalyzed Reactions: Practical Synthesis of L-Fructose and L-Iminocyclitols.* Adv. Synth. Catal., 2006. **348**(18): p. 2555-2559.
111. Garrabou, X., et al., *Highly efficient aldol additions of DHA and DHAP to N-Cbz-amino aldehydes catalyzed by l-rhamnulose-1-phosphate and l-fuculose-1-phosphate aldolases in aqueous borate buffer.* Org. Biomol. Chem., 2011. **9**(24): p. 8430-8436.
112. Schurmann, M.,G.A. Sprenger, *Fructose-6-phosphate aldolase is a novel class I aldolase from Escherichia coli and is related to a novel group of bacterial transaldolases.* The Journal of biological chemistry, 2001. **276**(14): p. 11055-61.
113. Samland, A.K., et al., *The transaldolase family: new synthetic opportunities from an ancient enzyme scaffold.* Chembiochem : a European journal of chemical biology, 2011. **12**(10): p. 1454-74.
114. Sánchez-Moreno, I., et al., *FSAB: A new fructose-6-phosphate aldolase from Escherichia coli. Cloning, over-expression and comparative kinetic characterization with FSAA.* J. Mol. Catal. B: Enzym., 2012. **84**(0): p. 9-14.
115. Schneider, S., et al., *Replacement of a phenylalanine by a tyrosine in the active site confers fructose-6-phosphate aldolase activity to the transaldolase of Escherichia coli and human origin.* The Journal of biological chemistry, 2008. **283**(44): p. 30064-72.
116. Thorell, S., et al., *Crystal Structure of Decameric Fructose-6-Phosphate Aldolase from Escherichia coli Reveals Inter-subunit Helix Swapping as a Structural Basis for Assembly Differences in the Transaldolase Family.* J. Mol. Biol., 2002. **319**(1): p. 161-171.
117. Samland, A.K.,G.A. Sprenger, *Transaldolase: From biochemistry to human disease.* Int. J. Biochem. Cell Biol., 2009. **41**(7): p. 1482-1494.
118. Schörken, U., et al., *Identification of catalytically important residues in the active site of Escherichia coli transaldolase.* Eur. J. Biochem., 2001. **268**(8): p. 2408-2415.
119. Jia, J., et al., *Crystal structure of the reduced Schiff-base intermediate complex of transaldolase B from escherichia coli: Mechanistic implications for class I aldolases.* Protein Sci., 1997. **6**(1): p. 119-124.
120. Lehwess-Litzmann, A., et al., *Twisted Schiff base intermediates and substrate locale revise transaldolase mechanism.* Nat. Chem. Biol., 2011. **7**(10): p. 678-684.
121. Castillo, J.A., et al., *A Mutant D-Fructose-6-Phosphate Aldolase (Ala129Ser) with Improved Affinity towards Dihydroxyacetone for the Synthesis of Polyhydroxylated Compounds.* Advanced Synthesis & Catalysis, 2010. **352**(6): p. 1039-1046.
122. Garrabou, X., et al., *Asymmetric Self- and Cross-Aldol Reactions of Glycolaldehyde Catalyzed by D-Fructose-6-phosphate Aldolase.* Angewandte Chemie International Edition, 2009. **48**(30): p. 5521-5525.

123. Rale, M., et al., *Broadening Deoxysugar Glycodiversity: Natural and Engineered Transaldolases Unlock a Complementary Substrate Space.* Chem. Eur. J., 2011. **17**(9): p. 2623-2632.

124. Schürmann, M., M. Schürmann,G.A. Sprenger, *Fructose 6-phosphate aldolase and 1-deoxy-d-xylulose 5-phosphate synthase from Escherichia coli as tools in enzymatic synthesis of 1-deoxysugars.* J. Mol. Catal. B: Enzym., 2002. **19–20**(0): p. 247-252.

125. Sugiyama, M., et al., *d-Fructose-6-Phosphate Aldolase-Catalyzed One-Pot Synthesis of Iminocyclitols.* J. Am. Chem. Soc., 2007. **129**(47): p. 14811-14817.

126. Meyer, H.-P., et al., *The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC).* Catal. Sci. Technol., 2013. **3**(1): p. 29-40.

127. Garrabou, X., et al., *Asymmetric self- and cross-aldol reaction of glycolaldehyde catalyzed by D-fructose-6-phosphate aldolase.* Angew. Chem. Int. Ed., 2009. **48**: p. 5521-5525.

128. Clapés, P.,X. Garrabou, *Current Trends in Asymmetric Synthesis with Aldolases.* Adv. Synth. Catal., 2011. **353**(13): p. 2263-2283.

129. Schneider, S., et al., *Redesigning the active site of transaldolase TalB from Escherichia coli: new variants with improved affinity towards nonphosphorylated substrates.* Chembiochem : a European journal of chemical biology, 2010. **11**(5): p. 681-90.

130. Concia, A.L., et al., *D-Fructose-6-phosphate Aldolase in Organic Synthesis: Cascade Chemical-Enzymatic Preparation of Sugar-Related Polyhydroxylated Compounds.* Chemistry – A European Journal, 2009. **15**(15): p. 3808-3816.

131. Castillo, J.A., et al., *Fructose-6-phosphate aldolase in organic synthesis: preparation of D-fagomine, N-alkylated derivatives, and preliminary biological assays.* Org. Lett., 2006. **8**(26): p. 6067-6070.

132. Gutierrez, M., et al., *Structure-guided redesign of d-fructose-6-phosphate aldolase from E. coli: remarkable activity and selectivity towards acceptor substrates by two-point mutation.* Chemical Communications, 2011. **47**(20): p. 5762-5764.

133. Gómez, L., et al., *d-Fagomine lowers postprandial blood glucose and modulates bacterial adhesion.* Br. J. Nutrit., 2012. **107**(12): p. 1739-1746.

134. Toscano, M.D., K.J. Woycechowsky,D. Hilvert, *Minimalist Active-Site Redesign: Teaching Old Enzymes New Tricks.* Angew. Chem. Int. Ed., 2007. **46**(18): p. 3212-3236.

135. Stemmer, W.P., *DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution.* Proc. Nat. Acad. Sci. USA, 1994. **91**(22): p. 10747-10751.

136. Drauz, K.,H. Waldmann, *Enzyme catalysis in organic synthesis: A comprehensive handbook, Volume III, 2nd Edition*. Vol. III. 2002, Weinheim: Wiley-VHC Verlag GmbH. 568.

137. Cadwell, R.C.,G.F. Joyce, *Randomization of genes by PCR mutagenesis.* PCR methods and applications, 1992. **2**(1): p. 28-33.

138. Williams, G.J., et al., *Creation of a pair of stereochemically complementary biocatalysts.* J. Am. Chem. Soc., 2006. **128**(50): p. 16238-16247.

139. Hsu, C.C., et al., *Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase.* Proc. Nat. Acad. Sci. USA, 2005. **102**(26): p. 9122-9126.

140. Bornscheuer, U.T.,M. Pohl, *Improved biocatalysts by directed evolution and rational protein design.* Curr. Opin. Chem. Biol., 2001. **5**(2): p. 137-143.

141. Heine, A., et al., *Observation of covalent intermediates in an enzyme mechanism at atomic resolution.* Science, 2001. **294**(5541): p. 369-74.

142. DeSantis, G., et al., *Structure-Based mutagenesis approaches toward expanding the substrate specificity of d-2-Deoxyribose-5-phosphate aldolase.* Bioorg. Med. Chem., 2003. **11**(1): p. 43-52.

143. Garrabou, X., et al., *Structure-guided minimalist redesign of L-fuculose-1-phosphate aldolase active site. Expedient synthesis of novel polyhydroxylated pyrrolizidines and their inhibitory properties against glycosidases and intestinal disaccharidases* Chem. Eur. J., 2010. **16**: p. 10691-10706.

144. Woodhall, T., et al., *Creation of a Tailored Aldolase for the Parallel Synthesis of Sialic Acid Mimetics.* Angewandte Chemie International Edition, 2005. **44**(14): p. 2109-2112.

145. Garrabou, X., et al., *Redesign of the Phosphate Binding Site of L-Rhamnulose- 1-Phosphate Aldolase towards a Dihydroxyacetone Dependent Aldolase.* Adv. Synth. Catal., 2011. **353**(1): p. 89-99.

146. Fong, S., et al., *Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars.* Chem. Biol., 2000. **7**(11): p. 873-883.

147. Lamble, H.J., et al., *Metabolic pathway promiscuity in the archaeon Sulfolobus solfataricus revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase.* The Journal of biological chemistry, 2003. **278**(36): p. 34066-72.

148. Lamble, H.J., et al., *Promiscuity in the part-phosphorylative Entner–Doudoroff pathway of the archaeon Sulfolobus solfataricus.* FEBS Lett., 2005. **579**(30): p. 6865-6869.

149. Lamble, H.J., et al., *Engineering stereocontrol into an aldolase-catalysed reaction.* Chem. Commun., 2005(1): p. 124-126.

150. Royer, S.F., et al., *Structurally Informed Site-Directed Mutagenesis of a Stereochemically Promiscuous Aldolase To Afford Stereochemically*

*Complementary Biocatalysts.* J. Am. Chem. Soc., 2010. **132**(33): p. 11753-11758.
151. Hao, J.,A. Berry, *A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents.* Protein engineering, design & selection : PEDS, 2004. **17**(9): p. 689-97.
152. Seebeck, F.P.,D. Hilvert, *Conversion of a PLP-Dependent Racemase into an Aldolase by a Single Active Site Mutation.* J. Am. Chem. Soc., 2003. **125**(34): p. 10158-10159.
153. Nicolaou, K.C.,H.J. Mitchell, *Adventures in Carbohydrate Chemistry: New Synthetic Technologies, Chemical Synthesis, Molecular Design, and Chemical Biology.* Angew. Chem. Int. Ed., 2001. **40**(9): p. 1576-1624.
154. Weymouth-Wilson, A.C., *The role of carbohydrates in biologically active natural products.* Nat. Prod. Rep., 1997. **14**(2): p. 99-110.
155. Bertozzi, C.R.,L.L. Kiessling, *Chemical glycobiology.* Science, 2001. **291**(5512): p. 2357-2364.
156. Hanessian, S., *Preparative carbohydrate chemistry*. 1997, New York: Marcel Dekker. xiii, 648 p.
157. Mahrwald, R., *Diastereoselection in Lewis-Acid-Mediated Aldol Additions.* Chem. Rev., 1999. **99**(5): p. 1095-1120.
158. Kazmaier, U., *Amino Acids—Valuable Organocatalysts in Carbohydrate Synthesis.* Angew. Chem. Int. Ed., 2005. **44**(15): p. 2186-2188.
159. Nicolaou, K.C., D.J. Edmonds,P.G. Bulger, *Cascade Reactions in Total Synthesis.* Angew. Chem. Int. Ed., 2006. **45**(43): p. 7134-7186.
160. Ricca, E., B. Brucher,J.H. Schrittwieser, *Multi-Enzymatic Cascade Reactions: Overview and Perspectives.* Adv. Synth. Catal., 2011. **353**(13): p. 2239-2262.
161. Mlynarski, J.,B. Gut, *Organocatalytic synthesis of carbohydrates.* Chem. Soc. Rev., 2012. **41**(2): p. 587-596.
162. Garcia-Junceda, E., *Multi-step enzyme catalysis : biotransformations and chemoenzymatic synthesis*. 2008, Weinheim

Chichester: Wiley-VCH ;

John Wiley distributor. xv, 241 p.
163. Mu, W., et al., *Recent advances on applications and biotechnological production of d-psicose.* Appl Microbiol Biotechnol, 2012. **94**(6): p. 1461-1467.
164. Granström, T., K. Izumori,M. Leisola, *A rare sugar xylitol. Part II: biotechnological production and future applications of xylitol.* Appl Microbiol Biotechnol, 2007. **74**(2): p. 273-276.
165. Levin, G.V., et al., *Sugar substitutes: their energy values, bulk characteristics, and potential health benefits.* The American journal of clinical nutrition, 1995. **62**(5 Suppl): p. 1161S-1168S.

166. Northrup, A.B., et al., *Enantioselective Organocatalytic Direct Aldol Reactions of α-Oxyaldehydes: Step One in a Two-Step Synthesis of Carbohydrates.* Angew. Chem. Int. Ed., 2004. **43**(16): p. 2152-2154.

167. Northrup, A.B.,D.W.C. MacMillan, *Two-Step Synthesis of Carbohydrates by Selective Aldol Reactions.* Science, 2004. **305**(5691): p. 1752-1755.

168. Córdova, A., et al., *Amino Acid Catalyzed Neogenesis of Carbohydrates: A Plausible Ancient Transformation.* Chem. Eur. J., 2005. **11**(16): p. 4772-4784.

169. Casas, J., et al., *Direct Amino Acid Catalyzed Asymmetric Synthesis of Polyketide Sugars.* Angew. Chem. Int. Ed., 2005. **44**(9): p. 1343-1345.

170. Borysenko, C.W., et al., *The synthesis of aldose sugars from half-protected dialdehydes using rabbit muscle aldolase.* J. Am. Chem. Soc., 1989. **111**(26): p. 9275-9276.

171. Alajarin, R., E. Garcia-Junceda,C.-H. Wong, *A Short Enzymic Synthesis of L-Glucose from Dihydroxyacetone Phosphate and L-Glyceraldehyde.* J. Org. Chem., 1995. **60**(13): p. 4294-4295.

172. Wong, C.-H., et al., *Enzymic Synthesis of L-Fucose and Analogs.* J. Org. Chem., 1995. **60**(22): p. 7360-7363.

173. Gijsen, H.J.M.,C.-H. Wong, *Sequential Three- and Four-Substrate Aldol Reactions Catalyzed by Aldolases.* Journal of the American Chemical Society, 1995. **117**(29): p. 7585-7591.

174. Mukherjee, S., et al., *Asymmetric Enamine Catalysis.* Chem. Rev., 2007. **107**(12): p. 5471-5569.

175. Sorensen, E.J.,G.M. Sammis, *A Dash of Proline Makes Things Sweet.* Science, 2004. **305**(5691): p. 1725-1726.

176. Fullerton, S.W.B., et al., *Mechanism of the Class I KDPG aldolase.* Bioorg. Med. Chem., 2006. **14**(9): p. 3002-3010.

177. Thorell, S., et al., *Crystal structure of decameric fructose-6-phosphate aldolase from Escherichia coli reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family.* J. Mol. Biol., 2002. **319**(1): p. 161-171.

178. Neuhaus, F.C., *The Enzymatic Synthesis of d-Alanyl-d-alanine : II. KINETIC STUDIES ON d-ALANYL-d-ALANINE SYNTHETASE.* J. Biol. Chem., 1962. **237**(10): p. 3128-3135.

179. Kajimoto, T., et al., *Palladium-mediated stereocontrolled reductive amination of azido sugars prepared from enzymic aldol condensation: a general approach to the synthesis of deoxy aza sugars [Erratum to document cited in CA115(13):136547j].* J. Am. Chem. Soc., 1991. **113**(23): p. 9009.

180. Pederson, R.L.,C.-H. Wong, *Enzymic aldol condensation as a route to heterocycles: synthesis of 1,4-dideoxy-1,4-imino-D-arabinitol, fagomine, 1-deoxynojirimycin and 1-deoxymannojirimycin.* Heterocycles, 1989. **28**(1): p. 477-80.

181. Hecht, M.-L., et al., *Recent advances in carbohydrate-based vaccines.* Curr. Opin. Chem. Biol., 2009. **13**(3): p. 354-359.

182. Wong, C.-H.,Editor, *Carbohydrate-Based Drug Discovery, Volume 1*. 2003. 459 pp.

183. Wong, C.-H.,Editor, *Carbohydrate-Based Drug Discovery, Volume 2*. 2003. 947 pp.

184. Hasehira, K., et al., *Development of a chemical strategy to produce rare aldohexoses from ketohexoses using 2-aminopyridine.* Carbohydr. Res., 2011. **346**(17): p. 2693-2698.

185. Gumina, G., G.-Y. Song,C.K. Chu, *l-Nucleosides as chemotherapeutic agents.* FEMS Microbiol. Lett., 2001. **202**(1): p. 9-15.

186. Butlerow, A., *Bildung einer zuckerartigen Substanz durch Synthese.* Justus Liebigs Annalen der Chemie, 1861. **120**(3): p. 295-298.

187. Loew, O., *Weiteres über die Condensation des Formaldehyds.* J. Prakt. Chem., 1886. **34**(1): p. 51-55.

188. Gabel, N.W.,C. Ponnamperuma, *Model for Origin of Monosaccharides.* Nature, 1967. **216**(5114): p. 453-455.

189. Kofoed, J., J.-L. Reymond,T. Darbre, *Prebiotic carbohydrate synthesis: zinc-proline catalyzes direct aqueous aldol reactions of [small alpha]-hydroxy aldehydes and ketones.* Org. Biomol. Chem., 2005. **3**(10): p. 1850-1855.

190. Lambert, J.B., S.A. Gurusamy-Thangavelu,K. Ma, *The Silicate-Mediated Formose Reaction: Bottom-Up Synthesis of Sugar Silicates.* Science, 2010. **327**(5968): p. 984-986.

191. Granström, T.B., et al., *Izumoring: A novel and complete strategy for bioproduction of rare sugars.* J. Biosci. Bioeng., 2004. **97**(2): p. 89-94.

192. Bednarski, M.D., H.J. Waldmann,G.M. Whitesides, *Aldolase-catalyzed synthesis of complex C8 and C9 monosaccharides.* Tetrahedron Lett., 1986. **27**(48): p. 5807-5810.

193. Bednarski, M.D., et al., *Rabbit muscle aldolase as a catalyst in organic synthesis.* J. Am. Chem. Soc., 1989. **111**(2): p. 627-635.

194. Concia, A.L., et al., *D-Fructose-6-phosphate aldolase in organic synthesis: cascade chemical-enzymatic preparation of sugar-related polyhydroxylated compounds.* Chem. Eur. J., 2009. **15**(15): p. 3808-3816.

195. Schoevaart, R., F. van Rantwijk,R.A. Sheldon, *Class I fructose-1,6-bisphosphate aldolases as catalysts for asymmetric aldol reactions.* Tetrahedron: Asymmetry, 1999. **10**(4): p. 705-711.

196. Fessner, W.-D., et al., *Enzymes in organic synthesis. Part 1. Diastereoselective, enzymatic aldol addition with L-rhamnulose- and L-fuculose-1-phosphate aldolases from E. coli.* Angew. Chem. Int. Ed., 1991. **30**(5): p. 555-558.

197. Stevens, R.V., et al., *Further studies on the utility of sodium hypochlorite in organic synthesis. Selective oxidation of diols and direct conversion of aldehydes to esters.* Tetrahedron Lett., 1982. **23**(45): p. 4647-4650.

198. Bonini, C., et al., *New Functionalised Hydroxymethyl Ketones from the Mild and Chemoselective KMnO4 Oxidation of Chiral Terminal Olefins.* Eur. J. Org. Chem., 2006. **2006**(1): p. 80-83.

199. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal. Biochem., 1976. **72**(1–2): p. 248-254.

200. Simmons, B., A.M. Walji,D.W.C. MacMillan, *Cycle-Specific Organocascade Catalysis: Application to Olefin Hydroamination, Hydro-oxidation, and Amino-oxidation, and to Natural Product Synthesis.* Angew. Chem. Int. Ed., 2009. **48**(24): p. 4349-4353.

201. Huang, Y., et al., *Enantioselective Organo-Cascade Catalysis.* J. Am. Chem. Soc., 2005. **127**(43): p. 15051-15053.

202. Gijsen, H.J.M.,C.-H. Wong, *Sequential Three- and Four-Substrate Aldol Reactions Catalyzed by Aldolases.* J. Am. Chem. Soc., 1995. **117**(29): p. 7585-91.

203. Herrmann, G.F., et al., *A new multi-enzyme system for a one-pot synthesis of sialyl oligosaccharides: Combined use of β-galactosidase and α(26)-sialyltransferase coupled with regeneration in situ of CMP-sialic acid.* Tetrahedron Lett., 1993. **34**(19): p. 3091-3094.

204. Northrup, A.B., et al., *Enantioselective organocatalytic direct aldol reactions of alpha-oxyaldehydes: Step one in a two-step synthesis of carbohydrates.* Angew. Chem. 2004, 116, 2204 –2206, 2004. **43**(16): p. 2152-2154.

205. Northrup, A.B.,D.W.C. MacMillan, *Two-Step Synthesis of Carbohydrates by Selective Aldol Reactions.* Science, 2004. **305**: p. 1752-1755.

206. Zhao, G.-L., et al., *Dynamic Kinetic Asymmetric Transformation (DYKAT) by Combined Amine- and Transition-Metal-Catalyzed Enantioselective Cycloisomerization.* Chem. Eur. J., 2010. **16**(5): p. 1585-1591.

207. Edin, M., J.-E. Bäckvall,A. Córdova, *Tandem enantioselective organo- and biocatalysis: a direct entry for the synthesis of enantiomerically pure aldols.* Tetrahedron Lett., 2004. **45**(41): p. 7697-7701.

208. Rulli, G., et al., *Direction of Kinetically versus Thermodynamically Controlled Organocatalysis and Its Application in Chemoenzymatic Synthesis.* Angew. Chem. Int. Ed., 2011. **50**(34): p. 7944-7947.

209. Baer, K., et al., *Sequential and Modular Synthesis of Chiral 1,3-Diols with Two Stereogenic Centers: Access to All Four Stereoisomers by Combination of Organo- and Biocatalysis.* Angew. Chem. Int. Ed., 2009. **48**(49): p. 9355-9358.

210. Kinnell, A., et al., *Development of an organo- and enzyme-catalysed one-pot, sequential three-component reaction.* Tetrahedron, 2012. **68**(37): p. 7719-7722.

211. Graber, M., et al., *Microbial polysaccharides containing 6-deoxysugars.* Enzyme Microb. Technol., 1988. **10**(4): p. 198-206.

212. Liu, H.,J.S. Thorson, *Pathways and Mechanisms in the Biogenesis of Novel Deoxysugars by Bacteria.* Annual Review of Microbiology, 1994. **48**(1): p. 223-256.

213. Trefzer, A., A. Bechthold,J. A. Salas, *Genes and enzymes involved in deoxysugar biosynthesis in bacteria.* Nat. Prod. Rep., 1999. **16**(3): p. 283-299.

214. Thibodeaux, C.J., C.E. Melançon,H.-w. Liu, *Natural-Product Sugar Biosynthesis and Enzymatic Glycodiversification.* Angew. Chem. Int. Ed., 2008. **47**(51): p. 9814-9859.

215. Boysen, M.M.K., *Carbohydrates as Synthetic Tools in Organic Chemistry.* Chem. Eur. J., 2007. **13**(31): p. 8648-8659.

216. Mendez, C., et al., *Deoxysugars in bioactive natural products: development of novel derivatives by altering the sugar pattern.* Current topics in medicinal chemistry, 2008. **8**(8): p. 710-24.

217. Johnson, D.A.,H.-w. Liu, *Mechanisms and pathways from recent deoxysugar biosynthesis research.* Curr. Opin. Chem. Biol., 1998. **2**(5): p. 642-649.

218. Thibodeaux, C.J., C.E. Melancon,H.-w. Liu, *Unusual sugar biosynthesis and natural product glycodiversification.* Nature, 2007. **446**(7139): p. 1008-1016.

219. Font, D., et al., *Toward an Artificial Aldolase.* Org. Lett., 2007. **10**(2): p. 337-340.

220. Gutierrez, M., et al., *Structure-guided redesign of* D*-fructose-6-phosphate aldolase from E. coli: remarkable activity and selectivity towards acceptor substrates by two-point mutation.* Chem. Commun., 2011. **47**(20): p. 5762-5764.

221. Giacalone, F., et al., *Low-loading asymmetric organocatalysis.* Chem. Soc. Rev., 2012. **41**(6): p. 2406-2447.

222. Mase, N.,I.C.F. Barbas, *In water, on water, and by water: mimicking nature's aldolases with organocatalysis and water.* Org. Biomol. Chem., 2010. **8**(18): p. 4043-4050.

223. Mlynarski, J.,S. Bas, *Catalytic asymmetric aldol reactions in aqueous media - a 5 year update.* Chem. Soc. Rev., 2014. **43**(2): p. 577-587.

224. Zheng, L., U. Baumann,J.-L. Reymond, *An efficient one-step site-directed and site-saturation mutagenesis protocol.* Nucleic Acids Res., 2004. **32**(14): p. e115.

225. Bimboim, H.C.,J. Doly, *A rapid alkaline extraction procedure for screening recombinant plasmid DNA.* Nucleic Acids Res., 1979. **7**(6): p. 1513-1523.

226. J. Sambrook, E.F.a.M.F., *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 1989.

227. Hanahan, D., *Studies on transformation of Escherichia coli with plasmids.* J. Mol. Biol., 1983. **166**(4): p. 557-580.

228. Altschul, S.F., et al., *Basic local alignment search tool.* J. Mol. Biol., 1990. **215**(3): p. 403-410.

229. Schrödinger, *Suite 2013 Update 2*. 2013, Schrödinger, LLC: New York, NY.

230. Schrödinger, *Maestro, version 9.5*. 2013, Schrödinger, LLC: New York, NY.

231. Schrödinger, *MacroModel, version 10.1*. 2013, Schrödinger, LLC: New York, NY.

232. Jorgensen, W.L., D.S. Maxwell,J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids.* J. Am. Chem. Soc., 1996. **118**(45): p. 11225-11236.

233. Still, W.C., et al., *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics.* J. Am. Chem. Soc., 1990. **112**(16): p. 6127-6129.

234. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res., 2000. **28**(1): p. 235-242.

235. Schrödinger, *Protein Preparation Wizard 2013-2; Epik version 2.4, Schrödinger, LLC, New York, NY, 2013; Impact version 5.9, Schrödinger, LLC, New York, NY, 2013; Prime version 3.2, Schrödinger, LLC, New York, NY, 2013.* 2013.

236. Sastry, G.M., et al., *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.* Journal of computer-aided molecular design, 2013. **27**(3): p. 221-34.

237. Olsson, M.H.M., et al., *PROPKA3: Consistent Treatent of Internal and Surface Residues in Empirical pKa predictions.* J. Chem. Theor. Comput., 2011. **7**: p. 525-537.

238. Kolossvary, I.,W.C. Guida, *Low-mode conformational search elucidated: application to C39H80 and flexible docking of 9-deazaguanine inhibitors into PNP.* J. Comput. Chem., 1999. **20**(15): p. 1671-1684.

239. Ryckaert, J.-P., G. Ciccotti,H.J.C. Berendsen, *Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes.* J. Comput. Phys., 1977. **23**: p. 327-341.

240. Schrödinger, *QSite, version 6.1*. 2013, Schrödinger, LLC: New York, NY.

241. Laemmli, U.K., *Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-685.

242. Simpson, R.J., *Proteins and Proteomics: A Laboratory Manual.* Cold Spring Harbor: NY, USA, 2003.

243. Frigerio, M., M. Santagostino,S. Sputore, *A User-Friendly Entry to 2-Iodoxybenzoic Acid (IBX).* J. Org. Chem., 1999. **64**(12): p. 4537-4538.

244. de Almeida, C.G., et al., *Synthesis and Antibacterial Activity of Aromatic and Heteroaromatic Amino Alcohols.* Chem. Biol. Drug D., 2011. **78**(5): p. 876-880.

245. Pathak, A.K., et al., *A facile method for deprotection of trityl ethers using column chromatography.* Tetrahedron Lett., 2001. **42**(44): p. 7755-7757.

246. Davies, S.G., R.L. Nicholson,A.D. Smith, *A SuperQuat glycolate aldol approach to the asymmetric synthesis of hexose monosaccharides.* Org. Biomol. Chem., 2005. **3**(2): p. 348-359.

247. Vargha, L., *Constitution of benzal-D-sorbitol. A synthesis of L-xylose.* Berichte der Deutschen Chemischen Gesellschaft [Abteilung] B: Abhandlungen, 1935. **68B**: p. 18.
248. Walls Jr, H.,D.E. Kiely, *The synthesis of 6,7-dideoxy-d-gluco-heptose and 6,7-dideoxy-l-ido-heptose.* Carbohydr. Res., 1973. **31**(2): p. 397-400.
249. Ko, S.Y., et al., *Total synthesis of the L-hexoses.* Tetrahedron, 1990. **46**(1): p. 245-264.
250. Weymouth-Wilson, A.C., et al., *Large scale synthesis of the acetonides of l-glucuronolactone and of l-glucose: easy access to l-sugar chirons.* Tetrahedron Lett., 2009. **50**(46): p. 6307-6310.