TECHNISCHE UNIVERSITÄT MÜNCHEN

Integrative Research Center Campus für Biotechnologie und Nachhaltigkeit

Lehrstuhl für Chemie Biogener Rohstoffe

Development of decarboxylases and dehydratases as valuable biocatalysts for the production of fine chemicals

Samuel Sutiono

Vollständiger Abdruck der von der promotionsführenden Einrichtung Campus Straubing für Biotechnologie und Nachhaltigkeit der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr.-Ing. Jakob Burger
Prüfer der Dissertation:	1.	Prof. Dr. Volker Sieber
	2.	Prof. Dr. Bastian Blombach
	3.	Prof. Dr. Sonja-Verena Albers

Die Dissertation wurde am 26.06.2020 bei der Technischen Universität München eingereicht und von der promotionsführenden Einrichtung Campus Straubing für Biotechnologie und Nachhaltigkeit am 20.11.2020 angenommen

Wahrlich es ist nicht das Wissen, sondern das Lernen, nicht das Besitzen, sondern das Erwerben, nicht das Dasein, sondern das Hinkommen, was den größten Genuss gewährt.

A Letter from Carl Friedrich Gauß to Wolfgang Bolyai, Göttingen, 2 September 1808.

Table of Contents

List of Publications

1. **Sutiono, S.**, Carsten, J. & Sieber, V. Structure-Guided Engineering of α-Keto Acid Decarboxylase for the Production of Higher Alcohols at Elevated Temperature. *ChemSusChem* **11**, 3335–3344 (2018).

2. **Sutiono, S.**, Satzinger, K., Pick, A., Carsten, J. & Sieber, V. To beat the heat – engineering of the most thermostable pyruvate decarboxylase to date. *RSC Adv.* **9**, 29743–29746 (2019).

3. **Sutiono, S.**, Teshima, M., Beer, B., Schenk, G. & Sieber, V. Enabling the direct enzymatic dehydration of D-glycerate to pyruvate as the key step in synthetic enzyme cascades used in the cell-free production of fine chemicals. *ACS Catal.* **10**, 3110–3118 (2020).

4. **Sutiono, S.**, Siebers, B. & Sieber, V. Characterization of highly active 2-keto-3-deoxy-L-arabinonate and 2-keto-3-deoxy-D-xylonate dehydratases in terms of biotransformation of hemicellulose sugars to chemicals. Applied Microbiology and Biotechnology. DOI: 10.1007/s00253-020-10742-5

Summary

First-generation biofuels and chemicals are produced from starch-containing crops, thus having strong competition with food production. In contrast, second-generation biofuels and chemicals are derived from non-edible biomass. Consolidated bioprocessing (CBP) is a strategy to perform biomass degradation, sugars fermentation, and product separation in one vessel, which has been developed with the primary goal to increase the feasibility of the second-generation approach. CBP strategy requires the development of an ideal microorganism that is capable to degrade biomass and metabolize the released sugars to produce chemicals of interest. Thermophillic microorganisms have been proposed as an ideal chasis due to their higher growth temperatures than mesophilic microorganisms. High temperatures (>50 °C) are prevailing conditions for biomass degradation. Higher temperatures is also beneficial in reducing the risk of contamination and assisting in the product removal, particularly for products with high vapor pressures. In a previous study, an *in vitro* metabolic engineering for the production of isobutanol and ethanol was already reported. From this work, two non-thermostable α-keto acid decarboxylases were used that have no reported thermophilic homologues. Therefore, improving thermostability of the two decarboxylases is a prerequisite for the success of bioalcohols production at elevated temperatures.

The first part of this study was focused on the first decarboxylase, the branched-chain α-keto acid decarboxylase (KDC). This enzyme catalyzes the decarboxylation of α-ketoisovalerate (KIV) to isobutyraldehyde, the penultimate step in the isobutanol pathway. Two previous works on one KDC isoenzyme from *Lactococcus lactis* (*Ll*KivD) using directed evolution coupled with *in silico* engineering failed to create any variants with an increased thermostability required for our thermoacidophilic microorganism of interest, *Sulfolobus acidocaldarius* (*Saci*), i.e. stable at 70 °C. By application of a structure-guided engineering approach in this study, the kinetic and thermodynamic stabilities of wild type (WT) *Ll*KdcA (another KDC isoenzyme) were able to be improved by 14.8 °C and 13.5 °C, respectively. The engineered variant (7M.D) showed a half-life ($t_{1/2}$) at 70 °C of 2 h. The variant was also able to withstand 4 vol% of isobutanol at 50 °C by showing a $t_{1/2}$ of 14.3 h (>600-fold increase in comparison to the WT). At this condition, no known microbial production of isobutanol has ever been reported; thus, the *in vitro* approach might serve as a promising alternative. With the improved stability in isobutanol, 7M.D will play a key role to drive forward the *in vitro* isobutanol biosynthesis.

In the second part of this study, another decarboxylase namely pyruvate decarboxylase (PDC), was studied. This enzyme catalyzes the decarboxylation of pyruvate to acetaldehyde, a common step in ethanol production in yeast and ethanologenic bacteria. In contrast to the works on KDC, two previous attempts to increase thermostability of PDC were done by rational designs (ancestral sequence reconstruction and Rosetta-based engineering). Neither of them yielded in a significant improvement of enzyme performance. Hence, in this work a directed evolution approach was employed. By a method called staggered-extension process (StEP), three DNAs encoding bacterial PDCs were shuffled by means of polymerase chain reaction (PCR). A significantly improved variant (PDC-Var.1) was obtained during screening. Introducing one beneficial substitution obtained from the previous study of *Ll*KdcA into PDC-Var.1 resulted in further increase in its stability (PDC-Var.2). The most thermostable variant showed half-lives at 70 °C and 75 °C of 10.7 h and 7.3 h, respectively or >500-fold and >2000-fold improvements in comparison with the most thermostable WT PDC. In addition to the production of ethanol, PDC has also been utilized in cell-free

syntheses of n-butanol. In the presence of 9 vol% n-butanol at 50°C (solubility limit of n-butanol in water), PDC-Var. 2 showed a half-life of 15 h (>5000-fold increase as compared with the WT PDC). Thus, PDC.Var.2 will be a suitable candidate in the future development of the cell-free biosyntheses of n-butanol.

After optimizing the two decarboxylases, this study was extended to [Fe-S]-dependent dehydratases. This enzyme is a key enzyme catalyzing three reactions in the previous cell-free approach to produce ethanol and isobutanol. One of the three reactions catalyzed, i.e. the dehydration of D-glycerate to pyruvate, also serves as a major bottleneck in the cascades. Prior to this study, a dihydroxy-acid dehydratase (DHAD) from *S. solftaricus* catalyzed this step with a k_{cat} of 0.01 s^{-1} and a total turnover number (TTN) of 900. To be industrially competitive, an enzyme should have at least a k_{cat} of 1 s^{-1} and a TTN of 10^4 to 10^5. In this study, no engineering approaches were employed as previous attempts to increase activity of *Ss*DHAD were proven to be very challenging. Instead, a sequence-based discovery approach was applied. Utilizing the vast number of genomes available, several novel [Fe-S]-dehydratases were discovered. From the sequence alignment and the substrate profiling, two distinct classes of [Fe-S]-dehydratase were identified. One enzyme in particular, a promiscuous sugar acid dehydratase from *Paralcaligenes ureilyticus* (*Pu*DHT), demonstrated the highest proficiency toward the dehydration of D-glycerate with a k_{cat} of 1.4 s^{-1} and a TTN of 9 x 10^4. *Pu*DHT also demonstrated a very high activity toward D-gluconate—the other sugar acid needed to be dehydrated in the cell-free approach—with a k_{cat} of 136 s^{-1} and a TTN of 10^7. Application of *Pu*DHT in two different cell-free approaches to produce ethanol from D-glucose and pyruvate from glycerol showed 10-fold and 5-fold increase in the production rate in comparison to the systems utilizing *Ss*DHAD.

In the final study, other classes of dehydratases, L-2-keto-3-deoxyarabinonate and D-2-keto-3-deoxyxylonate dehydratases (L-KdpD and D-KdpD) were focused. These two enzymes catalyze the third step of the Weimberg pathways—dehydrating L-KDP and D-KDP, respectively to α-ketoglutarate semialdehyde. The Weimberg pathways have gained increased attentions in the past years, particularly toward the utilization of hemicellulose sugars (L-arabinose and D-xylose) to produce chemicals. However, detailed characterizations of the dehydratases (L-KdpD and D-KdpD) were still lacking. Thus, in this study several novel dehydratases were cloned and characterized in detail. From this study, it was demonstrated that L-KdpD from *Cupriavidus necator* (*Cn*L-KdpD) and D-KdpD from *Pseudomonas putida* (*Pp*D-KdpD) as the most promising variants from each enzyme class concerning activity and kinetic stability. *Cn*L-KdpD and *Pp*D-KdpD are specific to their stereoisomer, they were not able to dehydrate the non-preferred stereoisomer. Both enzymes were able to catalyze the conversion of 500 mM L- and D-KDP reaching >95% and >90% conversion, respectively. Combination of both enzymes was able to dehydrate 500 mM racemic D,L-KDP reaching >80% conversion. These two enzymes are then expected to play an important role in the development of the Weimberg pathways for chemicals production.

In conclusion, four different lyases: α-keto acid decarboxylases, [Fe-S]-dehydratases, L-KdpDs, and D-KdpDs were studied. Engineering approaches were employed to increase the performance of the two decarboxylases, whereas sequence-based discovery was utilized to discover highly active dehydratases ([Fe-S]-dehydratases, L-KdpDs, and D-KdpDs). All the enzymes studied are anticipated to advance the utilization of lyases as valuable biocatalysts to produce second-generation biofuel and chemicals *in vivo* and *in vitro*.

Acknowledgements

First, I would like to express my gratitude to Prof. Volker Sieber who gave me the opportunity to pursue my PhD studies at his Chair. Your scientific guidance and belief in me were the main reasons I could finish this lifetime journey. Your continuous support and encouragement for me to think creatively and analytically have become a lifelong lesson.

Secondly, I would like to thank Prof. Bastian Blombach and Prof. Sonja Albers as my thesis examiners for your time and invaluable feedbacks and discssions that made this work better. I also would like to thank Prof. Jakob Burger as the Chairman of my thesis committee who ensured my defense went faultlessly amid the COVID-19 pandemic. Not to forget to my mentor, Prof. Robert Kourist, thank you for all valuable advices throughout my PhD studies.

To Dr. Jörg Carsten, thank you for the fruitful discussion. Moreover, my special recognitions to André Pick who always invigorated me with new ideas and inspired me to give 110% for science. I am waiting one day when you come and tell me you run out of ideas.

Furthermore, to my colleagues who I now proudly call the best people in Straubing: Edilberto V. Medina-Cabrera, Samed Güner, Ioannis Zachos, and Robert Genth. Without your company staying until late in the lab even sometimes on the weekend, I would be easily demotivated in the middle of this journey. Thank you to all of favors you did for me.

My special thanks to Ms. Elisabeth Aichner for having always greeted me with smile whenever I dropped by your office and put me at ease as you took care all of my non-lab matters. My gratitude also went to Ms. Petra Lommes, Ms. Anja Schmidt, Mr. Manuel Döring, and Ms. Magdalena Haslbeck. Thank you Petra and Anja for your invaluable introductions to the use of analytical instruments and always being there whenever I had problems. Thank you Manuel for your assistance in using automatic devices to screen many of my enzyme libraries. Thank you Megi for your patient and help to always make sure I handled plasmids and strains correctly.

During my studies I had the opportunity to work with so many great colleagues who I learned one thing or another. I thank Hendrik Hohagen, Marius Rüttering, Daniel Bauer, and Lisa Steiner for helping me inside and outside the lab since my Master thesis until their departure. Special thanks to Dr. Sumanth Ranganathan who proofread this work. Dr. Broder Ruhman, Dr. Barbara Beer, Dr. Josef Sperl, thank you for your time and advices during my PhD studies. To my present Red Lab mates: Vivian P. Willers, Mariko Teshima, and Tatjana Laudage your energy and excitement have made Red Lab to be my second home. To Gerassimos Kolaitis, thank you for the hospitality and fruitful discussions whenever I visited our lab in Garching.

It was also a great pleasure to have the chance to supervise and work with very talented Ms. Katharina Satzinger. Your diligent work not only resulted in our publication together, but also helped me to broaden my research interest. In addition, to the students whom I did not work with

them directly, but have significantly brightened the lab atmosphere: Ms. Nicole Schmeltzer, Ms. Arabella Essert, Ms. Daniela Zeitler, and Ms. Marlene Brünner, thank you all for the great and fun time. To Ms. Lisa Bock, thank you for the memes that keep me sane. Furthermore, to all diligent students who worked as HiWi at our chair, thank you for your assistances.

To my family, my parents: Inge Natalia and Wahyudi Sutiono and my little sister: Maria A. Sutiono; without your unconditional supports and love, this journey, starting from my Master studies in Sweden until finishing my PhD studies in Germany would have not been possible. To my sister from different country: Kim Haeyoung, your friendliness and countless invitations to Korean dinner were, without a doubt, among the top 10 reasons I could finish my PhD studies.

Last but not least, to two of my dearest friends: Vincensius S. P. Oetama and Ivan K. Santoso. Although we did not communicate on daily basis, your support when I needed the most have always been the tipping point to bounce me back to life.

1 Introduction

1.1 Enzymes for industrial applications

1.1.1 Enzyme as biocatalyst

Since the first discovery of a diastase by a French chemist Anselme Payen in the early 19 century and the cell-free fermentation of alcohol using yeast cell extract by a German chemist Eduard Buchner in the early 20 century, our understanding of enzymes has undergone massive development, in particular in the last half century.[1,2] From literatures and patents published in the field of enzymes, at least there were two major events that advanced enzyme studies rapidly (Figure 1.1). The first occurred in the late 1950s and the second in the late 1980s. There were so many seminal works between the 1950s and 1960s. However, during that period enzyme studies were carried out with the main focus to explain biochemistry.[3,4] The cause of the second rapid development in the late 1980s was probably accounted to the development and commercialization of the Polymerase Chain Reaction (PCR) technology by Kary Mullis in the early 1980s. This new technology—supported by the recombinant DNA technology that had been developed roughly a decade earlier—allowed scientists at that time to produce enzymes much easier and faster.[5,6]

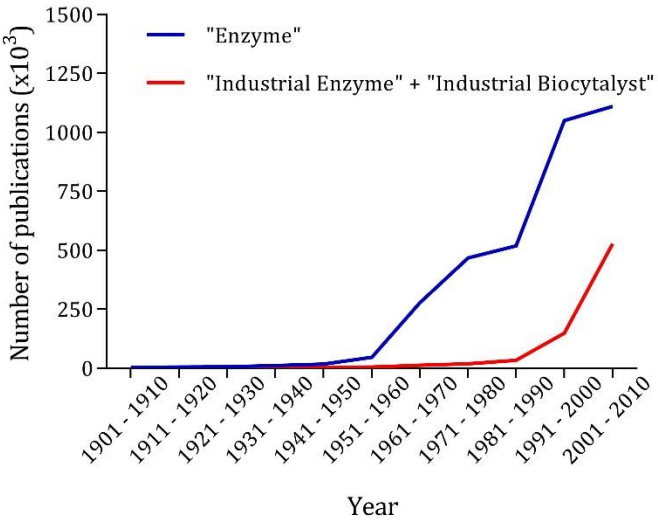


Figure 1.1. Scientific publications and patents published in every decade based on the keywords searching in Google Scholar. Keyword "Enzyme" was used as the representative of enzyme studies in general. Keywords "Industrial Enzyme" and "Industrial Biocatalyst" was used to represent the use of enzymes for industrial applications.

Enzyme applications in industry dates back almost 150 years, when a Danish chemist, Christian Hansen industrially produced rennet from calves or lambs stomach for milk coagulation, an important step in the cheese manufacturing. The main active ingredient of rennet is chymosin, a protease.[7] Pectinases and amylases have been two important enzymes for food industries since the early 20th century. Another notable industrial enzyme application was the utilization of an acylase for resolution of amino acids in the late 1960s by Tanabe Seiyaku, a Japanese company.[8]

Although the use of enzymes in industry has started quite some time ago, an unprecedented development took place only about two decades ago (Figure 1.1). Perhaps, the work of Frances Arnold introducing the concept of directed evolution to engineer enzymes´ characteristics was one of the catalysts to this development.[9,10] Prior to this era, mostly wild type enzymes, i.e. enzymes that are found in nature were used. With directed evolution developed in the early 1990s, any laboratory or company with a PCR machine could engineer enzymes within days.

1.1.1.1 Advantages and challenges of biocatalysts´ applications in industry

It is generally recognized that biocatalysts have several advantages over traditional chemical catalysts. In order to limit the discussion in this work, a biocatalyst is defined as *an enzyme that is used in vitro in purified or partially purified form as well as introduced to a foreign or wild type (WT) host to catalyze one or more chemical conversions of substrate(s) to desired product(s)*. There are at least three main advantages in using biocatalysts:[11,12]

Selectivity

As a protein, the biocatalyst has a rather intricate three-dimensional (3D) structure relative to its catalytic active site(s). This configuration allows an enzyme to catalyze reactions with a high degree of selectivity.[13] Three major types of enzyme selectivities are chemoselectivity, regioselectivity, and stereoselectivity.[14,15]

(I) **Chemoselectivity** is the ability of a biocatalyst to do a chemical reaction—introduce or break a chemical bond—to a specific functional group in the presence of different groups, e.g. selective amination of a C-H group of sulfonyl azides by a cytochrome P450 over a competing reduction to a sulfonamide.[16]

(II) **Regioselectivity** is the ability of a biocatalyst to carry out a chemical reaction in a specific region of a molecule, e.g. dehydration of D-glucaric acid to 5-keto-4-deoxy glucaric acid.[17]

(III) **Stereoselectivity** is the ability of a biocatalyst to perform a chemical reaction to yield a chiral compound from an achiral substrate or to recognize only one chiral molecule from racemic substrates. Oxidation of glycerol to D-glyceric acid by an alditol oxidase is an instance for the first case and selective hydrolysis of L-acylated amino acid in a racemic mixture to yield optically active L-amino acid by an acylase is for the latter case.[8,18]

Ample conditions

Since enzymes are produced by living (micro)organisms, most of them show functional activity in similar conditions of the microorganisms from which the enzymes originated. Most microorganisms thrive in the moderate conditions (20 to 45 °C, pH 6.5 to 7.5, ~1 atm). Microorganisms found in this condition are called mesophiles. The activity of enzymes at ambient condition allows bioconversion of a substrate carried out at ambient temperatures and

pressures.[13,19] Chemical catalysis, on the other hand, are usually performed at much higher temperatures and pressures.[20] High temperatures and pressures are energy intensive, thus less environmentally friendly.[14] Moreover, such processes possess greater risks meaning if there is an incident, the impact would be much worse compared to processes performed at ample temperatures and pressures. Ethanol production is a very good example. Ethanol can be chemically produced from the hydration of ethylene at 300 °C and 60 to 70 atm.[21] In contrast, fermentation of sugar to ethanol using yeast are performed at 30 °C and 1 atm.

Renewable and biodegradable

Catalysts are defined as molecules that help to speed up chemical reactions but not being consumed during the reactions. Thus, in theory catalysts can be used indefinitely. However, in practice catalysts have a limited lifespan before being inactivated. Some catalysts can be recycled and others are not possible; hence the necessity to synthesize new ones. It is not uncommon chemical catalysts consist of or are synthesized using rare metals.[20] These molecules are mined from the Earth´s crust making their availability limited.[13] Biocatalysts are produced by (micro)organisms; thus, they are renewable. Another aspect of being renewable is that the production of biocatalysts can be done in any part of the world and are not affected by geopolitical conditions. Rare metals, however, are only produced by some countries; thus, geopolitical conditions will play a big role in determining their prices and availability.[22] Competition with other industries, such as automotive and electronic industries will also influence the prices of rare metals.[13]

However, as enzymes have evolved over millions of years to sustain life, several traits make enzymes not readily suitable for industrial applications, such as:[11,12,23]

Compatibility with current production routes

It is undeniable that the field of biochemistry was developed much later in comparison to other chemistry sub disciplines, such as organic, inorganic, and physical chemistry. This resulted in many processes in chemical industries being established with the latter three as the backbones. One typical characteristic of such processes is the prevalent use of organic solvents.[13] Enzymes, on the other hand, are evolved to function in aqueous solutions. In consequence, most enzymes barring a few exceptions are not compatible with organic solvents.[24] Convincing industries to change the already established processes to work with aqueous solutions is not an easy task. This is understandable, because chemical reactions in aqueous solutions usually have poor yields. Low substrate solubility is also a major issue.[25] The latter issue could be addressed partially by increasing reaction temperatures. However, enzymes from mesophilic microorganisms would not be suitable to perform reactions at high temperatures.

Substrate or product inhibitions

As the main purpose of enzymes is to sustain life, enzymes are often cleverly regulated. One of prime examples of enzyme regulation is enzyme inhibition. Some enzymes can be inhibited by their product, some by their substrate, and some others by intermediates in biochemical pathways. This characteristic is important for metabolism in organisms. However, this could turn into a practical challenge in the industrial perspective.[26,27] It is always desired that catalysts are highly active. Inhibition of biocatalysts either by their substrates, or products, or intermediates implies slower reaction, hence lower productivity.

1.1.1.2 A good industrial biocatalyst

To be suitable for use in chemical industries, there are at least one or more characteristics described below that a biocatalyst should have[11,27,28]

High activity

Biocatalyst activity is defined as the amount of biocatalyst needed to generate certain amount of product per given time. A higher activity of a biocatalyst means less amount of the biocatalyst is required to generate the same amount of a product at a given time. As a biocatalyst needs to be produced or purchased, the amount of a biocatalyst used will add up to the production cost of a compound. For a case where the cost of a biocatalyst is almost negligible to the price of the final product, productivity of a process can be increased by simply adding more biocatalyst. However, this would not always be practical. High concentration of a biocatalyst used could lead to the decrease of mass transfer rate, thus overall process productivity.[19] As a biocatalyst is also a polymer (polypeptide), high concentrations of biocatalyst could lead to foam formation, which has been reported to deactivate many enzymes. High protein usage would also increase the complexity of the subsequent downstream processes.[29] Therefore, a highly active biocatalyst is almost always preferred.

High stability

Often industrial processes are more hostile than the biocatalysts´ normal operating environment in nature.[29] High temperature, extreme pH, high concentration of organic solvents are commonly used in industrial processes.[26] The stability of a biocatalyst will determine its lifetime or usability. When a biocatalyst is completely inactivated, a new biocatalyst must be added. Therefore, the use of a more stable biocatalyst in an industrial process will result in a lower production cost. There are three different parameters to represent the stability of a biocatalyst.[30] Thermodynamic stability relates to the intrinsic nature of an enzyme, i.e. folded or unfolded state. This parameter does not concern the activity of the enzyme. Kinetic stability indicates how a biocatalyst retain its activity over a period at a given condition. This parameter needs to be determined empirically and is often difficult to predict from the protein sequence. Process stability relates to the kinetic

stability, but this parameter is measured at a very specific process condition. This parameter is used to determine how much product can be generated from a given amount of biocatalyst until the biocatalyst is completely inactivated. All three parameters are important, but the last two have more practical applications, such as calculating production costs, designing (bio)reactors, and scaling up.[19] Thermodynamic stability could be a good indicator to kinetic or process stability, but quite often these parameters show no clear correlation.[30] Therefore, precaution is necessary in comparing thermodynamic stability data among biocatalysts.

High selectivity

High selectivity is one of the distinct features of biocatalysts that make them superior to chemical catalysts. This is true for many enzyme classes toward their natural substrates. However, biocatalysts are often used in industry to convert a compound, which is not its natural substrate. In this sense, the selectivity of the enzyme may be not at the same level as to its natural substrate.[29] Additionally, several enzymes naturally show broad substrate scopes. This type of enzymes is known as promiscuous enzymes.[31] When a selective conversion is intended, a biocatalyst with high degree of selectivity is required to perform the reaction without the necessity to add protecting groups. Adding a protecting group is to be avoided whenever possible as this approach is tedious and may generate more waste. The use of protecting groups will also ultimately increase the final production cost as well.[19]

High protein yield

The importance of this parameter is sometimes easily overlooked. The reason could be that this parameter is influenced by many factors and often hard to predict which factor is more important. At least factors as nucleotide sequences, enzyme class, expression system, and fermentation conditions will have an influence on the enzyme yield. Nevertheless, it is safe to argue that higher enzyme yield, either produced inside cells or expressed in a medium, implies a lower biocatalyst production cost, because more enzyme can be produced from the same production process.[32] However, there is exception to this rule: for instance, when it is preferred to produce an enzyme as inclusion bodies to refold them later into active biocatalysts.[33]

1.1.2 Strategies for industrial biocatalyst utilization

It is now clear that a biocatalyst is a more environmentally friendly catalyst than a chemical catalyst. However, a decision in which catalyst would be used in an industrial process cannot and perhaps will never be only be based on this factor alone. Production cost in general is the main consideration dictating the choice of a catalyst.[14,29] As discussed previously in the challenges and criteria of suitable biocatalysts for industrial applications, several strategies have been developed in recent years to promote the use of biocatalysts in chemical industries. Generally, these strategies can be divided into two major approaches.[10,11] The first approach is to adjust an

enzyme to be suitable for industrial processes and the second approach is to modify an industrial process to be enzyme compatible.

1.1.2.1 Tailoring an enzyme to fit an industrial process

This approach seems to be the simplest and the most straightforward one. A biocatalyst is modified to fit an already developed process in chemical industries. At least three strategies that can been applied are, [23]

Enzyme immobilization

Many enzymes with desired activity and selectivity suffer from poor stability in industrial processes. Immobilization techniques were developed to tackle this problem. Several immobilization techniques developed in recent years have improved enzymes´ tolerance toward hostile industrial processes.[34] Immobilization also allows easier biocatalyst separation from a final product. Separation of biocatalysts may permit the biocatalysts to be reused for the next process as well as reduce the production costs. Although immobilization is known to improve stability, it is not unusual to find that this comes at the expense of activity.[34] Furthermore, in recent years the production cost of enzymes has decreased considerably, thus outweighing the cost and need for enzyme immobilization.[19] However, technical as well as economic analyses are still needed to decide whether the immobilized version or rather the free form of the enzyme is better for a particular process.[35] Several immobilization strategies including binding enzymes to a support, cross-linking, and encapsulation exist each with its own pros and cons.[11]

Enzyme engineering

Unlike enzyme immobilization, this strategy can be used to improve an enzyme toward any characteristics desired.[29] This strategy comprises modifying randomly or systematically the DNA encoding an enzyme. The modification at the DNA level is called a mutation and the result of a mutation is a substitution or a change of an amino acid at the protein level of an enzyme. The general applicability of enzyme engineering makes this approach perhaps the most powerful strategy.[28,36] There are many examples of enzyme engineering used to improve enzyme performances toward industrial applications.[36] Additional details about the existing methods in protein engineering will be discussed in subchapter 1.2.

Enzyme discovery

Searching a suitable enzyme in nature for an industrial application perhaps is a standard approach used pre-enzyme engineering days. With rapid development in immobilization technology and protein engineering approaches, early researchers tend to shift to the latter strategies. However, with the recent advancement in genome sequencing, metagenomics, and high-throughput screening systems, finding of a readily-suitable biocatalyst for industrial application should have preceded all other strategies.[37,38] This concept is known as biocatalyst bioprospecting.[39] Recent

explorations of archaea and bacteria that are found living in extreme conditions (extremophiles) further facilitate the bioprospecting approach. These microorganisms are the "hotspot" to find new biocatalysts with characteristics that are not present in enzymes from mesophilic microorganisms.[40]

1.1.2.2 Tailoring an industrial process to support the use of an enzyme

In this approach, a biocatalyst is the center of the reaction and the process is developed to be suitable for a particular biocatalytic conversion of the desired substrate to its corresponding product.[11] Two strategies can be applied:

Medium engineering

In nature, enzymes typically function in aqueous solution; thus, conversion of water-insoluble organic compounds or reaction in non-aqueous media may not be readily suitable. In this strategy, the medium where the enzyme is used to perform a reaction can be modified. This modification can lead to a better enzyme stability, activity, and selectivity.[12,41] Medium engineering can also help a process by simply increasing the solubility of a substrate or assisting in the product separation, i.e. *via* crystallization. This strategy includes additives such as salts, oligosaccharides, polyols, DMSO, etc.[24] Performing the reaction in a completely new media such as ionic liquids, supercritical carbon dioxide, and deep eutectic solvents is another strategy that can be done and is being investigated extensively.[12,41]

Reaction engineering

This approach is a routine procedure in chemical industries when new processes are being developed. In this approach, a reactor is designed in such a way to accommodate a biocatalyst and to improve product yield. This strategy includes reactor designs (packed bed bioreactor, photo bioreactor, stirred-tank reactor, etc.) and its operational modes (batch, fed-batch, continuous).[42] For multiple enzymatic reactions or processes using combination of chemical and biocatalysts, reaction engineering is very crucial to determine the productivity, yield, and overall production cost.[43]

1.2 Methods in protein engineering

Enzyme engineering is a multidisciplinary field that was developed to answer modern challenges of enzymes, i.e. designing suitable biocatalysts for a wide range of applications. Enzyme engineering methods could broadly be categorized into two approaches: random and rational design. However, this division is not mutually exclusive. Several techniques can be categorized into both. A combination of both strategies is also often needed to achieve better biocatalysts with the desired properties.[44,45] Therefore, the division here is made to ease explanation.

1.2.1 Directed evolution

The development of this approach was perhaps motivated by a shrewd observation of nature. It has been long known that life on earth was formed through a combination of variation and natural selection. This principle was then applied in a laboratory to modify an enzyme. In a broad definition, directed evolution is a strategy to change the nucleotides of an enzyme randomly and rely on a screening system to find a handful of suitable variants from several hundreds to millions of variants generated, referred to as a library. Thus, the fundamental principle underlying this approach is, "you get what you screen for".[46] This principle means that from hundreds to millions of variants generated, the ability to find suitable variants depends on how the library is screened. The power of this approach is its simplicity, as no prior knowledge of the enzyme´s mechanism and structure is mandatory.[47] There are so significant numbers of methods in the directed evolution developed throughout the years. Four representative methods are discussed below.

1.2.1.1 Error prone PCR (epPCR)

This method explores the infidelity of a Taq-based DNA polymerase (DNA polymerase from *Thermus aquaticus*) due to a lack of 3´ to 5´exonuclease proofreading activity (Figure 1.2). The error rate of a Taq-based polymerase is calculated to be about 1 in 9000 nucleotides. In brief, PCR amplification by a Taq polymerase of a parental DNA encoding an enzyme will introduce base pair mutations at random positions.[47,48] Mutations can be silent, missense, or nonsense. A silent mutation is a mutation where a change of a nucleotide results in the same amino acid encoded. Missense is when a mutation results in an amino acid substitution. Nonsense is when a mutation is to a stop codon (TAA, TAG, TAG), thus prematurely truncating a protein translation. The result of this epPCR will be an amplicon containing different nucleotide sequences due to the mutations (a library). This amplicon is then cloned to an expression vector and a suitable host is used to express the library. Afterwards, a proper screening platform is developed to screen variants with improved traits. Depending on the need, one or more rounds of epPCR can be performed to obtain desired variants. Error rate of a Taq polymerase can be adjusted further to influence the mutation rate in the library development, such as by[49]

- Increasing the concentration of magnesium and/or magenese ion
- Unbalancing the ratio of dNTPs
- Increasing the extension time
- Increasing the number of PCR cycles
- Using different polymerases with lower fidelity

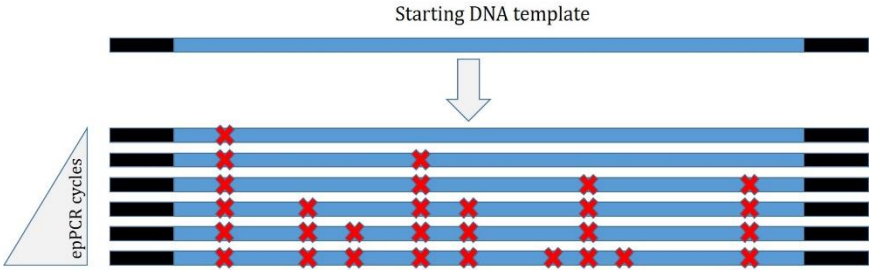


Figure 1.2. Schematic illustration of epPCR. Depending on the PCR conditions, each PCR cycle could introduce more mutations (red crosses).[49]

1.2.1.2 Sequence Saturation Mutagenesis (SeSaM)

This method was developed to overcome limitation in epPCR method described earlier, in particular the higher preference of a Taq-based DNA polymerase toward A to T and T to A mutation. This bias will limit the diversity of the library generated by epPCR.[50] In simplified explanation, SeSaM can be divided to four major steps (Figure 1.3).[51]

- The first step is a PCR of the parental DNA using a defined mixture of phosphorothioate and standard nucleotides. A combination of a forward primer with a biotin tag and a normal reverse primer is used. The ratio adjustment of these different nucleotides is important to allow an even incorporation of phosphorothioate-containing nucleotides. These special nucleotides will form a phosphorothioate bond instead of a phospodiester bond between two adjacent nucleotides. This bond is susceptible to cleavage upon treatment with iodine under alkaline conditions; thereby creating a pool of DNA with different lengths. This pool is then loaded into a streptavidin column and subsequent steps are done to remove unwanted fragments and complimentary strands (the reverse strand). The pool of single strand (ss) DNA (forward strand) with different length is then eluted from the strepavidin column.
- In the following step, the pool of the random sized ssDNA of the forward strands is elongated by a terminal transferase incorporating a universal base with one or more nucleotides addition are expected.
- Afterwards, the pool of the ssDNA forward strands containing one or more universal bases is elongated fully using a complementary sequence that is attached to a column to obtain a full-length gene. The full length of the forward strands containing the universal base(s) at random position are then purified out from the column and used as a template for the final PCR.
- In the final PCR step, a non-biased mutation (to A or T or G or C) takes place at the position where the universal bases were incorporated.

The key to success for this method depends heavily on the ability to create single strand templates in the first and third steps. Failure to do so will lead to PCR amplifications of the original sequence.[50]

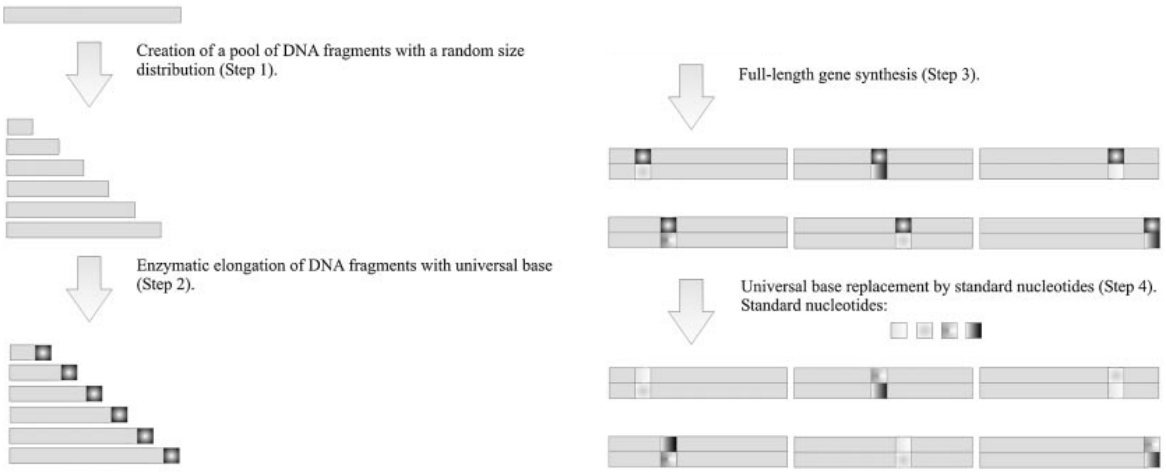


Figure 1.3. Simplified Workflow of SeSaM that comprises of 4 steps.[50]

1.2.1.3 Staggered Extension Process (StEP)

The first two methods described earlier is also known as an "asexual" approach at which only one parental DNA is used; thus, no crossover event occurs. Staggered Extension Process (StEP) was developed based on a "sexual" approach meaning two or more different parental DNAs were used mimicking natural crossover reaction (Figure 1.4. A). The sexual approach is also known as DNA shuffling. Unlike a standard PCR protocol consisting of denaturation, annealing, and elongation steps, StEP only consists of denaturation and annealing steps (Figure 1.4. A).[52] The use of a polymerase with low fidelity like Taq-based polymerase is still preferred in this method. In addition, the DNA polymerase with a slower activity is generally desired. During the denaturation step, two or more parental DNAs are denatured creating single strands. Forward and reverse primers will bind to each strand during the annealing step. At this step, the DNA polymerase will also start to incorporate nucleotides into the 3´ end of the annealed primers, albeit at very low speeds. This step will create short amplicons. Because there is no elongation step like in a normal PCR, the short amplicons generated will act as primers in the next cycle and so on. Due to the homologous parental DNAs used, the amplicons created can also anneal to different parental DNAs, thus creating hybrids. The cycles are continued until a full-length of amplicons are obtained. This chimeric genes library is subjected to selection steps. The diversity of a library obtained is determined by at least two factors below,[53]

- The higher the sequence similarity of the parental DNAs used, the higher the success of recombination will be. However, higher sequence identity will also lead to lower diversity. Parental DNAs with at least 70% sequence similarity is advised. Sequence similarity lower than 70% would lead to low crossover event, if at all.

- Because elongation happens at the same time as the annealing step, the temperature at this step dictates recombination events. Lower temperature in general will result in shorter amplicon being generated, thus yielding a higher diversity. However, very low annealing temperatures would result in failure to obtain a full-length amplicon. This parameter sometimes has to be adjusted empirically.

One or more rounds of StEP is sometimes needed to obtain variants with the desired characteristic.[52] DNaseI-assited DNA shuffling is another method that had been developed couples of years earlier than StEP based on the same "sexual" approach. DNaseI-assited DNA shuffling is based on the random fragmentation by DNAse I and followed with two PCR steps. The first PCR step is to recombine the fragments in part and the latter PCR step using terminal primers to get a full-length chimeric gene.[54]

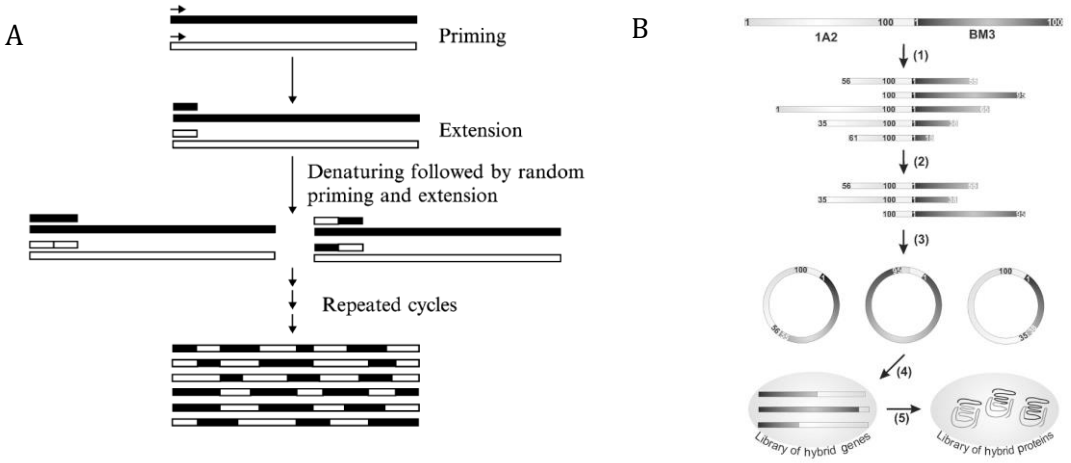


Figure 1.4. *In vitro* "sexual" approaches in directed evolution: Staggered Extension Process (A) and Sequence homology-independent protein recombination (SHIPREC) (B).[52,55]

1.2.1.1 Sequence homology-independent protein recombination (SHIPREC)

This approach was again developed to overcome the limitation of StEP at which recombination of low sequence similarity is not possible.[55] In this approach two parental DNAs with low sequence identity but with similar size is attached by a linker containing several restriction sites creating one big DNA fragment (Figure 1.4. B).[55,56] DNase I is then used to cut the fragment creating randomly fragmented DNAs with various sizes. The fragments with desired size (the size of the original parental DNA) are then selected by means of DNA electrophoresis. The linear fragments are then circularized by a T4 ligase. Appropriate restriction enzymes are used to cut the circular DNAs at the linker site and later cloned into an expression vector. Afterwards, a selection round is used to select variants with desired traits. The utilization of this method allowed one recombination event of two parental DNA that shared only 16% sequence similarity.[55] Incremental truncation for the creation of hybrid enzymes (ITCHY) is another similar approach

that relies on a PCR step rather than a restriction by a DNase I to create fragments.[57] In ITCHY there is also no size selection step occurring; thus, the chance to obtain non-functional variants is higher, hence a more powerful screening approach is needed.[55]

1.2.1 Rational design

Although the power of directed evolution has been proven throughout the years to tailor biocatalysts for industrial applications, this approach is not always suitable. One of the reasons is the need to screen a large number of variants to find only one or two variants with improved traits.[58] This drawback becomes more apparent when the cost or time-needed for screening is too high (low throughput screening). Factors such as expensive substrate, slow enzymatic turnover, product is not easily detected can be the reasons.[59,60] As an alternative, rational designs can be employed for such cases. To engineer an enzyme rationally at least knowledge of its crystal structure or catalytic mechanism is needed, depending on the target function. With the increasing number of crystal structures solved, this approach is getting more proponents.[60] Recent development in the computational approach to build a more reliable homology model also facilitates the popularity of this approach. Few representative strategies will be discussed to explain the general concept of rational design.[59]

1.2.1.1 Sequence-based approach

This approach explores a vast number of genomes that have been sequenced. In this approach, couples to thousands of sequences of a particular enzyme class are aligned. Several alignment algorithms have been developed in recent years, such as Clustal Omega, MUSCLE, Praline, etc.[61,62] From the multiple sequences alignment (MSA) obtained, several conserved amino acids can be identified.

Consensus-guided engineering is a concept that is developed with a hypothesis that modern proteins are evolved from a rather common ancestor with functionality and stability as the evolutionary drift. As a result, nature experimented millions of years to sample “the most suitable” amino acid at a particular position in an enzyme in respect to functional activity and stability. The stabilizing mutations would be retained as long as a necessary activity level was achieved. This means at any position a more conserved amino acid (found more frequently in nature) is more likely to be a better one, in particular in respect to stability.[63] Thus, after performing MSA, one or more consensus sequences can be retrieved. A parental sequence of an enzyme target is then compared to the consensus sequences. Site directed substitutions are performed at certain positions at which the original enzyme target has different amino acid sequences.[64] The success of this approach relies on several factor, such as:[65]

- The more sequences used, in particular with higher variability will result in better consensus sequences retrieved

- The more knowledge of an enzyme target (catalytic active site, cofactor binding site, substrate binding motifs, 3D structure), the better judgement could be taken to avoid substitutions with detrimental effects

A phytase and a glucose dehydrogenase were enginereed using this approach and significantly improved thermostabilty was achieved.[64,65] This concept can be extended further in an attempt to correlate an amino acid found at one position to another amino acid located at a distance in one enzyme through evolutionary events. This is known as correlated mutation analysis. This approach argues that amino acids located at a distant position in a sequence can interact in a 3D conformation. Thus, these amino acids may mutate coordinatively. This approach is used to better predict whether a mutation in a codon will compromise the enzyme integrity and thus activity.[66] For example, if there is a strong correlation between one amino acid to another one located in the distance, substitution of the first amino acid will likely affect its interaction to the corresponding amino acid. This approach is in particular useful to make a better decision in a sequence-based rational design where the crystal structure is unknown.[67] Databases such as 3DM provide such sequence based analyses.[68]

Ancestral sequence reconstruction is another approach in rational enzyme engineering based on the sequences. Originally, this approach was developed to study evolutionary event of DNA and protein.[69,70] However, this strategy has been recently used to create better biocatalysts.[70,71] The hypothesis is that life was started in a rather hot environment, thus ancient enzymes are thought to have higher stability than modern enzymes. Moreover, as enzymes underwent evolution to sustain more complex life, ancient enzymes are also thought to be less in number, hence broader in substrate promiscuity. Therefore, by applying this method, enzymes with higher thermostability and/or broader or new substrate specificity could be uncovered.[70,71]

1.2.1.2 Structure-based approach

In this approach, 3D structure information of a target enzyme is needed. It is always preferred the information is obtained from a crystal structure, in particular the one with good resolutions. However, when it is not possible, homology modelling can be performed. Several tools are available to perform homology modelling based on the protein sequence. Swiss-Model, Phyre2, and Robetta are three of the most known web-based homology modelers.[72–74] YASARA is another commercial software that can be used to perform homology modelling.[75] There are so many rational design strategies based on the enzyme´s structural information that have been shown to work in past years. To ease discussion, representative strategies will be discussed based on the purpose of the biocatalyst engineering.

Increasing stability

Biocatalyst stability is one of the important factors that countribute for the success of its application in industry. Several different types of stability have been described earlier (see subchapter 1.1.1.2). Two process conditions are often found in chemical industries: high temperature and presence of organic solvent. Enzyme stability toward both conditions are then needed.[11,24,26] Stability toward high temperature and organic solvent are not mutually exclusive.[30] It has been known that these two stabilities are correlated. Thus, improving stability toward one condition will likely to also increase stability toward the other.[76] At least, the following three strategies have been widely applied to increase a biocatalyst´s stability.

B-factor analysis—this method requires structural data from a protein crystal structure. A high quality of crystal structure can give more accurate and reliable B-factor information. In a simplified definition, B-factor of an amino acid residue is an average measurement of the respective residue position during X-ray diffraction. A residue that has higher flexibility will tend to vibrate more, thus higher B-factor value.[77] As higher stability can be associated with higher rigidity, the residues with higher B-factor would be good candidates for amino acid substitutions. This method has been used to improve thermostability of a lipase by more than 50 °C.[78]

Folding energy calculation is another approach that has been used in recent years. Many computer algorithms have been developed to calculate this value. The folding energy is calculated by the difference of Gibbs free energy (ΔΔG) before and after an amino acid substitution. Negative ΔΔG of the substitution means the new amino acid has stabilizing effect. Rosetta and FoldX are two algorithms that have been widely used in literature to calculate this value for improving enzyme stability.[79,80] The extension of this approach coupled with several other approaches resulted in the development of user-friendly and web-based servers to predict and create thermostable enzyme variants, such as FRESCO, PROSS, and FireProt[81–83].

Another approach combines empirical knowledge and tries to rationalize in the context of a known crystal structure or homology-modelled structure. Scientist have been trying to develop consensus in the factors contributing to protein stability. Optimization of protein surface charges, introduction of disulfide bonds, improving dimer interface interaction, rigidification of loop regions are among successful strategies reported.[84–88] Comparing the sequences of thermophilic enzyme is another strategy that has been proven successful in improving stability of mesophilic enzymes.[89]

Altering activity or selectivity

Besides high stability, high activity and/or selectivity are desired in biocatalyst for industrial application.[28] Rational designs toward improved activity and/or selectivity require detailed knowledge of catalytic mechanism of a target enzyme. Understanding how the substrate binds is also essential. This can be done by co-crystalizing the target enzyme with the substrate of choice. If this data is not available, *in silico* docking can be performed to a crystal structure or homology-

modelled structure of a target enzyme. Molecular docking can be perfromed using several programs, e.g. SwissDock, Glide, and RosettaDock.[90] More sophisticated docking protocols have also been developed throughout the years.[91]

After enzyme-ligand interaction has been reliably calculated and modelled, the interactions are carefully observed. The "hotspots" are then targeted for subsequent substitutions. Amino acid substitutions of the hotspots can be performed at one position at a time, either to one particular amino acid or to several amino acids at the same time (site saturation mutagenesis) prior to combining beneficial point substitutions. This approach is known as iterative saturation mutagenesis.[58,92] Another approach is to perform substitutions of the hotspots simultaneously creating a big library. This approach is known as multiple site saturation mutagenesis (MSSM).[93] MSSM will benefit for providing all possible combination of amino acid interactions, but ISM will benefit from a considerable smaller amount of variants need to be screened. Combinatorial active-site saturation test (CAST) concept was developed further to reduce the number of necessary variants to be screened in MSSM by dividing an enzyme active site into several subsites and perform MSSM to each subsites independently. A subsite can comprise one or more amino acids.[94,95]. The variants from CAST could be combined utilizing ISM approach to improve further the activity and/or selectivity of a biocatalyst.[96] Considering all the benefits and challenges of the approaches described above, the throughput and cost of a screening system will thus decide which strategy or combination of strategies above are the most suitable.

1.2.1.3 Targeted site directed/saturated mutations

Regardless of the approaches used in rational design, variants with targeted amino acid substitutions still need to be created. Overlap extension PCR and QuikChange are two most widely used techniques to alter nucleotide sequences. Several other molecular cloning techniques are also getting increased attentions in the past years.

Overlap Extension PCR

This technique was introduced already in the late 80s.[97] This technique comprises two separate PCR steps (Figure 1.5. A). In the first PCR, two set of primers are used in two separate reaction tubes.[97,98] The first primer set is an outer forward primer annealing to the outer part of a gene and an inner reverse primer containing a desired mutation to a target position in a gene. The second primer set is an inner forward primer containing a complementary mutation and an outer reverse primer. The key is that the inner reverse primer should contain an overlap region to the inner forward primer. After a standard PCR reaction, two PCR products (forward and reverse regions) are purified and mixed into one cup and a standard PCR reaction without primers is performed. During denaturation and annealing, two fragments will anneal to each other in the overlapping region bearing the desired mutation. Then, in the extension process, the 3´ of each

strand will be elongated to give a full amplicon. A set of the outer forward and reverse primers are added to amplify the whole gene. The gene with the mutation are then cloned to an expression vector and used to transform a desired expression host.

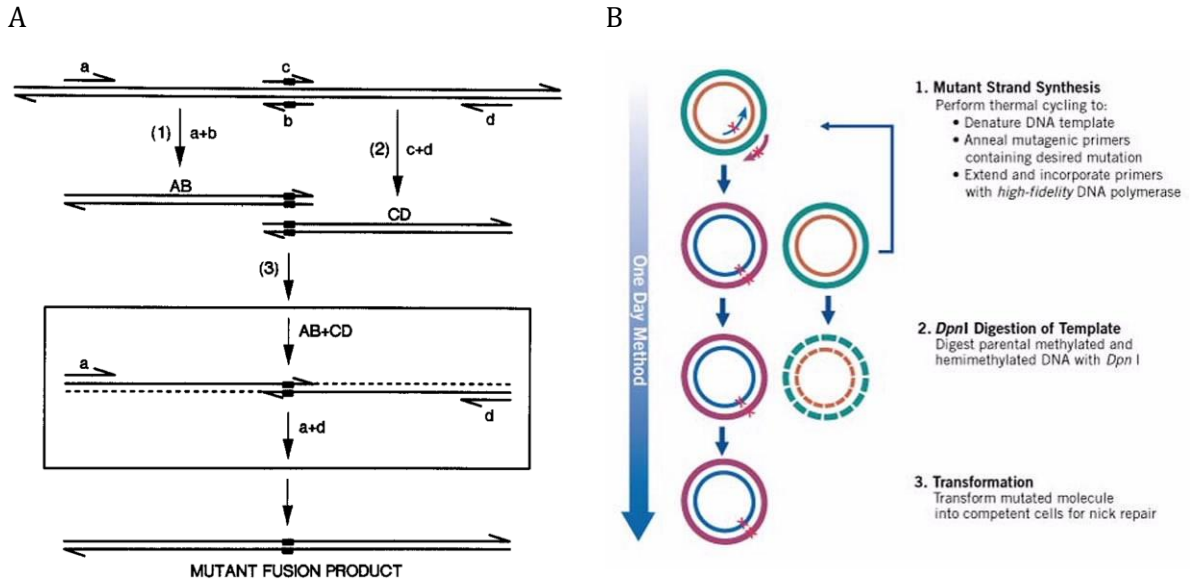


Figure 1.5. Schematic illustration of site directed mutagenesis by overlap extension PCR (A) and QuikChange (B).[97]

QuikChange

This technique was developed almost a decade later by Stratagene (La Jolla, USA) after overlap extension PCR (oePCR). The key to this technique is a restriction enzyme called DpnI.[99] This enzyme can digest methylated DNA, but is unreactive to non-methylated DNA. The methylated DNA is usually produced by microorganisms, such as *E. coli* while the non-methylated DNA can be obtained from PCR amplification. This technique begins with designing a set of primer bearing desired mutations to a gene in a plasmid. Several different primer design strategies have been reported in literatures.[100–102] A forward and a reverse primer bearing mutations to specific nucleotides of the gene can be designed to overlap completely (the original protocol by Stratagene) or partially. However, the latter design is reported to give a higher yield of correct PCR products.[100,101] Unlike oePCR, QuikChange (QC) PCR technique will amplify the entire plasmid. The QC amplifies the plasmid rather linearly than exponentially because the PCR product cannot be used as a new template. After QC PCR is finished, DpnI is added to remove parental DNA that is methylated. The plasmid product with the gene containing desired mutations is used to transform a desired microorganism. QC PCR bypasses the necessary cloning step in oePCR; thus, the entire QC process is faster and simpler (Figure 1.5. B). However, as the product of QC PCR is a plasmid containing nicks, competent cells with higher competency are needed.

Other methods

Development in molecular cloning techniques such as Golden Gate and Gibson Assembly can also be used to improve efficiency of a library generation, in particular when huge numbers of variants are expected.[103,104] However, all above-mentioned techniques are based on PCR amplification; thus, certain bias is expected, for example higher preference of certain primers with degenerate motifs (in particular the one that is the same as a WT codon) over other motifs.[105] This effect is more apparent when multiple sites saturation mutagenesis is performed. Chemical synthesis can be used instead to create such library.[106] However, chemical synthesis of a library costs much more than any above-mentioned techniques.

1.2.1.4 *De novo* protein design

This approach is relatively new in comparison to other techniques in Rational Design. This technique is based on the fundamental of protein folding to its lowest energy state. Thus, if the protein folding energy can be determined accurately, it would be possible to predict protein structure without the need of a homology template.[107] To design a biocatalyst *de novo*, in-depth knowledge of catalytic mechanisms and transition states are extremely crucial to predict which amino acids and their relative proximity in space needed to perform a chemical reaction. Following that, an artificial active site has to be arranged and fitted to a protein backbone to create a complete enzyme.[108] Many promising works have been reported so far.[109] However, most of them still perform worse than natural enzymes.[28,110] A combination of directed evolution and rational design of *de novo* designed protein has been shown to improve their activities.[111,112] Because this technology requires system with huge computing power (super computer) in combination with profound knowledge in protein biochemistry, biophysical chemistry, computational science, this approach has not yet to be widely applicable.[59,110] Perhaps, in the next decades along with the developments in better computational technology and methods as well as advancement of fundamental knowledge of protein, this approach will be more applicable for the development of industrially relevant biocatalysts.[10,107]

1.3 Lyases

Enzymes belonging to the class of Lyases are enzymes that can catalyze the breakdown of chemical bonds, such as C-C, C-O, C-N by means of elimination creating double bonds or cyclic forms. The reverse reaction, i.e. formation of the bonds facilitated by addition to an unsaturated group is also possible.

1.3.1 Lyases for application in second-generation chemicals production

Second-generation chemicals production is developed as an alternative to first-generation approach that utilizes carbohydrate-rich substrate, such as corns, sugar canes, and potatoes. In the second-generation approach, non-edible biomass is used, instead. This approach can be

carried out either *in vivo* or *in vitro*. In the *in vivo* approach, the production of the second-generation chemicals can be extended further to a concept called consolidated bioprocessing (CBP). CBP allows biomass degradation, production of chemicals, and product separation in one vessel.[113,114]. In the HotSysAPP project funded by BMBF (German Federal Ministry of Education and Research), the CBP concept is explored using *Sulfolobus acidocaldarius* (*Saci*) as a new industrial chassis (Figure 1.6).[115] *Saci* naturally grows at low pH and high temperature, a prevailing condition for biomass degradation.[116]

Cell-free production of chemicals has been developed in recent years as an alternative to the *in vivo* approaches. In this approach, a set of enzymes catalyzing only required reactions in a designed pathway is used, thus minimizing side products. Depending on the chemicals produced, a cell-free approach may also serve as a better alternative, in particular when the chemicals are toxic to the micoorganisms used.[117] Nevertheless, for both approaches—*in vivo* and *in vitro*—enzymes with high stability and activity are required. In this work, two subclasses of enzyme class Lyase namely decarboxylase (facilitating C-C elimination) and dehydratase (facilitating C-O elimination) were focused. The decarboxylases were targerted to be improved their stability, whereas the dehydratases were targerted to be increased their activity, either *via* engineering or *via* discovery of new enzymes. Both enzyme classes will be discussed in more detail hereinafter.

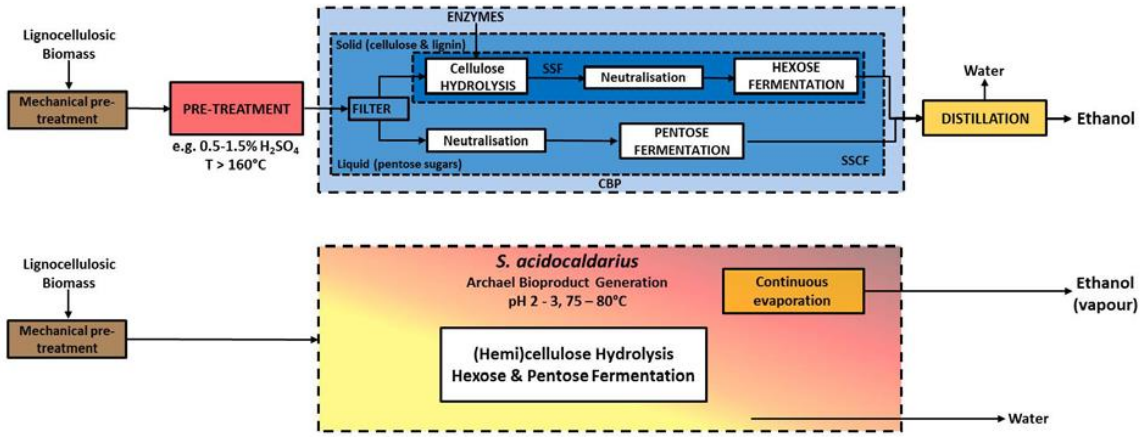


Figure 1.6. The project goal of HotSysAPP is to engineer *Sulfolobus acidocaldarius* toward consolidated bioprocessing of lignocellulose to bioalcohols. The picture was taken from the HotSysAPP proposal.

1.3.2 Decarboxylases

1.3.2.1 General mechanism and structure

Decarboxylases are one subclass of Lyases with EC number 4.1.1.x Decarboxylases facilitate removal or attachmentof a carboxyl group *via* a C-C elimination or addition. At ambient condition (1 atm, room temperature), the elimination reaction is greatly preferred.[118] This enzyme subclass is responsible to essentially all CO_2 produced during fermentation or respiration process of organisms.[119] There are more than one hundred different decarboxylases registered with the

Enzyme Commission. In general, decarboxylation reactions are performed with organic or inorganic cofactors as well as without cofactors. Common organic cofactors are biotin, flavin, glycyl radical, pyridoxal 5´-phosphate (PLP), pyruvoyl group, and thiamine diphosphate (ThDP). Inorganic cofactors include several divalent cations, such as Mg^{2+}, Fe^{2+}, Mn^{2+}, Zn^{2+}, Co^{2+}, and Cd^{2+}.[120] In addition to be one of fundamental reactions in biological systems, decarboxylases have gained interest as important biocatalysts. The reverse reaction, i.e. incorporation of carbon dioxide is viewed as a promising reaction toward utilization of CO_2 as the new carbon source.[121] Due to the vast diversity of decarboxylases, only relevant decarboxylases to this work will be of focus, specifically pyruvate decarboxylases (EC 4.1.1.1) and branched-chain α-keto acid decarboxylases (EC 4.1.1.72).

Pyruvate decarboxylases (PDC) and branched-chain α-keto acid decarboxylases (KDC) are decarboxylases that use ThDP to catalyze the decarboxylation reactions.[122] ThDP is a biologically active derivative of vitamin B1 that has broad catalytic functions in addition to decarboxylation. Besides ThDP, Mg^{2+} has been shown to be essential for the reactions catalyzed by PDC and KDC. In both decarboxylases, ThDP is stabilized by a conserved hydrogen bonding network in the ThDP-binding cleft that is built from interactions of two conserved domains (pyrimidine domain or PYR and pyrophosphate or PP domains).[123,124] The pyrophosphate (diphosphate) moiety of ThDP is coordinated further by Mg^{2+} which binds to conserved amino acids, aspartic acid and asparagine ($GDGX_{25\text{-}30}N$).[123,125] In addition to PYR and PP domains, both decarboxylases also have the third conserved domain called Regulatory domain (R domain). R domain connects PYR and PP domains.[123] In quaternary configuration, PDC enzymes are usually arranged as tetramer of two homodimer. Crystal structure of PDCs that have been solved are from *Saccharomyces cerevisiae* (*Sc*), *Zymomonas mobilis* (*Zm*), *Zymobacter palmae* (*Zp*), *Acetobacter pasteurianus* (*Ap*), *Gluconoacetobacter diazotrophicus* (*Gd*), and *Kluyveromyces lactis* (*Kl*) (see PDB). Only one crystal structure of KDC from *Lactobacillus lactis* (*Ll*KdcA) has been solved and it is a homodimer. An active site of both decarboxylases are located in an interface of the homodimer.[126]

Decarboxylation reactions of PDC and KDC follow these subsequent steps (Figure 1.7).[120,127,128] A conserved glutamic residue of either PDC or KDC abstracts a proton of N´1 of the pyrimidine of ThDP stabilizing the imino form of the pyrimidine ring. At this state, N´4 of the iminopyrimidine can deprotonate C2 of the thiazolium as a general base thus creating a highly nucleophilic ylide (ylide-ThDP). The ylide will later attack a carbonyl group of pyruvate or other α-keto acids that is present in the active site forming a covalent tetrahedral intermediate. Decarboxylation of this intermediate will create a C2-α-carbanion which stably resonate to its enamine form and a CO_2 molecule is released. This enamine configuration of the thiazolium ring is now acting as an electron sink. Protonation of the thiazolium C2 of the enamine state results in a second tetrahedral intermediate. Deprotonation of this intermediate is followed by concomitant

C-C cleavage between thiazolium C2 and intermediate Cα yielding an aldehyde. In the presence of another aldehyde, C2-α-carbanion can do nucleophilic attack to the aldehyde facilitating a carboligation forming a 2-hydroxyketone. In a WT enzyme, decarboxylation is greatly favored over carboligation even in the presence of another aldehyde.[129] Carboligation by ThDP-dependent decarboxylases, however, has been extensively researched to synthesize a number of building blocks.[122,128]

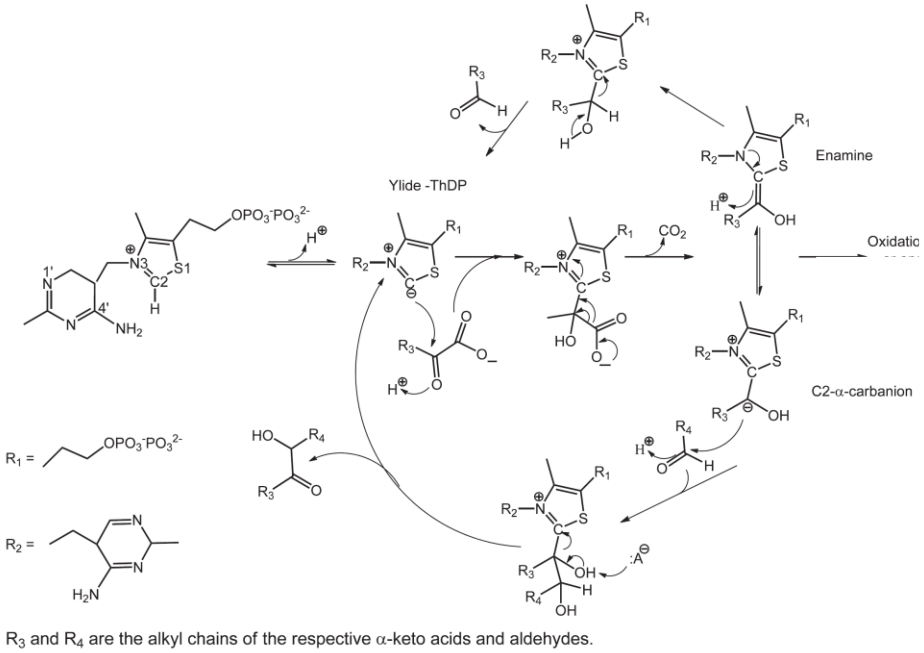


Figure 1.7. Reaction mechanisms of TDhP-dependent decarboxylases. The upper part represent standard decarboxylation reaction while the lower part represent carboligation.[120]

1.3.2.2 Carboligase activity of PDC and KDC

The first application of PDC toward synthesis of (R)-phenylacetylcarbinol *via* carboligation dated back almost a century ago. During that time, yeast fermentation was carried out in the presence of benzaldehyde.[130] Reductive amination of (R)-phenylacetylcarbinol ((R)-PAC) will result in (L)-ephedrine, a stimulant to prevent blood pressure drop during spinal anesthesia.[131] Two mostly studied pyruvate decarboxylases for carboligation are *Sc*PDC and *Zm*PDC.[128,131] When acetaldehyde is used as a substrate, *Sc*PDC will convert the aldehyde to (R)-acetoin (condensation of two acetaldehyde) while *Zm*PDC will result in (S)-acetoin. In fact all other bacterial PDCs (*Zp* and *Ap*PDC) will result in (S)-acetoin.[122] However, when acetaldehyde and benzaldehyde are used as substrates, all PDCs mentioned above will yield (R)-PAC (acetaldehyde acting as a donor wile benzaldehyde as an acceptor) with ee >88%. In addition to (R)-PAC, *Sc*PDC will also form (S)-acetoin with ee 20%. No formation of HPP (2-hydroxypropiophenone) as the possible product when benzaldehyde acting as donor and acetaldehyde as an acceptor is observed. Bezoin (condensation of two benzaldehyde) is also not formed.[122] This stereo control of carboligation in *Sc*PDC is attributed to the presence of a conserved isoleucine at position 476.[128,131]

As opposed to PDC, KDC (*Ll*KdcA) shows broader substrate specificity, in particular substrates with bulkier side chain. This makes KDC able to perform carboligation to more challenging substrates.[132] In the presence of acetaldehyde and benzaldehyde as substrates, 60% product is (R)-HPP with 93% ee and 40% is (R)-PAC with 92% ee. This could be explained partially that decarboxylation activity of *Ll*KdcA toward benzoylformate is faster than pyruvate, thus formation of "active benzoyl" is more preferred than "active acyl".[122] Interestingly, *Ll*KdcA could not catalyze self-condensation of isobutaryldehyde (the product of decarboxylation of its natural substrate: α-ketoisovalerate). In general, longer aliphatic aldehydes can push further selectivity toward (R)-PAC derivatives (aliphatic aldehydes as donor).[132]

1.3.2.3 Decarboxylation reaction of PDC toward CBP of second-generation biofuels

Pyruvate decarboxylase gene (pdc) is widely distributed in fungi and higher plants. A number of pdc genes are also present in bacteria. There is no pdc reported so far in thermophilic microorganisms or archaea. PDCs among bacteria, fungi, and plant share rather low protein sequence similarity (around 30%), but their 3D structures are rather conserved.[133] In metabolism, PDC is responsible for anaerobic fermentation for the production of ATP. PDC decarboxylates pyruvate produced from glycolysis to acetaldehyde, which later can be reduced to ethanol by an alcohol dehydrogenase. The ability of microorganisms to produce ethanol has been exploited for thousands of years, in particular in the production of alcoholic beverages. However, with the recent issues in energy crisis and climate change, ethanol has emerged as a sustainable alternative to gasoline. First-generation ethanol production began almost a half century ago with starchy substrates being used as the substrates for ethanol fermentation.[134] Recent developments in molecular biology enable engineering of a strain toward bioproduction of desired chemicals. This has opened up a possibility to produce ethanol not only in yeast and ethanologenic bacteria, but also in other microorganisms, such as thermophilic organisms.[134–136]

Ethanol production in thermophilic microorganisms have several advantages in comparison to mesophilic microorganisms. Utilization of high temperatures during ethanol fermentation will minimize bacterial contamination, e.g. by lactic acid bacteria, which is a prevalent issue in a standard ethanol fermentation at ambient temperatures. High temperature will allow a combination of product removal and ethanol fermentation. At 50 °C, ethanol readily evaporates and applying mild vacuum would be sufficient to strip ethanol from fermentation broth as opposed to energy intensive conventional distillation. This combined process will potentially reduce the cost of downstream processing. Substrate hydrolysis prior to ethanol fermentation is a necessary step as yeast or other ethanologenic bacteria are not able to degrade complex carbohydrate. This step is usually done at high temperature; thus, a cooling step is required prior to mesophilic fermentation, increasing further energy input. Ethanol fermentation

by thermophilic at constantly high temperature (>50 °C) will then significantly decrease the energy requirement in bioethanol production.[134,136]

A second-generation ethanol production aims to overcome food vs fuel debate that emerged alongside production of first-generation bioethanol. This dilemma is started by a concern that farmers may prefer to use their lands or to sell their crops for biofuel production than for food production. This concern is not without a solid base. The availability of farmlands is limited; thus, the price (profit) will dictate to which purpose the lands would be utilized. In the second-generation ethanol production, biomass and non-edible crops will be used instead. Development further of this approach results in consolidated bioprocessing concept.[113] This concept explores the possibility to perform biomass pretreatment, ethanol fermentation, and ethanol separation in one vessel. To realize this concept, a cellulosic and ethanologenic thermophilic microorganism will be required. In nature, this ideal microorganism has not yet to be found. Hence, this microorganism has to be designed in a laboratory.[114] The ability of such microorganisms to grow at lower pH will further improve compatibility with high activity of hydrolases to degrade biomass at low pH. Although powerful metabolic engineering techniques and our understanding in microorganisms have expanded significantly, there are still limitations in evolving microorganisms that we cannot address now. For example, engineering a microorganism´s ability to grow at significantly higher temperature or lower pH than its optimum conditions would be practically impossible. However, engineering a thermoacidophile, e.g. *Saci* to ferment ethanol from cellulosic feedstock (to make into a cellulosic and ethanologenic microorganism) is in theory within our reach. In relation to pyruvate decarboxylase, further discussion is focused on the development of ethanologenic thermophilic microorganisms.

In nature, production of ethanol from pyruvate can be facilitated by means other than *via* pyruvate decarboxylases. Two different enzymes, pyruvate dehydrogenase (EC 1.2.4.1) and pyruvate ferredoxin oxidoreductase (EC 1.2.7.1) can catalyze oxidative decarboxylation of pyruvate to yield acetyl CoA. The acetyl CoA formed can be reduced by a bifunctional aldehyde reductase (AdhE) to yield ethanol.[136] These oxidative pathways are frequently found in thermophilic bacteria. With these pathways, ethanol production is observed in thermophilic microorganisms, such as *Thermoanaerobacterium. saccharolyticum*, *Thermoanaerobacter mathranii*, *Thermoanaerobacter ethanolicus*, *Clostridium thermocellum*, and *Geobacillus thermoglucosidasius.* One common feature of the ethanol production by native thermophiles is its considerably low titer in comparison to yeast or ethanologenic bacteria. This phenomenon is thought to be due to competition with different pathways. For example, acetyl CoA produced can be diverted toward acetate, instead.[136] Moreover, lactate can also be produced simply by a reduction of pyruvate by a lactate dehydrogenase (EC 1.1.1.27). Thus, relative activity of PFOR or PDH in cytoplasm in comparison to LDH will determine whether pyruvate will be converted to acetyl CoA or to lactate.

Several engineering efforts have been devoted to improve ethanol titer in thermophilic microoganisms, such as by elimination of competitive pathways (toward lactate and acetate), heterologous expression of dehydrogenases, as well as overexpression of enzymes in native ethanol pathways.[136] Despite some success, the ethanol titer is still far below industrially relevant ethanol producers. One approach was proposed, to introduce a non-oxidative decarboxylase pathway *via* PDC in these thermophilic microorganisms. Production of ethanol *via* PDC is thermodynamically more favorable. Moreover, because less steps are required to produce ethanol with PDC than with PFOR, pathway competition can be minimized. Preliminary work in an introduction of *Ap*PDC to a relevant thermophilic ethanologenic microorganism, *C. thermocellum* resulted in promising improvement in ethanol titer and yield from cellulose in comparison to the WT strain at 55 °C.[137] Another work to introduce *Zm*PDC in *G. thermoglucosidasius*, however, did not result in higher ethanol titer. It was reasoned that the limited stability of *Zm*PDC prevents functional ethanol production.[138] The abovementioned motivations and promising results altogether suggest that improving thermostability of a pyruvate decarboxylase is an important step toward CBP of the second-generation bioethanol.

Prior to this work, several works in an attempt to improve thermostability of a PDC have been reported. One approach is to direct a thermostable acetolactate synthase (EC 2.2.1.6) to pyruvate decarboxylase.[139] Acetolactate synthase (ALS) is an enzyme that converts pyruvate to acetolactate playing a role in branched-chain amino acids (isoleucine, leucine, and valine) biosynthesis. ALS and PDC both belong to ThDP-dependent enzymes and undergo similar reaction mechanisms. An ALS from *Thermus thermophilus* (*Tt*ALS) was chosen in that study. Directed evolution of *Tt*ALS followed up by screening around 5000 colonies resulted in a variant bearing four amino acid changes. Further site directed studies revealed that a substitution, H474R is responsible to increased ratio of acetaldehyde-forming (decarboxylation) over acetolactate-forming activity. However, the acetolacte-forming activity is still favored.

Another attempt to increase thermostability of a PDC was done by ancestral sequence reconstruction of PDCs.[140] As it was discussed earlier, ASR techniques have been used to increase stability of a protein based on statistical calculation of multiple protein sequence alignment. However, this approach did not yield a more stable variant. An inferred PDC (ACN27) has rather lower stability in comparison to all bacterial PDC reported to date. The authors, however, reported a significant higher activity in comparison to the most active PDC, *Zm*PDC. ACN27 is calculated from 1.3 billion years ago. The inability to obtain a more stable variant is explained further by relating to progressive cooling theory that happened between 3.5 to 0.5 billion years ago.[141] Decrease in thermostability of reconstructed ancestral proteins over calculated evolutionary time is also observed for the case of *Bacillus* 3-isopropylmalate dehydrogenase (EC 1.1.1.85).[142]

Rosetta based rational approach was also performed to improve thermostability of *Zm*PDC. In this work, the authors used an iterative computational approach to evolve stability of *Zm*PDC.[143] The first approach is to do *in silico* SSM in Rosetta protein modelling software. From this approach, the authors identified side chains that showed poor packing. Structural rigidification was also performed by substitution of glycine to alanine and use Rosetta energy scoring function to filter beneficial substitutions. Optimization of surface net charge was also performed. Several variants that showed promising improved stability *in silico* were identified and tested *in vitro*. Activity data for all variants are reported to be similar to that of WT *Zm*PDC at room temperature without giving the actual numbers. Circular Dichroism (CD) of the best variant bearing 16 amino acid substitutions (PDC 2.03) indicated no change in molar ellipticity at 222 nm from 20 to 90 °C. Differential interference contrast (DIC) microscopy was used to determine the phase change over time as an indication of protein structural stability at 60 °C. The authors claim that any protein will undergo three distinct phase changes. First phase is where a protein is soluble in solution, thus only background solution observed. Then, protein starts to form an alternative conformation, but is still soluble in solution. The last phase is when a protein precipitates forming a distinct aggregated protein structure. The best variant entered phase 2 after 60 min and stayed in phase 2 longer than the observation duration (>90 min). WT *Zm*PDC is reported to enter phase 2 immediately at 60 °C and after 25 min it reached phase 3 indicating that the best variant is more stable than WT *Zm*PDC. The way of reporting improved thermostability of an enzyme in this study is rather unusual. Structural stability as measured by CD and DIC is an important parameter. However, as enzyme is a protein with catalytic activity, it is necessary to see the effect of temperatures toward its activity (kinetic stability).[30] Proper comparison to other PDCs in respect to kinetic stability is then needed to decide if the best variant reported in that study is also kinetically superior.

1.3.2.4 Decarboxylation of KDC toward synthesis of platform chemicals

KDC was first reported as an enzyme that was responsible for flavor generation during cheese production. Two isoenzymes from *Lactococcus lactis*, KdcA and KivD, have been characterized in detail. Both enzyme share 90% protein sequence similarity and showed similar substrate promiscuity.[144,145] In addition to KDC from *L. lactis*, two KDC from *Psychrobacter* (*P. cryohalolentis* and *P. arcticus*) and one from *Proteus mirabilis* have also been described.[146,147] However, two *L. lactis* KDCs still show higher thermostability than the other KDCs, thus are more relevant for industrial applications (see subchapter 1.1.1.2). Different WT KivD variants with different thermostability were also reported in the literature. Thus, precaution is needed when comparing these enzymes.[148]

This enzyme class was first utilized toward the synthesis of longer chain alcohols, which are a better substitute for gasoline than ethanol due to their higher energy density and lower

hygroscopy.[149] Introduction of KivD in *E. coli* facilitates production (>0.5 mM) of 1-propanol, isobutanol, 2-methyl-1-butanol, and 3-methyl-1-butanol in *E. coli.* KivD also performs as a better decarboxylase than a phenylpyruvate decarboxylase (EC 4.1.1.43) from *Saccharomyces cerevisiae* (Aro10). Higher alcohols are produced *via* decarboxylation of α-keto acids, intermediates in amino acid biosynthesis (the Ehrlich pathway).[149] Further work by integration of LeuABCD (enzymes responsible for leucine biosynthesis) in *E. coli* and engineering of a binding pocket of KivD (F381L/V461A) allows almost 10-fold increased titer of (S)-3-methyl-1-pentanol. The variant shows decreased activity toward 2-ketoisovalerate (precursor for isobutanol, a competing pathway for the production of (S)-3-methyl-1-pentanol) and (S)-2-keto-4-methyl-hexanoate (precursor for (S)-3-methyl-1-pentanol) in comparison to the WT. However, the catalytic efficiency ratio of (S)-2-keto-4-methyl-hexanoate over 2-ketoisovalearate is improved by almost 3-fold. Optimization of LeuA improved further the titer reaching almost 8 mM.[150] Further work on Rosetta-assisted engineering of KivD resulted in a variant (G402V/M538L/F542V) that improves 1-heptanol production. The variant in general exhibits lower activity to 2-ketooctanoate (precursor of 1-heptanol) in comparison to WT KivD. However, the specificity toward 2-ketooctanoate is increased over lower chain α-keto acids.[150] Other variants of KivD (V461G and V461A) are reported to improve their selectivity toward 1-pentanol production.[150] Nevertheless, one inherent drawback in the production of higher alcohols *in vivo* is the toxicity of such alcohols for the cells. Isobutanol at concentrations as low as 2 vol% already inhibits *E. coli*.[149] An alternative approach was proposed. A cell-free isobutanol production was achieved reaching >10 mM isobutanol titer. In that work, instead of KivD, KdcA was used.[151]

KivD and KdcA have also been used to produce other building blocks. Both enzymes have been reported to enable production of 1,2,4-butanetriol (BTO). Butanetriol has been proposed as a valuable plasticizer and a prospective building block for pharmaceutical application. In this study, KdcA was shown to be a better decarboxylase than KivD.[152] Slight modification of this pathways will allow production of 3,4-dihydroxybutyric acid, a precursor for 3-hydroxy-γ-butyrolactone, a versatile building block for polymer and pharmaceuticals.[153] In another study, KivD was used to produce 1,4-butanediol (BDO) *in vivo* from hemicellulose sugars (D-xylose and L-arabinose) *via* the modified Weimberg pathway. Engineering further KivD (V461I) allows higher production of BDO over BTO that is produced from a competing pathway.[154] 1,3-propanediol production from glucose was also facilitated by KivD. However, further study reveals that *Zm*PDC is a better enzyme to decarboxylate 4-hydroxy-ketobutyrate, a precursor for 1,3-propandiol production.[155] In a more complicated pathway, KivD has been utilized to produce 1,5-dipentadioic acid (glutarate) from glucose. Glutarate is an important building block for the synthesis of polyesters and polyamides (nylon-4,5 and nylon 5,5).[156] Another more sophisticated pathway was constructed *in vivo* to produce 6-aminocaproic acid from glucose. In this pathway

KivD and KdcA were used and compared. After 48 h fermentation, an engineered *E. coli* shows almost 4-fold higher titer of 6-aminocaproic acid when KdcA was used instead of KivD. 6-aminocaproic acid is a building block for nylon-6.[157]

The same as for ethanol, CBP of isobutanol production is an interesting approach. A possibility to produce bioisobutanol was first described about a decade ago; thus, much research is still on going. One important attempt toward CBP of bioisobutanol was performed by integration of a WT KivD to *G. thermoglucosidasius*. However, the titer is much lower than from *E. coli*. It was reasoned that the titer was hampered due to the stability of the KivD used. Hence, stability of a KDC (KivD or KdcA) is a prerequisite toward CBP of isobutanol. Prior to this work only limited efforts have been reported to improve thermostability of a KDC. Directed evolution combined with Rosetta Comparative Modelling resulted in a more stable variant (LLM4) bearing four amino acid substitutions (Q34H/V130I/A290V/ S386P). LLM4 shows a 13 °C improved T_m determined by CD as well as improved kinetic stability at 60 °C.[158] Another attempt to improve thermostability of KivD resulted in a variant with three amino acid substitutions (E156K/N351D/S385M). This variant was obtained by a combination of computational design and directed evolution. A FoldX force field analysis was used instead of Rosetta-based approach. The variant shows improved $T_{50}^{10\ min}$ (temperature at which a protein lost 50% of initial activity after 10 min incubation) by 9 °C.[159]

1.3.3 Dehydratases

Dehydratase, specifically hydro-lyase (EC 4.2.1.x), encompasses a diverse class of enzyme. This enzyme class cleaves C-O bonds *via* elimination thus producing a double bond and one water molecule. In brief, this enzyme class can be divided into three different subclasses, enzymes using an iron-sulfur cluster, enzymes using a metal ion, and enzymes using neither to catalyze a C-O elimination.[160]

1.3.3.1 [Fe-S]-dependent dihydroxy-acid dehydratase

Iron sulfur clusters [Fe-S] are perhaps the enzyme cofactor with the most diverse roles. This cluster has been reported to perform a single electron transfer, dehydration reaction by acting as a Lewis acids, stabilization of protein structure, regulation of metabolic pathways, biological sensors, and formation of protein bound radicals.[160,161] One class of [Fe-S]-dependent enzymes, dihydroxy-acid dehydratases (DHAD), in particular has gained increased attention in the last decades. This enzyme catalyzes the penultimate dehydration reaction in the branched-chain amino acids (BCAAs) biosynthesis (the Ehrlich pathway), thus widely spreads in organisms that are able to synthesize BCAAs, e.g. bacteria, archaea, fungi, and plant. Because animals lacking these pathways, this enzyme has been subjected as a promising target for inhibitor designs toward novel herbicides.[162,163] Recently, a new generation of antibiotics targeting this enzyme has also been proposed.[164,165]

As an [Fe-S]-dependent enzymes, DHADs have two different clusters: [2Fe-2S] and [4Fe-4S]. General consensus is that DHADs with [4Fe-4S] are instable in the presence of O_2 while the ones with [2Fe-2S] are stable.[160,161] Although different clusters are found, similar catalytic mechanisms are postulated. From isotopes and mechanistic inhibitor studies, it is proposed that dehydration mechanism is initiated by a proton abstraction of Cα (C2) by an amino acid of a DHAD acting as basic catalyst. The next step is an attack of [Fe-S] acting as a Lewis acid on the hydroxyl group of Cβ (C3). This then promotes formation of an enol intermediate, a double bond between Cα and Cβ, followed by concomitant loss of the hydroxyl group of Cβ. The enol intermediate then tautomerizes (keto-enol tautomerization) to a more stable α-keto group. Thus, the end of the reaction is the production of an α-keto acid and a molecular water.[162,166–168] Further crystallization of a DHAD from *Mycobaterium tuberculosis* and substrate docking studies support this proposal (Figure 1.8).[164]

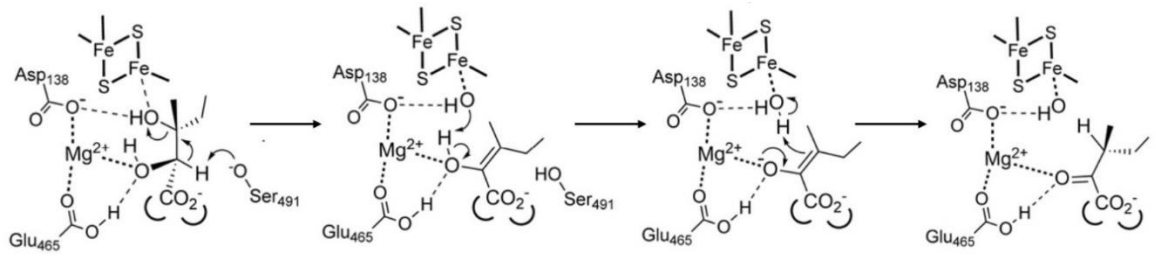


Figure 1.8. Proposed dehydration mechanisms of (2R,3R)-dihydroxy-3-methylvalerate based on docking study of the substrate to the crystal structure of *Mt*DHAD.[164]

In addition to the dehydration of (R)-2,3-dihydroxyisovalerate (DHIV) and (2R,3R)-dihydroxy-3-methylvalerate (DHMV), precursor for valine and isoleucine biosynthesis, other [Fe-S]-dehydratases catalyzing similar substrates have been reported in nature. Two other distinct classes include sugar acids and 6-phosphogluconate dehydratases (EC 4.2.1.12).[169,170] The first class is involved in pentose sugar metabolisms: the Dahm and the Weimberg pathways and the second class is in the Entner-Doudoroff (ED) pathway. *In vitro* and *in vivo* exploitation of the Dahm and the Weimberg pathways allow utilization of hemicellulose sugars toward fine chemicals synthesis.[171,172] In the classical ED pathway, modifications of the main pathway are also observed in nature. The main difference is at which branch phosphorylation occurs (Figure 1.9).[173] In the classical ED pathway, D-gluconate is phosphorylated to 6-phospho-D-gluconate which can be dehydrated by a 6-phospho-D-gluconate dehydratase. In the semi-phosphorylative ED pathway, 2-keto-3-deoxy-gluconate (a dehydration product of D-gluconate, either by a [Fe-S]-dependent sugar acid dehydratase or a metal dependent D-gluconate dehydratase) is phosphorylated, instead. The shortest pathway from D-glucose to reach pyruvate is the non-phosphorylative ED

(nED) pathway at which D-glycerate is phosphorylated. In the nED, only seven enzymes and two cofactors are needed.

Synthetic metabolic engineering of the nED pathway allows further reduction in the number of enzyme and cofactors requiring only four enzymes and one cofactor. Further combination of this new synthetic pathway with BCAA pathway facilitates production of isobutanol from D-glucose using eight enzymes and one cofactor. The key in the number of enzymes and cofactor reduction was achieved by utilization of the same dehydratase to dehydrate three intermediates: D-gluconate, D-glycerate, and DHIV (Figure 1.10).[151] All of the intermediates share similar configuration by having a hydroxyl group at R configuration in Cα and a hydroxyl in Cβ (S configuration for D-gluconate). A promiscuous DHAD from *Sulfolobus solfataricus* (also known as *Saccharolobus solfataricus*) was used to dehydrate the three substrates becoming the paramount enzyme in this cascade. This enzyme was proposed as a DHAD, rather than a [Fe-S] sugar acid dehydratase as *S. solfataricus* has another metal-dependent D-gluconate dehydratase (TIM-barrel enolase superfamily) catalyzing a dehydration of D-gluconate.[174,175]

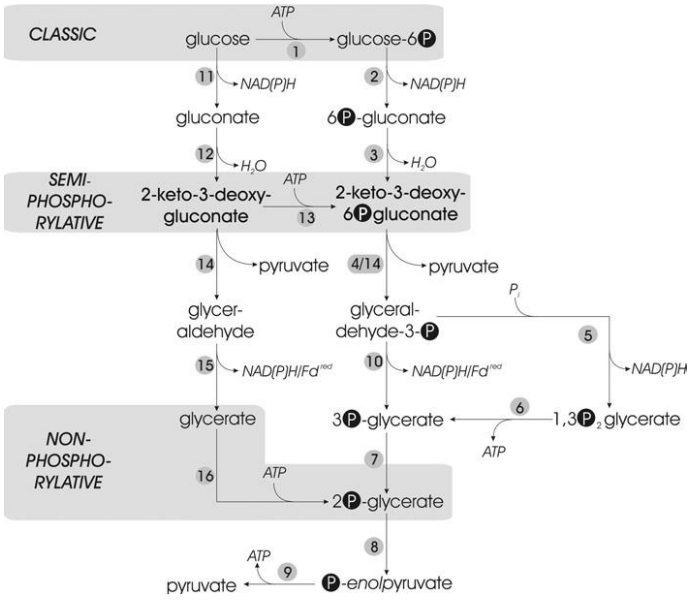


Figure 1.9. Different branches of the Entner-Doudoroff pathway for D-glucose metabolisms.[173]

This synthetic cascade later became the primary target to be introduced into *Saci* to realize bioalcohols production (see subchapter 1.3.1). The next challenge in introducing this pathway next to the stability of PDC and KDC (see subchapter 1.3.2.3 and 1.3.2.4) is the inherently low activity of the DHAD from *S. solfataricus* (*Ss*DHAD). Comparing activity of *Ss*DHAD toward the three substrates, dehydration of D-glycerate is by far the slowest by almost two-order of magnitude. Several attempts have been done to improve further activity of this enzyme *in vitro*, including enzyme engineering.[176–178] Thus, another approach namely enzyme discovery would be needed. There are two distinct enzyme classes that enable to perform abstraction of α-proton

and elimination of β-hydroxyl group: [Fe-S]-dehydratase and TIM-barrel sugar acid dehydratase. Dehydration of D-glycerate has not been identified in nature as this intermediate would be phosphorylated by a kinase prior to dehydration by an enolase (EC 4.2.1.11) to phosphoenol pyruvate (nED pathway). Thus, this reaction would be catalyzed rather as a side reactivity. Due to the diverse proficiency of [Fe-S] to sustain life, the hypothesis was that this class of dehydratase would have higher promiscuity than TIM-barrel sugar acid dehydratases. Study of *S. solfataricus* metabolism partly supported this hypothesis. *Ss*DHAD, a [Fe-S]-dependent dehydratase has higher promiscuity than *Ss*GAD (D-gluconate dehydratase), a TIM-barrel sugar acid dehydratase.[174–176,179]

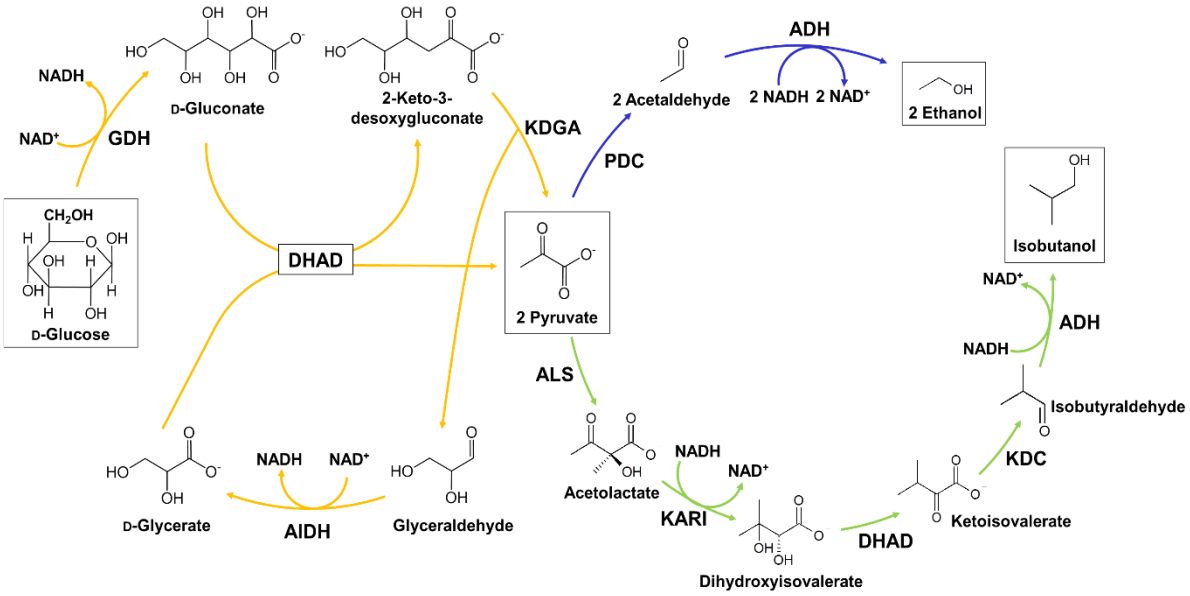


Figure 1.10. Synthetic metabolic engineering with minimized reactions toward production of alcohols. DHAD is used to dehydrate three substrates: D-gluconate, D-glycerate, and DHIV. The picture is provided by Dr. Jörg Carsten.

1.3.3.2 D,L-2-keto-3-deoxypentonate dehydratases

The oxidative pentose pathway was first proposed by Ralph Weimberg in the late 50s. In this pathway, a pentose sugar is first oxidized to its respective sugar acid. A dehydration step is then followed removing chirality at position C2 and C3. The result of this dehydration is D,L-2-keto-3-deoxypentonate. Due to the loss of functionality at C2 and C3, metabolism of ribose, arabinose, xylose, and lyxose will end up as the same intermediate, which now only differ as D- or L-stereoisomer. D and L functionality still persists as it is determined by the relative position of a hydroxyl group at position C4 (epimer). Dehydration further of D,L-2-keto-3-deoxypentonate at position C4 and C5 will results in α-ketoglutaric semialdehyde (KGSA) which is oxidized further to α-ketoglutaric acid (α-KG), an intermediate of TCA (tricarboxylic acid) cycle, or also known as citric acid cycle or Krebs cycle. This pathway facilitates metabolism of various pentose sugars to a common intermediate without the loss of any carbon.[180]

This Weimberg pathway has gained renewed interest among scientist, in particular toward valorization of hemicellulose sugars, D-xylose and L-arabinose.[171,172] Depending on the type of biomass, hemicellulose from softwood can contain up to 20% of L-arabinose while hardwood´s can contain up to 20% D-xylose.[181] For utilization and or modification of Weimberg pathway, the first oxidation step can be performed easily, either chemically or enzymatically.[182–184] The first dehydration step can be performed by a [Fe-S] dependent dehydratase or a TIM-barrel enolase sugar acid dehydratase discussed previously (see subchapter 1.3.3.1). The last oxidation step, oxidation of a terminal aldehyde of KGSA can be performed by a specific KGSA dehydrogenase or other promiscuous aldehyde dehydrogenases, including succinate semi aldehyde dehydrogenases.[185–187] The penultimate dehydration step, however, has only been reported by D-2-keto-3-deoxypetonate dehydratases (D-KdpDs) and L-KdpDs. Unequivocal reaction mechanism of this enzyme has not yet established. Nevertheless, from inhibition and isotopes studies of L-KdpD from *Pseudomonas saccharophila*, Schiff base-facilitated water elimination has been proposed for L-KdpD (Figure 1.11).[188] Crystallization of D-KdpD from *S. solfataricus* followed by substrate docking study proposed another reaction mechanisms for D-KdpD (*Figure* 1.12).[189] In the first study, no specific role of a metal ion is suggested, while in the second study a magnesium is reported to play role in substrate activation.[188,189]

Although D- and L- KDP only differ in the position of OH at position C4, the enzymes that catalyze the reaction D- and L-KdpD are reported to belong to different superfamilies. From sequence-based automated annotation, D-KdpD is classified to fumarylacetoacetate hydrolase (FAH) superfamily while L-KdpD belongs to dihydrodipicolinate synthase and N-acetylneuraminate lyase (DHDPS/NAL) superfamily.[190,191] So far, only few studies using purified enzymes of D- and L-Kdp dehydratases are reported in literatures. For D-KdpD, enzymes from *S. solfataricus*, *Caulobacter crescentus*, and *Paraburkholderia xenovorans* have been reported. No kinetic characterization of *Ss*D-KdpD, however, was reported although its crystal structure has been solved.[189,190] *Cc*D-KdpD (k_{cat}: 0.5 s^{-1} and K_M: 1.9 mM) shows almost 2-fold lower catalytic efficiency than *Bx*D-KdpD (k_{cat}: 4.7 s^{-1} and K_M: 9 mM).[154] For L-KdpD, only a dehydratase from *P. saccharophila*, *Azospirillum brasilense*, and *Bukholderia multivoransare* are reported. The first and elegant characterization of L-KdpD from *P. saccharophila* unfortunately does not provide its protein sequence.[192] A homologous enzyme from *Azospirillum brasilense* with protein sequence is reported to close the gap. *Ab*L-KdpD shows an activity of 20 U/mg (~11 s^{-1}), a much higher value than *Bm*L-KdpD, which shows k_{cat} of 0.2 s^{-1}.[154,191] In addition to α-KG, modification of Weimberg has been used toward production of BDO and mesaconate. Both compounds are important building blocks for the synthesis of several polymers.[154,193] Successful production of α-KG, BDO, and mesaconate from hemicellulose sugars will require highly active D- and L-Kdp dehydratases as high enzyme activity is prerequisite toward industrial applications (see subchapter 1.1.1.2).

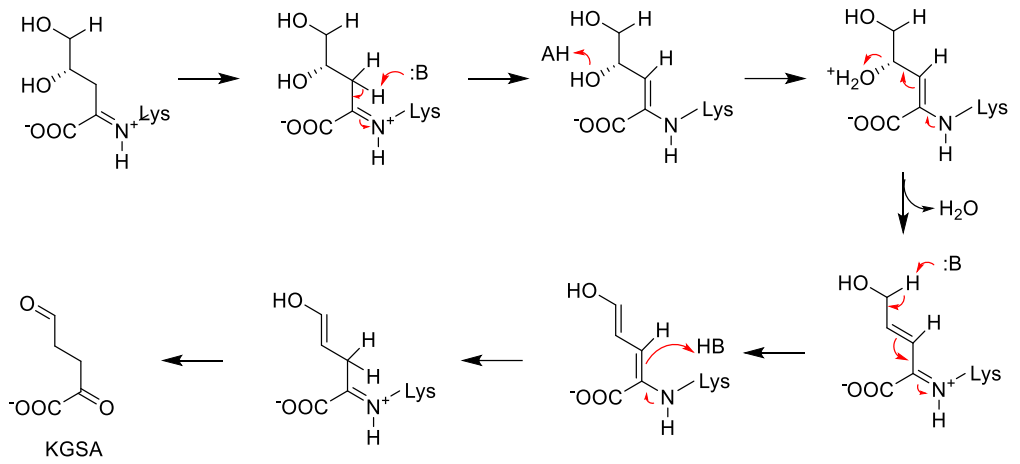


Figure 1.11. Proposed reaction mechanism of PsL-KdpD via formation of a Schiff base with a conserved lysine residue.[188]

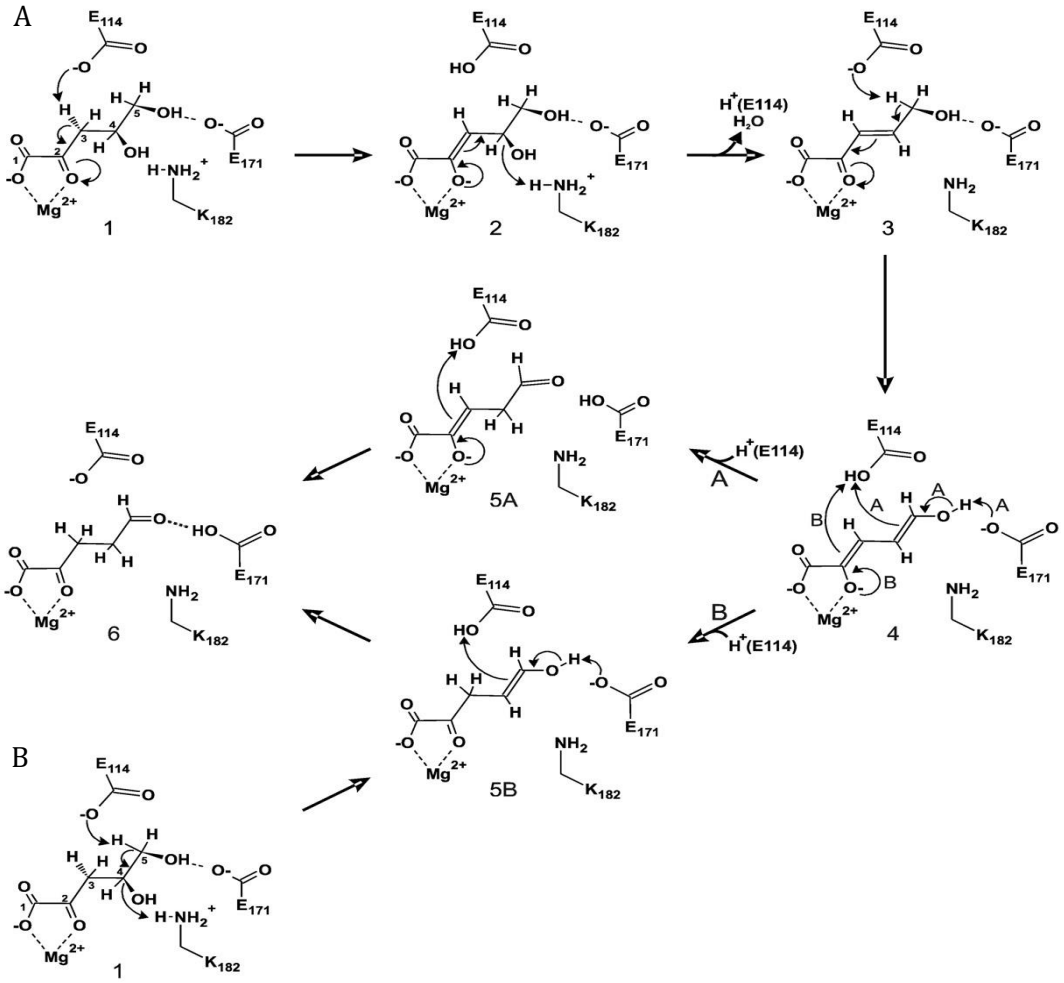


Figure 1.12. Proposed catalytic mechanisms of SsD-KdpD toward D-2-keto3-deoxypentonate. First scenario starts via $E_2\alpha\beta$ (A) and the second scenario starts via $E_2\beta\gamma$ (B).[189]

1.4 Aims of this work

1.4.1 Engineering thermostability of decarboxylases toward bioalcohols production in *Sulfolobus acidocaldarius*

Two ThDP-dependent decarboxylases, a PDC and a KDC, are only kown from mesophilic organisms. These enzymes have limited thermostability and thus direct integration into *Saci* is not readily possible. Unfortunately, there are no homologues of either decarboxylases found in thermophilic microorganisms. Therefore, enzyme engineering techniques should be applied to improve their stability.

One isoenzyme of KDC, KivD has been engineered using a combination of directed evolution and computational design (Rosetta Homology Modelling and FoldX force field). In these works, the stability of KivD were successfully improved. This significant improvement was not enough for application in *Saci* that grows at 70 °C. However, this important work might suggest that different approaches should be performed to increase further stability of a KDC. Crystal structures of another KDC isosenzyme, KdcA has been resolved with good resolutions. These crystal structures should be then used to evolve KdcA using semi rational approach. Several factors contributing to improved stability would be considered. A cheap and robust high-throughput screening (HTS) would be developed to find desired variants of KdcA stable at 70 °C. The variants with improved stability would be tested further in the context of stability in isobutanol. In addition to bioproduction of isobutanol by *Saci*, a cell-free approach is an interesting alternative. In order to function better in a cell-free approach, more thermostable and isobutanol-stable variants would be required.

Unlike KDCs, there were already several PDCs cloned and characterized prior to this study. However, several differences are also observed, in particular for their thermostability. Thus, several promising PDCs from the literatures would be cloned and analyzed under the same experimental conditions. Three or more PDCs would be used as the templates for directed evolution. StEP would be used to shuffle the parental DNAs due to its proven robustness to improve enzyme thermostability. Directed evolution was chosen as two previous works on evolving thermostability of a PDC relied on rather rational approaches (ASR and Rosetta-based prediction) resulted in no significant improvement. Again, HTS system would be developed to screen PDC libraries. Several semi rational design approaches would also be performed to increase further the thermostability of PDC variants. The same goal as KdcA, i.e. a stable enzyme at 70 °C was targeted. Moreover, a PDC has been applied to produce 1-butanol *in vitro*.[194,195] Therefore, the improved variants obtained would be subjected to stability test toward 1-butanol.

1.4.2 Discovery of new [Fe-S] dehydratases toward higher activity on sugar acids

For the synthesis of ethanol, two sugar acids: D-gluconate and D-glycerate need to be dehydrated. DHIV is the third sugar acid derivative that needs to be dehydrated for the synthesis of isobutanol. *Ss*DHAD that was used for the cell-free synthesis of ethanol and isobutanol shows very low activity, in particular for D-glycerate. Two sugar acid dehydratases from mesophilic microorganisms were reported to have higher activity than *Ss*DHAD on D-gluconate and other sugar acids. Due to limited work on this class of enzyme, it would be necessary to clone more enzymes to advance the knowledge prior to the engineering approach. Thus, several dehydratases would be cloned, expressed, and characterized for their activities toward D-glycerate, D-gluconate, and DHIV. From the enzymes characterized, signature motifs should be identified to guide the discovery of new enzymes with activity toward the sugar acids of at least 1 U/mg. The best variants would also be tested in cell-free cascades for the production of chemicals.

1.4.3 Identification and characterization of D-Kdp and L-Kdp dehydratases

So far, only limited number of D-Kdp and L-Kdp dehydratases were reported in literature. These enzymes have potential biotechnological applications for the valorization of hemicellulose sugars—D-xylose and L-arabinose. Hence, several new D-Kdp and L-Kdp dehydratases would be cloned and characterized in respect of activity, stability, intermediates inhibition, and pH profile. The best variants would later be tested toward one-step conversion of high substrate load (>500 mM).

2 Materials and Methods

Methods used in this study were derived directly or indirectly from Standard Operational Procedure (SOP) of the Chair of Chemistry of Biogenic Resources, TUM Campus Straubing for Biotechnology and Sustainability. In this section, only general materials and methods were presented. Detailed experiments for specific work will be explained in Chapter 3.

2.1 Materials

2.1.1 Chemicals

All common chemicals used in this study are presented below,

Chemicals	Manufacturer	Catalog number
Yeast extract	Carl Roth Gmbh & Co. Kg	2363
Peptone	Carl Roth Gmbh & Co. Kg	8952
NaCl	Carl Roth Gmbh & Co. Kg	P029
Agar	Carl Roth Gmbh & Co. Kg	5210
Glycerol 99.5%	Carl Roth Gmbh & Co. Kg	3783
KH_2PO_4	Carl Roth Gmbh & Co. Kg	3904
K_2HPO_4	Carl Roth Gmbh & Co. Kg	P749
Na_2PO_4	Carl Roth Gmbh & Co. Kg	P030
α-D-glucose monohydrate	SERVA Electrophoresis GmbH	22720
α-Lactose monohydrate	Carl Roth Gmbh & Co. Kg	6868
Kanamycin	Carl Roth Gmbh & Co. Kg	T832
Carbenicillin	Carl Roth Gmbh & Co. Kg	6344.3
IPTG (Isopropyl β- d-1-thiogalactopyranoside)	Carl Roth Gmbh & Co. Kg	CN08
Imidazole	Merck Millipore	1.047161
HCl 37 w/v%	Merck Millipore	30721
H_2SO_4 96 w/v%	Carl Roth Gmbh & Co. Kg	4623
TRIS (Tris(hydroxymethyl)aminomethane)	Sigma Aldrich	T1503
HEPES (4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid)	Carl Roth Gmbh & Co. Kg	HN78
Agarose	SERVA Electrophoresis GmbH	11404
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth Gmbh & Co. Kg	8040
Acetic acid glacial (100%)	Carl Roth Gmbh & Co. Kg	6755
Serva DNA Stain clear G	SERVA Electrophoresis GmbH	39804
$CaCl_2.2H_2O$	Carl Roth Gmbh & Co. Kg	52395
$MgCl_2.6H_2O$	Carl Roth Gmbh & Co. Kg	21891
Acrylamide/Bis-solution 30 % (37.5:1)	Carl Roth Gmbh & Co. Kg	3029
TEMED (N,N,N',N'-Tetramethylethylenediamine)	VWR International	M146
APS (Ammonium Persulfate)	Merck Millipore	1.01201
Coomassie Brilliant Blue G250	SERVA Electrophoresis GmbH	17524
SDS (Sodium Dodecyl Sulfate)	SERVA Electrophoresis GmbH	20760
Glycine	Carl Roth Gmbh & Co. Kg	3790
β-mercaptoethanol	Merck Millipore	8.05740
Bromophenol blue	Merck Millipore	1.08122
Sodium pyruvate	Carl Roth Gmbh & Co. Kg	8793
Sodium 2-oxo-butanoic acid	Sigma Aldrich	K0875
Sodium 2-oxo-3-methyl-butanoic acid (α-ketoisovalerate)	Sigma Aldrich	198994
Sodium 2-oxo-valeric acid	Sigma Aldrich	212768

(α-ketovalerate)		
NADH (Nicotinamide Adenine Dinucleotide reduced)	Carl Roth Gmbh & Co. Kg	AE12
ThDP (Thiamine Diphosphate)	Sigma Aldrich	C8754
D-Gluconate	Sigma Aldrich	S2054
D,L-Glycerate	TCI Deutschland	D0602
Ethanol absolute	VWR Chemicals	20821.321
Isobutanol (2-methyl-propanol)	Sigma Aldrich	33064
n-Butanol	Merck Millipore	1.01988
$(NH_4)HCO_3$	Sigma Aldrich	09830
BSA (Bovine Serum Albumin)	Ammresco	0332
Polymixin-B-sulfate	Applichem	A0890
SYPRO® Orange Protein Gel Stain	Sigma Aldrich	MKCH9337

2.1.2 Kits

All kits used in this study are presented below,

Kit	Manufacturer	Purpose
GeneJet Plasmid Miniprep Kit	Thermo Fischer Scientific	Plasmids isolation from E. coli
NucleoSpin Gel and PCR Clean up	Machery and Nagel	Agarose gel or PCR product purification
DNeasy UltraClean Microbial Kit	Qiagen	gDNA isolation

2.1.3 Enzymes

All commercial enzymes used in this study are presented below,

Enzymes	Manufacturer	Catalog number
DNase I	AppliChem	A3778
Yeast alcohol dehydrogenase	Sigma Aldrich	A7011-30KU
Lysozyme	Carl Roth Gmbh & Co. Kg	8259
T4 DNA Ligase	New England Biolabs GmbH	M0202
Taq DNA Polymerase	New England Biolabs GmbH	M0267
Phusion High Fidelity Polymerase	New England Biolabs GmbH	M0530
DpnI	New England Biolabs GmbH	R0176
XbaI	New England Biolabs GmbH	R0145
NheI	New England Biolabs GmbH	R0131
NdeI	New England Biolabs GmbH	R0111
XhoI	New England Biolabs GmbH	R0146
HindIII	New England Biolabs GmbH	R0104
EcoRI	New England Biolabs GmbH	R0101
Quick CIP (calf intestinal phosphatase)	New England Biolabs GmbH	M0525

2.1.4 Plasmids

All plasmids used and generated in this study are presented below,

Plasmid	Description	Reference
pET28a	A vector that carries an N-terminal Histag and an optional C-terminal Histag. Operated under T7 promoter with kanamycin as a selection marker.	Novagen
pET24a	A vector carrying a C-terminal Histaq. Operated under T7 promoter with kanamycin as a selection marker.	Novagen

pBAD	A vector carrying an N-terminal histag. Operated under arabinose promoter with ampicillin or carbenecillin as selection marker	Invitrogen
pET28Nhis_LlkdcA	A pET28a plasmid containing a N-terminal histag tag fused to a kdca gene from *Lactococcus lactis*	[132]
pET28Nhis_LlkdcA (variants)	For all kdcA mutant genes generated in this study, see subchapter 3.1	This study
pET24Chis_*Cg*pdc	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Candida glabrata*	This study
pET24Chis_*Zr*pdc	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Zygosaccharomyces rouxii*	This study
pET24Chis_*Ap*pdc	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Acetobacter pasteurianus*	This study
pET24Chis_*Zm*pdc	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Zymomonas mobilis*	This study
pET24Chis_*Zp*pdc	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Zymobacter palmae*	This study
pET24Chis_5npu	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from ASR reconstruction	This study
pET24Chis_5tma	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Zymomonas mobilis* bearing 16 point mutations	This study
pET24Chis_pdc-Var.1	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Acetobacter pasteurianus bearing 13 point mutations*	This study
pET24Chis_pdc-Var.2	A pET24a plasmid containing a C-terminal histag tag fused to a pdc gene from *Acetobacter pasteurianus bearing 14 point mutations*	This study
pET28Nhis_*Rl*aradht	A pET28a plasmid containing a N-terminal histag tag fused to a aradht gene from *Rhizobium leguminosarum*	This study
pET28Nhis_*Cc*xyldht	A pET28a plasmid containing a N-terminal histag tag fused to a xyldht gene from *Caulobacter crescentus*	This study
pET28Nhis_*Cn*H16dhad	A pET28a plasmid containing a N-terminal histag tag fused to a dhad gene from *Cupriavidus necator* H16	This study
pET28Nhis_*Cn*N1dhad	A pET28a plasmid containing a N-terminal histag tag fused to a dhad gene from *Cupriavidus necator* N-1	This study
pET28Nhis_*Cm*dhad	A pET28a plasmid containing a N-terminal histag tag fused to a dhad gene from *Cupriavidus metallidurans*	This study
pET28Nhis_*St*dhad	A pET28a plasmid containing a N-terminal histag tag fused to a dhad gene from *Schlegelella thermodepolymerans*	This study
pET28Nhis_*Ft*dhad	A pET28a plasmid containing a N-terminal histag tag fused to a dhad gene from *Fontimonas thermophila*	This study
pET28Nhis_*Tv*dhad	A pET28a plasmid containing a N-terminal histag tag fused to a dhad gene from *Thermosynecoccus vulcanus*	This study
pET28Nhis_*Rl*xyldht	A pET28a plasmid containing a N-terminal histag tag fused to a xyldht gene from *Rhizobium leguminosarum*	This study
pET28Nhis_*Hs*xyldht	A pET28a plasmid containing a N-terminal histag tag fused to a xyldht gene from *Herbaspirillum seropedicae*	This study
pET28Nhis_*Hs*aradht	A pET28a plasmid containing a N-terminal histag tag fused to a aradht gene from *Herbaspirillum seropedicae*	This study
pET28Nhis_*Hs*dht	A pET28a plasmid containing a N-terminal histag tag fused to a dht gene from *Herbaspirillum seropedicae*	This study
pET28Nhis_*Vp*dht	A pET28a plasmid containing a N-terminal histag tag fused to a dht gene from *Variovorax paradoxus*	This study
pET28Nhis_*Rf*dht	A pET28a plasmid containing a N-terminal histag tag fused to a dht gene from *Rhodanobacter fulvus*	This study
pET28Nhis_*Nm*dht	A pET28a plasmid containing a N-terminal histag tag fused to a dht gene from *Noviherbaspirillum massiliense*	This study
pET28Nhis_*Pu*dht	A pET28a plasmid containing a N-terminal histag tag fused to a aradht gene from *Paralcaligenes ureilyticus*	This study
pET28Nhis_*Rh*dht	A pET28a plasmid containing a N-terminal histag tag fused to a aradht gene from *Ramlibacter henchirensis*	This study
pET28Nhis_*Cn*kdad	A pET28a plasmid containing a N-terminal histag tag fused to a kdad gene from *Cupriavidus necator* H16	This study

pET24Chis_*Cn*kdad	A pET24a plasmid containing a C-terminal histag tag fused to a kdad gene from *Cupriavidus necator* H16	This study
pET28Nhis_*Ab*kdad	A pET28a plasmid containing a N-terminal histag tag fused to a kdad gene from *Azospirillum brasilense*	This study
pET28Nhis_*Hs*kdad	A pET28a plasmid containing a N-terminal histag tag fused to a kdad gene from *Herbaspirillum seropediceae*	This study
pET28Nhis_*Vp*kdad	A pET28a plasmid containing a N-terminal histag tag fused to a kdad gene from *Variovorax paradoxus* EPS	This study
pET24Chis_*Pp*kdxd	A pET24a plasmid containing a C-terminal histag tag fused to a kdxd gene from *Pseudomonas putida*	This study
pET24Chis_*Px*kdxd	A pET24a plasmid containing a C-terminal histag tag fused to a kdxd gene from *Paraburkholderia xenovorans*	This study
pET24Chis_*Hs*kdxd	A pET24a plasmid containing a C-terminal histag tag fused to a kdxd gene from *Herbaspirillum seropediceae*	This study
pET8Nhis_*Cc*kdxd1	A pET28a plasmid containing a N-terminal histag tag fused to a kdxd1 gene from *Caulobacter crescentus*	This study
pET24Chis_*Cc*kdxd1	A pET24a plasmid containing a C-terminal histag tag fused to a kdxd1 gene from *Caulobacter crescentus*	This study
pET24Chis_*Cc*kdxd2	A pET24a plasmid containing a C-terminal histag tag fused to a kdxd2 gene from *Caulobacter crescentus*	This study
pET24Chis_*Cn*kdxd	A pET24a plasmid containing a C-terminal histag tag fused to a kdxd gene from *Pseudomonas putida*	This study
pET28Nhis_*Bs*GDH	A pET28a plasmid containing a N-terminal histag tag fused to a gdh mutant gene from *Bacillus subtilis*	This study
pET28Nhis_*Pt*kdgA	A pET28a plasmid containing a N-terminal histag tag fused to a kdga gene from *Picrophilus torridus*	[196]
pET28Chis_*Ta*Aldh_M42	A pET28a plasmid containing a C-terminal histag tag fused to a aldh mutant gene from *Thermoplasma acidophilum*	[197]
pET28Chis_*Bst*ADH	A pET28a plasmid containing a C-terminal histag tag fused to a adh gene from *Bacillus stearothermophilus*	[151]
pQE80LNhis_*Sc*Aldo	A pQE80L plasmid containing a N-terminal histag fused to a aldo mutant gene from *Streptomyces coelicolor*	[18]

2.1.5 Single strand oligonucleotide (primers)

All primers were ordered from Eurofins Genomic (Ebesberg, Germany) or Biomer (Ulm, Germany) and presented below,

Primer name	Sequence (5´to 3´)	Purpose
T7Fwd_Mut	GAATTGTGAGCGGATAACAATTCCC	To amplify a gene under T7 promoter
T7Rev_Mut	CTTTGTTAGCAGCCGGATCTC	
pBAD_Fwd	ATGCCATAGCATTTTTATCC	To amplify a gene in a pBAD vector
pBAD_Rev	CCTTGAATACACCATGTAGTG	
QCs_kdcA	QuikChange primers for creating kdcA library are presented in subchapter 3.1	
QCs_pdc	QuikChange primers for creating PDC library are presented in subchapter 3.2	
Dehydratases cloning	Primers for amplifying dehydratases genes are presented in subchapter 3.3	
L-kdpD and D-kdaD cloning	Primers for amplifying genes encoding L-KdpD and D-KpdD are presented in subchapter 3.4	

2.1.6 Bacterial strains

Two *E coli* strains (Top10 and DH5α) were used for standard plasmid generation and one *E. coli* strain (BL21 (DE3)) was used for protein expression.

- *E. coli* Top 10 (Invitrogen, Germany): F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G

- *E. coli* DH5α (Invitrogen, Germany): F^- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-
- *E. coli* BL21 (DE3) (Novagen, Germany): *fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*

List of microorganisms used for cloning [Fe-S] dehydratases, KDA dehydratases, KDX dehydratases were presented below. All microorganisms were purchased from Leibniz Institute DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulture) GmbH, unless otherwise stated.

Microroganisms	Strains	DSMZ catalogue number
Cupriavidus necator	H16	428
Cupriavidus necator	N1	13513
Cupriavidus metallidurans	CH34	2839
Schlegelella thermodepolymerans	R-16654	15344
Fontimonas thermophila	HA-01	23609
Rhizobium leguminosarum	3D1K2	6040
Herbaspirillum seropediceae Z67	Z67	6445
Variovorax paradoxus	B4	21876
Rhodanobacter fulvus	Jip 02	18449
Noviherbaspirillum massiliense	JC206	25712
Paralcaligenes ureilyticus	GR24-5	24591
Ramlibacter henchirensis	TMB834	14656
Pseudomonas putida	KT2440	6125
Variovorax paradoxus	EPS	Obtained from Prof. Dirk Tischler (Ruhr-Universität Bochum)

2.1.7 Software

All software used are listed below,

Software	Producer	Description
ProtParam	Swiss Institute of Bioinformatics	Computation of various physical and chemical parameters based on a protein sequence
Clone Manager 9	Scientific & Educational Software	In silico DNA manipulation
PyMol 1.8.6.2	Schrodinger, LLC	Molecular visualization system
SigmaPlot 13.0	Systat Software, Inc	Statistical analysis software, including enzyme kinetics
Unicorn 7.0	GE Healtcare	Control software for ÄKTA protein purification systems.
Bio-Rad CFX Manager	Bio-Rad Laboratories GmbH	Real-time PCR data acquisition and analysis
Intas	Intas Science Imaging Instruments GmbH	Gel documentation
YASARA	YASARA Biosciences GmbH	Molecular-graphics, modelling, and simulation program
PDBePISA	European Bioinformatics Institute (EMBL-EBI)	Interactive tool for the exploration of macromolecular interfaces
Chromeleon	Thermo Fischer Scientific	Chromatography Data System
ChemDraw 12.0	PerkinElmer	Molecule Editor

2.1.8 Devices

All devices utilized in this study is presented below,

Devices	Model	Manufacturer
Benchtop centrifuge	Galaxy MiniStar	VWR
	Heraus Frisco 21	Thermo Fischer Scientific
	Heraus Pico 17	
Biological safety cabinet (sterile bench)	MSC Advantage	Thermo Fischer Scientific
Centrifuges and rotors	Sorvall RC-6+ F9-4x1000Y F10S-6x500Y SS-34	Thermo Fischer Scientific
	Sorvall Lynx 6000 F9-6x1000 LEX F12-6x500 LEX A27-8x50	Thermo Fischer Scientific
	Rotanta 460R	Hettich Lab Technology
Electric Balances	TE1502S	Sartorius AG
	TE6101	Sartorius AG
	Pioneer	Ohaus Europe GmbH
	AW320	Shimadzu Corporation
Electrophoresis (DNA)	Mini-Sub® Cell GT Cell	Bio-Rad Laboratories GmbH
	Wide Mini-Sub® Cell GT Cell	
Electrophoresis (protein)	Mini-PROTEAN® Tetra Vertical Electrophoresis Cell	Bio-Rad Laboratories GmbH
Freezer -20	Liebher Compact	Liebher
Freezer -80	Forma 900 series	Thermo
FPLC	Äkta Purifier UPC 900	GE Healtcare Life Sciences
FPLC columns	5 ml HisTrap FF Crude	GE Healtcare Life Sciences
	HiPrep 26/10 Desalting	GE Healtcare Life Sciences
GelDock	Intas	
Heating block/Thermoshaker	Eppendorf ThermoMixer® C	Eppendorf AG
	SLG digital dry bath	Süd-Laborbedarf GmbH
	Tmix	Analytik Jena AG
Homogenizer	Ultra Turrax T18 Basic	IKA
HPLC	Ultimate 3000	Dionex Corporation
HPLC Detectors	RS Diode Array	
	Ultimate 3000 Photodiode Array	
	Refractive Index Detector	Shodex
HPLC column	Metrosep A Supp 16-250	Metrohm
	Rezex ROA-Organic Acid H+(8%)	Phenomenex
Incubator	Heraus B12	Thermo Fischer Scientific
	Heraus Kelvitron kp	
	KBF 240	Binder
Incubation shakers	Multitron Standard	Infors
	New Brunswick™ Innova® 44/44R	Eppendorf AG
Microplate reader	IKA KS 4000 ic central	IKA Laboratory Equipment
	Multiskan Spectrum	Thermo Fischer Scientific
	Varioskan	Thermo Fischer Scientific
	Epoch 2	BioTek Instruments, Inc.
PCR thermocyclers	C1000 Thermal Cycler	Bio-Rad Laboratories GmbH
	MJ Mini	
	My Cycler	
	CFX96 Optical Reaction Module	
pH meter	FiveEasy	Mettler-Toledo GmbH
	FiveGo	
pH meter electrode	LE438	
	InLab® Micro Pro-ISM	
Pipettes (manual)		Brand
Pipettes (manual multichannel)		Brand
Pipettes (electric multichannel)		Eppendorf
Shaker	MaxQ 2000	Thermo
Shaker (nutating)	Rocking Platform	VWR International GmbH

Shaker (plate)	TiMix 5 control	Edmund Bühler GmbH
Spectrophotometer	NanoDrop 2000	Thermo Fischer Scientific
Sonicator	Ultrasonic Technology UIS 250 V	Hielscher Ultrasonics
Sonotrode	LS24d10	
Steam Steriliser (Autoclave)	Varioklav135S	Thermo Fischer Scientific
Titrator	TitroLine	SI Analytics
Transilluminator	DR 46B	Clara Chemical Research
Water bath	ED-Brü	Julabo Labortechnik GmbH
	CC1 Compatible Control	Huber
	Microprocessor Control MPC	Huber
Water purification system	Q-POD MiliQ	Merck Millipore

2.2 Methods

2.2.1 For DNA manipulation

2.2.1.1 Isolation of genomic DNA (gDNA)

To isolate gDNA, vegetative cells or growing cell were used. The gDNA was isolated using DNeasy UltraClean Microbial Kit using the protocols described by the manufacturer.

2.2.1.2 Isolation of plasmids from E.coli

To purify plasmids from *E. coli*, GeneJET Plasmid Miniprep Kit was used. The standard procedure described by the manufacturer was followed.

2.2.1.3 Amplification of a DNA fragment from gDNA

A 2-step PCR protocol described by NEB was followed. In brief, 50 ng of gDNA, 0.5 mM of forward and reverse primers, phusion polymerase 1 U, 1x GC buffer, ddH_2O up to 50 µl were mixed in a 200 µl PCR tube. PCR reaction was as follow: 98 °C for 30 s for initial denaturation, 30 cycles of 98 °C for 10 s and 72 °C for 1 min (30 s per 1 kb), 72 °C for 5 min for final extension, 16 °C for indefinite hold.

2.2.1.4 QuikChange (QC) PCR

For QC PCR, 50 ng of plasmid, 0.5 mM of forward and reverse primers, phusion polymerase 1 U, 1x HF buffer, 5 ml DMSO, ddH_2O up to 50 µl were mixed in a 200 µl PCR tube. PCR reaction was as follow: 98 °C for 30 s for initial denaturation, 18 cycles of 98 °C for 10 s, 64 °C for 30 s, 72 °C for 3.5 min (30 s per 1 kb), 72 °C for 5 min for final extension, and then kept at 16 °C until further use. After the PCR was completed, 20 U DpnI was added to the PCR solution and the PCR tube was incubated at least for 6 h at 37 °C. PCR product was purified using NucleoSpin® Gel and PCR Clean-up Kit using the protocol described by the manufacturer. Final elution volume is 25 µl. For chemically transforming a *E. coli* strain, 10 µl (about 500 ng DNA) of the elution was used.

2.2.1.5 Colony PCR

For standard cPCR, a standard protocol by NEB was adapted. In short, *E. coli* colonies were transferred to 200 µl PCR cups using toothpicks. A master mix of PCR reaction consisted of 0.5 mM

of forward and reverse primers, NEB Taq polymerase 20 U/ml, 1x ThermoPol buffer, ddH_2O up to desired volume was made. The PCR reaction was transferred to the PCR cups containing *E. coli* colonies. PCR reaction was as follow: 95 °C for 5 min for initial denaturation, 30 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 2 min (1 min per 1 kb), 72 °C for 5 min for final extension, and then kept at 16 °C until further use.

2.2.1.6 Restriction and ligation

Appropriate plasmid and PCR products (2000 ng) were restricted using appropriate restriction enzymes (20 U) at 37 °C for 3 h. Quick CIP (5 U) was added to the restriction solution of the plasmids to dephosphorylate the cut vector to reduce self-ligation background of the vector after 2 h incubation at 37 °C. After restriction was finished, the restricted DNAs (vector and insert) were run in 1% agarose gel. The correct DNA bands were excised using X-tracta gel extraction tool (Sigma Aldrich, Germany) and purified using NucleoSpin® Gel and PCR Clean-up Kit following the standard protocol described by the manufacturer.

2.2.1.7 DNA separation using electrophoresis

Agarose 1 w/v% was made by weighing certain amount of agarose powder and dissolved in 1x TAE buffer. Microwave-assisted heating was needed to completely dissolve the agar. The solution was stored at 65 °C prior to use.

50x TAE buffer consisted of 2 M TRIS, 0.05 M EDTA, 57.1 ml acetic acid glacial, and ddH_2O up to 1 L.

To perform DNA electrophoresis, 100 µl of the DNA staining dye was added to a beaker glass and followed by 45 ml agarose gel (200 µl of the DNA stain for 100 ml agarose gel) and mixed. The 40 to 50 ml agarose gel was poured to a small sample tray or 100 to 125 ml to a big tray. A special gel comb was attached on the sample tray to form wells with appropriate sample sizes. After the agar was solidified at room temperature, the tray was transferred to a DNA electrophoresis chamber. TAE buffer 1X (50X TAE stock was dissolved in ddH_2O) was added to the electrophoresis chamber and filled until covering the entire gel (few mm solution above the agarose gel). DNA samples were mixed with 6x purple loading dye (NEB, Germany) prior to loading to the wells. In total 10 µl to 55 µl of the DNA sample mix could be inserted to the wells depending on the type of the gel combs used. A 2-log DNA marker (NEB, Germany) was also loaded to at least one of the wells as reference. The electrophoresis chamber was then closed and electrophoresis was started by applying constant potential (120 volt) for 30 to 40 min. To observe the DNA separation profile, the agarose gel was transferred to a GelDock with a camera connected to a computer. UV light was switched on and DNA separation profiles were visualized by Intas Software in the computer. To excise DNA bands from the gel, a transilluminator was used to visualize the separation profiles, instead.

2.2.1.8 Transformation of chemical competent E. coli cells

An Eppendorf tube containing chemical competent cells was thawed in ice for 10 to 15 min. Appropriate amount of DNA was transferred to the tube and mix by gently tapping the tube. After 30 min incubation in ice, transformation was performed by incubating the tube in a water bath that had been set to 42 °C. The incubation was carried out for 1 min. After the heat shock, the tube was directly chilled in ice for another 5 min. LB media 0.9 ml was added to the tube for cell outgrowth. Outgrowing the cells by incubation at 37 °C, 150 rpm for 1 h prior to platting to a LB-agar plate containing a selection antibiotic, e.g. kanamycin.

2.2.1.9 DNA sequencing

Sequencing of plasmid DNA was performed by Eurofins Genomics GmbH, Germany.

2.2.1.10 Preparation of chemical competent cells

A single colony of desired *E. coli* strain was transferred to a 100 ml baffled Erlenmeyer flask containing 10 ml LB medium. The culture was grown overnight at 37 °C, 150 rpm. All culture was transferred to a 2 L baffled Erlenmeyer flask containing 500 ml LB media (1/50 dilution) under a sterile bench. The 500 ml culture was incubated at 37 °C, 120 rpm until OD reached 0.4. The culture was pelleted by centrifugation at 2000 xg for 10 min, at 4 °C. After this step, all subsequent steps were performed in ice. Following centrifugation, the supernatant was decanted and the pellet was resuspended with 50 ml of sterile and ice-cold 100 mM $CaCl_2$. The solution was transferred to a sterile 50 ml Falcon tube (Sarstedt, Germany) and incubated in ice for 1 h. Centrifugation at 4 °C, 800 xg for 10 min was performed to pellet the cells. After decanting the supernatant, the cells´ pellet was resuspended with 5 ml of sterile and ice-cold 85 mM $CaCl_2$ and 15 w/v% glycerol. Incubation in ice for 1 h was further performed. Afterwards, the cells were aliquoted as 100 µl solution in a 1.5 ml Eppendorf tube and directly flash frozen in liquid nitrogen. The frozen competent cells were stored at -80 °C until further use.

2.2.2 For protein production and analysis

2.2.2.1 Preparation of cultivation media and antibiotics

Luria-Bertani (LB) broth and agar

5 g/l yeast extract, 10 g/l peptone, 10 g/l NaCl, and ddH_2O until 1 L. For LB agar, 15 g/l agar was added in the mix. The medium was sterilely autoclaved at 121 °C, 15 min, 1 bar.

Terrific broth (TB)

24 g/l yeast extract, 12 g/l peptone, 5 w/v% glycerol, up to 0.9 l ddH_2O. The medium was sterilely autoclaved at 121 °C, 15 min, 1 bar.

10x TB salt contains 23.1 g/l and 125.4 g/l K_2HPO_4 in ddH_2O. The 10X salt was sterilely autoclaved at 121 °C, 15 min, 1 bar.

After the medium and the 10x salt were cooled reaching room temperature, 0.1 L of the salt was added to the medium making 1 L TB medium under sterile condition (sterile bench).

Auto induction (AI) medium

Auto induction medium was made by mixing four main components: 925 ml ZY medium, 50 ml 20x NPS, 20 ml 50x 5052, and 1 ml $MgSO_4$ 1 M. Each component was autoclaved at 121 °C, 15 min, 1 bar separately.

ZY medium contains 5 g/l yeast extract, 10 g/l peptone, and up to 925 ml ddH_2O.

20x NPS contains 66 g/l $(NH_4)_2SO_4$, 136 g/l KH_2PO_4, and 142 g/l Na_2PO_4 in ddH_2O.

50x 5052 contains 250 g/l glycerol, 25 g/l D-glucose monohydrate, 100 g/l lactose monohydrate in ddH_2O

Antibiotic

An antibiotic (kanamycin or carbenicillin) was dissolved in ddH_2O as 1000x stock solution (100 mg/ml). The solution was sterilely filtered using a sterile syringe through a 0.2 μm filter under a sterile bench. The solution was kept frozen at -20 °C prior to use.

2.2.2.2 Expression of recombinant genes in *E. coli*

All enzymes in this study except the [Fe-S]-dehydratases were expressed using AI medium. A single colony of *E. coli* BL21 (DE3) bearing plasmids with a gene of interest was grown over night in 100 ml baffled flask containing 10 ml LB medium supplemented with 100 μg/ml kanamycin at 37 °C, 150 rpm. In the following day, the 10 ml culture was transferred to a 2 L baffled flask containing 500 ml AI medium supplemented with 100 μg/ml kanamycin. The culture was shaken at 120 rpm. Different growth temperatures were used according to the respective enzymes being expressed. Detailed temperature profiles were presented in each work in Chapter 3. At the end of the protein expression, the cells were pelleted by centrifugation at 4000 xg for 30 min. The cell pellets were stored at -80 °C prior to enzyme purification.

For expressing [Fe-S}-dehydratases, a TB media was used. A single colony of *E. coli* BL21 (DE3) bearing pET28Nhis with a [Fe-S}-dehydratase gene was grown over night in 100 ml baffled flask containing 10 ml LB medium supplemented with 100 μg/ml kanamycin at 37 °C, 150 rpm. In the following day, the 10 ml culture was transferred to a 2 L baffled flask containing 500 ml TB medium supplemented with 100 μg/ml kanamycin. The culture was grown at 37 °C, 120 rpm until OD reached 0.8 to 1. Isopropyl β- d-1-thiogalactopyranoside (IPTG) 0.5 mM (end concentration) was used to induce the protein expression. The temperature for the protein expression was

shifted to 18 °C. The expression was performed for at least 16 h. At the end of the protein expression, the cells were pelleted by centrifugation at 4000 xg for 30 min. The cell pellets were stored at -80 °C prior to enzyme purification.

2.2.2.3 Preparation of buffers for Äkta purifier

Binding buffer: 50 mM KPi, 20 mM imidazole, 0.5 M NaCl, 10 w/v% glycerol in ddH_2O, adjusted with HCl to pH 8. The solution was filtered through 0.45 µm filter paper prior to use.

Elution buffer: 50 mM KPi, 500 mM imidazole, 0.5 M NaCl, 10 w/v% glycerol in ddH_2O, adjusted with HCl to pH 8. The solution was filtered through 0.45 µm filter paper prior to use.

Desalting buffer: KPi, HEPES, and TRIS buffers with different concentrations and pH were used throughout this study. Detailed buffers used were given in Chapter 3.

2.2.2.4 Cell disruption by sonication

All cell disruptions in this study were performed by sonication. Cell pellet (10 to 20 %w/v) was dissolved in the binding buffer. DNase 5 µg/ml and 1 mM $MgCl_2$ were added to the solution. The cell suspension in a 50 ml falcon tube was placed in ice and a sonotrode was inserted to the cell suspension. Sonication was performed for 20 min (0.5 s cycle and 80%). Upon sonication, the cell suspension was centrifuged at 20.000 xg for 30 min at room temperature. The supernatant was filtered through a 0.45 µm filter before application to an Äkta purifier.

2.2.2.5 His-tag proteins purification *via* IMAC

All enzymes used in this study carried a hexahistidine tag either at the N or C-terminus. Immobilized Metal Affinity Chromatography (IMAC) was used to purify all enzymes. Purification was performed with an Äkta purifier. The protein supernatant was loaded onto a 5 ml His-tag column (HisTrap HP) with 10 ml/min flow rate. The column was washed with the binding buffer until online UV detector at 280 nm showed absorption <50 mAU (approximately 25 ml of the binding buffer) using 7 ml/min flow rate. The principle of a His-tag purification is immobilized nickel ions in a His-tag column will interact with six histidines located at the terminal of the protein of interest; thus, the protein is retained in the column. Other *E. coli* proteins would be washed away from the column during the washing step. To elute enzyme of interest, the elution buffer was used with 7 ml/min flow rate. Higher concentration of imidazole will break the interaction between the nickel ions and the hexahistidine tag thus allowing a protein of interest to be eluted from the column. A single fraction was collected. The column was equilibrated again with binding buffer using 10 ml/min of flow rate. Protein purity was analyzed by means of SDS-PAGE.

2.2.2.6 Size exclusion chromatography (SEC)

SEC was used to remove imidazole and salt in the eluted protein. A 50 ml desalting column (HiPrep 26/10) was applied to the Äkta purifier. Buffer of interest (desalting buffer)—depending on the enzymes—were used to equilibrate the column. Approximately 100 ml of the desalting buffer (two column-volume) was needed. Maximum volume of sample that could be loaded to allow good separation is 15 ml. Upon sample loading, the desalting buffer was pumped into the column with 10 ml/min flow rate. A single fraction of protein in a desired desalting buffer was collected. Salts and imidazole were eluted from the column after the protein due to their lower molecular weights.

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

The gel of SDS PAGE consisted of two parts: a separation gel (lower gel) and a collection gel (upper gel). The composition of the gel is as follows (for 14 gel):

Chemicals	Collection gel 5% (ml)	Separation gel 12% (ml)
Bis acrylamide	6.64	32
Upper buffer	10	-
Lower buffer	-	20
10% APS	0.4	0.8
TEMED	0.04	0.08
Water	Until 25 ml	Until 80 ml

4x Lower buffer for separation gel consisted of 8 g/l SDS, 1.5 M TRIS. The pH was adjusted with HCl to 8.8. The solution was stored at 4 °C.

4x Upper buffer for collection gel consisted of 8 g/l SDS, 0.5 M TRIS. The pH was adjusted with HCl to 6.8. The solution was stored at 4 °C.

10x Running buffer consisted of 10 g/l SDS, 2.5 M TRIS, 144 g/l glycine. The pH was adjusted with HCl to 8.5. The solution was stored at room temperature.

5x SDS loading dye consisted of 12.5 g/l β-mercaptoethanol, 75 g/l SDS, 500 g/l glycerol, 0.25 g/l bromphenol blue. The pH was adjusted with HCl to 6.8

Coomassie staining solution was made by dissolving 80 mg of coomassie brilliant blue G250 in 1 l ddH_2O and stirred for 3 h prior to addition of 3 ml of HCl 37 v/v %.

Gels (14 x 1 ml) were prepared using a Mini-Protean®3 Multi-Casting Chamber (Bio-Rad Laboratories GmbH) according to a recipe above and the protocol by the manufacture. Prior to loading protein sample to the wells of a SDS-PAGE gel, the sample was mixed with 5x SDS loading dye and heated at 95 °C for 5 min. For a gel with 15 wells, 10 µg of protein (end concentration) could be loaded or a maximum of 20 µl to allow good protein visualization. Mini-PROTEAN® Tetra Vertical Electrophoresis Cell was used to perform the electrophoresis. The electrophoresis cell was filled with 1x running buffer (10x running buffer dissolved in ddH_2O). A constant current of

40 mA per gel for 1 h was set as the parameters. Upon completion, the SDS gel was transferred to a staining chamber. Coomassie staining solution was poured to the chamber and heated by Microwave for 5 min. After removal of the staining solution, water was added and destaining was performed in a nutating shaker for 1 h. A paper towel was added to assist the removal of the remaining dye.

2.2.3 For enzyme characterization

2.2.3.1 UV-based photometric assay

An α-keto acid group shows maximum absorbance in the range 310 to 320 nm. This characteristic was used to measure activity of decarboxylases (Kdca and PDC) as the consumption of substrates. Detailed procedures were presented in subchapter 3.1 (page 59 and 74) and 3.2 (page 84).

2.2.3.2 Coupled assay based on NADH

Kinetic characterization of PDC was performed using a coupled assay. Yeast ADH was used to reduce acetaldehyde, decarboxylation product of pyruvate by a PDC, in the presence of NADH. Detailed procedure was presented in subchapter 3.2 (page 84).

Kinetic characterization of L-KdpD and D-KdpD also used a coupled assay. *Pp*KGSADH was used to oxidize KGSA, a dehydration product of L-KDP or D-KDP, in the presence of NAD^+. Detailed procedure was presented in subchapter 3.4 (page 130).

2.2.3.3 HPLC for enzyme analysis

High Performance Liquid Chromatography (HPLC) was used to measure enzymatic activity of [Fe-S]-dehydratases (page 118). Another HPLC method was also used to measure substrates, intermediates, and products for D-glucose and glycerol in *in vitro* cascades. Detailed procedures were presented in subchapter 3.3 (page 119-122). HPLC was also used to analyse the conversion of L-KDP and D-KDP overtime by L-KdpD and D-KdpD. Detailed experimental procedures are presented in subchapter 3.4 (page 131).

2.2.3.4 Thermal shift assay

Melting temperature (T_m) of enzymes in this study was determined indirectly by monitoring the interaction of a dye (Sypro orange) that will fluoresce when binding with exposed hydrophobic regions of an enzyme upon enzyme unfolding. Detailed experimental procedures are presented in Chapter 3 (page 59, 85, 119, and 131).

3 Results

3.1 Structure-guided engineering of α-keto acid decarboxylase for the production of higher alcohols at elevated temperature

In this publication, a KdcA from *Lactococcus lactis* (*Ll*KdcA) was engineered toward improved stability. KdcA catalyzes decarboxylation of α-ketoisovalerate (KIV) to isobutyraldehyde, the penultimate reaction in a synthetic isobutanol production. A step-wise semi-rational approach was used after previous attempts to increase stability of an isoenzyme (*Ll*KivD) using directed evolution coupled with computational predictions had failed to increase kinetic stability beyond 10 °C. *Ll*KdcA and *Ll*KivD wild types (WT) showed a T_{50}^{1h} (temperature at which 50% of protein is deactivated after 1 h) of 55°C. Thus, a 15 °C improvement would be needed to reach the project (see subchapter 1.4.1). Prior to libraries generation, a simple high-throughput screening (HTS) was developed. The HTS relied on the decrease of absorbance of KIV at 310 nm upon decarboxylation reaction.

The stepwise approach combined strategies such as stabilization of catalytic center, sequence homology, surface engineering, and optimization of dimer interaction were used. A variant with seven point substitutions (7M.A) was created using this approach. Characterization of 7M.A determined that this variant was inhibited severely by isobutanol. Further analysis of this variant successfully identified that a change (A290M) contributed to the isobutanol inhibition. Thus, an SSM was performed to find a compromise between stability and isobutanol inhibition. Two variants M290V and M290C were recovered from library screening. M290C variant, however, demonstrated higher initial activity. The improved variant (7M.D) was further characterized and compared to WT *Ll*KdcA. 7M.D demonstrated an increased of T_{50}^{1h} by 14.8° C and T_m by 13.5 °C. The amino acid changes introduced did not affect substrate specificity of 7M.D. Half-life of 7M.D at 70 °C reached 2 h, or >400-fold improvement in comparison to the WT. At 50°C and 4% isobutanol where most *in vivo* production of isobutanol would be halted, a cell-free approach has been developed and shown to be functional. At this condition, 7M.D registered a half-life of 14.3 h corresponding to >600-fold improvement as compared with the WT.

The author and the co-author Prof. Sieber conceived the study. The author designed the HTS in this study and applied a stepwise engineering approach. The author also analyzed the data and wrote the manuscript. The co-author Dr. Carsten contributed to content and language of this publication. The co-author Prof. Sieber critically reviewed the submitted manuscript.

Structure-guided engineering of α-keto acid decarboxylase for the production of higher alcohols at elevated temperature

Samuel Sutiono, Dr. Jörg Carsten, Prof. Volker Sieber

ChemSusChem

2018

DOI: 10.1002/cssc.201800944

ChemPubSoc Europe

DOI: 10.1002/cssc.201800944

CHEMSUSCHEM Full Papers

Check for updates

Structure-Guided Engineering of α-Keto Acid Decarboxylase for the Production of Higher Alcohols at Elevated Temperature

Samuel Sutiono,[a] Jörg Carsten,[b] and Volker Sieber*[a, b, c, d]

Branched-chain keto acid decarboxylases (KDCs) are a class of enzymes that catalyze the decarboxylation of α-keto acids. They are key enzymes for production of higher alcohols in vivo and in vitro. However, the two most active KDCs (KivD and KdcA) have only moderate thermostability (< 55 °C), which hinders the production of alcohols at high temperatures. Herein, structure-guided engineering toward improved thermostability of KdcA is outlined. Strategies such as stabilization of the catalytic center, surface engineering, and optimization of dimer interactions were applied. With seven amino acid substitutions, variant 7M.D showed an increase of the temperature at which 50% of activity remains after one-hour incubation T_{50}^{1h} by 14.8 °C without compromising its substrate specificity. 7M.D exhibited greater than 400-fold improvement of half-life at 70 °C and greater than 600-fold increase in process stability in the presence of 4% isobutanol at 50 °C. 7M.D is more promising for the production of higher alcohols in thermophiles (> 65 °C) and in cell-free applications.

Introduction

There has been a concerted effort to reduce our dependence on fossil fuels.[1] Alcohol production has been proposed as a suitable and sustainable replacement for the traditional gasoline derived from fossil fuels.[2] Higher branched-chain alcohols such as 2-propanol, isobutanol, and 1-butanol have properties similar to those of gasoline in terms of energy density and volatility, in contrast to their shorter and linear counterpart ethanol.[2] Isobutanol was produced in vivo for the first time in *Escherichia coli*.[3] Since then, similar studies have been reported.[2,4] Although a high theoretical yield could be achieved, a low concentration of isobutanol (1.5–2%) is already toxic to *E. coli*.[3] In an effort to avoid this in vivo toxicity of isobutanol, in vitro alcohol production has been reported to withstand isobutanol concentrations up to 4%.[5] Biofuel production at elevated temperatures in vivo or in vitro is highly desirable. Higher temperatures avoid the risk of contamination and may allow biomass pretreatment and alcohol fermentation to be performed simultaneously in one pot while speeding up the production process and lowering costs.[6] Isobutanol production in the thermophilic bacterium *Geobacillus thermoglucosidasius* has been reported recently; however, the yield was much lower than in the process performed with *E. coli*.[7] One explanation suggests that one of the key enzymes, namely, the branched-chain keto acid decarboxylase from *Lactoccocus lactis* (KivD), which catalyzes the decarboxylation of 3-methyl-2-oxobutanoic acid (ketoisovalerate) to 2-methylpropanal (isobutanal), is not thermostable.[7] Its isoenzyme KdcA (80% sequence identity) used in the cell-free production of isobutanol by Guterl et al. has a similar moderate thermostability, which impedes the enzymatic cascade at temperatures higher than 50 °C.[5] Recent work on KivD by directed evolution (epPCR) and computational design was able to increase the temperature at which 50% of activity remains after one-hour incubation T_{50}^{1h} from 56 to 60 °C (LLM4 mutant).[6]

Sulfolobus acidocaldarius (*Saci*) is a thermoacidophilic archaea that is able to grow and prosper at 75 °C and at pH 2–3.[8] High temperature and low pH are ideal conditions for biomass pretreatment, and hence *Saci* is a promising chassis for isobutanol production from renewable resources by one-pot biomass pretreatment and alcohol fermentation.[9] High temperature may also facilitate isobutanol removal, which decreases further the toxicity problems encountered in vivo.[10] One obvious hurdle of isobutanol production in *Saci* is that no branched-chain keto acid decarboxylases from *Sulfolobus* or other thermophilic organisms have been identified, and the most active mesophilic KDCs (KivD, KdcA, and LLM4) are not

[a] S. Sutiono, Prof. Dr. V. Sieber
Chair of Chemistry of Biogenic Resources
TUM Campus Straubing for Biotechnology and Sustainability
Technical University of Munich
Schulgasse 16, 94315 Straubing (Germany)
E-mail: sieber@tum.de
Homepage: http://www.cs.tum.de/

[b] Dr. J. Carsten, Prof. Dr. V. Sieber
Catalytic Research Center
Technical University of Munich
Ernst-Otto-Fischer-Straße 1, 85748 Garching (Germany)

[c] Prof. Dr. V. Sieber
Straubing Branch BioCat
Fraunhofer IGB
Schulgasse 11a, 94315 Straubing (Germany)

[d] Prof. Dr. V. Sieber
School of Chemistry and Molecular Biosciences
The University of Queensland
68 Copper Road, St. Lucia 4072 (Australia)

Supporting Information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/cssc.201800944.

ChemPubSoc Europe

CHEMSUSCHEM Full Papers

yet suitable for heterologous expression due to their limited thermostability. Therefore, engineering a member of this enzyme class toward higher thermostability is critical to enable isobutanol production in *Saci*.

Several strategies for stabilizing enzymes toward higher temperatures have been described previously.[11–13] Directed evolution, a method that relies on the generation of a random gene library and screening a large number of such libraries to find variants with increased thermostability, is promising, and many success stories have been reported.[12,14] Another approach, rational design, relies more on the knowledge of the crystal structure of the enzyme and results in a smaller library to screen. Several factors contributing to protein thermostability and possible strategies to rationally improve it have been discussed recently.[13] However, until today, all of the strategies described are not yet able to encompass a single universal method to improve protein thermostability.[11] Therefore, a combination of these strategies could be the best approach to evolve a mesophilic branched-chain keto acid decarboxylase to be suitable for isobutanol production in *Saci*.[15,16] In this work, KdcA was engineered semirationally, on the basis of knowledge of its crystal structure. The strategy and the suitability of the improved variants for in vivo and in vitro isobutanol production are discussed.

Results and Discussion

Enzyme thermostability engineering

A primary consideration of protein engineering is the selection of a suitable starting template. To the best of our knowledge, no enzymes from thermophilic organisms are able to catalyze the decarboxylation of ketoisovalerate with similar activity to that of KivD or KdcA.[7] Both enzymes have been described previously in the literature.[17,18] In addition, the crystal structure of KdcA has been solved with good resolution (1.8 Å) and is available from the Protein Data Bank (PDB ID: 2vbf and 2vbg).[19] Thus, KdcA was selected as the starting template in this study. The first approach to increase thermostability was by considering the B factor.[20] B-FITTER was used to extract the B-factor values from the KdcA PDB file (PDB ID: 2vbg).[21] Four amino acids—181E and 188N from chain A, and 342K and 345E from chain B—showed the highest B-factor values and were chosen for site saturation mutagenesis (SSM) with the degenerate codon NNK. Libraries were constructed by QuikChange. High-throughput screening was performed by measuring the remaining activity of the cell lysate after a ten-minute heat challenge. All activity measurements were conducted at 30 °C. None of the resulting variants exhibited a higher remaining activity after heat treatment.

The second approach was to improve the stability of the catalytic center.[22] It was hypothesized that the catalytic center is in a portion of the enzyme that is inherently flexible enough to allow conformational changes preceding and throughout catalysis; therefore, it is prone to destabilization during heat treatment.[23] Active-center stabilization can sometimes lead to decreased catalytic activity.[22] However, the gain in conformational stability could facilitate higher activity at higher temperature as a result of a shift in the optimum temperature.[22] Six amino acids that shape the catalytic center of KdcA were selected (377Q, 381F, 382F, 461V, 465I, and 542F). The codons for these residues were saturated by using the degenerate codon NDT. Variant V461I, renamed 1M, showed an increase in its post-heat-treatment activity during cell-lysate screening. 1M was purified by using a HisTrap column and further characterized. The T_{50}^{1h} value of 1M was determined to be 59 °C, a 3.3° increase compared to wild-type (WT) KdcA, with a T_{50}^{1h} value of 55.7 °C (Figure 1). The temperature at which 50% of the protein is in the unfolded state T_m was determined by Thermofluor analysis, as previously described.[24] The T_m value for 1M

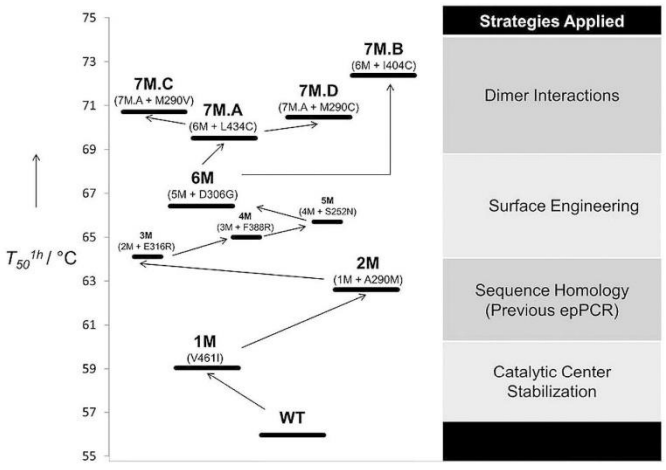


Figure 1. Effect of single amino acid exchanges on enhanced T_{50}^{1h} of KdcA. Approaches considered are presented on the right. T_{50}^{1h} is the incubation temperature resulting in a decrease to 50% residual activity after 1 h incubation of the enzyme (for detailed graph, see Figure S1).

ChemPubSoc Europe

CHEMSUSCHEM Full Papers

Table 1. Melting temperatures and kinetic characteristics of the purified WT and variant enzymes.[a]

Abbreviation	Mutation	T_m[b] [°C]	k_{cat} [s^{-1}][b]	K_m[b] [mmol L^{-1}]
WT	–	61 ± 0	145 ± 1.1	5.1 ± 0.9
1M	V461I	64 ± 0	120 ± 1.2	5.6 ± 0.2
2M	1M_A290M	66.5 ± 0	118 ± 1.4	9.3 ± 0.4
6M	2M_Q252N, D306G, E316R, F388Y	70.5 ± 0	149 ± 1.8	14.6 ± 0.7
7M.A	6M_L434C	74.5 ± 0	95 ± 1.0	8.6 ± 0.4
7M.B	6M_I404C	75.5 ± 0	5.2 ± 0.1	2.6 ± 0.2
7M.B (heat activated)[c]	6M_I404C	n.d.[e]	8.8 ± 0.1	2.7 ± 0.2
7M.C	7M.A_M290V	74.5 ± 0	43 ± 0.7	6.6 ± 0.4
7M.D	7M.A_M290C	74.5 ± 0	69 ± 0.7	5.2 ± 0.3
7M.D (heat activated)[c]	7M.A_M290C	n.d.[d]	101 ± 1.7	4.2 ± 0.3

[a] For detailed Thermofluor and Michaelis–Menten plots, see Figures S2 and S3. [b] The results were obtained as three independent replicates. [c] Heat activation was done by incubating the enzymes at 60 °C for 1 h. [e] Not determined.

of 64 °C corresponds to a 3 °C increase relative to the WT enzyme (Table 1). The value of T_{50}^{1h} is related to that of T_m, but the two values are not necessarily identical, because the former is time-dependent and describes the kinetic stability, and the latter is a thermodynamic descriptor of thermostability.[13,25] This substitution likely increases the hydrophobic packing in the catalytic cavity and leads to an increase in thermostability (Figure 2 A). This substitution was reported earlier to decrease specific activity by more than 60% relative to WT KdcA.[26] However, our results showed that 1M exhibits a smaller decrease (< 20%) in activity (Table 1). This discrepancy could be caused by the different position of the hexahistidine tag. In this study, the N-terminal hexahistidine tag of KdcA was used, whereas other studies utilized the C-terminal hexahistidine tag.[26] As earlier reported, the position of the hexahistidine tag influences the substrate scope of KdcA.[27]

During this study, Liao et al. published similar findings on KivD, an isoenzyme of KdcA, and reported four positions that are involved in improved thermostability.[6] It was logical to assume that these amino acid residues would have a similar effect on KdcA, since both have similar protein sequence identities (80%). On the basis of a sequence alignment of the two enzymes, Q34, L130, A290, and T386 were found to be homologous. The corresponding codons were saturated by using the degenerate codon NNK, and the four libraries were screened.[6] After a second screening, several variants with increased thermostability were identified: Q34H, A290M, A290L, A290V, and T386P. No improved variant was detected in the L130 library. A single exchange at position 130 has been reported to increase activity rather than thermostability.[6] All variants, except A290M and A290L, had the same amino acid substitutions as reported by Liao et al.[6] Instead of purifying each variant, stepwise SSM was performed. 1M was chosen as the template for SSM and the codon for the residue in position 290 was saturated. Position 290 was chosen because, in the cell lysate screening, all improved variants from that library showed a higher residual activity than Q34H and T386P. SSM of 1M revealed several variants with improved thermostability, namely, 1M_A290M, 1M_A290C, and 1M_A290V, of which 1M_A290M and 1M_A290V appeared twice during the screening. The T_{50}^{1h} values of these three variants were 62.6 °C (Figure 1 and Figure S4 in the Supporting Information). However, the initial activity of the double variant 1M_A290M was the highest compared with 1M_A290V and 1M_A290C (Figure S4 in the Supporting Information). Therefore, 1M_A290M was selected as the template for the next rational design and renamed 2M. An exchange of residue 290 suggested an increase in the hydrophobic interactions with 541L and 545F (Figure 2).[6] In the next SSM, the gene of 2M was saturated at the codon encoding residue 34. Variant 2M_Q34H was detected, and after purification it showed an increase of T_m by 0.5 °C without significant improvement of T_{50}^{1h} (data not shown). This triple variant was then saturated further at the codon encoding residue 386; however, no improved variants were detected. Due to an insignificant increase in thermostability of the triple variant, 2M was chosen for the next stage of semirational design.

The next strategy to increase thermostability was an investigation into surface engineering. Amino acid residues on a protein surface have been previously shown to play an important role in enzyme thermostability.[28,29] Swiss PDB viewer (https://spdbv.vital-it.ch/) was accessed to visualize and list amino acids that are exposed to the solvent. The default value of 30% solvent exposure was selected. Additionally, amino acids located in the surface loop regions were given priority over the amino acids located in α-helices or β-sheets. It was reasoned that surface loops are regarded as weak points for affecting the thermostability of an enzyme because their flexibility could allow water to penetrate into the enzyme core and thus cause the enzyme to unfold.[30] Reetz et al. also reported that most of the mutations they obtained from the B-factor approach are located on the surface loop of their model enzyme (lipase).[20] Consequently, ten amino acids located on the surface loop were chosen for SSM (see Table S10 in the Supporting Information). 2M was used as the template. Two variants, D306G and E316R, were found to have increased thermostability in the cell-lysate screening. These variants were purified to homogeneity. 2M_D306G and 2M_E316R showed T_{50}^{1h} values of 63.3 and 64.2 °C, respectively, and T_m values of 68 and 68.5 °C, respectively (Figure S5). We propose that an exchange of residue 306 may relax the loop region and thus increase hydrophobic interac-

ChemPubSoc Europe

CHEMSUSCHEM Full Papers

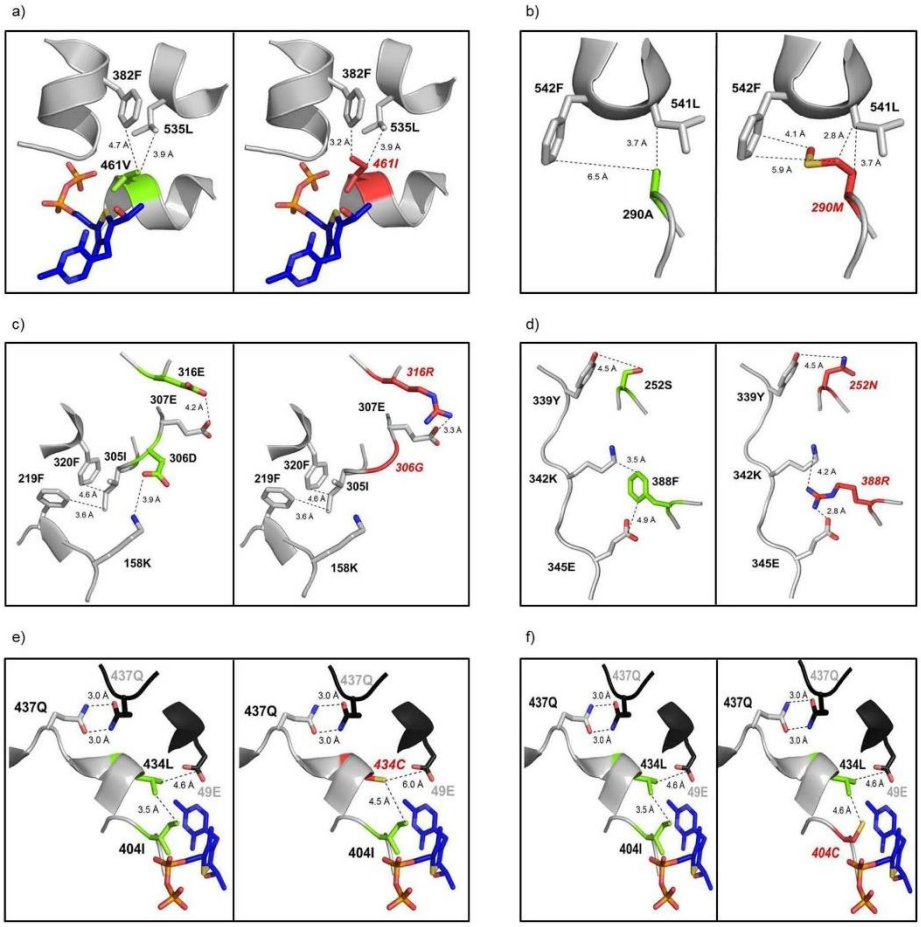


Figure 2. Illustrative views of the amino acid exchanges leading to increased thermostability. Native amino acids (green) on the left and the corresponding exchanges (red) on the right are shown. Surrounding amino acids that may interact with the respective native or substituted amino acids are colored gray. Distances of possible interactions were calculated by using Pymol. A) Exchange in 1M of position 461 from valine to isoleucine near the catalytic center. The cofactor is shown as a blue stick. B) Exchange at position 290 (2M) of alanine to methionine that is located near the surface of the enzyme. C) Exchanges on the surface of the enzyme at position 306 from aspartic acid to glycine and at position 316 from glutamic acid to arginine (6M). D) Substitutions at position 252 from serine to asparagine and at position 388 from phenylalanine to arginine may bring the second region of the extended loop closer to the main chain (globular protein core, 6M). E) Substitution at residue 434 (7M.A) from leucine to cysteine also increased thermostability of the enzyme. The second chain of the homodimer KdcA is colored black, as are the amino acids of the second monomer. F) Substitution at position 404 to cysteine increased thermostability but led to a significant decrease in the initial activity, likely due to perturbation of the interaction with the thiazole ring of thiamine diphosphate. The crystal structure was obtained from PDB (2vbg) and was visualized with Pymol.

tions of 305I with 219F and 320F (Figure 2C) to give slightly increased thermostability. The substitution of a negatively charged glutamic acid by a positively charged arginine residue at position 316 may increase electrostatic interactions with 307E and thus increase thermostability (Figure 2C). The relation between surface salt bridges and thermostability has been described previously.[31]

Several reports on the influence of loop flexibility on thermostability have been published.[11,32,33] Wintrode and Arnold argue that a loop could serve as the initiation site of protein unfolding.[12] After several attempts at exchanging or shortening amino acids in extended loop regions of KdcA with no significant improvement in thermostability (for detailed experiments, see Supporting Information S7), we shifted our ap-

ChemPubSoc Europe

CHEMSUSCHEM Full Papers

proach to the amino acids that are located in the main region (globular protein core) of KdcA.[34,35] In the second loop region (330L to 355S), there were several candidates of amino acids that showed potential for increasing thermostability, such as S252 and F388. Both of these are located on the loop of the main chain (Figure 2D). The gene encoding 2M was saturated at these positions independently and two variants, 2M_S252N and 2M_F388R, were identified from the cell-lysate screening. Variants 2M_S252N and 2M_F388R exhibited increases of T_{50}^{1h} by 0.9 and 1.4 °C, respectively. Both variants also showed an increase in T_m in comparison with 2M, by 1.5 and 2 °C, respectively (Figure S5). Substitutions at position 252 and 388 were hypothesized to increase the thermostability by providing new electrostatic interactions to 339Y in the case of S252N. An arginine residue substituted at position 388 could form a new electrostatic interaction with 345E and thus anchor the second extended loop region to the main chain (Figure 2D).

Having identified four beneficial single mutations located on the surface and loops of 2M (D306G, E316R, S252N, and F388R), we explored the possibility of the cumulative effect of these point substitutions. Although it is generally believed that the stabilizing effect of point substitutions is likely to be additive, we opted to saturate each position sequentially.[36,37] It is argued that the additive effects of beneficial substitutions also depends on the distance between the exchanges.[37,38] 2M_E316R (3M) was used as the first template because it showed the highest increase in T_{50}^{1h} relative to the other mutations (Figure S5 in the Supporting Information). The gene of 3M was saturated at the codon encoding residue 388 to obtain a variant with four exchanges. Variant 3M_F388R was selected and renamed 4M. The gene of 4M was saturated at the codon encoding residue 252 to obtain a variant with five exchanges. Variant 4M_S252N was selected and renamed 5M. The gene of 5M was saturated further at the codon encoding residue 306 to result in a variant bearing six substitutions. Variant 5M_D306G was selected and renamed 6M. In each SSM round a slightly improved T_{50}^{1h} was observed (Figure 1). The stepwise SSM of each position found reliably the previously determined amino acid exchange to be the best performing amino acid at the respective position. Compared with 2M, 6M improved T_{50}^{1h} and T_m by 4 and 3.5 °C, respectively (Figure 1 and Table 1).

The crystal structure of KdcA (PDB ID: 2vbg) suggests the enzyme is a homodimer.[19] In the next approach this property was explored as a target to improve thermostability. Optimization of the amino acid interactions at the oligomer interfaces has led to increased thermostability, as described in the case of 3-epimerase from *P. cichorii* (*Pc*DTE).[39] In this reported study, most of the amino acids contributing to an increase in thermostability were located in the loop of the interfaces.[39] For this approach, PDBePISA was used to list the amino acids that are located at the dimer interface and therefore have interactions between the monomers of KdcA (PDB ID: 2vbg).[40] Eight positions were chosen for SSM. Except position 434, all eight residues are located in the loops. Residue 434 is situated in an α-helix and is in close proximity to the catalytic center. This position was targeted for saturation given the thermostability increase identified in the 1M variant (V461I) because of substitution of the residue located in an α-helix that is also in close proximity to the catalytic center. 6M was used as the template, and from the cell-lysate screenings, four beneficial substitutions were identified. One is a change at position 404, from isoleucine to cysteine, and the other three are changes at position 434 from leucine to threonine, cysteine, or glutamine. Variant 6M_L434C showed the highest T_{50}^{1h} and T_m values in comparison with 6M_L434T and 6M_L434Q (for detailed characterization, see Figure S6). Variant 6M_L434C was then renamed 7M.A, with increases of 3.4 and 4 °C in T_{50}^{1h} and T_m, respectively, relative to 6 M. Variant 6M_I404C, renamed 7M.B, showed the largest improvement with increases in T_{50}^{1h} and T_m of 6.4 and 5 °C, respectively, relative to 6M. Unfortunately, variant 7 M.B showed a substantial decrease (>90%) in initial activity at 30 °C (Table 1). We proposed that a substitution at position 434 increased thermostability by decreasing its hydrophobic interaction to 404I, thus relaxing the adjacent loops, and ultimately increasing the electrostatic interactions of 437Q between the two monomers (Figure 2E and Figure S6 for a more detailed explanation).

Variants 7M.A and 7M.B showed increases of T_{50}^{1h} by 13.8 and 16.8 °C, respectively, relative to WT KdcA (Figure 1), and T_m was improved by 13.5 and 14.5 °C for 7M.A and 7M.B, respectively (Table 1). SSM at the codon encoding residue 404 of the 7M.A gene with the degenerate NNS codon did not yield any variants with improved thermostability. Because 7M.B lost considerable initial activity even after heat activation (Table 1), 7M.A was selected for further characterization.

Relation between thermostability and inhibition by isobutanol: Further engineering

To determine whether the mutants produced are suitable for isobutanol production, the IC_{50} values for isobutanol were determined. The IC_{50} is defined as the concentration of isobutanol that inhibits 50% of the initial activity of the enzyme. The IC_{50} for WT KdcA at 50 °C was 3.5 vol%, similar to a previous report (Figure 3A).[5] Unexpectedly, the IC_{50} for 7M.A showed a dramatic decrease in comparison to the WT enzyme. More than 50% of the activity of 7M.A was inhibited in the presence of just 1 vol% isobutanol (Figure 3A).

Examining the IC_{50} plots of the WT enzyme and 7M.A indicated different inhibition mechanisms of isobutanol. The activity of WT KdcA decreased almost linearly with increasing concentration of isobutanol, and this suggests that isobutanol may inhibit the WT enzyme competitively. The activity of 7M.A decreased significantly (>70%) in the presence of 2 vol% isobutanol. The decrease continued gradually until the concentration of isobutanol reached 9 vol%, at which a two-phase solution began to form. This result prompted the hypothesis that the inhibition was likely noncompetitive for 7M.A. To test this hypothesis, the IC_{50} values for each additional mutation were analyzed. The IC_{50} for 1M was similar to that of WT KdcA (Figure 3A). The additional exchange in position 290, from alanine to methionine (2M), gave a lower IC_{50} (Figure 3A) compared with the 1M and the WT enzyme. Four substitutions on the surface of 2M (6M) and the seventh exchange (7M.A) did not

ChemPubSoc Europe
CHEMSUSCHEM Full Papers

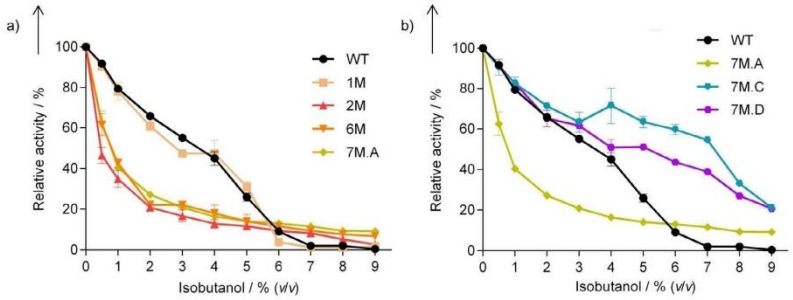


Figure 3. Initial activity of WT and mutant KdcAs was measured in the presence of different isobutanol concentrations to determine IC_{50}. A) The effect of each additional mutation to the decrease in IC_{50} of 7M.A. B) Further engineering of 7M.A resulted in an increase of IC_{50}. 100% refers to the activity of enzymes variants in the absence of isobutanol measured for the decarboxylation of KIV (50 mM) at 50 °C.

affect the IC_{50} significantly (Figure 3A). These results suggest that substitution at position 290, from an alanine (small amino acid) to a methionine residue (large nonpolar amino acid), increases the interaction between the enzyme and isobutanol, which leads to a significant decrease in the IC_{50} value. We were aware of the possible reversion of the substitution to alanine, which might increase the IC_{50} value, but this reversion could also come at the expense of thermostability (3.6 °C increase in T_{50}^{1h} after exchanging alanine residue at position 290 of 1M to methionine). On the basis of this finding, the gene of 7M.A was saturated at the codon encoding residue 290 to explore whether another amino acid residue would provide a larger increase in IC_{50} while retaining the thermostability of 7M.A.

In this screening, we added a third plate, which measured the initial activity in the presence of 4 vol% isobutanol. After library screening, we identified two variants that showed the desired characteristics. The first variant was 7M.A_M290V (renamed 7M.C), and the second was 7M.A_M290C (renamed 7M.D). The IC_{50} values of 7M.C and 7M.D at 50 °C were 7 and 5 vol%, respectively, which represent more than a fivefold increase in comparison with 7M.A (Figure 3B). 7M.C and 7M.D exhibited slight increases in T_{50}^{1h} relative to 7M.A (Figure 1), but there was no difference in the T_m values determined by Thermofluor (Table 1).

The exchange in position 290 to a valine or cysteine residue was shown earlier to also contribute to increases in thermostability (Figure S4 in the Supporting Information). The T_{50}^{1h} values of 70.8 °C for 7M.C and 70.5 °C for 7M.D correspond to improvements of 15.1 and 14.8 °C, respectively, relative to the WT enzyme (Figure 1). 7M.D was selected for additional studies because it showed higher activity than 7M.C (Table 1) and the difference in T_{50}^{1h} was less than 0.5 °C (Figure 1).

During preparation of this manuscript, Maier et al. reported an improvement of T_{50}^{10min} by more than 6 °C for WT KivD after exchanging serine at residue 385 to methionine.[41] However, several efforts to transfer this beneficial substitution to 7M.D did not yield significant improvements (see Supporting Information S7 for detailed experiments)

Effect of mutations on substrate specificity

KdcA has been reported to show promiscuous activity toward different α-keto acids.[18] To check whether the increased thermostability of 7M.D led to a change in substrate range, several α-keto acids (their respective salts) were tested. 2-Oxopropanoic acid (pyruvate), 2-oxobutanoic acid (ketobutyrate), and 2-oxopentanoic acid (ketovalerate) are the α-keto acids (salts) for the production of ethanol, 1-propanol, and 1-butanol, respectively.[3] In general, the substrate specificity of 7M.D was similar to that of WT KdcA (Figure 4). The initial activity of the WT

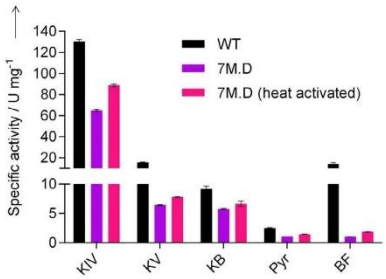


Figure 4. Specific activity of WT KdcA and 7M.D toward different keto acids at 30 °C. Concentrations of all keto acids were 50 mM. KIV is ketoisovalerate, KV is ketovalerate, KB is ketobutyrate, Pyr is pyruvate, and BF is benzoylformate. Heat activation of 7M.D was performed by incubating at 60 °C for 1 h.

enzyme at 30 °C was higher for all α-keto acids tested (see Supporting Information S8 for further explanations). Heat activation of 7M.D performed at 60 °C for 1 h led to an almost 50% increase in initial activity but did not change the substrate specificity. At 60 °C, the WT enzyme completely lost activity after 1 h of incubation (Figure S9). 7M.D exhibited decreased activity toward 2-oxo-2-phenylacetic acid (benzoylformate). This is likely due to the substitution at residue 461 from valine to isoleucine. The mutation at 461 to a larger amino acid residue would decrease the size of the binding pocket and thus lead to a severe decrease in activity toward bulky

ChemPubSoc Europe
CHEMSUSCHEM Full Papers

substrates such as benzoylformate.[26] This result showed that 7M.D would be a suitable candidate for in vivo or in vitro production of other higher alcohols at elevated temperatures (>65 °C).

Applicability of 7M.D for in vivo and in vitro isobutanol production

To introduce KdcA into *Saci*, the enzyme must be active at 70 °C, which is the ideal temperature for *Saci* to grow.[8] At 70 °C WT KdcA lost 50% of its activity after incubation for 15 s, whereas 7M.D showed an increase in activity of more than 50% in the first 2 min, but the activity quickly dropped after 5 min (Figure 5). Subsequently, the activity of 7M.D decreased slowly and reached 52% of the initial activity after 1 h of incubation at 70 °C, which was maintained for the next hour to retain 50% activity after 2 h of incubation

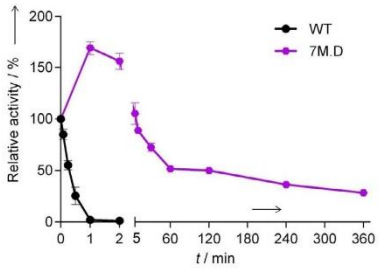


Figure 5. Stability of WT KdcA and 7M.D at 70 °C. Remaining activity was measured at 30 °C.

7M.D preserved 20% of the initial activity even after 6 h of incubation at 70 °C. Compared with WT KdcA, 7M.D had an increased half-life at 70 °C from 17.3 s to 2 h (>400-fold). Mutant 7M.D is the most thermostable α-keto acid decarboxylase reported to date.[6,27,41] 7M.D was expected to be suitable for isobutanol production by integration into *Saci* or any other thermophiles that grow at temperatures higher than 65 °C.

The in vitro production of isobutanol is also highly desirable.[5] The IC_{50} plot of 7M.D showed that the enzyme variant was still active in the presence of 9% isobutanol, at which point the two-phase system of water and isobutanol started to form (Figure 3). The two-phase behavior of water/isobutanol will facilitate simple separation processes leading to lower production costs of isobutanol.

To check the applicability of 7M.D for in vitro isobutanol production, the concentration of isobutanol at which the enzyme activity decreased by 50% after 1 h incubation at 50 °C ($Iso_{50}^{50\,°C,\,1\,h}$) was determined. The amount of active WT KdcA decreased by more than 50% after incubation with 2.5% isobutanol for 1 h at 50 °C, whereas 7M.D retained more than 50% of its initial active enzyme after incubation with 6% isobutanol (Figure 6A). A similar stability profile was shown by 7M.A (Figure 6A), although IC_{50} at 50 °C was significantly lower (Figure 3B).

This finding suggested 7M.A was inhibited, rather than deactivated or denatured, in the presence of high isobutanol concentrations (>2%). 7M.C showed the highest stability in isobutanol with an $Iso_{50}^{50\,°C,\,1\,h}$ value of 7.3% (Figure 6A). However, the initial activity of 7M.D was 150% higher than that of 7M.C (Table 1), and this makes the 7M.D variant slightly more favorable for in vitro isobutanol production. The process stability of WT KdcA and 7M.D in the presence of 4 vol% isobutanol at 50 °C was investigated in the next step. Process stability is an important parameter for industrial application because it dictates the production cost. An isobutanol concentration of 4 vol% was deemed an "ideal concentration" to study enzyme stability for in vitro isobutanol production. At this concentration no in vivo production could take place due to the toxicity of isobutanol to the cells.[42,43] WT KdcA lost more than 50% of its activity in the first 2 min, whereas 7M.D was still highly active (>60% initial activity) even after 6 h of incubation in 4 vol% isobutanol (Figure 6B). WT KdcA appeared to follow first-order kinetics of deactivation, whereas 7M.D seemed to follow second-order kinetics. Deactivation constants *k* for WT KdcA and 7M.D were $-32.432\ h^{-1}$ and $-0.0007\ \%A^{-1}\,h^{-1}$ (Figure 6B). Based on the deactivation constant, the half-life of WT KdcA in the presence of 4 vol% isobutanol at 50 °C was

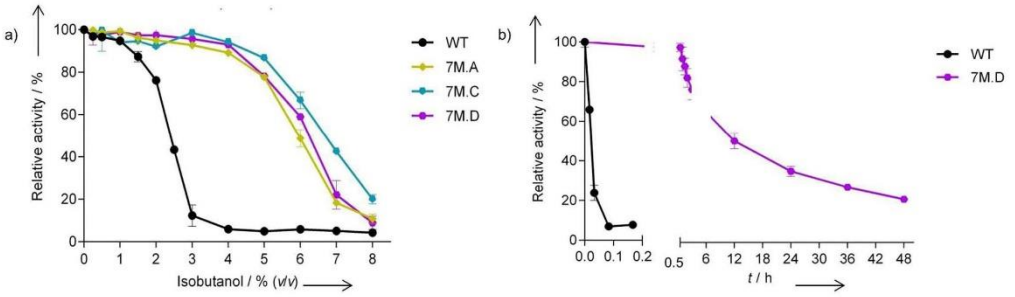


Figure 6. Stability of the WT KdcA and variants in the presence of isobutanol. A) The effect of different isobutanol concentrations on the stability of WT KdcA, 7M.A, 7M.C, and 7M.D after 1 h incubation at 50 °C. B) Process stability of WT KdcA and 7M.D in 4 vol% isobutanol at 50 °C. Deactivation of WT KdcA followed first-order kinetics, whereas that of 7M.D followed second-order kinetics with calculated deactivation constants *k* for WT KdcA and 7M.D of $-32.432\ h^{-1}$ and $-0.007\ \%A^{-1}\,h^{-1}$ respectively, assuming activity *A* at t_0 was 100%.

ChemPubSoc Europe

CHEMSUSCHEM Full Papers

1.3 min, and that of 7M.D was 14.3 h. These results indicated that the increase in thermostability of 7M.D provided superior process stability at 4 vol% isobutanol (>600-fold improvement).[44]

Conclusions

We engineered a mesophilic keto acid decarboxylase (KdcA) from *Lactococcus lactis* by structure-guided engineering. In total we screened fewer than 3000 colonies to achieve a more than 14 °C improvement in $T_{50}^{1\,h}$ (7M.D). The melting temperature ™ of 7M.D also increased by 13.5 °C. 7M.D showed improved stability at 70 °C with a half-life of 2 h, whereas wild-type (WT) KdcA lost more than 95% activity after 1 min incubation at 70 °C (>400-fold improvement). The engineering of 7M.D toward higher stability came at the expense of less than 30% decrease of k_{cat} (number of substrate molecule converted per second per enzyme monomer) at ambient temperature (30 °C). After heat activation, 7M.D still showed 100 U mg^{-1} at 30 °C; thus, 7M.D is the most active and thermostable KDC reported to date.[6,27,41] In addition, lost activity is completely regained when the enzyme is used at 14 °C higher process temperature. At 44 °C, 7M.D exhibited an activity of 133 U mg^{-1}, which is similar to that of the WT enzyme measured at 30 °C (data not shown). Increased thermostability did not influence substrate specificity, so that 7M.D will be suitable for the production of other higher alcohols, such as 1-propanol and 1-butanol. The integration of 7M.D into *Saci* to facilitate isobutanol production in vivo is currently under investigation. Although 7M.D was only intended for heterologous expression in *Saci*, in vitro studies revealed that the improved thermostability also translated into increased process stability in the presence of isobutanol. 7M.D showed an IC_{50} value almost two times higher than that of WT KdcA. 7M.D could tolerate almost threefold higher isobutanol concentrations at 50 °C with an $Iso_{50}^{50\,°C,1\,h}$ value of 6.2 vol% as opposed to just 2.3 vol% for WT KdcA. At 4 vol% isobutanol, at which no known microorganism could proliferate, 7M.D showed a half-life of 14 h at 50 °C, which corresponds to a more than 600-fold improvement relative to WT KdcA.[42,43] 7M.D was still highly active (>20% initial activity) at 9 vol% isobutanol, for which phase separation began to occur. The two-phase behavior of isobutanol and water will facilitate easy product removal. This improved process stability of 7M.D together with a simple separation step would lead to lower production costs and make in vitro isobutanol synthesis more feasible.[5] We anticipate that our study on evolving KdcA will further advance in vivo and in vitro production of isobutanol to serve as a sustainable candidate for fossil-fuel replacement.

Experimental Section

Reagents

All chemicals used in this work were of analytical grade or higher. Phusion polymerase and *Dpn*I were purchased from New England Biolabs (Germany). *E. coli* BL21 (DE3) was obtained from Invitrogen (Germany). All keto acid substrates were purchased from Sigma-Aldrich (Germany). Isobutanol was purchased from VWR (Germany).

Library development

All libraries were developed using a standard QuikChange protocol with slight modifications in primer designs.[45] WT KdcA was cloned to pET28a with N-terminal hexahistidine tag according to Göcke et al.[27]

Engineering approach

All primers are listed in the Supporting Information S10.

Cell-lysate screening

E. coli BL21 (DE3) containing the libraries was grown on lysogeny broth (LB) agar plates with kanamycin (50 μg mL^{-1}) overnight. Single colonies were picked and inoculated in 96-deep-well plates (GBO, Germany) containing auto-induction media with kanamycin 100 μg mL^{-1} and the cultures were grown overnight at 30 °C.[46] *E. coli* harboring the WT or respective mutant plasmids were used as positive controls and distributed randomly in eight wells throughout the plate, while four wells containing media were used as negative controls. On the following day, 100 μL of the culture was transferred to a sterile transparent F-bottom microplate (GBO, Germany) containing 100 μL of glycerol, which was stored at −80 °C as back-up plate. The other 200 μL of the culture was transferred to a U-bottom microplate (Sarstedt, Germany), and subsequently 50 μL of lysis buffer was added (final concentration after mixing was lysozyme 1 mg mL^{-1}, polymyxin B sulfate 0.5 mg mL^{-1}, DNAse 0.1 mg mL^{-1}, thiamine diphosphate (TDP) 0.1 mM, and $MgCl_2$ 2.5 mM). Lysis was performed at 37 °C for 1 h with an aluminum seal (Corning, USA) covering the microplate. Cell debris was pelleted by centrifugation at 4000 *g* for 15 min and clear supernatants were obtained. 20 μL of the supernatant was transferred to a new transparent F-bottom microplate to measure the initial activity, and a further 50 μL was transferred to a 96-well PCR plate (Brand, Germany) for the heat challenge (10 min at a different temperature, depending on the positive control used to give a residual activity <20%). After the heat challenge, the PCR plate was centrifuged at 4000 *g* for 15 min. 20 μL of the supernatant was transferred to a new F-bottom microplate to measure the residual activity.

Initial and residual activities were measured by adding 180 μL of the substrate solution containing phosphate buffer (50 mM pH 6.5), KIV (10 mM), TDP (0.1 mM) and $MgCl_2$ (2.5 mM). Substrate consumption was monitored at 310 nm for 15 min at 30 °C with a Multiskan spectrophotometer (Thermo Scientific, Germany; see Supporting Information S11 for detailed explanation of direct assay). For screening of the improved IC_{50}, a further 20 μL of supernatant was transferred to a third F-bottom microplate plate. 180 μL of the substrate solution with addition of 4% isobutanol (final concentration) was added, and the activity was monitored at 310 nm for 15 min at 45 °C. Variants that appeared to have at least three times higher residual activity compared to the positive controls were subjected to a second screening. *E. coli* cells harboring plasmids encoding the respective variants from the glycerol stock were grown overnight for plasmid purification for the second screening. New *E. coli* BL21 (DE3) competent cells were transformed with the respective plasmids encoding the variants and the whole screening procedure was repeated. Eight or more colonies were picked in

ChemPubSoc Europe

CHEMSUSCHEM Full Papers

the second screening, and the plasmids encoding the variants that retained their thermostability were sent for sequencing for final confirmation (Eurofins Genomics, Germany).

Enzyme purification

E. coli BL21 (DE3) was transformed with the WT or mutant plasmids and grown overnight on LB agar with 50 μg mL^{-1} kanamycin. A single colony was inoculated in LB (50 mL) with kanamycin (50 μg mL^{-1}) overnight. The overnight culture was then transferred to a 2 L baffled flask containing auto-induction medium (450 mL) with kanamycin (100 μg mL^{-1}).[46] The main culture was grown at 37 °C for 2 h at 120 rpm and the temperature was reduced to 18 °C and the incubation continued overnight. The cells were harvested by centrifugation at 4000 *g* for 15 min. The pellet was re-suspended in binding buffer (40 mL, phosphate buffer 50 mM/pH 8, NaCl 500 mM, imidazole 20 mM, and glycerol 10 vol%) and the final concentration of DNAse (0.1 μg mL^{-1}) was added. Cell disruption was performed by ultrasonication (Hielscher, Germany) in ice in 0.5 s cycles with an amplitude of 70% for 20 min. Cell debris was removed by centrifugation at 20000 *g* for 15 min at room temperature. The clear supernatant was applied to an Äkta Purifier equipped with a 1 mL His-trap column (GE Healthcare, Germany), and the column was washed with binding buffer (5 mL). The WT or variant proteins were eluted with elution buffer (phosphate buffer 50 mM/pH 8, NaCl 500 mM, imidazole 500 mM, and glycerol 10%) by monitoring the UV signal at 280 nm. The eluted fractions were subjected to buffer exchange on a HiTrap Desalting Column (GE Healthcare, Germany). Phosphate buffer (50 mM pH 6.5) was used as the desalting buffer.

Protein concentration determination

All enzymes were measured with a NanoPhotometer (Implen, Germany). Molecular weight and extinction coefficients were determined by ProtParam (https://web.expasy.org/protparam/).

Determination of T_{50}^{1h}

Enzyme solution (20 μL, 1 mg mL^{-1}) containing TDP (0.1 mM) and $MgCl_2$ (2.5 mM) were transferred to a PCR plate (Brand, Germany). The PCR plate was then incubated in a thermocycler (Bio-Rad, Germany) at a temperature gradient from 50 to 75 °C for 1 h. After incubation, the enzymes were diluted tenfold with dilution buffer (phosphate buffer, 50 mM/pH 6.5, TDP 0.1 mM, and $MgCl_2$ 2.5 mM). Diluted enzyme (20 μL) was transferred to an F-bottom microplate. Substrate solution (180 μL, phosphate buffer 50 mM/pH 6.5, TDP 0.1 mM, $MgCl_2$ 2.5 mM, and KIV 50 mM) was added, and the reaction was monitored with Multiskan (Thermo Scientific, Germany) at 310 nm at 30 °C for 15 min. A linear decrease was observed. Three independent biological replicates were performed. An average of the relative activities and the standard deviation were plotted (Figure S1). For T_{50}^{1h} determination, third-degree polynomials were used. The stepwise increase of T_{50}^{1h} is shown in Figure 1.

Measurement of melting temperature by Thermofluor

The procedure for T_m determination followed a previously described protocol.[24] Enzyme (2 μL, 1 mg mL^{-1}) was added to a clear PCR plate (Bio-Rad, Germany). Dilution buffer (21 μL, phosphate buffer 50 mM/pH 6.5, TDP 0.1 mM, and $MgCl_2$ 2.5 mM) was added followed by 2 μL of diluted SYPRO Orange (1 μL of stock solution in 79 μL water). The PCR plate was then sealed with a transparent seal (Bio-Rad, Germany) and centrifuged at 800 *g* for 1 min to mix the solution. The change in florescence was monitored by the preset program FRET in Bio-Rad RT-PCR CFX96 from 5 to 95 °C. The heating rate was 0.5 °C s^{-1}. The relative fluorescence unit (RFU) and derivative curve were calculated by Biorad CFX Manager. Three independent biological replicates were performed.

Kinetic characterization

Enzyme (20 μL) was added to a 96-well F-bottom microplate (GBO, Germany). Preheated reaction buffer (180 μL) containing different concentrations of KIV (end concentration 0–100 mM), TDP (0.1 mM), and $MgCl_2$ (2.5 mM) in phosphate buffer (50 mM/pH 6.5 were added to the enzyme. Substrate preheating was done by incubation of the reaction buffer at 30 °C for 30 min. Specific enzyme activity was monitored directly with the Multiskan every 30 s as the decrease of the keto acid group of KIV at 310 nm for 10 min. The temperature during measurement was kept at 30 °C. Specific enzyme activity was defined as the amount of enzyme (per monomer) required to decarboxylate 1 μmol of KIV per minute. Turnover number k_{cat} was defined as the number of substrate molecule converted per second per enzyme monomer. Three independent technical replicates were performed.

IC_{50} and $Iso_{50}^{50\,°C,\,1h}$ determination

For isobutanol IC_{50}, the reaction solution (180 μL, phosphate buffer 50 mM/pH 6.5, TDP 0.1 mM, $MgCl_2$ 2.5 mM, and KIV 50 mM) containing different concentrations of isobutanol (end concentration 0–9 vol% was preheated in a 96-well U-bottom microplate (Sarstedt, Germany) and covered with an aluminum seal (Corning, USA) at 50 °C for 30 min. The reaction solution was then added to a transparent 96-well F-bottom microplate (GBO, Germany) that had been filled with 20 μL of the WT or mutant enzyme. The reaction was monitored by using a BioTek Spectrophotometer Epoch 2 (Germany) at 310 nm, at 50 °C for 15 min. Three independent technical replicates were performed. Figure 3 shows the average relative activity and standard deviations.

For isobutanol $Iso_{50}^{50\,°C,\,1h}$, enzyme solution (20 μL, 1 mg mL^{-1}) containing TDP (0.1 mM) and $MgCl_2$ (2.5 mM) was incubated in a PCR plate (Brand, Germany). Isobutanol (100%) was added to the respective PCR wells to give end isobutanol concentrations of 0–8 vol%. Incubation was performed at 50 °C for 1 h. After incubation, the enzymes were diluted tenfold with dilution buffer as described previously. Diluted enzyme (20 μL) was transferred to a transparent 96-well F-bottom microplate (GBO Germany). substrate solution (180 μL) was added to each well and the reaction was monitored with the Multiskan (Thermo Scientific, Germany) at 310 nm, and 30 °C for 15 min. A linear decrease was observed. Three independent technical replicates were performed. The average relative activity and standard deviations are shown in Figure 6.

Substrate range determination

WT KdcA, 7M.D, and 7M.D after heat activation (20 μL), were transferred to a transparent 96-well F-bottom microplate (GBO Germany). Reaction buffer (180 μL) containing different α-keto acids (end concentration 50 mM) was added. For KIV, KV, and KB the enzymatic activities were monitored at 310 nm, consumption of pyruvate was monitored at 320 nm, and for BF monitoring was performed at 375 nm (see Supporting Information S11 for details of the

ChemPubSoc Europe

CHEMSUSCHEM Full Papers

direct). The decrease in the keto acid groups was measured at 30 °C with the Multiskan for 15 min. Specific enzyme activity was defined as the amount of enzyme (per monomer) to decarboxylate 1 μmol of respective keto acid per minute. Three independent technical replicates were performed.

Kinetic stability

For kinetic stability at 70 °C, WT and 7M.D enzyme solutions (100 μL, 1 mg mL^{-1}) containing TDP (0.1 mM) and $MgCl_2$ (2.5 mM) were incubated in a 96-well PCR plate (Brand, Germany) at 70 °C. 5 μL aliquots of both enzyme solutions were taken at several points in time and diluted tenfold with dilution buffer. Initial activity at t_0 and the residual activity after incubation at different points in time were measured at 30 °C as described previously. Three independent technical replicates were conducted. Average activity and standard deviations are presented in Figure 5.

For kinetic stability in 4 vol% isobutanol, a similar procedure was done, except that isobutanol was added to the enzyme solution to give an end concentration of 4 vol% was added. Incubation was performed at 50 °C. 5 μL aliquots of both enzyme solutions were taken at certain points in time. Three independent technical replicates were conducted. Figure 6B shows the average activity and standard deviations.

Acknowledgements

The authors acknowledge financial support by German Federal Ministry of Education and Research (BMBF) through HotsysAPP project (grant No. 031L0078F) and COST action CM1303 Systems Biocatalysis. Fruitful discussions with André Pick, Ioannis Zachos, and Samed Güner are greatly appreciated. The authors thank Hendrik Hohagen for his comments and Scott Bottoms for proofreading the manuscript. The authors declare no conflict of interest.

Conflict of interest

The authors declare no conflict of interest.

Keywords: alcohols · decarboxylation · enzyme catalysis · enzymes · high-throughput screening

[1] S. Sorrell, *Renewable Sustainable Energy Rev.* **2015**, *47*, 74 – 82.
[2] M. R. Connor, J. C. Liao, *Curr. Opin. Biotechnol.* **2009**, *20*, 307 – 315.
[3] S. Atsumi, T. Hanai, J. C. Liao, *Nature* **2008**, *451*, 86 – 89.
[4] Y. J. Choi, J. Lee, Y. Jang, S. Y. Lee, *MBio* **2014**, *5*, e01524-14.
[5] J. K. Guterl, D. Garbe, J. Carsten, F. Steffler, B. Sommer, S. Reiße, A. Philipp, M. Haack, B. Rühmann, A. Koltermann, U. Kettling, T. Brück, V. Sieber, *ChemSusChem* **2012**, *5*, 2165 – 2172.
[6] L. M. J. Soh, W. S. Mak, P. P. Lin, L. Mi, F. Y. H. Chen, R. Damoiseaux, J. B. Siegel, J. C. Liao, *ACS Synth. Biol.* **2017**, *6*, 610 – 618.
[7] P. P. Lin, K. S. Rabe, J. L. Takasumi, M. Kadisch, F. H. Arnold, J. C. Liao, *Metab. Eng.* **2014**, *24*, 1 – 8.
[8] T. D. Brock, K. M. Brock, R. T. Belly, R. L. Weiss, *Arch. Mikrobiol.* **1972**, *84*, 54 – 68.
[9] J. Quehenberger, L. Shen, S. V. Albers, B. Siebers, O. Spadiut, *Front. Microbiol.* **2017**, *8*, 1 – 13.
[10] F. Ollivier, P. Rousseaux, C. Cellier, (Metabolic Explorer), WO 2011003962A2, **2011**.
[11] V. G. H. Eijsink, S. Gåseidnes, T. V. Borchert, B. Van Den Burg, *Biomol. Eng.* **2005**, *22*, 21 – 30.
[12] B. P. L. Wintrode, F. H. Arnold in *Advances in Protein Chemistry, Vol. 55* (Ed.: F. H. Arnold), Academic Press, Sand Diego, **2001**, pp. 161 – 225.
[13] H. Yang, L. Liu, J. Li, J. Chen, G. Du, *ChemBioEng Rev.* **2015**, *2*, 87 – 94.
[14] V. Sieber, A. Plückthun, F. X. Schmid, *Nat. Biotechnol.* **1998**, *16*, 291 – 294.
[15] R. A. Chica, N. Doucet, J. N. Pelletier, *Curr. Opin. Biotechnol.* **2005**, *16*, 378 – 384.
[16] M. T. Reetz, S. Prasad, D. Carballeira, *J. Am. Chem. Soc.* **2010**, *132*, 9144 – 9152.
[17] M. De La Plaza, P. Fernández De Palencia, C. Peláez, T. Requena, *FEMS Microbiol. Lett.* **2004**, *238*, 367 – 374.
[18] B. A. Smit, J. E. T. V. H. Vlieg, W. J. M. Engels, L. Meijer, J. T. M. Wouters, G. Smit, *Appl. Environ. Microbiol.* **2005**, *71*, 303 – 311.
[19] C. L. Berthold, D. Gocke, M. D. Wood, F. J. Leeper, M. Pohl, G. Schneider, *Acta Crystallogr. Sect. D* **2007**, *63*, 1217 – 1224.
[20] M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem. Int. Ed.* **2006**, *45*, 7745 – 7751; *Angew. Chem.* **2006**, *118*, 7909 – 7915.
[21] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891 – 903.
[22] X. F. Zhang, G. Y. Yang, Y. Zhang, Y. Xie, S. G. Withers, Y. Feng, *Sci. Rep.* **2016**, *6*, 33797.
[23] M. Kokkinidis, N. M. Glykos, V. E. Fadouloglou in *Advances in Protein Chemistry and Structural Biology, Vol. 87* (Eds.: T. Karabencheva, C. Christov), Academic Press, San Diego, **2012**.
[24] S. Boivin, S. Kozak, R. Meijers, *Protein Expression Purif.* **2013**, *91*, 192 – 206.
[25] A. S. Bommarius, M. F. Paye, *Chem. Soc. Rev.* **2013**, *42*, 6534.
[26] A. Yep, G. L. Kenyon, M. J. McLeish, *Bioorg. Chem.* **2006**, *34*, 325 – 336.
[27] D. Gocke, C. L. Nguyen, M. Pohl, T. Stillger, L. Walter, M. Müller, *Adv. Synth. Catal.* **2007**, *349*, 1425 – 1435.
[28] V. G. H. Eijsink, A. Bjørk, S. Gåseidnes, R. Sirevåg, B. Synstad, B. Van Den Burg, G. Vriend, *J. Biotechnol.* **2004**, *113*, 105 – 120.
[29] A. Martin, V. Sieber, F. X. Schmid, *J. Mol. Biol.* **2001**, *309*, 717 – 726.
[30] P. L. Wintrode, D. Zhang, N. Vaidehi, F. H. Arnold, W. A. Goddard, *J. Mol. Biol.* **2003**, *327*, 745 – 757.
[31] S. S. Strickler, A. V. Gribenko, A. V. Gribenko, T. R. Keiffer, J. Tomlinson, T. Reihle, V. V. Loladze, G. I. Makhatadze, *Biochemistry* **2006**, *45*, 2761 – 2766.
[32] C. Ó'Fágáin, *Enzyme Microb. Technol.* **2003**, *33*, 137 – 149.
[33] B. M. Nestl, B. Hauer, *ACS Catal.* **2014**, *4*, 3201 – 3211.
[34] A. D. Nagi, L. Regan, *Folding Des.* **1997**, *2*, 67 – 75.
[35] S. Ahmad, M. Z. Kamal, R. Sankaranarayanan, N. M. Rao, *J. Mol. Biol.* **2008**, *381*, 324 – 340.
[36] I. Wu, F. H. Arnold, *Biotechnol. Bioeng.* **2013**, *110*, 1874 – 1883.
[37] M. Lehmann, L. Pasamontes, S. F. Lassen, M. Wyss, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2000**, *1543*, 408 – 415.
[38] C. Hoppe, D. Schomburg, *Protein Sci.* **2005**, *14*, 2682 – 2692.
[39] A. Bosshart, S. Panke, M. Bechtold, *Angew. Chem. Int. Ed.* **2013**, *52*, 9673 – 9676; *Angew. Chem.* **2013**, *125*, 9855 – 9858.
[40] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774 – 797.
[41] M. Maier, C. P. Radtke, J. Hubbuch, C. M. Niemeyer, K. S. Rabe, *Angew. Chem. Int. Ed.* **2018**, *57*, 5539 – 5543; *Angew. Chem.* **2018**, *130*, 5638 – 5642.
[42] S. S. Atsumi, T. Y. Wu, I. M. P. MacHado, W. C. Huang, P. Y. Chen, M. Pellegrini, J. C. Liao, *Mol. Syst. Biol.* **2010**, *6*, 1 – 11.
[43] M. Kanno, T. Katayama, H. Tamaki, Y. Mitani, X. Y. Meng, T. Hori, T. Narihiro, N. Morita, T. Hoshino, I. Yumoto, N. Kimura, S. Hanada, Y. Kamagata, *Appl. Environ. Microbiol.* **2013**, *79*, 6998 – 7005.
[44] E. Vazquez-Figueroa, V. Yeh, J. M. Broering, J. F. Chaparro-Riggers, A. S. Bommarius, *Protein Eng. Des. Sel.* **2008**, *21*, 673 – 680.
[45] L. Zheng, *Nucleic Acids Res.* **2004**, *32*, e115.
[46] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207 – 234.

Manuscript received: April 30, 2018
Revised manuscript received: June 15, 2018
Accepted manuscript online: June 28, 2018
Version of record online: July 31, 2018

CHEMISTRY & SUSTAINABILITY
CHEMSUSCHEM
ENERGY & MATERIALS

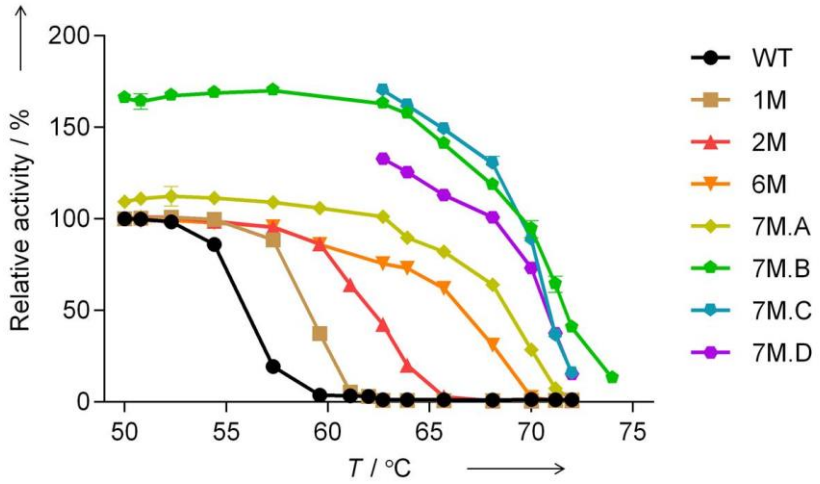
Supporting Information

Structure-Guided Engineering of α-Keto Acid Decarboxylase for the Production of Higher Alcohols at Elevated Temperature

Samuel Sutiono,[a] Jörg Carsten,[b] and Volker Sieber*[a, b, c, d]

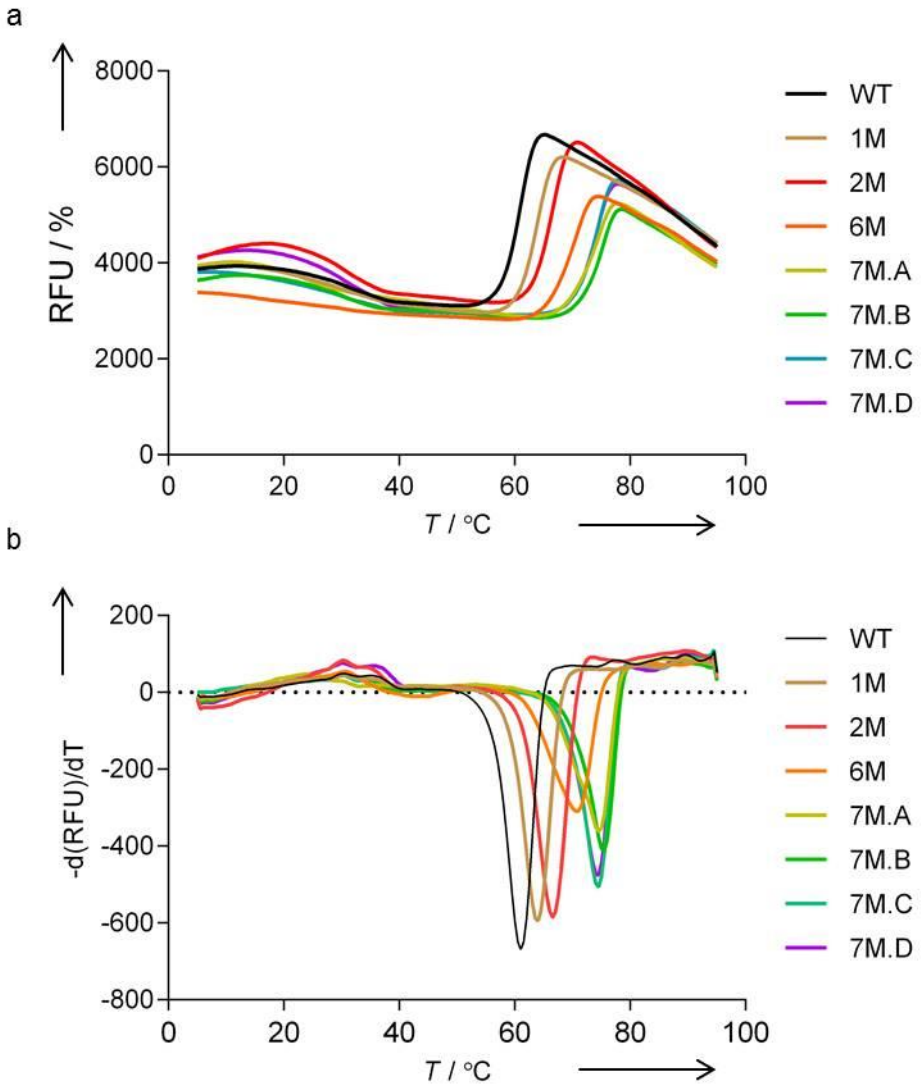
cssc_201800944_sm_miscellaneous_information.pdf

Determination of T_{50}^{1h}



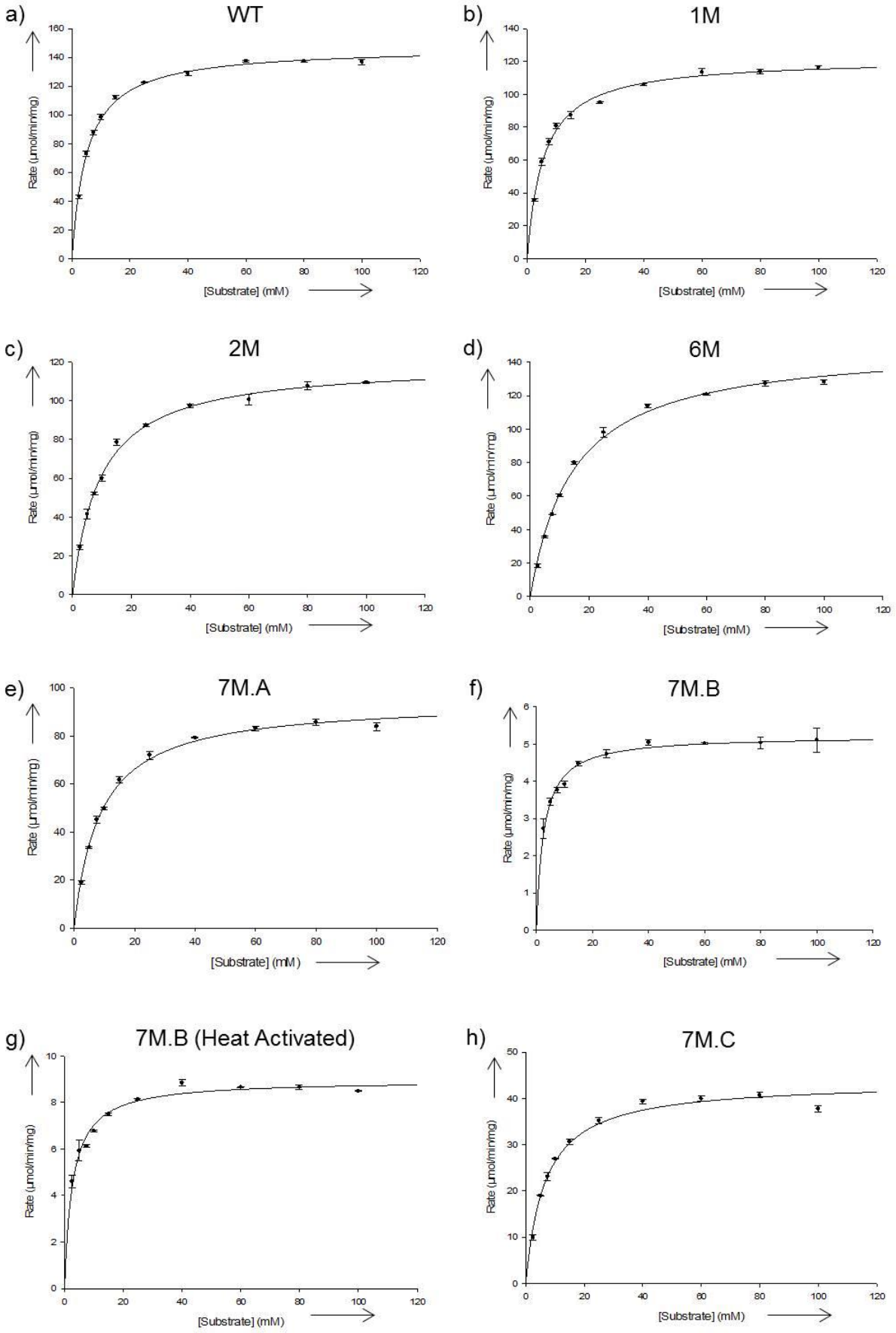
S1. Stability of WT KdcA and all the variants as measured by the decrease in initial activity after 1 h incubation at different temperatures. 100 % activity is the activity of each enzyme incubated at 30 °C for 1 h. Variants 7M.B, 7M.C, and 7M.D showed heat activation profiles. Three independent repeats were performed

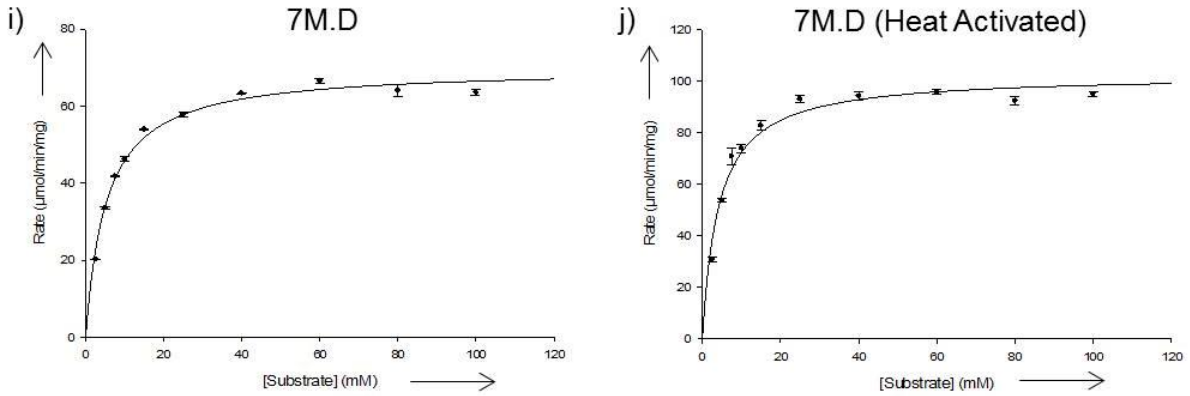
Melting Curve determined by Thermofluor



S2. Melting curves of WT and all the variants presented in Table 1 obtained from Thermofluor assays. (A) Primary data is given as plot of RFU (Relative Fluorescence Unit) vs *T* (°C). (B) Derivative data is plotted against *T* (°C). The lowest point in S2 B is the melting temperature (T_m) as presented in Table 1. Three independent repeats were performed.

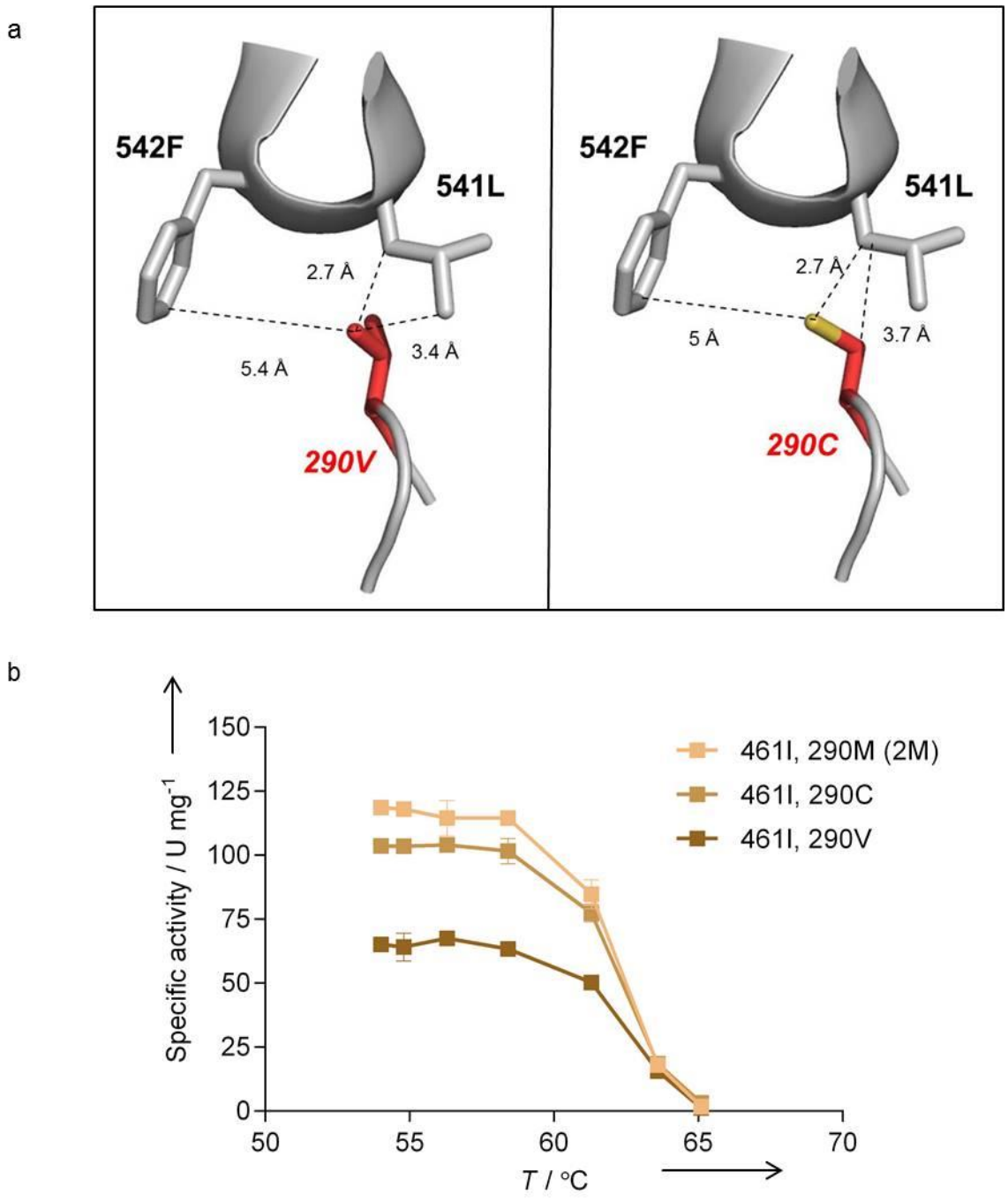
Kinetic Characterization of the WT and variants





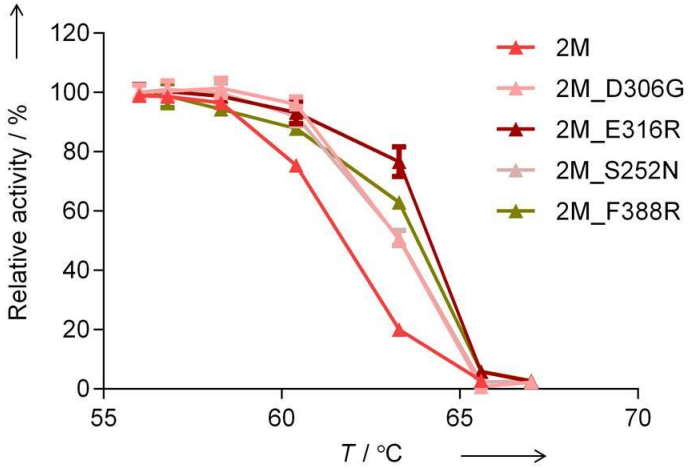
S3. Michaelis Menten plot of WT and mutants as presented in Table 1 calculated by Sigma Plot. Heat activation was carried out by incubating the respective enzymes at 60 °C for 1 h. Activity was measured at 30 °C as the consumption of ketoisovalerate (KIV) over time. Detail assay conditions were presented in the Experimental Section. Three independent repeats were performed..

Engineeering based on sequence homology approach: Substitution at position 290



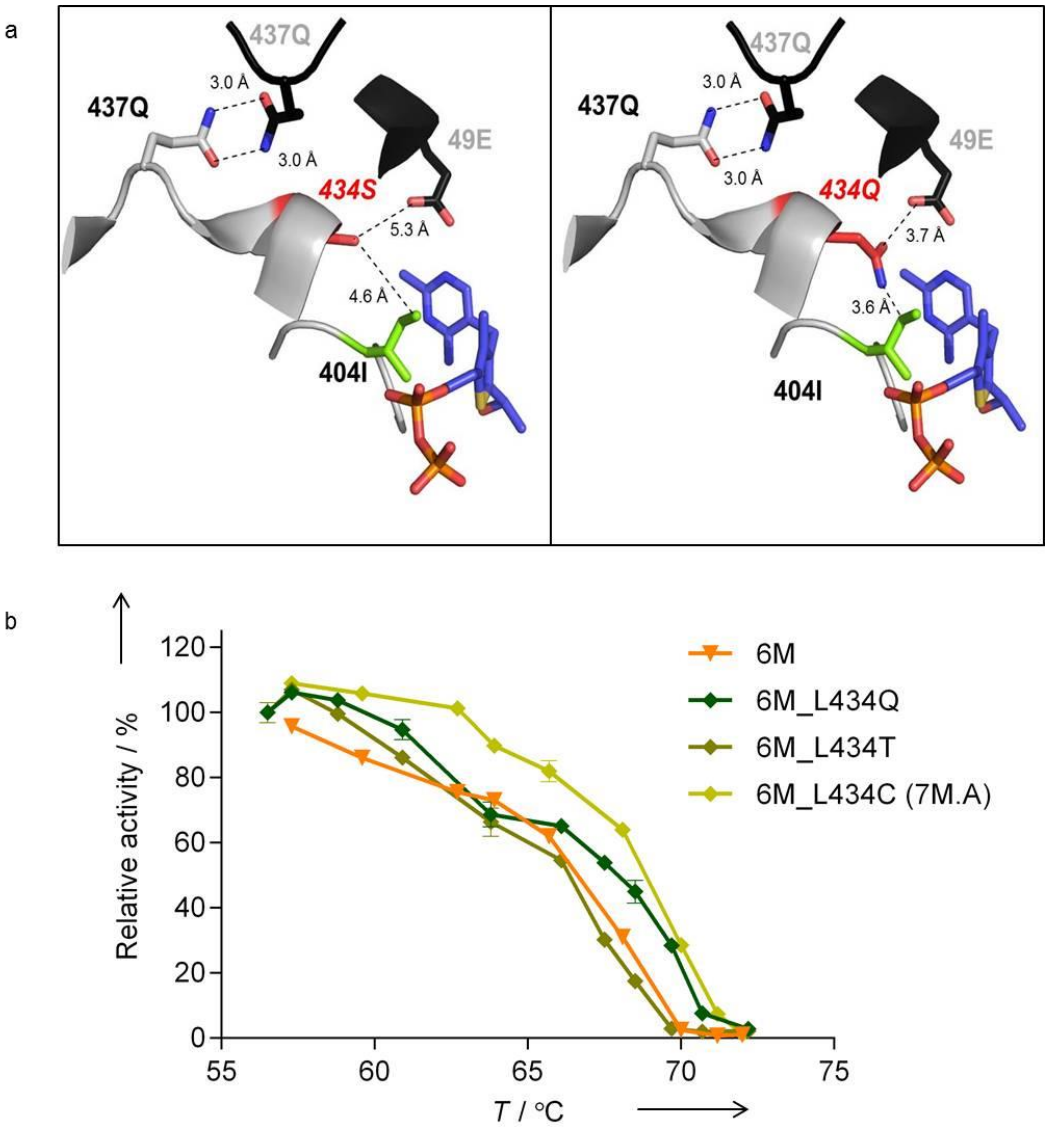
S4. Different substitutes at position 290 led to similar thermostability. (A) Substitution of alanine to valine (left) and to cysteine (right) at position 290. (B) Residual activity of 2M compared to 1M_A290C and 1M_A290V after 1 h incubation at different temperatures. Activity was measured at 30 °C with 50 mM KIV concentration. Three independent repeats were performed.

Surface engineering



S5. The effect of single amino acid exchanges on the protein surface on the increased stability. Activity was measured at 30 °C with 50 mM KIV concentration. 100% activity was the activity of each variant after incubation at 30 °C for 1 h. Three independent repeats were performed.

Dimer optimization: Substitution at position 434



S6A. Different substitutions at position 434 led to different stabilities. (A) Substitution of leucine at position 434 by threonine (left) and by glutamine (right). (B) Heat tolerance profile of the different variants as compared to 6M. Activity was measured at 30°C with 50 mM KIV concentration. 100% activity was the activity of each variant after incubation at 30°C for 1 h. Three independent repeats were performed.

S6B. Further explanations on how the substitution could lead to increased thermostability

An increase in thermostability by exchanging a bigger to a smaller amino acid at position 434 was not initially expected. We had hypothesized that a mutation to a larger amino acid at this position would have increased the hydrophobic compaction, as was the case for 1M. In contrast to our first hypothesis, we proposed that a substitution at position 434 increased thermostability by decreasing its hydrophobic interaction to 404I; thus the adjacent loops is relaxed. This could lead to increased the electrostatic interactions of 437Q between the two monomers (Figure 2E). This was also supported by the fact that other beneficial exchanges at position 434 (to T or Q) would result in a reduced interaction to the non-polar isoleucine at position 404 (Figure S6A). Although it could be argued that the substitution by threonine or glutamine may also increase electrostatic interactions to 49E from the second monomer (Figure 2E and Figure S6A). However, the contributions of these new electrostatic interactions were not as impactful as decreasing the non-polar interaction to 404I, since the improvement in thermostability of 6M_L434T and 6M_L434Q were lower than for 6M_L434C (7M.A) (Figure S6A). The possibility of disulfide bond formation in 7M.A was eliminated by rationalizing that the distance between the cysteine at position 434 from the first monomer to another cysteine from the second monomer is beyond 10 Å, thus too far for meaningful interaction. The substitution

of isoleucine at position 404 by cysteine (7M.B) leading to an increase in thermostability may further support this argument (Figure 2F).

Further Experiments

S7A. Experiment on the two extended loop regions

There are two extended loop regions in the crystal structure of KdcA, from 170A to 188N and from 330L to 355S. Previous SSM experiments based on B-factor values of two amino acids in the first and second region (181E and 188N, 342K and 345E) did not result in any variants with higher thermostability. However, to investigate the assumption that loop flexibility is responsible for low thermostability we shortened the loops by performing a single amino acid deletion. It was reported previously that there is an indirect proportionality between loop length and temperature stability.[1] Δ177A and Δ352A deletions were constructed using the 2M as template. 2M_Δ177A resulted in almost complete loss of activity, while 2M_Δ352A resulted in almost 80 % reduction in activity, with no improvement in thermostability as compared to the 2M variant. To explore the influence of loop flexibility further, we proposed that a mutation in the main chain amino acids (globular protein core) could bring the two loops closer to the main chain. Increasing the interactions between the main chain and the loop regions has been shown to improve thermostability successfully, as reported in the case of a lipase.[2] However, on revisiting the crystal structure (PDB ID: 2vbg), we were unable to determine the amino acids surrounding the first loop region that could be exchanged to achieve this goal. The amino acids that form the α-helix and β-sheet were not considered in this approach. In the second loop region, we identified several potential candidates, such as S252 and F388. Substitution at these residues led to improved thermostability as presented in the Discussion (Figure 2C).

S7B. Attempts to transfer beneficial mutation reported by Maier et al.[3]

The work by Maier and cowokers on KivD showed an improvement of T_{50}^{10min} by more than 6°C for WT KivD after exchanging serine at residue 385 to methionine.[3] It is worth noting that the Liao group also described a substitution of the adjacent amino acid (S386P) giving rise to an increase in thermostability of WT KivD.[4] We opted to determine whether the beneficial mutation at the residue 385 in KivD could be transferred to our KdcA variant to further improve its stability. Accordingly, the gene of 7M.D was saturated at the codon encoding residue 385 (homologues position) with the degenerate codon NNS. Unfortunately, there were no variants with increased thermostability or improved initial activity in the presence of isobutanol found. This result led to prompt assumption that 7M.D might have been trapped in what is called the “local minima”.[5] After evaluation of all mutations in 7M.D, it was reasonable to assume that substitution at the residue 388 (F388R) was likely responsible for the "local minima" trap. It was reasoned that F388R could have altered adjacent amino acid networks, thus discriminating the role of the serine residue at 385 in enzyme stability.

To escape from the “local minima”, a set of primers was designed to mutate the gene of 7M.D at the codon encoding residue 388 back to the original phenylalanine that is found in WT KdcA as well as in WT KivD and at the same time to substitute serine residue at 385 to methionine (double substitution). By QuikChange® mutagenesis, variant 7M.E (7M.D_R388F_S385M) was created. After purifying 7M.E into homogeneity, the T_m was determined with Thermofluor and it exhibited melting peak at 74 °C (data not shown), which was 0.5 °C lower than 7M.D (Table 1). Saturation at the codon encoding residue 385 of 7M.E gene with the degenerate codon NNS failed to detect any variants with improved thermostability or increased in initial activity in the presence of isobutanol. 7M.E was not characterized further due to lower T_m in comparison to 7M.D.

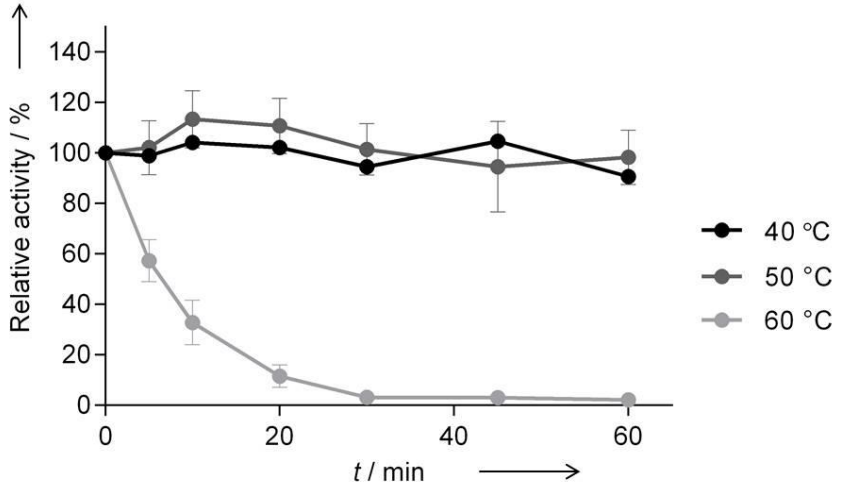
Isoenzymes are often structurally and evolutionary closely related. Therefore, it has been assumed that stabilizing mutations from one enzyme can be transferred to the isoenzyme.[6] We find, that this was not the case for KdcA and KivD. Some stabilizing substitutions reported in KivD did not give significant effects when the homologue positions in KdcA were exchanged. This may suggest that the stabilizing mutations described in one isoenzyme do not necessary give the same effect on another isoenzyme.

S8. Effect of mutations on activity

In nature, compared to the enzymes from mesophiles, most enzymes found in thermophilic organisms will show lower activity at ambient temperature.[7] This is reasoned because thermophiles do not need highly active enzymes at ambient temperature. Likewise, highly active enzymes at ambient temperatures from mesophiles or psychrophiles are often not stable at high temperatures.[8] This concept is known as the activity–stability trade-off.[7] Enzymes that diverge from this principle were generated usually through laboratory evolution.[8,9] 7M.D showed a lower k_{cat} at 30 °C as compared to the WT (Table 1). To understand, which substitutions contribute to this decrease in activity at ambient temperature, each generation of variant was characterized in detail (Table 1). The substitution at the catalytic center (1M) showed an almost 20 % decrease in k_{cat} without significant change in affinity (K_m). An additional substitution at position 290, near the surface (2M), did not decrease k_{cat} further, but the K_m nearly doubled. The four surface substitution (6M) restored initial activity, making it comparable to the WT, but the K_m was increased further suggesting that surface rigidification decreases substrate affinity but not activity. Further substitution at position 434 (7M.A) and 404 (7M.B) decreased activity by almost 40 % and by >90 % respectively (Table 1). The substitution at 434 was hypothesized to increase dimer interactions. The catalytic sites of KdcA are located at the interface of the two monomers, indicating modifications in dimer interaction may either increase dimer affinity, which in turn allows less conformational movement of the enzyme dimer, or change the shape or size of the catalytic center leading to a slight decrease in activity.[10] Increased interactions between monomers led to increased thermostability, but decreased activity has been reported before in the case of *Pc*DTE.[11] The mutation at position 404, from isoleucine to cysteine, will alter its interaction with the thiazole ring of thiamine diphosphate, the cofactor of KdcA (Figure 2F) causing 7M.B to suffer a robust decrease in activity. As 7M.A, 7M.C and 7M.D carried the same 6 mutations, differing only in the amino acid substitution at position 290. The mutation to valine (7M.C) and to cysteine (7M.D) at position 290 led to an increased isobutanol IC_{50} (Figure 3B) in both cases. In prior observations, different amino acid substitutions at position 290 affected the initial activities but not the thermostability (Figure S4).

In conclusion, the relationship between thermostability, k_{cat}, and K_m remains unclear. Unless thorough experiments to determine any changes in the thermodynamic parameters (ΔG, ΔH, and ΔS) upon each additional mutation are conducted, (beyond the scope of this paper), these relationships cannot be easily elucidated.[9]

Stability of WT KdcA



S9. Stability of WT KdcA at different temperatures over 1 h.

List of the degenerate primers used

The primers are grouped based on the engineering approach. Degenerate codons and or mutations are written in italics.

Table S10. List of primers used to develop the QuikChange® libraray

Strategies	Mutation	Fwd/Rev	Sequence (5´- 3´)
B-factor	L180NNK	Fwd	GCATTATCT*NNK*GAAAAAGAAAGCTCTACAACAAATACAACTGAAC
		Rev	CTTTTTC*MNN*AGATAATGCAGGCTTCTCTGCTTTTGC
	E181NNK	Fwd	GCATTATCTTTA*NNK*AAAGAAAGCTCTACAACAAATACAACTGAAC
		Rev	CTTTCTTT*MNN*TAAAGATAATGCAGGCTTCTCTGCTTTTGC
	K342NNK	Fwd	CAATATATTGAT*NNK*CAATATGAAGAATTTATTCCATCAAGTGC
		Rev	CTTCATATTG*MNN*ATCAATATATTGTCCTTCATATTCTATTCCTTTTAATTC
	E345NNK	Fwd	GATAAGCAATAT*NNK*GAATTTATTCCATCAAGTGCTCCCTTATC
		Rev	GAATAAATTC*MNN*ATATTGCTTATCAATATATTGTCCTTCATATTCC
Catalytic center	Q377NDT	Fwd	GCTGAA*NDT*GGAACCTCATTTTTTGGAGCTTCAAC
		Rev	GGTTCC*AHN*TTCAGCAACGATTGTTTCATTGCTTTGAGTC
	F381NDT	Fwd	GGAACCTCA*NDT*TTTGGAGCTTCAACAATTTTCTTAAAATCAAATAGTCG
		Rev	CTCCAAA*AHN*TGAGGTTCCTTGTTCAGCAACGATTGTTTC
	F382NDT	Fwd	GGAACCTCATTT*NDT*GGAGCTTCAACAATTTTCTTAAAATCAAATAGTCG
		Rev	GAAGCTCC*AHN*AAATGAGGTTCCTTGTTCAGCAACG
	V461NDT	Fwd	GGTTATAC*NDT*GAAAGAGAAATCCACGGACCTACTC
		Rev	GGATTTCTCTTTC*AHN*TGTATAACCATCATTATTTATGATAAAACAAATTGG
	I465NDT	Fwd	GAGAA*NDT*CACGGACCTACTCAAAGTTATAACGAC
		Rev	GGTCCGTG*AHN*TTCTCTTTCAACTGTATAACCATC
	F542NDT	Fwd	GGGTAAATTA*NDT*GCTGAGCAAAATAAATAAGAATTCGAGCTCCG
		Rev	GCTCAGC*AHN*TAATTTACCCATTTTTTTCAGTAATTTTGGCGC
Sequence homology	Q34NNK	Fwd	CAATTTTTAGAT*NNK*ATTATTTCACGCGAAGATATGAAATGGATTG
		Rev	GAAATAAT*MNN*ATCTAAAAATTGTAAGTTATAGTCACCAGGAACTC
	L130NNK	Fwd	GAACCT*NNK*ACAGCAGCGCGGACTTTAC
		Rev	GCTGT*MNN*AGGTTCATGCATCTTCATAAAGTGTTTAAAATC
	A290NNK	Fwd	CAGGT*NNK*TTCACACATCATTTAGATGAAAATAAAATGATTTC
		Rev	GTGAA*MNN*ACCTGTTGAGGAGTCCGTAAG
	T386NNK	Fwd	GCTTCA*NNK*ATTTTCTTAAAATCAAATAGTCGTTTTATTGGACAAC
		Rev	GAAAAT*MNN*TGAAGCTCCAAAAAATGAGGTTCCTTGTTC
Surface Engineering	E39NNS	Fwd	TTCACGC*NNS*GATATGAAATGGATTGGAAATGCTAATG
		Rev	TTCATATC*SNN*GCGTGAAATAATTTGATCTAAAAATTGTAAGTTATAGTC
	Q206NNS	Fwd	AAATGCC*NNS*AAACCAGTAGTGATTGCAGGAC
		Rev	TGGTTT*SNN*GGCATTTTTCAAACTTTCTTCAATCTTACTC
	Y257NNS	Fwd	AGGAATA*NNS*AACGGGAAACTTTCAGAAATCAGTCTTAAAAATTTTG
		Rev	CCCGTT*SNN*TATTCCTAAAAATGAGGGCAAAGATTCATC
	E316NNS	Fwd	GTGGTA*NNS*GATTTTGATTTTAGAGCAGTGGTTTCTTC
		Rev	AAAATC*SNN*TACCACTTTATTGAAAATTATTCCTTCATCTATGTTTAGTG
	D306NNS	Fwd	TAAACATA*NNS*GAAGGAATAATTTTCAATAAAGTGGTAGAAGATTTTGATTTTAG
		Rev	TTCCTTC*SNN*TATGTTTAGTGAAATCATTTTATTTTCATCTAAATGATGTG
	E420NNS	Fwd	GATAAA*NNS*AGCAGACACCTTTTATTTATTGGTGATGG
		Rev	TCTGCT*SNN*TTTATCCGCAATTTGGCTTCC
	E491NNS	Fwd	AGCAACA*NNS*GATCGTGTAGTATCAAAAATTGTTAGAACAG
		Rev	CACGATC*SNN*TGTTGCTCCAAATGTTTCTGG
	D492NNS	Fwd	AACAGAA*NNS*CGTGTAGTATCAAAAATTGTTAGAACAGAGAATG
		Rev	TACACG*SNN*TTCTGTTGCTCCAAATGTTTCTGG
	N517NNS	Fwd	GATGTC*NNS*AGAATGTATTGGATAGAACTAGTTTTGGAAAAAG
		Rev	ACATTCT*SNN*GACATCTGCTTGGGCTTCTTTCATG
Loop engineering	Δ177A	Fwd	GAAGCCTTTATCTTTAGAAAAAGAAAGCTCTACAACAAATAC
		Rev	AAAGATAAAGGCTTCTCTGCTTTTGCTGCAGC
	Δ352A	Fwd	ATCAAGTCCCTTATCACAAGACCGTCTATGG
		Rev	ATAAGGGACTTGATGGAATAAATTCTTCATATTGCTTATCAATATATTG
	252	Fwd	TTTGCCC*NNS*TTTTTTAGGAATATATAACGGGAAACTTTCAG
		Rev	TAAAAA*SNN*GGGCAAAGATTCATCAACAGCACTTTTACC
	388	Fwd	AACAATT*NNS*TTAAAATCAAATAGTCGTTTTATTGGACAACCTTTATGG
		Rev	ATTTTAA*SNN*AATTGTTGAAGCTCCAAAAAATGAGGTTCCTTG
Dimer interaction	V23NNS	Fwd	TTTTGGA*NNS*CCTGGTGACTATAACTTACAATTTTTAGATC
		Rev	ACCAGG*SNN*TCCAAAAATTTCTTCAATTCCCAACTC
	F291NNS	Fwd	AGGTGCA*NNS*ACACATCATTTAGATGAAAATAAAATGATTTCACTAAAC
		Rev	ATGTGT*SNN*TGCACCTGTTGAGGAGTCCG
	I404NNS	Fwd	GGTTCT*NNS*GGATATACTTTTCCAGCGGCTTTAGG
		Rev	TATATCC*SNN*AGAACCCCATAAAGGTTGTCCAATAAAACG
	L434NNS	Fwd	ACTTCAA*NNS*ACCGTACAAGAATTAGGACTATCAATCAGAG
		Rev	TACGGT*SNN*TTGAAGTGAACCATCACCAATAAATAAAAGG
	P476NNS	Fwd	GACATT*NNS*ATGTGGAATTACTCGAAATTACCAGAAAC
		Rev	TCCACAT*SNN*AATGTCGTTATAACTTTGAGTAGGTC
	F487NNS	Fwd	AGAAACA*NNS*GGAGCAACAGAAGATCGTGTAGTATC
		Rev	TTGCTCC*SNN*TGTTTCTGGTAATTTCGAGTAATTCCAC
Local minima	S385NNS	Fwd	TGGAGCT*NNS*ACAATTTTCTTAAAATCAAATAGTCGTTTTATTGGAC
		Rev	GAAAATTGT*SNN*AGCTCCAAAAAATGAGGTTCCTTGTTC
	S385M_ R388F	Fwd	TGGAGCT*ATG*ACAATT*TTC*TTAAAATCAAATAGTCGTTTTATTGGACAACC
		Rev	GATTTTAA*GAA*AATTGT*CAT*AGCTCCAAAAAATGAGGTTCCTTGTTCAGCAAC

Direct assay

S11A Direct assay for high-throughput screening

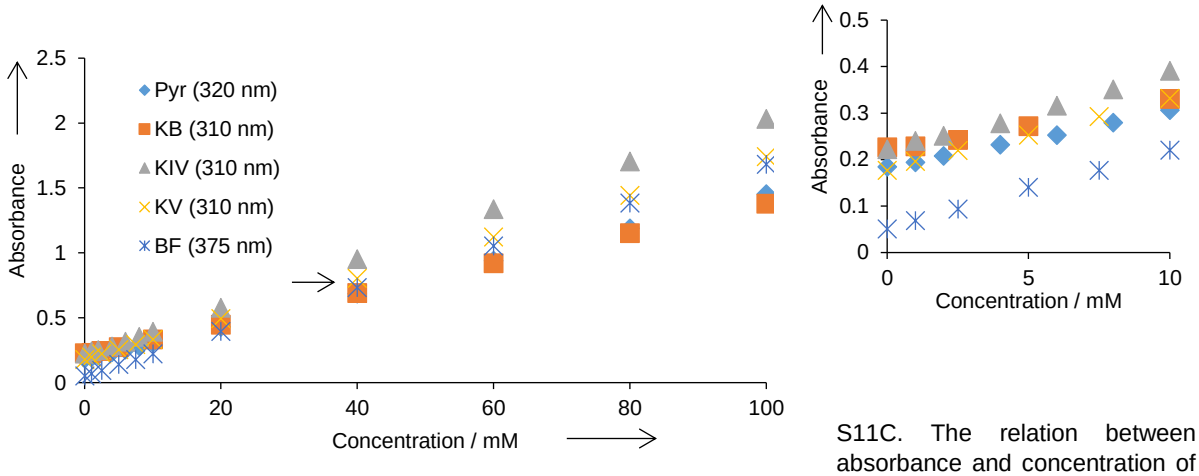
In developing high-throughput screening, there are three important factors that need to be considered: sensitivity, simplicity, and cost. As it was reported earlier, ketoisovalerate (KIV) gives absorption maximum at around 315 nm.[4,12] In this work, this characteristic was utilized. We developed a screening platform based on the absorption at 310 nm. At this wavelength, the decarboxylation product of KIV, isobutanal, does not give meaningful absorption. Thus, we can follow the decrease of absorption upon consumption of KIV.

In order to minimize the background of different cell density during screening, the cells were grown over night in the 96 deep well plate (GBO, Germany). After over night culture, cells in all wells reach saturation; thus there were no substantial differences of OD_{600}. After the cells were lysed with lysozyme, cell debris was pelleted by centrifugation and clear supernatant was obtained. Applying this supernatant for screening increased slightly the background absorption at 310 nm, but overall activity (the decrease of absorbance) was not affected. Supernatant from media controls and *Escherichia coli* harboring empty plasmid did not show any decrease of absorbance during the measurement (15 min). The supernatants subjected to heat challenge were also pelleted to obtain clear solution for the measurement of residual activity. Therefore, with direct assay at 310 nm, we could develop a robust yet simple screening platform for KdcA toward improved stability. This approach had been used earlier in other studies with slightly difference in the treatment of cell lysate.[4]

S11B Direct assay for kinetic and substrate scopes determination

Although some people argue that direct assay is not sensitive enough for determination of decarboxylase activity, we have successfully adapted the assay; thus the results obtained are comparable to the ones obtained by coupled assay.[12] The other reason we used direct assay over coupled assay for kinetic characterization was because we wanted to maintain the same assay procedure for all experiments, hence the results could be easily compared. Coupled assay with alcohol dehydrogenases (ADH) will not be suitable for measuring isobutanol IC_{50} because the outcomes will be dependent not only to how isobutanol inhibit KdcA's activity but also how isobutanol inhibit ADH's activity, thus will lead to incorrect results.

Other α-keto acids (corresponding salts) can also be measured in the similar manner. Ketobutyrate and ketovalerate showed maximum wavelength at 310 nm, the same as KIV. Meanwhile, pyruvate absorbed maximum at 320 nm. Benzoylformate did not show distinctive maximum wavelength during wavelength scanning due to the presence of two absorbing groups (benzene and keto acid). In order to detect only the keto acid group, the wavelength had to be shifted to 375 nm. At this wavelength, benzene group does not give any absorption, only the keto acid group of benzoylformate. Thus, direct assay could be used reliably for measuring enzyme activity toward different α-keto acids with slight adjustment on the wavelength to achieve desired sensitivity.



S11C. The relation between absorbance and concentration of different α-keto acids measured at specific wavelength. The results presented as the average of three replications. On the upper right side, graph at smaller concentration range is shown. Coefficient of determination (R^2) of each graph is >0.99.

References

[1] A. D. Nagi, L. Regan, *Fold. Des.* **1997**, *2*, 67–75.

[2] S. Ahmad, M. Z. Kamal, R. Sankaranarayanan, N. M. Rao, *J. Mol. Biol.* **2008**, *381*, 324–340.

[3] M. Maier, C. P. Radtke, J. Hubbuch, C. M. Niemeyer, K. S. Rabe, *Angew. Chemie. Int. Ed.* **2018**, *57*, 5539–5543.

[4] L. M. J. Soh, W. S. Mak, P. P. Lin, L. Mi, F. Y. H. Chen, R. Damoiseaux, J. B. Siegel, J. C. Liao, *ACS Synth. Biol.* **2017**, *6*, 610–618.

[5] G. Lois, J. Blawzdziewicz, C. S. O'Hern, *Biophys. J.* **2008**, *95*, 2692–2701.

[6] E. Vázquez-Figueroa, J. Chaparro-Riggers, A. S. Bommarius, *Chembiochem.* **2007**, 8, 2298–2301

[7] K. S. Siddiqui, R. Cavicchioli, *Annu. Rev. Biochem.* **2006**, *75*, 403–433.

[8] B. P. L. Wintrode, F. H. Arnold, in *Adv. Protein Chem.* (Ed.: F.H. Arnold), Academic Press, **2001**, pp. 161–225.

[9] S. R. Miller, *Evolution (N. Y).* **2017**, *71*, 1876–1887.

[10] C. L. Berthold, D. Gocke, M. D. Wood, F. J. Leeper, M. Pohl, G. Schneider, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *63*, 1217–1224.

[11] A. Bosshart, S. Panke, M. Bechtold, *Angew. Chemie.* **2013**, 125, 9855–9858; *Angew. Chemie. Int. Ed.* **2013**, *52*, 9673–9676.

[12] D. Gocke, C. L. Nguyen, M. Pohl, T. Stillger, L. Walter, M. Müller, *Adv. Synth. Catal.* **2007**, *349*, 1425–1435.

3.2 To beat the heat – engineering of the most thermostable pyruvate decarboxylase to date

In this publication, a pyruvate decarboxylase (PDC) was engineered to increase its stability. Two previous works had been focused to engineer stability of a PDC *via* rational design, e.g. ancestral sequence reconstruction (ASR) and Rosetta-based prediction. There have been a number of PDC enzymes reported in literatures. Several PDCs described in the literature were cloned and characterized in this work. In addition, two new yeast PDCs (*Cg*PDC and *Zr*PDC) were also cloned and characterized. Similar to the literature, three bacterial PDCs: *Ap*PDC, *Zm*PDC, and *Zp*PDC demonstrated the highest thermostability and activity. *Cg*PDC and *Zr*PDC showed typical substrate cooperation behavior for yeast PDCs and rather much lower activity and thermostability. Three bacterial PDCs were used as templates for directed evolution. Staggered extension process was chosen due to its simplicity and proven effectiveness toward generating efficient libraries in general and specifically improving enzyme stability.

After screening about 800 possible variants, one variant showed significant improvement after heat challenge at 75 °C. This variant (PDC-Var.1) was shown to bear 13 point substitutions. PDC-Var.1 has 98% sequence similarity to *Ap*PDC. SSM at seven homologous positions in PDC-Var.1 from the previous work with *Ll*KdcA (variant 7M.D), one variant—PDC-Var.1_L441V (PDC-Var.2)—showed further increased in thermostability. PDC-Var.2 demonstrated an increase of T_{50}^{1h} and T_m by 13.6 °C and 12.0 °C in comparison with WT *Ap*PDC, respectively. At 65 °C, 70 °C, and 75 °C, PDC-Var.2 showed half-lives of 18 h, 10.7 h, and 7.3 h, corresponding to 19-fold, >500, and >2400-fold increase as compared with WT *Ap*PDC, respectively. PDC has also been used in two cell-free approaches of n-butanol production. PDC-Var.2 showed a half-life of 15 h at 9 vol% n-butanol, 50 °C (maximum solubility of n-butanol in water). The half-life is >5000-fold improved in comparison with WT *Ap*PDC. The variant also showed to be able to withstand ethanol 40 vol% at 50 °C, a 2-fold improvement to WT *Ap*PDC.

The author designed the whole experiments and devised the engineering strategy under the supervision of the co-author Prof. Sieber. Together, the author with the co-author Ms. Satzinger developed the library *via* StEP resulted in two libraries. One library was screened by Ms. Satzinger as part of her Master´s thesis and the other one was screened by the author. Variants with similar stability were obtained independently. The author continued with the variant obtained from his screening (PDC-Var.1) to obtain a more stable variant (PDC-Var.2). The other co-author Mr. Pick together with the author cloned two new yeast PDCs (*Cg*PDC and *Zr*PDC). The author characterized all the enzymes cloned and the variants obtained in this study. The author wrote the manuscript. The other co-author Dr. Carsten contributed to content and language of this publication. The co-author Prof. Sieber critically reviewed the submitted manuscript.

To beat the heat – engineering of the most thermostable pyruvate decarboxylase to date

Samuel Sutiono, Katharina Satzinger, André Pick, Dr. Jörg Carsten, Prof. Volker Sieber

RSC Advances

2019

DOI: 10.1039/C9RA06251C

RSC Advances

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *RSC Adv.*, 2019, **9**, 29743

To beat the heat – engineering of the most thermostable pyruvate decarboxylase to date†

Samuel Sutiono,[a] Katharina Satzinger,‡[a] André Pick,[a] Jörg Carsten[ab] and Volker Sieber*[abcd]

Received 11th August 2019
Accepted 3rd September 2019

DOI: 10.1039/c9ra06251c

rsc.li/rsc-advances

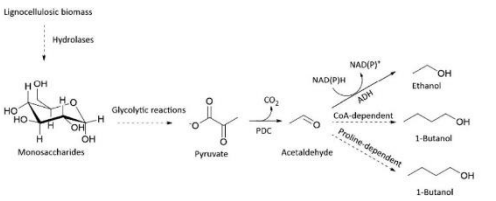
Pyruvate decarboxylase (PDC) is a key enzyme for the production of ethanol at high temperatures and for cell-free butanol synthesis. Thermostable, organic solvent stable PDC was evolved from bacterial PDCs. The new variant shows >1500-fold-improved half-life at 75 °C and >5000-fold-increased half-life in the presence of 9 vol% butanol at 50 °C.

For many years, bioethanol has been used widely as an additive in gasoline or as a fuel by itself.[1] To avoid competition with food, production of ethanol from plant-derived starch or sugar (first generation) has shifted to lignocellulosic biomass (second generation). To date, most of the bioethanol produced industrially still utilises yeasts as the main fermentation organism.[1–3] However, in recent years, several efforts have been made to use thermophilic microorganisms as the new framework.[1,3] The main motivating forces behind this are the volatility of ethanol at higher temperatures, which facilitates easy product removal, reduced risk of contamination and feasibility of consolidated bioprocessing (CBP). In this process, cellulolytic ethanologenic thermophilic microorganisms are applied, thus reducing the energy consumption required for heat exchange between biomass pre-treatment, fermentation and product separation and to avoid cellulase inhibition by accumulating glucose.[3,4]

There are two major pathways for producing ethanol from pyruvate, the central metabolite in glycolysis: non-oxidative decarboxylation *via* pyruvate decarboxylase (PDC) and oxidative decarboxylation *via* pyruvate dehydrogenase (PDH) or pyruvate ferredoxin oxidoreductase (PFOR). Non-oxidative ethanol productions are commonly found in yeasts and other ethanologenic bacteria, while PDH and PFOR routes are described in thermophilic microorganisms.[3,4] PDC mode may offer advantages compared with PDH and PFOR, as it is thermodynamically more favourable and will not lead to any other organic acid by-products.[3–5] In fact, all industrially relevant ethanol producing organisms use PDC. Thermostable PDCs are, therefore, suggested to play an important role in creating homoethanologenic, thermophilic microorganisms.[5,6]

Until recently, there were only limited efforts described in literature for improving the thermostability of PDC. Ancestral sequence reconstruction, supposedly a powerful method of resurrecting ancient thermostable enzymes, did not yield more thermostable PDC.[7] A computational Rosetta-based design resulted in several variants with improved stability, determined by differential interference contrast (DIC) microscopy; unfortunately, the work was not corroborated by kinetic activity and stability data.[8] Another effort to redirect the specificity of a thermostable, acetolactate synthase to perform a PDC-like reaction by conventional directed evolution did not yield variants either, with activity similar to that of bacterial PDC.[9] Thus, creating thermostable PDC variants remains as the main challenge towards second generation ethanol production by thermophilic microorganisms.

Besides acting as an important enzyme for ethanol production, PDC has also been applied in the *in vitro* production of *n*-butanol (Scheme 1).[10,11] Butanol together with other longer-chain alcohols, such as 1-propanol, isobutanol and



Scheme 1 Simplified pathway to produce ethanol and 1-butanol from lignocellulosic biomass *via* pyruvate decarboxylase (PDC). There are two major pathways for producing 1-butanol *via* pyruvate: CoA-dependent or proline-dependent condensation. ADH is an alcohol dehydrogenase.

[a]*Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, Schulgasse 16, 94315 Straubing, Germany. E-mail: sieber@tum.de*
[b]*Catalytic Research Center, Technical University of Munich, Ernst-Otto-Fischer-Straße 1, 85748 Garching, Germany*
[c]*Straubing Branch BioCat Fraunhofer IGB, Schulgasse 11a, 94315 Straubing, Germany*
[d]*School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Copper Road, St. Lucia 4072, Australia*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9ra06251c
‡ Current address: Pharma Stulln GmbH, Werksstraße 3, 92551 Stulln.

isopentanol, is regarded as the next-generation biofuel, due to its closer resemblance to traditional gasoline.[12] Thus, there is an increasing demand to produce longer-chain alcohols sustainably.[12] Industrially, butanol is still produced from petroleum-derived propene utilizing Co-catalyst, H_2, and CO *via* oxo synthesis and Reppe process.[13] Albeit its efficiency, production of butanol from petroleum is not sustainable.

In a more sustainable way, butanol can be produced *via* an acetone–butanol–ethanol (ABE) process, utilising *Clostridium acetobutylicum*.[12,14] Recent work on metabolic engineering has demonstrated the possibility of producing butanol in other microorganisms.[15,16] However, one of the major challenges in butanol fermentation is still the toxicity of butanol to many microorganisms at relatively low concentrations (<2 vol%).[17,18] This effect, however, is less pronounced for the enzymes catalysing butanol production *in vitro*.[10,19] Many enzymes can already tolerate higher organic solvent concentrations than the corresponding microorganisms.[20] Furthermore, engineering enzyme stability towards organic solvents is less challenging than evolving microorganisms to withstand the same concentration of organic solvents.[17,21] Therefore, *in vitro* butanol production emerges as a very promising alternative to traditional butanol fermentation. Thus, in this work, we focused on the engineering of PDC towards higher thermal stability for prospective ethanol production in thermophilic microorganisms and improved stability in butanol for applications in *in vitro*, artificial butanol synthesis.

Prior to engineering a PDC, we characterised two new yeast PDCs from *Candida glabrata* (*Cg*PDC) and *Zygosaccharomyces rouxii* (*Zr*PDC), as well as other PDCs reported in the literature, to determine the most suitable template. As presented in Table 1, the two new PDCs from the yeast are not thermostable and showed typical substrate cooperativity, as described for other yeast PDCs.[22] Characterisation of bacterial PDCs showed similar findings to the previous report, with PDC from *Acetobacter pasteurianus* (*Ap*PDC) demonstrating the highest kinetic stability ($T_{50}{}^{1\ h}$) and melting temperature (T_m).[22] $T_{50}{}^{1\ h}$ is defined as the temperature at which 50% of the enzymes remained active after 1 h incubation, while T_m is defined as the temperature at which 50% of the enzymes are in an unfolded state.[23] In general, bacterial PDCs showed higher activity at 25 °C and overall stability ($T_{50}{}^{1\ h}$ and T_m) than newly characterised *Cg*PDC and *Zr*PDC.

Another PDC that was reported to have improved stability, called 5TMA (Rosetta-designed *Zm*PDC), was also characterised in this study to fill the gap remaining due to the lack of data on activity and kinetic stability in the original study.[8] It was claimed that 5TMA did not show any loss of molar ellipticity, measured by circular dichroism and that it was stable at 60 °C, observed by DIC microscopy. However, our kinetic stability studies based on activity suggested otherwise. We observed that 5TMA was less stable than wild-type (WT) *Zm*PDC, the parental PDC, in respect to $T_{50}{}^{1\ h}$ and T_m (Table 1). We could not explain this major discrepancy, except that we used a buffer of pH 6.5, while the previous work used pH 7.5.[8] It is worth mentioning that during our $T_{50}{}^{1\ h}$ experiment, we observed that 5TMA did not show any aggregation in contrast to other WT PDCs. This could explain why the authors did not observe a phase transition in DIC of 5TMA, but did for WT *Zm*PDC.

Our previous work on evolving thermostable, branched-chain α-keto acid decarboxylase from *Lactococcus lactis* (*Ll*KdcA) showed that improved thermostability was also translated to increased stability against isobutanol.[24] The direct relationship between thermostability and organic solvent stability has also been described in other enzyme classes.[21,25,26] Therefore, to save time and expense in finding PDC variants with increased thermostability and butanol-stability, we developed a simple screening platform to screen only for improved thermostability. For library development, staggered extensions process (StEP) was utilised.[27] StEP has been shown in previous work to be a robust and relatively easy approach for increasing thermostability.[27] Three PDCs from *Zymomonas mobilis* (*Zm*PDC), *Zymobacter palmae* (*Zp*PDC) and *Ap*PDC with relatively similar stability and activity, were chosen as parental templates.

After screening about 800 colonies, we obtained several variants that showed improved thermostability in comparison with our control (WT *Ap*PDC). The most thermostable variant (renamed PDC-Var. 1) was selected, and the mutations were determined by DNA sequencing. PDC-Var. 1 showed the highest sequence similarity to *Ap*PDC (98%), bearing 13 substitutions (Fig. S2 and S3†). In comparison with *Ap*PDC, the most thermostable PDC reported to date, PDC-Var. 1 exhibited increases of 8.2 °C and 7.0 °C in $T_{50}{}^{1\ h}$ and T_m, respectively. Surprisingly, PDC-Var. 1 also showed improved activity in comparison with WT *Ap*PDC, without significant change in K_m (Table 1). This indicated that increased thermostability is not always accompanied by a decrease in activity.[28] We later used this new variant as a new template to further increase its thermostability.

In a previous work with *Ll*KdcA, we successfully improved its thermal and isobutanol-stability by introducing seven amino acid substitutions (Var. 7M.D).[24] As *Ap*PDC and *Ll*KdcA belong to the same α-keto acid decarboxylase family, we intended to

Table 1 Summary of kinetic characterisations and thermostability of WT PDCs and variants

Enzyme	V_{max} (U mg^{-1})	K_m (mM)	$T_{50}{}^{1\ h}{}^{a}$ (°C)	$T_m{}^{b}$ (°C)
*Cg*PDC	22.2 ± 0.4	9.1 ± 0.2 h: 2.02[c]	49.6	55.5
*Zr*PDC	16.7 ± 0.4	11.6 ± 0.4 h: 2.39[c]	42.9	50.5
*Ap*PDC	80.3 ± 0.6	1.9 ± 0.1	64.9	70.0
*Zm*PDC	92.5 ± 0.7	1.3 ± 0.1	62.4	66.5
*Zp*PDC	79.7 ± 0.4	0.8 ± 0.1	61.2	65.0
5NPU[7]	20.7 ± 0.8	4.8 ± 0.4 h: 1.28[c]	51.9	55.5
5TMA[8]	87.0 ± 0.7	1.8 ± 0.1	54.7	59.5
PDC-Var. 1	91.7 ± 0.7	2.1 ± 0.1	73.1	77.0
PDC-Var. 2	71.6 ± 0.7	1.5 ± 0.1	78.5	82.0

[a] $T_{50}{}^{1\ h}$ is defined as the temperature at which the enzyme gives 50% remaining activity after 1 h incubation (Fig. S1).[23] [b] Melting temperature was determined using Thermofluor assay (Fig. S1). [c] The Hill coefficient was determined according to the equation: $V = \frac{V_{max}[S]^h}{[K_m]^h + S^h}$. *Cg* is *Candida glabrata*, *Zr* is *Zygosaccharomyces rouxii*, *Ap* is *Acetobacter pasteurianus*, *Zm* is *Zymomonas mobilis* and *Zp* is *Zymobacter palmae*.

transfer the seven beneficial substitutions of Var. 7M.D to our new PDC-Var. 1. A similar screening procedure was applied. From the seven libraries, only one variant showed improved stability. We named the new variant PDC-Var. 2, which displays a change at position 441 of PDC-Var. 1 from isoleucine to valine (Fig. S3†).[24] PDC-Var. 2 exhibited $T_{50}{}^{1\ h}$ and T_m of 78.5 °C and 82.0 °C, respectively, meaning an improvement of 5.4 °C and 5.0 °C compared with PDC-Var. 1 or an increase of 13.6 °C and 12.0 °C compared with WT *Ap*PDC.

To check the stability of PDC-Var. 2 in comparison with WT *Ap*PDC for ethanol fermentation applications *via* thermophilic microorganisms, both enzymes were incubated at 65 °C, 70 °C and 75 °C (Fig. 1). At 65 °C, WT *Ap*PDC showed a first-order deactivation kinetics, with a half-life of 57 min, while PDC-Var. 2 showed a half-life of 18 h. This improvement represented an almost 19-fold increase. At higher temperatures, the stability effect of PDC-Var. 2 became even more apparent. At 70 °C, WT *Ap*PDC completely lost its activity after 10 min of incubation, while PDC-Var. 2 retained 25% of initial activity after 24 h incubation. At 75 °C, WT *Ap*PDC was completely deactivated after 2 min, while PDC-Var. 2 retained more than 20% of initial activity after 12 h incubation. Based on a first-order kinetic deactivation, WT *Ap*PDC exhibited half-lives of 1.2 min and 10.8 s at 70 °C and 75 °C, respectively. For PDC-Var. 2, the half-lives at 70 °C and 75 °C are 10.7 h and 7.3 h, respectively. As such, the improved stability of PDC-Var. 2 corresponded to a 522-fold and a 2438-fold increased half-life at 70 °C and 75 °C (Fig. 1), respectively. PDC-Var. 2 also demonstrated improved stability towards ethanol by tolerating ethanol up to 40 vol% at 50 °C in comparison with WT *Ap*PDC that showed stability up to only 22 vol% (Fig. S4†). We stress that this new improved characteristic is important for ethanol production *in vitro*.[19]

Biotechnological production of butanol is challenging, mainly due to its high toxicity for microbial cells. Until now, no microorganism has been found in nature that can grow at butanol concentrations higher than 4 vol%.[29] Recently, production schemes were proposed that utilise the solubility limit of butanol in water (9 vol% at 50 °C).[10,11] Using solvent-stable enzymes and hands-on *in vitro* enzyme cascade, cell-free butanol production within a two-phase water/butanol system can be envisioned. Some enzymes from thermophilic microorganisms can already tolerate higher alcohol concentrations, such as aldolase from *Saccharolobus solfataricus* (previously known as *Sulfolobus solfataricus*), which was shown to retain full activity in a two-phase system.[19]

At 50 °C, the most thermostable WT PDC, *Ap*PDC, showed stability up to only 5.6 vol% butanol after incubation for 1 h (Fig. 2A). In 7 vol% butanol, WT *Ap*PDC was completely deactivated. As expected, PDC-Var. 2 demonstrated exceptional stability, by retaining more than 70% of its initial activity in 9 vol% butanol after 1 h incubation at 50 °C (Fig. 2A). Further kinetic stability studies revealed that WT *Ap*PDC was completely deactivated after 30 s incubation at 9 vol% butanol at 50 °C (Fig. 2B). PDC-Var. 2, however, maintained more than 40% of its initial activity after 24 h incubation and 15% of its initial activity after 48 h incubation (Fig. 2B). This corresponds to a half-life of 15 h and thus an over 5000-fold increase in half-life in comparison with WT *Ap*PDC.

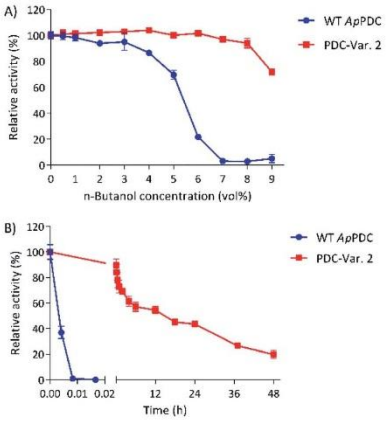


Fig. 2 Stability of WT *Ap*PDC and PDC-Var. 2 in the presence of butanol. (A) Stability of the WT *Ap*PDC and the variant after incubation in different concentrations of butanol for 1 at 50 °C. Lines are drawn to ease reading. (B) Stability of the WT *Ap*PDC and the variant over time in 9 vol% butanol at 50 °C. Standard deviations are shown from three independent technical replications.

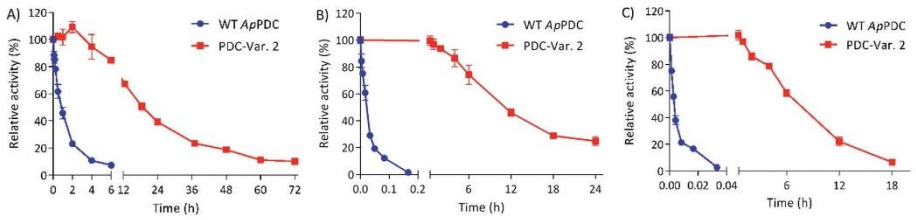


Fig. 1 Kinetic stability of WT *Ap*PDC in comparison with PDC-Var. 2 at 65 °C (A), 70 °C (B) and 75 °C (C). Thermal denaturations of the WT and PDC-Var. 2 followed first order kinetics at any given temperature. Half-lives of WT *Ap*PDC are 57 min, 1.2 min and 10.8 s at 65, 70 and 75 °C, respectively. Half-lives of Variant 2 are 18, 10.7 and 7.3 h at 65, 70 and 75 °C, respectively. Standard deviations are shown from three independent technical replications.

When a higher temperature is desired, *e.g.* 60 °C, as indicated by previous work, WT *Ap*PDC could retain 50% of its activity only up to 2 vol% of butanol at 60 °C, while PDC-Var. 2 started to lose 50% of initial activity in the presence of 8 vol% butanol (Fig. S5†).[10,11] Hence, with this new improved process stability and high activity (Table 1), PDC-Var. 2 rises as a better candidate than any PDCs reported to date and holds promise for the application of *in vitro* butanol production at its solubility limit.

In conclusion, we demonstrated that organic solvent-stable PDC variants could be found only by developing a screening platform focused on improved thermostability. In general, developing a screening platform for improved thermostability is easier than for organic solvent tolerance, as some organic solvents may possess incompatibility issues with enzymatic assays, and general handling is more difficult. We expect our new thermostable PDC variant (PDC-Var. 2) to enhance the feasibility of consolidated bioprocessing approach for ethanol production in cellulolytic thermophilic organisms.[3,5] Improved butanol stability of PDC-Var. 2 should drive forward the already developed *in vitro* butanol production, either *via* CoA dependent or proline-facilitated aldehyde condensation pathways (Scheme 1).[10,11] Additionally, improved thermostability and stability in the presence of butanol of PDC-Var. 2 will facilitate simpler product removal at high temperatures in cell-free butanol synthesis.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the German Research Foundation (DFG) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program. Fruitful discussions with Samed Güner and Ioannis Zachos are greatly appreciated. We thank BMBF funding (Grant No. 031L0078F) and COST action CM1303 Systems Biocatalysis.

Notes and references

1 D. Barnard, A. Casanueva, M. Tuffin and D. Cowan, *Environ. Technol.*, 2010, **31**, 871–888.
2 S. Yang, X. Liu and Y. Zhang, in *Bioprocessing for Value-Added Products from Renewable Resources: New Technologies and Applications*, ed. S. Yang, Elsevier, Amsterdam, 1st edn, 2007, pp. 73–118.
3 D. G. Olson, R. Sparling and L. R. Lynd, *Curr. Opin. Biotechnol.*, 2015, **33**, 130–141.
4 M. P. Taylor, K. L. Eley, S. Martin, M. I. Tuffin, S. G. Burton and D. A. Cowan, *Trends Biotechnol.*, 2009, **27**, 398–405.
5 L. Tian, S. J. Perot, S. Hon, J. Zhou, X. Liang, J. T. Bouvier, A. M. Guss, D. G. Olson and L. R. Lynd, *Microb. Cell Fact.*, 2017, **16**, 171.
6 A. H. Thompson, D. J. Studholme, E. M. Green and D. J. Leak, *Biotechnol. Lett.*, 2008, **30**, 1359–1365.
7 L. Buddrus, E. S. V. Andrews, D. J. Leak, M. J. Danson, V. L. Arcus and S. J. Crennell, *Acta Crystallogr., Sect. F: Struct. Biol. Commun.*, 2018, **74**, 179–186.
8 D. W. Sammond, N. Kastelowitz, B. S. Donohoe, M. Alahuhta, V. V. Lunin, D. Chung, N. S. Sarai, H. Yin, A. Mittal, M. E. Himmel, A. M. Guss and Y. J. Bomble, *Biotechnol. Biofuels*, 2018, **11**, 1–13.
9 M. Cheng, H. Yoshiyasu, K. Okano, H. Ohtake and K. Honda, *PLoS One*, 2016, 1–13.
10 S. Reiße, M. Haack, D. Garbe, B. Sommer, F. Steffler, J. Carsten, F. Bohnen, V. Sieber and T. Brück, *Front. Bioeng. Biotechnol.*, 2016, **4**, 1–10.
11 B. Krutsakorn, K. Honda, X. Ye, T. Imagawa, X. Bei, K. Okano and H. Ohtake, *Metab. Eng.*, 2013, **20**, 84–91.
12 M. R. Connor and J. C. Liao, *Curr. Opin. Biotechnol.*, 2009, **20**, 307–315.
13 M. Uyttebroek, W. Van Hecke and K. Vanbroekhoven, *Catal. Today*, 2013, **239**, 7–10.
14 K. Karimi, M. Tabatabaei, I. S. Horváth and R. Kumar, *Biofuel Res. J.*, 2015, **8**, 301–308.
15 E. I. Lan and J. C. Liao, *Metab. Eng.*, 2011, **13**, 353–363.
16 S. Atsumi, A. F. Cann, M. R. Connor, C. R. Shen, K. M. Smith, M. P. Brynildsen, K. J. Y. Chou, T. Hanai and J. C. L. Ã, *Metab. Eng.*, 2008, **10**, 305–311.
17 L. H. Reyes, M. P. Almario, J. Winkler, M. M. Orozco and K. C. Kao, *Metab. Eng.*, 2012, **14**, 579–590.
18 K. A. Zingaro and E. Terry, *Metab. Eng.*, 2013, **15**, 196–205.
19 J. K. Guterl, D. Garbe, J. Carsten, F. Steffler, B. Sommer, S. Reiße, A. Philipp, M. Haack, B. Rühmann, A. Koltermann, U. Kettling, T. Brück and V. Sieber, *ChemSusChem*, 2012, **5**, 2165–2172.
20 S. Torres, M. A. Martínez, A. Pandey and G. R. Castro, *Bioresour. Technol.*, 2009, **100**, 896–902.
21 T. Koudelakova, R. Chaloupkova, J. Brezovsky, Z. Prokop, E. Sebestova, M. Hesseler, M. Khabiri, M. Plevaka, D. Kulik, I. K. Smatanova, P. Rezacova, R. Ettrich, U. T. Bornscheuer and J. Damborsky, *Angew. Chem., Int. Ed.*, 2013, **52**, 1959–1963.
22 D. Gocke, T. Graf, H. Brosi, I. Frindi-Wosch, L. Walter, M. Müller and M. Pohl, *J. Mol. Catal. B: Enzym.*, 2009, **61**, 30–35.
23 A. S. Bommarius and M. F. Paye, *Chem. Soc. Rev.*, 2013, **42**, 6534–6565.
24 S. Sutiono, J. Carsten and V. Sieber, *ChemSusChem*, 2018, **11**, 3335–3344.
25 E. Vazquez-Figueroa, V. Yeh, J. M. Broering, J. F. Chaparro-Riggers and A. S. Bommarius, *Protein Eng., Des. Sel.*, 2008, **21**, 673–680.
26 M. T. Reetz, P. Soni and L. Ferna, *Chem. Commun.*, 2010, **46**, 8657–8658.
27 H. Zhao, L. Giver, Z. Shao, J. A. Affholter and F. H. Arnold, *Nat. Biotechnol.*, 1998, **16**, 258–261.
28 B. P. L. Wintrode and F. H. Arnold, in *Advances in Protein Chemistry*, ed. F. H. Arnold, Academic Press, 1st edn, 2001, vol. 55, pp. 161–225.
29 M. Kanno, T. Katayama, H. Tamaki, Y. Mitani, X. Y. Meng, T. Hori, T. Narihiro, N. Morita, T. Hoshino, I. Yumoto, N. Kimura, S. Hanada and Y. Kamagata, *Appl. Environ. Microbiol.*, 2013, **79**, 6998–7005.

Supporting Information

Table of contents

1. Materials and methods

Cloning, expression and purification

Acetobacter pasteurianus pyruvate decarboxylase (*Ap*PDC), *Zymomonas mobilis* PDC (*Zm*PDC), *Zymobacter palmae* PDC (*Zp*PDC), PDC inferred from ASR (PDB: 5NPU) and Rosetta-based *Zm*PDC mutant (PDB: 5TMA) are cloned into pET24 using *Nde*I and *Xho*I bearing a C-terminal hexahistidine tag. Two, new PDCs from *Candida glabrata* (*Cg*) and *Zygosaccharmoyces rouxii* (*Zr*) are cloned into pET24 using the same procedures. All PDCs, except *Zm*PDC, are ordered codon-optimised for *Escherichia coli* and purchased from either GeneArt (Germany) or ATG Biosynthesis (Germany).

After cloning all PDCs into pET24, chemically competent *E. coli* BL21 (DE3) cells were transformed with plasmids encoding different PDCs, and the cells were plated on LB agar containing Kanamycin 50 µg/ml. The plates were incubated overnight at 37 °C. Pre-cultures were prepared by inoculating a single colony from the plates into 10-ml LB containing Kanamycin 50 µg/ml and shaken overnight at 150 rpm, 37 °C. On the following day, the 10 ml pre-cultures were transferred to 250 ml autoinduction media containing Kanamycin 100 µg/ml in 1 L baffled flasks.[1] The cultures were incubated at 37 °C, 120 rpm, for 3 h before the temperature was adjusted to room temperature (25 °C) for overnight protein expression. Protein purifications followed the same procedures as described previously, and 50 mM phosphate buffer, pH 6.5, was used as the final buffer.[2] Each purified protein was flash frozen in liquid nitrogen and stored at −80 °C before further characterisations.

Kinetic characterisation

For kinetic characterisation, a standard coupled assay, with yeast alcohol dehydrogenase (Sigma), was used. The reaction solution contained different sodium pyruvate concentrations (0–50 mM), 50 mM phosphate buffer, pH 6.5, 1 U/ml yeast ADH (Sigma), 1 mM NADH, 0.1 mM thiamine diphosphate (TDP) and 0.5 mM $MgSO_4$. 180 µl of reaction solution was transferred to a 96-well, F-bottom plate (Greiner, Germany) that had been pre-filled with 20 µl of the enzyme solution to start the reaction. The reaction was monitored at 340 nm, 25 °C using Multiskan (Thermo-scientific). Specific enzyme activity (U/mg) was defined as the amount of enzyme required for the consumption of 1 µmol NADH per minute. Three technical replications were performed. The kinetic data are presented in Table 1.

Kinetic stability

There are two parameters used to measure kinetic stability. The first one is T_{50}^{1h}, defined as the temperature at which 50% of initial activity remained after 1 h incubation. To determine the T_{50}^{1h}, 1 mg/ml of wild-type (WT) enzymes or variants containing 0.1 mM TDP and 5 mM $MgSO_4$ in 50 mM phosphate buffer, pH 6.5, were incubated in a 96 PCR plate at different temperatures for 1 h. After the heat challenge, dilution buffer (50-mM phosphate buffer, pH 6.5, 0.1 mM TDP and 5 mM $MgSO_4$) was added. Twenty microliters of the mixture was transferred to a new 96 MTP. 180 µl of substrate solution (50 mM phosphate buffer, pH 6.5, 0.1 mM TDP, 5 mM $MgSO_4$ and 25 mM sodium pyruvate) was added to determine residual activity by direct measurement of pyruvate consumption at 315 nm, at 25°C. One hundred percent activity is defined as the activity of the enzyme after incubation at 25°C for 1 h. Three technical replications were conducted. The T_{50}^{1h} data are presented in Table 1.

The second parameter is the half-life of the enzyme ($t_{1/2}$). For this parameter, 1 mg/ml of WT *Ap*PDC and Variant 2, containing 0.1 mM TDP and 5 mM $MgSO_4$ in 50 mM phosphate buffer, pH 6.5, were transferred to several 8-strip, PCR cups. The PCR cups were incubated at 65, 70 and 75 °C. At different, each PCR cup was removed from the PCR, and dilution buffer was

added. Residual activity was measured by direct assay at 315 nm. Half-life is defined as the time at which the enzyme retained 50% of its initial activity. The temperature-dependent, deactivation data are presented in Figure 1.

Thermodynamic stability by Thermofluor

Melting temperatures (T_m) were measured, as described previously.[2] T_m data are presented in Table 1

Stability in the presence of butanol

There are two parameters evaluated for the stability of WT *Ap*PDC and Variant 2. The first is stability in the presence of 9 vol% butanol at 50 °C. At this concentration and temperature, butanol will start to form two phases with water. Both enzymes were diluted (end concentration 1 mg/ml) in 50 mM phosphate buffer, pH 6.5, containing 9 vol% butanol, 0.1 mM TDP and 5 mM $MgSO_4$). The solutions were then aliquoted 20 µl in 8-strip, PCR cups and incubated in a PCR machine. At different intervals, each PCR cup was removed from the PCR, and dilution buffer was added. Residual activity was measured by direct assay at 315 nm, as described above. Half-life is defined as the time at which the enzyme retained 50% of its initial activity. Three independent technical replications were performed. The butanol and temperature-dependent deactivations are presented in Figure 2A.

In the second parameter, different concentrations of butanol (0–9 vol%) were added to the enzyme solution (end concentration 1 mg/ml) in 50 mM phosphate buffer, pH 6.5, 0.1 mM TDP and 5 mM $MgSO_4$. The 20 µl aliquots were transferred to a 96 PCR plate (Brand, Germany) and incubated at 60 °C for 1h. After incubation, dilution buffer was added, and 20 µl of the solution was transferred to a 96 MTP. The remaining activity was measured by direct assay after the addition of 180 µl substrate solution, as described above. Three independent technical replications were performed. The butanol stabilities of WT *Ap*PDC and Variant 2 are presented in Figure 2B.

Stability in the presence of ethanol

The phosphate buffer of WT PDCs or variants was exchanged for 50 mM MOPS, pH 6.5, using PD-10 desalting column (GE Healthcare, Germany) prior to the stability experiments because phosphate buffer precipitates in the presence of high ethanol concentrations. Different volumes of anhydrous ethanol were added to each enzyme solution, in the presence of 0.1 mM TDP and 5 mM $MgSO_4$, to make a final enzyme concentration of 1 mg/mL with ethanol concentrations ranging from 0–50 vol%. Twenty microliters of the mixture was transferred to a 96 PCR plate and incubated at 50°C for 1 h. After incubation, dilution buffer (50 mM MOPS, pH 6.5, 0.1 mM TDP and 5 mM $MgSO_4$) was added to the PCR plate, and 20 µl of the solution was transferred to a 96 MTP. The remaining activity was measured by direct assay after the addition of 180 µl substrate solution (50 mM MOPS, pH 6.5, 0.1 mM TDP, 5 mM $MgSO_4$ and 25 mM sodium pyruvate).

Library development (StEP)

The three most stable PDC, *Ap*PDC, *Zm*PDC and *Zp*PDC, were used as the parental templates for the staggered extension process (StEP). Prior to the StEP, all three PDCs were re-cloned into a modified pBAD plasmid with *Xba*I and *Xho*I. Standard StEP protocol was used, with slight modification.[3] In brief, 10 ng/µl of each plasmid was mixed with dNTPs 2 mM, GoTaq polymerase 1 U, GoTaq green buffer 1x, forward primer 15 mM, reverse primer 15 mM and double-distilled water, to make a 100 µl PCR mix. The mixture was divided into 8 x 12.5 µl aliquots and transferred to 8-strip, PCR cups. The PCR programme used was as follows: 5 min initial denaturation at 95 °C, 30 s denaturation at 94 °C, 7 s annealing at 50 °C, 5 s elongation

at 72 °C and repetition of the denaturation step to elongation step 130 times before holding indefinitely at 16 °C. The PCR products were then separated using 1% agarose gel (Serva, Germany). A strong band, with a size of about 1800 bp, was observed. The DNA band was excised and purified using a gel purification kit (Macherey Nagel). Purified DNA was digested with *Xba*I and *Xho*I and cloned into pET24. *E. coli* BL21 (DE3) cells were transformed, with the pET24 harbouring the library and plated on LB agar containing Kanamycin 50 µg/ml. The plates were incubated overnight at 37°C.

High-throughput screening

Single colonies from the library plates were inoculated into 96 deep-well plates (DWP) containing 800 µl autoinduction media with Kanamycin 100 µg/ml. *E. coli* BL21 (DE3) cells harbouring *Ap*PDC were used as a positive control and inoculated randomly into 8 wells in each of the 96 DWPs, because *Ap*PDC has the highest thermostability compared with *Zm*PDC and *Zp*PDC (Table 1). The 96 DWPs were shaken in a plate shaker at 1000 rpm, 30°C, and incubated overnight. On the following day, 50 µl of the culture was transferred to a 384 MTP (Brand, Germany) that had been pre-filled with glycerol 50% using a Tecan, liquid-handling station for backup. The plates were kept at −80°C. Another 200 µl of the culture was transferred to a 96 U-bottom plate (Sarstedt, Germany), and the plate was centrifuged at 4000 xg for 15 min to remove the supernatant. Two hundred microliters of lysis solution (50 mM phosphate buffer, pH 6.5, 0.1 mM TDP, 5 mM $MgSO_4$, 1 mg/ml lysozyme, 0.5 mg/ml Polymyxin B sulfate and 10 µg/ml DNAse) was added, and the plates were shaken at 1000 rpm, 37°C for 1 h. After the lysis step, a heat challenge was performed by incubating the plates in a water bath, at 75°C, for 30 min. Several colonies showing residual activity higher than three times the standard deviation, were subjected to a second screening. About 800 colonies were selected in the first screening.

In the second screening, the colonies that appeared to have higher stability in the first screening were inoculated into LB liquid containing Kanamycin 50 µg/ml from the backup plates for plasmid purification. *E. coli* Bl21 (DE3) cells were transformed with the plasmids and grown on LB agar containing 50 µg/ml Kanamycin. Eight or more colonies were selected from each variant, and all the screening steps described above were repeated. Plasmid from the variants that retained stability was sent for sequencing (Eurofins, Germany).

Site saturation mutagenesis with QuikChange

For QuikChange, the same protocol, as described previously, was followed.[2] pET24a containing gene encoding Variant 1 was used for the templates. The Primers used are listed in the Section 7 of Supporting Information. The same first and second screening procedures described above were employed. *E. coli* BL21 (DE3) cells harbouring Variant 1 plasmid were used as a positive control, and the heat challenge was undertaken at 80°C for 30 min in the water bath.

2. Kinetic and thermodynamic stability

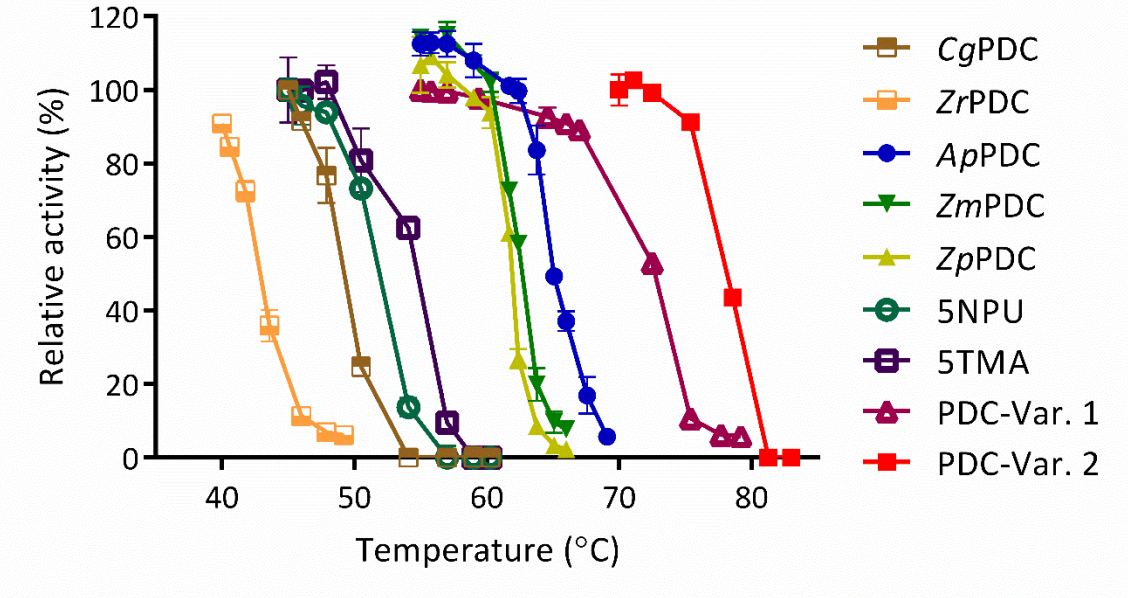


Figure S1A. Stability of WT PDCs and variants after 1 h incubation at different temperatures. 100% activity is defined as the activity after 1 h incubation at 25 °C. *Cg* is *Candida glabrata*, *Zr* is *Zygosaccharomyces rouxii*, *Ap* is *Acetobacter pasteurianus*, *Zm* is *Zymomonas mobilis* and *Zp* is *Zymobacter palmae*. 5NPU was obtained by ancestral sequence reconstruction.[4] 5TMA is derived from *Zm*PDC by Rosetta design.[5]

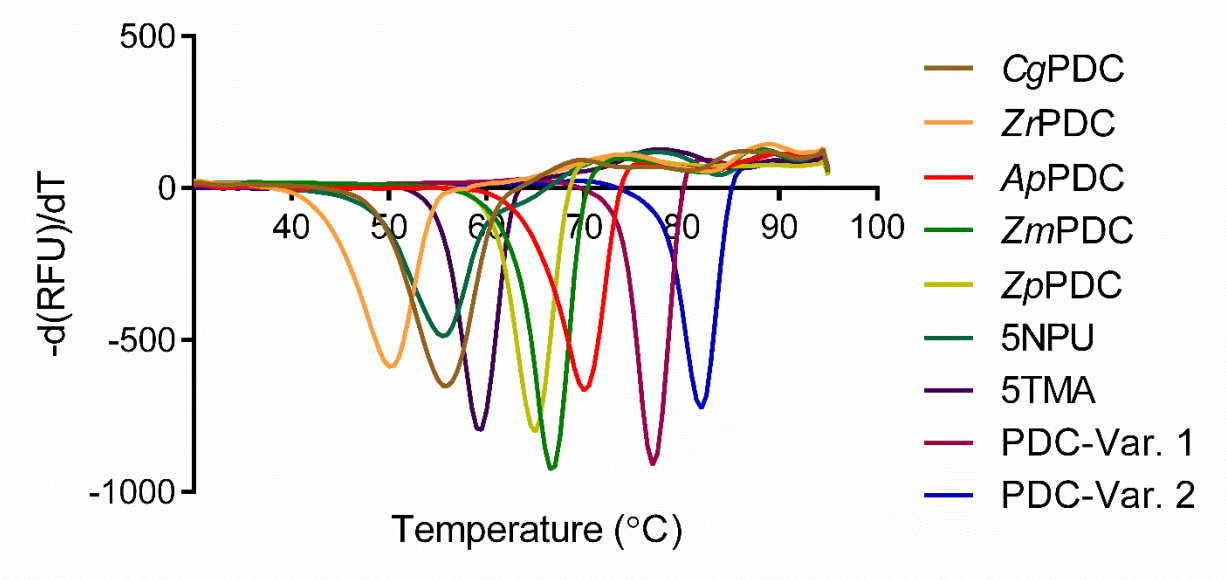


Figure S1B. First-derivative curve of T_m of PDCs measured by Thermofluor assay. *Cg* is *Candida glabrata*, *Zr* is *Zygosaccharomyces rouxii*, *Ap* is *Acetobacter pasteurianus*, *Zm* is *Zymomonas mobilis* and *Zp* is *Zymobacter palmae*. 5NPU was obtained by ancestral sequence reconstruction.[4] 5TMA is derived from *Zm*PDC by Rosetta design.[5]

3. Crossover of bacterial PDCs

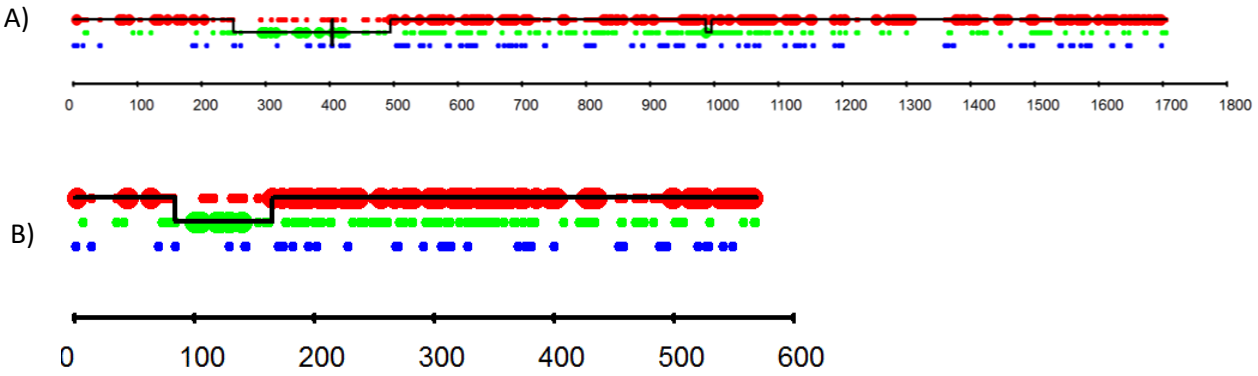


Figure S2. Crossover of Variant 1 in comparison with its parental PDC at nucleotide level (A) and amino acid level (B). Red is *Ap*PDC, green is *Zp*PDC and blue is *Zm*PDC. Large dots represent 100% parental match and small dots represent more than one parental match.

Crossover was calculated using Xover 3.0.[6] The results suggested that crossover on the nucleotide level happened at least six times, while on the amino acid level, it happened only twice. This result indicated that some mutations are silent. From the amino acid level, Variant 1 showed major changes in comparison with *Ap*PDC, starting from amino acid at position 83 up to 143.

4. Comparison between wild-type pyruvate decarboxylase from *Acetobacter pasteurianus* and thermostable variants

Amino acid sequence of Variant 2. The substitutions, in comparison with WT *Ap*PDC, are marked yellow.

MTYTVGMYLAERLVQIGLKHHFAVAGDYNLVLLDQLLLNKDMKQIYCCNELNCGFSAEGYARSNGAA
AAVVTFSVGAISAMNAIGGAYAENLPVILISGSPNTNDYGTGHILHHTIGTTDYNYQLEMVKHVTCAAES
IVSAEEAPAKIDHVIRTALRERKPAYLDIACNIASEPCVRPGPVSSLLSEPEIDHTSLKAAVDATVALLEK
SASPVMLLGSKLRAANALAATETLADKLQCAVTIMAAAKGFFPEDHAGFRGLYWGEVSNPGVQELVE
TSDALLCIAPVFNDYSTVGWSAWPK
GPNVILAEPDRVTVDGRAYDGFTLRAFLQALAEKAPARPASAQKSSVPTCSLTATSDEAGLTNDEIVR
HINALLTSNTTLVAETGDSWFNAMRMTLPRGARVELEMQWGHIGWSVPSAFGNAMGSQDRQHVVM
VGDGSFQVTAQEVAQMVRYELPVIIFLINNRGYVIEIAIHDGPYNYIKNWDYAGLMEVFNAGEGHGLGL
KATTPKELTEAIARAKANTRGPTLIECQIDRTDCTDMLVQWGRKVASTNARKTTLA

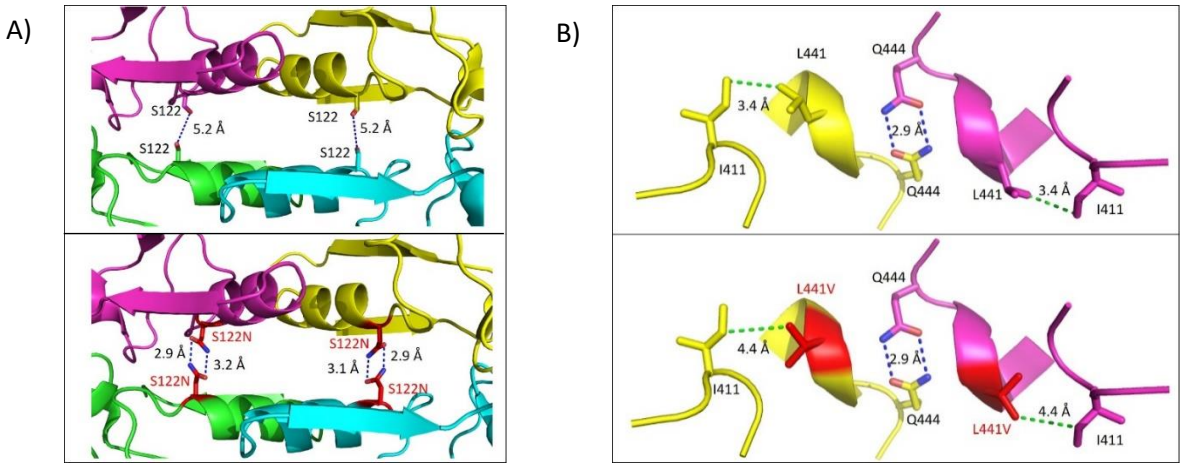


Figure S3 Amino acid substitutions in Variant 2 that led to improved thermostability. The illustration was obtained from the crystal structure of *Ap*PDC (PDB:2vbi). The amino acid substitutions are coloured in red, while the original amino acids are coloured according to the backbone. 2vbi is depicted as a homotetramer and consists of two homodimers (purple-yellow and green-cyan). (**A)** Substitution from serine, at position 122, to asparagine may provide four, new electrostatic interactions between two dimers in PDC tetrameric form. (**B)** Change from leucine to valine, at position 441, may decrease its previous hydrophobic interaction to I411, leading to better electrostatic interaction between Q444 from each monomer.

Variant 1 obtained from StEP contained 13 point substitutions. Until the crystal structure is elucidated, the explanation for the individual substitutions may not be straightforward. However, one notable change is S122N, presented in Figure S3.

5. Ethanol stability

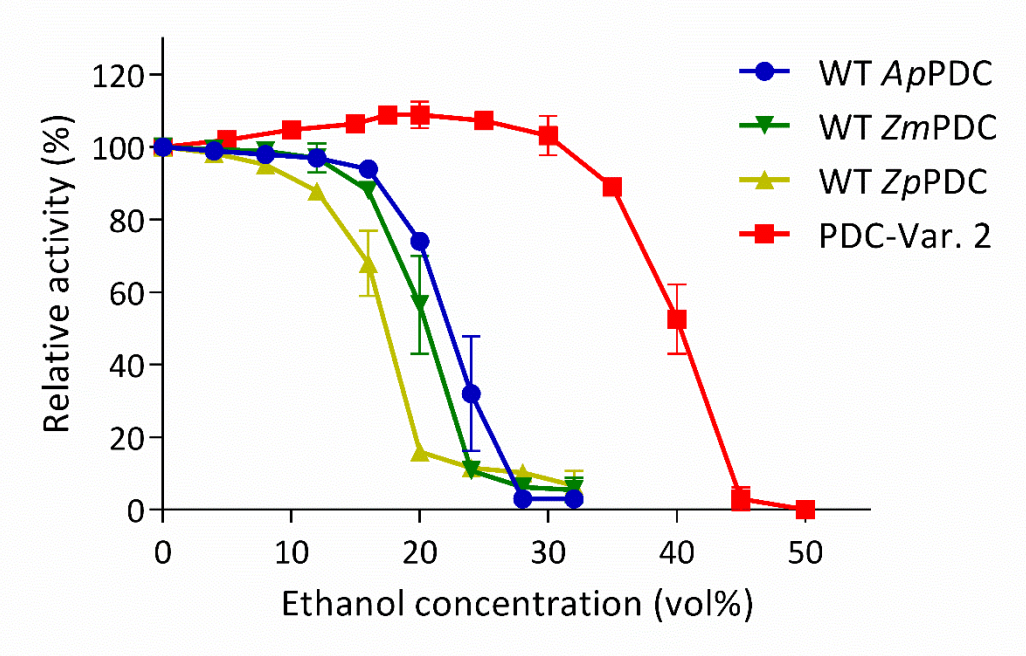


Figure S4. Stability of different PDC variants in the presence of ethanol, at 50°C, after 1 h incubation. 100% activity is defined as the activity after 1 h incubation, at 50 °C, of the enzymes at 0 vol% ethanol. *Ap* is *Acetobacter pasteurianus*, *Zm* is *Zymomonas mobilis* and *Zp* is *Zymobacter palmae*.

6. Butanol stability

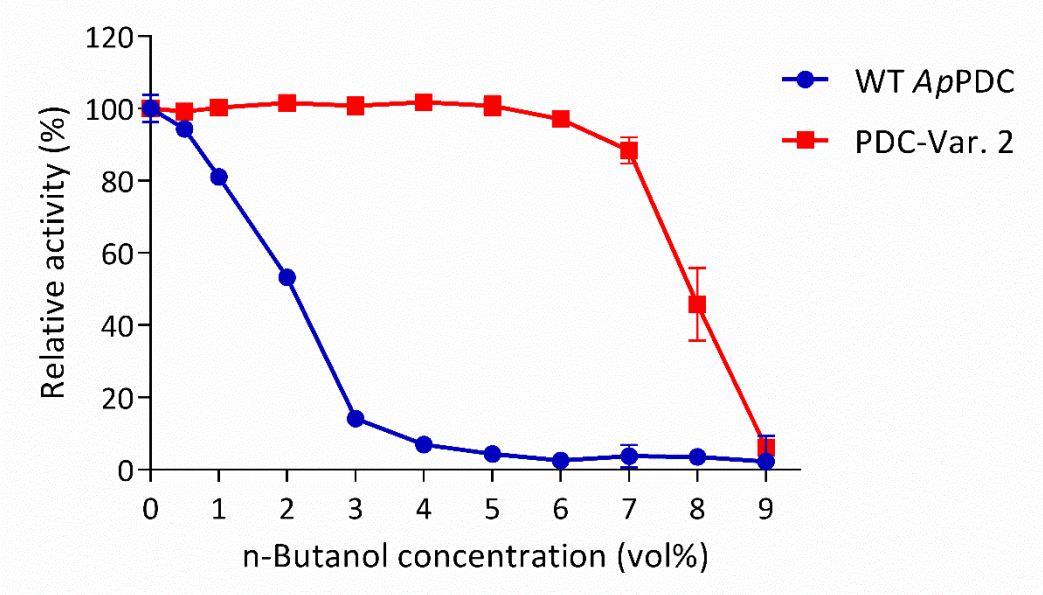


Figure S5. Stability of different WT *Ap*PDC variants in the presence of butanol, at 60°C, after 1 h incubation. 100% activity is defined as the activity after 1 h incubation, at 60 °C, of the enzymes at 0 vol% n-butanol. *Ap* is *Acetobacter pasteurianus.*

7. Primer used in this study

Table S6. List of the primers used. Mutations are written in italics.

Primer	Sequence (5´→ 3´)
pBAD_Fwd	ATGCCATAGCATTTTTATCC
pBAD_Rev	CCTTGAATACACCATGTAGTG
pBADQC_XbaI.F	TACCCG*TCTAGA*GGGCTAACAGGAGGAATTAACC
pBADQC_XbaI.R	TTAGCCC*TCTAGA*CGGGTATGGAGAAACAGTAGAG
PDC-Var.1_D120NNS.F	ACCACC*NNS*TACAATTATCAGCTGGAAATGGTGAAAC
PDC-Var.1_D120NNS.R	AATTGTA*SNN*GGTGGTGCCAATGGTATGATGCAGAATATG
PDC-Var.1_G256NNS.F	ATCATGCA*NNS*TTTCGTGGTCTGTATTGGGGTG
PDC-Var.1_G256NNS.R	ACGAAA*SNN*TGCATGATCTTCCGGAAAAAAACCTTTTGC
PDC-Var.1_G294NNS.F	ACCGTT*NNS*TGGAGCGCATGGCCGAAAGGTC
PDC-Var.1_G294NNS.R	GCTCCA*SNN*AACGGTGCTATAATCATTAAAAACCGG
PDC-Var.1_D310NNS.F	GAACCG*NNS*CGTGTTACCGTTGATGGTCGTG
PDC-Var.1_D310NNS.R	AACACG*SNN*CGGTTCGGCCAGAATAACATTCG
PDC-Var.1_D320NNS.F	GCATAC*NNS*GGTTTTACCCTGCGTGCCTTTCTG
PDC-Var.1_D320NNS.R	AAAACC*SNN*GTATGCACGACCATCAACGGTAACACG
PDC-Var.1_T395NNS.F	CGTATG*NNS*CTGCCTCGTGGTGCACGTGTTG
PDC-Var.1_T395NNS.R	AGGCAG*SNN*CATACGCATTGCATTAAACCAGCTATC
PDC-Var.1_L441NNS.F	TTTCAG*NNS*ACCGCACAAGAGGTTGCACAGATG
PDC-Var.1_L441NNS.R	TGCGGT*SNN*CTGAAAGCTACCATCACCAACCATAAC
PDC-Var.1_I465NNS.F	TTACGTG*NNS*GAAATTGCCATTCATGATGGTCCGTAC
PDC-Var.1_I465NNS.R	AATTTC*SNN*CACGTAACCGCGATTGTTAATCAGGAAG

8. References

1 F. W. Studier, *Protein Expr. Purif.*, 2005, **41**, 207–234.

2 S. Sutiono, J. Carsten and V. Sieber, *ChemSusChem*, 2018, **11**, 3335–3344.

3 H. Zhao, L. Giver, Z. Shao, J. A. Affholter and F. H. Arnold, *Nat. Biotechnol.*, 1998, **16**, 258–261.

4 L. Buddrus, E. S. V. Andrews, D. J. Leak, M. J. Danson, V. L. Arcus and S. J. Crennell, *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, 2018, **74**, 179–186.

5 D. W. Sammond, N. Kastelowitz, B. S. Donohoe, M. Alahuhta, V. V. Lunin, D. Chung, N. S. Sarai, H. Yin, A. Mittal, M. E. Himmel, A. M. Guss and Y. J. Bomble, *Biotechnol. Biofuels*, 2018, **11**, 1–13.

6 W. Huang, W. A. Johnston, M. Boden and E. M. G. Gillam, *Biotechniques*, 2016, **60**, 91–94.

3.3 Enabling the direct enzymatic dehydration of D-glycerate to pyruvate as the key step in synthetic enzyme cascades used in the cell-free production of fine chemicals

In this work, several [Fe-S]-dehydratases were cloned and characterized in regard to their activity toward D-glycerate and D-gluconate. In particular, dehydration of D-glycerate has been a major bottleneck for many cell-free approaches from D-glucose and glycerol to chemicals. *Ss*DHAD was used in all cell-free approaches. Protein sequences of two newly described sugar acid dehydratases (*Rl*AraDHT and *Cc*XylDHT) together with *Ss*DHAD were used to discover a distinct class of [Fe-S]-dehydratases. This class, D-glycerate-converting, showed almost 50-fold higher activity on D-glycerate in comparison to *Ss*DHAD, the most active DHAD prior to this work. However, all DHADs of the D-glycerate-converting class showed very low reactivity to D-gluconate (k_{cat} <0.02 s^{-1}). *Ss*DHAD showed k_{cat} of 0.6 s^{-1} toward D-gluconate. Using the protein sequence of *Cn*DHAD (D-glycerate-converting DHAD from *Cupriavidus necator*) with *Rl*AraDHT and *Cc*XylDHT allowed to uncover another distinct class of dehydratase that are active on D-glycerate and D-gluconate. This new class was named promiscuous-class dehydratases.

Total turnover number (TTN) determination of all enzymes studied demonstrated *Pu*DHT (promiscuous-class dehydratase from *Paralcaligenes ureilyticus*) as the most promising variant by showing a TTN of >90.000 toward D-glycerate (100-fold higher than *Ss*DHAD) and of 10^6 with D-gluconate (>300-fold higher than *Ss*DHAD). Activities of *Pu*DHT toward D-glycerate and D-gluconate were 1.3 s^{-1} and 135 s^{-1} at 50 °C (>100-fold higher than that of *Ss*DHAD). *Pu*DHT was then applied in a biotransformation of glycerol and D-glucose to validate its performance in the context of multiple enzymatic cascade reactions. For biotransformation of D-glucose to ethanol at 50 °C, a higher ethanol titer could be achieved in comparison to when *Ss*DHAD was used. Higher production rate (>10 fold) was also achieved. Coupling *Pu*DHT to an alditol oxidase allowed conversion of glycerol to pyruvate. Due to lower stability of the alditol oxidase, the biotransformation was performed at 25 °C. Application of *Pu*DHT in a cascade from glycerol to pyruvate facilitated >5-fold increased production rate in comparison when *Ss*DHAD was used at 50 °C. 97% theoretical yield of pyruvate from 50 mM glycerol could be achieved in the cascade with *Pu*DHT. The ability of *Pu*DHT to convert crude-glycerol was also verified.

The author and the co-author Prof. Sieber conceived the study. The author designed all the experiments, performed the cell-free cascades, and analyzed the results under the supervision of the co-author Prof. Sieber. Mr. Tristan Rath was helping to pretreat the crude glycerol used in this study. The other co-authors Ms. Teshima and Dr. Beer contributed to content and language of the manuscript. The co-authors Prof. Schenk and Prof. Sieber cowrote and critically reviewed the manuscript.

Enabling the Direct Enzymatic Dehydration of d-Glycerate to Pyruvate as the Key Step in Synthetic Enzyme Cascades Used in the Cell-Free Production of Fine Chemicals

Samuel Sutiono, Mariko Teshima, Dr. Barbara Beer, Prof. Gerhard Schenk, Prof. Volker Sieber

ACS Catalysis

2020

DOI: 10.1021/acscatal.9b05068

ACS Catalysis

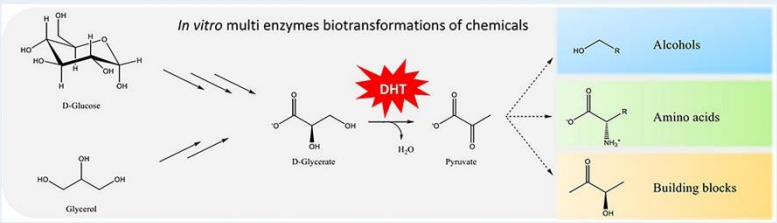
pubs.acs.org/acscatalysis Research Article

Enabling the Direct Enzymatic Dehydration of D-Glycerate to Pyruvate as the Key Step in Synthetic Enzyme Cascades Used in the Cell-Free Production of Fine Chemicals

Samuel Sutiono, Mariko Teshima, Barbara Beer, Gerhard Schenk, and Volker Sieber*

Cite This: *ACS Catal.* 2020, 10, 3110−3118 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: A high degree of dependence on fossil fuels is one of the major problems faced by modern societies. D-Glucose and glycerol have emerged in recent years as prospective replacements for fossil fuels used in the production of high-value chemicals, and cell-free bioproduction routes are expected to play a crucial role in such processes. Recently, several synthetic cascades used for the cell-free biotransformations of D-glucose and glycerol to pyruvate and beyond have been described. However, these were limited by the very slow dehydration step of D-glycerate to pyruvate catalyzed by a dehydratase from *Sulfolobus solfataricus* (*Ss*DHAD), making this step by far the major bottleneck. By combining the vast number of available genomes with a sequence-based discovery approach, we have identified signature sequences leading to the discovery of two distinct classes of dehydratases which exhibit promising activity and total turnover number (TTN) toward D-glycerate. In particular, the dehydratase from *Paralcaligenes ureilyticus* (*Pu*DHT) demonstrated >100-fold higher activity and TTN for D-glycerate in comparison to *Ss*DHAD. In addition, *Pu*DHT showed exceptionally high activity and TTN toward D-gluconate. The replacement of *Ss*DHAD by *Pu*DHT in our model cascade from D-glucose to ethanol enhanced the production rate 10-fold, reaching a 92% theoretical yield at 50 °C. *Pu*DHT was also suitable for the conversion of glycerol to pyruvate at ambient temperature, leading to a >5-fold improvement in production rate in comparison to the system utilizing *Ss*DHAD at 50 °C and attaining 97% of the theoretical yield. *Pu*DHT was also compatible when crude glycerol was used as the substrate, and it no longer caused a bottleneck in the enzymatic cascade.

KEYWORDS: *biocatalyst, dehydration, dehydratases, DHAD, D-glycerate, pyruvate, chemicals*

INTRODUCTION

Since the start of the industrial revolution in the late 18th century, human consumption of fossil fuels has increased by a staggering 1300-fold.[1,2] Given that fossil fuels represent a finite resource, their depletion is inevitable, and their utilization is also a major contributor to greenhouse gas emissions.[2−4] Nevertheless, global demand for fossil fuels is by no means slowing down.[5,6] Consequently, there is an urgent need to develop sustainable and clean alternatives to fossil fuels used to produce value-added chemicals for the food, feed, fiber, and fuel industries.[7,8] D-Glucose and glycerol are ideal renewable resources for the synthesis of such chemicals.

D-Glucose can be derived from lignocellulosic biomass.[9,10] More than 200 billion tons of dried biomass are produced annually.[10] Lignocellulose typically consists of 40−50% cellulose, 25−35% hemicellulose, and 15−20% lignin.[9,10] Acting as the main building block of cellulose, D-Glucose is the most abundant monomer on Earth, and recent studies on the production of glucose from cellulose have demonstrated that this molecule is an ideal starting material for industrial biotransformations.[10,11]

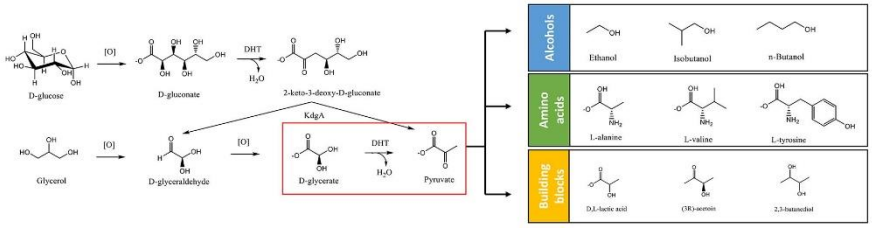
Glycerol is produced as the main byproduct of the synthesis of biodiesel. With an annual production of ∼3.5 million tons, glycerol has shifted in status from that of an alternative product in the late 1990s to that of a waste product in recent years.[12,13] Consequently, upgrading its value has since been viewed as a

Received: November 22, 2019
Revised: February 1, 2020
Published: February 3, 2020

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

ACS Catalysis pubs.acs.org/acscatalysis Research Article

Scheme 1. Cell-Free Production of a Variety of Chemicals from D-Glucose and Glycerol via Pyruvate[a]



[a]DHT is dehydratase, KdgA is 2-keto-3-deoxygluconate aldolase, and [O] is an oxidation reaction. The major bottleneck of the cascades, i.e. the dehydration step of D-glycerate to pyruvate, is highlighted in red. The cascades were adapted from multiple works.[21,22,24−26] Two exemplary cascade reactions performed in this study are presented in Schemes S1 and S2 in the Supporting Information.

key component influencing the economic viability of biodiesel production.[12,13] The transformation of glycerol to useful chemicals will lead to waste reduction and provide a renewable source able to be used as an alternative to fossil fuels.[12]

D-Glucose and glycerol can be transformed to a variety of chemicals (e.g., alcohols, diols, organic acids, amino acids, etc.), via either microbial fermentations or cell-free applications.[14,15] With regard to the production of at least some of these chemicals, cell-free applications have attracted increasing attention because they possess a number of advantages in contrast to microbial fermentations, including a lower degree of complexity and easier optimization, as well as greater tolerance and flexibility with respect to various process parameters (i.e., temperature, pH, organic solvents) than most industrially relevant microorganisms. A higher yield and fewer side products may also be achieved.[16−18]

Numerous cell-free synthetic approaches starting from D-glucose and glycerol have been reported in the past few years.[19] Most of them have been built upon two major biochemical pathways: e.g., the Embden−Meyerhof−Parnas (EMP) and Entner−Doudoroff (ED) pathways. A number of chemicals can be derived from the intermediates of these pathways or from the end product, pyruvate.[19] The cell-free production of chemicals via pyruvate, which mimics the EMP or ED pathways would require at least 10 enzymes. The use of so many enzymes significantly increases the complexity of the reaction design as well as the associated enzyme production costs, thus limiting the merit of such a cell-free approach.[16−18] The discovery of the so-called nonphosphorylative ED (nED) pathway in the late 1980s has reduced the number of required enzymes to 7, including two cofactors (NAD(P)H and ATP).[20] A further modification of this pathway has enabled the elimination of 3 more enzymes and the ATP. As a result, only 4 enzymes and the cofactor NADH are required in this minimized pathway to convert D-glucose to pyruvate.[21] One key step in this synthetic pathway is the use of a dehydratase to directly dehydrate D-glycerate to pyruvate. Furthermore, the coupling of this minimal pathway with alditol oxidase and catalase also facilitates the production of pyruvate from glycerol.[22,23] Among the chemicals that have been produced from these new routes are alcohols, amino acids, and chemical building blocks for the synthesis of additional high-value products (Scheme 1).[21,22,24−26]

The direct enzymatic dehydration of D-glycerate to pyruvate plays a key role in facilitating the design of the minimal cell-free cascades but is also its major bottleneck.[21,22,25,26] This is mainly due to the enzyme that catalyzes this step in the cascades, which is the dihydroxy-acid dehydratase from *Sulfolobus solfataricus*, also known as *Saccharolobus solfataricus* (*Ss*DHAD), exhibiting very low activity (k_{cat} = 0.01 s^{-1}). In addition, *Ss*DHAD has poor process stability in this reaction (total turnover number/TTN = 950).[27] In order to be economically attractive for the production of chemicals, a biocatalyst should have at least an activity (k_{cat}) of 1 s^{-1} and a TTN of 10000 to 100000.[28] The success of the industrial production of chemicals from D-glucose and glycerol via pyruvate therefore depends to a large extent on this dehydratase-catalyzed reaction.

Two novel dehydratases from mesophilic microorganisms, D-xylonate dehydratase from *Caulobacter crescentus* (*Cc*XylDHT) and L-arabinonate dehydratase from *Rhizobium leguminosarum* (*Rl*AraDHT), have been described recently.[29] These two new dehydratases belong to the same family of [Fe-S]-containing dehydratases as *Ss*DHAD. However, in contrast to the latter, *Cc*XylDHT and *Rl*AraDHT have high activity toward C5 and C6 sugar acids (almost 1 order of magnitude higher than that of *Ss*DHAD). These two dehydratases, however, were not tested for their efficiency in dehydrating D-glycerate.[29] Due to the significance of the dehydration of D-glycerate to pyruvate, we thus focused our attention toward the discovery of proficient catalysts that may remove, or at least improve, the bottleneck in the conversion of D-glucose to products. Several dehydratases with vastly improved catalytic properties and process stability relative to *Ss*DHAD were discovered and characterized. In order to illustrate the strong effect of the most proficient of these catalysts in bioproduction processes, we demonstrated its efficiency in greatly enhancing the biotransformation of (i) D-glucose to ethanol at elevated temperatures as well as (ii) pure and crude glycerol to pyruvate at ambient temperature. Considering that this catalyst has not yet been optimized with respect to its catalytic efficiency and stability, it provides an ideal platform to establish highly efficient cell-free biomanufacturing pipelines for the production of a diverse range of products (Scheme 1).

RESULTS AND DISCUSSION

Conservative Approach to Uncover a Distinct Class of D-Glycerate-Converting Dehydratase. Since the time of our initial work on the minimal enzyme cascade to convert D-glycerate to pyruvate, several synthetic cascade reactions have been developed from this synthetic route. However, the dehydratase-catalyzed step has remained as the major bottle-

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

ACS Catalysis pubs.acs.org/acscatalysis Research Article

Table 1. Activity (k_{cat}) as well as Process (TTN) and Thermodynamic (T_m) Stability of Several New Dehydratases Using D-Glycerate and D-Gluconate as Substrates[e]

no.	enzyme[a,b]	activity on D-gly (s^{-1})	activity on D-gluc (s^{-1})	TTN on D-gly (x10^3)	TTN on D-gluc (×10^6)	T_m (°C)
1	*Ss*DHAD[27]	0.01[c]	0.40[c]	0.94 ± 0.03	0.03 ± 0.01	N.D.
2	*Cc*XylDHT	<0.01 ± 0.00[d]	43.75 ± 2.61[29]	N.D.	N.D.	59.5
3	*Rl*AraDHT	<0.01 ± 0.00[d]	7.18 ± 0.67[29]	N.D.	N.D.	62.0
4	*Rl*XylDHT	<0.01 ± 0.00[d]	26.51 ± 0.45[d]	N.D.	N.D.	62.5
5	*Hs*XylDHT	<0.01 ± 0.00[d]	36.85 ± 1.52[d]	N.D.	N.D.	63.5
6	*Hs*AraDHT	0.05 ± 0.00[d]	2.45 ± 0.35[d]	N.D.	N.D.	49.5
7	*Cn*H16DHAD	0.65 ± 0.02	0.01 ± 0.00	21.51 ± 2.22	N.D.	75.5
8	*Cn*N1DHAD	0.73 ± 0.01	0.01 ± 0.00	22.05 ± 0.22	N.D.	77.0
9	*Cm*DHAD	0.75 ± 0.04	0.02 ± 0.00	24.22 ± 1.01	N.D.	74.5
10	*St*DHAD	0.55 ± 0.03	0.01 ± 0.00	82.02 ± 4.71	N.D.	72.0
11	*Ft*DHAD	0.49 ± 0.00	0.01 ± 0.00	94.43 ± 0.73	N.D.	79.0
12	*Tv*DHAD	0.22 ± 0.00	0.01 ± 0.00	62.52 ± 6.14	N.D.	85.5
13	*Hs*DHT	0.58 ± 0.02	85.89 ± 5.16	77.73 ± 3.52	6.26 ± 0.47	71.0
14	*Vp*DHT	0.36 ± 0.01	55.14 ± 1.58	17.92 ± 1.43	3.47 ± 0.03	75.0
15	*Rf*DHT	1.20 ± 0.06	23.49 ± 0.73	37.74 ± 1.92	2.49 ± 0.13	61.5
16	*Nm*DHT	0.66 ± 0.00	36.36 ± 1.26	30.41 ± 0.34	8.44 ± 0.44	72.0
17	*Pu*DHT	1.03 ± 0.03	135.80 ± 4.90	94.52 ± 0.33	10.06 ± 0.59	70.5
18	*Rh*DHT	0.34 ± 0.00	10.95 ± 1.55	18.12 ± 1.51	2.06 ± 0.19	62.5

[a]For abbreviations of the microorganisms, see the text or Table S1 in the Supporting Information. SDS-PAGE of all dehydratases are presented in Figure S1. All activities presented in Table 1 were determined by linear regression of at least three different time points, of which two replicates were performed. All TTNs were obtained by analyzing samples withdrawn after 0, 24, 48, and 96 h from two replicates. Errors in the table represent standard deviation. [b]Enzymes were ordered according to different classes of dehydratases based on sequence alignments. Nos. 1−6 is the pentonate class, 7−12 is the D-glycerate class, and 13−18 is the promiscuous class. [c]Depending on the enzyme preparation and [Fe-S] loading, activities between 0.01 and 0.03 s^{-1} for D-glycerate (D-gly) and between 0.4 and 1.2 s^{-1} for D-gluconate (D-gluc) have been achieved. [d]Activity determined at 37 °C due to the lower thermal stability of these enzymes. [e]All measurements were performed in HEPES 50 mM, pH 7.5 at 50 °C, unless otherwise stated. N.D. denotes not determined.

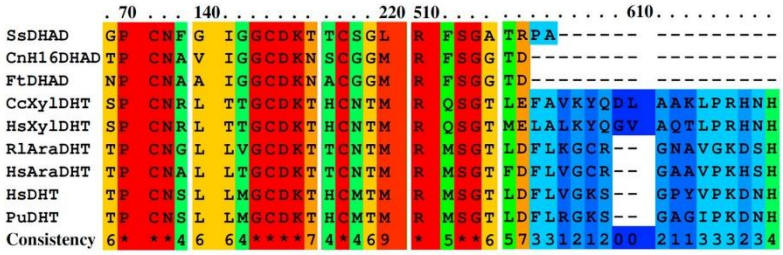


Figure 1. Multiple sequence alignment (MSA) of several dehydratases revealed signature sequence motifs for activity toward D-glycerate and D-gluconate. Three conserved cysteines (positions 71, 144, and 217) are reported to coordinate to the [Fe-S] cluster.[29] Serine at position 512 is proposed as the catalytically active amino acid.[35] The amino acid at the C-terminal of the D-glycerate class dehydratases (*Cn*H16DHAD and *Ft*DHAD), which only exhibited activity toward D-glycerate, is aspartic acid. The C-terminus of the promiscuous class dehydratases (*Hs*DHT and *Pu*DHT), which were active toward both D-glycerate and D-gluconate, is elongated and ends with a histidine. The residue numbering is based on the whole sequence alignment with gaps. For the complete sequence alignment, see Figure S2 in the Supporting Information. MSA was performed by PRALINE.[34]

neck of the cascades.[21,22,24−26] We thus tested two pentonate dehydratases (*Rl*AraDHT and *Cc*XylDHT) for their catalytic efficiency toward D-glycerate.[29] Both of them exhibited minimal activity toward this substrate (k_{cat} = 5 × 10^{-3} s^{-1} for *Rl*AraDHT and 3 × 10^{-3} s^{-1} for *Cc*XylDHT). Nevertheless, this observation suggested that activity toward D-glycerate was not an exclusive property of archaeal DHAD.[27] We thus embarked on a more thorough search for potential D-glycerate dehydratases with enhanced catalytic efficiency.

In one approach, we used the sequences of *Ss*DHAD, *Rl*AraDHT, and *Cc*XylDHT as queries to search for related dehydratases in the sequence database (NCBI and UniProt). Several annotated DHADs from *Cupriavidus necator* H16 (*Cn*H16) were thus selected and recombinantly expressed and purified.[30] Only one of these exhibited significantly higher activity toward D-glycerate in comparison to *Ss*DHAD, *Rl*AraDHT, and *Cc*XylDHT (Table 1). Subsequently, we refined our sequence database search by using *Cn*H16DHAD as the new query. Several new DHADs from *C. necator* N1 (*Cn*N1), *C. metallidurans* (*Cm*), *Schlegelella thermodepolymerans* (*St*), *Fontimonas thermophila* (*Ft*), and *Thermosynecoccocus vulcanus* (*Tv*) were retrieved, cloned, and expressed. Among

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

these new DHADs, the two *Cupriavidus* DHADs have the highest activity toward D-glycerate, followed by *St*DHAD and *Ft*DHAD. Importantly, this new class of dehydratases is almost 50-fold more active toward D-glycerate than the pentonate class dehydratases *Ss*DHAD, *Rl*AraDHT, and *Cc*XylDHT.

We were also interested in assessing the activity of these new dehydratases toward D-gluconate as an essential step for their use in a cascade that aims to convert D-glucose to pyruvate.[21] However, none of them exhibit significant activity toward this substrate (Table 1). Collectively, we thus designated this new class of dehydratases as the "D-glycerate class".

A Sequence-Driven Approach Led to the Discovery of a Novel Class of Dehydratases That Is Active toward both D-Glycerate and D-Gluconate. The ability to directly dehydrate D-glycerate to pyruvate has long been believed to be a reaction that "did not exist" in nature and was an exclusive feature of archaeal dehydratases.[27,31] However, the observation that *Cc*XylDHT and *Rl*AraDHT possess some measurable activity toward D-glycerate (Table 1), together with our discovery of the D-glycerate class of dehydratases, encouraged us to adapt our sequence database search to find promiscuous dehydratases that exhibit catalytic efficiency simultaneously toward D-glycerate and D-gluconate before making any effort to engineer these dehydratases.[32,33] We combined *Cn*H16DHAD (active toward D-glycerate), *Rl*AraDHT, and *Cc*XylDHT (both active toward D-gluconate) in a sequence database search but omitted *Ss*DHAD. *Ss*DHAD is the least reactive of these enzymes, a fact that is likely to be linked to the rapid loss of its [Fe-S] cluster.[27]

Several sequences were thus identified and included in a multiple sequence alignment (MSA) used to rationally identify signature amino acid sequences associated with activity toward D-glycerate and/or D-gluconate (Figure 1).[34] On the basis of the MSA analysis, several common features in *Cn*H16DHAD, *Rl*AraDHT, and *Cc*XylDHT are apparent. All contain three conserved cysteines and a catalytic serine residue. The cysteine residue coordinates to the [Fe-S] cluster in the active site.[29] One notable difference, however, is the C-terminal amino acid. *Rl*AraDHT and *Cc*XylDHT sequences end with histidine, whereas *Cn*H16DHAD ends with aspartic acid. In fact, the C-terminal aspartic acid is conserved in all D-glycerate dehydratases. We thus hypothesized that a C-terminal histidine was a signature fingerprint for dehydratases exhibiting activity toward C5 and C6 sugar acids. In order to test this hypothesis, several new dehydratases from *Rhizobium leguminosarum* (*Rl*) and *Herbaspirillum seropedicae* (*Hs*) with a histidine in the C-terminal position were cloned and characterized.

Two dehydratases from the protein sequence alignment and activity profile—one from *R. leguminosarum* and one from *H. seropedicae*—belonged to the D-xylonate dehydratase family (~75% identity to *Cc*XylDHT). These two dehydratases were thus named *Rl*XylDHT and *Hs*XylDHT, respectively. Two additional dehydratases from *H. seropedicae* also had high protein sequence similarity to *Rl*AraDHT (64% identity) but shared only ~40% identity to *Cc*XylDHT. The activity profile of one of these two enzymes indicated that it belonged to the group of L-arabinonate dehydratases (data not shown) and was thus named *Hs*AraDHT (Table 1). The other enzyme, in contrast, demonstrated a preference for D-gluconate rather than L-arabinonate (data not shown). In order to reflect its more promiscuous substrate selection, we named this enzyme *Hs*DHT (Table 1).

*Rl*XylDHT, *Hs*XylDHT, and *Hs*AraDHT had a low activity toward D-glycerate which was similar to that of *Cc*XylDHT and *Rl*AraDHT (Table 1). However, *Hs*DHT did exhibit noticeable activity toward this substrate (comparable to the D-glycerate class dehydratases), in addition to exhibiting very high activity toward D-gluconate (2-fold higher than that of *Cc*XylDHT). Hence, on the basis of its activity, substrate promiscuity, and protein sequence, *Hs*DHT represented another new class of dehydratases. Using its sequence, we have identified several related dehydratases, which we labeled as "promiscuous class" dehydratases. All of these enzymes had relatively high activity toward D-glycerate and D-gluconate (Table 1). The following section describes our investigation of the suitability of the dehydratases identified in this study as biocatalysts for the conversion of D-glucose and glycerol to fine chemicals.

Process Stability and Activity of the D-Glycerate and Promiscuous-Class Dehydratases. The total turnover number (TTN) is a parameter widely used in chemical industries to determine the robustness and total costs of (bio)catalysts in relevant processes.[28,36] TTN is defined as the amount (in moles) of product formed per mole of biocatalyst used.[28,36] A high TTN is desirable for most industrial processes, with suitable values ranging from 10000 to 100000.[28] The highest TTN (>90000) for the dehydration of D-glycerate was achieved by *Ft*DHAD and *Pu*DHT (*Paralcaligenes ureilyticus* dehydratase), followed by *St*DHAD, *Hs*DHT, and *Tv*DHAD, respectively (Table 1). Surprisingly, DHADs from *Cupriavidus* (*Cn*H16-, *Cn*N1-, and *Cm*DHAD), although having relatively high activities toward D-glycerate and high T_m values, exhibited relatively modest TTNs (~20000). This result is in contrast to our previous studies with α-keto acid decarboxylases, in which we observed a direct relationship between T_m and kinetic stability.[37,38] The melting temperature (T_m) may indicate the general stability (thermodynamic stability) of an enzyme.[36] Thus, in the case of proteins with similar initial activities and T_m values, a lower TTN could be linked to rapid loss of activity over time rather than to thermal denaturation. Since the activity of [Fe-S]-dehydratases is often associated with the stability of their [Fe-S] clusters, the modest TTN of *Cupriavidus* DHAD relative to that of the D-glycerate-converting DHAD may be linked to an unstable [Fe-S] cluster in their active sites. A similar behavior was also observed in the promiscuous-class *Vp*DHT (*Variovorax paradoxus* dehydratase), which exhibited the lowest TTN in our comparison despite having the highest T_m value (Table 1). Furthermore, *Rh*DHT (*Ramlibacter henchirensis* dehydratase) possessed an activity toward and TTN for D-glycerate similar to those of *Vp*DHT, despite having a significantly lower T_m. These combined results thus indicated that the stability of an [Fe-S] cluster in dehydratases is a characteristic feature of individual enzymes rather than that of a specific class of dehydratases.

For the dehydration of D-gluconate, all promiscuous-class dehydratases had TTNs >10^6, i.e., 2 orders of magnitude above the minimum requirement for industrial applications. The highest TTN was achieved by *Pu*DHT (~10 million), followed by *Nm*DHT (*Noviherbaspirillum massilense* dehydratase) and *Hs*DHT (6–8 million). The high TTNs for D-gluconate were a reflection of the high catalytic activity of all promiscuous-class dehydratases toward that substrate (Table 1). The consistently high proficiency toward D-gluconate by the promiscuous-class dehydratases identified their activity toward D-glycerate as the

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

decisive factor used in identifying the most optimal catalysts for use in enzyme cascades. A reasonable biocatalyst should have a minimum activity (k_{cat}) of ~1 s^{-1} toward a substrate of choice.[28] Only *Rf*DHT (*Rhodonobacter fulvus* dehydratase) and *Pu*DHT matched this criterion (Table 1), but *Pu*DHT was a better candidate for process applications due to its 4-fold higher TTN.

While *Pu*DHT appeared to be the most suitable catalyst identified in this study, the D-glycerate class *Ft*DHAD and promiscuous class *Hs*DHT also have properties that may be relevant for their potential in industrial applications. *Ft*DHAD had a very high TTN (~94500) for D-glycerate, while *Hs*DHT has a high TTN for both D-glycerate and D-gluconate while maintaining a T_m slightly higher than that of *Pu*DHT. The three biocatalysts were subjected to a more detailed kinetic characterization to assess their aptitude in a cascade process. Kinetic rate measurements of the three enzymes as a function of the D-glycerate concentration revealed that these enzymes were obeying standard Michaelis–Menten behavior (see Figure S3 in the Supporting Information). *Ft*DHAD had the highest affinity for D-glycerate (lowest K_m) followed by *Pu*DHT and *Hs*DHT. *Pu*DHT has, however, the highest k_{cat} value followed by *Hs*DHT and *Ft*DHAD, which translates into *Pu*DHT being the most efficient biocatalyst (on the basis of the k_{cat}/K_m ratio; Table 2). In combination with its high TTN,

Table 2. Kinetic Characterization of Three Efficient Dehydratases using D-Glycerate as a Substrate[a]

enzyme[b]	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})
*Ft*DHAD	0.61 ± 0.01	5.48 ± 0.29	0.11 ± 0.01
*Hs*DHT	1.01 ± 0.02	14.85 ± 0.81	0.07 ± 0.00
*Pu*DHT	1.40 ± 0.04	7.92 ± 0.91	0.18 ± 0.02

[a]Reactions were performed in HEPES 50 mM, pH 7.5 at 50 °C. [b]Errors in the table represent standard deviation from two replicates. Nonlinear regression of the enzyme activity as a function of substrate concentration is presented in Figure S3

*Pu*DHT thus emerged as the most efficient [Fe-S]-dehydratase known to date. In subsequent parts of this study, we thus focused on assessing the performance of *Pu*DHT in cascade reactions starting from D-glucose and glycerol, respectively.

A Dehydratase Active toward D-Glycerate Is the Key for the Complete Biotransformation of D-Glucose to Chemicals. The dehydration of D-glycerate to pyruvate was first applied for the synthesis of ethanol and isobutanol.[21] Subsequently, several *in vitro* cascades were developed starting from the conversion of D-glucose to pyruvate (Scheme 1). Most notable are the biotransformations of D-glucose to L-alanine and L-lactic acid.[25,26] Common to all these cascades is the fact that the rate-limiting bottleneck reaction is the dehydration of D-glycerate to pyruvate as performed by *Ss*DHAD.[25,26] In this case, we tested whether the replacement of *Ss*DHAD by the newly discovered *Pu*DHT led to a significant improvement in the performance of these cascades. The synthesis of ethanol was chosen as the model cascade because this biotransformation pathway is both a cofactor and is pH balanced (a detailed cascade reaction presented in Scheme S1). In contrast, the production of L-alanine and L-lactate is accompanied by a net decrease in pH.[25,26]

In a preliminary experiment, the amount of *Pu*DHT was varied in the nonoptimized ethanol cascade reaction. A clear relationship between the amount of *Pu*DHT used and ethanol yield was observed after running the biotransformation for 16 h (Figure S4 in the Supporting Information). In addition, the higher the concentration of *Pu*DHT used in the reaction, the lower the concentration of the remaining D-glycerate at the end of the process. This result is in agreement with the dehydration of D-glycerate representing the main bottleneck in the biotransformation of D-glucose. This nonoptimized cascade was repeated using 2.3 U/mL of *Pu*DHT (1.8 mg/mL). Initially, a steady increase in the concentrations of D-glyceraldehyde and D-glycerate was observed. However, D-glyceraldehyde quickly oxidized to D-glycerate and was no longer observed after 1 h. The concentration of D-glycerate increased over the first 2 h but then steadily decreased until a minimum was reached after 12 h (Figure 2).

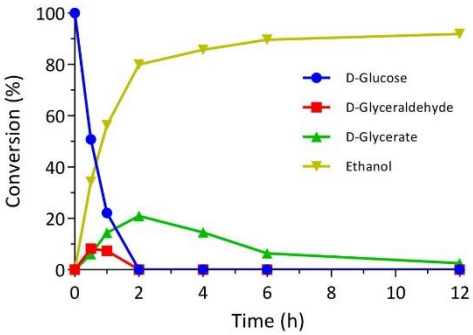


Figure 2. Cell-free biotransformation of D-glucose to ethanol utilizing *Pu*DHT as the dehydratase. A significantly faster reaction rate and higher yield were achieved in comparison to the reaction with *Ss*DHAD.[21] One mole of D-glucose yields up to 1 mol of D-glyceraldehyde or D-glycerate as intermediates or 2 mol of ethanol as the end product. The reaction mixture contained D-glucose 25 mM in HEPES 50 mM, pH 7.5, at 50 °C. Error bars represent standard deviation from two replicates. In this experiment, all standard deviations were <3%; thus, they are covered by the data markers of Figure 2.

In this nonoptimized cascade reaction, almost 80% of the theoretical ethanol yield was achieved in the first 2 h of the process and then there was a steady increase to ~90% after 6 h. In comparison, in the equivalent cascade but with *Ss*DHAD, only 57% of the possible ethanol yield was obtained after 19 h.[21] Similarly, in the production cascade for L-alanine where *Ss*DHAD and *Cc*XylDHT were used in combination to address the low activity of *Ss*DHAD toward D-gluconate, the total L-alanine yield only reached 52% after 12 h.[25] A higher product yield (up to 90%) with *Ss*DHAD was reported for the biotransformation of D-glucose to L-lactate.[26] However, 20 mg/mL of *Ss*DHAD had to be used in order to achieve such a high yield. Such a high required concentration for a biocatalyst limits the practical scope of the cell-free biosynthesis of chemicals from D-glucose and sharply contrasts with the 1.8 mg/mL of *Pu*DHT needed in the cascade described above.[26]

In addition to obtaining a higher yield, employing *Pu*DHT also increased the overall productivity of the reaction. The ethanol production rate estimated from the data in Figure 2 reached 30 mM/h in the first hour and 20 mM/h in the next hour. In comparison, in the equivalent reaction with *Ss*DHAD

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

the rate was only 2.5 mM/h, and for the production of L-lactate in which 20 mg/mL *Ss*DHAD was used, the highest production rate achieved was 10 mM/h.[21,26] The significant improvement in the production rate was achieved not only due to the improved activity toward D-glycerate but also because of the considerably higher activity of *Pu*DHT toward D-gluconate (130 U/mg) in comparison to *Ss*DHAD (0.6 U/mg) (Table 1).[27] This could also be the reason for the higher production rate achieved in this study in comparison to other study that attempted to bypass the direct dehydration of D-glycerate to pyruvate with three additional enzymes and ATP (nED pathway).[39] In summary, by application of *Pu*DHT in the model cascade, the higher TTNs and activities of *Pu*DHT toward D-glycerate and D-gluconate significantly improved the yield and production rate for the biotransformation of D-glucose.

An Active Dehydratase Facilitated a Faster Conversion of Pure and Crude Glycerol to Pyruvate. The increased production of biodiesel in recent years has also led to a sharp accumulation of the byproduct glycerol, which needs to be disposed and which limits the economic viability of biodiesel production.[12,13] The utilization of glycerol to produce value-added chemicals is thus a pivotal factor for the feasibility of commercial biodiesel production. Several cascades starting from glycerol have been developed in recent years, and the dehydration of D-glycerate to pyruvate has been identified as the main bottleneck.[22,24]

In cell metabolism, glycerol is usually converted to pyruvate via ATP-dependent pathways, requiring at least eight enzymes.[22] The coupling of an alditol oxidase with *Ss*DHAD reduced the number of required enzymes to three (including a catalase needed to remove H_2O_2) for the conversion of glycerol to pyruvate (a detailed cascade reaction is presented in Scheme S2).[22] Both *Ss*DHAD and the alditol oxidase have low activities (k_{cat} < 1 s^{-1}), with *Ss*DHAD being approximately 1 order of magnitude less reactive.[27,40] Furthermore, in recent work the activity of the alditol oxidase from *Streptomyces coelicolor* (*Sc*Aldo) has been enhanced, leading to an improvement in the catalytic efficiency of glycerol consumption.[40] We used this oxidase in our trial cascade, but due to its mesophilic nature, the reactions were run at 25 °C using both pure and crude glycerol as a starting material.

The production of pyruvate from pure glycerol happened relatively instantaneously. However, accumulation of D-glyceraldehyde and D-glycerate were observed over the initial 2 and 4 h, respectively (Figure 3). The presence of the intermediate D-glycerate indicated that *Sc*Aldo was faster than *Pu*DHT in the beginning of the reaction. However, pyruvate formation increased steadily, reaching 90% of the theoretical yield after only 12 h. When *Ss*DHAD was used instead of *Pu*DHT, a 90% conversion was reached only after 18 h.[22] Additionally, in the cascade with *Pu*DHT, a 5-fold higher substrate load was used, and the production rate of pyruvate reached 5.3 mM/h (Figure 3), an almost 8-fold improvement in comparison to the equivalent cascade using *Ss*DHAD (0.7 mM/h).[22] It is also worth reiterating that the reaction with *Pu*DHT was performed at ambient temperature, while that with *Ss*DHAD was carried out at 50 °C. At 25 °C *Ss*DHAD would have <50% of its activity at 50 °C, thus reducing the efficiency of that cascade even further in comparison with the *Pu*DHT-containing process.[27]

Next, the *Pu*DHT-containing cascade was tested using crude glycerol as the substrate. The substrate load was reduced by

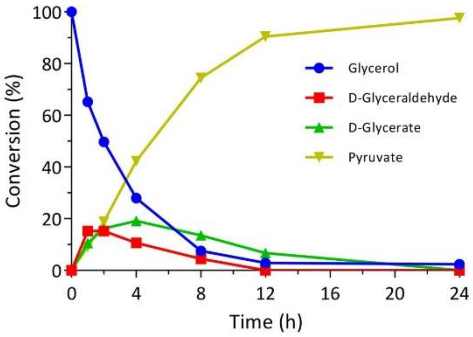


Figure 3. *In vitro* conversion of pure glycerol to pyruvate utilizing *Pu*DHT as the dehydratase. Significantly higher product yield and faster reaction rates were achieved in comparison to an equivalent cascade using *Ss*DHAD.[22] The reaction mixture contained glycerol 50 mM in HEPES 250 mM, pH 8.0, at 25 °C. Error bars represent the standard deviation from two replicates. In this experiment, all standard deviations were <3%; thus, they are covered by the data markers of Figure 3.

half (25 mM instead of 50 mM) due to the presence of possible inhibitors in the crude glycerol.[15,41] We also doubled the amount of *Sc*Aldo while retaining the concentration of *Pu*DHT. Crude glycerol was pretreated via a simple acidification method with HCl.[42] Two fractions were formed at a lower pH (~2). After removal of the insoluble fraction by means of filtration, the soluble fraction was neutralized by KOH. Further treatment using methanol and activated carbon could improve the purity of the substrate sample. However, these steps were not performed in order to minimize the costs of the process and to minimize potential health threats resulting from the use of methanol. In general, the conversion of pretreated crude glycerol progressed rather quickly during the first hour and slowed down over time, thus following a logarithmic function (Figure 4). The conversion reached 60% and 75% after 18 and 45 h, respectively. This reaction was thus

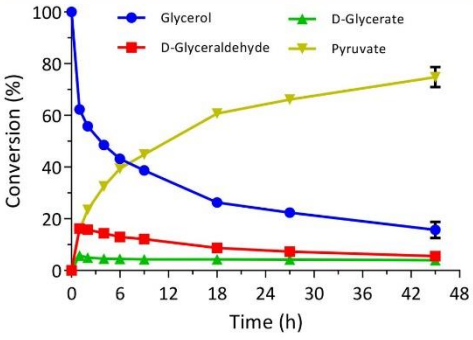


Figure 4. Conversion of crude glycerol to pyruvate. A significantly slower conversion rate was observed in comparison to a process using pure glycerol (Figure 3). The reaction mixture contained crude glycerol 25 mM in HEPES 250 mM, pH 8.0 at 25 °C. Error bars represent standard deviation from two replicates.

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

much slower in comparison to the reaction with pure glycerol (see above). It is nevertheless noteworthy that, in contrast to the reaction with pure glycerol, the concentration of D-glycerate remained very low (<1.5 mM). The amount of D-glyceraldehyde was similar in both processes, reaching ~3.5 mM after 2 h. This observation indicated that *Sc*Aldo was more affected by the presence of inhibitors in crude glycerol than was *Pu*DHT, an interpretation that is further supported by the fact that, after 45 h of reaction time, the remaining amount of glycerol was still 4-fold higher than that of D-glycerate.

Challenges and Future Prospects of the New Dehydratases. Although *Pu*DHT performed much better than *Ss*DHAD in our model *in vitro* biotransformations, its aptitude for the industrial synthesis of fine chemicals from D-glucose and glycerol still shows potential for further improvement, particularly with respect to the TTN for and activity toward D-glycerate. Depending on which chemicals need to be synthesized, a further improvement of 10−100-fold may be desirable.[36] At the present time, we are applying DNA shuffling and epPCR using some of the D-glycerate and promiscuous-class dehydratases that we identified in this work in order to engineer variants which are catalytically more efficient with respect to TTN and D-glycerate activity.[43,44] Innovative selection platforms are being employed to facilitate the screening of large numbers of variants within a short time frame.[45] Furthermore, in order to use (semi)rational strategies for the design of enzyme variants with higher TTN and activity toward D-glycerate, we are currently also optimizing crystallization conditions for *Pu*DHT and *Ft*DHAD so as to solve their structures and to identify amino acid residues and structural elements relevant to the binding of substrates and the catalytic mechanism.[46,47]

■ CONCLUSIONS

In this work, two novel classes of [Fe-S]-dehydratases (i.e., the D-glycerate and promiscuous classes) were discovered using signature sequences based on previously characterized dehydratases (the pentonate dehydratases). Both classes demonstrated significantly higher activities for the dehydration of D-glycerate to pyruvate in comparison to *Ss*DHAD. This step represents the major bottleneck in cell-free biotransformations of D-glucose and glycerol to fine chemicals (Scheme 1).[21,22,25,26] Over the past decade, *Ss*DHAD has been used with limited success to catalyze this reaction.[21,22,24−26] We have demonstrated herein that the promiscuous class *Pu*DHT has *in vitro* catalytic properties which far exceed those of *Ss*DHAD and exhibits high activities toward both D-glycerate and D-gluconate at 50 °C (>100-fold higher than those of *Ss*DHAD). *Pu*DHT also has >100-fold and >300-fold higher TTNs for D-glycerate and D-gluconate, respectively, in comparison with *Ss*DHAD.[27] The replacement of *Ss*DHAD by *Pu*DHT in our model cascade starting from D-glucose resulted in a 10-fold increase in production rate and almost doubled the product yield at 50 °C.[21] Furthermore, employing *Pu*DHT in a cascade used to convert pure glycerol to pyruvate at 25 °C resulted at in an nearly 8-fold increase in production rate over the use of *Ss*DHAD at 50 °C.[22] *Pu*DHT was also effective in the conversion of crude glycerol to pyruvate, and consequently, this step no longer represented the bottleneck of the cascade. It is anticipated that the relatively simple sequence-based approach to the discovery of *Pu*DHT and other dehydratases will inspire similar strategies used to discover other classes of enzymes which catalyze seemingly "nonexisting" reactions. It is also expected that the superior performance of *Pu*DHT in cell-free biotransformations of D-glucose and glycerol will be able to drive the sustainable production of a number of chemicals, thus alleviating our dependence on fossil fuels.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.9b05068.

SDS-PAGE of all dehydratases purified, multiple sequence alignment of representative dehydratases, and protein annotations, as well as the source of the microorganisms, Michaelis−Menten kinetics of *Pu*DHT, *Hs*DHT, and *Ft*DHAD, preliminary work on an ethanol cascade as well as detailed auxiliary enzymes used in the biotransformations of D-glucose and glycerol, chromatograms of substrates, intermediates, and product separation as well as a comparison of pure and crude glycerol, and a detailed experimental section (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Volker Sieber − *Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability and Catalytic Research Center, Technical University of Munich, 94315 Straubing, Germany; School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Queensland 4072, Australia; Straubing Branch BioCat, Fraunhofer IGB, 94315 Straubing, Germany;* orcid.org/0000-0001-5458-9330; Email: sieber@tum.de

Authors

Samuel Sutiono − *Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, 94315 Straubing, Germany*

Mariko Teshima − *Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, 94315 Straubing, Germany*

Barbara Beer − *Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, 94315 Straubing, Germany*

Gerhard Schenk − *School of Chemistry and Molecular Biosciences and Sustainable Minerals Institute, The University of Queensland, St. Lucia, Queensland 4072, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.9b05068

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.S. was funded by the BMBF (German Federal Ministry of Education and Research) through the HotSysAPP project (Grant No. 031L0078F), and M.T. was funded by the TUM IGSSE (International Graduate School of Science and Engineering). B.B. was funded by the BMBF through the PFIFF project (Grant No. 031B0351C). Crude glycerol was obtained from a biodiesel processing plant. The authors thank Tristan Rath for the preparation of crude glycerol. The plasmid of *Sc*Aldo was a kind gift from Prof. Bornscheuer (University of Greifswald). S.S. and V.S. wish to acknowledge Prof. Victor

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

Guallar and Gerard Santiago for fruitful discussions of this project during S.S.'s internship at the Barcelona Supercomputing Centre (BSC). Discussions with Samed Güner, Ioannis Zachos, and André Pick are also greatly appreciated. The authors also thank Edilberto V. Medina-Cabrera for his help in statistical analysis and Samed Güner for his invaluable feedback and for proofreading the manuscript.

■ REFERENCES

(1) Klass, D. L. Energy Consumption, Reserves, Depletion, and Environmental Issues. In *Biomass for Renewable Energy, Fuels, and Chemicals*; Klass, D. L., Ed.; Academic Press: San Diego, 1988; pp 1−27.
(2) Höök, M.; Tang, X. Depletion of Fossil Fuels and Anthropogenic Climate Change — A Review. *Energy Policy* **2013**, *52*, 797−809.
(3) Lelieveld, J.; Klingmüller, K.; Pozzer, A.; Burnett, R. T.; Haines, A.; Ramanathan, V. Effects of Fossil Fuel and Total Anthropogenic Emission Removal on Public Health and Climate. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 7192−7197.
(4) Intergovernmental Panel on Climate Change. Drivers, Trends and Mitigation. In *Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the IPCC Fifth Assessment Report*; Cambridge University Press: Cambridge, 2015; pp 351−412.
(5) Mohr, S. H.; Wang, J.; Ellem, G.; Ward, J.; Giurco, D. Projection of World Fossil Fuels by Country. *Fuel* **2015**, *141*, 120−135.
(6) Covert, T.; Greenstone, M.; Knittel, C. R. Will We Ever Stop Using Fossil Fuels? *J. Econ. Perspect.* **2016**, *30*, 117−138.
(7) García-Olivares, A. Substitutability of Electricity and Renewable Materials for Fossil Fuels in a Post-Carbon Economy. *Energies* **2015**, *8*, 13308−13343.
(8) Zou, C.; Zhao, Q.; Zhang, G.; Xiong, B. Energy Revolution: From a Fossil Energy Era to a New Energy Era *. *Nat. Gas Ind. B* **2016**, *3*, 1−11.
(9) Isikgor, F. H.; Becer, C. R. Lignocellulosic Biomass: A Sustainable Platform for the Production of Bio-Based Chemicals and Polymers. *Polym. Chem.* **2015**, *6*, 4497−4559.
(10) Huang, Y.-B.; Fu, Y. Hydrolysis of Cellulose to Glucose by Solid Acid Catalysts. *Green Chem.* **2013**, *15*, 1095−1111.
(11) Buffiere, J.; Ahvenainen, P.; Borrega, M.; Svedstrom, K.; Sixta, H. Supercritical Water Hydrolysis: A Green Pathway for Producing Low-Molecular-Weight Cellulose. *Green Chem.* **2016**, *18*, 6516−6525.
(12) Lari, G. M.; Pastore, G.; Haus, M.; Ding, Y.; Papadokonstantakis, S.; Mondelli, C.; Perez-Ramirez, J. Environmental and Economical Perspectives of a Glycerol Biorefinery. *Energy Environ. Sci.* **2018**, *11*, 1012−1029.
(13) Zhou, C. C.; Beltramini, J. N.; Fan, Y.; Lu, G. Q. M. Chemoselective Catalytic Conversion of Glycerol as a Biorenewable Source to Valuable Commodity Chemicals. *Chem. Soc. Rev.* **2008**, *37*, 527−549.
(14) Taylor, R.; Nattrass, L.; Alberts, G.; Robson, P.; Chudziak, C.; Bauen, A.; Libelli, I. M.; Lotti, G.; Prussi, M.; Nistri, R.; Chiaramonti, D.; López-Contreras, A. M.; Bos, H. L.; Eggink, G.; Springer, J.; Bakker, R.; Ree Van, R. *From the Sugar Platform to Biofuels and Biochemicals: Final Report for the European Commission Directorate-General Energy*; E4tech/Re-CORD/Wageningen UR: 2015.
(15) Sivasankaran, C.; Ramanujam, P. K.; Balasubramanian, B.; Mani, J. Recent Progress on Transforming Crude Glycerol into High Value Chemicals: A Critical Review. *Biofuels* **2019**, *10*, 309−314.
(16) Claassens, N. J.; Burgener, S.; Vo, B.; Erb, T. J.; Bar-Even, A. A Critical Comparison of Cellular and Cell-Free Bioproduction Systems. *Curr. Opin. Biotechnol.* **2019**, *60*, 221−229.
(17) Wilding, K. M.; Schinn, S.; Long, E. A.; Bundy, B. C. The Emerging Impact of Cell-Free Chemical Biosynthesis. *Curr. Opin. Biotechnol.* **2018**, *53*, 115−121.
(18) Guterl, J.; Sieber, V. Biosynthesis " Debugged ": Novel Bioproduction Strategies. *Eng. Life Sci.* **2013**, *13*, 4−18.
(19) Dudley, Q. M.; Karim, A. S.; Jewett, M. C. Cell-Free Metabolic Engineering: Biomanufacturing beyond the Cell. *Biotechnol. J.* **2015**, *10*, 69−82.
(20) Budgen, N.; Danson, M. J. Metabolism of Glucose via a Modified Entner-Doudoroff Pathway in the Thermoacidophilic Archaebacterium. *FEBS Lett.* **1986**, *196*, 207−210.
(21) Guterl, J. K.; Garbe, D.; Carsten, J.; Steffler, F.; Sommer, B.; Reiße, S.; Philipp, A.; Haack, M.; Rühmann, B.; Koltermann, A.; Kettling, U.; Brück, T.; Sieber, V. Cell-Free Metabolic Engineering: Production of Chemicals by Minimized Reaction Cascades. *ChemSusChem* **2012**, *5*, 2165−2172.
(22) Gao, C.; Li, Z.; Zhang, L.; Wang, C.; Li, K.; Ma, C.; Xu, P. An Artificial Enzymatic Reaction Cascade for a Cell-Free Bio-System Based on Glycerol. *Green Chem.* **2015**, *17*, 804−807.
(23) van Hellemond, E. W.; Vermote, L.; Koolen, W.; Sonke, T.; Zandvoort, E.; Heuts, D. P. H. M.; Janssen, D. B.; Fraaije, M. W. Exploring the Biocatalytic Scope of Alditol Oxidase from Streptomyces Coelicolor. *Adv. Synth. Catal.* **2009**, *351*, 1523−1530.
(24) Li, Z.; Yan, J.; Sun, J.; Xu, P.; Ma, C.; Gao, C. Production of Value-Added Chemicals from Glycerol Using in Vitro Enzymatic Cascades. *Commun. Chem.* **2018**, *1*, 1−7.
(25) Gmelch, T. J.; Sperl, J. M.; Sieber, V. Optimization of a Reduced Enzymatic Reaction Cascade for the Production of L-Alanine. *Sci. Rep.* **2019**, *9*, 11754.
(26) Xie, L.; Wei, X.; Zhou, X.; Meng, D.; Zhou, R. Conversion of D -Glucose to L -Lactate via Pyruvate by an Optimized Cell-Free Enzymatic Biosystem Containing Minimized Reactions. *Synth. Syst. Biotechnol.* **2018**, *3*, 204−210.
(27) Carsten, J. M.; Schmidt, A.; Sieber, V. Characterization of Recombinantly Expressed Dihydroxy-Acid Dehydratase from Sulfobus Solfataricus — A Key Enzyme for the Conversion of Carbohydrates into Chemicals. *J. Biotechnol.* **2015**, *211*, 31−41.
(28) Bommarius, A. S.; Schwarm, M.; Drauz, K. Comparison of Different Chemoenzymatic Process Routes to Enantiomerically Pure Amino Acids. *Chimica* **2001**, *55*, 50−59.
(29) Andberg, M.; Aro-Kärkkäinen, N.; Carlson, P.; Oja, M.; Bozonnet, S.; Toivari, M.; Hakulinen, N.; Donohue, M. O. Characterization and Mutagenesis of Two Novel Iron - Sulphur Cluster Pentonate Dehydratases. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 7549−7563.
(30) Pohlmann, A.; Fricke, W. F.; Reinecke, F.; Kusian, B.; Liesegang, H.; Schwartz, E.; Strittmatter, A.; Cramm, R.; Eitinger, T.; Ewering, C.; Po, M.; Bowien, B.; Vo, I.; Gottschalk, G.; Steinbu, A. Genome Sequence of the Bioplastic-Producing '" Knallgas "' Bacterium Ralstonia Eutropha H16. *Nat. Biotechnol.* **2006**, *24*, 1257−1262.
(31) Reher, M.; Fuhrer, T.; Bott, M.; Schonheit, P. The Nonphosphorylative Entner-Doudoroff Pathway in the Thermoacidophilic Euryarchaeon Picrophilus Torridus Involves a Novel 2-Keto-3-Deoxygluconate- Specific Aldolase. *J. Bacteriol.* **2010**, *192*, 964−974.
(32) Bommarius, A. S. Check Nature First, Then Evolve. *Nat. Chem. Biol.* **2010**, *6*, 793−794.
(33) Höhne, M.; Schätzle, S.; Jochens, H.; Robins, K.; Bornscheuer, U. T. Rational Assignment of Key Motifs for Function Guides in Silico Enzyme Identification. *Nat. Chem. Biol.* **2010**, *6*, 807−813.
(34) Simossis, V. A.; Heringa, J. PRALINE: A Multiple Sequence Alignment Toolbox That Integrates Homology-Extended and Secondary Structure Information. *Nucleic Acids Res.* **2005**, *33*, W289−W294.
(35) Rahman, M. M.; Andberg, M.; Thangaraj, S. K.; Parkkinen, T.; Penttila, M.; Koivula, A.; Rouvinen, J.; Hakulinen, N. The Crystal Structure of a Bacterial L-Arabinonate Dehydratase Contains a [2Fe-2S] Cluster. *ACS Chem. Biol.* **2017**, *12*, 1919−1927.
(36) Bommarius, A. S.; Paye, M. F. Stabilizing Biocatalysts. *Chem. Soc. Rev.* **2013**, *42*, 6534−6565.
(37) Sutiono, S.; Carsten, J.; Sieber, V. Structure-Guided Engineering of α-Keto Acid Decarboxylase for the Production of Higher Alcohols at Elevated Temperature. *ChemSusChem* **2018**, *11*, 3335−3344.

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

(38) Sutiono, S.; Satzinger, K.; Pick, A.; Carsten, J.; Sieber, V. To Beat the Heat - Engineering of the Most Thermostable Pyruvate Decarboxylase to Date. *RSC Adv.* **2019**, *9*, 29743−29746.
(39) Okano, K.; Zhu, Q.; Honda, K. In Vitro Reconstitution of Non-Phosphorylative Entner e Doudoroff Pathway for Lactate Production. *J. Biosci. Bioeng.* **2019**, DOI: 10.1016/j.jbiosc.2019.09.010.
(40) Gerstenbruch, S.; Wulf, H.; Mußmann, N.; Connell, T. O.; Maurer, K.; Bornscheuer, U. T. Asymmetric Synthesis of D -Glyceric Acid by an Alditol Oxidase and Directed Evolution for Enhanced Oxidative Activity towards Glycerol. *Appl. Microbiol. Biotechnol.* **2012**, *96*, 1243−1252.
(41) Yang, F.; Hanna, M. A.; Sun, R. Value-Added Uses for Crude Glycerol - a Byproduct of Biodiesel Production. *Biotechnol. Biofuels* **2012**, *5*, 13.
(42) Yuan, Z.; Qin, W.; Ma, P. Purification of Crude Glycerol Using Acidification: Effects of Acid Types and Product Characterization. *Austin Chem. Eng.* **2018**, *1*, 1−7.
(43) Tracewell, C. A.; Arnold, F. H. Directed Enzyme Evolution: Climbing Fitness Peaks One Amino Acid at a Time. *Curr. Opin. Chem. Biol.* **2009**, *13*, 3−9.
(44) Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. Molecular Evolution by Staggered Extension Process (StEP) in Vitro Recombination. *Nat. Biotechnol.* **1998**, *16*, 258−261.
(45) Xiao, H.; Bao, Z.; Zhao, H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015**, *54*, 4011−4020.
(46) Reetz, M. T.; Prasad, S.; Carballeira, D. Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity: Rigorous Comparison with Traditional Methods. *J. Am. Chem. Soc.* **2010**, *132*, 9144−9152.
(47) Lutz, S. Beyond Directed Evolution — Semi-Rational Protein Engineering and Design. *Curr. Opin. Biotechnol.* **2010**, *21*, 734−743.

https://dx.doi.org/10.1021/acscatal.9b05068
ACS Catal. 2020, 10, 3110−3118

Supporting Information

Enabling the Direct Enzymatic Dehydration of D-Glycerate to Pyruvate as the Key Step in Synthetic Enzyme Cascades Used in the Cell-Free Production of Fine Chemicals

Samuel Sutiono, [a] *Mariko Teshima,* [a] *Barbara Beer,* [a] *Gerhard Schenk,* [d,e] *Volker Sieber* *[a,b,c,d]

a. Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, Schulgasse 16, 94315 Straubing, Germany.

b. Catalytic Research Center, Technical University of Munich, Ernst-Otto-Fischer-Straße 1, 85748 Garching, Germany

c. Straubing Branch BioCat, Fraunhofer IGB, Schulgasse 11a, 94315 Straubing, Germany

d. School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Copper Road, St. Lucia 4072, Australia

e. Sustainable Minerals Institute, The University of Queensland, 47 Staff House Road, St. Lucia 4072, Australia

*Email: sieber@tum.de

Table of Contents

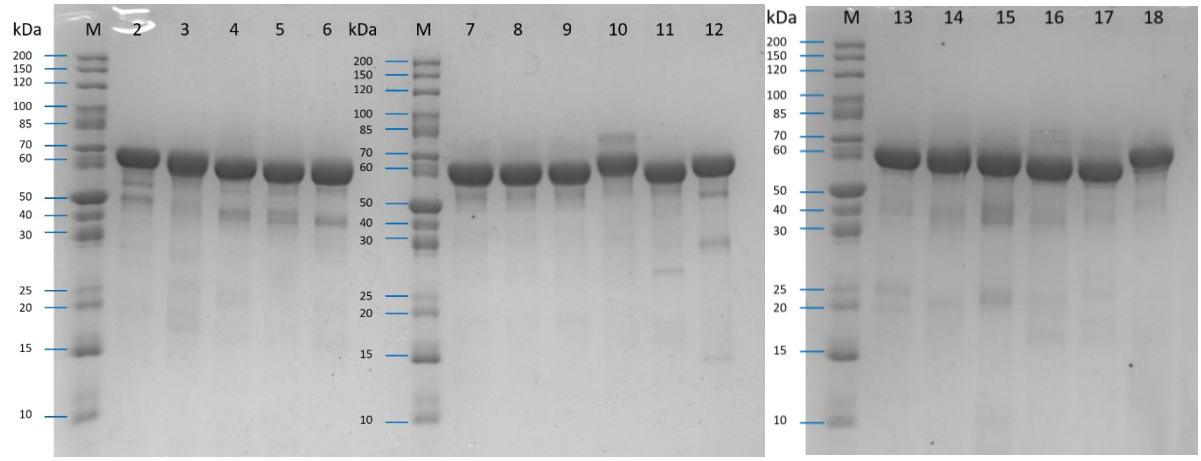


Figure S1. SDS-PAGE gel of all dehydratases purified in this study. The number in the gel corresponds to the enzyme listed in Table 1 in manuscript.

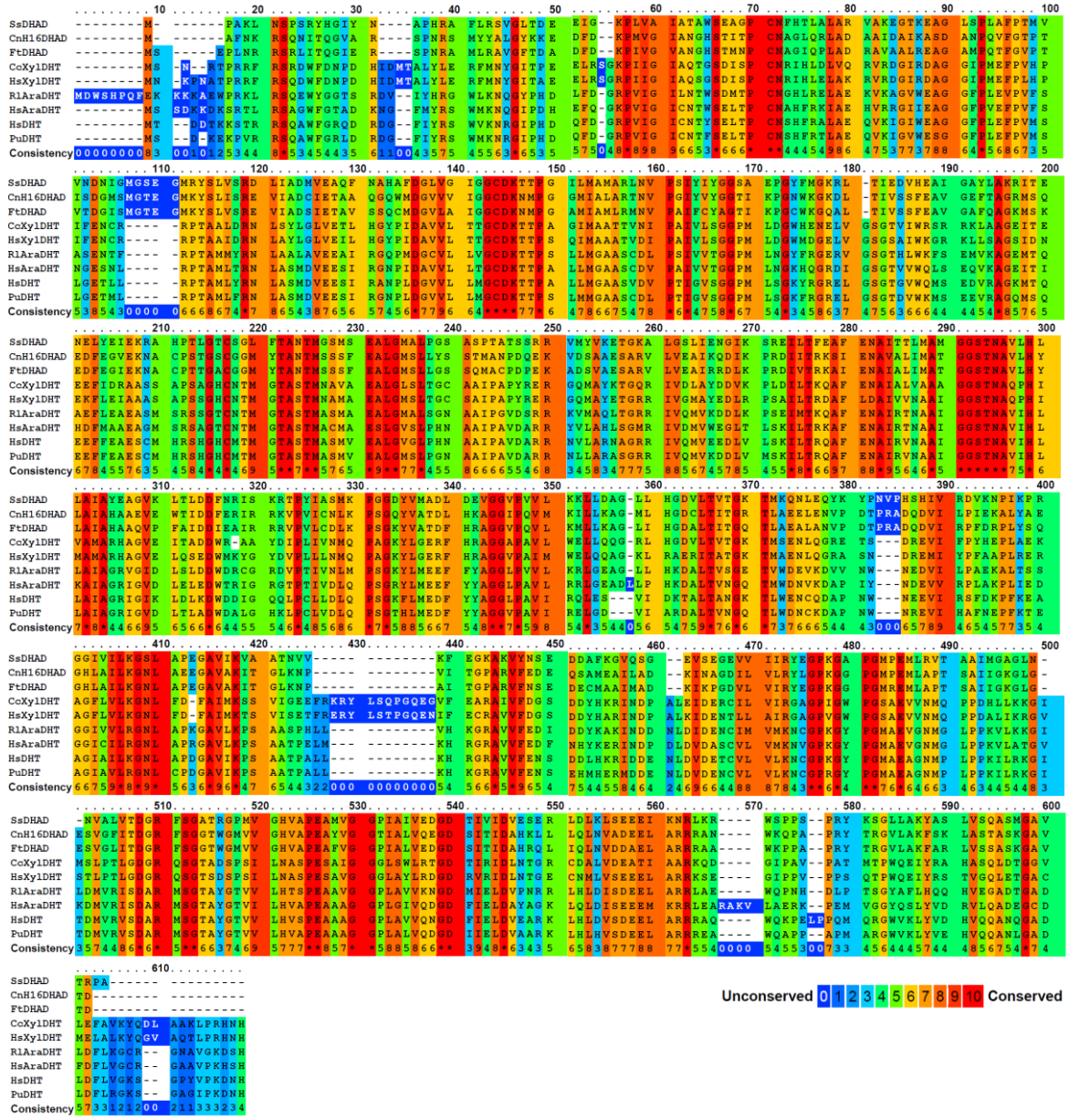


Figure S2. Multiple sequence alignment of several dehydratases from petonate-, D-glycerate-, and promiscuous-class dehydratases. The alignment was performed by PRALINE (http://www.ibi.vu.nl/programs/pralinewww/).

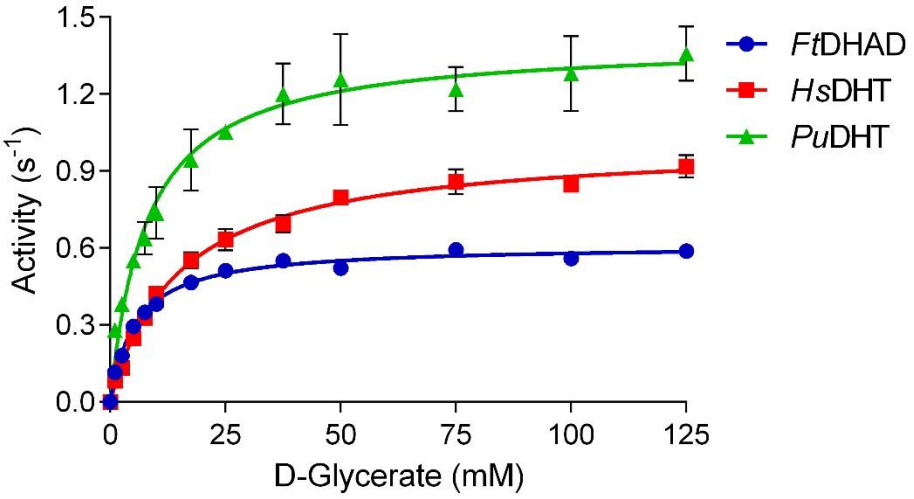


Figure S3. Michaelis-Menten kinetics of the three most promising dehydratases toward D-glycerate. The kinetic parameters presented in Table 2 in the main text were calculated using non-linear regression with least-square fit. All activities presented in Figure S3 were determined by linear regression of three different time points (0, 15, 60 min), of which two replicates were performed. Error bars represent standard deviation.

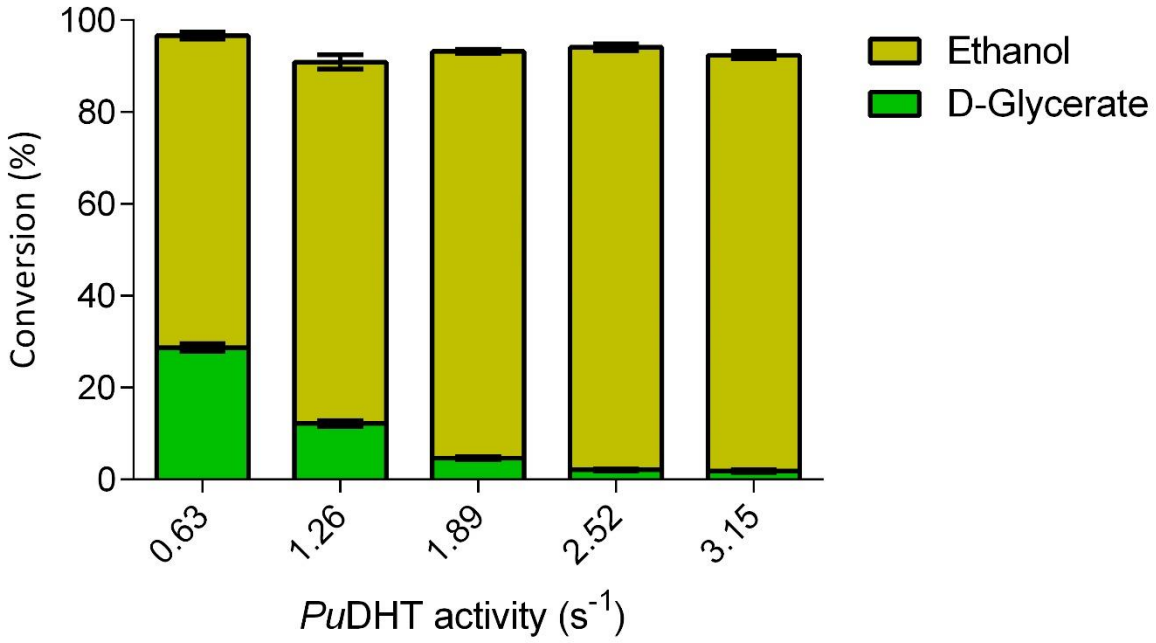


Figure S4. Effect of *Pu*DHT concentration on the ethanol yield and the concentration of remaining D-glycerate after 16 h biotransformation at 50 °C. Higher activity (increased concentration) of *Pu*DHT toward D-glycerate resulted in higher ethanol yield and lower remaining D-glycerate at the end of biotransformation. The reaction was performed with D-glucose 25 mM in HEPES 50 mM, pH 7.5 at 50 °C for 16 h at 500 rpm. Conversion (%) is defined as the amount of ethanol formed or D-glycerate remaining per theoretical ethanol concentration that was able to be achieved. One mole of D-glucose will give a maximum yield of one mole D-glycerate intermediate or two moles of ethanol as final product. Negative control consisted of all components except for D-glucose. Negative control yielded neither D-glycerate nor ethanol. Error bars represent standard deviation from two replicates.

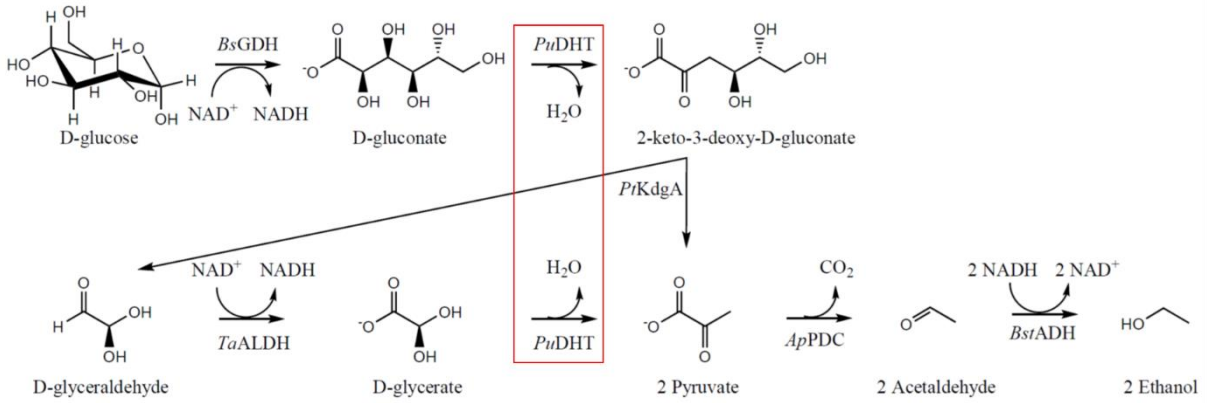
Table S1. List of dehydratases cloned and expressed in this study. *Rl*AraDHT, *Cc*XylDHT, and *Tv*DHAD were ordered as gene codon-optimized for E.coli

Microorganisms	DSM number	Enzymes	NCBI Ref. Seq.	Restriction enzymes	Primers (5´→ 3´)
Rhizobium leguminosarum	-	*Rl*AraDHT	PDB: 5j84	NdeI/XhoI	Fwd: CACCAG*CATATG*AAGAAAAAAGCAGAATGGCCTCGTAAACTG Rev: CTGGTG*CTCGAG*CTATTAATGGCTATCTTTACCAACGGCATTACC
Caulobacter crescentus	-	*Cc*XylDHT	PDB: 5oyn	NdeI/XhoI	Fwd: TATATA*CATATG*AGCAATCGTACACCGCG Rev: TATATA*CTCGAG*TTAATGATTATGGCGAGGC
Cupriavidus necator H16	428	*Cn*H16DHAD	WP_010814934.1	NdeI/XhoI	Fwd: CAGCAG*CATATG*GCATTCAACAAACGCTCG Rev: CAGCAG*CTCGAG*TTAGTCCGTCACTGCCCCCTTG
Cupriavidus necator N1	13513	*Cn*N1DHAD	WP_013957820.1	NdeI/XhoI	Fwd: CACGAG*CATATG*GCATTCAACAAACGCTCGCAG Rev: CACGAG*CTCGAG*TTAATCCGTCACCGCACCCTTGCTC
Cupriavidus metallidurans	2839	*Cm*DHAD	WP_011517373.1	NdeI/HindIII	Fwd: CACGAG*CATATG*GCTTACAACAAACGTTCCCAGC Rev: CACGAGA*AGCTT*TTAGTCCGTCACGGCGCCCTTGC
Schlegelella thermodepolymerans	15344	*St*DHAD	WP_104358646.1	NdeI/XhoI	Fwd: CACGAG*CATATG*GACACCAAGACGATCCGGATC Rev: CACGAG*CTCGAG*TTACTCGAACTTGTCGAGCACCGCAC
Fontimonas thermophila	23609	*Ft*DHAD	WP_091533200.1	NdeI/XhoI	Fwd: CACGAG*CATATG*TCCGAGCCGTTGAATCGC Rev: CACGAG*CTCGAG*TTAGTCGGTCACCGCGCCCTTG
Thermosynecoccus vulcanus	-	*Tv*DHAD	WP_126985616.1	NdeI/HindIII	Fwd: CAGCAG*CATATG*GCAGAAAATTGGCGTAGCCGTATTATTACCGAAGGTG Rev: GAGTTA*AAGCTT*TTAATCGGTAACTGCACCCAGGCTGCTGCTGCTCACC
Rhizobium leguminosarum	6040	*Rl*XylDHT	WP_112904778.1	NdeI/HindIII	Fwd: CAGCAG*CATATG*ACGGACAGCCATTCGCCG Rev: CAGCAG*AAGCTT*TTAATGATTGTGCCGCGGCGGGTTTTTC
Herbaspirillum seropediceae Z67	6445	*Hs*XylDHT	WP_013236417.1	NdeI/HindIII	Fwd: CAGCAG*CATATG*AACAAGCCAAACGCCACCCC Rev: CAGCAG*AAGCTT*TTAGTGATTGTGCCGGGGCAGGG
		*Hs*AraDHT	WP_013233070.1	NdeI/XhoI	Fwd: CAGCAG*CATATG*AGTAGCGACAAAAAAGACAAGAGCCGC Rev: CAGCAG*CTCGAG*TTAGTGCGAGTGCTTGGGCACGGC
		*Hs*DHT	WP_013235280.1	NdeI/XhoI	Fwd: CAGCAG*CATATG*ACTGACGATACCAAGAAATCCACCCG Rev: CAGCAG*CTCGAG*TTAGTGATTGTCCTTGGGCACGTAAGGCC
Variovorax paradoxus B4	21876	*Vp*DHT	WP_021008351.1	NdeI/HindIII	Fwd: CAGCAG*CATATG*AGCACTTCCCCCAAGAAAAAACCCGAAGACCTGC Rev: CAGCAG*AAGCTT*TTA GTGATTGTCGCGCGGCACGAAAGCCCC
Rhodonobacter fulvus	18449	*Rf*DHT	WP_007081210.1	NdeI/HindIII	Fwd: CAGCAG*CATATG*ACTTCTCCCAAAGACACGAAAAGACGCAGCCAATCC Rev: CAGCAG*AAGCTT*TTAGTGCGAATCGCGCGGCACCTCGCTGC
Noviherbaspirillum massiliense	25712	*Nm*DHT	WP_019140327.1	NdeI/HindIII	Fwd: CAGCAG*CATATG*ACCGATTCCAAGCGCCCATTGCGCC Rev: CAGCAG*AAGCTT*TTAGTGATTGTCCTTCGGCACGCCCGCGC
Paralcaligenes ureilyticus	24591	*Pu*DHT	WP_132585145.1	NheI/EcoRI	Fwd: CAGCAG*GCTAGC*ATGAGTGACAAAGAAAAGCCTGTGCGTCGCAGCCAG Rev: CAGCAG*GAATTC*TTAGTGATTGTCTTTGGGTATGCCCGCGCCGCTTTTAC
Ramlibacter henchirensis	14656	*Rh*DHT	WP_135262325.1	NdeI/HindIII	Fwd: CAGCAG*CATATG*ACTGATCGCAAACCCCGCAAAAAGCCCGAAG Rev: CAGCAG*AAGCTT*TTAGTGGTTGTCCCGAGGCACGAACGCCC

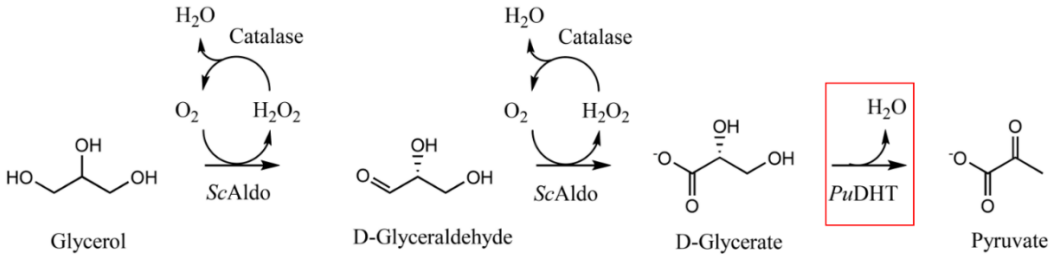
Table S2. Auxiliary enzymes for biotransformation of D-glucose to ethanol and conversion glycerol to pyruvate

Enzyme	Microorganisms	NCBI Ref. Seq.*	Substitutions	Reference
*Bs*GDH	*Bacillus subtilis*	WP_003246720.1	E170K, Q252L	[1]
*Pt*KdgA	*Picrophilus torridus*	WP_048059513.1	Wild type	[2]
*Ta*ALDH	*Thermoplasma acidophillum*	WP_010901221.1	8 points substitutions	Publication submitted
PDC-Var.2	*Acetobacter pasteurianus*	WP_141376382.1	See publication	[3]
*Bst*ADH	*Bacillus stearothermophilus*	KFL15473.1	Wild type	[4]
*Sc*Aldo	*Streptomyces coelicolor*	WP_011030685.1	V125M/A244T/V133M/ G399R	[5]

*The NCBI reference sequences presented in Table S2 corresponds to the wild type sequences.

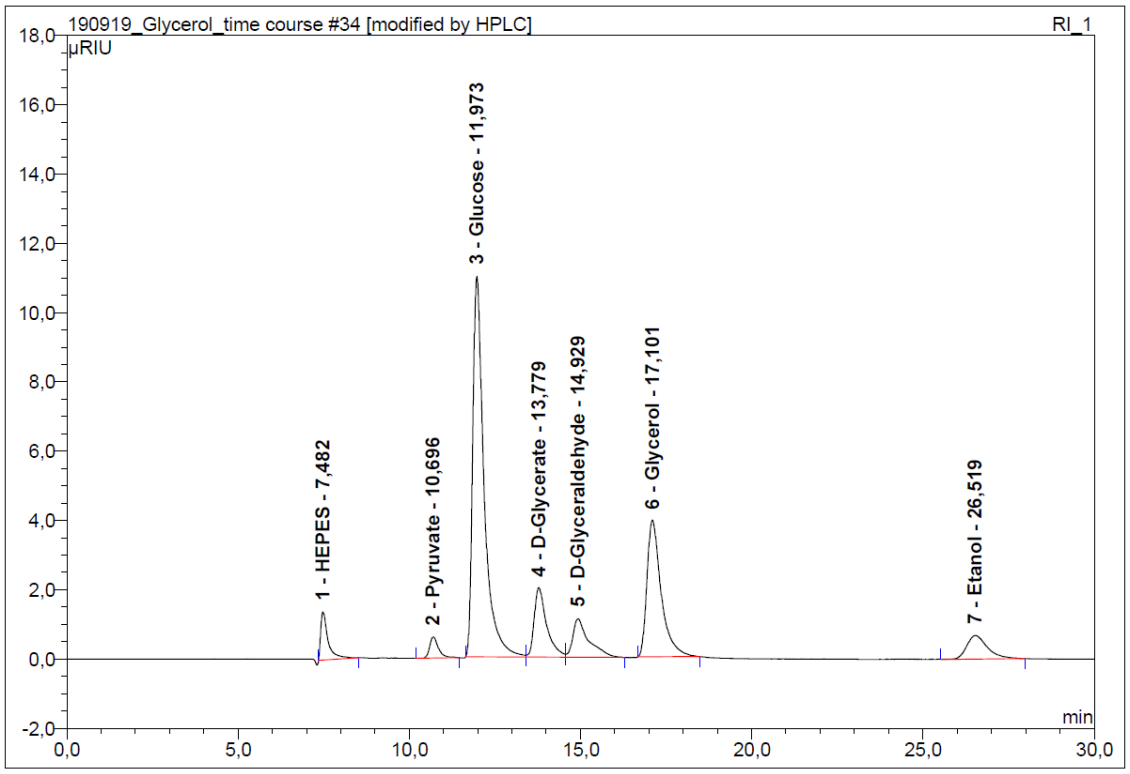


Scheme S1. Biotransformation of D-glucose to ethanol. Dehydration reactions are highlighted red.



Scheme S2. Conversion of glycerol to pyruvate. Dehydration reaction is highlighted red.

A)



B)

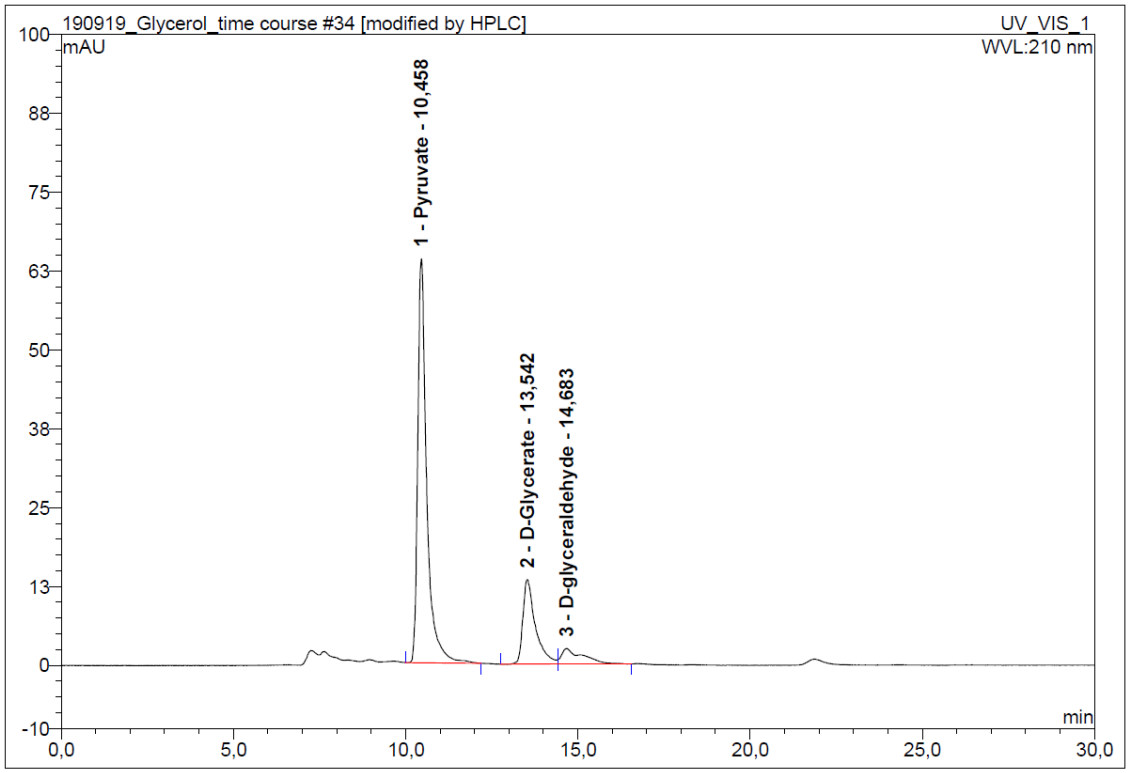
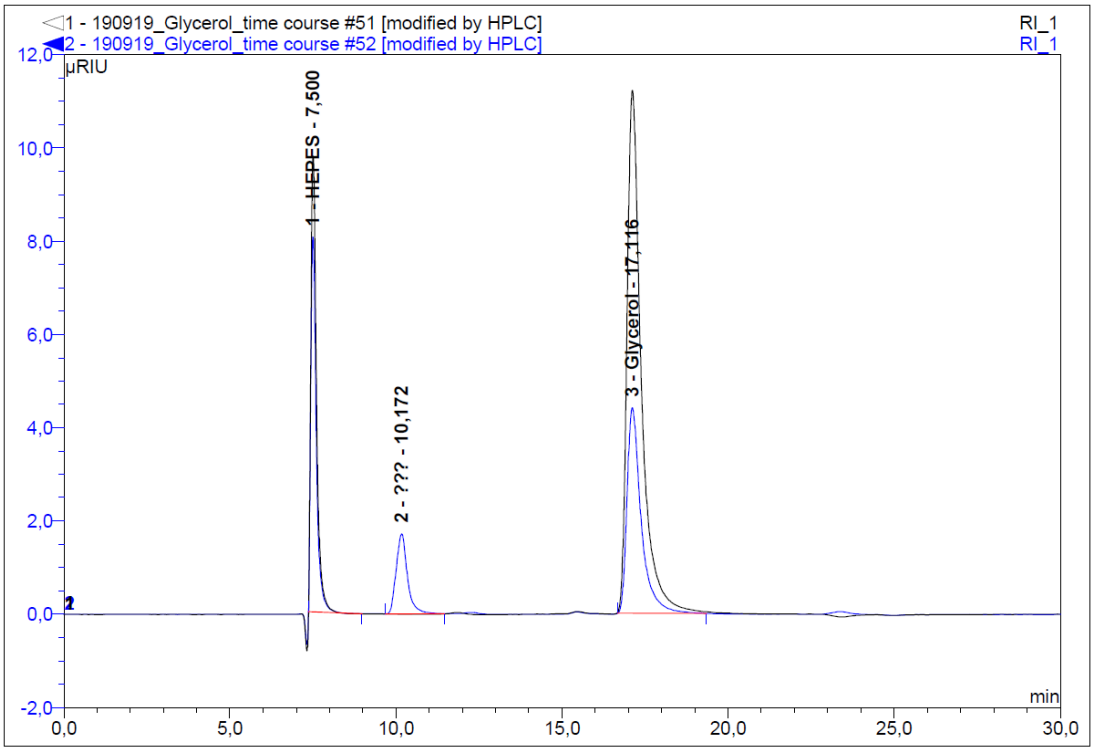


Figure S5. Chromatogram of various substrates, intermediates, products in cascades of D-glucose and glycerol detected via Refractive Index detector (A) and UV detector (B).

A)



B)

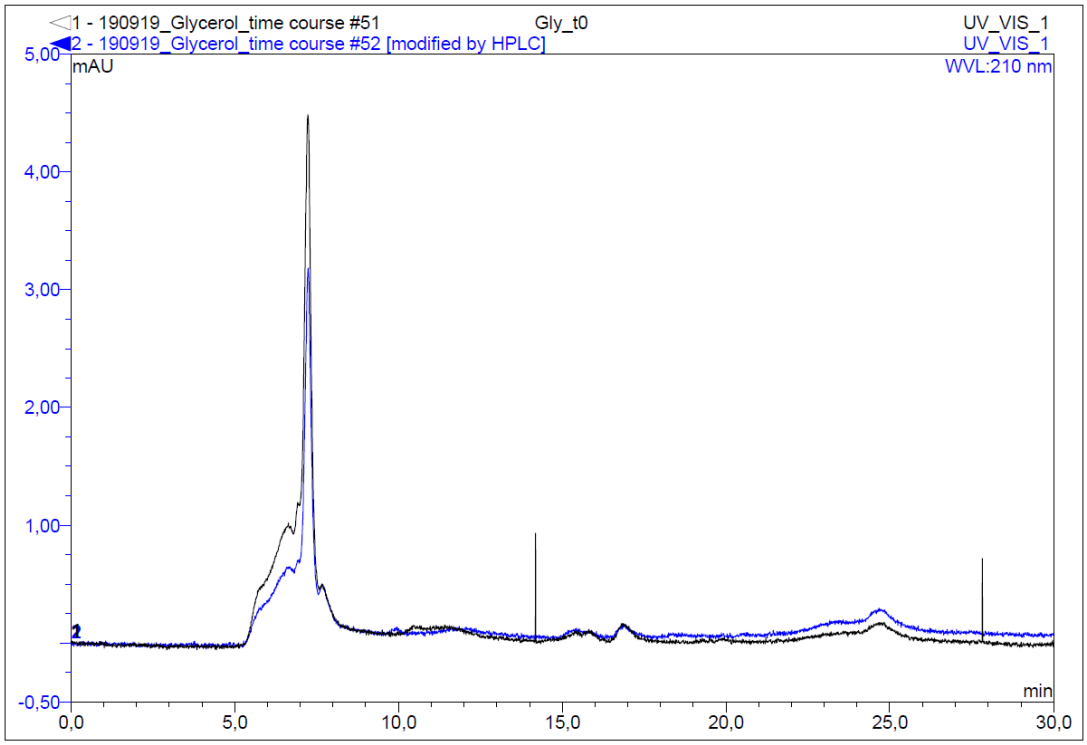


Figure S6. Crude glycerol after pretreatment with acid detected via Refractive Index detector (A) and UV detector (B). The black chromatogram is pure glycerol dissolved in HEPES, and that in blue is pre-treated crude glycerol in HEPES. Unknown peak at retention time 10.2 min observed via RI detector but not UV detector could be a potential inhibitor causing slower production rate when crude glycerol was used (Figure 4).

S1. EXPERIMENTAL SECTIONS

S1.1. Reagents

All chemicals were purchased as analytical grade or higher. Sodium gluconate was purchased from Sigma Aldrich (Darmstadt, Germany), D,L-glyceric acid was purchased from TCI Deutschland (Eschborn, Germany), and glycerol was purchased from Roth (Kalrsruhe, Germany). All restriction enzymes were purchased from New England Biolab (Frankfurt, Germany). Microorganism strains and genomic DNAs (gDNAs) were obtained from German Collection of Microorganisms and Cell Cultures (DSMZ). *Rl*AraDHT, *Cc*XylDHT, and *Tv*DHAD was ordered as a gene codon-optimized for *E. coli* from ATG Biosynthetics (Merzhausen, Germany).

S1.2. Cloning, expression, and purification

For the microbial strains ordered, gDNA was isolated using DNeasy UltraClean Microbial Kit (Qiagen, Germany) using the protocols described by the manufacturer. Two-step PCR described by NEB (PCR without annealing step) was performed to amplify the desired dehydratase using the primers described in Table S4. Each PCR product was purified and digested using the combination of restriction enzymes described in Table S4. The dehydratase genes were then ligated to pET28a plasmid (Novagen, Germany) bearing an N-terminal hexahistidine-tag. The ligated plasmids were used to transform *E. coli* BL21 (DE3) (Invitrogen, Germany) and plated on LB media containing kanamycin 50 µg/ml. The sequence of each dehydratase was confirmed by sequencing (Eurofins, Germany).

For expression, a single colony of a positive transformant was picked and grown in 25 ml LB media containing kanamycin 50 µg/ml at 150 rpm, 30 °C overnight. On the following morning, 10 ml of the overnight culture was transferred to a 2 L baffled flask containing TB (Terrific Broth) media 500 ml supplemented with kanamycin 100 µg/ml. The culture was incubated at

37 °C and shaken at 120 rpm until the OD_{600} reached 0.8 to 1. Isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM end concentration) was then added to induce protein expression. The temperature was lowered to 20 °C, and the protein expression was carried out for 16 h. The culture was pelleted at 4000 xg for 15 min, and the cell pellet was kept at -80 °C until further purification.

The cell pellet was dissolved in a binding buffer (potassium phosphate 50 mM, pH 8, imidazole 20 mM, sodium chloride 500 mM, and glycerin 10 vol%). The cells were disrupted using sonication (0.5 s, 75%) for 20 min in ice. The cell debris was removed by means of centrifugation at 40000 xg for 30 min. The supernatant was filtered through a 0.45 µm filter before application to Äkta Purifier. A Histrap column FF Crude 5 mL (GE Healthcares, Germany) was used to retain the target enzyme. A binding buffer was used to wash the column before elution with elution buffer (potassium phosphate 50 mM, pH 8, imidazole 500 mM, sodium chloride 500 mM, and glycerin 10 vol%). A yellow to red eluate was obtained for all dehydratases, indicating the presence of the [Fe-S] cluster. The buffer was then changed to HEPES 50 mM, pH 7.9, at 25 °C using HiPrep desalting column 26/10 50 ml (GE Healthcare, Germany). All enzymes were flash frozen in liquid nitrogen and kept at -80°C until further use.

S1.3. Activity on D-glycerate and D-gluconate

All activities were determined at 50 °C, unless otherwise stated. Reaction solutions contained $MgCl_2$, D,L-glycerate 50 mM (i.e. D-glycerate) or D-gluconate 25 mM in HEPES 50 mM, pH 7.5, at 50 °C (pH 7.9 at 25 °C). BSA (1 mg/ml end concentration) was added for the activity measurement of promiscuous class dehydratases toward D-gluconate due to the diluted enzymes used. The reactions were started by the addition of the respective enzyme and were performed in 1.5 ml Eppendorf tubes and shaken at 300 rpm. Aliquots were taken during at least three different sampling periods and diluted ten times in water. The aliquot was then

transferred to a 10 kDa filter (VWR, Germany) and centrifuged at 12000 xg for 3 min to stop the reaction. The formation of the products, pyruvate or 2-keto-3-deoxygluconate (KDG), were detected via HPLC coupled to a UV-detector at 210 nm. An anion exchange column (Metrosep A Supp 16-250, Metrohm, Germany) was used as the stationary phase, and an ammonium bicarbonate buffer (ABc) was used as the mobile phase. The column temperature was set to 65 °C, and a flow of 0.2 ml/min was applied. The elution profile was ABc 12 mM, pH 10.02, for 30 min, followed by 0.2 ml/min of ABc 30 mM, pH 10.40, for 10 min. The column was then re-equilibrated with ABc 12 mM, pH 10.02, for 10 min. All activities were determined as an average of two technical duplications. The catalytic activity (k_{cat}) is defined as the formation of 1 µmol of pyruvate or KDG per µmol of enzyme per second.

S1.4. Determination of Total Turnover Number (TTN)

The Total Turnover Number (TTN) was defined as the total amount of product formed in mole per mole of enzyme used.[6] Reaction solutions contained $MgCl_2$ 5 mM, D,L-glycerate 250 mM or D-gluconate 1 M, BSA 1 mg/ml in HEPES 50 mM, pH 7.5, at 50 °C. The reactions were started by addition of the respective enzymes. Aliquots were taken at 0, 24, 48, and 96 h. The reaction was stopped by centrifugation of diluted aliquot through a 10 kDa filter. The product formed was detected by HPLC as described above. No additional product was observed after 96 h for all enzymes. The TTN was calculated according to the formula described elsewhere.[6,7]

S1.5. Determination of the melting temperature (T_m)

Thermodynamic stability was determined by a thermofluor assay using the SyproOrange dye as described previously.[8]

S1.6. Kinetic analysis of the most promising dehydratases toward D-glycerate

The three most promising dehydratases were further characterized in order to determine their kinetic parameters at 50 °C. Reaction solutions contained $MgCl_2$ 5 mM with different concentrations of D,L-glycerate in HEPES 50 mM, pH 7.5, at 50 °C. The reactions were started by adding the respective enzyme. Two independent repeats were performed. Aliquots were taken after 0, 15 and 60 minutes of the reaction. Centrifugations of the diluted aliquots at 12000 xg for 3 min were performed to stop the reactions. Formation of pyruvate was quantified using HPLC as described above. A linear increase of pyruvate was observed over the course of 1 h for all enzymes. Catalytic activity (k_{cat}) is defined as the formation of 1 µmol of pyruvate or KDG per µmol of enzyme per second.

S1.7. Biotransformation of ethanol from D-glucose using *Pu*DHT

The synthesis of ethanol from D-glucose was used as the model cascade to test the applicability of *Pu*DHT in an enzymatic chain reaction. We adapted the previously established enzyme cascade.[9] In a preliminary run, we tested the effect of the *Pu*DHT concentration on the ethanol yield. We used *Bs*GDH (20 U/ml), *Pt*KdgA (2 U/ml), increasing amounts of *Pu*DHT, engineered *Ta*ALDH (2 U/ml), PDC-Var.2 (12 U/ml) and *Bst*ADH (12 U/ml) in HEPES 50 mM, pH 7.5, at 50 °C containing D-glucose 25 mM, thiamine diphosphate (TDP) 0.1 mM, $MgCl_2$ 5 mM, and NAD^+ 5 mM (see Table S5 for enzyme abbreviations). The cascade was initiated by the addition of D-glucose. The negative control contained the same reaction mix, but without D-glucose. The cascade was run in 250 µl aliquots in 1.5 ml Eppendorf tubes at 50 °C and for 16 h using an Eppendorf shaker (500 rpm, ThermoMixer C, Eppendorf, Germany). For analysis of the samples, 200 µl of each solution was withdrawn and diluted 50:50 with H_2SO_4 5 mM. The mixtures were then filtered through 10 KDa filters. The filtrates were subjected to HPLC analysis. D-glucose, D-gluconate, D-glyceraldehyde, D-glycerate, pyruvate and ethanol were separated using an ion-exclusion column (Rezex ROA-Organic Acid

H+(8%), Phenomenex, Germany), run isocratically using H_2SO_4 2.5 mM at 70 °C for 30 min (see Figure S6 for separation of substrates, intermediates, and products). D-gluconate and pyruvate were not detected at the end of the biotransformations. Two independent repeats were carried out. The results are presented in Figure S3. No ethanol formation in the negative controls was able to be detected.

After identifying the optimum concentration of *Pu*DHT (1.8 mg/ml), three different mixtures were made which consisted of the same components as described above. Each mixture was then divided into seven 1.5 ml Eppendorf tubes. The reactions were started immediately. After different reaction times (0, 0.5, 1, 2, 4, 6, 12 h), each tube from the three different starting mixtures (representing three independent repeats) was subjected to the same pre-treatment prior to analysis by HPLC. The results are presented in Figure 2.

S1.8. Conversion of pure and crude glycerol to pyruvate using *Pu*DHT

The conversion of glycerol to pyruvate was performed at 25 °C. The reaction mixture contained *Sc*Aldo (0.36 U/ml), *Pu*DHT (0.53 U/ml), catalase from *Corynebacterium glutamicum* (1000 U/ml), and glycerol 50 mM in HEPES 250 mM, pH 8. The catalase was purchased from Sigma Aldrich (Darmstadt, Germany). The reaction was started by addition of *Sc*Aldo. The biocatalysis was performed in 2.5 ml solutions in 50 ml Falcon tubes placed on a nutating shaker (VWR Rocking Platform, VWR, Germany) at 50 rpm. At certain time points, 100 µl of the solution was withdrawn and diluted 50:50 with H_2SO_4 5 mM. The reaction was stopped by enzyme removal using 10 KDa filter column. Because 10 KDa filter columns contained glycerol as a storage solution, the columns were washed with 4 x 500 µl ddH_2O prior to use. The eluates from different sampling times were analyzed using the same HPLC method as for the ethanol cascade. The negative control without glycerol did not yield any pyruvate. Two independent repeats were performed. The results are presented in Figure 3.

For the conversion of crude glycerol to pyruvate, the crude glycerol was first subjected to pre-treatment. After dissolving the crude glycerol in water, the pH of the solution was lowered to 2 with HCl. At pH 2, the solution produced two distinct fractions. The insoluble fraction was removed by filtration through standard filter paper. The soluble fraction was then neutralized with KOH. The initial concentration of pre-treated crude glycerol was measured with HPLC. No D-glyceraldehyde, D-glycerate, or pyruvate could be observed (Figure S7). For the biotransformation of crude glycerol, the same reaction set-up described above was used. The only differences were that the amount of *Sc*Aldo was doubled, and the amount of glycerol was adjusted to 25 mM. At different time points, 100 µl of the solution was subjected to further analysis by HPLC. The results are presented in Figure 4.

REFERENCES

(1) Vázquez-Figueroa, E.; Chaparro-Riggers, J.; Bommarius, A. S. Development of a Thermostable Glucose Dehydrogenase by a Structure-Guided Consensus Concept.Pdf. *ChemBioChem* **2007**, 2295 – 2301.

(2) Reher, M.; Fuhrer, T.; Bott, M.; Scho, P.; Ju, F. The Nonphosphorylative Entner-Doudoroff Pathway in the Thermoacidophilic Euryarchaeon Picrophilus Torridus Involves a Novel 2-Keto-3-Deoxygluconate- Specific Aldolase □. *J. Bacteriol.* **2010**, *192*, 964–974.

(3) Sutiono, S.; Satzinger, K.; Pick, A.; Carsten, J.; Sieber, V. To Beat the Heat – Engineering of the Most Thermostable Pyruvate Decarboxylase to Date †. *RSC Adv.* **2019**, *9*, 29743–29746.

(4) Guagliardi, A.; Martino, M.; Iaccarino, I.; Rosa, M. D. E.; Bartoluccil, S. Purification and Characterization of the Alcohol Dehydrogenase from a Novel Strain of Bacillrcs Stearotherrnophilus Growing at 70 ° C. *Int. J. Biochem. Cell Biol.* **1996**, *28*, 239–246.

(5) Gerstenbruch, S.; Wulf, H.; Mußmann, N.; Connell, T. O.; Maurer, K.; Bornscheuer, U. T. Asymmetric Synthesis of D -Glyceric Acid by an Alditol Oxidase and Directed Evolution for Enhanced Oxidative Activity towards Glycerol. *Appl. Microbiol. Biotechnol.* **2012**, *96*, 1243–1252.

(6) Bommarius, A. S.; Paye, M. F. Stabilizing Biocatalysts. *Chem. Soc. Rev.* **2013**, *42*, 6534–6565.

(7) Bommarius, A. S.; Schwarm, M.; Drauz, K. Comparison of Different Chemoenzymatic Process Routes to Enantiomerically Pure Amino Acids. *Chimica* **2001**, *55*, 50–59.

(8) Sutiono, S.; Carsten, J.; Sieber, V. Structure-Guided Engineering of α-Keto Acid Decarboxylase for the Production of Higher Alcohols at Elevated Temperature. *ChemSusChem* **2018**, *11*, 3335–3344.

(9) Guterl, J. K.; Garbe, D.; Carsten, J.; Steffler, F.; Sommer, B.; Reiße, S.; Philipp, A.; Haack, M.; Rühmann, B.; Koltermann, A.; Kettling, U.; Brück, T.; Sieber, V. Cell-Free Metabolic Engineering: Production of Chemicals by Minimized Reaction Cascades. *ChemSusChem* **2012**, *5*, 2165–2172.

3.4 Characterization of highly active 2-keto-3-deoxy-L-arabinonate and 2-keto-3-deoxy-D-xylonate dehydratases in terms of biotransformation of hemicellulose sugars to chemicals.

L-Arabinose and D-xylose are two major constituents of hemicellulose. The utilization of hemicellulose as an alternative carbon source has attracted increased attentions, in particular for the production of biofuels and chemicals (second-generation approach). Unlike D-glucose, these two sugars are not readily fermentable by most of the industrial strains. Three pathways to metabolize L-arabinose and D-xylose have been described in microorganisms. One of the pathway—the Weimberg pathway—can convert L-arabinose and D-xylose to α-ketoglutarate (α-KG) *via* four enzymatic steps. In recent years, the Weimberg pathway has been subsequently modified toward the production of 1,4-butanediol (BDO), glutamate, and mesaconate. Most of these compounds are still obtained from fossil resources. One of the key enzymes in the Weimberg pathway—L- and D-2-keto-3-deoxypentonate dehydratases (KdpDs), however, have not been characterized in detail in literature, and thus became the focus of this study.

From kinetic characterizations of a number of L-KdpDs and D-KdpDs studied, it was shown that L-KdpD from *Cupriavidus necator* (*Cn*) and *Herbaspirillum seropedicae* (*Hs*) and D-KdpD from *Pseudomonas putida* (*Pp*) and *H. seropedicae* have the highest activity (k_{cat}) and catalytic efficiencies (k_{cat}/K_M). These enzymes were then subjected to kinetic and thermodynamic stability measurement. The study revealed that *Cn*L-KdpD and *Pp*D-KdpD have higher kinetic stability than the two dehydratases from *H. seropedicae*. This study also showed that albeit both classes of dehydratases catalyze very similar reactions, both enzyme classes are rather different. D-KdpDs require Mg^{2+} for their activity and stability, whereas L-KdpDs do not. *Pp*D-KdpD also showed an optimum pH in an alkaline condition (8-9), while *Cn*L-KdpD showed an optimum pH under a slightly acidic condition (6-7). *Cn*L-KdpD was not inhibited by L-arabinose and L-arabinonate, while *Pp*D-KdpD was inhibited by D-xylonate (I_{50}: 75 mM), but not by D-xylose. I_{50} is defined as the concentration of an inhibitor to give 50% inhibition effect. Both enzyme also showed to be compatible for high substrate load conversion by being able to convert 500 mM L-KDP and D-KDP reaching >95% conversion in 2 h for *Cn*L-KdpD and >90% conversion in 4 h for *Pp*D-KdpD, respectively. Both enzymes are rather specific as each enzyme was not able to convert the other non-preferred stereoisomer. A combination of both enzymes, however, was able to convert 500 mM racemic D,L-KDP to reach >83% in 4 h.

The author with the co-author Prof. Sieber conceived and designed the study. The author performed the experiments and analyzed the data. All authors wrote and approved the final version of the manuscript. This work was accepted for publication in Applied Microbiology and Biotechnology.

Characterization of highly active 2-keto-3-deoxy-L-arabinonate and 2-keto-3-deoxy-D-xylonate dehydratases in terms of biotransformation of hemicellulose sugars to chemicals

Samuel Sutiono, Prof. Bettina Siebers, Prof. Volker Sieber

Applied Microbiology and Biotechnology

2020

DOI: 10.1007/s00253-020-10742-5

Applied Microbiology and Biotechnology
https://doi.org/10.1007/s00253-020-10742-5

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

Characterization of highly active 2-keto-3-deoxy-L-arabinonate and 2-keto-3-deoxy-D-xylonate dehydratases in terms of the biotransformation of hemicellulose sugars to chemicals

Samuel Sutiono[1] · Bettina Siebers[2] · Volker Sieber[1,3,4,5]

Received: 8 April 2020 / Revised: 4 June 2020 / Accepted: 9 June 2020

Abstract

2-keto-3-L-arabinonate dehydratase (L-KdpD) and 2-keto-3-D-xylonate dehydratase (D-KdpD) are the third enzymes in the Weimberg pathway catalyzing the dehydration of respective 2-keto-3-deoxy sugar acids (KDP) to α-ketoglutaric semialdehyde (KGSA). The Weimberg pathway has been explored recently with respect to the synthesis of chemicals from L-arabinose and D-xylose. However, only limited work has been done toward characterizing these two enzymes. In this work, several new L-KdpDs and D-KdpDs were cloned and heterologously expressed in *Escherichia coli*. Following kinetic characterizations and kinetic stability studies, the L-KdpD from *Cupriavidus necator* (*Cn*L-KdpD) and D-KdpD from *Pseudomonas putida* (*Pp*D-KdpD) appeared to be the most promising variants from each enzyme class. Magnesium had no effect on *Cn*L-KdpD, whereas increased activity and stability were observed for *Pp*D-KdpD in the presence of Mg^{2+}. Furthermore, *Cn*L-KdpD was not inhibited in the presence of L-arabinose and L-arabinonate, whereas *Pp*D-KdpD was inhibited with D-xylonate (I_{50} of 75 mM), but not with D-xylose. Both enzymes were shown to be highly active in the one-step conversions of L-KDP and D-KDP. *Cn*L-KdpD converted > 95% of 500 mM L-KDP to KGSA in the first 2 h while *Pp*D-KdpD converted > 90% of 500 mM D-KDP after 4 h. Both enzymes in combination were able to convert 83% of a racemic mixture of D,L-KDP (500 mM) after 4 h, with both enzymes being specific toward the respective stereoisomer.

Key points

- *L-KdpDs and D-KdpDs are specific toward L- and D-KDP, respectively.*
- *Mg^{2+} affected activity and stabilities of D-KdpDs, but not of L-KdpDs.*
- *CnL-KdpD and PpD-KdpD converted 0.5 M of each KDP isomer reaching 95 and 90% yield.*
- *Both enzymes in combination converted 0.5 M racemic D,L-KDP reaching 83% yield.*

Keywords L-Arabinose · D-Xylose · Dehydratase · Weimberg · Biotransformation · Chemicals

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00253-020-10742-5) contains supplementary material, which is available to authorized users.

✉ Volker Sieber
sieber@tum.de

[1] Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, Schulgasse 16, 94315 Straubing, Germany

[2] Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Centre for Water and Environmental Research (CWE), University of Duisburg-Essen, Universitätsstraße 5, 45117 Essen, Germany

[3] Catalytic Research Center, Technical University of Munich, Ernst-Otto-Fischer-Straße 1, 85748 Garching, Germany

[4] Straubing Branch BioCat, Fraunhofer IGB, Schulgasse 11a, 94315 Straubing, Germany

[5] School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Copper Road, St. Lucia 4072, Australia

Published online: 21 June 2020

Introduction

Petrochemicals have been utilized by humans for almost a century in producing chemicals and other building blocks that are used in everyday life. The utilization of fossil resources, which are the primary source of petrochemicals, has long been a concern as regards sustainability and climate change. The use of plant-derived raw materials has been proposed in order to alleviate our dependency on petrochemicals. In the first generation, starchy materials were utilized, mainly for bioethanol fermentation and lactic acid production (Naik et al. 2010; Mohr and Raman 2015). This approach, while being more environmentally friendly, is still under debate, in particular in relation to competition with food sources (Naik et al. 2010; Mohr and Raman 2015; Rulli et al. 2016). The second-generation bio-production of fuel and chemicals has been developed in order to avoid the potential of competition. In this approach, lignocellulosic biomass is utilized. With annual production of dried biomass exceeding 220 billion tons, it provides us with an enormous amount of truly renewable raw materials (Huang and Fu 2013; Isikgor and Becer 2015).

Unlike the first-generation raw materials, lignocellulosic biomass consists of the following three different components: cellulose, hemicellulose, and lignin. Cellulose can be hydrolyzed to yield D-glucose. Fermentation of D-glucose to chemicals is a straightforward process because it is a universal substrate for most industrially relevant microorganisms. During degradation of the second carbohydrate polymer, hemicellulose will result in several pentose and hexose sugars, of which D-xylose and L-arabinose are two major constituents (Isikgor and Becer 2015). In contrast to D-glucose, these hemicellulose sugars are not easily fermentable. Several works have thus been focused on introducing pentose metabolism pathways in industrially relevant hosts in order to establish hexose and pentose co-fermentation and, therefore, to increase the feasibility of biomass utilization (Fernandes and Murray 2010; Chandel et al. 2011).

There are three major pathways found in nature that metabolize D-xylose and L-arabinose. The first pathway is the pentose phosphate pathway, which is accessed after isomerization of the pentose sugars from the aldolase to the ketose form. The second and third pathways are called non-phosphorylated oxidative pathways (Jagtap and Rao 2018; Valdehuesa et al. 2018). In the latter pathways, pentose sugars are oxidized to their respective sugar acids. Dehydration at position C2 and C3 results in 2-keto-3-deoxy-D-xylonate and 2-keto-3-deoxy-L-arabinonate, respectively. At this point, the two pathways are branched. In one pathway, aldolase will split the respective intermediates to pyruvate and glycoaldehyde. This pathway is known as the Dahms pathway (Stephen Dahms 1974). The alternative takes place via further dehydration of the 2-keto-3-deoxy-D-xylonate and 2-keto-3-deoxy-L-arabinonate at positions C4 and C5 by means of a dehydratase producing α-ketoglurate semialdehyde (KGSA). Oxidation of the terminal aldehyde of KGSA will yield α-ketoglutarate (α-KG), which is an intermediate of the TCA cycle. This pathway is known as the Weimberg pathway (Weimberg 1961). Because the intermediates at the branch point (2-keto-3-deoxy-D-xylonate and 2-keto-3-deoxy-L-arabinonate) lose their chirality at C2 and C3, all D-pentose sugars (D-xylose, D-ribose, D-arabinose, and D-lyxose) will result in the same intermediate after undergoing one oxidation followed up by the subsequent dehydration at C2 and C3. This is also true for their respective L-sugars. D and L functionality is maintained because it is dictated by the position of a hydroxyl group at C4. Therefore, 2-keto-3-deoxy-D-xylonate and 2-keto-3-deoxy-L-arabinonate are referred to herein as 2-keto-3-deoxy-D-pentonate (D-KDP) and 2-keto-3-deoxy-L-pentonate (L-KDP), respectively, as used in the nomenclature from a previous study (Watanabe et al. 2006). The enzymes that catalyze the dehydration reactions are thus called D-KdpD and L-KdpD, respectively.

Introduction of the Weimberg pathway in industrially relevant microorganisms will allow the production of α-KG and its subsequent derivatives in the TCA cycles from D-xylose and L-arabinose (Tai et al. 2016; McClintock et al. 2017). A-KG is a promising building block for the synthesis of heterocyclic compounds (Stottmeister et al. 2005). A-KG has also widely been used in animal feed and for humans as a food and medicine additive (Wu et al. 2016; Liu et al. 2018). The longevity effect of α-KG in *Caenorhabditis elegans* was also recently shown (Chin et al. 2014). The Weimberg pathway has also been explored further in recent years in order to determine its potential toward the synthesis of non-natural chemicals from the hemicellulose sugars. Several platform chemicals that have been produced from modified Weimberg pathways include 1,4-butanediol (BDO), mesaconate, glutamate, and succinate (Fig. 1) (Bai et al. 2016; Tai et al. 2016; McClintock et al. 2017). These are important platform chemicals for the production of several polymers. The dehydration of BDO results in tetrahydrofuran, which is a building block for Spandex and an industrial solvent. Decarboxylation of mesaconate will yield methacrylic acid, which is a precursor of vinyl ester resin (Yadav et al. 2018). Succinate has been utilized in the production of polyester, surfactants, detergent, food additives, and pharmaceuticals (Saxena et al. 2016).

In order to design efficient pathways in a microorganism, the kinetic characterization of responsible enzymes is indispensable. This parameter is equally important when an in vitro approach is preferred. In vitro approaches to producing chemicals have gained increasing attention in recent years as an alternative to in vivo (Guterl and Sieber 2013; Dudley et al. 2015; Claassens et al. 2019). In contrast to the other enzymes of the Weimberg pathway, D-KdpD and L-KdpD have only been characterized partially. It has been reported that

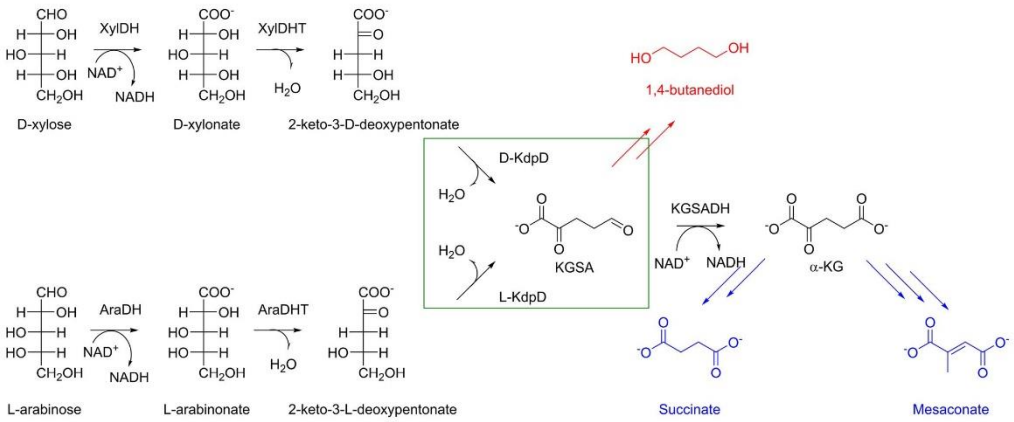


Fig. 1 The Weimberg pathway and its subsequent modification for the conversion of L-arabinose and D-xylose to added-value compounds/chemicals. The natural pathway is shown in black. The conversions catalyzed by L-KdpD and D-KdpD that are the focus of this study are highlighted by a green box. The pathways are adapted from previous studies (Bai et al. 2016; Tai et al. 2016)

Caulobacter crescentus and *Paraburkholderia xenovorans* are able to encode functional D-KdpD. However, their activities are rather low (k_{cat}: 0.5 s^{-1} and K_M: 1.9 mM for *Cc*D-KdpD; k_{cat}: 4.7 s^{-1} and K_M: 9 mM for *Px*D-KdpD) (Tai et al. 2016). With regard to *Cc*D-KdpD in particular, its activity seems to be much lower than the two preceding enzymes of the Weimberg pathway in *C. crescentus* (D-xylose dehydrogenase and D-xylonate dehydratase) (Andberg et al. 2016). Regarding L-KdpD, one elegant work from the 1960s previously characterized a functional enzyme from *Pseudomonas saccharopila* (also known as *Pelomonas saccharophila*) (Stoolmiller and Abeles 1966). Possible catalytic mechanisms have also been proposed on the basis of isotopes and inhibitor studies (Portsmouth et al. 1967). However, no protein sequence is available, which one would be necessary for easy heterologous protein expression. The protein sequence of a homologous enzyme from *Azospirillum brasilense* (*Ab*L-KdpD) has been reported recently and exhibits an activity of 20 U/mg (k_{cat} ~ 11 s^{-1}) (Watanabe et al. 2006). Another homologous enzyme from *Bukholderia multivorans* (*Bm*L-KdpD) has also been reported, but it exhibits considerably lower activity (k_{cat} of 0.2 s^{-1}) (Tai et al. 2016).

Therefore, several D-KdpDs and L-KdpDs have been characterized in detail in this study. *Cc*D-KdpD and *Ab*L-KdpD were used as guide sequences to mine homologous enzymes. A chemoenzymatic approach was applied to synthesize the substrates D-KDP and L-KDP from D-xylose and L-arabinose, respectively. In addition to kinetic parameters, pH optimum, temperature stability, and inhibitory effects of substrates and intermediates in the Weimberg pathway were determined. The best enzyme variants with respect to activity and stability were applied in the conversion of high substrate load (> 0.5 M).

Materials and methods

Cloning, expression, and enzyme purification

Caulobacter crescentus (DSM 4727), *Pseudomonas putida* KT 2440 (DSM 6125), and *Cupriavidus necator* H16 (DSM 428) were purchased from DSMZ (Germany). All cells were grown according to the protocol described by DSMZ for isolating their gDNA. The isolation of gDNA was performed using the DNeasy UltraClean Microbial Kit (Qiagen, Germany). gDNA of *Herbaspirillum seropedicae* Z67 (DSM 6445) was purchased directly from DSMZ. D-kdpD and L-kdpD genes were amplified from the corresponding gDNAs using a two-step PCR protocol (PCR without annealing step). In brief, the PCR mix consisted of the following: Phusion high-fidelity polymerase 1 U (NEB, Germany), 2 mM of dNTPs (VWR; Germany), 0.5 mM of forward and reverse primers, 25 ng of gDNA, 1x GC buffer (NEB, Germany), and ddH_2O up to 50 μL. The PCR reactions were as follows: 98 °C for 30 s, 30 cycles at 98 °C for 10 s, 72 °C for 1 min, and 72 °C for 5 min, and an indefinite hold at 16 °C. Upon PCR, the products obtained were separated using agarose gel electrophoresis. Bands with correct sizes were excised and purified using the NucleoSpin Gel and PCR Clean up kit (Machery and Nagel, Germany). Purified DNAs were then digested and ligated into either pET28a or pET24a (Invitrogen, Germany) to give corresponding N- or C-terminal hexa-histidine tags. Another kdpD1 gene from *Caulobacter crescentus* (*Cc*D-kdpD1), D-kdpD gene from *Paraburkholderia xenovorans* (*Px*D-kdpD), as well as L-kdpD genes from *Azospirillum brasilense* (*Ab*L-kdpD) were ordered from ATG Biosynthesis (Germany) as optimized codons for expression in *E. coli*. The *Cc*DkdpD1 gene was synthesized according to the sequence described in the previous study (Tai et al. 2016). These

genes were also cloned to either pET28a or pET24a using appropriate restriction enzymes. The ligated plasmids were used to transform *E. coli* BL21 (DE3) (Invitrogen, Germany). Forward and reverse primers used to amplify the genes as well as restriction enzymes used for cloning to pET vectors are presented in Table S1 in the Supporting Information.

To express the enzymes, a single colony of *E. coli* BL21 (DE3) possessing the correct gene was grown overnight in 10 mL LB medium supplemented with 100 μg/mL kanamycin in a 100 mL baffled flask at 37 °C, 150 rpm. The overnight culture was then transferred to a 2 L baffled flask containing 500 mL autoinduction media supplemented with 100 μg/mL kanamycin (Studier 2005). The culture was incubated at 30 °C, 120 rpm for 16 h. The cells were pelleted by centrifugation at 4000 ×*g* for 15 min. After decanting the supernatant, the cell pellet was transferred to a 50 mL falcon tube for subsequent purification.

A binding buffer (50 mM KPi pH 8, 20 mM imidazole, 500 mM NaCl, and 10 vol% glycerin) was added up to 40 mL in the 50 mL falcon tube. A total of 5 μg/mL DNase (AppliChem) and 2 mM $MgCl_2$ were added to the cell suspension. Cells were disrupted by sonication (80% and cycle 0.5 s) in ice for 20 min. The solution was then cleared by centrifugation at 20,000 ×*g* for 30 min. The supernatant was filtered through a 0.45 μm cellulose filter (VWR, Germany) before application to an Äkta purifier (GE Healthcare, Germany). A 5 mL HisTrap FF Crude (GE Healthcare, Germany) was used for purification of His-tagged proteins. The column was washed with 30 mL binding buffer to remove *E. coli* proteins. An elution buffer (50 mM KPi, pH 8, 500 mM imidazole, 500 mM NaCl, and 10 vol% glycerin) was used to elute the His-tagged protein from the column. The buffer was exchanged using a HiPrep desalting column (GE Healthcare, Germany). A total of 50 mM HEPES pH 7.5 was used as the final buffer for all enzymes in this study. After the buffer exchange, all enzymes were flash-frozen in liquid nitrogen and stored at − 80 °C until further use. The purity and size of the enzymes were analyzed via SDS-PAGE.

Synthesis of 2-keto-3-deoxy-L-pentonate and 2-keto-3-deoxy-D-pentonate

A chemoenzymatic approach was applied for synthesizing the substrates (L-KDP and D-KDP) from their respective aldoses. This approach was modified from previously published work (Sperl et al. 2016). L-Arabinose and D-xylose were oxidized separately using a 0.5% gold catalyst (Evonik, Germany) in the presence of saturated oxygen. Oxidation was performed in an automatic titrator (SI Analytics, Germany). The pH and temperature were set to 8 and 50 °C, respectively. The gold catalyst was used at a ratio of 0.033 g/g sugar. The oxygen flow rate used was 40 mL/min, and the reaction was mixed using the magnetic stirrer of the automatic titrator. The oxidation was started by the addition of the gold catalyst. After the oxidation was finished (yield > 97% for L-arabinose and D-xylose), the catalyst was pelleted by centrifugation at 4000 ×*g* for 15 min. The sugar acid (L-arabinonate and D-xylonate) solutions were filtered through a 0.45 μm cellulose filter.

Dehydration of the sugar acids was performed as follows. *Cc*XylDHT was used to dehydrate D-xylonate and *Rl*AraDHT to dehydrate L-arabinonate (Andberg et al. 2016). The expression of each enzyme was performed as described previously (Sutiono et al. 2020). Dehydration of 37.5 mmol of the sugar acids was achieved by using 5 mg of each enzyme. Magnesium chloride (5 mM) was added to the reaction as a cofactor for both enzymes. No buffer was used in the reaction. After an overnight reaction, > 99% conversion of D-xylonate and L-arabinonate was achieved. Given that this was an HPLC analysis and no sugar acid or other peaks were observed beyond a single product peak, it was assumed that > 99% conversion of the sugar acids resulted only in 2-keto-3-deoxy-D-pentonate and 2-keto-3-deoxy-L-pentonate. The final product of the dehydration reaction was a clear solution for both KDPs. The solutions were filtered using an Amicon Ultra Centrifugal Filters 10K (Sigma Aldrich, Germany) to remove the enzymes. The solutions were stored at − 20 °C until further use. No apparent substrate decomposition after being stored > 3 months was observed by HPLC. The standard HPLC method to detect sugar acids and their subsequent dehydrated product was described in earlier research (Guterl et al. 2012).

Kinetic characterization of the dehydratases

To determine dehydratase activity, a coupled-assay with *Pp*KGSADH as auxiliary enzyme was used to link the oxidation of KGSA to α-KG using NAD^+, as was described in a previous study (Beer et al. 2017). The measurement was performed in a 96-well F-bottom plate (Greiner Bio-One, Germany). 20 μL of a diluted enzyme was added to a 96-well plate. A total of 180 μL of reaction solution containing different concentrations of either D-KDP or L-KDP with 0.5 U/mL *Pp*KGSADH, 2 mM NAD^+, 5 mM $MgCl_2$, 50 mM HEPES pH 7.5 (end concentration) were added subsequently. The measurement was performed with a Multiskan spectrophotometer (Thermo Scientific, Germany) by measuring the development of NADH at 340 nm at 25 °C. Enzyme activity (k_{cat}) was defined as the number of NADH molecules formed per molecule of enzyme per second. Measurements were performed in triplicate.

Effect of Mg^{2+} on activity, kinetic, and thermodynamic stability

Two different enzyme stock solutions were used to determine activities. The first stock was a standard enzyme solution,

whereas the second stock was an enzyme solution that had been pre-incubated with 5 mM EDTA at 25 °C for 1 h. A total of 20 μL of a diluted enzyme solution from the first stock was added to a 96-well F-bottom plate. A total of 180 μl of reaction mix containing 5 mM of each substrate (L-KDP for *Cn*L-KdpD and *Hs*L-KdpD; D-KDP for *Pp*D-KdpD and *Hs*D-KdpD), 0.5 U/mL *Pp*KGSADH, 2 mM NAD^+, and 50 mM HEPES pH 7.5, without and with 5 mM $MgCl_2$ was added. The development of NADH was monitored at 340 nm using Multiskan spectrophotometry. The same reaction mixture without $MgCl_2$ was also used to measure activity of the second enzyme stock. All measurements were performed at 25 °C.

To measure $T_{50}{}^{1h}$ (temperature at which an enzyme loses 50% of its initial activity after 1 h incubation), 20 μL of enzyme solution was transferred to a 96-PCR plate (Brand, Germany). The enzyme solution contained 1 mg/mL enzyme in 50 mM HEPES pH 7.5 with 5 mM EDTA, with or without 5 mM $MgCl_2$. The PCR plate was incubated in a thermal cycler (MyCycler, Bio-Rad, Germany) at gradient temperatures from 30 to 55 °C. The incubation was performed for 1 h. Following this heating step, the enzyme solution was diluted accordingly with 50 mM HEPES pH 7.5. A total of 20 μL of diluted solution was transferred to a 96-well plate. A total of 180 μL of reaction solution containing 5 mM of each substrate (L-KDP for *Cn*L-KdpD and *Hs*L-KdpD and D-KDP for *Pp*D-KdpD and *Hs*D-KdpD), 0.5 U/mL *Pp*KGSADH, 2 mM NAD^+, 2.5 mM $MgCl_2$, 50 mM HEPES pH 7.5 (end concentration) was added. The development of NADH was monitored at 340 nm using a Multiskan spectrophotometer. The activities obtained were normalized to the highest value. Three independent repeats were performed.

The thermodynamic stability is displayed as the melting temperature (T_m) and was analyzed by the Thermofluor assay (Boivin et al. 2013; Sutiono et al. 2018). The buffer used in the assay was 50 mM HEPES pH 7.5 with 5 mM EDTA, with or without 5 mM $MgCl_2$.

Determination of pH optimum

A total of 20 μL of diluted enzyme (*Cn*L-KdpD with N-terminal His-tag or *Pp*D-KdpD) was added to a 96-well plate. A total of 180 μL reaction solution containing 5 mM of each substrate (L-KDP for *Cn*L-KdpD and D-KDP for *Pp*D-KdpD), 0.5 U/mL *Pp*KGSADH, 2 mM NAD^+, 5 mM $MgCl_2$, and 50 mM of buffer with corresponding pH (end concentration) was added. The reaction was monitored at 340 nm using Multiskan spectrophotometer to detect NADH formation. For pH 6 to 8, KPi buffer and for pH 7 to 9, Tris-HCl was used. Activities obtained were normalized to the highest value. Measurements were performed in triplicate. *Pp*KGSADH was confirmed to be active in the entire pH range tested (data not shown).

Effect of substrates and selected intermediates of the Weimberg pathway

D-Xylose and D-xylonate were used to check their inhibitory effects on *Pp*D-KdpD. L-Arabinose and L-arabinonate were used for *Cn*L-KdpD with a N-terminal fused His-tag. The experiments were performed as follows. 20 μL of diluted enzyme (*Pp*D-KdpD and *Cn*L-KdpD) was added to a 96-well plate along with 180 μL of reaction solution containing 5 mM of each substrate (L-KDP for *Cn*L-KdpD and D-KDP for *Pp*D-KdpD), 0.5 U/mL *Pp*KGSADH, 2 mM NAD^+, 5 mM $MgCl_2$, 50 mM HEPES pH 7.5, and increasing concentration of the respective inhibitors (end concentration). The formation of NADH was monitored at 340 nm using a Multiskan spectrophotometer. Measurements were performed in triplicate. The lack of inhibition of *Pp*KGSADH by the highest compound concentration tested (100 mM) was confirmed.

One-step conversion of high substrate loads

Regarding the *Cn*L-KdpD with a N-terminal His-tag and *Pp*D-KdpD, this was analyzed if high substrate load (L-KDP and D-KDP) concentrations would have an inhibitory effect. Each enzyme was used to convert each substrate in a reaction mix. The reaction mix contained 0.1 mg/mL enzyme, 5 mM $MgCl_2$, 500 mM of respective substrate in 50 mM HEPES pH 7.5. The reaction mix was incubated in a thermoshaker (ThermoMixer C, Eppendorf, Germany) at 300 rpm and 25 °C. At certain times, 5 μL of the solution was removed from the reaction mix and diluted using 995 μL of ddH_2O. A total of 500 μL of the diluted solution was transferred to a 10 KDa centrifugal filter (VWR, Germany), and the filter column was centrifuged at 12,000 ×*g* for 2 min. This step was used to stop the reaction by removing protein from the solution. The filtrate was used to analyze the formation of KGSA from either D-KDP or L-KDP by a previously established method (Beer et al. 2017). In brief, Metrosep A Supp 16–250 was used as a stationary phase, and 30 mM ammonium bicarbonate pH 10.4 was used as the mobile phase. The isocratic flow was set to 0.2 mL/min, and D-KDP, L-KDP, and KGSA were monitored by a UV detector at 210 nm. In this system, D-KDP and L-KDP showed the same retention time. KGSA eluted directly after D-KDP and L-KDP (Fig. S4 in the Supporting Information).

Results

Cloning, expression, and purification of the dehydratases

The position of a hexa-histidine tag (His-tag) sometimes has an effect on enzyme yield, stability, and activity. Therefore,

*Cn*L-KdpD was cloned as an N-terminally as well as a C-terminally fused His-tag protein. After expression and purification, the protein yield of the N-Histag variant was almost 3 times higher (data not shown). In subsequent L-KdpDs, all variants were therefore cloned using N-terminally fused His-tags (Table 1). All variants were expressed using autoinduction media. A total of 20 to 50 mg of protein were able to be obtained from the 500 mL culture (except for *Ab*L-KdpD). To improve expression of *Ab*L-KdpD, a more enriched media (Terrific broth media) was used (Watanabe et al. 2006). From 500 mL culture, only 5 mg of this enzyme having lower protein purity were able to be obtained (Fig. S1). As a result, due to the difficulty with expression and the other L-KdpDs demonstrated having a similar magnitude of activity (Table 2), this enzyme would be of less interest with regard to application.

*Cc*D-KdpD1 was used as a template for analyzing which His-tag variant (N- or C-terminally fused His-tag) would be better with respect to activity and stability of D-KdpDs. Both His-tag variants of *Cc*D-KdpD1 were able to be expressed and purified, and similar protein yields were also obtained. However, the N-terminal variant precipitated rather quickly during enzyme characterization. With respect to all D-KdpDs, all variants therefore were cloned using C-terminally fused His-tags (Table 1). We could not obtain reasonable yield for *Vp*D-KdpD, therefore this variant was not pursued further.

Kinetic characterization of the respective dehydratases

Regarding the enzymes purified, L-KdpD showed higher a preference for L-KDP than for D-KDP by at least 45-fold (Table 2). The activity (k_{cat}) and K_M-values of all L-KdpDs are also of the same magnitude as reported for the first L-KdpD from *Pseudomonas saccharophila* published in the late 1960's (Stoolmiller and Abeles 1966). Regarding D-KdpDs, all enzymes exhibited a higher a preference for D-KDP over L-KDP by over 60-fold, except for *Cn*D-KdpD, which showed only a 35-fold higher preference. The activities of *Cc*D-KdpD1 and *Px*D-KdpD were considerably higher than reported in earlier research. The significant difference is presumably due to the fact that the authors in the previous study expressed *Cc*D-KdpD1 and *Px*D-KdpD with a N-terminally fused His-tag (Tai et al. 2016). The in vivo production of chemicals using *Cc*D-KdpD1 and *Px*D-KdpD from D-xylose has been demonstrated and no accumulation of D-KDP was reported. This could indicate that the enzyme introduced in *E. coli* did not bear any His-tag. Two enzymes from each class showing the highest activity and catalytic efficiency were then selected for further studies. *Cn*L-KdpD (Nhis) and *Hs*L-KdpD were chosen for L-KdpD and *Pp*D-KdpD and *Hs*KdpD for D-KdpD conversion.

Effect of magnesium on activity, and the kinetic and thermodynamic stabilities of the most promising L-KdpDs and D-KdpDs

Magnesium has been reported as playing a role in D-KdpD from *Sulfolobus solfataricus* (*Ss*D-KdpD) (Brouns et al. 2006; Brouns et al. 2008). The cation is also observed in the crystal structure of *Ss*D-KdpD. However, *Ps*L-KdpD, which is the enzyme catalyzing the respective dehydration of the L-stereoisomer, was reported to be active without the addition of magnesium ions (Stoolmiller and Abeles 1966; Portsmouth et al. 1967). Magnesium ions have also been reported for the activity and stability of several other lyases, e.g., Fe-S dependent dehydratases and decarboxylases (De La Plaza et al. 2004; Andberg et al. 2016). This work studied the effect of magnesium ions on the most promising L-KdpDs and D-KdpDs. There was no apparent effect of magnesium on the activity of *Cn*L-KdpD and *Hs*L-KdpD (Fig. 2a). Nor did pretreatment of the enzymes with EDTA affect their activity. A different result was observed for *Pp*D-KdpD and *Hs*D-KdpD. When

Table 1 List of L-KdpDs and D-KdpDs cloned in this study

Microorganism	Protein	NCBI accession number
Caulobacter crescentus	*Cc*D-KdpD1	WP_010918708.1
	*Cc*D-KdpD2	WP_012640070.1
Paraburkholderia xenovorans	*Px*D-KdpD	WP_011494434.1
Pseudomonas putida KT12440	*Pp*D-KdpD	WP_010953745.1
Herbaspirillum seropedicae Z67	*Hs*D-KdpD	WP_013235815.1
Variovorax paradoxus EPS	*Vp*D-KdpD	WP_013538688.1
Cupriavidus necator N-1	*Cn*D-KdpD	WP_011616492.1
Azospirillum brasilense	*Ab*L-KdpD	PDB: 3FKK
Cupriavidus necator N-1	*Cn*L-KdpD	WP_010809845.1
Herbaspirillum seropedicae Z67	*Hs*L-KdpD	WP_013233389.1
Variovorax paradoxus EPS	*Vp*L-KdpD	ADU35339.1

Table 2 Kinetic characterization of L-Kdp and D-Kdp dehydratases toward their natural isomer. All measurements were done in triplicate at 25 °C in 50 mM HEPES pH 7.5[a]

Substrate	Enzymes	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	Substrate preference[b]
L-KDP	*Ab*L-KdpD	13.6 ± 0.3	0.19 ± 0.02	71.3 ± 0.3	73
	*Cn*L-KdpD (Nhis)	23.4 ± 0.1	0.17 ± 0.00	137.2 ± 0.2	46
	*Cn*L-KdpD (Chis)	22.5 ± 0.2	0.16 ± 0.01	137.5 ± 0.2	80
	*Hs*L-KdpD	29.6 ± 0.2	0.18 ± 0.00	163.9 ± 0.2	63
	*Vp*L-KdpD	21.8 ± 0.2	0.15 ± 0.01	146.0 ± 0.2	62
D-KDP	*Cc*D-KdpD1	45.6 ± 0.3	0.25 ± 0.01	185.7 ± 0.2	65
	*Cc*D-KdpD2	55.8 ± 0.4	0.29 ± 0.01	190.6 ± 0.2	78
	*Px*D-KdpD	63.6 ± 0.4	0.28 ± 0.01	229.8 ± 0.2	62
	*Pp*D-KdpD	75.9 ± 0.6	0.35 ± 0.01	219.9 ± 0.2	69
	*Hs*D-KdpD	58.2 ± 0.4	0.17 ± 0.01	347.9 ± 0.2	65
	*Cn*D-KdpD	13.8 ± 0.2	0.23 ± 0.02	58.9 ± 0.4	35

[a] Error bars represent standard deviation from three replicates. Nonlinear regression of the enzyme activity as a function of substrate concentration is presented in Fig. S2

[b] The substrate preference was calculated as the ratio between the catalytic efficiency of the natural substrate and the non-preferred stereoisomer presented in Table S2

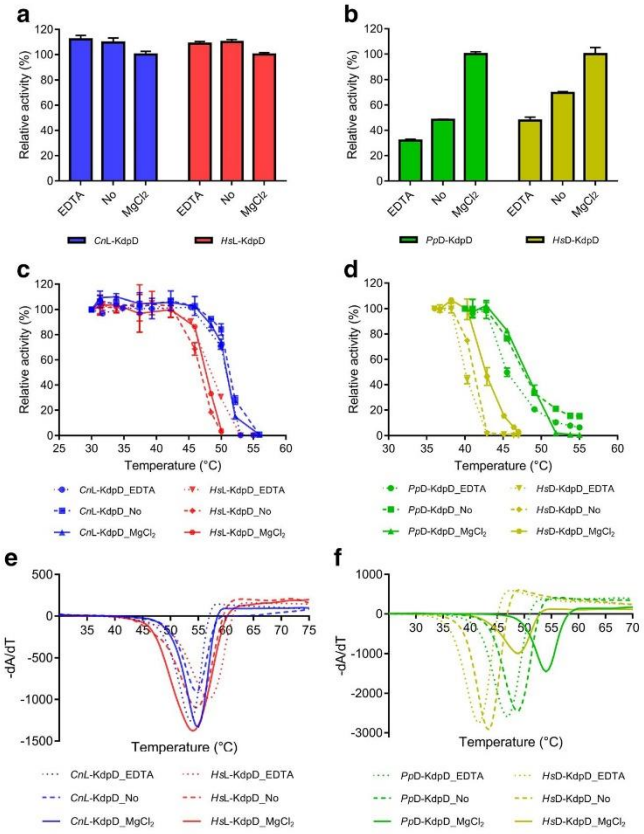


Fig. 2 Influence of $MgCl_2$ and EDTA on the activity, kinetic, and thermodynamic stabilities of L-KdpDs and D-KdpDs. L-KdpDs (**a**) and D-KdpDs (**b**) were pre-incubated with 5 mM EDTA (labeled with EDTA) prior to activity measurement using the coupled enzyme assay in the absence of $MgCl_2$. In addition, activities were also determined without EDTA pre-incubation in the absence (labeled with "No") and presence of 5 mM $MgCl_2$. The kinetic stability of L-KdpDs (**c**) and D-KdpDs (**d**) was assayed as $T_{50}{}^{1h}$ (temperature at which 50% of kinetic activity is lost after 1 h of incubation). The thermodynamic stability of L-KdpDs (**e**) and D-KdpDs (**f**) was determined using Thermofluor and is represented as the melting temperature (T_m). The effect of the addition of EDTA (5 mM), the absence of $MgCl_2$, and the addition of $MgCl_2$ (5 mM) in the enzyme solution was investigated for both enzymes. The numerical values are presented in Table S3

magnesium ions were not present during the activity measurement, a significant reduction of activity was observed (Fig. 2b), and the initial activity also seemed to be delayed (Fig. S3). Pretreatment with EDTA further decreased the initial activity.

Kinetic stability is also an important parameter in determining the applicability of an enzyme. In this study, T_{50}^{1h} was used as an indicator of kinetic stability. This value can be used to reliably compare kinetic stability of enzymes given the same experimental conditions. *Cn*L-KdpD appeared to have higher kinetic stability (T_{50}^{1h} of 53.2 °C) in comparison to *Hs*L-KdpD (T_{50}^{1h} of 47.1 °C) (Fig. 2c). Again, it was shown that the absence of magnesium had barely any effect in regard to their kinetic stability. The presence of 5 mM metal chelator EDTA (> 250 equivalent) also did not decrease nor increase the kinetic stability of the two L-KdpDs. Regarding the two D-KdpDs, *Pp*D-KdpD revealed a T_{50}^{1h} of 48.0 °C in the presence of magnesium compared to a T_{50}^{1h} of 42.5 °C for *Hs*D-KdpD. The absence of magnesium did not significantly affect the kinetic stability of *Pp*D-KdpD. However, in the presence of 5 mM EDTA a slight decrease in kinetic stability (T_{50}^{1h} lowered by 2.5 °C) was observed. The T_{50}^{1h} of *Hs*D-KdpD was decreased by 1.3 °C and 2.4 °C in the absence of magnesium and presence of 5 mM EDTA, respectively.

In the present study, the thermodynamic stability is represented by the melting temperature (T_m) and was determined by a thermal shift assay using Sypro orange dye. No apparent effect of magnesium in either of the representative L-KdpDs was found. The presence of 5 mM EDTA (>2500 equivalent) did not alter the T_m of *Cn*L-KdpD significantly, but it slightly increased the T_m of *Hs*L-KdpD (Fig. 2e). *Cn*L-KdpD and *Hs*L-KdpD appeared to have similar T_m, although *Cn*L-Kdpd showed a higher T_{50}^{1h}. The stabilizing effect of magnesium was observed for the representative D-KdpDs. An addition of 5 mM magnesium increased the T_m of *Pp*D-KdpD and *Hs*D-KdpD by 5.5 and 5.1 °C, respectively. The presence of 5 mM EDTA (>2500 equivalent) decreased the T_m of both enzymes by about 2 °C, further supporting the effect of magnesium on their thermodynamic stability. Given that kinetic stability is a more relevant parameter in practical enzyme application in vivo and in vitro than thermodynamic stability, this parameter was used to further select the most promising enzyme variant from each dehydratase class. *Cn*L-KdpD and *Pp*D-KdpD were chosen for the determination of their pH profiles, the effect of selected intermediates of the Weimberg pathway, as well as suitability for high substrate conversion.

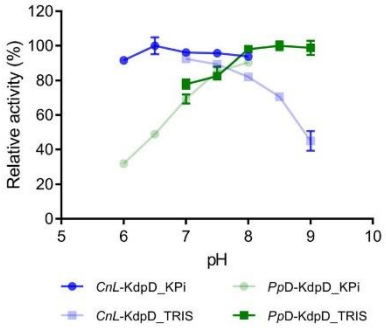


Fig. 3 Optimum pH of *Cn*L-KDAD and *Pp*D-KdpD. *Cn*L-KDAD revealed an optimum pH of between 6.5 and 7 (darker blue), whereas *Pp*D-KdpD possessed an optimum pH of between 8 and 9 (darker green). Error bars represent the standard deviation of the three replicates

Optimum pH for *Pp*D-KdpD and *Cn*L-KdpD

The pH profile of an enzyme is an important parameter for its application in vitro and in vivo. In particular, the pH of a reaction can be adjusted easily in a cell-free approach. Both enzymes, *Cn*L-KdpD and *Pp*D-KdpD, showed broad pH activity (Fig. 3). However, the main difference was that *Cn*L-KdpD showed higher activity under acidic conditions (optimum pH at 6.5 to 7) and showed significantly reduced activity (> 50% reduction) at pH 9. *Pp*D-KdpD, on the other hand, showed a preference for alkaline conditions (optimum pH at 8 to 9), and its activity was significantly reduced at pH 6 (70% reduction). There was no apparent effect on either enzyme activity by the buffer used, KPi or TRIS. A slightly acidic pH optimum was previously reported for *Ps*L-KdpD (Stoolmiller and Abeles 1966).

Effect of upstream intermediates of the Weimberg pathways on *Cn*L-KdpD and *Pp*D-KdpD

In a multi-enzymatic reaction, intermediates can be accumulated depending on varying enzyme activities in a cascade reaction. Therefore, it would be important to determine whether the substrates or intermediates of the Weimberg pathway would have an inhibitory effect on *Cn*KdpD and *Pp*D-KdpD. L-arabinose and L-arabinonate were selected for the oxidative L-arabinose pathway and D-xylose and D-xylonate were selected for the oxidative D-xylose pathway. KGSA and α-KG were excluded in this study because the activities of *Cn*KdpD and *Pp*D-KdpD were determined by a coupled assay using *Pp*KGSADH as auxiliary enzyme to oxidize the formed KGSA to α-KG in the presence of NAD^+. As for α-KG, not all modified pathways of D-xylose and L-arabinose will end or go via α-KG, e.g., toward 1,4-BDO formation (Fig. 1). Therefore, the effect of α-KG and other intermediates in the synthetic pathways would need to be determined later based on the desired target products.

*Cn*LKdpD demonstrated no inhibitory effects in the presence of L-arabinose and L-arabinonate up to 100 mM (Fig. 4). These characteristics would make *Cn*LKdpD a suitable enzyme for application in the non-phosphorylative oxidative conversion of L-arabinose (McClintock et al. 2017). As for *Pp*D-KdpD, the enzyme was inhibited in the presence of D-

xylonate with an I_{50} value (the concentration of a compound that gives 50% inhibition to the initial activity of an enzyme) of 75 mM, whereas the *Pp*D-KdpD was not inhibited by D-xylose (Fig. 4). *Pp*KGSADH, the auxiliary enzyme in the coupled assay, was also confirmed not to be inhibited by D-xylonate. Therefore, the inhibition effect shown was only for *Pp*D-KdpD.

One-step conversion of L-KDP and D-KDP to KGSA using *Cn*L-KdpD and *Pp*KdpD

A high substrate load is often desirable for the biotransformation of chemicals. To determine the feasibility of both enzymes respecting high substrate load conversion, each enzyme was used to transform 500 mM D-KDP and L-KDP. A previously established HPLC analysis was utilized for this experiment (Beer et al. 2017). For the conversion of D-KDP, *Pp*D-KdpD was able to reach 90% conversion to KGSA after 4 h. For L-KDP, *Cn*L-KdpD reached > 95% conversion after only 2 h (Fig. 5a). In the first 10 min, *Pp*D-KdpD showed a higher turnover rate than *Cn*L-KdpD, but the conversion of *Pp*D-KdpD slowed over time. When the concentration of KGSA reached 300 mM, the activity of *Pp*D-KdpD was only 17 U/mg, representing only 12% of its initial activity (in the first 10 min). As for *Cn*L-KdpD, its activity was 38 U/mg in the presence of 300 mM KGSA, representing 45% of its initial activity. A combination of both enzymes was also able to convert a 500 mM racemic mixture of D,L-KDP (250 mM of each isomer). In this experimental set-up, > 80% conversion was observed after 4 h (Fig. 4b). The concentrations of KGSA decreased after 8 h, and a further decrease after 24 h of incubation was noticed in both set-ups (Fig. 4a and b). The degradation of KGSA was likely due to interaction of its terminal aldehyde group with amino acids on the protein surface (Robert and Penaranda 1954; Uchida 2003; Grimsrud et al. 2008). This is further supported by the yellow color of the reaction solutions observed after 24 h of reaction, which is indicative of a typical aldehyde protein interaction.

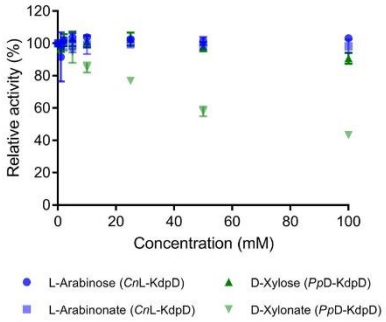


Fig. 4 Effect of substrates and intermediates of the Weimberg pathway on the activity of *Cn*L-KdpD and *Pp*D-KdpD. There was no inhibitory effect of D-xylose up to 100 mM on the activity of *Pp*D-KdpD, as well as L-arabinose and L-arabinonate on the activity of *Cn*L-KdpD. D-xylonate inhibited *Pp*D-KdpD activity at an I_{50}-value of 75 mM

During the kinetic characterization, it was shown that both enzymes were also reactive toward the other non-preferred stereoisomer (*Pp*D-KdpD toward L-KDP and *Cn*L-KdpD toward D-KDP) although at a much lower efficiency. A similar finding was previously reported for the D-KdpD from *Herbaspirillum huttiense* and L-KdpD from *Paraburkholderia mimosarum* (Watanabe et al. 2019). The application of only one enzyme for the conversion of both stereoisomers would be of interest for practical applications such as the conversion of mixture of sugars, e.g., L-arabinose and D-xylose, into certain chemicals. Therefore, in order to confirm their promiscuity, each enzyme was used to convert 500 mM of the non-preferred stereoisomer. As shown in Fig. 4c, both enzymes were active toward the other non-preferred stereoisomer. However, the reaction appeared to have halted rather quickly. After 21 and 45 h of reaction, only a slightly higher formation of KGSA was observed for both enzymes. HPLC analysis indicated that the product formed was indeed KGSA because the same retention time was observed (Fig. S4). When the solution of each enzyme after 21 h reaction with the other non-preferred isomer (Fig. 5c) was transferred to a new solution containing 500 mM of their natural isomer, the formation of KGSA was immediately initiated (Fig. 5d). This indicates that the presence of high concentration of the non-preferred isomer did not have any negative effect for both enzymes. *Cn*L-KdpD reached 40% of the theoretical yield of KGSA (80% if calculated from L-KDP only), whereas *Pp*D-KdpD reached 33% of the theoretical yield of KGSA (66% from D-KdpD) after 3 h.

Discussion

D-xylose and L-arabinose are the two most abundant pentose sugars in hemicellulose. Utilization of these sugars to produce platform chemicals would allow valorization of lignocellulose biomass. The enzymatic cascade reactions of the Weimberg pathways have been used to transform these pentose sugars into several chemicals, e.g., BDO, α-KG, succinate, glutamate, and mesaconate (Fig. 1). These chemicals are still industrially produced or derived from petrochemicals. Enzymes for the key steps of the Weimberg pathway, specifically L-KdpD and D-KdpD, have only been partially characterized with major focus on their kinetic properties. These specific enzymes catalyze the dehydration of L-KDP and D-KDP to KGSA.

In this study, several L-KdpDs and D-KdpDs were cloned and heterologously expressed in *E. coli*. For the *Cn*L-KdpD, the N-terminally fused His-tag resulted in higher protein yields compared to the C-histag variant. For D-KdpD,

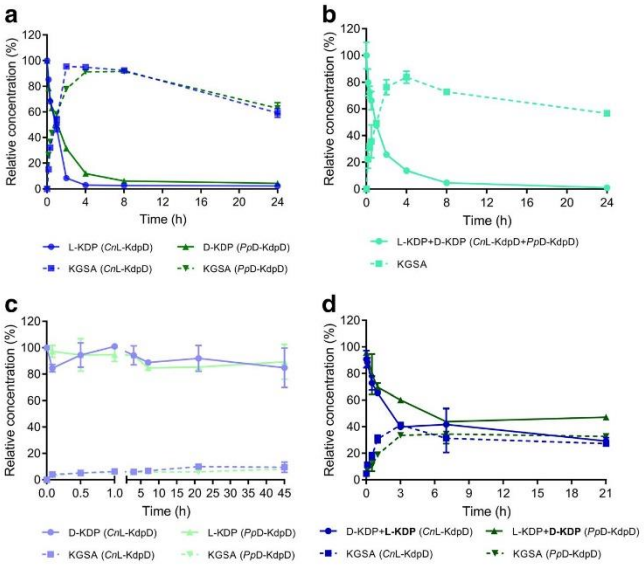


Fig. 5 Time-dependent conversion of L-KDP and D-KDP to KGSA using *Cn*L-KdpD and *Pp*D-KdpD. A total of 0.1 mg/mL of each enzyme was used to convert 500 mM of the respective KDP (**a**,**c**). A combination of *Cn*L-KdpD and *Pp*D-KdpD (0.05 mg/mL each) was used to convert 500 mM of a racemic mixture of D,L-KDP (**b**). After each enzyme was used to convert the non-preferred isomer for 21 h (in panel c), 100 μL of each solution was transferred to 100 μL with 500 mM of the respective natural isomer (labeled with bold) to see if both enzymes were still able to convert their natural substrate (**d**). All experiments were performed in HEPES pH 7.5 at 25 °C

however, the C-histag was the better variant. The fact of the N-terminally fused variant of *Cc*D-KdpD1 quickly precipitating and the higher activity of *Cc*D-KdpD2, which is a homologous enzyme of *Cc*KdpD1 (58 amino acids shorter at the N-terminus, Fig. S1) as compared to *Cc*D-KdpD1, could suggest that the N-terminus for D-KdpD is important to enzyme activity and/or stability, and that the addition of further amino acids N-terminally would impede either. However, in a recent study, N-terminally fused His-tag of *Cc*KdpD2 (bearing additional 20 amino acids in N-terminus in comparison to C-histag of *Cc*KdpD2 in this study) showed a magnitude of activity similar to *Cc*KdpD2 (Shen et al. 2020). Further analysis using BLAST searches in the NCBI and UniProt database suggest that the previously cloned and characterized *Cc*D-KdpD1 might be cloned with the wrong start codon (Tai et al. 2016), since most of the homologs do not show the 58 amino acids extension at the N-terminus. The amino acid downstream of the current start codon (the 59th amino acid encoded by GTG) is a second possible start site. Therefore, future studies have to be completed in order to confirm whether the truncated version is more stable and/or active and to decide whether the length of the additional amino acids is also an important factor.

All L-KdpDs revealed a high catalytic efficiency (> 100-fold) with L-KDP, whereas all D-KdpDs showed high efficiency (> 150-fold) with D-KDP. The catalytic efficiency of the KdpDs is an important parameter for in vitro and in vivo pathway design. In particular, the catalytic efficiency for D-KdpD measured in vitro has been shown to have a strong correlation with enzyme efficiency in vivo (Tai et al. 2016). In the previous study, it was shown that the *E. coli* mutant expressing *Px*D-KdpD produced BDO faster than the *E. coli* mutant expressing *Cc*D-KdpD1 (Tai et al. 2016). The authors (as well as this study) have demonstrated a higher catalytic efficiency for *Px*D-KdpD than for *Cc*D-KdpD1 (Table 2). Furthermore, in the recent study combining enzymatic and modeling approach, the authors were able to predict and simulate the activity of all five Weimberg enzymes, including *Cc*D-KdpD2 in cell-free extracts based on their kinetic model (Shen et al. 2020).

A study on the effect of magnesium on activity, kinetic, and thermodynamic stability revealed that this cation showed effects on members of the D-KdpD class, while no apparent effects were observed on the enzymes of the L-KdpD class. The presence of magnesium increased the activity and kinetic ($T_{50}{}^{1h}$) and thermodynamic (T_m) stabilities of *Pp*D-KdpD and *Hs*D-KdpD. No significant difference was observed when comparing the thermodynamic stability of *Cn*L-KdpD and *Hs*L-KdpD, although the kinetic stability suggested otherwise. Apparently, the relationships between the kinetic and thermodynamic stabilities were not able to be observed for L-KdpD, which stands in contrast to our previous studies of two α-keto acid decarboxylases (Sutiono et al. 2018; Sutiono et al. 2019).

The detailed characterization with respect to pH profiles revealed an optimum pH at 6.5 to 7 for *Cn*L-KdpD, whereas *Pp*D-KdpD showed the highest activity at pH 8 to 9. Together with the differing effect of magnesium toward L- and D-

KdpD, these results suggest that L-KdpD and D-KdpD may adopt different catalytic mechanisms for performing the dehydration at position C4 and C5 of L-KDP and D-KDP, respectively. This outcome could also imply that different amino acid residues are acting during the catalysis, since the pH level would determine the protonation states of charged amino acids. This hypothesis is further supported by the fact that, from the automated gene annotation in these two enzyme classes, although they catalyze similar reactions, they belong to different enzyme families. All D-KdpDs are annotated as members of the fumarylacetoacetate hydrolase (FAH) family, whereas all L-KdpDs are annotated as members of the dihydrodipicolinate synthase (DHDPS) family. The relative similarity of each dehydratase in their corresponding family is presented in Fig. S5.

The effects of upstream substrates and intermediates in the Weimberg pathway on each enzyme were also studied. L-arabinose and L-arabinonate up to 100 mM did not show any inhibitory effect on *Cn*L-KdpD. Nor did the same concentration of D-xylose inhibit the activity of *Pp*-KdpD. However, D-xylonate showed a significant inhibition of *Pp*-KdpD at an I_{50}-value of 75 mM. This concentration seems to be quite high (the higher the I_{50}, the less significant the inhibitory effect). However, due to the low activity of D-xylonate dehydratase (Fig. 1), D-xylonate could accumulate to this relevant range. The [Fe-S]-dependent sugar acid dehydratase has been reported as being one of the major bottlenecks in in vitro and in vivo utilization of the Weimberg pathway (Salusjärvi et al. 2017; Boer et al. 2019; Shen et al. 2020). Therefore, a more active D-xylonate dehydratase would be needed to avoid an accumulation of D-xylonate that would inhibit *Pp*D-KdpD. In addition, due to the high k_{cat} value exhibited by *Pp*D-KdpD (75 s^{-1}), it would be unlikely for this enzyme to be the bottleneck in the non-phosphorylative oxidative conversion of D-xylose (Fig. 1). A stronger inhibition effect of D-xylonate was reported for *Cc*D-KdpD2 (Shen et al. 2020).

In the one-step conversion, *Cn*L-KdpD was able to convert 500 mM of L-KDP to KGSA, reaching > 95% yield after only 2 h. When the same amount of *Pp*D-KdpD (higher initial activity) was used to convert D-KDP, the enzyme needed 4 h to reach > 90% yield. This indicates that the KGSA formed has a higher inhibitory effect on *Pp*D-KdpD than on *Cn*L-KdpD. The combination of both enzymes can convert 500 mM of a racemic mixture of D,L-KDP, reaching 83% yield after 4 h. An initial activity test of both enzymes suggests that each enzyme could convert the non-preferred stereoisomer of its natural substrate (Table S2). However, time-based experiments of the non-preferred isomer (Fig. 4c) reveal that both enzymes were only able to convert < 5%, even after 45 h of reaction. This could indicate that the non-preferred isomers inhibit each of the two enzymes. Another, more plausible, explanation is that most (if not all) of the KGSA was formed due to contamination with the natural stereoisomer. An impurity could be derived from the initial substrates used, e.g., aldoses or an unwanted epimerization at C4 which could occur during preparation of the D- and L-KDP. Further experiments, in which the non-preferred stereoisomer was titrated to each enzyme reaction, demonstrated that a greater substrate addition leads to a greater formation of KGSA (Fig. S6). This result suggests that each enzyme is somewhat more specific to the respective stereoisomer than appeared to be the case from the initial activity measurement (Table S2) (Stoolmiller and Abeles 1966). Apparently, the initial activity measurements are not enough to determine the promiscuity of these enzyme classes.

In conclusion, we have cloned, expressed, and characterized a number of L-KdpDs and D-KdpDs. On the basis of catalytic efficiency, activity, and kinetic stability, we demonstrated that *Cn*L-KdpD and *Pp*D-KdpD are the most promising variants from each enzyme class. Both enzymes possess a broad pH range. However, although *Cn*L-KdpD revealed the highest activity at a slightly acidic pH, *Pp*D-KdpD was more active at an alkaline pH. Both enzymes are suitable for high-substrate load conversion.

Acknowledgments The authors would like to acknowledge funding by German Federal Ministry for Education and Research (BMBF) through HotSysAPP project (Grant Nos. 031L0078F, 031L0078A). The authors thank Prof. Dirk Tischler (Ruhr University Bochum) for providing *Variovorax paradoxus* (EPS) cells. The authors also wish to thank Alexander Benson for proofreading the manuscript.

Author contribution statement SS and VS conceived and designed the study. SS performed the experiments and analyzed the data. SS, BS, and VS wrote the manuscript. All authors read and approved the final version of the manuscript.

Funding information Open Access funding provided by Projekt DEAL. This study was funded by German Federal Ministry for Education and Research (BMBF) through HotSysAPP project (Grant Nos. 031L0078F, 031L0078A).

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Andberg M, Aro-Kärkkäinen N, Carlson P, Oja M, Bozonnet S, Toivari M, Hakulinen N, Donohue MO (2016) Characterization and mutagenesis of two novel iron–sulphur cluster pentonate dehydratases. Appl Microbiol Biotechnol 100:7549–7563. https://doi.org/10.1007/s00253-016-7530-8

Bai W, Tai YS, Wang J, Wang J, Jambunathan P, Fox KJ, Zhang K (2016) Engineering nonphosphorylative metabolism to synthesize mesaconate from lignocellulosic sugars in *Escherichia coli*. Metab Eng 38:285–292. https://doi.org/10.1016/j.ymben.2016.09.007

Beer B, Pick A, Sieber V (2017) In vitro metabolic engineering for the production of α-ketoglutarate. Metab Eng 40:5–13

Boer H, Andberg M, Pylkkänen R, Maaheimo H, Koivula A (2019) In vitro reconstitution and characterisation of the oxidative d-xylose pathway for production of organic acids and alcohols. AMB Express 9:48. https://doi.org/10.1186/s13568-019-0768-7

Boivin S, Kozak S, Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif 91:192–206. https://doi.org/10.1016/j.pep.2013.08.002

Brouns SJJ, Walther J, Snijders APL, Van De Werken HJG, Willemen HLDM, Worm P, De Vos MGJ, Andersson A, Lundgren M, Mazon HFM, Van Den Heuvel RHH, Nilsson P, Salmon L, De Vos WM, Wright PC, Bernander R, Van Der Oost J (2006) Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment. J Biol Chem 281:27378–27388. https://doi.org/10.1074/jbc.M605549200

Brouns SJJ, Barends TRM, Worm P, Akerboom J, Turnbull AP, Salmon L, van der Oost J (2008) Structural insight into substrate binding and catalysis of a novel 2-keto-3-deoxy-d-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily. J Mol Biol 379:357–371. https://doi.org/10.1016/j.jmb.2008.03.064

Chandel AK, Chandrasekhar G, Radhika K, Ravinder R, Ravindra P (2011) Bioconversion of pentose sugars into ethanol: a review and future directions. Biotechnol Mol Biol Rev 6:8–20. https://doi.org/10.1079/9781845936662.0101

Chin RM, Fu X, Pai MY, Vergnes L, Hwang H, Deng G, Diep S, Lomenick B, Meli VS, Monsalve GC, Hu E, Whelan SA, Wang JX, Jung G, Solis GM, Fazlollahi F, Kaweeteerawat C, Quach A, Nili M, Krall AS, Godwin HA, Chang HR, Faull KF, Guo F, Jiang M, Trauger SA, Saghatelian A, Braas D, Christofk HR, Clarke CF, Teitell MA, Petrascheck M, Reue K, Jung ME, Frand AR, Huang J (2014) The metabolite α-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. Nature 510:397–401. https://doi.org/10.1038/nature13264

Claassens NJ, Burgener S, Vo B, Erb TJ, Bar-even A (2019) A critical comparison of cellular and cell-free bioproduction systems. Curr Opin Biotechnol 60:221–229. https://doi.org/10.1016/j.copbio.2019.05.003

De La Plaza M, Fernández De Palencia P, Peláez C, Requena T (2004) Biochemical and molecular characterization of α-ketoisovalerate decarboxylase, an enzyme involved in the formation of aldehydes from amino acids by *Lactococcus lactis*. FEMS Microbiol Lett 238:367–374. https://doi.org/10.1016/j.femsle.2004.07.057

Dudley QM, Karim AS, Jewett MC (2015) Cell-free metabolic engineering: biomanufacturing beyond the cell. Biotechnol J 10:69–82. https://doi.org/10.1002/biot.201400330

Fernandes S, Murray P (2010) Metabolic engineering for improved microbial pentose fermentation. Bioeng Bugs 1:424–428. https://doi.org/10.4161/bbug.1.6.12724

Grimsrud PA, Xie H, Griffin TJ, Bernlohr DA (2008) Oxidative stress and covalent modification of protein with bioactive aldehydes. J Biol Chem 283:21837–21841. https://doi.org/10.1074/jbc.R700019200

Guterl J, Sieber V (2013) Biosynthesis "debugged": novel bioproduction strategies. Eng Life Sci 13:4–18. https://doi.org/10.1002/elsc.201100231

Guterl JK, Garbe D, Carsten J, Steffler F, Sommer B, Reiße S, Philipp A, Haack M, Rühmann B, Koltermann A, Kettling U, Brück T, Sieber V (2012) Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. ChemSusChem 5:2165–2172. https://doi.org/10.1002/cssc.201200365

Huang Y-B, Fu Y (2013) Hydrolysis of cellulose to glucose by solid acid catalysts. Green Chem 15:1095–1111. https://doi.org/10.1039/c3gc40136g

Isikgor FH, Becer CR (2015) Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. Polym Chem 6:4497–4559. https://doi.org/10.1039/c5py00263j

Jagtap SS, Rao CV (2018) Microbial conversion of xylose into useful bioproducts. Appl Microbiol Biotechnol 102:9015–9036. https://doi.org/10.1007/s00253-018-9294-9

Liu S, He L, Yao K (2018) The antioxidative function of alpha-ketoglutarate and its applications. Biomed Res Int 2018:1–6. https://doi.org/10.1155/2018/3408467

McClintock MK, Wang J, Zhang K (2017) Application of nonphosphorylative metabolism as an alternative for utilization of lignocellulosic biomass. Front Microbiol 8:1–6. https://doi.org/10.3389/fmicb.2017.02310

Mohr A, Raman S (2015) Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. Effic Sustain Biofuel Prod Environ Land-Use Res 63:281–310. https://doi.org/10.1016/j.enpol.2013.08.033

Naik SN, Goud VV, Rout PK, Dalai AK (2010) Production of first and second generation biofuels: a comprehensive review. Renew Sust Energ Rev 14:578–597. https://doi.org/10.1016/j.rser.2009.10.003

Portsmouth D, Toolmiller AC, Abeles RH (1967) Studies on the mechanism of action of 2-keto-3-deoxy-l-arabonate dehydratase. J Biol Chem 242:2751–2759. https://doi.org/10.1094/phyto-62-239

Robert L, Penaranda FS (1954) Studies on aldehyde-protein interactions. I. Reaction of amino acids with lower aldehydes. J Polym Sci 12:337–350. https://doi.org/10.1002/pol.1954.120120128

Rulli MC, Bellomi D, Cazzoli A, De Carolis G, D'Odorico P (2016) The water-land-food nexus of first-generation biofuels. Sci Rep 6:1–10. https://doi.org/10.1038/srep22521

Salusjärvi L, Toivari M, Vehkomäki ML, Koivistoinen O, Mojzita D, Niemelä K, Penttilä M, Ruohonen L (2017) Production of ethylene glycol or glycolic acid from D-xylose in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 101:8151–8163. https://doi.org/10.1007/s00253-017-8547-3

Saxena RK, Saran S, Isar J, Kaushik R (2016) Production and applications of succinic acid. Curr Dev Biotechnol Bioeng Prod Isol Purif Ind Prod:601–630. https://doi.org/10.1016/B978-0-444-63662-1.00027-0

Shen L, Kohlhaas M, Enoki J, Meier R, Schönenberger B, Wohlgemuth R, Kourist R, Niemeyer F, van Niekerk D, Bräsen C, Niemeyer J, Snoep J, Siebers B (2020) A combined experimental and modelling approach for the Weimberg pathway optimisation. Nat Commun 11:1–13. https://doi.org/10.1038/s41467-020-14830-y

Sperl JM, Carsten JM, Guterl JK, Lommes P, Sieber V (2016) Reaction design for the compartmented combination of heterogeneous and enzyme catalysis. ACS Catal 6:6329–6334. https://doi.org/10.1021/acscatal.6b01276

Stephen Dahms A (1974) 3-Deoxy-D-pentulosonic acid aldolase and its role in a new pathway of D-xylose degradation. Biochem Biophys Res Commun 60:1433–1439. https://doi.org/10.1016/0006-291X(74)90358-1

Stoolmiller AC, Abeles RH (1966) Formation of α-ketoglutaric semialdehyde from L-2-Keto-3-deoxyarabonic acid and isolation of dehydratase from *Pseudomonas saccharophila*. J Biol Chem 241:5764–5771

Stottmeister U, Aurich A, Wilde H, Andersch J, Schmidt S, Sicker D (2005) White biotechnology for green chemistry: fermentative 2-oxocarboxylic acids as novel building blocks for subsequent chemical syntheses. J Ind Microbiol Biotechnol 32:651–664. https://doi.org/10.1007/s10295-005-0254-x

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207–234. https://doi.org/10.1016/j.pep.2005.01.016

Sutiono S, Carsten J, Sieber V (2018) Structure-guided engineering of α-keto acid decarboxylase for the production of higher alcohols at elevated temperature. ChemSusChem 11:3335–3344. https://doi.org/10.1002/cssc.201800944

Sutiono S, Satzinger K, Pick A, Carsten J, Sieber V (2019) To beat the heat—engineering of the most thermostable pyruvate decarboxylase to date †. RSC Adv 9:29743–29746. https://doi.org/10.1039/c9ra06251c

Sutiono S, Teshima M, Beer B, Schenk G, Sieber V (2020) Enabling the direct enzymatic dehydration of D-glycerate to pyruvate as the key step in synthetic enzyme cascades used in the cell-free production of fine chemicals. ACS Catal acscatal.9b05068. https://doi.org/10.1021/acscatal.9b05068

Tai YS, Xiong M, Jambunathan P, Wang J, Wang J, Stapleton C, Zhang K (2016) Engineering nonphosphorylative metabolism to generate lignocellulose-derived products. Nat Chem Biol 12:247–253. https://doi.org/10.1038/nchembio.2020

Uchida K (2003) Histidine and lysine as targets of oxidative modification. Amino Acids 25:249–257. https://doi.org/10.1007/s00726-003-0015-y

Valdehuesa KNG, Ramos KRM, Nisola GM, Bañares AB, Cabulong RB, Lee WK, Liu H, Chung WJ (2018) Everyone loves an underdog: metabolic engineering of the xylose oxidative pathway in recombinant microorganisms. Appl Microbiol Biotechnol 102:7703–7716. https://doi.org/10.1007/s00253-018-9186-z

Watanabe S, Shimada N, Tajima K, Kodaki T, Makino K (2006) Identification and characterization of L-arabonate dehydratase, L-2-keto-3-deoxyarabonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose metabolism: novel evolutionary insight into sugar metabolism. J Biol Chem 281:33521–33536. https://doi.org/10.1074/jbc.M606727200

Watanabe S, Fukumori F, Nishiwaki H, Sakurai Y, Tajima K, Watanabe Y (2019) Novel non-phosphorylative pathway of pentose metabolism from bacteria. Sci Rep 9:1–12. https://doi.org/10.1038/s41598-018-36774-6

Weimberg R (1961) Pentose oxidation by *Pseudomonas fragi*. J Biol Chem 236:629–635

Wu N, Yang M, Gaur U, Xu H, Yao Y, Li D (2016) Alpha-ketoglutarate: physiological functions and applications. Biomol Ther 24:1–8. https://doi.org/10.4062/biomolther.2015.078

Yadav SK, Schmalbach KM, Kinaci E, Stanzione JF, Palmese GR (2018) Recent advances in plant-based vinyl ester resins and reactive diluents. Eur Polym J 98:199–215. https://doi.org/10.1016/j.eurpolymj.2017.11.002

Supplementary Materials

Characterization of highly active 2-keto-3-deoxy-L-arabinonate and 2-keto-3-deoxy-D-xylonate dehydratases in terms of the biotransformation of hemicellulose sugars to chemicals

Samuel Sutiono[a], Bettina Siebers[b], Volker Sieber[a,c,d,e*]

[a] Chair of Chemistry of Biogenic Resources, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, Schulgasse 16, 94315 Straubing, Germany

[b] Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Centre for Water and Environmental Research (CWE), University of Duisburg-Essen, Universitätsstraße 5, 45117 Essen, Germany

[c] Catalytic Research Center, Technical University of Munich, Ernst-Otto-Fischer-Straße 1, 85748 Garching, Germany

[d] Straubing Branch BioCat, Fraunhofer IGB, Schulgasse 11a, 94315 Straubing, Germany

[e] School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Copper Road, St. Lucia 4072, Australia

*Email: sieber@tum.de; ORCID: 0000-0001-5458-9330; Tel: +49 (0) 9421 187-300; Fax: +49 (0) 9421 187-310

Table S1. List of primers and restriction enzymes used to clone L-KdpDs and D-KdpDs in this study*

Protein	NCBI Accession number	Restriction enzymes	Primers (5´to 3´)
*Cc*D-KdpD1	WP_010918708.1	NdeI, XhoI	Ordered as an optimized gene for *E. coli* expression (GenBank Accession number: MT550669)
*Cc*D-KdpD2	WP_012640070.1	NdeI, HindIII	Fwd: CAGCAGCATATGGGCGTGAGTGAATTCCTGCCGGAAGATTG Rev: CAGCAGAAGCTTGAGGAGGCCGCGGCCGGC
*Px*D-KdpD	WP_011494434.1	NdeI, XhoI	Ordered as an optimized gene for *E. coli* expression (GenBank Accession number: MT550670)
*Pp*D-KdpD	WP_010953745.1	NdeI, XhoI	Fwd: CAGCAGCATATGACTGATCGAACGAACGCCAATC Rev: CAGCAGCTCGAGAAGCGGGCGGGCGGTGG
*Hs*D-KdpD	WP_013235815.1	NdeI, XhoI	Fwd: CAGCAGCATATGGCACACACCTTTTCGCTCCAGGCGC Rev: CAGCAGCTCGAGAATGAGTTTGCGGTCGGCCAGATTTTTGAACAGCGC
*Vp*D-KdpD	WP_013538688.1	NdeI, EcoRI	Fwd: CAGCAGCATATGGCCCTGAACCTCTCGCCCGCGAC Rev: CAGCAGGAATTCCAACAGGCCGCGCGCGCCCAGGTTG
*Cn*D-KdpD	WP_011616492.1	NdeI, XhoI	Fwd: CAGCAGCATATGGCATTGCCGCTCACGCTCAGCG Rev: CAGCAGCTCGAGGAAAGTGGTGTGGCCACGTGCGGCC
*Ab*L-KdpD	PDB: 3FKK	NdeI, XhoI	Ordered as an optimized gene for *E. coli* expression (GenBank Accession number: MT550671)
*Cn*L-KdpD	WP_010809845.1	NdeI, XhoI	Fwd: CAGCAGCATATGACCCGCACCCCATCCC Rev1: CAGCAGCTCGAG**TTATCA**GCGGGCCCAGCGCAG Rev2: CAGCAGCTCGAGGCGGGCCCAGCGCAG
*Hs*L-KdpD	WP_013233389.1	NdeI, XhoI	Fwd: CAGCAGCATATGAGCACCCCTCTCTACCGCGGCGTATTCCC Rev: CAGCAGCTCGAG**TTA**CCTGCCCCAGCTCAGCACCAGCGGATC
*Vp*L-KdpD	ADU35339.1	NdeI, XhoI	Fwd: CAGCAGCATATGCCCAAGCCCATGAACCACAACCTCTCC Rev: CAGCAGCTCGAG**TTA**CTTCCCCCAGCGCAGCACCAGCGGG

* Restriction site in the primers is underlined. Stop codon is highlighted bold.

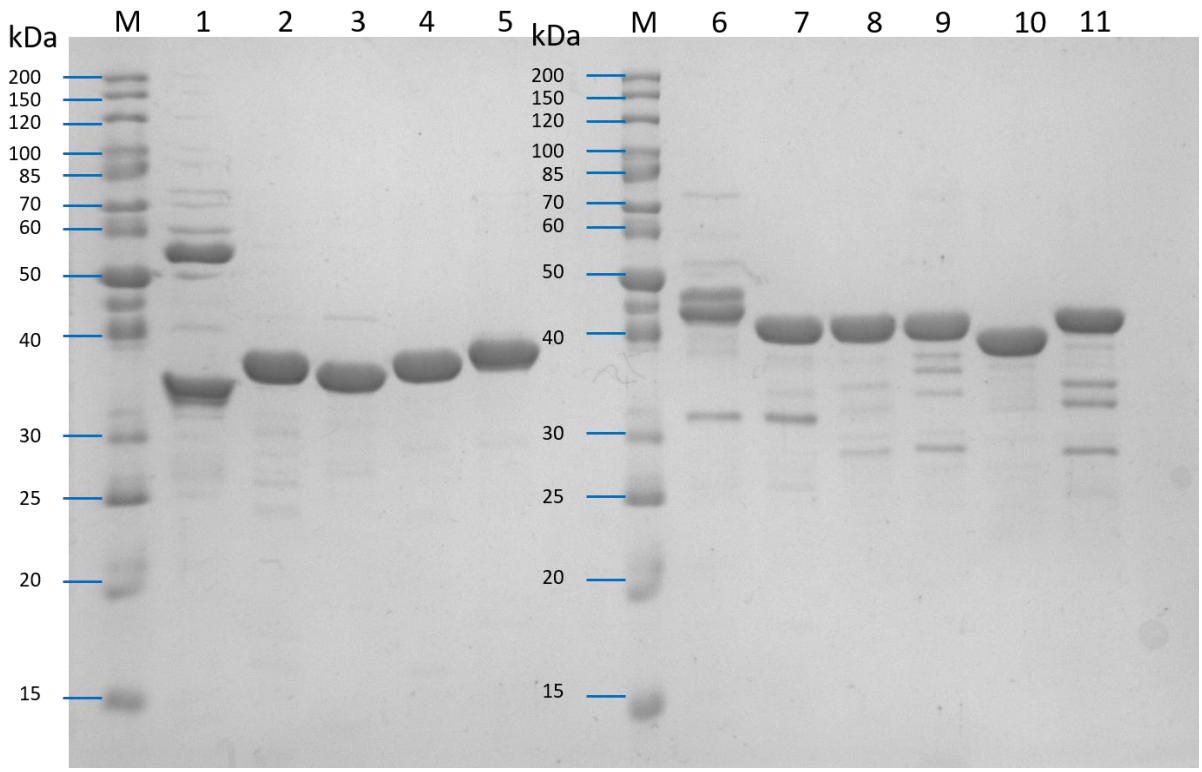


Figure S1. SDS-PAGE of all L-KdpDs (left) and D-KdpDs (right) after purification and desalting. From left to right: Marker, *Ab*L-KdpD, *Cn*L-KdpD with N-terminal His-tag, *Cn*-L-KdpD with C-terminal His-tag, *Hs*L-KdpD, *Vp*L-KdpD, Marker, *Cc*D-KdpD1, *Cc*D-KdpD2, *Px*D-KdpD, *Pp*D-KdpD, *Hs*D-KdpD, *Cn*D-KdpD.

Table 2. Apparent (APP) kinetic characterization of L-Kdp and D-Kdp dehydratases toward their non-preferred stereoisomer. All measurements were performed in triplicate at 25 °C, in 50 mM HEPES pH 7.5.*

Substrate	Enzymes	k_{cat}^{APP} (s^{-1})	K_M^{APP} (mM)	k_{cat}/K_m ($mM^{-1}s^{-1}$)
D-KDP	*Ab*L-KdpD	6.65 ± 0.10	6.81 ± 0.35	0.98 ± 0.23
	*Cn*L-KdpD (Nhis)	12.04 ± 0.16	4.08 ± 0.21	2.95 ± 0.23
	*Cn*L-KdpD (Chis)	11.01 ± 0.08	6.40 ± 0.17	1.72 ± 0.16
	*Hs*L-KdpD	11.38 ± 0.06	4.41 ± 0.09	2.58 ± 0.14
	*Vp*L-KdpD	11.78 ± 0.19	4.99 ± 0.30	2.36 ± 0.25
L-KDP	*Cc*D-KdpD1	7.99 ± 0.08	2.81 ± 0.12	2.84 ± 0.21
	*Cc*D-KdpD2	9.67 ± 0.08	3.96 ± 0.13	2.44 ± 0.18
	*Px*D-KdpD	18.73 ± 0.15	5.08 ± 0.17	3.69 ± 0.18
	*Pp*D-KdpD	19.64 ± 0.17	6.12 ± 0.21	3.21 ± 0.19
	*Hs*D-KdpD	16.92 ± 0.21	3.18 ± 0.18	5.32 ± 0.23
	*Cn*D-KdpD	4.24 ± 0.07	2.49 ± 0.20	1.70 ± 0.28

* Error bars represent standard deviation from three replicates. Nonlinear regression of the enzyme activity as a function of substrate concentration is presented in Figure S2.

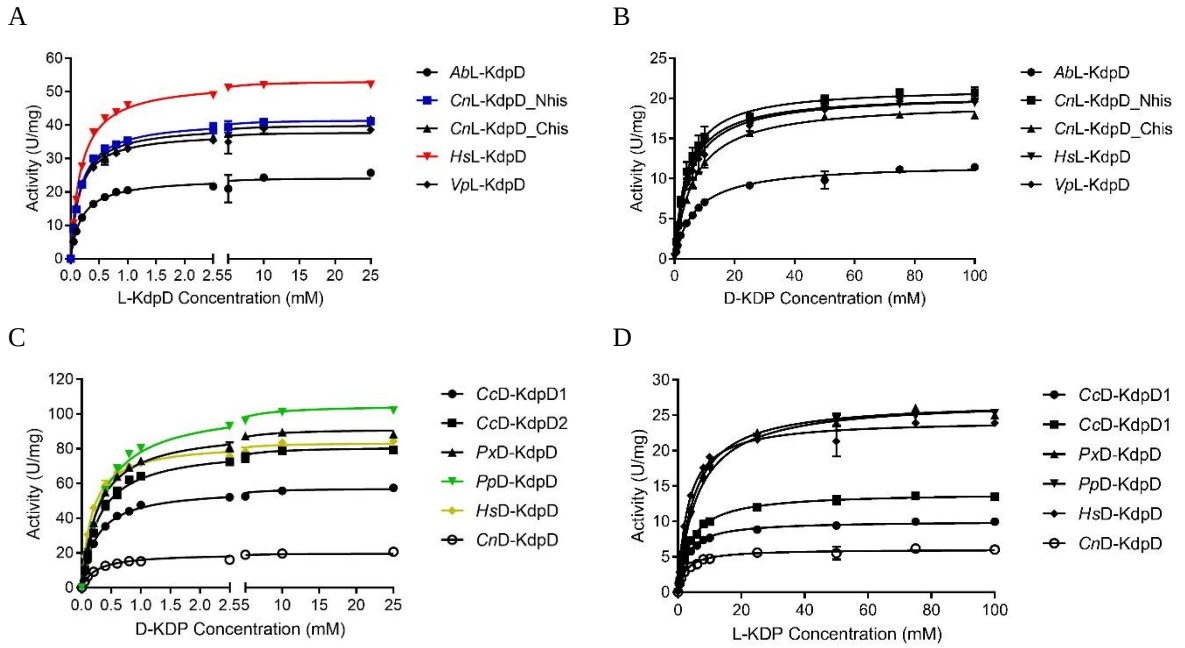


Figure S2. Michaelis-Menten kinetics of all dehydratases used in this study. Error bars represent standar deviation from three independent replicates. Activity was determined by a coupled-assay with *Pp*KGSADH. One unit is defined as the amount of enzyme to convert 1 µmol of NAD^+ per minute. The k_{cat}-values presented in Table 1 were calculated based on the specific activity (U/mg) obtained from this graph divided by the enzyme used. All experiments were performed in HEPES 50 mM pH 7.5, 25 °C

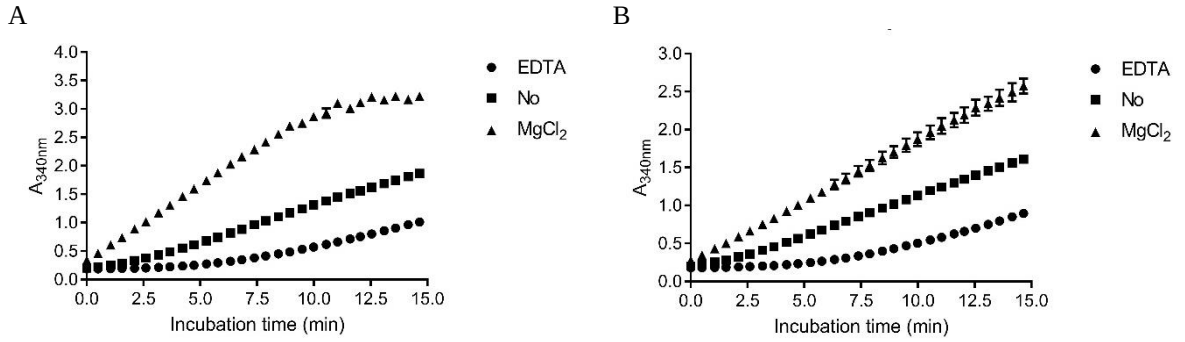
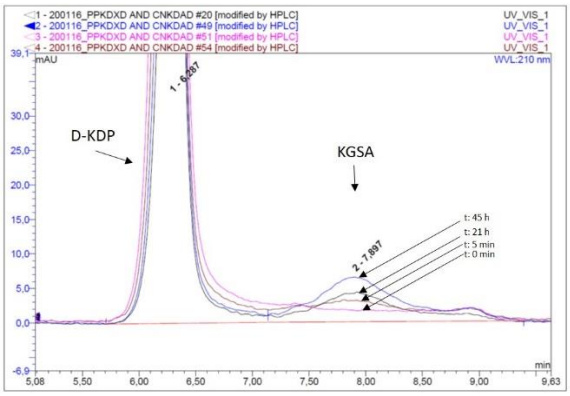


Figure S3. Effect of pre-incubation with EDTA as well as absence, and presence of $MgCl_2$ on the initial activity of *Pp*D-KdpD (A) and *Hs*D-KdpD (B). There is apparent delay in the initial activity of *Pp*D-KdpD and *Hs*D-KdpD after pretreatment with EDTA (no magensium presents in the assay) and when magnesium was absent during the assay. Thus, for the initial activity determination, the enzymatic activity (EDTA and No) was calculated from minute 8 to 12. Error bars represent standard deviation for three replicates.

Table S3. Effect of the presence of EDTA, absence of Mg^{2+}, and presence of Mg^{2+} on the kinetic (T_{50}^{1h}) and thermodynamic (T_m) stabilities of respective L-KdpDs and D-KdpDs as presented in Figure 1 D-F

Enzyme	T_{50}^{1h} (°C)			T_m (°C)		
	EDTA	No	Mg	EDTA	No	Mg
*Cn*L-KdpD	53.2 ± 0.1	53.2 ± 0.0	53.1 ± 0.0	53.5 ± 0.0	54.5 ± 0.0	55.0 ± 0.0
*Hs*L-KdpD	48.0 ± 0.1	47.3 ± 0.1	47.1 ± 0.1	57.0 ± 0.0	54.5 ± 0.0	54.0 ± 0.0
*Pp*D-KdpD	45.5 ± 0.4	47.7 ± 0.0	48.0 ± 0.2	46.8 ± 0.3	48.8 ± 0.3	54.3 ± 0.6
*Hs*D-KdpD	40.1 ± 0.1	41.2 ± 0.0	42.5 ± 0.7	41.3 ± 0.3	43.7 ± 0.3	48.8 ± 0.3

A



B

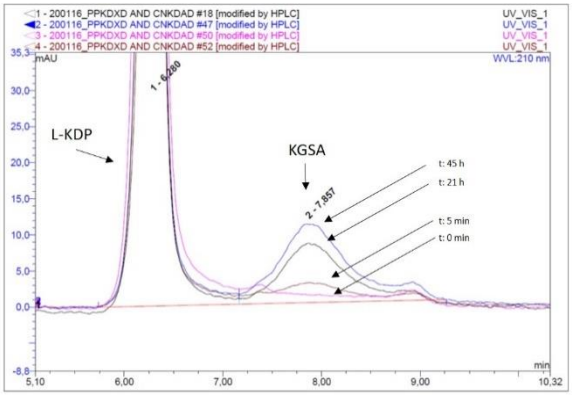


Figure S4. Chromatogram of the conversion of D-KDP (non-preferred stereoisomer) to KGSA by *Cn*L-KdpD (A) and L-KDP (non-preferred stereoisomer) to KGSA by *Pp*D-KdpD (B). D-KDP and L-KDP showed the same retention time. The conversion of D-KDP to KGSA by *Cn*L-KdpD appeared to be severely slow in comparison to the initial activity presented in Table 2. Similar results were observed for the conversion of L-KDP to KGSA by *Pp*D-KdpD. Retention times of D- or L-KDP are the same at 6.28 min while retention time of KGSA is at 7.86 min.

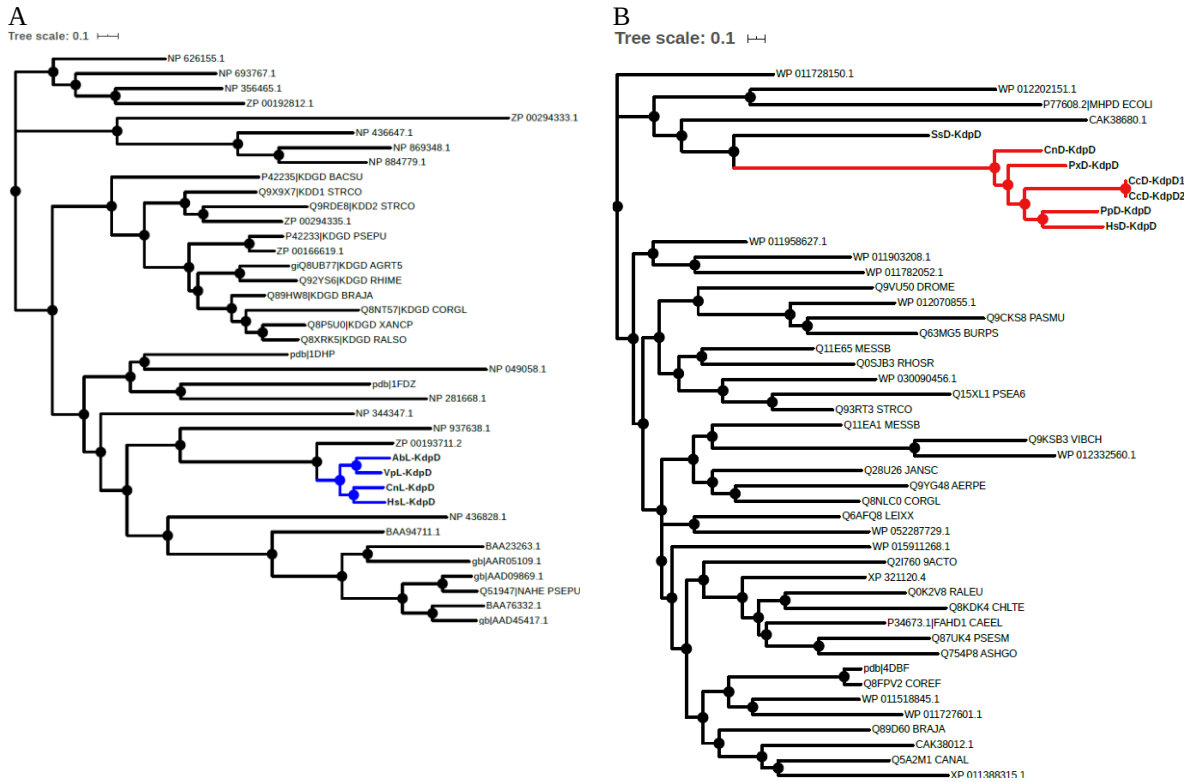


Figure S5. Phylogenetic tree of L-KdpDs (blue clades) in relation to other enzymes in dihydrodipicolinate synthase super family (DHDPS) (A) and D-KdpDs (red clades) in relation to enzymes in fumarylacetocatate hydrolase super family (FAH) (B). The sequences of other members of each family are retrieved from the Conserved Domain Database (Marchler-Bauer et al. 2016). Sequence alignment and phylogenetic tree construction were performed using MEGA-X software (Kumar et al. 2018). Sequence alignment was performed using the MUSCLE algorithm and the phylogenetic tree was developed using Maximum Likelihood algorithm embedded in MEGA-X using default parameters (Jones et al. 1992; Edgar 2004). The trees in Newick format were visualized using iTOL (Interactive Tree Of Life) (Letunic and Bork 2019).

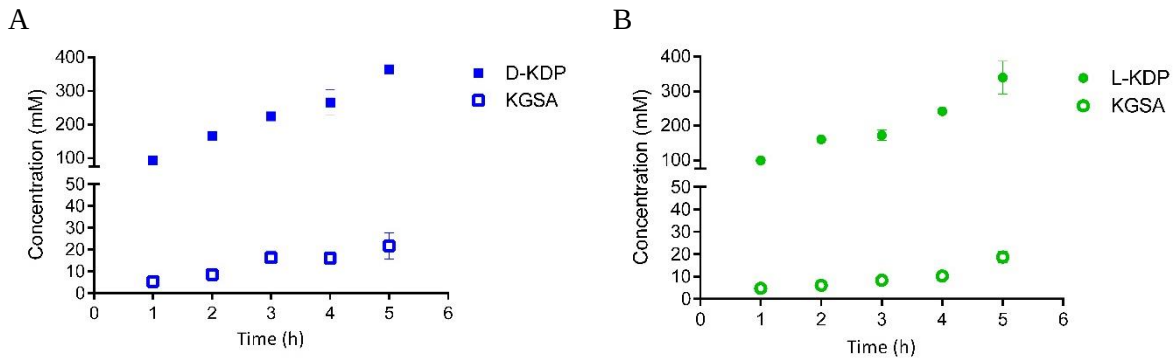


Figure S6. Effect of titration of D-KDP to *Cn*L-KdpD (A) and L-KDP to *Pp*D-KdpD (B). After each addition of the substrates, both solutions were incubated for 1 h, prior to analysis of KGSA formation via HPLC. Higher formation of KGSA was observed when more substrate was added for both enzyme indicating the KGSA formed could be from the impurity, i.e. the other KDP isomer.

References

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113 . https://doi.org/10.1186/1471-2105-5-113

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275–282 . https://doi.org/10.1093/bioinformatics/8.3.275

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549 . https://doi.org/10.1093/molbev/msy096

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259 . https://doi.org/10.1093/nar/gkz239

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203 . https://doi.org/10.1093/nar/gkw1129

4 Discussion and Outlook

4.1 Evolving protein stability: lesson from decarboxylases

An enzyme´s stability is one of the prerequisites for its industrial utilization (see subchapter 1.1.1.2). There are numerous different types of stability, such as thermostability, stability in organic solvents, stability against chaotropic agents, and stability from protease degradation. These different stabilities are often related. Increasing stability toward one parameter, will likely increase stability toward others.[76,198,199] In this work, thermostability of two thiamine diphosphate (ThDP)-dependent decarboxylases needed to be improved to realize the production of ethanol and isobutanol at high temperatures. 70 °C was the target temperature to be achieved as this temperature is an ideal temperature for *Sulfolobus acidocaldarius* (*Saci*) to grow.[116] This temperature is closer to the boiling point of ethanol; thus, producing ethanol at this temperature will ease product separation.

Several strategies are described in literature to improve protein thermostability. The easiest approach probably is to find an already thermostable enzyme (thermozyme) from thermophilic microorganisms or to engineer an enzyme from mesophilic microorganisms using homologous enzyme found in thermophilic microorganisms as a guide. There are distinct features of enzymes found in thermophiles in comparison to the ones found from mesophiles, such as less serine and cysteine, more arginine and tyrosine, increased salt-bridges, and increased hydrogen bonds.[200] Unfortunately, prior to this study, there were no homologues of branched-chain α-keto acid decarboxylase (KDC) and pyruvate decarboxylase (PDC) reported from thermophilic microorganisms. Thus, both enzymes had to be engineered. Both enzymes belong to the same ThDP-dependent decarboxylase, but each enzyme shares only 30% protein identity. In this study, the first work was focused to engineer KDC instead of PDC as for ethanol production in *Saci*, there are alternative pathways other than *via* PDC.[136] However, for isobutanol production, the pathway *via* KDC was the only option available (the modified Ehrlich Pathway).

The question of engineering an enzyme toward improved thermal stability often is not “if” but more “how”. The mostly employed strategies can be divided roughly in two categories.[86,201] Directed evolution is a strategy that is based on the random creation of variants and relies on a screening system to select one or several variants with improved stability. Rational design, on the other hand, is an approach that is based on the knowledge of a crystal structure of an enzyme to design variants with improved stability rationally. During this study, one isoenzyme of KDC from *Lactococcus lactis* (KivD) has been engineered toward improved stability.[158] Another work on KivD also resulted in a variant with improved stability.[159] In these two important works, directed evolution coupled with computational prediction were employed. Although significant improvements have been achieved in there, the stability of the variants reported did not reach the

target temperature of 70 °C. Thus, in this study another approach was taken with the motivation to achieve higher thermostability than the two previous works.

Structure-based engineering was applied in this work. In this approach, structure of another isoenzyme of KDC from *L. lactis* (KdcA) was used to rationally target amino acid substitution.[126] In order to maximize the chance to find stabilizing substitutions, site saturation mutagenesis (SSM) was performed.[59] After several rounds of SSM, a variant (7M.A) with improved T_{50}^{1h} (temperature at which an enzyme retains 50% of its initial activity after 1 h incubation) and T_m (melting temperature) by 13.8 and 13.5°C, respectively was obtained. Albeit its increased thermostability, 7M.A was surprisingly severely inhibited by isobutanol. Retrospective study of this variant found that a substitution at position 290 from alanine to methionine (A290M) contributed strongly to the inhibitory effect of isobutanol. SSM at this position with an additional screening plate to measure initial activities of the library in the presence of isobutanol resulted in two variants (7M.C and 7M.D) with similar level of thermostability to 7M.A, but not to be inhibited by isobutanol. 7M.C and 7M.D have a substitution to valine and cysteine at position to 290, respectively. This result underlines the fundamental principle in protein engineering: "you get what you screen for". As the plate to measure initial activity in the presence of isobutanol had not been used in the first screening rounds, such subtitutions—to valine and to cysteine—were overlooked by substitution to methionine due to a higher initial activity of the latter.

Amino acid at position 290 is located near the surface of KdcA. Hence, the inhibitory effect of isobutanol to the methionine substitute (7M.A) can be rationalized by stronger interaction of isobutanol to a large non-polar amino acid in comparison with a smaller non-polar amino acid, e.g. valine (7M.C) or cysteine (7M.D), thus promoting a structural disturbance. 7M.C and 7M.D demonstrated IC_{50}^{1h} (concetration of inhibitor that gives 50% inhibotry effect on enzyme initial activity) of isobutanol at 50 °C of 7 vol% and 5 vol%, or >7-fold and >5-fold higher than 7M.A. 7M.D was chosen for further characterization due to its higher initial activity than 7M.C. In addition, 7M.D demonstrated slightly higher T_{50}^{1h} than 7M.A (1 °C higher). This variant showed a half-life of 2 h at 70 °C. From this stepwise approach, it was demonstrated that all stabilizing substitutions identified are additive. This means that combination of stabilizing substitutions resulted in variants with further improved thermostability.[202] 7M.D also showed increased stability in the presence of isobutanol. At 4 vol% isobutanol at 50 °C, 7M.D showed a half-life of 14.3 h or >600-fold improvement in comparison with WT *Ll*KdcA.

Contrary to KDC, two significant attempts have been described in literature to improve stability of a PDC based on rational design. The first work utilized a sequence-based approach to increase stability of PDC *via* ancestral sequence reconstruction (ASR).[140] This approach, however, did not result in variants with improved thermostability. Another work utilized Rosetta-based design and reported significant improvement of a PDC from *Zymomonas mobilis* (*Zm*PDC) after

introducing 16 point substitutions.[143] However, this variant was cloned and characterized in this study and shown not to be thermostable. Both T_{50}^{1h} and T_m analyses of the variant (5TMA) resulted in the same conclusion that the variant was less stable than WT *Zm*PDC. It was difficult to rationalize this major discrepancy. One of the possible reason is that the authors in the original study used rather an uncommon approach to determine the stability of the variant. The authors used differential interference contrast (DIC) microscopy that monitors different states of a protein in solution, e.g. soluble or aggregated.[143] This analysis approach could potentionally result in an incorrect conclusion to be drawn as it is generally known that not all enzymes when destabilized will form an aggregation/precipitation. It is worth noting that in this study, 5TMA did not show any precipitation during heat challenges, although its remaining activities were completely eliminated. Thus, DIC analysis of 5TMA failed to reflect its actual kinetic stability, a more important parameter when comparing enzymes for applications at elevated temperatures. Consequently, prior to this study there was no PDC reported in literature with kinetic thermostability high enough to withstand the growth temperatures of *Saci*.

Instead of a structure-based approach like the engineering of KdcA, another approach was performed. StEP (staggered extension process) was chosen to shuffle the most stable PDC variants.[52] Comparing yeast to bacterial PDCs, the latter showed significantly higher activity and thermostability; thus, three bacterial PDCs were used.[122,203] The hypothesis underlying this strategy was that throughout evolution, stabilizing substitutions are often retained. There would be no benefit for nature to evolve an enzyme toward decreased stability. Even in mesophilic microorganism when activity was perhaps the evolutionary drift in enzyme evolution, having a stable enzyme is not disadvantageous. Thus, stability would only be given up to gain activity.[63]

After performing StEP of *Zm*PDC, *Zp*PDC (PDC from *Zymobacter palmae*), and *Ap*PDC (PDC from *Acetobacter pasteurianus*), a more stable variant was obtained. This variant (PDC-Var.1) showed an improved T_{50}^{1h} and T_m by 8.2 and 7.0 °C, respectively. PDC-Var.1 shares 98% protein identity to *Ap*PDC. Crossover analysis indicated that all beneficial substitutions are obtained from *Zp*PDC (subchapter 3.2). Recombinant event is rather low, possibly due to additional elongation time in comparison to the original StEP protocol.[53] Without this step, it was not possible to obtain any PCR product due to low sequence similarity among the three PDCs (<70%). Although, this seemed to diverge from the principle of StEP itself, the finding of a more stable variant using this modification may suggest that the compromise between recombinant event and sequence diversity worked. Thus, an elongation step could be added to allow recombination of proteins with low protein similarity (<70%).

Since PDC and KDC belong to the same ThDP-dependent decarboxylase, seven point stabilizing substitutions previously identified in KDC were used to improve stability of PDC-Var.1 further. Surprisingly, only one position seems to be important toward increased thermostability

in both enzymes. I434C in *Ll*KdcA and I441V in *Ap*PDC are the homologous substitution contributing to increased thermostability. This amino acid seems to be conserved across KDC and PDC as this amino acid is in the conserved ThDP binding region.[126] This position would have never been found *via* a sequence alignment-based approach to improve stability. This position was identified during structure-based engineering approach in *Ll*KDcA as a possible amino acid that could improve stability *via* increased dimer interaction.[202] This result may suggest that careful examination of a crystal structure coupled with educated guess could be a viable approach to improve the stability of a protein. The fact that different substitutions in the homologous position between *Ll*KdcA and *Ap*PDC observed further establishes SSM as a better approach than a site directed substitution.[59] The most stable PDC variant (PDC-Var.2) showed half-lives of 10.7 h and 7.3 h at 70 °C and 75 °C, respectively.[203]

Of all characterized and identified KDC and PDC, a relationship between T_{50}^{1h} and T_m was analyzed. T_m is a descriptor for thermodynamic stability while T_{50}^{1h} is for kinetic stability. It has been shown in literature that these parameters sometimes do not correlate to each other.[30] Since determination of T_m based on thermofluor is much faster than determination of T_{50}^{1h}, a simpler screening system could be build given the case of direct relation between these two parameters. For KdcA, a strong correlation between T_{50}^{1h} and T_m of the WT and variants obtained from structure-based engineering approach was observed (R^2=0.988) (blue circles). A similar strong correlation is also observed for PDC (R^2=0.992) (red square). When KDC and PDC are combined, a strong correlation could still be observed (R^2=0.991) (Figure 4.1). This suggests the direct relationship was able to be established for ThDP-dependent decarboxylases. The importance of this analysis becomes more apparent as for KDC and PDC, improved kinetic stability was also translated to improved stability in organic solvents.[202,203]

Isobutanol stability of the enzymes used in *in vitro* cell-free isobutanol production is perhaps one major advantage of using cell-free approach for the production of this chemical. *E. coli* and many other microorganisms are inhibited in the presence of as low as 1 vol% isobutanol.[117,149,151] Although the most thermostable KDC variant (7M.D) showed significantly improved stability toward isobutanol, the projection is to evolve this variant further to withstand 9 vol% isobutanol. At this concentration, a two-phase isobutanol/water will start to occur. Production of isobutanol at its solubility limit will ease product removal, thus lowering production cost of isobutanol. When it is achieved, it will serve as a “game-changer” in isobutanol production. Therefore, with direct relationship between T_{50}^{1h} and T_m described in KDC, future screening system can benefit T_m as the predictor to find variants with increased thermostability and stability toward isobutanol. PDC-Var.2 already showed remarkable stability in the presence of 9 vol% n-butanol. PDC has been applied toward production of n-butanol.[194,195] This variant could also be further improved its stability toward even higher thermostabilty using the same screening

approach. It is always of scientific interest to learn to what extend a protein could be improved its thermostability.[204] Proteins that are stable at boiling point have been reported in literature, so going beyond 100 °C is already within our reach.[205]

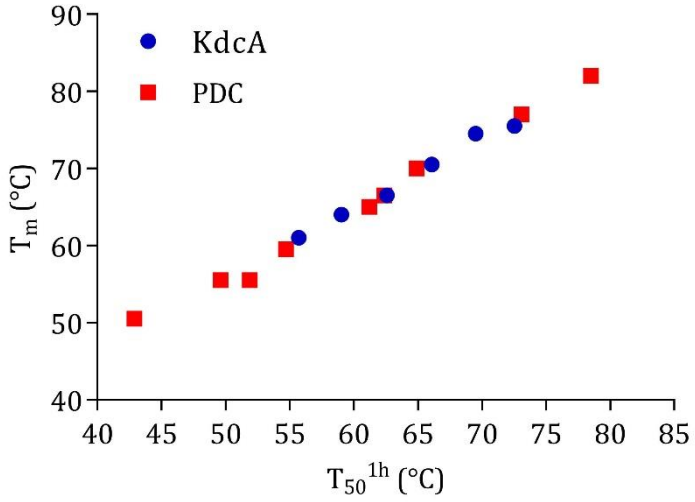


Figure 4.1. Relation between T_{50}^{1h} and T_m in ThDP-dependent decarboxylases. Blue circles represent LlKdcA WT and its more thermostable variants generated in this study. Red squares represent all WT PDC variants and two PDC variants with increased thermostability evolved in this study.

4.2 New [Fe-S]-dependent dehydratases: sequence and promiscuity comparison

It has been generally known that protein sequences are more conserved than nucleotide sequences encoding the proteins. It is also already a consensus that in a protein family, a three dimentional structure is more conserved than a protein sequence. However, not all protein structures of a family are readily available. Sequence information, on the other hand, is widely available due to the rapid development in genome sequencing technology. Thus, it is always beneficial to examine how sequences relate to activities in a protein family. [Fe-S]-dependent dehydratases, in particular IlvD/EDD superfamily comprise three to four distinct subfamilies that are thought to be evolutionary related.[169] The first subfamily is EDD that includes all 6-phosphogluconate dehydratases (6-PGDHT), enzymes responsible in the classic Entner-Doudoroff (ED) pathway (Figure 1.9). The second class is dihydroxy-acid dehydratases (DHAD), enzymes catalyzing the dehydration of (R)-2,3-dihydroxyisovalerate (DHIV) and (2R,3R)-dihydroxy-3-methylvalerate (DHMV) in branched-chain amino acids (BCAA) biosynthesis. The third subfamilies have been recently established. This subfamily consists of enzymes catalyzing dehydration of sugar acids. D-xylonate and L-arabinonate dehydratases are two prime examples of this subfamily, that are recently characterized.[169] Their crystal structures have also been solved.[206,207] The last class is another D-xylonate dehydratase with XylDHT from *E. coli* being the only member of this subfamily characterized. Due to only limited number of reports on this

subfamily, more work would be needed to establish this enzyme into a subfamily of IlvD/EDD superfamily.

In this study, several new [Fe-S]-dehydratases have been cloned, purified, and partially characterized. These dehydratases were then compared to other dehydratases belonging to IlvD/EDD superfamily and a phylogenetic tree was built. The first class of the identified dehydratases appears as a new branch in a supposed [2Fe-2S]-dependent DHAD (denoted by dark green branch). This class consists of all dehydratases that were denoted as D-glycerate-converting class in subchapter 3.3 due to their much higher activity toward D-glycerate than toward D-gluconate.[208] Later work identified the most promising member, *Ft*DAHD as a DHAD due to its significantly higher activity toward DHIV than toward D-glycerate (Table 4.1). Thus, based on activity and their relative position on the IlvD/EDD phylogenetic tree (Figure 4.2), *Ft*DHAD and all homologous enzymes in D-glycerate-converting class is reclassified as a new branch of [Fe-S]-dependent DHAD.

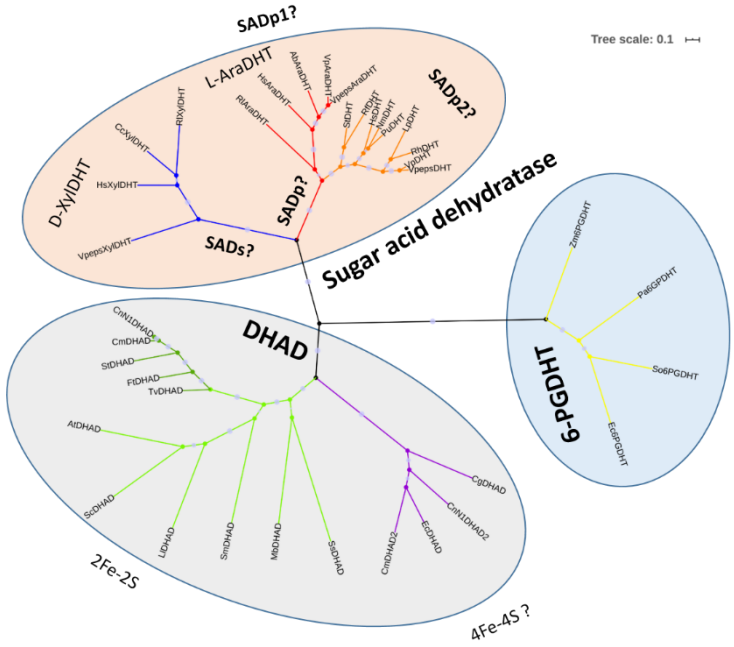


Figure 4.2. Phlyogenetic tree of IlvD/EDD superfaymily. Different colors of branch was used to help visualization. The tree is produced using MEGA and visualized using iTOL.[209,210] Abbreviation of the organisms are presented in Appendix 6.3. DHAD is dihydroxy-acid dehydratase, 6-PGDHT is 6-phosphogluconate dehydratase, SADs is a proposed specific sugar acid dehydratase class, and SADp is a proposed promiscuous sugar acid dehydratase class.

The second class of dehydratases identified is called promiscuous class dehydratase in subchapter 3.3.[208] A notable example is a dehydratase from *Paralcaligenes ureilyticus* (*Pu*DHT). *Pu*DHT demonstrated the highest activity toward D-glycerate and D-gluconate in this study. In the phylogenetic tree, this enzyme belongs to the sugar acid dehydratase branch. The sugar acid dehydratase subfamily was divided further into two classes depending on the substrate preference toward epimers at C4 of pentonates in the previous work. D-xylonate dehydratase

(blue branch) is an enzyme class that prefers D-xylonate to L-arabinonate with D-xylonate dehydratase from *Caulobacter crescentus* (*Cc*XylDHT) as the prime member of this class.[169] The other class is L-arabinonate dehydratase (red branch) with a reverse substrate preference. L-Arabinonate dehydratase from *Rhizobium leguminosarum* (*Rl*AraDHT) is an example of this class.[169]

Table 4.1. Activity summary of different [Fe-S]-dependent dehydratases

Enzyme*	**k_{cat} (s^{-1})**				
	D-Gluconate	D-Xylonate	L-Arabinonate	DHIV	D-glycerate
***Rl*AraDHT**	2.32 ± 0.21	3.74 ± 0.06	8.9 ± 1.03	N.D.	<0.01
***Cc*XylDHT**	70.11 ± 0.41	38.70 ± 0.02	0.93 ± 0.00	N.D.	<0.01
***Ss*DHAD**	0.40	5.07 ± 0.13	2.68 ± 0.39	0.31	0.01
***Ft*DHAD**	0.01 ± 0.00	0.37 ± 0.00	0.66 ± 0.01	8.05 ± 0.199	0.67 ± 0.01
***Pu*DHT**	130.32 ± 4.7	72.91 ± 3.7	33.22 ± 1.81	0.03 ± 0.00	0.99 ± 0.03

*All activities were measured at 50°C in HEPES 50 mM, pH 7.5 in the presence of 25 mM substrate except for *Rl*AraDHT and *Cc*XylDHT that were measured at 37 °C due to their limited stability.

*Pu*DHT showed higher preference for D-xylonate than to L-arabinonate (Table 4.1). However, from the phylogenetic tree this enzyme and its subsequent homologues share the same node as of L-arabinonate dehydratase (Figure 4.2). Further analysis of the tree revealed that although this enzyme class may share the same ancestor (emerged from the same node) to L-AraDHT class, this class formed a distinct branch. Thus, in L-arabinonate branch (red branch) two distinct nodes are observed. The first node (red branch) encompasses all homologous enzymes of *Rl*AraDHT and the second node emerges all homologous enzymes of *Pu*DHTs. Substrate promiscuity analysis further demonstrates that indeed *Pu*DHT is closer to *Rl*AraDHT than to *Cc*XylDHT. *Pu*DHT activity toward L-arabinonate is only two-fold lower than to D-xylonate. This is also true for *Rl*AraDHT that showed only 2-fold lower activity toward D-xylonate than L-arabinonate. *Cc*XylDHT activity toward L-arabinonate, however, is 40-fold lower than to D-xylonate (Table 4.1). This may indicate, in addition to protein sequence, comparing relative substrate promiscuity is more descriptive than comparing absolute substrate preference in sugar acid dehydratases subfamily. Comparison to previously published phylogenetic tree of IlvD/EDD superfamily, *Pu*DHT and its subsequent homologues (described as promiscuous class in subchapter 3.3) may be located in the magenta branch that is assigned as a class with unknown substrate promiscuity (Appendix 6.1).[169] Because L-AraDHT and the promiscuous class have emerged from the same node, it is suggested that the branch that has emerged from the node is renamed as promiscuous sugar acid dehydratase (SADp) due to their promiscuous nature. Hence, L-AraDHT can be renamed as SADp1 and the promiscuous class (DHT) can be renamed as SADp2. D-Xylonate dehydratase from *C. crescentus* and its homologues could thus be named as specific sugar acid dehydratase (SADs). More detailed characterization of this enzyme class should reveal

if all members of this class is only specific to D-xylonate or there is another branch that is specific to L-arabinonate.

After analysis of the phylogenetic tree (Figure 4.2), it would be interesting if from sequence alignment, signature motifs could be deduced to predict activity and or promiscuity as well as state of the [Fe-S] state and in turn oxygen susceptibility. [Fe-S]-cluster is postulated to be coordinated by three cysteines.[160] In the sequence alignment (Figure 4.3), three conserved cysteines are observed: C57, C125, and C198 (*Pu*DHT numbering). In the first cysteine, three distinct motifs are observed **P<u>C</u>N**, **PGH**, and **SAH**. **SAH** is the motif identified to 6-PGDHT (Figure 4.3). **PGH** is identified to DHAD that is reported to have [4Fe-4S], with *E. coli* DHAD being characterized best.[211] Homology modelling of *Ec*DHAD to a recently solved crystal structure of DHAD from *Mycobaterium tuberculosis* (*Mt*DHAD) demonstrated that from all twelve cysteines in *E. coli* DHAD, only two cysteines are located within 10 Å of the [Fe-S] cluster.[164] These cysteines are homologous to C125 and C198. Thus, it was proposed that the third coordinating amino acid to the [Fe-S]-cluster is the histidine in **PG<u>H</u>** motif. Histidine appears to be also conserved in 6-PGDHT further supporting this hypothesis. Coordination of a histidine to a [Fe-S]-cluster has been reported for several enzymes.[164] Another enzyme with **PG<u>H</u>** motif that was partially characterized in this study is *Cn*N1DHAD2 (DHAD2 from *Cupriavidus necator* N-1). This enzyme is not the same as *Cn*N1DHAD (part of D-glycerate-converting class in subchapter 3.3). *C. necator* N-1 has multiple annotated dehydratases. *Cn*N1DHAD2 was successfully purified under aerobic condition and the typical yellow color was relatively stable during purification. This enzyme, however, showed >10-fold lower activity toward D-glycerate, L-arabinonate, and D-xylonate than *Cn*N1DHAD. This enzyme was not checked toward DHIV, a natural substrate of a DHAD. Thus, its significantly lower activity could be caused by the instability of its [Fe-S]-cluster further supporting the hypothesis that [4Fe-4S]-cluster in a DHAD is susceptible to oxygen or simply by the fact that not a correct substrate was checked.

All dehydratases bearing **P<u>C</u>N** motif are reported to contain [2Fe-2S], except L-arabinonate dehydratase from *Azospirillum brasilense* (*Ab*AraDHT, member of proposed SADp1).[191] This enzyme was shown to bear [4Fe-4S] instead. This enzyme was reported to be relatively stable as purification was performed aerobically and the enzyme was shown to be active. This result seems to contradict the general belief that [4Fe-4S] cluster is oxygen instable. One explanation could be that [4Fe-4S] of *Ab*DHT was actually degraded over time to form a stable [2Fe-2S] and this cluster was the one that was active during enzyme characterization. Oxidation of [4Fe-4S] to yield a stable [2Fe-2S] was reported for a [Fe-S]-dependent transcriptional regulator and several regulatory enzymes in *Mycobacterium tuberculosis*.[212–214] Another explanation, that perhaps is more likely, is that the stability of a [Fe-S] actually depends on the protein structure, thus the protein sequence and not the type of a cluster. If a protein structure allows protection of the cluster from direct

contact to atmospheric or dissolved oxygen, the [Fe-S]-cluster would be more stable.[211] The latter hypothesis is also partially supported by the fact that although *Mt*DHAD was reported to have [2Fe-2S], the cluster is susceptible to oxygen.[164]

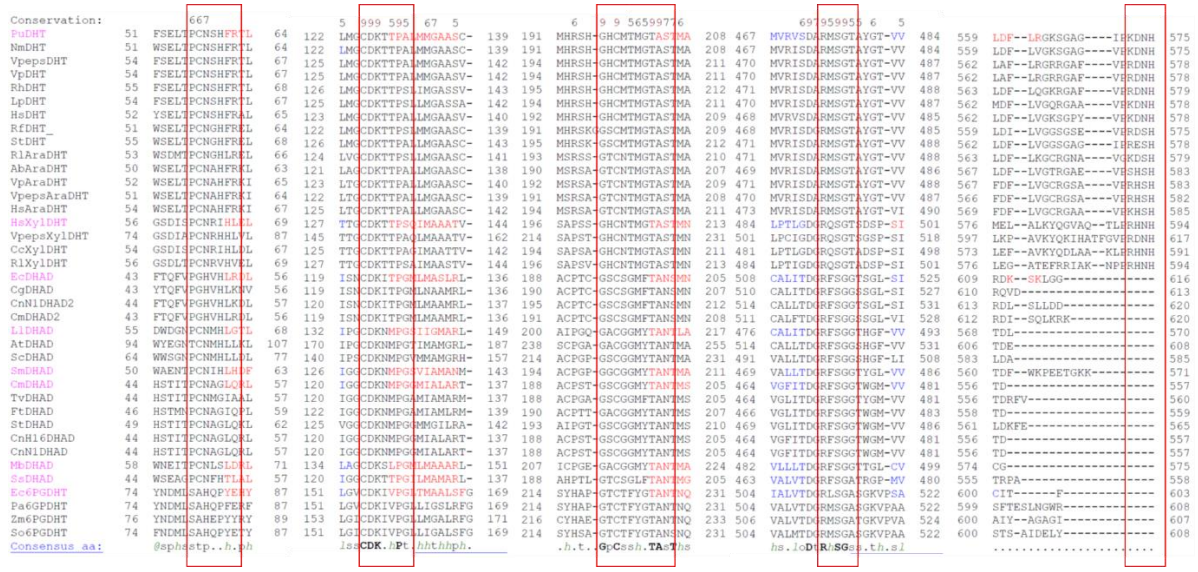


Figure 4.3. Sequence alignment of all dehydratases in this study in comparison to other dehydratases described in literatures. Possible conserved motifs are highlighted red. The full sequence alignment is presented in Appendix 6.2. Abbreviation of the organisms are presented in Appendix 6.3.

Closer analysis of sequence alignment, it could be concluded that all XylDHTs (SADs) that were characterized in this study shows **PCNRXH** while promiscuous sugar acid dehydratases (SADp) shows **PCNXHXR** (X represent any amino acids, unless otherwise stated). No clear distinction between SADp1 and SADp2 can be identified from the extended **PCN** motif. *Ft*DHAD that was active toward DHIV and reasonably active toward D-glycerate (Table 4.1) seems to have **PCNXG** motif when compared to its homologous enzymes (D-glycerate-converting class in subchapter 3.3). The glycine residue appeared to be conserved in this subclass as other DHADs have histidine at this homologous positons (except for *Mt*DHAD that has serine). In the second cysteine, C150, **CDK** motif appears to be conserved in all IlvD/EDD superfamily. All sugar acid dehydratases show **CDKTTP** motif while all DHADs except *Mt*DHAD have **CDKNMPG**. From this sequence motif, SADs can be differentiated from SADp as the first bears **TGCDKTTP** and the latter has **TXCDKTTP**. The X in SADp1 includes Val, Ala, or Thr while for SADp2, a conserved M is observed. In the last conserved cysteine, C200, several consensus motifs can be summarized. SADs has **GHCNTMG**, SADp1 has **GTCNTMG**, while SADp2 shows **GHCMTMG** (except for *St*DHT and *Rf*DHT with the histidine is replaced by serine, instead). All DHADs with supposed [4Fe-4S] shows motif of **GSCSGMF** and all DHADs with supposed [2Fe-2S] shows **GXCGGMY**. Motif of **GTCTFYG** was observed for 6-PGDHT. Next to this motif, three following conserved amino acid sequences, **TAS** are observed in all sugar acid dehydratases while other DHADs and EDDs show **TAN** motif.

IlvD/EDD superfamily is postulated to use a serine to deprotonate hydrogen located at Cα and [Fe-S]-cluster as a lewis acid to attack hydroxyl group of Cβ.[160,167,168] This serine is located at position 476 in *Pu*DHT and its relative position is conserved in all IlvD/EDD superfamily. Again, a distinct motif around this conserved serine could be identified. SADs has **Q<u>S</u>GT** motif while SADp has **M<u>S</u>GT**. All DHAD enzymes have **F<u>S</u>GG** and EDD enzymes show **X<u>S</u>GG**. The last differentiation that can be drawn is the histidine located at the C-terminus of all sugar acid dehydratases. This histidine does not appear in [2Fe-2S]-dependent DHADs. In fact, based on the sequence alignment, it appeared that all sugar acid dehydratases seem to have "elongated" C-terminis. Crystal structures of *Cc*XylDHT (a member of SADs) and *Rl*AraDHT (a member of SADp1) showed that this C-terminis protrudes into the catalytic active site.[206,207] The histidine is likely to play a role in catalytic activity and substrate stabilization.

From all analysis above, it was able to be concluded that there are indeed differences and conservations among different subclasses. It was rather straightforward to differentiate whether an enzyme would be a sugar acid dehydratase or a DHAD. Closer analysis also reveals promiscuity of sugar acid dehydratases can be further predicted based on the sequence to discriminate SADs from SADp. However, to predict whether an SADp enzyme would be more active toward D-xylonate or L-arabinonate is proven to be more challenging from the sequence only. Substrate docking studies would be required to predict which amino acids have a role to discriminate the two epimers. *Ft*DHAD and its subsequent homologues are promising DHADs in respect to their activity toward D-glycerate. However, other DHADs have not been checked their activity toward D-glycerate; thus, its relevance in the sequence based prediction to discover "D-glycerate-converting" DHAD awaits further studies on other DHADs.[164,215] Whether the type of [Fe-S] cluster could be predicted based on the few sequence motifs described above is not yet conclusive. However, characterizations of *Ab*AraDHT and *Mt*DHAD possibly suggest that the stability toward oxygen cannot be based solely on the type of [Fe-S]-cluster.[164,191] Protein structure may have a bigger role in determining the stability of a [Fe-S]-cluster.[211]

4.3 Enzymatic biotransformation of D-glucose and glycerol to chemicals *via* dehydration of D-glycerate to pyruvate: challenges and future perspectives

4.3.1 From D-glucose or glycerol to pyruvate: "Short glycolysis"

D-glucose is a relatively cheap raw material and can be obtained sustainably from biomass. In a previous work, a minimized cascade reaction had been developed to transform D-glucose to chemicals *via* dehydration of D-glycerate to pyruvate. The first two products synthesized were ethanol and isobutanol.[151] The first pathway to ethanol has been modified to produce other chemicals, such as L-lactic acid and L-alanine.[196,216] For both pathway it was identified that the

dehydration of D-glycerate to pyruvate as the major bottleneck.[151,196,216] In this work, *Pu*DHT was identified as a better biocatalyst than *Ss*DHAD, an enzyme that has been used in the last decade to catalyze this reaction. *Pu*DHT activity toward D-gluconate is also remarkably higher than *Ss*DHAD. The second enzyme in the cascade that appeared to be a bottleneck is an aldehyde dehydrogenase from *Thermoplasma acidophilum* (ALDH). Recent work on this enzyme could increase its activity and specificity toward NAD^+ (M42).[197] When both enzymes—M42 and *Pu*DHT—were compared, *Pu*DHT still appeared to be a bottleneck, in particular when low substrate concentration was used (no D-glyceraldehyde accumulation in the scenario 1 and 2 in Table 4.2). However, when higher substrate load was used, low activity of M42 relative to other enzymes in a cascade started to be a bottleneck (scenario 3 vs. 4 in Table 4.2). M42 is responsible for oxidation of D-glyceraldehyde to D-glycerate; thus, in an event M42 serves as a bottleneck, accumulation of D-glyceraldehyde will start to occur. D-glyceraldehyde like any other aldehydes can react with amino acids of proteins, thus destabilizing the enzymes in the cascade.[217–219] This limitation can be partially alleviated by adjusting the rate of the cascade. One way would be to use less amount of D-glucose dehydrogenase (GDH). The only drawback in this approach perhaps is the stability of *Pu*DHT at 50°C.

In subchapter 3.3, it was shown that TTN of *Pu*DHT toward D-gluconate and D-glycerate reaches 10^6 and 10^3, respectively. However, these values were obtained at substrate saturations. In the cascade, the amount of the substrates, i.e. D-gluconate and D-glycerate are limited to the amount of cofactor (NAD^+) used. D-glycerate can be accumulated up to the total amount of the cofactor used, i.e. 5 mM but this concentration is far less from saturation. Thus, in actual cascade reactions stability of *Pu*DHT at 50 °C would be significantly less (higher ethanol conversion at 37 °C than at 50 °C, scenario 1 vs. 7 and 2 vs. 8 in Table 4.2). Hence, future work should be focused on improving the stability of *Pu*DHT. Because several new dehydratases that belong to the same class as *Pu*DHT (SADp2) have been characterized, DNA shuffling could be performed. PCR-based DNA shuffling (StEP) has been used also in this study to improve the stability of a PDC (subchapter 3.2).[203] A screening system to select variants with improved thermostability as well as activity toward D-glycerate simultaneously needs to be carefully designed. Microfluidic systems can be utilized for such purposes.[220,221] Another alternative is to run the cascade at lower temperature. Nevertheless, activity of *Pu*DHT (k_{cat}) of 1.3 s^{-1} will still need further improvement to increase further the feasibility of the cell-free approach.[151]

The minimized cascade described can also be used to produce pyruvate, a versatile building block with wide applications in pharmaceutical, chemical, and food industries.[151,222] Because in the production of pyruvate there is no equivalent reduction step as in ethanol, L-alanine, and L-lactic acid, NAD^+ reduced (NADH) will need to be regenerated. NADH regeneration can be performed enzymatically by NADH oxidase, chemically by iron (III) porphyrin, or

electrochemically.[223–225] However, the fact that no reduction step is required, it opens up a possibility to develop cofactor free approaches (Figure 4.4). Avoiding the use of NAD^+ will significantly reduce the production cost.[226] Oxygen can be used to replace NAD^+ to oxidize D-glucose and D-glyceraldehyde in a cell-free pyruvate production. Two different catalysts can be used to perform the oxidation.

Table 4.2. Effect of *Pu*DHT and M42 concentrations as well as temperature to the conversion of D-glucose to ethanol and remaining intermediates after 16 h biotransformation.

No	Temp (°C)	D-Glucose (mM)	*Pu*DHT (U/ml)	M42 (U/ml)	Conversion (%)			
					D-Glucose	D-GLA	D-GlyA	Ethanol
1	50	25	0.47	0.67	4.74	0.00	30.99	61.20
2	50	25	2.34	0.67	0.00	0.00	13.20	82.17
3	50	100	0.47	0.67	53.11	8.41	2.29	26.17
4	50	100	2.34	0.67	25.94	15.60	3.42	37.60
5	50	25	2.34	1.33	0.00	0.00	2.81	86.36
6	50	25	2.34	2.00	0.00	0.00	3.09	90.94
7	37	25	0.47	0.67	0.00	0.00	24.72	76.83
8	37	25	2.34	0.67	0.00	0.00	0.00	97.42

Gold catalyst has been used at our group for the oxidation of sugar acids and has also been shown to oxidize D-glyceraldehyde. The main challenge in this approach, however, is the compatibility between gold catalyst and enzymes. From the previous work, it is reported that gold catalyst is poisoned by the [Fe-S]-cluster leaked from *Ss*DHAD. H_2O_2 produced during oxidation will also oxidize [Fe-S]-cluster, thus in turn deactivating the dehydratase.[184] *Pu*DHT that seems to coordinate its cluster more stably than *Ss*DHAD was tried in one pot with the gold catalyst to produce 2-keto-3-deoxy-D-gluconate (KDG) from D-glucose. Similar catalyst poisoning, unfortunately, was also observed. Clever design to separate oxidation and dehydration steps in the previous study was able to tackle this incompatibility. 2-keto-3-deoxy-sugar acids (KDS) can be produced from their corresponding sugars using this design *via* semi-continuous approach.[184]

Adaptation of this design toward production of pyruvate, however, would be very challenging, if not practically impossible. The main reason is due to the nature of retro aldol cleavage of 2-keto-3-deoxy-gluconate (KDG) to D-glyceraldehyde and pyruvate. This reaction is an equilibrium reaction, thus complete formation of the latter products is not possible.[227] It means semi-continuous approach in the previous design would not be able to be used. Hence, the reaction solution needs to be recirculated between a chamber containing gold catalyst and a column containing an aldolase. Site production of H_2O_2 during D-glyceraldehyde oxidation catalyzed by gold catalyst will directly decarboxylate pyruvate produced in the reaction solution if there is no catalase present. Addition of a catalase in the oxidation chamber will impair the gold catalyst activity. That was the reason that in the previous study a dedicated column containing

catalase was used prior to pumping sugar acids to a column containing *Ss*DHAD—to remove H_2O_2 produced.[184] Although this approach may offer advantage for not using a cofactor, complexity of the reaction design would certainly contribute to the production cost.

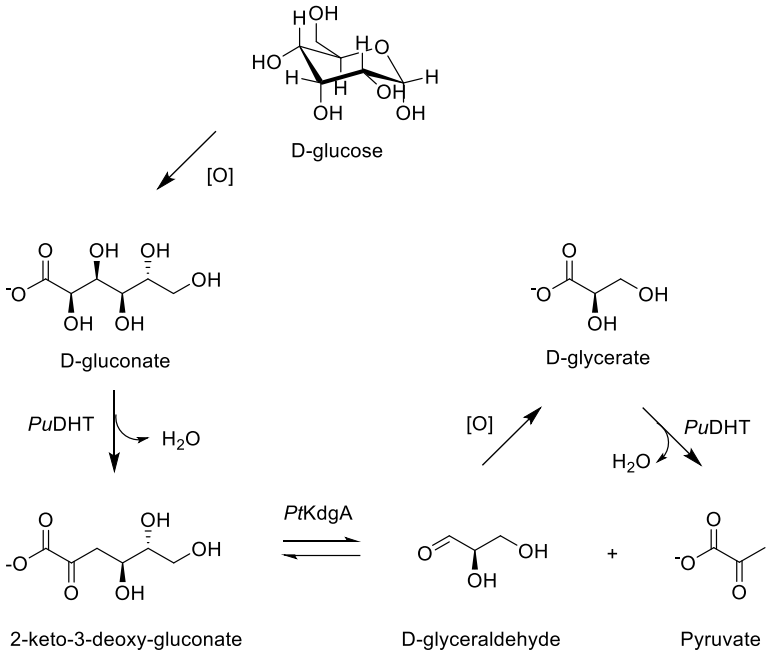


Figure 4.4. Cell-free and cofactor-free production of pyruvate from D-glucose

Another approach that appears to be simpler than chemoenzymatic approach is by performing oxidation of D-glucose and D-glyceraldehyde using enzymes. Oxidation of D-glucose by a glucose oxidase has been known for decades. The remaining challenge is oxidation of D-glyceraldehyde. So far, only an alditol oxidase from *Streptomyces coelicolor* is described that is able to catalyze such reaction.[18] A variant (V125M/A244T/V133M/ G399R) of this enzyme was used in this study to perform biotransformation of glycerol to pyruvate (subchapter 3.3). Oxidation of glycerol using this enzyme variant yielded D-glycerate. However, when D-glyceraldehyde was used no activity could be detected based on a coupled assay using a horseradish peroxidase and DA-64 dye (N-(carboxymethyl-aminocarbonyl)-4,4'-bis(dimethylamino)-diphenylamine). This could be caused by the fact that oxidation of D-glyceraldehyde to D-glycerate was much slower in comparison to double oxidation of glycerol to D-glycerate.[228] Perhaps, the variant used is also less reactive than its WT in respect to oxidation of D-glyceraldehyde. Another challenge in this approach is how to design a reaction chamber to allow optimum oxygen transfer rate to the reaction solution without deactivating the alditol oxidase. Bubbling oxygen directly to a reaction solution resulted in rapid discoloration of the alditol oxidase used. This phenomenon is also known for another FAD-dependent oxidase.[229] Therefore, future work on the discovery and engineering of an alditol oxidase to improve activity toward D-glyceraldehyde together with an attempt to increase activity of *Pu*DHT toward D-

glycerate would be necessary to challenge and possibly replace an existing pyruvate production *via* fermentation.[222] A more active and stable variant of the alditol oxidase will also benefit cell-free conversion of glycerol to pyruvate (see subchapter 3.3).[230,231]

4.3.2 From pyruvate to isobutanol: Toward cell-free isobutanol biosynthesis

In the second pathway of the minimized cascade reaction, D-glucose was transformed to isobutanol. Four additional enzymes were required. Among these enzymes, dehydration of DHIV to α-ketoisovalerate (KIV) was identified as the main bottleneck.[151] *Ss*DHAD showed an activity of 0.4 s^{-1} toward this reaction. *Pu*DHT that demonstrated 100-fold higher activity than *Ss*DHAD toward D-gluconate and D-glycerate showed 10-fold lower activity than *Ss*DHAD. This was rather surprising indicating that sugar acid dehydratases would not accept DHIV as their substrate. Activity of *Cc*XylDHT and *Rl*AraDHT toward DHIV needs to be determined to support this hypothesis. Next to *Pu*DHT, an SADp from *Herbaspirillum seropedicae* (*Hs*DHT) and *Ft*DHAD are the next promising enzymes identified in the conversion of D-glycerate to pyruvate. *Hs*DHT belongs to the same enzyme class as *Pu*DHT; thus, its activity toward DHIV is equally low. *Ft*DHAD, on the other hand, showed an activity of 9 s^{-1} toward DHIV (Table 4.1). *Ec*DHAD has higher activity but its instability in the presence of oxygen will hinder its utilization in a cell-free approach.[211] *Ft*DHAD showed almost no activity toward D-gluconate, hence cannot be used solely to replace *Ss*DHAD. Instead, two dehydratases *Pu*DHT and *Ft*DHAD would need to be used in combination. Considering significantly higher activity of *Pu*DHT toward D-gluconate and D-glycerate in comparison to *Ss*DHAD and *Ft*DHAD toward D-glycerate and DHIV as compared to *Ss*DHAD, combination of these two enzymes is expected to improve the performance of the cell-free isobutanol production. In addition to isobutanol, L-valine can also be produced from this pathway by replacing two last enzymes: a KDC and an ADH in the isobutanol production by an amino acid dehydrogenase, e.g. L-valine dehydrogenase (EC 1.4.1.23). L-valine is an essential amino acid, thereby important nutrient for livestock and human.[232–234]

Although combination of *Pu*DHT and *Ft*DHAD will have improved the cell-free isobutanol production in comparison to *Ss*DHAD, it is always of scientific and practical challenge to engineer an enzyme to reach its optimum catalytic fitness, in this case a dehydratase that can take all three substrates with similar magnitude of activity. This will not only reduce production cost further by only producing one enzyme instead of two, but will also ease cascade modelling and optimization. Therefore, preliminary work was performed in an attempt to explain the difference in substrate preference between *Pu*DHT and *Ft*DHAD. Substrate docking studies performed by YASARA allowed all three substrates to enter the catalytic site of both enzymes (Figure 4.5). From the docking studies, it appeared that D-gluconate and DHIV are coordinated correctly in the catalytic

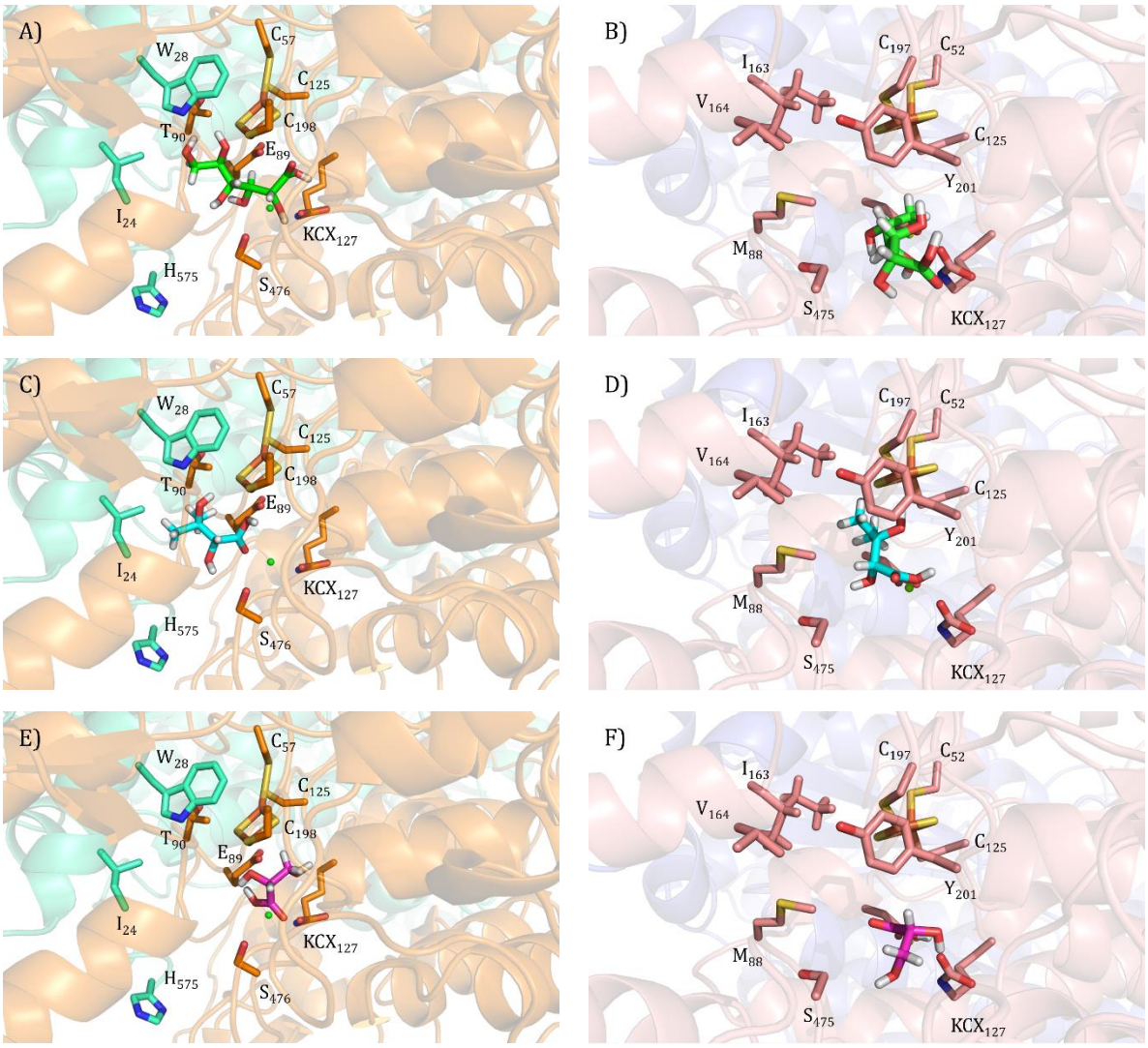


Figure 4.5. Docking studies of D-gluconate (A, B), DHIV (C,D), and D-glycerate (E.F) to *Pu*DHT (left) and *Ft*DHAD (right). D-gluconate, DHIV, D-glycerate are showed in stick conformation with green, cyan, and purple colors, respectively. Different color of each monomer for both enzymes are used. Catalytic site of *Pu*DHT appears to be orchestrated between its monomers while for *Ft*DHAD, its catalytic active site appears to be arranged only by one monomer. Swiss-MODEL was used to homology model *Pu*DHT and *Ft*DHAD. YASARA was used to perform substrates´ docking.

active site of *Pu*DHT. Hence, it is perhaps not straightforward to rationalize the very low activity of *Pu*DHT toward DHIV (Table 4.1). One possible explanation is that the hydrophobic group of DHIV interacts better with hydrophobic amino acids in the catalytic active site of *Pu*DHT, e.g. 24I and 28W. Another possible explanation is the electronic destabilizing effect of DHIV caused by the presence of two methyl groups at C3 of DHIV. Methyl group is viewed to have tendency to donate electron, while hydroxyl group tends to withdraw electron. Having an electron withdrawing group may make C3 of a substrate more susceptible, thus easier to giving up its hydroxyl group to be attacked by the [Fe-S] cluster. *Pu*DHT showed an activity of 10 s^{-1} toward L-theronate (data not shown), a >300-fold higher than toward DHIV. DHIV is different to L-theronate by having an additional methyl group at C3 and no hydroxyl group at C4. The docking studies indicate that the

active site of *Pu*DHT is large enough to accommodate an additional methyl group of DHIV (Figure 4.5 C).

Another possible explanation of the low ativity of *Pu*DHT to DHIV is in the relation to the putative role of its terminal histidine (H575). From the homology model of *Pu*DHT, it shows that its C-terminal histidine from the other monomer protrudes into the active site (Figure 4.5 B), the same as other sugar acid dehydratases.[206,207] This terminal histidine may have a role in the substrate stabilization by interacting with the terminal hydroxyl group of sugar acids. This hypothesis is supported by the kinetic data showing that activity of *Pu*DHT is higher with longer sugar acids (Table 4.1). DHIV that lacks any terminal hydroxyl group cannot be stabilized by the histidine, thus may occupy a wrong position for an efficient catalysis. This hypothesis needs to be validated further by testing if extending the C-terminus of *Pu*DHT by one or two amino acids prior to the histidine would increase activity of *Pu*DHT toward shorter sugar acids, e.g. L-threonate and D-glycerate. Validation of the putative role of H575 would partly explain the reason of *Pu*DHT´s low activity toward DHIV.

The docking studies in *Ft*DHAD showed clearer results. DHIV is arranged correctly in the catalytic active of *Ft*DHAD (Figure 4.5 B) to allow catalysis according to the proposed mechanism (Figure 1.8). Catalytic active site of *Ft*DHAD appeared to have a hydrophobic pocket arranged by 88M, 163I, and 164V. This could be the reason *Ft*DHAD accepts D-gluconate rather poorly. This hydrophobic binding pocket may push D-gluconate into occupying wrong arrangement for an efficient catalysis (Figure 4.5 B). The same as in *Pu*DHT, the docking studies on *Ft*DHAD suggested that D-glycerate is a small compound compared to D-gluconate and DHIV, thus cannot be stabilized by amino acids in the catalytic active sites to allow efficient catalysis (Figure 4.5 E and F). Another interesting result from the docking studies is that the distance between Fe atom of the [Fe-S] to oxygen atom of the catalytic serine in *Ft*DHAD (10.3 Å) is farther than in *Pu*DHT (7.1 Å) (data not shown). This could explain that in *Pu*DHT and other sugar acid dehydratases configuration of hydroxyl group at C2 and C3 have to be (R) and (S).[169] For a DHAD, only (R) configuration of hydroxyl group in C2 is important. Spinach DHAD (and most likely *Ft*DHAD) can dehydrate DHMV that has hydroxyl groups at C2 and C3 in (R,R) configuration as well as the epimer of DHMV that has (R,S) configuration.[168] Whether it would be easier to engineer *Pu*DHT to accept DHIV or *Ft*DHAD to accept D-gluconate cannot be concluded at this stage of study. In this study, these two enzymes have been characterized and based on their activity, their combination is predicted to perform better in a cell-free isobutanol or L-valine production than *Ss*DHAD.

4.4 Cell-free production of chemicals from hemicellulose sugars: combination of dehydratases, dehydratases, and decarboxylases

Hemicellulose constitutes almost 20% of dry weight of lignocellulosic biomass. Unlike cellulose that is constituted mainly of D-glucose, degradation of hemicellulose will result in hexoses and pentoses. Several attempts have been developed to convert pentose sugars, in particular D-xylose and L-arabinose to chemicals. These two pentose sugars are not readily fermentable by most industrially relevant microorganisms.[171,172,235] Two major approaches could be developed. The first approach is to introduce pentose pathway to the microorganisms and the second approach is to design cell-free enzymatic cascades. The latter approach is considered as an emerging technology. Several advantages of the latter approach have been discussed in literatures.[117,236,237] Among them are that cell-free production is easier to be optimized due to only a number of enzymes used in comparison to entire intricate metabolic pathways. Side products can also be minimized in a cell-free approach, as only necessary enzymes to convert a certain product are present. Faster production rate can be expected as there is no cell membrane acting as boundary that could possibly limit substrates to diffuse into the cells or product to diffuse out the cells. Engineering of enzymes to resist toxic substrates, intermediates, or products is practically less challenging than engineering the entire microorganisms to withstand such toxicity. However, the main challenge of an *in vitro* approach is to achieve cofactor, e.g. ATP and NAD(P)H balance in a designed pathway. Stoichiometric use of this cofactor would make the production cost too high, thus limiting the merit of such approach.

In this work, the Weimberg pathway was explored as a possible pathway to convert D-xylose and L-arabinose into chemicals. In Figure 4.6, several chemicals that can be derived from the Weimberg pathway and its subsequent modifications include α-ketoglutarate (α-KG), succinate, 1,4-butanediol (BDO), (S)- or (R)-1,2,4-butanetriol (BTO), and (S)- or (R)-3-hydroxy-γ-butyrolactone (3-HBL). BTO and 3-HBL are building blocks for pharmaceuticals in the synthesis of cholesterol-lowering drug, antibiotics, and HIV inhibitor.[152,238] In addition to pharmaceutical application, BTO can also be used as plasticizer and a precursor for a safer propellant (butantriol trinitrate).[239] Succinate and α-KG have several applications as additives in food, cosmetic, and pharmaceuticals.[240,241] Succinate and BDO are building block for every-day-use polymers, such as Spandex and polyurethane.[153] BDO is also an important solvent in industry.[153]

Several of these compounds are still produced or derived from fossil fuels. BTO is produced *via* reduction of esterified malic acid using sodium borohydride.[242] Malic acid is industrially produced from butene *via* maleic anhydride. Butene is obtained by catalytic cracking of petroleum and its oxidation to maleic anhydride is energy intensive—performed at >300 °C.[242] BDO is produced in industry from condensation of acetylene and two equivalent of formaldehyde *via* Reppe process, named after a German Chemist. This process is also energy intensive and similar

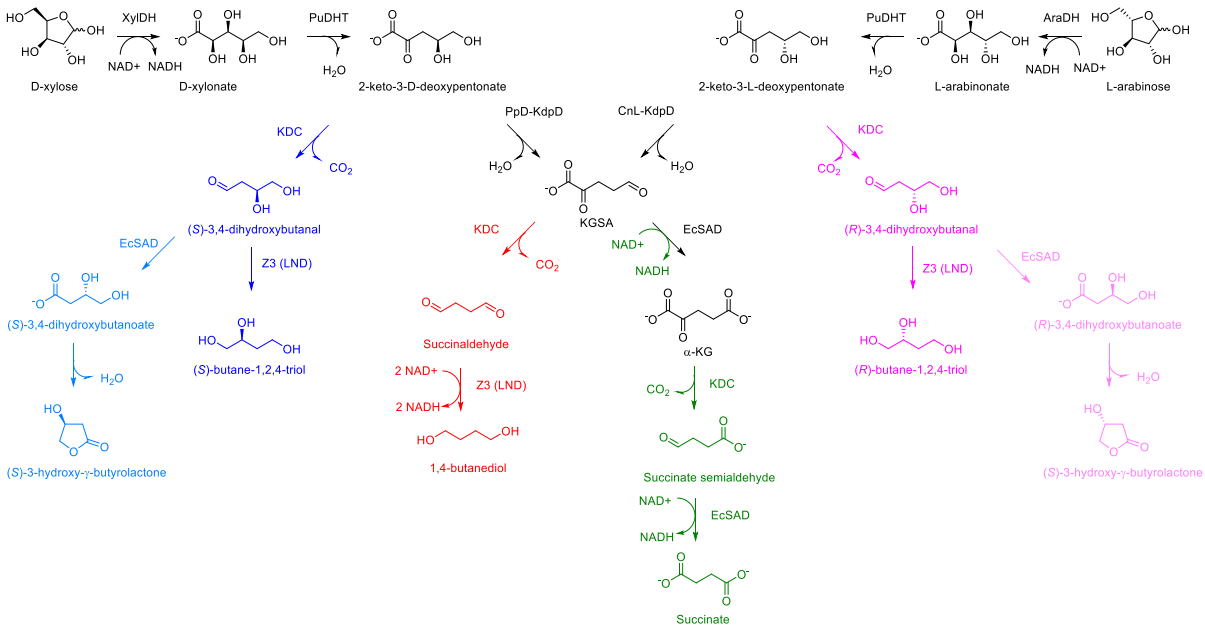


Figure 4.6. The Weimberg pathway and its modifications toward synthesis of different chemicals. XylDH is a D-xylose dehydrogenase, AraDH is a L-arabinose dehydrogenase, Z3 (LND) is a variant of a *E. coli* alcohol dehydrogenase that has been engineered at our chair toward higher selectivity to NAD, *Ec*SAD is a succinaldehyde dehydrogenase from *E. coli*.

to butane, acetylene is produced *via* hydrocarbon cracking.[243] Therefore, production of such chemicals from renewable materials such as D-xylose and L-arabinose obtained from hemicellulose is of great scientific and industrial interests toward bio-based economy (bioeconomy).

All lyases that have been engineered and identified in this study play significant roles in the synthetic Weimberg pathway described above (Figure 4.6). *Pu*DHT has been shown to be active toward D-xylonate and L-arabinonate. This enzymes also showed higher activity than previously described sugar acid dehydratases: *Rl*AraDHT and *Cc*XylDHT (Table 4.1).[169] When hydrolysates of hemicellulose containing both sugars is used, utilization of *Pu*DHT solely is enough to convert both sugar acids to desired chemicals. *Rl*AraDHT although showed similar degree of promiscuity to *Pu*DHT, its activities are significantly lower than those of *Pu*DHT. Thus, the promiscuity and high activity of *Pu*DHT would contribute to lower the production cost of this pathway. D-KdpD and L-KdpD will be needed when products other than BTO and 3-HBL are desired. It was important to identify which enzymes are better variants for cell-free applications. Thus, in this study several new D-KdpDs and L-KdpDs were cloned and characterized. In regard to activity and kinetic stability, *Pp*D-KdpD and *Cn*L-KdpD emerged as more superior variants than other D-KdpDs and L-KdpDs studied. Further study revealed that both enzymes are rather specific toward each isomer of 2-keto-3-deoxy-pentonate (KDP) intermediates. Consequently, when both pentose sugars are present and one product is desired, e.g. α-KG both enzymes are still needed to be produced. Future work could be directed toward improved promiscuity of one of the enzymes. However, if two product are desired, e.g. (S)-BTO and BDO or (R)-BTO and α-KG from mixed

pentose sugars (L-arabinose and D-xylose), one of these enzymes could be used to convert D,L-KDP intermediates enantioselectively toward desired product.

The third lyase in this study, KDC, in fact is the main reason to allow production of the majority products from the modified Weimberg pathways. Chemicals like 3-HBL, BTO, and BDO are not naturally produced from the Weimberg pathway.[180] Intermediates for production of such chemicals are also not natural substrates of KDC. However, due to promiscuity of KDC, decarboxylation of such intermediates are facilitated to some extent. WT of *Ll*KdcA already showed minor activity toward all intermediates in the modified pathways (Table 4.3). From partial characterization of WT KdcA, intermediates with more hydroxyl groups are proven to be more challenging to be decarboxylated. Several mutations in the catalytic active sites will be required to improve activity toward such substrates. Strategies like combinatorial active site saturation test (CAST) or epPCR can be applied.[46,95] In this study, *Ll*KdcA was engineered toward improved thermostability. A more stable enzyme variant will allow to resist substitutions introduced, in particular toward improved catalytic fitness.[244] Promiscuity of the most thermostable variant (7M.D) is predicted to be similar to that of the WT as it was shown for a number of α-keto acids (pyruvate, α-ketobutyrate, α-ketoisovalerate, and α-ketoisovalerate) (subchapter 3.1).[244] In addition to 7M.D, PDC-Var.2 that was evolved from *Ap*PDC in this study could also be used as a starting point. This variant is significantly more stable than 7M.D. A point substitution in WT *Zm*PDC (I476A) has already been shown to improve its activity significantly toward longer α-keto acids.[245]

Table 4.3. Summary of activity of *Ll*KdcA toward α-keto acid intermediates in the Weimberg pathway*

Substrate	V_{max} (U/mg)	K_M (mM)	k_{cat}/K_M (x 10^{-3} $mM^{-1}s^{-1}$)
L-KDP	1.2 ± 0.1	118.1 ± 20.5	10.2 ± 0.2
D-KDP	1.0 ± 0.1	36.4 ± 4.3	27.5 ± 0.2
KGSA	2.4 ± 0.2	8.2 ± 1.2	292.7 ± 0.2
α-KG	0.3 ± 0.0	132.8 ± 37.1	2.3 ± 0.3

*Activity was determined at 30°C in 10 mM KPi buffer, pH 6.5 using previously described pH shift assay.[246]

As it was mentioned earlier that a design of cofactor balance reaction is important in a cell-free approach. In these modified Weimberg pathways, only the production toward BTO from either L-arabinose or D-xylose is cofactor neutral. To produce BDO, one additional oxidation step is required to balance two reduction of succinaldehyde to BDO. This could be achieved by using formate dehydrogenase/FDH (EC 1.2.1.2) or phosphite dehydrogenase/PTDH (EC 1.20.1.1).[247,248] FDH will oxidize formate to CO_2. To close the loop, CO_2 can be reduced to formate electrochemically.[249] PTDH is an enzyme that has been used as regeneration system as this enzyme can oxidize phosphite to phosphate, a common buffer. For the production of 3-HBL, α-KG, and succinate, there is no reduction step; thus, the strategy required is less complicated than for BDO. To recycle NADH produced from the oxidation steps, NADH oxidase (NOX) or iron (III)

porphyrin can be used.[223,225] The challenge in the utilization of NOX is the same as for alditol oxidase described earlier. A smart reaction vessel design is required to allow maximum diffusion of oxygen into solution, while still maintaining the stability of the NOX.[229]

5 References

1. Jaenicke, L. Centenary of the award of a Nobel Prize to Eduard Buchner, the father of biochemistry in a test tube and thus of experimental molecular bioscience. *Angew. Chemie - Int. Ed.* **46**, 6776–6782 (2007).
2. Demirjian, D. C., Shah, P. C. & Morís-Varas, F. Screening for Novel Enzymes. in *Biocatalysis - From Discovery to Application* (eds. Fessner, W.-D. et al.) 1–29 (Springer Berlin Heidelberg, 1999).
3. Quastel, J. H. The development of biochemistry in the 20th century. *Mol. Cell. Biochem.* **69**, 17–26 (1985).
4. Reetz, M. T. Biocatalysis in organic chemistry and biotechnology: Past, present, and future. *J. Am. Chem. Soc.* **135**, 12480–12496 (2013).
5. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3240–3244 (1973).
6. Saiki, R. K. *et al.* Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* vol. 230 1350–1354 (1985).
7. May, O. Industrial Enzyme Applications – Overview and Historic Perspective. in *Industrial Enzyme Applications* (eds. Vogel, A. & Oliver, M.) 1–24 (Wiley-VCH, 2019).
8. Ichiro, C., Tetsuya, T. & Noriko, E. Resolution of Racemic Amino Acids. (1968)
9. Turner, N. J. & Truppo, M. D. Biocatalysis enters a new era. *Curr. Opin. Chem. Biol.* **17**, 212–214 (2013).
10. Vogel, A. Enzyme Development Technologies. in *Industrial Enzyme Applications* (eds. Vogel, A. & Oliver, M.) 25–45 (Wiley-VCH, 2019).
11. Kara, S. & Liese, A. Process Considerations for the Application of Enzymes. in *Industrial Enzyme Applications* (eds. Vogel, A. & Oliver, M.) 71–93 (Wiley-VCH, 2019).
12. Bommarius, A. S. & Riebel-Bommarius, B. R. *Biocatalysis: Fundamentals and Applications*. *Biocatalysis* (John Wiley & Sons, Ltd, 2005).
13. Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **8**, 476–480 (2017).
14. Thomas, S. M., DiCosimo, R. & Nagarajan, V. Biocatalysis: Applications and potentials for the chemical industry. *Trends Biotechnol.* **20**, 238–242 (2002).
15. Rozzell, J. D. Biocatalysis at commercial scale: Myths and realities. *Chim. Oggi* **17**, 42–47 (1999).
16. Dydio, P., Key, H. M., Hayashi, H., Clark, D. S. & Hartwig, J. F. Chemoselective, enzymatic C-H bond amination catalyzed by a cytochrome P450 containing an Ir(Me)-PIX cofactor. *Journal of the American Chemical Society* vol. 139 1750–1753 (2017).
17. Palmer, D. R. J. & Gerlt, J. A. Evolution of enzymatic activities: Multiple pathways for generating and partitioning a common enolic intermediate by glucarate dehydratase from Pseudomonas putida. *J. Am. Chem. Soc.* **118**, 10323–10324 (1996).
18. Gerstenbruch, S. *et al.* Asymmetric synthesis of D -glyceric acid by an alditol oxidase and directed evolution for enhanced oxidative activity towards glycerol. *Appl. Microbiol. Biotechnol.* **96**, 1243–1252 (2012).
19. Woodley, J. M. Accelerating the implementation of biocatalysis in industry. *Appl. Microbiol. Biotechnol.* **103**, 4733–4739 (2019).
20. Chapman, J., Ismail, A. E. & Dinu, C. Z. Industrial applications of enzymes: Recent advances, techniques, and outlooks. *Catalysts* **8**, 20–29 (2018).
21. Britton, R. A. Direct Hydration of Ethylene to Ethanol. (1972).
22. Zhou, B., Li, Z. & Chen, C. Global potential of rare earth resources and rare earth demand from clean technologies. *Minerals* **7**, (2017).
23. Sheldon, R. A. & Brady, D. The limits to biocatalysis: Pushing the envelope. *Chem. Commun.* **54**, 6088–6104 (2018).
24. Stepankova, V. *et al.* Strategies for stabilization of enzymes in organic solvents. *ACS Catal.* **3**, 2823–2836 (2013).

25. Woodley, J. M. Protein engineering of enzymes for process applications. *Curr. Opin. Chem. Biol.* **17**, 310–316 (2013).
26. Jemli, S., Ayadi-Zouari, D., Hlima, H. Ben & Bejar, S. Biocatalysts: Application and engineering for industrial purposes. *Crit. Rev. Biotechnol.* **36**, 246–258 (2016).
27. Abdelraheem, E. M. M., Busch, H., Hanefeld, U. & Tonin, F. Biocatalysis explained: From pharmaceutical to bulk chemical production. *React. Chem. Eng.* **4**, 1878–1894 (2019).
28. Bommarius, A. S., Blum, J. K. & Abrahamson, M. J. Status of protein engineering for biocatalysts: How to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.* **15**, 194–200 (2011).
29. Luetz, S., Giver, L. & Lalonde, J. Engineered enzymes for chemical production. *Biotechnol. Bioeng.* **101**, 647–653 (2008).
30. Bommarius, A. S. & Paye, M. F. Stabilizing biocatalysts. *Chem. Soc. Rev.* **42**, 6534–6565 (2013).
31. Schoemaker, H. E., Mink, D. L. & WubboLts, M. G. Dispelling the myths - Biocatalysis in industrial synthesis. *Science (80-.).* **299**, 1694–1697 (2003).
32. Liu, L. *et al.* How to achieve high-level expression of microbial enzymes: Strategies and perspectives. *Bioengineered* **4**, 212–223 (2013).
33. Slouka, C., Kopp, J., Spadiut, O. & Herwig, C. Perspectives of inclusion bodies for bio-based products: curse or blessing? *Appl. Microbiol. Biotechnol.* **103**, 1143–1153 (2019).
34. Garcia-Galan, C., Berenguer-Murcia, Á., Fernandez-Lafuente, R. & Rodrigues, R. C. Potential of different enzyme immobilization strategies to improve enzyme performance. *Adv. Synth. Catal.* **353**, 2885–2904 (2011).
35. Eş, I., Vieira, J. D. G. & Amaral, A. C. Principles, techniques, and applications of biocatalyst immobilization for industrial application. *Appl. Microbiol. Biotechnol.* **99**, 2065–2082 (2015).
36. Wang, M., Si, T. & Zhao, H. Biocatalyst development by directed evolution. *Bioresour. Technol.* **115**, 117–125 (2012).
37. Kaul, P. & Asano, Y. Strategies for discovery and improvement of enzyme function: State of the art and opportunities. *Microb. Biotechnol.* **5**, 18–33 (2012).
38. Lorenz, P. & Eck, J. Screening for novel industrial biocatalysts. *Eng. Life Sci.* **4**, 501–504 (2004).
39. Ferrer, M. *et al.* Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. *Microb. Biotechnol.* **9**, 22–34 (2016).
40. Elleuche, S., Schröder, C., Sahm, K. & Antranikian, G. Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. *Curr. Opin. Biotechnol.* **29**, 116–123 (2014).
41. Wang, S. *et al.* Enzyme stability and activity in non-aqueous reaction systems: A mini review. *Catalysts* **6**, 1–16 (2016).
42. Lima-Ramos, J., Neto, W. & Woodley, J. M. Engineering of biocatalysts and biocatalytic processes. *Top. Catal.* **57**, 301–320 (2014).
43. Yuryev, R., Strompen, S. & Liese, A. Coupled chemo(enzymatic) reactions in continuous flow. *Beilstein J. Org. Chem.* **7**, 1449–1467 (2011).
44. Bornscheuer, U. T. & Pohl, M. Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **5**, 137–143 (2001).
45. Dalby, P. A. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **21**, 473–480 (2011).
46. Schmidt-Dannert, C. & Arnold, F. H. Directed evolution of industrial enzymes. *Tibtech* **17**, 135–136 (1999).
47. Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5618–5622 (1993).
48. Chen, K. & Arnold, F. H. Enzyme Engineering for Nonaqueous Solvents: Random Mutagenesis to Enhance Activity of Subtilisin E in Polar Media. *Nat. Biotechnol.* **9**, 1073–1077 (1991).

49. McCullum, E. O., Williams, B. A. R., Zhang, J. & Chaput, J. C. Random Mutagenesis by Error-Prone PCR. in *In Vitro Mutagenesis Protocols: Third Edition* (ed. Braman, J.) 103–109 (Humana Press, 2010).
50. Wong, T. S., Tee, K. L., Hauer, B. & Schwaneberg, U. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.* **32**, 26e–26 (2004).
51. Ruff, A. J., Kardashliev, T., Dennig, A. & Schwaneberg, U. The Sequence Saturation Mutagenesis (SeSaM) Method. in *Directed Evolution Library Creation: Methods and Protocols* (eds. Gillam, E. M. J., Copp, J. N. & Ackerley, D.) 45–68 (Springer New York, 2014).
52. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).
53. Zhao, H. & Zha, W. In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nat. Protoc.* **1**, 1865–1871 (2006).
54. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).
55. Sieber, V., Martinez, C. A. & Arnold, F. H. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19**, 456–460 (2001).
56. Udit, A. K., Silberg, J. J. & Sieber, V. Sequence Homology-Independent Protein Recombination (SHIPREC). in *Directed Evolution Library Creation: Methods and Protocols* (eds. Arnold, F. H. & Georgiou, G.) 153–163 (Humana Press, 2003).
57. Ostermeier, M., Shim, J. H. & Benkovic, S. J. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17**, 1205–1209 (1999).
58. Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y. & Bocola, M. Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity : Rigorous Comparison with Traditional Methods. *J. Am. Chem. Soc.* **132**, 9144–9152 (2010).
59. Lutz, S. Beyond directed evolution-semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21**, 734–743 (2010).
60. Steiner, K. & Schwab, H. Recent advances in rational approaches for enzyme engineering. *Comput. Struct. Biotechnol. J.* **2**, e201209010 (2012).
61. Bo, C. B. & Katoh, K. Protein Multiple Sequence Alignment. in *Methods in molecular biology* (eds. Thompson, J. D., Schaeffer-Reiss, C. & Ueffing, M.) vol. 484 379–413 (Humana Press, 2008).
62. Notredame, C. Recent progress in multiple sequence alignment: A survey. *Pharmacogenomics* **3**, 131–144 (2002).
63. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng. Des. Sel.* **29**, 245–251 (2016).
64. Lehmann, M. *et al.* From DNA sequence to improved functionality: Using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng.* **13**, 49–57 (2000).
65. Vázquez-Figueroa, E., Chapparro-Riggers, J. & Bommarius, A. S. Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. *ChemBioChem* **8**, 2295–2301 (2007).
66. Deane, C. & de Oliveira, S. Co-evolution techniques are reshaping the way we do structural bioinformatics. *F1000Research* **6**, 1–6 (2017).
67. Franceus, J., Verhaeghe, T. & Desmet, T. Correlated positions in protein evolution and engineering. *J. Ind. Microbiol. Biotechnol.* **44**, 687–695 (2017).
68. Kuipers, R. K. *et al.* 3DM: Systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins Struct. Funct. Bioinforma.* **78**, 2101–2113 (2010).
69. Jermann, T. M., Opitz, J. G., Stackhouse, J. & Benner, S. A. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374**, 57–59 (1995).
70. Thornton, J. W. Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).
71. Gumulya, Y. & Gillam, E. M. J. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: The 'retro' approach to protein engineering. *Biochem.*

J. **474**, 1–19 (2017).
72. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, 526–531 (2004).
73. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
74. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2016).
75. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA - A self-parameterizing force field. *Proteins Struct. Funct. Genet.* **47**, 393–402 (2002).
76. Vazquez-Figueroa, E., Yeh, V., Broering, J. M., Chaparro-Riggers, J. F. & Bommarius, A. S. Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media. *Protein Eng. Des. Sel.* **21**, 673–680 (2008).
77. Karplus, P. A. & Schulz, G. E. Prediction of chain flexibility in proteins - A tool for the selection of peptide antigens. *Naturwissenschaften* **72**, 212–213 (1985).
78. Reetz, M. T., Carballeira, J. D. & Vogel, A. Iterative saturation mutagenesis on the basis of b factors as a strategy for increasing protein thermostability. *Angew. Chemie - Int. Ed.* **45**, 7745–7751 (2006).
79. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins Struct. Funct. Bioinforma.* **79**, 830–838 (2011).
80. Schymkowitz, J. *et al.* The FoldX web server: An online force field. *Nucleic Acids Res.* **33**, 382–388 (2005).
81. Wijma, H. J. *et al.* Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27**, 49–58 (2014).
82. Musil, M. *et al.* FireProt: web server for automated design of thermostable proteins. *Nucleic Acids Res.* **45**, W393–W399 (2017).
83. Goldenzweig, A. *et al.* Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **63**, 337–346 (2016).
84. Strickler, S. S. *et al.* Protein stability and surface electrostatics: A charged relationship. *Biochemistry* **45**, 2761–2766 (2006).
85. Martin, A., Sieber, V. & Schmid, F. X. In-vitro selection of highly stabilized protein variants with optimized surface. *J. Mol. Biol.* **309**, 717–726 (2001).
86. Eijsink, V. G. H. *et al.* Rational engineering of enzyme stability. *J. Biotechnol.* **113**, 105–120 (2004).
87. Nestl, B. M. & Hauer, B. Engineering of flexible loops in enzymes. *ACS Catal.* **4**, 3201–3211 (2014).
88. Bosshart, A., Panke, S. & Bechtold, M. Systematic optimization of interface interactions increases the thermostability of a multimeric enzyme. *Angew. Chemie - Int. Ed.* **52**, 9673–9676 (2013).
89. Russell, R. J. M. & Taylor, G. Engineering thermostability: lessons from thermophilic proteins. *Curr. Opin. Biotechnol.* **6**, 370–374 (2001).
90. Pagadala, N. S., Syed, K. & Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* **9**, 91–102 (2017).
91. Salmaso, V. & Moro, S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Front. Pharmacol.* **9**, 1–16 (2018).
92. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).
93. Funke, S. A. *et al.* Directed evolution of an enantioselective Bacillus subtilis lipase. *Biocatal. Biotransformation* **21**, 67–73 (2003).
94. Wang, J. bo, Li, G. & Reetz, M. T. Enzymatic site-selectivity enabled by structure-guided directed evolution. *Chem. Commun.* **53**, 3916–3928 (2017).
95. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. Expanding the range of

substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chemie - Int. Ed.* **44**, 4192–4196 (2005).
96. Reetz, M. T., Wang, L. W. & Bocola, M. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angew. Chemie - Int. Ed.* **45**, 1236–1241 (2006).
97. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
98. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2**, 924–932 (2007).
99. Lacks, S. & Greenberg, B. A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA. *J. Biol. Chem.* **250**, 4060–4066 (1975).
100. Zheng, L., Baumann, U. & Reymond, J.-L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32**, e115–e115 (2004).
101. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, (2008).
102. Wang, W. & Malcolm, B. A. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange(TM) Site-Directed Mutagenesis. *BioTechniques* vol. 26 680–682 (1999).
103. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
104. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, (2008).
105. Acevedo-Rocha, C. G., Reetz, M. T. & Nov, Y. Economical analysis of saturation mutagenesis experiments. *Sci. Rep.* **5**, 1–12 (2015).
106. Hoebenreich, S., Zilly, F. E., Acevedo-Rocha, C. G., Zilly, M. & Reetz, M. T. Speeding up directed evolution: Combining the advantages of solid-phase combinatorial gene synthesis with statistically guided reduction of screening effort. *ACS Synth. Biol.* **4**, 317–331 (2015).
107. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
108. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De novo enzyme design using Rosetta3. *PLoS One* **6**, 1–12 (2011).
109. Khoury, G. A., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.* **32**, 99–109 (2014).
110. Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817–1819 (2010).
111. Khersonsky, O. *et al.* Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407**, 391–412 (2011).
112. Khersonsky, O. *et al.* Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the ke07 series. *J. Mol. Biol.* **396**, 1025–1042 (2010).
113. Olson, D. G., McBride, J. E., Joe Shaw, A. & Lynd, L. R. Recent progress in consolidated bioprocessing. *Curr. Opin. Biotechnol.* **23**, 396–405 (2012).
114. Parisutham, V., Kim, T. H. & Lee, S. K. Feasibilities of consolidated bioprocessing microbes: From pretreatment to biofuel production. *Bioresour. Technol.* **161**, 431–440 (2014).
115. Bundesministerium für Bildung und Forschung. HotSysAPP - Heiße Angewandte System Biologie - Sulfolobus acidocaldarius als neues archaeelles, thermoacidophiles Chassis für die industrielle Biotechnologie. https://www.gesundheitsforschung-bmbf.de/de/hotsysapp-heisse-angewandte-system-biologie-sulfolobus-acidocaldarius-als-neues-archaeelles-5190.php (2016).
116. Quehenberger, J., Shen, L., Albers, S. V., Siebers, B. & Spadiut, O. Sulfolobus - A potential key organism in future biotechnology. *Front. Microbiol.* **8**, 1–13 (2017).
117. Claassens, N. J., Burgener, S., Vo, B., Erb, T. J. & Bar-even, A. A critical comparison of cellular and cell-free bioproduction systems. *Curr. Opin. Biotechnol.* **60**, 221–229 (2019).
118. Martin, J., Eisoldt, L. & Skerra, A. Fixation of gaseous CO2 by reversing a decarboxylase for

the biocatalytic synthesis of the essential amino acid l-methionine. *Nat. Catal.* **1**, 555–561 (2018).
119. S., O. Biological mechanisms of carboxylation and decarboxylation. *Physiol. Rev.* **31**, 56–106 (1951).
120. Li, T., Huo, L., Pulley, C. & Liu, A. Decarboxylation mechanisms in biological system. *Bioorg. Chem.* **43**, 2–14 (2012).
121. Kourist, R., Guterl, J. K., Miyamoto, K. & Sieber, V. Enzymatic decarboxylation-An emerging reaction for chemicals production from renewable resources. *ChemCatChem* **6**, 689–701 (2014).
122. Gocke, D. *et al.* Comparative characterisation of thiamin diphosphate-dependent decarboxylases. *J. Mol. Catal. B Enzym.* **61**, 30–35 (2009).
123. Lindqvist, Y. & Schneider, G. Thiamin diphosphate dependent enzymes: transketolase, pyruvate oxidase and pyruvate decarboxylase. *Curr. Opin. Struct. Biol.* **3**, 896–901 (1993).
124. Muller, Y. A. *et al.* A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase. *Structure* **1**, 95–103 (1993).
125. Hawkins, C. F., Borges, A. & Perham, R. N. A common structural motif in thiamin pyrophosphate-binding enzymes. *FEBS Lett.* **255**, 77–82 (1989).
126. Berthold, C. L. *et al.* Structure of the branched-chain keto acid decarboxylase (KdcA) from Lactococcus lactis provides insights into the structural basis for the chemoselective and enantioselective carboligation reaction. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 1217–1224 (2007).
127. Killenberg-jabs, M., Wikner, C., Schneider, G. & Hubner, G. How Thiamine Diphosphate Is Activated in Enzymes. *Science (80-.).* **275**, 67–70 (1997).
128. Iding, H., Siegert, P., Mesch, K. & Pohl, M. Application of α-keto acid decarboxylases in biotransformations. *Biochim. Biophys. Acta* **1385**, 307^322 (1998).
129. Meyer, D. *et al.* Conversion of pyruvate decarboxylase into an enantioselective carboligase with biosynthetic potential. *J. Am. Chem. Soc.* **133**, 3609–3616 (2011).
130. Neuberg, C. & Hirsch, J. Über ein Kohlenstoffketten knüpfendes Ferment (Carboligase). *Biochem. Zeitschr.* **115**, 282 (1921).
131. Pohl, M. Protein design on pyruvate decarboxylase (PDC) by site-directed mutagenesis. Application to mechanistical investigations, and tailoring PDC for the use in organic synthesis. *Adv. Biochem. Eng. Biotechnol.* **58**, 15–43 (1997).
132. Gocke, D. *et al.* Branched-chain keto acid decarboxylase from Lactococcus lactis (KdcA), a valuable thiamine diphosphate-dependent enzyme for asymmetric C - C bond formation. *Adv. Synth. Catal.* **349**, 1425–1435 (2007).
133. Eram, M. S. & Ma, K. Decarboxylation of pyruvate to acetaldehyde for ethanol production by hyperthermophiles. *Biomolecules* **3**, 578–596 (2013).
134. Taylor, M. P. *et al.* Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends Biotechnol.* **27**, 398–405 (2009).
135. Barnard, D., Casanueva, A., Tuffin, M. & Cowan, D. Extremophiles in biofuel synthesis. *Environ. Technol.* **31**, 871–888 (2010).
136. Olson, D. G., Sparling, R. & Lynd, L. R. Ethanol production by engineered thermophiles. *Curr. Opin. Biotechnol.* **33**, 130–141 (2015).
137. Tian, L. *et al.* Enhanced ethanol formation by Clostridium thermocellum via pyruvate decarboxylase. *Microb. Cell Fact.* 1–10 (2017)
138. Thompson, A. H., Studholme, D. J., Green, E. M. & Leak, D. J. Heterologous expression of pyruvate decarboxylase in Geobacillus thermoglucosidasius. *Biotechnol. Lett.* **30**, 1359–1365 (2008).
139. Cheng, M., Yoshiyasu, H., Okano, K., Ohtake, H. & Honda, K. Redirection of the Reaction Specificity of a Thermophilic Acetolactate Synthase toward Acetaldehyde Formation. *PLoS One* **11**, 1–13 (2016).
140. Buddrus, L. *et al.* Crystal structure of an inferred ancestral bacterial pyruvate decarboxylase. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **74**, 179–186 (2018).

141. Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707 (2008).
142. Hobbs, J. K. *et al.* On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of Bacillus. *Mol. Biol. Evol.* **29**, 825–835 (2012).
143. Sammond, D. W. *et al.* An iterative computational design approach to increase the thermal endurance of a mesophilic enzyme. *Biotechnol. Biofuels* **11**, 1–13 (2018).
144. Smit, B. A. *et al.* Identification , Cloning , and Characterization of a Lactococcus lactis Branched-Chain a-Keto Acid Decarboxylase Involved in Flavor Formation. *Appl. Environ. Microbiol.* **71**, 303–311 (2005).
145. De La Plaza, M., Fernández De Palencia, P., Peláez, C. & Requena, T. Biochemical and molecular characterization of α-ketoisovalerate decarboxylase, an enzyme involved in the formation of aldehydes from amino acids by Lactococcus lactis. *FEMS Microbiol. Lett.* **238**, 367–374 (2004).
146. Wei, J., Timler, J. G., Knutson, C. M. & Barney, B. M. Branched-chain 2-keto acid decarboxylases derived from Psychrobacter. *FEMS Microbiol. Lett.* **346**, 105–112 (2013).
147. Wang, B., Bai, Y., Fan, T., Zheng, X. & Cai, Y. Characterisation of a thiamine diphosphate-dependent alpha-keto acid decarboxylase from Proteus mirabilis JN458. *Food Chem.* **232**, 19–24 (2017).
148. Lin, P. P. *et al.* Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius. *Metab. Eng.* **24**, 1–8 (2014).
149. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–90 (2008).
150. Zhang, K., Sawaya, M. R., Eisenberg, D. S. & Liao, J. C. Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20653–20658 (2008).
151. Guterl, J. K. *et al.* Cell-free metabolic engineering: Production of chemicals by minimized reaction cascades. *ChemSusChem* **5**, 2165–2172 (2012).
152. Wang, X. *et al.* D-1,2,4-Butanetriol production from renewable biomass with optimization of synthetic pathway in engineered Escherichia coli. *Bioresrouce Technol.* **250**, 406–412 (2017).
153. Yim, H. *et al.* Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7**, 445–452 (2011).
154. Tai, Y. S. *et al.* Engineering nonphosphorylative metabolism to generate lignocellulose-derived products. *Nat. Chem. Biol.* **12**, 247–253 (2016).
155. Zhong, W., Zhang, Y., Wu, W., Liu, D. & Chen, Z. Metabolic Engineering of a Homoserine-Derived Non-Natural Pathway for the de Novo Production of 1,3-Propanediol from Glucose. *ACS Synth. Biol.* **8**, 587–595 (2019).
156. Wang, J. *et al.* De Novo Biosynthesis of Glutarate via α-Keto Acid Carbon Chain Extension and Decarboxylation Pathway in Escherichia coli. *ACS Synth. Biol.* **6**, 1922–1930 (2017).
157. Turk, S. C. H. J. *et al.* Metabolic Engineering toward Sustainable Production of Nylon-6. *ACS Synth. Biol.* **5**, 65–73 (2016).
158. Soh, L. M. J. *et al.* Engineering a Thermostable Keto Acid Decarboxylase Using Directed Evolution and Computationally Directed Protein Design. *ACS Synth. Biol.* **6**, 610–618 (2017).
159. Maier, M., Radtke, C. P., Hubbuch, J., Niemeyer, C. M. & Rabe, K. S. On-Demand Production of Flow-Reactor Cartridges by 3D Printing of Thermostable Enzymes. *Angew. Chemie - Int. Ed.* **57**, 5539–5543 (2018).
160. Flint, D. H. & Allen, R. M. Iron-sulfur proteins with nonredox functions. *Chemical Reviews* vol. 96 2315–2334 (1996).
161. Imlay, J. A. Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol.* **59**, 1073–1082 (2006).
162. Pirrung, M. C., Ha, H. joon & Holmes, C. P. Purification And Inhibition of Spinach α,β-dihydroxyacid Dehydratase. *J. Org. Chem.* **54**, 1543–1548 (1989).
163. Yan, Y. *et al.* Resistance-gene-directed discovery of a natural-product herbicide with a

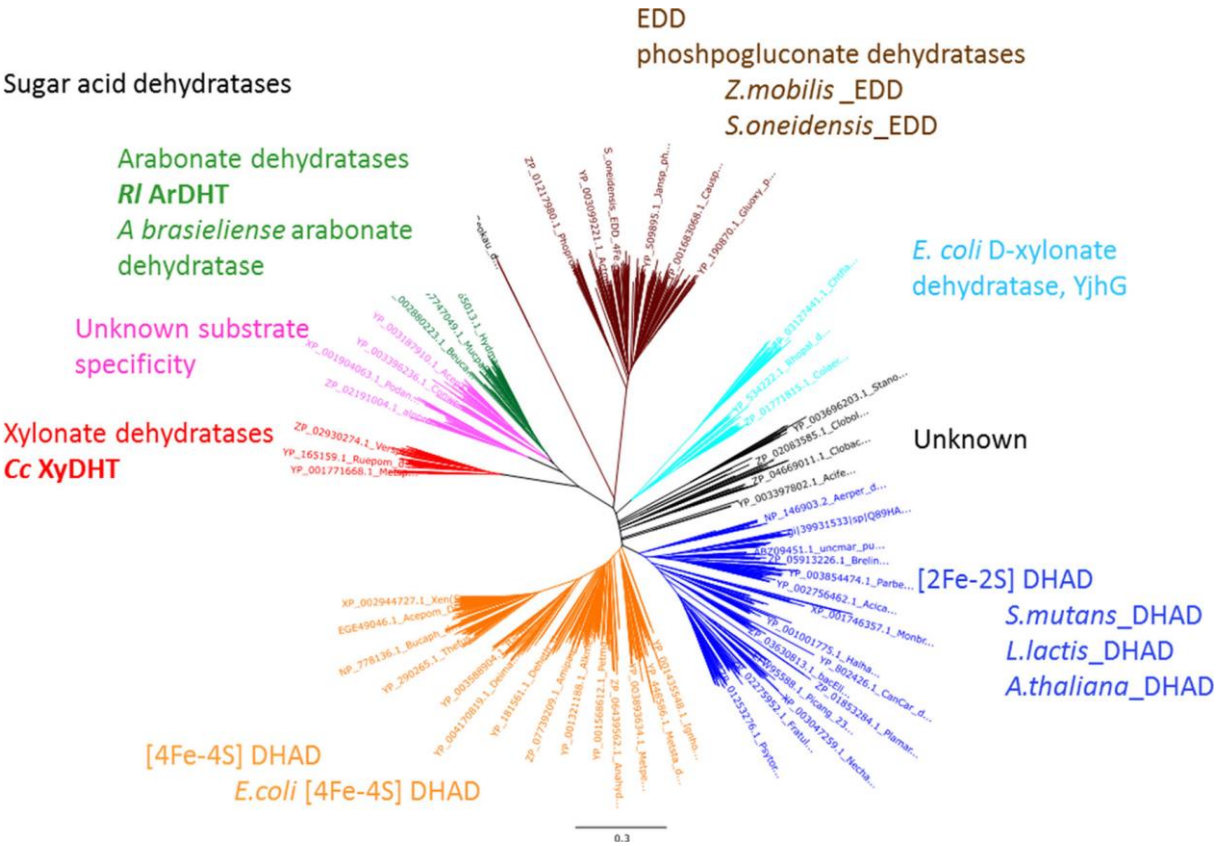
new mode of action. *Nature* **559**, 415–418 (2018).
164. Bashiri, G. *et al.* The active site of the Mycobacterium tuberculosis branched-chain amino acid biosynthesis enzyme dihydroxyacid dehydratase contains a 2Fe-2S cluster. *J. Biol. Chem.* **294**, 13158–13170 (2019).
165. Amorim Franco, T. M. & Blanchard, J. S. Bacterial Branched-Chain Amino Acid Biosynthesis: Structures, Mechanisms, and Drugability. *Biochemistry* **56**, 5849–5865 (2017).
166. Arfin, S. M. Evidence for an enol intermediate in the enzymatic conversion of alpha,beta-dihydroxyisovalerate to alpha-ketoisovalerate. *Journal of Biological Chemistry* vol. 244 2250–2251 (1969).
167. Flint, D. H. & Nudelman, A. Studies on the Active Site of Dihydroxy-acid Dehydratase. *Biorgan. Chem.* **21**, 367–385 (1993).
168. Pirrung, M. C., Holmes, C. P., Horowitz, D. M. & Nunn, D. S. Mechanism and Stereochemistry of α,β-Dihydroxyacid Dehydratase. *J. Am. Chem. Soc.* **113**, 1020–1025 (1991).
169. Andberg, M. *et al.* Characterization and mutagenesis of two novel iron – sulphur cluster pentonate dehydratases. *Appl. Microbiol. Biotechnol.* **100**, 7549–7563 (2016).
170. Scopes, R. K. & Griffiths-Smith, K. Use of differential dye-ligand chromatography with affinity elution for enzyme purification: 6-Phosphogluconate dehydratase from Zymomonas mobilis. *Anal. Biochem.* **136**, 530–534 (1984).
171. Jagtap, S. S. & Rao, C. V. Microbial conversion of xylose into useful bioproducts. *Appl. Microbiol. Biotechnol.* **102**, 9015–9036 (2018).
172. Valdehuesa, K. N. G. *et al.* Everyone loves an underdog: metabolic engineering of the xylose oxidative pathway in recombinant microorganisms. *Appl. Microbiol. Biotechnol.* **102**, 7703–7716 (2018).
173. Ahmed, H. *et al.* The semi-phosphorylative Entner – Doudoroff pathway in hyperthermophilic archaea : a re-evaluation. *Biochem. J* **540**, 529–540 (2005).
174. Kim, S. & Lee, S. B. Identification and characterization of Sulfolobus solfataricus D-gluconate dehydratase: A key enzyme in the non-phosphorylated Entner-Doudoroff pathway. *Biochem. J.* **387**, 271–280 (2005).
175. Kim, S. & Lee, S. B. Catalytic promiscuity in dihydroxy-acid dehydratase from the thermoacidophilic archaeon Sulfolobus solfataricus. *J. Biochem.* **139**, 591–596 (2006).
176. Carsten, J. M., Schmidt, A. & Sieber, V. Characterization of recombinantly expressed dihydroxy-acid dehydratase from Sulfobus solfataricus — A key enzyme for the conversion of carbohydrates into chemicals. *J. Biotechnol.* **211**, 31–41 (2015).
177. Begander, B., Huber, A., Döring, M., Sperl, J. & Sieber, V. Development of an improved peroxidase-based high-throughput screening for the optimization of d-glycerate dehydratase activity. *Int. J. Mol. Sci.* **21**, (2020).
178. Carsten, J. Charakterisierung, Optimierung und Anwendung einer thermostabilen Dehydratase. (Technical University of Munich, 2014).
179. Lamble, H. J., Milburn, C. C., Taylor, G. L., Hough, D. W. & Danson, M. J. Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in Sulfolobus solfataricus. *FEBS Lett.* **576**, 133–136 (2004).
180. Weimberg, R. Pentose oxidation by Pseudomonas fragi. *J. Biol. Chem.* **236**, 629–635 (1961).
181. Walker, J. C. F. Basic wood chemistry and cell wall ultrastructure. in *Primary Wood Processing: Principles and Practice* 23–67 (Springer Netherlands, 2006).
182. Zdarta, J., Pinelo, M., Jesionowski, T. & Meyer, A. S. Upgrading of Biomass Monosaccharides by Immobilized Glucose Dehydrogenase and Xylose Dehydrogenase. *ChemCatChem* **10**, 5164–5173 (2018).
183. Vuong, T. V., Foumani, M., MacCormick, B., Kwan, R. & Master, E. R. Direct comparison of gluco-oligosaccharide oxidase variants and glucose oxidase: Substrate range and H2O2 stability. *Sci. Rep.* **6**, 15–17 (2016).
184. Sperl, J. M., Carsten, J. M., Guterl, J. K., Lommes, P. & Sieber, V. Reaction Design for the

Compartmented Combination of Heterogeneous and Enzyme Catalysis. *ACS Catal.* **6**, 6329–6334 (2016).
185. Watanabe, S., Kodaki, T. & Makino, K. A novel α-ketoglutaric semialdehyde dehydrogenase: Evolutionary insight into an alternative pathway of bacterial L-arabinose metabolism. *J. Biol. Chem.* **281**, 28876–28888 (2006).
186. Esser, D. *et al.* Unraveling the function of paralogs of the aldehyde dehydrogenase super family from Sulfolobus solfataricus. *Extremophiles* **17**, 205–216 (2013).
187. Adams, E. & Rosso, Gl. α-Ketoglutaric semialdehyde dehydrogenase of Pseudomonas. *J. Biol. Chem.* **242**, 1802–1814 (1967).
188. Portsmouth, D., Stoolmiller, A. C. & Abeles, R. H. Studies on the Mechanism of Action of 2-Keto-3-deoxy-l-arabonate Dehydratase. *J. Biol. Chem.* **242**, 2751–2759 (1967).
189. Brouns, S. J. J. *et al.* Structural Insight into Substrate Binding and Catalysis of a Novel 2-Keto-3-deoxy-d-arabinonate Dehydratase Illustrates Common Mechanistic Features of the FAH Superfamily. *J. Mol. Biol.* **379**, 357–371 (2008).
190. Brouns, S. J. J. *et al.* Identification of the missing links in prokaryotic pentose oxidation pathways: Evidence for enzyme recruitment. *J. Biol. Chem.* **281**, 27378–27388 (2006).
191. Watanabe, S., Shimada, N., Tajima, K., Kodaki, T. & Makino, K. Identification and characterization of L-arabonate dehydratase, L-2-Keto-3-deoxyarabonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose metabolism: Novel evolutionary insight into sugar metabolism. *J. Biol. Chem.* **281**, 33521–33536 (2006).
192. Stoolmiller, A. C. & Abeles, R. H. Formation of α-Ketoglutaric Semialdehyde from L-2-Keto-3- deoxyarabonic Acid and Isolation of Dehydratase from Pseudomonas saccharophila. *J. Biol. Chem.* **241**, 5764–5771 (1966).
193. Bai, W. *et al.* Engineering nonphosphorylative metabolism to synthesize mesaconate from lignocellulosic sugars in Escherichia coli. *Metab. Eng.* **38**, 285–292 (2016).
194. Reiße, S. *et al.* In Vitro Bioconversion of Pyruvate to n-Butanol with Minimized cofactor Utilization. *Front. Bioeng. Biotechnol.* **4**, 1–10 (2016).
195. Krutsakorn, B. *et al.* In vitro production of n -butanol from glucose. *Metab. Eng.* **20**, 84–91 (2013).
196. Gmelch, T. J., Sperl, J. M. & Sieber, V. Optimization of a reduced enzymatic reaction cascade for the production of L-alanine. *Sci. Rep.* **9**, 11754 (2019).
197. Gmelch, T. J., Sperl, J. M. & Sieber, V. Molecular dynamics analysis of a rationally designed aldehyde dehydrogenase gives insights into improved activity for the non-native cofactor NAD + . *ACS Synth. Biol.* (2020)
198. Sieber, V., Plückthun, A. & Schmid, F. X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 291–294 (1998).
199. Koudelakova, T. *et al.* Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angew. Chemie - Int. Ed.* **52**, 1959–1963 (2013).
200. Kumar, S., Tsai, C. J. & Nussinov, R. Factors enhancing protein thermostability. *Protein Eng.* **13**, 179–191 (2000).
201. Eijsink, V. G. H., GÅseidnes, S., Borchert, T. V. & Van Den Burg, B. Directed evolution of enzyme stability. *Biomol. Eng.* **22**, 21–30 (2005).
202. Sutiono, S., Carsten, J. & Sieber, V. Structure-Guided Engineering of α-Keto Acid Decarboxylase for the Production of Higher Alcohols at Elevated Temperature. *ChemSusChem* **11**, 3335–3344 (2018).
203. Sutiono, S., Satzinger, K., Pick, A., Carsten, J. & Sieber, V. To beat the heat-engineering of the most thermostable pyruvate decarboxylase to date. *RSC Adv.* **9**, 29743–29746 (2019).
204. Daniel, R. M. The upper limits of enzyme thermal stability. *Enzyme Microb. Technol.* **19**, 74–79 (1996).
205. Arnold, F. H. Enzyme engineering reaches the boiling point. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2035–2036 (1998).
206. Rahman, M. M. *et al.* The Crystal Structure of a Bacterial L-Arabinonate Dehydratase

Contains a [2Fe-2S] Cluster. *ACS Chem. Biol.* **12**, 1919–1927 (2017).
207. Rahman, M. M., Andberg, M., Koivula, A., Rouvinen, J. & Hakulinen, N. The crystal structure of D-xylonate dehydratase reveals functional features of enzymes from the Ilv/ED dehydratase family. *Sci. Rep.* **8**, 23–25 (2018).
208. Sutiono, S., Teshima, M., Beer, B., Schenk, G. & Sieber, V. Enabling the direct enzymatic dehydration of D-glycerate to pyruvate as the key step in synthetic enzyme cascades used in the cell-free production of fine chemicals. *ACS Catal.* **10**, 3110–3118 (2020).
209. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
210. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
211. Flint, D. H., Emptage, M. H., Finnegan, M. G., Fu, W. & Johnson, M. K. The role and properties of the iron-sulfur cluster in Escherichia coli dihydroxy-acid dehydratase. *J. Biol. Chem.* **268**, 14732–14742 (1993).
212. Saini, V., Farhana, A. & Steyn, A. J. C. Mycobacterium tuberculosis WhiB3: A novel iron-sulfur cluster protein that regulates redox homeostasis and virulence. *Antioxidants Redox Signal.* **16**, 687–697 (2012).
213. Crack, J. C. *et al.* Influence of the Environment on the [4Fe-4S]2+ to [2Fe-2S]2+ Cluster Switch in the Transcriptional Regulator FNR. *J. AM. CHEM. SOC.* **130**, 1749–1758 (2008).
214. Duin, E. C. *et al.* [2Fe-2S] to [4Fe-4S] cluster conversion in Escherichia coli biotin synthase. *Biochemistry* **36**, 11811–11820 (1997).
215. Gao, H. *et al.* Function and maturation of the Fe-S center in dihydroxyacid dehydratase from Arabidopsis. *J. Biol. Chem.* **293**, 4422–4433 (2018).
216. Xie, L., Wei, X., Zhou, X., Meng, D. & Zhou, R. Conversion of D -glucose to L -lactate via pyruvate by an optimized cell-free enzymatic biosystem containing minimized reactions. *Synth. Syst. Biotechnol.* **3**, 204–210 (2018).
217. Robert, L. & Penaranda, F. S. Studies on aldehyde-protein interactions. I. Reaction of amino acids with lower aldehydes. *J. Polym. Sci.* **12**, 337–350 (1954).
218. Grimsrud, P. A., Xie, H., Griffin, T. J. & Bernlohr, D. A. Oxidative stress and covalent modification of protein with bioactive aldehydes. *J. Biol. Chem.* **283**, 21837–21841 (2008).
219. Uchida, K. Histidine and lysine as targets of oxidative modification. *Amino Acids* **25**, 249–257 (2003).
220. Weng, L. & Spoonamore, J. E. Droplet Microfluidics-Enabled High-Throughput Screening for Protein Engineering. *Micromachines* **10**, 734 (2019).
221. Chiu, F. W. Y. & Stavrakis, S. High-throughput droplet-based microfluidics for directed evolution of enzymes. *Electrophoresis* **40**, 2860–2872 (2019).
222. Maleki, N. & Eiteman, M. A. Recent progress in the microbial production of pyruvic acid. *Fermentation* **3**, 1–17 (2017).
223. Nowak, C. *et al.* A water-forming NADH oxidase from Lactobacillus pentosus suitable for the regeneration of synthetic biomimetic cofactors. *Front. Microbiol.* **6**, 1–9 (2015).
224. Rao, T. N., Yagi, I., Miwa, T., Tryk, D. A. & Fujishima, A. Electrochemical oxidation of NADH at highly boron-doped diamond electodes. *Anal. Chem.* **71**, 2506–2511 (1999).
225. Maid, H. *et al.* Iron catalysis for in situ regeneration of oxidized cofactors by activation and reduction of molecular oxygen: A synthetic metalloporphyrin as a biomimetic NAD(P)H oxidase. *Angewandte Chemie - International Edition* vol. 50 2397–2400 (2011).
226. Zachos, I., Nowak, C. & Sieber, V. Biomimetic cofactors and methods for their recycling. *Curr. Opin. Chem. Biol.* **49**, 59–66 (2019).
227. Reher, M., Fuhrer, T., Bott, M., Scho, P. & Ju, F. The Nonphosphorylative Entner-Doudoroff Pathway in the Thermoacidophilic Euryarchaeon Picrophilus torridus Involves a Novel 2-Keto-3-Deoxygluconate- Specific Aldolase □. *J. Bacteriol.* **192**, 964–974 (2010).
228. Hellemond, E. W. Van *et al.* Exploring the Biocatalytic Scope of Alditol Oxidase from Streptomyces coelicolor. *Adv. Synth. Catal.* **351**, 1523 – 1530 (2009).
229. Beer, B., Pick, A. & Sieber, V. In vitro metabolic engineering for the production of α-ketoglutarate. *Metabolic Engineering* vol. 40 5–13 (2017).

230. Gao, C. *et al.* An artificial enzymatic reaction cascade for a cell-free bio-system based on glycerol. *Green Chem.* **17**, 804–807 (2015).
231. Li, Z. *et al.* Production of value-added chemicals from glycerol using in vitro enzymatic cascades. *Commun. Chem.* **1**, 1–7 (2018).
232. Zhang, S., Zeng, X., Ren, M., Mao, X. & Qiao, S. Novel metabolic and physiological functions of branched chain amino acids: A review. *J. Anim. Sci. Biotechnol.* **8**, 4–15 (2017).
233. Tavernari, F. C. *et al.* Valine needs in starting and growing cobb (500) broilers. *Poult. Sci.* **92**, 151–157 (2013).
234. Corzo, A. *et al.* Nutritional feasibility of L-valine inclusion in commercial broiler diets. *J. Appl. Poult. Res.* **20**, 284–290 (2011).
235. McClintock, M. K., Wang, J. & Zhang, K. Application of nonphosphorylative metabolism as an alternative for utilization of lignocellulosic biomass. *Front. Microbiol.* **8**, 1–6 (2017).
236. Guterl, J. & Sieber, V. Biosynthesis " debugged ": Novel bioproduction strategies. *Eng. Life Sci.* **13**, 4–18 (2013).
237. Dudley, Q. M., Karim, A. S. & Jewett, M. C. Cell-free metabolic engineering : Biomanufacturing beyond the cell. *Biotechnol. J.* 69–82 (2015)
238. Martin, C. H. *et al.* A platform pathway for production of 3-hydroxyacids provides a biosynthetic route to 3-hydroxy-γ-butyrolactone. *Nature Communications* vol. 4 (2013).
239. Niu, W., Molefe, M. N. & Frost, J. W. Microbial Synthesis of the Energetic Material Precursor 1,2,4-Butanetriol. *J. Am. Chem. Soc.* **125**, 12998–12999 (2003).
240. Liu, S., He, L. & Yao, K. The antioxidative function of alpha-ketoglutarate and its applications. *Biomed Res. Int.* **2018**, (2018).
241. Saxena, R. K., Saran, S., Isar, J. & Kaushik, R. Production and Applications of Succinic Acid. *Curr. Dev. Biotechnol. Bioeng. Prod. Isol. Purif. Ind. Prod.* 601–630 (2016)
242. Cao, Y., Niu, W., Guo, J., Xian, M. & Liu, H. Biotechnological production of 1,2,4-butanetriol: An efficient process to synthesize energetic material precursor from renewable biomass. *Sci. Rep.* **5**, 1–9 (2015).
243. Platz, K. Submerged-flame cracking of crude oils for acetylene production. *6th World Petroleum Congress* 6 (1963).
244. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5869–5874 (2006).
245. Pohl, M., Siegert, P., Mesch, K., Bruhn, H. & Grotzinger, J. Active site mutants of pyruvate decarboxylase from Zymomonas mobilis. *Eur. J. Biochem.* **257**, 538–546 (1998).
246. Pick, A. *et al.* Identification and characterization of two new 5-keto-4-deoxy-D-Glucarate Dehydratases / Decarboxylases. *BMC Biotechnol.* 1–10 (2016)
247. Vrtis, J. M., White, A. K., Metcalf, W. W. & Van Der Donk, W. A. Phosphite dehydrogenase: A versatile cofactor-regeneration enzyme. *Angewandte Chemie - International Edition* vol. 41 3257–3259 (2002).
248. Shaked, Z. & Whitesides, G. M. Enzyme-Catalyzed Organic Synthesis: NADH Regeneration by Using Formate Dehydrogenase. *J. Am. Chem. Soc.* **102**, 7104–7105 (1980).
249. Ávila-Bolívar, B., García-Cruz, L., Montiel, V. & Solla-Gullón, J. Electrochemical reduction of CO2 to formate on easily prepared carbon-supported Bi nanoparticles. *Molecules* **24**, 1–15 (2019).

6 Appendices



Appendix 6.1. A phylogenetic tree of IlvD/EDD dehydratase related proteins divided into six groups or branches according to the previous study.[167] The red-magenta-green branch contains the sugar acid dehydratases with *Cc*XylDHT found in the red subgroup and *Rl*AraDHT. PuDHT and its subsequence homologous are expected to be located in the magenta branch. The brown branch contains the *Z. mobilis* and *S. oneidensis* 6-PGDHT (EDD) dehydratases. The cyan branch contains *E. coli* D-xylonate dehydratase YjhG, the blue branch contains *S. mutans*, *A. thaliana* and *L. lactis* [2Fe–2S] dihydroxyacid dehydratases (DHAD), and the orange branch *E. coli* [4Fe–4S] DHAD. The two black branches are undefined due to the lack of any biochemically characterized members.

```
Conservation:          5   699   6   8   5  5     75 6 5                                                  55     7
PuDHT           316 HLMEDFYYAGGVPAVIRELGD---VIARDALTVNGQTLWDNCKDAPNW-----------------------------------NREVIHAFNEPFKT 374
NmDHT           316 YLMEDFFYAGGLPAVIRELKP---VINGDTLTVNGRTLWDNCKDASNW-----------------------------------NRDVIHAYEHPFKE 374
VpepsDHT        319 FLMEDFCYAGGLPVVIKEIAE---HLHKDILTVTGQSLWDNVKDAENY-----------------------------------GPQVIRPLAEPFKE 377
VpDHT           319 FLMEDFCYAGGLPVVIKEIAQ---YLHKDAITANGQTLWDNVKDVENY-----------------------------------NPQVIRPLAEPFKE 377
RhDHT           320 FLMEDFCYAGGLPAVMKEIRS---HLHLDSITANGATVGQNIETAQNF-----------------------------------NAEVIQPLEKPFKD 378
LpDHT           319 HLMEDFCYAGGLPVIMKEIEH---LLHKDIITVTGKSVSENIAGAVNY-----------------------------------DPRVIKTFEAPFKE 377
HsDHT           317 FLMEDFYYAGGLPAVIRQLES---VIDKTALTANGKTLWENCQDAPNW-----------------------------------NEEVIRSFDKPFKE 375
RfDHT           317 HLMEDYYYAGGLPAVLREISA---HLHLGALTVNGKTLGENVASAPCW-----------------------------------NRDVIMTTSEPFKP 375
StDHT           320 FLMEDFFYAGGLPAVMREIQH---LLHLQALTVNGRTLGENIASAPCW-----------------------------------NREVIRPYTEPFKP 378
RlAraDHT        318 YLMEEFFYAGGLPVVLKRLGEAG-LLHKDALTVSGETVWDEVKDVVNW-----------------------------------NEDVILPAEKALTS 378
AbAraDHT        315 FPMEEFYYAGGLPAVLRRLGEGGLLPNPDALTVNGKSLWDNVREAPNY-----------------------------------DEEVIRPLDRPLIA 376
VpAraDHT        317 FLMEEFYYAGGLPAVLRRLGESGLLPHPGALTVNGRSIWDNVREAPSL-----------------------------------NDEVIRPLDKPLIA 378
VpepsAraDHT     316 FLMEEFYYAGGLPAVLRRLGENGLLPHPGALTVNGQSIWDNVREAPSL-----------------------------------NDEVIRPLDNPLIA 377
HsAraDHT        319 YLMEEFYYAGGLPAVLRRLGEADLLPHKDALTVNGQTMWDNVKDAPIY-----------------------------------NDEVVRPLAKPLIE 380
HsXylDHT        321 YLGERFHRAGGVPAIMWELQQAG-KLRAERITATGKTMAENLQGRASN-----------------------------------DREMIYPFAAPLRE 381
VpepsXylDHT     339 YLGEDYYRAGGVPAVVAELIAKG-KIQ-PALTVNGKTIIENCEGRFTE-----------------------------------DRKVIFPYAKPMMK 398
CcXylDHT        318 YLGERFHRAGGAPAVLWELLQQG-RLHGDVLTVTGKTMSENLQGRETS-----------------------------------DREVIFPYHEPLAE 378
RlXylDHT        321 FLGERFHRAGGTPAIMWELLQAG-KLDGNCRTVTGKTIAENLQGKEAR-----------------------------------DREVIKPFAEPLKE 381
EcDHAD          321 YHMEDVHRAGGVIGILGELDRAG-LLNRDVKNVLGLTLPQTLEQYDVMLTQDDAVKNMFRAG-PAGIRTTQAFSQDCRWDTLDDDRANGCIRSLEHAYSK 418
CgDHAD          322 YHMEDVHRAGGIPALLGELNRGG-LLNKDVHSVHSNDLEGWLDDWDIRSGKTTEEATELFHAAPGGIRTTEAFSTENRWDELDTDAAKGCIRDVEHAYTA 420
CnN1DHAD2       327 YHIEDVHRAGGIMAILGELERAG-KLHTDVPTVHAPTLKDALEQWDISRTQDEAVRNFYLAG-PAGIPTQVAFSQNTRWPSLDLDRAEGCIRSYEHAFSK 424
CmDHAD2         324 YHIEDVHRAGGILGILGELARAE-LLDLSCGNVHSGTLGNAISQWDVVGGAGEDAQKFFRAA-PGGIPTTVAFSQAATFESLDTDRKTGCIRNKQNAYSK 421
LlDHAD          325 YMMEDLHKIGGLPAVLKYLLKEG-KLHGDCLTVTGKTLAENVETALDLD-----------------------------------FDSQDIMRPLKNPIKA 388
AtDHAD          363 YVMEDIHKIGGTPAVLRYLLELG-LMDGDCMTVTGQTLAQNLENVPSLT-----------------------------------EGQEIIRPLSNPIKE 425
ScDHAD          339 YVMADLINVGGTQSVIKYLYENN-MLHGNTMTVTGDTLAERAKKAPSLP-----------------------------------EGQEIIKPLSHPIKA 401
SmDHAD          319 YVFQDLYKVGGVPAVMKYLLKNG-FLHGDRITCTGKTVAENLKAFDDLT-----------------------------------PGQKVIMPLENPKRE 381
CmDHAD          313 YVATDLHRAGGIPQVMKILLNAG-LLHGDCLTITGRTLAEELEHVPNEP-----------------------------------RTDQDVILPISQALYK 376
TvDHAD          313 YVTADLHRAGGIPQVMKMLLNAG-LLHGDCLTITGETIAERLRHVPDTP-----------------------------------DPNQDVIRPFDQPLYA 376
FtDHAD          315 YVATDFHRAGGVPQVLKMLLKAG-LIHGDALTITGQTLAEALANVPDTP-----------------------------------RADQDVIRPFDRPLYS 378
StDHAD          318 YLAVDLHRAGGIPQVMKILLNAG-LLHGDCITITGKTIAETLADVPDQP-----------------------------------PVDQDVIRPIDKPLYK 381
CnH16DHAD       313 YVATDLHKAGGIPQVMKILLKAG-MLHGDCLTITGRTLAEELENVPDTP-----------------------------------RADQDVILPIEKALYA 376
CnN1DHAD        313 YVATDLHKAGGIPQVMKILLKAG-MLHGDCLTITGRTLAEELENVPDTP-----------------------------------RADQDVILPIEKALYK 376
MbDHAD          332 HVMSDVDHIGGVPVVMKALLDAG-LLHGDCLTVTGHTMAENLAAITPPD-----------------------------------PDGKVLRALANPIHP 394
SsDHAD          313 YVMADLDEVGGVPVVLKKLLDAG-LLHGDVLTVTGKTMKQNLEQYKYPN-----------------------------------VPHSHIVRDVKNPIKP 376
Ec6PGDHT        342 ADINHFQAAGGVPVLVRELLKAG-LLHEDVNTVAGFGLSRYTLEPWLNNGE-----------------------LDWREGAEKSLDSNVIASFEQPFSH 416
Pa6GPDHT        342 ADINHFQAAGGMSFLIRQLLDGG-LLHEDVQTVAGPGLRRYTREPFLEDGR-----------------------LVWREGPERSLDEAILRPLDKPFSA 416
Zm6PGDHT        344 RDINEFQNAGGMAYVIKELLSAN-LLNRDVTTIAKGGIEEYAKAPALNDAG-----------------------ELVWKPAGEPG-DDTILRPVSNPFAK 418
So6PGDHT        342 ADINHFHAAGGMAFLIKELLDAG-LLHEDVNTVAGYGLRRYTQEPKLLDGE-----------------------LRWVDGPTVSLDTEVLTSVATPFQN 416
Consensus_aa:       @hhpch..hGGhs.lh.bL.psG.hlc.sh.TlsG.sl.c.hp.....................................sppll.sh.psh..
Consensus_ss:       hhhhhhhh   hhhhhhhhhh              hhhhh                                          ee  hhhh

Conservation:        765 57 967     6 7                     6 8 85              5      5 5 7  99   9 975 7         6
PuDHT           375 EAGIAVLRGNLCPDGAVIKPSAATP----------ALLKHKGRAVVFENSEHMHERMDDENLDVDEN--CVLVLKNCGPR--GYPGMAEAGNMPLPPKI 459
NmDHT           375 RAGIAVLRGNLCPDGAVIKPSAATP----------ALLKLRGRAVVFENSDHMHERLDDESLDVDEN--CVLVLKNCGPK--GYPGMAEAGNMPLPPKV 459
VpepsDHT        378 NAGICVLRGNLAPNGAIIKPSAATP----------ELLVHKGRAVVFESADDLHKRIDDEDLDIDEH--CIMVLKNCGPK--GYPGMAEAGNMPLPPKV 462
VpDHT           378 KAGICVLRGNLAPNGAIIKPSAATP----------ELLVHKGRAVVFESADDLHKRIDDENLDIDEH--CVMVLKNCGPR--GYPGMAESGNMPLPPKV 462
RhDHT           379 KAGIAVLRGNLAPRGAVIKPSAATP----------KLMVHKGRAVVFENIEDFHKRIDDENLDVDEN--CVMVLKNCGPK--GYPGMAEVGNMPLPPKV 463
LpDHT           378 KAGIAILRGNLAPRGAVIKPSAATP----------SLMVHTGRAVVFEDSTDFHKRIDDESLDVDEN--CILVLKNCGPK--GYPGMAEVGNMPLPPKV 462
HsDHT           376 AAGIAILKGNLAPDGAVIKPSAATP----------ALLKHRGRAVVFENSDDLHKRIDDENLDVDET--CVLVLKNCGPK--GYPGMAEAGNMPLPPKI 460
RfDHT           376 EAGIAVLRGNLAPDGAVIKPSAAST----------HLLQHRGRAVVFENIEDLNARIDDDALDIDEH--CVMVLKNCGPR--GYPGMAEVGNMPLPPKL 460
StDHT           379 AAGIAVLRGNLAPDGAVIKPSAASP----------HLLRHRGRAVVFENIEDFRARIDDESLDVDET--CVLVLKNCGPR--GYPGMAEVGNMPLPPKL 463
RlAraDHT        379 SGGIVVLRGNLAPKGAVLKPSAASP----------HLLVHKGRAVVFEDIDDYKAKINDDNLDIDEN--CIMVMKNCGPK--GYPGMAEVGNMGLPPKL 463
AbAraDHT        377 DGGIRILRGNLAPRGAVLKPSAASP----------ELLKHRGRAVVFENLDHYKATINDEALDIDAS--SVMVLKNCGPR--GYPGMAEVGNMGLPPKL 461
VpAraDHT        379 DGGIRILRGNLSPRGAVLKPSAASP----------ELLKHRGRAVVFENLEHYKERIVDESLEVDAN--SVLVMKNCGPK--GYPGMAEVGNMGLPPKL 463
VpepsAraDHT     378 DGGIRILRGNLSPRGAVLKPSAATP----------ELLKHRGRAVVFENLEHYKERIVDESLDIDAS--CVMVLKNCGPK--GYPGMAEVGNMGLPPKL 462
HsAraDHT        381 DGGICILRGNLAPRGAVLKPSAATP----------ELMKHRGRAVVFEDFNHYKERINDPDLDVDAS--CVLVMKNVGPK--GYPGMAEVGNMGLPPKV 465
HsXylDHT        382 RAGFLVLKGNLFD-FAIMKTSVISETFRERYLSTPGQENIFECRAVVFDGSDDYHARINDPALKIDEN--TLLAIRGAGPV--GWPGSAEVVNMQPPDAL 476
VpepsXylDHT     399 NAGFLDFSGNLFD-SAIMKTSVITPEFHNRYLEDPADPMAFEGTAVVFDGPEDYHRRIDDPKMKIDAH--SILFMRGVGPV--GYPGAAEVVNMRAPAYL 493
CcXylDHT        379 KAGFLVLKGNLFD-FAIMKSSVIGEEFRKRYLSQPGQEGVFEARAIVFDGSDDYHKRINDPALEIDER--CILVIRGAGPI--GWPGSAEVVNMQPPDHL 473
RlXylDHT        382 RAGFLVLKGNLFD-FAIMKMSVVSEDFRRRYLEEPGHEGVFKGKAVVFDGSEDYHKRINDPELGIDEN--TILVIRGAGPI--GWPGSAEVVNMQPPDHL 476
EcDHAD          419 DGGLAVLYGNFAENGCIVKTAGVDD-----------SILKFTGPAKVYESQDDAVEAILGGK--VVAG--DVVVIRYEGPK--GGPGMQEML--YPTSFL 499
CgDHAD          421 DGGLVVLRGNISPDGAVIKSAGIEE-----------ELWNFTGPARVVESQEEAVSVILTKT--IQAG--EVLVVRYEGPS--GGPGMQEML--HPTAFL 501
CnN1DHAD2       425 EGGLAVLRGNIALDGCVVKTAGVDE-----------SILVFEGTAHVTESQDEAVENILNDK--VKAG--DVVIVRYEGPK--GGPGMQEML--YPTSYI 505
CmDHAD2         422 DGGLAVLYGNLAEKGCIVKTAGVDE-----------SQWVFSGRARVFESQDEAVDAILGDK--VVAG--DVVVIRYEGPK--GGPGMQEML--YPTSYL 502
LlDHAD          389 TGHLQILYGNLAQGGSVAKISGKE-------------GEFFKGTARVFDGEQHFIDGIESGR--LHAG--DVAVIRNIGPV--GGPGMPEML--KPTSAL 467
AtDHAD          426 TGHIQILRGDLAPDGSVAKITGKE-------------GLYFSGPALVFEGEESMLAAISADP-MSFKG--TVVVIRGEGPK--GGPGMPEML--TPTSAI 505
ScDHAD          402 NGHLQILYGSLAPGGAVGKITGKE-------------GTYFKGRARVFEEEGAFIEALERGE--IKKGEKTVVVIRYEGPR--GAPGMPEML--KPSSAL 482
SmDHAD          382 DGPLIILHGNLAPDGAVAKVSGVK-------------VRRHVGPAKVFNSEEEAIEAVLNDD--IVDG--DVVVVRFVGPK--GGPGMPEML--SLSSMI 460
CmDHAD          377 QGHLAILKGNLAEEGAVAKITGLK-------------NPVITGPARVFEDEQSAMEAILADK--INAG--DILVLRYLGPK--GGPGMPEML--APTSAI 455
TvDHAD          377 TGHLAILKGNLASEGAVAKISGVK-------------NPQITGPARVFDSEEACLDAILAGK--INPG--DVIVIRYEGPV--GGPGMREML--APTSAI 455
FtDHAD          379 QGHLAILKGNLAPEGAVAKITGLK-------------NPAITGPARVFNSEDECMAAIMADK--IKPG--DVLVIRYEGPK--GGPGMREML--APTSAI 457
StDHAD          382 EGHLAILKGNLSPEGAVAKITGLK-------------NPVITGPARVFNDEQSALQAILDGK--IKPG--DVMVLRYLGPK--GGPGMPEML--APTGAL 460
CnH16DHAD       377 EGHLAILKGNLAEEGAVAKITGLK-------------NPVITGPARVFEDEQSAMEAILADK--INAG--DILVLRYLGPK--GGPGMPEML--APTSAI 455
CnN1DHAD        377 EGHLAILKGNLAEEGAVAKITGLK-------------NPVITGPARVFEDEQSAMEAILADK--INAG--DILVLRYLGPK--GGPGMPEML--APTSAI 455
MbDHAD          395 SGGITILHGSLAPEGAVVKTAGFD-------------SDVFEGTARVFDGERAALDALEDGT--ITVG--DAVVIRYEGPK--GGPGMREML--AITGAI 473
SsDHAD          377 RGGIVILKGSLAPEGAVIKVAATN-------------VVKFEGKAKVYNSEDDAFKGVQSGE--VSEG--EVVIIRYEGPK--GAPGMPEML--RVTAAI 455
Ec6PGDHT        417 HGGTKVLSGNLGRAVMKTSAVPVE-------------NQVIEAPAVVFESQHDVMPAFEAGL--LDRD--CVVVVRHQGPKANGMPELHKLM--PPLGVL 497
Pa6GPDHT        417 EGGLRLMEGNLGRGVMKVSAVAPE-------------HQVVEAPVRIFHDQASLAAAFKAGE--LERD--LVAVVRFQGPRANGMPELHKLT--PFLGVL 497
Zm6PGDHT        419 DGGLRLLEGNLGRAMYKASAVDPK-------------FWTIEAPVRVFSDQDDVQKAFKAGE--LNKD--VIVVVRFQGPRANGMPELHKLT--PALGVL 499
So6PGDHT        417 NGGLKLLKGNLGRAVIKVSAVQPQ-------------HRVVEAPAVVIDDQNKLDALFKSGA--LDRD--CVVVVKGQGPKANGMPELHKLT--PLLGSL 497
Consensus_aa:       pt.l.lLpGNLt..stlhK.ss.p.............hpt.A.VFcs.pph...l.s....lp.s..slhVl+..GPp..G.PGh.Eh.....ss.l
Consensus_ss:          eeeee       eeeeee              eeeeeeeee  hhhhhhhhh           eeee             hhhh  h  hhh
```

Appendix 6.2. Mulitple Sequence Alignment of the representative dehydratases in IlvD/EDD superfamily. List of the protein sequences are presented in Appendix 6.3. The MSA was performed using PROMALS3D. Amino acids colored magenta are predicted to form α-helix and amino acids colored blue to form β-sheet.

Protein	Organism source	NCBI Accession Number
***Pu*DHT**	*Paralcaligenes ureilyticus*	WP_132585145.1
***Nm*DHT**	*Noviherbaspirillum massiliense*	WP_019140327.1
***Vp*epsDHT**	*Variovorax paradoxus* (eps)	WP_013542644.1
***Vp*DHT**	*Variovorax paradoxus*	WP_021008351.1
***Rh*DHT**	*Ramlibacter henchirensis*	WP_135262325.1
***Lp*DHT**	*Limnohabitans parvus*	WP_108311471.1
***Hs*DHT**	*Herbaspirillum seropedicae*	WP_013235280.1
***Rf*DHT**	*Rhodanobacter fulvus*	WP_007081210.1
***St*DHT**	*Schlegelella thermodepolymerans*	WP_104358697.1
***Rl*AraDHT**	*Rhizobium leguminosarum*	WP_003592857.1
***Ab*AraDHT**	*Azospirillum brasilense*	WP_039340365.1
***Vp*AraDHT**	*Variovorax paradoxus*	WP_021005766.1
***Vp*epsAraDHT**	*Variovorax paradoxus* (eps)	WP_013539583.1
***Hs*AraDHT**	*Herbaspirillum seropedicae*	WP_013233070.1
***Hs*XylDHT**	*Herbaspirillum seropedicae*	WP_013236417.1
***Vp*epsXylDHT**	*Variovorax paradoxus* (eps)	WP_013543739.1
***Cc*XylDHT**	*Caulobacter crescentus*	WP_012640069.1
***Rl*XylDHT**	*Rhizobium leguminosarum*	WP_112904778.1
***Ec*DHAD**	*Escherichia coli*	WP_147715748.1
***Cg*DHAD**	*Corynebacterium glutamicum*	WP_074494098.1
***Cn*N1DHAD2**	*Cupriavidus necator* (N1)	WP_013952014.1
***Cm*DHAD2**	*Cupriavidus metallidurans*	WP_011519031.1
***Ll*DHAD**	*Lactococcus lactis*	WP_010905837.1
***At*DHAD**	*Arabidopsis thaliana*	NP_189036.1
***Sc*DHAD**	*Saccharomyces cerevisiae*	NP_012550.1
***Sm*DHAD**	*Streptococcus mutans*	WP_002265627.1
***Cm*DHAD**	*Cupriavidus metallidurans*	WP_011517373.1
***Tv*DHAD**	*Thermosynechococcus vulcanus*	WP_126985616.1
***Ft*DHAD**	*Fontimonas thermophila*	WP_091533200.1
***St*DHAD**	*Schlegelella thermodepolymerans*	WP_104358646.1
***Cn*H16DHAD**	*Cupriavidus necator* (H16)	WP_010814934.1
***Cn*N1DHAD**	*Cupriavidus necator* (N1)	WP_013957820.1
***Mb*DHAD**	*Mycobacterium tuberculosis*	WP_031680691.1
***Ss*DHAD**	*Saccharolobus solfataricus*	WP_009990927.1
***Ec*6-PGDHT**	*Escherichia coli*	WP_001548644.1
***Pa*6-PGDHT**	*Pseudomonas aeruginosa*	WP_128691989.1
***Zm*6-PGDHT**	*Zymomonas mobilis*	WP_011240289.1
***So*6-PGDHT**	*Shewanella oneidensis*	WP_011072452.1

Appendix 6.3. List of proteins used in the multiple sequence alignment presented in Appendix 6.2.

7 Abbreviations

(R)-	Rectus
(S)-	Sinister
[Fe-S]	Iron-sulfur cluster
[O]	Oxidation step
°C	Degree Celsius
3D	Three-dimensional
Å	Ångstrom (1 x 10^{-10} m)
APS	Ammonium persulfate
Ara	L-Arabinonate
BCAA	Branched-chain amino acids
BDO	1,4-Butanediol
BSA	Bovine serum albumin
BTO	1,2,4-Butantriol
C	Carbon
CAST	Combinatorial Active-site Saturation Test
CBP	Consolidated Bioprocessing
cPCR	Colony PCR
D-	Dextrorotary
DHAD	Dihydroxy-acid dehydratase
DHIV	(R)-2,3-dihydroxyisovalerate
DHMV	(2R,3R)-dihydroxy-3-methylvalerate
DHT	Dehydratase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotides
ED	Entner-Doudoroff
EDTA	Ethylenediaminetetraacetic acid
epPCR	Error-prone PCR
g	Gram
h	Hour
H_2O_2	Hydrogen peroxide
HBL	Hydroxy-γ-butyrolactone

HEPES	4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid
HotSysAPP	Hot System Biology Applied
HPLC	High Performance Liquid Chromatography
HTS	High-throughput screening
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β- d-1-thiogalactopyranoside
ISM	Iterative Saturation Mutagenesis
k_{cat}	Turnover numer
k_{cat}/K_M	Catalytic efficiency
KDC	Branched-chain α-keto acid decarboxylase
KdcA	An isosenzyme of KDC from *Lactococcus lactis*
KDG	2-Keto-3-deoxy-gluconate
KDP	2-Keto-3-deoxypentonate
KdpD	2-Keto-3-deoxypentonate dehydratase
KGSA	2-Ketoglutarate semialdehyde
KIV	α-ketoisovalerate
KivD	An isosenzyme of KDC from *Lactococcus lactis*
K_M	Michaelis-Menten constant
KPi	Pottasium phosphate buffer
L-	Levorotary
LB	Lysogeny broth
m	Meter
M	Molar
mg	Milligram
min	Minute
ml	Milliliter
mM	Millimolar
MSA	Multiple Sequence Alignment
MSSM	Multiple Site Saturation Mutagenesis
N	Nitrogen
NAD^+	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
NOX	NADH Oxidase

O	Oxygen
oePCR	Overlap extension PCR
PCR	Polymerase Chain Reaction
PDC	Pyruvate decarboxylase
QC	QuikChange
R^2	Coefficient of determination
RNA	Ribonucleic acid
rpm	Rotation per minute
s	Second
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
SeSaM	Sequence Saturation Mutagenesis
SHIPREC	Sequence Homology-Independent Protein RECombination
ssDNA	Single strand DNA
StEP	Staggered Extension Process
T_{50}^{x}	Temperature at which an enzyme losses 50% of its initial activity after incubation for a given unit time (x)
TAE	TRIS Acetic acid EDTA buffer
TB	Terrific broth
TEMED	N,N,N,N-Tetramethylethane-1,2-diamine
ThDP	Thiamine diphosphate
TIM	Triosephosphate isomerase
T_m	Melting temperature
TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
vol%	volume per volume (1 vol% is 1 ml of compound in 100 ml water)
w/v%	weight per volume (1 w/v% is 1 g of a compound in 100 ml water)
WT	Wild type
xg	Earth´s gravitational force (10 N/kg)
α-KG	2-Ketoglutarate

8 List of Tables

9 List of Figures