TECHNISCHE UNIVERSITÄT MÜNCHEN
Integrative Research Center Campus Straubing für Biotechnologie und Nachhaltigkeit

Development of a cell-free enzymatic cascade towards the production of L-alanine

Tobias Gmelch *M.Sc.*

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. Rubén Dario Costa Riquelme
Prüfer der Dissertation:	1. Prof. Dr. Volker Sieber
	2. Prof. Dr. Iris Antes

Die Dissertation wurde am 07.07.2020 bei der Technischen Universität München eingereicht und von der promotionsführenden Einrichtung Campus Straubing für Biotechnologie und Nachhaltigkeit am 20.11.2020 angenommen.

Acknowledgements

First and foremost I want to thank my adviser Prof. Dr. Sieber who gave me this opportunity to work on enzyme cascades. You trusted me as a chemist to be efficient in biotechnology! Thank you for the continuous support and advise.

Prof. Dr. Rubén Dario Costa Riquelme and Prof. Dr. Iris Antes I want to thank for the acceptance to the thesis committee.

My mentor Dr. Elisabeth Jacobsen from the Norwegian University of Science and Technology (NTNU) in Trondheim I want to thank for career advises and the possibility to be part of the organizing committee of the 4th Multistep Enzyme Catalyzed Processes Congress in Trondheim.

A special thank goes to Dr. Josef Sperl who always took time to discuss my problems and teach me new techniques in the lab. Without your encouragement, patience and motivation, this work would not have been possible. Thank you for proofreading all my articles and this thesis.

Dr. Broder Rühmann, I want to thank for the fruitful discussions about analytic problems and his patience during HPLC issues. For the support with technical issues in the lab, I would like to thank Magdalena Haslbeck, Manuel Döring, Petra Lommes and Anja Schmidt and Elisabeth Aichner I thank for the ongoing support with all the bureaucracy!

Thanks to all colleagues at the Chair of Chemistry of Biogenic Resources who made this a great time for me. Especially, I would like to thank Janine Simon, Dr. Sumanth Ranganathan, Torben Hüsing and Christoph Schilling for the amazing time I had in Straubing and the great moments we spent together. Torben and Christoph, I really miss our hiking trips and the awesome discussions we had. Additionally I want to thank Samuel Sutiono, Ioannis Zachos and Samed Güner for all the scientific hints and discussions, which made my lab work a lot more efficient.

Finally but most importantly, I would like to thank my family.
Mama und Papa ihr seid die Besten! Vielen Dank für all die Möglichkeiten und eure Unterstüzung! Ich vermisse dich Papa.
Claudia, you are my rock I can always count on! Thank you for all the motivation and encouragement in every possible way – I love you.

Contents

Summary

Cell free enzymatic reaction cascades combine isolated enzymes from various organisms to *in vitro* pathways, in order to overcome the limits of classical fermentation approaches. The major advantage of systems biocatalysis, as this approach is termed nowadays, is the freedom in pathway design which means that biocatalysts from different natural pathways can directly be combined to artificial pathways leading to novel products. The process parameters (e.g. pH or salt concentration) of such reaction cascades can be easily fine tuned since the enzymes are not compartmentalized within a regulatory cell environment.

Within this work, a cell free enzymatic cascade was designed for the production of L-alanine from commonly available D-glucose which serves as model substrate for biomass. Deduced from the archeal Entner-Doudoroff pathway, this minimized glycolytic degradation combines an oxidative and a reductive module to a redox balanced cascade. The oxidative module is composed of five enzymes yielding two pyruvate equivalents together with two redox-equivalents NADH from one molecule of glucose. Cofactor regeneration is then coupled to L-alanine formation via L-alanine dehydrogenase within the reductive module. In sum, D-glucose shall be converted atom efficiently to two equivalents of L-alanine, including an internal regeneration of the sole cofactor NAD^+.

Starting from a previously designed reaction cascade towards ethanol, the oxidative module was investigated and an in detail literature review and extensive screening effort resulted in novel promising candidates with improved properties. This new set of biocatalysts demonstrated increased activity and specificity under the predefined cascade conditions. The kinetic characterization also included saturated alanine conditions, since very high product titers were to be achieved. While the oxidative module seemed to be generally unaffected, the alanine dehydrogenase showed a drastic decrease in the overall activity.

Once the enzymes were identified and kinetically characterized, an optimization of the enzyme ratio and the process parameters (such as buffer and cofactor concentration and ammonium source) was conducted in order to identify the most economic values. For this analysis, several quantitative alanine assays were compared according to their compatibility with the complex reaction matrix. While the rather slow HPLC analysis as well as ninhydrin or *o*-phthaldialdehyde assays gave only unsatisfying results, fluorescamine was able to selectively derivatize alanine within the complex reaction matrix. Combining the developed fluorescence based high-throughput assay with an enzyme titration study, an optimized enzyme ratio yielded the maximum cascade yield already at 50% of the original enzyme loading.

In the second part of the thesis, the aldehyde dehydrogenase from *Thermoplasma acidophilium* was engineered towards increased NAD^+ acceptance. This highly selective and thermostable enzyme can be applied for the cofactor dependent oxidation of D-glyceraldehyde within the cascade. Unfortunately, the naturally existing enzyme prefers $NADP^+$ over NAD^+ as demonstrated by an approximately 1000 times large K_M for NAD^+, which diminishes its other benefits. Since a previous random mutagenesis approaches did not significantly increase the NAD^+ acceptance, a rational design approach on the basis of

the available crystal structure was conducted. Alteration of four key residues in the proximity of the cofactor binding site could increase the specificity towards NAD^+ by a factor of 14 and slightly increase the activity without influencing the substrate specificity for its second substrate D-glyceraldehyde. Additional modifications at the substrate tunnel entrance could successfully erase the D-glyceraldehyde substrate inhibition, double the activity and further decrease the K_M for NAD^+ by a factor of two. In order to understand the effects of the mutations on a fundamental basis, molecular dynamics simulations have been conducted. Enzyme-glyceraldehyde-NAD^+ complexes of original template and final mutated variant were simulated for 10 ns and analysis of the Δ-RMSF revealed a subtle increase in flexibility within cofactor binding areas. Concerning the D-glyceraldehyde inhibition, the increase of the substrate tunnel opening was found to be responsible for the obliteration of the inhibition. Overall the k_{cat}/K_M for NAD^+ could be increased by a factor > 100 with only minor effects on the temperature stability. Additional improvements in solvent stability as well as in artificial cofactor acceptance, render this mutant biocatalyst a promising variant for further cascade designs.

List of Publications

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).

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. 9, 920−929 (2020).

1 Introduction

1.1 Systems biocatalysis

Catalysis is the process of decreasing the activation barrier of a chemical reaction in order to speed up the turnover. Within this field, biocatalysis focuses on catalytically active proteins called enzymes, which are usually produced within living systems and are commonly biodegradable. During natural evolution, combinations of those highly functional biocatalysts have been developed in order to sustain the metabolism of every organism. Due to the (multistep) cascade set up of these metabolic pathways, the concentration of unstable intermediates can be kept at a minimum. Furthermore, the additional pull of downstream reactions can drive thermodynamically unfavorable reactions to completion. Fermentation or whole cell catalysis approaches apply those established pathways directly and convert substrates to the desired products *in vivo* and with the present set of enzymes. Although metabolic engineering for well described (and sequenced) organisms led to highly specified production strains which are still applied industrially for e.g. fermentation of glutamate,[1] laborious design-build-test cycles are commonly required to identify suitable gene knock outs or gene replacements.[2] Despite the use of computer based pathway predictions and decades of experience, issues like substrate channeling or the elimination of side products remain challenging for *in vivo* systems.[3]

In contrast to this, the emerging concept of systems biocatalysis deals with the *in vitro* combination of isolated enzymes towards cell free reaction cascades. Such operations are free from competing metabolic pathways, from kinetic restrictions by physical barriers and regulating circuits, and from toxicity problems with reactive foreign substrates, which are notorious problems in whole-cell systems.[4] Due to the absence of a regulatory cell environment, fine tuning of the reaction conditions is easier compared to *in vivo* approaches. Although the required enzymes need to be purified, the level of complexity is drastically reduced and novel products are enabled by artificial cascades. Economic applications for systems biocatalysis are still in the research stage, but with the prizes for enzyme purification decreasing[5] and the awareness for sustainable, oil independent processes increasing, the time of change might be getting closer.

Cell-free metabolic engineering[6], synthetic biochemistry system[7] and *in vitro* metabolic engineering[3] are other terms for the concept of systems biocatalysis.

1.1.1 Redox neutral pathway design

The major advantage of systems biocatalysis is the freedom in pathway design because enzymes can be directly combined *in vitro* in order to obtain the desired substrate to product flux. Economically feasible production pathways can be obtained by minimizing the amount of enzymatic steps and the types of cosubstrates (e. g. only NAD/H as redox shuttle). Especially cost intensive redox cofactors such as nicotinamide or flavine based nucleotides are commonly required in stochiometric amounts and need to be regenerated in order to keep the production costs low. Clearly, this regeneration needs to be very

accurate in order to avoid degradation of the cofactor. If 50% of the original cofactor activity is to be remained after 100 turnovers, the regeneration efficiency must be 99.3%.[9] Regeneration can be achieved electrochemically,[10] photochemically[11] and enzymatically.[12] Enzymatic regeneration is preferred for industrial processes due to its high selectivity and efficiency.[13] Two different approaches have been realized:

A) Addition of a regeneration enzyme which converts a cheap scavanger substrate (e.g. glucose dehydrogenase converting glucose and NAD^+ to gluconate and NADH). This approach can be easily integrated into pathways, since it is completely decoupled from the other reaction. With an excess of scavenger substrate, a rapid regeneration can be guaranteed, thus providing constantly high concentrations of regenerated cofactor. On the downside an extra enzyme is needed and the scavenger substrate as well as the reaction product need to be removed during the downstream process. Nevertheless, this approach is widely pursued in pharmaceutical industry for the synthesis of high value chiral compounds.[14]
B) Self sufficient cascade design. In this case, the regeneration enzyme is part of the production pathway and equal amounts of reduced and oxidized cofactors are produced during the reaction turnover. With this redox neutral strategy, scavenger substrates as well as additional enzymes for the regeneration can be avoided completely. Drawbacks here are the increased complexity during cascade design and the direct coupling of the cofactor regeneration to the production pathway. If the cofactor dependent reactions are not directly following each other in the multistep pathway, an efficient regeneration can only be guaranteed when the intermediate reactions are sufficiently fast in order to generate the substrate for the recycling reaction. An outstanding example for this approach is the cell free production of monoterpenes from glucose, where 27 enzymes have been combined *in vitro* (including internal NAD(P)H and ATP recycling) in order to reach product titres > 15 g/L.[15]

Another interesting example for the cofactor regeneration is the concept of molecular purge valves, where the regeneration of the cofactor is decoupled from the substrate flux (like A) but also balanced (like B). This is particularly useful in redox neutral cascades, when the amount of NADH is decreased by side reactions. Opgenorth *et al.* demonstrated the balanced production and consumption of NADPH and NADH by using two different pyruvate dehydrogenases that selectively accept either $NADP^+$ or NAD^+ in combination with an NADH oxidase that does not oxidize NADPH.[7] Within their PHB or isoprene producing test systems, they could maintain almost quantitative yields and high levels of NADPH completely independent to up to 10 fold variations in the intial cofactor concentration. With this proof of concept, other examples for this novel recycling strategy might follow in the future.

1.1.2 *In vitro* enzymatic cascade reaction from D-glucose to ethanol as blueprint for the L-alanine cascade

Ethanol from biogenic resources is used as a fuel additive, slowing down the depletion of fossil resources.[16] The biocatalytic production is commonly done with an *in vivo* fermentation process applying various microorganisms (e.g. *saccharomyces* strains).[17] The major disadvantage of this process is the cell toxicity of the produced ethanol, limiting the product titer to around 20 vol-%. Applying the concepts of systems biocatalysis, Guterl *et al*. could demonstrate molecular yields of up to 57 % from glucose.[6] Their artificial glycolytic pathway can be devided into two modules: (1) The pyruvate module, where one molecule of glucose is converted into two molecules of pyruvate, and (2) the ethanol formation module (Figure 1).

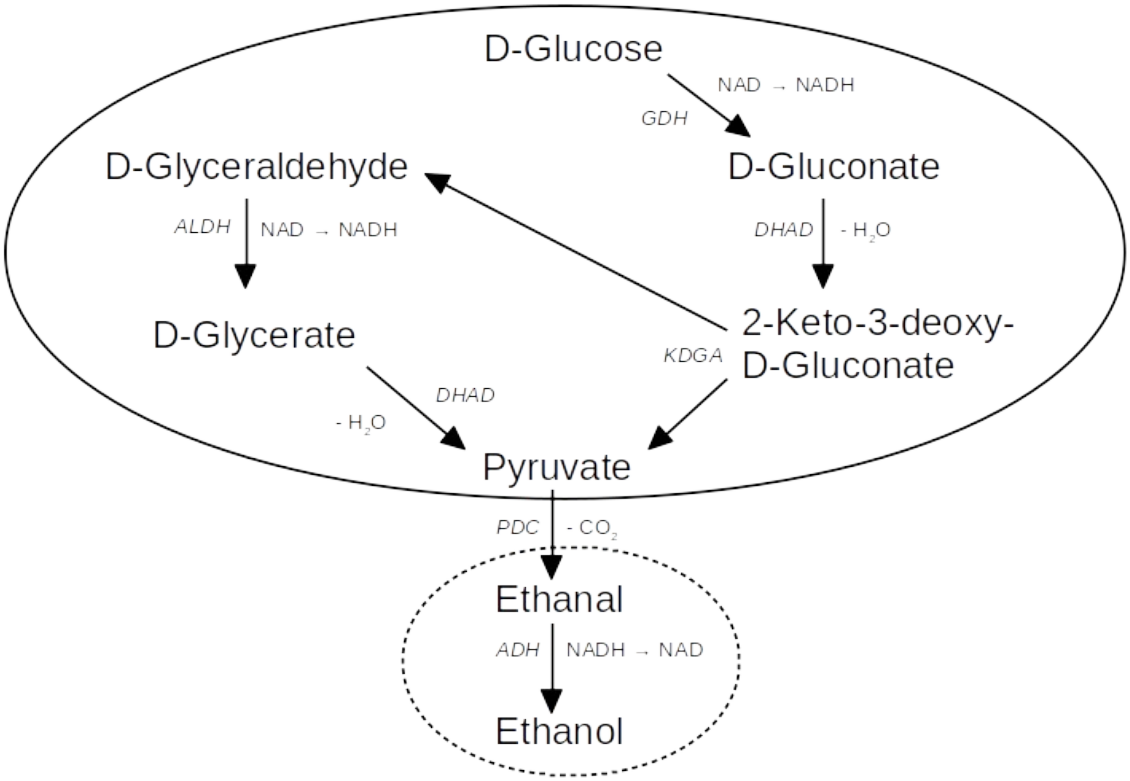


Figure 1: Cell free enzymatic cascade design from D-glucose to ethanol. Adapted from Guterl et al.[6]

The pyruvate module starts with the oxidation of glucose towards gluconate which is catalyzed by an NAD-dependent glucose dehydrogenase (GDH). A dihydroxyacid dehydratase (DHAD) was used for the subsequent dehydration towards 2-keto-3-deoxy-gluconate which is then broken up into D-glyceraldehyde and pyruvate by an aldolase (KDGA). D-glyceraldehyde is further oxidized to D-glycerate by an NAD-dependent aldehyde dehydrogenase (ALDH) from *Thermoplasma acidophilum*, since this variant was found to selectively oxidize D-glyceraldehyde and not ethanal, which is also an intermediate of the cascade. The dehydration of D-glycerate finally results in the second pyruvate equivalent. Due to a substrate promiscuity of the applied DHAD from *Sulfolobus solfataricus*, both dehydration reactions could be performed by the same enzyme. In sum,

the pyruvate module generated two equivalents of pyruvate from one equivalent of D-glucose and additionally two redox equivalents NADH.
The second module uses two additional enzymes in order to produce the final product ethanol with simultaneous regeneration of the two redox equivalents. In the first step pyruvate is decarboxylated by pyruvate decarboxylase (PDC) before the NADH-dependent alcohol dehydrogenase (ADH) reduces the acetaldehyde to ethanol.
In total this artificial reaction cascade consists of six enzymes and is redox neutral due to an internal NAD^+ recycling. Additionally, an increased solvent stability of the cell-free system compared to the whole cell process was pointed out with the applied aldehyde dehydrogenase being least solvent stable. Minimization of the ordinary glycolytic pathway was mainly achieved by the substrate promiscuity of the DHAD, eliminating the need for phosphorylation. Although a molecular yield greater than 50 % indicates a successful redirection of the produced D-glyceraldehyde towards pyruvate, especially the dehydration of D-glycerate needs further investigation. Nevertheless, this proof of concept demonstrates a brilliant cascade design with internal regeneration and high atom economy, which can be used as platform for future cell free approaches.
On the basis of this artificial ethanol production pathway, a novel L-alanine producing cascade was to be developed within this work. Therefore, the pyruvate module should be combined with a suitable alanine dehydrogenase and an ammonia source in order to produce the desired product L-alanine and simultaneously regenerating the consumed cofactor NAD^+.

1.2 Methods for the production of L-alanine

L-alanine is a natural, non essential amino acid, which forms a white solid at room temperature. It is the smallest, chiral amino acid and mainly used as food or feed additive because of its sweetening properties[18] or in health industry as supplement for nutrition therapies.[19] In order to evaluate the novel enzymatic cascade, the industrially relevant production processes of L-alanine are briefly explained in the following sections.

1.2.1 Extraction from protein hydrolysates

With 29.7 % L-alanine[20] in the average protein backbone, the recovery from hydrolysates seems promising on the first glimpse. The enantiomeric excess from naturally derived amino acids is exceptionally high, since nature developed a preference for L-amino acids and thereby eliminated the need of an enantio-separation. On the downside, the desired L-alanine is always accompanied by other, chemically similar, amino acids which need to be depleted.
Before any separation can take place, the peptide bonds of naturally occurring protein sources such as silk fibroin, gelatin or zein need to be hydrolyzed. In the second step, the amino acids are separated based on different solubilities of the corresponding copper[21], chromium[22] or sulfonic acid[23] salts. Due to interfering effects of other amino acids, L-alanine cannot directly be obtained within a single extraction step, but an excellent recovery of 96 % pure L-alanine was achieved by stepwise extraction of strontium-picrate complexes.[24] Within this process, cysteine and tyrosine are extracted from the solidified hydrolysate with water. Then, the aromatic and basic amino acids are removed with

charcoal-adsorbtion[25] and glycine is precipitated as glycin-copper-picrate complex. Finally, alanine and proline are coprecipitated as strontium-picrate complexes. After acidic decomposition of both complexes, proline is extracted with ethanol and pure L-alanine is obtained via recrystallization.
Although multiple, enantiopure amino acids can be obtained during this process, the up-scaling of such extraction processes for industrial production is rather uneconomical. Furthermore is the yield of the specific amino acid depending on the occurrence in the hydrolysate.

1.2.2 Chemical synthesis

In contrast to the extraction of amino acids from hydrolysates, chemical synthesis can be designed to result directly in the desired product. The key factors for a successful chemical production include cheap and easily available raw materials, high product yields, sufficient purity and facile scalability. For amino acid synthesis, excellent asymmetric induction is necessary in order to avoid a tedious separation of the enantiomeres.
The most relevant synthesis of α-amino acids was accidentally discovered in 1850, when Adolph Strecker tried to produce lactic acid from acetaldehyde, aqueous ammonia and hydrogen cyanide.[26] Instead of lactic acid, D/L-alanine was obtained after acidic hydrolysis. This one pot, three component reaction is since known as the Strecker reaction.[27] Due to the extreme toxicity of HCN, improved methods use alkali cyanides such as KCN or NaCN as substitute. The reaction mechanism (Figure 2) starts with the nucleophilic addition of the ammonium **II** (or primary/ secondary amine) which leads to the imine **IV**. In the next step, the addition of the cyanide produces the α-amino nitrile **V** which is then hydrolyzed to the α-amino acid. Although asymmetric induction is possible with either chiral amines[28,29] or chiral metal- or organocatalysts,[30,31] the total production costs (including catalyst, purification and toxic waste disposal) are too high for standard amino acids like unsubstituted L-alanine.

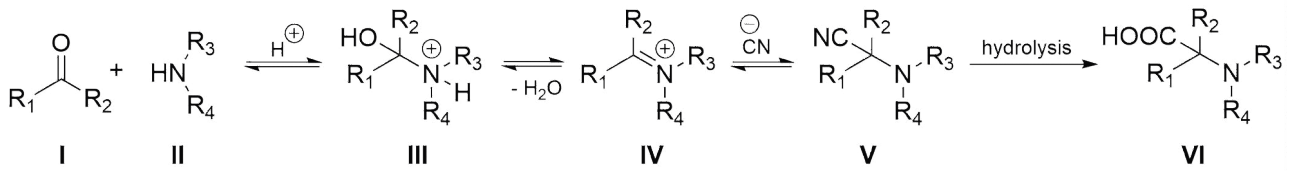


Figure 2: Mechanism of the Strecker reaction

1.2.3 Fermentation

L-alanine is essential for protein synthesis, hence every living organism needs to acquire them by either biosynthesis or uptake from the environment. Microorganisms, such as *E.coli*, are independent from the amino acid uptake and can synthesize all proteinogenic amino acids on their own, only requiring ammonia as the sole source of nitrogen.[32] Although the energy demanding synthesis of amino acids is normally regulated by a feedback loop, several strains of bacteria, yeast and fungi were found to achieve an overproduction resulting in an accumulation of specific amino acids.[33] Once produced, the

amino acids usually remain inside the cell, where they can be used by the organism for e.g. protein synthesis. Surprisingly, recent reports identified an L-alanine transport system[34] and speculate about an additional D-alanine transporter[35] for the transfer of both enantiomers into the extracellular space. In order to further increase the permeability of the cell walls, biotin limitation,[36] high concentration of $C_{16\text{-}18}$ fatty acids[37] or β-lactam antibiotics (e.g. penicillin)[38] were applied to increase amino acid concentrations within the media. Concerning the L-alanine fermentation, one major problem is the alanine-racemase activity, which is present in most wildtype strains.[39] Extensive research resulted in genetically engineered organisms (e.g. *Vibrio natriegens,*[40] *Lactococcus lactis*[41] or *Zymomonas mobilis*[42]) which are tailored towards an efficient production of optically enriched L-alanine. This work focuses on engineering of *Escherichia coli*, since it is one of the best investigated and understood microorganisms.

As depicted in the metabolic overview (Figure 3), sugar metabolism leads to the central intermediate pyruvate, which is then converted by an NADH-dependent alanine dehydrogenase (*alaD*).

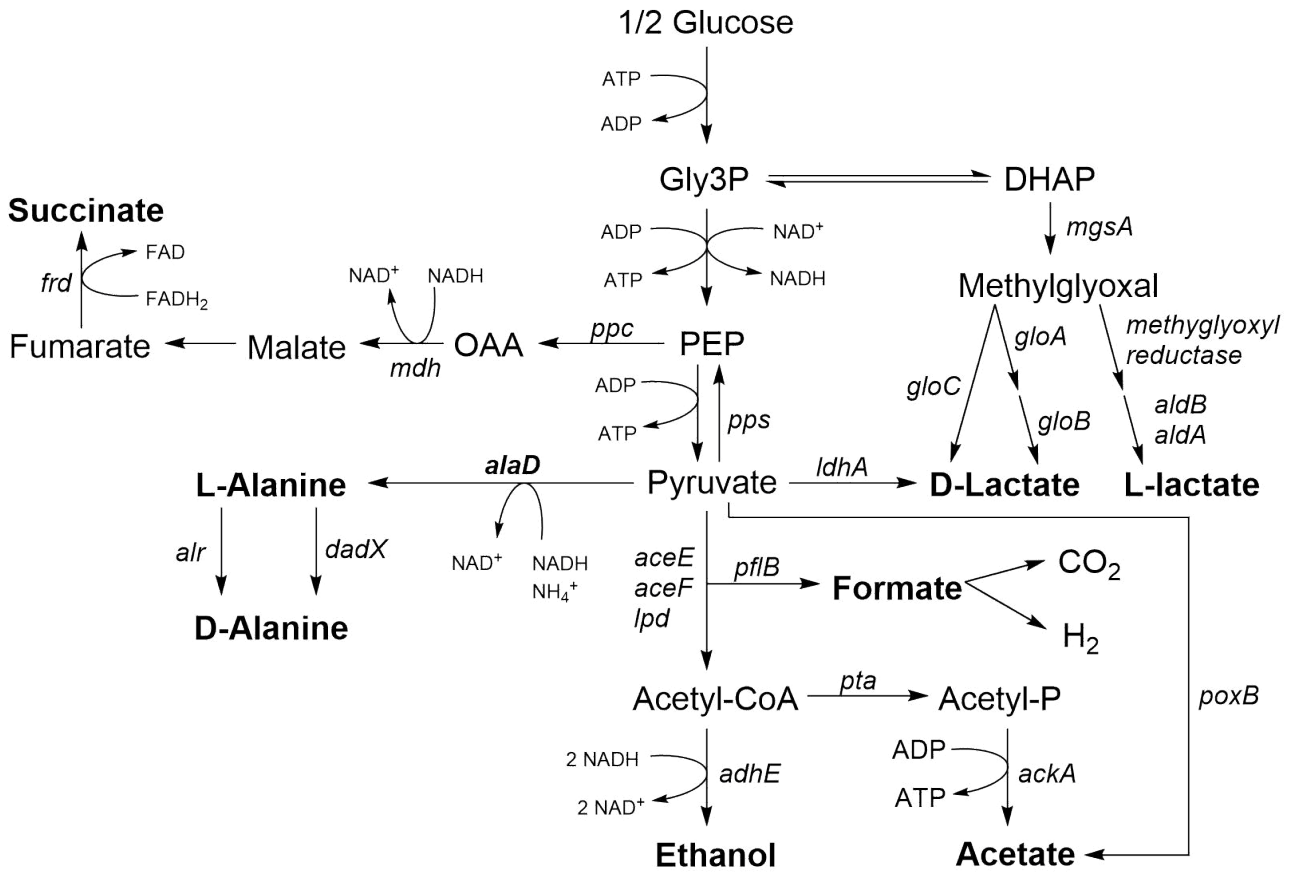


Figure 3: Metabolic overview of the L-alanine fermentation pathway in recombinant *E. coli,* according to Ingram et al*.* [43]

Since the *alaD* activity is not naturally present in *E. coli,* variants from either *Arthrobacter* or *Bacillus* have been coexpressed.[44] The resulting redox balanced pathway can theoretically produce two eq. of alanine from one eq. of D-glucose (theoretical mass yield of 0.978). NADH produced during the catabolic conversion of glucose to pyruvate is used for the reductive amination towards L-alanine. The equillibrium of this reaction strongly

favors the L-alanine formation with an equilibrium constant of about 10^{-17} M.[45] Due to the rather low NADH/NAD^+ ratio in aerobic culture of 0.02,[46] the final alanine yield is directly reduced if the NADH concentration is decreased by side reactions. Most engineering approaches thus focus on the elimination of such side reactions or on the channeling of the central intermediate pyruvate towards the L-alanine production.

The predominant carbon flux from pyruvate to acetyl CoA is catalyzed by the pyruvate dehydrogenase complex which is encoded in *aceF*, *aceE* and *lpd*. A knockout of the dihydrolipoyl transacetylase Δ*aceF* enabled the accumulation of nearly 40 g/L pyruvate under aerobic conditions.[47] Another direct competitor for both, pyruvate and NADH is the lactate dehydrogenase *ldhA*.[48] Pyruvate-formate lyase (*pfl*),[49] expressed under anaerobic conditions, as well as the pyruvate oxidase (*poxB*) and the PEP synthase (*pps*) can also reduce the pyruvate pool.[50] Knockouts of these genes can increase the concentration of available pyruvate and therefore directly increase the L-alanine production rate.

During normal growth of *E.coli*, NADH is consumed by oxygen via oxidative phosphorylation and NADH oxidases. Lee et *al*. could demonstrate a 6 fold increase of the alanine yield (up to 0.63 g/g glucose) by reducing the oxygen mass transfer from 109 h^{-1} to 7 h^{-1}.[51] Smith et al[50] separated the growth phase from the production phase because their engineered *E.coli Δpfl*, *ΔldhA* did not grow under anaerobic conditions.[52] The applied fed-batch approach limited the side product acetate which possibly occurred as a result of an "overflow mechanism" due to high growth rates.[53–55] Overall, they achieved a complete conversion of glucose to racemic alanine during the anaerobic production phase resulting in a productivity of 4 g/Lh and a final concentration of 88 g/L. From the two alanine racemase genes in *E. coli*, *alr* is constitutive and less abundant than the predominate racemase encoded by *dadX*.[56] In order to increase the optical purity, a knockout of the *dadX* gene is necessary, while the ALR racemase is still able to produce the essential D-alanine.[43] In combination with the knockouts of acetate kinase *ackA*, alcohol dehydrogenase *adhE* and methylglyoxylate synthase *mgsA*, Ingram et al. developed a metabolic evolution platform where the cell growth is coupled to alanine production.[43] They further optimized their aerobic process by integrating the less active alanine dehydrogenase from *G. stearothermophilus* under the regulation of the native lactate dehydrogenase (*ldhA*) and finally obtained 1.28 M optically pure L-alanine from 120 g/L glucose within 48 h during a batch fermentation in mineral salts medium. Another approach, published by Yokota et al., applied an H^+-ATPase- and ldhA defective *E. coli* mutant which showed increased pyruvate production.[57] With 20 g/L DL-alanine from 50 g/L glucose after 24 h of aerobic fermentation, the yield of 41 % stayed behind the expectations.[58] The increase of the glycolytic flux led to 16 g/L pyruvate, which could possibly not be converted to alanine due to the aerobic conditions.[59] An overview of the described strains is shown in Table 1.

Finally a study from Zhou et al.[60] found a significant inhibition of the cell growth by L-alanine, why they focused again on a two step process. They integrated the *alaD* from *G. stearothermophilus* under the thermo-regulated p_R-p_L promotor hence separating the aerobic growth phase from the oxygen-limited alanine production phase. After optimization of the induction procedure they reached 121 g/L L-alanine due to the increased biomass.

Downstream processing of the L-alanine fermentation broth normally starts with separation of the cells by either centrifugation or filtration. Since the formation of side products is reduced via knockouts of competing pathways, the L-alanine content of the broth easily exceeds 90 wt-%. Residual media compounds or byproducts such as succinate, formate or acetate can then be reduced by crystallization or ion exchange.[32]

Table 1: Overview of alanine-producing strains, n.r. = not reported

Organism	Modification	Time [h]	Alanine [g/L]	Yield [%]	L-Alanine purity [%]
A. oxydans DAN 75[61]	Alanine racemase deficient	120	77	51	98
Z. mobilis CP4 (pZY73)[42]	Plasmid with *B. sphaericus* IFO3525 *alaD*	26	8	16	n.r.
L. lactis NZ3950 (pNZ2650)[41]	Plasmid with *B. sphaericus* IFO3525 *alaD,* Δ*ldhA*, Δ*alr*	17	13	70	85-90
E. coli ALS887 (pTrc99A-alaD)[51]	Plasmid with *B. sphaericus* IFO3525 *alaD,* Δ*ldhA*, Δ*aceF*	27	32	63	n.r.
E. coli ALS929 (pTrc99A-alaD)[50]	Plasmid with *B. sphaericus* IFO3525 *alaD,* Δ*pfl,* Δ*pps,* Δ*poxB,* Δ*ldhA*, Δ*aceEF*	48	88	100	n.r.
E. coli XZ132[43]	Integrated *G. stearothermophilus alaD*, Δ*pfl,* Δ*ackA,* Δ*adhE,* Δ*ldhA,* Δ*mgsA,* Δ*dadX*	48	114	95	>99
E. coli TBLA-1[58]	Plasmid with *G. stearothermophilus alaD,* Δ*ldhA, (atpA401, bgl$^+$)*	24	20	41	n.r
E. coli B0016-060BC[60]	Δ*ack-pta,* Δ*pflB,* Δ*adhE,* Δ*frdA,* Δ*ldhA* Δ*dadX::cI ts 857-p R -p L -alaD-FRT*	39	121	97	n.r

1.2.4 Enzymatic production

In contrast to the fermentation of living microorganisms where the entire metabolism is involved in the product formation, enzymatic production relies on selected biocatalysts for the conversion of precursors. Approaches via isolated alanine dehydrogenase systems comparable to the fermentation pathway (e.g. using malate as substrate)[62] are depending on expensive cofactors and cannot easily be upscaled. Efficient L-alanine production is thus achieved from aspartic acid using aspartate β-decarboxylase (EC 4.1.1.12) which was found in multiple organisms such as *Clostridia*,[63] *Nocardia*[64] or *Achromobacter*.[65] Chibata et al. identified *Pseudomonas dacunhae* (PdAspD) as the most advantageous organism which can directly be used in a whole cell process for the accumulation of L-alanine from aspartic acid.[66] Within their initial report, they found an induction of the enzyme expression by ammonium fumarate with led to 24 times increased volumetric activity. Furthermore the presence of surfactants could again double the conversion rate, resulting in a stoichiometric conversion of 40 % (w/v) L-aspartic acid at 37 °C and 72 h.[66] An analysis of the crystalline enzyme showed an allosteric activation by a-ketoglutarate[67] as well as a dependence on pyridoxal 5'-phosphate[68] which is in accordance to other mechanistic studies.[69,70]

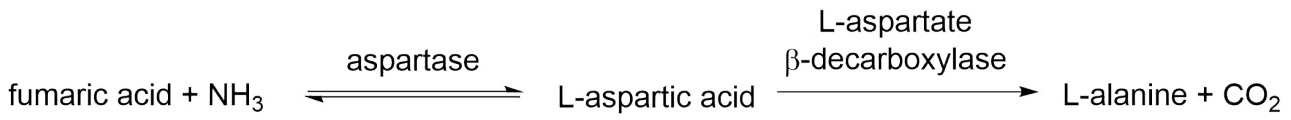


Figure 4: Enzymatic L-alanine production from fumaric acid

Since the required aspartic acid is also produced enzymatically from fumarate (Figure 4), the combination of these two reactions can further facilitate the process. For the first step, naturally occurring aspartase from *E.coli* (EcAsp) was selected because of its high activity and stability.[71] Unfortunately the whole cell application of *E. coli* also results in side reactions towards L-malate and D-alanine due to fumarase and alanine racemase activity. In order to avoid these side reactions, the first continuous L-alanine process from fumarate was done with two consecutive membrane reactors, containing purified lysates of PdAspD and EcAsp.[72] Although the side product formation could be eliminated, especially the second biocatalyst decayed rather fast due to shear stress by liberated CO_2 bubbles. Pressurized reactors show the advantage of increased reaction velocity and can contain the liberated CO_2 in solution which resulted in 50 % higher efficiency.[73] Nevertheless the purification of biocatalysts is an economical disadvantage and immobilization of the cells could further enhance the stability of the two biocatalysts. From the different immobilizing agents like polyurethane,[74] gelatin[75] and κ-carragenan, the latter was the most efficient.[76] Furthermore Chibata et al. could remove side reactivities of fumarase and alanine racemase by an acid treatment of the immobilized cells.[77] The final immobilized and treated biocatalysts demonstrated a half live of two weeks at 37 °C and reached space time yields of 13.4 g/(Lh). In order to identify the most efficient combination of the immobilized catalysts, simulations of various continuous bioreactors were conducted.[78]
The standard sequential process with two in line reactors and strictly separated aspartase/ decarboxylase reaction (both at pH 7) was set to 100% efficiency. Comparison to one pot approaches and mixed reactor set ups showed similar to decreased efficiency within the simulations. Ultimately, a jump in efficiency to 140% was observed by simulating the sequential process with an intermediate pH adjustment to account for the different pH optima of the two reactions. Apparently, the facilitated approach with pH 7 for both reactions decreases the overall reaction velocity and the efficiency for the aspartase reaction increases significantly at pH 8.5. Although the necessary change of the reaction volume due to the intermediate pH adjustment was neglected, this method was found to be most efficient and is still applied for the industrial production of L-alanine.[78]

1.3 Aldehyde dehydrogenase

As a subclass of oxidoreductases, aldehyde dehydrogenases (ALDHs) are ubiquitous in all live forms, acting as detoxification enzymes against endogenous and exogenous aldehydes.[79] They catalyze the oxidation of aldehydes to carboxylic acids and like 80% of all oxidoreductases usually require nicotinamide adenine dinucleotides $NAD(P)^+$ as cofactors.[80] These nicotinamide based cofactors have an identical chemical active functionality and differ only in a phosphorylation at the ribose moiety. Although this minimalistic modification is sufficiently distant to the reactive site of the molecule, it

enabled the evolution of a natural specificity. Hence ALDHs tend to exhibit a strong preference for one derivative and significantly reduced activity for the other. Because of this, anabolic and catabolic pathways can coexist in living cells without influencing each other. For the mammalian metabolism, ALDHs are key players and were hence divided into four groups.[81] These groups vary substantially in the presence of divalent ions, the pH-optimum and the rate-limiting step of catalysis.[82] While the release of the reduced cofactor is the rate limiting step for ALDH1, the hydride transfer is limiting for ALDH3. Non-mammalian ALDHs can be assigned to the above mentioned groups based on sequence homology. The nomenclature system for the ALDH superfamiliy suggests the root symbol ALDH for all genes followed by an arabic number denoting the family, a letter representing the given subfamily and another number signifying the individual gene within that subfamily.[83] An ALDH family is composed of sequences with more than ~40% identity and subfamilies share more than ~60% identity. Despite their varying polypeptide chain length and oligomerisation states (dimeric and tetrameric), ALDHs possess a conserved architecture of three domains (catalytic domain, $NAD(P)^+$ binding domain and the oligomerization domain (Figure 5)).[84]

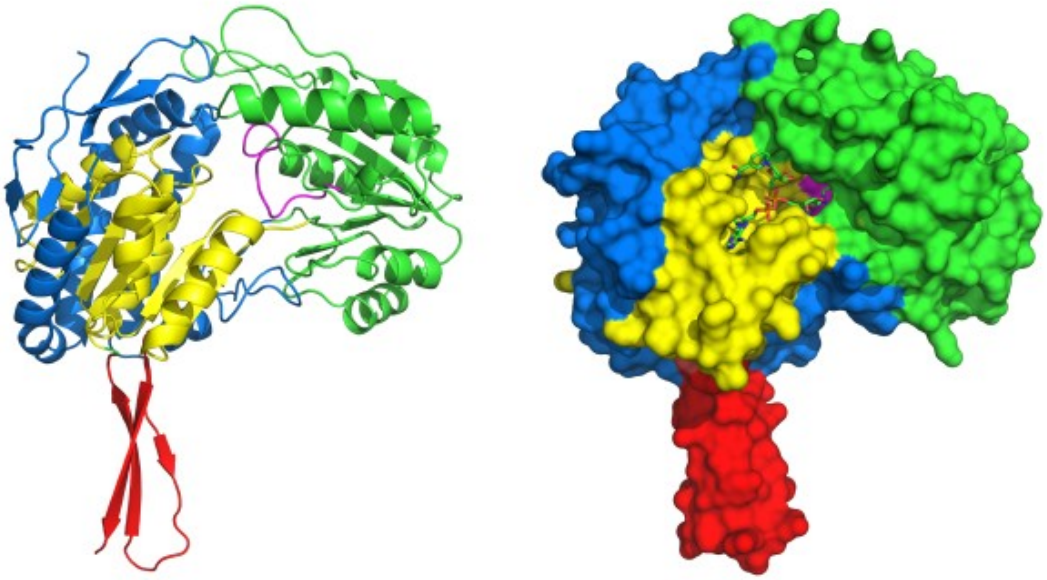


Figure 5: Crystal structure (2IMP) of the monomer of lactaldehyde dehydrogenase from E. coli. The active-site loop containing the catalytic nucleophile Cys285 is magenta. The Rossmann fold (yellow), catalytic domain (green), oligomerization domain (red), and cofactor-binding domain (blue) are indicated. Right: Surface model including bound NADH (disordered between HY and HT-conformation).[85]

The focus of this work will be on glyceraldehyde dehydrogenases especially on the variant from *Thermoplasma acidophilum* (TaALDH) since it was previously applied for the cell free ethanol production.[86] Since there is no complete crystal structure of the TaALDH enzyme, the structure of the closely related lactaldehyde dehydrogenase from *E. coli* (2IMP) (39 % sequence-identity) is shown below. The surface model on the right side in Figure 5 shows the protein-NADH complex with the disordered nicotinamide ring.[85] While the cofactor is

bound to the Rossmann fold (see section 1.3.2), the substrate (in this case lactaldehyde) reaches the nucleophilic Cystein via a hydrophobic tunnel from the opposite side of the protein.[87] The residues lining this tunnel and pointing towards the active site are presumed to determine the substrate specificity of the ALDH.[88]

1.3.1 Catalytic mechanism of ALDHs

The underlying catalytic mechanism of ALDHs (Figure 6) is family-independent. In the first step, the cofactor $NAD(P)^+$ binds to the Rossman fold, forming hydrogen bonds to highly conserved residues (e.g. Lys172 and Glu381, residue numbers according to TaALDH).[89] This leads to a structural change of the protein[90] and the proton abstraction at the sulfhydryl group of Cys281 by the essential Glu248 enables substrate binding.[91] In the second step, the carbonyl group of the substrate is attacked by the nucleophilic thiolate and a thiohemiacetal intermediate is formed. The negatively charged oxygen atom is partially stabilized by the Asn149 residue, which is essential for catalysis. From this intermediate, the hydride shift to $NAD(P)^+$ forms a thioester bond between the aldehyde and the catalytic Cys281. Finally, this bond is hydrolyzed with water (activated by a Glu248 residue) releasing the carboxylic acid. In some fermentation pathways, CoA esters are produced instead of free carboxylic acids.[92]

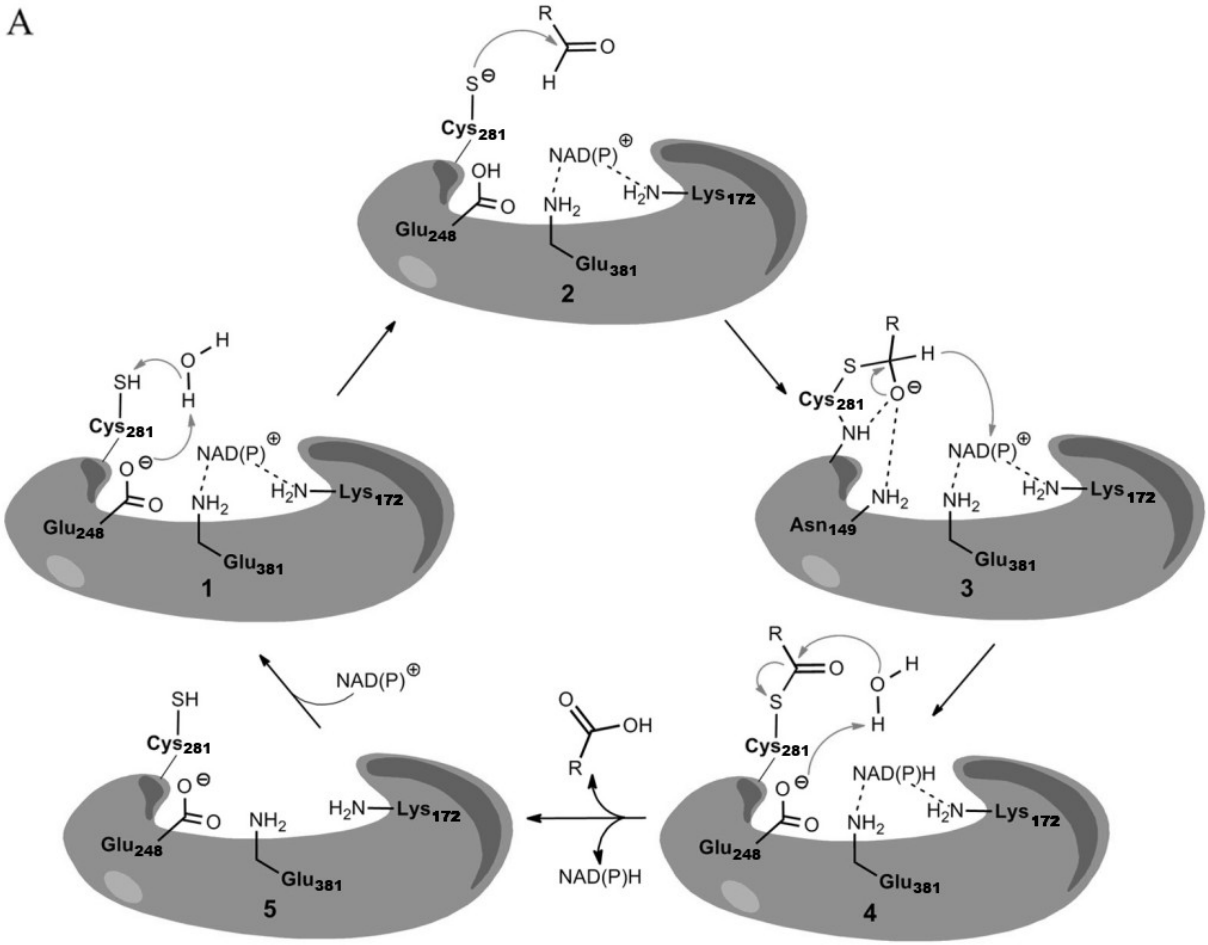


Figure 6: Catalytic mechanism of ALDHs adapted from Vasiliou et al.[82] (residue numbering according to TaALDH).

1.3.2 The Rossmann fold of ALDHs

From the comparison of multiple dehydrogenases, Rossmann et al. could identify a binding domain responsible for the dinucleotide fixation, which is since referred to as the Rossmann fold.[93] Within all originally analyzed crystal structures, the orientation of the cofactor in respect to the protein chain was identical.[94] The binding domain is usually built up from two β-α-β-α-β-units, joined in such a way, that the first strands of each unit are adjacent resulting in a strand order 654 123 (Figure 7, left).[95] $NAD(P)^+$ binds with the pyrophosphate moiety to the cleft in between the 4-1 strand.[96] Usually $NADP^+$ depending enzymes present positively charged residues in positions able to interact with the 2'-phosphate group of the adenosine ribose moiety e. g. establishing hydrogen bonds.[97] On the right site of Figure 7, the zoom on the first β-α-β-α-β-unit shows the highly conserved fingerprint sequence $G_IX_{1\text{-}2}G_{II}XXG_{III}$ typical for Rossmann folds, which is located in between β1-α1. These essential glycine residues stabilize the entire domain and enable the cofactor binding. G_I allows a tight turn of the main-chain from the β-strand into the loop, G_{II} permits close contact of the main-chain to the pyrophosphate of the nucleotide and G_{III} facilitates close packing of the helix with the β-strand.[98] Additionally, the $G_{III}XXXG/A_{IV}$ motif in the first helix strengthens the interaction between the helix and the first β-strand.[99] This stabilization is possible due to the four residue distance of the glycyl residues, aligning G_{III} and G/A_{IV} on one side of the helix and enabling hydrogen bonds to the adjacent helix.[100] For some NADP-dependent dehydrogenases, G_{III} is replaced by alanine, serine or proline to increase the distance between the helix and the strand, hence providing space for the additional phosphorylation.[101]

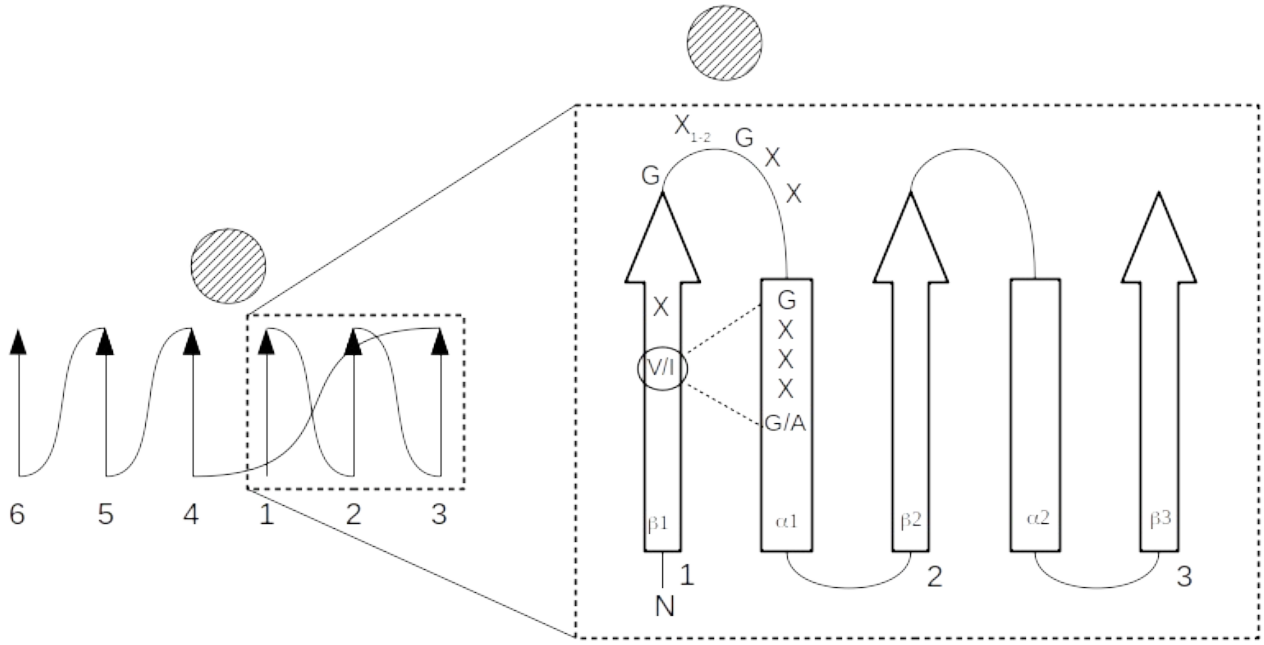


Figure 7: Left: Schematic representation of sheet structure in the NAD(P)-binding domain of dehydrogenases with indicated NAD(P) binding (circle).[96] Right: Zoom on secondary structure of the FAD or NAD(P)-binding Rossmann fold (strand 1-3). The loop connecting β-strand1 and α-helix 1 is referred to as the ligand-binding loop.[99] Rectangles represent α-helices and arrows are β-strands.

For ALDHs, a variant Rossmann fold was observed, showing unique dinucleotide binding features compared to other dehydrogenases.[102] Opposed to the classical two β-α-β-α-β-units, the ALDH Rossmann fold is composed of only five β-strands connected by four α-helices. Additionally, the pyrophosphate moiety of the dinucleotide does not appear to interact with any helix dipole or the glycine rich fingerprint motif, as it is usually found in classical Rossmann fold binding. Instead, the NAD-binding seems to be stabilized by a β4-αD-β2 motif, forming hydrogen bonds between the adenine ribose and an essential glutamic acid (located in β2), as well as other hydrophobic interactions between the adenine ring and the surrounding hydrophobic residues (such as valine or leucine).[102] One consequence of this special binding mode is, that the ALDH enters less contacts with the cofactor upon binding, allowing multiple conformations.[103]
While the identification of $NAD(P)^+$ binding sites was previously done via homolog analysis, recent methods use machine learning algorithms in order to predict cofactor binding sites for uncharacterized proteins based on their amino acid sequence.[104]

1.3.3 Modifying the cofactor specificity of dehydrogenases

As described above, $NAD(P)^+$ utilizing enzymes normally show a strong preference towards one cofactor, with significantly reduced activity for the other. Although this specificity enables the separation of opposed natural pathways within living systems, it can be a major drawback for the design of artificial cascades. Concerning metabolic engineering a variety of examples can be found, where an optimization of the cofactor specificity could drastically improve the efficiency of the biotechnological process.[105] For instance, Arnold et al. could increase the anaerobic fermentation yield of 2-methylpropan-1-ol up to the theoretical maximum after engineering an alcohol dehydrogenase and the keol-acid reductoisomerase in *E. coli.*[106] Commonly, the change of the cofactor specificity is more attractive towards NAD^+, since the phosphorylated derivative is much more expensive. The first report of such a cofactor switch date back to 1990, when Scrutton *et al.* reported the redesign of the coenzyme specificity of an *E. coli* gluthathione reductase (GSR) by protein engineering.[107] Due to the absence of a valid crystal structure for the GSR of *E. coli* (EcGSR), human gluthathione reductase (GSR) was used as a structural model.[108] The major focus of this investigation was the stabilization of the NADPH 2'-phosphate group within the Rossmann fold. A comparison of the crystal structures of the NADPH-dependent human GSR and the NADH-dependent human dihydrolipoamide dehydrogenase (LPD), gave some promising leads. Positively charged Arg218, His219 Dihydroxyand Arg224 (numbering for human GSR) were hypothesized to be responsible for the stabilization of the additional negative charge of the 2'-phosphate group, hence confering NADPH specificity. Amino acid exchanges R218M or R224L of the EcGSR led to a large rise (~25 fold) in K_M for NADPH and a modest fall of the K_M for NADH (~4 fold). This loss of NADP-specificity after exchange of the positively charged groups (here arginine) next to the 2'-phosphate was also found for other examples.[109] Additionally, Ala179 and Ala183 (EcGSR numbering) within the first helix of the Rossmann fold were presumed to open it up in order to incorporate the larger NADPH. These two residues are located on the same side of the helix and their methyl groups occupy more space than the corresponding glycine residues of the NAD^+ binding fold. By exchanging A179 to G, the

specificity of EcGSR for NADH increased ~40 fold compared to the wildtyp while k_{cat} as well as the kinetic parameters for NADPH remained unchanged. The A183G exchange had only minor effects on the NADH kinetics but increased k_{cat} and K_M for the NADPH. In sum, the alteration of all identified residues yielded a 70 times increased k_{cat} to K_M ratio for the EcGSR with NADH.[107] This initial success could demonstrate the power of rational design and laid the foundation of cofactor specificity redesign. Key learnings from this initial study include the importance of the stabilization of the additional 2'-phosphate of NADP(H) and the necessary decrease in size for the smaller NAD(H). Nevertheless, homology based methods are still limited to structural related sequences and cannot provide a general procedure.[110]

Systematic approaches for the cofactor switch, cannot be based on static methods like the comparison of crystal structures, since the dynamic cofactor binding involves conformational changes.[111,112] This protein motion during the binding process is not limited to the Rossmann fold, hence mutations in any domain (binding, catalytic or bridging) can alter this folding process and therefore affect the reaction kinetics[113] or the substrate specificity[114] Furthermore, Rossman folds show a huge diversity and although some structural motifs are conserved, NAD(P)H binding is controlled through various residues conferring specificity for one of the dinucleotides.[115,116] Successful alteration of the cofactor specificity normally requires multiple simultaneous mutations which are non-additive in their effect.[117,118] This results in a large combinatorial space of mutations, which needs to be analyzed systematically in order to identify the selectivity determining key mutations.

One of the first systematic procedures for the modification of cofactor specificity was based on the iterative protein redesign and optimization alogirthm (IPRO).[119] Its workflow includes the stepwise approximation of the detailed cofactor binding energy for every mutation *in silico.* Requiring intensive computational power, this approach was successfully established for the redesign of *Candida boidinii* xylose reductase (CbXR) with the non-native cofactor NADH.[120] Out of ten predicted variants, seven had significant activity with NADH, and the best variant showed 27 fold increased activity. Although the CbXR binds the cofactor via a Tim-Barrel structure, the method could be in theory transferable to Rossmann fold dehydrogenases.

The next step towards a systematic cofactor switch was done by Brinkmann-Chen *et al.* focusing on NADPH-dependent ketol-acid reductoisomerases (KARIs).[121] Although KARIs are usually NADPH-dependent, a purely NADH-dependent variant was developed via enzyme engineering.[106] In detail structural analysis revealed, that three out of four mutations necessary for this specificity reversal were located in the loop connecting the β2 sheet with the αB helix of the Rossmannn fold (termed: $\beta 2\alpha$B loop). Residues within this loop showed direct contact to the 2'-phosphate moiety. From an alignment with 643 reviewed KARI structures, three different $\beta 2\alpha$B-loop lengths could be identified containing either six (14%), seven (68%) or 12 (18%) residues. Moreover, subalignments with the same loop length showed conserved patterns, indicating an important function.[122] Their developed workflow suggests mutations based on the identified loop type, which is followed by random mutagenesis in order to restore the overall activity. Following this protocol, they could increase the k_{cat}/K_M ratio for NADH by a factor of ~2.5 while NADPH is barely accepted anymore.[121] While this approach is limited to KARIs, the key learning from

this work is, that the stabilization of the 2'-phosphate and the available space within the cofactor binding pocket have a major impact on the dinucleotide selectivity.
These findings were also considered by Cui et al, who implemented a computational approach based on molecular dynamics simulations for the reversal of cofactor specificity.[123] For the NADH-dependent Gox2181, the strength of the hydrogen bonds between cofactor and enzyme were calculated from MD simulations and compared to *in silico* designed variants. Since NADPH should be accommodated inside the binding pocket instead of NADH, exchanges towards positively charged residues were preferred. Although the final mutant demonstrated high specificity for the non-natural NADPH, extensive computational power is necessary for the MD simulations, reducing the throughput of mutants and rendering this approach rather unattractive for general applications.
The most recent, general tool for the engineering of the NAD(P) cofactor preference of oxidoreductases termed: "Cofactor Specificity Reversal - Structural Analysis and Library Design (CSR-Salad)" was published by Cahn et al.[124] This semirational strategy comprises three steps: structural analysis, cofactor switch and recovery of the catalytic activity. Using a crystal structure or homology model of the desired protein with bound cofactor as input, the automated online-tool suggests mutagenesis targets, which are normally in the immediate vicinity of the 2'-motif of the NAD(P). Additional residues for activity recovery as well as residues that can interact with the cofactor after mutagenesis are also suggested. Furthermore the mutations are classified according to a modified system of Carugo and Argos,[116] accounting for the residues role within the binding pocket (Figure 8).

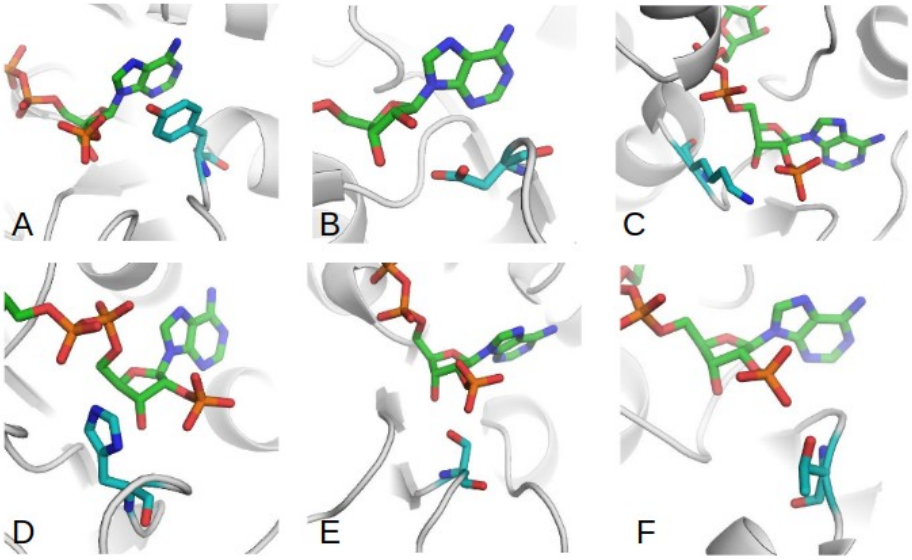


Figure 8: Classification of residues important for cofactor specificity used by CSR-SALAD shown with either NADH or NADPH. A) Face, B) Edge, C) Motif D) Pyrophosphate, E) Bidentate and F) Other. Adapted from Cahn et al.[124]

Face residues (A) make contact with the plane of the adenine moiety and Edge residues (B) lie along the edge of the adenine residue, continuing parallel towards the ribose moiety. Motif residues (C) are part of the highly conserved $G_IX_{1-2}G_{II}XXG_{III}$ motif common for Rossmann folds and Pyrophosphate residues (D) make contact to the O2' moiety as well

as the bridging phyrophosphate motif of NAD(P). Bidentate residues (E) make contact with the O2' hydroxyl or phosphate and the O3' hydroxyl. Finally the other residues (F) do not fit into any other group. The selection of potential phosphate binding residues is then done by a distance based approach, focusing on residues within 4.2 Å of the 2'-phosphate. Additionally only residues pointing towards the cofactor (applying the concept of the side-chain pseudocenter)[125] are considered for the mutagenesis. After the classification, amino acids for cofactor stabilization are selected and translated into degenerated codons. Within the initial report, the method was validated experimentally, reversing the cofactor specificity of four structurally diverse NADP-dependent enzymes: glyoxylate reductase, cinnamyl alcohol dehydrogenase, xylose reductase, and iron-containing alcohol dehydrogenase.[124] The broad applicability as well as the user-friendly online platform which provides a comprehensive report render the CSR-SALAD a valuable tool for the engineering of dinucleotide cofactor preference.

1.3.4 Molecular dynamics simulations of ALDHs

Molecular dynamics (MD) simulations allow for atomic-level characterization of biomolecular processes such as the conformational transitions associated with protein function.[126] The internal molecular motion of enzymes is calculated stepwise in order to track e.g. binding or folding processes. Multiple force fields such as CHARM,[127] AMBER[128] and GROMOS[129] have been developed for the accurate treatment of internal forces. Nevertheless those classical potentials often fail to faithfully capture key quantum effects and currently machine-learning algorithms are deployed in order to develop improved force fields to achieve spectroscopic accuracy in molecular simulations.[130] For the set up of an MD simulation of a protein-ligand complex, the system needs to be parameterized against the desired force field and solvated with an explicit solvent (e.g. TIP3P[131] water). By the addition of ions, the system is neutralized before energy minimization algorithms (e.g. steepest descent) relax it. A temperature and pressure equilibration is necessary before the production run can be performed. The analysis of the resulting trajectories usually includes root-mean-square difference (RMSD) and root-mean square fluctuation (RMSF) in relation to the input structure. While the RMSD represents the time-dependent movement of the entire system, the RMSF shows the average movement of each atom (or residue). Additionally, a principal component analysis can reveal the essential modes relevant for the protein motion.[132] Furthermore time-dependent distances, angles and charges can be calculated in order to analyze e.g. substrate binding.

The first MD simulations for ALDHs (class 3) have been published by Wymore et al. with the focus on the positioning of the NAD^+ cofactor.[133] Although the 200 ps simulation time was too short to adequately sample all motions of the nicotinamide ring, they postulated a flexible active center were the nicotinamide moiety moves quickly in and out. From crystal structures and NMR experiments it is known, that the nicotinamide and phosphate groups of NAD^+ exist in two different conformations: The active "hydride transfer" form (HT), where the nicotinamid is in the vicinity of the nucleophilic Cys in order to receive the hydride from the thiohemiacetal intermediate and the "hydrolysis" form (HY), where the nicotinamid is further away from the acitve site (Figure 5, right).[103] Although it was assumed that the HT position is more favorable for thiolate stabilization,[134] an MD analysis

suggested the HY position for substrate binding.[135] This finding is supported by the formation of dead-end products from the reaction of thiolates and nicotinamides,[136] which exhibit a broad absorption maximum around 319 nm.[137] Instead of a stabilization of the HT-positioned NAD, the MD analysis predicted a proton transfer from the main chain amide in order to avoid the breakdown of the oxyanion intermediate.[135] Furthermore it was shown for pyrroline-carboxy dehydrogenases, that the activation of the nucleophilic Cys is repositioning the substrate, finally enabling catalysis.[138] Additional studies on the substrate binding predicted a preference for the formation of the R-thiohemiacetal which would also facilitate the hydride transfer due to a closer contact of the hydrid to the NAD^+ cofactor.[139] Another set of simulations on the betaine ALDH from rice found that the NAD-binding domain is the most flexible within the protein and a principal component analysis revealed a scissor like motion of the protein.[140] Most recently Talfournier et al. could fill the knowledge gap of how the dinucleotide cofacor can flip from the HY to the HT position while bound to the flexible Rossmann fold.[141] Combining FRET-based kinetics in order to determine the cofactor conformation with targeted MD simulations, they could identify a conserved ELGG active site loop which acts as a gatekeeper for the cofactor flip. While the pre- and post-flip backbone structure exhibit a similar conformation,[142] the flip relies on the flexibility of the identified ELGG loop. Although various details of the ALDH mechanism could be explained by MD simulations, further questions like what is triggering the cofactor flip or is this currently accepted mechanism valid for all ALDHs remain to be answered in future experiments.

In sum, MD approaches were successfully deployed for the elucidation of the ALDH mechanism. The information gathered from simulations concluded the analysis of crystal structures but additionally gave insights into the protein motion revealing essential changes of the protein structure during catalysis.

1.4 Objectives of this work

The aim of this work is the development and optimization of a cell free enzymatic cascade for the conversion of D-glucose to L-alanine in order to have a sustainable alternative for the current, fossil resources based approach. To realize such an enzymatic cascade, the first step is to set target conditions under which the reaction would take place. These conditions should enable an efficient turnover and simultaneously allow an economic process. In a next step, suitable biocatalysts need to be identified and expressed in a dedicated host strain. To keep the production of the biocatalysts rather simple, the focus will be on enzymes, which could already be heterologously expressed. After characterizing different enzymes under cascade conditions, the most active variants should be applied for the cell free cascade. Next to activity, the specificity for the desired substrate (and cofactor) as well as sufficient temperature and storage stability are key factors for successful candidates. The combination of the identified biocatalysts to a cell free enzymatic cascade demands a thorough process optimization in order to obtain the maximum production rate with a minimum amount of biocatalyst. For a direct relation of the process parameters to the production rate, a suitable assay needs to be established.

This mainly economically driven optimization can enable a cost efficient alternative for a future amine production.
While wild type enzymes or characterized mutants are tested for cascade use in the first step, crucial enzymes, which match the cascade requirements only partly, are going to be engineered in the second step. The method of choice here is the rational protein design, which is done based on crystal structures, homology models and docking studies. For a comprehensive understanding of beneficial mutations, molecular dynamics analysis of initial and final variant are conducted and analyzed under identical conditions in order to identify design principals for future engineering approaches.

2 Materials and Methods

2.1 Materials

The following materials were used throughout this thesis for conducting experiments. All materials were used as received, unless stated otherwise.

2.1.1 Instruments

Table 2: Overview of used equipment

Equipment	Manufacturer / Model
Autoclave	Thermo Scientific (Ulm) Varioklav 135 S
Centrifuge	Thermo Scientific (Ulm) Sorvall RC-6 Plus
Clean bench	Thermo Scientific (Ulm) MSC-Advantage
Climate cabinet	Binder (Tuttlingen) KBF 240 E5.1/C
Drying oven	Thermo Scientific (Ulm) Function Line T12
Micro scale	Ohaus Europe (Nänikon) Pioneer™
Freezer -20 °C	Liebherr-Hausgeräte (Ochsenhausen)
Freezer -80 °C	Thermo Scientific (Ulm) Forma 906 -86 °C ULT
Liquid handling station	BRAND (Wertheim) LHS
Liquid handling system	Tecan (Männedorf) Freedom EVO
Incubator StoreX IC	LiCONiC Services GmbH (Montabauer)
Multi-channel-arm (MCA96)	Tecan (Männedorf)
Robotic manipulator (RM)	Tecan (Männedorf)
Spectrophotometer	Tecan (Männedorf) Infinite 200 pro
Magnetic stirrer	Thermo Scientific (Ulm) Variomag Telesystem
Microplate shaker	Edmund Bühler (Hechingen) TiMix 5 control and TH15
Multichannel pipette	Eppendorf AG (Hambug) Research pro 8x 1200 µL
pH-meter and electrode	Mettler-Toledo (Giessen) Five Easy™ and InLab® Expert Pro
Pipettes	Brand (Wertheim) Transferpetten
Rotor	Thermo Scientific (Ulm) SS-34,SH-3000 and F9-4x1000y
Scale	Satorius (Göttingen) TE1502S and TE6101
Shaker	Thermo Scientific (Ulm) MaxQ 2000
Spectrophotometer	Thermo Scientific (Ulm) Varioscan Flash and Multiscan BioTek GmbH (Bad Friedrichshall) Epoch2
Table centrifuge	Thermo Scientific (Ulm) Heraeus Fresco 21
UHPLC	Dionex(Idstein) Ultimate 3000RS

Degasser	Dionex(Idstein) SRD 3400
Pump module	Dionex(Idstein) HPG 3400RS
Autosampler	Dionex(Idstein) WPS 3000TRS
Column compartment	Dionex(Idstein) TCC 3000RS
Diode array detector	Dionex(Idstein) DAD 3000RS
Fluorescence detector	Dionex(Idstein) FLD 3100
Column	Metrohm GmbH (Filderstadt) Metrosep A Supp 10-250/4.0
Column	YMC Europe GmbH Triart C18 12nm, S-1,9µm, 100 x 2,0 mm
Ultrapure water system	ELGA LabWater (Celle) PURELAB classic
Vortexer	Scientific Industries (Bohemia) Vortex Genie 2
Water bath	Huber (Offenburg) CC1
FPLC	GE Healthcare (Freiburg) ÄKTA™ purifier
Pump module	GE Healthcare (Freiburg) P-900
Sample pump module	GE Healthcare (Freiburg) UP-960
Control unit	GE Healthcare (Freiburg) UPC-900
column	GE Healthcare (Freiburg) HisTrap FastFlow 5 ml
column	GE Healthcare (Freiburg) HiPrep 26/10 Desalting
Agarose-electrophoresis	Bio-Rad GmbH (München) Mini-Sub Cell GT System
Dispenser	BioTek GmbH (Bad Friedrichshall) MicroFlow select
Gel documentation-system	Intas Imaging Instruments GmbH, (Göttingen) Gel iX Imager
Colony picker	Hudson Robotics Inc. (Springfield NJ, USA) Hudson Rapid Pick Lite
Thermo cycler	Bio-Rad GmbH (München) Mini-Sub Cell GT System MJ Mini™ and MyCycler™
Ultrasonic unit	Hielscher Ultrasonics GmbH (Teltow) UIS250L and VialTweeter sonotrode
Microwave	ECG (Prag, Czech republic) MH 25 ED
Heating block	Analytik Jena AG (Jena) Thermostat Tmix

2.1.2 Software

Table 3: List of softwares

Software	**Purpose**
ChemDraw 12.0	Drawing chemical structures and schemes
Microsoft Excel (2017)	Calculations
Microsoft Powerpoint (2017)	Preparing schemes
Microsoft Word (2017)	Preparing manuscripts for publication
Endnote X8	Manage citations

Gromacs (2019)	Molecular dynamics simulation
Yasara 15.11.18	Protein visualization and docking
Avogadro 1.2.0	Protein visualization
Pymol 1.8.4.0	Protein visualization
Chromeleon 6.8	Control LC station
Clonemanager 9	Manage dna sequences
Bioedit	Analyze sequencing results
Matlab R2016a	Scripting for automatization
Brand Liquid Handling	Control liquid handling station
Sigmaplot 13	Prepare plots and fits

2.1.3 Special consumables

Table 4: List of special consumables

Name of the Material	Manufacturer	Catalog number
96 well deep well plate 2.0 mL (DWP)	Greiner Bio-One	780271
96-well micro titer plate F-Bottom (MTP)	Greiner Bio-One	655101
96-well silicon cap mat	Whatmann	7704-0105
Aluminum sealing film	Axygen	PCR-AS-200
Breathable sealing film	Axygen	BF-400-S
HisGraviTrap	GE Healthcare Europe GmbH	11003399
PD-10 column	GE Healthcare Europe GmbH	17085101
96-well micro titer plate F-Bottom Nunc black	Thermo Scientific	237107
384-well micro titer plate F-Bottom	Greiner Bio-One	781101

2.1.4 Media and Buffer

Heat-stable cultivation media, buffers, and solutions were autoclaved for 20 min at 121 °C and 2 bar. Heat-unstable solutions were filtered through a sterile syringe filter holder (0.2 µm, VWR International GmbH). For cultivation media consisting of various solutions, all solutions were sterilized before mixing. Solid media were obtained by adding 1.5 % (w/v) agar-agar before autoclaving. For selective medium, sterilized antibiotics were added after the medium was cooled down. Purified water, ddH_2O, (PURELAB Classic) was used for all media, buffers, and solutions. All solutions were stored at room temperature, if not otherwise specified. The final concentrations of the anitbiotics were: Kanamycin 50 µg/mL, chloramphenicol 30 µg/mL and carbenizilin 100 µg/mL.

Luria-Berani (LB)-media[143]

0.5 % (w/v) yeast extract, 1.0 % (w/v) tryptone, 1 % (w/v) NaCl

Autoinduction medium[144]

ZY-solution

0.5 % (w/v) yeast extract, 1.0 % (w/v) tryptone

5052-solution (50x)

25 % (v/v) glycerol, 2.5 % (w/v) D-glucose monohydrat, 10 % (w/v) α-lactose

NPS-solution (20x)

1 M Na_2HPO_4, 1 M KH_2PO_4, 500 mM $(NH_4)_2SO_4$

ZYP-5052 autoinduction media

928 mL ZY, 50 mL 20xNPS, 20 mL 50x5052, 2 mL 1M $MgSO_4$

Lysis buffer

1 g/L lysozyme and 10 mg/L DnaseI in 100 mM HEPES pH 7.35

2.1.5 Enzymes and reagents

All substances were used as received.

Table 5: Overview of the used enzymes and reagents

Name of Enzyme or reagent	Manufacturer	Catalog number
α-D-Glucose (monohydrate)	Carl Roth GmbH	6887.5
Sodium D-gluconate	Sigma-Aldrich	S2054
Sodium pyruvate	Carl Roth GmbH	8793.1
D-Glyceraldehyde	Sigma-Aldrich	49800
DL-Glyceric acid solution 20 % in H_2O	TCI Deutschland GmbH	D0602
L-Alanine	Sigma-Aldrich	A7627
NAD (free acid)	Carl Roth GmbH	AE11.3
NADH (disodium salt)	Carl Roth GmbH	AE12.3
NADP (disodium salt)	Carl Roth GmbH	AE13.3
Isobutanol	Sigma-Aldrich	33064
HEPES	Carl Roth GmbH	HN78.3
Ammonium sulfate	Applichem	A1032,5000
Fluorescamine	Alfa Aesar	43749.MB
Acetonitrile	VWR	83640320
Sodium hydroxide	Carl Roth GmbH	6771.2
ammonium nitrate	Carl Roth GmbH	K299
ammonium chloride	Carl Roth GmbH	K298.1
diammonium phosphate	Carl Roth GmbH	0268.2
Roti®-Nanoquant	Carl Roth GmbH	K015.3
Bovine serum albumin	Sigma-Aldrich	A2153
Imidazol	Merck	1047160250

Manganese chloride dihydrate	Merck	1.0593410100
sodium borate	Sigma-Aldrich	71996
Phusion polymerase	New England Biolabs	M0530
Phusion high fidelity buffer	New England Biolabs	M0530
Dimethyl sulfoxide	Carl Roth GmbH	AE02.1
dNTP mix	VWR	733-1363
DpnI	New England Biolabs	R0176
Glycerol	Carl Roth GmbH	3783.2
Lysozyme	Carl Roth GmbH	8259.3
DNaseI	Applichem	A3778
SYPRO Orange	Sigma-Aldrich	S5692
Yeast extract	Carl Roth GmbH	2363
Tryptone	Carl Roth GmbH	8952.5
Sodium chloride	Carl Roth GmbH	P029.3
α-Lactose	Carl Roth GmbH	6868.1
Magnesiumsulfate	Carl Roth GmbH	8283.2
Disodiumphosphate	Carl Roth GmbH	P030.3
Monopotassiumphosphate	Carl Roth GmbH	3904.3

KDG was synthesized as described by Carsten *et al*.[145]

2.1.6 Kits

Table 6: Ready to use Kits applied throughout this work following manufactures recommendations

Kit	Manufacturer
GeneJET™ Plasmid Miniprep Kit	Thermo Scientific, Fermentas, St. Leon-Rot
NucleoSpin® Gel and PCR Clean-up Kit	Macherey-Nagel GmbH & Co. KG, Düren
CloneJET PCR Cloning Kit	Thermo Scientific, Fermentas, St. Leon-Rot

2.1.7 Bacteria strains

Table 7: Bacterial strains used within this work

Strain	Genotype	Reference
E. coli BL21(DE3)	F ompT gal dcm hsdS B (r B m B) λ(DE3)	Novagen (Darmstadt)
E. coli DH5α	fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Invitrogen (Carlsbad, USA)
BL21-BsGDH	*E. coli* BL21(DE3) with pACYC-Duet-BsGDH encoding	This work

	Glucose dehydrogenase (E170K/Q252L) from *Bacillus subtilis*	
BL21-CcDHAD	*E. coli* BL21(DE3) with pET28a-CcDHAD encoding Dihydroxyacid dehydratase from *Caulobacter crescentus*	This work
BL21-PtKDGA	*E. coli* BL21(DE3) with pET28a-PtKDGA encoding 2-keto-3-deoxygluconate aldolase from *Picrophilus torridus*	This work
BL21-AfAlaDH	*E. coli* BL21(DE3) with pET24a-AfAlaDH encoding alanine dehydrogenase from *Archaeoglobus fulgidus*	This work
BL21-MjALDH	*E. coli* BL21(DE3) with pET28a-MjALDH encoding aldehyde dehydrogenase from *Methanocaldococcus jannaschii*	This work
BL21-SsDHAD	*E. coli* BL21(DE3) with pCBR-SsDHAD encoding dihydroxyacid dehydratase from *Sulfolobus solfataricus*	Carsten *et al.*[145]
BL21-TaALDH-M0	*E. coli* BL21(DE3) with pCBR-TaALDH encoding aldehyde dehydrogenase (F34M-Y399C-S405N) from *Thermoplasma acidophilum*	Steffler *et al.*[146]

2.2 Methods

2.2.1 Shaking flask expression

Enzyme expression was performed in E. coli BL21(DE3) as host strain in shaking flask cultures using ZYP-5052 autoinduction media, supplemented with 100 µg/mL kanamycin or 10 µg/mL chloramphenicol. After 1 %-inoculation of the autoinduction media with an overnight grown culture of E. coli BL21(DE3) harboring the corresponding plasmid, expressions were carried out over night at 30°C and 120 rpm.

2.2.2 Protein purification

The detailed protein purification method is described by Gmelch et al [147] (see this work page 35). Briefly, cells were harvested, resuspended in 100 mM HEPES buffer pH 7.35 and disrupted by ultrasonication in an ice cooled water bath. Lysates were heat treated and after clarifying via centrifugation samples were either purified via a His GraviTrap column (GE Healthcare) equilibrated with 100 mM HEPES pH 7.35 containing 10 mM imidazole following the manufacturer's recommendations or directly desalted using a PD-10 column (GE Healthcare) equilibrated with 100 mM HEPES pH 7.35. Purified proteins were flash frozen in liquid nitrogen, stored at -80 °C and thawed freshly every day. SsDHAD purification was performed as described previously.
Protein concentration was measured by a Bradford protein assay using the Roti-Nanoquant reagent (Carl Roth GmbH) according to the manufacturer's recommendations with bovine serum albumin as standard.

2.2.3 Kinetic characterization of Proteins

Kinetic experiments contained 100 mM HEPES buffer pH 7.35, 100 mM ammonium sulfate pH 7 (both titrated with NaOH) and were preheated for 10 min at 50 °C before the reaction was started with the addition of the appropriate amount of enzyme diluted in 100 mM HEPES pH 7.35 containing 0.2 mg/mL BSA.
All enzymes were analyzed photometrically following NADH consumption or production at 340 nm (ε_{NADH} = 6.22 $mM^{-1}cm^{-1}$) with an Epoch 2 Microplate Spectrophotometer (BioTek

GmbH) at 50 °C. If not mentioned otherwise, reactions were carried out in triplicates using 96 well plates (Greiner f-bottom) containing a final volume of 200 µL. Michaelis-Menten constants were determined for all enzymes by measuring eight different substrate concentrations, plotting initial velocities vs. substrate concentration and fitting the parameters with the online tool ic50.tk. One unit (1 U) is defined as the amount of enzyme that consumes or releases 1 µmol of NADH per minute.

2.2.4 Fluorescence based L-alanine quantification

Alanine assays were performed in a 96 f-bottom well plate (NUNC black) using a Varioscan plate reader (Thermo Fisher) according to a modified protocol from Bantan-Polak *et al*.[148] 10 µL of diluted sample were mixed with 45 µL of a 0.1 % fluorescamine solution in dry acetonitrile. After the addition of 8 µL 100 mM sodium borate buffer pH 10 and 95 µL of water, the fluorescence was measured at λ_{em} = 486 nm (λ_{ex} = 396 nm). Signal calibration was done using L-alanine concentrations ranging from 1-30 mM and the complex matrix of the reactions.

2.2.5 Construction of sequence saturation libraries

Site-directed mutagenesis was performed in a C1000 Thermo Cycler (BIORAD) with Phusion polymerase from New England Biolabs (NEB) with the supplied high fidelity buffer in a final volume of 50 µL. Oligonucleotides were obtained desalted from Eurofins Genomics (Germany)

A 2 min initial denaturation at 98 °C was followed by 26 cycles of 10 sec at 98 °C, 30 sec at 60 °C and 3.5 min at 72 °C. Final elongation was done for 10 min at 72 °C before 1 µL of 20 U/L DpnI (NEB) was added. Digestion of template dna was done at 37 °C for at least 12 h followed by heat inactivation at 80 °C for 20 min and PCR cleanup with a Macherey Nagel PCR cleanup kit according to manufactures recommendations. Chemical competent *E. coli* BL21 (DE3) were transformed with the according libraries and spread onto LB agar plates containing 50 µg/mL kanamycin.

2.2.6 HT-MTP oxidoreductase expression

All cultivation media were supplemented with 100 µg/ml kanamycin. Colonies of *E. coli* BL21 (DE3) containing the TaALDH variants were transferred into f-bottom microtiter plates (Greiner) with a RapidPick Lite colony picker (Hudson Robotics, Springfield, NJ, USA) containing 150 µL LB media. After an overnight incubation at 37 °C and 1200 rpm, V-bottom deepwell plates (Greiner) containing 600 µL ZYP-5052 autoinduction media were inoculated with 1.7 vol% of preculture and shaken over night at 30 °C and 1200 rpm. Backup plates were prepared by mixing 30 µL expression culture with 30 µL 50 % glycerol and freezing at -80 °C. The expression culture was then harvested by centrifugation at 4200 × g for 15 min and the supernatant was removed. The cell pellets were frozen at -20 °C before 200 µL lysis buffer (1 g/L lysozyme and 10 mg/L DNaseI in 100 mM HEPES pH 7.35) were added followed by an incubation at 37 °C for 60 min and an additional heat step at 60 °C for 60 min. Insoluble cell debris was removed by centrifugation at 4200 × g for 15 min and the supernatant containing the soluble variant enzymes of TaAlDH were tested for activity as described below. The two most active variants per position showing at

least 20 % increased activity were regrown from backup plates, plasmid purified and sent for sequencing (Eurofins Genomics, Germany). For further characterization with 1 mM NAD, 5 mM NAD, 3 % Isobutanol or 150 µM NADP, ten colonies harboring the corresponding mutant plasmid were randomly picked and treated as described above. The heat purified enzyme was diluted three or five times for the activity assays.

2.2.7 Screening for improved oxidoreductase variants

Experiments were carried at room temperature using 96/384 well plates (Greiner f-bottom) containing a final volume of 200/60 µL. All reactions contained 100 mM HEPES buffer pH 7.35, 100 mM ammonium sulfate pH 7 (both titrated with NaOH) and were started with the addition of the appropriate amount of enzyme. Reaction velocity was monitored photometrically following NADH release at 340 nm with an Epoch 2 Microplate Spectrophotometer (BioTek GmbH). The mutant screening assay solution contained 1.3 mM D-glyceraldehyde and 1 mM NAD^+ and the additional assays for the cofactor acceptance contained either 1 mM or 5 mM NAD^+ or 0.15 mM NADP. Reactions with 3 vol % Isobutanol contained 1 mM NAD. One unit (1 U) is defined as the amount of enzyme activity producing 1 µmol of substrate per minute.

2.2.8 Thermofluor-Assay

The melting point of the proteins was determined with a SYPRO Orange Protein Gel Stain (Sigma-Aldrich) thermofluor assay in a CFX 96 Real-Time PCR thermocycler (Biorad) containing 25 µL total volume according to a protocol of Meijers *et al*.[149]
Briefly, the 5000x concentrated SYPRO Orange solution is prediluted with water to a 62x working solution. Then the sample is directly prepared in the 96 well plate according to the following mixing protocol: 18 µL 100 mM HEPES buffer pH 7.35, 2 µL protein stock solution and 2 µL of the SYPRO Orange working solution. After sealing the plate with an optical-clear sealing foil it is transferred to the thermocycler, where the plate is heated from 20 °C to 95 °C at one degree per minute. At every temperature a fluorescence readout at 485/20 nm (E_x) and 530/30 nm (E_m) is conducted.

3 Results

3.1 Optimization of a reduced enzymatic reaction cascade for the production of L-alanine

The following publication shows the development of a minimized, enzymatic reaction cascade for the direct production of L-alanine from D-glucose and ammonium. The artificial pathway design eliminates the need for phosphorylation and only requires NAD^+ as cofactor. An efficient, activity based enzyme selection is demonstrated and the resulting redox neutral cascade is composed of only six biocatalysts. Next to the specific activity, thermostability, specificity for the desired substrate and activity under saturated L-alanine conditions were taken into account for the selection process. Proper kinetic characterization as well as an analysis of the influence of the ammonium source on the selected enzymes were done in order to define a starting point for the optimization.
For the following process optimization, a fluorescence based assay was developed in order to directly monitor the produced L-alanine within the complex reaction matrix. From a variety of tested methods, the derivatization with fluorescamine was the most accurate and reliable assay. Due to the excellent sensitivity, the assay could be transferred to MTP scale requiring only 10 µL of diluted sample. The process time for 96 samples could be reduced to less than 10 min.
A stepwise enzyme titration approach was performed for all biocatalysts resulting in an ideal enzyme ratio for the maximum production rate. The process parameters such as cofactor concentration, buffer concentration and ammonium concentration were optimized accordingly. In the end a comparison of the optimized to the initial conditions revealed a significant reduction of the total enzyme load for a similar production rate and a yield > 95 %. Finally, critical steps and future experiments concerning the dehydratase reactions were discussed.

Tobias Gmelch designed the experiments, selected the biocatalysts and conditions and did all the experiments, measurements, calculations and analysis and wrote the manuscript. Josef Sperl supported in designing the experiments and proofread the manuscript. Tobias Gmelch designed the research project and experimental approach guided by Volker Sieber, who also proofread the manuscript.

Supporting Information to this manuscript can be found in the appendix.

Optimization of a reduced enzymatic reaction cascade for the production of L-alanine

Tobias J. Gmelch, Josef M. Sperl, Volker Sieber

Scientific Reports
2019

Open Source
DOI: 10.1038/s41598-019-48151-y

SCIENTIFIC REPORTS
nature research

OPEN

Optimization of a reduced enzymatic reaction cascade for the production of L-alanine

Received: 6 March 2019
Accepted: 25 July 2019
Published online: 13 August 2019

Tobias J. Gmelch[1], Josef M. Sperl[1] & Volker Sieber[1,2,3,4]

Cell-free enzymatic reaction cascades combine the advantages of well-established *in vitro* biocatalysis with the power of multi-step *in vivo* pathways. The absence of a regulatory cell environment enables direct process control including methods for facile bottleneck identification and process optimization. Within this work, we developed a reduced, enzymatic reaction cascade for the direct production of L-alanine from D-glucose and ammonium sulfate. An efficient, activity based enzyme selection is demonstrated for the two branches of the cascade. The resulting redox neutral cascade is composed of a glucose dehydrogenase, two dihydroxyacid dehydratases, a keto-deoxy-aldolase, an aldehyde dehydrogenase and an L-alanine dehydrogenase. This artificial combination of purified biocatalysts eliminates the need for phosphorylation and only requires NAD as cofactor. We provide insight into in detail optimization of the process parameters applying a fluorescamine based L-alanine quantification assay. An optimized enzyme ratio and the necessary enzyme load were identified and together with the optimal concentrations of cofactor (NAD), ammonium and buffer yields of >95% for the main branch and of 8% for the side branch were achieved.

Sustainability is a major focus of modern day research, fighting a world governed by the demand on oil. To avoid a final depletion of fossil resources, renewable alternatives need to be identified and integrated into industrial production pathways. Promising directions are the bio-mimicking of natural pathways[1] or the de novo design of artificial enzyme cascades[2]. They are considered powerful tools since natural evolution already found ways to convert a huge amount of substrates under nearly identical conditions[3,4]. Integrating multiple approaches for the conversion of biomass into biorefineries can further enhance their potential as a renewable platform for energy, fuels or chemicals[5,6]. Tailoring of enzymes towards specific problems additionally increases their versatility. In the last decades two inherently different process types evolved where either metabolically engineered whole cell approaches[7] are applied or purified enzymes from various organisms are combined to *in vitro* enzyme reaction cascades. The latter approach has been termed "(*in vitro*) synthetic biosystems"[8,9], "*in vitro* metabolic engineering"[10,11] or "synthetic pathway design"[12,13]. Isolated biocatalysts are combined with the necessary cofactors in a controlled environment to produce valuable products[3,4,10]. With this approach the advantages of well-established *in vitro* biocatalysis are combined with the power of multi-step *in vivo* pathways to yield sophisticated biomanufacturing platforms[14]. These are less effected by cytotoxic compounds, do not have a membrane limited mass transport and can be directly controlled in contrast to classical microbial fermentations[15]. With the application of thermostable enzymes the time consuming protein purification can be reduced and the process stability at elevated temperatures is enhanced[9]. A wealth of successful examples that deploy cell-free enzymatic cascades for the production of e.g. alcohols[16], chiral amines[17], saccharides[18] or hydrogen[12] has thus been reported. The absence of a regulatory cell environment enables a direct process control and further optimization of such *in vitro* enzyme networks can dramatically improve the economy[8]. To identify the appropriate enzyme ratio and thereby avoid unnecessary expenses, one approach is a thorough kinetic analysis of the required biocatalysts including inhibition constants and pH effects as demonstrated by Beer *et al.*[19]. Next to this knowledge based strategy, Liu

[1]Chair of Chemistry of Biogenic Resources, Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, Schulgasse 16, 94315, Straubing, Germany. [2]Catalysis Research Center, Technical University of Munich, Garching, Germany. [3]Fraunhofer Institute of Interfacial Biotechnology (IGB), Bio-, Electro- and Chemo Catalysis (BioCat) Branch, Straubing, Germany. [4]School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD, Australia. Correspondence and requests for materials should be addressed to V.S. (email: sieber@tum.de)

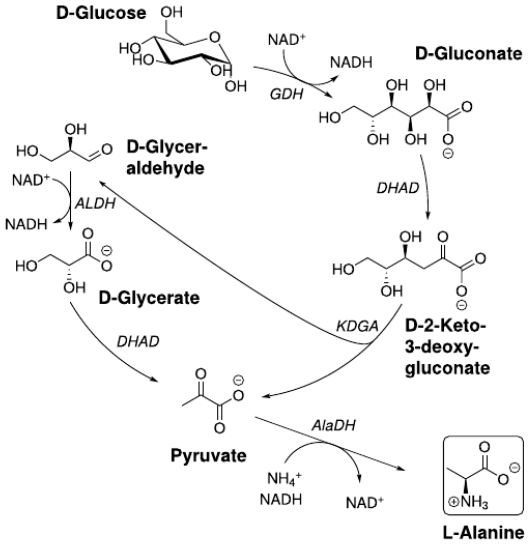


Figure 1. Schematic representation of the developed cell-free reaction cascade to L-alanine. Redox neutrality is achieved via combination of D-glucose and D-Glyceraldehyde oxidation with the reductive amination of pyruvate.

et al. described an empirical enzyme titration method, varying one enzyme at a time, in order to identify the best enzyme ratio for maximum productivity[20].

In this work, we designed a cell-free enzymatic cascade for the production of L-alanine directly from glucose and ammonium sulfate. As one of the smallest chiral compounds, L-alanine is used in various areas for example as food/ feed additive (L-alanine is the only L-amino acid with a sweet taste[21]), in health industry for nutrition therapies (e.g. main ingredient of Travasol®) and possibly as future feedstock for tailored thermoplastics[22,23]. Although various organisms have been shown to accumulate L-alanine[24], the current industrial production relies on a resting cell biotransformation by *Pseudomonas dacunhae*[25]. κ-carragenan immobilized cells rendered this process quite feasible from an economic point of view, since the half-life of the derived biocatalyst is 260 d at 37 °C and an isolated L-alanine yield of >90% can be reached by a simple ion exchange resin purification step[26]. Nevertheless the substrate L-aspartic acid couples this process to the petroleum based production of ammonium fumarate[27]. In order to render this process sustainable, the design of our cell-free reaction cascade was inspired by the non-phosphorylative Entner-Doudoroff-Pathway derived from hyperthermophilic archaea[28]. We have previously described an artificial glycolytic cascade for the conversion of D-glucose to ethanol comprising only six enzymes without any need for phosphorylation and only NAD as redox shuttle[16]. Continuing this work we identified improved enzyme variants for the conversion of glucose via gluconate, 2-keto-3-deoxy-gluconate, glyceraldehyde and glycerate towards pyruvate. In combination with an L-alanine dehydrogenase from *Archaeoglobus fulgidus* (EC 1.4.1.1, AfAlaDH) a redox neutral, L-alanine producing reaction cascade was obtained using only five to six enzymes (Fig. 1). For the final optimization of the enzyme ratio we combined the two approaches mentioned above and first determined the kinetic constants and then titrated the biocatalysts based on their maximum activity to achieve an overall increased efficiency.

Results and Discussion

Cascade design. A cell-free enzymatic reaction cascade for the conversion of D-glucose to L-alanine was developed from our previously published *in* vitro enzyme cascade to ethanol/ isobutanol[16]. In the first step, D-glucose is oxidized to D-gluconate under formation of NADH. D-gluconate then undergoes dehydration forming 2-keto 3-deoxy D-gluconate (KDG) followed by a retro aldol reaction resulting in one molecule pyruvate and one molecule D-glyceraldehyde. The aldehyde is oxidized producing another equivalent of NADH together with D-glycerate which is further converted to a second pyruvate molecule. While pyruvate was decarboxylated to acetaldehyde and further reduced to ethanol in the previously published cascade, we modified the last steps and directly integrated a reductive amination of pyruvate to yield L-alanine and recycle NAD simultaneously. With this redox neutral cascade two molecules of L-alanine can be formed from each molecule of glucose with NAD as the exclusive cofactor (Fig. 1). While the effects of ethanol or isobutanol on the enzymes can now be neglected, ammonium needs to be present for the formation of L-alanine and possible inhibitory effects on the biocatalysts need to be investigated. Inhibitory effects of L-alanine were not expected since most free amino acids tend to stabilize proteins[29,30].

Enzyme	K_M (substrate) [mM]	K_M (cofactor) [mM]	V_{max} [U/mg]
BsGDH	7.9	0.5 (NAD)	218 (15)
CcDHAD	3.2 (7.8)	—	29 (0.7)
PtKDGA	1.1 (14)	—	12 (4)
MjALDH	0.3	0.3 (17)	2 (1)
AfAlaDH	0.2	0.06 (NADH) 113 ($(NH_4)_2SO_4$)	26
SsDHAD[32]	7.8 (D-gluconate)	—	0.7 (D-gluconate) 0.01 (D-glycerate)

Table 1. Michaelis Menten kinetics at 50 °C in 100 mM HEPES pH 7.35 with 100 mM ammonium sulfate; values in brackets represent previously published values for ethanol cascade enzymes[16].

Activity based enzyme selection. Learning from our ethanol cascade[16] about the bottlenecks and drawbacks, our goal was to identify more effective enzyme variants for the pyruvate synthesis module. The combination of such an improved set of enzymes with a highly active L-alanine dehydrogenase should result in an optimized redox neutral, enzymatic cascade (Fig. 1).

At time point zero the cascade is started via the addition of D-glucose and NAD and the Glucose dehydrogenase (GDH) converts the initial substrates to D-gluconate and NADH. Previously, the GDH from the extremophile *Sulfolobus solfataricus* was applied because of its stability at the process temperature (50 °C). The v_{max} of SsGDH (15 U/mg) is in the lower range compared to other GDHs. Usually enzymes from mesophilic hosts show higher activity at moderate temperatures, but normally lack stability and degrade relatively fast. Bommarius *et* al. could increase the thermostability of a GDH from *Bacillus subtilis* by introducing two point mutations, while retaining its high specific activity[31]. Under the final process conditions this double mutant BsGDH was 14x more active towards D-glucose compared to our previously applied SsGDH, enabling a reduced enzyme load which resulted in an improved economy of the system. The following dehydration of D-gluconate towards KDG is catalyzed by a dihydroxyacid dehydratase (DHAD). In the ethanol cascade DHAD from *Sulfolobus solfataricus* was applied because of its activity for both relevant substrates D-gluconate and D-glycerate in combination with its high thermo-tolerance and excellent expression in *E. coli*. In another study we found that the preferred D-gluconate reaction is inhibited by D-glycerate, thus slowing down the entire cascade[32]. By applying two independent dehydratases, we tried to circumvent this problem and boost the overall cascade turnover. Since SsDHAD is to date the only biocatalyst that is able to catalyze the dehydration of D-glycerate, a substitution was not possible. We thus focused on the identification of an improved enzyme for the conversion of D-gluconate. Unfortunately we encountered solubility issues during the expression of archaeal gluconate dehydratases that are reportedly very active and stable against thermal degradation[33,34]. During our experiments a highly active mesophilic DHAD from *Caulobacter crescentus* (CcDHAD) was published[35]. In our hands CcDHAD was roughly 40x more active on D-gluconate than SsDHAD corresponding to a specific activity of 29 U/mg albeit at the cost of reduced thermostability. The product of the dehydration (KDG) is further subjected to a reversible retro aldol reaction. At this point the pathway is branched by an aldolase forming D-glyceraldehyde and pyruvate. Previously applied KDGA from *S. acidocaldarius* has a relatively high K_M for KDG (~14 mM) in combination with a moderate v_{max} (4 U/mg). From a range of candidate enzymes the highly specific aldolase from extremophile *P. torridus* (PtKDGA)[33] with roughly 3 times higher activity was selected. With one pyruvate equivalent obtained from the KDG-splitting, one equivalent of NAD can directly be regenerated in the following reductive amination. For this last step a variety of L-alanine dehydrogenases is described in literature. Of these the enzymes from *Bacillus subtilis* and the extremophile *Archaeoglobus fulgidus* were selected for further investigations based on their described properties like activity and pH dependence[36,37]. While both biocatalysts can be produced in a soluble form in *E. coli* and have a pH optimum around pH 9, AfAlaDH was selected because of its excellent long term heat stability. Defined as the main branch this pathway via GDH, DHAD, KDGA and AlaDH can effectively run as a redox neutral cascade producing L-alanine and D-glyceraldehyde. Regarding the economy of our process, D-glyceraldehyde needs to undergo further conversion in order to eliminate the byproduct and double the theoretical L-alanine yield. The oxidation of the aldehyde was previously done by a mutated aldehyde dehydrogenase (ALDH) from *Thermoplasma acidophilum*[38]. With the original enzyme being strictly NADP dependent, a directed evolution mutagenesis resulted in some activity for NAD[39]. Although the K_M for NAD remained relatively high (~17 mM), the outstanding advantage of this enzyme for the ethanol cascade was its exclusive activity for D-glyceraldehyde with no side reactivity for the also occurring acetaldehyde. Since the alanine cascade does not depend on this exclusive activity anymore, we decided to use the ALDH of *Methanocaldococcus jannaschii* instead due to its higher expression yields and an improved specifity for NAD. The second step in the side branch is the dehydration of the D-glycerate to obtain another equivalent of pyruvate. While the natural glycolytic pathway requires phosphorylation we could previously demonstrate a successful deployment of SsDHAD for the direct dehydration of D-glycerate[16]. Although the activity is rather low (~10 mU/mg)[32], this reaction enables our kinase-free strategy. Since D-glycerate is a strong competitive inhibitor for CcDHAD (K_I of ~0.5 mM), we tried to avoid its accumulation by applying the maximum possible amount of SsDHAD. The kinetic characteristics of all biocatalysts are summarized in Table 1.

[U/mL]	BsGDH	CcDHAD	PtKDGA	AfAlaDH
H_2O	$(4.4 \pm 0.1)*10^3$	12.3 ± 1.6	41.6 ± 1.3	2.9 ± 1.5
NH_4Cl	$(5.8 \pm 0.7)*10^3$	0	21.4 ± 2.8	91.0 ± 0.2
NH_4NO_3	$(5.7 \pm 0.1)*10^3$	0	4.8 ± 0.4	77.8 ± 1.0
$(NH_4)_2HPO_4$	$(5.0 \pm 0.3)*10^3$	2.5 ± 0.4	31.5 ± 2.8	77.5 ± 3.1
$(NH_4)_2SO_4$	$(5.0 \pm 0.1)*10^3$	8.3 ± 0.5	31.5 ± 0.2	86.2 ± 4.3

Table 2. Activity of cascade enzymes depending on the ammonium source. Reactions were carried out in triplicates at 50 °C containing 100 mM HEPES pH 7.35 and either 5 mM NAD or 0.4 mM NADH and a final concentration of 200 mM ammonium (except H_2O control). For BsGDH 100 mM glucose, for PtKDGA 1 mM KDG and an excess of MjALDH, for CcDHAD 20 mM gluconate and excess of PtKDGA/ MjALDH and for AfAlaDH 1 mM pyruvate were used. Average values are given in U/mL ± standard deviation.

Analysis of different ammonium sources. L-alanine formation was achieved via reductive amination of pyruvate with an L-alanine dehydrogenase from *Archaeoglobus fulgidus*. In our hands the K_M of AfAlaDH for ammonium was roughly 10x higher than reported in literature[36]. Because of that we had to identify a concentration of ammonium that promotes a high activity of AfAlaDH and simultaneously does not inhibit the other biocatalysts. In addition we realized different effects of the counterions on the activity of the biocatalysts of our cascade. To identify the ammonium source with least inhibition, activity assays were performed using ammonium chloride and ammonium nitrate as monovalent compounds and ammonium sulfate and diammonium phosphate respectively as bivalent ammonium sources. As shown in Table 2, BsGDH is generally activated by ammonium (+33% in case of ammonium chloride), while AfAlaDH is barely influenced by the type of ammonium source as long as ammonium is present. PtKDGA and CcDHAD show drastically reduced or no activity in presence of monovalent ammonium sources. Overall, ammonium sulfate is tolerated by all enzymes, retaining 60% and 70% activity for CcDHAD and PtKDGA. A final concentration of 200 mM ammonium was selected to be in the range of K_M for AlaDH while still not promoting further inhibition of the other enzymes.

L-alanine detection. The detection of amino acids (including L-alanine) in complex matrices is normally done via HPLC using precolumn derivatization with e.g. *o*-phtaldialdehyde (OPA)[40]. Since the run time per sample was relatively long, our aim was to identify a fast colorimetric assay, which is compatible with our enzyme and ammonium matrix. We compared a modified Berthelot reaction[41], OPA, ninhydrin and fluorescamine in order to obtain the best product monitoring platform for our system. The Berthelot reaction refers to the alkaline derivatization of ammonium by phenol and hypochlorite in order to quantify free ammonium. From the residual ammonium concentration, we wanted to infer the alanine concentration but these measurements were unsatisfying. The other three compounds represent common reagents for the derivatization and quantification of amines. Especially primary amino groups can react with these detection agents, producing chromo- or fluorophores. Unfortunately, ammonia also reacts with these compounds and forms similar products, which interfere the alanine detection. For OPA and ninhydrin the excess of ammonium in the matrix had a significant influence on the measurement and prevented an accurate quantification of L-alanine. In contrast, fluorescamine[42] derivatization showed sufficient selectivity for L-alanine over ammonium and was therefore selected to directly quantify the produced L-alanine without any prepurification (Fig. S1). Calibration was linear in the range of 1 to 30 mM L-alanine in complex enzyme/ ammonium matrix (Fig. S1). Further avoiding NADH fluorescence influence ($\lambda_{ex} = 366$ nm (in living culture), $\lambda_{em} = 460$ nm[43]), only endpoint measurements after 12 h have been taken.

Initial L-Alanine production. Since we knew from the ethanol cascade that the D-glycerate dehydration is relatively slow, we decided to start the optimization of the cascade with the main branch including the enzymes BsGDH, CcDHAD, PtKDGA and AfAlaDH while D-glyceraldehyde was accumulated. A separate optimization should avoid the risk of an NAD sink, where every consumed NAD molecule is temporarily trapped until pyruvate is formed from glycerate.

The starting conditions of 50 °C, 100 mM HEPES pH 7.35, 25 mM D-glucose and 5 mM NAD have been adapted from our previous cascade setup[16]. Together with 100 mM ammonium sulfate the enzyme reaction was incubated at 50 °C for 12 h. For the initial L-alanine production the activity of the biocatalysts was adjusted to be in a 1:1 ratio to each other and the sum of these activities was defined as total units within the cascade. Stepwise increase of the total enzyme load results in a linear increase of the L-alanine yield up to about 90% for 5.6 U/ml total enzyme (Fig. 2).

Optimization of L-alanine production. To improve L-alanine production from D-glucose while using a minimal amount of enzymes, an optimization of the enzyme ratio was performed followed by optimizations of cofactor, buffer and ammonium concentration. Applying the above mentioned starting conditions, the concentration of each enzyme was set to 0.4 U/mL and CcDHAD was increased incrementally. As depicted in Fig. 3, L-alanine yields increased steadily up to 90% with no further increase for volumetric activities higher 2 U/mL. Hence this enzyme concentration was used for the further optimization reactions. A similar optimization of PtKDGA resulted in roughly quantitative yields at 0.2 U/mL and AfAlaDH was ideal at 0.4 U/mL. This led to an optimized enzyme ratio of BsGDH: CcDHAD: PtKDGA: AfAlaDH = 2/10/1/2. Interestingly, the cofactor recycling enzymes BsGDH and AfAlaDH are present in a 1:1 ratio while only half of an equivalent PtKDGA is necessary. The required excess of CcDHAD is possibly due to its lower thermostability. In the next step we used the optimized enzyme ratio to determine the buffer, NAD and ammonium optima. The maximum L-alanine yield

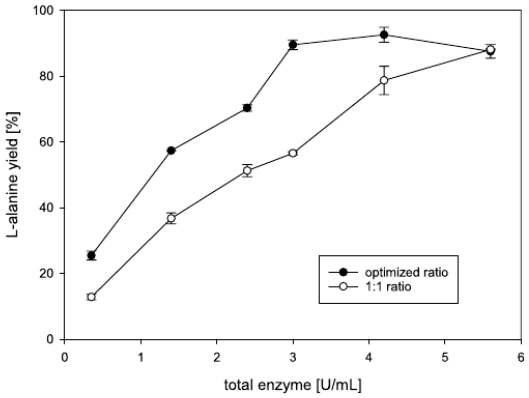


Figure 2. Comparison of L-alanine yields of different enzyme loads of the optimized enzyme ratio and a 1:1 enzyme ratio. The optimized ratio shows improved performance and higher L-alanine yields at identical enzyme load. Reaction contained either the optimized enzyme ratio BsGDH:CcDHAD:PtKDGA:AfAlaDH = 2:10:1:2 or a 1:1 ratio of all enzymes.

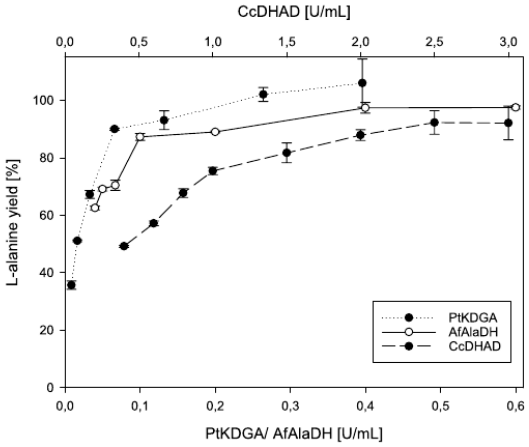


Figure 3. Single enzyme titration of CcDHAD, PtKDGA and AfAlaDH. The curves show the dependence of the L-alanine yield on the amount of single cascade enzymes.

(97%) was reached for 3 mM NAD and 75 mM ammonium sulfate (Fig. 4). For the HEPES buffer optimization 50% of the above mentioned enzyme load were applied, to temporarily increase the buffer influence on the system. The results correlate to our other experiments and we found a linear increase of product yield until 100 mM, which is corresponding to a 4 fold excess compared to the 25 mM glucose load. This excess is necessary due to acid formation during the glucose oxidation pathway. Finally we repeated the total enzyme load experiment comparing the optimized to the 1:1 enzyme ratio (Fig. 2). Although both curves increase linearily up to above 90% of L-alanine yield, the optimized enzyme ratio reaches its maximum already at 3 U/mL, while almost the double amount of enzyme (5.6 U/mL) is necessary for the classic 1:1 ratio to obtain the same L-alanine yield. Time dependent experiments revealed that the production of L-alanine ceased after 6-8 h indicating enzyme inactivation. A possible explanation could be the modification of enzymes by the accumulated D-glyceraldehyde.

To avoid this problem we included MjALDH and SsDHAD into our set up trying to redirect the D-glyceraldehyde towards pyruvate. With these additional two enzymes the theoretical yield doubles to 50 mM L-alanine from 25 mM D-glucose. Due to the very poor activity of SsDHAD for D-glycerate, we decided to use the maximum possible amount of the enzyme resulting in 4 mU/mL final activity. Applying the optimized conditions from above, we stepwise increased the concentration of MjALDH. Interestingly a very low ALDH concentration of 25 mU/mL promotes the formation of 29 mM L-alanine, clearly indicating some activity in the side branch of the cascade. Increasing the concentration of MjALDH up to 0.2 U/mL resulted in a steady decrease of the

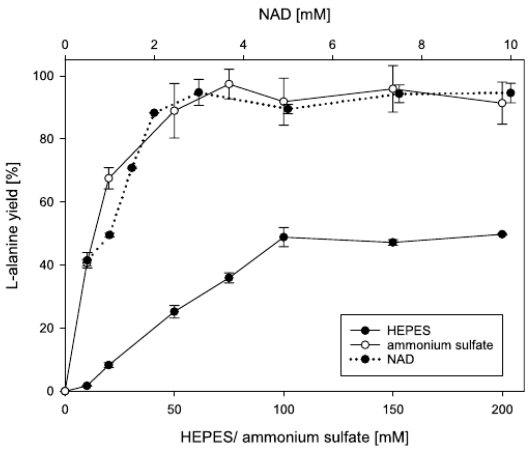


Figure 4. Dependence of the L-alanine yield on the concentrations of HEPES buffer, ammonium sulfate and NAD.

L-alanine yield reaching a plateau around 12 mM. Since almost quantitative yields (24 mM L-alanine) could be demonstrated in the main branch, this additional pathway seems to have a negative impact on the entire cascade. With increasing ALDH concentrations, the D-glycerate and NADH production speeds up and an inhibition of CcDHAD by accumulated D-glycerate could explain the influence on the main branch (for experimental analysis of CcDHAD inhibition see Figure S2). Furthermore the regeneration of the now accumulated NADH is also limited to the D-glycerate dehydration since the *in situ* cofactor recycling via AlaDH does require Pyruvate. This leads to a steady decrease of the NAD pool, finally slowing down the entire cascade to the speed of the D-glycerate dehydration. A twostep process, where the main branch was run for 12 h before MjALDH and SsDHAD were added did not produce any additional L-alanine.

Nevertheless we could establish a redox neutral, *in vitro* enzymatic reaction cascade from D-glucose to L-alanine comprising six enzymes and NAD as the only cofactor. After identification and characterization of suitable biocatalysts, a thorough optimization of the enzyme ratio, total enzyme load, buffer-, NAD- and ammonium concentration doubled the cascade's effectiveness resulting in a more cost efficient enzyme loading. By splitting the cascade into two branches, we obtained nearly quantitative yields (>95%) in the main branch via BsGDH, CcDHAD, PtKDGA and AfAlaDH plus additional 8% in the less effective side branch via MjALDH and SsDHAD and again AfAlaDH. During the reaction time of 12 h we obtained a space-time-yield of 0.17 g/(Lh) with a catalyst loading of 0.05 g of catalyst per g of L-alanine. This space-time-yield is still far behind current industrial alanine production from fumarate, where 13.4 g/(Lh) can be reached with immobilized cells from *E. coli* and *P.dacunhae*[44]. Engineered *E. coli* can form up to 121 g/L L-alanine during a 39 h fermentation resulting in a space time yield of 3.1 g/(Lh)[45]. Although we could demonstrate a successful redirection of the accumulated D-glyceraldehyde towards pyruvate, the dehydration of D-glycerate is still considered as the main bottleneck of this reaction cascade. In order to further increase the efficiency of the cascade and enable preparative scale production of L-alanine, identification of improved enzymes as well as enzyme engineering will be necessary.

Materials and Methods

Chemicals. D-glucose (monohydrate), sodium pyruvate, NAD and NADH (disodium salt) were obtained from Carl Roth GmbH and ammonium sulfate was purchased from Applichem. Sodium-D-gluconate and L-alanine were purchased from Sigma-Aldrich and fluorescamine was bought from Alfa Aesar. All chemicals were used as received and 2-keto-3-deoxy D-gluconate (KDG) was synthesized using 200 mM D-gluconate and the dihydroxyacid dehydratase from S*ulfolobus solfataricus* (SsDHAD) as described previously[32].

Strains and plasmids. *E. coli* BL21(DE3)(F^- ompT hsdSB (rB^- mB^-) gal dcm (DE3)) cells were purchased from Novagen and used for protein production. For cloning *E. coli* DH5α (fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17) was obtained from Invitrogen. The genes encoding for Dihydroxyacid dehydratase from *Caulobacter crescentus* (CcDHAD), 2-keto 3-deoxygluconate aldolase from *Picrophilus torridus* (PtKDGA), aldehyde dehydrogenase from *Methanocaldococcus jannaschii* (MjALDH) and alanine dehydrogenase from A*rchaeoglobus fulgidus* (AfAlaDH) were obtained codon optimized for *E. coli* from GeneArt Thermo Scientific and cloned into pET vectors via NdeI/ XhoI restriction sites or NcoI/ XhoI for MjALDH (Table 3). The plasmids coding for D-glucose dehydrogenase from *Bacillus subtilis* (BsGDH E170 K/ Q252L)[31] and SsDHAD[32] have been prepared previously in our lab. All sequences can be found in the supplement.

Enzyme expression. Enzyme expression was performed in *E. coli* BL21(DE3) as host strain in shaking flask cultures using ZYP-5052 autoinduction media[46], supplemented with 100 µg/mL kanamycin or 10 µg/mL

NO.	Plasmid	Enzyme	Abbrev.	Source
1	pACYC-Duet-BsGDH[31]	Glucose dehydrogenase E170K/Q252L	BsGDH	*Bacillus subtilis*
2	pET28a-CcDHAD	Dihydroxyacid dehydratase	CcDHAD	*Caulobacter crescentus*
3	pET28a-PtKDGA	2-keto-3-deoxygluconate aldolase	PtKDGA	*Picrophilus torridus*
4	pET24a-AfAlaDH	Alanine dehydrogenase	AfALaDH	*Archaeoglobus fulgidus*
5	pET28a-MjALDH	Aldehyde dehydrogenase	MjALDH	*Methanocaldococcus jannaschii*
6	pCBR-SsDHAD[32]	Dihydroxyacid dehydratase	SsDHAD	*Sulfolobus solfataricus*

Table 3. Overview of the applied enzymes and plasmids.

chloramphenicol. After 1%-inoculation of the autoinduction media with an overnight grown culture of *E. coli* BL21(DE3) harboring the corresponding plasmid, expressions were carried out over night at 30 °C and 120 rpm.

Enzyme purification. Unless stated otherwise, all steps were carried out at room temperature (25 °C), since some proteins showed sensitivity towards lower temperatures. Protein concentration was measured by a Bradford protein assay using the Roti-Nanoquant reagent (Carl Roth GmbH) according to the manufacturer's recommendations with bovine serum albumin as standard.

Cells were harvested, resuspended in 100 mM HEPES buffer pH 7.35 and disrupted by ultrasonication in an ice cooled water bath. Lysates of BsGDH, AfAlaDH and MjALDH were heat treated at 70 °C for 15 min, lysates of PtKDGA at 65 °C for 30 min. After clarifying the lysates via centrifugation, CcDHAD and PtKDGA samples were loaded onto a His GraviTrap column (GE Healthcare) equilibrated with 100 mM HEPES pH 7.35 containing 10 mM imidazole following the manufacturer's recommendations. The elution buffer contained 500 mM imidazole and subsequently the eluted protein and the heat treated lysates were desalted using a PD-10 column (GE Healthcare) equilibrated with 100 mM HEPES pH 7.35. Purified proteins were flash frozen in liquid nitrogen, stored at -80 °C and thawed freshly every day. 17.1 µmol $MnCl_2$ was added per mg of purified CcDHAD before freezing. SsDHAD purification was performed as described previously[32].

Activity assays. If not stated otherwise, all experiments contained 100 mM HEPES buffer pH 7.35, 100 mM ammonium sulfate pH 7 (both titrated with NaOH) and were preheated for 10 min at 50 °C before the reaction was started with the addition of the appropriate amount of enzyme diluted in 100 mM HEPES pH 7.35 containing 0.2 mg/mL BSA.

All enzymes were analyzed photometrically following NADH consumption or production at 340 nm ($\varepsilon_{NADH} = 6.22\,mM^{-1}cm^{-1}$) with an Epoch 2 Microplate Spectrophotometer (BioTek GmbH) at 50 °C. If not mentioned otherwise, reactions were carried out in triplicates using 96 well plates (Greiner f-bottom) containing a final volume of 200 µL. Michaelis-Menten constants were determined for all enzymes by measuring eight different substrate concentrations, plotting initial velocities vs. substrate concentration and fitting the parameters with the online tool *ic50.tk*. One unit (1 U) is defined as the amount of enzyme that consumes 1 µmol of substrate per minute.

Kinetic measurements of BsGDH contained either 0.06–2.5 mM NAD (100 mM D-glucose) or 1–100 mM D-glucose (3 mM NAD). Reaction mixtures for determining the dependence of PtKDGA activity on the concentrations of 2-keto 3-deoxy D-gluconate (KDG) contained 1 mM NAD and KDG concentrations ranging between 0.03 and 3 mM together with sufficient amounts of purified MjALDH.

Kinetic parameters of AfAlaDH were determined in cuvettes (0.9 mL final volume) with reaction mixtures containing either 0.005-0.3 mM NADH (1 mM pyruvate, 100 mM $(NH_4)_2SO_4$) or 0.1–1 mM pyruvate (0.2 mM NADH, 100 mM $(NH_4)_2SO_4$) or 15–175 mM $(NH_4)_2SO_4$ (1 mM pyruvate, 0.2 mM NADH).

Reaction mixtures for the determination of kinetic parameters of CcDHAD contained 0.2–20 mM D-gluconate, 1 mM NAD and an excess of purified PtKDGA and MjALDH.

Analysis of different ammonium sources. To identify the most suitable ammonium source, reactions were carried out as described above (at maximum substrate concentrations) with a final concentration of 200 mM ammonium. 1 M stock solutions of NH_4Cl, NH_4NO_3, $(NH_4)_2SO_4$ and $(NH_4)_2HPO_4$ were titrated with NaOH to a pH of 7 (at 25 °C) prior to use.

L-alanine detection. Alanine assays were performed in a 96 f-bottom well plate (NUNC black) in a Varioscan plate reader (Thermo Fisher) according to a modified protocol from Bantan-Polak *et al.*[42]. 10 µL of diluted sample were mixed with 45 µL of a 0.1% fluorescamine solution in dry acetonitrile. After the addition of 8 µL 100 mM sodium borate buffer pH 10 and 95 µL of water, the fluorescence was measured at $\lambda_{em} = 486$ nm ($\lambda_{ex} = 396$ nm). Signal calibration was done using L-alanine concentrations ranging from 1-30 mM and the complex matrix of the reactions.

Optimization of L-alanine production. One pot L-alanine synthesis reactions were set up in PCR-tubes containing 25 µL final volume and incubated in a C1000 Thermo Cycler (BIORAD) at 50 °C for 12 h. The initial reaction setup contained 25 mM D-glucose, 5 mM NAD, 100 mM HEPES pH 7.35, 100 mM ammonium sulfate pH 7 and 0.4 U/mL of each enzyme (BsGDH, CcDHAD, PtKDGA and AfAlaDH). BsGDH was kept constant while the activity of the other enzymes was varied. The tested enzyme concentration for the optimization was CcDHAD 0.4–3 U/mL, PtKDGA 0.013-0.8 U/mL and AfAlaDH 0.04-0.8 U/mL using the optimized values for following optimizations. In addition the total enzyme load (0.35-5.6 U/mL), the NAD concentration (0.5-10 mM),

the ammonium sulfate concentration (0–200 mM) and the HEPES buffer concentration (0-200 mM) were optimized accordingly.

Reactions for investigating the full cascade contained the optimized conditions for the main branch (0.4 U/mL BsGDH, 2 U/mL CcDHAD, 0.2 U/mL PtKDGA, 0.4 U/mL AfAlaDH, 5 mM NAD, 100 mM HEPES buffer pH 7.35, 100 mM ammonium sulfate and 25 mM D-glucose). MjALDH was used at levels between 13-800 mU/mL in combination with 0.42 mg/mL SsDHAD.

Data Availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

References

1. Kim, E.-J., Wu, C.-H., Adams, M. W. W. & Zhang, Y. H. P. Exceptionally High Rates of Biological Hydrogen Production by Biomimetic *In Vitro* Synthetic Enzymatic. *Pathways. Chem. Eur. J* **22**, 16047–16051 (2016).
2. France, S. P., Hepworth, L. J., Turner, N. J. & Flitsch, S. L. Constructing Biocatalytic Cascades: *In Vitro* and *in Vivo* Approaches to de Novo Multi-Enzyme Pathways. *ACS Catal* **7**, 710–724 (2017).
3. Sperl, J. M. & Sieber, V. Multi-enzyme cascade reactions – status and recent advances. *ACS Catal* **8**, 2385–2396 (2018).
4. Schrittwieser, J. H., Velikogne, S., Hall, M. & Kroutil, W. Artificial Biocatalytic Linear Cascades for Preparation of Organic Molecules. *Chem. Rev.* **118**, 270–348 (2018).
5. Venkata Mohan, S. *et al*. Waste biorefinery models towards sustainable circular bioeconomy: Critical review and future perspectives. *Bioresour. Technol.* **215**, 2–12 (2016).
6. Hasunuma, T. *et al*. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresour. Technol.* **135**, 513–522 (2013).
7. de Carvalho, C. C. C. R. Enzymatic and whole cell catalysis: Finding new strategies for old processes. *Biotechnol. Adv.* **29**, 75–83 (2011).
8. Zhang, Y.-H. P. Production of biofuels and biochemicals by *in vitro* synthetic biosystems: Opportunities and challenges. *Biotechnol. Adv.* **33**, 1467–1483 (2015).
9. You, C. & Percival Zhang, Y. H. Biomanufacturing by *in vitro* biosystems containing complex enzyme mixtures. *Process Biochem.* **52**, 106–114 (2017).
10. Hodgman, C. E. & Jewett, M. C. Cell-free synthetic biology: thinking outside the cell. *Metab. Eng.* **14**, 261–269 (2012).
11. Jung, G. Y. & Stephanopoulos, G. A Functional Protein Chip for Pathway Optimization and *in Vitro* Metabolic Engineering. *Science* **304**, 428–431 (2004).
12. Martin del Campo, J. S. *et al*. High-yield production of dihydrogen from xylose by using a synthetic enzyme cascade in a cell-free system. *Angew. Chem. Int. Ed. Engl.* **52**, 4587–4590 (2013).
13. Medema, M. H., van Raaphorst, R., Takano, E. & Breitling, R. Computational tools for the synthetic design of biochemical pathways. *Nat Rev Microbiol* **10**, 191–202 (2012).
14. Morgado, G., Gerngross, D., Roberts, T. M. & Panke, S. In *Synthetic Biology – Metabolic Engineering* (eds Huimin Zhao & An-Ping Zeng) 117–146 (Springer International Publishing, 2018).
15. Hold, C., Billerbeck, S. & Panke, S. Forward design of a complex enzyme cascade reaction. *Nat Commun* **7**, 12971 (2016).
16. Guterl, J. K. *et al*. Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem* **5**, 2165–2172 (2012).
17. Altenbuchner, J., Siemann-Herzberg, M. & Syldatk, C. Hydantoinases and related enzymes as biocatalysts for the synthesis of unnatural chiral amino acids. *Curr. Opin. Biotechnol.* **12**, 559–563 (2001).
18. Fessner, W. D. Systems Biocatalysis: Development and engineering of cell-free "artificial metabolisms" for preparative multi-enzymatic synthesis. *N Biotechnol* **32**, 658–664 (2015).
19. Beer, B., Pick, A. & Sieber, V. *In vitro* metabolic engineering for the production of alpha-ketoglutarate. *Metab. Eng.* **40**, 5–13 (2017).
20. Liu, Z. *et al*. *In Vitro* Reconstitution and Optimization of the Entire Pathway to Convert Glucose into Fatty Acid. *ACS Synth Biol* **6**, 701–709 (2017).
21. Solms, J., Vuataz, L. & Egli, R. H. The taste of L- and D-amino acids. *Experientia* **21**, 692–694 (1965).
22. Mallakpour, S. & Behranvand, V. The influence of acid-treated multi-walled carbon nanotubes on the surface morphology and thermal properties of alanine-based poly(amide–imide)/MWCNT nanocomposites system. *Colloid. Polym. Sci.* **293**, 333–339 (2014).
23. Mallakpour, S. & Dinari, M. Progress in Synthetic Polymers Based on Natural. *Amino Acids. J MACROMOL SCI A* **48**, 644–679 (2011).
24. Zhang, X., Jantama, K., Moore, J. C., Shanmugam, K. T. & Ingram, L. O. Production of l-alanine by metabolically engineered Escherichia coli. *Appl. Microbiol. Biotechnol.* **77**, 355–366 (2007).
25. Chibata, I., Kakimoto, T. & Kato, J. Enzymatic Production of l-Alanine by Pseudomonas dacunhae. *Appl. Microbiol.* **13**, 638–645 (1965).
26. Çalık, G., Savaşçı, H., Çalık, P. & Özdamar, T. H. Growth and κ-carrageenan immobilization of Pseudomonas dacunhae cells for l-alanine production. *Enzyme Microb. Technol.* **24**, 67–74 (1999).
27. Takamatsu, S., Tosa, T. & Chibata, I. Industrial production of L-alanine from ammonium fumarate using immobilized microbial cells of two kinds. *J. Chem. Eng. Japan* **19**, 31–36 (1986).
28. Ahmed, H. *et al*. The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem. J.* **390**, 529–540 (2005).
29. Taneja, S. & Ahmad, F. Increased thermal stability of proteins in the presence of amino acids. *Biochem. J.* **303**(Pt 1), 147–153 (1994).
30. Bozorgmehr, M. R. & Monhemi, H. How Can a Free Amino Acid Stabilize a Protein? Insights from Molecular Dynamics Simulation. *J. Solution Chem.* **44**, 45–53 (2015).
31. Vazquez-Figueroa, E., Chaparro-Riggers, J. & Bommarius, A. S. Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. *ChemBioChem* **8**, 2295–2301 (2007).
32. 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).
33. 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.* **192**, 964–974 (2010).
34. 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).
35. Andberg, M. *et al*. Characterization and mutagenesis of two novel iron-sulphur cluster pentonate dehydratases. *Appl. Microbiol. Biotechnol.*, (2016).
36. Schroder, I., Vadas, A., Johnson, E., Lim, S. & Monbouquette, H. G. A novel archaeal alanine dehydrogenase homologous to ornithine cyclodeaminase and μ-crystallin. *J. Bacteriol.* **186**, 7680–7689 (2004).

37. Yoshida, A. & Freese, E. Enzymic properties of alanine dehydrogenase of Bacillus subtilis. *Biochim Biophys Acta Gene Regul Mech* **96**, 248–262 (1965).
38. Steffler, F. & Sieber, V. Refolding of a thermostable glyceraldehyde dehydrogenase for application in synthetic cascade biomanufacturing. *PLoS One* **8**, e70592 (2013).
39. Steffler, F., Guterl, J.-K. & Sieber, V. Improvement of thermostable aldehyde dehydrogenase by directed evolution for application in Synthetic Cascade Biomanufacturing. *Enzyme Microb. Technol.* **53**, 307–314 (2013).
40. Hill, D. W., Walters, F. H., Wilson, T. D. & Stuart, J. D. High performance liquid chromatographic determination of amino acids in the picomole range. *Anal. Chem.* **51**, 1338–1341 (1979).
41. Rhine, E. D., Mulvaney, R. L., Pratt, E. J. & Sims, G. K. Improving the Berthelot Reaction for Determining Ammonium in Soil Extracts and Water. *Soil Sci. Soc. Am. J.* **62**, 473–480 (1998).
42. Bantan-Polak, T., Kassai, M. & Grant, K. B. A comparison of fluorescamine and naphthalene-2,3-dicarboxaldehyde fluorogenic reagents for microplate-based detection of amino acids. *Anal. Biochem.* **297**, 128–136 (2001).
43. Beyeler, W., Einsele, A. & Fiechter, A. On-line measurements of culture fluorescence: Method and application. *Appl. Microbiol. Biotechnol.* **13**, 10–14 (1981).
44. Takamatsu, S., Tosa, T. & Chibata, I. Production of L-alanine from ammonium fumarate using two microbial cells immobilized with k-carrageenan. *J. Chem. Eng. Japan* **18**, 66–70
45. Zhou, L., Deng, C., Cui, W. J., Liu, Z. M. & Zhou, Z. M. Efficient L-Alanine Production by a Thermo- Regulated Switch in Escherichia coli. *Appl. Biochem. Biotechnol.* **178**, 324–337
46. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif.* **41**, 207–234 (2005).

Acknowledgements

This work was supported by the Federal Ministry of Education and Research (Germany) under grant agreement number 031B0071A and a COST Action CM1303 Systems Biocatalysis grant to T.G. Publishing of this work was supported by the German Research Foundation (DFG) and the Technical University of Munich within the funding programme Open Access Publishing. The authors would like to thank Magdalena Haslbeck for assistance in molecular cloning and kinetic experiments.

Author Contributions

T.G., J.S. and V.S. designed the study. T.G. performed the experiments and T.G., J.S. and V.S. analyzed the data. T.G., J.S. and V.S. wrote and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-48151-y.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.2 Molecular Dynamics Analysis of a Rationally Designed Aldehyde Dehydrogenase Gives Insights into Improved Activity for the Non-Native Cofactor NAD^+

In this publication, the enzyme engineering of the thermostable glyceraldehyde dehydrogenase from *Thermoplasma acidophilum* (TaALDH) towards an improved acceptance of the desired cofactor NAD^+ is described. From a homology model based on incomplete crystal structures, seventeen positions of interest were identified and targeted with a sequence saturation approach allowing all canonical amino acids. By screening more than 80 colonies per saturated position, six positions showed an improved volumetric activity. While three positions were next to the cofactor binding site (S206K, S175E, D176S) and one position was on the surface (M262I), the greatest impact on the activity was detected for the two mutations at the substrate binding site (W271 and W275). Further kinetic analysis could demonstrate an obliteration of the rather strong substrate inhibition (Ki = 1.25 mM) by these mutations. Additionally the manuscript describes the crux of identifying the most effective amino acid at those two vital positions, since they depended on the genetic background and could not be predicted beforehand. To identify the best combination of these beneficial mutations, the specific point mutations S175E, D176S, S206K and M262I were assembled in a combinatorial way, before saturating the most active variant simultaneously at positions 271 and 275. Various measurements with different NAD^+ concentrations showed increasing volumetric activity for combined mutants indicating K_M as well as k_{cat} improvements. Finally the combination of five point mutations reduced the K_M for NAD^+ from 18 mM to 0.6 mM and increased the activity for the desired substrate D-glyceraldehyde from 0.4 U/mg to 1.5 U/mg. This final variant was then compared to the initial variant via a 10 ns molecular dynamics simulation. Under identical conditions, the final variant showed an increased RMSF especially in the cofactor binding region. Additionally, the reactive nicotinamide increased in flexibility, while the average distance to the active site stayed constant. Summarizing those findings into a novel design principal, the manuscript combines the practical world of enzyme engineering with the theoretical world of molecular dynamics.

Tobias Gmelch designed the screening, selected the positions of interest and did all the experiments, molecular dynamics simulations, calculations and analysis and wrote the manuscript. Josef Sperl supported in designing the experiments and proofread the manuscript. Tobias Gmelch designed the research project and experimental approach guided by Volker Sieber, who also proofread the manuscript.

Supporting Information to this manuscript can be found in the appendix.

Molecular Dynamics Analysis of a Rationally Designed Aldehyde Dehydrogenase Gives Insights into Improved Activity for the Non-Native Cofactor NAD^+

Tobias J. Gmelch, Josef M. Sperl, Volker Sieber

ACS Synthetic Biology
2020

DOI: https://doi.org/10.1021/acssynbio.9b00527

ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

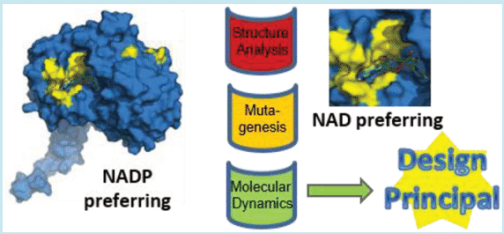
Molecular Dynamics Analysis of a Rationally Designed Aldehyde Dehydrogenase Gives Insights into Improved Activity for the Non-Native Cofactor NAD+

Tobias J. Gmelch, Josef M. Sperl, and Volker Sieber*

Cite This: ACS Synth. Biol. 2020, 9, 920–929 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The aldehyde dehydrogenase from *Thermoplasma acidophilum* was previously implemented as a key enzyme in a synthetic cell-free reaction cascade for the production of alcohols. In order to engineer the enzyme's cofactor specificity from NADP+ to NAD+, we identified selectivity-determining residues with the CSR-SALAD tool and investigated further positions based on the crystal structure. Stepwise combination of the initially discovered six point mutations allowed us to monitor the cross effects of each mutation, resulting in a final variant with reduced K_M for the non-native cofactor NAD+ (from 18 to 0.6 mM) and an increased activity for the desired substrate D-glyceraldehyde (from 0.4 to 1.5 U/mg). Saturation mutagenesis of the residues at the entrance of the substrate pocket could eliminate substrate inhibition. Molecular dynamics simulations showed a significant gain of flexibility at the cofactor binding site for the final variant. The concomitant increase in stability against isobutanol and only a minor reduction in its temperature stability render the final variant a promising candidate for future optimization of our synthetic cell-free enzymatic cascade.

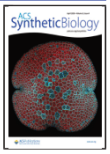


KEYWORDS: *protein engineering, molecular dynamics, cofactor specificity, oxidoreductase, semirational engineering*

Rational protein design is a powerful tool to improve desired enzyme functions. On the basis of a template DNA, carefully designed point mutations can drastically improve activity or stability and enable novel or artificial reactions. The combination of such design approaches with molecular dynamics (MD) can visualize the effects of point mutations and facilitate structure–function analysis.[1] With increasing computational power, the calculation of protein motions allows us to differentiate structurally flexible areas from rigid ones, which is a key factor to understand the thermostability of biocatalysts, for example.[2] Although current MD results strongly depend on the selected force field, recent developments show that machine learning algorithms can help to define force fields that will enhance the predictive capacity.[3] Nevertheless, comparing identically treated systems can already provide insights into enzyme functionality and render MD simulations a valuable tool for the analysis of rationally designed mutations.

Oxidoreductases are nature's way of catalyzing redox reactions in living systems. The biological electron transfer is usually done *via* a hydride shift to a nicotinamide-based cofactor, namely, nicotinamide adenine dinucleotide (NAD+) or its phosphorylated derivative (NADP+). With nicotinamide as the chemically active functionality, they only differ in a phosphorylation at the ribose moiety, which is sufficiently distant from the active end of the molecule. Nevertheless, a natural specificity evolved and biocatalysts usually show a strong preference for one derivative with significantly reduced activity for the other. Because of this, different anabolic and catabolic pathways can coexist in living cells without influencing each other. For an artificial cascade design, this exclusiveness can be rather challenging if the desired biocatalyst favors the "wrong" cofactor. For this reason, there are numerous enzyme engineering examples for the realization of a cofactor switch.[4−6] Although various tools and procedures have been suggested for this task,[7−10] switching an enzyme's cofactor preference is still challenging and needs to be investigated on the single enzyme level. We have run into this situation previously when developing a cell-free enzymatic cascade from D-glucose to ethanol or isobutanol.[11] Representing an artificial Entner–Doudoroff pathway common for archaea, the *in vitro* cascade does not require any phosphorylation step and relies on NAD+ as the sole redox cofactor. From a screening of various aldehyde dehydrogenases, only the enzyme from *Thermoplasma acidophilum* (TaALDH) matched the cascade requirements regarding

Received: December 23, 2019
Published: March 25, 2020



https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920–929

temperature stability, efficient heterologous expression in *E. coli*, and exclusive activity for D-glyceraldehyde (not active on acetaldehyde). The remaining drawback was its cofactor preference toward $NADP^+$ resulting in high K_M for NAD^+ and a rather low overall activity under cascade conditions. Previous attempts to improve this enzyme included refolding for increased protein yield[12] and a directed evolution approach.[13] Although the expression yield and total activity could be increased, the specificity toward NAD^+ remained mostly unchanged. In this study, we continued our work, aiming at a high activity of TaALDH with the cofactor NAD^+. The most active variant from the previous screening F34M, Y399C, S405N (here termed M0) was structurally analyzed, and positions of interest were saturated with an NNK-motif. Mutations resulting in increased activity were combined, and the final variant was then compared to M0 *via* a molecular dynamics simulation in order to understand the molecular basis of the activation.

RESULTS AND DISCUSSION

Identification of Mutagenesis Targets. We previously reported the results of a random mutagenesis approach for the aldehyde dehydrogenase of *Thermoplasma acidophilum* (TaALDH), where we could increase volumetric activity and slightly improve NAD^+ acceptance.[13] Based on the selected triple mutant F34M, Y399C, S405N (M0), we designed this study to further improve activity for NAD^+ as well as overall activity while retaining substrate specificity and thermostability. The three available crystal structures of the TaALDH variants in the PDB database (5IZD, wild type; 5M4X, triple mutant M0; and 5J77, F34M, S405N), show rather poor quality with multiple amino acids missing in the structures. Since a full structure containing all residues was essential for our molecular dynamics approach, we decided to model the full sequence of the triple mutant TaALDH (M0) with SWISS-MODEL against the 5M4X template in order to include all residues. The quality of the resulting model was checked by Ramachandran structure validation using RAMPAGE,[14] resulting in 94.1% residues in a favored region, 5.1% residues in an allowed region, and 0.8% residues in an outlier region, which indicates a good structural quality.[15,16] In comparison to 5M4X, the model structure M0 shows three additional residues in outlier regions but gives coordinates for 32 amino acids, which were not represented in the crystal structure solution (see the Supporting Information). Further validation of the model structure M0 was conducted using the ERRAT server.[17] This online tool provides a statistical method analyzing the non-bonded atomic interactions in order to differentiate between correctly and incorrectly predicted protein structures. Both uploaded structures obtained quality factors above 90, indicating reliable atomic interaction (5M4X, 97.5 and M0, 92.4). Next, the Verify 3D server was used, which evaluates accuracy by comparing the input structure to its own amino acid sequence, using a three-dimensional profile generated from atomic coordinates of the 3D structure.[18] The analysis revealed that 90.67% of the residues of the M0-model have an average 3D−1D score ≥0.2, indicating a good reliability of the model. As expected, 5M4X performed poorly within this analysis, with only 72.89% of the residues reaching a score ≥0.2. This is possibly due to the missing amino acids in the crystal structure. A similarity search of the constructed model within the PDB database yielded lactaldehyde dehydrogenase from *E. coli* as the most similar structure (2IMP: 39% identity, 60% positives, and 0.976 rms), which shows identical orientation of the highly conserved ALDH residues K172, E247, and C281 (see the Supporting Information).[19] With this result, we could further confirm the quality of our model. The structure 2IMP and partially 5M4X contained information on the cofactor position, which was used for the docking of NAD^+ with YASARA. As depicted in the energy minimized structure (Figure 1), the binding motive for NAD^+ is located on a surface-exposed binding pocket.

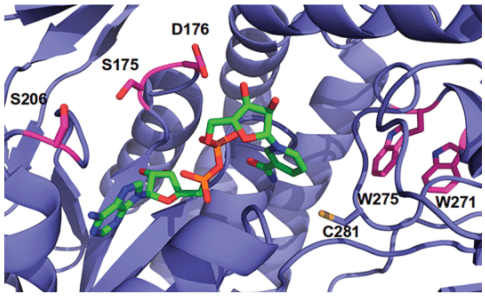


Figure 1. NAD^+ binding pocket model of TaALDH (M0) with docked NAD^+ (shown as green sticks).

The substrate can access the active center from the opposite side of the protein *via* another cleft, and the catalytically active C281 is placed directly in between the two substrates showing an initial distance to DGA of 4.2 Å and to NAD^+ of 4.7 Å. Our model of M0 with docked NAD^+ was then subjected to the CSR-SALAD tool. Originally, this tool was designed to predict cofactor specificity determining residues using crystal structures with bound native cofactor as input.[8] We were only aiming on enhancing the activity for the non-native cofactor NAD^+ and not necessarily on reversing the selectivity, and thus, we used our model with docked NAD^+ as input. The results suggested to simultaneously target the residues G205 with an RSA codon, S174 with an RSC codon, and K172 with an AVK codon in order to switch the reactivity. For activity recovery, the residues G205, G209, T229, R232, and I233 should be targeted with medium priority and D176 as well as D210 should be targeted with low priority (see the Supporting Information for the detailed CSR-SALAD report). Deviating from the standard CSR method, we decided to screen all positions predicted from the NAD structure (except I233, where we discovered PCR issues) separately with an NNK-motif in order to identify the effect of each individual position. Additionally, positions W271 and W275 were included in the screening, since Steffler *et al.* had previously identified a major impact on the overall activity without a detailed characterization.[13] Furthermore, positions S206 and E207 at the cofactor entrance site and positions S175 and T177 next to the $NADP^+$ phosphate binding site were selected. Finally, the surface-exposed positions M262, R306, and S349 were found to be beneficial in a previous directed evolution experiment and were also included in this study. A full list of the corresponding oligonucleotides can be found in the Supporting Information.

Screening Coverage and Hit Identification. Library quality and coverage are important aspects to consider for sequence saturation methods. Reetz *et al.* investigated mutagenesis yields for NNK saturation and found values ranging from 52 to 72% depending on the supplier of the

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

degenerated oligonucleotides.[20] By us assuming a 50% mutagenesis yield, the screening of 80 or 2500 variants for the single or double NNK libraries resulted in a codon coverage of roughly 70%, while P_{T2} is greater than 95% (for a detailed description, see the Supporting Information). Due to codon diversity, most amino acids are encoded in multiple codons (*e.g.*, alanine = GCN). Since our focus was on the protein level, we did not want to screen all codons but identify the amino acid resulting in the highest activity. With our selected economic library size, we almost certainly identified at least the second best variant within each library. Prior to the investigation of mutant enzymes, the screening platform was evaluated applying the Z-factor as a measure for assay quality, resulting in an excellent screening with a value of 0.86.[21]

The screening of most variant enzymes showed reduced activity or none at all for the single NNK-saturated positions, and only variants with mutations at 175, 176, 206, and 262 as well as 271 and 275 exceeded the hit limit. As depicted in the variant landscape (Figure 2), mutations at 206 and 262 led to

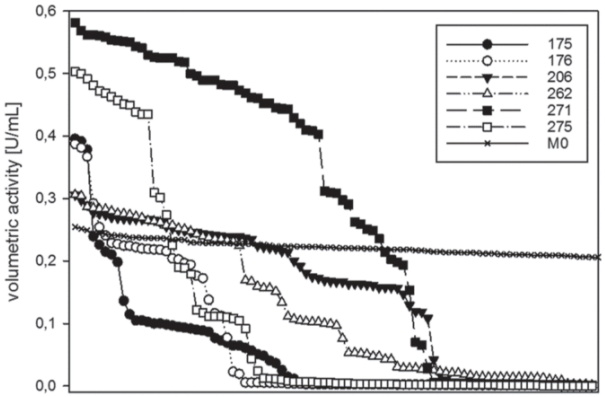


Figure 2. Variant landscape for single NNK screening compared to template activity (M0); the reactions were carried out at room temperature with 100 mM HEPES pH 7.35, 100 mM ammonium sulfate pH 7, 1.3 mM DGA, and 1 mM NAD^+ in a final volume of 200 μL.

the smallest activity increase, barely exceeding the hit limit. Mutations at 175 and 176 had roughly 70% more activity, showing in each case four mutants with similar activity. Mutations at positions 271 and 275 more than doubled the activity of M0, with a rather high proportion of variants showing higher activity than the template.

Sequencing revealed that a specific amino acid exchange leads to the improved activity for the variants S175E, D176S, S206K, and M262I. While the first three positions are located around the NAD^+ binding site, M262 is surface exposed and significantly distant from the substrate binding areas. On the contrary, W275 is directly facing the DGA pocket on one side and aligns with W271 on the other. Sequencing showed mutational potential at those two positions, with the most active variants being W271Y and W271L or W275T and W275L. In short, we realized that many mutations at positions 271 or 275 increase the activity but that the specific amino acid exchange (which results in the highest gain of activity) needs to be identified for every single genetic background. While D176 was directly identified by CSR-SALAD as low priority activity recovery, the other successful positions around the NAD^+ pocket (S175 and S206) were directly next to predicted positions. Unfortunately, the specificity determining residues predicted by the tool (K172, S174, and G205) did not lead to an activity increase, which might be due to our sequential screening protocol (instead of the suggested combinatorial approach). Nevertheless, CSR-SALAD proved to be an efficient tool for the prediction of selectivity-determining residues.

Sensitivity of Positions 271 and 275. Since the results of the first screening suggested a large mutational potential of positions W271 and W275, the four identified variants S175E, D176S, S206K, and M262I were tested individually with an NNK-motif at either W271 or W275. By introducing mutations at 271, all variants demonstrated double or nearly triple template enzyme activity, with the exception of variant S175E, which could not be improved further (Supporting Information). The most active combinations were S206K with W271T/N and M262I with W271L/I. Since tyrosine was the preferred amino acid at position 271 for M0 and D176S, the exchange of tryptophan at position 271 seems beneficial within the class of hydrophobic amino acids. Concerning mutations at position 275, the templates S175E and S206K showed the highest activity in combination with W275N/T/L, while D176S and M262I preferred W275S/I and W275M/L (Supporting Information). With polar uncharged as well as hydrophobic residues, there is no obvious preference for a specific type of amino acid at position 275. Overall, the most active variant from this screening of 271/275 NNK variants was the M262I + W271L/I variant, with roughly five times higher activity compared to that of the initial template M0. Since the single M262I mutation resulted in 30% higher activity and the best single 271 mutation just doubled the activity of M0, there were obviously synergistic effects. This amino acid flexibility of positions 271 and 275 reveals a strong dependence on the genetic background of the template and is summarized in the Supporting Information. To identify the best combination of our set of mutations, we decided to combine the specific point mutations S175E, D176S, S206K, and M262I in a combinatorial way first, before saturating the most active variant simultaneously at positions 271 and 275 to obtain the overall best variant.

Analysis of Combined Mutants. A combinatorial assembly of the four identified point mutations resulted in 15 variants that were cloned directly *via* site-directed mutagenesis. From each mutant, we analyzed 10 independently grown colonies based on their volumetric activity (Table 1). Under standard screening conditions (1 mM NAD^+), the single mutant activities were comparable to the first screening round, with S175E showing the highest activity. The four double mutants containing either S206K or M262I, together with either S175E or D176S, had an activity within the range 0.45−0.56 U/mL, hence exceeding double the activity of M0. Interestingly, the combination of the two most active single mutants S175E and D176S resulted in a lower activity similar to that of the combination of the two least active mutations S206K and M262I. Apparently, the combination of S175E and D176S is unfavorable in all mutants. Regarding the triple mutants, three out of four could further increase the activity of the double mutants, with M33 (D176S, S206K, and M262I) being the most active variant. Under standard conditions (1 mM NAD^+), M33 demonstrated a 3-fold increase of the volumetric activity compared to that of M0 (up to 0.77 U/mL). This high activity could not be exceeded by the quadruple mutant, which showed slightly reduced activity.

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

Table 1. Volumetric Activity of Combined TaALDH Variants[a]

[U/mL]		1 mM NAD^+	3% isobutanol	5 mM NAD^+	150 μM $NADP^+$
media	media	0.00	0.00	0.00	0.00
pET28a	pET28a	0.00	0.00	0.01	0.01
template	M0	0.23	0.08	0.86	0.77
S175E	M12	0.46	0.19	1.07	0.06
D176S	M17	0.39	0.20	0.89	0.52
S206K	M13	0.29	0.10	1.02	0.73
M262I	M14	0.28	0.10	1.11	0.72
S175E S206K	M18	0.45	0.17	1.07	0.10
D176S S206K	M19	0.48	0.25	1.01	0.57
D176S M262I	M20	0.57	0.28	1.30	0.68
S175E M262I	M26	0.56	0.24	1.36	0.09
S20 K M262I	M29	0.35	0.14	1.38	0.65
S175E D176S	M31	0.33	0.11	0.65	0.13
S175E S206K M262I	M28	0.56	0.22	1.42	0.15
S175E D176S M262I	M32	0.61	0.24	1.28	0.24
D176S S206K M262I	M33	0.77	0.44	1.63	0.88
S175E D176S S206K	M34	0.23	0.09	0.55	0.18
S175E D176S S206K M262I	M35	0.49	0.19	1.18	0.37
D176S S206K M262I W271Y W275V	M42	2.00	1.18	3.03	1.46

[a]Reactions were carried out at room temperature with 100 mM HEPES pH 7.35, 100 mM ammonium sulfate pH 7, 1.3 mM DGA, and either 1 mM NAD^+, 5 mM NAD^+, or 150 μM $NADP^+$ in a final volume of 200 μL. Reactions with 3% v/v isobutanol contained 1 mM NAD^+. The values represent the average activity of ten independently grown colonies.

In order to distinguish between k_{cat} and K_M improvements, measurements containing 5 mM NAD^+ were conducted. Since the activity of all mutants could be further increased with a higher cofactor concentration, we concluded that 1 mM NAD^+ is rather far from saturating conditions. Furthermore, measurements with 150 μM $NADP^+$ indicated an unchanged strong preference for $NADP^+$, since the activity level of most mutants remained above the experiments with 1 mM NAD. Exceptions are variants bearing the S175E mutation, which seems to have a severe impact on the $NADP^+$ binding. The rather poor $NADP^+$ activity of the single mutant could be restored to some extent in combination with the D176S mutation. Since S175E is predicted to be part of the phosphate binding motive, D176S might partially substitute the missing serine functionality.

Since the final mutant should also be applied in the previously published cell-free enzymatic cascade toward isobutanol,[11] the solvent stability was analyzed with measurements containing 3% v/v of isobutanol in addition to the standard conditions (1 mM NAD^+). A general reduction of the overall activity was observed for all mutants to roughly 30−50% of the activity without isobutanol. Remarkably, M33 demonstrated increased stability, retaining 60% of the initial activity. Overall, the most promising variant M33 showed the highest activity during all experiments and was therefore selected for the simultaneous saturation of 271/275 NNK.

The results of the double NNK saturation at 271 and 275 are presented in Figure 3, visualizing the activity data of more than 2500 variants.

Roughly 450 mutants did not show any activity, and around 400 mutants had an activity below the template (M33) threshold, which is clearly clustering around 0.75 U/mL. Considering the results from the single NNK screening for

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

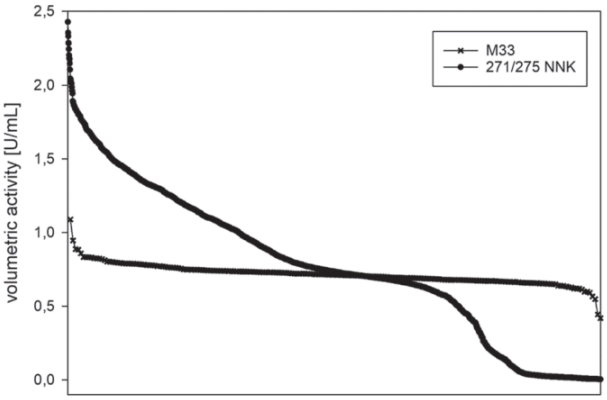


Figure 3. Results of the simultaneous NNK saturation of variant M33 at positions 271 and 275. Reactions were carried out at room temperature in 384 MTP with 100 mM HEPES pH 7.35, 100 mM ammonium sulfate pH 7, 1.3 mM DGA, and 1 mM NAD^+ in a final volume of 60 μL. In total, 2752 variants and 224 templates were analyzed.

either 271 or 275, we expected a high number of improved variants and found more than 1000 mutants (>40% of the library) that were more active than the template (M33). From this variety, only seven mutants had an activity greater than 2.1 U/mL, corresponding to more than three times the activity of M0. These seven mutants were regrown from backup plates and measured again in 10 replicates similar to the 1 mM NAD^+ measurements of the combined mutants. The four most active variants from this rescreen were sent for sequencing, and all four showed the mutations W271Y and W275V (M42). Characterization of M42 revealed an activity of 2 U/mL with 1 mM NAD^+ (~10× M0), 3 U/mL with 5 mM NAD^+, similar isobutanol tolerance compared to M33, and almost equal activities with 1 mM NAD^+ or 150 μM $NADP^+$.

Characterization of Key Variants. This present screening was primarily designed to improve the cofactor activity of the TaALDH toward NAD^+ in order to get a promising candidate for the cell-free enzymatic cascade toward isobutanol. Additional requirements such as stability at 50 °C, sufficient solvent stability toward isobutanol, and a high K_i/K_M ratio for DGA were also important for a successful candidate. The kinetic characterization of the key variants M0, M33, and M42 (summarized in Table 2) revealed a rather strong substrate inhibition of M0 and M33 by DGA, which is not present in M42.

Although the K_M of DGA increased up to 0.4 mM for the M42 variant, the kinetic curve is always above the M0 and M33 curve (see the Supporting Information). This is due to the increased V_{max} and the obliterated substrate inhibition. Furthermore, the K_M for NAD^+ could be decreased from 18 mM for M0 to 1.3 mM for the preliminary variant M33 and 0.6 mM for final variant M42. This increase of activity with the NAD^+ cofactor confirmed our screening method designed for this purpose. The temperature and storage stability were also included in the screening with a freeze/thaw cycle of the harvested cell pellet and a heat step of 60 °C for 1 h of the lysed cells. Consequently, M0 and M33 showed an almost identical melting point around 79 °C, while the thermofluor assay indicated a slight reduction of the thermostability of M42. Nevertheless, M42 is sufficiently stable for the enzymatic cascade, which was run at 50 °C.

Without especially screening for solvent stability, isobutanol stability increased from M0 to M33 and even more to M42 (Table 1). The most striking difference between M0 and M42 is the enlarged substrate pocket, which is realized by mutations at position W271 and W275, thereby enabling DGA to have a closer contact to the binding site. A previous report of Damborsky *et al.* illustrated that the gatekeeper residue (in this case W275) at the entrance of a substrate pocket greatly affects substrate inhibition.[22] This is in accordance with our results since substrate inhibition was drastically reduced by mutations at this position. For surface mutation M262I, we expected an influence on the solubility, and the exchanges of D176S and S206K might influence the cofactor binding since they are located at this binding site. To get further insights into the molecular alterations causing these improvements, we decided to continue the analysis with molecular dynamics simulations.

Molecular Dynamics. The analysis of the activation of M42 in comparison to that of M0 was done *via* 10 ns molecular dynamics simulations. An identical set of parameters was selected, including the solvation in water, neutralization with sodium ions, temperature of 323 K, and pressure of 1 bar. To explore structural changes and flexibility, we computed the root-mean-square deviations (RMSDs) and fluctuations (RMSFs) of the three-dimensional TaALDH trajectories. RMSDs were calculated by comparing the movement of atoms to initial coordinates at $t = 0$. In both cases, Cα RMSD (Figure 4) indicated that the core protein structure became stable after 4 ns.

A repeating, systematic shift of the Cα RMSD was detected for M42, changing from 0.3 to 0.5 nm. This behavior, potentially indicating a shift between two protein states, was less dominant for M0. Furthermore, the cross RMSD (Supporting Information) revealed a constant change toward another state for M0. For M42, the systematic shift between a closely related state and at least one other state was enhanced. The analysis of the DGA binding revealed a significant decrease of the distance between the DGA Cα and the nucleophilic sulfur of Cys281, dropping from 1.0 ± 0.2 nm for M0 to 0.5 ± 0.1 nm for M42 (Figure 4). Together with the reduced RMSD of DGA compared to that of the protein backbone (see the Supporting Information) and the increased size of the binding pocket, this finding is a plausible

Table 2. Kinetic Constants and Melting Points for Purified TaALDH Variants.[a]

	K_M (NAD^+) [mM]	K_M (D-glycer-aldehde) [mM]	V_{max} (D-Gly) [U/mg]	K_i (D-glycer-aldehyde) [mM]	T_m [°C]
M0	18.4	0.16	0.4	1.25	79.5
M33	1.3	0.14	0.6	3.8	78.5
M42	0.6	0.4	1.5	31.2	72.5

[a]Kinetic measurements were carried out in triplicate at 50 °C containing 100 mM HEPES pH 7.35 and 100 mM ammonium sulfate pH 7. NAD^+ saturation kinetics were determined with 1.3 mM DGA, and DGA saturation measurements contained 1 mM NAD^+ in a final volume of 200 μL. Melting points were detected *via* a thermofluor assay.

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

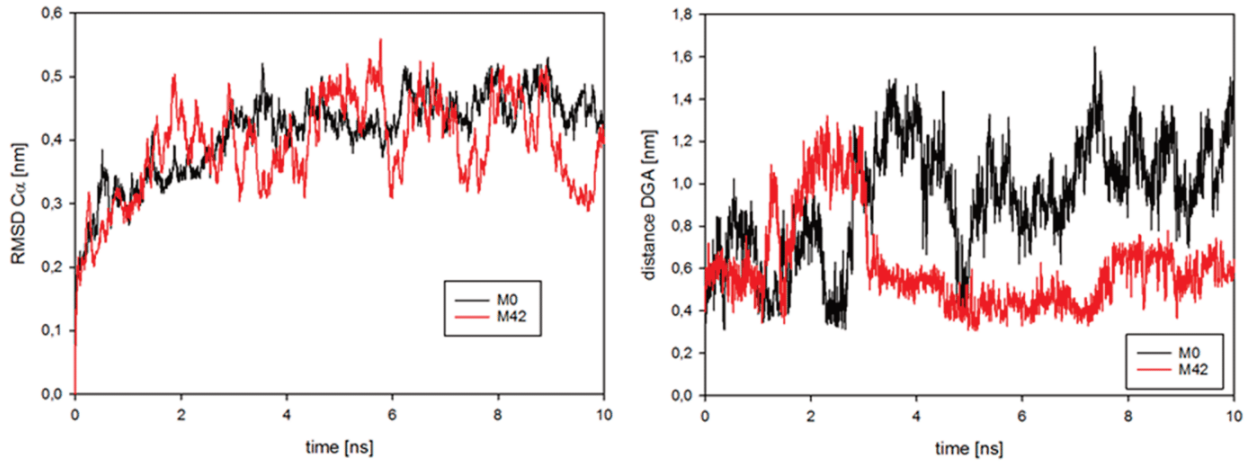


Figure 4. Left: Cα RMSD of M0 and M42 during the molecular dynamics simulation. Right: Distance between Cα of DGA and the nucleophilic sulfur of Cys281 over the course of the simulation.

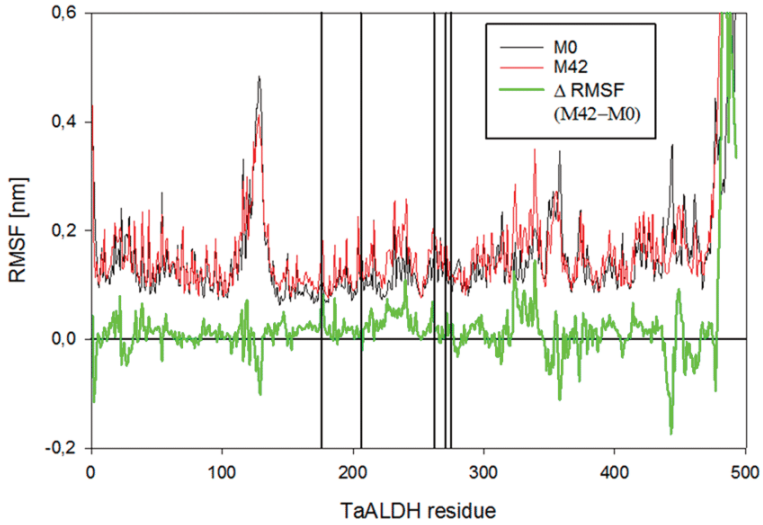


Figure 5. RMSF comparison of M0 and M42 with a residual plot. Vertical lines indicate mutated positions of M42.

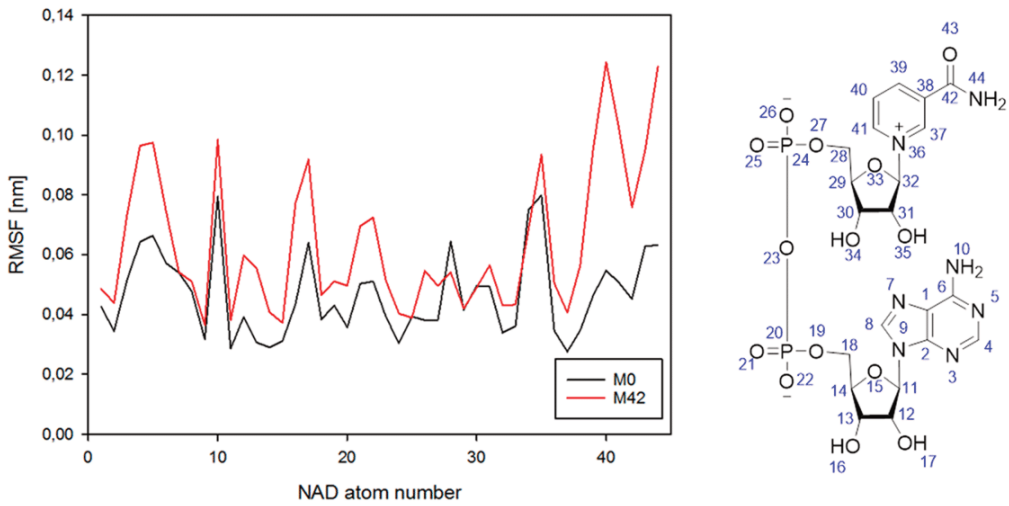


Figure 6. Atom resolved RMSF of NAD^+

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

explanation for the loss of substrate inhibition. In order to compare equilibrated systems, all RMSF values were calculated based on the last 6 ns of the simulation. The two variants show a very similar fluctuation with only subtle changes in the residual plot (Figure 5).

Ignoring the very flexible C-terminal loop region, the general difference is that M42 has increased RMSF in the NAD^+ binding regions and decreased RMSF in the DGA binding regions (as depicted in the Supporting Information). In addition, the mutations at positions 176, 262, 271, and 275 led to a direct increase of the local RMSF, indicating a gain in flexibility. A recent report by Talfournier and co-workers demonstrated the need of a cofactor flip for successful ALDH functionality, which is controlled by a shift of a conserved ELGG loop.[23] This motive can also be found in the TaALDH (positions 247 to 250) and shows increased RMSF after mutation of M0 to M42. The increased flexibility from positions 217 to 251 includes parts of the bottom NAD^+ binding site, hence facilitating a possible cofactor flip. Other MD studies with betaine aldehyde dehydrogenases from *Oryza sativa* also found that the NAD^+ binding region is the most flexible within the entire protein.[24,25] While the distance of the NAD^+ C38 to the catalytically active sulfur of Cys281 stayed at 1.3 ± 0.1 nm for both variants, the RMSD of NAD^+ slightly increased for M42 in comparison to the backbone (see the Supporting Information). Moreover, the RMSF of the NAD^+ cofactor increased in the area of the reactive nicotinamide (Figure 6) due to the mutations.

In summary, our results demonstrate an increased flexibility of the NAD^+ binding site and the cofactor itself for the M42 variant. Our future design principle for aldehyde dehydrogenases will therefore aim at an increase of the cofactor flexibility.

CONCLUSION

In this study, we effectively engineered TaALDH toward increased NAD^+ activity by mutagenesis of rationally selected positions. The 17 positions of interest were identified by a crystal structure analysis and the CSR-SALAD tool. Although designed for a complete cofactor switch, CSR-SALAD demonstrated its applicability in the prediction of specificity determining residues. Individual sequence saturation of each position *via* an NNK motif resulted in six single mutants showing increased volumetric activity. The comprehensive combination of these mutations led to M42 (D176S, S206K, M262I, W271Y, and W275V) as the most active variant, with close to 10-fold increased activity under standard conditions. Comparing the initial variant M0 with M42 by means of molecular dynamics, we identified an increased flexibility at the NAD^+ binding site as a potential cause of the 30-fold improved activity with NAD^+. Furthermore, our analysis revealed initial evidence for another binding state, which seems worth investigating in future studies. The next step will be the transfer of this flexibility-based design toward other oxidoreductases in order to guide future engineering approaches. The kinetic characterization of the initial and final variant also revealed an obliteration of the substrate inhibition by DGA. Based on its improved properties, the new variant M42 is a promising candidate for cell-free metabolic engineering approaches.

METHODS

Chemicals, Strains, and Plasmids. NAD^+ and NADH (disodium salt) were obtained from Carl Roth GmbH (Karlsruhe, Germany), and D-glyceraldehyde (DGA) was bought from Sigma-Aldrich (Darmstadt, Germany). All chemicals were used as received without any further purification.

E. coli BL21 (DE3) (F^- ompT hsdSB (rB^- mB^-) gal dcm (DE3)) was obtained from Novagen (Darmstadt, Germany) and used for library construction, protein production, and screening. The construction of the plasmid coding for the triple mutant aldehyde dehydrogenase from *Thermoplasma acidophilum* (TaALDH) was reported previously.[11,13]

Structural Modeling and Docking. The 3D structure of TaALDH (M0) was modeled by SWISS-MODEL[26−28] with the available crystal structure of the triple mutant TaALDH F34M, Y399C, S405N (PDB code: 5M4X) as a template. Structure validation was done using RAMPAGE for the Ramachandran analysis and ERRAT and Verify 3D from the Structural Analysis and Verification Server (SAVES).[14,17,18] To each tool, the final model of M0 and the input structure 5M4X (without cofactor) were uploaded in order to get comparable data. The docking of NAD^+ and DGA was done with Yasara 15.11.18[29] by applying the docking method VINA. The model structure of M42 for the molecular dynamics analysis was generated by mutating M0 *in silico* with Yasara. For the prediction of mutagenesis targets, the M0 model with docked NAD^+ was uploaded to the online tool CSR-SALAD, which was developed by the Arnold lab for the cofactor switch of $NADP^+$ toward NAD^+.[8] Additionally, M0 with docked NADP was uploaded to the CSR-SALAD in order to compare the both outputs, The maximum library size was set to 850, and the options "rossman fold only", "exclude pyrophosphate-binding residues", and "exclude peripheral residues" were set to false. Together with additional positions of interest from the crystal structure, all mutagenesis targets have been addressed separately with an NNK codon in order to monitor the effects of the individual position.

Construction of Variant Libraries *via* Site-Directed Mutagenesis. Site-directed mutagenesis was performed in a C1000 Thermo Cycler (BIORAD) with Phusion polymerase from New England Biolabs (Frankfurt am Main, Germany) with the supplied high-fidelity buffer in a final volume of 50 μL. Oligonucleotides were obtained desalted from Eurofins Genomics (Ebersberg, Germany). A typical reaction is summarized in the Supporting Information. The PCR conditions were 2 min of initial denaturation at 98 °C, followed by 26 cycles of 10 s at 98 °C, 30 s at 60 °C, and 3.5 min at 72 °C. Final elongation was done for 10 min at 72 °C before 1 μL of 20 U/L DpnI (New England Biolabs, Frankfurt am Main, Germany) was added. The digestion of template DNA was done at 37 °C for at least 12 h, followed by heat inactivation at 80 °C for 20 min and PCR cleanup with a Macherey Nagel PCR cleanup kit (Macherey Nagel, Düren, Germany) according to manufacturer's recommendations. Chemically competent *E. coli* BL21 (DE3) were transformed with the prepared libraries and spread onto LB agar plates containing 50 μg/mL kanamycin.

Screening Procedure for TaAlDH Variants. The expression of variant enzymes was performed according to a modified protocol by Steffler *et al.*[13] All media were supplemented with 100 μg/mL kanamycin, and the screening

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

was done in 96 MTPs containing 86 mutant variants, 7 template enzymes, and 3 media controls. Colonies of *E. coli* BL21 (DE3) containing the TaALDH variants were transferred into F-bottom microtiter plates (Greiner, Frickenhausen, Germany) with a RapidPick Lite colony picker (Hudson Robotics, Springfield, NJ, USA) containing 150 μL LB media. After overnight incubation at 37 °C and 1200 rpm, V-bottom deep well plates (Greiner, Frickenhausen, Germany) containing 600 μL ZYP-5052 autoinduction media[30] were inoculated with 1.7% v/v of preculture and shaken overnight at 30 °C and 1200 rpm. Prior to harvesting, backup plates were prepared by mixing 30 μL of expression culture with 30 μL of 50% glycerol and freezing at −80 °C. The expression cultures were centrifuged at 4200*g* for 15 min, and the supernatant was removed. The cell pellets were frozen at −20 °C before 200 μL of lysis buffer (1 g/L lysozyme and 10 mg/L DNaseI in 100 mM HEPES pH 7.35) was added, followed by incubation at 37 °C for 60 min and an additional heat step at 60 °C for 60 min. Insoluble cell debris was removed by centrifugation at 4200*g* for 15 min, and the supernatants containing the soluble variant enzymes of TaAlDH were tested for activity, as described below. The single saturation mutagenesis libraries were analyzed with at least 80 clones; more than 2500 clones were screened for the double saturation W275 + W271. For the simultaneous saturation of 271 and 275, the colonies were grown as described above and the lysates of four 96 MTPs were combined into one 384 MTP for the activity assay.

The two most active variants per position showing at least 20% increased volumetric activity were regrown from backup plates, plasmid purified, and sent for sequencing (Eurofins Genomics, Ebersberg, Germany). For the lysate tests of combined variants with 1 mM NAD^+, 5 mM NAD^+, 3% v/v isobutanol, or 150 μM $NADP^+$, 10 independently grown colonies harboring the corresponding mutant plasmid were randomly picked and treated as described above. For the activity assay, the heat purified enzyme was diluted three or five times.

Activity Assays. Unless mentioned otherwise, the experiments were carried out in triplicate at room temperature using 96-well plates (Greiner F-bottom) containing a final volume of 200 μL. The necessary ammonia for the reductive amination of pyruvate was supplied with 100 mM ammonium sulfate, and all reactions contained 100 mM HEPES buffer pH 7.35 and various amounts of NAD^+ and DGA and were started with the addition of the appropriate amount of enzyme. The reaction velocity was monitored photometrically following NADH release at 340 nm with an Epoch 2 Microplate spectrophotometer (BioTek GmbH, Bad Friedrichshall, Germany). The mutant screening assay solution contained 1.3 mM DGA and 1 mM NAD^+, and the additional assays for the cofactor acceptance contained either 1 mM or 5 mM NAD^+ or 0.15 mM $NADP^+$. Reactions with 3% v/v of isobutanol contained 1 mM NAD^+. The final screening of M33 with 271/275 NNK was performed in 384 MTPs (F-bottom Greiner) in a final volume of 60 μL with the above-described substrate concentrations. Michaelis−Menten constants were determined at 50 °C by measuring eight different substrate concentrations, plotting initial velocities *vs* substrate concentration, and fitting the parameters with Sigmaplot 13.0 (Systat Software GmbH, Erkrath, Germany). The parameters for DGA dependence were fitted with the equation for substrate inhibited enzymes. One unit (1 U) is defined as the amount of enzyme activity producing 1 μmol of substrate per minute.

Kinetic measurements of TaALDH contained either 0.5−30 mM NAD^+ (1.3 mM DGA) or 0.05−5 mM DGA (1 mM NAD^+).

Shaking Flask Expression and Upscaled Experiments. The upscaled expression and purification of M0, M33, and M42 was done as described above in 10 mL incubation tubes containing 5 mL of media. Protein concentration was determined *via* a Bradford protein assay by applying the Roti-Nanoquant reagent (Carl Roth GmbH, Karlsruhe, Germany) according to the manufacturer's recommendations with bovine serum albumin as standard. The melting point of the protein was determined with a SYPRO Orange Protein Gel Stain (Sigma-Aldrich, Darmstadt, Germany) thermofluor assay in a CFX 96 Real-Time PCR thermocycler (Biorad, Feldkirchen, Germany) containing 25 μL of total volume.[31]

Molecular Dynamics. The simulations were carried out with GROMACS 2019[32−34] with the modified gromos54a7 force field from the automated topology builder (atb) and repository version 3.0.[35,36] The parameter sets for the ligands NAD^+ and DGA were also obtained from atb. The protein−ligand complex was centered in a dodecahedral box with 1 nm distance and periodic boundaries. Solvation was done with TIP3P water molecules before the system was neutralized with five or three sodium ions for M0 or M42, respectively. The particle mesh Ewald (PME)[37] method with cubic interpolation, a Fourier spacing of 0.16 nm, and a short-range electrostatic cutoff of 1.2 nm were used for electrostatic treatment. Steepest descent energy minimization was done to remove bad contacts until the maximum force was <10 kJ/mol. The constant temperature ensemble simulation (NVT) and the constant pressure ensemble simulation (NPT) were carried out for 100 ps, and the non-hydrogen atoms of the protein and the two ligands were restrained by a force constant of 1000 kJ mol^{-1} nm^{-2}. The bonds to hydrogen were restrained in all simulations by the lincs[38] algorithm with iter 1 and order 4. The temperature coupling was done using a v-rescale[39] thermostat at 323 K with a coupling constant of 0.1 ps and the coupling groups protein−ligands and water−ions. The Berendsen algorithm at 1 bar with a coupling constant of 2 ps was used for pressure coupling. In the final 10 ns production run, the Parrinello−Rahman algorithm[40] was used for pressure coupling, time steps of 1 fs were used for integration, and the coordinates were recorded every 5 ps. Simulation trajectories were analyzed using GROMACS tools. XRMSD plots were generated using MDplot.[41]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.9b00527.

Discussions of calculations of screening coverage, DNA sequences, application of CSR-SALAD, tables of library coverage, comparison of CSR-SALAD results, results of Ramachandran analysis, typical site-directed mutagenesis protocol, lists of oligonucleotides, and amino acids at positions 271 or 275, and figures of CSR-SALAD output, overlay of M0 model, mutant landscapes, 3D structural models, kinetic curves, RMSD plots, and structural visualization (PDF)

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

AUTHOR INFORMATION

Corresponding Author

Volker Sieber − *Chair of Chemistry of Biogenic Resources, Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, D-94315 Straubing, Germany; Catalysis Research Center, Technical University of Munich, Garching 85748, Germany; Bio-, Electro- and Chemocatalysis (BioCat) Branch, Fraunhofer Institute of Interfacial Biotechnology (IGB), Straubing 94315, Germany; School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD 4072, Australia;* orcid.org/0000-0001-5458-9330; Email: sieber@tum.de

Authors

Tobias J. Gmelch − *Chair of Chemistry of Biogenic Resources, Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, D-94315 Straubing, Germany*

Josef M. Sperl − *Chair of Chemistry of Biogenic Resources, Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, D-94315 Straubing, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.9b00527

Author Contributions

T.J.G., J.M.S., and V.S. designed the study. T.J.G. performed the experiments and simulations. T.J.G., J.M.S., and V.S. analyzed the data. T.J.G., J.M.S., and V.S. wrote and reviewed the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research under grant agreement number 031B0071A. The authors would like to thank Magdalena Haslbeck and Manuel Döring for their technical assistance during the screening.

REFERENCES

(1) Rouhani, M., Khodabakhsh, F., Norouzian, D., Cohan, R. A., and Valizadeh, V. (2018) Molecular dynamics simulation for rational protein engineering: Present and future prospectus. *J. Mol. Graphics Modell. 84*, 43−53.

(2) Parra-Cruz, R., Jäger, C. M., Lau, P. L., Gomes, R. L., and Pordea, A. (2018) Rational Design of Thermostable Carbonic Anhydrase Mutants Using Molecular Dynamics Simulations. *J. Phys. Chem. B 122*, 8526−8536.

(3) Chmiela, S., Sauceda, H. E., Müller, K.-R., and Tkatchenko, A. (2018) Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun. 9*, 3887.

(4) Beier, A., Bordewick, S., Genz, M., Schmidt, S., van den Bergh, T., Peters, C., Joosten, H.-J., and Bornscheuer, U. T. (2016) Switch in Cofactor Specificity of a Baeyer-Villiger Monooxygenase. *ChemBioChem 17*, 2312−2315.

(5) Borlinghaus, N., and Nestl, B. M. (2018) Switching the cofactor specificity of an imine reductase. *ChemCatChem 10*, 183−187.

(6) Pick, A., Ott, W., Howe, T., Schmid, J., and Sieber, V. (2014) Improving the NADH-cofactor specificity of the highly active AdhZ3 and AdhZ2 from Escherichia coli K-12. *J. Biotechnol. 189*, 157−165.

(7) Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., Meinhold, P., Zhang, L., and Arnold, F. H. (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proc. Natl. Acad. Sci. U. S. A. 110*, 10946−10951.

(8) Cahn, J. K. B., Werlang, C. A., Baumschlager, A., Brinkmann-Chen, S., Mayo, S. L., and Arnold, F. H. (2017) A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases. *ACS Synth. Biol. 6*, 326−333.

(9) Cui, D., Zhang, L., Jiang, S., Yao, Z., Gao, B., Lin, J., Yuan, Y. A., and Wei, D. (2015) A computational strategy for altering an enzyme in its cofactor preference to NAD(H) and/or NADP(H). *FEBS J. 282*, 2339−2351.

(10) King, Z. A., and Feist, A. M. (2013) Optimizing Cofactor Specificity of Oxidoreductase Enzymes for the Generation of Microbial Production Strains—OptSwap. *Ind. Biotechnol. 9*, 236−246.

(11) 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., and Sieber, V. (2012) Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem 5*, 2165−2172.

(12) Steffler, F., and Sieber, V. (2013) Refolding of a thermostable glyceraldehyde dehydrogenase for application in synthetic cascade biomanufacturing. *PLoS One 8*, No. e70592.

(13) Steffler, F., Guterl, J.-K., and Sieber, V. (2013) Improvement of thermostable aldehyde dehydrogenase by directed evolution for application in Synthetic Cascade Biomanufacturing. *Enzyme Microb. Technol. 53*, 307−314.

(14) Lovell, S. C., Davis, I. W., Arendall, W. B., Bakker, P. I. W. d., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Struct., Funct., Genet. 50*, 437−450.

(15) Bertini, I., Cavallaro, G., Luchinat, C., and Poli, I. (2003) A use of Ramachandran potentials in protein solution structure determinations. *J. Biomol. NMR 26*, 355−366.

(16) Honarparvar, B., and Skelton, A. A. (2015) Molecular dynamics simulation and conformational analysis of some catalytically active peptides. *J. Mol. Model. 21*, 100.

(17) Colovos, C., and Yeates, T. (1993) ERRAT: an empirical atom-based method for validating protein structures. *Protein Sci. 2*, 1511−1519.

(18) Bowie, J. U., Luthy, R., and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*, 164−170.

(19) Koppaka, V., Thompson, D. C., Chen, Y., Ellermann, M., Nicolaou, K. C., Juvonen, R. O., Petersen, D., Deitrich, R. A., Hurley, T. D., and Vasiliou, V. (2012) Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. *Pharmacol. Rev. 64*, 520−539.

(20) Acevedo-Rocha, C. G., Reetz, M. T., and Nov, Y. (2015) Economical analysis of saturation mutagenesis experiments. *Sci. Rep. 5*, 10654.

(21) Zhang, Chung, and Oldenburg (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screening 4*, 67−73.

(22) Brezovsky, J., Babkova, P., Degtjarik, O., Fortova, A., Gora, A., Iermak, I., Rezacova, P., Dvorak, P., Smatanova, I. K., Prokop, Z., Chaloupkova, R., and Damborsky, J. (2016) Engineering a de Novo Transport Tunnel. *ACS Catal. 6*, 7597−7610.

(23) Rahuel-Clermont, S., Bchini, R., Barbe, S., Boutserin, S., André, I., and Talfournier, F. (2019) Enzyme Active Site Loop Revealed as a Gatekeeper for Cofactor Flip by Targeted Molecular Dynamics Simulations and FRET-Based Kinetics. *ACS Catal. 9*, 1337−1346.

(24) Baicharoen, A., Vijayan, R., and Pongprayoon, P. (2018) Structural insights into betaine aldehyde dehydrogenase (BADH2) from Oryza sativa explored by modeling and simulations. *Sci. Rep. 8*, 12892.

(25) Jiamsomboon, K., Treesuwan, W., and Boonyalai, N. (2012) Dissecting substrate specificity of two rice BADH isoforms: Enzyme kinetics, docking and molecular dynamics simulation studies. *Biochimie 94*, 1773−1783.

(26) Bienert, S., Waterhouse, A., Beer, T. A. P. d., Tauriello, G., Studer, G., Bordoli, L., and Schwede, T. (2017) The SWISS-MODEL

Repository-new features and functionality. *Nucleic Acids Res. 45*, D313−D319.
(27) Guex, N., Peitsch, M. C., and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis 30*, S162−S173.
(28) Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, T. A. P. d., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res. 46*, W296−W303.
(29) Krieger, E., and Vriend, G. (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics 30*, 2981−2982.
(30) Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif. 41*, 207−234.
(31) Boivin, S., Kozak, S., and Meijers, R. (2013) Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression Purif. 91*, 192−206.
(32) Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1−2*, 19−25.
(33) Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model. 7*, 306−317.
(34) Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 845−854.
(35) Malde, A. K., Zuo, L., Breeze, M., Stroet, M., Poger, D., Nair, P. C., Oostenbrink, C., and Mark, A. E. (2011) An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput. 7*, 4026−4037.
(36) Stroet, M., Caron, B., Visscher, K. M., Geerke, D. P., Malde, A. K., and Mark, A. E. (2018) Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. *J. Chem. Theory Comput. 14*, 5834−5845.
(37) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys. 103*, 8577−8593.
(38) Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem. 18*, 1463−1472.
(39) Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684−3690.
(40) Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. (Melville, NY, U. S.) 52*, 7182−7190.
(41) Margreitter, C., and Oostenbrink, C. (2017) MDplot: Visualise molecular dynamics. *R J. 9*, 164.

https://dx.doi.org/10.1021/acssynbio.9b00527
ACS Synth. Biol. 2020, 9, 920−929

4 Discussion

4.1 *In Vitro* vs. *in vivo* L-alanine production from D-glucose

In vivo, the standard glycolysis (Embden-Meyerhof-Parnas pathway, black in Figure 9 A), requires ten enzymes for the degradation of D-glucose towards pyruvate. L-alanine can then be obtained from an aminotransferase reaction involving glutamate as amino donor. This pathway evolved naturally in order to supply the organism with energy in form of adenosine triphosphate (ATP) and reduction equivalents NADH.

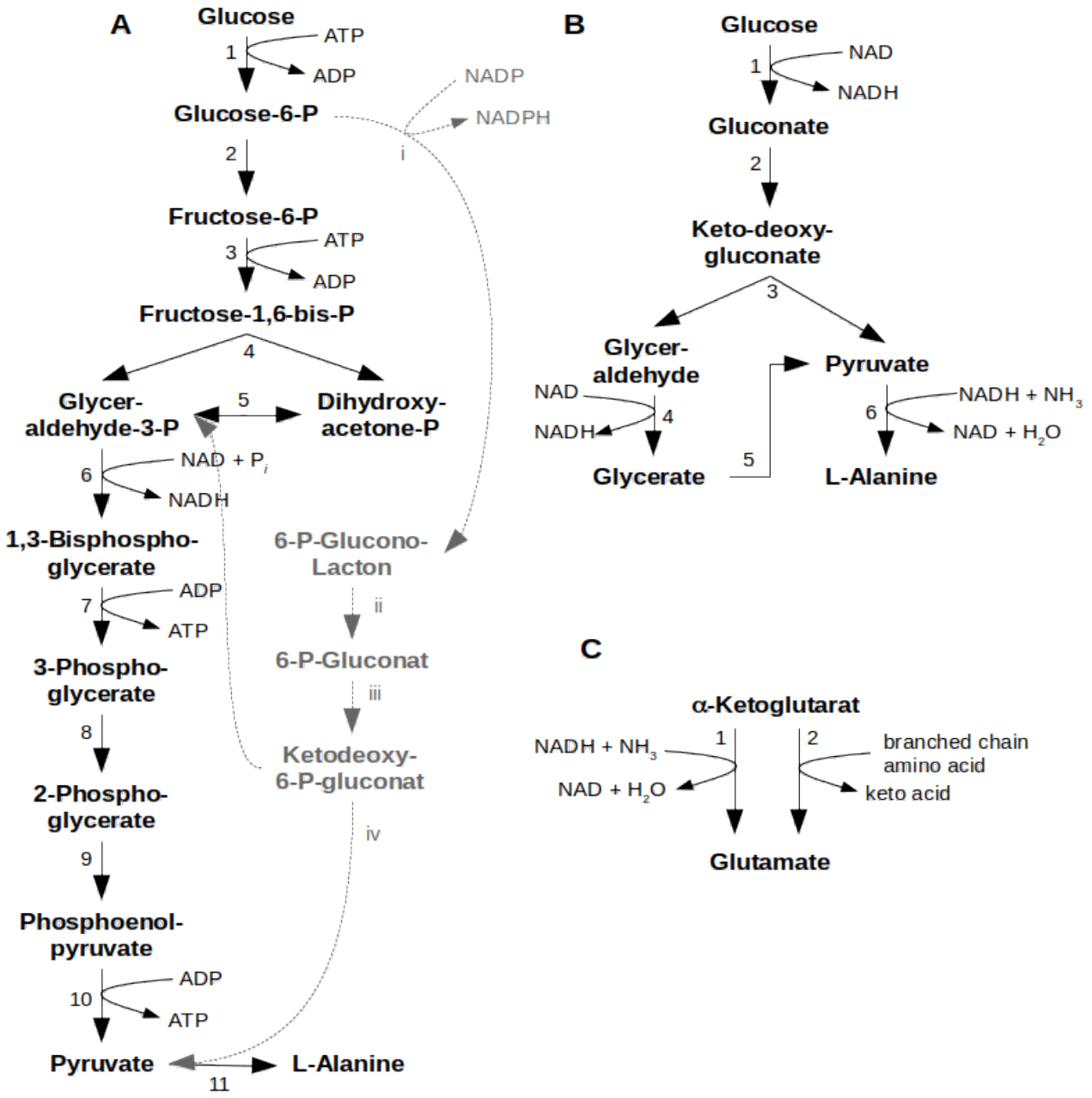


Figure 9: Natural and artificial pathways from D-glucose to L-alanine. (A) The natural pathway in mammals (in black) consisting of at least eleven enzymes: 1 hexokinase, 2 phosphoglucose isomerase, 3 Phosphofructokinase, 4 Fructosebisphosphate aldolase, 5 Triosephosphate isomerase, 6 Glyceraldehyde-3-phosphate dehydrogenase, 7 Phosphoglycerate kinase, 8 Phosphoglycerate mutase, 9 Enolase, 10 Pyruvate kinase, 11 Alanine transaminase requiering glutamate as amine donor. In light grey is the Entner-Doudoroff pathway common archea: i glucose-6-phosphate dehydrogenase, ii 6-Phosphoglucolactonase, iii Phosphogluconate dehydratase, iv Ketodeoxyphosphate aldolase, (B) The cell free artificial pathway established in this work involves only six enzymes: 1 Glucose dehydrogenase, 2 Dihydroxyacid dehydratase, 3 Ketodeoxy aldolase, 4 Glyceraldehyde dehydrogenase, 5 Dihydroxyacid dehydratase, L-Alanine dehydrogenase. (C) Formation of the amin-donor glutamate from either 1 dehydrogenase or 2 transaminase reaction.

Although two ATP-consuming phosphorylation steps are necessary in the beginning of the glycolysis, the breakdown into pyruvate returns 4 equivalents ATP and two equivalents NADH for each glucose molecule.

Alternatively, some bacteria and archaea developed a pathway, that avoids the second phosphorylation step by an oxidation of the glucose-6-phosphate. Termed Entner-Doudoroff pathway (grey in Figure 9 A), this modified glycolysis reduces the overall ATP yield from two to one equivalent. Obviously less efficient, the major advantage of this pathway is the possibility to metabolize gluconate or similar organic acids.

In both cases the alanine biosynthesis requires glutamate as amino-shuttle, which is commonly derived from α-ketoglutarate by either dehydrogenase or transferase activity (Figure 9 C). This dependence on glutamate further complicates the production pathway and metabolic engineering approaches usually rely on an alanine dehydrogenase (*alaD*) for the direct reductive amination of pyruvate instead of a transferase. Further genetic engineering is then required in order to channel the carbon flux towards pyruvate and eliminate common side products such as acetate, ethanol or the D-alanine.

The cell free enzymatic cascade designed in this work, is not limited by such drawbacks. As presented in Figure 9 B, only six enzymes and a single cofactor are necessary for the L-alanine synthesis from glucose. Since no metabolism needs to be maintained, the ATP yield is of no interest. Au contraire, without phosphorylation of the intermediates, the cascade steps can be drastically reduced. Derived from the semi-phosphorylated Entner-Doudoroff pathway, the reaction sequence could in theory occur within living organisms.[150] The only artificial reaction is the direct dehydration of glycerate towards pyruvate. The applied dihydroxyacid dehydratase from *Sulfolobus solfataricus* (SsDHAD) was found to catalyze this reaction with a maximum velocity of about 10 mU/mg.[145] Because the same enzyme also catalyzes the dehydration of gluconate, the first experiments contained only this dehydratase (similar to the ethanol cascade of Guterl *et al*).[6] Due to the rather slow turnover rate of the SsDHAD for both reactions, archaeal gluconate dehydratases of *Sulfolobus solfataricus (SsGAD)* and *Picrophilus torridus (PtGAD)* were investigated as possible substitute. The thermophilic origin of these enzymes combined with the high activity on gluconate, render these variants very interesting.[151,152] Unfortunately, an expression of the active protein from the preferred expression host *E. coli* was not possible and the proteins were always found in the

insoluble cell debris with no detectable activity. During this work, a dehydratase from the mesophilic organism *Caulobacter crescentus* (CcDHAD) was identified,[153] which was finally chosen for the gluconate dehydration due to the 40 times increased activity compared to the SsDHAD. Since the CcDHAD did not show significant activity for the bottleneck substrate glycerate a double dehydratase set up (CcDHAD for gluconate and SsDHAD for glycerate) was established for this cascade. After this decision was done, Sutiono *et al.* reported dehydratase activity on glycerate by the dehydratase from *Paralcaligenes ureilyticus* (PuDHT). With > 100-fold higher activity for glycerate compared to SsDHAD and more than 3 times increased activity on gluconate compared to CcDHAD, this enzyme appears to be an ideal choice for future experiments.[154]

Once the dehydratase system was established, the focus was shifted towards the cofactor acceptance of the aldehyde dehydrogenase. The initially deployed variant from *Thermoplasma acidophilum* (TaALDH) showed a preference for the phosphorylated cofactor NADP.[146,155] Since NAD^+ was to be used as single cofactor, a two-ponged strategy was followed in order to optimize this enzyme: A) random screening of ALDHs for glyceraldehyde and NAD^+ activity and B) engineering the cofactor specificity of the TaALDH (see section 4.2). Screening of 21 enzyme variants available at our chair, identified the ALDH of *Geobacillus stearothermophilus* (GsALDH) as interesting alternative for the TaALDH.[156] Although showing improved specificity towards the cofactor NAD^+, further investigation revealed a rather high K_M for the substrate glyceraldehyde. In the end, the variant from methanocaldococcus jannaschii (MjALDH) was identified from literature,[157] tested and applied within the cell free cascade due to increased expression yield and cofactor specificity.

With finalization of the identification of the crucial biocatalysts, the next step was to analyze the reaction cascade and develop the process for the alanine production. For this, a reliable, fast and quantitative detection method for alanine was required. Usually precolumn derivatization with e.g. o-phthalaldehyde and separation/detection via HPLC-DAAD are the state of the art for amino acid quantification.[158] Nevertheless several colorimetric assays have been investigated in order to increase the sample throughput. Due to the presence of ammonium in the matrix, satisfying results were only obtained with fluorescamine derivatization, where the increased reaction velocity for the formation of the alanine intermediate led to a selective derivatization.[148] The initial process parameters for pH, buffer, temperature, cofactor and substrate concentration were adapted from the ethanol cascade of Guterl *et al.*[6] From there an empirical approach with step wise variation of single parameters resulted in the reported optimized parameters.[147]

The direct comparison of the previously published ethanol cascade to the herein developed alanine cascade reveals identical product yields of 58 %. Although the isolated biocatalysts of the novel alanine cascade show improved kinetic parameters, the combination could not improve the product yield. Since the only unchanged enzyme was the SsDHAD, the bottleneck situation of the glycerate dehydration was demonstrated again.

After all, with a yield greater than 50% a successful redirection of the produced glyceraldehyde towards pyruvate could be demonstrated. With space time yields of 0.17 g/(Lh) and a catalyst load of 0.05 g/g alanine, the productivity of the presented cascade is

still far behind the industrial production via immobilized cells of *E. coli* and *P. dacunhae*, where 13.4 g/(Lh) are achieved.[76] Nevertheless, with further optimization of the glycerate dehydration, the next goal will be to beat the fermentation yield of 121 g/L alanine in 39 h, which corresponds to a space time yield of 3.1 g/(Lh).[60]

4.2 Engineering the cofactor specificity of the glyceraldehyde dehydrogenase of *Thermoplasma acidophilum*

In addition to the screening of available ALDHs, the initially deployed TaALDH was engineered towards the unnatural cofactor NAD^+. Previously, Steffler *et al.* increased the protein yield from the heterolog expression in *E. coli* by refolding[155] and achieved a 55 fold increase of volumetric activity after several rounds of directed evolution.[146] The final triple mutant F34M-Y399C-S405 (here termed “M0”) was the basis for the rational design presented in this work.

First, the screening method used by Steffler *et al.* was evaluated and optimized towards improved reliability. This was done recording wildetype profiles[159] and calculating the Z-factor.[160] Although Steffler *et al.* directly inoculated autoinduction media with a single colony, experiments with a preculture in LB media showed an improved activity distribution on wildtype plates. For the cell lysis, a mixture of DNaseI and Lysozyme was used instead of the ready to use BPER-reagent. Together with a novel tip-wash program for the Tecan automated liquid handling platform, the costs of consumables for the screening could be pushed below 0.3 €/mutant. For further throughput optimization, the final activity measurement was adapted to the 384 mtp format which led to a maximum screening capacity of around 3000 mutants (plus controls) per 8 h.

In order to identify positions of interest for a rational design, an accurate 3D-model of the target protein is essential. Although the crystal structures of the TaALDH wildtype, M0 and another mutant variant were published before,[161] the 3D-model deployed here was obtained by alignment of the full amino acid sequence to the crystal structures via SWISS-MODEL. Due to 32 non represented amino acids in the crystal structure, this step was essential for a high quality model. In order to assess the quality of the obtained structure for M0, a comparison to the template (5M4X) as well as to the most similar crystal structure 2IMP was conducted (Table 8). First the Ramachandran analysis with the RAMPAGE online tool gave one outlier residue for 5M4X, two outliers for 2IMP and 4 outlier residues for M0.[162] Generally the results of the Ramachandran investigation is comparable to other values obtained for ALDHs.[163] In the next step, the acceptance criteria (quality factor > 90) for the non-bonded atomic interactions analysis with the ERRAT server was exceeded for all three structures.[164] Although, the surrounding of the active center (I282 - A284) was found to be of poor quality, the specificity determining residues are sufficiently distant and in the high quality range. Finally, the Verify 3D server was used, which evaluates accuracy by comparing the input structure to its own amino acid sequence, using a three dimensional profile generated from atomic coordinates of the 3D structure.[165] Here, the acceptance criteria is > 80% of all residues need to score ≥ 0.2 which was only met for the M0 model. Possibly the SWISS 3D modeling included changes into the template structure in order to optimize this folding score.

Table 8: Comparison of the M0 model to the template crystal structure 5M4X and the most similar crystal structure 2IMP (39% identity, 60% positives compared to M0)

	5M4X crystal structrue	M0 model	2IMP crystal structure
Represented amino acids	461/ 493	493/ 493	479/ 483
RAMPAGE	97.8% favored region 2.0% allowed region 0.2% outlier region	94.1% favored region 5.1% allowed region 0.8% outlier region	94.9% favored region 4.7% allowed region 0.4% outlier region
ERRAT server	Quality factor 97.5	Quality factor 92.4	Quality factor 96.1
Verify 3D server	73 % residues with score ≥ 0.2	91 % residues with score ≥ 0.2	77 % residues with score ≥ 0.2
Containing cofactor	none	NAD	2x SO4, lactaldehyde 2 conformations for NAD
RMS to M0	0.860	n.a.	0.976

In the next step, the Rossmann fold was identified within the structure. As described in section 1.3.2, the cofactor binding site is composed of five β-strands connected by four α-helices.[102] The strand order is 3-2-1-4-5 and a modified version of the typical Rossmann fingerprint motive $G_IX_{1-2}G_{II}XXG_{III}$ is found in the loop region right after β3. The insertion of an additional amino acid in front of G_{III} of the fingerprint to G_{203}RGSEIG might further enhance the NADP specificity.[101] Further proof of the NADP preference is the absence of glutamic acid within the binding pocket, which commonly forms hydrogen bonds to 2' and 3'- hydroxyls of the adenine ribose.[166]

The docking of the cofactor was done with yasara and the final structure was uploaded to the CSR-SALAD[124] tool in order to predict the specificity determining residues important for the cofactor switch towards NAD. Although this platform was initially tested with NADP containing structures, the potential of this tool was further investigated by uploading structures containing either NAD or NADP. A comparison of the results showed, that most of the positions have been found for both cofactors. While the key selectivity determining residues for the NAD containing structure where K172, S174 and G205, for die NADP containing structure it was L145 and K172. Since the K172 position is reported to be essential in ribose binding, it was rather surprising that it was identified in both structures.[82] Most likely the algorithm suggested this position in order to remove the only positively charged residue within reach of the phosphorylated ribose motive of NADP. Even more interesting is the recommendation of G205 which is part of the modified fingerprint motive. Potentially the tool recognized the impact on the NADP specificity by this modified form and suggested the mutation towards the typical fingerprint. Unfortunately, there are no detailed explanations on why specific positions are selected by the tool, or why the selectivity determining residues are dependent on the cofactor which is present in the uploaded structure.

Nevertheless, the residues recommended for the NAD containing structure (including the positions for activity recovery) were saturated separately with an NNK-codon. Together with further hand picked positions, in total 17 positions have been investigated. From these, only mutations at the positions S175, D176, S206, M262, W271 and W275 could

increase the volumetric activity during the screening. Interestingly, the screening results could be categorized into two groups: While S175, D176, S206, M262 show one to three hits exceeding the hit limit with similar activity, the two tryptophan mutations resemble an activity distribution with multiple hits of increased activity. Sequencing of the two most active variants per plate further proved that the first group is based on specific amino acid exchanges, where only one amino acid was found to be beneficial (S175E, D176S, S206K and M262I). With the introduction of glutamic acid at position 175, the for NAD preferring ALDHs essential residue which forms hydrogen bonds to 2' and 3'- hydroxyls of the adenine ribose is established.[166] Analog to this discovery, Zhang *et al,* who were working on ALDH from *Vibrio harveyi* also identified beneficial effects on the NAD acceptance by introducing the glutamic acid residue in the vicinity of the ribose moiety. With their T175E mutation, they could realize a 7-fold increase of the NAD-specificity and simultaneously reduce the $NADP^+$ specificity by a factor of 2000.[167] Interestingly, the introduction of D176S resulted in the second most active single mutant, but in combination with the S175E, the activity is almost restored to the initial activity of M0. Apparently, the double serine functionality enables similar cofactor binding than the introduction of glutamic acid, but with a neighboring serine the hydrogen bonding capacity of glutamic acid seems to be reduced.

The second group of beneficial positions, W271 and W275, showed even higher potential for increasing the volumetric activity. Several amino acids were found to improve the enzyme and by combining this group to the first group with specific mutations, a dependence on the genetic background was identified. For example the D176S mutant showed highest activity with W271Y and W275S, while M0 was most active by including W271Y and W275T. This promiscuity was already discovered by Steffler *et al.* but the mutants containing either W271S or W271R were not further characterized.[146] Located on the entrance to the substrate tunnel, these residues might act as gate keeper for glyceraldehyde. Changing size and the chemistry of these positions effectively modulates the substrate specificity and the increased volumetric activity is possibly the result of changes in the substrate inhibition. Damborsky *et al*. also identified a gatekeeper tryptophan for their ALDH, which showed direct influence on the inhibition pattern.[168] Furthermore, Keller *et al.* proposed a tremendous effect of a C-terminal helix on the substrate specificity due to trans-protein effects.[169] Since the C-terminal region tends to crystallize poorly, this effect cannot be properly investigated for the TaALDH. The two large tryptophanes (W271 and W275) might interact with the C-terminus of another protein within the tetramer, hence enabling substrate control.

With this background, the most active combination of the identified mutations was obtained by combining the positions of the first group in a combinatorial way and saturate this variant simultaneously at W271 and W275. With this approach the mutant D176S S206K M262I W271Y W275V (termed M42) was found to be most active, showing three-fold increased specific activity, thirty fold decreased K_M for NAD^+ and a complete loss of the glyceraldehyde inhibition. Furthermore, this engineered variant TaALDH shows more than 10 times improved activity in the presence of 3% isobutanol which renders it a promising candidate for the cell free ethanol/ isobutanol cascade.

4.3 Molecular dynamics analysis for beneficial mutations

Molecular dynamics (MD) analysis is a useful tool for the investigation of molecular motion of e.g. binding or folding processes. Here, the simulation of the engineered TaALDH M42 was compared to the template M0 in order to identify the beneficial effects of the introduced mutations and relate them to the kinetic improvements.

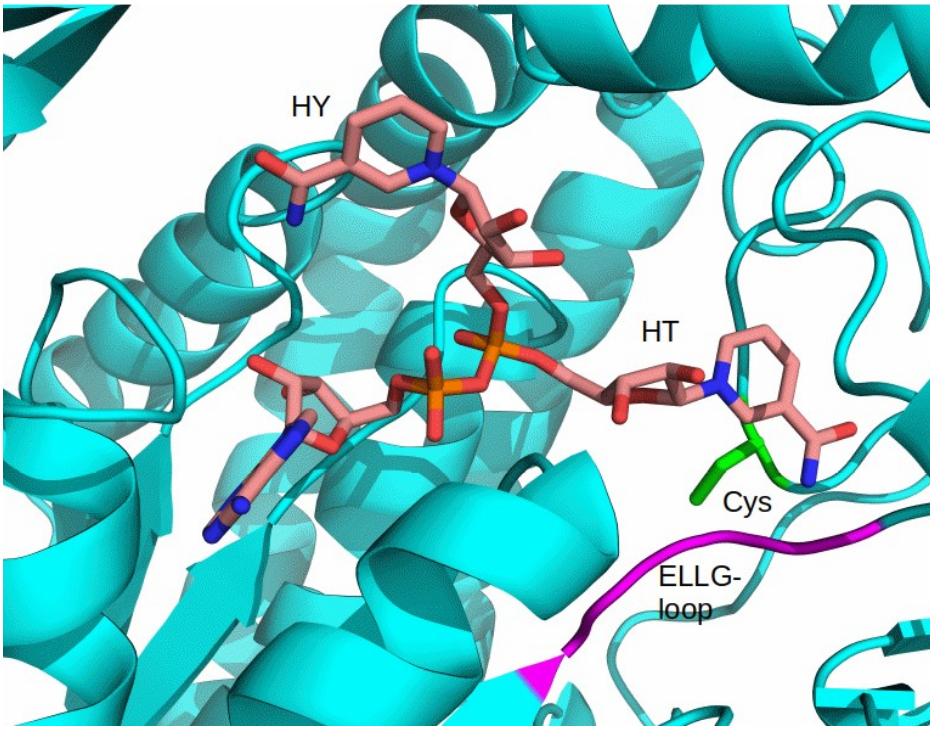


Figure 10: Representation of the hydrolysis (HY) and the hydrid transfer (HT) mode of NAD in the crystal structure of lactaldehyde dehydrogenase from E. coli (2IMP). The catalytically active cystein is colored in green and the ELLG-loop in magenta.

For the analysis of the ALDH-NAD interaction it is important to consider the two different conformations of bound NAD. The binding process is considered to start in the "hydrolysis" mode (HY), where the nicotinamid is facing away from the acitve site. A subsequent change to the active "hydride transfer" mode (HT) enables the reception of the hydride from the thiohemiacetal intermediate. Finally, another conformational change to the HY mode needs to occur before the reduced cofactor is released.[135] Figure 10 shows the most similar, available crystal structure (2IMP) compared to M0 with cocrystalized NAD in the two possible binding conformations. As described above, the specificity determining adenosine-ribose stays within the binding pocket and the reactive nicotinamide is flipped in- and out of the reactive center.

Since NAD was accepted to some extend by the initial variant M0 of the TaALDH, the binding step was omitted and the focus was on the active state with NAD in the HT mode. With this approach, a tedious comparison of hydrogen bond patterns was avoided and the attention was shifted towards the hydride transfer itself. For the MD simulations, the 3D-model for M0 containing NAD was engineered *in silico* towards M42 and both structures were simulated for 10 ns with an identical set of parameters. A similar approach of comparing two enzyme variants via molecular dynamics was previously done by e.g. Boonyalai *et al.* who were working on the substrate specificity of betaine aldehyde

dehydrogenases. In their case, the bonding step had the highest priority and with the identification of specific hydrogen bonding patterns they could explain differences in substrate specificity.[170] The more general approach in this work should result in a basic understanding of the introduced mutations in the first step and possibly predict beneficial mutations in future experiments.

Because M42 is a variant of M0, the 10 ns MD simulations of the two structures with the NAD cofactor showed to be similar. The RMSF revealed subtle changes which pointed towards an increased flexibility of the NAD binding regions and decreased flexibility in the glyceraldehyde (DGA) binding region of M42. Furthermore, the nicotinamide ring shows increased RMSF, also indicating higher fluctuation of the active site of the cofactor. Since the cofactor needs to flip in and out of the active center, the increased flexibility of the binding pocket is basically supporting this movement and generally indicates an increased reaction turnover. Talfournier *et al*. were also working on MD simulations of ALDHs, and they identified the ELGG loop, which possibly controls the cofactor switch between HT and HY mode.[141] This loop is also present in TaALDH (starting with E247) and shows increased flexibility for M42, further supporting the reaction turnover. Summarizing this, the focus on the bound NAD in the active HT state resulted in overall increased molecular motion of the binding pocket for M42 containing the unnatural cofactor.

Concerning the substrate binding site, the gain in rigidity of M42 is accompanied by a closer and less fluctuating position of DGA to the active C281. In contrast to the cofactor, which needs to flip during the reaction, the substrate needs only minimal reorganization. Hence this finding indicates a stabilization of the active center. Similar results were obtained by Madhusudhana Rao *et al.* who identified increased activity for a more rigid mutant of their lipase.[171]

With this simulation based analysis of the initial and the engineered variant, first indications for the improved activity of M42 could be identified. Nevertheless, the differences between the simulations of M0 and M42 are rather small, suggesting only minor changes in the cofactor reactivity. Future experiments should therefore include the binding process with a focus on the specific hydrogen bonds for the HY state of NAD.

4.4 Future perspectives

The cell free enzymatic cascade from D-glucose to L-alanine was successfully optimized towards efficient enzyme loading. Thereby, the most economic ratio of the involved cascade enzymes of the main branch could be identified, which was found to be GDH/DHAD/KDGA/AlaDH = 2/10/1/2. In contrast to the highly productive main branch via gluconate, keto-deoxy-gluconate and pyruvate which gave almost quantitative L-alanine yield, the side branch via glyceraldehyde, glycerate and pyruvate needs further investigation. It was shown, that the bottleneck here is the very poor activity of SsDHAD for the unnatural substrate glycerate. Alternatively, the artificial glycerate dehydration could be avoided by the redesign of the side branch towards L-serin. In this case, the separation of L-serin and L-alanine would arise as new problem. With the recently discovered DHADs that show increased activity for the dehydration of glycerate, this developed reaction cascade should be reinvestigated with a focus on the side branch productivity.

In terms of further enzyme optimization, the applied glucose dehydrogenase should be used as benchmark concerning the activity and heat stability. Although the glucose specificity could be further increased, as the first enzyme of the reaction sequence, the substrate concentration is far above a limiting level. Since the identified DHAD shows increased activity and substrate specificity, the major limitation is the long term temperature stability, which could be improved in order to further reduce enzyme loading. For the applied KDGA, ALDH and the AlaDH, temperature stability is sufficient, but a general speed up of the reaction kinetics would also improve the economy of this cascade. Although AlaDH could demonstrate high reaction turnover, this will certainly change with increasing Alanine concentrations during upscaling of the cascade. Together with the ammonium specificity, the product inhibition is to be targeted in the next round of engineering.

All of these enzyme optimizations can be targeted by mutagenesis based on a combination of molecular dynamics and rational design, which was already successfully applied here for engineering of an aldehyde dehydrogenase. Increasing the scope of this approach beyond the cofactor specificity of ALDHs, the design guidelines obtained from molecular dynamics should be validated by extensive engineering experiments in the first step. With the refined engineering guidelines, beneficial mutations for e.g. increased NAD^+ specificity of ALDHs could possibly be predicted *in silico* from targeted molecular dynamics simulations. The further development of an analysis server with automated structure handling could further contribute to insights into e.g. redox cofactor binding and possibly be another step towards the *de novo* design of enzymes.

Finally, the efficiency of the presented (and in future optimized) L-alanine process, needs to be upscaled and the reactor design as well as the downstream processing should be included into the picture. Increasing the product concentration above to the solubility limit could be the next step in order to facilitate the product separation and purification, which is essential for an economically feasible process. Further process parameters like temperature, pH, feeding strategy or dwell time need to be investigated in order to establish a competitive process. Additionally, the designed cascade core module to the central intermediate pyruvate could also be adapted towards a variety of other products such as L-valin, for example.

5 Bibliography

1. Kusumoto, I. Industrial Production of L-Glutamine. *J. Nutr.* **131**, 2552S-2555S (2001).
2. Choi, K. R. *et al.* Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering. *Trends Biotechnol.* **37**, 817–837 (2019).
3. Wang, L., Dash, S., Ng, C. Y. & Maranas, C. D. A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol.* **2**, 243–252 (2017).
4. Fessner, W. D. Systems Biocatalysis: Development and engineering of cell-free 'artificial metabolisms' for preparative multi-enzymatic synthesis. *N Biotechnol* **32**, 658–664 (2015).
5. Zhang, Y.-H. P., Sun, J. & Ma, Y. Biomanufacturing: history and perspective. *J. Ind. Microbiol. Biotechnol.* **44**, 773–784 (2017).
6. Guterl, J. K. *et al.* Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem* **5**, 2165–2172 (2012).
7. Opgenorth, P. H., Korman, T. P. & Bowie, J. U. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nat Commun* **5**, 4113 (2014).
8. Nowak, C. *et al.* A water-forming NADH oxidase from Lactobacillus pentosus suitable for the regeneration of synthetic biomimetic cofactors. *Front Microbiol* **6**, 957 (2015).
9. Chenault, H. K., Simon, E. S. & Whitesides, G. M. Cofactor Regeneration for Enzyme-Catalysed Synthesis. *Biotechnol. Genet. Eng. Rev.* **6**, 221–270 (1988).
10. Ali, I., Khan, T. & Omanovic, S. Direct electrochemical regeneration of the cofactor NADH on bare Ti, Ni, Co and Cd electrodes: The influence of electrode potential and electrode material. *J. Mol. Catal. A Chem.* **387**, 86–91 (2014).
11. Willner, I. & Mandler, D. Enzyme-catalysed biotransformations through photochemical regeneration of nicotinamide cofactors. *Enzyme Microb. Technol.* **11**, 467–483 (1989).
12. Hollmann, F., Hofstetter, K. & Schmid, A. Non-enzymatic regeneration of nicotinamide and flavin cofactors for monooxygenase catalysis. *Trends Biotechnol.* **24**, 163–171 (2006).
13. Liu, W. & Wang, P. Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnol. Adv.* **25**, 369–384 (2007).

14. Shi, T., Han, P., You, C. & Zhang, Y.-H. P. J. An in vitro synthetic biology platform for emerging industrial biomanufacturing: Bottom-up pathway design. *Synth. Syst. Biotechnol.* **3**, 186–195 (2018).

15. Korman, T. P., Opgenorth, P. H. & Bowie, J. U. A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun.* **8**, 15526 (2017).

16. Wyman, C. *Handbook on Bioethanol*. (2018).

17. Bai, F. W., Anderson, W. A. & Moo-Young, M. Ethanol fermentation technologies from sugar and starch feedstocks. *Biotechnol. Adv.* **26**, 89–105 (2008).

18. Yamamoto, T., Kato, T., Matsuo, R., Kawamura, Y. & Yoshida, M. Gustatory reaction time to various sweeteners in human adults. *Physiol. Behav.* **35**, 411–415 (1985).

19. Mundy, H. R., Williams, J. E., Cousins, A. J. & Lee, P. J. The effect of L-alanine therapy in a patient with adult onset glycogen storage disease type II. *J Inherit Metab Dis* **29**, 226–229 (2006).

20. Jakubke, H.-D. & Hans, J. *Aminosäuren, Peptide, Proteine*. (Verlag Chemie, Weinheim, 1982).

21. Brazier, M. A. A new method for the separation of the products of protein hydrolysis. *Biochem. J.* **24**, 1188–1198 (1930).

22. Bergmann, M. & Niemann, C. On the structure of silk fibroin. *J. Biol. Chem.* **122**, 577–596 (1938).

23. Stein, W. H., Moore, S., Stamm, G., Chou, C.-Y. & Bergmann, M. Aromatic sulfonic acids as reagents for amino acids: The preparation of L-serine, L-alanine, L-phenylalanine, and L-leucin from protein hydrolysates. *J. Biol. Chem.* **143**, 121–129 (1942).

24. Selim, A. S. M., Ramadan, M. E. A. & El-Sadr, M. M. A new method for the isolation of L-alanine and L-proline from protein hydrolysates. *J. Biol. Chem.* **227**, 871–878 (1957).

25. Robson, W. & Selim, A. S. M. A new technique for the estimation and isolation of the hexone bases in protein hydrolysates. *Biochem. J.* **53**, 431–436 (1953).

26. Strecker, A. Ueber die künstliche Bildung der Milchsäure und einen neuen, dem Glycocoll homologen Körper; *Justus Liebigs Ann. Chem.* **75**, 27–45 (1850).

27. Strecker, A. Ueber einen neuen aus Aldehyd - Ammoniak und Blausäure entstehenden Körper. *Justus Liebigs Ann. Chem.* **91**, 349–351 (1854).

28. Duthaler, R. O. Recent developments in the stereoselective synthesis of α-aminoacids. *Tetrahedron* **50**, 1539–1650 (1994).

29. Davis, F. A., Reddy, R. E. & Portonovo, P. S. Asymmetric strecker synthesis using enantiopure sulfinimines: A convenient synthesis of α-amino acids. *Tetrahedron Lett.* **35**, 9351–9354 (1994).

30. Iyer, M. S., Gigstad, K. M., Namdev, N. D. & Lipton, M. Asymmetric catalysis of the Strecker amino acid synthesis by a cyclic dipeptide. *Amino Acids* **11**, 259–268 (1996).

31. Gröger, H. Catalytic Enantioselective Strecker Reactions and Analogous Syntheses. *Chem. Rev.* **103**, 2795–2828 (2003).

32. van Balken, J. A. M. & Kamphuis, J. Amino acids: their production and uses in the food industry. in *Biotechnological Innovations in Food Processing* 211–251 (Butterworth-Heinemann, 1991).

33. Kinoshita, S., Udaka, S. & Shimono, M. Studies on the amino acid fermentation. *J. Gen. Appl. Microbiol.* **3**, 193–205 (1957).

34. Hori, H., Ando, T., Isogai, E., Yoneyama, H. & Katsumata, R. Identification of an l-alanine export system in Escherichia coli and isolation and characterization of export-deficient mutants. *FEMS Microbiol. Lett.* **316**, 83–89 (2011).

35. Katsube, S., Sato, K., Ando, T., Isogai, E. & Yoneyama, H. Secretion of D-alanine by escherichia coli. *Microbiol. (United Kingdom)* **162**, 1243–1252 (2016).

36. Shiio, I., Tsuka, S.-I. & Takahasi, M. Effect of Biotin on the Bacterial Formation of Glutamic Acid. *J. Biochem.* **51**, 56–62 (1962).

37. Takinami, K., Okada, H. & Tsunoda, T. Biochemical Effects of Fatty Acid and its Derivatives on L-Glutamic Acid Fermentation. *Agric. Biol. Chem.* **28**, 114–119 (1964).

38. Somerson, N. L. & Phillips, T. Production of glutamic acid. (1961).

39. Leuchtenberger, W. Amino Acids – Technical Production and Use. *Biotechnology* (2008).

40. Hoffart, E. *et al.* High substrate uptake rates empower Vibrio natriegens as production host for industrial biotechnology. *Appl Env. Microbiol* (2017).

41. Hols, P. *et al.* Conversion of Lactococcus lactis from homolactic to homoalanine fermentation through metabolic engineering. *Nat. Biotechnol.* **17**, 588–592 (1999).

42. Uhlenbusch, I., Sahm, H. & Sprenger, G. A. Expression of an L-alanine dehydrogenase gene in Zymomonas mobilis and excretion of L-alanine. *Appl. Environ. Microbiol.* **57**, 1360–1366 (1991).

43. Zhang, X., Jantama, K., Moore, J. C., Shanmugam, K. T. & Ingram, L. O. Production of l-alanine by metabolically engineered Escherichia coli. *Appl. Microbiol. Biotechnol.* **77**, 355–366 (2007).

44. Katsumata, R. & Hashimoto, S.-I. Process for producing alanine. (1996).

45. Ohshima, T., Sakane, M., Yamazaki, T. & Soda, K. Thermostable alanine dehydrogenase from thermophilic Bacillus sphaericus DSM 462. *Eur. J. Biochem.* **191**, 715–720 (1990).

46. De Graef, M. R., Alexeeva, S., Snoep, J. L. & Teixeira De Mattos, M. J. The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in Escherichia coli. *J. Bacteriol.* **181**, 2351–2357 (1999).

47. Tomar, A., Eiteman, M. A. & Altman, E. The effect of acetate pathway mutations on the production of pyruvate in Escherichia coli. *Appl. Microbiol. Biotechnol.* **62**, 76–82 (2003).

48. Futai, M. & Kimura, H. Inducible membrane-bound L-lactate dehydrogenase from Escherichia coli. Purification and properties. *J. Biol. Chem.* **252**, 5820–5827 (1977).

49. Knappe, J. & Sawers, G. A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of Escherichia coli. *FEMS Microbiol. Rev.* **6**, 383–398 (1990).

50. Smith, G. M., Lee, S. A., Reilly, K. C., Eiteman, M. A. & Altman, E. Fed-batch two-phase production of alanine by a metabolically engineered Escherichia coli. *Biotechnol. Lett.* **28**, 1695–1700 (2006).

51. Lee, M., Smith, G. M., Eiteman, M. A. & Altman, E. Aerobic production of alanine by Escherichia coli aceF ldhA mutants expressing the Bacillus sphaericus alaD gene. *Appl. Microbiol. Biotechnol.* **65**, 56–60 (2004).

52. Mat-Jan, F., Alam, K. Y. & Clark, D. P. Mutants of Escherichia coli deficient in the fermentative lactate dehydrogenase. *J. Bacteriol.* **171**, 342–348 (1989).

53. Luli, G. W. & Strohl, W. R. Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. *Appl. Environ. Microbiol.* **56**, 1004–1011 (1990).

54. Vemuri, G. N., Altman, E., Sangurdekar, D. P., Khodursky, A. B. & Eiteman, M. A. Overflow metabolism in Escherichia coli during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl. Environ. Microbiol.* **72**, 3653–3661 (2006).

55. Kleman, G. L. & Strohl, W. R. Acetate metabolism by Escherichia coli in high-cell-density fermentation. *Appl. Environ. Microbiol.* **60**, 3952–3958 (1994).

56. Wild, J., Hennig, J., Lobocka, M., Walczak, W. & Klopotowski, T. Identification of the dadX gene coding for the predominant isozyme of alanine racemase in Escherichia coli K12. *Mol. Gen. Genet. MGG* **198**, 315–322 (1985).

57. Causey, T. B., Shanmugam, K. T., Yomano, L. P. & Ingram, L. O. Engineering Escherichia coli for efficient conversion of glucose to pyruvate. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2235 LP – 2240 (2004).

58. Wada, M., Narita, K. & Yokota, A. Alanine production in an H+-ATPase- and lactate dehydrogenase-defective mutant of Escherichia coli expressing alanine dehydrogenase. *Appl. Microbiol. Biotechnol.* **76**, 819–825 (2007).

59. Orencio-Trejo, M. *et al.* Engineering the Escherichia coli Fermentative Metabolism BT - Biosystems Engineering II: Linking Cellular Networks and Bioprocesses. in (eds. Wittmann, C. & Krull, R.) 71–107 (Springer Berlin Heidelberg, 2010).

60. Zhou, L., Deng, C., Cui, W. J., Liu, Z. M. & Zhou, Z. M. Efficient L-Alanine Production by a Thermo-Regulated Switch in Escherichia coli. *Appl Biochem Biotechnol* **178**, 324–337 (2016).

61. Hashimoto, S.-I. & Katsumata, R. l-Alanine fermentation by an alanine racemase-deficient mutant of the dl-alanine hyperproducing bacterium Arthrobacter oxydans HAP-1. *J. Ferment. Bioeng.* **86**, 385–390 (1998).

62. Suye, S., Kawagoe, M. & Inuta, S. Enzymatic production of L-alanine from malic acid with malic enzyme and alanine dehydrogenase with coenzyme regeneration. *Can. J. Chem. Eng.* **70**, 306–312 (1992).

63. Meister, A., Sober, H. A. & Tice, S. V. Enzymatic decarboxlation of aspartic acid to α-alanine. *J. Biol. Chem.* **189**, 577–590 (1951).

64. Crawford, L. V. Studies on the aspartic decarboxylase of Nocardia globerula. *Biochem. J.* **68**, 221–225 (1958).

65. Wilson, E. M. & Kornberg, H. L. Properties of crystalline L-aspartate 4-carboxy-lyase from achromobacter sp. *Biochem. J.* **88**, 578–587 (1963).

66. Chibata, I., Kakimoto, T. & Kato, J. Enzymatic Production of l-Alanine by Pseudomonas dacunhae. *Appl. Microbiol.* **13**, 638–645 (1965).

67. Chibata, I., Kakimoto, T., Kato, J., Shibatani, T. & Nishimura, N. On the activation mechanism of L-aspartate β-decarboxylase from Pseudomonasdacunhae by α-ketoglutarate. *Biochem. Biophys. Res. Commun.* **32**, 375–379 (1968).

68. Chibata, I., Kakimoto, T., Kato, J., Shibatani, T. & Nishimura, N. Crystalline aspartic β-decarboxylase of pseudomonas dacunhae. *Biochem. Biophys. Res. Commun.* **26**, 662–667 (1967).

69. Novogrodsky, A., Nishimura, J. S. & Meister, A. Transamination and β-Decarboxylation of Aspartate Catalyzed by the Same Pyridoxal Phosphate-Enzyme. *J. Biol. Chem.* **238**, PC1903–PC1905 (1963).

70. Novogrodsky, A. & Meister, A. Control of Aspartate β-Decarboxylase Activity by Transamination. *J. Biol. Chem.* **239**, 879–888 (1964).

71. Rudolph, F. B. & Fromm, H. J. The purification and properties of aspartase from Escherichia coli. *Arch. Biochem. Biophys.* **147**, 92–98 (1971).

72. Jandel, A.-S., Hustedt, H. & Wandrey, C. Continuous production of L-alanine from fumarate in a two-stage membrane reactor. *Eur. J. Appl. Microbiol. Biotechnol.* **15**, 59–63 (1982).

73. Furui, M. & Yamashita, K. *Pressurized reaction method for continuous production of L-alanine by immobilized Pseudomonas dacunhae cells*. *J. Ferment. Technol.* **61**, (1983).

74. Fusee, M. C. & Weber, J. E. Immobilization by Polyurethane of Pseudomonas dacunhae Cells Containing l-Aspartate β-Decarboxylase Activity and Application to l-Alanine Prod. *Appl. Environ. Microbiol.* **48**, 694 LP – 698 (1984).

75. Burdick, B. A. & Schaeffer, J. R. Production of specialty chemicals using immobilized whole cells in spiral-wound bioreactors: I. Aspartic acid and alanine production using cells immobilized in gelatin on nylon mesh. *Biotechnol. Bioeng.* **31**, 390–395 (1988).

76. Takamatsu, S., Tosa, T. & Chibata, I. Production of L-alanine from ammonium fumarate using two microbial cells immobilized with k-carrageenan. *J. Chem. Eng. Japan* **18**, 66–70 (1985).

77. Takamatsu, S. *et al.* Production of l-alanine from ammonium fumarate using two immobilized microorganisms. *Eur. J. Appl. Microbiol. Biotechnol.* **15**, 147–152 (1982).

78. Takamatsu, S., Tosa, T. & Chibata, I. Industrial production of L-alanine from ammonium fumarate using immobilized cells of two kinds. *J. Chem. Eng. Japan* **19**, 31–36 (1986).

79. Ahmed Laskar, A. & Younus, H. Aldehyde toxicity and metabolism: the role of aldehyde dehydrogenases in detoxification, drug resistance and carcinogenesis. *Drug Metab. Rev.* **51**, 42–64 (2019).

80. Paul, P. E. V., Sangeetha, V. & Deepika, R. G. Emerging Trends in the Industrial Production of Chemical Products by Microorganisms. in (ed. Buddolla, V. B. T.-R. D. in A. M. and B.) 107–125 (Academic Press, 2019).

81. Crabb, D. W., Matsumoto, M., Chang, D. & You, M. Overview of the role of alcohol dehydrogenase and aldehyde dehydrogenase and their variants in the genesis of alcohol-related pathology. *Proc. Nutr. Soc.* **63**, 49–63 (2004).

82. Koppaka, V. *et al.* Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. *Pharmacol Rev* **64**, 520–539 (2012).

83. Vasiliou, V., Bairoch, A., Tipton, K. F. & Nebert, D. W. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenet. Genomics* **9**, (1999).

84. Hayes, K. *et al.* The quaternary structure of Thermus thermophilus aldehyde dehydrogenase is stabilized by an evolutionary distinct C-terminal arm extension. *Sci. Rep.* **8**, 13327 (2018).

85. Di Costanzo, L., Gomez, G. A. & Christianson, D. W. Crystal Structure of Lactaldehyde Dehydrogenase from Escherichia coli and Inferences Regarding Substrate and Cofactor Specificity. *J. Mol. Biol.* **366**, 481–493 (2007).

86. Jung, J. H. & Lee, S. B. Identification and characterization of Thermoplasma acidophilum glyceraldehyde dehydrogenase: a new class of NADP+-specific aldehyde dehydrogenase. *Biochem J* **397**, 131–138 (2006).

87. Hempel, J. *et al.* Aldehyde Dehydrogenase Catalytic Mechanism BT - Enzymology and Molecular Biology of Carbonyl Metabolism 7. in (eds. Weiner, H., Maser, E., Crabb, D. W. & Lindahl, R.) 53–59 (Springer US, 1999).

88. Moore, S. A. *et al.* Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* **6**, 1541–1551 (1998).

89. Ni, L., Sheikh, S. & Weiner, H. Involvement of Glutamate 399 and Lysine 192 in the Mechanism of Human Liver Mitochondrial Aldehyde Dehydrogenase. *J. Biol. Chem.* **272**, 18823–18826 (1997).

90. Stiti, N. *et al.* Nicotinamide Cofactors Suppress Active-Site Labeling of Aldehyde Dehydrogenases. *ACS Chem Biol* **11**, 1578–1586 (2016).

91. Wang, X. & Weiner, H. Involvement of Glutamate 268 in the Active Site of Human Liver Mitochondrial (Class 2) Aldehyde Dehydrogenase As Probed by Site-Directed Mutagenesis. *Biochemistry* **34**, 237–243 (1995).

92. Clark, D. P. The fermentation pathways of Escherichia coli. *FEMS Microbiol. Lett.* **63**, 223–234 (1989).

93. Rossmann, M. G., Moras, D. & Olsen, K. W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250**, 194–199 (1974).

94. Kessel, A. & Ben-Tal, N. *Introduction to Proteins. Stucture Function and Motion*. (2018).

95. Rao, S. T. & Rossmann, M. G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**, 241–256 (1973).

96. Brändén, C.-I. Relation between structure and function of α/β–protejns. *Q. Rev. Biophys.* **13**, 317–338 (1980).

97. Chánique, A. M. & Parra, L. P. Protein Engineering for Nicotinamide Coenzyme Specificity in Oxidoreductases: Attempts and Challenges. *Front. Microbiol.* **9**, 194 (2018).

98. Wierenga, R. K., Terpstra, P. & Hol, W. G. J. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187**, 101–107 (1986).

99. Kleiger, G. & Eisenberg, D. GXXXG and GXXXA Motifs Stabilize FAD and NAD(P)-binding Rossmann Folds Through Cα–H···O Hydrogen Bonds and van der Waals Interactions. *J. Mol. Biol.* **323**, 69–76 (2002).

100. Senes, A., Ubarretxena-Belandia, I. & Engelman, D. M. The Calpha ---H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9056–9061 (2001).

101. Bellamacina, C. R. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.* **10**, 1257–1269 (1996).

102. Liu, Z.-J. *et al.* The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat Struct Mol Biol* **4**, 317–326 (1997).

103. Hammen, P. K., Allali-Hassani, A., Hallenga, K., Hurley, T. D. & Weiner, H. Multiple Conformations of NAD and NADH When Bound to Human Cytosolic and Mitochondrial Aldehyde Dehydrogenase. *Biochemistry* **41**, 7156–7168 (2002).

104. Ansari, H. R. & Raghava, G. P. S. Identification of NAD interacting residues in proteins. *BMC Bioinformatics* **11**, 160 (2010).

105. Wang, Y., San, K.-Y. & Bennett, G. N. Cofactor engineering for advancing chemical biotechnology. *Curr. Opin. Biotechnol.* **24**, 994–999 (2013).

106. Bastian, S. *et al.* Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metab. Eng.* **13**, 345–352 (2011).

107. Scrutton, N. S., Berry, A. & Perham, R. N. Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38–43 (1990).

108. Rice, D. W., Schulz, G. E. & Guest, J. R. Structural relationship between glutathione reductase and lipoamide dehydrogenase. *J. Mol. Biol.* **174**, 483–496 (1984).

109. Yaoi, T., Miyazaki, K., Oshima, T., Komukai, Y. & Go, M. Conversion of the Coenzyme Specificity of Isocitrate Dehydrogenase by Module Replacement1. *J. Biochem.* **119**, 1014–1018 (1996).

110. Campbell, E., Chuang, S. & Banta, S. Modular exchange of substrate-binding loops alters both substrate and cofactor specificity in a member of the aldo-keto reductase superfamily. *Protein Eng. Des. Sel.* **26**, 181–186 (2012).

111. Naylor, C. E. *et al.* NADP+ and NAD+ binding to the dual coenzyme specific enzyme Leuconostoc mesenteroides glucose 6-phosphate dehydrogenase: Different interdomain hinge angles are seen in different binary and ternary complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 635–648 (2001).

112. Plapp, B. V. Conformational changes and catalysis by alcohol dehydrogenase. *Arch. Biochem. Biophys.* **493**, 3–12 (2010).

113. Cahn, J. K. B., Baumschlager, A., Brinkmann-Chen, S. & Arnold, F. H. Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)H-dependent enzymes. *Protein Eng. Des. Sel.* **29**, 31–38 (2016).

114. Maddock, D. J., Patrick, W. M. & Gerth, M. L. Substitutions at the cofactor phosphate-binding site of a clostridial alcohol dehydrogenase lead to unexpected changes in substrate specificity. *Protein Eng Des Sel* **28**, 251–258 (2015).

115. Cahn, J. K. *et al.* Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases. *Biochem J* **468**, 475–484 (2015).

116. Carugo, O. & Argos, P. NADP-Dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins Struct. Funct. Bioinforma.* **28**, 10–28 (1997).

117. Rosell, A. *et al.* Complete Reversal of Coenzyme Specificity by Concerted Mutation of Three Consecutive Residues in Alcohol Dehydrogenase. *J. Biol. Chem.* **278**, 40573–40580 (2003).

118. Rodríguez-Arnedo, A., Camacho, M., Llorca, F. & Bonete, M.-J. Complete Reversal of Coenzyme Specificity of Isocitrate Dehydrogenase from Haloferax volcanii. *Protein J.* **24**, 259 (2005).

119. Saraf, M. C. *et al.* IPRO: An Iterative Computational Protein Library Redesign and Optimization Procedure. *Biophys. J.* **90**, 4167–4180 (2006).

120. Khoury, G. A. *et al.* Computational design of Candida boidinii xylose reductase for altered cofactor specificity. *Protein Sci* **18**, 2125–2138 (2009).

121. Brinkmann-Chen, S. *et al.* General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proc. Natl. Acad. Sci.* **110**, 10946–10951 (2013).

122. Wong, S.-H., Lonhienne, T. G. A., Winzor, D. J., Schenk, G. & Guddat, L. W. Bacterial and Plant Ketol-Acid Reductoisomerases Have Different Mechanisms of Induced Fit during the Catalytic Cycle. *J. Mol. Biol.* **424**, 168–179 (2012).

123. Cui, D. *et al.* A computational strategy for altering an enzyme in its cofactor preference to NAD(H) and/or NADP(H). *FEBS J.* **282**, 2339–2351 (2015).

124. Cahn, J. K. *et al.* A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases. *ACS Synth Biol* **6**, 326–333 (2017).

125. Bahar, I. & Jernigan, R. L. Coordination geometry of nonbonded residues in globular proteins. *Fold. Des.* **1**, 357–370 (1996).

126. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **19**, 120–127 (2009).

127. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

128. Weiner, P. K. & Kollman, P. A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* **2**, 287–303 (1981).

129. Scott, W. R. P. *et al.* The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **103**, 3596–3607 (1999).

130. Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **9**, 3887 (2018).

131. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).

132. Balsera, M. A., Wriggers, W., Oono, Y. & Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **100**, 2567–2572 (1996).

133. Wymore, T., Nicholas, H. B. & Hempel, J. Molecular dynamics simulation of class 3 aldehyde dehydrogenase. *Chem. Biol. Interact.* **130**–**132**, 201–207 (2001).

134. Marchal, S. & Branlant, G. Evidence for the Chemical Activation of Essential Cys-302 upon Cofactor Binding to Nonphosphorylating Glyceraldehyde 3-Phosphate Dehydrogenase from Streptococcus mutans. *Biochemistry* **38**, 12950–12958 (1999).

135. Wymore, T., Deerfield, D. W. & Hempel, J. Mechanistic Implications of the Cysteine−Nicotinamide Adduct in Aldehyde Dehydrogenase Based on Quantum Mechanical/Molecular Mechanical Simulations. *Biochemistry* **46**, 9495–9506 (2007).

136. Dittmer, D. C. & Kolyer, J. M. Addition Compounds of Thiols and 1-Substituted Nicotinamides1,2. *J. Org. Chem.* **28**, 1720–1722 (1963).

137. Tsybovsky, Y., Donato, H., Krupenko, N. I., Davies, C. & Krupenko, S. A. Crystal Structures of the Carboxyl Terminal Domain of Rat 10-Formyltetrahydrofolate

Dehydrogenase: Implications for the Catalytic Mechanism of Aldehyde Dehydrogenases. *Biochemistry* **46**, 2917–2929 (2007).

138. Ion, B. F., Aboelnga, M. M. & Gauld, J. W. Insights from molecular dynamics on substrate binding and effects of active site mutations in Δ1-pyrroline-5-carboxylate dehydrogenase. *Can. J. Chem.* **94**, 1151–1162 (2016).

139. Wymore, T., Deerfield, D. W., Field, M. J., Hempel, J. & Nicholas, H. B. Initial catalytic events in class 3 aldehyde dehydrogenase: MM and QM/MM simulations. *Chem. Biol. Interact.* **143**–**144**, 75–84 (2003).

140. Baicharoen, A., Vijayan, R. & Pongprayoon, P. Structural insights into betaine aldehyde dehydrogenase (BADH2) from Oryza sativa explored by modeling and simulations. *Sci. Rep.* **8**, 12892 (2018).

141. Rahuel-Clermont, S. *et al.* Enzyme Active Site Loop Revealed as a Gatekeeper for Cofactor Flip by Targeted Molecular Dynamics Simulations and FRET-Based Kinetics. *ACS Catal.* **9**, 1337–1346 (2019).

142. D'Ambrosio, K. *et al.* The First Crystal Structure of a Thioacylenzyme Intermediate in the ALDH Family: New Coenzyme Conformation and Relevance to Catalysis. *Biochemistry* **45**, 2978–2986 (2006).

143. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J. Bacteriol.* **62**, 293–300 (1951).

144. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

145. 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).

146. Steffler, F., Guterl, J.-K. & Sieber, V. Improvement of thermostable aldehyde dehydrogenase by directed evolution for application in Synthetic Cascade Biomanufacturing. *Enzyme Microb. Technol.* **53**, 307–314 (2013).

147. 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).

148. Bantan-Polak, T., Kassai, M. & Grant, K. B. A comparison of fluorescamine and naphthalene-2,3-dicarboxaldehyde fluorogenic reagents for microplate-based detection of amino acids. *Anal. Biochem.* **297**, 128–136 (2001).

149. Boivin, S., Kozak, S. & Meijers, R. Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif.* **91**, 192–206 (2013).

150. Budgen, N. & Danson, M. J. Metabolism of glucose via a modified Entner-Doudoroff pathway in the thermoacidophilic archaebacterium Thermoplasma acidophilum. *FEBS Lett.* **196**, 207–210 (1986).

151. 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* **192**, 964–974 (2010).

152. 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).

153. Andberg, M. *et al.* Characterization and mutagenesis of two novel iron-sulphur cluster pentonate dehydratases. *Appl Microbiol Biotechnol* (2016).

154. 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).

155. Steffler, F. & Sieber, V. Refolding of a thermostable glyceraldehyde dehydrogenase for application in synthetic cascade biomanufacturing. *PLoS One* **8**, e70592 (2013).

156. Cramer, B. Verbesserung einer zellfreien Enzymkaskade. (Technical University of Munich, 2016).

157. Grochowski, L. L., Xu, H. & White, R. H. Identification of lactaldehyde dehydrogenase in Methanocaldococcus jannaschii and its involvement in production of lactate for F420 biosynthesis. *J Bacteriol* **188**, 2836–2844 (2006).

158. Hill, D. W., Walters, F. H., Wilson, T. D. & Stuart, J. D. High performance liquid chromatographic determination of amino acids in the picomole range. *Anal. Chem.* **51**, 1338–1341 (1979).

159. Salazar, O. & Sun, L. Evaluating a Screen and Analysis of Mutant Libraries BT - Directed Enzyme Evolution: Screening and Selection Methods. in (eds. Arnold, F. H. & Georgiou, G.) 85–97 (Humana Press, 2003).

160. Zhang, J.-H., Chung, T. D. Y. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **4**, 67–73 (1999).

161. Iermak, I., Degtjarik, O., Steffler, F., Sieber, V. & Kuta Smatanova, I. Crystallization behaviour of glyceraldehyde dehydrogenase from Thermoplasma acidophilum. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **71**, (2015).

162. Lovell, S. C. *et al.* Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins Struct. Funct. Bioinforma.* **50**, 437–450 (2003).

163. Jimenez-Lopez, J. C. *et al.* Genome-Wide Identification and Functional Classification of Tomato (Solanum lycopersicum) Aldehyde Dehydrogenase (ALDH) Gene Superfamily. *PLoS One* **11**, e0164798–e0164798 (2016).

164. Colovos, C. & Yeates, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519 (1993).

165. Eisenberg, D., Lüthy, R. & Bowie, J. U. B. T.-M. in E. VERIFY3D: Assessment of protein models with three-dimensional profiles. in *Macromolecular Crystallography Part B* **277**, 396–404 (Academic Press, 1997).

166. Perozich, J., Kuo, I., Lindahl, R. & Hempel, J. Coenzyme specificity in aldehyde dehydrogenase. *Chem. Biol. Interact.* **130**–**132**, 115–124 (2001).

167. Zhang, L., Ahvazi, B., Szittner, R., Vrielink, A. & Meighen, E. Change of Nucleotide Specificity and Enhancement of Catalytic Efficiency in Single Point Mutants of Vibrio harveyi Aldehyde Dehydrogenase. *Biochemistry* **38**, 11440–11447 (1999).

168. Bendl, J. *et al.* HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. *Nucleic Acids Res* **44**, W479-87 (2016).

169. Keller, M. A. *et al.* A gatekeeper helix determines the substrate specificity of Sjögren–Larsson Syndrome enzyme fatty aldehyde dehydrogenase. *Nat. Commun.* **5**, 4439 (2014).

170. Jiamsomboon, K., Treesuwan, W. & Boonyalai, N. Dissecting substrate specificity of two rice BADH isoforms: Enzyme kinetics, docking and molecular dynamics simulation studies. *Biochimie* **94**, 1773–1783 (2012).

171. Kamal, M. Z., Mohammad, T. A. S., Krishnamoorthy, G. & Rao, N. M. Role of active site rigidity in activity: MD simulation and fluorescence study on a lipase mutant. *PLoS One* **7**, e35188–e35188 (2012).

6 Appendix

6.1 Supplemental information: Optimization of a reduced enzymatic reaction cascade for the production of L-alanine

Supplementary Material

Optimization of a reduced enzymatic reaction cascade for the production of L-alanine

Tobias J. Gmelch[1], Josef M. Sperl[1], Volker Sieber[1,2,3,4]*

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

[2] Catalysis Research Center, Technical University of Munich, Garching, Germany

[3] Fraunhofer Institute of Interfacial Biotechnology (IGB), Bio-, Electro- and Chemo Catalysis (BioCat) Branch, Straubing, Germany

[4] School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD, Australia

* Correspondence should be addressed to
Prof. Volker Sieber
E-Mail: sieber@tum.de
Tel: +499421187301

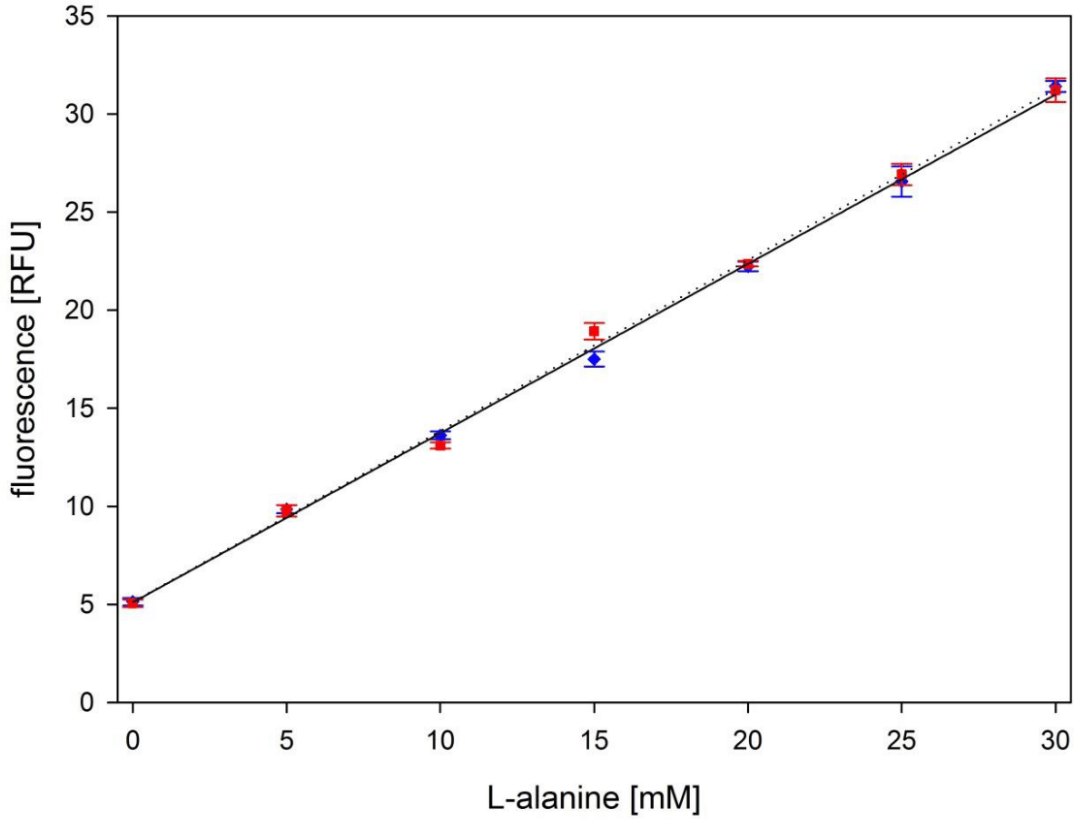


Figure S1. Fluorescamine calibration curves in different buffer matrices. Ammonium is consumed during the reductive amination of pyruvate to L-alanine. To exclude effects of reduced ammonium concentration on the quantification of L-alanine, we compared standards containing either 200 mM ammonium or an ammonium concentration that is reduced by the concentration of L-alanine in the corresponding standard sample. Blue diamonds: 100 mM HEPES pH 7.35, 200 mM ammonium; red squares: 100 mM HEPES pH 7.35, (200 mM – [L-alanine]) ammonium. Fluorescamine assays were performed with 100-fold diluted samples in total reaction volumes of 158 µL, average values of technical triplicates are shown with error bars indicating the standard deviation.

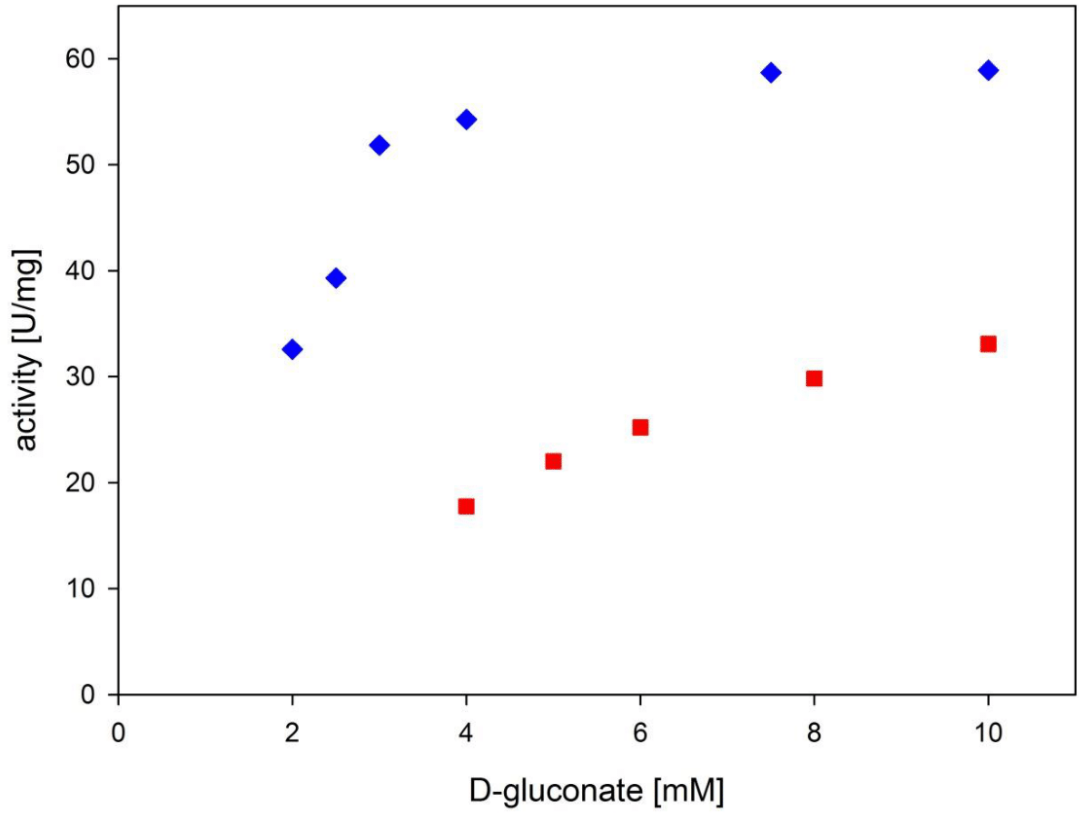


Figure S2. Kinetic measurements of CcDHAD containing 2.2 mM glycerate (red squares) compared to standard conditions (blue diamonds). Reactions were performed at 50 °C in 100 mM HEPES buffer pH 7.35 and stopped after 10 min by enzyme removal. Activities were calculated from the residual amount of D-gluconate detected by HPLC.

Expressed DNA Sequences

Bold nucleotides represent additional tags from the expression plasmid.

BsGDH (NCBI-Prot. ID: NP_388275.1)

ATGTATCCGGATTTAAAAGGAAAAGTCGTCGCTATTACAGGAGCTGCTTCAGGGCTCGGAA
AGGCGATGGCCATTCGCTTCGGCAAGGAGCAGGCAAAAGTGGTTATCAACTATTATAGTAA
TAAACAAGATCCGAACGAGGTAAAAGAAGAGGTCATCAAGGCGGGCGGTGAAGCTGTTGT
CGTCCAAGGAGATGTCACGAAAGAGGAAGATGTAAAAAATATCGTGCAAACGGCAATTAAG
GAGTTCGGCACACTCGATATTATGATTAATAATGCCGGTCTTGAAAATCCTGTGCCATCTCA
CGAAATGCCGCTCAAGGATTGGGATAAAGTCATCGGCACGAACTTAACGGGTGCCTTTTTA
GGAAGCCGTGAAGCGATTAAATATTTCGTAGAAAACGATATCAAGGGAAATGTCATTAACAT
GTCCAGTGTGCACGAAGTGATTCCTTGGCCGTTATTTGTCCACTATGCGGCAAGTAAAGGC
GGGATAAAGCTGATGACAAAGACATTAGCGTTGGAATACGCGCCGAAGGGCATTCGCGTC
AATAATATTGGGCCAGGTGCGATCAACACGCCAATCAATGCTGAAAAATTCGCTGACCCTA
AACAGAAAGCTGATGTAGAAAGCATGATTCCAATGGGATATATCGGCGAACCGGAGGAGAT

CGCCGCAGTAGCAGCCTGGCTTGCTTCGAAGGAAGCCAGCTACGTCACAGGCATCACGTT
ATTCGCGGACGGCGGTATGACACTCTATCCTTCATTCCAGGCAGGCCGCGGTTAA

CcDHAD (NCBI-Prot. ID: WP_012640069.1)

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCA
TATGAGCAATCGTACACCGCGTCGTTTTCGTAGCCGTGATTGGTTTGATAATCCGGATCATA
TTGATATGACCGCACTGTATCTGGAACGCTTTATGAATTATGGTATCACACCGGAAGAACTG
CGTAGCGGTAAACCGATTATTGGTATTGCACAGACCGGTAGCGATATTAGCCCGTGTAATC
GTATTCATCTGGATCTGGTTCAGCGTGTTCGTGATGGTATTCGTGATGCCGGTGGTATTCC
GATGGAATTTCCGGTTCATCCGATTTTTGAAAATTGTCGTCGTCCGACCGCAGCACTGGAT
CGTAATCTGAGCTATCTGGGTTTAGTTGAAACCCTGCATGGTTATCCGATTGATGCAGTTGT
TCTGACCACCGGTTGTGATAAAACCACACCGGCAGGTATTATGGCAGCAACCACCGTTAAT
ATTCCGGCAATTGTTCTGAGCGGTGGTCCGATGCTGGATGGTTGGCATGAAAATGAACTGG
TTGGTAGCGGCACCGTTATTTGGCGTAGCCGTCGTAAACTGGCAGCCGGTGAAATTACCGA
AGAGGAATTTATTGATCGTGCAGCAAGCAGCGCACCGAGCGCAGGTCATTGCAATACAATG
GGCACCGCAAGCACCATGAATGCAGTTGCAGAAGCACTGGGTCTGAGCCTGACCGGTTGC
GCAGCAATTCCGGCACCGTATCGTGAACGTGGTCAGATGGCATATAAAACAGGTCAGCGTA
TTGTTGATCTGGCCTATGATGATGTTAAACCGCTGGATATTCTGACCAAACAGGCATTTGAA
AATGCAATTGCCCTGGTTGCAGCAGCAGGCGGTAGCACCAATGCACAGCCGCATATTGTTG
CAATGGCACGTCATGCCGGTGTTGAAATTACAGCAGATGATTGGCGTGCAGCCTATGATAT
TCCGCTGATTGTTAATATGCAGCCTGCAGGTAAATATCTGGGTGAACGTTTTCATCGTGCCG
GTGGCGCTCCGGCAGTTCTGTGGGAACTGTTACAGCAGGGTCGTCTGCATGGTGATGTGC
TGACCGTTACCGGCAAAACCATGAGCGAAAATCTGCAGGGTCGCGAAACCAGCGATCGTG
AAGTTATTTTTCCGTATCATGAACCGCTGGCAGAAAAAGCAGGTTTTCTGGTTCTGAAAGGT
AACCTGTTTGATTTTGCCATCATGAAAAGCAGCGTTATCGGTGAAGAATTTCGTAAACGTTA
TCTGAGCCAGCCTGGTCAAGAAGGTGTTTTTGAAGCACGTGCAATTGTGTTTGATGGTAGC
GACGATTATCATAAGCGTATTAATGATCCGGCACTGGAAATTGATGAACGTTGCATTCTGGT
TATTCGTGGTGCAGGTCCGATTGGTTGGCCTGGTAGTGCCGAAGTTGTGAACATGCAGCCA
CCGGACCATCTGCTGAAAAAAGGCATTATGAGCCTGCCGACCTTAGGTGATGGTCGTCAGA
GTGGCACCGCAGATAGCCCGAGCATTCTGAATGCAAGTCCGGAAAGCGCAATTGGTGGTG
GTCTGAGCTGGCTGCGTACCGGTGATACAATTCGTATTGATCTGAATACCGGTCGTTGTGA
TGCACTGGTTGATGAAGCAACCATTGCAGCACGTAAACAGGATGGCATTCCTGCAGTTCCG
GCAACCATGACACCGTGGCAAGAAATTTATCGTGCACATGCAAGCCAGCTGGATACCGGTG
GTGTTCTGGAATTTGCCGTTAAATATCAGGATCTGGCAGCAAAACTGCCTCGCCATAATCAT
TAA

PtKDGA (NCBI-Prot. ID: 4UXD_A)

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCA
TATGTACAAGGGAATCGTTTGTCCGATGATTACCCCGCTGGACGCACACGGCAACATCGAC
TATAACGCAACGAACATTCTGATCAAATACCTGGAAGGCATCAACGTGGACTACCTGTTCCC
GATGGGCTCTACCGGTGTGTTTCCGTATTTCACGCTGAAAGAACGTAAAGACTTTCTGAAAT
TCGTTCGCGAAAACAGTAAAAAACCGATCATGGCTGGCGTCGGTAGCTCTAGTATCAACGA
AGTGAACGAACTGATGAAATTCAGCATGGATATCGGTATCGAAGCGGCCGTGCTGATGCCG
CCATATTACATTAAACTGAACCAGGAAGCAATCTACCATTACTACAAAGAAATCCTGTCCTC
AAATGACATGGATCTGCTGATTTACAACATCCCGCAATTTACCAATAAAATCGATCCGGAAA

CGGTCAAAAACCTGAAATCTGAATTTAGCAGCGTGAAAGGTGTTAAAGACTCTAGTGCAGAT
ATTCGTGGCTTTATGGAAATGCTGTCCCTGTCAGATGACGATTTTGCTGTTTTCCAGGGTCA
AGACGATCTGCTGTTTACCTCGCTGGAACTGGGTGCCAGCGGCGGTGTTTGTGGTACCAC
GAATTTTAGCGACGGCATCGTCCGTCTGTATCACGAATACAAAAACAATCGCGAAATGGCC
CTGAAAATTGAGAAAAACGATGTTATCCCGCTGATGAAAAAACTGGGCAAATATCAGTTTCC
GAATGCATATTACGAATATTTCTACAAGAAAAACAATATCAATGGCGGCTACCGTCCGCCGA
TGTATCGCGTTGGCATCGAAATCTAA

MjALDH (NCBI-Prot. ID: WP_010870928)

ATGTTTATTGATGGCAAATGGATTAACCGCGAGGATATGGATGTTATTAATCCGTATAGCCT
GGAAGTGATCAAAAAAATCCCTGCACTGAGCCGTGAAGAAGCAAAAGAAGCCATTGATACC
GCAGAGAAATACAAAGAAGTCATGAAAAATCTGCCGATCACCAAACGCTATAACATCCTGAT
GAATATTGCCAAGCAGATCAAAGAAAAAAAAGAAGAACTGGCCAAAATCCTGGCAATTGAT
GCAGGTAAACCGATTAAACAGGCACGTGTTGAAGTTGAACGTAGCATTGGCACCTTTAAAC
TGGCAGCCTTTTATGTGAAAGAACACCGTGATGAAGTTATTCCGAGTGATGATCGTCTGATT
TTTACCCGTCGTGAACCGGTTGGTATTGTTGGTGCAATTACCCCGTTTAACTTTCCGCTGAA
TCTGAGCGCACATAAAATTGCACCGGCAATTGCAACCGGTAATGTTATTGTTCATCATCCGA
GCAGCAAAGCACCGCTGGTTTGTATTGAACTGGCAAAGATTATTGAGAACGCCCTGAAAAA
ATACAATGTTCCGCTGGGTGTTTATAATCTGCTGACCGGTGCCGGTGAAGTTGTTGGTGAT
GAAATTGTTGTGAACGAGAAGGTGAACATGATTAGCTTTACCGGTAGCAGTAAAGTGGGCG
AACTGATTACCAAAAAAGCCGGTTTTAAGAAAATCGCCCTGGAATTAGGTGGTGTGAATCC
GAATATTGTTCTGAAAGATGCCGATCTGAATAAAGCAGTTAACGCGCTGATTAAAGGCAGCT
TTATCTATGCAGGTCAGGTTTGCATTAGCGTTGGTATGATTCTGGTTGATGAAAGCATTGCC
GATAAGTTCATCGAAATGTTTGTGAACAAAGCCAAAGTGCTGAATGTTGGTAATCCGCTGGA
TGAAAAAACCGATGTTGGTCCGCTGATTAGCGTGGAACATGCAGAATGGGTTGAAAAAGTT
GTGGAAAAAGCGATTGATGAAGGTGGCAAACTGCTGTTAGGTGGTAAACGTGATAAAGCAC
TGTTTTATCCGACCATTCTGGAAGTGGATCGTGATAATATTCTGTGCAAAACCGAAACCTTT
GCACCGGTGATTCCGATTATTCGTACCAATGAAGAAGAAATGATCGATATTGCGAACAGCA
CCGAATATGGTCTGCATAGCGCAATTTTTACCAACGACATTAACAAGAGCCTGAAATTTGCC
GAAAATCTGGAATTTGGCGGTGTTGTGATTAATGATAGCAGCCTGTTTCGTCAGGATAACAT
GCCGTTTGGTGGCGTGAAAAAAAGCGGTCTGGGTCGTGAAGGTGTTAAATATGCAATGGAA
GAGATGAGCAACATCAAAACCATCATCATCAGCAAATAA

AfAlaDH (NCBI-Prot. ID: WP_010879161.1)

ATGGAAACCCTGATCCTGACGCAAGAAGAAGTTGAAAGCCTGATTAGCATGGATGAAGCAA
TGAATGCAGTTGAAGAGGCATTTCGTCTGTATGCACTGGGTAAAGCACAGATGCCTCCGAA
AGTTTATCTGGAATTTGAAAAAGGTGATCTGCGTGCAATGCCTGCACATCTGATGGGTTATG
CAGGTCTGAAATGGGTTAATAGCCATCCGGGTAATCCGGATAAAGGTCTGCCGACCGTTAT
GGCACTGATGATTCTGAATAGTCCGGAAACAGGTTTTCCGCTGGCAGTTATGGATGCAACC
TATACCACCAGTCTGCGTACCGGTGCAGCCGGTGGTATTGCAGCAAAATATCTGGCACGTA
AAAATAGCAGCGTGTTTGGTTTTATTGGTTGTGGCACCCAGGCATATTTTCAGCTGGAAGCA
CTGCGTCGTGTTTTTGATATTGGTGAAGTTAAAGCGTATGACGTGCGTGAAAAAGCAGCCA
AAAAATTCGTTAGCTATTGCGAAGATCGTGGTATTAGCGCAAGCGTTCAGCCTGCCGAAGA
GGCAAGCCGTTGTGATGTTCTGGTTACCACCACACCGAGCCGTAAACCGGTTGTTAAAGCA
GAATGGGTAGAAGAGGGCACCCATATTAATGCAATTGGTGCAGATGGTCCGGGTAAACAAG

AACTGGATGTTGAAATTCTGAAGAAAGCCAAAATCGTGGTGGATGATCTGGAACAGGCAAA
ACATGGTGGTGAAATTAATGTTGCAGTTAGCAAAGGTGTGATTGGCGTTGAAGATGTTCATG
CAACCATTGGCGAAGTTATTGCTGGCCTGAAAGATGGTCGTGAAAGTGATGAAGAAATCAC
CATTTTTGATAGCACCGGTCTGGCAATTCAGGATGTTGCCGTTGCAAAAGTTGTTTATGAAA
ATGCCCTGAGCAAAAACGTGGGTAGCAAAATCAAATTTTTCCGCATC

SsDHAD (NCBI-Prot. ID: WP_012953192.1)

ATGCCTGCAAAACTGAATAGCCCGAGCCGTTATCATGGTATTTATAATGCACCGCATCGTG
CATTTCTGCGTAGCGTTGGTCTGACCGATGAAGAAATTGGTAAACCGCTGGTTGCAATTGC
CACCGCATGGTCTGAAGCCGGTCCGTGTAATTTTCATACCCTGGCACTGGCACGTGTTGCA
AAAGAAGGCACCAAAGAAGCCGGTCTGTCTCCGCTGGCATTTCCGACCATGGTTGTGAATG
ATAATATTGGCATGGGTAGCGAAGGTATGCGTTATAGCCTGGTTAGCCGTGATCTGATTGC
AGATATGGTTGAAGCACAGTTTAATGCCCATGCATTTGATGGTCTGGTTGGTATTGGTGGTT
GTGATAAAACCACACCGGGTATTCTGATGGCAATGGCACGTCTGAATGTTCCGAGCATTTA
TATTTATGGTGGTAGCGCAGAACCGGGTTATTTTATGGGTAAACGCCTGACCATTGAAGAT
GTTCATGAAGCCATTGGTGCATATCTGGCAAAACGCATTACCGAAAATGAACTGTATGAAAT
TGAAAAACGTGCACATCCGACCCTGGGCACCTGTAGCGGTCTGTTTACCGCAAATACCATG
GGTAGCATGAGCGAAGCACTGGGTATGGCACTGCCTGGTAGCGCATCTCCGACCGCAACC
AGCAGCCGTCGTGTTATGTATGTTAAAGAAACCGGTAAAGCCCTGGGTAGCCTGATTGAAA
ATGGCATTAAAAGCCGTGAAATTCTGACCTTTGAAGCCTTTGAAAATGCAATTACAACCCTG
ATGGCGATGGGTGGTAGCACCAATGCAGTTCTGCATCTGCTGGCAATTGCTTATGAAGCCG
GTGTTAAACTGACCCTGGATGATTTTAATCGCATTAGCAAACGCACCCCGTATATTGCAAGC
ATGAAACCGGGTGGTGATTATGTTATGGCCGATCTGGATGAAGTTGGTGGTGTTCCGGTTG
TTCTGAAAAAACTGCTGGATGCCGGTCTGCTGCATGGTGATGTTCTGACCGTTACCGGTAA
AACCATGAAACAGAATCTGGAACAGTATAAATATCCGAATGTGCCGCATAGCCATATTGTTC
GTGATGTGAAAAATCCGATTAAACCGCGTGGTGGTATTGTTATTCTGAAAGGTAGCCTGGC
ACCGGAAGGTGCAGTTATTAAAGTTGCAGCCACCAATGTGGTTAAATTTGAAGGCAAAGCC
AAAGTGTATAATAGCGAAGATGATGCCTTTAAAGGTGTTCAGAGCGGTGAAGTTAGCGAAG
GTGAAGTGGTGATTATTCGCTATGAAGGTCCGAAAGGTGCACCGGGTATGCCGGAAATGCT
GCGCGTTACCGCAGCGATTATGGGTGCCGGTCTGAATAATGTTGCACTGGTTACCGATGGT
CGTTTTAGCGGTGCAACCCGTGGTCCGATGGTTGGTCATGTTGCACCGGAAGCAATGGTT
GGTGGTCCGATTGCAATTGTTGAAGATGGCGATACCATTGTGATTGATGTGGAAAGCGAAC
GTCTGGATCTGAAACTGAGCGAAGAAGAAATTAAAAATCGCCTGAAACGTTGGAGCCCGCC
GTCACCGCGTTATAAAAGCGGTCTGCTGGCAAAATATGCAAGCCTGGTTTCTCAGGCAAGC
ATGGGTGCAGTTACCCGTCCGGCATAA

6.2 Supplemental information: Molecular Dynamics Analysis of a Rationally Designed Aldehyde Dehydrogenase Gives Insights into Improved Activity for the Non-Native Cofactor NAD^+

Molecular dynamics analysis of a rationally designed aldehyde dehydrogenase gives insights into improved activity for the non-native cofactor NAD^+

Tobias J. Gmelch[1], Josef M. Sperl[1], Volker Sieber[1,2,3,4*]

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

[2] Catalysis Research Center, Technical University of Munich, Garching, Germany

[3] Fraunhofer Institute of Interfacial Biotechnology (IGB), Bio-, Electro- and Chemo Catalysis (BioCat) Branch, Straubing, Germany

[4] School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD, Australia

*Correspondence should be addressed to

Volker Sieber, Prof.

E-Mail: sieber@tum.de

Tel: +499421187301

Supporting Information

Calculation of screening coverage

Without further investigation, we assumed a yield of 50% for our mutagenesis approach. The codon-based coverage was then calculated applying the formula given below, with "x" representing the colonies analyzed and "a" the amount of possible codons (1xNNK = 32).[1]

$$Codoncoverage[\%] = 1 - (1 - \frac{1}{a})^{0.5x} \cdot 100\%$$

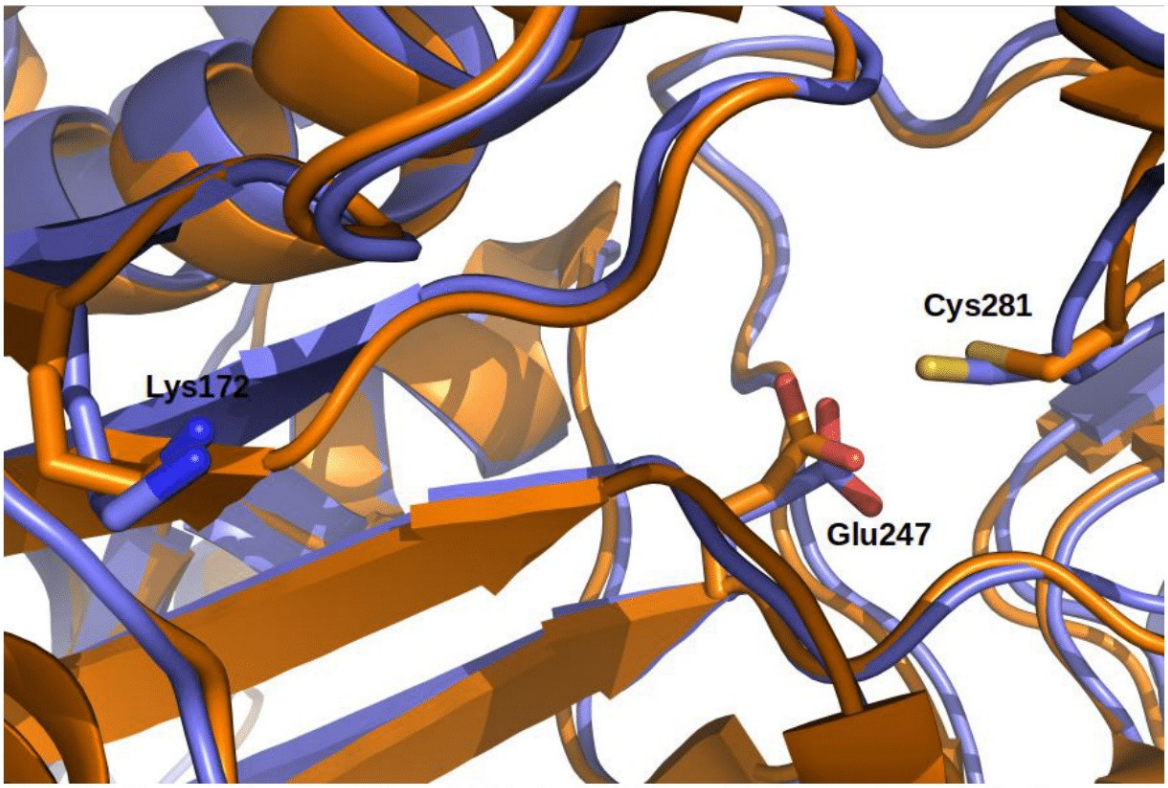
Another approach to describe the coverage is to calculate the probability of the occurrence of the most active variants within the library.[2] The probabilities P_{T1} and P_{T2}, which correspond to the detection probability of the best or one of the two best variants of the library, were calculated by

Supplemental Table 2: Comparison of CSR-SALAD results depending on the type of cofactor docked within the input structure. Residues in italics are suggested for mutagenesis for both input structures.

Docked cofactor	Selectivity determining	Medium priority	Low priority
NAD	*Lys172*	Gly205	*Asp176*
	Codon: AVK	Gly209	*Asp210*
	Ser174	*Thr229*	
	Codon: RSC	*Arg232*	
	Gly205	*Ile233*	
	Codon: RSA		
NADP	Leu145	*Asp210*	*Asp176*
	Codon: VWA	Val213	Glu207
	Lys172	*Thr229*	*Asp210*
	Codon: AVK	*Arg232*	
		Ile233	

Supplemental Table 3: Result of Ramachandran analysis via RAMPAGE. Underlined residues are found in both structures.

	5M4X	Model structure of M0
Total residues in structure	461	493
Allowed region	G140, K251, N315, E339, S340, Y361, D404, L409, E466,	K92, W114, L118, T126, K136, V143, A237, A239, K251, D261, Q279, A283, G311, E336, Y361, N370, I377, S393, D404, L409, R429, R451, M452, A456, S460
Outlier region	G408	P93, P178, P313, G408



Supplemental Figure 2: Overlay of M0 model (blue) with 2IMP crystal structure (orange). Position numbering based on TaALDH

Supplemental Table 4: Typical site-directed mutagenesis protocol

Vol [µL]	Stock	compound	final conc.
10	5x	HF-Phusion buffer	1x
1	10 mM	dNTPs	0.2 mM
2.5	10 µM	Each oligonucleotide	0.5 µM
0.5	< 300 ng/µL	DNA template	<3 ng/µL
0.5	2 U/µL	Phusion polymerase	0.02 U/µL
2.5	100%	DMSO	5 %

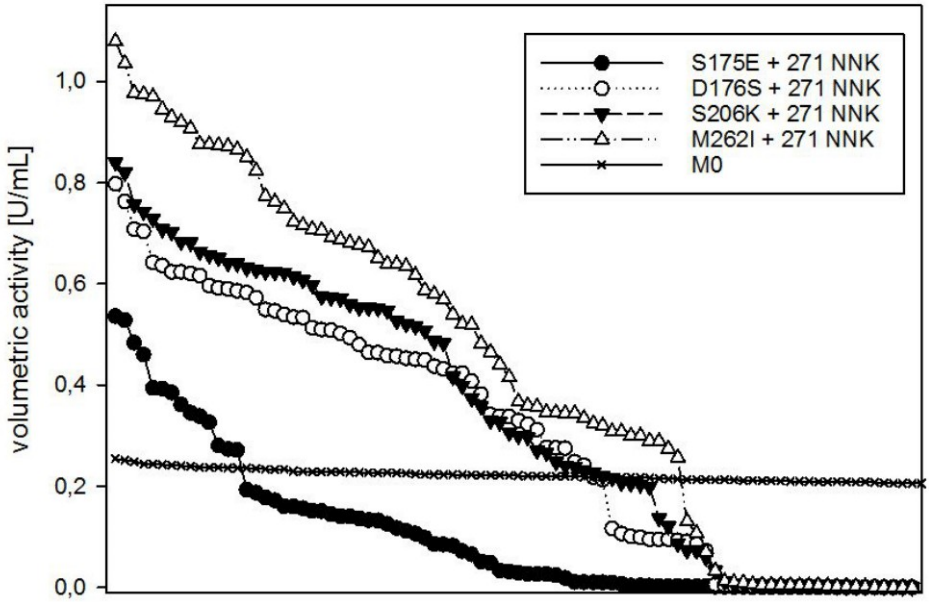
Supplemental Table 5: List of oligonucleotides. Bold letters indicate the modified position

#	Pos	Name	5'-3' sequence
1	172	172 NNK-f	GTAATACCGTTGTTCTG**NNK**CCGAGCAGCGATACAC
2	172	172 NNK-r	GTGTATCGCTGCTCGG**MNN**CAGAACAACGGTATTAC
3	174	174 NNK-f	GTTGTTCTGAAACCG**NNK**AGCGATACACCGGGTAG
4	174	174 NNK-r	CTACCCGGTGTATCGCT**MNN**CGGTTTCAGAACAAC
5	175	175 NNK-f	GTTGTTCTGAAACCGAGC**NNK**GATACACCGGGTAGC
6	175	175-NNK-r	GCTACCCGGTGTATC**MNN**GCTCGGTTTCAGAACAAC
7	176	176 NNK-f	CTGAAACCGAGCAGC**NNK**ACACCGGGTAGCGC
8	176	176 NNK-r	GCGCTACCCGGTGT**MNN**GCTGCTCGGTTTCAG
9	177	177 NNK-f	GAAACCGAGCAGCGAT**NNK**CCGGGTAGCGCAGAATG
10	177	177 NNK-r	CATTCTGCGCTACCCGG**MNN**ATCGCTGCTCGGTTTC
11	205	205 NNK-f	GAATTTTATTACCGGTCGT**NNK**AGCGAAATTGGCG
12	205	205 NNK-r	CGCCAATTTCGCT**MNN**ACGACCGGTAATAAAATTC
13	206	206 NNK-f	GAATTTTATTACCGGTCGTGGT**NNK**GAAATTGGCGATTAC
14	206	206 NNK-r	GTAATCGCCAATTTC**MNN**ACCACGACCGGTAATAAAATTC
15	207	207 NNK-f	CCGGTCGTGGTAGC**NNK**ATTGGCGATTACATTG
16	207	207 NNK-r	CAATGTAATCGCCAAT**MNN**GCTACCACGACCGG
17	209	209 NNK-f	CGTGGTAGCGAAATT**NNK**GATTACATTGTGGAAC
18	209	209 NNK-r	GTTCCACAATGTAATC**MNN**AATTTCGCTACCACG
19	210	210 NNK-f	CGTGGTAGCGAAATTGGC**NNK**TACATTGTGGAAC
20	210	210 NNK-r	GTTCCACAATGTA**MNN**GCCAATTTCGCTACCACG
21	229	229 NNK-f	GACCGGTAGCACCGCA**NNK**GGTCAGCGCATTATGC
22	229	229 NNK-r	GCATAATGCGCTGACC**MNN**TGCGGTGCTACCGGTC
23	232	232 NNK-f	GCACCGCAACAGGTCAG**NNK**ATTATGCAGAAAGC
24	232	232 NNK-r	GCTTTCTGCATAAT**MNN**CTGACCTGTTGCGGTGC
25	262	262-NNK-f	GGAAAGATGCCGAT**NNK**GATAATGCACTGAAAACCC
26	262	262-NNK-r	GGGTTTTCAGTGCATTATC**MNN**ATCGGCATCTTTCC
27	275	275 NNK-f	CCCTGCTGTGGGCAAAATAT**NNK**AATGCCGGTCAGAGC
28	275	275-NNK-r	GCTCTGACCGGCATT**MNN**ATATTTTGCCCACAGCAGGG
29	306	306 NNK-f	GTTTTGTTGAACTGAGC**NNK**AAACTGGCACTGGGTG

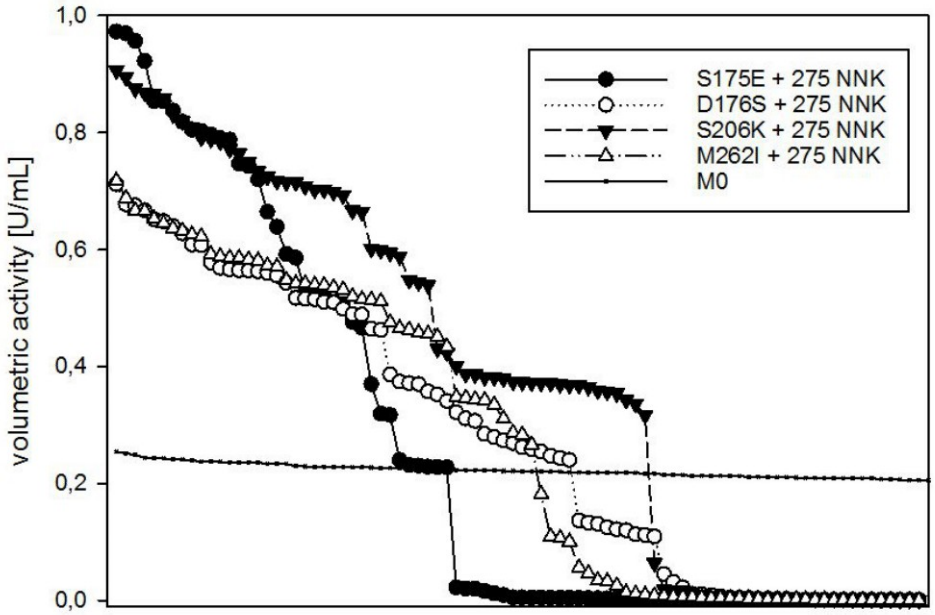
30	306	306 NNK-r	CACCCAGTGCCAGTTT**MNN**GCTCAGTTCAACAAAAC
31	349	349 NNK-f	CTGTTTGGTGGT**NNK**CAGCCGAGCCTGAGCGGTC
32	349	349 NNK-r	GACCGCTCAGGCTCGGCTG**MNN**ACCACCAAACAG
33	271	271-NNK-f	GCACTGAAAACCCTGCTG**NNK**GCAAAATATTGGAATGCC
34	271	271-NNK-r	GGCATTCCAATATTTTGC**MNN**CAGCAGGGTTTTCAGTGC
35	271 + 275	271+275 NNK-f	CCCTGCTG**NNK**GCAAAATAT**NNK**AATGCCGGTCAG
36	271 + 275	271+275 NNK-r	CTGACCGGCATT**MNN**ATATTTTGC**MNN**CAGCAGGG
37	175	S175E-f	GTTGTTCTGAAACCGAGC**GAG**GATACACCGGGTAGC
38	175	S175E-r	GCTACCCGGTGTATC**CTC**GCTCGGTTTCAGAACAAC
39	176	D176S-f	CTGAAACCGAGCAGC**TCT**ACACCGGGTAGCGC
40	176	D176S-r	GCGCTACCCGGTGT**AGA**GCTGCTCGGTTTCAG
41	206	S206K-f	GAATTTTATTACCGGTCGTGGT**AAG**GAAATTGGCGATTAC
42	206	S206K-r	GTAATCGCCAAT**TTC**CTTACCACGACCGGTAATAAAATTC
43	262	M262I-f	GGAAAGATGCCGAT**ATT**GATAATGCACTGAAAACCC
44	262	M262I-r	GGGTTTTCAGTGCATTATC**AAT**ATCGGCATCTTTCC
45	175 + 176	S175E + D176S-f	CTGAAACCGAGC**GAGTCT**ACACCGGGTAGCGC
46	175 + 176	S175E + D176S-r	GCGCTACCCGGTGT**AGACTC**GCTCGGTTTCAG

Supplemental Table 6: Amino acid at position 271 or 275 of the most active variant depending on the genetic background. All positions were mutated separately at either 271 or 275 and the two most active variants were sequenced, *M33 was mutated simultaneously at both positions

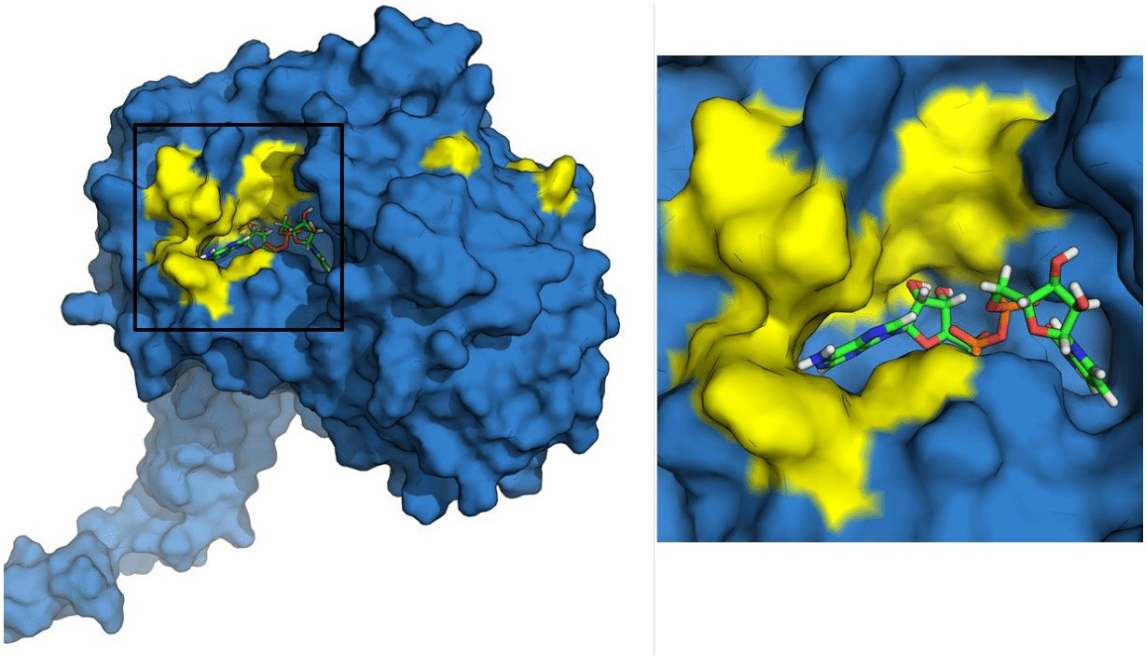
Template	W271	W275
M0	Y/L	T/L
S175E	W	N/T
D176S	Y	S/I
S206K	T/N	L
M262I	L/I	M/L
M33*	Y	V



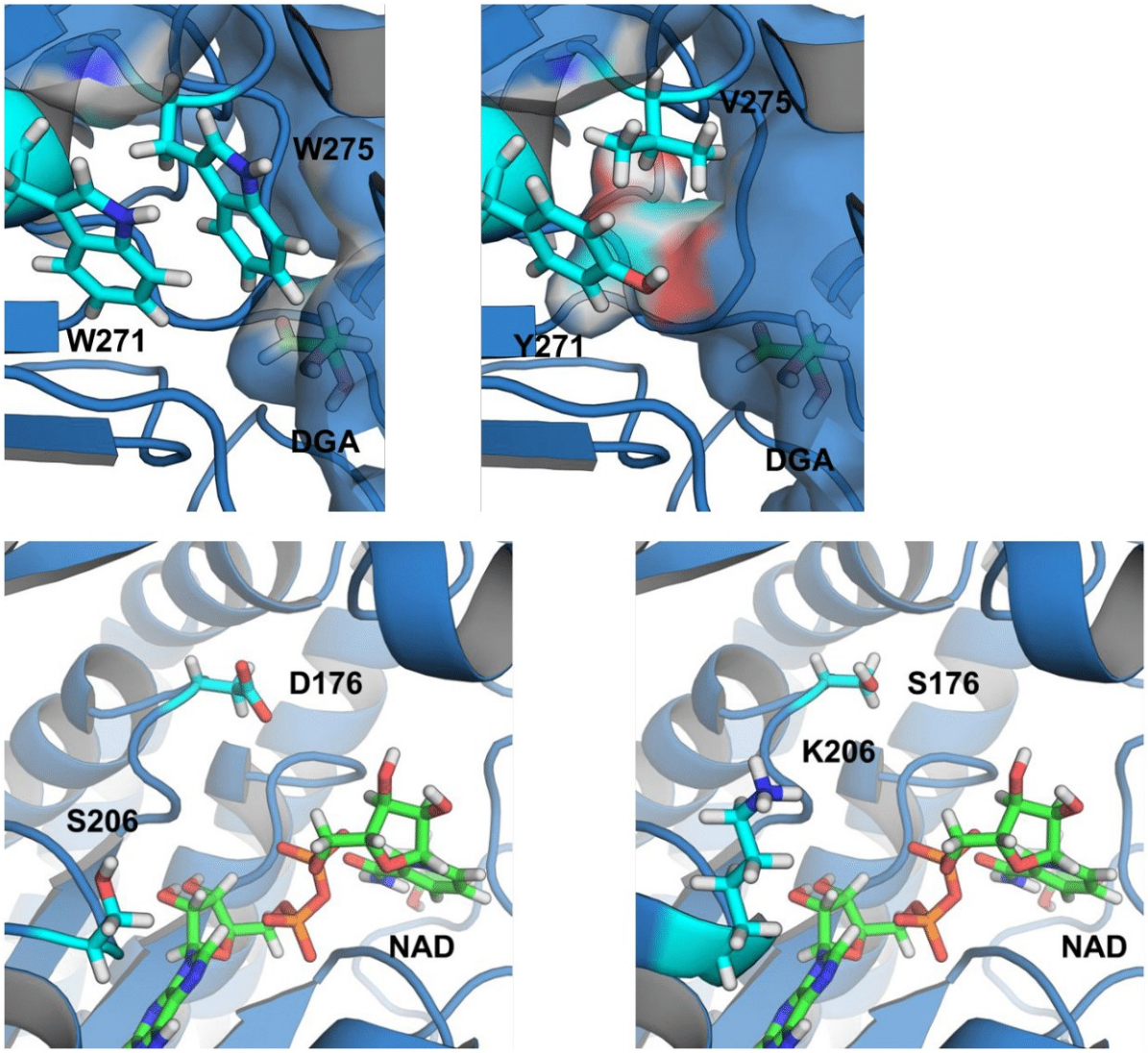
Supplemental Figure 3: Mutant landscape for single mutant in combination with 271 NNK compared to template activity (M0). The reactions were carried out at room temperature with 100 mM HEPES pH 7.35, 100 mM ammonium sulfate pH 7, 1.3 mM DGA and 1 mM NAD^+ in a final volume of 200 µL.



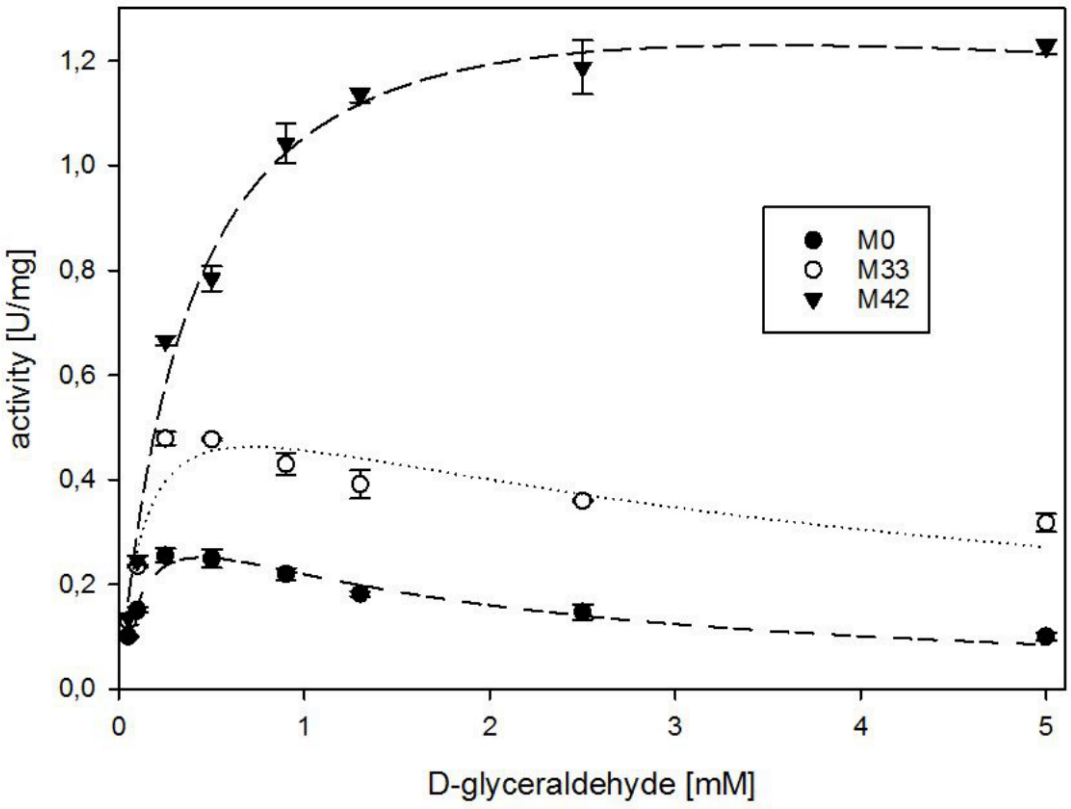
Supplemental Figure 4: Mutant landscape for single mutant in combination with 275 NNK compared to template activity (M0). The reactions were carried out at room temperature with 100 mM HEPES pH 7.35, 100 mM ammonium sulfate pH 7, 1.3 mM DGA and 1 mM NAD^+ in a final volume of 200 µL.



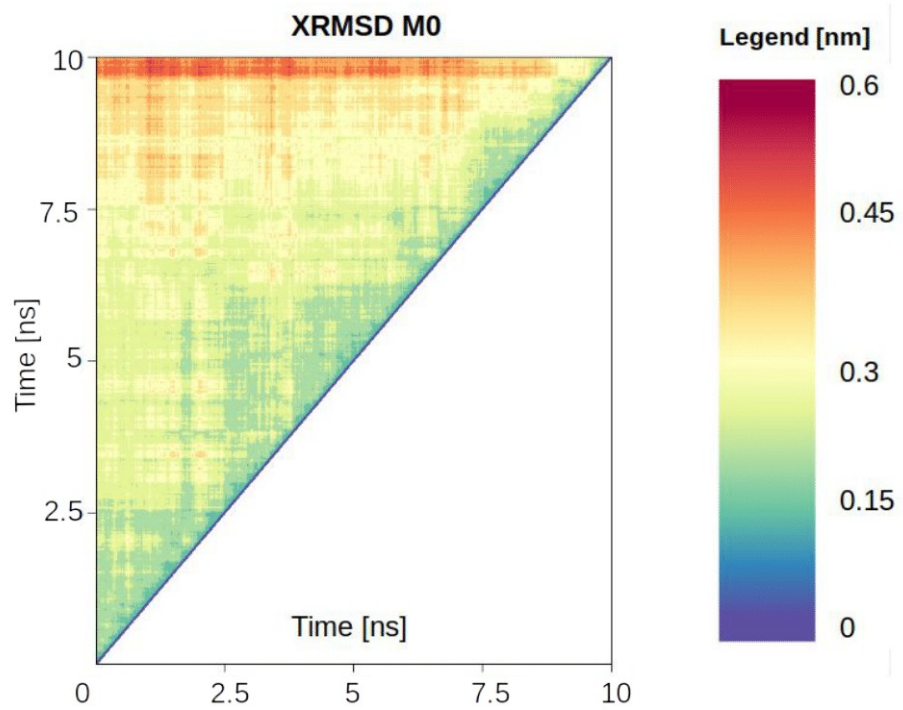
Supplemental Figure 5: 3D structure model of TaALDH (M0) with docked NAD^+. Positions selected for mutations are colored in yellow.



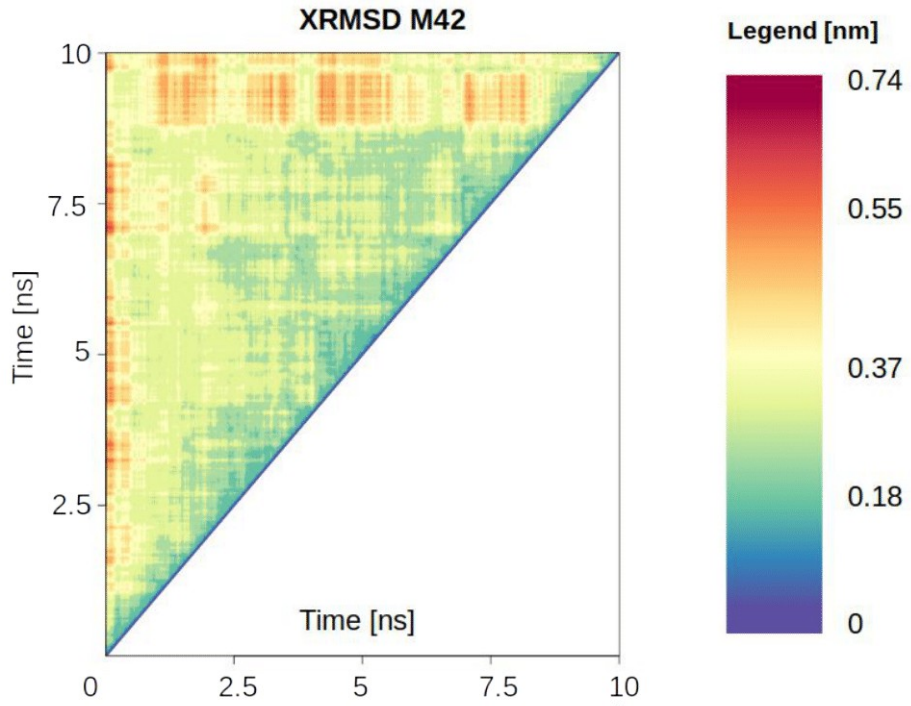
Supplemental Figure 6: Illustrative view of amino acid exchanges with the template M0 on the right side and the mutant M42 on the left. The exchanged amino acids are colored in light blue and NAD^+ and DGA are in green.



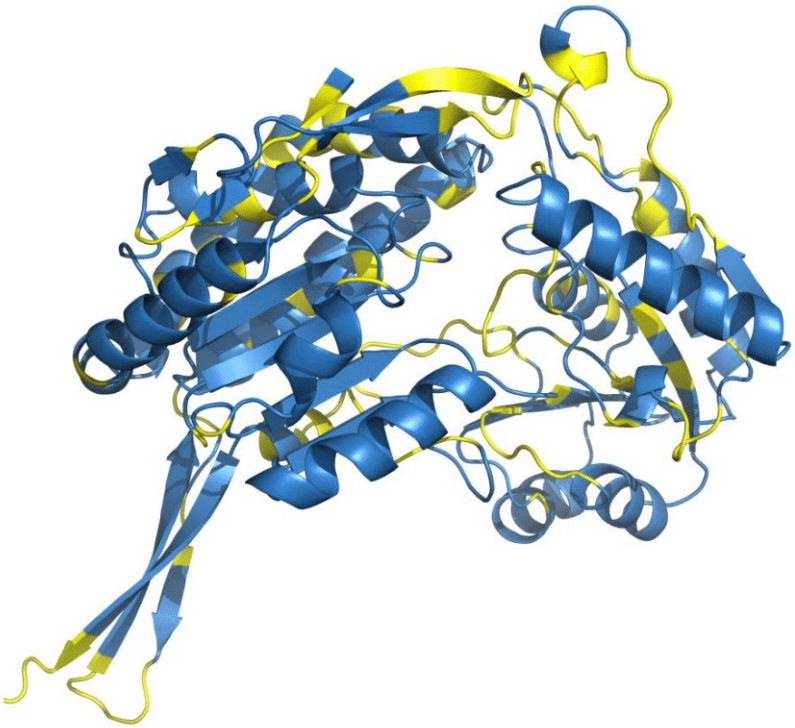
Supplemental Figure 7: Kinetic curves for template (M0) and engineered variants of TaALDH. The measurements were carried out in triplicate at 50 °C containing 100 mM HEPES pH 7.35 and 100 mM ammonium sulfate pH 7. NAD^+ saturation kinetics were determined with 1.3 mM DGA and DGA saturation measurements contained 1 mM NAD^+ in a final volume of 200 µL.



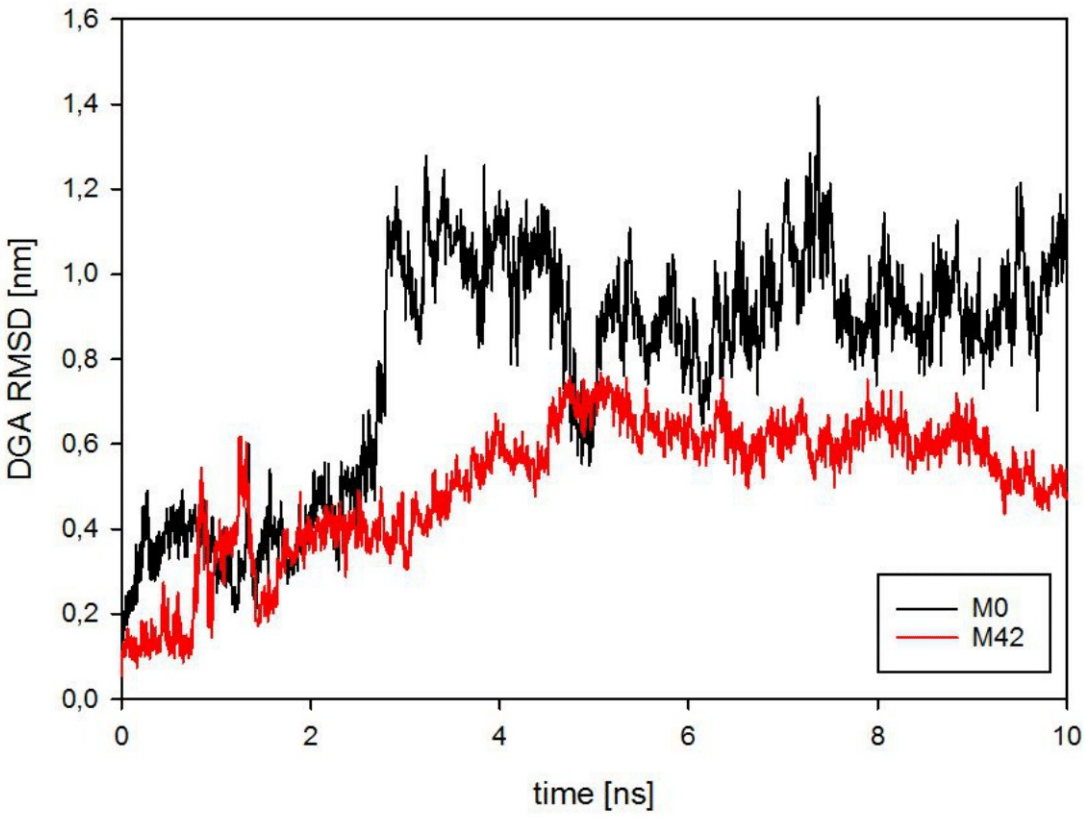
Supplemental Figure 8: Cross RMSD of M0 plotted via MDplot.[5] For the calculations all non-hydrogen atoms were considered.



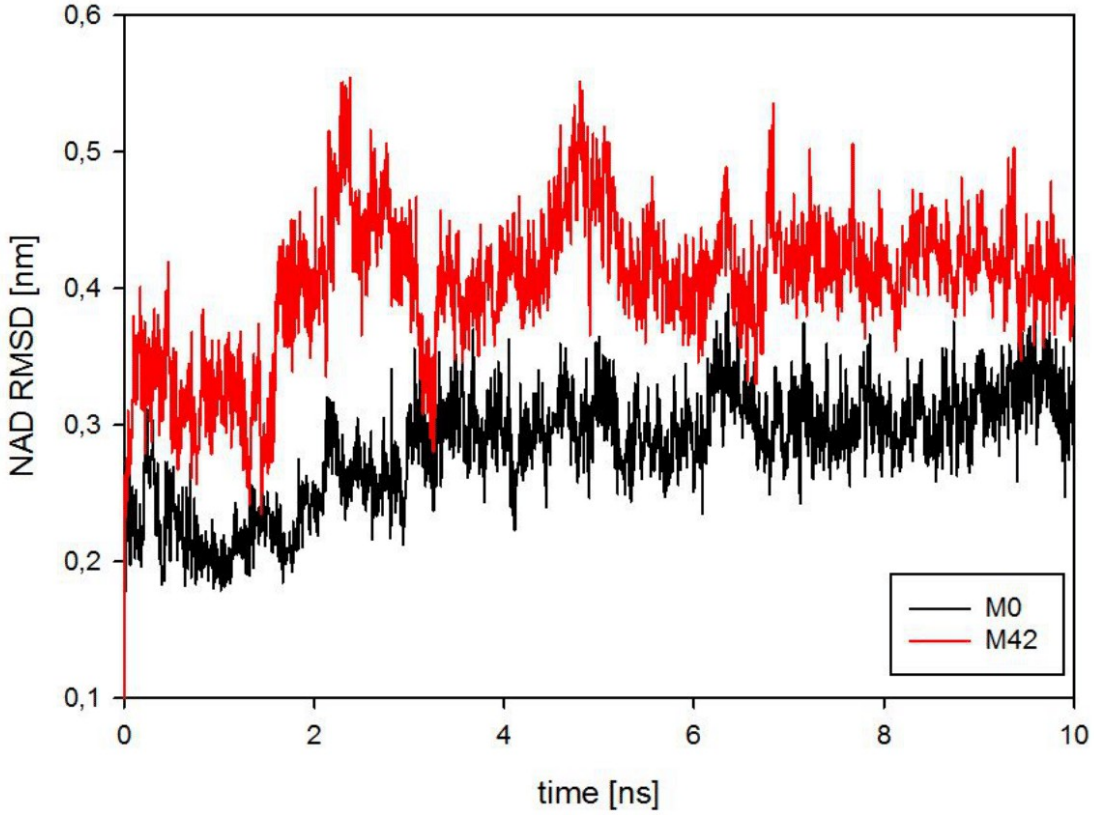
Supplemental Figure 9: Cross RMSD of M42 plotted via MDplot.[5] For the calculations all non-hydrogen atoms were considered.



Supplemental Figure 10: Graphical visualization of the delta RMSF plot. Regions where the RMSF of M0 exceeds the RMSF of M42 are indicated in yellow.



Supplemental Figure 11: RMSD of DGA in comparison to the protein backbone.



Supplemental Figure 12: RMSD of NAD^+ in comparison to the protein backbone.

Supplemental References

[1] Bosley, A. D., and Ostermeier, M. (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries, *Biomol. Eng. 22*, 57–61.

[2] Nov, Y. (2012) When second best is good enough: another probabilistic look at saturation mutagenesis, *Appl. Environ. Microbiol. 78*, 258–262.

[3] Cahn, J. K. B., Werlang, C. A., Baumschlager, A., Brinkmann-Chen, S., Mayo, S. L., and Arnold, F. H. (2017) A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases, *ACS Synth. Biol. 6*, 326–333.

[4] Koppaka, V., Thompson, D. C., Chen, Y., Ellermann, M., Nicolaou, K. C., Juvonen, R. O., Petersen, D., Deitrich, R. A., Hurley, T. D., and Vasiliou, V. (2012) Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application, *Pharmacol. Rev. 64*, 520–539.

[5] Margreitter, C., and Oostenbrink, C. (2017) MDplot: Visualise molecular dynamics, *R J. 9*, 164.

6.3 Sequences of gluconate dehydratase

Sequences were cloned into pET28a-vectors in order to obtain: NHis-TAG-TEVsite-GAD

SsGAD

ATGAGAATCAGAGAAATAGAACCAATAGTACTCACCTCGAAAGAGAAAGGAAGTGCAACTTGGGCATCTA
TAATGATTGTCACAAGGGTCATTACGGAAAATGGGGAAGTAGGCTATGGTGAGGCAGTACCCACACTAAG
AGTTATATCTGTATATAACGCAATTAAACAAGTTAGTAAGGCTTATATAGGGAAAGAGGTAGAGGAAGTTGA
GAAGAACTATCATGAATGGTATAAACAAGATTTCTATTTAGCTAGGTCTTTTGAATCAGCAACTGCAGTAAG
TGCAATCGATATAGCCTCATGGGATATAATAGGGAAAGAGCTTGGAGCACCAATTCATAAATTATTAGGAGG
AAAAACCAGGGATAGGGTACCAGTCTACGCAAACGGATGGTATCAGGACTGCGTAACTCCAGAGGAATTT
GCGGAAAAGGCAAAAGACGTTGTAAAGATGGGATATAAGGCTTTAAAATTTGATCCGTTTGGTCCATATTA
CGATTGGATAGATGAGAGAGGTCTAAGAGAAGCTGAGGAGAGAGTAAAGGCTGTTAGAGAGGCAGTTGG
AGACAACGTGGATATTTTAATAGAGCATCACGGTAGGTTTAATGCGAATTCGGCTATTATGATAGCGAAAAG
ATTGGAAAAATACAATCCGGGATTTATGGAGGAACCGGTACATCATGAGGACGTAATTGGTTTAAGAAAGT
ATAAAGCCAGTACTCATTTAAGGGTTGCATTGGGAGAAAGACTGATAAGTGAAAAGGAAACTGCGTTTTAC
GTTGAGGAAGGTCTTGTAAACATATTGCAACCAGATTTAACTAATATAGGTGGTGTAACAGTAGGTAGGAG
TGTTATAAAAATAGCTGAAGCTAATGATGTAGAGGTGGCTTTTCACAACGCCTTTGGTTCAATACAGAATG
CAGTTGAAATACAACTAAGTGCAGTTACACAGAATTTGTATTTACTTGAGAACTTCTATGATTGGTTCCCTC
AGTGGAAAAGGGATTTAGTATATAATGAAACGCCAGTTGAAGGAGGTCACGTTAAGGTTCCATACAAGCC
TGGACTAGGTGTTTCAATTAATGAAAAAATAATAGAACAGCTAAGAGCTGAACCAATACCATTAGATGTAAT
TGAAGAACCGGTTTGGGTCGTCAAGGGAACCTGGAAGAATTATGGTGTTTGA

PtGAD

ATGGAGACAATAAAAAGCGTAGATATATACGAGCTTGGATCTCCCGGGGAAAAATCATCTCCATGGAGCTC
AACAATCCTGATAGTTAAGCTGACATCGTCCAATGGCAACATTGGTTATGGAGAGGCACCAACAACGTTC
ATGACGCTTCCTGTTAAGGAAAGCATGCGCGAGGTTGAGCGTGTCTTTAAGGACCAGAATTATTTTAATAT
AGAAAAGAACATGCGCGAGTTTTATAAGCATTCATTTTACCTATCAAGATCCATGGAGGCAACATCAGCCC
TAAGTGCATTTGAAATAGCCTCATGGGATTTAATAGGCAAGGATCTTGGTACGCCCGTGTATAACCTGCTT
GGTGGTGAATACAATTCTGAGCTCAGGGCCTATGCCAATGGCTGGTACTCTGACTGTTTGGAACCTGATG
ACTTTGTTTCAAGGGCAAAGGAATACATAAAAAAGGGATATACCGCATTTAAGTTCGATCCATTTAGAAACA
ACTTTGACAGGATAGGCAACGATGGTATAAAAAAGGCCGTTGACATAGTCTCGGCAATGAGGTCAGAGCT
TGGTGAAAATATAGATCTTTTAATAGAATGCCATGGAAGGTTCTCAACAAAGTATGCAATAAAGGTTGGCCA
GGCACTTGATGAGTTCAATCCGTTATTTATAGAGGAGCCAATACATCCTGAGATGGAGCTGGGCCTCTTT
GATTTTAAAAGGTATGTAAATACGCCGGTTGCACTTGGTGAGAGGCTTTTAAACAAGGAGGATTTTGCAAG
GTATATATCACAGGGCATGGTCGACATAGTACAGGCAGATCTAACAAATTCAAAGGGAATACTTGAGGCAA
AGAAGATCTCTGCAATAGTTGAATCCTTTGGAGGCCTCATGGCATTTCATAATGCCTTTGGACCGGTTCAG
ACTGCCGCAACGTTAAACGTTGATTACACACTGACCAATTTTTTAATACAGGAAAGCTTTGAGGACTCATG
GCCTGACTGGAAGAGAAATCTTTTCTCAGGATATAGGATAGAAAACGGTCATTTCAAACTTTCAGGGAAAC
CGGGGCTTGGCATAACAGCAGATGAAAAATTAATGGAAAAATTAATTTATGATGGCATGGAGGAATTCAAT
AAAAACGAGCCATCATGGGTTGTCTCTGGAACATACAAA

His-GST-Tag

ATGAAACATCACCATCACCATCACCCCATGAGCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCA
ACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTG
ATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATG
TTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAA
AAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGC
ATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGA
AGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCT
CTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGT
ATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTG
GCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA

His-MBP-Tag

ATGGCTCATCACCATCACCATCACCCCATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCG
ATAAAGGCTATAACGGCCTGGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGT
TGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTAT
CTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAA
AGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTAC
CCGATCGCTGTTGAAGCGTTATCGCTGATTTACAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGG
GAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAA
GAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAG
TACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCT
GATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAA
CAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAA
CGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACG
CCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGG
AAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAG
ATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTC
CGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAG
CCCTGAAAGACGCGCAGACT

His-SUMO-Tag

ATGGGAATGAATTGGAGCCACCCGCAGTTCGAAAAAAGCAGCGGCAGCAGCGGCGGTCATCACCATCA
TCATCACGGCGGCAGCGGCGGCAGCGGGTCGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTC
AAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCA
AGATCAAAAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGG
ACTCCTTAAGATTCTTGTACGACGGTATTAGAATTCAAGCTGATCAGGCCCCTGAAGATTTGGACATGGAG
GATAACGATATTATTGAGGCTCACCGCGAACAGATTGGAGGT

TEV-cleavage site

GAGAATCTTTATTTTCAGGGC

7 Abbreviations

%	Per cent
°C	Degree celsius
ADH	Alcohol dehydrogenase
AfAlaDH	Alanine dehydrogenase from *Archaeoglobus fulgidus*
Ala/ A	Alanine
ALDH	Aldehyde dehydrogenase
Arg/ R	Arginine
Asn/ N	Asparagine
Asp/ D	Aspartic acid
AspD	Aspartate decarboxylase
ATP	Adenosine triphosphate
BsGDH	Glucose dehydrogenase from *Bacillus subtilis*
CcDHAD	Dihydroxyacid dehydratase from *Caulobacter crescentus*
CO_2	Carbon dioxide
CoA	Coenzyme A
CSR-SALAD	Cofactor Specificity Reversal–Structural Analysis and Library Design
Csy/ C	Cysteine
ddH2O	Double destilled water
DGA	D-glyceraldehyde
DHAD	Dihydroxyacide dehydratase
E. coli	*Escherichia coli*
e.g.	For example
EC	Enzyme class
ε_{NADH}	Absorption coefficient of NADH
eq	equivalent
FAD	Flavin Adenine Dinucleotide
FRET	Förster resonance energy transfer
g	gram
g/g	Gram per gram
g/L	Gram per liter
g/Lh	Gram per liter and hour
Gln/ Q	glutamine

Glu/ E	Glutamic acid
Gly/ G	glycine
GSR	Gluthathione reductase
h	hour
h^{-1}	Per hour
H2O	water
HCN	Hydrogen cyanide
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic acid
His/ H	Histidine
HPLC	High-performance liquid chromatography
HT	Hydride transfer
HY	Hydrolysis
Ile/ I	Isoleucine
KARI	Ketol-acid reductoisomerase
k_{cat}	Catalytic constant
KCN	Potassium cyanide
KDG	keto-deoxygluconate
KDGA	Keto-deoxygluconate aldolase
K_M	Michaelis–Menten Constant
LB	Lysogeny broth
λ_{em}	Wavelength of emission
Leu/ L	Leucine
λ_{ex}	Wavelength of absorption
Lys/ K	Lysine
M	Molar or mol per liter
MBP	Maltose binding protein
MD	Molecular dynamics
Met/ M	Methionine
MjALDH	Aldehyde dehydrogenase of Methanocaldoccocus janaschii
mm	Milli meter
mM	Milimolar or millimole per liter
MTP	Microtiter plate
NaCN	Sodium cyanide
NAD	Nicotinamide Adenine Dinucleotide
NADH	Dihydro Nicotinamide Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate

NADPH	Dihydro Nicotinamide Adenine Dinucleotide Phosphate
NaOH	Sodium hydroxide
NEB	New England Biolabs
NH3	ammonia
nm	Nano meter
ns	Nano second
NTP	Nucleoside triphosphate
o-	ortho-
OPA	ortho-phthaldialdehyde
PCR	Polymerase chain reaction
PDC	Pyruvate decarboxylase
PHB	polyhydroxybutyrate
Phe/ F	Phenylalanine
Pro/ P	Proline
ps	Pico second
PtKDGA	Keto-deoxygluconate aldolase from *Picrophilus torridus*
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
rpm	Rounds per minute
Ser/ S	Serine
SsDHAD	Dihydroxyacid dehydratase from *Sulfolobus solfataricus*
SUMO	Small ubiquitin-related modifier
TaALDH	Aldehyde dehydrogenase from *Thermoplasma acidophilum*
Thr/ T	Threonine
Trp/ W	Tryptophane
Tyr/ Y	Tyrosine
U	Unit / Enzyme Activity µmol/min
Val/ V	Valine
v_{max}	Maximum Reaction Rate
Vol-%	Volume per cent
VWR	VWR International GmbH
w/v	Weight per volume
wt-%	Weight per cent
YASARA	Yet Another Scientific Artificial Reality Application
µL	Micro liter

8 List of Figures

9 List of Tables