Development of a Highly Selective NAD+-Dependent Glyceraldehyde Dehydrogenase and its Application in Minimal Cell-Free Enzyme Cascades

Mariko Teshima,[a] Samuel Sutiono,[a, b] Manuel Döring,[a] Barbara Beer,[a, c] Mikael Boden,[d] Gerhard Schenk,[d, e, f] and Volker Sieber*[a, d, g, h]

Anthropogenic climate change has been caused by over-exploitation of fossil fuels and CO_2 emissions. To counteract this, the chemical industry has shifted its focus to sustainable chemical production and the valorization of renewable resources. However, the biggest challenges in biomanufacturing are technical efficiency and profitability. In our minimal cell-free enzyme cascade generating pyruvate as the central intermediate, the NAD+-dependent, selective oxidation of D-glyceraldehyde was identified as a key reaction step to improve the overall cascade flux. Successive genome mining identified one candidate enzyme with 24-fold enhanced activity and another whose stability is unaffected in 10% (v/v) ethanol, the final product of our model cascade. Semi-rational engineering improved the substrate selectivity of the enzyme up to 21-fold, thus minimizing side reactions in the one-pot enzyme cascade. The final biotransformation of D-glucose showed a continuous linear production of ethanol (via pyruvate) to a final titer of 4.9% (v/v) with a molar product yield of 98.7%. Due to the central role of pyruvate in diverse biotransformations, the optimized production module has great potential for broad biomanufacturing applications.

Introduction

Fossil resources are finite and non-renewable on a human time scale.[1] However, according to the U.S. Energy Information Administration's (EIA) World Primary Energy Consumption report, the global economy remains heavily dependent on fossil fuels, which are estimated to account for 83.5% of the total primary energy consumption in 2021.[2] Although recent technological advances have led to a positive trend in the renewable energy sector, with 21 PWh of renewable energy produced in 2021, fossil fuel consumption has only decreased by 2.3% compared to 2001. This is largely due to the fact that total global primary energy consumption has increased from 120 to 180 PWh over the last 20 years.[2] Due to this excessive energy demand, anthropogenic climate change has led to a 1.1 °C increase in global surface temperature (the Synthesis Report for the Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change) and it calls for urgent near-term action to achieve net-zero CO_2 emissions.[3]

The technology of the 20th century was based on the efficient utilization of fossil resources, which were abundantly available at low costs.[4] In addition to electricity, heat, and

[a] M. Teshima, Dr. S. Sutiono, M. Döring, Dr. B. Beer, Prof. Dr. V. Sieber
Chair of Chemistry of Biogenic Resources
Technical University of Munich
Campus Straubing for Biotechnology and Sustainability, Schulgasse 16
94315 Straubing (Germany)
E-mail: Sieber@tum.de
Homepage: https://cbr.cs.tum.de/

[b] Dr. S. Sutiono
Current address:
CarboCode Germany GmbH
Byk-Gulden-Straße 2, 78467 Constance (Germany)

[c] Dr. B. Beer
Current address:
CASCAT GmbH
Europaring 4, 94315 Straubing (Germany)

[d] Prof. Dr. M. Boden, Prof. Dr. G. Schenk, Prof. Dr. V. Sieber
School of Chemistry and Molecular Biosciences
The University of Queensland
68 Cooper Rd, St. Lucia, 4072 Brisbane (Australia)

[e] Prof. Dr. G. Schenk
Australian Institute for Bioengineering and Nanotechnology
The University of Queensland
Corner of College and Cooper Rds, St. Lucia, 4072 Brisbane (Australia)

[f] Prof. Dr. G. Schenk
Sustainable Minerals Institute
The University of Queensland
Corner of College and Staff House Rds, St. Lucia, 4072 Brisbane (Australia)

[g] Prof. Dr. V. Sieber
SynBioFoundry@TUM
Technical University of Munich
Schulgasse 22, 94315 Straubing (Germany)

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

▢ Supporting information for this article is available on the WWW under https://doi.org/10.1002/cssc.202301132

ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

Chemistry
Europe
European Chemical
Societies Publishing

transportation fuels, more than 90% of the carbon-containing materials of daily use (solvents, paints, plastics, and other polymers) were produced from coal, natural gas, and crude oil.[4–5] However, in the post-peak oil era, sustainable alternatives to the conventional petrochemical industry are required to prevent the over-exploitation of natural resources.[1,5a] The concept of second-generation biorefinery to valorize biomass waste streams, especially lignocellulosic biomass, is one such alternative that also promises a sustainable socio-economic growth.[6] For instance, the substitution of conventional fossil fuels by bioethanol is estimated to reduce greenhouse gas emissions by up to 80%.[7]

The whole-cell approach is currently the most widely used technology used in the biomanufacturing industry. However, establishing a synthetic pathway in microbial strains has remained challenging due to the complexity of the endogenous metabolism.[8] As an alternative, cell-free enzyme-based cascades (dating back to the ethanol production with yeast cell lysates by Buchner and coworkers in 1897)[9] have gained attention for the bioproduction of an increasing number of chemicals.[9–10] A major advantage of this approach is that it overcomes physiological constraints associated with living cells, such as the redirection of substrates or intermediates to cell metabolism or low tolerance to reaction products.

A cell-free artificial minimal enzyme cascade for the valorization of glucose *via* the central intermediate, pyruvate, to ethanol or isobutanol was first described by Sieber and coworkers in 2012 (Figure 1).[10b] One main characteristic of the system is the minimal pyruvate generation module that forms the core of the cascade, using only four enzymes and NAD+/NADH as the sole redox cofactor. This significant reduction in the number of required enzymes and cofactors required (by comparison, the Embden-Meyerhof-Parnas (EMP) pathway employs ten enzymes and two cofactors) was achieved by

Figure 1. Cell-free minimal cascades for D-glucose valorization. (a) Ethanol and isobutanol cascades proposed by Guterl and coworkers (the pathways branching to the left (black) and to the right (gray), respectively).[10b] (b) GADH performing the key reaction of the cascade (wild-type GADH from *Thermoplasma acidophilum* with NADP+; PDB: 5IZD). This enzyme redirects the reactive by-product D-glyceraldehyde, derived from the aldol cleavage by KDGA, *via* a DHT reaction to pyruvate (green circle).

exploiting the "non-phosphorylative" Entner-Doudoroff pathway and enzyme promiscuity.[10d,11] Recently, Bowie and coworkers[10c] presented an opposite approach using a 16-membered cell-free enzyme cascade for isobutanol production. This huge cascade reconstituted the EMP pathway for in vitro pyruvate generation and extended it with the isobutanol module.[12] Of the two expensive cofactors used (NADP+/NADPH and ATP), ATP was depleted during operation, requiring the introduction of an ATP rheostat system to keep the cascade in balance. In addition, competitiveness has remained critical due to the high cost of enzyme production and the still small scale of the process. Motivated by these challenges, we reconsidered the great potential of the minimal cascade, which is independent of phosphorylation and ATP as an energy supply. We sought to unlock the key to the efficient operation of the minimal pyruvate generation module. Since pyruvate is a central molecule in the formation of alcohols, as well as carboxylic acids and amino acids, our future perspective is to adapt this module for the production of a wide range of chemicals.

However, the initial cell-free minimal cascades displayed modest yields, converting 25 mM D-glucose to 57% and 41% of ethanol and isobutanol, respectively. The oxidation of D-glyceraldehyde by aldehyde dehydrogenase (ALDH) was identified as a critical step in the cascade.[13] Firstly, ALDH is involved in the oxidation of the highly reactive intermediate D-glyceraldehyde to D-glycerate,[14] which is further converted to pyruvate by a sugar acid-specific dehydratase (DHT).[15] Secondly, the NAD+-dependent reaction catalyzed by ALDH is tightly involved in the redox cofactor regeneration system, thus influencing the overall flux of the cascade.

ALDH (EC 1.2.1.3/1.2.1.4/1.2.1.5) is an enzyme class belonging to the oxidoreductases, oxidizing a broad range of aliphatic and aromatic aldehydes to form the corresponding carboxylic acids.[16] Members of this class are ubiquitously found across prokaryotes, eukaryotes and archaea, in various subcellular locations, and are involved in diverse physiological functions, such as detoxification, biosynthesis, antioxidant functions, non-catalytic binding of endogenous and chemotherapeutic compounds, structural and regulatory mechanisms, but also in other catalytic functions (*e.g.* ester hydrolysis, nitrate reduction).[17] Currently, nearly 400,000 ALDH entries are available in the UniProt database.[18] Since ALDH malfunctions have been linked to several clinically relevant diseases, human ALDHs and their roles in endogenous and xenobiotic metabolism have been extensively investigated.[19] In contrast, prokaryotic ALDHs are still less characterized, which limits the design of novel variants by (semi-)rational methods.[16] Of particular interest for applications in one-pot enzyme cascades are ALDH variants with high specificity for the oxidation of D-glyceraldehyde. Previous studies identified candidate glyceraldehyde dehydrogenases (GADHs) in the thermoacidophilic euryarchaeota *Picrophilus torridus* and *Thermoplasma acidophilum*.[20] However, their activity and producibility were low, limiting their applicability.

Here, we employ an iterative approach using genome mining and enzyme engineering to develop NAD+-dependent GADH with significantly improved D-glyceraldehyde selectivity
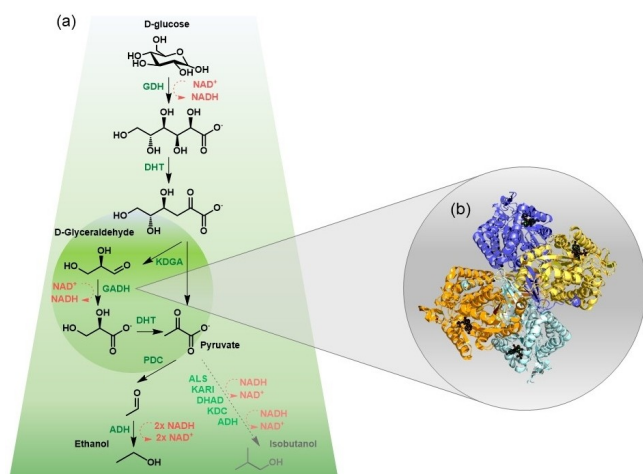
and suppressed activity toward the naturally favored substrate acetaldehyde. We also demonstrate that the integration of improved GADH variants into the multi-enzyme cascade greatly increases product (*i.e.*, ethanol) yield and titer, thereby significantly enhancing the potential of this reaction for industrial applications.

Results and Discussion

Challenges of GADH Engineering

For a one-pot multi-enzymatic cascade with several aldehyde species as intermediates, undesired side reactions need to be minimized. In the case of the minimal ethanol cascade, acetaldehyde is a highly favored substrate for ALDHs derived from the alcohol detoxification pathway, leading to the formation of acetic acid, which, once formed, cannot easily re-enter the cascade.[13b]

Our initial cascade experiments using the ALDH variant M42 from *Thermoplasma acidophilum* (*Ta*ALDH),[13c] achieved only a 53% molar ethanol yield (Figure S1). Furthermore, only 70% of the cumulative molar mass could be traced by HPLC at the end of the reaction, indicating a significant unspecific carbon loss. During the first four hours of the reaction, 50 mM D-glucose was rapidly consumed (~10 mM h^{-1}). However, the increase in the ethanol concentration was only 8 mM h^{-1}, which is only 40% of the maximum expected molar product conversion rate (two molecules of ethanol are generated from one molecule of glucose). In addition, the concentration of D-glyceraldehyde peaked at 17 mM after four hours, which is ~34% of the maximum theoretically possible D-glyceraldehyde concentration. After four hours, D-glucose consumption and ethanol production slowed down significantly, and after another three to four hours, ethanol production ceased altogether. The concentration of D-glyceraldehyde decreased slowly by 13 mM to 4 mM 1 hours after the cascade was initiated, while the concentration of D-glycerate increased by 11 mM. The observation of a relatively rapid consumption of D-glucose, followed by an accumulation of D-glyceraldehyde and a significantly slower formation of D-glycerate also highlights the difference in efficiency between glucose dehydrogenase (GDH) and ALDH (Figure S1). The GDH-catalyzed reaction is significantly favored, as GDH has a 15-fold lower K_m value than ALDH for NAD$^+$ (0.08 mM for GDH and 1.20 mM for *Ta*ALDH M42, respectively) and a high V_{max} (318 U mg^{-1}) (Tables S10 and S13). In contrast, the *Ta*ALDH variant M42 has lower solubility and lower activity (~3 U mg^{-1}) and is inhibited by substrate at pH values above 7.0 (Table S10 and Figure S2). To improve the performance of the cascade, it is therefore necessary to identify and engineer an ALDH with improved specific activity, a preference for D-glyceraldehyde and a higher affinity for NAD$^+$. GADHs represent a suitable starting point for this endeavor, although only a limited number of GADHs have been studied to date, all from thermophilic archaeal organisms, and all showing a strict preference for NADP$^+$ as a cofactor.[20]

ALDH Superfamily Genome Mining

In an initial search, *Ta*ALDH was used as a probe to identify homologs within this family. Five unique sequences annotated as GADHs and an additional ALDH sequence from *Sulfodiicoccus acidiphilus* (*Sa*ALDH) were retrieved from the NCBI and UniProt databases (Table S2 *GADH Screening*). Their genes were cloned and expressed in *Escherichia coli* (*E. coli*; the GADHs from *Acidianus hospitalis* (*Aho*ALDH_Nhis) and *Picrophilus torridus* (*Pto*ALDH_Chis) were poorly soluble), and the enzyme activity towards 5 mM D-glyceraldehyde was measured in the presence of 5 mM NAD$^+$ at 50°C. The best performing candidate was *Aho*ALDH_Nhis, with a specific activity of 0.1 U mg^{-1}, only 9% as active as *Ta*ALDH M42 (1.3 U mg^{-1}). Subsequently, the search for GADH homologs was extended to less closely related ALDH families, including those with a more promiscuous use of both substrates and cofactors, but still of thermophilic origin. Five different ALDH species from thermophilic archaea and bacteria, such as *Methanocaldococcus jannaschii* (*Mj*ALDH) and *Geobacillus stearothermophilus* (*Bst*ALDH), were investigated (Table S2 *Screening of ALDH from thermophilic organisms*). The heat-purified *Mj*ALDH exhibited the highest specific activity (0.9 U mg^{-1}) among these candidates, 72% of the activity of *Ta*ALDH M42. *Mj*ALDH has been previously reported, and its suitability for applications in *in vitro* enzyme cascades for the production of lactate and amino acids has been demonstrated.[13a,21] The natural preference of *Mj*ALDH for NAD$^+$ over NADP$^+$ is an advantage, with a K_m value for NAD$^+$ that is six-fold lower than that of *Ta*ALDH M42 (0.2 mM for *Mj*ALDH and 1.2 mM for *Ta*ALDH M42, respectively; Table S10).[13c] However, despite a high melting temperature (T_m ~90°C), *Mj*ALDH decays rapidly when incubated at 50°C with an almost complete loss of activity within five hours (Figure S4a), whereas *Ta*ALDH M42 retains ~25% of the initial activity after 36 hours of incubation (Figure S3). Moreover, ~50% of the initial activity is lost upon storage at 4°C in 100 mM HEPES buffer, pH 8.0 (Figure S4b), and *Mj*ALDH displays reduced tolerance to organic solvents such as isobutanol compared to *Ta*ALDH M42 (IC$_{50}$ values of ~4.5% for *Mj*ALDH and $\geq$7.0% for *Ta*ALDH M42, respectively; Figure S4c). Two non-phosphorylating glyceraldehyde-3-phosphate dehydrogenases from *Thermococcus kodakarensis* (*Tk*GAPN)[22] and *Clostridium cellulovorans* (*Cc*GAPN) and the phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Oryctolagus cuniculus* (*Oc*GAPDH, purchased from Sigma-Aldrich; belongs to the GAPDH superfamily, illustrated as an outgroup in Figure 2a), were also evaluated using the same assays (Table S2 *GAPN/GAPDH*). Although these enzymes have a strong preference for the substrate D-glyceraldehyde-3-phosphate, they all show detectable activity towards D-glyceraldehyde (0.002–0.06 U mg^{-1}).

Subsequently, we expanded our genome mining search to include mesophilic bacteria (Figure 2). Specifically, we searched for ALDHs in the genomes of three mesophiles, *Herbaspirillum seropedicae Z67* (*Hs*), *Variovorax paradoxus EPS* (*Vp*), and *Paralcaligenes ureilyticus* (*Pu*). Several sequences were retrieved from the NCBI database that share moderate identity (below 40%) with *Ta*ALDH but have high query coverage ($\leq$95%).

ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

**Chemistry
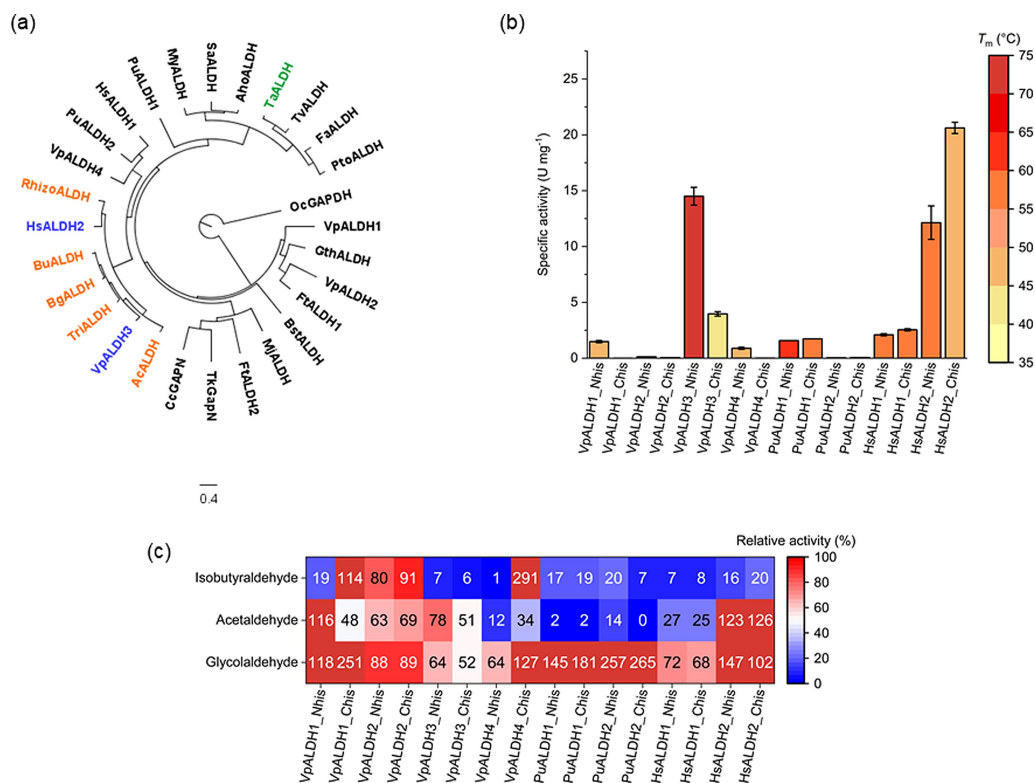Europe**
European Chemical
Societies Publishing

Figure 2. Stepwise genome mining. (a) Reconstruction of a phylogenetic tree using representative ALDH sequences from the genome mining. The starting point, *Ta*ALDH (green), *Hs*ALDH2 and *Vp*ALDH3 from the first genome mining (blue), and ALDH sequences from the second genome mining (orange) are highlighted. (b) Screening of homologous ALDHs from the mesophilic bacteria *V. paradoxus EPS* (*Vp*), *P. ureilyticus* (*Pu*) and *H. seropedicae Z67* (*Hs*). Specific activities (bar heights) were measured at 50 °C using 5 mM D-glyceraldehyde, 5 mM NAD$^+$, 100 mM HEPES, pH 7.5 and melting temperatures (T_m; red color intensity according to the scale) were determined by the Thermofluor assay in 100 mM HEPES, pH 7.5. Both measurements were performed in triplicate (n = 3) and the error bars represent the SD (standard deviation). (c) Analysis of substrate scope. The relative activities of ALDHs from the mesophilic bacteria towards 5 mM isobutyraldehyde, acetaldehyde and glycolaldehyde were determined (compared to the specific activity of the enzymes towards 5 mM D-glyceraldehyde, *i.e.*, 100 % activity; Equation 1) and depicted according to the blue-red color scale (right). Activity measurements were performed at 50 °C in 5 mM NAD$^+$, 100 mM HEPES, pH 7.5 in triplicate (n = 3).

Eight sequences with the highest similarity to *Ta*ALDH and no redundancy with each other (sequences sharing ≥ 95 % identity were omitted) are shown in Table S3 (four from *V. paradoxus* and two each from *P. ureilyticus* and *H. seropedicae*) and were selected for an activity screen. Sequence identity among these eight candidates ranged from 32–62 % (Table S4), and half of these sequences were annotated as succinate-semialdehyde dehydrogenases (SSDHs; Table S3). The sequences were cloned with either an N- or C-terminal His-tag (*i.e.*, Nhis or Chis in Figure 2b) and heterologously expressed in *E. coli* (16 recombinant proteins in total). Specific activity measurements with 5 mM D-glyceraldehyde at 50 °C revealed that eight of these recombinant enzymes had significantly improved activity compared to the *Ta*ALDH M42 variant, with *Hs*ALDH2_Chis being the most efficient catalyst (20.6 U mg^{-1}, ~16-fold higher than *Ta*ALDH M42; Figure 2b and Table S5). The T_m values of these 16 candidates range from 35.2 °C to 73.7 °C; notably, *Vp*ALDH3_Nhis not only has a higher T_m (73.7 °C) than *Ta*ALDH M42, but also shows significantly higher specific activity (14.5 U mg^{-1}). The activity and stability screens also revealed that the location of the His-tag affects both properties. For example, moving the His-tag of *Vp*ALDH3 from the N-terminus to the C-terminus results in a decrease in T_m from 73.7 °C to

43.4 °C and an almost four-fold decrease in specific activity. The unfavorable effect of the C-terminal His-tag may be due to its interference with the "arm-like" oligomerization domain of this enzyme.[16,23] In contrast, while shifting the His-tag from the N- to the C-terminus of *Hs*ALDH2 also resulted in a decrease in T_m (by 7.7 °C), the specific activity increased by 170 % from 12.1 U mg^{-1} to 20.6 U mg^{-1}.

Exploiting the High Similarity between *Vp*ALDH and *Hs*ALDH

Our next aim was to combine the beneficial properties of both ALDH candidates (*i.e.*, the high catalytic activity and good expression of *Hs*ALDH2_Chis with the high stability of *Vp*ALDH3_Nhis).

Residues within three signature motifs play a critical role for members of the ALDH superfamily (Table S5 and Figure S5). *Hs*ALDH2 and *Vp*ALDH3, the most active ALDHs from the initial genome mining, have very similar sequences in each of the three motifs (although the sequence identity is only 62 %). To combine the beneficial properties, we focused on the amino acid residue at position 9 of the Cys site, immediately adjacent to the catalytic cysteine, *i.e.*, Val284 in *Hs*ALDH2 and Ile287 in

ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

**Chemistry
Europe**

European Chemical
Societies Publishing

*Vp*ALDH3. These residues are the only difference between these two enzymes within 5 Å of the catalytic cysteine. Structural models of the active sites of the two enzymes (using Phyre2) also suggest that the catalytic Cys residues may be oriented differently (Figure S6). In *Vp*ALDH3, the thiol group of the catalytic Cys (corresponding to the nucleophilic Cys283 of *Hs*ALDH) points towards the entrance of the substrate binding pocket and is well positioned to interact with the glutamate residue (corresponding to Glu249 of *Hs*ALDH) from the Glu site. In contrast, in *Hs*ALDH2, the thiol group of the catalytic Cys283 points away from the active site cavity. To test whether the variations in position 9 and the orientations of the catalytic Cys contribute to the observed differences in properties between *Hs*ALDH2 and *Vp*ALDH3, Val284 was replaced by Ile in *Hs*ALDH2, and Ile287 by Val in *Vp*ALDH3. The variant *Hs*ALDH2_V284I_Chis showed an increase in thermostability with a gain of 4 °C in T_m (Table S6). At the same time, the variant retained the high catalytic activity of the wild-type enzyme towards 5 mM D-glyceraldehyde at 37 °C ($\sim$25 U mg^{-1}). The half-life ($t_{1/2}$) of the enzyme at 45 °C also increased from $\sim$20 minutes to almost two hours, and accordingly the total turnover number (TTN$_{3\text{ h, 5 mM}}$) at 45 °C increased almost threefold to $\sim$130,000 (Table S7 and Figure S7). When the variant was incubated in the presence of reactive aldehyde (10 mM D-glyceraldehyde), its stability decreased drastically (with an almost four-fold shorter $t_{1/2}$), but it still reached an almost two-fold higher TTN$_{\text{h, 5 mM}}$ compared to the wild-type enzyme. In contrast, the Ile287Val mutation in *Vp*ALDH3_Nhis did not result in improved catalytic properties (Table S6).

A Focused Second Round of Genome Mining

In another attempt to combine the advantageous characteristics of *Hs*ALDH2_Chis and *Vp*ALDH3_Nhis, a second sequence-based genome mining approach was employed. Using

*Hs*ALDH2 and *Vp*ALDH3 as the best performing candidates from the first round of genome mining individually, approximately 5,000 homologous enzymes for each query sequence were retrieved from an extensive Protein BLAST® search (https://blast.ncbi.nlm.nih.gov/Blast.cgi).[24] A total of 251 sequences co-occurred in both sequence collections and shared ≥65% identity with both *Hs*ALDH2 and *Vp*ALDH3 (Figure S8). From these, five candidates from *Burkholderia glumae* (*Bg*ALDH), *B. ubonensis* (*Bu*ALDH), *Acidovorax citrulli* (*Ac*ALDH), *Trinickia sp. 7GSK02* (*Tri*ALDH) and *Rhizobiaceae bacterium* (*Rhizo*ALDH) were selected for further characterization (sequence identity of 62–88%; Table S8), showing different distances to both parental sequences and covering a wide sequence variability within the collection of the top candidates (Figure S8). *Rhizo*ALDH is the only candidate that is more similar to *Hs*ALDH2 (75% and 65% identity with *Hs*ALDH2 and *Vp*ALDH3, respectively; however, some terminal residues are apparently missing), while the others share an identity of 62–68% with *Hs*ALDH2 and 75–80% with *Vp*ALDH3 (Table S8). Also, the signature sequence motifs of each candidate except *Rhizo*ALDH are identical to those of *Vp*ALDH3, whereas *Rhizo*ALDH has a Val at position 9 of the Cys site similar to *Hs*ALDH2 (see previous section; Table S9). Several candidates show both high activity and optimized stability at 50 °C (Figure 3a), and except for *Rhizo*ALDH, all of the novel candidates have high specific activities, up to 115% of that of *Hs*ALDH2_Chis (*Bu*ALDH_Chis). Similarly, the melting points are comparable to that of *Vp*ALDH3_Nhis. Residual enzyme activities were also recorded after three hours of incubation at 50 °C to assess the process stability of these enzymes. Of the best candidates from the first round of genome mining and protein engineering (see above), *Hs*ALDH2_V284I_Chis was virtually inactive, while *Vp*ALDH3_Nhis still retained 73% of its initial activity after three hours (Figure 3b). For the ALDH candidates from the second round of genome mining, residual activities as high as 78% (for *Bg*ALDH_Nhis) were recorded. Even the presence of 1% (v/v) of isobutanol did not drastically alter their
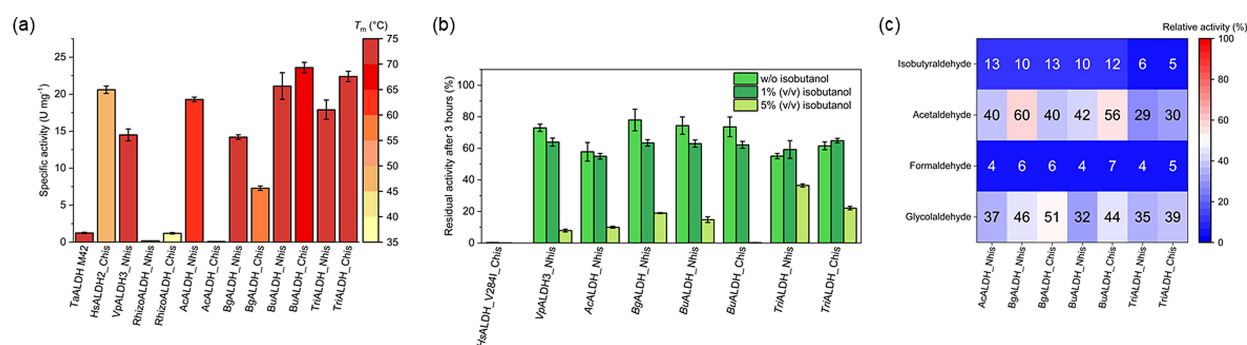

Figure 3. Second round of genome mining. (a) Screening of "second generation" ALDHs. Specific activities (bar heights) were measured at 50 °C using 5 mM D-glyceraldehyde, 5 mM NAD$^+$, 100 mM HEPES, pH 7.5 and melting temperatures (T_m; red color intensity according to the scale) were determined by the Thermofluor assay in 100 mM HEPES, pH 7.5. Both measurements were performed in triplicate ($n=3$) and the error bars represent the SDs. (b) Analysis of enzyme stability in different concentrations of isobutanol. Residual activities (compared to the initial activities before incubation) of the selected ALDHs from the first and second round of genome mining were measured after incubation in isobutanol concentrations of 0 (light green), 1 (dark green), and 5% (v/v) (lime green) for three hours at 50 °C using 5 mM D-glyceraldehyde, 5 mM NAD$^+$, 100 mM HEPES, pH 7.5. Measurements were performed in triplicate ($n=3$) and the error bars represent the SDs. (c) Analysis of substrate scope. The relative activities of the ALDHs from the second round of genome mining towards 5 mM of different C1 to C4 aldehydes were determined (compared to the specific activity of the enzymes towards 5 mM D-glyceraldehyde, *i.e.*, 100% activity; Equation 1) and depicted according to the blue-red color scale (right). Activity measurements were performed at 50 °C in 5 mM NAD$^+$, 100 mM HEPES, pH 7.5 in triplicate ($n=3$).

ChemSusChem Research Article
doi.org/10.1002/cssc.202301132

Chemistry
Europe
European Chemical
Societies Publishing

activities. In the case of *Tri*ALDH_Nhis and *Tri*ALDH_Chis, the presence of 1% solvent appeared to slow down the structural disintegration of the enzymes. However, increasing the isobutanol concentration to 5% (v/v) was detrimental for most of the enzymes, except for *Tri*ALDH_Nhis, which still retained nearly 40% of its initial activity.

Kinetic assays with D-glyceraldehyde concentrations as high as 100 mM showed that there is no significant substrate inhibition for these "second generation" ALDHs (Table S10 and Figure S9). It should be noted that while *Hs*ALDH2_Chis is still the most active ALDH (77 U mg^{-1} at 37°C), all ALDHs except *Rhizo*ALDH from the second round of genome mining are at least 12-fold more active than *Ta*ALDH M42 and more stable than *Hs*ALDH2_Chis (Table S10).

*Hs*ALDH2_Chis from the first round of genome mining was proven to be promiscuous in its substrate selection, preferring small aldehydes, such as acetaldehyde (126% activity; Equation 1 and Figure 2c), whose conversion is an undesired reaction in the ethanol cascade. *Vp*ALDH3_Nhis, on the other hand, prefers D-glyceraldehyde over small C$_1$ and C$_2$ aldehydes (e.g., 78% activity towards acetaldehyde; Equation 1 and Figure 2c). ALDHs selected from the second round of genome mining have further shifted their preference towards D-glyceraldehyde. In particular, *Tri*ALDH_Nhis has reduced the relative activity towards acetaldehyde to 29% (Equation 1) and also shows only a low activity towards isobutyraldehyde (6%; undesired reaction in the isobutanol cascade; Equation 1 and Figure 3c).

Optimization of Substrate Selectivity

*Hs*ALDH2_V284I_Chis from the first round of genome mining and *Tri*ALDH_Nhis from the second round were selected to further optimize substrate selectivity towards D-glyceraldehyde. Using molecular docking experiments with Gibbs free energy (ΔG) calculations, ancestral sequence reconstruction (ASR), and structure analysis, seven amino acid residues were selected for mutagenesis studies (Table S11). Among them, the residue corresponding to position 446 in *Hs*ALDH2_V284I_Chis, is a conserved Phe among all the ALDH candidates (including SSDH) investigated in this study. In *Hs*ALDH2_V284I_Chis, this residue may be involved in cation-π interactions with Arg159 (moderately conserved; 54% of the enzymes analyzed), a residue that may interact directly with the substrate (Figure S10). However, in GADHs, the residue corresponding to Phe446 is mostly a histidine, and the phylogenetic analysis (using GRASP[25] and EMBOSS Cons) also indicated the possibility of integrating the F446H mutation (Figure S11). Considering the high selectivity of GADH for D-glyceraldehyde, we first focused on Phe446 in *Hs*ALDH2_V284I_Chis to prove its role in substrate binding.

Site saturation mutagenesis was performed using a degenerate NNK codon (Figure S12). Screening of the Phe446 NNK library was performed in three 96-well microtiter plates (with a total sampling of 252 clones, 18 positive and 18 negative controls; Figure S13).[26] Activities were determined towards

5 mM D-glyceraldehyde and 5 mM acetaldehyde to identify variants with a higher preference for D-glyceraldehyde. In the most promising double mutant, Phe446 was replaced by a histidine (*Hs*ALDH2_V284I_F446H_Chis); relevant catalytic parameters are summarized in Table S12. Compared to both wild-type and single mutant *Hs*ALDH2, the double mutant showed an improved binding for D-glyceraldehyde (K_m~3 mM), while at the same time exhibiting a drastically reduced affinity for acetaldehyde (K_m ~180 mM). Therefore, the catalytic efficiency (k_{cat}/K_m) for the conversion of D-glyceraldehyde decreased by only ~10% (from 5.2 to 4.7 s^{-1} mM^{-1}) compared to the wild-type enzyme, despite a significant decrease in the turnover number k_{cat} (from 68 to 15 s^{-1}). In contrast, for acetaldehyde, the double mutation resulted in a massive decrease in catalytic efficiency (from 4.0 to 0.2 s^{-1} mM^{-1}). Thus, the double mutation had the desired effect of shifting the substrate preference of *Hs*ALDH2 strongly towards D-glyceraldehyde, and the selectivity for D-glyceraldehyde over acetaldehyde improved more than 21-fold (Equation 2, Table S12, and Figure 4c).

Wild-type *Tri*ALDH_Nhis is inherently more stable than *Hs*ALDH2 (T_m of 79.2°C for *Tri*ALDH_Nhis and 49.5°C for *Hs*ALDH2_Chis; Table S12) and more selective for D-glyceraldehyde over acetaldehyde (Figures 4a and 4b). We, therefore, replaced the phenylalanine (Phe450) corresponding to Phe446 in *Hs*ALDH2 with a histidine as described for *Hs*ALDH2_V284I_F446H_Chis (note that the I284 introduced in *Hs*ALDH2 to increase its stability is naturally present in *Tri*ALDH and other SSDHs analyzed). In the resulting single mutant, *Tri*ALDH_F450H_Nhis, there was no significant impact on substrate selectivity (a small reduction from 3.3 of the wild-type *Tri*ALDH
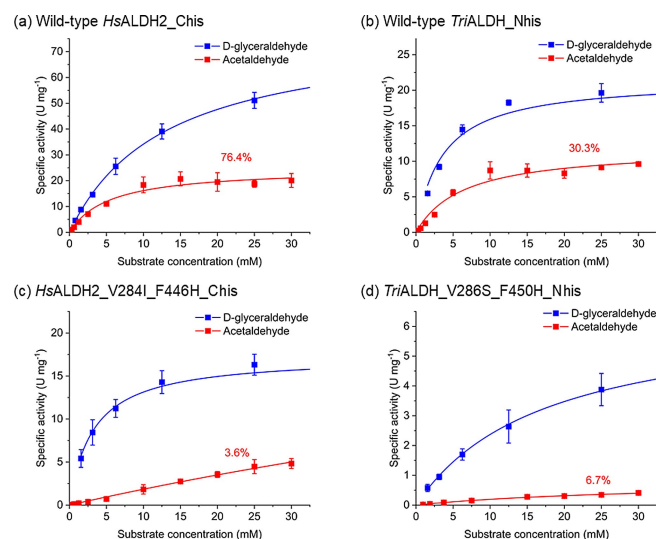

Figure 4. Selectivity of the engineered ALDH variants for D-glyceraldehyde over acetaldehyde. Catalytic rates of wild-type *Hs*ALDH2_Chis (a), wild-type *Tri*ALDH_Nhis (b), *Hs*ALDH2_V284I_F446H_Chis (c), and *Tri*ALDH_V286S_F450H_Nhis (d) were analyzed as a function of D-glyceraldehyde (blue) and acetaldehyde (red) concentrations. Measurements were performed photometrically in 100 mM HEPES, pH 7.5, 5 mM NAD$^+$ at 37°C with different substrate concentrations in triplicate (n = 3) and the error bars represent the SDs. The unspecific conversion of acetaldehyde is indicated as the percentage catalytic efficiency of D-glyceraldehyde conversion (Equation 3).

ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

**Chemistry
Europe**

European Chemical
Societies Publishing

to 3.2; Equation 2 and Table S12). However, the K_m value for acetaldehyde conversion increased by ~2.5-fold (6.4 mM for the wild-type enzyme and 15.7 mM for the variant). In addition, this variant further decreased the unspecific conversion of another small aldehyde, formaldehyde, by ~two-fold compared to HsALDH2_V284I_F446H_Chis (Figure S14). The mutation also had minimal effect on the T_m (78 °C) and the high tolerance of the wild-type enzyme to ethanol was not affected (Figures 5b and 5d).

In a further attempt to increase the selectivity of TriALDH for D-glyceraldehyde, two additional mutants were analyzed, both in combination with the F450H mutation. The first mutation, A212T near the cofactor binding site of TriALDH, was recently found to be beneficial for increasing activity in the screening campaign of a randomized library of HsALDH (Teshima et al., unpublished data). Here, the resulting double mutant, TriALDH_A212T_ F450H_Nhis, showed an increase in V_{max} (from 11.3 to 14.3 U mg^{-1}) and a slightly improved binding of D-glyceraldehyde (K_m reduced from 5.5 to 4.8 mM; Table S12). Furthermore, the catalytic efficiency (k_{cat}/K_m) for acealdehyde decreased to 50 % (from 0.6 to 0.3 s^{-1} mM^{-1}). As a result, the substrate selectivity shifted towards D-glyceraldehyde (Equation 2 and Figure S15a). The second mutation, V286S, involves a residue upstream of the catalytic cysteine. Multiple sequence alignments indicated that a valine or threonine at this position is more common in the SSDH family, whereas a hydrophilic serine is highly conserved in the D-glyceraldehyde-specific GADH family members (Figure S16). We speculated that a serine at this position might form a hydrogen bond with D-glyceraldehyde, thus favoring its binding. The catalytic efficiency of the TriALDH_V286S_F450H_Nhis double mutant is significantly reduced for both substrates, but the selectivity for

D-glyceraldehyde increased from 3.2 of the single mutant to 15.6 (Equation 2 and Table S12). This nearly five-fold improvement in substrate selectivity is largely driven by the suppression of the acetaldehyde activity to a minimal level with a k_{cat}/K_m of 0.02 s$^-$ mM^{-1} (Figure 4d). Note that the combination of the A212T and V286S mutations to form the triple mutant TriALDH_A212T_V286S_F450H_Nhis achieved the D-glyceraldehyde selectivity of only 9.9 due to the increased catalytic efficiency towards acetaldehyde compared to TriALDH_V286S_F450H_Nhis (Equation 2, Table S12, and Figure S15b).

Stability in Ethanol

The tolerance of the wild-type HsALDH2_Chis and TriALDH_Nhis, and their variants HsALDH2_V284I_F446H_Chis and TriALDH_F450H_Nhis to ethanol, the ultimate product of the multi-enzyme cascade described below,[27] was assessed by incubating the enzymes for up to eight hours at 37 °C in the presence of 0, 1, 5 and 10 % (v/v) ethanol (Figure 5). Residual activities were measured with 5 mM D-glyceraldehyde and 5 mM NAD$^+$. For HsALDH, the wild-type enzyme and the double mutant retained ~80 % and ~45 % of their respective activities after eight hours of incubation in the absence of ethanol. The presence of 1 % (v/v; ~0.2 M) ethanol had a minimal effect on these two enzymes. However, the presence of 5 % (v/v; ~0.9 M) ethanol significantly accelerated the decrease in their specific activities. Wild-type HsALDH2_Chis has a $t_{1/2}$ of ~eight hours, while the double mutant has a $t_{1/2}$ of only ~three hours. At 10 % (v/v; ~1.7 M) ethanol, both variants of HsALDH2 were rapidly inactivated within one hour. In contrast, TriALDH (both wild-type and its variant) is more tolerant to ethanol, even at concentrations as high as 10 % (v/v). In fact, enzyme inactivation over the eight-hour period was largely independent of the presence of ethanol and the TriALDH_F450H_Nhis variant also retained ~60 % activity until the end of the incubation.

The introduction of the second mutation significantly reduced the stability of TriALDH_A212T_F450H_Nhis, which showed only ~10 % activity after eight hours of incubation at 37 °C in the absence of ethanol ($t_{1/2}$ = ~three hours; Figure S17a). In contrast, TriALDH_V286S_F450H_Nhis retained ~60 % activity up to eight hours ($t_{1/2}$ = ~nine hours), similar to the single mutant. However, the addition of 10 % (v/v) ethanol affected the stability of the TriALDH_V286S_F450H_Nhis variant ($t_{1/2}$ = ~two hours) (Figure S17b), with activity rapidly decreasing below 60 % immediately after ethanol addition (within the first one hour). However, the further decline in activity during incubation was significantly slower compared to the TriALDH_A212T_F450H_Nhis variant. Along with these advantages in functional stability and D-glyceraldehyde selectivity of the TriALDH_V286S_F450H_Nhis variant, the test cascades (see 1.1 in the Supplementary Information) also supported the selection of this variant together with HsALDH_V284I_F446H_Chis for use in the final upscaling of the ethanol cascade.
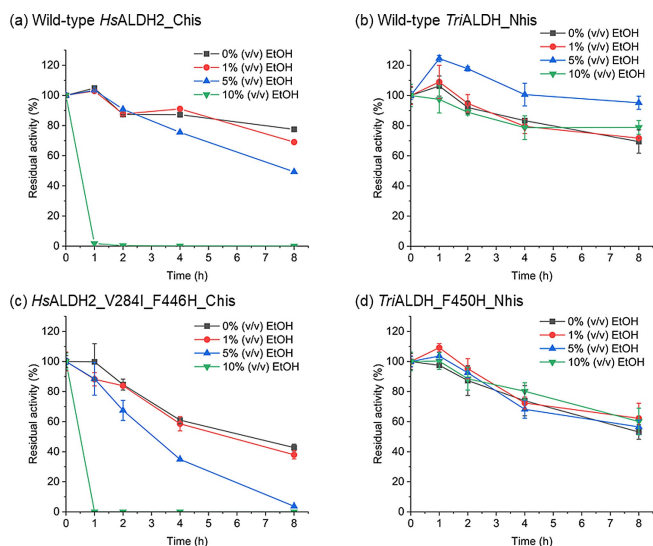

Figure 5. Ethanol tolerance of wild-type HsALDH2_Chis (a), wild-type TriALDH_Nhis (b), HsALDH2_V284I_F446H_Chis (c), and TriALDH_F450H_Nhis (d). Enzymes were incubated at 37 °C in 100 mM HEPES, pH 7.5 supplemented with different concentrations of ethanol from 0–10 % (v/v), and residual activities were measured using 5 mM D-glyceraldehyde and 5 mM NAD$^+$ in 100 mM HEPES, pH 7.5 in triplicate (n = 3). The error bars represent the SDs.

Validation of ALDH Variants in the Ethanol Cascade

Among the best ALDH variants available, one has high activity and specificity, while another one has slightly reduced specificity but higher stability (see above). We tested their performance in the minimal cascade for the production of ethanol via pyruvate (Figure 1). First, the kinetic parameters of each of the six enzymes used in the cascade were measured under conditions corresponding to those of the cascade (100 mM HEPES buffer, pH 7.5, containing additives such as thiamine pyrophosphate (TPP) and magnesium chloride (MgCl$_2$); Table S13). Side reactions were also analyzed, including not only the unspecific oxidation of acetaldehyde to acetic acid by ALDH (see the previous section; Figure 4), but also the reduction of D-glyceraldehyde by ADH (Figure S18). The experimental data were used to model the cascade performance using COPASI 4.36 software (Equations S1–S13).[28] An interactive and iterative optimization process between computational cascade modeling and experimental analysis was performed to refine the cascade setup.

It was challenging to balance the three NAD$^+$/NADH regeneration reactions catalyzed by GDH, ALDH and ADH (Figure 1) to optimize cofactor circulation. It should be noted that the ethanol cascade first goes through two oxidation steps (catalyzed by GDH and ALDH for pyruvate generation) to generate two molecules of NADH, and the oxidizing equivalents are regenerated only in the last reaction step (catalyzed by ADH, requiring a two-fold higher TTN). The best outcomes were predicted by using only small amounts of the exceptionally active GDH to prevent the accumulation of D-glyceraldehyde. Further fine-tuning of the enzyme composition in the cascade was performed experimentally, for instance, to adjust the long-term activity of ADH, since its activity is significantly susceptible to the end product, ethanol (Figure S19).

A closed system with continuous shaking was found to be the most suitable reaction system for our purposes (see 1.2 in the Supplementary Information). In addition, the HEPES buffer concentration was increased from 100 to 200 mM to address the previously observed pH drops (from pH 7.5 to ~4.0) and concomitant enzyme inactivation/precipitation.[10b] The optimized system was used to run an ethanol cascade with an increased substrate load of 77 g L^{-1} (~430 mM; Table S14). Ethanol production was monitored over two days at 37 °C with shaking (600 rpm). Four different cascade mixtures containing one of the following ALDH variants were compared (Figure S20): HsALDH_V284I_F446H_Chis (higher selectivity and activity but lower stability) and TriALDH_V286S_F450H_Nhis (higher stability at elevated temperatures and in organic solvents). In two additional cascades, the TaALDH M42 variant replaced the HsALDH and TriALDH variants.[13c]

D-glucose was consumed linearly at a rate of 1.7 g L^{-1} h^{-1} (9.6 mM h^{-1}; Figure S22a) for 39 hours or 1.4 g L^{-1} h^{-1} (7.7 mM h^{-1}; Figure S22b) for 44 hours in the cascades using HsALDH or TriALDH, respectively (Figures 6a and S21a). During the same period, ethanol concentrations increased linearly with a productivity of 1.0 g L^{-1} h^{-1} (21.0 mM h^{-1}; Figure S22c) and 0.8 g L^{-1} h^{-1} (16.3 mM h^{-1}; Figure S22d), respectively, to reach

more than ~95% final conversion. This indicates highly efficient cascade fluxes from D-glucose to the product, ethanol, without a rate-limiting step. In comparison, when HsALDH was replaced by the TaALDH M42 variant (Figure 6b), the ethanol productivity was lower at only ~60% of that of the HsALDH cascade. Surprisingly, after 14 hours, while the ethanol productivity remained almost constant at 13.2 mM h^{-1} (Figure S23c), the D-glucose consumption increased significantly from 5.2 mM h^{-1} (Figure S23a) to ~19.6 mM h^{-1} (Figure S23b) and the total carbon mass decreased. When TriALDH was replaced by the TaALDH M42 variant in the cascade, similar profiles were observed (Figure S21b). This increase in the rate of glucose consumption was accompanied by a sudden increase in the concentration of D-glyceraldehyde and the appearance of substantial amounts of glycerol (up to 8%). Apparently, upon a threshold concentration of glyceraldehyde, ADH is able to reduce this intermediate while regenerating NAD$^+$.

In the cascades with HsALDH and TriALDH, almost full conversion was achieved within 48 hours (up to 98.7% molar ethanol), and the final ethanol titer was recorded as high as 843 mM (4.9% (v/v)). However, only up to ~55% molar ethanol was achieved when the TaALDH M42 variant was used instead. Furthermore, while the carbon mass loss was negligible when the HsALDH and TriALDH variants were used in the cascade, ~22% of the total carbon mass was lost when these enzymes were replaced by the TaALDH M42 variant.

The most prominent intermediate recorded for the HsALDH and TriALDH cascades was D-glycerate, whose concentration increased steadily to up to 5% after 39–44 hours (Figures 6c and S21c). This is consistent with its dehydration by PuDHT being the rate-limiting step of the cascade (Figure 1). However, the concentration of D-glycerate decreased continuously towards the end of the reaction when the conversion of D-glucose (and D-gluconate, another substrate of PuDHT) was almost complete. Furthermore, the concentration of reactive D-glyceraldehyde never exceeded 1% (in the presence of HsALDH, a peak of 0.5% is reached after 27 hours, with TriALDH a maximum concentration of 0.9% was recorded after 39 hours). In contrast, D-glyceraldehyde was the most abundant intermediate in the control cascades using TaALDH M42, reaching a concentration of up to 14% (Figures 6d and S21d). Presumably, the catalytic activity of TaALDH M42 was insufficient due to its high K_m and the limited cofactor availability at the expense of the accelerated D-glucose consumption by GDH (after 14 hours of cascade reaction). Other by-products (acetoin, acetate, and lactate) were generally present at very low concentrations, if any, in each of the cascades and are not expected to play an important role in the efficiency of ethanol production.

Optimizing the Operating Temperature

The observed ethanol production rates were significantly slower than expected. Computational modeling (Equations S1–S13) predicted that the cascades using HsALDH or TriALDH would be completed within three or four hours, respectively. One reason for the difference between experimental and simulated produc-
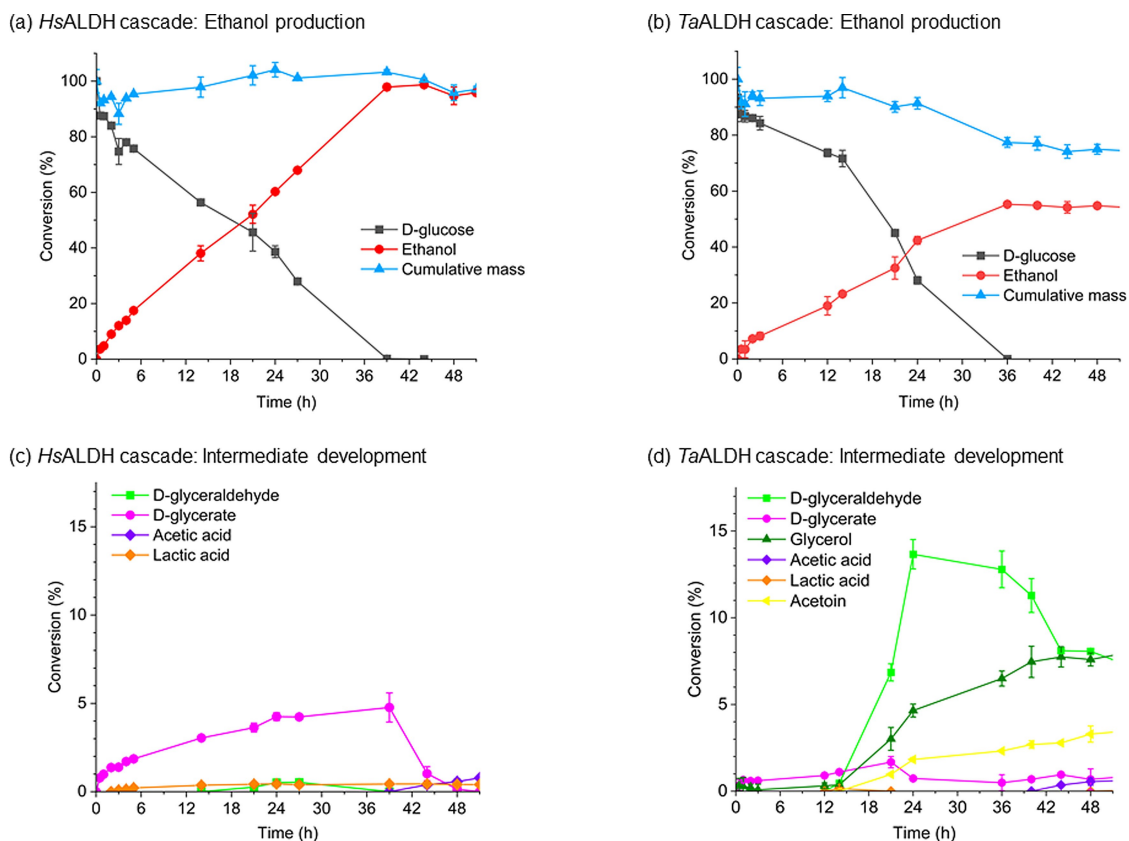
ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

**Chemistry
Europe**
European Chemical
Societies Publishing

Figure 6. Cell-free biotransformation of 77 g L^{-1} D-glucose to ethanol with HsALDH2_V284I_F446H_Chis (a and c) and the previously best available ALDH TaALDH M42 (b and d). The consumption of the substrate D-glucose (gray), the production of the final product ethanol (red) and the cumulative mass (blue) (a and b) and the development of the intermediates (c and d) were monitored at 37 °C for 48 hours. Reactions were performed in triplicate (n = 3) and the error bars indicate the SDs.

tion rates is the complexity of simulating cofactor fluxes and regeneration in the modeling. Furthermore, the reaction temperature was kept at 37 °C to ensure stable enzyme activities throughout the run, and initial experiments indicated that the selectivity of ALDH variants for D-glyceraldehyde over acetaldehyde at 5 mM concentrations is higher at moderate temperatures (Table S15). Therefore, to evaluate the effect of temperature on ethanol productivity, the cascades were run at elevated temperatures of 45 and 50 °C, and ethanol yields were quantified after 24 hours instead of 48 hours as in the previous experiments. A cascade without ALDH was also run as a negative control.

The cascade with TriALDH_V286S_F450H_Nhis produced the highest ethanol yield (>95 %) at 45 °C in one day (Figure 7a). Increasing the process temperature by 8 °C accelerated the ethanol production almost two-fold to a space-time yield (SPY) of ~39 g L^{-1} d^{-1} (in good agreement with the Arrhenius equation, which predicts that an increase of 10 °C will double the reaction rate).[29] At the same time, the molar ethanol yield was not significantly affected (97.1 % from 77 g L^{-1} D-glucose at 37 °C and 95.5 % from 80 g L^{-1} at 45 °C). However, by further increasing the temperature to 50 °C, the ethanol titer decreased by ~10 %, suggesting that some of the enzymes may be gradually inactivated or the mass fluxes were imbalanced, probably due to different temperature optima for different enzymes. In contrast, while the cascade with HsALDH achieved a ~60 % ethanol yield after 24 hours at 37 °C, it dramatically decreased by 88 % at 45 °C (raising the temperature to 50 °C had no further impact), consistent with the reduced thermal stability of this ALDH variant. For the cascade run with TaALDH (with a temperature optimum of ~63 °C),[30] increasing the temperature significantly improved the ethanol yield after 24 hours from 29 % at 37 °C to 61 % at 50 °C. The negative control cascade lacking ALDH achieved ~30 % ethanol conversion at both 45 °C and 50 °C, comparable to the yield of the cascade using TaALDH at 37 °C. In addition, D-glyceraldehyde accumulated up to ~17 % in the negative control cascade and ~4 % was detected in the TaALDH cascade (Figure 7b). When the HsALDH or TriALDH variant was used instead, D-glyceraldehyde was not detectable.

Conclusions

In the present work, we have demonstrated an iterative and interactive approach between sequence-based genome mining, enzyme engineering and cascade optimization to, firstly, develop an NAD$^+$-dependent D-glyceraldehyde dehydrogenase with remarkably high selectivity and, secondly, enable the efficient conversion of D-glucose to pyruvate with only four
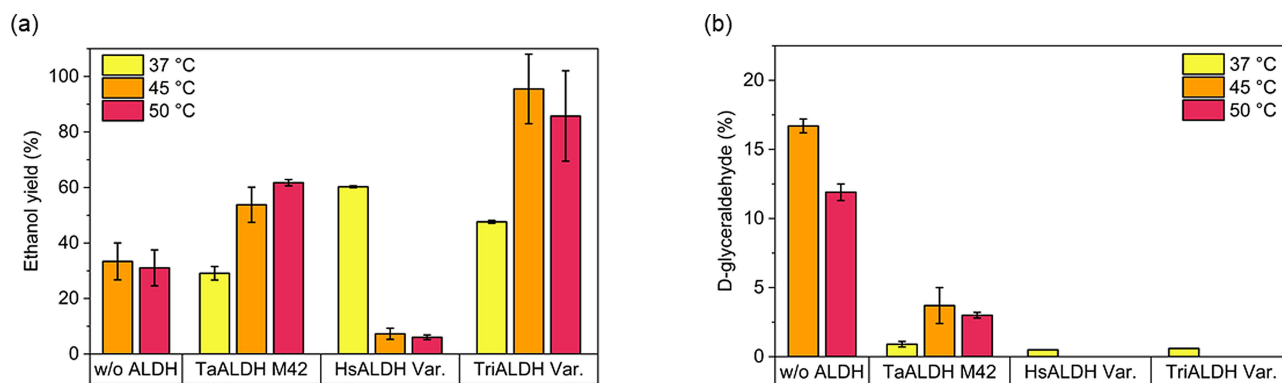
ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

Chemistry
Europe
European Chemical
Societies Publishing

Figure 7. Effect of reaction temperature on the ethanol production. (a) Ethanol yield from 80 g L^{-1} D-glucose was determined after a cascade reaction of 24 hours at 45 or 50 °C using the *Ta*ALDH M42 variant, *Hs*ALDH2_V284I_F446H_Chis, *Tri*ALDH_V286S_F450H_Nhis or without using ALDH, and compared with the ethanol yield obtained at 37 °C after 24 hours (see the previous section). (b) D-glyceraldehyde concentration was determined after a cascade reaction of 24 hours at 45 or 50 °C using different ALDH variants and compared with the concentration obtained at 37 °C after 24 hours. All cascade reactions were performed in triplicate (n = 3) and the error bars represent the SDs.

enzymes and NAD$^+$. Successive genome mining identified *Hs*ALDH2, which had 24-fold higher activity than the *Ta*ALDH M42 variant in the first superfamily-wide search, and *Tri*ALDH with higher thermostability and ethanol tolerance in the second focused search. ASR-assisted semi-rational engineering of the enzymes at their active sites yielded several ALDH variants with up to 21-fold improved D-glyceraldehyde selectivity over acetaldehyde, thus minimizing its unspecific conversion to a minimum for an application in a one-pot system. In the final multi-enzyme cascade, both tested enzyme variants, *Hs*ALDH2_V284I_F446H_Chis and *Tri*ALDH_V286S_F450H_Nhis, showed continuous linear ethanol production up to 4.9% (v/v) with 98.7% molar ethanol yield. The optimized cofactor and mass fluxes improved the final ethanol titer by more than 37-fold compared to our previous work.[15a] The high ethanol yield is a clear advantage over the fermentative ethanol production process using microorganisms, which generally suffers from low substrate utilization for product formation.[31] Furthermore, a high product titer and minimal production of by-products and intermediates are desirable to reduce the cost and effort of the future downstream process.

A comparison with the control cascade using the original ALDH variant suggested that three cofactor-dependent enzymes, GDH, ALDH, and ADH, must be balanced to avoid by-product formation triggered by undesired cofactor fluxes. Optimizing the stability of the *Tri*ALDH variant enabled a cascade process at elevated temperatures to further accelerate ethanol production. Together with the *Hs*ALDH variant with higher activity at a lower temperature, this variant provides different options for a flexible process design, depending on ecological, technical, and production cost requirements (*e.g.*, cooling/heating and time). This work demonstrated the great potential of enzyme engineering to produce designer enzymes for specific cascade processes. Although the challenge of further increasing the activity and robustness of the enzymes remains for further upscaling, enhancing the efficiency of the minimal cascade for ethanol production provided an excellent

example of the continuous product formation desired for an industrial process.

Experimental Section

Reagents

All chemicals used in this work were purchased from Carl Roth (Karlsruhe, Germany), Thermo Fisher Scientific (Waltham, MA, USA), Merck (Darmstadt, Germany), Millipore (Darmstadt, Germany), New England Biolabs (NEB) (Frankfurt am Main, Germany), Promega (Walldorf, Germany), Qiagen (Hilden, Germany), Serva (Heidelberg, Germany), Sigma-Aldrich (Darmstadt, Germany), and VWR (Darmstadt, Germany). Oligonucleotides were synthesized by Eurofins Genomics (Ebersberg, Germany) and codon-optimized full-length genes were synthesized by Geneart (Regensburg, Germany).

Strains and Plasmids

Escherichia coli (*E. coli*) XL1-Blue or NEB® Turbo from NEB were used as cloning strains and *E. coli* BL21(DE3) from Novagen (Nottingham, UK) was used for heterologous gene expression. The bacterial strains *Herbaspirillum seropedicae Z67* (DSM 6445) and *Paralcaligenes ureilyticus* (DSM 24591) were purchased from DSMZ German Collection of Microorganisms and Cell Cultures GmbH (Braunschweig, Germany) for the extraction of genomic DNA (gDNA). The bacterial strain *Variovorax paradoxus EPS* was kindly provided by Prof. Dirk Tischler (Ruhr University Bochum). pET24a and pET28a were obtained from Novagen.

Plasmid Construction

gDNA was isolated from *H. seropedicae Z67* (DSMZ), *P. ureilyticus* (DSMZ), and *V. paradoxus EPS* (Ruhr University Bochum) using the DNeasy UltraClean Microbial Kit (Qiagen). The genes of interest were amplified using oligonucleotides (Eurofins genomics) and Phusion® High-Fidelity DNA Polymerase (NEB) by polymerase chain reaction (PCR) (C1000 Touch™ Thermal Cycler, Bio-Rad, Feldkirchen, Germany). The PCR products were inserted into the pET28a or pET24a vector (Novagen) *via* NdeI/XhoI and HindIII/XhoI restriction sites (NEB) using T4 DNA Ligase (NEB). After transformation, the

ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

**Chemistry
Europe**

European Chemical
Societies Publishing

plasmid constructs were propagated in an *E. coli* cloning strain and extracted using the QIAprep Spin Miniprep Kit (Qiagen).

Site-Directed Mutagenesis

QuikChange™ site-directed mutagenesis (Agilent, Waldbronn, Germany) was performed with a partially overlapping primer pair inserting a mutation in the overlapping region.[32] After the whole plasmid amplification using Phusion® High-Fidelity DNA Polymerase (NEB), the parental template plasmid was digested with DpnI (NEB) at 37 °C overnight, followed by its inactivation at 80 °C for 20 minutes.

Modified site-directed mutagenesis and site saturation mutagenesis were performed using the Gibson assembly method.[33] Two fragments with 20–25 bp overlapping terminal sequences were generated in separate PCR reactions and assembled by Gibson assembly. The mutation was introduced at one of the assembly sites by a primer sequence. In the NNK primer, the mutation site was placed outside of the overlapping region. The other assembly site was located at the kanamycin resistance gene. Equimolar amounts of the purified PCR products were mixed in 3.3 % (w/v) PEG-8000, 66.7 mM Tri/Cl, pH 7.5, 6.7 mM MgCl$_2$, 6.7 mM DTT, 0.1 mM dATP, 0.1 mM dCTP, 0.1 mM dGTP, 0.1 mM dTTP, 0.7 mM NAD$^+$, 0.03 U T5 exonuclease (NEB), 0.17 U Phusion® High-Fidelity DNA Polymerase (NEB), and 26.67 U *Taq* DNA Ligase (NEB) in a total volume of 10 µL, incubated in a C1000 Touch™ Thermal Cycler (Bio-Rad) at 50 °C for 1 hour, and used for transformation. Screening of the NNK library was performed on three 96-well deep well plates.[26] The photometric assay was performed on a 384-well plate.

Protein Expression

20 mL of LB broth preculture containing 100 µg mL^{-1} kanamycin was inoculated with a single colony of freshly transformed *E. coli* BL21(DE3) cells[34] and cultivated for 16 hours at 37 °C and 160 rpm (Heraeus® and MaxQ™ 2000, Thermo Scientific, Schwerte, Germany). The main culture of 1 L ZYP-5052 autoinduction medium containing 100 µg mL^{-1} kanamycin was inoculated to an OD$_{600}$ of 0.07 and shaken at 37 °C and 95 rpm (Multitron, Infors, Sulzemoos, Germany) until an OD$_{600}$ of 0.5–0.6 was reached. The temperature was decreased to 20 °C and the protein expression was performed for 20 hours. Cells were harvested by centrifugation at 4 °C and 3,000 g for 30 minutes (Sorvall™ LYNX™ 6000, Thermo Scientific) and stored at −20 °C.

Protein purification

The thawed cell pellet was resuspended to a maximum cell concentration of 20 % (w/v) in the binding buffer (20 mM NaP$_i$, 0.5 M NaCl, 20 mM imidazole, 10 % (v/v) glycerol, pH 7.5) supplemented with 250 µg mL^{-1} lysozyme and 2 µg mL^{-1} Dnas I (Applichem, Darmstadt, Germany). Cell lysis was performed using an ultrasonicator UIS250 V (Hielscher Ultrasonics, Brandenburg, Germany) in an ice-water bath at 80 % amplitude with a 0.6-second pulse for 3×5 minutes. The cell lysate was centrifuged at 16 °C, 27,000 g for 40 minutes (Sorvall™ LYNX™ 6000, Thermo Scientific). The soluble protein fraction was filtered through a 0.45 µm syringe filter cellulose acetate membrane (VWR).

Immobilized-metal affinity chromatography (IMAC) was performed using the AEKTA Purifier system (GE HealthCare, Chicago, IL, USA) equipped with a 5 mL HisTrap FF Crude column (Cytiva, Freiburg im Breisgau, Germany). The purification column was equilibrated with 10 column volumes (CV) of the binding buffer and the clarified cell lysate was applied. The column was washed with ~10 CV of the binding buffer to remove unspecific host cell proteins. The imidazole concentration was increased with the elution buffer (20 mM NaPi, 0.5 M NaCl, 0.5 M imidazole, 10 % (v/v) glycerol, pH 7.5) at a gradient of 10–20 mM mL^{-1} for the elution of the bound protein. The eluate fractions containing the target protein were pooled.

The eluate buffer was exchanged to the storage buffer (100 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), 10 % (v/v) glycerol, pH 7.5) using a HiPrep26/10 desalting column (Cytiva). Desalted protein solutions were flash-frozen in liquid nitrogen and stored as small protein beads at −80 °C.

Determination of DNA and Protein Concentrations

DNA and protein concentrations were measured using a Nano-Photometer® P-Class (Implen, Munich, Germany) at $\lambda = 260$ and 280 nm, respectively. Background signals were abstracted by preceding blank measurements. Molecular weights MW and molar extinction coefficients ε_{280} were determined using the ProtParam tool (https://web.expasy.org/protparam/).

Enzyme Activity Assay

The catalytic activity of the aldehyde dehydrogenases was measured photometrically using a BioTek Epoch 2 microplate spectrophotometer (Agilent). 20 µL of enzyme solution was added to the reaction mixture in a total reaction volume of 200 µL or 100 µL for 96- or 384-well microplates, respectively. Standard assays contained 5 mM D-glyceraldehyde and 5 mM NAD$^+$ in 100 mM HEPES, pH 7.5–8.0 and were performed at 37 or 50 °C. The stoichiometric increase in NADH concentration with product formation was monitored at $\lambda = 340$ nm (ε_{340} (NADH) = 6.22 L mmol^{-1} cm^{-1}). One unit (U) of enzyme activity was defined to convert 1 µmol of substrate per minute. K_m values for the substrate and V_{max} were determined at a defined NAD$^+$ concentration of 5 mM by varying the substrate concentration. OriginPro 2021b was used for the evaluation.

Substrate Scope, Selectivity, and Unspecific Selectivity

To determine the substrate scope, the specific activities of the enzymes were measured against D-glyceraldehyde and other aldehyde species (*e.g.* isobutyraldehyde, acetaldehyde) at 5 mM concentrations at 50 °C. Substrate preference was expressed as the relative activity towards an unspecific aldehyde based on the activity towards D-glyceraldehyde:

$$\textit{Relative activity } (\%) = \frac{\textit{Specific activity with unspecific aldehyde}}{\textit{Specific activity with D} - \textit{glyceraldehyde}} \times 100 \quad (1)$$

The substrate selectivity and unspecific selectivity of the enzymes were determined as the ratio of the catalytic efficiencies k_{cat}/K_m towards D-glyceraldehyde and acetaldehyde:

$$\textit{Substrate selectivity} = \frac{k_{cat}/K_{m \, D-glyceraldehyde}}{k_{cat}/K_{m \, Acetaldehyde}} \quad (2)$$

$$\textit{Unspecific selectivity } (\%) = \frac{\frac{k_{cat}}{K_{m \, Acetaldehyde}}}{\frac{k_{cat}}{K_{m \, D-glyceraldehyde}}} \times 100 \quad (3)$$

Melting Point Analysis

Melting temperatures T_m were determined by Thermofluor analysis[35] with 25 μL reaction mixtures containing 2 μL of 1:80 diluted SYPRO™ Orange (Invitrogen, Waltham, MA, USA) and 80 μg mL^{-1} protein in 100 mM HEPES, pH 7.5. Samples were covered with a transparent seal (Bio-Rad) and incubated in a clear PCR plate (Bio-Rad) in the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Fluorescence intensity was detected after each 0.5 °C increase in temperature per 5 seconds from 4.0 °C to 100.0 °C. The negative derivative of relative fluorescence values versus temperature ($-$d-(RFU)/dT) was calculated by the Bio-Rad CFX Manager software, and the minimum was defined as T_m.

Determination of Enzyme Functional Stability

1 mg mL^{-1} enzyme in 100 mM HEPES, pH 7.5 (or pH 8.0 with 0.5 M NaCl for *Mj*ALDH) were supplemented with 5 mg mL^{-1} bovine serum albumin and incubated in PCR strip tubes in the CFX96 Touch Real-Time PCR Detection System (Bio-Rad) at a defined temperature. For stability measurements in ethanol or isobutanol, corresponding organic solvents were added to defined final concentrations. The time-dependent decrease in enzyme activity was monitored photometrically. The half-life $t_{1/2}$ was determined as the time point with the half-maximum of the initial enzyme activity.

IC$_{50}$ Determination

Enzyme activity was measured photometrically in isobutanol concentrations of 0–7% at 50 °C and the isobutanol concentration of the half maximal enzyme activity was determined as IC$_{50}$.[36]

In vitro Biotransformation of D-Glucose to Ethanol

Enzyme mixtures were prepared according to Table S14 and concentrated using 10 kDa centrifugal filters with modified PES (VWR). The reaction mixtures contained 5 mM NAD$^+$, 0.5 mM thiamine pyrophosphate, 5 mM MgCl$_2$, and 200 mM HEPES, pH 7.5. Cascade reactions were initiated by the addition of D-glucose.

Small-scale experiments (30–50 μL) for the conversion of 200 mM D-glucose were performed in PCR tubes in the C1000 Touch™ Thermal Cycler (Bio-Rad) with endpoint analysis, and the final cascades of 250 μL were performed in 2.0 mL reaction tubes (Eppendorf, Hamburg, Germany) at 600 rpm in a ThermoMixer C (Eppendorf) covered with a plastic lid. For the latter, 7 μL of the samples were taken at each time point for HPLC analysis. Samples were diluted in 1:25–60 in 2.5 mM H$_2$SO$_4$ and filtered through 10 kDa centrifugal columns (VWR). HPLC measurements (Ultimate-3000 HPLC system equipped with an autosampler, a diode array detector (Dionex Softron, Germering, Germany) and a refractive index detector (RI-101, Shodex, Gersthofen, Germany)) using an ion exclusion column (Rezex ROA-Organic Acid H + (8%), Phenomenex, Aschaffenburg, Germany) were performed with isocratic elution with 2.5 mM H$_2$SO$_4$ at 70 °C for 30 minutes.[15a] D-glucose, D-gluconate, 2-keto-3-deoxy-gluconate, D-glyceraldehyde, D-glyceraldehyde, pyruvate, ethanol, glycerol, acetic acid, lactic acid, acetoin, and 2,3-butanediol were detected as shown in Figures S26, S27 and S28. Quantification was performed using Chromeleon Software (Thermo Fisher Scientific, Waltham, US).

Ancestral Sequence Reconstruction

ASR (http://grasp.scmb.uq.edu.au/guide) was performed with 5,000 homologous enzyme sequences each of *Hs*ALDH2 (WP_013233459.1) and *Vp*ALDH3 (WP_013542910.1), and the *Ta*ALDH sequence (WP_010901221.1) from the NCBI databases (Protein BLAST®; https://blast.ncbi.nlm.nih.gov/Blast.cgi). MAFFT[37] was used for sequence alignments and FastTree[38] to infer phylogenetic trees by maximum likelihood. The *Ta*ALDH sequence was used to root the tree. GRASP (version 2020.05.05) was used to reconstruct ancestral sequences and perform mutation analysis.[25] Alternatively, sequences of homologous enzymes with ≥70% identity to *Hs*ALDH2 were collected from Protein BLAST®/UniProtKB and a consensus sequence was generated by EMBOSS Cons (http://www.ebi.ac.uk/Tools/msa/emboss_cons/).

Protein Structure Analysis and Docking Experiments

Predictions of protein structures were performed using Phyre2,[39] AlphaFold2,[40] or Robetta (https://robetta.bakerlab.org/). Graphical representations of protein structures were performed by the open-source PyMol™ Molecular Graphics System Version 2.5.0 (Schrö-dinger). Molecular docking experiments were performed by YASARA Version 13.9.8 (Elmar Krieger, IMBM, Austria, CMBI, The Netherlands) with FoldX Suite 4.0[41] for energy minimization or Autodock Vina/Autodock Version 4.2.

Supporting Information

The authors have cited additional references within the Supporting Information.[42–48]

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

ChemSusChem

Research Article
doi.org/10.1002/cssc.202301132

Chemistry
Europe
European Chemical
Societies Publishing

[1] I. Capellán-Pérez, M. Mediavilla, C. de Castro, Ó. Carpintero, L. J. Miguel, *Energy* **2014**, *77*, 641–666.

[2] U. S. Energy Information Administration, *Total energy production, Total energy consumption: Primary energy.* http://www.eia.gov/international/data/world (accessed December 17, 2022).

[3] The Intergovernmental Panel on Climate Change, *Synthesis report of the IPCC sixth assessment report (AR6); Summary for Policymakers.* https://report.ipcc.ch/ar6syr/pdf/IPCC_AR6_SYR_SPM.pdf (accessed March 21, 2023).

[4] C. H. Christensen, J. Rass-Hansen, C. C. Marsden, E. Taarning, K. Egeblad, *ChemSusChem* **2008**, *1*, 283–289.

[5] a) S. Fernando, S. Adhikari, C. Chandrapal, N. Murali, *Energy Fuels* **2006**, *20*, 1727–1737; b) B. Berge, *The Ecology of Building Materials*, Routledge, **2009**, pp. 150–167; c) B. von Vacano, H. Mangold, G. W. Vandermeulen, G. Battagliarin, M. Hofmann, J. Bean, A. Künkel, *Angew. Chem. Int. Ed.* **2022**, *62*, e202210823.

[6] a) S. Takkellapati, T. Li, M. A. Gonzalez, *Clean Technol. Environ. Policy* **2018**, *20*, 1615–1630; b) R. A. Sheldon, *Green Chem.* **2014**, *16*, 950–963.

[7] R. Muktham, S. K. Bhargava, S. Bankupalli, A. S. Ball, *J. Sustain. Bioenergy Syst.* **2016**, *6*, 72–92.

[8] C. E. Hodgman, M. C. Jewett, *Metab. Eng.* **2012**, *14*, 261–269.

[9] E. Buchner, *Ber. Dtsch. Chem. Ges.* **1897**, *30*, 1110–1113.

[10] a) M. Teshima, V. P. Willers, V. Sieber, *Curr. Opin. Biotechnol.* **2023**, *79*, 102868; b) J. K. Guterl, D. Garbe, J. Carsten, F. Steffler, B. Sommer, S. Reisse, A. Philipp, M. Haack, B. Ruhmann, A. Koltermann, U. Kettling, T. Bruck, V. Sieber, *ChemSusChem* **2012**, *5*, 2165–2172; c) S. Sherkhanov, T. P. Korman, S. Chan, S. Faham, H. Liu, M. R. Sawaya, W. T. Hsu, E. Vikram, T. Cheng, J. U. Bowie, *Nat. Commun.* **2020**, *11*, 4292; d) S. Sutiono, A. Pick, V. Sieber, *Green Chem.* **2021**, *23*, 3656–3663; e) P. H. Opgenorth, T. P. Korman, J. U. Bowie, *Nat. Chem. Biol.* **2016**, *12*, 393–395; f) M. W. Ullah, W. A. Khattak, M. Ul-Islam, S. Khan, J. K. Park, *Biochem. Eng. J.* **2014**, *91*, 110–119.

[11] N. Budgen, M. J. Danson, *FEBS Lett.* **1986**, *196*, 207–210.

[12] S. Atsumi, T. Hanai, J. C. Liao, *Nature* **2008**, *451*, 86–89.

[13] a) T. J. Gmelch, J. M. Sperl, V. Sieber, *Sci. Rep.* **2019**, *9*, 11754; b) F. Steffler, J. K. Guterl, V. Sieber, *Enzyme Microb. Technol.* **2013**, *53*, 307–314; c) T. J. Gmelch, J. M. Sperl, V. Sieber, *ACS Synth. Biol.* **2020**, *9*, 920–929.

[14] a) J. E. Hodge, *J. Agric. Food Chem.* **1953**, *1*, 928–943; b) F. J. Tessier, V. M. Monnier, L. M. Sayre, J. A. Kornfield, *Biochem. J.* **2003**, *369*, 705–719.

[15] a) S. Sutiono, M. Teshima, B. Beer, G. Schenk, V. Sieber, *ACS Catal.* **2020**, *10*, 3110–3118; b) O. Melse, S. Sutiono, M. Haslbeck, G. Schenk, I. Antes, V. Sieber, *ChemBioChem* **2022**, *23*, e202200088; c) T. Bayaraa, T. Lonhienne, S. Sutiono, O. Melse, T. B. Brück, E. Marcellin, P. V. Bernhardt, M. Boden, J. R. Harmer, V. Sieber, L. W. Guddat, G. Schenk, *Chemistry.* **2023**, *29*, e202203140.

[16] K. Shortall, A. Djeghader, E. Magner, T. Soulimane, *Front. Mol. Biosci.* **2021**, *8*, 659550.

[17] V. Vasiliou, D. W. Nebert, *Hum. Genomics* **2005**, *2*, 138–143.

[18] UniProt, http://www.uniprot.org/ (accessed September 29, 2023).

[19] V. Vasiliou, D. W. Nebert, *Hum. Genomics* **2005**, *2*, 138–143.

[20] M. Reher, P. Schonheit, *FEBS Lett.* **2006**, *580*, 1198–1204.

[21] a) L. L. Grochowski, H. Xu, R. H. White, *J. Bacteriol.* **2006**, *188*, 2836–2844; b) B. Begander, A. Huber, J. Sperl, V. Sieber, *Catalysts* **2020**, *11*:31.

[22] a) B. Jia, L. T. Linh, S. Lee, B. P. Pham, J. Liu, H. Pan, S. Zhang, G.-W. Cheong, *Extremophiles* **2011**, *15*, 337–346; b) V. P. Willers, B. Beer, V. Sieber, *ChemSusChem* **2023**, *16*, e202202122.

[23] Z.-J. Liu, Y.-J. Sun, J. Rose, Y.-J. Chung, C.-D. Hsiao, W.-R. Chang, I. Kuo, J. Perozich, R. Lindahl, J. Hempel, *Nat. Struct. Biol.* **1997**, *4*, 317–326.

[24] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–410.

[25] G. Foley, A. Mora, C. M. Ross, S. Bottoms, L. Sützl, M. L. Lamprecht, J. Zaugg, A. Essebier, B. Balderson, R. Newell, R. E. S. Thomson, B. Kobe, R. T. Barnard, L. Guddat, G. Schenk, J. Carsten, Y. Gumulya, B. Rost, D. Haltrich, V. Sieber, E. M. J. Gillam, M. Bodén, *PLOS Comput Biol.* **2022**, *18*, e1010633.

[26] Y. Nov, *Appl. Environ. Microbiol.* **2012**, *78*, 258–262.

[27] a) K. K. Podkaminer, X. Shao, D. A. Hogsett, L. R. Lynd, *Biotechnol. Bioeng.* **2011**, *108*, 1268–1278; b) A. C. Ross, G. Bell, P. J. Halling, *J. Mol. Catal. B* **2000**, *8*, 183–192; c) H. Yoshikawa, A. Hirano, T. Arakawa, K. Shiraki, *Int. J. Biol. Macromol.* **2012**, *50*, 865–871.

[28] F. T. Bergmann, S. Hoops, B. Klahn, U. Kummer, P. Mendes, J. Pahle, S. Sahle, *J. Biotechnol.* **2017**, *261*, 215–220.

[29] S. Arrhenius, *J. Phys. Chem.* **1889**, *4*, 226–248.

[30] F. Steffler, V. Sieber, *PLoS One* **2013**, *8*, e70592.

[31] S. H. Mohd Azhar, R. Abdulla, S. A. Jambo, H. Marbawi, J. A. Gansau, A. A. Mohd Faik, K. F. Rodrigues, *Biochem. Biophys. Rep.* **2017**, *10*, 52–61.

[32] L. Zheng, U. Baumann, J. L. Reymond, *Nucleic Acids Res.* **2004**, *32*, e115.

[33] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343–345.

[34] H. Inoue, H. Nojima, H. Okayama, *Gene* **1990**, *96*, 23–28.

[35] S. Boivin, S. Kozak, R. Meijers, *Protein Expression Purif.* **2013**, *91*, 192–206.

[36] S. Sutiono, J. Carsten, V. Sieber, *ChemSusChem* **2018**, *11*, 3335–3344.

[37] K. Katoh, K. Misawa, K. i. Kuma, T. Miyata, *Nucleic Acids Res.* **2002**, *30*, 3059–3066.

[38] a) M. N. Price, P. S. Dehal, A. P. Arkin, *Mol. Biol. Evol.* **2009**, *26*, 1641–1650; b) M. N. Price, P. S. Dehal, A. P. Arkin, *PLoS One* **2010**, *5*, e9490.

[39] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **2015**, *10*, 845–858.

[40] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.

[41] a) R. Guerois, J. E. Nielsen, L. Serrano, *J. Mol. Biol.* **2002**, *320*, 369–387; b) J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, *Nucleic Acids Res.* **2005**, *33*, W382–388.

[42] a) J. Farres, T. T. Wang, S. J. Cunningham, H. Weiner, *Biochemistry* **1995**, *34*, 2592–2598; b) Y. Tsybovsky, S. A. Krupenko, *J. Biol. Chem.* **2011**, *286*, 23357–23367.

[43] E. Vázquez-Figueroa, J. Chaparro-Riggers, A. S. Bommarius, *ChemBioChem* **2007**, *8*, 2295–2301.

[44] M. Reher, T. Fuhrer, M. Bott, P. Schönheit, *J. Bacteriol.* **2010**, *192*, 964–974.

[45] S. Sutiono, K. Satzinger, A. Pick, J. Carsten, V. Sieber, *RSC Adv.* **2019**, *9*, 29743–29746.

[46] A. Guagliardi, M. Martino, I. Iaccarino, M. D. Rosa, M. Rossi, S. Bartolucci, *Int. J. Biochem. Cell Biol.* **1996**, *28*, 239–246.

[47] S. Wolterink-van Loo, A. Van Eerde, M. A. Siemerink, J. Akerboom, B. W. Dijkstra, J. Van der Oost, *Biochem. J.* **2007**, *403*, 421–430.

[48] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, *Nat. Methods* **2022**, *19*, 679–682.