To be submitted to Journal of Bacteriology

TITLE: Cofactor specificity of the bifunctional alcohol and aldehyde dehydrogenase (AdhE) in wild-type and mutants of *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*

AUTHORS: Tianyong Zheng,[a,f] Daniel G. Olson,[b,f] Liang Tian,[b,f] Yannick J. Bomble,[c,f] Michael E. Himmel,[c,f] Jonathan Lo, [a,f] Shuen Hon,[b,f] A. Joe Shaw,[d] Johannes P. van Dijken,[e] Lee R. Lynd[a,b,f,#]

Department of Biological Sciences, Dartmouth College, Hanover, NH, USA[a]; Thayer School of Engineering, Dartmouth College, Hanover, NH, USA[b]; Biosciences Center, National Renewable Energy Laboratory, Golden, CO, USA[c]; Novogy Inc., Cambridge, MA, USA[d]; Delft University of Technology, Delft, the Netherlands[e]; BioEnergy Science Center, Oak Ridge, TN, USA[f]

RUNNING TITLE: Comparison of AdhE in *C. thermocellum* and *T. saccharolyticum*.

#Corresponding author: Lee R. Lynd, Lee.R.Lynd@Dartmouth.edu

ABSTRACT

Clostridium thermocellum and *Thermoanaerobacterium saccharolyticum* are thermophilic bacteria that have been engineered to produce ethanol from the cellulose and hemicellulose fractions of biomass respectively. Although engineered strains of *T. saccharolyticum* produce ethanol with a yield of 90% of the theoretical maximum, engineered strains of *C. thermocellum* produce ethanol at lower yields (~50% of theoretical maximum). In the course of engineering these strains, a number of mutations have been discovered in their *adhE* genes, which encode both alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) enzymes. To understand the effect of these mutations, the *adhE* genes from six strains of *C. thermocellum* and *T. saccharolyticum* were cloned and expressed in *Escherichia coli*, followed by purification by affinity chromatography and enzyme activity measurement. In wild type strains of both organisms, NADH was the preferred cofactor for both ALDH and ADH activity. In high ethanol-producing ("ethanologen") strains of *T. saccharolyticum*, both ALDH and ADH activities showed increased NADPH-linked activity. Interestingly, the ethanologenic strain of *C. thermocellum*, AdhE has acquired high NADPH-linked ADH activity while maintaining NADH-linked ALDH and ADH activity at wild-type levels. When single amino acid mutations in AdhE, that caused increased NADPH-linked ADH activity, were introduced into *C. thermocellum* and *T. saccharolyticum*, ethanol production increased in both organisms. Structural analysis was performed for the wild-type and mutant AdhE proteins to provide explanations for the cofactor specificity change on a molecular level.

IMPORTANCE

This work describes the characterization of the AdhE enzyme from different strains of *C. thermocellum* and *T. saccharolyticum*. *C. thermocellum* and *T. saccharolyticum* are thermophilic anaerobes that have been engineered to make high yields of ethanol and can solubilize components of plant biomass and ferment the sugars to ethanol. In the course of engineering these strains, several mutations arose in the bifunctional alcohol/aldehyde dehydrogenase AdhE, changing both enzyme activity and cofactor specificity. We show that changing AdhE cofactor specificity from mostly NADH-linked to mostly NADPH-linked resulted in higher ethanol production in *C. thermocellum* and *T. saccharolyticum*.

INTRODUCTION

Thermophilic organisms, *Clostridium thermocellum* in particular, hold great promise for production of ethanol from lignocellulosic feedstocks (1, 2). *C. thermocellum* is a thermophilic, gram-positive obligate anaerobe that rapidly consumes cellulose. Engineered strains of *Thermoanaerobacterium saccharolyticum* (3), a thermophilic anaerobe that ferments xylan and other sugars derived from biomass, have been shown to produce over 50 g/L ethanol at near-theoretical yield (4). While comparable concentrations of ethanol are tolerated by selected strains of *C. thermocellum* (5–7), the maximum reported concentration of ethanol produced by this organism is 23.6 g/L (8) and the maximum ethanol yield achieved to date is 51% of theoretical maximum (9) versus 92% in *T. saccharolyticum* (10). It is of interest to understand why ethanol production in *T. saccharolyticum* is thus far superior to that in *C. thermocellum*, in order to facilitate engineering of the *C. thermocellum* ethanol-production pathway.

In microorganisms, fermentation of pyruvate to ethanol can proceed either with or without acetyl-CoA as an intermediate. In yeasts and *Zymomonas mobilis*, pyruvate is decarboxylated directly to acetaldehyde, which is then reduced to ethanol (11). In many other organisms, pyruvate is oxidatively decarboxylated to acetyl-CoA, which is reduced to acetaldehyde, which is further reduced to ethanol. This two-step conversion of acetyl-CoA to ethanol can be catalyzed by one protein: a bifunctional alcohol dehydrogenase AdhE. AdhE consists of a C-terminal alcohol dehydrogenase (ADH) domain and an N-terminal aldehyde dehydrogenase (ALDH) domain: the ADH domain is usually part of the iron-containing ADH superfamily (Fig. 1) (12). AdhE is present in a variety of

mesophilic and thermophilic anaerobic bacteria capable of producing ethanol as a fermentation product (13–16). AdhE has also been found in parasitic eukaryotes (17), anaerobic fungi (18) and algae (19). In all organisms investigated thus far, the deletion of *adhE* is associated with a loss of ethanol formation. When *adhE* was deleted in *C. thermocellum* and *T. saccharolyticum*, nearly 100% of ethanol production was eliminated (20), demonstrating the importance of AdhE in ethanol formation in these two organisms.

Numerous mutations in *adhE* have appeared during the course of engineering *C. thermocellum* and *T. saccharolyticum* for higher ethanol production and tolerance (Figure 1, Table 1). In the *C. thermocellum* strain LL346 (a.k.a "*adhE**"), the mutations P704L and H734R in AdhE were associated with higher ethanol tolerance (21); in the strain LL350, the D494G mutation in AdhE was associated with higher ethanol production (22). In the *T. saccharolyticum* strain LL1040 (a.k.a "ALK2") (10) and LL1049 (23), higher ethanol yield was achieved and mutations were also observed in AdhE. In wild-type *C. thermocellum* and *T. saccharolyticum*, the ADH activity in cell extracts was largely NADH-linked (10, 21). In all of the engineered strains mentioned above, there was an increase in NADPH-linked ADH activity in cell extracts (10, 21, 22). However, it has not been unequivocally established whether this cofactor specificity change must be ascribed to mutations in AdhE, as cells contain multiple alcohol dehydrogenases and measurements with cell extracts cannot distinguish between isoenzymes.

It was therefore of interest to investigate purified preparations of AdhE and its mutant forms to address the following questions:

1. Whether ALDH and/or ADH activity has changed in cofactor specificity.

97 2. Whether it is a general trend that the change of AdhE cofactor specificity
98 from NADH to NADPH is associated with more favorable features in ethanol production.
99

MATERIALS AND METHODS

Plasmid and strain construction. The *adhE* genes from strains LL1004 (wt *C. thermocellum*), LL346, LL350, LL1025 (wt *T. saccharolyticum*), LL1040, LL1049 (See Table 1 for strain descriptions, accession numbers for *adhE* genes in the order above: KR632757, KR632758, KR632759, KR632761, KR632760, KR632762) were cloned into plasmid pEXP5-NT/TOPO (Invitrogen) with standard molecular biology techniques, generating the respective *E. coli* expression plasmids (Table S1). Cloning the *adhE* genes into pEXP5-CT/TOPO plasmids instead of pEXP5-NT/TOPO generated native AdhE proteins without His-tags. The plasmids were Sanger sequenced (Genewiz) to confirm correct insertion of the target gene, and were then transformed into chemically competent *lysY/I*^{q} (New England Biolabs) *E. coli* cells. The control plasmid pNT-CALML3 (Invitrogen) was also transformed into *E. coli*. The resulting *E. coli* strains were used for protein expression. *C. thermocellum* strains LL1160 and LL1161 were constructed by transforming the respective integration plasmids pSH016 and pSH019 into strain LL1111 (Table 1, Table S1); transformation and colony selection were carried out as previously described (24). *T. saccharolyticum* strains LL1193 and LL1194 were constructed by transforming the respective vectors pCP14 and pCP14* into wild-type *T. saccharolyticum* using a natural competence based system (25) (Table 1, Table S1), and transformants were selected by resistance to the antibiotic kanamycin.

Media and growth conditions. For biochemical characterization and transformation, *C. thermocellum* and *T. saccharolyticum* strains were grown anaerobically to exponential phase (OD_{600}~0.5) in the appropriate medium: for *C. thermocellum*, CTFUD rich medium at pH 7.0 as previously described (24); for *T.*

123 *saccharolyticum*, CTFUD rich medium at pH 6.0. *E. coli* strains were grown in LB broth
124 Miller (Acros) with the appropriate antibiotic. Fermentation end-products were measured
125 using high-pressure liquid chromatography as previously described (26). For end-product
126 analysis, *C. thermocellum* and *T. saccharolyticum* strains were grown in the appropriate
127 medium: for *C. thermocellum*, chemically defined MTC medium as previously described
128 (27); for *T. saccharolyticum*, the MTC medium was modified as follows: thiamine was
129 added to a final concentration of 4 mg/L, and urea was replaced with ammonium chloride
130 at a final concentration of 1.5 g/L. In preparation for fermentation end-product analysis,
131 cultures were grown at 55°C in 150 mL serum bottles with 50 mL working volume and
132 100 mL headspace for 72 h. Ethanol concentrations were calculated from biological
133 duplicates and reported in Table 2, other end-products are reported in Table S2.

134 **Expression of various *adhE* genes.** 500 µl of *E. coli* culture containing a plasmid
135 with the *adhE* gene of interest was inoculated into 100 mL sterile LB broth Miller
136 (Acros) with the appropriate antibiotic, and were grown aerobically to OD_{600} = 0.5 with
137 shaking at 200 RPM at 37°C (Eppendorf Innova 42 shaker). The *E. coli* strain harboring
138 the pNT-CALML3 control plasmid (Invitrogen) was used as a negative control to
139 measure native *E. coli* ADH or ALDH activity. Because the *E. coli* AdhE was shown to
140 be sensitive to oxygen (13), and *C. thermocellum* cell extracts lost ADH activity after
141 exposure to air for 30 minutes (data not shown), AdhE protein expression and all
142 subsequent experiments were carried out anaerobically. The *E. coli* cultures were then
143 transferred to sterile serum bottles, and 40 mM IPTG (Isopropyl β-D-1-
144 thiogalactopyranoside) was used to induce protein expression. The serum bottles were

purged with nitrogen to generate an anaerobic protein expression environment, and the cells were cultured for an additional 2 h at 37°C before harvesting.

Preparation of cell extracts. *C. thermocellum*, *T. saccharolyticum* and *E. coli* cultures were grown as described above. Cells were harvested by centrifugation at 3000 ×*g* for 30 min at 4°C, the supernatant was decanted and the pellet stored anaerobically at -80°C. All cell extracts were generated anaerobically. Prior to generating cell extracts, the frozen pellets were thawed on ice and resuspended in 0.5 mL Lysis Buffer: 1X BugBuster reagent (EMD Millipore) at pH 7.0 in phosphate buffer (100 mM) with 5 μM $FeSO_4$. Dithiothreitol (DTT) was added to a final concentration of 0.1 mM. For *T. saccharolyticum* cell extracts used in ALDH activity measurements, ubiquinone-0 (Sigma catalog number D9150) was added to the final concentration of 2 mM to relieve the possible inhibition of ALDH activity as previously reported (20). The cells were lysed with Ready-Lyse Lysozyme (Epicentre), and DNase I (New England Biolabs) was added to reduce viscosity. The resulting solution was centrifuged at 10,000 ×*g* for 5 min at room temperature, and the supernatant was used as cell-free-extract for enzyme assays.

Protein purification. The *E. coli* crude extracts described above were incubated at 50°C anaerobically for 20 min to denature *E. coli* proteins, and the denatured proteins were separated by centrifugation. The ADH activity in *E. coli* cell extracts expressing AdhE from strains LL346 and LL1040 was sensitive to heat, and lost activity after 50°C incubation, so these cell extracts were applied directly to the purification column without heating. *E. coli* cells expressing the control plasmid pNT-CALML3 were subject to the same treatment as above, and its ALDH and ADH activities before and after heat treatment were measured (Table S3). The cell extracts containing His-tagged AdhE were

then subjected to anaerobic affinity column purification (Ni-NTA spin columns, Qiagen). The purification was carried out according to the Qiagen protocol "Ni-NTA Agarose Purification of 6xHis-tagged Proteins from *E. coli* under Native Conditions" with some modifications as described below. The column was first equilibrated with Equilibrium Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 5 μM $FeSO_4$, pH 7), then cell extracts were applied to the column and the column was washed twice with Wash Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 50 mM imidazole, 20% ethanol, 5 μM $FeSO_4$, pH 7). The His-tagged AdhE was eluted by addition of 200 μL Elution Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, 5 μM $FeSO_4$, pH 7), this is "Eluent 1" in Table S3 and S4. Repeating this elution step sequentially generated more-purified "Eluent 2" and "Eluent 3". For *C. thermocellum* and *T. saccharolyticum* AdhE, activity was measured at various stages during purification (Table S4). Electrophoresis results showed that Eluent 3 had the least amount of contaminating bands, thus Eluent 3 was used for enzyme assays. The degree of protein purity was estimated by gel densitometry using the image analysis software ImageJ, where the density of each visible gel band from Eluent 3 was plotted against its size in the gel. The area of each resulting peak was then integrated and compared to the total of all peaks to estimate the percentage of protein that can be attributed to each band (Table S5). *E. coli* cell extracts with native AdhE expressed (i.e. without the His-tag) were used directly without purification (Table S6).

ALDH and ADH activity assays. All the ALDH activity measurements mentioned in this study refer to the reaction in the acetaldehyde-producing direction. All the ADH activity measurements mentioned in this study refer to the reaction in the ethanol-producing direction. For ADH (acetaldehyde reduction) reactions, the anaerobic

reaction mixture contained 0.24 mM NADH or NADPH, 17.6 mM acetaldehyde, 1 mM DTT, 100 mM Tris-HCl, 5 µM $FeSO_4$ and cell extract or purified protein solution (protein amount indicated separately for each assay). The final volume was 850 µL, the assay temperature was 55°C and the assay was started by the addition of acetaldehyde. For ALDH (acetyl-CoA reduction) reactions, the acetaldehyde in the above anaerobic reaction mixture was substituted with 0.35 mM acetyl-CoA. Background activity was recorded before start of reaction (addition of acetaldehyde or acetyl-CoA), and was subtracted from the reaction activity recorded. In the inhibition assays (Table S8), the inhibitor was added before the start of the reaction at the following concentrations: 1M ethanol or 2.35mM $NAD(P)^+$. Decrease in absorbance at 340 nm caused by NAD(P)H oxidation was monitored by an Agilent 8453 UV-Vis spectrophotometer with Peltier temperature control (28). Protein concentration was determined using the Bradford method with bovine serum albumin (Thermo Scientific) as the standard. Specific activities are expressed as units per mg of protein. One unit of activity = formation of 1 µmol of product per min. Specific activities in Table 2 and Table 3 are reported for at least two biological replicates. T-tests were performed to analyze the cofactor specificity in Table 2 and Table 3: NADH-linked and NADPH-linked activities were analyzed using the unpaired two-tailed T-test, and differences between the two cofactors were considered significant if the p-value was <0.05. Standard deviations and raw data for all enzyme assays are presented in Table S7. The software Visual Enzymics (SoftZymics) was used for non-linear regression to calculate the apparent K_m and k_{cat} values in Table 4. k_{cat} was calculated on the basis of a molecular weight of 96-kDa for LL1004 AdhE and LL350 AdhE.

Homology modeling and molecular dynamics. The homology models corresponding to the ADH domains of the AdhE from LL1004, LL346, LL350, LL1025, LL1040 and LL1049 were constructed using the bioinformatics toolkit SWISS-MODEL (Swiss Institute of Bioinformatics). The recent 2.5 Å resolution X-ray structure of the alcohol dehydrogenase of *Geobacillus thermoglucosidasius* (3ZDR) (12) and 1.3 Å resolution X-ray structure of the alcohol dehydrogenase from *Thermatoga maritima* (1O2D) (29) were used as templates for their high level of homology and presence of NADP cofactor and iron ion, respectively. The resulting structures were inspected for proper phi/psi angles. All resulting structures were submitted to molecular dynamics simulations using the program CHARMM with the CHARMM 36 force field and the TIP3P water model (30). The systems were generated via the CHARMM-GUI web server (31) and the parameters for NADP were generated by ParamChem. The structures were initially minimized in-vacuo with steepest decent for 1000 steps and then solvated in a cubic water box with a minimum of 10 Å from the edge of the box, sodium cations were added to neutralize the system. These resulting systems were minimized using steepest decent for 1000 steps followed by newton-raphson minimization for 100 steps. They were then submitted to 1-ns equilibration in the NPT ensemble at 298K and 1 bar followed by 10 ns simulation in the NVT ensemble with an integration time step of 2 fs. All simulations were conducted in duplicate with different starting seeds and analyzed using carma (32).

Journal of Bacteriology

RESULTS

Cofactor specificity of wild-type and mutant AdhE. ALDH and ADH activity was determined in cell extracts of *C. thermocellum* and *T. saccharolyticum* strains, as well as in the affinity-purified AdhE expressed in *E. coli*. In purified preparations of AdhE, gel densitometry results showed that the proteins were all > 70% pure (Table S5). Furthermore, negative *E. coli* controls all showed <0.4 U/mg specific activity for ALDH and ADH (Table S3), indicating that the small amount of contaminating protein observed on the gel did not substantially interfere with ADH or ALDH activity measurements. With respect to cofactor preference, the cell extracts of the *T. saccharolyticum* high-ethanol-producers (LL1040, LL1049) has changed their ADH and ALDH cofactor specificity to mostly NADPH-linked, as compared to mostly NADH-linked in wild-type (LL1025) (Table 2). The difference between NADH and NADPH-linked activity was significant (P-value <0.05) for all three strains except for ADH activity in LL1049. In *C. thermocellum* cell extracts, wild-type (LL1004) and the ethanol-tolerant strain (LL346) both showed more preference for NADH as a cofactor, only the preference was statistically significant for wild-type and non-significant for the ethanol-tolerant strain (LL346). The *C. thermocellum* moderate-ethanol-producer (LL350) had significantly increased NADPH-linked ADH activity, while maintaining significant NADH-linkage in ALDH activity. The results for affinity-purified AdhE enzymes (Table 3) also showed higher NADPH-linked AdhE activity in evolved strains. The purified AdhE enzymes were all significantly linked to either NADH or NADPH (P-value < 0.05), with the exception of the LL350 AdhE where, as in cell extracts, NADPH-linked ADH activity increased to an amount comparable to its NADH-linked ADH activity. In all cases, the

cofactor specificity change towards NADPH was much more complete in *T. saccharolyticum* AdhE than in *C. thermocellum* (Table 3). Additionally, strains exhibiting this increase in NADPH-linked cofactor specificity in AdhE also generally showed increased ethanol production compared with their parent strains (Table 2 and 3).

Because the Asp-494-Gly (D494G) mutation (22) in the *C. thermocellum* moderate-ethanol-producer (LL350) AdhE enabled the enzyme to use both NADH and NADPH as cofactors, the apparent k_{cat} and K_m values were measured with purified protein from LL350 and wild-type *C. thermocellum* expressed in *E. coli* (Table 4). Unpaired t-tests were conducted to compare the kinetic parameters in wild-type AdhE and the D494G mutant. In terms of K_m, k_{cat} and catalytic efficiency (k_{cat}/K_m), the newly acquired NADPH-linked activity in the D494G mutant was not significantly different from NADH-linked ADH activity in wild-type (p-values>0.05). For NADH-linked activity, however, a significant decrease (p-value <0.05) in catalytic efficiency was observed for NADH in the D494G mutant compared to wild-type.

Product inhibition (ethanol or $NAD(P)^+$) of purified AdhE proteins from *C. thermocellum* and *T. saccharolyticum* was measured (Table S8). The AdhE of the *C. thermocellum* ethanol-tolerant strain (LL346) was significantly different from other AdhE proteins in both ethanol and $NAD(P)^+$ inhibition. It retained 98% of ADH activity and 92% of ALDH activity in the presence of 2.35 mM NAD^+, while the other five AdhE proteins on average retained only 30% of ADH activity and 33% of ALDH activity. Interestingly, in the presence of 1M ethanol, the LL346 AdhE showed a 2-fold *increase* in ADH activity, while the other five AdhE proteins on average only retained 40% of ADH activity.

Journal of Bacteriology

Effect of AdhE mutations on ethanol production. The physiological effect of two selected point mutations was investigated by re-introducing those mutations into either *C. thermocellum* or *T. saccharolyticum*. For *C. thermocellum*, the D494G mutation (22) was chosen. This mutation could not be introduced directly into the wild-type strain (due to limits of existing genetic tools), so instead *adhE* was deleted (strain LL1111) and replaced by the D494G mutant *adhE* (strain LL1161). A control strain (LL1160) was made by re-introduction of the wild-type *adhE* into strain LL1111. Fermentation of 5 g/L cellobiose resulted in ethanol yield of 0.14 g/g cellobiose for strain LL1160 and 0.24 g/g cellobiose for strain LL1161, a 1.7-fold increase. This increase is larger than the 1.4-fold increase in ethanol yield from *C. thermocellum* wild-type to moderate-ethanol-producer LL350, which has the D494G AdhE mutation but also other genetic modifications.

For *T. saccharolyticum,* the AdhE G544D mutation (23) was chosen. In this organism, the mutation *could* be introduced directly into the wild-type strain, although a kanamycin (*kan*) antibiotic resistance marker had to be added downstream of *adhE*. The resulting strain was LL1194. A control strain (LL1193) was made by inserting only the *kan* marker downstream of *adhE*. Fermentation of 5 g/L cellobiose resulted in ethanol production of 0.21 g/L for strain LL1193 and 0.32 g/L for strain LL1194, a 1.5-fold increase. This increase is comparable to the 1.7-fold increase in ethanol production from *T. saccharolyticum* wild-type to high-ethanol-producer LL1049, which has the G544D AdhE mutation but also other genetic modifications.

AdhE protein structure prediction. To understand the impact of mutations on cofactor specificity we performed homology modeling and docking. The average structure of the ADH domains from wild type and D494G mutants of the *C.*

thermocellum AdhE were compared to identify potential explanations for the switch in cofactor specificity (Fig. 2). In wild-type *C. thermocellum* AdhE, the Asp-494 interferes with the 2'-phosphate group of NADPH because of electrostatic repulsion (both negatively charged) and steric hindrance. Molecular dynamics simulation was conducted to compare the average structures of the six different ADH domains (Fig. 3) including the previously mentioned mutant D494G to evaluate if the observation from homology modeling and docking were correct. The conformation of NADPH in the binding pocket of the ADH domain varied in the AdhE mutants. In the case of *C. thermocellum*, the behavior of NADPH is similar in wild-type *C. thermocellum* (LL1004) AdhE and the ethanol-tolerant *C. thermocellum* (LL346) AdhE where the 2'-phosphate group of NADPH does not have access to the binding pocket (Figure 3). In the *C. thermocellum* moderate-ethanol-producer (LL350) AdhE, the D494G mutation changed NADPH binding significantly as mentioned above, allowing access for the 2'-phosphate group of NADPH to the binding pocket. A similar trend was observed in the case of *T. saccharolyticum* AdhE where the mutations in the high-ethanol-producers LL1049 and LL1040 seem to change the conformation of NADPH in the binding pocket, allowing NADPH to bind.

Journal of Bacteriology

DISCUSSION

Primary structure of AdhE. The ALDH and ADH domains of AdhE are highly conserved and connected by a linker sequence. There is a disagreement in the literature on the number of NADH binding sites in AdhE. Some studies predict a single NADH binding site located within or near the linker region of AdhE (13, 33–37), suggesting that the ALDH and ADH domains share one nicotinamide binding site. Other studies predict an additional NADH binding site in the ALDH domain (19, 28, 38, 39). Fungal AdhE enzymes have been shown to have three putative NADH binding sites (18). In our analysis, we focused on glycine-rich regions and relied on structural information from our homology models or closely-related structural homologs. In both *C. thermocellum* and *T. saccharolyticum* AdhE proteins, we found two strong NADH binding sites (Fig. 1A) and an additional putative nucleotide binding region with a GXGXXG motif in the linker region between the ALDH and ADH domains, which has been reported in previous studies as a potential recognition locus (13, 18, 19, 28, 33–39). This putative binding region within the linker is almost identical to one identified in another iron-dependent alcohol dehydrogenase from *E. coli* (40). In this study, the glycine at the center of the locus was mutated but only resulted in a marginal loss of NAD^+ binding. However, mutations in the other glycine rich locus with the motif GGG located in the ALDH domain resulted in complete loss of NAD^+ binding (40), indicating that the GGG motif is important in NAD^+ binding. The GXGXXG motif located in the linker is conserved in several ADHE enzymes but does not seem to contribute to nucleotide binding directly; however, it might be important in nucleotide channeling or recognition before entering the binding pocket. The NADH binding site in the ALDH domain appears to be a

345 combination of a conventionally accepted binding motif GXGXG and another glycine
346 rich helix turn (Fig 1B). The NADH binding site in the ALDH domain has also been
347 suggested to have acetyl-CoA binding abilities (28). We see a high level of conservation
348 for both strong binding sites across different organisms (Fig 1B and 1C). The prediction
349 of two binding sites (one in each domain) agrees with our observation that the D494G
350 mutant AdhE gained NADPH specificity for ADH activity but not for ALDH activity. If
351 the D494G mutant AdhE shared a single NADH binding site in the linker region, then
352 one would expect to find NADPH cofactor specificity in ALDH activity as well.

353 **The effect of AdhE cofactor specificity on ethanol production.** Thus far, most
354 bifunctional AdhE enzymes investigated are NADH-linked: however, there are some
355 exceptions. The *Thermoanaerobacter mathranii* AdhE showed a small amount of
356 NADPH-linked activity in addition to NADH-linked activity for both ALDH and ADH
357 (37), and the *Thermoanaerobacter ethanolicus* JW200 AdhE showed NADH-linked
358 ALDH activity and small amounts of NADPH-linked ADH activity (36). In all *T.*
359 *saccharolyticum* strains investigated in this study, higher ethanol production was
360 associated with an increase in NADPH specificity and decrease in NADH specificity for
361 both ADH and ALDH activity (Tables 2 and 3). In *C. thermocellum* this association is not
362 as consistent: although NADPH-linked ADH activity significantly increased, NADH-
363 linked activity was maintained at wild-type levels for both ADH and ALDH activities in
364 the moderate-ethanol-producer LL350 strain (Tables 2 and 3). When mutations that were
365 previously shown to increase NADPH-linked ADH activity were re-introduced into wild-
366 type *C. thermocellum* and *T. saccharolyticum* AdhE, ethanol production increased in the
367 resulting strains (LL1161 and LL1194), showing that these mutations (both of which

caused changes in cofactor specificity) are responsible for at least some of the increased ethanol production in the high-ethanol-producing strains. The effect of reintroducing the *adhE* mutations depends somewhat on medium composition. Another study comparing the same mutants found almost no difference in ethanol production (M2202 vs M2203, Table 3, Shaw et al. (23)). We suspect that the difference between our results and their results is due to the presence of large quantities of yeast extract in the medium of Shaw et al.

The physiological reason for the increase in NADPH-linked activity in strains LL350, LL1040 and LL1049 remains to be elucidated. NADPH is generally thought to be the reducing equivalent for anabolism whereas NADH is generally thought to be the reducing equivalent in anaerobic catabolism. Apparently these strains use NADPH for both anabolic and catabolic processes. This may be related to changes in the fluxes of NADH and NADPH generation elsewhere in metabolism. There are several possible sources in *C. thermocellum* and *T. saccharolyticum* that can provide NADPH for ethanol production. The *nfnAB* genes, which are present in both *C. thermocellum* and *T. saccharolyticum*, encode the NfnAB complex that catalyzes the reaction: $2NADP^+ + NADH + Ferredoxin_{red}^{2-} + H^+ \rightarrow 2NADPH + NAD^+ + Ferredoxin_{ox}$ (41). Other *T. saccharolyticum* enzymes that generate NADPH for catabolic purposes include the glucose-6-phosphate dehydrogenase and the phosphogluconate dehydrogenase. These *T. saccharolyticum* genes are both highly expressed (42). Although *C. thermocellum* does not have the above two enzymes present in the pentose phosphate cycle, the malic enzyme in *C. thermocellum*, which catalyzes the formation of pyruvate from L-malate and generates NADPH, is very active (26).

Enzymes in the *T. saccharolyticum* ethanol production pathway. Although the *T. saccharolyticum* LL1040 and LL1049 strains are able to produce ethanol at high yield, purified AdhE proteins from these mutant strains showed comparable or lower ADH activity (Table 3). Furthermore, one-way ANOVA (Analysis of Variance) showed that the ADH activity in cell extracts of LL1040 and LL1049 (Table 2) is not significantly different from that in the *T. saccharolyticum adhE* deletion strain LL1076 (P-value 0.56). This suggests that the *T. saccharolyticum* high-ethanol-producers do not largely rely on the ADH activity from AdhE for ethanol production, and that another alcohol dehydrogenase may be the main ADH in these strains. The cell extract activity measurements in Table 2 suggests that this other ADH is NADPH-linked and may have higher ADH activity than AdhE. It has been reported that an NADPH-linked primary alcohol dehydrogenase AdhA is present in the *Thermoanaerobacter* species, and may be part of the ethanol production pathway (43, 44). Sequence analysis shows *T. saccharolyticum* JW/SL-YS485 has a gene (Tsac_2087) encoding an alcohol dehydrogenase that is 86% identical (at the protein level) to the *T. mathranii* and *T. ethanolicus* AdhA. Other reported NADPH-linked alcohol dehydrogenases involved in ethanol production include the AdhB enzyme, such as the secondary alcohol dehydrogenase reported in *T. ethanolicus* 39E (45). However, sequence analysis showed that *T. saccharolyticum* does not possess an *adhB* gene. Therefore AdhA may be responsible for the observed NADPH-linked ADH activity in *T. saccharolyticum* cell extracts, and also may be important in ethanol production in the *T. saccharolyticum* high-ethanol-production strains LL1040 and LL1049.

Unique characteristics of the ethanol-tolerant LL346 *C. thermocellum* strain. The mutations introduced into the AdhE protein from the ethanol-tolerant strain of *C. thermocellum*, LL346 (aka adhE*) are notably different from the other mutations that we have described thus far with respect to inhibition. It has been reported that low amounts of NAD^+ and ethanol inhibit ADH activity in cell extracts of *C. thermocellum* (46). High inhibition by $NAD(P)^+$ in purified AdhE proteins in *C. thermocellum* was observed (at least 70% activity was inhibited), with the exception of LL346, in which inhibition by NAD^+ was less than 10% (Table S8). Another unexpected property of the AdhE from strain LL346 was increased activity in the presence of ethanol. A similar phenomenon has been observed with the *Z. mobilis* ZADH-2 enzyme, which was also stimulated by ethanol (47). The authors proposed ethanol-induced acceleration of NAD^+ dissociation as a mechanism for the observed activation by ethanol, because nicotinamide dissociation is presumed to be the rate-limiting step in most dehydrogenases.

It has been reported by Brown et al. (21) that mutations in the *adhE* gene of *C. thermocellum* LL346 (P704L and H734R) are the sole basis for the alcohol tolerance of this mutant. As the mutations coincided with a change from NADH-linked to NADPH-linked ADH activity in cell-free extracts, they concluded that these mutations are responsible for a change of the cofactor specificity from NADH to NADPH in the ADH part of AdhE. An increase in NADPH-linked ADH activity was observed in the cell-free extracts of LL346 compared to wild-type (Table 2), but in assays done with the purified AdhE from LL346, nearly 100% of the activity for both ALDH and ADH was NADH-linked (Table 3). This changes the interpretation of the effect of the mutation, and suggests that it reduced enzyme activity instead of changing cofactor specificity. Thus the

Journal of Bacteriology

small increase of NADPH-linked ADH activity observed in the cell extracts of LL346 may be the result of another enzyme.

AdhE cofactor specificity at the molecular level. Several factors can explain the changes in cofactor specificity described in the *C. thermocellum* moderate-ethanol-producer AdhE (from strain LL350). It is clearly not energetically favorable to accommodate the extra 2'-phosphate group in the wild-type *C. thermocellum* AdhE because of the negative charge of Asp-494. This 2'-phosphate group is absent in NADH, which may in fact be stabilized by hydrogen bonding interactions with this residue. This evidence suggests that Asp-494 is important in distinguishing nicotinamide cofactors as previously described. As shown in Figure 2, the substitution of Asp-494 to glycine removes the interference between Asp-494 and NADPH thus enabling the ADH to use both NADH and NADPH as a cofactor. The low K_m value for NADPH in the D494G mutant AdhE (Table 4) agrees with the structural prediction, as it suggests that this mutation resulted in an increase in affinity of the enzyme to NADPH. Aspartic acid residues have been shown to play an important role regulating the binding of NADH over NADPH and are potential targets for mutations to change cofactor specificity. For example, the mutation D38N in the NADH recognition motif of an NADH dependent ADH from a *Drosophila* allowed the enzyme to use both NADH and NADPH (48). A similar study was conducted on an alcohol dehydrogenase yielding the same results (40). The positions of these aspartic acids are almost identical to that of D494 in wild-type *C. thermocellum* (LL1004) AdhE.

Regarding LL346, the mutations would likely lead to a loss of enzymatic activity in AdhE. Even though the LL346 mutations H734R and P704L both occurred in the ADH

Journal of Bacteriology

459 domain, the ALDH activity may also be affected. The H734R mutation has been studied
460 in *E. histolytica* AdhE (a.k.a EhADH2), where it resulted in reduced ALDH and ADH
461 activity (28). Their results suggested that alterations in the ADH domain, especially
462 within the putative iron-binding domain where H734R resides, could affect ALDH
463 domain activity.

464 Helical assemblies of AdhE proteins named "spirosomes" have been observed in
465 many other organisms (12, 28, 34, 49), as well as in recombinant AdhE following His-
466 purification (50). The formation of such structures has been suggested to influence
467 enzyme activity (28). The formation of this quaternary structure offers a potential
468 explanation for how mutations in one domain of AdhE could impact the activity of the
469 other domain.

470 In the wild-type *T. saccharolyticum* AdhE (from strain LL1025), Asp-486 is the
471 equivalent of Asp-494 in the *C. thermocellum* AdhE and as mentioned above may
472 selectively mediate the binding of NADH over NADPH. The G544D mutation in LL1049
473 replaces a glycine residue by a charged aspartic acid across from Asp-486 and the 2'-
474 phosphate group of NADPH appears sandwiched between these two amino acid residues
475 (Figure 3E). There are several hydrogen bonds shared between this phosphate group and
476 the two aspartic acids that could help relieve their overall repulsion based on their
477 respective charges. In the case of the LL1040 variant, there is a large loop of 13 amino
478 acids introduced in the ADH domain, and given its flexibility and close proximity to the
479 NADH binding site in the linker sequence, could induce subtle changes to the binding site
480 that would result in the observed cofactor specificity change (Figure 3F).

Regarding cofactor change in the ALDH domain of the LL1040 and the LL1049 mutants, this domain either possess a mutation far from the NADH binding site (LL1040) or lacks such a mutation (LL1049). It is possible that spirosome formation (12, 28, 34, 49) not only influences enzyme activity, but also affects cofactor specificity: thus cofactor changes in the ADH domain may cause cofactor changes in the ALDH domain through the formation of such superstructures.

Journal of Bacteriology

ACKNOWLEDGMENTS

We thank the Mascoma Corporation for their gift of the *T. saccharolyticum* strains LL1076 (aka M3223), LL1040 (aka ALK2 or M0001), LL1049 (aka M1442 or MO1442), LL1193 (aka M2203) and LL1194 (aka M2202). We thank Sean Jean-Loup Murphy for his contribution in end-product analysis.

The genome sequencing work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The manuscript has been authored by Dartmouth College under sub-contract no. 4000115284 and contract no. DE-AC05-00OR22725 with the U.S. Department of Energy.

REFERENCES

1. **Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS**. 2002. Microbial cellulose utilization : fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. **66**:506–577.

2. **Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski J V, Conway JM, Adams MWW, Kelly RM**. 2014. Thermophilic lignocellulose deconstruction. FEMS Microbiol. Rev. **38**:393–448.

3. **Mai V, Lorenz WW, Wiegel J**. 1997. Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance. FEMS Microbiol. Lett. **148**:163–167.

4. **Shaw AJ, Covalla SF, Miller BB, Firliet BT, Hogsett DA, Herring CD**. 2012. Urease expression in a *Thermoanaerobacterium saccharolyticum* ethanologen allows high titer ethanol production. Metab. Eng. **14**:528–532.

5. **Shao XJ, Raman B, Zhu MJ, Mielenz JR, Brown SD, Guss AM, Lynd LR**. 2011. Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of *Clostridium thermocellum*. Appl. Microbiol. Biotechnol. **92**:641–652.

6. **Williams TI, Combs JC, Lynn BC, Strobel HJ**. 2007. Proteomic profile changes in membranes of ethanol-tolerant *Clostridium thermocellum*. Appl. Microbiol. Biotechnol. **74**:422–432.

7. **Sudha Rani K, Swamy M, Sunitha D, Haritha D, Seenayya G**. 1996. Improved ethanol tolerance and production in strains of *Clostridium thermocellum*. World J. Microbiol. Biotechnol. **12**:57–60.

8. **Sato K, Tomita M, Yonermura S, Goto S, Sekine K, Okuma E, Takagi Y, Hon-nami K, Saiki T**. 1993. Characterization of and ethanol hyper-production by *Clostridium thermocellum* I-1-B. Biosci. Biotechnol. Biochem. **57**:2116–2121.

9. **Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, Caiazza NC**. 2011. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. Appl. Environ. Microbiol. **77**:8288–8294.

10. **Shaw AJ, Podkaminer KK, Desai SG, Bardsley JS, Rogers SR, Thorne PG, Hogsett DA, Lynd LR**. 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. Proc. Natl. Acad. Sci. U. S. A. **105**:13769–13774.

11. **König S**. 1998. Subunit structure, function and organisation of pyruvate decarboxylases from various organisms. Biochim. Biophys. Acta **1385**:271–286.

12. **Extance J, Crennell SJ, Eley K, Cripps R, Hough DW, Danson MJ**. 2013. Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in *Geobacillus thermoglucosidasius*. Acta Crystallogr. Sect. D Biol. Crystallogr. **69**:2104–2115.

13. **Membrillo-Herna J, Echave P, Cabiscol E, Tamarit J, Ros J, Lin ECC**. 2000. Evolution of the *adhE* gene product of *Escherichia coli* from a functional reductase to a dehydrogenase. J. Biol. Chem. **275**:33869–33875.

14. **Yao S**. 2008. Ph.D. thesis. Technical University of Denmark, Denmark.

15. **Peng H, Wu G, Shao W**. 2008. The aldehyde/alcohol dehydrogenase (AdhE) in relation to the ethanol formation in *Thermoanaerobacter ethanolicus* JW200. Anaerobe **14**:125–127.

16. **Bhandiwad A, Shaw AJ, Guss A, Guseva A, Bahl H, Lynd LR**. 2014. Metabolic engineering of *Thermoanaerobacterium saccharolyticum* for *n*-butanol production. Metab. Eng. **21**:17–25.

17. **Pineda E, Encalada R, Olivos-García A, Néquiz M, Moreno-Sánchez R, Saavedra E**. 2013. The bifunctional aldehyde–alcohol dehydrogenase controls ethanol and acetate production in *Entamoeba histolytica* under aerobic conditions. FEBS Lett. **587**:178–184.

18. **Boxma B, Voncken F, Jannink S, van Alen T, Akhmanova A, van Weelden SWH, van Hellemond JJ, Ricard G, Huynen M, Tielens AGM, Hackstein JHP**. 2004. The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. Mol. Microbiol. **51**:1389–1399.

19. **Atteia A, van Lis R, Mendoza-Hernández G, Henze K, Martin W, Riveros-Rosas H, González-Halphen D**. 2003. Bifunctional aldehyde/alcohol dehydrogenase (ADHE) in chlorophyte algal mitochondria. Plant Mol. Biol. **53**:175–188.

20. **Lo J, Zheng T, Hon S, Olson DG, Lynd LR**. 2015. The bifunctional alcohol and aldehyde dehydrogenase gene, *adhE*, is necessary for ethanol production in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. J. Bacteriol. **JB02450-14**. (In press)

21. **Brown SD, Guss AM, Karpinets T V, Parks JM, Smolin N, Yang S, Land ML, Klingeman DM, Bhandiwad A, Rodriguez Jr. M, Raman B, Shao X, Mielenz JR, Smith JC, Keller M, Lynd LR**. 2011. Mutant alcohol dehydrogenase leads to

improved ethanol tolerance in *Clostridium thermocellum*. Proc. Natl. Acad. Sci. **108**:13752–13757.

22. **Biswas R, Zheng T, Olson DG, Lynd LR, Guss AM**. 2015. Elimination of hydrogenase active site assembly blocks H_2 production and increases ethanol yield in *Clostridium thermocellum*. Biotechnol. Biofuels **8**:1–8.

23. **Shaw AJ, Miller BB, Rogers SR, Kenealy WR, Meola A, Bhandiwad A, Sillers WR, Shikhare I, Hogsett DA, Herring CD**. 2015. Anaerobic detoxification of acetic acid in a thermophilic ethanologen. Biotechnol. Biofuels **8**. (In press)

24. **Olson DG, Lynd LR**. 2012. Transformation of *Clostridium thermocellum* by electroporation.Methods in Enzymology, 1st ed. Elsevier Inc.

25. **Shaw AJ, Hogsett DA, Lynd LR**. 2010. Natural competence in *Thermoanaerobacter* and *Thermoanaerobacterium* species. Appl. Environ. Microbiol. **76**:4713–4719.

26. **Zhou J, Olson DG, Argyros DA, Deng Y, van Gulik WM, van Dijken JP, Lynd LR**. 2013. Atypical glycolysis in *Clostridium thermocellum*. Appl. Environ. Microbiol. **79**:3000–3008.

27. **Zhang Y, Lynd LR**. 2003. Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation : development of an enzyme-linked immunosorbent assay-based method with application to *Clostridium thermocellum* batch cultures. Anal. Chem. **75**:3131–3139.

28. **Espinosa A, Yan L, Zhang Z, Foster L, Clark D, Li E, Stanley Jr. SL**. 2001. The bifunctional *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) protein is necessary for amebic growth and survival and requires an intact C-terminal domain for both alcohol dehydrogenase and acetaldehyde dehydrogenase activity. J. Biol. Chem. **276**:20136–20143.

29. **Schwarzenbacher R, von Delft F, Canaves JM, Brinen LS, Dai X, Deacon AM, Elsliger MA, Eshaghi S, Floyd R, Godzik A, Grittini C, Grzechnik SK, Guda C, Jaroszewski L, Karlak C, Klock HE, Koesema E, Kovarik JS, Kreusch A, Kuhn P, Lesley SA, Mcmullan D, Mcphillips TM, Miller MA, Miller MD, Morse A, Moy K, Ouyang J, Page R, Robb A, Rodrigues K, Selby TL, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, Wang X, West B, Wolf G, Hodgson KO, Wooley J, Wilson IA**. 2004. Crystal structure of an iron-containing 1,3-propanediol dehydrogenase (TM0920) from *Thermotoga maritima* at 1.3 Å resolution. PROTEINS Struct. Funct. Bioinforma. **54**:174–177.

30. **Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML**. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. **79**:926–935.

31. **Jo S, Kim T, Iyer VG, Im W**. 2008. Software news and updates CHARMM-GUI : A web-based graphical user interface for CHARMM. J. Comput. Chem. **29**:1859–1865.

32. **Koukos PI, Glykos NM**. 2013. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. J. Comput. Chem. **34**:2310–2312.

33. **Arnau J, Jørgensen F, Madsen SM, Vrang A, Israelsen H**. 1998. Cloning of the *Lactococcus lactis adhE* gene, encoding a multifunctional alcohol dehydrogenase, by complementation of a fermentative mutant of *Escherichia coli*. J. Bacteriol. **180**:3049–3055.

34. **Bruchhaus I, Tannich E**. 1994. Purification and molecular characterization of the NAD(+)-dependent acetaldehyde/alcohol dehydrogenase from *Entamoeba histolytica*. Biochem. J. **303**:743–748.

35. **Dailly Y, Bunch P, Clark D**. 2000. Comparison of the fermentative alcohol dehydrogenases of *Salmonella typhimurium* and *Escherichia coli*. Microbios.

36. **Pei J, Zhou Q, Jiang Y, Le Y, Li H, Shao W, Wiegel J**. 2010. *Thermoanaerobacter* spp. control ethanol pathway via transcriptional regulation and versatility of key enzymes. Metab. Eng. **12**:420–428.

37. **Yao S, Mikkelsen MJ**. 2010. Identification and overexpression of a bifunctional aldehyde/alcohol dehydrogenase responsible for ethanol production in *Thermoanaerobacter mathranii*. J. Mol. Microbiol. Biotechnol. **19**:123–133.

38. **Koo OK, Jeong DW, Lee JM, Kim MJ, Lee JH, Chang HC, Kim JH, Lee HJ**. 2005. Cloning and characterization of the bifunctional alcohol/acetaldehyde dehydrogenase gene (*adhE*) in *Leuconostoc mesenteroides* isolated from kimchi. Biotechnol. Lett. **27**:505–510.

39. **Chen M, Li E, Stanley SL**. 2004. Structural analysis of the acetaldehyde dehydrogenase activity of *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2), a member of the ADHE enzyme family. Mol. Biochem. Parasitol. **137**:201–205.

40. **Montella C, Bellsolell L, Perez-Luque R, Badia J, Baldoma L, Coll M, Aguilar J**. 2005. Crystal structure of an iron-dependent group III dehydrogenase that interconverts L-lactaldehyde and L-1,2-propanediol in *Escherichia coli*. J. Bacteriol. **187**:4957–4966.

41. **Huang H, Wang S, Moll J, Thauer RK**. 2012. Electron bifurcation involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H_2 plus CO_2. J. Bacteriol. **194**:3689–3699.

42. **Currie DH, Guss AM, Herring CD, Giannone RJ, Johnson CM, Lankford PK, Brown SD, Hettich RL, Lynd LR**. 2014. Profile of secreted hydrolases, associated proteins, and SlpA in *Thermoanaerobacterium saccharolyticum* during the degradation of hemicellulose. Appl. Environ. Microbiol. **80**:5001–5011.

43. **Verbeke TJ, Zhang X, Henrissat B, Spicer V, Rydzak T, Krokhin O V., Fristensky B, Levin DB, Sparling R**. 2013. Genomic evaluation of *Thermoanaerobacter* spp. for the construction of designer co-cultures to improve lignocellulosic biofuel production. PLoS One **8**:1–18.

44. **Radianingtyas H, Wright PC**. 2003. Alcohol dehydrogenases from thermophilic and hyperthermophilic archaea and bacteria. FEMS Microbiol. Rev. **27**:593–616.

45. **Burdette D, Zeikus JG**. 1994. Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase (2^0 Adh) as a bifunctional alcohol dehydrogenase-acetyl-CoA reductive thioesterase. Biochem. J. **302**:163–170.

46. **Lamed R, Zeikus JG**. 1980. Ethanol production by thermophilic bacteria: relationship between fermentation product yields of and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. J. Bacteriol. **144**:569–578.

47. **Neale AD, Scopes RK, Kelly JM, Wettenhall REH**. 1986. The two alcohol dehydrogenases of *Zymomonas mobilis* purification by differential dye ligand chromatography, molecular characterisation and physiological roles. Eur. J. Biochem. **154**:119–124.

48. **Chen Z, Lee WR, Chang SH**. 1991. Role of aspartic acid 38 in the cofactor specificity of *Drosophila* alcohol dehydrogenase. Eur. J. Biochem. **202**:263–267.

49. **Kessler D, Herth W, Knappe J**. 1992. Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. J. Biol. Chem. **267**:18073–18073.

50. **Extance JP**. 2012. Ph.D thesis. University of Bath, United Kingdom.

51. **Lee JM, Venditti R a., Jameel H, Kenealy WR**. 2011. Detoxification of woody hydrolyzates with activated carbon for bioconversion to ethanol by the thermophilic anaerobic bacterium *Thermoanaerobacterium saccharolyticum*. Biomass and Bioenergy **35**:626–636.

678

679

Figure legends

FIGURE 1 Primary structure of AdhE from wild-type *C. thermocellum* and *T. saccharolyticum*. (A) The ALDH domain is shown in light grey (1-423 for *C. thermocellum* and 1-420 for *T. saccharolyticum*), ADH domain in white (463-873 for *C. thermocellum* and 460-860 for *T. saccharolyticum*) and linker sequence in dark grey (424-462 for *C. thermocellum* and 421-459 for *T. saccharolyticum*). The NADH binding sites are shown as red boxes ("NADH binding site 1" is 200-221 for *C. thermocellum* and 199-220 for *T. saccharolyticum*; "NADH binding site 2" is 551-553 for *C. thermocellum* and 543-545 for *T. saccharolyticum*) and Fe^{2+} binding site as green boxes. Mutated residues discussed in this study are annotated at the appropriate positions: D494G in LL350; P704L and H734R in LL346; V52A and K451N and a 13 amino acid insertion in LL1040; G544D in LL1049. All elements are drawn to scale. (B) and (C) show the sequence conservation of the NADH binding motifs (highlighted in yellow in the "Consensus" sequence) of AdhE from *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter mathranii*, *T. saccharolyticum*, *Entamoeba histolytica*, *E. coli*, *C. thermocellum*, *Leuconostoc mesenteroids*, *Lactococcus lactis*, *Oenococcus oeni* and *Streptococcus equinus*. Residues highlighted in red are of the highest conservation, and blue are of the least. The numbering of amino acids is based on the AdhE sequence from *C. thermocellum*.

FIGURE 2 Homology modeling and docking analysis of the phosphate in NADPH interacting with the *C. thermocellum* wild-type AdhE (A) and its D494G mutant (B). The dotted lines represent hydrogen bonds. The red residue is D494, yellow residue is G494 and blue residues are N495 and F496.

Journal of Bacteriology

FIGURE 3 Average structure of the ADH domains of AdhE from *C. thermocellum* wild-type LL1004 (A), the ethanol-tolerant LL346 (B) strain, moderate-ethanol-producer LL350 (C), *T. saccharolyticum* wild-type LL1025 (D), high-ethanol-producer LL1049 (E), and high-ethanol producer LL1040 (F). The amino acids of interest are shown in blue and red, NADPH cofactor is shown color-coded by elements with its 2'-phosphate group highlighted (green open circle), and the iron ions (green) are also shown. Additionally, the 39 base pair insertion in the high-ethanol producer LL1040 is shown in orange (F). The yellow filled circle designates the binding pocket where the additional 2'-phosphate of NADPH is commonly found in NADPH dependent AdhEs. The location of D494 and D486 is usually the recognition site for NADH and doesn't allow the 2'-phosphate group of NADPH (green circle) to access the preferred binding pocket (yellow circle). When the green open circle and the yellow filled circle overlap, that indicates that the NADPH molecule is able to access its preferred binding pocket. This is present in panels C, E and F, but not in the other panels. These panels (C, E and F) correspond to enzymes that can use NADPH as a cofactor. Note that NADPH was used for modeling purposes and does not reflect the actual cofactor specificity of the enzymes but was rather used to explain the level of affinity observed of the enzymes for NADPH.

721

722 **TABLE 1** Strains used in this study

Organism	Strain name	Description	Accession No. [a]	Source or reference
C. thermocellum	LL1004	Wild-type *C. thermocellum* strain DSM 1313, low-ethanol-producer [b]	CP002416	DSMZ [c]
C. thermocellum	LL346	A.k.a "*adhE** (EA)" from Brown et al. (20). Evolved *C. thermocellum* strain, tolerant to 40g/L ethanol, have mutations P704L and H734R in AdhE, low-ethanol-producer.	SRX030164.1	Brown et al. (21)
C. thermocellum	LL350	A.k.a "Δ*hydG*" from Biswas et al. (22). *C. thermocellum* Δ*hpt* Δ*hydG* strain with mutation D494G in AdhE, moderate-ethanol-producer [d].	N/A [f]	Biswas et al. (22)
C. thermocellum	LL1111	*C. thermocellum* Δ*hpt* Δ*adhE*, no ethanol production.	SRX744221	Lo et al. (20)
C. thermocellum	LL1160	LL1111 *C. thermocellum* strain with *adhE* reintroduced to the original *adhE* locus, low-ethanol-producer.	N/A	This study
C. thermocellum	LL1161	LL1160 with mutation D494G in AdhE, moderate-ethanol-producer.	N/A	This study
T. saccharolyticum	LL1025	Wild-type *T. saccharolyticum* strain JW/SL-YS485, moderate-ethanol-producer.	CP003184	Mai et al. (3)
T. saccharolyticum	LL1049	A.k.a M1442 or MO1442 from Lee et al. and Shaw et al. (23, 51). Evolved *T. saccharolyticum* Δ(*pta-ack*) Δ*ldh* Δ*or796 ure metE* Δ*eps* strain with mutation G544D in AdhE, high-ethanol-producer [e].	SRA233073	Shaw et al. (23)
T. saccharolyticum	LL1040	A.k.a strain "ALK2" from Shaw et al. (10). Evolved *T. saccharolyticum* Δldh::erm	SRA233066	Shaw et al. (10)

		Δ(*pta-ack*)::*kan* strain with mutations V52A, K451N and a 13-amino-acid insertion in AdhE, high-ethanol-producer.		
T. saccharolyticum	LL1076	Δ*adhE*::(*pta-ack kan*), no ethanol production.	SRX744220	Mascoma Corp.
T. saccharolyticum	LL1193	*adhE*::*kan*, differs from wild-type only with *kan* marker downstream of AdhE, moderate-ethanol-producer. A.k.a M2203.	N/A	Shaw et al. (23)
T. saccharolyticum	LL1194	LL1193 with the G544D mutation in AdhE. A.k.a M2202.	N/A	Shaw et al. (23)

[a] Accession numbers starting with CP refer to finished genome sequences in Genbank, accession numbers starting with SR refer to raw sequencing data from JGI

[b] Produces ethanol at 0-40% theoretical yield.

[c] German Collection of Microorganisms and Cell Cultures. Leibniz-Institute, Germany.

[d] Produces ethanol at 40-80% theoretical yield.

[e] Produces ethanol at 80-90% theoretical yield.

[f] Not applicable.

732

733 **TABLE 2** ALDH and ADH activities in *C. thermocellum* and *T. saccharolyticum* cell

734 extracts

Strain characteristics			ALDH activity [b]		ADH activity	
Strain name	Description	Ethanol yield [a]	NADH	NADPH	NADH	NADPH
C. thermocellum LL1004	Wild-type	0.16	**2.20** ± 0.05 [c]	**0.21** ± 0.03	**6.73** ± 0.72	**0.04** ± 0.00
C. thermocellum LL346	Ethanol-tolerant	0.11	**0.27** ± 0.13	**0.05** ± 0.03	**0.66** ± 0.20	**0.38** ± 0.02
C. thermocellum LL350	Moderate-ethanol-producer	0.22	**2.00** ± 0.49	**0.05** ± 0.01	**6.71** ± 0.93	**5.90** ± 0.27
C. thermocellum LL1111	*adhE* deletion	0.01	**0.05** ± 0.00	**0.13** ± 0.02	**0.10** ± 0.09	**0.22** ± 0.18
T. saccharolyticum LL1025	Wild-type	0.26	**0.41** ± 0.13	**0.05** ± 0.04	**7.06** ± 0.50	**0.95** ± 0.58
T. saccharolyticum LL1049	High-ethanol-producer	0.43	**0.09** ± 0.02	**0.50** ± 0.05	**0.18** ± 0.14	**1.10** ± 0.42
T. saccharolyticum LL1040	High-ethanol-producer	0.45	**0.08** ± 0.02	**0.30** ± 0.02	**0.08** ± 0.05	**1.55** ± 0.72
T. saccharolyticum LL1076	*adhE* deletion	0.01	**0.00** ± 0.18 [d]	**0.018** ± 0.09	**0.04** ± 0.18	**1.78** ± 0.56

735 [a] Ethanol yield is in g/g cellobiose, produced from 5 g/L cellobiose.

736 [b] Activity units are in U/mg protein (see Material and Methods).

737 [c] Error represents one standard deviation, n = 4 to 8.

738 [d] When activity is very low and background activity is higher than measured activity,

739 this results in a negative value (see Material and Methods) and is shown here as

740 zero activity.

741

742

743 **TABLE 3** Cofactor specificity of purified AdhE

Source of AdhE		ALDH activity [a]		ADH activity	
Strain name	Description	NADH	NADPH	NADH	NADPH
C. thermocellum LL1004	Wild-type	**18.02** ± 2.45 [b]	**2.03** ± 0.46	**42.23** ± 3.13	**1.96** ± 1.25
C. thermocellum LL346	Ethanol-tolerant	**13.50** ± 3.54	**0.43** ± 0.42	**4.02** ± 1.48	**0.03** ± 0.34
C. thermocellum LL350	Moderate-ethanol-producer	**31.50** ± 5.80	**5.76** ± 0.16	**42.67** ±7.46	**42.30** ± 4.28
T. saccharolyticum LL1025	Wild-type	**10.63** ± 1.15	**4.46** ± 1.63	**17.43** ± 2.45	**2.43** ± 2.43
T. saccharolyticum LL1049	High-ethanol-producer	**1.63** ± 0.30	**11.13** ± 1.35	**0.66** ± 1.97	**12.70** ± 2.18
T. saccharolyticum LL1040	High-ethanol-producer	**0.07** ± 0.55	**5.03** ± 2.05	**0.01** ± 1.34	**12.96** ± 4.54

744

745 [a] Activity units are in U/mg protein (see Material and Methods)

746 [b] Error represents one standard deviation, n = 4 to 8

747

748

749

750

751

752

753

754

755

756

757

758

759 **TABLE 4** Apparent K_m and k_{cat} values of purified *C. thermocellum* wild-type and LL350
760 AdhE

761

Source of AdhE [a]	Strain characteristics	Reaction	Substrate	K_m (mM)	k_{cat} (S^{-1})	k_{cat}/K_m ($S^{-1}M^{-1}$)
C. thermocellum LL1004	Wild-type	ALDH	Acetyl-CoA	**0.0084** ± 0.0002 [e]	**280** ± 2	**3.3E+07**
			NADH	**0.052** ± 0.003	**150** ± 3	**3.0E+06**
		ADH	Acetaldehyde	**1.6** ± 0.05	**240** ± 4	**1.5E+05**
			NADH	**0.088** ± 0.006	**240** ± 6	**2.7E+06**
C. thermocellum LL350	Moderate-ethanol-producer	ALDH	Acetyl-CoA	**0.0042** ± 0.0004*[d]	**230** ± 2**	**5.4E+07**
			NADH	**0.026** ± 0.0002*	**62** ± 0.1**	**2.4E+06***
		ADH	Acetaldehyde [b]	**1.4** ± 0.05	**220** ± 2*	**1.5E+05**
			Acetaldehyde [c]	**1.4** ± 0.04	**210** ± 3*	**1.5E+05**
			NADH	**0.17** ± 0.01*	**220** ± 5	**1.3E+06***
			NADPH	**0.075** ± 0.01	**280** ± 7	**3.7E+06**

762 [a] AdhE was expressed in *E. coli* and affinity-purified

763 [b] Measurements were done with NADH as a cofactor

764 [c] Measurements were done with NADPH as a cofactor

765 [d] Compared to respective values in wild-type AdhE, two-tailed p-value <0.05 designated
766 as *, <0.01 as **.

767 [e] Error represents one standard error, at a confidence level of 0.95, n=10 to 16 depending
768 on the sample.

769