Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*

Steven D. Brown[a,b,1], Adam M. Guss[a,b,c], Tatiana V. Karpinets[a,b], Jerry M. Parks[a,b], Nikolai Smolin[a], Shihui Yang[a,b], Miriam L. Land[a,b], Dawn M. Klingeman[a,b], Ashwini Bhandiwad[b,c], Miguel Rodriguez, Jr.[a,b], Babu Raman[a,b], Xiongjun Shao[b,c], Jonathan R. Mielenz[a,b], Jeremy C. Smith[a,b,d], Martin Keller[a], and Lee R. Lynd[b,c]

[a]Biosciences Division and [b]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [c]Thayer School of Engineering, Dartmouth College, Hanover, NH 03755; and [d]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996

Edited* by Arnold L. Demain, Drew University, Madison, NJ, and approved July 14, 2011 (received for review February 24, 2011)

Clostridium thermocellum **is a thermophilic, obligately anaerobic, Gram-positive bacterium that is a candidate microorganism for converting cellulosic biomass into ethanol through consolidated bioprocessing. Ethanol intolerance is an important metric in terms of process economics, and tolerance has often been described as a complex and likely multigenic trait for which complex gene interactions come into play. Here, we resequence the genome of an ethanol-tolerant mutant, show that the tolerant phenotype is primarily due to a mutated bifunctional acetaldehyde-CoA/alcohol dehydrogenase gene (*adhE*), hypothesize based on structural analysis that cofactor specificity may be affected, and confirm this hypothesis using enzyme assays. Biochemical assays confirm a complete loss of NADH-dependent activity with concomitant acquisition of NADPH-dependent activity, which likely affects electron flow in the mutant. The simplicity of the genetic basis for the ethanol-tolerant phenotype observed here informs rational engineering of mutant microbial strains for cellulosic ethanol production.**

bioenergy | genomics | inhibitor | resequencing | 454

Fuels from cellulosic biomass are among the leading options to meet sustainability and energy security challenges associated with fossil fuels, and conversion processes featuring biological fermentation are among the leading options for producing cellulosic biofuels. Among fermentation-based conversion processes, use of cellulose-fermenting microorganisms without added enzymes—consolidated bioprocessing—has strong potential (1), and a variety of microorganisms are under development (2).

Clostridium thermocellum is a thermophilic bacterium that can rapidly solubilize biomass and use cellulose as a carbon and energy source. Wild-type (WT) strains produce ethanol as well as organic acids, but growth is inhibited by relatively low ethanol concentrations (<10 g/L; refs. 3 and 4). Cultures of *C. thermocellum* have been adapted to tolerate ethanol concentrations as high as 80 g/L (5), and although greater ethanol production has been reported for tolerant strains, the highest concentration of ethanol production reported for this organism is <30 g/L (6).

Ethanol tolerance and inhibition are typically complex, incompletely understood traits, although some studies point to compromised membrane integrity as a key factor (7, 8). In the case of *C. thermocellum*, Williams et al. (5) used MALDI-TOF analyses to obtain membrane proteomic profiles for WT and ethanol-adapted (EA) strains, which suggested that membrane-associated proteins were less abundant in EA strains or had issues related to incorporation into the cell membrane. Recently, Timmons et al. (9) showed that *C. thermocellum* EA had more fatty acids with chain lengths of >16:0 and significantly more 16:0 plasmalogens compared with WT and proposed that ethanol tolerance was due to fatty acid alterations that increased membrane rigidity to counteract the fluidizing effect of ethanol. However, the genetic basis for the enhanced ethanol tolerance for EA strains of *C. thermocellum* has not been determined. We hypothesized that the genome of *C. thermocellum* EA accumulated one or more mutations that permitted its growth in higher concentrations of ethanol compared with the WT strain.

Results and Discussion

Resequencing Analyses. To determine the genetic basis of the *C. thermocellum* ethanol-tolerant phenotype, the genomes of EA and WT strains were resequenced by using a pyrosequencing approach (see Table 1 and *Materials and Methods* for details). A total of 72 and 500 454-pyrosequencing "high-confidence differences" ("HCDiffs"), as defined by the GSMapper software (454 Life Sciences), were detected for the WT and EA mutant (5), respectively (*SI Appendix*, Tables S1 and S2). The *C. thermocellum* WT and EA genomes were also resequenced by using a microarray-based comparative genome sequencing (CGS) approach, which revealed 410 putative differences (for details, see *Materials and Methods* and *SI Appendix*, Table S3). A summary of the combined pyrosequencing and CGS analyses is presented in Table 2 and *SI Appendix*, Fig. S1.

The number of putative differences identified by either pyrosequencing or CGS alone decreased dramatically as more stringent filtering criteria were applied (*SI Appendix*, Fig. S2). In contrast, most differences identified by both independent technologies were retained as confidence stringency was increased, and 230 genetic differences were validated by both resequencing platforms after filtering (*SI Appendix*, Fig. S2*B*). Repetitive DNA elements were identified in several lower-confidence differences and were filtered out (*SI Appendix*, Fig. S3), a finding that may point to areas for future bioinformatics improvements. Analysis of pyrosequencing data indicated that two large regions (~9 and 21.5 kb) encoding hypothetical or phage-related proteins were deleted in EA, and CGS data supported that these regions were deleted (*SI Appendix*, Fig. S1 *vii* and *viii*).

The ability to characterize strains through multiple high-throughput resequencing approaches not only provides rapid insights into adaptive evolution of strains with important biological or industrial traits, but also helps overcome the limitations associated with respective resequencing technologies. This study and previous pyrosequencing studies (10, 11) suggest that putative high-confidence differences with variation values of <80% should

Author contributions: S.D.B., A.M.G., J.M.P., S.Y., J.C.S., and L.R.L. designed research; S.D.B., T.V.K., N.S., S.Y., D.M.K., A.B., M.R., B.R., and X.S. performed research; M.L.L. contributed new reagents/analytic tools; S.D.B., A.M.G., T.V.K., J.M.P., N.S., S.Y., J.R.M., J.C.S., M.K., and L.R.L. analyzed data; and S.D.B., A.M.G., T.V.K., J.M.P., J.C.S., and L.R.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession nos. SRX030163 and SRX030164).

[1]To whom correspondence should be addressed. E-mail: brownsd@ornl.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1102444108/-/DCSupplemental.

Table 1. Summary statistics for genome resequencing via GS FLX system

Genome	No. of reads	No. of bases	No. of contigs (>500 bp)	No. of assembled bases in contigs (>500 bp)	Genome coverage
C. the[†]	687,158	139,392,584	126	3,711,170	~38
C. the EtOH[‡]	451,118	105,817,828	122	3,677,593	~29

[†]*C. thermocellum* WT strain ATCC 27405.
[‡]*C. thermocellum* ethanol-tolerant mutant strain derived from ATCC 27405.

be interpreted with caution. We suggest that, although it is possible to rank putative differences based on possible technical limitations, lower-ranked differences cannot be disregarded entirely.

C. thermocellum EA SNPs were detected in genes that encode proteins previously reported as differentially expressed compared with the WT strain (*SI Appendix*, Tables S2 and S3 and ref. 5). These genes included Cthe_0578 (glycoside hydrolase, family 9), Cthe_2263 (H^+-transporting two-sector ATPase, C subunit), Cthe_1939 (magnesium transporter), Cthe_0858 (protein of unknown function DUF1432), Cthe_1385 (protein translocase subunit SecA), Cthe_1285 (metal-dependent phosphohydrolase), Cthe_3171 (S-layer domain-like protein), Cthe_2341 (glycosyl transferase, family 2), Cthe_0912 (glycoside hydrolase, family 10), Cthe_1284 (glycogen/starch synthases, ADP-glucose type), and Cthe_2664 (2-octaprenylphenol hydroxylase). In addition, mutations were detected at many other loci that did not have corresponding differences at the protein level and included important genes such as Cthe_0423 [encoding AdhE, bifunctional acetaldehyde-CoA/alcohol dehydrogenase (ADH)]. An analysis of distribution of the identified mutations across the *C. thermocellum* genome in terms of their type and the predicted operons and metabolic pathways was used to gain further insights into the 500 pyrosequencing differences.

Nonrandom Distribution of Mutations Across the Genome and Their Link to EA Phenotypes. The number of insertions, deletions, SNPs (synonymous and nonsynonymous), and multiple substitutions in coding and noncoding parts of the genome is presented in *SI Appendix*, Fig. S4. Single nucleotide substitutions are the dominant type of mutation that occurs in EA, and nonsynonymous substitutions were approximately twice as abundant as synonymous substitutions. Multiple substitutions, however, mainly targeted coding sequences of the genome, and insertions and deletions were overrepresented in noncoding sequences.

Further statistical analysis of the distribution of mutations revealed an increased number of mutations in EA at several locations, when considering the frequencies of different types of mutations, their total number, and the presence of neighboring genes or genes comprising an operon (*SI Appendix*, Fig. S5). A manual curation of the distribution pattern of the mutations identified 16 putative hot spots for mutation (Table 3). Genes linked to these hot spots included 72 out of 500 putative mutations identified (14%). Many EA membrane-associated proteins for carbohydrate transport and metabolism, including enzymatic and structural components of the organism's cellulosome, have been reported to be less abundant compared with the WT strain (5). It is therefore not surprising that 7 of the 16 mutation hot spots (hot spot IDs 1, 3, 6, 7, 9, 13, and 16) were in genes related to cellulose degradation, consistent with a previously observed poor growth phenotype for EA on crystalline cellulose (5). Poor growth of EA was confirmed, and acetate and ethanol were produced (*SI Appendix*, Fig. S6). Most of the hypothetical gene mutations found in hot spots in this analysis were adjacent to phage/transposase genes (Table 3).

Importantly, a putative operon (DOOR Database operon 295062; ref. 12) containing 10 adjacent genes (Cthe_0422 to Cthe_0431) likely involved in ethanol production was significantly affected by different mutations (hot spot ID 9). The bifunctional acetaldehyde-CoA/ADH (AdhE; Cthe_0423) is an important enzyme for ethanol production and had two mutations, indicating an increased probability of importance for altered metabolism of EA. The EA AdhE contains two nonsynonymous mutations resulting in predicted amino acid changes. We pursued the hypothesis that this mutation may affect ethanol metabolism and therefore ethanol tolerance.

Mutant *adhE* Allele Alone Confers Increased *C. thermocellum* Ethanol Tolerance. To determine whether the mutant AdhE plays a role in EA's enhanced ethanol tolerance, the WT and mutant alleles of the *adhE* gene were cloned into a replicating plasmid and transformed into *C. thermocellum* DSM 1313 WT strain (i.e., *adhE*$^+$). The resulting strains were then assayed for their ability to grow in medium with cellobiose as the sole carbon source and elevated levels of ethanol (Fig. 1). The strain carrying the mutant allele showed marked improvement in growth in the presence of 20 and 24 g/L added ethanol, and it was the only strain able to grow in the presence of 40 g/L added ethanol. Although 500 putative mutations were identified in EA (*SI Appendix*, Table S2), a single mutated gene from this culture was able to confer most of the ethanol-tolerant phenotype of EA (Fig. 1). However, the possibility that other mutations could also confer the EA phenotype cannot be ruled out. In this study, increased ethanol tolerance was not linked to higher ethanol yields (*SI Appendix*, Fig. S6), which is in keeping with a recent study that examined *Escherichia coli* isobutanol tolerance and productivity (13).

Plasmid DNA was unable to be isolated from cultures containing the mutant *adhE* allele grown with 40 g/L added ethanol during routine strain verification. However, when the chromosomal *adhE* was sequenced from this strain, a gene conversion event was discovered in which the mutant allele replaced the WT allele on the chromosome (*SI Appendix*, Fig. S7). This strain, which provides a clean genetic background for studying the effect of the mutant *adhE*, is herein called *C. thermocellum adhE**(EA).

Microbial ethanol tolerance has generally been thought to be a complex and likely multigenic trait (5, 7, 9). There have been suggestions that no single gene can endow microbes with tolerance to ethanol and other toxic compounds (14), and until recently little progress had been made in identification of key genetic changes that confer enhanced ethanol tolerance (15). Global transcription machinery engineering has been used to improve tolerance to both glucose and ethanol and to increase productivity in *Saccharomyces cerevisiae* by altering expression of

Table 2. CGS differences supported by pyrosequencing

CGS confidence values	No. of CGS differences	Subset supported by 454[†]
High	234	230 (98%)
Medium	85	3 (4%)
Low	91	9 (10%)

[†]Variation values reported by the GSMapper software are shown in parentheses. A higher score indicates more reads were in agreement, with the maximum value being 100.

Table 3. Mutational hot spots identified in the EA mutant genome

Hot spot ID	Genes with mutations	No. of mutations	Description[†]	Products
1	Cthe_2825 Cthe_2826	11	Two genes with multiple mutations in one TU	Hypothetical proteins
2	Cthe1806 Cthe1807	6	Two adjacent TU, two genes with INS	Cellulosome enzyme, dockerin type I
3	Cthe2018	6	One gene with multiple mutations	Hypothetical protein
4	Cthe0056 Cthe0059	5	Two adjacent TU and an intergenic regions	Ig-like, group 2 and cellulose-binding
5	the1235 Cthe1237	5	Two adjacent TU with INS	Cellulose 1,4-beta-cellobiosidase and leucyl-tRNA synthetase
6	Cthe1890	5	One gene with multiple mutations	Cellulosome enzyme, dockerin type I
7	Cthe2854 Cthe2855	5	Two genes with multiple mutations in one TU	Hypothetical proteins
8	Cthe3077 Cthe3078	4	Two genes with multiple mutations in one TU	Cellulosome anchoring protein, cohesin region
9	Cthe0422 Cthe0423	4	Two genes with two NON in each in one TU	Redox-sensing transcriptional repressor Rex and bifunctional acetaldehyde-CoA/ADH
10	Cthe2427	3	One gene with multiple mutations	Hypothetical protein
11	Cthe2559 Cthe2560	3	Two genes with STOP in one TU	DegT/DnrJ/EryC1/StrS aminotransferase and dTDP-4-dehydrorhamnose 3,5-epimerase
12	Cthe2611 Cthe2612	3	Two genes with two NON in one TU	Fibronectin, type III
13	Cthe1114 Cthe1115	3	Three genes in one TU with DEL	Tn7-like transposition protein C and HMG-I and HMG-Y, DNA-binding
14	Cthe0996 Cthe0997	3	Three total mutations and two NON including INS in two genes in one TU	DNA polymerase III PolC and 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
15	Cthe0659 Cthe0660 Cthe0661	3	Three adjacent genes and two NON i ncluding DEL	Hypothetical protein and glycoside hydrolase family protein
16	Cthe1885 Cthe1886	3	Two adjacent genes with INS	Phage integrase-like SAM-like and integrase catalytic subunit

[†]TU, transcription unit. Types of mutations are referred to as insertions (INS), deletions (DEL), and nonsynonymous substitutions (NON).

many genes simultaneously through a single genetic modification (16). In contrast, Hong et al. (15) used an inverse metabolic engineering approach to demonstrate that overexpression of endogenous *S. cerevisiae* genes (INO1, DOG1, HAL1, or a truncated MSN2) individually can confer improved alcohol tolerance, higher titers, higher volumetric productivities, and increased

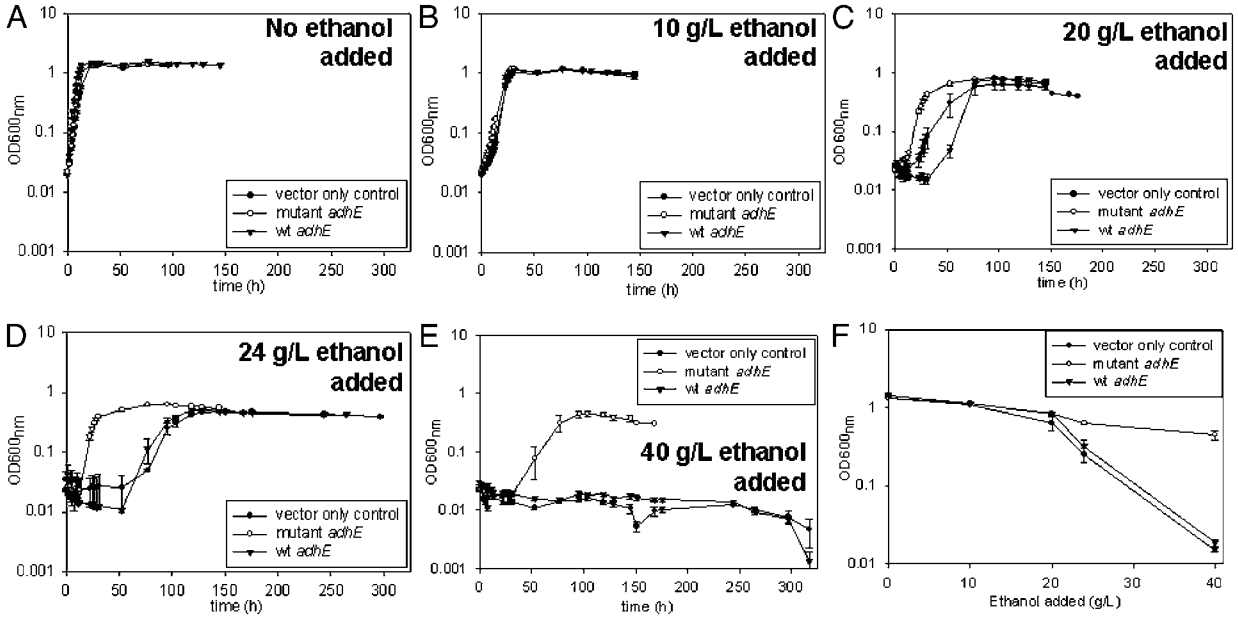

Fig. 1. Mutant *C. thermocellum* ADHs confers enhanced ethanol tolerance. (*A–E*) Growth of *C. thermocellum* DSM 1313 strains with different plasmids that provided a vector-only control, an additional copy of the WT version of *adhE*, and the mutant *adhE* gene was monitored by measuring culture turbidity ($\log_{10}$ OD_{600nm}). Strains were grown at 55 °C in increasing amounts of added ethanol. (*F*) The final culture turbidities after 96 h of growth are presented as a function of ethanol added to the culture from the beginning the experiment. The graphs show the mean and the SE (bars) for three independent dose–response curves, and this experiment was repeated twice.

specific growth rate. However, none of the *S. cerevisiae* genes that confer improvements in alcohol tolerance are similar to the *C. thermocellum adhE* gene, and their products perform quite different functions. Therefore, we investigated the mechanism of alcohol tolerance conferred by this *C. thermocellum* gene.

Protein Structural Modeling. Structural models of WT and double-mutant *C. thermocellum* AdhE ADH domain were constructed by homology using protein sequence data and the X-ray structures of the *Thermotoga maritima* Fe-containing ADH [1.3-Å resolution; Protein Data Bank (PDB) ID 1O2D] and the *Klebsiella pneumoniae* 1,3-propanediol dehydrogenase (2.7-Å resolution; PDB ID 3BFJ) proteins. These structures were chosen as templates based on their similarity to AdhE and inclusion of NAD or NADP and Fe^{2+} in their structures. The model structures inform hypotheses concerning the possible effects of the mutations on the enzyme function. The residues in the active site, their 3D orientations, and the iron coordination sphere were strictly conserved. The two mutations occurred in different regions of the protein. The Pro-704–Leu mutation was located at the external terminus of a surface α-helix, far from the active site (~23 Å) and cofactor binding sites (Fig. 2). Hence, this mutation is less likely to affect cofactor binding or catalysis directly. The His-734–Arg mutation, conversely, was close to the active site iron (~9 Å) and cofactor binding sites (Fig. 2) and likely involves a change in net charge that might alter the relative cofactor binding specificity. The presence of an arginine residue facing the adenine plane has been suggested to interact with the phosphomonoester group of NADP and may be a requirement for NADP recognition, although not a mechanism for NAD and NADP discrimination (17). Indeed, arginine is substituted for histidine and serine resides on occasions in interactions mediated by a water molecule (17), and a moderately polar hydrogen bond change was sufficient to ensure mutated *Bacillus stearothermophilus* glyceraldehyde-3-phosphate dehydrogenase had altered cofactor specificity (18). It is also possible that a functional effect arises in which mutation of histidine to arginine could block putative proton shuttle activity or that, as the His-734–Arg site lies at a possible entrance channel, substrate access and/or product release may be affected.

Localization of the His-734–Arg mutation near the NADH cofactor binding site suggested possible alteration of cofactor binding specificity as a mechanism for increased ethanol tolerance of *C. thermocellum adhE** and strain EA. Similar levels of NADH-dependent and NADPH-dependent ADH activities have been reported for *C. thermocellum* strain ATCC 27405 (19), which is quite different from an earlier report that *C. thermocellum* strains LQRI and AS39 had ≥200-fold higher NADH-dependent ADH activities compared with NADPH-dependent ADH activities (20).

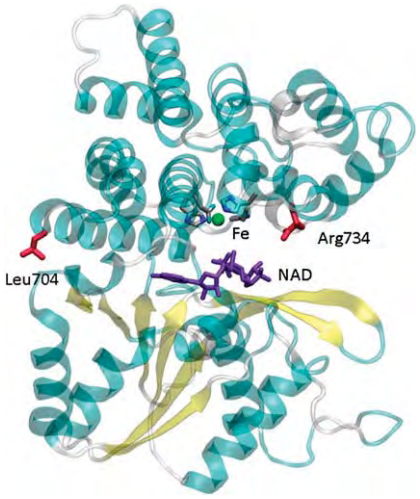

Fig. 2. Homology/MD model of *C. thermocellum* AdhE double mutant (P704L, H734R). Mutation sites (Leu-704 and Arg-734) and the NAD cofactor and Fe are labeled. The configuration was taken from the end of a 10-ns MD trajectory. Figure was rendered by using VMD (57).

ADH Cofactor Specificity Has Shifted from NADH to NADPH. WT, EA mutant, and *adhE**(EA) in vitro ADH activity was tested in crude cell extracts (Table 4). In this study, *C. thermocellum* WT ADH activity is predominantly NADH-dependent, with only very low levels of NADPH-dependent activity, whereas the specificity is different for the mutant allele, with NADH-dependent activity abolished and a concomitant increase in NADPH-dependent activity. These biochemical data, in combination with the loss of the WT allele in strain *adhE**(EA), raise the possibility that not only is the mutant allele beneficial, but perhaps the WT allele is also harmful at high ethanol concentrations.

NAD or NADP differ with respect to presence or absence of a phosphate group esterified at the 2′ position of the adenosine ribose and are similar at the level of structure. Rosell et al. (21) have shown complete reversal of ADH cofactor specificity in crystallography studies. However, NAD is typically used in oxidative, ATP-generating degradation reactions, and NADP usually acts as a reductant in reductive biosynthetic reactions. An ethanol cycle with two ADH isozymes catalyzing opposite reactions (i.e., ethanol oxidation or ethanol synthesis) has been proposed for *Thermoanaerobacter pseudoethanolicus* (formerly *Clostridium thermohydrosulfuricum*) (22) and similarly for the naturally more ethanol-tolerant *Zymomonas mobilis* (23). We did not test the ethanol cycling hypothesis in this study, but net ethanol oxidation did not appear to be a major detoxification mechanism (*SI Appendix*, Fig. S6).

The reduction in specific activity with respect to NADH was far greater (~25-fold less activity) than the increase with respect to NADPH when the WT and *adhE**(EA) ADH activities were compared (Table 4). Although total ADH activity dropped by 25-fold, ethanol production did not drop significantly between strains under these conditions (*SI Appendix*, Fig. S6). A recent study has shown that Cthe_0423 ADH transcripts are among the most abundant *C. thermocellum* transcripts (24). Although there was a decrease in total ADH activity in *adhE**(EA), it was still sufficient for ethanol production. The NADH/NAD ratio of WT *T. pseudoethanolicus* was raised, but not in an EA strain 39E (22), and an earlier study with the same mutant strain found that it lacked one of two ADHs (25). This study and earlier studies suggest that there may be similarities for ethanol tolerance mechanisms and redox homeostasis within the Clostridiales, thermophilic anaerobes, or some naturally ethanol-tolerant microorganisms. Further enzymatic and protein structural studies are required to elucidate how possible differences in the maintenance of NADP/NADPH pools links to membrane changes and the molecular mechanism of *C. thermocellum* ethanol tolerance.

Table 4. Cofactor specificity of WT and mutant *C. thermocellum* ADH

	NADH	NADPH
WT	2.7 (0.18)	0.025 (0.005)
EA	<0.005[†]	0.052 (0.007)
*adhE**(EA)	<0.005	0.12 (0.03)

Shown is the specific activity in μg of NAD(P)H oxidized per mg of crude extract protein per min. SD values are in parentheses.
[†]Below assay detection limit.

Although ethanol tolerance generally correlates with membrane alterations, ethanol tolerance may be limited by electron flux and central metabolism; however, further studies are required to examine carbon and electron flow, and possible linkages to changes in membrane composition. It is clear from this study that approaches to genetically modify *C. thermocellum* and possibly other microorganisms for biofuel production from cellulosic feedstocks must be reconsidered. Indeed, recent deletion of the *pta* gene, required for acetate production, resulted in the elimination of acetate as a fermentation end product but did not increase ethanol yield (26). Hence, not only ethanol tolerance but also ethanol production might be limited by electron flow as the ethanol concentration begin to rise. The use of a *C. thermocellum* strain with altered ADH cofactor specificity might help overcome issues related to carbon and electron flow. Aside from ethanol, the breadth of compounds tolerated by *C. thermocellum* strains EA and *adhE**(EA) is unclear. Future determination of compounds resisted by these strains may reveal the selective pressures that led to evolution of altered cofactor specificity of AdhE and suggest further paths for metabolic engineering of this organism for industrial biofuel production. Finally, the ability to identify and characterize sets of biological components linked to desired phenotypes, such as the mutated AdhE gene in this study, or overexpression of endogenous genes (11) offers the prospect for improved rational design of systems in the future that will be best suited to particular feedstocks and desired processes.

Materials and Methods

Detailed methods are available in the *SI Appendix, SI Materials and Methods*.

Strains and Culture Conditions. *C. thermocellum* WT strains ATCC 27405 and DSM 1313 were obtained from their respective culture collections. The *C. thermocellum* EA mutant culture was derived from strain ATCC 27405 and likely represents a mixed population of ethanol-tolerant strains. EA is the same culture that has been used in previous proteomics (5) and lipid (9) studies. The ethanol tolerance phenotype of EA has been described as being stable and retained after growth of >2,000 generations in the absence of ethanol.

Resequencing. Resequencing was conducted essentially as described (27). Briefly, genomic DNA from *C. thermocellum* WT ATCC 27405 or EA mutant cultures was sent to the NimbleGen facility for CGS service following the company's procedure. Pyrosequencing using the Roche 454 GS FLX System (454 Life Sciences) was carried out by using both shotgun and paired-end DNA library preparation methods, and sequences have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession nos. SRX030163.2 and SRX030164.1, respectively). The GSMapper application in the 454 GS FLX software package 1.1.03 (454 Life Sciences) was used to map the reads generated from GS FLX onto the *C. thermocellum* ATCC 27405 reference genome (GenBank accession no. CP000568).

Analysis of Distribution of Mutations Inferred from Pyrosequencing Data. Distribution of the mutations across the genome was analyzed by calculating quantities of each type of mutation (total mutations, nonsynonymous, synonymous, indels, and number of mutations in the intergenic regions) within each 1,000 and 5,000 nucleotides across the genome. For each position in the genome (L_i) we also calculated the local mutation frequency (LMP) as LMP = $1/[(M_{up} - L_i) + (L_i - M_{down})]/2*100$, where L_{up} is the start position of the closest upstream mutation and L_{down} is the start position of the closest downstream mutation. Mutation hot spots were identified by manual curation of the distributions and by analysis of potential functional relationships between genes comprising each hot spot. In bacteria, the neighboring gene or genes may comprise an operon and are likely involved in the same biological process or metabolic pathway. To find out whether mutations target such functionally related genes, we calculated the number of mutations in each pair of genes and in genes that belong to an operon with subsequent manual analysis of the affected protein products and their annotation with MetaCyc pathways (28) using *C. thermocellum* pathway genome database available in the BioEnergy Science Center KnowledgeBase (http://cricket.ornl.gov/cgi-bin/beocyc_home.cgi). At least one nonsynonymous mutation in two adjacent genes or in a gene of the operon was required to consider mutations as functionally related. The intergenic region was considered as functionally related to a hot spot if the distance from a mutation in the region to a mutated gene/operon was <400 nucleotides.

Plasmid and Strain Construction. The WT and EA *adhE* alleles were subcloned into plasmid pAMG205 (29), deleting *pyrF* and creating an artificial operon with the antibiotic resistance gene, *cat*, for expression in *C. thermocellum*. Empty vector control plasmid pAMG226 (pAMG205Δ*pyrF*) was constructed via restriction digestion of pAMG205 (29) with ZraI and SmaI, followed by self-ligation of the 6.9-kb fragment and transformation into *E. coli* Top10. Plasmids were then transformed into *C. thermocellum* DSM 1313 through electroporation as described (29) with minor modifications.

ADH Enzyme Assays. The ADH enzyme assays were based on described methods (30, 31). Briefly, half-liter cultures of each strain were grown to OD_{600} = 0.6 and centrifuged at 4 °C at 6,000 × *g* in a Beckman Coulter Avanti J-25 centrifuge with a JA-10 rotor. The culture was brought into the anaerobic chamber, and the supernatant was removed. All further steps were carried out in the anaerobic chamber. The cell pellet was resuspended in 4 mL of 100 mM Tris·HCl (pH 7.6)/0.1 mM DTT buffer, transferred to a 10-mL glass beaker, and sonicated for 8 min with 10-s pulses and 10-s pauses at 50% of the max intensity by using a Misonix Sonicator 4000 with a microtip. Crude cell extract was centrifuged at 14,000 × *g* for 25 min and stored on ice until assayed. The anaerobic reaction mixture contained 100 mM Tris·HCl (pH 7.6)/0.1 mM DTT buffer, 0.5 mM NAD(P)H, 55 mM acetaldehyde, and 2–50 μL of cell extract in 1.2-mL total volume. Decrease in absorbance was monitored at 340 nm to follow NAD(P)H oxidation (extinction coefficient 6.22 $mM^{-1}·cm^{-1}$) by using an Agilent 8453 UV-Vis spectrophotometer with Peltier controlled heating set at 55 °C. Protein concentration was determined by using the Bradford method.

Homology Modeling of *C. thermocellum* AdhE. By using the protein sequence of the ADH (AdhE) domain of *C. thermocellum* ALDH/ADH, homology models of WT and double-mutant AdhE were constructed. The HHPRED webserver (32, 33), part of the Bioinformatics Toolkit webserver (34), was used to perform multiple sequence alignments of the *C. thermocellum* AdhE sequence to potential structural templates available in the PDB (35). The 2.7-Å resolution X-ray structure of 1,3-propanediol dehydrogenase from *Klebsiella pneumoniae* (PDB ID 3BFJ; ref. 36) and the 1.3-Å X-ray structure of Fe-containing ADH from *Thermotoga maritima* Tm0920 (PDB ID 1O2D; ref. 37) were selected as templates based on their homology to AdhE and inclusion of NAD(P), Fe, or both in the structures. The program MODELLER (38–41) was used to construct homology structures including the NAD cofactor and Fe ligand, and resulting models were assessed by using the Discrete Optimized Protein Energy (DOPE) method (42). The fully automated I-TASSER webserver (43) was also used to construct a homology model, and it was found that MODELLER produced structures with a more favorable (i.e., lower) DOPE score if the I-TASSER model was included as an additional template structure. To provide an additional level of validation of the homology structures, the MolProbity server (44, 45) was used to perform a Ramachandran analysis on the final models.

Molecular Dynamics (MD) Simulations. MD simulations were performed on the homology models on WT and double mutant (P704L, H734R) AdhE by using the program GROMACS (46) with the CHARMM 27 force field (47) and TIP3P water model (48). Previously reported Lennard–Jones parameters for Fe(II) were used (49). Energy minimization was performed on the homology structures by using the steepest descent method for 1,000 steps, and then each protein was solvated in a rectangular water box of with a minimum of 10 Å from the surface of the protein to the edge of the solvent box. Sodium cations were added to neutralize the charge of the system. Periodic boundary conditions were imposed, and the particle mesh Ewald method (50, 51) was used to describe long-range electrostatic interactions. MD simulations were carried out with an integration time step of 2 fs. To reach the target temperature (298 K) and pressure (1 bar), the Berendsen method was used with a relaxation time of 0.1 ps (52). After a 1-ns equilibration, production simulations were performed in the NPT ensemble by using the Nosé–Hoover thermostat (53, 54) and the Parrinello–Rahman barostat (55, 56) with relaxation times of 1.0 ps. The production run was carried out for 10 ns, and coordinates were saved every 1 ps for analysis.

ACKNOWLEDGMENTS. We thank Herb Strobel (formerly of the University of Kentucky) for the generous gift of strains ATCC 27405 and EA; Marybeth Maloney for technical assistance in plasmid construction; Jessica Whitaker for strain confirmation; Hao-Bo Guo for assistance with modeling; Choo Hamilton for HPLC analysis; and Tatiana Vishnivetskaya for gDNA preparation. This work was supported by the Office of Biological and

Environmental Research in the Department of Energy (DOE) Office of Science through the BioEnergy Science Center, a DOE Bioenergy Research Center. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the DOE under Contract DE-AC05-00OR22725.

1. Lynd LR, van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: An update. *Curr Opin Biotechnol* 16:577–583.
2. la Grange DC, den Haan R, van Zyl WH (2010) Engineering cellulolytic ability into bioprocessing organisms. *Appl Microbiol Biotechnol* 87:1195–1208.
3. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66:506–577.
4. Herrero AA, Gomez RF (1980) Development of ethanol tolerance in *Clostridium thermocellum*: Effect of growth temperature. *Appl Environ Microbiol* 40:571–577.
5. 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.
6. Rani KS, Swamy MV, Sunitha D, Haritha D, Seenayya G (1996) Improved ethanol tolerance and production in strains of *Clostridium thermocellum*. *World J Microbiol Biotechnol* 12:57–60.
7. Alper H, Stephanopoulos G (2009) Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat Rev Microbiol* 7:715–723.
8. Ingram LO, Buttke TM (1984) Effects of alcohols on micro-organisms. *Adv Microb Physiol* 25:253–300.
9. Timmons MD, Knutson BL, Nokes SE, Strobel HJ, Lynn BC (2009) Analysis of composition and structure of *Clostridium thermocellum* membranes from wild-type and ethanol-adapted strains. *Appl Microbiol Biotechnol* 82:929–939.
10. Yang S, et al. (2009) Improved genome annotation for *Zymomonas mobilis*. *Nat Biotechnol* 27:893–894.
11. Yang SH, et al. (2010) Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 107:10395–10400.
12. Mao FL, Dam P, Chou J, Olman V, Xu Y (2009) DOOR: A database for prokaryotic operons. *Nucleic Acids Res* 37(Database issue):D459–D463.
13. Atsumi S, et al. (2010) Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. *Mol Syst Biol* 6:449.
14. Stephanopoulos G (2007) Challenges in engineering microbes for biofuels production. *Science* 315:801–804.
15. Hong M-E, et al. (2010) Identification of gene targets eliciting improved alcohol tolerance in *Saccharomyces cerevisiae* through inverse metabolic engineering. *J Biotechnol* 149:52–59.
16. Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* 314: 1565–1568.
17. Carugo O, Argos P (1997) NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins* 28:10–28.
18. Didierjean C, et al. (1997) A crystallographic comparison between mutated glyceraldehyde-3-phosphate dehydrogenases from Bacillus stearothermophilus complexed with either NAD+ or NADP+. *J Mol Biol* 268:739–759.
19. Rydzak T, Levin DB, Cicek N, Sparling R (2009) Growth phase-dependant enzyme profile of pyruvate catabolism and end-product formation in *Clostridium thermocellum* ATCC 27405. *J Biotechnol* 140:169–175.
20. 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.
21. Rosell A, et al. (2003) Complete reversal of coenzyme specificity by concerted mutation of three consecutive residues in alcohol dehydrogenase. *J Biol Chem* 278:40573–40580.
22. Lovitt RW, Shen GJ, Zeikus JG (1988) Ethanol production by thermophilic bacteria: Biochemical basis for ethanol and hydrogen tolerance in *Clostridium thermohydrosulfuricum*. *J Bacteriol* 170:2809–2815.
23. Kalnenieks U, et al. (2006) Respiratory behaviour of a *Zymomonas mobilis adhB:kan*(^{r}) mutant supports the hypothesis of two alcohol dehydrogenase isoenzymes catalysing opposite reactions. *FEBS Lett* 580:5084–5088.
24. Riederer A, et al. (2011) Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Appl Environ Microbiol* 77:1243–1253.
25. Lovitt RW, Longin R, Zeikus JG (1984) Ethanol production by thermophilic bacteria: Physiological comparison of solvent effects on parent and alcohol-tolerant strains of *Clostridium thermohydrosulfuricum*. *Appl Environ Microbiol* 48:171–177.
26. Tripathi SA, et al. (2010) Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. *Appl Environ Microbiol* 76:6591–6599.
27. Burdette D, Zeikus JG (1994) Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase (2 degrees Adh) as a bifunctional alcohol dehydrogenase—acetyl-CoA reductive thioesterase. *Biochem J* 302:163–170.
28. Caspi R, et al. (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 38(Database issue):D473–D479.
29. Olson DG, et al. (2010) Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *Proc Natl Acad Sci USA* 107:17727–17732.
30. Durre P, Kuhn A, Gottwald M, Gottschalk G (1987) Enzymatic investigations on butanol dehydrogenase and butyraldehyde dehydrogenase in extracts of *Clostridium acetobutylicum*. *Appl Microbiol Biotechnol* 26:268–272.
31. Shaw AJ, Jenney FE, Adams MWW, Lynd LR (2008) End-product pathways in the xylose fermenting bacterium, *Thermoanaerobacterium saccharolyticum*. *Enzyme Microb Technol* 42:453–458.
32. Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951–960.
33. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue): W244-8.
34. Biegert A, Mayer C, Remmert M, Söding J, Lupas AN (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* 34(Web Server issue):W335-9.
35. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242.
36. Marçal D, Rêgo AT, Carrondo MA, Enguita FJ (2009) 1,3-Propanediol dehydrogenase from Klebsiella pneumoniae: Decameric quaternary structure and possible subunit cooperativity. *J Bacteriol* 191:1143–1151.
37. Schwarzenbacher R, et al. (2004) Crystal structure of an iron-containing 1,3-propanediol dehydrogenase (TM0920) from Thermotoga maritima at 1.3 angstrom resolution. *Proteins Struct Funct Bioinf* 54:174–177.
38. Eswar N, et al. (2006) Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics*, eds Baxevanis AD, et al. (Wiley, New York), Vol 15, pp 5.6.1–5.6.30.
39. Fiser A, Do RKG, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9: 1753–1773.
40. Martí-Renom MA, et al. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291–325.
41. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.
42. Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524.
43. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–738.
44. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21.
45. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375-83.
46. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
47. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30:1545–1614.
48. 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.
49. Babu CS, Lim C (2006) Empirical force fields for biologically active divalent metal cations in water. *J Phys Chem A* 110:691–699.
50. Darden T, York D, Pedersen L (1993) Particle mesh Ewald—an N.log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
51. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103: 8577–8593.
52. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690.
53. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A* 31:1695–1697.
54. Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Mol Phys* 50:1055–1076.
55. Nosé S (1984) A molecular-dynamics method for simulations in the canonical ensemble. *Mol Phys* 52:255–268.
56. Parrinello M, Rahman A (1981) Polymorphic transitions in single-crystals—a new molecular-dynamics method. *J Appl Phys* 52:7182–7190.
57. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:33–38, 27–28.