

Engineering pyruvate decarboxylase-mediated ethanol production in the thermophilic host *Geobacillus thermoglucosidasius*

L. J. Van Zyl, M. P. Taylor, K. Eley, M. Tuffin and D. A. Cowan

Abstract

This study reports the expression, purification, and kinetic characterization of a pyruvate decarboxylase (PDC) from *Gluconobacter oxydans*. Kinetic analyses showed the enzyme to have high affinity for pyruvate ($120 \mu\text{M}$ at pH 5), high catalytic efficiency ($4.75 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ at pH 5), a pH_{opt} of approximately 4.5 and an in vitro temperature optimum at approximately 55°C . Due to in vitro thermostability (approximately 40 % enzyme activity retained after 30 min at 65°C), this PDC was considered to be a suitable candidate for heterologous expression in the thermophile *Geobacillus thermoglucosidasius* for ethanol production. Initial studies using a variety of methods failed to detect activity at any growth temperature ($45\text{--}55^\circ\text{C}$). However, the application of codon harmonization (i.e., mimicry of the heterogeneous host's transcription and translational rhythm) yielded a protein that was fully functional in the thermophilic strain at 45°C (as determined by enzyme activity, Western blot, mRNA detection, and ethanol productivity). Here, we describe the first successful expression of PDC in a true thermophile. Yields as high as $0.35 \pm 0.04 \text{ g/g}$ ethanol per gram of glucose consumed were detected, highly competitive to those reported in ethanologenic thermophilic mutants. Although activities could not be detected at temperatures approaching the growth optimum for the strain, this study highlights the possibility that previously unsuccessful expression of pdc's in *Geobacillus* spp. may be the result of ineffective transcription/translation coupling.

Introduction

Pyruvate decarboxylase (PDC, EC 4.1.1.1) is responsible for the non-oxidative decarboxylation of pyruvate to acetaldehyde and carbon dioxide. PDCs are common in the plant and fungal kingdoms and at least in the latter, together with alcohol dehydrogenase (ADH, EC 1.1.1.1) form part of an ethanol fermentation pathway (Konig et al. 1998). Several plant and yeast PDCs have been isolated and characterized, but as yet only four of bacterial origin has been described— from *Zymomonas mobilis*, *Zymobacter palmae*, *Acetobacter pasteurianus*, and *Sarcina ventriculi* (Raj et al. 2002). Bacterial PDCs participate in ethanol production via the Entner– Doudoroff pathway and not through glycolysis for pyruvate production.

There has been an increased interest in the use of thermophiles, such as *Geobacillus thermoglucosidasius*, for ethanol production, primarily because of their catabolic promiscuity, an important benefit for a second-generation bioprocess design (Cripps et al. 2009; Taylor et al. 2009). Other advantages include improved product removal, reduced incidence of

contamination, and high ethanol yields in selectively mutated strains (Taylor et al. 2009). Ethanol production in *G. thermoglucosidasius* and mutants with enhanced ethanogenic phenotypes relies on endogenous metabolic pathways, generating acetyl CoA via pyruvate dehydrogenase and its subsequent conversion to acetaldehyde and ethanol by aldehyde dehydrogenase and alcohol dehydrogenase, respectively (Cripps et al. 2009). An alternative to further develop *G. thermoglucosidasius* as an ethanogenic strain is to engineer the expression of the PDC pathway. The in vitro high specificity and thermostability (half-life of 30 min at 60 °C) of the *Z. mobilis* PDC (ZmoPDC) has made it the main candidate for such engineering (Pohl et al. 1995). However, both the ZmoPDC and *Z. palmae* PDC (ZpaPDC) have been expressed in *G. thermoglucosidasius* but do not function at temperatures exceeding 55 °C, despite good in vitro thermostability at these temperatures. The reasons for the low levels of activity are not fully understood (Taylor et al. 2008; Thompson et al. 2008), but protein misfolding resulting in inactive protein has been proposed (Thompson et al. 2008). Attempts to express these proteins in mesophilic Gram-positive hosts (notably lactic acid bacteria and *Bacillus megaterium*) have also had limited success (Gold et al. 1996; Bongers et al. 2005; Kaczowka et al. 2005; Talarico et al. 2005; Liu et al. 2005, 2006, 2007; Orencio-Trejo et al. 2008; Bi et al. 2009).

The role that codon usage plays in heterologous protein expression has been recognized, but is not well understood (Gustafsson et al. 2004). It has been demonstrated that the position and usage frequency of codons, together, play a role in correct protein folding and that “codon harmonization” could be used to overcome poor expression, at least in *Escherichia coli* (Angov et al. 2008; Rosano et al. 2009). Incompatibilities in codon usage and their effect on expression of PDCs have been reported (Talarico et al. 2001, 2005; Lowe et al. 1992). Two examples of the effect of codon usage on PDC production include the five- to tenfold increase in soluble *S. ventriculi* PDC (SvePDC) when expressed in an *E. coli* strain with or without accessory tRNAs (specifically those which are rarely used in *E. coli*), as well as the superior production of this PDC relative to those from *A. pasteurianus* and *Z. mobilis* in *B. megaterium* (Raj et al. 2002; Talarico et al. 2001, 2005).

Despite the rarity of prokaryotic PDCs, we have identified a *pdh*-like gene sequence in the genome sequence of a Gram-negative acetic acid bacterium, *Gluconobacter oxydans*. A PDC enzyme from *G. oxydans* has previously been characterized (King et al. 1954). *G. oxydans* is often associated with sugar-rich environments such as ripe fruit, honey, and cider as well as in a variety of soil types and is used industrially to produce L-sorbose from D-sorbitol, D-gluconic acid and 5-keto- and 2-ketogluconic acids from D-glucose, and dihydroxyacetone from glycerol (Gupta et al. 2001). This organism uses the PDC as part of the well-characterized lactate oxidation and acetate excretion pathways (Raj et al. 2001; Peters et al. 2012).

The aim of this study was to evaluate the ethanogenic potential of the *G. oxydans* PDC (GoxPDC) expression in *G. thermoglucosidasius*. We report the cloning, expression, and characterization of GoxPDC as well as its “codon harmonization” for enhanced expression in this Gram-positive thermophilic host.

Materials and methods

Media, bacterial strains, and plasmids

Bacterial strains and plasmids used in this study are shown in Table 1. *E. coli* strains were grown in Luria–Bertani (LB) broth (Sambrook et al. 1989) with 200 µg/ml ampicillin or 50 µg/ml kanamycin added as required. *G. thermoglucosidasius* strains were cultured either in LB, 2TY, TGP media, or modified urea sulphates medium (USM). In general, *E. coli* DH5α was used for plasmid construction.

One liter of TGP broth contained 17 g tryptone, 3 g soy peptone, 2.5 g K₂HPO₄, and 5 g NaCl. The pH was adjusted to 7.3 before autoclaving, after which 4 g Na-pyruvate and 4 mL glycerol were added in the form of filter-sterilized 10× concentrates. For solid media, 15 g/L agar was added before autoclaving. LB and TGP were used during genetic manipulation and general maintenance of cultures.

Per liter, 2TY medium contained 10 g yeast extract, 5 g NaCl, 20 g tryptone, and 15 g agar (where applicable), with a final pH of 7.0. USM supplemented with yeast extract (USMYE) contained 10 g glucose, 0.42 g citric acid, 0.31 g MgSO₄, 3.1 g NaH₂PO₄, 3.5 g K₂SO₄, 3 g urea, 2.2 mg CaCl₂, 0.4 mg Na₂MoO₄, 1 g yeast extract, 1 g typtone, 8.36 g Bis-Tris, 12.08 g PIPES, 10.4 g HEPES, 1 ml silicone antifoam, and 5 ml trace elements solution per liter. The trace element solution contained (per liter) 1.44 g ZnSO₄·7H₂O, 0.56 g CoSO₄·6H₂O, 0.25 g CuSO₄·5H₂O, 5.56 g FeSO₄·6H₂O, 0.89 g NiSO₄·6H₂O; 1.69 g MnSO₄, and 5.0 ml 12 M H₂SO₄. The trace elements solution and glucose (50 ml of a 20 % w/v solution) were added aseptically after autoclaving. The pH of USM was adjusted to pH of 7 using 10 M NaOH. Cultures were incubated at 45, 52, or 60 °C as required with vigorous aeration.

G. oxydans was cultured in medium containing, per liter: 8 g yeast extract, 15 g peptone, 10 g glucose, 0.5 % (w/v) ethanol, and 0.3 % (w/v) acetic acid. The final pH was between 3.5 and 4. Ethanol, acetic acid, and glucose were added after autoclaving. Cultures were incubated at 25 °C.

DNA manipulations and sequencing

Restriction endonuclease digestion, gel electrophoresis, and ligation performed using standard methods or following the manufacturers' recommendations (Sambrook et al. 1989). Ultrapure plasmid DNA was obtained using the Wizard Plus SV miniprep DNA purification system (Promega™). Total DNA from all bacterial strains was prepared as described (Kotze et al. 2006).

Table 1 Bacterial strains, plasmids, and primers used in this study. Underlined sections in primer sequences indicate restriction endonuclease sites

Strain or plasmid	Genotype or description	Source or reference
Strains		
<i>G. thermoglucosidasius</i> TM89	<i>ldhA</i> ⁻ variant of <i>G. thermoglucosidasius</i> NCIMB 11955	TMO renewables
<i>G. oxydans</i> (DSMZ7145)	Isolated from beer	DSMZ
<i>E. coli</i> DH5 α	F' <i>endA1 hsdR17</i> (r_{κ}^{-} m_{κ}^{+}) <i>supE44 thi-1 reacA1 gyrA</i> (Nal ^r) <i>relA1</i> Δ (<i>lacZYA-argF</i>)U169 (ϕ 80 <i>dlac</i> Δ (<i>lacZ</i>)M15)	Promega Corp.
<i>E. coli</i> BL21-DE3	<i>E. coli</i> B F ⁻ <i>dcn ompT hsdS</i> (r_{B}^{-} m_{B}^{-}) <i>gal λ</i> (DE3)	Invitrogen Corp.
<i>E. coli</i> JM109	F' <i>traD36 proA⁺B⁺ lacIq</i> Δ (<i>lacZ</i>)M15/ Δ (<i>lac-proAB</i>) <i>glnV44 e14⁻ gyrA96 recA1 relA1 endA1 thi hsdR17</i>	New England Biolabs, Beverly, MA, USA
Plasmids		
pET28a	Kan ^r ; ColE1 replicon, HIS-tag expression vector	Novagen Corp.
pGO	Kan ^r ; ColE1 replicon; <i>G. oxydans pdc</i> gene cloned into pET28a	This study
pldhGO	Kan ^r ; ColE1 replicon; lactate dehydrogenase (<i>Pldh</i>) gene promoter region (\pm 170 bp <i>Nco</i> I- <i>Nde</i> I) from <i>G. thermoglucosidasius</i> NCA1503 cloned upstream of the <i>G. oxydans pdc</i> gene in pET28a	This study
pTM049	Derivative of pUB190 containing the <i>ldh</i> promoter from <i>G. stearothermophilus</i> NCA1503.	TMO Renewables (Cripps 2009)
pTMO111	Amp ^r , Kan ^r (in <i>G. thermoglucosidasius</i>) ColE1 replicon, pUB110 IncA replicon, <i>E. coli</i> - <i>G. thermoglucosidasius</i> shuttle-suicide (>55 °C) vector containing a truncated <i>pflB</i> gene	TMO Renewables (Cripps 2009)
pGO111 (GoxPDC _{WT})	Amp ^r , Kan ^r (in <i>G. thermoglucosidasius</i>), 3603 bp <i>Dra</i> III- <i>Eco</i> RV fragment, blunted at the <i>Dra</i> III end, from pldhGO cloned into the unique <i>Swa</i> I site of pTMO111	This study
pGOF111 (GoxPDC _{OPT})	Amp ^r , Kan ^r (in <i>G. thermoglucosidasius</i>), 1,887 bp <i>Not</i> I- <i>Not</i> I fragment, containing the fully codon optimized <i>G. oxydans</i> PDC with the <i>Pldh</i> upstream, cloned into the unique <i>Not</i> I site of pTMO111	This study
Primers		
LDHF	5'-TATACCATGGGCGGGACGGGGAGCTGAGTGCTC-3'	Cripps (2009)
LDHR	5'-GCCGCATATGATTCATCCTCCCTCAATAT-3'	Cripps (2009)
GOPDC _p ETF	5'-GGAATTCATATGACTTATACTGTCCG-3'	This study
GOPDC _p ETR	5'-CCGCTCGAGTCAGACGCTCTGCCG-3'	This study

The QIAGEN plasmid midi kit was used for large-scale plasmid preparations. DNA was sequenced using an ABI Prism 377 automated DNA sequencer and sequences were analyzed with DNAMAN (version 4.1, LynnonBioSoft). Codon usage in *G. thermoglucosidasius* NCIMB 11955 (with particular reference to the PDC genes from *G. oxydans*) were analyzed using the web servers <http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=1426&aa=1&style=N> and <http://gcua.schoedl.de/>.

Polymerase chain reaction

Polymerase chain reaction (PCR) was performed using Phusion DNA polymerase (New England BiolabsTM). Generally, 50 ng DNA were used in a 50 μ l reaction volume containing 2 mM MgCl₂, 0.125 μ M of each primer, 0.2 mM of each deoxynucleoside triphosphate, and 1 U DNA polymerase. Reactions were carried out in a Hybaid Sprint thermocycler, with an initial denaturation at 94 °C for 60 s, followed by 30 cycles of denaturation (30 s at 94 °C), annealing (30 s), and variable elongation (72 °C), where annealing temperatures and elongation times were adjusted as required. Primers are listed in Table 1.

Cloning of the *G. oxydans* pdc gene

The pdc gene was amplified using genomic DNA isolated from *G. oxydans* DSMZ7145 using primers GoxPDCpETF and GoxPDCpETR (Table 1). The gene encoding the *G. oxydans* pyruvate decarboxylase was cloned into the pET28a expression vector in two parts. The 5' 913 bp fragment was cloned by digesting both PCR product and pET28a with NdeI and XhoI. The 779 bp 3'-fragment was first cloned into pBluescriptSK by digesting the GoxPDC PCR product with Xho I, and then excised by Xho I digestion and cloned into Xho I-digested pET28-GoxPDC-5'. Clones with the correct orientation of this 779 bp fragment were identified by restriction enzyme digest of the final construct (pGO) using Sph I, and confirmed by sequencing. The nucleotide sequences of the wild type and codon-optimized pdc genes have been submitted to the EMBL-GenBank database and are available under accession numbers KF650838 and KF650839, respectively.

Purification of PDC protein

An overnight culture of pGO (Table 1) in *E. coli* BL21-DE3 with kanamycin was used to inoculate fresh LB broth (1/ 100 ml) and then incubated with aeration (120 rpm) overnight at room temperature to express the protein without IPTG induction. The cells were collected by centrifugation (3, 214×g for 10 min). BugBuster™ was used to lyse cells (3 ml/g of wet cells) and the suspension incubated at room temperature for 20 min with shaking. After centrifugation to remove cell debris (12,857×g for 20 min), DNaseI and RNaseA (Fermentas) were added (10 U/ml) to the lysate to reduce the viscosity and incubated at room temperature with shaking for 30 min. The HisBind™ resin and buffer kit (Novagen) was used to purify the protein. After elution with 9 ml imidazole buffer (100 mM), the protein was dialysed against 200 volumes of buffer (50 mM MES pH 6.4) containing 1 mM TPP and 1 mM MgCl₂. Purity was estimated by reducing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel (12 %) and the protein concentrations determined using Bradford reagent (Bio-Rad) with bovine serum albumin as the standard (Laemmli et al. 1970).

Steady state kinetics and substrate specificity

GoxPDC activity was measured using a coupled assay with baker's yeast ADH (Sigma-Aldrich) as described (Conway et al. 1987), and represents a different assay to that used in the initial *G. oxydans* PDC characterization in the King and Cheldelin study (King et al. 1954). The enzyme was purified independently four times and kinetic measurements performed with each preparation to give the average ($n \geq 3$) of the results shown. The reaction mixture (1 ml final volume) contained 0.25 mM NADH, 5 mM MgCl₂, 0.1 mM TPP, 5 mM pyruvate (unless stated otherwise), and 10 U of ADH in 50 mM MES or 200 mM Na citrate pH 6.4 or 6.0, respectively. For substrate range determination, ADH was replaced with 1 U/ml baker's yeast aldehyde dehydrogenase (ALDH, Sigma-Aldrich). β -mercaptoethanol was added to a final concentration of 3 mM and NADH replaced with NAD⁺. Assays were performed in 100 mM citric acid/K₂HPO₄ buffer pH7.0 (Vuralhan et al. 2005). Activities were recorded at 25 °C unless otherwise indicated using a Cary 50 temperature controlled spectrophotometer (Varian). In the case where aldehydes produced by the decarboxylation of certain substrates by PDC were not recognized as a substrate for ALDH, high-performance liquid

chromatography (HPLC) analysis was used to determine activity on that substrate. Reactions were run on a Rezex RHM monosaccharide column (Phenomex), using 5 mM H₂SO₄ as mobile phase under isocratic elution (0.6 ml/min, 40 °C) on a Dionex Ultimate 3000 machine. Samples (20 µl) were injected by autosampler and the components detected using either a refractive index detector or a UV/ Vis photodiode array at 215 nm. For kinetic data, initial rates were measured over the substrate range of 0.1 to 30 mM for pyruvate or 0.1 to 50 mM for other 2-keto acids. Kinetic parameters were determined by nonlinear fitting of data to hyperbolic curves according to Michaelis–Menten (GraphPad Prism v. 4.00, GraphPad Software, San Diego, CA, USA). One unit of enzyme activity corresponds to the amount of enzyme that generates 1 µmol of acetaldehyde per minute. k_{cat} values were calculated based on the MW of the monomer with one active site.

Construction of the pdc expression for transformation in *G. thermoglucosidasius*

For expression of PDC in *G. thermoglucosidasius*, the 170-bp promoter region of the lactate dehydrogenase gene *Pldh* from *G. thermoglucosidasius* NCA1503 was cloned upstream of the *Goxpdc* gene. This promoter has been shown to be induced under microaerobic conditions in *G. thermoglucosidasius* (Cripps et al. 2009). The promoter region was amplified from pTMO49 using the LDHF and LDHR primers and cloned into pGEM-T Easy. Sequencing confirmed that no DNA base changes had occurred. The *pldh* was cloned into pGO (Table 1) using the *Nco*I and *Nde*I sites such that the promoter was functional for *pdc* expression. This construct was digested with *Dra* III and the ends filled in using T4 DNA polymerase (Fermentas, ThermoFisher). Digestion with *Eco* RV yields a 3,603 bp *Dra* III (blunt)–*Eco* RV fragment. The plasmid TMO111 was digested with *Swa* I and treated with rAPid™ alkaline phosphatase (Roche) to prevent self-ligation. The *Dra* III (blunt)–*Eco* RV fragment and *Swa* I digested pTMO111, were ligated using T4 DNA ligase (Fermentas). Insertion at the *Swa*I site leaves 809 and 436 bp of the *pflB* gene on either side of the *ldh-pdc* for recombination with its chromosomal counterpart.

The constructs were passaged through *E. coli* JM109 for DNA methylation prior to transformation to prevent endonuclease degradation in *G. thermoglucosidasius*. *G. thermoglucosidasius* competent cells were prepared and transformed (Cripps et al. 2009).

Fermentative product profile quantification

Cultures expressing *pdc* gene were grown overnight at 37 °C for 16 h and 200 rpm in LB media (*E. coli*) and 2TY media (*G. thermoglucosidasius*) at 60 °C. A volume of 0.5 ml of this culture was transferred to 10 ml of USMYE media, contained separately, in 50 and 15 ml screw-cap universal tubes. This effectively generated 40 ml and 5 ml headspaces, respectively, mimicking aerobic and microaerobic or fermentative culture conditions. These cultures were grown overnight at 37 (*E. coli*), 45, or 52 °C (*G. thermoglucosidasius*) for 16 h and 200 rpm and the supernatant removed by centrifugation (2, 057×g for 10 min). Metabolite concentrations in culture supernatants were determined by HPLC (see above) and products were compared to suitable standards of known concentration and against the media in which the cultures

were grown. Experiments were carried out in triplicate for *E. coli* and at least in duplicate for *G. thermoglucosidasius*.

Western blotting

Rabbit anti-GoxPDC polyclonal antibodies were made by Antibodies Incorporated (Davis, CA, USA) using His-tag purified GoxPDC protein. Cells were harvested directly after fermentation by centrifugation at $3,214\times g$ for 10 min. The cell pellet was resuspended in MES buffer pH6.5 and sonicated using five pulses of 30 s each. Cell debris was removed by centrifugation ($15,682\times g$ for 20 min), and the supernatant decanted. Protein concentrations were determined by Bradford assay. Forty micrograms of total protein was loaded for each sample and run on a 12 % SDS-PAGE gel. Protein was transferred from the gel to Biotrace™PVDF membrane by semidry blotting. For signal detection, the anti-rabbit Super Signal West Femto Chemiluminescent Substrate (Pierce) kit was used and the signal visualized using a chemiluminescent camera.

Results

Amino acid sequence considerations in the *G. oxydans* PDC

The cloned and sequenced GoxPDC gene differed from the published genome sequence (NC_006677.1) (Prust et al. 2005) at 22 positions, resulting in 5 amino acid changes (Y163F, S207N, A209T, I469M, and D517E) but no frame shifts or deleterious events. The sequence alignment (Fig. 1) indicated that none of the affected residues have been shown to be directly involved in catalysis, substrate, or co-factor binding. Most changes are conservative (F163, N207, and E517) and/or are located distantly from the active site (F163, N207, T209, and E517). Met469 is, however, located in a region of the enzyme which may be sensitive to changes. It is positioned adjacent to Glu468, important in catalytic activity; Ile467, involved in substrate recognition; and Ile471, crucial for substrate positioning (Prust et al. 2005; Pohl et al. 1998; Siegert et al. 2005). However, equivalent residues in other PDCs have not been associated with catalysis or substrate recognition (Fig. 1). Phe163 (Fig. 1) is an interesting change, as it is occupied by a tyrosine in all the PDCs (including the *G. oxydans* NC_006677.1 PDC, except for SvePDC). GoxPDC, unlike ZpaPDC and ApaPDC, also contains an extra six amino acid loop from position 498-503 (EESGKY), which is similar to positions 503-508 in ZmoPDC (DSGAGK), but the residues are not conserved. As the encoding gene was amplified using a polymerase with 3'-5' exonuclease activity and the differences were consistent in several independent clones, we infer that the changes represent natural variations in the GoxPDC and are not artifacts of the cloning procedure.

The protein sequence demonstrates all the features typical of ThDP-binding enzymes including the conserved ThDP binding motif GDGS-XXX-NN as well as several conserved residues required for substrate binding and catalysis (indicated in Fig. 1). It does, however, lack a cysteine residue equivalent to Cys221 in ScePDC, as shown for all the bacterial PDCs, thought to be involved in allosteric substrate activation (Lu et al. 2000).

Kinetic characterization of the GoxPDC enzyme

GoxPDC was purified to homogeneity by affinity chromatography, and the estimated molecular weight of the protein, at ± 60 kDa (Fig. 2), corresponds well to the theoretical molecular mass of 60.8 kDa. The predicted pI value is 6.0. Conventional enzyme characterization was performed using pyruvate as a substrate (kinetic data for GoxPDC are summarized in Table 2). The K_M value for pyruvate was found to be in the range of those determined for other PDCs from Gram-negative bacteria assayed under similar conditions. The enzyme also displayed a ± 20 -fold decrease in the K_M for pyruvate with a decrease in pH from 7 to 5, without an equivalent change in the catalytic rate (k_{cat} showed an approximate twofold decrease). This is in line with previous observations in related enzymes (Raj et al. 2002; Siegert et al. 2005) and supports the interpretation that PDCs require a protonated residue for efficient binding of the substrate, in which the ionizable group is thought to be the aminopyrimidine ring of the ThDP coenzyme (Meyer et al. 2010). The GoxPDC enzyme displayed normal Michaelis–Menten kinetics with pyruvate as the substrate and was not subject to allosteric substrate activation as has been reported for PDCs from plants, yeasts, and the SvePDC (Konig et al. 2009). The GoxPDC has a lower catalytic rate than ApaPDC but the catalytic efficiencies were similar to those reported for ZmoPDC and SvePDC. The pH optimum of GoxPDC was determined to lie between 4.5 and 5.0, similar to ApaPDC and slightly lower than for other PDCs from Gram-negative bacteria (Fig. 3; Gocke et al. 2009). The temperature optimum of GoxPDC was between 50 and 55 °C. Thermal inactivation studies demonstrated that the enzyme was stable at the experimental T_{opt} , with no loss of activity after an hour of incubation. However, at temperatures ≥ 60 °C, moderate to rapid loss of activity was recorded, retaining 30 % of the initial activity at 65 °C after an hour of incubation (Fig. 4). These data demonstrate that GoxPDC thermostability is equivalent to the Gram-negative homologs.

ZmoPDCMSYTVGIMLAGRLVCIQLPHFPAVGDMLVLDLILLNKNMEQVYCCNEINCGF 55
 GdiPDCMTYTVGMYIARLACIGLPHFPAVGDMLVLDLILLNLMQYVCSNEINCGF 55
 ZpaPDCMTYTVGMYIARLACIGLPHFPAVGDMLVLDLILLNKNMEQVYCCNEINCGF 54
 ApaPDCMTYTVGMYIARLACIGLPHFPAVGDMLVLDLILLNKNMKQYVCCNEINCGF 55
 GoxPDCMTYTVGMYIARLACIGLPHFPAVGDMLVLDLILLNKNMKQYVCCNEINCSF 55
 ZmaPDC METLLAGNPANGVAKPTCNVGVGALFVANSHAIATPAAAAATLAPAGPTIGRHLARPLVCI GASDVFAVGDMLVLDLILIAEPGLTLVGCNEINAGY 100
 ScePDCMSEITIGKMLFBRKCVNNTVFGLEFDENLSIDIKIYEVEGMWAGNANEINAGY 56
 SvePDCMKITIAEMLRLRKEVNVHEHFGVGDMLVLDLILVDESKDIEVWGSQNEINAGY 55
 Consensus t l rl f gd nl ld neln

ZmoPDC SAFCYARAFGAAAAVTIFSVGALSFAFTAIGGAYENIFVIIISGFFNNNDHAAGHVIHHALGKTDYHYQLEMARNTAAAEFTYTPESAFA<IDHVIRKTA 155
 GdiPDC SAFCYARAFGAAAIVTIFSVGALSFAFNALGGAYENIFVIIISGFFNNAND<GTGHIIHHITIGTIDYGYQLEMARHITCAAES<VAEDAFA<IDHVIRTA 155
 ZpaPDC SAFCYARAFGAAAIVTIFSVGALSFAFNALGGAYENIFVIIISGFFNSND<GTGHIIHHIRIGKIDYSYQLEMARQVTCARES<TDAHSAFA<IDHVIRTA 154
 ApaPDC SAFCYARAFGAAAIVTIFSVGALSFAFNALGGAYENIFVIIISGFFNSND<GTGHIIHHIRIGKIDYSYQLEMARQVTCARES<TDAHSAFA<IDHVIRTA 155
 GoxPDC SAFCYARAFGAAAIVTIFSVGALSFAFNALGGAYENIFVIIISGFFNSND<GSGHVIHHIRIGTIDYSYQMEMAKHVTCARES<TSAETAFA<IDHVIRTM 155
 ZmaPDC SAFCYARAFGAAAVTIFSVGALSFAFNALGGAYENIFVIIISGFFNSNDY<GTRNRIHHITIGLEDFSQELRCFQTITICYQATINLLNLDHED<IDTAIATA 200
 ScePDC SAFCYARAFGAAAVTIFSVGALSFAFNALGGAYENIFVIIISGFFNSNDY<GTRNRIHHITIGLEDFSQELRCFQTITICYQATINLLNLDHED<IDTAIATA 156
 SvePDC SAFCYARAFGAAAVTIFSVGALSFAFNALGGAYENIFVIIISGFFNSNDY<GTRNRIHHITIGLEDFSQELRCFQTITICYQATINLLNLDHED<IDTAIATA 155
 Consensus a gyar g t vg s g e g p hh i a

ZmoPDC IPEKRPVYLEIACNIAIMPCAAAPGASALFNDEASDEASLNA...AVEETIKFIANRDKVAVLVSGSKIFAAAEAAVAFADALGGAVATMPAAKSF<FEE 253
 GdiPDC IPEKRPVYLEIACNIVAGACVVRPGGIDALLSPPAPDEASLKA...A<DAALAFIEQRGSVIMLVGSRIFAAGAAQAVAIADALGGAVTITMPAAKSF<FED 253
 ZpaPDC IPEKRPVYLEIACNIVAGACVVRPGPINSLLRELEVDTQSVTA...A<DAAVEWLQDRQNVMLVSGSKIFAAGAAEQAVAIADALGGCAVITMPAAKGF<FED 252
 ApaPDC IPEKRPVYLEIACNIASEPCCVRPGPVSSLLSEPEIDHTSLKA...A<DATVALLKNRPAFVMLLGSKIFAANA<LAATETIADKLQCAVITMPAAKGF<FED 253
 GoxPDC IPEKRPVYLEIACNISAAPCVVRPGPVSSLLHAHPRPDEASLKA...A<LDESLSFLNKTNKVAIVLVGKIFAATA<LKETVELADKLGCPVTVMPAAKSY<FET 253
 ZmaPDC IRESKPVYISVSNLGLSHPTFSRDPVPMFISPRLSNKANLEYAVEAAAD<FLNKAVKPVVMVGGPKIRVA<AREAFAAVADASGYPPFVAVMAA<KGLVEEH 300
 ScePDC YVTRCPVYGLPANLVDLANPAKLQTFIDMSLKPNDASEKEVIDTILALV<KDAKNPILALACCSRH<DVKAETFKLIDLTFQPAFVTPNGKGSID<EQ 255
 SvePDC KYQLPGYIELFVDIVSKIEIEMKPLNLTMRSEKTELEKFNVDKEMVASS<KQH...ILADYEVIFA<KAKELEGFINEAKIPVNTLSIG<KTAVSES. 251
 Consensus r

ZmoPDC NEHYIGTSWGAIVSYPGVEKTMKEDAVIALAPVENVYSTIGTWDI<PDPKRLVLAEPRSVVVNGIRFVS<VHLKDYLTRLAQVSKKGTALDFFK<SLNAGEL 353
 GdiPDC HEGYFGHYWGIVSSPGAQAVECD<DEVICLAPVENVYAVVGSANWFKGDNVMVLDHRTVTVGGVAYAGIM<DRDFLTRAHTVRRDATARGA<VQTPP 353
 ZpaPDC HNFYFGHYWGIVSSPGAQAVECD<DEVICLAPVENVYAVVGSANWFKGDNVMVMDTRVTVFAGQ<SFEGLSLSTFAA<LAEKAPARPA<SAQKTSQVPLGIE 352
 ApaPDC HAGFGLYWGIVSNPGVQELVETS<DALLCLAPVENVYSTVGSNMGFKGPNVILAE<PDRVTVDGRAYLGF<TLRAFLLQALAEKAPARPA<SAQKTSQVPLGIE 353
 GoxPDC HEGYFGHYWGIVSSPGAQAVECD<DEVICLAPVENVYSSGGNKSVVVR<KEVLEVDNPRVTVNGKT<FEGFRLKEFVKALTEKAPK<KSAALTGEYKPVMLPK 353
 ZmaPDC HBRFICVYKGVSTFTFCATIVES<PDAYLFCAGIENYSSVLYSLLIKREKAVIV<QDPGPAFCIIMPE<FLRALAKRLRNTTAYINRRI<FVPD 400
 ScePDC HPRYGGVYWGIVSKPEVKEAVE<SDILLSVGLLSDENTGSFSYSYKTKNIVE<FSDHMKIRNATFF<GVQMKFVLQKLLT<TIADAARKYKPVAV<PARTPA 355
 SvePDC NEYFAGLESQTS<SDLVKELCKASD<IVLLFGVKEIDTITAGFRYINKD<VRMIEIGLFD<CRIGETIY<TGLYTKDVIKAL<TDAKIKFHN<DVKVEREAVE<KEFV 351
 Consensus g g s d

ZmoPDC KKAAPADPSAPLVNAEIRQVEALLTPNTTVIPE<TGHSWFNA<FRMKLPHGARVEYEM<CWGHIGW<SVPAAF<GNA<VGAPE...RRN<ILMV<GIGSG<QIT<PCE 449
 GdiPDC AAAPT...A<PLNNAEMARQIGALLTPRTTLLT<PETGHSWFNA<FRMKLPHGARVELEM<CWGHIGW<SVPAAF<GNA<LAPE...RQH<VIMV<GIGSG<QIT<PCE 445
 ZpaPDC AAEPN...A<PLTNDMETRQIQSLITS<DITLLT<PETGHSWFNA<FRMPICPGGARVELEM<CWGHIGW<SVPAAF<GNA<VGSPE...RRH<IMMV<GIGSG<QIT<PCE 444
 ApaPDC ATSD...A<GLTNDIEIVRHINALLTSNTTLLV<PETGHSWFNA<FRMILA<GARVELEM<CWGHIGW<SVPAAF<GNA<VGSQD...RQH<VIMV<GIGSG<QIT<PCE 444
 GoxPDC ADPSK...PLSNDMETRQINELVDGNTTLF<PETGHSWFNA<FRMHLPEGA<VETEM<CWGHIGW<SVPMF<GNA<TASPE...RKH<VIMV<GIGSG<QIT<PCE 444
 ZmaPDC REPPNGKPNELPRVNVLFKHIKGLSGDSAVV<PETGHSWFN<CQKRL<PEG<GYEFQ<CYGSG<VSVGATL<GVA<QAAD...KRV<IACI<GIGSG<QIT<PCE 496
 ScePDC NAAVPA<STFLKQEWMMNQLGNFLQEGDVVI...<PETGHSE<GINQ<TFFPNNTY<GISV<VLWGS<ICFTTGATL<GAA<FAAEE<IDPKKRV<LLF<GIGSG<QIT<PCE 453
 SvePDC PTDARLQDRYFKQMEAFKPNNDVLV...<ETG<SYSGA<NMR<FPEG<SSVFG<QSS<MSTIGYAT<PAVIG<PHLAD<KSRN<ILLS...<GIGSG<QIT<PCE 441
 Consensus etg s ig g gcgs q t q

ZmoPDC VACMRRLKLPVIFL<LNNGY<IEV<IH<EGEYNNIKN<WYACIMEVFN<GNGYDSGAGK<LKA...TGGELAAE<IKVALANTDGP<TILIE<CFIGRED<CTEE 547
 GdiPDC VACMRRLKLPVIFL<LNNGY<IEV<IH<EGEYNNIKN<WYACIMEVFNAGE<N...<GLGLRARTGGELAAE<IEQARANRNGP<TILIE<CTDRD<DCTQE 539
 ZpaPDC VACMRRLKLPVIFL<LNNGY<IEV<IH<EGEYNNIKN<WYAGLIDVFN<DEGDH...<GLGLKASTGAE<LEGA<IKKALDNRRGP<TILIE<CNIAQDD<CTET 538
 ApaPDC VACMRRLKLPVIFL<LNNGY<IEV<IH<EGEYNNIKN<WYACIMEVFNAGE<H...<GLGLKATTPKELTEA<IARAKANTRG<PILIE<CQDRD<DCTET 538
 GoxPDC VACMRRLKLPVIFL<LNNGY<IEV<IH<EGEYNNIKN<WYAA<LMQC<FNQGV<PGEES<GKYGLHAT<TGAE<LAEA<IARAKANTRG<PILIE<CKLDRD<DCTET 544
 ZmaPDC VSTMRCG<KSI<FL<LNNGY<IEV<IH<EGEYNNIKN<WY<TGLVNA<IHNS<EGNCWIM<KRVTEE<QLKEA<IATV<GAKKDC<LCP<IEVIV<HKDDT<SKELLEW 596
 ScePDC ISTMRCG<KLPY<FL<LNNGY<IEV<IH<EGEYNNIKN<WY<TGLVNA<IHNS<EGNCWIM<KRVTEE<QLKEA<IATV<GAKKDC<LCP<IEVIV<HKDDT<SKELLEW 553
 SvePDC VSTMRCG<KLNIV<FL<LNNGY<IEV<IH<EGEPERE<Y<NHQ<WY<AELV<KTLATER<DIQ<PTCFK<VTE<KELAA<AMEE<IN<KGT<EGIA<FV<EVM<DRM<DAP<KSLRQ 541
 Consensus m r f nn gy ie ih

ZmoPDC LVKWKGRVAANSRKPVNKL 567
 GdiPDC LVTWGRVAANARQPRAG. 558
 ZpaPDC LIAWGRVAAINSRKPPQA.. 556
 ApaPDC LVQWGRVASTNARKTTLA. 557
 GoxPDC LVEWGRVAANSRKPPQSV. 563
 ZmaPDC SRVS...AANSRPPNPQ. 610
 ScePDC KLTAAATNAKQ..... 563
 SvePDC EASLFSSQNNY..... 552
 Consensus

Fig. 1 Multiple sequence alignment of the protein sequences from selected PDC proteins: *GdiPDC*, *G. diazotrophicus* (YP_001600462.1); *GoxPDC*, *G. oxydans*; *ZpaPDC*, *Z. palmarum* (AF474145_1); *ApaPDC*, *A. pasteurianus* (AF368435_1); *ZmoPDC*, *Z. mobilis* (YP_163095.1); *ZmaPDC*, *Z. mays* (P28516.1); *ScePDC*, *S. cerevisiae* (EGA85775.1); *SvePDC*, *S. ventriculi* (AF354297_1). The alignment was generated using the “full alignment” feature in DNAMAN with default setting. Residues shaded in *black* are 100 % conserved while those in *grey* are 75 % conserved. The *underlined region* shows the conserved ThDP-binding motif and *triangles* indicate those residues which bind ThDP. *Arrows* indicate Mg²⁺ binding residues. *Circles* indicate residues which line the catalytic pocket and are thought to play a role in catalysis. *Asterisk* indicates the residue involved in substrate specificity. The *star* shows a residue thought to be needed for positioning of the substrate for catalysis. The *square* indicates the arginine residue involved in substrate activation of *ScePDC* and *SvePDC*. The *open triangle* shows the position of the unique phenylalanine residue in *GoxPDC*.

The following substrates were shown by HPLC analysis to be decarboxylated by *GoxPDC*; 2-ketopropanoate (pyruvate), 2-ketobutanoate, 2-ketopentanoate, 2-keto-4-methylpentanoate, and 4-hydroxy-phenyl-pyruvate. Calculated specific activities were 32 U/mg for 2-keto-butanoate, 1.2 U/mg for 2-keto-pentanoate, and 0.2 U/mg for 2-keto-4-methylpentanoate. No activity was recorded on 3-phenyl-2-oxopropanoate, benzoyl formate, 4-hydroxy-phenyl pyruvate, and indole-3-pyruvate.

Native *pdc* (*GoxPDC*WT) expression in *E. coli* and *G. thermoglucosidasius*

The expression of *GoxPDC*WT in *E. coli* produced 0.5 ± 0.005 g/g ethanol per gram of glucose consumed under fermentative conditions, substantially higher than the control (DH5α-pTMO111) which produced only 0.1 ± 0.01 g/g. Cell densities were of the same order of magnitude, demonstrating that the higher ethanol concentrations were not simply the result of higher biomass levels in *GoxPDC*-positive strains. It is noted that these elevated ethanol yields were achieved in the absence of a recombinant *adh II*, which has previously been shown to be essential for enhanced ethanol production in *E. coli* strains expressing *ZmoPDC* (Lawford et al. 1991). This constitutes the first report demonstrating ethanol production as a direct result of the functional expression of only a *pdc* gene in *E. coli* (Liu et al. 2005; Talarico et al. 2005; Lowe et al. 1992).

Expression in *G. thermoglucosidasius* under microaerobic conditions did not result in ethanol production more than the control strain (Fig. 5). RT-PCR confirmed that the gene was transcribed (data not shown); however, no soluble protein could be detected by Western blotting for cultures grown at 45 °C (Fig. 6) and no PDC activity was detected using cell free extracts from the same cultures. Together, these data suggest a failure at the level of translation, possibly with the generation of misfolded protein which would be targeted for intracellular proteolysis.

Assessing codon usage and predicting a gene sequence for harmonization of *GoxPDC* expression in *G. thermoglucosidasius*

A relatively new concept, termed codon harmonization (Angov et al. 2008, 2011a, b), was proposed as a technique to assist protein folding during heterologous expression. Co-don harmonization, as opposed to other codon optimization strategies, involves mimicking the

translation rates of the native host in a heterologous strain. The translation rate is predominantly determined by codon usage frequencies, where the presence of infrequently used codons forces a reduction of the translation rate, allowing the protein being translated to fold in phases. Both the frequency and positioning of infrequently used codons is critical for protein folding (Clarke et al. 2008).

An analysis of the codon usage frequencies in GoxPDC for expression in *G. thermoglucosidasius*, *E. coli*, and *G. oxydans* was conducted (Table 3, Fig. 7). There are five codons which are rarely used in *G. oxydans*, the native host, one of which (CGA) is also rarely used in *G. thermoglucosidasius*. However, in *G. thermoglucosidasius*, 20 codons in the GoxPDC gene (CCC and CTC) are recognized as rare, which are frequently used (>20 %) in the native host. It is thought that the inclusion of so many rare codons when expressing the gene in *G. thermoglucosidasius*, is detrimental to correct folding of the protein (Kane et al. 1995; Kim et al. 2006; Rosano et al. 2009). Similarly, in *E. coli*, the CTC and CGA codons are also rare. However, unlike in *G. thermoglucosidasius*, the CCC codon is infrequently used (Table 3). Based on the rare codon analysis, *E. coli* is expected to be a more suitable expression host for the WT GoxPDC relative to *G. thermoglucosidasius*. This was the case experimentally; however, the temperature had to be reduced (25 °C) to enable soluble expression of the GoxPDC protein in *E. coli*.

In order to codon harmonize the GoxPDC gene for *G. thermoglucosidasius* expression, 416 bp changes (348 codon substitutions) were made to the wild-type gene sequence so as to match the usage frequencies found in the native host for every codon position while maintaining the amino acid composition of the wild-type protein. Figure 7 shows a 20 amino acid section of the protein to demonstrate how the harmonization was performed.

This harmonization strategy was chosen to demonstrate whether the naturally evolved translation frequency alone would enable correct folding of the protein in *G. thermoglucosidasius*, without the need to calculate link/end segments (Angov et al. 2008; Thanaraj et al. 1996).

Fig. 2 A denaturing SDS-PAGE gel showing purified GoxPDC protein. 1 Molecular weight marker (#SM0671), 2 HIS-tag purified GoxPDC protein. The GoxPDC protein was approximately 59 kDa in size

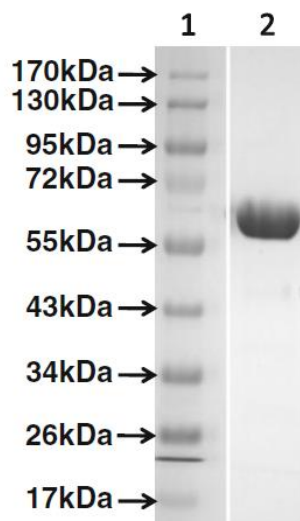


Table 2 Steady state kinetic constants for GoxPDC compared with those from other Gram negative bacteria. Values given in brackets indicate the pH at which measurements were made

PDC	K_M (mM)	Specific activity in (U/mg)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1} s^{-1}$)	T_{opt} ($^{\circ}C$)	pH _{opt}	$T_{1/2}$ at $^{\circ}C$
GoxPDC _R	0.12 (5.0)±0.005	57 (5.0)	57 (5.0)	4.75×10^5 (5.0)	53	4.5–5.0	10 min at 65 $^{\circ}C$
	1.2 (6.5)±0.2	47 (6.5)	47 (6.5)	3.6×10^4 (6.5)			
	2.8 (7.0)±0.4	125 (7.0)	125 (7.0)	4.2×10^4 (7.0)			
GoxPDC _N	0.74 (6.0)	4.4 (6.0) ^g	N/A	N/A	N/A	6.0	^f
ApaPDC	2.8 (6.5) ^a /0.39 (5.0) ^e	110 (6.5) ^a /97 (5.0) ^e	341–508 ^e	N/A	65 ^a	3.5–6.5 ^a	24 min at 70 $^{\circ}C$ ^a
ZpaPDC	2.5 (6.5) ^a /0.24 (6.0) ^e	116 (6.5) ^a /130 (6.0) ^e	341–508 ^e	N/A	55 ^a	7.0 ^a	24 min at 60 $^{\circ}C$ ^a
ZmoPDC	1.3 (6.5) ^a /0.31 (6.0) ^b /1.1 ^c	120 (6.5) ^a /120 ^c	150 (6.0) ^b /486 (6.5) ^c	4.8×10^5 (6.0) ^b / 4.4×10^5 (6.5) ^c	60 ^a	6.0–6.5 ^a	30 min at 60 $^{\circ}C$ ^a
SvePDC	13 ^d	103 ^d	412 ^d	3.2×10^{4d} / 0.87×10^4 (7.0)	N/A	6.3–6.7 ^d	30 min at 50 $^{\circ}C$

For all experiments done on GoxPDC_R, $n \geq 3$. Values in brackets indicate pH

^N native, ^R recombinant, ^f stable for 5 min at 55 $^{\circ}C$ but completely inactivated by incubating for 3 min at 80 $^{\circ}C$, ^g crude extract

^aGocke et al. (2009)

^bMeyer et al. (2010)

^cSiegert et al. (2005)

^dLowe et al. (1992)

^eRaj et al. (2002)

Expression of GoxPDCOPT in *G. thermoglucosidasius*

Comparative expression of the wild-type (GoxPDCWT) and codon harmonized (GoxPDCOPT) GoxPDC in *G. thermoglucosidasius* TM89 was evaluated at 45 and 52 $^{\circ}C$ (Fig. 5). Expression of GoxPDCOPT at 45 $^{\circ}C$ produced 0.35 ± 0.04 ethanol per gram of glucose consumed, compared to 0.26 ± 0.04 g/g for TM89 alone or 0.24 ± 0.02 g/g for GoxPDCWT (Fig. 5). This result clearly demonstrates that translational discord was, at least in part, a significant limitation to the functional expression of GoxPDC in *G. thermoglucosidasius*. However, at 52 $^{\circ}C$ strains expressing either GoxPDCWT (0.26 ± 0.01 g/g) or GoxPDCOPT (0.25 ± 0.04 g/g) produced lower ethanol yields than the control TM89 (0.32 ± 0.05 g/g; Fig. 5). We speculate that the reduced performance of TM89 when expressing either the unmodified or codon-harmonized GoxPDC at 52 $^{\circ}C$ may be due to the metabolic load

imposed by the presence of the shuttle vector and/or the metabolic load imposed by misfolded PDC protein (proteotoxic stress).

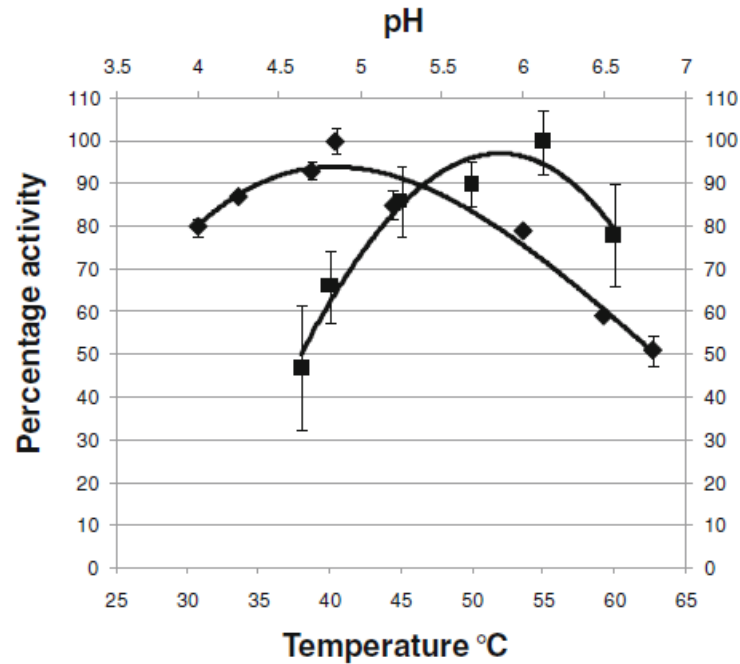


Fig. 3 Effect of pH and temperature (in degree Celsius) on the activity of GoxPDC (black diamond) when using pyruvate as substrate. For all data points, $n \geq 3$. The assay buffer used was 100 mM Na_2HPO_4 /citrate buffer. The 100 % activity was analogous to a specific activity of 72 U/mg for T_{opt} and 160 U/mg for pH_{opt}

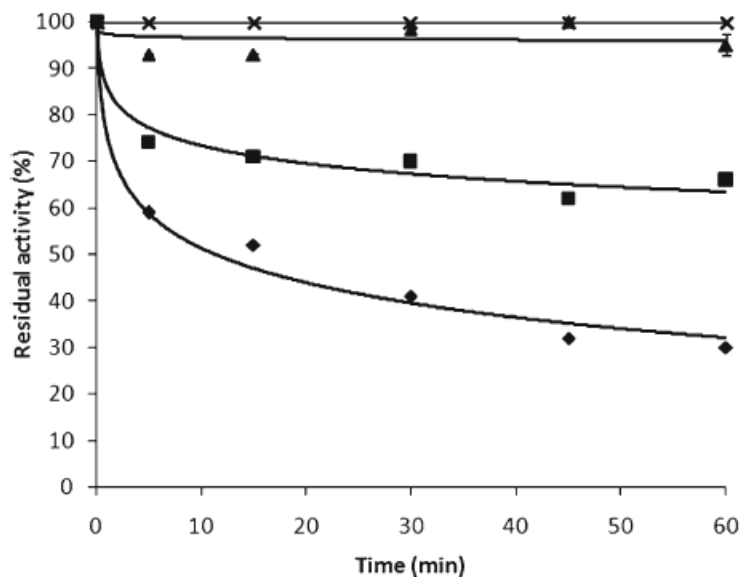


Fig. 4 Thermal inactivation profile of GoxPDC at 25 °C (multiplication symbol), 55 °C (black triangle), 60 °C (black square), and 65 °C (in black diamond). Activity is expressed as a percentage of that at time zero in the standard assay at 25 °C. The activity at 100 % correlates to a specific activity of 26 U/mg. Assays were performed in 50 mM MES buffer at pH6.5

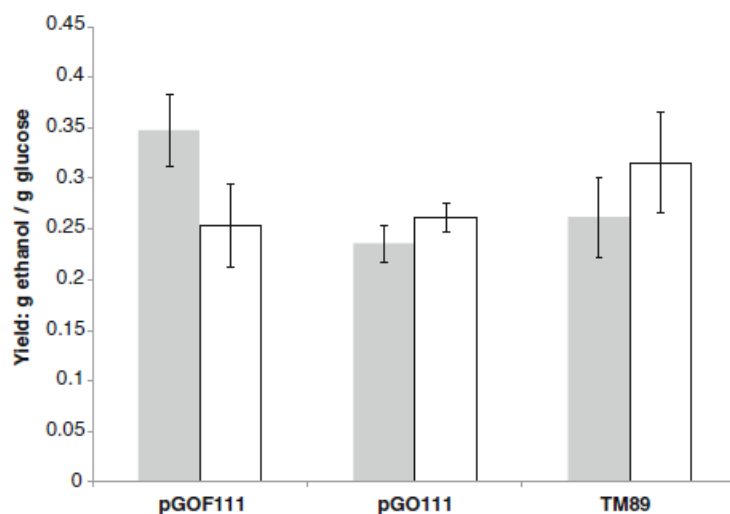


Fig. 5 Comparison of the yield of ethanol produced per gram of glucose consumed during 10/15 model fermentations in *G. thermoglucosidasius*. Grey bars represent fermentations performed at 45 °C and white bars 52 °C. The data represents an average of three independent fermentations (batches of medium) with three repeats of each culture in each fermentations after 48H at the given temperature

Intracellular PDC activity was assayed from GoxPDCOPT–*G. thermoglucosidasius* cultures grown at 45 °C and 52 °C. A specific activity of 0.22 U/mg was determined for GoxPDCOPT cultures grown at 45 °C, consistent with the observed phenotype. No detectable activity could be demonstrated for the control and GoxPDCWT cultures under similar conditions. No PDC activity was detectable in GoxPDCOPT–*G. thermoglucosidasius* cultures grown at 52 °C. Furthermore, through Western blot analysis, a GoxPDC signal was only detectable in soluble protein extracts from *G. thermoglucosidasius* cultures expressing GoxPDCOPT at 45 °C (Fig. 6). Extracts from cells grown at 52 °C (GoxPDCWT, GoxPDCOPT, and control cultures) were reproducibly negative. Given that in vitro thermostability does not necessarily equate to in vivo stability, the lack of improved ethanol production at 52 °C is not unexpected. These results, taken together, indicate that although codon harmonization enhanced the expression of GoxPDC in *G. thermoglucosidasius* TM89 at 45 °C, a second limitation, probably thermally related, prevents functional enzyme accumulation at higher temperatures. The lack of detectable protein by Western blotting either indicates that a second limitation is present at the transcription– translation interface or that the protein produced is extremely unstable and the resultant aggregates or proteolysis products are not suitable for antibody binding.

Nevertheless, the increased ethanol yield at 45 °C represents a significant improvement in comparison to other engineered thermophiles (Cripps et al. 2009; Shaw et al. 2008). It has been demonstrated that effective partitioning of carbon into biosynthesis and fermentation is critical in achieving optimal production of ethanol under fermentative conditions (Underwood et al. 2002b). Typically, *G. thermoglucosidasius* TM89

fermentations are characterized by the production of formate (average of 40.5 ± 7.2 mM after 48 h at 45 °C) and acetate with ethanol (Cripps et al. 2009). However, it was noted that for TM89-GoxPDC_{OPT} fermentations at 45 °C, no formate was produced but low level (± 185 μ M) fumarate accumulation was detected in culture supernatants. This may be due to a reduced metabolic flux through pfl where the active GoxPDC_{OPT} may outcompete the pfl enzyme for pyruvate (Orencio-Trejo et al. 2008; Tolan et al. 1987; Feldmann et al. 1989; Underwood et al. 2002a).

Discussion

The pyruvate decarboxylase from the acetic acid bacterium *G. oxydans* has been described in this study. The substrate recognition and decarboxylation range of the enzyme is similar to that of the other four Gram-negative PDCs identified to date, showing a preference for short-chain aliphatic 2-keto acids (Gocke et al. 2009). The value of k_{cat}/K_M for pyruvate compared to those for 2-ketobutanoate and 2-ketopentanoate, the nearest analogs, and the retention of Ile468, thought to be involved in substrate specificity, suggests that this enzyme favors pyruvate as its physiological substrate (Pohl et al. 1998; Gocke et al. 2009). The GoxPDC kinetics are also similar to the other Gram-negative bacterial PDCs, displaying the same pH dependent increase in k_{cat}/K_M while catalytic efficiency (k_{cat}) remaining largely unchanged.

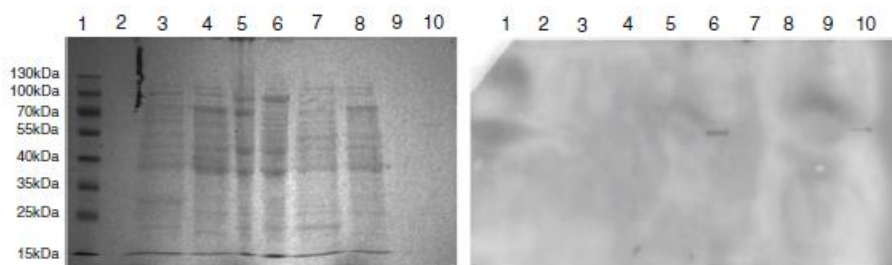


Fig. 6 SDS-PAGE and Western blots of cell extracts from *G. thermoglucosidasius* TM89, containing either empty vector pTMO111 (5 and 8) or pGO111 (GoxPDC_{WT}; 4 and 7), or pGOF111 (GoxPDC_{OPT}; 3 and 6). Cultures were grown at either 45 °C (6, 7, and 8) or 52 °C (3, 4, and 5), respectively. Purified GoxPDC protein served as positive control (10). Lanes 2 and 9 are empty

Table 3 The correspondence of rare codons (<10 % usage) for PDCs which have been expressed in *G. thermoglucosidasius* and in *E. coli*, between their native host and *G. thermoglucosidasius*

	All codons with <10 % usage in the native host (number of codons in respective <i>pdc</i> genes)	All codons with <10 % usage in <i>G. thermoglucosidasius</i> (amino acid position)	All codons with <10 % usage in <i>E. coli</i> (amino acid position)	% codon usage for selected codons in their native host
ZmoPDC	CTC (11)	CTC (2, 18, 30, 56, 164, 174, 215, 236, 306, 320, 348, 362, 400, 436, 567, 568), CCC (174, 320), AGT (2, 56, 362)	CGG (12), CTC (18, 30, 164, 215, 236, 306, 348, 400, 436, 567, 568, 569)	CCC 10–20 %, AGT 10–20 %
ZpaPDC	CTA (4), TTG (3), ACA (3), CTC (3), GTG (2), CTT (2), GCG (2), TCA (2), CCC (1), AGT (1), GGA (1), TCG (1), AAG (1), CCA (1)	CTC (35, 187, 517), CCC (356), CTA (12, 155, 231, 509), AGT (193)	CTC (35, 187, 517), CTA (12, 155, 231, 509),	N/A
GoxPDC	ACT (2), CGA (1), TCT (1), TCA (1)	CGA (12), CTC (30, 33, 84, 95, 164, 206, 226, 305, 431, 505, 531, 536, 545), CCC (174, 239, 251, 255, 309, 348, 396)	CGA (12), CTC (30, 33, 84, 95, 164, 206, 226, 305, 431, 505, 531, 536, 545)	CTC >20 %, CCC >20 %

N/A not applicable

Despite being an elusive enzyme in the bacterial kingdom, PDCs are sought after enzymes for the generation of ethanologens, and the engineering of a PDC-expressing pathway in fermentative bacteria is now a well-established procedure to achieve increased ethanol yields from mesophilic organisms (Taylor et al. 2008, Thompson et al. 2008, Talarico et al. 2005, Bi et al. 2009, Tolan et al. 1987, Ingram et al. 1987, Correa et al. 2011). In this study, expression of GoxPDC in *E. coli* resulted in a fivefold ethanol production increase, with a final yield of 0.5 g/g, similar to the best results reported for other recombinant *E. coli* strains, and moreover, without the co-expression of a heterologous *adh* (Ohta et al. 1991).

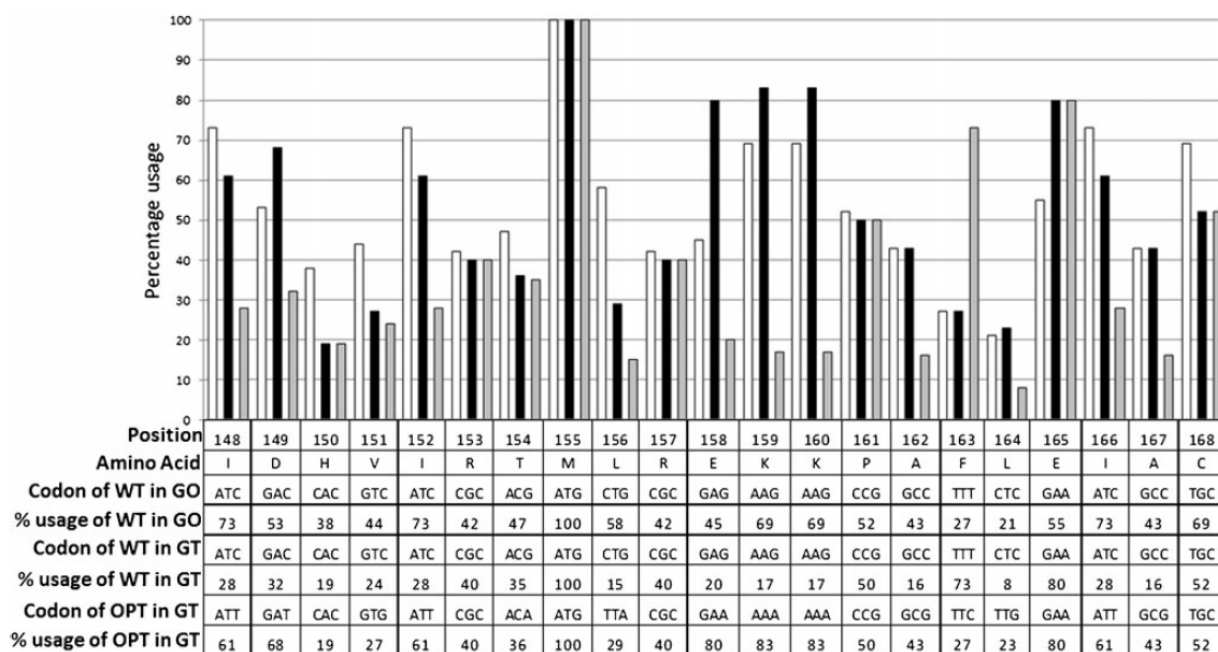


Fig. 7 A comparison of codon usage frequencies of GoxPDC in the native host (*G. oxydans*—white bars), the heterologous host (*G. thermoglucosidarius*—grey bars), and after harmonization (black bars), for the amino acid residues 148–168. Disparities exist between the percent usage of codons in *G. oxydans* for the wild-type PDC and those used in *G. thermoglucosidarius* (large differences between the green and

red bars). The rarely used codons in *G. thermoglucosidarius* (<10%), such as amino acid 164, would represent likely loci where a hiatus would be reached in translation of the gene in this strain. This does not correlate with a locus of low codon usage in *G. oxydans*. Post harmonization, it can be seen that the black and white bars are closer to each other in terms of percentage codon usage

In *G. thermoglucosidarius* expression of the codon, harmonized GoxPDC resulted in a significant improvement in ethanol production. This represents the first account of in vivo PDC-mediated ethanol production in a Gram positive organism, and in a thermophile. While expression of several PDCs has been demonstrated in *Geobacillus* and *Bacillus megaterium*, none of these studies could demonstrate increased cellular ethanol production (Taylor et al. 2008, Thompson et al. 2008, Talarico et al. 2001). Considering that *G. thermoglucosidarius* is currently employed in a commercially viable bioethanol technology, this finding could represent a significant advancement in the engineering of a thermophilic ethanologen. The upregulation of pyruvate dehydrogenase and subsequent conversion of acetyl-CoA to acetaldehyde is stoichiometrically equivalent to a PDC intervention, and is the engineering strategy employed for the bioethanol process strain *G. thermoglucosidarius* TM242 (Cripps et al. 2009). It would be of value therefore for future studies to compare the performance of PDC- vs PDH-engineered *G. thermoglucosidarius*

in a pfl-negative background, in order to maximize the flux of pyruvate to ethanol. However, despite the significance of these findings, PDC-mediated ethanol production in this study was only possible at a maximum temperature of 48 °C, and therefore the lack of PDC thermostability continues to limit the viability of the PDC route for a thermophilic commercial process.

It has long been known that codon usage differences during heterologous protein expression can result in low expression or formation of insoluble aggregates. The tendency has been to replace the rare codons in the protein of interest for codons used more frequently by the expression host. However, only recently has the suggestion been made that the frequency and positioning of infrequently used codons is critical for protein folding, and that the standard codon optimization approach is flawed. Instead, the principal of codon harmonization (Angov et al. 2008) involving the substitution of synonymous codons from the heterologous host such that the codon usage frequency, positioning, and therefore rhythm of translation follows that of the native host, was recently proposed. Codon harmonization has been exclusively applied to expression in *E. coli*, therefore this study provides further evidence that codon harmonization may provide a general strategy for improving the expression of soluble, functional proteins in a wide range of bacterial hosts. Precedence for this already exists if one analyses the expression of two other Gram-negative PDCs (ZmoPDC and ZpaPDC) in *G. thermoglucosidasius*. Production of soluble ZmoPDC in cell free extracts of *G. thermoglucosidasius* grown at 52, 54, 56, and 58 °C was observed to decrease with increased temperature, and PDC activity was undetectable above 52 °C (Thompson et al. 2008). For ZpaPDC, activity was absent at growth temperatures above 45 °C (Taylor et al. 2008). An analysis of the codon usage pattern of these PDCs expressed in *G. thermoglucosidasius* revealed that they have a higher coincidence of both frequency-of-usage and positioning of rare codons for expression in *G. thermoglucosidasius* (Table 3), which does not represent a large deviation from the frequency in the native host. Therefore, this provides further support for the codon harmonization concept, and this correlation may be responsible for the reported variations in PDC-expression efficiency in *G. thermoglucosidasius*: ZmoPDC > ZpaPDC > GoxPDC.

Despite the improvement in GoxPDC expression as a result of codon harmonization, there are still other factors which continue to play a major role in the functionality of the enzyme at higher growth temperatures. When correctly folded the GoxPDC protein displays relatively high thermostability when assayed *in vitro*. However, this does not necessarily translate to the ability to fold correctly at elevated temperatures, offering a possible explanation for the apparent failure of functional expression at 52 °C, unlike ZmoPDC when expressed in the same host (Thompson et al. 2008). We suggest that although codon harmonization contributes to the correct folding of a nascent protein during translation of the mRNA, it cannot necessarily compensate for the kinetics involved in protein folding in temperature ranges outside those for which the protein had been selected for or evolved under. The further stabilization of the enzyme therefore represents an area of future improvement for the use of PDC in engineering superior homo-ethanolic pathways in *G. thermoglucosidasius*.

The significant role that codon harmonization played in the correct processing of GoxPDC protein when expressed in *G. thermoglucosidasius* serves to reiterate the importance of codon usage in heterologous protein expression. This study represents the first account of improved expression of a protein of mesophilic origin in a thermophilic host using this technique, and demonstrates the potential benefits for microbial biotechnology. Due to its metabolic versatility, *Geobacillus* is a suitable platform organism for the synthesis of additional industrial products. In this light, codon harmonization should play a pivotal role in enabling and improving its development and in expanding its biotechnological repertoire.

Acknowledgments

The *G. oxydans* strain was a kind gift from the Institute for Wine Biotechnology (IWBT) at the University of Stellenbosch. This work was supported by the National Research Foundation of South Africa.

References

- Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J* 66:650–659
- Angov E, Hillier CJ, Kincaid RL, Lyon JA (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE* 3:e2189
- Angov E, Legler PM, Mease RM (2011) Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods Mol Biol* 705:1–13
- Bi C, Zhang X, Ingram LO, Preston JF (2009) Genetic engineering of *Enterobacter asburiae* strain JDR-1 for efficient production of ethanol from hemicellulose hydrolysates. *Appl Environ Microbiol* 75:5743–5749
- Bongers RS, Hoefnagel MH, Kleerebezem M (2005) High-level acetaldehyde production in *Lactococcus lactis* by metabolic engineering. *Appl Environ Microbiol* 71:1109–1113
- Clarke TF IV, Clark PL (2008) Rare codons cluster. *PLoS ONE* 3:e3412
- Conway T, Osman YA, Konnan JI, Hoffmann EM, Ingram LO (1987) Promoter and nucleotide sequences of the *Zymomonas mobilis* pyruvate decarboxylase. *J Bacteriol* 169:949–954
- Correa A, Oppezzo P (2011) Tuning different expression parameters to achieve soluble recombinant proteins in *E. coli*: advantages of high-throughput screening. *Biotechnol J* 66:715–730
- Cripps RE, Eley K, Leak DJ, Rudd B, Taylor M, Todd M, Boakes S, Martin S, Atkinson T (2009) Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metab Eng* 11:398–408
- Feldmann S, Sprenger GA, Sham H (1989) Ethanol production from xylose with a pyruvate-formate-lyase mutant of *Klebsiella planticola* carrying a pyruvate-decarboxylase gene from *Zymomonas mobilis*. *Appl Microbiol Biotechnol* 31:152–157
- Gocke D, Graf T, Brosi H, Frindi-Wosch I, Walter L (2009) Comparative characterisation of thiamin diphosphate-dependent decarboxylases. *J Mol Catal B Enzym* 61:30–35
- Gold RS, Meagher MM, Tong S, Hutkins RW, Conway T (1996) Cloning and expression of the *Zymomonas mobilis* “production of ethanol” genes in *Lactobacillus casei*. *Curr Microbiol* 33:256–260
- Gupta A, Singh VK, Qazi GN, Kumar A (2001) *Gluconobacter oxydans*: its biotechnological applications. *J Mol Microbiol Biotechnol* 33: 445–456
- Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol* 22:346–353
- Ingram LO, Conway T, Clark DP, Sewell GW, Preston JF (1987) Genetic engineering of ethanol production in *Escherichia coli*. *Appl Environ Microbiol* 53:2420–2425
- Kaczowka SJ, Reuter CJ, Talarico LA, Maupin-Furlow JA (2005) Recombinant production of *Zymomonas mobilis* pyruvate decarboxylase in the haloarchaeon *Haloferax volcanii*. *Archaea* 15: 327–334
- Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol* 65: 494–500
- Kim S, Lee SB (2006) Rare codon clusters at 5'-end influence heterologous expression of archaeal gene in *Escherichia coli*. *Protein Expr Purif* 50:49–57

- King TE, Cheldelin VH (1954) Pyruvic carboxylase of *Acetobacter suboxydans*. *J Biol Chem* 208:821–831
- Konig S (1998) Subunit structure, function and organisation of pyruvate decarboxylases from various organisms. *Biochim Biophys Acta* 1385:271–286
- Konig S, Spinka M, Kutter S (2009) Allosteric activation of pyruvate decarboxylases. A never-ending story? *J Mol Catal B Enzym* 61: 100–110
- Kotze AA, Tuffin IM, Deane SM, Rawlings DE (2006) Cloning and characterization of the chromosomal arsenic resistance genes from *Acidithiobacillus caldus* and enhanced arsenic resistance on conjugal transfer of ars genes located on transposon TnAtcArs. *Microbiology* 152:3551–3560
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685
- Lawford HG, Rousseau JD (1991) Ethanol production by recombinant *Escherichia coli* carrying genes from *Zymomonas mobilis*. *Appl Biochem Biotechnol* 28–29:221–236
- Liu S, Dien BS, Cotta MA (2005) Functional expression of bacterial *Zymobacter palmae* pyruvate decarboxylase gene in *Lactococcus lactis*. *Curr Microbiol* 50:324–328
- Liu S, Nichols NN, Dien BS, Cotta MA (2006) Metabolic engineering of a *Lactobacillus plantarum* double ldh knockout strain for enhanced ethanol production. *J Ind Microbiol Biotechnol* 33:1–7
- Liu S, Dien BS, Nichols NN, Bischoff KM, Hughes SR, Cotta MA (2007) Coexpression of pyruvate decarboxylase and alcohol dehydrogenase genes in *Lactobacillus brevis*. *FEMS Microbiol Lett* 274: 291–297
- Lowe SE, Zeikus JG (1992) Purification and characterization of pyruvate decarboxylase from *Sarcina ventriculi*. *J Gen Microbiol* 138:803– 807
- Lu G, Dobritsch D, Baumann S, Schneider G, Konig S (2000) The structural basis of substrate activation in yeast pyruvate decarboxylase. A crystallographic and kinetic study. *Eur J Biochem* 267:861– 868
- Meyer D, Neumann P, Parthier C, Friedemann R, Nemeria N, Jordan F, Tittmann K (2010) Double duty for a conserved glutamate in pyruvate decarboxylase: evidence of the participation in stereoelectronically controlled decarboxylation and in protonation of the nascent carbanion/enamine intermediate. *Biochemistry* 49: 8197–8212
- Ohta K, Beall DS, Mejia JP, Shanmugam KT, Ingram LO (1991) Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of *Zymomonas mobilis* genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl Environ Microbiol* 57:893–900
- Orencio-Trejo M, Flores N, Escalante A, Hernandez-Chavez G, Bolivar F, Gosset G, Martinez A (2008) Metabolic regulation analysis of an ethanologenic *Escherichia coli* strain based on RT-PCR and enzymatic activities. *Biotechnol Biofuels* 1:8
- Peters B, Junker A, Brauer K, Mühlthaler B, Kostner D, Mientus M, Liebl W, Ehrenreich A (2012) Deletion of pyruvate decarboxylase by a new method for efficient markerless gene deletions in *Gluconobacter oxydans*. *Appl Microbiol Biotechnol* 97:2521–2530
- Pohl M, Mesch K, Rodenbrock A, Kula MR (1995) Stability investigations on the pyruvate decarboxylase from *Zymomonas mobilis*. *Biotechnol Appl Biochem* 22:95–105

- Pohl M, Siegert P, Mesch K, Bruhn H, Grotzinger J (1998) Active site mutants of pyruvate decarboxylase from *Zymomonas mobilis*—a site-directed mutagenesis study of L112, I472, I476, E473, and N482. *Eur J Biochem* 257:538–546
- Prust C, Hoffmeister M, Liesegang H, Wiezer A, Fricke WF, Ehrenreich A, Gottschalk G, Deppenmeier U (2005) Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. *Nat Biotechnol* 23:195–200
- Raj KC, Ingram LO, Maupin-Furlow JA (2001) Pyruvate decarboxylase: a key enzyme for the oxidative metabolism of lactic acid by *Acetobacter pasteurianus*. *Arch Microbiol* 176:443–451
- Raj KC, Talarico LA, Ingram LO, Maupin-Furlow JA (2002) Cloning and characterization of the *Zymobacter palmae* pyruvate decarboxylase gene (*pdc*) and comparison to bacterial homologues. *Appl Environ Microbiol* 68:2869–2876
- Rosano GL, Ceccarelli EA (2009) Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb Cell Fact* 8:41
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbour Press, Cold Spring Harbor, NY
- 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* 105:13769–13774
- Siegert P, McLeish MJ, Baumann M, Iding H, Kneen MM, Kenyon GL, Pohl M (2005) Exchanging the substrate specificities of pyruvate decarboxylase from *Zymomonas mobilis* and benzoylformate decarboxylase from *Pseudomonas putida*. *Protein Eng Des Sel* 18:345–357
- Talarico LA, Ingram LO, Maupin-Furlow JA (2001) Production of the Gram-positive *Sarcina ventriculi* pyruvate decarboxylase in *Escherichia coli*. *Microbiology* 147:2425–2435
- Talarico LA, Gil MA, Yomano LP, Ingram LO, Maupin-Furlow JA (2005) Construction and expression of an ethanol production operon in Gram-positive bacteria. *Microbiology* 151:4023–4031
- Taylor MP, Esteban CD, Leak DJ (2008) Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. *Plasmid* 60: 45–52
- Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, Cowan DA (2009) Thermophilic ethanologeneses: future prospects for second-generation bioethanol production. *Trends Biotechnol* 27:398–405
- Thanaraj TA, Argos P (1996) Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci* 5:1973–1983
- Thompson AH, Studholme DJ, Green EM, Leak DJ (2008) Heterologous expression of pyruvate decarboxylase in *Geobacillus thermoglucosidasius*. *Biotechnol Lett* 30:1359–1365
- Tolan JS, Finn RK (1987) Fermentation of D-xylose to ethanol by genetically modified *Klebsiella planticola*. *Appl Environ Microbiol* 53:2039–2044
- Underwood SA, Buszko ML, Shanmugam KT, Ingram LO (2002a) Flux through citrate synthase limits the growth of ethanologenic *Escherichia coli* KO11 during xylose fermentation. *Appl Environ Microbiol* 68:1071–1081

- Underwood SA, Zhou S, Causey TB, Yomano LP, Shanmugam KT, Ingram LO (2002b) Genetic changes to optimize carbon partitioning between ethanol and biosynthesis in ethanologenic *Escherichia coli*. *Appl Environ Microbiol* 68:6263–6272
- Vuralhan Z, Luttik MA, Tai SL, Boer VM, Morais MA, Schipper D, Almering MJ, Kotter P, Dickinson JR, Daran JM, Pronk JT (2005) Physiological characterization of the ARO10-dependent, broad- substrate-specificity 2-oxo acid decarboxylase activity of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 71:3276–3284