2-Oxoacid dehydrogenase multienzyme complexes in the halophilic Archaea? Gene sequences and protein structural predictions

Keith A. Jolley, Deborah G. Maddocks, Shan L. Gyles, Zoe Mullan, Sen-Lin Tang, Michael L. Dyall-Smith, David W. Hough and Michael J. Danson

Author for correspondence: Michael J. Danson. Tel: +44 1225 826509. Fax: +44 1225 826779. e-mail: M.J.Danson@bath.ac.uk

All Archaea catalyse the conversion of pyruvate to acetyl-CoA via a simple pyruvate oxidoreductase. This is in contrast to the Eukarya and most aerobic bacteria, which use the pyruvate dehydrogenase multienzyme complex [PDHC], consisting of multiple copies of three component enzymes: E1 (pyruvate decarboxylase), E2 (lipoate acetyl-transferase) and E3 (dihydrolipoamide dehydrogenase, DHLipDH). Until now no PDHC activity has been found in the Archaea, although DHLipDH has been discovered in the extremely halophilic Archaea and its gene sequence has been determined. In this paper, the discovery and sequencing of an operon containing the DHLipDH gene in the halophilic archaeon *Haloferax volcanii* are reported. Upstream of the DHLipDH gene are 3 ORFs which show highest sequence identities with the E1α, E1β and E2 genes of the PDHC from Gram-positive organisms. Structural predictions of the proposed protein product of the E2 gene show a domain structure characteristic of the E2 component in PDHCs, and catalytically important residues, including the lysine to which the lipoic acid cofactor is covalently bound, are conserved. Northern analyses indicate the transcription of the whole operon, but no PDHC enzymic activity could be detected in cell extracts. The presence in the E2 gene of an insertion (equivalent to approximately 100 aa) not found in bacterial or eukaryal E2 proteins, might be predicted to prevent multienzyme complex assembly. This is the first detailed report of the genes for a putative 2-oxoacid dehydrogenase complex in the Archaea, and the evolutionary and metabolic consequences of these findings are discussed.

Keywords: halophile, central metabolism, pyruvate dehydrogenase complex, dihydrolipoamide dehydrogenase

INTRODUCTION

The 2-oxoacid dehydrogenase complexes [ODHCs: the pyruvate dehydrogenase complex (PDHC), the 2-oxoglutarate dehydrogenase complex (OGDHC) and the branched-chain 2-oxoacid dehydrogenase complex (BCODHC)] are multienzyme systems catalysing the general reaction:

\[
2\text{-oxoacid} + \text{CoA} + \text{NAD}^+ \rightarrow \text{acyl-CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+
\]

They are all three component systems consisting of multiple copies of enzymes E1 (2-oxoacid decarboxylase), E2 (lipoate acetyl-transferase) and E3 (dihydrolipoamide dehydrogenase, DHLipDH). Whilst E1 and E2 are specific to the individual complexes, the E3 component is the same gene product in each complex and catalyses an identical reaction (Perham, 1991, 1996). Central to the catalytic mechanism of these complexes...
(Fig. 1a) is the acyl-carrying cofactor, lipoic acid, which is covalently attached to E2 and serves to connect the three active sites and channel substrate through the complex. In addition to forming the catalytic core of PDHC, the E2 component also comprises the structural core, to which the E1 and E3 components are non-covalently bound.

These complexes are found in the Eukarya and most aerobic bacteria, whereas the Archaea possess simpler 2-oxoacid oxidoreductases [reviewed by Danson (1993) and references therein]. In the halophilic Archaea, the pyruvate and 2-oxoglutarate oxidoreductases have an αβ2 structure (Plaga et al., 1992) and the catalytic mechanism does not involve the participation of a lipoic acid moiety (Fig. 1b). Instead, the hydroxyethyl group is transferred directly from thiamine pyrophosphate (TPP) to Coenzyme-A and the reducing equivalents are passed via an iron–sulphur centre to ferredoxin (Kerscher et al., 1981).

No 2-oxoacid dehydrogenase activity has ever been found in the Archaea. It was therefore a surprise to discover the presence of DHLipDH and its substrate, lipoic acid, in the halophilic Archaea (Danson et al., 1984; Pratt et al., 1989). The only known function of DHLipDH is as the third enzyme component of the ODHCs along with the glycine cleavage system, the activity of which has also not been reported in the Archaea (reviewed by Danson, 1988, 1993). The gene encoding DHLipDH from Haloferax volcanii has been cloned and sequenced (Vettakkorumakan & Stevenson, 1992), and from sequence alignments it is clearly related to the DHLipDHs from bacterial and eukaryal ODHCs. However, consistent with these complex activities not being present in this organism, a DHLipDH-negative mutant of H. volcanii was unaffected in its ability to grow on a variety of carbon sources that would require metabolism of pyruvate, 2-oxoglutarate and glycine (Jolley et al., 1996). Neither wild-type nor mutant strains grew on the branched-chain amino acids.

The function of DHLipDH in the halophilic Archaea is therefore still a mystery. However, in this paper we report the sequencing of the region upstream of the DHLipDH gene in H. volcanii and find that the gene appears to be the fourth ORF in an operon containing other genes whose putative protein products are homologous to the E1x, E1β and E2 components of the PDHC from Gram-positive bacteria. The domain structure of these proteins is predicted and the conservation of functional amino acids is identified. Finally, we show that the operon is transcribed and discuss possible reasons why this does not lead to a detectable 2-oxoacid dehydrogenase activity.

**METHODS**

**Materials.** Restriction endonucleases, Vent DNA polymerase and bovine serum albumin were obtained from New England Biolabs. [γ³²-P]dCTP was supplied by Amersham. The λ DNA preparation kit was from Qiagen, DNA primers were synthesized by PE- Applied Biosystems and the Random Primed DNA Labelling Kit was from Boehringer Mannheim.

**Strains and plasmids.** H. volcanii (WFD 11) was grown at 37 °C in 18% salt-water modified growth medium (SW-MGM), consisting of 14.4% NaCl, 1.8% MgCl₂, 2H₂O, 2.1% MgSO₄, 7H₂O, 0.42% KCl, 0.5% peptone and 0.1% yeast extract, pH 7.2 (Nuttall & Dyall-Smith, 1993). _Escherichia coli_ XL-1 Blue MRA(P2) and the _βEMBL3_ replacement vector were obtained from Stratagene. Plasmid pNAT82, which contains the cloned DHLipDH gene from _H. volcanii_, was a kind gift of Professor K. J. Stevenson (University of Calgary, Alberta, Canada).

**Construction of a genomic DNA library.** _H. volcanii_ genomic DNA was prepared as described by Jolley et al. (1996) and a genomic DNA library was then constructed in _βEMBL3_. Partially digested chromosomal DNA (200 µg), after Sau3A1 treatment at 37 °C for 30 min, was fractionated by density gradient centrifugation using 5–25% NaCl and inserts of approximately 15 kb were pooled and ligated to 1 µg _BamHI_ arms of _βEMBL3_. The mixture was packaged according to the manufacturer’s instructions and _E. coli_ XL-1 Blue MRA(P2) was infected with the packaging mix.

**Construction of the DHLipDH gene replacement mutant.** The wild-type K. _volcanii_ was infected with the packaging mix.
DNA sequencing. Direct sequencing of pNAT82 and selected 3′ DNA clones was performed on a Perkin Elmer ABI Prism 377 DNA Sequencer using gene-specific primers.

Sequence alignments and secondary structural predictions. Sequence alignments were performed using the program BLAST-P at the European Bioinformatics Server (http://www2.ebi.ac.uk/blast2). Secondary structural predictions were made using the Predict-Protein Server at EMBL (http://www2.ebi.ac.uk/predictprotein), using the program PHDsec (Rost, 1996).

Total RNA isolation and Northern blotting. H. volcanii was cultured aerobically at 37°C in 18% SW-MGM, and anaerobically under N2 in the same medium but including 10 mM NaNO3. At mid-exponential phase (A550 = 0.5–0.8), a 1 ml sample from each culture was quickly removed to a pre-chilled 1.5 ml microcentrifuge tube and placed on ice. The cells were pelleted by centrifugation (11000 × g, 1 min, 4°C), the supernatant carefully removed using a micropipette and the cells resuspended and lysed in 80 µl lysis buffer (25 mM NaOH, 0.5% SDS, 5 mM EDTA, 8% sucrose, 5 µM 1,2-cyclohexanediamine N,N,N′,N′-tetraacetic acid). After incubation at 37°C for 15 min, the tubes were placed on ice for 2 min and then 30 µl pre-cooled sodium acetate (3 M, pH 5.6) was added and the solution vortexed for a few seconds. The lysate was centrifuged (11000 × g, 30 min, 4°C) to remove precipitated proteins and the supernatant removed to a fresh tube. The RNA was precipitated from the supernatant by adding 2 vols ethanol and the subsequent RNA pellet was washed twice with 70% ethanol before being air-dried (30 min, 25°C, under a slight vacuum) and finally dissolved in diethylpyrocarbonate-treated water.

Agarose gel electrophoresis was performed using a modified version of the method described by Goda & Minton (1995). RNA was denatured by adding one-fifth volume of 60 mM sodium phosphate buffer, pH 6.8, containing 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol, 12% SDS, and heating at 75°C for 5 min. Samples were immediately loaded into the wells of a 10 cm long 1% agarose gel and electrophoresed until the bromophenol blue dye had reached within 2 cm of the bottom. The gel was prepared by adding guanidine thiocyanate (final concentration 20 mM) and 0.6 µl ethidium bromide (10 mg ml−1) to 35 ml molten (cooled to 60°C) agarose immediately prior to pouring into a casting tray pre-soaked for 2 h in 1% SDS.

After electrophoresis, the RNA was transferred to a nylon membrane by capillary-elution. The membrane was then prehybridized (30 min, 65°C) in hybridization solution (0.25M NaH2PO4, pH 7.2, 7% SDS) before the radiolabelled DNA probe (the 4.3 kb insert from pNAT82; see Fig. 2) was added and hybridization allowed to proceed at 68°C overnight. The membrane was then rinsed in 0.2×SSC/0.1% SDS at room temperature (Sambrook et al., 1989) before washing (with agitation) at 65°C in 0.2×SSC/0.1% SDS. Membranes were then briefly air-dried at room temperature, covered in plastic wrap and exposed to X-ray film (12–24 h, −70°C, with an intensifier screen).

RESULTS

Sequence of plasmid pNAT82

The gene encoding H. volcanii DHLipDH has previously been cloned as part of a 4.3 kb genomic MboI restriction fragment and incorporated into pBlueScript KS+ to give plasmid pNAT82. The DHLipDH gene (1425 bp) was sequenced from this plasmid and was shown to lie at one end of the cloned fragment in pNAT82 (Vettakkorumakankav & Stevenson, 1992). PCR amplification, subcloning and homologous expression of this gene in the parent organism required the use of a halophilic archaeal rRNA promoter to achieve the production of active enzyme, suggesting that the DHLipDH gene does not have its own promoter immediately upstream of the coding region (Jolley et al., 1996). Moreover, N. N. Vettakkorumakankav and K. J. Stevenson (personal communication) noticed that the 76 upstream bases of their published sequence contained part of an ORF that showed high similarity to the E2 component of bacterial and eukaryal ODHCs.

Therefore, we decided to sequence the whole of the DHLipDH upstream region in pNAT82 and in doing so found two further complete ORFs (1563 and 984 bp, respectively), plus 207 bp at the 3′ end of a third, but incomplete ORF (Fig. 2). The complete nucleotide sequence of pNAT82 is thus 4.25 kb. The proximity of the 4 ORFs in pNAT82 (Fig. 2), with no recognizable promoters preceding ORFs 2, 3 and 4, suggest an operon structure with the DHLipDH gene being the final component. It was therefore necessary to return to genomic DNA to obtain the chromosomal sequence upstream of that cloned into pNAT82.

The complete operon sequence

A genomic DNA library of H. volcanii was constructed in λEMBL3, as described in Methods. To probe this library, a 600 bp region of pNAT82, corresponding to the 3′ end of ORF1 and the 5′ end of ORF2, was PCR-amplified and the product, after purification, was radiolabelled by random priming with [32P]dCTP. In all, 28 positive colonies were identified in the primary screen, two of which (clones 8 and 24) gave the strongest hybridization signals on the secondary screens. Using sequencing primers corresponding to that part of ORF1 identified in pNAT82, it was found that clone 24 contained this ORF, and the complete sequence of this gene, plus the upstream region, was subsequently obtained.

As summarized in Fig. 2, the four ORFs are all in the same direction and are tightly spaced: ORF1 and ORF2 coding regions overlap by 4 nt, with the ATG start codon of ORF2 sharing bases with the ORF2 TGA stop codon; ORF2 and ORF3 are in-frame with each other, separated by three bases; ORF3 and ORF4 (the DHLipDH gene) share a single nucleotide, with the TTA stop codon of ORF3 overlapping the ATG start codon of the DHLipDH gene.

The start codon of ORF1 appears to be GTG. This prediction is based on sequence alignments with homologous genes in the Bacteria and Eukarya (see below), the lack of any nearby in-phase ATG and the presence of a good potential Shine–Dalgarno sequence 8 bp upstream of the GTG codon (i.e. CAGGAGG, which is complementary to the first 7 nt of the 3′ end of H. volcanii 16S rRNA). Furthermore, approximately 30–
40 bp further upstream of the proposed Shine–Dalgarno sequence are two AT-rich sequences, each of which may act as a promoter (Palmer & Daniels, 1995). Interestingly, these two sequences are inverted repeats of each other. Downstream of ORF4 is a good potential transcriptional stop signal (poly-dT tract) (Fig. 2).

The upstream regions of ORFs 2, 3 and 4 do not appear to have any clear promoter motifs, suggesting that the four ORFs may function as an operon. Also, there are no obvious Shine–Dalgarno sites upstream of these ORFs and the tight (mainly overlapping) spacing of start and stop codons could well indicate that the translation of these genes is coupled (Normark et al., 1983). These features are consistent with our previous attempts to express the subcloned DHLipDH gene (ORF4), where expression was only successful when a halophilic rRNA promoter was added upstream (Jolley et al., 1996).

Identification of homologues in the sequence databases

A blast search of the sequence databases showed that the proteins encoded by the four ORFs from H. volcanii are highly similar to the components of the ODHCs from Bacillus steardo thermophilus. The five highest identity scores, with their corresponding P-values, are given in Table 1 and it can be seen that the pyruvate dehydrogenase components were most consistently identified, although the list also included the homologous enzymes from the OGDHC and BCODHC. Significantly, the Haloferax proteins from ORFs 1, 2, 3 and 4 corresponded to the Bacillus E1α, E1β, E2 and E3 (DHLipDH) components, and the gene order is the same in the two genomes.

Predicted structural domains and motifs of the E2 component of PDHC

The E2 component of the ODHCs of the Bacteria and Eukarya serves as both the catalytic centre and the structural core of these multi-enzyme systems. Moreover, it has a number of sequentially arranged structural domains associated with these various functions (Perham, 1991, 1996). Thus, in PDHC, the E2 polypeptide consists of one or more N-terminal lipoyl domains followed by a peripheral E1 and E3 subunit-binding domain and then the C-terminal structural core and acetyltransferase domain. Each of these motifs is connected by extended, flexible linker regions. The number of lipoyl domains depends on the species of origin.

From the sequence alignments, it is predicted that H. volcanii ORF3 encodes an E2-like protein. If this is indeed the case, then the boundaries of typical E2 structural domains and motifs should be identifiable within the predicted amino acid sequence. However, before this can be investigated, account must be taken of the sequence alignments which show that the H. volcanii gene contains an extra 306 bp within the middle of the sequence that are not present in any other of the E2 sequences in the databases. This ‘insert’ is in-frame with the rest of the sequence and would give an extra 102 aa (residues 155–256 in the 521 aa complete sequence) to the polypeptide. The possible significance and origin of this segment is considered in the Discussion.

With respect to the protein (minus the insert) that would be produced from H. volcanii ORF3, secondary structural predictions were made using the PHDsec program on the EMBL Predict-Protein Server (see Methods). This program constructs a multiple sequence alignment from the databases (in the current case, 36 sequences were selected by the program) and this is used as input for secondary structural predictions by a system of neural networks. Three structured domains are predicted, separated by two regions of 36 and 37 aa that have mean probabilities of loop assignment (on a scale of 0–9) of 8.1 and 7.8, respectively. This is the same arrangement of domains as a typical E2 polypeptide of an ODHC and therefore the domains of the halophilic protein have tentatively been assigned the same names and functions (Fig. 3).

Analysing known E2 sequences (e.g. from B. steardo thermophilus) with the PHDsec program gives almost
The lipoyl domain. Lipoyl domains consist of approximately 80 residues arranged in two $\beta$-sheets forming a flattened $\beta$-barrel, with a core of hydrophobic residues (Green et al., 1995). The lipoyl domain of the E2 from *H. volcanii*, extending from residues 1 to 82, is evident by the conserved DKA motif around the presumed lipoyl lysine, K44 (Dardel et al., 1993), as well as high homology throughout the region. The secondary structural predictions indicate the presence of two $\beta$-sheets and there is also a large proportion of hydrophobic residues in the core area. The structure is similar to the *B. steatorrhomphilus* PDHC E2 in having a single lipoyl domain.

**The peripheral E1 and E3 subunit-binding domain.** Peripheral subunit-binding domains of E2 chains consist of approximately 35 residues and form two parallel $\alpha$-helices, separated by a short strand, a helix-like turn and an irregular loop, with a hydrophobic core (Robien et al., 1992; Kalia et al., 1993). From the sequence alignments, the domain in *H. volcanii* consists of residues 119–154 and the secondary structural predictions indicate the $\alpha$-helical and $\beta$-strand regions in the expected positions. Many of the core hydrophobic residues are conserved, whereas others are exchanged for different hydrophobic groups. The ‘insert’ referred to above (155–256) is thus immediately adjacent to the C-terminal side of this domain.

It has been suggested that an aspartate and a threonine residue in the PDHC E2 of *B. steatorrhomphilus* form a hydrogen bond crucial to domain stability (Kalia et al., 1993). The aspartate is conserved throughout all known E2 sequences, including the *H. volcanii* E2 (D152), whereas the threonine is replaced by a serine in many

### Table 1. Sequence identities to proteins in databases

Identification of database homologues was performed using BLAST p. $P$-score, probability.

<table>
<thead>
<tr>
<th><em>H. volcanii</em> ORF</th>
<th>Database match</th>
<th>Species</th>
<th>Function</th>
<th>Length (aa)</th>
<th>Accession no.</th>
<th>Identity (%)</th>
<th>$P$-score</th>
<th>Overlap (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (373 aa)</td>
<td>Bacillus steatorrhomphilus E1α PDHC</td>
<td>E1α PDHC</td>
<td>368</td>
<td>P21873</td>
<td>44</td>
<td>−75</td>
<td>347</td>
<td></td>
</tr>
<tr>
<td>2 (328 aa)</td>
<td>Bacillus subtilis E1α PDHC</td>
<td>E1α PDHC</td>
<td>370</td>
<td>P21881</td>
<td>43</td>
<td>−73</td>
<td>347</td>
<td></td>
</tr>
<tr>
<td>3 (419 aa)</td>
<td>Acholeplasma laidlawii E1α PDHC</td>
<td>E1α PDHC</td>
<td>345</td>
<td>P35483</td>
<td>45</td>
<td>−72</td>
<td>343</td>
<td></td>
</tr>
<tr>
<td>4 (475 aa)</td>
<td>Mycobacterium tuberculosis E1α PDHC</td>
<td>E1α PDHC</td>
<td>367</td>
<td>O06161</td>
<td>45</td>
<td>−72</td>
<td>344</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Streptomyces avermitilis E1α BCODHC</td>
<td>E1α BCODHC</td>
<td>381</td>
<td>Q53592</td>
<td>43</td>
<td>−68</td>
<td>329</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacillus steatorrhomphilus E1β PDHC</td>
<td>E1β PDHC</td>
<td>324</td>
<td>P21874</td>
<td>56</td>
<td>−94</td>
<td>322</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacillus subtilis E1β PDHC</td>
<td>E1β PDHC</td>
<td>324</td>
<td>P21882</td>
<td>54</td>
<td>−92</td>
<td>322</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acholeplasma laidlawii E1β PDHC</td>
<td>E1β PDHC</td>
<td>327</td>
<td>P35488</td>
<td>50</td>
<td>−85</td>
<td>318</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aeropyrum pernix E1β PDHC</td>
<td>E1β PDHC</td>
<td>325</td>
<td>Q9YBC3</td>
<td>52</td>
<td>−84</td>
<td>322</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Deinococcus radiodurans E1β ODHC</td>
<td>E1β ODHC</td>
<td>344</td>
<td>AAF09622</td>
<td>48</td>
<td>−84</td>
<td>324</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacillus steatorrhomphilus E2 PDHC</td>
<td>E2 PDHC</td>
<td>427</td>
<td>P11961</td>
<td>48</td>
<td>−93</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Staphylococcus aureus E2 PDHC</td>
<td>E2 PDHC</td>
<td>430</td>
<td>Q59821</td>
<td>47</td>
<td>−88</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Myxococcus xanthus E2 BCODHC</td>
<td>E2 BCODHC</td>
<td>416</td>
<td>Q9X6X2</td>
<td>48</td>
<td>−86</td>
<td>395</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aeropyrum pernix E2 PDHC</td>
<td>E2 PDHC</td>
<td>412</td>
<td>Q9YBC6</td>
<td>45</td>
<td>−79</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acholeplasma laidlawii E2 PDHC</td>
<td>E2 PDHC</td>
<td>544</td>
<td>P35489</td>
<td>40</td>
<td>−73</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacillus steatorrhomphilus E3 PDHC</td>
<td>E3 PDHC</td>
<td>470</td>
<td>P11959</td>
<td>49</td>
<td>−116</td>
<td>474</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacillus subtilis E3 PDHC</td>
<td>E3 PDHC</td>
<td>470</td>
<td>P21880</td>
<td>48</td>
<td>−115</td>
<td>470</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Staphylococcus aureus E3 PDHC</td>
<td>E3 PDHC</td>
<td>468</td>
<td>Q59822</td>
<td>47</td>
<td>−108</td>
<td>471</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Deinococcus radiodurans E3 PDHC</td>
<td>E3 PDHC</td>
<td>467</td>
<td>AAF11916</td>
<td>43</td>
<td>−98</td>
<td>465</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Escherichia coli E3 PDHC/OGDHC</td>
<td>E3 PDHC/OGDHC</td>
<td>473</td>
<td>P00391</td>
<td>44</td>
<td>−97</td>
<td>454</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Predicted structural domains of the E2 protein that would be produced from *H. volcanii* ORF3. The domain boundaries were deduced from secondary structural predictions made using the PHDsec program of the EMBL Predict-Protein Server, as described in Methods. The three domains are shown as open boxes (with the corresponding number of amino acids contained within), and the inter-domain segments and proposed insert are shown as filled boxes. The position of the conserved lysine residue (K44) that is thought to be the residue to which a lipoic acid is attached, is indicated by an arrow.
structures. In *H. volcanii* E2 it appears as S142 (Fig. 4a). Since the two residues are similar in structure, it is predicted that the stabilizing hydrogen bond is present in the *H. volcanii* structure.

With respect to the domain’s possible subunit-binding function, most of the residues implicated in binding the E3 dimer in the *B. stearothermophilus* structure are conserved in *H. volcanii* (detailed in Mande et al., 1996). Thus, S124, R126, R130 and R147 are all conserved, whereas R127 is a similarly charged lysine residue in the *B. stearothermophilus* structure. G141 is conserved and has been implicated in the binding of E1 (Kalia et al., 1993), although E150 is either a lysine or an arginine in most other sequences and has been implicated in electrostatic interactions with E1. These sequence alignments are illustrated in Fig. 4(a).

**The catalytic and structural core domain.** The catalytic and structural core domain of the *H. volcanii* E2, from residues 294 to 521, is highly conserved with the *B. stearothermophilus* E2 protein, especially around the proposed active site histidine (H493) and aspartate (D497) (Hendle et al., 1995).

**The inter-domain segments.** The inter-domain regions in the *E. coli* PDHC are dominated by Ala and Pro residues, interspersed with charged and hydrophilic amino acids; the conformation of one of these linkers has been studied by NMR spectroscopy and shown to be conformationally flexible, although not a random coil (Perham, 1991, and references therein). The proposed inter-domains segments of the *H. volcanii* protein (residues 83–118 and 257–293) are predicted not to have a defined secondary structure (see above for loop prediction probabilities). They possess, respectively, 10 Ala plus 16 Asp/Glu, and 11 Ala, 4 Pro plus 6 Asp/Glu residues, and thus fit with their being flexible linker regions as found in other E2 enzymes from non-archaeal 2-oxoacid complexes. The proposed insert (residues 155–256), which is situated between the peripheral subunit binding domain and its adjacent loop region (Fig. 3), also appears not to possess a well-defined secondary structure.

When the insert sequence is compared with the amino acid sequence databases, the best matches are with E2 components of the ODHCs, even though they do not themselves contain this extra sequence. However, the matches are partial and of a low identity. For example, there is partial alignment of 20 residues of the insert with amino acids 120–139 of its own E2 sequence (within the peripheral subunit binding domain), giving a 60% identity score (Fig. 4b).

**The E1α and E1β components**

These components of the ODHCs do not contain a domain structure that can be used to characterize the homologous *H. volcanii* proteins. However, amino acid sequence comparisons of TPP-binding enzymes have identified a common sequence motif of approximately 30 residues in length, beginning with the highly conserved sequence GDG and concluding with a second highly conserved NN (Hawkins et al., 1989). Approximately 10 aa to the C-terminal side of the GDG sequence there is usually an E or D residue, followed about 5 and 11 residues further on by a generally conserved A and P residue, respectively. Immediately preceding the NN sequence is a cluster of hydrophobic amino acids.

In PDHCs, the binding site for TPP is thought to reside in the E1α and not the E1β subunits (Stepp & Reed, 1985) and it is in the E1α enzyme that the conserved motif is found. The proposed E1α protein of *H. volcanii* contains this motif, the level of identity (Fig. 5) arguing that it has retained the functional property of TPP binding thought to be associated with this structural element.

**Enzyme activities**

In the context of our discovering genes for a putative ODHC, it is worth reiterating that we and others have never detected enzymic activities of PDHC, OGDHC or the BCODHCs in *H. volcanii* or any other halophilic archaeae. However, DHLipDH activity is present in all strains tested. Assays for the ODHCs have been repeated with the strain of *H. volcanii* used in this work and the absence of enzymic activity has been reconfirmed. Positive controls for these activities have been performed with cell extracts of *E. coli*.
Northern blots

Total RNA prepared from *H. volcanii* was subjected to agarose gel electrophoresis and transferred to nylon membranes as described in Methods. Northern analysis using the *H. volcanii* insert from pNAT82 (Fig. 2) highlighted a single RNA band at 5.2 kb from both the aerobically and anaerobically grown cells.

**DISCUSSION**

The sequence data presented in this paper, the structural predictions of the protein products of the four ORFs discovered and the previous identification of DHLipDH as the enzymically active product of what is now ORF4, all strongly suggest that *H. volcanii* possesses an operon encoding an ODHC. Moreover, sequence identities to proteins in the databases support the tentative conclusion that this system is a PDHC.

Northern analysis of *H. volcanii* RNA strongly indicates that the whole operon is transcribed as a single message in both aerobically and anaerobically grown cells; that is, a 5.2 kb RNA species hybridizing to the cloned insert in pNAT82 has been detected, and this is very close to the predicted length (5.4 kb) from the combined E1α, E1β, E2 and E3 gene sequences. The known translation of the fourth ORF (DHLipDH), the proximity of all four genes in the operon and the occurrence of overlapping start and stop codons (between ORFs 1 and 2, and between ORFs 3 and 4) suggest translational coupling and therefore the production of enzymes E1α, E1β and E2.

However, neither PDHC nor any ODHC activity has been detected in *H. volcanii*, nor have they been found in any archaeon (Danson, 1988, 1993). Assuming that all four protein products are produced, a number of explanations are possible. First, a functional ODHC may be produced but the substrate is not pyruvate, 2-oxoacids nor any of the branched-chain 2-oxoacids. Alternatively, ORFs 1, 2 and 3 may produce non-functional proteins for an ODHC even though an active DHLipDH (ORF 4) is synthesized. Consistent with this, halophilic and other archaea possess 2-oxoacid oxidoreductases as alternative enzymes for the required catabolic functions. However, many essential features of the proposed E1α and E2 proteins have been conserved (E1β structural and functional data, apart from primary sequences, are not available in non-archaeal organisms for comparison), suggesting retention of function. These include the E2 lysine residue that is lipoylated in non-archaeal E2 components, and we have previously shown the presence of lipoic acid in the halophilic *Archaea* (Pratt et al., 1989), although its cellular location was not established.

An added complicating factor to the question of functional proteins being produced from the operon is the nature of the ‘insert’ thought to be present in the E2 component. It is not clear if this sequence is of importance to the function in *H. volcanii*, or if it has arisen through aberrant crossing-over at the gene level (see Fig. 4b), resulting in a non-functional E2 protein that either is inactive or cannot assemble into the normal core structure that binds E1α, E1β and DHLipDH.

Clearly, our findings suggest a number of experiments to probe further the possibilities, although none are technically easy due to the requirement of halophilic enzymes for high (> 2 M) concentrations of salt; therefore they are outside the scope of the current paper. However, it is interesting to note that a DHLipDH gene is present in the genomes of *Methanococcus janaschii* and *Methanobacterium thermoautotrophicum*. In contrast to *M. janaschii*, which has only the DHLipDH gene, *Mb. thermoautotrophicum* also has genes that show homology with the proposed E1α, E1β and E2 genes from *H. volcanii*, although the identities are low (17, 23 and 18%, respectively) and they are not in an operon structure (intergenic distances of 105, 25 and 145 kbp, respectively). The proposed E1α of *Mb. thermoautotrophicum* shows do the TPP-binding mo-
tif. However, prediction of structural motifs of the E2 protein failed to demonstrate the characteristic E2 domain structure and sequence alignments show that very few of the functional residues conserved in the \textit{H. volcanii} E2 are present in the protein from the methanogen. Interestingly, no homologues to E1\text{a}, E1\text{b}, E2 or DH\text{LipDH} could be identified in the genomes of \textit{Archaeoglobus fulgidus} and \textit{Pyrococcus horikoshii}.

Since the completion of the work reported in this paper, four genes have been identified within the genome of \textit{Aeropyrum pernix} that show homology to the components of PDHC. In turn, they are 40–50\% identical at the amino acid level with the \textit{H. volcanii} sequences given here, but there are no reported enzymological studies from this particular thermophilic archaeon. It is hoped that our analyses will point the way to future studies in \textit{A. pernix} and a wide range of other \textit{Archaea} to investigate the origin and evolution of the ODHCs in particular and of the pathways of central metabolism in general. The presence of ODHC genes in the halophilic and thermophilic \textit{Archaea} adds further weight to the idea that archaean metabolism shares many common features with that of the \textit{Bacteria} and the \textit{Eukarya}, and that the pathways of central metabolism may have been laid down before the divergence of the three domains (Danson \textit{et al.}, 1998).

\textbf{ACKNOWLEDGEMENTS}

We thank the Biotechnological and Biological Research Council, UK (research grants to M. J. D., D. W. H. and Garry Taylor, and to Julian Chaudhuri, D. W. H. and M. J. D., and Earmarked Studentships to K. A. J. and D. G. M.), the Royal Society and NATO (travel grants to M. J. D. and D. W. H.) and the Australian Research Council (research grant to M. D. S.) for generous financial support. We are also indebted to K. J. Stevenson and N. N. Vettakkorumakan (University of Calgary, Alberta, Canada) for the supply of plasmid pNAT82 and for helpful discussions concerning their initial sequence data.

\textbf{REFERENCES}


Archaeal 2-oxoacid dehydrogenase complexes


Received 24 November 1999; revised 14 February 2000; accepted 25 February 2000.