Extremophiles (2001) 5:61–69
Digital Object Identifier (DOI) 10.1007/s007920000174


ORIGINAL PAPER

**Benjamin N. Mijts · Bharat K.C. Patel**

# Random sequence analysis of genomic DNA of an anaerobic, thermophilic, halophilic bacterium, *Halothermothrix orenii*

Received: June 19, 2000 / Accepted: December 1, 2000 / Published online: February 1, 2001

**Abstract** A pBluescriptSK+ vector library consisting of 3,360 clones with an average insert size of 3.5 kb was constructed from the genome of *Halothermothrix orenii*, a halophilic and thermoanaerobic member of the family *Haloanaerobiaceae*. From both ends, 77 clones were sequenced using T3 and T7 vector primers generating 154 sequence tags, representing approximately 85 kb of the genome. Comparison of sequence tags against the GenBank database using BLASTX identified 66 known proteins and 15 conserved hypothetical proteins. The putative proteins included a V-ATPase, hydrogenases, and enzymes with potential for industrial applications. The overall G + C% of the codons used was 42.9% with a third-position G + C content of 38.6%. High levels of excess acidic amino acids were not detected in the putative proteins of *H. orenii* as compared to the mesophilic haloanaerobes. This lack may be the result of reduced activity of acidic, halophilic enzymes at high temperatures and intermediate salt concentrations.

**Key words** Halophile · Thermophile · *Halothermothrix orenii* · Genomic sequence · V-ATPase · Sucrose phosphate synthase · Codon bias

## Introduction

Moderately halophilic microbes (optimum growth between 5% and 20% NaCl) of the domain *Bacteria* are distinguished from nonhalophiles by a range of diverse molecular adaptations that promote their survival and proliferation in high-salt-containing environments. The aerobic and facultative anaerobic members of the moderate halophiles are phylogenetically diverse and are distributed in numerous phyla of the domain *Bacteria*, whereas the less diverse halophilic obligate anaerobes are generally restricted to the order *Haloanaerobiales* in the low G+C DNA-containing phylum. These two groups of moderate halophiles are also fundamentally different in their strategies for survival in high salt concentrations (Madigan and Oren 1999). Members of the order *Haloanaerobiales* studied to date counteract a high-salt environment by maintaining comparable high salt levels within their cytoplasm (Oren 1986; Rengpipat et al. 1988), in contrast to the aerobic and facultative anaerobic moderate halophiles, which maintain low intracellular salt levels and synthesize compatible solutes such as betaine. The 'salt-in' strategy is also found in members of the order *Halobacteriales*, within the domain *Archaea* (Rengpipat et al. 1988). This approach to maintaining cell integrity requires both extracellular and intracellular enzymes and cell structures to be active or stable under high-salt conditions.

To date, two thermophilic moderate halophiles have been described, both of which are members of the low G+C DNA-containing gram-positive phylum. *Halothermothrix orenii* is a member of the family *Haloanaerobiaceae*, order *Haloanaerobiales* (Cayol et al. 1994), and grows at temperatures to 68°C in the presence of NaCl to 20% (optimal growth at 68°C and 10% NaCl), whereas *Thermohalobacter berrensis*, a member of the order *Clostridiales*, is more thermophilic than *Halothermothrix orenii* and grows readily at 70°C in the presence of 15% NaCl (Cayol et al. 2000).

Very little is known about genes and enzymes from anaerobic moderate halophiles of the domain *Bacteria*. The only protein sequence that has been determined for members of the order *Haloanaerobiales* is the ribosomal A-protein from *Haloanaerobium praevalens* (Matheson et al. 1987). In this article, we report on the analysis of sequence tags determined from randomly selected clones of a *Halothermothrix orenii* genomic DNA library to gain information about the genetic structure of this bacterium. An

Communicated by J. Wiegel

B.N. Mijts · B.K.C. Patel (✉)
Microbial Discovery Research Unit, School of Biomolecular and Biomedical Sciences, Faculty of Science, Griffith University, Brisbane, Queensland 4111, Australia
Tel. +61-41772-6671; Fax +61-73875-7800
e-mail: bharat@genomes.sci.gu.edu.au

objective of this study is to investigate the specific cellular and molecular adaptations that allow *H. orenii* to survive at high temperatures and salt concentrations, in particular any distinctive trend in overall amino acid usage in *H. orenii* proteins. Additionally, sequence tags corresponding to any proteins of potential in industrial applications are identified.

## Materials and methods

### Bacterial strains and plasmids

*Halothermothrix orenii* strain H168 was obtained from Dr. Bernard Ollivier (Institut de Recherche pour le Developpement [IRD], Universite de Provence, France). *Escherichia coli* strain DH10B [F- *mcr*A Δ(*mrr-hsd*RM*S-mcr*BC) ϕ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara-leu*)7697 *ara*D139 *gal*U *gal*K *nup*G *rps*L λ-] was used for construction of the genomic library in pBluescriptSK+ vector (Stratagene, La Jolla, CA, USA) using routine molecular biology methods (Sambrook et al. 1989).

### Culturing of *Halothermothrix orenii* and DNA extraction

*Halothermothrix orenii* was grown anaerobically at 60°C in tryptone-yeast extract-glucose (TYEG) medium (Patel et al. 1985a) modified by the addition of 100 g $l^{-1}$ NaCl. The techniques for preparing anaerobic media have been described previously (Patel et al. 1985b). A 100-ml culture was grown overnight and cells harvested by centrifugation at 8,000 *g* for 15 min. The cell pellet was resuspended in 3.95 ml 10 mM Tris-HCl, 1.5 M NaCl (pH 8.0), and 100 µl 50 mg/ml lysozyme was added. The cell suspension was incubated on ice for 5 min, and 0.6 ml 0.5 M ethylenediaminetetraacetic acid (EDTA) (pH 7.5) and 0.25 ml 20% (w/v) sodium dodecyl sulfate (SDS) were added. Following incubation on ice for a further 5 min, 100 µl 20 mg/ml proteinase K was added and the cell lysate incubated at 55°C overnight. Genomic DNA was then extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol precipitated. Genomic DNA was spooled, dissolved in Tris-EDTA (TE) buffer, treated with RNase, and stored at –20°C.

### Construction of the *Halothermothrix orenii* genomic DNA library

Conventional recombinant DNA techniques were used (Sambrook et al. 1989). Briefly, *Halothermothrix orenii* genomic DNA, partially digested with *Sau*IIIA restriction enzyme, was size fractionated using polyacrylamide gel electrophoresis and 2- to 10-kb fragments were electroeluted. Eluted DNA fragments were ligated to *Bam*HI-digested and dephosphorylated pBluescriptSK+ vector. An aliquot of the ligation mixture was electroporated into DH10B electrocompetent cells (Gibco BRL, Rockville, MD, USA), and white recombinant transformants were selected after overnight growth on MacConkey agar plates containing 100 µg/ml ampicillin (Jennings and Beacham 1989). Recombinant plasmid-containing colonies were inoculated into 150 µl Luria-Bertani (LB) ampicillin media in microtiter plates, grown overnight at 37°C, and preserved by storing at –70°C after addition of sterile glycerol to 40%.

### DNA sequencing

Plasmids from randomly selected clones were purified using QIAprep (Qiagen, Chatsworth, CA, USA) spin columns and used as the template for sequencing. Sequencing reactions were prepared in thin-walled 0.2-ml polymerase chain reaction (PCR) microfuge tubes with purified plasmid, BigDye Terminator Ready Reaction Mix, and T3 or T7 universal primers as recommended by the manufacturer (Perkin-Elmer, Norwalk, CT, USA). The reactions were placed in an Idaho Rapid Cycler (Idaho Technologies, Idaho Falls, ID, USA) and cycled using the parameters recommended for use with the BigDye terminator kit. Nucleotide sequences were determined using an ABI 377 automated DNA sequencer at the Griffith University Molecular Biology Facility.

### Sequence analysis

*Halothermothrix orenii* genomic DNA sequence data obtained were edited to remove vector DNA sequences and analyzed by similarity searches against the nonredundant Genbank database (Benson et al. 1999) using NCBI BLASTX software found on the World Wide Web at http://www.ncbi.nlm.nih.gov/BLAST/. A number of genomic sequence tags that generated long, high amino acid matches were selected and the nucleotide sequences used to determine codon bias using CodonW software (http://www.molbiol.ox.ac.uk/cu/codonW.html). Amino acid sequences from BLASTX alignments with high matches to characterized enzymes were extracted and an amino acid profile determined using BioEdit software (http://www.mbio.ncsu.edu/RNaseP/info/programs/BIOEDIT/bioedit.html)

## Results and discussion

### Construction of a *Halothermothrix orenii* genomic library and genes from *Halothermothrix orenii*

An initial *Halothermothrix orenii* genomic DNA library consisting of 3,360 clones with an average insert size of approximately 3.5 kb was prepared in pBluescriptSK+ vector. T3 and T7 primer sequencing of 77 plasmids generated 154 sequence tags, totaling approximately 85 kb of *H. orenii* genome sequence. Sequence tags were submitted for

analysis by searching for sequence homology against the nonredundant Genbank database using BLASTX. In general, sequences with bit scores greater than 80 were considered to be high matching, although a number of sequences with lower scores considered to be significant on examination of individual alignments are also included. In total, 66 sequence tags with significant similarity to proteins with assigned function were identified, and a further 15 sequence tags with homology to conserved hypothetical proteins without assigned functions were identified.

Detailed descriptions of the 81 sequence tags identified in this study with significant homology to characterized and hypothetical proteins are listed in Table 1. Based on BLASTX search results, the sequence tags were classified into protein functional categories involved in cell processes such as cell structures (e.g., flagella), intermediate metabolism (e.g., enzymes involved in amino acid metabolism), information pathways (e.g., DNA helicase, DNA gyrase, and exonucleases), and a number of enzymes involved in fermentative energy production (e.g., hydrogenases).

Sequence tag matches for histidyl-tRNA synthetase, imidazoleglycerol-phosphate dehydratase, acetolactate synthase, and periplasmic (Ni-Fe-Se) hydrogenase were observed for inserts from two different clones. In each case, the corresponding nucleotide sequences did not match, and in the case of the hydrogenase gene matches, overlapping sequences were obtained (see following). The relatively small number of duplicates obtained in the analysis does not suggest any significant bias in the constructed genomic DNA library and may be the result of multiple gene copies being present in the *Halothermothrix orenii* genome.

The numerical distribution of highest sequence matches among different organisms (Table 2 ) is biased toward those with complete, annotated genome sequences in databases. 16S rRNA gene sequence analysis places *H. orenii* as a member of the family *Haloanaerobiaceae* of the phylum represented by the low G+C DNA-containing gram-positive bacteria. As would be expected, the distribution of sequence tag matches is biased toward the only completely sequenced member of the low G+C DNA gram-positive phylum, *Bacillus subtilis* (Kunst et al. 1997). Another interesting aspect of the analysis is that high matches were also obtained against the slightly halophilic, thermophilic members of domains *Bacteria* (e.g., *Aquifex* and *Thermotoga*) and *Archaea* (e.g., *Desulfurococcus*, *Pyrococcus*, and *Archaeoglobus*). Whether this is the result of the halophilic nature or the thermophilic nature of *Halothermothrix orenii* proteins can only be elucidated with certainty when more sequence data are available for the genomes of moderate halophiles.

A number of sequence tags that provide insight into the poorly characterized genetics and biochemistry of *H. orenii* were identified and are discussed in further detail here. These sequence tags include those with homology to V-type ATPase, a number of hydrogenases, and sucrose phosphate synthase. Additionally, a number of sequence tags with homology to enzymes of potential for industrial applications are discussed.

## V-type ATPase

Sequence tag 23D6/T7 was found to have very high homology to the conserved catalytic A subunit of V-type ATPase, a class of enzymes found primarily in vacuolar vesicles in domain *Eukarya* and in plasma membranes from Archaea. Most members of the domain *Bacteria* express only distantly related F-type ATPases. However, V-ATPases have been identified and characterized in a few members of the domain *Bacteria*, including the $Na^+$-transporting ATPases from *Enterococcus hirae* (Takase et al. 1993) and *Caloramator fervidus* (formerly *Clostridium fervidus*) (zu Bentrup et al. 1997) and $H^+$-transporting ATPases from members of the genus *Thermus* (Tsutsumi et al. 1991). The highest match for the *Halothermothrix orenii* V-ATPase sequence was with the V-ATPase from the archaeon *Desulfurococcus* (Shibui et al. 1997), although as the sequence tag encodes an extremely well conserved region, more sequence data are required to draw detailed phylogenetic conclusions. High matches were also obtained with V-ATPases from other members of the domain *Bacteria* and other members of the domain *Archaea* (known as A-ATPases), including extreme halophiles such as *Halobacterium salinarium* (Fig.1).

In members of the order *Haloanaerobiales* studied to date, high intracellular KCl levels osmotically balance high extracellular NaCl concentrations. Intracellular $Na^+$ concentrations were found to be about 0.5 to 0.6 times the extracellular $Na^+$ concentrations, remaining proportional at most external salt concentrations. Intracellular $Cl^-$ was found to be very similar to that of the external environment, and $K^+$ and $Mg^{2+}$ ions were found to be incorporated into cells against a concentration gradient (Rengpipat et al. 1988). Oren (1986) suggested that members of the order *Haloanaerobiales* generate a membrane potential by ATP hydrolysis to drive intracellular potassium accumulation. In the related thermophilic, fermentative anaerobe *Caloramator fervidus*, a V-ATPase was shown to act as an active primary ion pump, generating an electrochemical $Na^+$ membrane gradient for transport of substances into the cell. An $Na^+$-stimulated V-ATPase has also been detected in the membrane of the facultatively anaerobic, halophilic, alkaliphilic bacterium *Amphibacillus* sp. M12 (Kaieda et al. 1998) and is also thought to be involved in generating a transmembrane $Na^+$ gradient for amino acid transport and motility. On the basis of these data, ATP-powered $Na^+$ translocation to generate a membrane potential for $K^+$ accumulation provides an appealing hypothesis for the likely role of V-ATPase in *Halothermothrix orenii*.

## Hydrogenases

Several different sequence tags with homology to hydrogenase and hydrogenase-associated proteins were identified. Hydrogenases catalyze the reversible oxidation of $H_2$ gas and, as *Halothermothrix orenii* is a hydrogen-producing fermentative organism, the presence of these genes would be expected. The sequence tag corresponding to clone 23B08/

**Table 1.** List of *Halothermothrix orenii*-derived sequence tags having matches to known gene products in other organisms

| Accession number | Gene description | Organism | Length of ORF | Score (bits) | Clone / Primer |
|---|---|---|---|---|---|
| Intermediate metabolism | | | | | |
| gi\|2621546 | Acetolactate synthase | *Methanobacterium thermoautotrophicum* | 92 | 113 | 10B09 / T3 |
| dbj\|BAA17984\| | Acetolactate synthase | *Synechocystis* sp. | 89 | 81.5 | 23A03 / T7 |
| sp\|P29726\|PURA_BACSU | Adenylosuccinate synthetase | *Bacillus subtilis* | 125 | 158 | 10D11 / T3 |
| emb\|CAB49570.1\| | Alpha-glucan phosphorylase | *Pyrococcus abyssi* | 197 | 104 | 10B03 / T7 |
| dbj\|BAA30246\| | Aminomethyltransferase | *Pyrococcus horikoshii* | 107 | 100 | 10E06 / T3 |
| sp\|P36772\|LON_BACBR | ATP-dependent protease La | *Bacillus brevis* | 144 | 138 | 23D04 / T3 |
| gi\|2984001 | C-terminal fumarate hydratase, class I | *Aquifex aeolicus* | 80 | 102 | 10B10 / T3 |
| gb\|AAC84034.1\| | Diapophytoene dehydrogenase CrtN | *Heliobacillus mobilis* | 182 | 157 | 23C12 / T3 |
| emb\|CAA04932\| | Esterase* | *Thermotoga maritima* | 97 | 94.4 | 23A12 / T7 |
| sp\|P19210\|FPG_BACFI | Formamidopyrimidine-DNA glycosylase | *Bacillus firmus* | 136 | 129 | 23D03 / T3 |
| ref\|NP_000145.1\|\| | Galactokinase 1 | *Homo sapiens* | 174 | 155 | 23B05 / T7 |
| sp\|O22493\|GSH1_LYCES | Gamma-glutamylcysteine synthetase | *Lycopersicon esculentum* | 146 | 79.6 | 10A06 / T3 |
| gi\|2983123 | Glutamyl-tRNA amidotransferase subunit B | *Aquifex aeolicus* | 58 | 72.6 | 23A07 / T7 |
| emb\|CAB50351.1\| | Glycerate dehydrogenase | *Pyrococcus abyssi* | 153 | 173 | 23C05 / T7 |
| sp\|Q9X1E4\|GLK2_THEMA | Glycerol kinase 2 | *Thermotoga maritima* | 54 | 69.9 | 23D12 / T7 |
| emb\|CAB10270.1\| | Imidazoleglycerol-phosphate dehydratase | *Arabidopsis thaliana* | 134 | 136 | 10A04 / T7 |
| gi\|2982780 | Imidazoleglycerol-phosphate dehydratase | *Aquifex aeolicus* | 114 | 143 | 10A07 / T3 |
| emb\|CAB38894.1\| | Integrase | *Streptomyces coelicolor* | 199 | 95.9 | 23B07 / T7 |
| sp\|P32099\|LPLA_ECOLI | Lipoate-protein ligase A | *Escherichia coli* | 63 | 71.4 | 10D10 / T7 |
| gi\|2342551 | Methylcobamide:CoM methyltransferase isozyme A | *Methanosarcina barkeri* | 160 | 74.5 | 10C12 / T3 |
| gi\|2293201 | Peptidase | *Bacillus subtilis* | 120 | 139 | 10D02 / T3 |
| gb\|AAF05093.1\|AF150930_2 | Phosphoribosyl formimino-5-aminoimidazole isomerase | *Thermoanaerobacter ethanolicus* | 231 | 136 | 10A04 / T3 |
| sp\|P12039\|PUR2_BACSU | Phosphoribosylamine-glycine ligase | *Bacillus subtilis* | 148 | 100 | 10D03 / T7 |
| sp\|P77836\|PDP_BACST | Pyrimidine-nucleoside phosphorylase PynP | *Bacillus stearothermophilus* | 178 | 220 | 10B06 / T7 |
| dbj\|BAA13561.1\| | Pz-peptidase | *Bacillus licheniformis* | 150 | 133 | 10D02 / T7 |
| pir\|\|E53402 | Serine *O*-acetyltransferase | *Bacillus stearothermophilus* | 120 | 186 | 10D03 /T3 |
| gi\|1001295 | Sucrose phosphate synthase* | *Synechocystis* sp. | 72 | 80.4 | 23E03 / T7 |
| sp\|P31014\|TNA1_SYMTH | Tryptophanase 1 | *Symbiobacterium thermophilum* | 41 | 60.9 | 10A01 / T3 |
| Information pathways | | | | | |
| sp\|O32038\|SYD_BACSU | Aspartyl-tRNA synthetase | *Bacillus subtilis* | 175 | 182 | 10B07 / T3 |
| sp\|P94604\|GYRB_CLOAB | DNA gyrase subunit B | *Clostridium acetobutylicum* | 209 | 209 | 10D12 / T7 |
| sp\|P23479\|SBCD_BACSU | Exonuclease SbcD homologue | *Bacillus subtilis* | 177 | 127 | 10D09 / T3 |
| sp\|P50830\|YPRA_BACSU | Helicase | *Bacillus subtilis* | 148 | 218 | 10D04 / T7 |
| gi\|2982856 | Histidyl-tRNA synthetase | *Aquifex aeolicus* | 146 | 68.7 | 10A07 / T7 |
| sp\|O32422\|SYH_STAAU | Histidyl-tRNA synthetase | *Staphylococcus aureus* | 150 | 156 | 10B07 / T7 |
| gb\|AAD35415.1\|AE001714_6 | m4C-Methyltransferase | *Thermotoga maritima* | 91 | 72.6 | 10B01 / T7 |
| sp\|P05651\|RECF_BACSU | Recombination protein RecF | *Bacillus subtilis* | 185 | 172 | 10D12 / T3 |
| gb\|AAF10128.1\|AE001913_2 | Replicative DNA helicase | *Deinococcus radiodurans* | 164 | 120 | 10D11 / T7 |
| Bioenergetics | | | | | |
| gi\|2983520 | Hydrogenase accessory protein HypA | *Aquifex aeolicus* | 64 | 107 | 23A04 / T7 |
| sp\|Q43950\|HYPF_AZOCH | Hydrogenase maturation protein HypF | *Azotobacter chroococcum* | 95 | 84.7 | 23A04 / T3 |
| gb\|AAD33071.1\|AF148212_1 | Hydrogenase-1* | *Clostridium thermocellum* | 200 | 181 | 23B08 / T7 |
| sp\|P13063\|PHSS_DESBA | Periplasmic (Ni-Fe-Se) hydrogenase* | *Desulfomicrobium baculatum* | 149 | 102 | 23B03 / T7 |
| gi\|3420957 | Periplasmic (Ni-Fe-Se) hydrogenase* | *Desulfovibrio baculatus* | 160 | 101 | 23B03 / T3 |
| pdb\|1CC1\|S | Periplasmic (Ni-Fe-Se) hydrogenase* | *Desulfomicrobium baculatum* | 66 | 68.3 | 30A02 / T3 |
| gb\|AAB64416.1\| | V-ATPase A subunit* | *Desulfurococcus* sp. *SY* | 150 | 257 | 23D06 / T7 |
| Transmembrane transport | | | | | |
| gb\|AAB85112.1\| | Manganese transport system membrane protein | *Methanobacterium thermoautotrophicum* | 177 | 108 | 23C02/T7 |

**Table 1.** *(Continued)*

| Accession number | Gene description | Organism | Length of ORF | Score (bits) | Clone / Primer |
|---|---|---|---|---|---|
| Signal transduction | | | | | |
| gi\|2688139 | Acid-inducible protein | *Borrelia burgdorferi* | 101 | 61.6 | 10C07 / T3 |
| emb\|CAA53764\| | Chemotaxis response regulator CheY | *Listeria monocytogenes* | 33 | 44.1 | 10C02 / T7 |
| sp\|Q56310\|CHEA_THEMA | Chemotaxis sensor histidine kinase CheA | *Thermotoga maritima* | 181 | 141 | 10C08 / T3 |
| sp\|P54663\|DEGS_BACBR | Sensor protein DegS | *Brevibacillus brevis* | 169 | 119 | 23C05 / T3 |
| sp\|Q45831\|REGA_CLOAB | Transcription regulatory protein RegA | *Clostridium acetobutylicum* | 142 | 113 | 10A01 / T7 |
| sp\|P06964\|CYTR_ECOLI | Transcriptional repressor CytR | *Escherichia coli* | 118 | 101 | 23A09 / T7 |
| emb\|CAB12514\| | Two-component sensor histidine kinase YesN | *Bacillus subtilis* | 152 | 89.3 | 10E05 / T7 |
| sp\|P39779\|CODY_BACSU | Vegetative protein CodY | *Bacillus subtilis* | 179 | 123 | 10C08 / T7 |
| Cell structure and processes | | | | | |
| gb\|AAD36435.1\|AE001790_7 | Flagellar basal-body rod protein FlgC | *Thermotoga maritima* | 102 | 117 | 10C02 / T3 |
| sp\|P35528\|FLIP_BACSU | Flagellar biosynthetic protein FliP | *Bacillus subtilis* | 88 | 109 | 23C04 / T3 |
| gi\|3322960 | Flagellar hook-associated protein 1 FlgK | *Treponema pallidum* | 124 | 65.6 | 10C05 / T7 |
| sp\|Q57511\|FLIM_BORBU | Flagellar motor switch protein FliM | *Borrelia burgdorferi* | 156 | 144 | 23C04 / T7 |
| gi\|2983377 | Flagellar hook basal-body protein FlgG | *Aquifex aeolicus* | 195 | 224 | 10C01 / T3 |
| pir\|\|JC4505 | Flagellar motor switch protein FliY | *Treponema pallidum* | 42 | 58.2 | 30A04 / T3 |
| sp\|P30725\|DNAJ_CLOAB | Heat-shock protein DnaJ | *Clostridium acetobutylicum* | 170 | 170 | 23D05 / T3 |
| sp\|Q45551\|DNAK_BACST | Heat-shock protein DnaK | *Bacillus stearothermophilus* | 131 | 214 | 23D05 / T7 |
| sp\|Q59240\|GRPE_BACST | Heat-shock protein GrpE | *Bacillus stearothermophilus* | 171 | 125 | 10D05 / T7 |
| dbj\|BAA12652\| | Penicillin-binding protein DacF | *Bacillus subtilis* | 177 | 136 | 10B06 / T3 |
| sp\|P16919\|RHSD_ECOLI | RHSD protein precursor | *Escherichia coli* | 398 | 71.4 | 10A05 / T7 |
| emb\|CAB50070.1\| | UTP-glucose-1-phosphate uridylyltransferase | *Pyrococcus abyssi* | 72 | 73 | 10D07 / T7 |
| sp\|Q07833\|WAPA_BACSU | Wall-associated protein precursor | *Bacillus subtilis* | 107 | 68.7 | 10A05 / T3 |
| Hypothetical proteins without assigned function | | | | | |
| gi\|2984375 | Hypothetical protein | *Aquifex aeolicus* | 126 | 112 | 10D10 / T3 |
| gi\|2649934 | Hypothetical protein | *Archaeoglobus fulgidus* | 171 | 94.8 | 10C12 / T7 |
| sp\|P50840\|YPSC_BACSU | Hypothetical 43.5-kDa protein | *Bacillus subtilis* | 80 | 82.7 | 10C04 / T7 |
| sp\|P50837\|YPRB_BACSU | Hypothetical protein | *Bacillus subtilis* | 159 | 80 | 10D04 / T3 |
| sp\|Q08794\|YCXC_BACSU | Hypothetical 34.1-kDa protein | *Bacillus subtilis* | 150 | 93.6 | 10C04 / T3 |
| sp\|P37565\|YACC_BACSU | Hypothetical 31.8-kDa protein | *Bacillus subtilis* | 135 | 122 | 10A08 / T7 |
| sp\|Q99171\|YMFB_BACSU | Hypothetical 26.4-kDa protein | *Bacillus subtilis* | 169 | 180 | 10A09 / T3 |
| sp\|P46347\|YQFG_BACSU | Hypothetical 17.8-kDa protein | *Bacillus subtilis* | 62 | 75.7 | 23A10 / T3 |
| emb\|CAB49194.1\| | Hypothetical protein | *Pyrococcus abyssi* | 148 | 157 | 23D12 / T3 |
| emb\|CAA20296\| | Hypothetical protein SC9A10.09 | *Streptomyces coelicolor* | 144 | 158 | 10A09 / T7 |
| dbj\|BAA18743\| | Hypothetical protein | *Synechocystis* sp. | 145 | 180 | 23A05 / T3 |
| dbj\|BAA10817\| | Hypothetical protein | *Synechocystis* sp. | 136 | 105 | 23A02 / T7 |
| dbj\|BAA17400\| | Hypothetical protein | *Synechocystis* sp. | 145 | 146 | 23A05 / T7 |
| dbj\|BAA17405\| | Hypothetical protein | *Synechocystis* sp. | 105 | 92.4 | 23B08 / T3 |
| gb\|AAD36177.1\|AE001769_7 | Conserved hypothetical protein | *Thermotoga maritima* | 144 | 102 | 10E08 / T3 |

Sequence tags discussed in detail in the text are indicated with an asterisk

**Table 2.** Sequences from the genera most frequently matching *Halothermothrix orenii* sequence tags

| Genus | Number of matching clones | Genome sequence available |
|---|---|---|
| *Bacillus* | 25 | Yes |
| *Aquifex* | 7 | Yes |
| *Thermotoga* | 6 | Yes |
| *Pyrococcus* | 5 | Yes |
| *Synechocystis* | 5 | Yes |
| *Clostridium* | 4 | No |

T7 was found to be homologous to Fe hydrogenase from the anaerobic thermophile *Thermotoga maritima* (Verhagen et al. 1999). The sequence tag from clone 23B03/T7, and partially overlapping sequences from 23B03/T3 and 30A02/T3, spanned a region of the *H. orenii* genome with coding regions homologous to large and small Ni-Fe-Se hydrogenase subunits from sulfate-reducing bacteria. The highest matches were obtained with Ni-Fe-Se hydrogenase from *Desulfomicrobium baculatum* (Garcin et al. 1999). This result indicates that, as in sulfate-reducing bacteria, more than one class of hydrogenase is present in *H. orenii*.

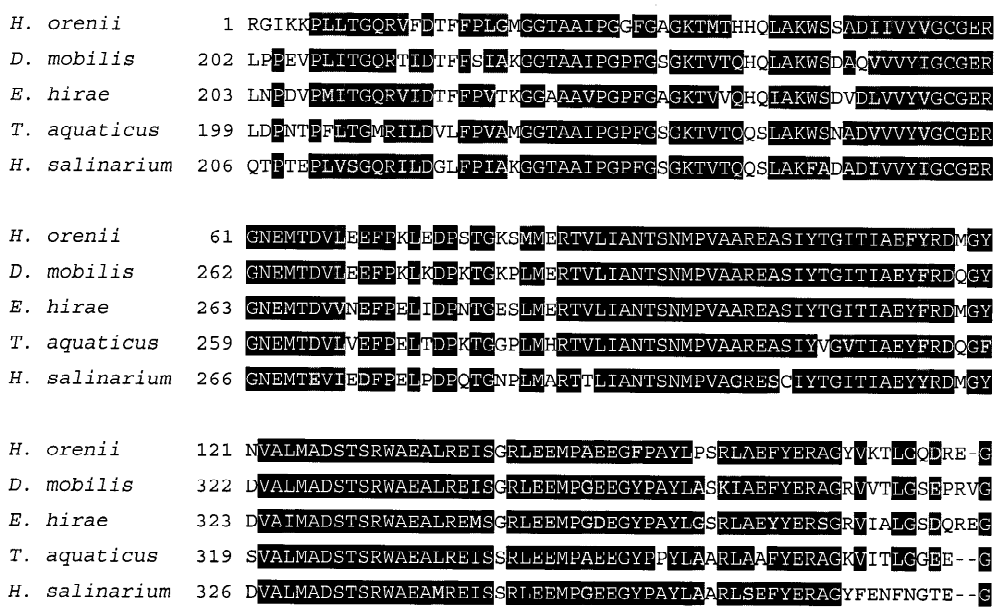


**Fig. 1.** *Halothermothrix orenii* putative V-ATPase partial sequence aligned to V-ATPase alpha-subunits of *Desulfurococcus mobilis*, *Enterococcus hirae*, *Thermus aquaticus*, and *Halobacterium salinarium*

## Sucrose phosphate synthase

Sequence tag 23E3/T7 was found to have homology to a sucrose phosphate synthase (SPS). SPS, a key enzyme in sucrose biosynthesis, has been identified in plant tissue and some unicellular eukaryotic algae. The only prokaryotes shown to date to possess SPS activity are the cyanobacteria *Anabaena* sp. PCC7119 and *Synechocystis* sp. PCC 6803. The SPS gene from *Synechocystis* sp. PCC 6803 has been identified in whole-genome sequencing, cloned, and expressed in an active form in *E. coli* (Curatti et al. 1998). BLAST searches against the Genbank unfinished microbial genome sequence database indicated that a region of the *Thiobacillus ferrooxidans* genome encoding a 705-amino-acid open reading frame (ORF) showed high homology to the *Synechocystis* SPS sequence. The SPS homologous ORF in *H. orenii* clone 23E3 was sequenced using a primer walking strategy and found to encode a complete, 496-amino-acid protein. Approximately 200 C-terminal amino acids were truncated from the putative *H. orenii* SPS compared to the *Synechocystis* and putative *T. ferrooxidans* proteins, which are both about 700 amino acids in length. Although the missing C-terminus region in the *H. orenii* enzyme may indicate it has a different, related catalytic activity, the N-terminal amino acid region, thought to be involved in binding of the substrate fructose-6-phosphate, is reasonably well conserved in all three bacterial SPS homologues (Fig.2). The other proposed substrate-binding region (UDP [uridine diphosphate] glucose) is poorly conserved in all three sequences, suggesting it may have been incorrectly identified or that each enzyme has a different catalytic activity or specificity.

Interestingly, biosynthesis of sucrose, along with other potential osmoprotectant compounds such as trehalose and glycinebetaine, has been observed under salt or osmotic stress conditions in Cyanobacteria. These compounds appear to play a role similar to that of compatible solutes in aerobic bacterial moderate halophiles, although sucrose itself has also been suggested as an intercellular transport compound. Lunn et al. (1999) were unable to observe any change in SPS gene expression in salt-stressed *Synechocystis* cells and therefore concluded that SPS did not play a significant role in salt-stress adaptation. Why bacterial genes encoding SPS homologues have been detected in *Synechocystis* sp. and now in *Thiobacillus ferroxidans* and *Halothermothrix orenii* (as described in this article) and not other bacterial genomes studied thus far is not clear. There appears to be no common survival requirement for sucrose synthesis in these bacteria, because they have very different phylogeny, physiology, and ecology. Purification and biochemical characterization of *T. ferroxidans* and *H. orenii* putative SPSs will allow the properties and physiological role of these enzymes to be investigated.

## Industrial enzymes

A number of genes with potential for industrial applications were also identified. These enzymes would be of particular utility in applications in which activity or stability at a range of salt levels and high temperatures is required. These genes include a number of peptidases (clones 10D2, 23D4/T3) and an esterase (23A12 /T7). Figure 3 is an alignment of the putative esterase region from *H. orenii* with the equivalent region of the closely related *Thermotoga maritima* enzyme.

## Codon usage

*Halothermothrix orenii* codon bias was determined using CodonW software from ten sequence tags with high


**Fig. 2.** Putative substrate-binding regions of *Halothermothrix orenii*, *Thiobacillus ferroxidans,* and *Syncehocystis* sp. SPS homologues. *Boxed regions* indicate putative substrate-binding domains identified by Curatti et al. (1998). *Box 1* is the fructose-6-phosphate-binding regions; *box 2* is the UDP glucose-binding region


**Fig. 3.** Alignment of *Halothermothrix orenii* putative esterase partial amino acid sequence to *Thermotoga maritima* esterase

matches to known proteins and ORFs longer than 150 amino acids. A number of sequenced genes related to sugar and polysaccharide metabolism identified from *Halothermothrix orenii* genomic DNA library activity screening, but not reported in this article, were also used to determine overall codon bias. A total of 3,554 codons were used for *H. orenii* codon bias analysis and these results, along with the codon bias for *Clostridium thermocellum* and *E. coli*, are listed in Table 3 .

The codon bias for *H. orenii* is quite distinct from that of *C. thermocellum* and *E. coli*. The overall G+C% of the codons was 42.9% with a third position G+C content of 38.6%. The overall G+C% of the *H. orenii* genome was found to be 39.6% (Cayol et al. 1994). The slightly higher G+C% of 42.9% observed for *H. orenii* coding sequences suggests the noncoding regions of the genome are relatively A+T rich.

## Amino acid usage

Previous studies have indicated a number of broad trends in amino acid usage of enzymes from extreme halophiles of the domain *Archaea*. The most distinct characteristic is a relatively high level of acidic amino acids observed in both individual proteins and amino acid composition of total cellular proteins (Lanyi 1974). Crystal structures have indicated that these acidic residues normally form part of the protein surface (Madern et al. 2000). Studies on the anaerobic moderate halophiles *Halobacteroides halobius* and *Haloanaerobium praevalens* by Oren (1986) indicated that these bacteria show similar trends in amino acid usage. Although a slightly increased level of positively charged amino acids has been observed in thermophilic enzymes when compared to equivalent enzymes from mesophiles, most of the adaptations in thermophilic enzymes are specific structural adaptations such as increased salt bridges. These structures have little discernible influence on the total amino acid content of both individual proteins and entire proteomes.

To investigate the amino acid usage of *Halothermothrix orenii*, amino acid sequences from high matching genes identified in BLASTX searches were pooled and used to generate an amino acid composition profile using BioEdit software. The excess acidic amino acid content of *H. orenii*, based on the analysis of approximately 8,000 amino acids from partial sequences of more than 60 proteins, is presented in Table 4 . The results indicate that *H. orenii* lacks the acidic amino acid excess observed in other members of the order *Haloanaerobiales*, such as *Haloanaerobium praevalens*, and extreme halophiles of the domain *Archaea* such as *Haloferax volcanii*. Although the method used cannot be considered a completely accurate means of determining acidic amino acid excess, the size and random nature of the sample should be sufficient to ascertain any distinctive trends. Comparative analysis of amino acid sequences of individual enzymes from *Halothermothrix orenii*, including a number of secreted enzymes that have been subcloned into *E. coli* and shown to be active at high temperatures and salt concentrations, also confirms this lack of elevated acidic amino acid levels (unpublished results).

A possible explanation for this difference lies in the thermophilic and moderately halophilic nature of *H. orenii*. Many enzymes with an acidic amino acid excess from mesophilic extreme halophiles of the domain *Archaea* are unusually thermophilic at high salt concentrations (Keradjopoulos and Holldorf 1977). However, this thermophilic character is often lost at low and intermediate salt concentrations. Excellent examples of this characteristic are the isocitrate dehydrogenase of *Haloferax volcanii* (Camacho et al. 1995) and glucose dehydrogenase of *Haloferax mediterranei* (Bonete et al. 1996). Both these enzymes are enzymatically active in the absence of salt but only show significant thermostability at salt concentrations greater

**Table 3.** Codon bias in *Halothermothrix orenii*, *Escherichia coli*, and *Clostridium thermocellum*

| Residue | Codon | *H. orenii* | | *C. thermocellum* (%)[a,b] | *E. coli* (%)[a,b] | Residue | Codon | *H. orenii* | | *C. thermocellum* (%)[a,b] | *E. coli* (%)[a,b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | No. of codons | %[a] | | | | | No. of codons | %[a] | | |
| Phe | TTT | 110 | 72 | 60 | 51 | | GCG | 10 | 4 | 16 | 34 |
| | TTC | 42 | 28 | 40 | 49 | Tyr | TAT | 131 | 78 | 65 | 53 |
| Leu | TTA | 75 | 24 | 13 | 11 | | TAC | 37 | 22 | 35 | 47 |
| | TTG | 25 | 8 | 28 | 11 | His | CAT | 47 | 75 | 59 | 52 |
| | CTT | 60 | 19 | 32 | 10 | | CAC | 16 | 25 | 41 | 48 |
| | CTC | 37 | 12 | 8 | 34 | Gln | CAA | 16 | 19 | 35 | 70 |
| | CTA | 20 | 6 | 2 | 3 | | CAG | 67 | 81 | 65 | 30 |
| | CTG | 99 | 31 | 18 | 55 | Asn | AAT | 134 | 67 | 50 | 39 |
| Ile | ATT | 114 | 46 | 42 | 47 | | AAC | 66 | 33 | 50 | 61 |
| | ATC | 54 | 22 | 15 | 46 | Lys | AAA | 139 | 62 | 62 | 76 |
| | ATA | 81 | 32 | 42 | 7 | | AAG | 87 | 38 | 38 | 24 |
| Met | ATG | 61 | 100 | 100 | 100 | Asp | GAT | 180 | 76 | 52 | 59 |
| Val | GTT | 105 | 41 | 69 | 29 | | GAC | 56 | 24 | 48 | 41 |
| | GTC | 45 | 18 | 8 | 20 | Glu | GAA | 165 | 65 | 69 | 70 |
| | GTA | 62 | 24 | 36 | 17 | | GAG | 89 | 35 | 31 | 30 |
| | GTG | 43 | 17 | 17 | 34 | Cys | TGT | 27 | 75 | 47 | 43 |
| Ser | TCT | 35 | 32 | 13 | 19 | | TGC | 9 | 25 | 53 | 57 |
| | TCC | 33 | 30 | 17 | 17 | Trp | TGG | 48 | 100 | 100 | 100 |
| | TCA | 35 | 32 | 27 | 12 | Arg | CGT | 16 | 8 | 10 | 42 |
| | TCG | 7 | 6 | 10 | 13 | | CGC | 14 | 7 | 6 | 37 |
| Pro | CCT | 55 | 41 | 31 | 16 | | CGA | 7 | 3 | 1 | 5 |
| | CCC | 40 | 30 | 9 | 10 | | CGG | 51 | 25 | 4 | 8 |
| | CCA | 16 | 12 | 20 | 20 | Ser | AGT | 56 | 75 | 14 | 13 |
| | CCG | 22 | 17 | 40 | 55 | | AGC | 19 | 25 | 19 | 27 |
| Thr | ACT | 40 | 24 | 27 | 21 | Arg | AGA | 37 | 18 | 55 | 4 |
| | ACC | 75 | 45 | 19 | 43 | | AGG | 77 | 38 | 24 | 3 |
| | ACA | 43 | 26 | 41 | 13 | Gly | GGT | 108 | 37 | 32 | 38 |
| | ACG | 9 | 5 | 13 | 23 | | GGC | 36 | 12 | 19 | 40 |
| Ala | GCT | 78 | 34 | 27 | 19 | | GGA | 83 | 29 | 44 | 9 |
| | GCC | 102 | 45 | 16 | 25 | | GGG | 63 | 22 | 6 | 13 |
| | GCA | 37 | 16 | 41 | 22 | Total codons | 3,554 | | | | |

[a]Individual codon percentage for each amino acid type
[b]Data obtained from the KDRI Codon Usage Database (http://www.kazusa.or.jp/codon/)

**Table 4.** Excess amino acids in *H. orenii* and other members of domains *Bacteria* and *Archaea*

| Organism | *Halothermothrix orenii* | *Haloanaerobium praevalens*[b] | *Haloferax volcanii*[c] | *Aquifex aeolicus*[c] |
|---|---|---|---|---|
| Excess acidic amino acids[a] | –11.46 | 120.0 | 84.41 | –19.5 |

[a]Excess acidic over basic amino acids calculated according to the following (Asp+Glu)–(His+Lys+Arg) and expressed as amino acid frequency per 1,000 residues
[b]Determined chemically from whole-cell proteins (Oren 1986)
[c]Determined from the KDRI Codon Usage Database (http://www.kazusa.or.jp/codon/)

than 1 M. The acidic surface amino acid approach to stability and activity at high salt levels may therefore be advantageous for mesophilic, moderately halophilic anaerobes that do not require thermostable enzymes. *Halothermothrix orenii* requires enzymes that are stable and active at high temperatures and a range of salt concentrations (5%–20%) and therefore appears to have developed enzymes with distinct structural adaptations for stability and activity under these conditions.

## References


Benson DA, Boguski MS, Lipman DJ, Ostell J, Ouellette BF, Rapp, BA, Wheeler DL (1999) GenBank. Nucleic Acids Res 27(1):12–17

Bonete MJ, Pire C, Lorca FI, Comacho ML (1996) Glucose dehydrogenase from the halophilic Archaeon *Haloferax mediterranei*: enzyme purification, characterisation and N-terminal sequence. FEBS Lett 383(3):227–229

Camacho ML, Brown RA, Bonete MJ, Danson MJ, Hough DW (1995) Isocitrate dehydrogenases from *Haloferax volcanii* and *Sulfulobus solfataricus*: enzyme purification, characterisation and N-terminal sequence. FEMS Microbiol Lett 134(1):85–90

Cayol JL, Ollivier B, Patel BKC, Prensier G, Guezennec J, Garcia JL (1994) Isolation and characterization of *Halothermothrix orenii* gen. nov., sp. nov., a halophilic, thermophilic, fermentative, strictly anaerobic bacterium. Int J Syst Bacteriol 44(3):534–540

Cayol JL, Ducerf S, Patel BK, Garcia JL, Thomas P, Ollivier B (2000) *Thermohalobacter berrensis* gen. nov., sp. nov., a thermophilic, strictly halophilic bacterium from a solar saltern. Int J Syst Evol Microbiol 50(pt 2):559–564

Curatti L, Folco E, Desplats P, Abratti G, Limones V, Herrera-Estrella L, Salerno G (1998) Sucrose-phosphate synthase from *Synechocystis* sp. strain PCC 6803: identification of the *spsA* gene and characterisation of the enzyme expressed in *Escherichia coli*. J Bacteriol 180(24):6776–6779

Garcin E, Vernede X, Hatchikian EC, Volbeda A, Frey M, Fontecilla-Camps JC (1999) The crystal structure of a reduced [NiFeSe] hydrogenase provides an image of the activated catalytic center. Struct Fold Des 7(5):557–566

Jennings MP, Beacham IR (1989) MacConkey agar as an alternative to X-gal in the detection of recombinant plasmids. Biotechniques 7(10):1082

Kaieda N, Wakagi T, Koyama N (1998) Presence of $Na^+$ stimulated V-type ATPase in the membrane of a facultatively anaerobic and halophilic alkaliphile. FEMS Microbiol Lett 167: 57–61

Keradjopoulos D, Holldorf AW (1977) Thermophilic character of enzymes from extreme halophilic bacteria. FEMS Microbiol Lett 1:179–182

Kunst Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Cummings NJ, Daniel RA, Denizot F, Devine KM, Dusterhoft A, Ehrlich SD, Emmerson PT, Entian KD, Errington J, Fabret C, Ferrari E, Foulger D, Fritz C, Fujita M, Fujita Y, Fuma S, Galizzi A, Galleron N, Ghim S-Y, Glaser P, Goffeau A, Golightly EJ, Grandi G, Guiseppi G, Guy BJ, Haga K, Haiech J, Harwood CR, Henaut A, Hilbert H, Holsappel S, Hosono S, Hullo M-F, Itaya M, Jones L, Joris B, Karamata D, Kasahara Y, Klaerr-Blanchard M, Klein C, Kobayashi Y, Koetter P, Koningstein G, Krogh S, Kumano M, Kurita K, Lapidus A, Lardinois S, Lauber J, Lazarevic V, Lee S-M, Levine A, Liu H, Masuda S, Mauel C, Medigue C, Medina N, Mellado RP, Mizuno M, Moestl D, Nakai S, Noback M, Noone D, OReilly M, Ogawa K, Ogiwara A, Oudega B, Park S-H, Parro V, Pohl TM, Portetelle D, Porwollik S, Prescott AM, Presecan E, Pujic P, Purnelle B, Rapoport G, Rey M, Reynolds S, Rieger M, Rivolta C, Rocha E, Roche B, Rose M, Sadaie Y, Sato T, Scanlan E, Schleich S, Schroeter R, Scoffone F, Sekiguchi J, Sekowska A, Seror SJ, Serror P, Shin B-S, Soldo B, Sorokin A, Tacconi E, Takagi T, Takahashi H, Takemaru K, Takeuchi M, Tamakoshi A, Tanaka T, Terpstra P, Tognoni A, Tosato V, Uchiyama S, Vandenbol M, Vannier F, Vassarotti A, Viari A, Wambutt R, Wedler E, Wedler H, Weitzenegger T, Winters P, Wipat A, Yamamoto H, Yamane K, Yasumoto K, Yata K, Yoshida K, Yoshikawa H-F, Zumstein E, Yoshikawa H, Danchin A (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature (Lond) 390(6657):249–256

Lanyi JK (1974) Salt dependent properties of proteins from extremely halophilic bacteria. Bacteriol Rev 38(3):272–290

Lunn JE, Price GD, Furbank RT (1999) Cloning and expression of a prokaryotic sucrose-phosphate synthase gene from the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Mol Biol 40:297–305

Madern D, Ebel C, Zaccai G (2000) Halophilic adaptations of enzymes. Extremophiles 4: 91–98

Madigan MT, Oren A (1999) Thermophilic and halophilic extremophiles. Curr Opin Microbiol 2:265–269

Matheson AT, Louie KA, Tak BD, Zuker M (1987) The primary structure of the ribosomal A-protein (L12) from the halophilic eubacterium *Haloanaerobium praevalens*. Biochimie (Paris) 69(10):1013–1020

Oren A (1986) Intracellular salt concentrations of the anaerobic halophilic eubacteria *Haloanaerobium praevalens* and *Halobacteroides halobious*. Can J Microbiol 32:4–9

Patel BKC, Morgan HW, Daniel RM (1985a) *Fervidobacterium nodosum* gen. nov., a new chemoorganotrophic caldoactive anaerobic bacterium. Arch Microbiol 14:63–69

Patel BKC, Morgan HW, Daniel RM (1985b) A simple and efficient method for preparing anaerobic media. Biotechnol Lett 7:227–228

Rengpipat S, Lowe SE, Zeikus JG (1988) Effect of extreme salt concentrations on the physiology and biochemistry of *Halobacteroides acetoethylicus*. J Bacteriol 170(7):3065–3071

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Shibui H, Hamamoto T, Yohda M, Kagawa Y (1997) The stabilising residues and the functional domains in the hyperthermophilic V-ATPase of *Desulfurococcus*. Biochem Biophys Res Commun 234:341–345

Takase K, Yamato L, Kakinuma Y (1993) Cloning and sequencing of the genes coding for the A and B subunits of vacuolar-type $Na^+$-ATPase from *Enterococcus hirae*. J Biol Chem 269:11037–11044

Tsutsumi S, Denda K, Yokoyama K, Oshima T, Date T, Yoshida M (1991) Molecular cloning of the genes encoding the major two subunits of a eubacterial V-type ATPase from *Thermus thermophilis*. Biochim Biophys Acta 1098:13–20

Verhagen MF, O'Rourke T, Adams MW (1999) The hyperthermophilic bacterium, *Thermotoga maritima*, contains an unusually complex iron-hydrogenase: amino acid sequence analyses versus biochemical characterization. Biochim Biophys Acta 1412(3):212–229

zu Bentrup K, Ubbink-Kok T, Lolkema JS, Konings WN (1997) An $Na^+$-pumping $V_1V_0$-ATPase complex in the thermophilic bacterium *Clostridium fervidus*. J Bacteriol 179(4):1274–1279