Molecular Evolution of H^+-ATPases. I. *Methanococcus* and *Sulfolobus* are Monophyletic with Respect to Eukaryotes and Eubacteria

Johann Peter Gogarten, Thomas Rausch, Paul Bernasconi, Henrik Kibak, and Lincoln Taiz

Biology Department, Thimann Laboratories, University of California, Santa Cruz, California 95064

Z. Naturforsch. **44c**, 641–650 (1989); received April 6, 1989

Molecular Evolution, *Methanobacteria*, Archaebacteria, ATPase, Polymerase Chain Reaction

The classification of methanogenic bacteria as archaebacteria based on 16 s rRNA sequence analysis is currently in dispute. To provide an alternative molecular marker, the polymerase chain reaction technique was used to amplify a 930 bp fragment of *Methanococcus thermolithotrophicus* genomic DNA corresponding to the catalytic domain of the membrane H^+-ATPase. The deduced amino acid sequence was 54–58% identical to the ~ 70 kDa subunits of *Sulfolobus acidocaldarius* and the eukaryotic vacuolar-type H^+-ATPase, and only 29% identical to the beta subunit of the eubacterial-type F_0F_1-ATPases. Interestingly, a highly conserved aspartate residue in the phosphorylation domain of E1E2-ATPases (P-type) is conserved in the *Methanococcus* sequence, but is absent from all other known vacuolar and F_0F_1-ATPases. This suggests that the H^+-ATPase of *M. thermolithotrophicus*, like that of *M. voltae*, may have a phosphorylated intermediate, despite belonging to the vacuolar-type class of proton pumps. Phylogenetic analysis using Felsenstein's maximum likelihood method and Lake's evolutionary parsimony method confirmed that the H^+-ATPases of the two archaebacteria, *Methanococcus* and *Sulfolobus*, when compared to eukaryotic vacuolar-type ATPases and eubacterial F_0F_1-ATPases, form a monophyletic group.

Introduction

As first proposed by Woese, Fox and colleagues [1–3], the Archaebacteria include the methanogens, extreme halophiles and sulfur metabolizing thermoacidophiles. Archaebacteria are distinguished from the Eubacteria on the basis of their 16 s rRNA sequences, the presence of ether-linked lipids in their membranes and other biochemical criteria. A third, equally distant branch from the common ancestor, or progenote, was hypothesized to lead to the eukaryotes [4]. Recently, Lake has challenged the concept of a monophyletic archaebacterial kingdom based on studies or ribosome morphology [5] and a re-analysis of 16 s rRNA sequences using a rate-independent method [6, 7]. According to the Lake tree, the methanogens and halobacteria group with the other bacteria, while the sulfur-dependent thermoacidophiles and the eukaryotic host cell form a separate "eocyte" branch diverging from the common ancestor [7]. Given the conflicting results from the analysis of 16 s rRNA, phylogenetic studies of other molecular markers are needed.

H^+-pumping ATPases are highly conserved, probably ubiquitous enzymes which are proving to be useful for evolutionary comparisons of highly divergent species [8–10]. H^+-ATPases have been divided into three main categories: P (or E1E2-type), F (or F_0F_1-type) and V (or vacuolar type) [11, 12]. P-ATPases function as cation pumps on the plasma membranes of eukaryotes and eubacteria (*e.g. E. coli* [13] and *Streptococcus faecalis* [14]). The diagnostic feature of the P-ATPases is the formation of a phosphoenzyme intermediate during catalysis. F-ATPases are present in mitochondria and chloroplasts of eukaryotes, and in the plasma membranes of most eubacteria. V-ATPases, the most recently discovered class of H^+-ATPases, are present on membranes of the vacuolar system in eukaryotic cells where they serve to acidify intracellular compartments. Until now, neither F- or the V-type ATPases have been found to form a phosphoenzyme catalytic intermediate [11, 12].

Abbreviations: 60's, ~60 kDa noncatalytic subunits of V-ATPases; 70's, ~70 kDa catalytic subunits of V-ATPases; F-ATPases, H^+ pumping ATPases of the eubacterial type (F_0F_1-ATPases) P-ATPases, ATPases that have a phosphorylated catalytic intermediate; PCR, polymerase chain reaction; *Taq*-polymerase, heat resistant DNA polymerase from *Thermus aquaticus*; V-ATPases, H^+ pumping ATPases of the vacuolar type.

Reprint requests to Dr. J. P. Gogarten.

Verlag der Zeitschrift für Naturforschung, D-7400 Tübingen
0341–0382/89/0700–0641 $ 01.30/0

The membrane ATPases of a number of archaebacteria have been characterized biochemically. These studies have indicated that the H^+-ATPases of *Halobacterium halobium* [15], *H. saccharovorum* [16], *Sulfolobus acidocaldarius* [17, 18] and *Methanosarcina barkeri* [19] are in some respects more similar to V-ATPases than to F-ATPases. An exception seems to be the *Methanococcus voltae* ATPase, which forms a phosphoenzyme intermediate and is therefore classified as a P-type ATPase [20].

The sequences of the α and β subunits of *S. acidocaldarius* [21, 22] were recently compared with those of the β and α subunits, respectively, of the F-ATPases, and the 70 and 60 kDa subunits, respectively, of the V-ATPases [10]. The *Sulfolobus* subunits were 50% identical to the V-ATPase subunits, and only 25% identical to the F-ATPase subunits, clearly indicating that the *Sulfolobus* ATPase is a V-type H^+-ATPase. Moreover, since the two types of subunits (60's and 70's in V-, and α and β in F-ATPases) are homologous to each other, they presumably arose by gene duplication prior to the divergence of Archaebacteria, Eubacteria and Eukaryotes from the common ancestor. This view is supported by the phylogenetic analysis of the available sequences. Based on this analysis we have proposed that the progenote diverged into two main branches: one giving rise to the Eubacteria, the other leading to *Sulfolobus* and the eukaryotic host cell [10].

While the above scheme is consistent with Lake's tree, it leaves unanswered the question of the relationship between *Sulfolobus* and the other Archaebacteria. Using oligonucleotide primers directed to the highly conserved flanking sequences of the core region of the V-ATPase 70 kDa subunit, we amplified and subcloned a 930 nucleotide fragment of *Methanococcus thermolithotrophicus* genomic DNA by the polymerase chain reaction (PCR) technique. The results show that *Methanococcus* and *Sulfolobus* belong to a monophyletic group with respect to the eubacteria and eukaryotes.

Materials and Methods

DNA source

Genomic DNA from *Methanococcus thermolithotrophicus* was a gift from Dr. Jordan Konisky, University of Illinois.

Primer synthesis

The design of oligonucleotide primers representing consensus sequences in highly conserved regions of other already known H^+-ATPase 70 kDa subunits is explained in detail in the Results section.

Polymerase chain reaction (PCR)

The reaction volume was 100 µl containing 0.2 µg genomic DNA (template), 1 µM each primer, 10 mM Tris/HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, 1.25 mM of each nucleotide (dATP, dCTP, dTTP, dGTP) and 2.5 units of *Taq*-polymerase. The mix was overlayered with mineral oil. The temperature program for amplification (run in a Perkin-Elmer Cetus DNA Thermal Cycler, Norwalk, CT, U.S.A.) was as follows: 5 min at 94 °C (initial melting) followed by 10 amplification cycles with a slow temperature ramp between 37 °C and 72 °C (1 min 94 °C – ramp 1 °C/sec – 2 min 37 °C – ramp 7 °C/min – 3 min 72 °C – ramp 1 °C/sec), and by 20 more amplification cycles omitting the slow ramp. The final 72 °C step was extended to 20 min.

The reaction mixtures were directly separated by electrophoresis on 1% agarose gels, and the amplified products were visualized with ethidium bromide. The DNA was eluted by adsorption to glass powder according to [23].

Cloning and sequencing

After ApaI digestion the eluted amplified product was cloned into ApaI/EcoRV-cut pBluescript (pBluescript DNA sequencing system, Stratagene, La Jolla CA, U.S.A.). Subclones were generated using the restriction sites for HindIII (261), EcoRI (580) and PstI (750). The whole sequence was covered by at least two different subclones. Double stranded sequencing was performed with sequenase (U.S. Biochem. Corp.) [24] using [α^{35}S]thiodATP as label [25]. Subclones were sequenced in both directions such that the final sequence was the result of four independent sequencing reactions. To assess the fidelity of the *Taq*-polymerase the products of two other independent PCRs were partially sequenced (see Results section).

Chemicals

Taq-polymerase from Perkin-Elmer Cetus. Restriction enzymes from Bethesda Research Laboratories. [α^{35}S]ThiodATP (specific activity 1000 Ci/mmol) from New England Nuclear.

```
D. carota              408 ACACAAGCCCCTATCTGTGGAGTTGGGTCCTGG
N. crassa              830 AGGCAAGCCTCTCTCTGTCGGACTTGGCCCTGG
S. acidocaldarius      201 TGGAGCCCCATTATCAGTAGAATTAGGTCCTGG

LEFT PRIMER:            5' AGGCGGGCCCCTATCTGTCGAATTTGGTCCTGG 3'
                               ApaI


D. carota             1387 TGCTGTTTCACCTCCAGGAGGAGATTTTTCT
N. crassa             1805 TGCTGTCTCTCCCCCCGGTGGTGATTTCTCT
S. acidocaldarius     1167 TGCAGTTTCTCCTCCTGGAGGTGATTTTACA

                        5' TGCTGTTTCTCCTCCCGGAGGTGATTTTAAA 3'
RIGHT PRIMER:           3' ACGACAAAGAGGAGGGCCTCCACTAAAATTT 5'
                                                   Dra I
```

Fig. 1. Sequences of the left and right primer used for the specific amplification of genomic DNA from *Methanococcus thermolithotrophicus* coding for part of the catalytic subunit of the "archaebacterial-vacuolar ATPase". The primer sequences are shown aligned to the DNA sequences from *Sulfolobus acidocaldarius*, *Neurospora crassa* and *Daucus carota* that were used to construct the primers. The complementary strand (bottom line) was used in the case of the right primer in order to prime the synthesis of the strand in the direction of the left primer. The conserved regions selected to design the primers are approx. 950 bp apart in the cDNA sequences. Mismatches between primers and cDNA sequences are emphasized by bold type face. A restriction site (ApaI/DraI) was included at the end of each primer in order to facilitate subcloning. Note that most of the mismatches are closer to the 5'end of the primer.

Results and Discussion

Using the set of primers depicted in Fig. 1, we were able to amplify a 994 bp long DNA fragment (including the primers) from *M. thermolithotrophicus* genomic DNA, which was subcloned and sequenced as indicated in Fig. 2. The DNA sequence of this fragment and the encoded amino acid sequence are given in Fig. 3.

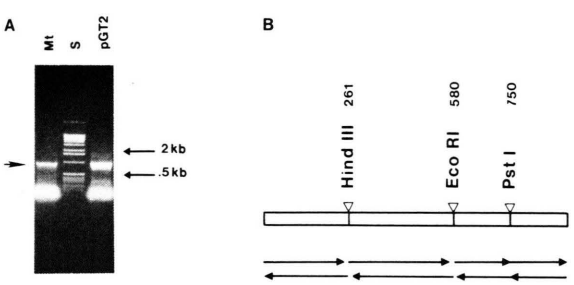

Fig. 2. Result of the amplification by PCR.
A: Separation on a 1% agarose ethidium bromide stained gel.
Mt: PCR started with 200 ng of genomic *M. thermolithotrophicus* DNA.
pGT2: PCR run with approx. 1 ng of a plasmid bearing the cDNA coding for the carrot 70's subunit.
S: DNA size standards (BRL).
Arrow indicates the amplified product.
B: Restriction map of and sequencing strategy for the cloned Mt sequence. Numbering does not include the primer sequences. Arrows indicate direction and extent of sequencing.

Fidelity of the DNA polymerase

As the *Taq*-polymerase lacks exonuclease activity, it has no proof reading activity and therefore exhibits a relatively high error rate [26]. We assessed the fidelity of the enzyme under the chosen reaction conditions by re-sequencing 420 nucleotides from the ends, using fragments derived from two other independent amplification reactions. Since the resulting sequences were identical, the likely error rate of the enzyme is $< 0.24\%$ with regard to the final sequence and $< 2.7 \times 10^{-5}$ per incorporated base during each amplification step.

The amino acid sequence

The peptide encoded by the amplified region is very similar to the core region of the catalytic subunit of V-ATPases and even more similar to that of the *S. acidocaldarius* enzyme (Fig. 4). The similarity to the β-subunits of F-ATPases is much less, although still recognizable (Fig. 4). The *M. thermolithotrophicus* sequence contains most of the residues that have been implicated in ATP binding and ATP hydrolysis [27, 28] (*e.g.* the "G-K-T-V"-region, residues 159–162 in Fig. 3). The amplified fragment

```
5'  AGGCGGGCCCCTATCTGTCGAATTTGGTCCTGG  ATGATGAAAGCGATGTACGATGGTATTCAAAGGCCATTA  39
          LEFT PRIMER                      M  M  K  A  M  Y  D  G  I  Q  R  P  L  13

ACTGCAATCGAAACCGAAACTGGATCAATATTTATTCCAAGAGGTGTACATGTTCCTTCATTACCAAGGGATGTAAAA 117
 T  A  I  E  T  E  T  G  S  I  F  I  P  R  G  V  H  V  P  S  L  P  R  D  V  K  39

TGGGACTTCACACCAACCGTAAACGTTGGGGATATAGTTAGCGAAGGAGACGTTATAGGTACAGTACCTGAAACCCCA 195
 W  D  F  T  P  T  V  N  V  G  D  I  V  S  E  G  D  V  I  G  T  V  P  E  T  P  65

TCCATTATCCACAAAATAATGGTTCCAGTTGGTGTAAACGGTAAAGTAAAAGAAATAAAATCAGGAAGCTTTACCGTA 273
 S  I  I  H  K  I  M  V  P  V  G  V  N  G  K  V  K  E  I  K  S  G  S  F  T  V  91

GAAGAAACCGTTGCTGTTATTGAAACAGAAAACGGAGATAAAGAAGTTACAATGATGCAAAAGTGGCCAGTAAGAAAG 351
 E  E  T  V  A  V  I  E  T  E  N  G  D  K  E  V  T  M  M  Q  K  W  P  V  R  K  117

CCAAGACCACACAAAGGAAAATTGCCACCAATTATCCCATTAATTACAGGTCAGAGGGTAGAAGATACATTCTTTGGT 429
 P  R  P  H  K  G  K  L  P  P  I  I  P  L  I  T  G  Q  R  V  E  D  T  F  F  G  143

CTTGCTAAAGGAGGTACTTCAGCTATTCCTGGTCCATTCGGAAGTGGTAAAACTGTTACACAGCACCAGCTTGCTAAG 507
 L  A  K  G  G  T  S  A  I  P  G  P  F  G  S  G  K  T  V  T  Q  H  Q  L  A  K  169

TGGTCAGATGTTGACGTTGTGGTTTACATCGGATGTGGGGAAAGAGGAAACGAGATGACTGAAGTTATTGAAGAATTC 585
 W  S  D  V  D  V  V  V  Y  I  G  C  G  E  R  G  N  E  M  T  E  V  I  E  E  F  195

CCACACTTGGAAGACATTAAAACAGTAAACAAGTTAATGGATAGGACTGTTTTAATTGCAAACACCTCAAACATGCCG 663
 P  H  L  E  D  I  K  T  V  N  K  L  M  D  R  T  V  L  I  A  N  T  S  N  M  P  221

GTTGCTGCGAGGGAAGCATCAGTTTACACAGGAATTACAATTGCAGAATACTTTAGAGACATGGGATACGGTGTTCTC 741
 V  A  A  R  E  A  S  V  Y  T  G  I  T  I  A  E  Y  F  R  D  M  G  Y  G  V  L  247

TTAACTGCAGACTCAACCTCAAGATGGGCAGAGGCAATGAGAGAGATTTCAGGAAGATTGGAAGAAATGCCAGGGGAA 819
 L  T  A  D  S  T  S  R  W  A  E  A  M  R  E  I  S  G  R  L  E  E  M  P  G  E  273

GAAGGTTACCCTGCTTACCTTGCCTCAAGACTTGCACAGTTCTACGAAAGAGCTGGAAGAGTTAACTGTTTAGGTTCC 797
 E  G  Y  P  A  Y  L  A  S  R  L  A  Q  F  Y  E  R  A  G  R  V  N  C  L  G  S  299

                                ACGACAAAGAGGAGGGCCTCCACTAAAATTT  5'
GATGACAAGCAAGGATTTATCTGTATTGTTGGT                                3'          930
 D  D  K  Q  G  F  I  C  I  V  G          RIGHT PRIMER                       310
```

Fig. 3. Nucleotide sequence and encoded amino acid sequence of the PCR amplified DNA fragment. The sequences given for the primers are the one S used for amplification, *i.e.* the sequence of the right primer corresponds to the complementary strand.

also contains the so-called non-homologous region (residues 10–110 in Fig. 3). This region is present in the catalytic subunit (70's) of V-ATPases [27, 28], including the membrane ATPase of *Sulfolobus acidocaldarius* [10, 21], but is absent from the homologous [27] catalytic subunits (β) of F-ATPases, and it is also absent from the noncatalytic subunits of both types of enzymes (60's in V- and α's in F-ATPases) [10]. The latter are also derived from the same ancestral gene in the progenote (paralogous subunits) [10, 30].

The sequence immediately in front of the right primer (298–309 in Fig. 3) corresponds to the region in the F-ATPase that has been suggested to be homologous (analogous?) to the phosphorylation domain of the P-ATPases [31]. This region is particularly conserved among the various V-ATPases as well as between F- and V-ATPases, but is surprisingly dissimilar between the *M. thermolithotrophicus* sequence and other V-ATPases (Fig. 4 and Table I). Furthermore, it contains an aspartate residue followed by lysine in the same positions where the phosphorylated aspartate is located in the P-ATPases (Table I). Together with the observation that the H^+-ATPase of *Methanococcus voltae*, which is known to be a P-type ATPase, has some properties that are more typical of V- or F-ATPases than of P-ATPases (*e.g.* dissociation of the catalytic subunit by high salt buffers [20]), the conservered aspartate suggests the possibility that the *Methanococci* might have a V-type ATPase that becomes phosphorylated during the catalytic cycle.

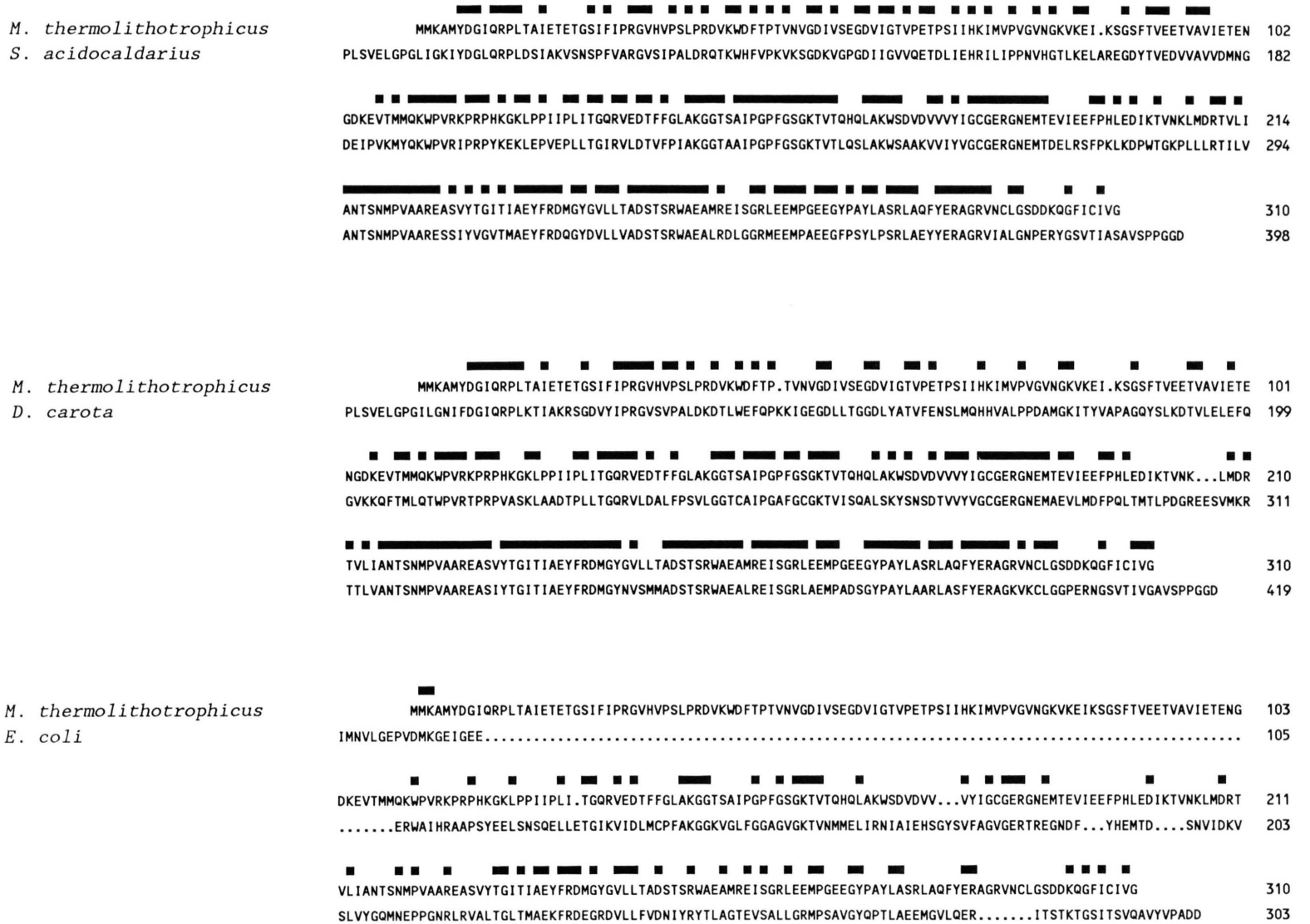

Fig. 4. Binary alignments of the *Methanococcus thermolitotrophicus* sequence with the homologous sequences from an archaebacterium (A) [21], an eukaryote (B) [27], and an eubacterium (C) [28]. Solid bars above the alignments indicate identical residues. The large gap in the *E. coli* sequence is discussed in the text and in reference [27].

Table I. Comparison between the sequences around the aspartate residue that is phosphorylated in P-ATPases and sequences from F- and V-ATPases that have been proposed to be homologous to this sequence. The aspartate that is phosphorylated during the catalytic cycle in P-ATPases and the aspartate in the corresponding position of the *Methanococcus* sequence is emphasized by bold type face. C/M denote subunits from chloroplast and mitochondrial ATPases respectively.

P-ATPases:												
Escherichia KdP [32]	L	L	**D**	K	T	G	T	I	T	L	G	315
Neurospora PM [33]	C	S	**D**	K	T	G	T	L	T	K	N	386
Homo Na^+/K^+ [34]	C	S	**D**	K	T	G	T	L	T	Q	N	384
Oryctolagus Ca^{2+} [35]	C	S	**D**	K	T	G	T	L	T	T	N	359
Archaebacterial and V-ATPases:												
70-subunits:												
M. thermolithotr.	S	D	**D**	K	Q	G	F	I	C	I	V	309
Sulfolobus [21]	N	P	E	R	Y	G	S	V	T	I	A	389
Daucus [27]	G	P	E	R	N	G	S	V	T	I	V	410
Neurospora [29]	S	P	P	R	E	G	S	V	S	I	V	403
60-subunits:												
Saccharomyces [36]	V	E	G	R	N	G	S	I	T	Q	I	332
Neurospora [37]	V	A	G	R	N	G	S	I	T	Q	I	326
Arabidopsis [38]	I	E	G	R	K	G	S	I	T	Q	I	333
Sulfolobus [22]	V	I	G	K	K	G	S	I	T	Q	M	305
F-ATPases:												
β-subunits:												
Anabaena [39]	T	S	T	T	E	G	S	I	T	S	I	303
Escherichia [28]	T	S	T	K	T	G	S	I	T	S	V	294
Rhodospirillum [40]	T	S	T	K	K	G	S	I	T	S	V	303
Bos [41]	T	T	T	K	K	G	S	I	T	S	V	307
Nicotiana C [42]	T	S	T	K	E	G	S	I	T	S	I	324
Zea C [43]	T	S	T	K	K	G	S	I	T	S	I	324
α-subunits:												
Nicotiana C [44]	S	S	L	G	E	G	S	M	T	A	L	317
Saccharomyces M [45]	E	K	E	G	S	G	S	L	T	A	L	361
Rhodospirillum [46]	D	D	N	G	A	G	S	L	T	A	L	325
Zea M [47]	D	Q	T	G	A	G	S	L	T	A	L	327
Escherichia [48]	V	K	G	K	T	G	S	L	T	A	L	327

Phylogenetic relationships

Based on the alignments given in Fig. 4, it is obvious that the *M. thermolithotrophicus* sequence is much more similar to the V-ATPases of eukaryotes than to the eubacterial enzyme. Closer inspection reveals that the *M. thermolithotrophicus* sequence is slightly more similar to the *Sulfolobus* sequence (58% identical residues) than to the eukaryotic subunit (54% identical residues). However, one cannot infer the phylogenetic relationships directly from the degree of similarity. This is because sequences which evolve more slowly might be more similar to each other than to more rapidly evolving sequences, even though they might have a more recent common ancestor with the latter.

To deduce the phylogenetic relation among the different ATPase subunits, and to assess the significance of the derived branching order, we employed a number of analytical programs. Since none of the programs that allow for statistical evaluation of the various possible branching orders is currently able to use protein sequences, the nucleotide sequences were aligned according to the protein sequence alignment.

Protein sequences from a number of ATPase subunits (ref. Fig. 5) were aligned to the *M. thermolithotrophicus* sequence using Feng and Doolittle's multiple alignment program [49]. In order to faithfully preserve the non-homologous region with respect to the α and β subunits, it was necessary to set the GAPPEN parameter to 16. This led to a mis-

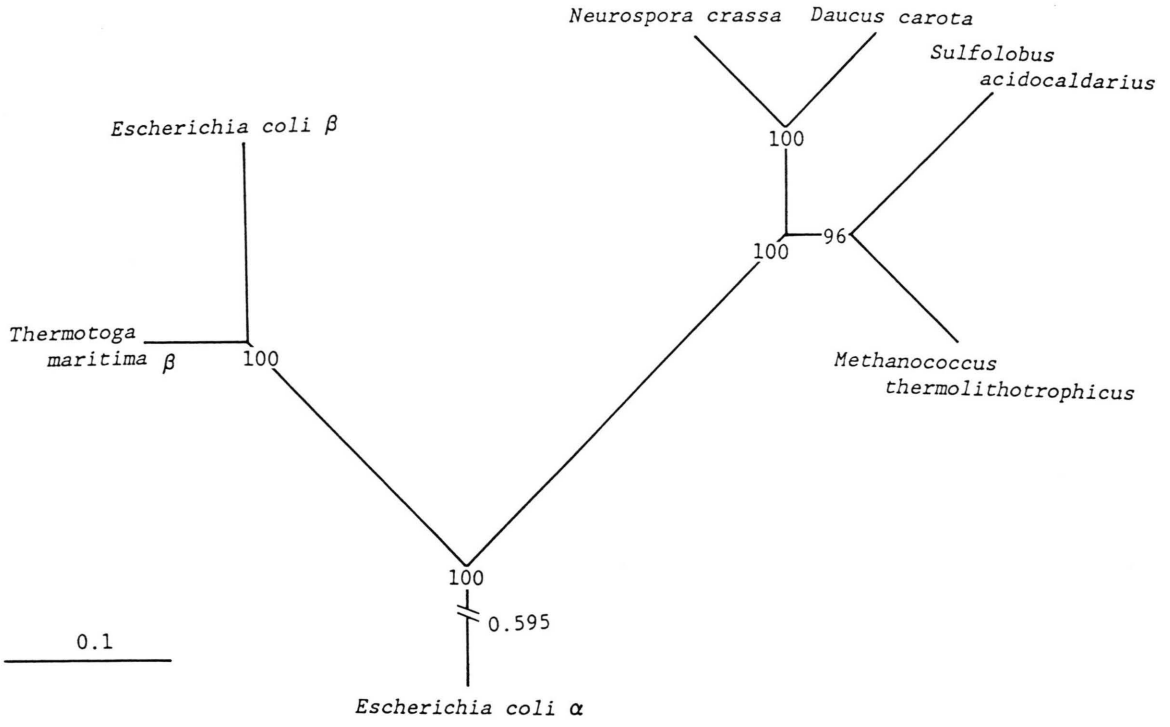

Fig. 5. Phylogenetic tree for ATPase subunits. The topology and the branch lengths of the tree were calculated using Felsenstein's maximum likelihood method as implemented in [55]. The model parameters were chosen according to the evolution of the beta subunits of F-ATPases encoded in chloroplasts [52], *i.e.* the relative probability for substitution of the three codon positions (C) was set to 2-1-10 and the transition/transversion ratio (T) at 3.3; however, the same result (topology and significantly positive branches) was obtained with widely different parameters (*e.g.* C: 1-1-4 and T: = 2). The branch lengths are scaled in terms of expected numbers of substitutions, and in terms of the probability for change of the first base of the codon. All branches were determined to be significantly positive at a level of $p \leq 1\%$. The numbers written at the bifurcations give the percentage of bootstrapped samples in which the same group of species as in the depicted tree is found attached to one node, although, not necessarily with the same topology inside this group. 100 bootstrapped samples were evaluated with DNA parsimony as implemented in [55].

alignment of a small region conserved in all catalytic subunits which was corrected by hand (see Fig. 4 *M. t.* 177-V-Y-I-G-C-G-E-R-G-N-E-187). This region normally aligns correctly at traditional (lower) gap penalties.

The tree for the evolution of the ATPase subunits displayed in Fig. 5 was calculated using Felsenstein's maximum likelihood method [50]. If one excludes the possibility of lateral gene transfer, the topology of this tree can be directly translated into the evolution of organisms, although the branch lengths probably become meaningless since little is known about how the rate of substitution has changed during evolution [53]. It seems that the deeper branches in the ATPase trees are exaggerated when compared to trees obtained with other sequences (*e.g.* 16 s rRNA [7, 54] or ribosomal A protein [55, 56]) and are probably also exaggerated with respect to time. One possible explanation for this is that the ribosome at the time of the progenote was already highly optimized by evolution, whereas the optimization of semipermeable membranes energized by ATP-hydrolyzing enzymes is probably more recent. However, as the gene duplication that gave rise to the noncatalytic α and 60 subunits had already occurred in the progenote [10], the location where the α-subunit joins the tree for the catalytic subunits (Fig. 5) gives the location of the progenote in the tree for the phylogeny of organisms.

All branches, including the one connecting the two Archaebacteria to the branch leading to the eukaryotes, were significantly positive at the $P = 0.01$ level

according to the likelihood ratio test performed with Felsenstein's DNAML program [51].

Due to the redundancy of the genetic code the third base of each codon is subjected to much less evolutionary pressure and therefore has a much higher rate of change than the other nucleotides. Therefore, in order to increase the signal to noise ratio, we deleted all third base positions from the input files for programs that did not allow for different weights (*i.e.* DNABOOT [51] and evolutionary parsimony [6, 57]).

When analyzed by bootstrapping [51] (= random sampling of positions from the aligned sequences and evaluation of the sampled sequences), in 97% of the bootstrapped samples *Methanococcus* and *Sulfolobus* grouped together when analyzed by DNA-parsimony [51] (Fig. 5).

If data are analyzed by normal parsimony methods, grossly unequal rates of change in different lineages can lead to consistently wrong results, *i.e.* the more sequence data are used, the more surely the result of the analysis is wrong [6, 54, 58].

We used Lake's method of invariants or evolutionary parsimony [6] to assess the effect of unequal rates on the grouping of the two archaebacterial sequences. Since this method uses only part of the information that is present in the sequences, the determined groupings are usually less significant than, for example, with maximum likelihood methods. However, evolutionary parsimony is insensitive to the effects of unequal rates [58, 6]. The outcome of the analysis with this algorithm is summarized in Fig. 6. As was the case with the other algorithms, the archaebacterial tree is supported, confirming [10] that in the case of the catalytic ATPase subunits the analysis is not hampered by unequal rate effects.

Future perspectives for the archaebacterial kingdom

One intriguing question is whether other methanogens and halobacteria belong to the same monophyletic group als *Sulfolobus* and *Methanococcus*. The fact that ATPases from *Sulfolobus* and *Halobacteria* immunologically crossreact is no unequivocal crite-

Tree Topology	Σ Invariants	P
(*Daucus carota* *Neurospora crassa*) — *Methanococcus thermolithotrophicus* / (*Escherichia coli* β *Thermotoga maritima* β) — *Sulfolobus acidocaldarius*	19	0.0198
	archaebacterial tree	
(*Daucus carota* *Neurospora crassa*) — (*Escherichia coli* β *Thermotoga maritima* β) / *Methanococcus thermolithotrophicus* — *Sulfolobus acidocaldarius*	9	0.2528
(*Daucus carota* *Neurospora crassa*) — *Methanococcus thermolithotrophicus* / *Sulfolobus acidocaldarius* — (*Escherichia coli* β *Thermotoga maritima* β)	-1	1.0000
	eocyte tree	

Fig. 6. Evidence for the archaebacterial tree hypothesis. The different tree topologies were tested with Lake's method of evolutionary parsimony as implemented in [57]. All possible four-species comparisons were performed with one species each from the following four groups 1: *M. thermolithotrophicus* 2: *S. acidocaldarius* 3: β-subunits from *E. coli* and *T. maritima* [59] and 4. eukaryotic 70-subunits from *N. crassa* and *D. carota*. The sum of the invariants for each tree topology is given in the second column and the probability that this or a higher sum might arise by chance alone is given in the last column. The first and the third tree topology have been proposed in the literature as eocyte tree (*Sulfolobus* has a more recent common ancestor together with eukaryotes) [7] and as archaebacterial tree (*Sulfolobus* and *Methanococcus* have more recent common ancestors) [3]. Only the topology of the archaebacterial tree is supported by this analysis.

rion, since crossreactivity has been also reported between the ATPases from *E. coli* and *Sulfolobus* [60]. The molecular sizes of the subunits of the halobacterial enzyme are quite different from those of either V- or F-ATPases [61, 62]. *Halobacteria* are the group of the Archaebacteria which according to Lake's analysis of the 16 s rRNA are least related to the *Eocytes* and which are most eubacterial like, a view which is supported by several other markers such as ferredoxin types [63] and gas vacuole protein [64].

Because they allow the placement of the progenote in the phylogenetic tree, F- and V-type ATPases are very useful markers for early organismal evolution [10]. Furthermore, since the sequences of the eubacterial F-ATPases and eukaryotic V-ATPases are homologous but sufficiently different, we expect that sequence information on the halobacterial H^+-ATPase subunits will make it possible to also decipher the phylogenetic position of *Halobacteria* presently in dispute.

Acknowledgements

We thank Jordan Konisky (Univ. of Illinois) for the generous gift of genomic DNA from various *Methanococci* and for stimulating discussions; Johann Peter Gogarten and Thomas Rausch are recipients of fellowships from the Deutsche Forschungsgemeinschaft, Paul Bernasconi is recipient of a fellowship from the Swiss National Foundation.

[1] C. R. Woese and G. E. Fox, J. Mol. Evol. **10,** 1–6 (1977).
[2] C. R. Woese and G. E. Fox, Proc. Natl. Acad. Sci. U.S.A. **74,** 5088–5090 (1977).
[3] G. E. Fox, E. Stackebrandt, R. B. Hespel, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese, Science **209,** 457–463 (1980).
[4] C. R. Woese, L. J. Magrum, and G. E. Fox, J. Mol. Evol. **11,** 245–252 (1978).
[5] J. A. Lake, E. Henderson, M. Oakes, and M. W. Clark, Proc. Natl. Acad. Sci. U.S.A. **81,** 3786–3790 (1984).
[6] J. A. Lake, Mol. Biol. Evol. **4,** 167–191 (1987).
[7] J. A. Lake, Nature **331,** 184–186 (1988).
[8] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay, EMBO J. **1,** 945–951 (1982).
[9] R. Amann, W. Ludwig, and K. H. Schleifer, J. Gen. Microbiol. **134,** 2815–2821 (1988).
[10] J. P. Gogarten, H. Kibak, P. Dittrich, L. Taiz, E. J. Bowman, B. J. Bowman, M. F. Manolson, R. J. Poole, T. Date, T. Oshima, J. Konishi, K. Denda, and M. Yoshida, Proc. Natl. Acad. Sci. U.S.A. (1989) [In press].
[11] P. L. Pedersen and E. Carafoli, Trends Biochem. Sci. **12,** 146–150 (1987).
[12] P. L. Pedersen and E. Carafoli, Trends Biochem. Sci. **12,** 186–189 (1987).
[13] J. E. Hesse, L. Wieczorek, K. Altendorf, A. S. Reicin, E. Dorus, and W. Epstein, Proc. Natl. Acad. Sci. U.S.A. **81,** 4746–4750 (1984).
[14] M. Solioz, S. Mathews, and P. Furst, J. Biol. Chem. **262,** 7358–7362 (1987).
[15] T. Nanba and Y. Mukohata, J. Biochem. **102,** 591–598 (1987).
[16] H. Kristjansson, M. H. Sadler, and L. I. Hochstein, FEMS Microbiol. Rev. **39,** 151–157 (1986).
[17] M. Lübben and G. Schäfer, Eur. J. Biochem. **164,** 533–540 (1987).
[18] J. Konishi, T. Wakagi, T. Oshima, and M. Yoshida, J. Biochem. **102,** 1379–1387 (1987).
[19] K. Inatomi, J. Bacteriol. **167,** 837–841 (1986).
[20] R. M. Dharmavaram and J. Konisky, J. Bacteriol. **169,** 3921–3925 (1987).
[21] K. Denda, J. Konishi, T. Oshima, T. Date, and M. Yoshida, J. Biol. Chem. **263,** 6012–6015 (1988).
[22] K. Denda, J. Konishi, T. Oshima, T. Date, and M. Yoshida, J. Biol. Chem. **263,** 17251–17254 (1988).
[23] B. Vogelstein and D. Gillespie, Proc. Natl. Acad. Sci. U.S.A. **76,** 615–619 (1979).
[24] S. Tabor and C. C. Richardson, Proc. Natl. Acad. Sci. U.S.A. **84,** 4767–4771 (1987).
[25] M. D. Biggin, T. J. Gibson, and G. F. Hong, Proc. Natl. Acad. Sci. U.S.A. **80,** 3963–3965 (1983).
[26] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich, Science **239,** 487–491 (1988).
[27] L. Zimniak, P. Dittrich, J. P. Gogarten, H. Kibak, and L. Taiz, J. Biol. Chem. **263,** 9102–9112 (1988).
[28] ATPase beta chain – Escherichia coli, M. Saraste, N. J. Gay, A. Eberle, M. J. Runswick, and J. E. Walker, Nucl. Acids Res. **9,** 5287–5296 (1981).
[29] E. J. Bowman, K. Tenney, and B. J. Bowman, J. Biol. Chem. **263,** 13994–14001 (1988).
[30] M. A. Ragan, BioSystems **21,** 177–188 (1988).
[31] L. Ernster, T. Hundal, B. Norling, G. Sandri, L. Wojtczak, V. A. Grinkevich, N. N. Modyanov, and Yu. A. Ovchinnikov, Chem. Scr. **26B,** 273–279 (1986).
[32] J. E. Hesse, L. Wieczorek, K. Altendorf, A. S. Reicin, E. Dorus, and W. Epstein, Proc. Natl. Acad. Sci. U.S.A. **81,** 4746–4750 (1984).
[33] R. Addison, J. Biol. Chem. **261,** 14896–14901 (1986).
[34] K. Kawakami, T. Ohta, H. Nojima, and K. Nagano, J. Biochem. **100,** 389–397 (1986).
[35] D. H. MacLennan, C. J. Brandl, B. Korczak, and N. M. Green, Nature **316,** 696–700 (1985).
[36] B. J. Bowman, R. Allen, M. A. Wechser, and E. J. Bowman, J. Biol. Chem. **263,** 14002–14007 (1988).
[37] H. Nelson, S. Mandiyan, and N. Nelson, J. Biol. Chem. **264,** 1775–1778 (1989).
[38] M. F. Manolson, B. F. Ouellette, M. Filion, and R. J. Poole, J. Biol. Chem. **263,** 17987–17994 (1988).

[39] S. E. Curtis, J. Bact. **169,** 80–86 (1987).
[40] G. Falk, A. Hampe, and J. E. Walker, Biochem. J. **228,** 391–407 (1985).
[41] M. J. Runswick and J. E. Walker, J. Biol. Chem. **258,** 3081–3089 (1983).
[42] K. Shinozaki, H. Deno, A. Kato, and M. Sugiura, Gene **24,** 147–155 (1983).
[43] E. T. Krebbers, I. M. Larrinua, L. McIntosh, and L. Bogorad, Nucl. Acids Res. **10,** 4985–5002 (1982).
[44] H. Deno, K. Shinozaki, and M. Sugiura, Nucl. Acids Res. **11,** 2185–2191 (1983).
[45] M. Takeda, W. J. Chen, J. Saltzgaber, and M. G. Douglas, J. Biol. Chem. **261,** 15126–15133 (1986).
[46] G. Falk, A. Hampe, and J. E. Walker, Biochem. J. **228,** 391–407 (1985).
[47] C. J. Braun and C. S. Levings III, Plant Physiol. **79,** 571–577 (1985).
[48] N. J. Gay and J. E. Walker, Nucl. Acids Res. **9,** 2187–2194 (1981).
[49] D. Feng and R. F. Doolittle, J. Mol. Evol. **25,** 351 (1987).
[50] J. Felsenstein, J. Mol. Evol. **17,** 368–376 (1981).
[51] J. Felsenstein, Phylogeny Inference Package, Version 3.1, Department of Genetics SK-50, University of Washington, Seattle, Washington 98195, U.S.A.
[52] G. Zurawski and M. T. Clegg, Ann. Rev. Plant Physiol. **38,** 391–418 (1987).
[53] M. Goodman, J. Czelusniak, B. F. Koop, D. A. Tagle, and J. L. Slighton, Cold Spring Harb. Symp. Quant. Biol. **52,** 875–890 (1987).
[54] G. Olsen, Cold Spring Harbor Symp. Quant. Biol. **52,** 825–837 (1987).
[55] M. Yaguchi, A. T. Matheson, L. P. Visentin, and M. Zuker, in: Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes, pp. 585–599, University of Tokyo Press (1980).
[56] T. Itoh, Eur. J. Biochem. **176,** 297–303 (1988).
[57] D. Swofford, Parsimony and U Analysis (PAUP) beta test version, Illinois State Natural History Survey, 607 E. Peabody, Champaign, IL 61820, U.S.A. (1989).
[58] J. Felsenstein, Nature **335,** 118 (1988).
[59] K. H. Schleifer, personal communication.
[60] M. Luebben, H. Luensdorf, and G. Schaefer, Eur. J. Biochem. **167,** 211–219 (1987).
[61] H. Stan-Lotter and L. I. Hochstein, Eur. J. Biochem. **170,** 155–160 (1989).
[62] T. Nanba and Y. Mukohata, J. Biochem. **102,** 591–598 (1987).
[63] D. Oesterhelt, L. Kerscher, S. Mathews, and F. Lottspeich, in: Proceedings of the International Symposium on Frontiers in Bioorganic Chemistry and Molecular Biology, pp. 327–334 (Elsevier Science Publishers, 1984).
[64] J. E. Walker, P. K. Hayes, and A. E. Walsby, J. Gen. Microbiol. **130,** 2709–2715 (1984).

Note added in proof: Meanwhile in cooperation with Drs. Helga and Friedrich-Wilhelm Bentrup from the University of Giessen we obtained a part of the corresponding sequence from *Methanococcus voltae*. The sequence around the putative phosphorylation site (see Table I) is: K-D-G-R-Q-G-A-V-T-A-I.