Phylogenetic analyses of the homologous transmembrane channel-forming proteins of the F_0F_1 -ATPases of bacteria, chloroplasts and mitochondria

Alan Blair, Linh Ngo, James Park, Ian T. Paulsen and Milton H. Saier, Jr

Author for correspondence: Milton H. Saier, Jr. Tel: +1 619 534 4084. Fax: +1 619 534 7108.
e-mail: msasier@ucsd.edu

Department of Biology,
University of California at
San Diego, La Jolla,
CA 92093-0116, USA

Sequences of the three integral membrane subunits (subunits a, b and c) of the F_0 sector of the proton-translocating F-type (F_0F_1 -) ATPases of bacteria, chloroplasts and mitochondria have been analysed. All homologous-sequenced proteins of these subunits, comprising three distinct families, have been identified by database searches, and the homologous protein sequences have been aligned and analysed for phylogenetic relatedness. The results serve to define the relationships of the members of each of these three families of proteins, to identify regions of relative conservation, and to define relative rates of evolutionary divergence. Of these three subunits, c-subunits exhibited the slowest rate of evolutionary divergence, b-subunits exhibited the most rapid rate of evolutionary divergence, and a-subunits exhibited an intermediate rate of evolutionary divergence. The results allow definition of the relative times of occurrence of specific events during evolutionary history, such as the intragenic duplication event that gave rise to large c-subunits in eukaryotic vacuolar-type ATPases after eukaryotes diverged from archaea, and the extragenic duplication of F-type ATPase b-subunits that occurred in blue-green bacteria before the advent of chloroplasts. The results generally show that the three F_0 subunits evolved as a unit from a primordial set of genes without appreciable horizontal transmission of the encoding genetic information although a few possible exceptions were noted.

Keywords: phylogenetic relationships, ATP synthase, F_0F_1 -ATPases, proton transport, membrane proteins

INTRODUCTION

Proton-translocating ATP synthases of mitochondria, chloroplasts and bacteria, also called F_0F_1 -ATPases or F-type ATPases, catalyse the reversible synthesis of ATP in response to a proton electrochemical gradient or proton-motive force by coupling the electrophoretic flux of protons through the F_0 membrane sector of the ATPase complex to the peripheral membrane F_1 -catalysed synthesis of ATP (Mitchell, 1979; Serrano, 1988; Futai *et al.*, 1989; Senior, 1990; Hatefi, 1993; Pedersen & Amzel, 1993). The proposed reaction sequence coupling ATP synthesis by the F_1 sector to H^+ flux via the F_0 sector of the ATP synthase is:

- (3) Energy transduction from F_0 to F_1
- (4) Release of bound ATP

When the proton-motive force is generated by oxidative electron flow in mitochondria or in bacteria, the process is termed oxidative phosphorylation, but when it is generated by light-dependent electron activation followed by electron flow in chloroplasts or in photosynthetic bacteria, the process is termed photosynthetic phosphorylation (Mitchell, 1979; Malmström, 1989). F-type ATPases are so efficient that they can turn over an organism's body weight in ATP several times in a single day (Pedersen & Amzel, 1993). Some bacterial ATPases can transport Na^+ instead of or in addition to H^+ , but the process appears to be mechanistically the same (Laubinger & Dimroth, 1988; Gogarten *et al.*, 1989a, b; Kluge & Dimroth, 1992; 1993; Dmitriev *et al.*, 1993).

Vacuolar or V-type ATPases are evolutionarily distantly

Table 1. Homologous sequenced c-subunits of F- and V-type ATPases

Abbreviation used*	Biological source	No. of residues	Accession no.
Eco-B	<i>Escherichia coli</i>	79	ATPL_ECOLI
Val-B	<i>Vibrio alginolyticus</i>	84	ATPL_VIBAL
Tba-B	Thermophilic bacterium PS-3	72	ATPL_THEP3
Bme-B	<i>Bacillus megaterium</i>	70	ATPL_BACME
	<i>Bacillus firmus</i>	69†	ATPL_BACFI
	<i>Bacillus alcalophilus</i>	71†	ATPL_BACAO
Sfa-B	<i>Streptococcus faecalis</i>	71	STRATPEFHA_1
Ssp-B	<i>Synechococcus</i> sp. (PCC 6301)	81	ATPL_SYNP6
	<i>Synechocystis</i> sp. (PCC 6803)	81†	PWYBLB
	<i>Anabaena</i> sp. (PCC 7120)	81†	ATPL_ANASP
Pmo-B	<i>Propionigenium modestum</i>	89	ATPL_PROMO
Rru-B	<i>Rhodospirillum rubrum</i>	75	RSPF0ATP_4
Egr-C	<i>Euglena gracilis</i> ; flagellate	77	EGRCPATPH_1
Sol-C	<i>Spinacea oleracea</i> ; spinach	81	ATPH_SPIOL
	<i>Marchantia polymorpha</i> ; liverwort	81†	ATPH_MARPO
	<i>Glycine max</i> ; soybean	81†	ATPH_SOYBN
	<i>Nicotiana tabacum</i> ; tobacco	81†	ATPH_TOBAC
	<i>Pisum sativum</i> ; garden pea	81†	ATPH_PEA
Ani-M	<i>Aspergillus nidulans</i> ; fungus	74	S28794
Ani-N	<i>Aspergillus nidulans</i> ; fungus	143	P16000
Ncr-M	<i>Neurospora crassa</i> ; fungus	74	MINCO1_1
Ncr-N	<i>Neurospora crassa</i> ; fungus	147	ATP9_NEUCR
	<i>Podospora anserina</i> ; fungus	144†	S17915
Sce-M	<i>Saccharomyces cerevisiae</i> ; yeast	76	YSCMTOL1
Spo-M	<i>Schizosaccharomyces pombe</i> ; fission yeast	74	ATP9_SCHPO
Pte-M	<i>Paramecium tetraurelia</i> ; ciliate	75	ATP9_PARTE
Nta-M	<i>Nicotiana tabacum</i> ; tobacco	77	MINTATP9_1
	<i>Nicotiana tabacum</i> ; tobacco	74†	S19862
	<i>Petunia hybrida</i> ; petunia	77†	MIPHAT9_1
Osp-M	<i>Oenothera</i> sp.; evening primrose	78	MIOATP9_1
	<i>Helianthus annuus</i> ; sunflower	83†	NNMTHAATP1
Osa-M	<i>Oryza sativa</i> ; rice	87	ATP9_ORYSA
	<i>Oryza sativa</i> ; rice	87†	MIOSATP_9
Bvu-M	<i>Beta vulgaris</i> ; sugarbeet	88	ATP9_BETVU
	<i>Vicia faba</i> ; broadbean	74†	MIVFATP9_2
	<i>Pisum sativum</i> ; garden pea	74†	ATP9_PEA
Tac-M	<i>Triticum aestivum</i> ; wheat	80	ATP9_WHEAT
	<i>Zea mays</i> ; maize	74†	MZEMTATP91
	<i>Marchantia polymorpha</i> ; liverwort	74†	MPOMTCG_19
Hsa-N	<i>Homo sapiens</i> ; human	100	HUMATPD_1
	<i>Bos taurus</i> ; cow	136†	ATPM_BOVIN
	<i>Bos taurus</i> ; cow	143†	ATPM_BOVIN
	<i>Ovis aries</i> ; sheep	75†	ATPL_SHEEP
Sac-A	<i>Sulfolobus acidocaldarius</i>	101	A33351
Sce-V1	<i>Saccharomyces cerevisiae</i> ; yeast	1–80	VATL_YEAST
Sce-V2	<i>Saccharomyces cerevisiae</i> ; yeast	81–160	VATL_YEAST
Sce-?	<i>Saccharomyces cerevisiae</i> TP3; yeast	102	A35665
Alu-V1	<i>Ascaris lumbricoides</i> ; nematode	1–80	NEMG12_1
Alu-V2	<i>Ascaris lumbricoides</i> ; nematode	81–160	NEMG12_1
Bta-V1	<i>Bos taurus</i> ; cow	1–82	VATL_BOVIN
Bta-V2	<i>Bos taurus</i> ; cow	83–155	VATL_BOVIN
	<i>Homo sapiens</i> ; human	155†	UMPCHSUC1
	<i>Torpedo marmorata</i> ; marbled electric ray	154†	S08261

Table 1.(cont.)

Abbreviation used*	Biological source	No. of residues	Accession no.
	<i>Drosophila melanogaster</i> ; fruit fly	159†	S13117
	<i>Mus musculus</i> ; mouse	155†	MUSMVP_1
	<i>Avena sativa</i> ; oat	165†	A40814
	<i>Manduca sexta</i> ; tobacco hornworm	156†	S19985

* The abbreviation used refers to the genus (first upper-case letter) and species (second two lower-case letters) followed by an upper-case letter as follows: B, bacterial protein; C, chloroplast protein; M, mitochondrially encoded mitochondrial protein; N, nuclearly encoded mitochondrial protein; A, archaeobacterial protein; V, vacuolar-type protein; ?, protein of unknown function; a '1' or '2' following the source designation is included either when two such proteins are found within a single species (bacterial or mitochondrial proteins) or when the protein has two homologous repeat sequences, each of which is separately analysed (vacuolar-type). Abbreviations are indicated only for those proteins that were included in this study.

† This sequence was not used in the comparisons reported because it proved to exhibit greater than 90 % identity with the sequence above it in the table.

related to F-type ATPases, and they exhibit similar structural features (Anraku *et al.*, 1989; Nelson, 1989, 1994; Gogarten *et al.*, 1989a, b; Nelson & Taiz, 1989; Stone *et al.*, 1989; Cross & Taiz, 1990; Kibak *et al.*, 1992). They energize the endomembranes of eukaryotes and are found in both archaea and bacteria (Denda *et al.*, 1990; Yokoyama *et al.*, 1990; Saier *et al.*, 1993; Olsen *et al.*, 1994). These enzyme complexes, like F-type ATPases, are multisubunit enzymes (molecular masses of about 500 kDa) with peripheral membrane catalytic sectors and integral membrane cation channel complexes (Nelson & Taiz, 1989; Foster & Fillingame, 1992; Fraga *et al.*, 1994). It is believed that the V-type ATPases diverged from F-type ATPases relatively early, before individual F-type or V-type ATPases diverged from each other (Kibak *et al.*, 1992; Saier *et al.*, 1993).

Topological analyses of integral membrane subunits of the F₀ sector of the F-type ATPases have been reported (Vik & Dao, 1992). The a-, b- and c-subunits probably span the membrane six, one and two times, respectively. Fourier analyses of sequence variation and hydrophobicity led to the suggestion that two of the six transmembrane spanners of the a-subunit, the single spanner in the b-subunit, and the C-terminal spanner in the c-subunit each have a face in direct contact with membrane lipids. The remaining spanners were predicted to be more shielded from the lipid by protein-protein interactions (Vik & Dao, 1992). A recent electron-spectroscopic imaging study of the *Escherichia coli* F₀ sector revealed three projections. Simulations with different models of the *E. coli* F₀ sector and the a₂c₁₀-subunit complex revealed that these projections would only be obtained by tilting and rotating the model in which the a- and b-subunits are located outside the c-subunit oligomer (Pedersen & Amzel, 1993; Birkenhager *et al.*, 1995).

The F₀ sector of the ATPase is known to provide a pathway for proton transport through the membrane, but

this pathway has not been characterized. All three subunits are required for reconstitution of functional F₀ transport activity (Schneider & Altendorf, 1984, 1985; Brusilow, 1993). Various residues in the a- and c-subunits are believed to constitute the pathway for proton transport from one side of the membrane to the other (Schneider & Altendorf, 1987) in a fashion reminiscent of the bacteriorhodopsin H⁺-conducting pathway (Zimányi *et al.*, 1992; Krebs & Khorana, 1993). The polar domain in the dimeric b-subunit appears to interact with the F₁ sector (Dunn, 1992). X-ray diffraction analyses of F-type ATPases are beginning to reveal the detailed structures of these enzyme complexes (see Stokes & Nakamoto, 1994, and Walker, 1994, for reviews).

In this study we have attempted to identify the sequenced, homologous a-, b- and c-subunits of F-type ATPases that were available when this study was initiated. Of these three subunits, only the c-subunits proved to be demonstrably homologous to those of V-type ATPases. Moreover, homologues of bacterial c- and a-subunits were found to be present in both mitochondrial and chloroplast F-type ATPases, but homologues of bacterial b-subunits were identified only in chloroplast ATPases. Mitochondrial b-subunits may have diverged unrecognizably from the bacterial and chloroplastic homologues. With the exception of a single functionally uncharacterized c-subunit homologue in *Saccharomyces cerevisiae* and the V-type ATPase c-subunits, all proteins included within these three protein families are derived from F-type ATPases. However, c-subunits of V-type ATPases are intragenically duplicated while b-subunits of chloroplast and blue-green bacteria are extragenically duplicated. The intragenic duplication event giving rise to the c-subunits of V-type ATPases is predicted to have occurred shortly after the divergence of proeukaryotes from archaea, while the extragenic duplication of b-subunits is predicted to have occurred after cyanobacteria diverged

from other bacteria but before the divergence of cyanobacteria from chloroplasts.

In this report, we present the multiple alignments and phylogenetic trees for these three families of integral membrane ATPase subunits, and the relative rates of evolutionary divergence of members of these three families are evaluated. The results show that the phylogenetic groupings of each of the subunits roughly follows the phylogenies of the organisms from which they were derived with only a few possible exceptions. These observations lead to the conclusion that, in general, the three F_0 -ATPase subunits evolved in parallel, but at different rates, throughout divergence of the species with little or no horizontal transmission of genetic information.

METHODS

The available databases were screened in homology searches initially using selected a-, b- and c-integral membrane subunits of the *E. coli* F_0F_1 -ATPases using the FASTA (Pearson & Lipman, 1988) or BLAST (Altschul *et al.*, 1990) programs. Subsequently, representatives of each divergent group of these subunits were similarly screened in order to identify all homologous proteins in the database. The ALIGN program (Dayhoff *et al.*, 1983) with 500 random shuffles was used to evaluate the significance of each binary alignment. Protein sequence alignments were generated using the TREE (Feng & Doolittle, 1990) and PILEUP (Devereux *et al.*, 1984) programs. The relative evolutionary distances for construction of phylogenetic trees were determined using the TREE program (Feng & Doolittle, 1990), and in the case of the c-subunits, these results were verified using the PAPA program (Doolittle & Feng, 1990). These programs were modified for execution in the UCSD VAX/VMS DNA system (Smith, 1988). Evaluation of these methods has been considered in previous publications from this laboratory (Van Rosmalen & Saier, 1993; Reizer *et al.*, 1994; Saier, 1994).

RESULTS

c-subunits

All F-type and V-type ATPases contain c-subunits (also called proteolipids or lipoproteins) which can be shown to be homologous (see below). The proteins of this family, indicated by abbreviation, are cited in Table 1 together with their biological sources, their numbers of amino acid residues, and the accession numbers used to identify their sequences. The standard abbreviations used in this study indicate the genus designation of the organism of origin (first letter, upper-case) as well as the species designation (second and third letters, lower-case), followed by a letter to indicate whether the subunit is from a bacterium (B), a chloroplast (C), a mitochondrion and mitochondrially encoded (M), a mitochondrion but nuclearly encoded (N), a eukaryotic vacuole (V) or an archaeon (A) (Olsen *et al.*, 1994). These subcategories of c-subunits include 12 bacterial proteins (B), 6 chloroplast proteins (C), 25 mitochondrial proteins (M and N), 1 archaeal protein (A), and several vacuolar proteins (V). The vacuolar proteins analysed were generated by an internal genetic duplication event, and they consequently possess two dissimilar copies

of the sequence analysed. The F-type c-subunits are generally of 71–89 residues in length unless they are mitochondrial sequences that are encoded within the nucleus, in which case their larger sizes reflect the presence of targeting leader sequences that are cleaved following import into the mitochondria (Saier *et al.*, 1989; Pao & Saier, 1994). The archaeal c-type subunit (Sac-A) is 101 residues long and contains, in addition to a single repeat unit, a unique N-terminal hydrophobic sequence which on the basis of hydropathy analyses is long enough to span the membrane once. All homologous V-type ATPase lipoproteins contain two repeat units. These proteins are from 154 to 160 residues long. The Sce-V3 protein has not been functionally characterized and consequently is designated here as Sce-?. Some sequenced V-type ATPase subunits were not included in our study because of their high degree of similarity with an included protein. No abbreviation for these proteins has been provided (see Table 1 and footnotes to Table 1).

Fig. 1 presents the multiple alignment of all of the c-type subunit sequences listed in Table 1 that exhibit less than 90% identity with each other. These sequences were divided into two groups, those from F-type ATPases (top) and those from V-type and archaeal ATPases (bottom). Because of the very considerable sequence divergence of these two groups of proteins, their consensus sequences (consensus) were determined independently. A number (0–9) above a particular alignment (below the line indicating the alignment position) specifies that the residues in this position are conserved in all but 0–9 sequences. Conserved residues which are common to both the F-type and the V-type ATPase subunit consensus sequences (common consensus) are provided at the very bottom of the multiple alignment.

In the first half of the alignment (Fig. 1a), only five residues are included in the common consensus sequence, and four of these are glycyl residues. They are undoubtedly of structural significance. By contrast, in the second half of the alignment (Fig. 1b), all but one of the commonly conserved residues are hydrophobic in nature (L, F, I and A). The single exception is the essential glutamyl residue that is known to be susceptible to derivatization by dicyclohexylcarbodiimide (Hassinen & Vuokila, 1993). In two of the included F-type ATPases (those from *E. coli* and *Vibrio alginolyticus*) this last residue is an aspartyl residue. In the first half of each duplicated eukaryotic V-type ATPase, this essential glutamyl residue is replaced by a glycyl residue (Fig. 1).

Fig. 2 shows the phylogenetic trees for all c-subunit sequences included in this study as determined with the programs TREE and PAPA. The proteins fall into four major clusters (A–D), regardless of the program used to generate the phylogenetic tree. Cluster A (left) includes all mitochondrial c-subunits, regardless of whether they are encoded within the nucleus or the mitochondrion as well as a single bacterial c-subunit (Rru-B). There are several subclusters within cluster A. The plant proteins all cluster tightly together while the fungal proteins cluster loosely together. The nuclearly encoded fungal proteins cluster

Fig. 1. Complete multiple alignment of 23 c-subunits from F-type ATPases (top) and eight c-subunits from archaeal or vacuolar ATPases (bottom). The residue number in each protein is provided at the beginning and end of each line. The alignment position (alignment #) is provided at the top of the alignment. The digits 0, 1, 2, 3–9 above a particular residue in the multiple alignment (variability) indicate that that residue is conserved with 0, 1, 2, 3–9 exceptions, respectively. The consensus sequence (consensus) is shown below the alignment for the proteins from both the F-type ATPases (top) and the V-type ATPases (bottom), and residues common to both consensus sequences are indicated at the very bottom of the alignment (common consensus). Abbreviations are as indicated in Table 1.

relatively distantly from the mitochondrially encoded proteins. A single animal mitochondrial c-subunit sequence (Hsa-N) was included in this study, and as expected, it comprises a distinct subcluster. The most distantly related mitochondrial c-subunit is that from *Paramecium tetraurelia* (Pte-M). It branches from the trunk of the tree at about the same position as the c-subunit of the F-type ATPase from the α -type purple bacterium, *Rhodospirillum rubrum* (Rru-B). In the tree generated with the PAPA program (Fig. 2b), the *Para. tetraurelia* c-subunit clusters very loosely with the Sce-? protein of *S. cerevisiae*.

Cluster B (upper right) includes all bacterial and chloroplast sequences with the sole exception of the protein from *R. rubrum* (Rru-B). In both trees (Fig. 2a, b) the branching order within cluster B is the same. Thus, as expected from the phylogenies of the organisms from which these subunits derive [based on rRNA sequence data (Olsen *et al.*, 1994)], c-subunits from cyanobacteria and chloroplasts comprise one subcluster, other Gram-negative bacterial proteins comprise a second, and the Gram-positive bacterial proteins comprise a third. The c-subunit from the Na⁺-transporting ATPase of the Gram-negative eubacterium, *Propionigenium modestum* (Laubinger & Dimroth, 1988; Kluge & Dimroth, 1992, 1993),

clusters loosely with the cyanobacterial–chloroplast sub-group on both trees (Fig. 2a, b).

The *Sulfolobus* subunit (Sac-A) and the vacuolar subunits comprise cluster C (lower right in Fig. 2a, b). As with cluster B, the branching order within cluster C is the same in both trees. The first halves of the eukaryotic vacuolar subunits all cluster together as do the second halves, and the archaeal sequence [which does not contain a duplication but does possess an unusual N-terminal hydrophobic stretch of about 25 residues (unpublished results)] alone comprises the third subcluster. These facts suggest that the intragenic duplication that occurred during the evolution of eukaryotic V-type ATPase c-subunits (Mandel *et al.*, 1988) occurred after the divergence of archaea from eukaryotes, but before the divergence of the eukaryotic V-type c-subunits from each other (see Discussion).

Cluster D (lower centre) consists of a single protein in Fig. 2(a). This protein is the functionally uncharacterized protein from *S. cerevisiae* (Sce-?). It may have evolved for a function distinct from that of proton translocation through an F-type ATPase. Possibly it functions as a subunit in a V-type ATPase since it exhibits a high binary

Fig. 2. Phylogenetic trees for homologous c-subunits of F- and V-type ATPases obtained using the TREE program (a) or the PAPA program (b) (see Methods). Abbreviations are as indicated in Table 1. The four clusters are: A, all mitochondrial proteins plus the (distant) c-subunit of the *R. rubrum* ATPase (Rru-B); B, all chloroplast and remaining bacterial proteins; C, all c-subunits from vacuolar and archaeal ATPases; D, the functionally unidentified protein from *S. cerevisiae* (Sce-?) (both trees) as well as the mitochondrial protein from *Para. tetraurelia* (Pte-M) [part (b) only]. Relative branch lengths are shown adjacent to the branches. The percentage SD values calculated from the distance matrix were 7.99 (a) and 5.52 (b).

Table 2. Intercluster comparisons of representative sequences for the c-subunits of F- and V-type ATPases

Cluster*	Protein no. 1†	Cluster*	Protein no. 2†	Comparison score (SD)‡
A	Ani-M	B	Sol-C	9.3
A	Spo-M	C	Bta-V2	5.0
A	Sce-M	A	Pte-M	9.3
B	Sol-C	C	Bta-V2	7.8
B	Sol-C	A	Pte-M	4.5
D	Sce-?	C	Bta-V1	11.8

* The cluster designation (A–D) refers to the phylogenetic grouping in which the protein is found (see Fig. 2).

† Abbreviations for the proteins are provided in the footnotes to Table 1.

‡ The comparison score was determined using the ALIGN program with 500 random shuffles (see Methods).

comparison score with Bta-V1 (see below and Table 2). In the tree depicted in Fig. 2(b), Sce-? clusters loosely with Pte-M, the mitochondrial c-subunit from *Para. tetraurelia*. This latter protein is distant from the other mitochondrial proteins in both trees and is not closely related to any other sequence included in our study.

In separate phylogenetic analyses using the TREE and PAPA programs (see Methods), the proteins (or protein segments) within individual clusters were analysed for their phylogenetic relationships (data not shown). The results were in general agreement with those depicted in Fig. 2. Thus, for example, in one series of tree constructions, the proteins in cluster C (V-type and archaeal c-subunits) as well as the Sce-? and Pte-M proteins were analysed. In these trees, Alu-V1, Bta-V1 and Sce-V1 clustered together, Alu-V2, Bta-V2 and Sce-V2 clustered together, and Sac-A formed a branch which was closer to the second repeat units than to the first repeat units of the vacuolar ATPase c-subunits as expected. Interestingly, however, Sce-? branched off from the trunk of the tree close to the V1 cluster and very distant from the V2 cluster. Moreover, the Pte-M branch emanated from a point closest to the branch for the archaeal Sac-A protein. These observations suggest that Sce-? may be a vacuolar ATPase subunit and are in agreement with the phylogenetic assignments of archaea and ciliates as common ancestors of the progenitors of early eukaryotes.

When the proteins within each major cluster in Fig. 2(a, b) were evaluated for phylogenetic relatedness using the ALIGN program with 500 random shuffles to establish statistical significance, comparison scores always exceeded 9 SD. This fact established that the proteins within each of the four clusters are homologous. When proteins between clusters were examined, values obtained were frequently less than 9 SD. Representative results are presented in Table 2. When Spo-M (cluster A) was compared with Bta-

V2 (cluster C), a comparison score of only 5.0 sd was obtained, and when Sol-C (cluster B) was compared with Pte-M (cluster A or D in Fig. 2a or b, respectively) a value of 4.5 sd was obtained. However, Spo-M (cluster A) gave a comparison score with Pte-M (cluster A or D) of 9.3 sd, and Sce-? (cluster D) gave a comparison score with Bta-V1 (cluster C) of 11.8 sd. Thus, clusters A and B as well as clusters C and D are clearly homologous. Comparison of Sol-C (cluster B) with Bta-V2 (cluster 3) gave a value of 7.8 sd (Table 2), highly suggestive of homology. Employing the superfamily principle (Doolittle, 1986; Saier, 1994), it is probable that all of the proteins included in Table 1 are homologous.

a-subunits

Table 3 lists the a-subunits of F-type ATPases which proved to be demonstrably homologous (ALIGN comparison score of > 9 sd with other a-subunits). a-subunits listed in Table 3 are all of bacterial (B), mitochondrial (M) or chloroplastic (C) origin. Forty-nine protein sequences are listed, and 38 of these were included in our study. These proteins were between 199 residues (Cel-M or Asu-M from worm mitochondria) and 395 residues (Nta-M from tobacco mitochondria) in length. Excluding the worm and plant mitochondrial sequences which are unusually short and long, respectively, all sequences fall within the range 214 (Tbr-M, the mitochondrial a-subunit of *Trypanosoma brucei*) to 289 (Pmo-B, the a-subunit of *P. modestum*).

The multiple alignment of the major portion of the 38 dissimilar a-subunit sequences is shown in Fig. 3. The N-terminal portions of these proteins (alignment positions 1–95) exhibit poor sequence similarity; positions 95–185 exhibit intermediate sequence similarity, and positions 185–218 exhibit the highest degree of sequence similarity with no gaps in the aligned sequences. Within this last region are three residues in close proximity to each other that are fully conserved. These residues comprise the R L L -N sequence motif at alignment positions 203–207. These are the only residues that are fully conserved in all a-subunits. This region of high conservation is followed by a very poorly conserved region with many gaps (positions 218–250) followed by a second relatively well-conserved region with no gaps in the aligned sequences (positions 250–270). Examination of the consensus sequence reveals that the regions conserved are, in general, amphipathic with interspersed hydrophobic, acidic, basic and semipolar residues.

A phylogenetic tree of the a-subunits is presented in Fig. 4. Cluster A (left) includes only animal mitochondrial sequences with the protein phylogenies corresponding well with the organismal phylogenies. Cluster B (top centre) consists of plant and fungal mitochondrial proteins as well as the *R. rubrum* sequence (Rru-B). Cluster C (right) includes only chloroplastic and bacterial sequences, with the *P. modestum* sequence proving to be more closely related to the sequences from chloroplasts and blue-green bacteria than to the remaining bacterial proteins as was noted in the Fig. 2 trees of the c-subunits. The single

mitochondrial sequence from *T. brucei* is found as a distant relative of the other a-subunit sequences, and this sequence alone comprises cluster D (bottom centre).

b-subunits

Screening the *E. coli* F-type ATPase b-subunit as well as homologous proteins against the protein databases revealed homologues only of bacterial and chloroplastic origins. Mitochondrial F-type ATPase subunits that were demonstrably homologous did not appear. However, subunits of these mitochondrial enzyme complexes of similar size and properties, and exhibiting comparison scores of less than or equal to 4 sd, as obtained with the ALIGN program, could be detected (Walker *et al.*, 1982; R. C. Wagenknecht & M. H. Saier, Jr, unpublished results). Because homology with the bacterial and chloroplastic b-subunits could not be established, the putative b-subunits of mitochondrial F-type ATPases were not analysed.

A major portion of the multiple alignment of the b-subunits of bacteria and chloroplasts is shown in Fig. 5. The b-subunits proved to be noticeably less well-conserved than the c- or a-subunits. Thus, within the alignment, shown in Fig. 5 (representing the most conserved portions of these proteins), 29 out of 159 positions (19%) appeared in the consensus sequence. This value is to be compared with 43 out of 122 positions (35%) for the c-subunits and 63 out of 274 positions (23%) for the a-subunits (compare Figs 1, 3 and 5).

Examination of residues conserved in the b-subunits revealed that almost all of the residues occurring in the consensus sequence are hydrophobic (L, I, V and A). Only four residues in the consensus sequence (two Rs and two Es) are hydrophilic. The first well-conserved arginyl residue (alignment position 31) is fully conserved in all b-subunits except the duplicated halves of the cyanobacterial and chloroplast proteins. Further, all but three residues at alignment position 43 are acidic (E or D). These two residues presumably play important functional roles. The other two hydrophilic residues in the consensus sequence (R at alignment position 93 and E at position 106) are less well-conserved (Fig. 5).

Fig. 6 shows the phylogenetic tree of the 30 bacterial and chloroplastic b-subunit sequences analysed. At the lower left-hand portion of the tree, one finds exclusively bacterial proteins. Within this first major cluster, four subclusters are found. The first consists of four Gram-negative bacterial proteins while a second consists of six Gram-positive bacterial proteins. In this respect the branching pattern follows the phylogenies of the organisms from which these proteins derive.

The third major branch (Rru-B) is for the Gram-negative photosynthetic bacterial b-subunit of *R. rubrum*, while the fourth major branch is for the cell wall-less *Mycoplasma gallisepticum*, normally classified with Gram-positive bacteria. It should be noted that the *R. rubrum* proteins are phylogenetically distant from other bacterial subunits on the trees for subunits c, a and b on the trees shown in Figs 2, 4 and 6, respectively, consistent with the classification

Table 3. Homologous sequenced a-subunits of F-type ATPases

Abbreviation used*	Biological source	No. of residues	Accession no.
Rno-M	<i>Rattus norvegicus</i> ; rat	226	S04752
	<i>Mus musculus</i> ; mouse	266†	MUSMT_3
Csp-M	<i>Cricetulus</i> sp.; Chinese hamster	226	CRUMT1_2
Hsa-M	<i>Homo sapiens</i> ; human	226	HUMMTCG_2
Bta-M	<i>Bos taurus</i> ; cow	226	ATP6_BOVIN
	<i>Phoca vitulina</i> ; harbour seal	226	ATP6_PHOVI
Bph-M	<i>Balaenoptera physalus</i> ; finback whale	226	ATP6_BALPH
Gga-M	<i>Gallus gallus</i> ; chicken	227	CHKMTATP82
Cca-M	<i>Cyprinus carpio</i> ; carp	227	CYIMTCCCG5
Xla-M	<i>Xenopus laevis</i> ; frog	226	XELMTCG_6
Poc-M	<i>Pisaster ochraceus</i> ; sea star	228	ATP6_PISOC
Pli-M	<i>Paracentrotus lividus</i> ; sea urchin	232	PALMTCG_7
Lmi-M	<i>Locusta migratoria</i> ; locust	225	ATP6_LOCFI
Dme-M	<i>Drosophila melanogaster</i> ; fruit fly	224	PWFF6
Ame-M	<i>Apis mellifera ligustica</i> ; bee	226	AMFATPASE2
Cel-M	<i>Caenorhabditis elegans</i> ; worm	199	ATP6_CAEEL
Asu-M	<i>Ascaris suum</i> ; roundworm	199	S26017
Osa-M	<i>Oryza sativa</i> ; rice	336	RICMTOSAP1
	<i>Zea mays</i> ; maize	291†	MZEMTATP61
Tae-M	<i>Triticum aestivum</i> ; wheat	386	PWWT6
	<i>Brassica napus</i> ; radish/rape	261†	BNAATP6B1
Nta-M	<i>Nicotiana tabacum</i> ; tobacco	395	TOBMTATP61
	<i>Oenothera berteriana</i> ; primrose	361†	OBEMTATP61
Bvu-M	<i>Beta vulgaris</i> ; beet	250	S21339
Ani-M	<i>Aspergillus nidulans</i> ; fungus	256	EMEMTAN052
Che-M	<i>Cochliobolus heterostrophus</i> ; mushroom	257	ATP6_SCHPO
Pan-M	<i>Podospira anserina</i> ; fungus	264	S02157
	<i>Neurospora crassa</i> ; fungus	252†	NEUMTOLI11
Spo-M	<i>Schizosaccharomyces pombe</i> ; fission yeast	257†	ATP6_SCHPO
Sce-M	<i>Saccharomyces cerevisiae</i> ; yeast	259	PWBY3
Cpa-M	<i>Candida parapsilosis</i> ; imperfect fungus	246	S15378
Tbr-M	<i>Trypanosoma brucei</i> ; protozoan	214	A35349
Sol-C	<i>Spinacia oleracea</i> ; spinach	247	PWSPA6
	<i>Pisum sativum</i> ; pea	247†	A24621
	<i>Zea mays</i> ; maize	247†	PWZMA
	<i>Oryza sativa</i> ; rice	247†	LWRZ6
	<i>Triticum aestivum</i> ; wheat	247†	S14127
	<i>Nicotiana tabacum</i> ; tobacco	247†	LWNT6
Mpo-C	<i>Marchantia polymorpha</i> ; liverwort	248	MPOPCPG17
Osi-C	<i>Odontella sinensis</i>	242	S23356
Rru-B	<i>Rhodospirillum rubrum</i>	241	ATP6_RHORU
Pmo-B	<i>Propionigenium modestum</i>	289	S12619
Ssp-B	<i>Synechococcus</i> sp.	261	PWYCA6
Asp-B	<i>Anabaena</i> sp.	251	ATP6_ANASP
Eco-B	<i>Escherichia coli</i>	271	LWEC6
Val-B	<i>Vibrio alginolyticus</i>	270	S06076
Bfi-B	<i>Bacillus firmus</i>	237	S17720
Tba-B	Thermophilic bacterium	238	S02063
Bme-B	<i>Bacillus megaterium</i>	236	B31482

* The abbreviation used refers to the genus (first upper-case letter) and species (second two lower-case letters) followed by an upper-case letter as follows: B, bacterial protein; C, chloroplast protein; M, mitochondrially encoded mitochondrial protein. Abbreviations are indicated only for those proteins that were included in this study.

† This sequence was not used in the comparisons reported because it proved to exhibit greater than 80 % identity with the sequence above it in the table.

Table 4. Homologous sequenced b-subunits of F-type ATPases

Abbreviation used*	Biological source	No. of residues	Accession no.
Asp-B	<i>Anabaena</i> sp.	187	ATPF_ANASP
Asp-B1	<i>Anabaena</i> sp.	163	ATPX_ANASP
Eco-B	<i>Escherichia coli</i>	156	ATPF_ECOLI
Bme-B	<i>Bacillus megaterium</i>	172	ATPF_BACME
Bfi-B	<i>Bacillus firmus</i>	153	ATPF_BACFI
Efa-B	<i>Enterococcus faecalis</i>	174	ATPF_ENTFA
Pmo-B	<i>Propionigenium modestum</i>	163	S12620
Val-B	<i>Vibrio alginolyticus</i>	156	ATPF_VIBAL
Tba-B	Thermophilic bacterium	163	ATPF_THEP3
Spn-B	<i>Streptococcus pneumoniae</i>	164	SPATPAS1_4
Sor-B	<i>Streptococcus oralis</i>	127	SPOTPAS4_2
Rru-B	<i>Rhodospirillum rubrum</i>	161	ATPX_RHORU
Ssp-B	<i>Synechococcus</i> sp.	171	ATPF_SYNP6
Ssp-B1	<i>Synechococcus</i> sp.	158	ATPX_SYNP6
Scp-B	<i>Synechocystis</i> sp.	179	ATPF_SYNP3
Scp-B1	<i>Synechocystis</i> sp.	143	ATPX_SYNP3
Mga-B	<i>Mycoplasma gallisepticum</i>	198	S24335
Tfe-B	<i>Thiobacillus ferrooxidans</i>	159	TFEUNCAH_2
Mpo-C	<i>Marchantia polymorpha</i> ; liverwort	184	ATPF_MARPO
Osi-C	<i>Odontella sinensis</i>	177	ATPF_ODOSI
Osi-C1	<i>Odontella sinensis</i>	156	ATPX_ODOSI
Asp-C	<i>Antithamnion</i> sp.	192	ATPF_ANTSP
Asp-C1	<i>Antithamnion</i> sp.	159	PLASATPG_4
Egr-C	<i>Euglena gracilis</i> ; flagellate	183	ATPF_EUGGR
Sol-C	<i>Spinacia oleracea</i> ; spinach	184	ATPF_SPIOL
Osa-C	<i>Oryza sativa</i> ; rice	180	ATPF_ORYSA
Tae-C	<i>Triticum aestivum</i> ; wheat	183	ATPF_WHEAT
Nto-C	<i>Nicotiana tabacum</i> ; tobacco	184	ATPF_TOBAC
Psa-C	<i>Pisum sativum</i> ; pea	172	ATPF_PEA
Sol-C1	<i>Spinacia oleracea</i> ; spinach	222	STATPGMR_1

* The abbreviation used refers to the genus (first upper-case letter) and species (second two lower-case letters) followed by an upper-case letter as follows: B, bacterial protein; C, chloroplast protein; a '1' following the source designation is included when two b-subunits have been sequenced from a single organism.

of *R. rubrum* as the sole representative of the α -group of purple bacteria.

All remaining proteins are either from cyanobacteria or from eukaryotic chloroplasts, and these proteins are intermixed, reflective of the relatively recent endosymbiotic origin of chloroplasts from cyanobacteria. Many, and possibly all of these organisms and organelles, possess two distinct genes rather than a single gene encoding the two b-subunits of the F₀ channel. These b-subunits thus probably comprise a heterodimer rather than a homodimer within the F₀ protein complex. Each of the representative members of the two dissimilar b-subunits (B and C versus B1 and C1; see Fig. 6) cluster together. Thus, the six sequenced B1 or C1 b-subunits are found together on one distant branch, while the twelve sequenced B and C b-subunits cluster together on a distinct branch. It is noteworthy that the branching order

for the six pairs of b-subunits for which complete sequence data are available for both members of the b-subunit heterodimer are nearly the same, and that the relative branch lengths never differ by more than twofold. The B and C b-subunits generally exhibit longer branch lengths than those for the B1 and C1 b-subunits. This fact suggests that the latter subunits have undergone slower evolutionary divergence than the former b-subunits following the duplication event that gave rise to these two subunits.

Comparison of phylogenetic trees for subunits c, a and b

Fig. 7 shows phylogenetic trees for the three subunits of the F₀ portions of F-type ATPases for which complete sequence data are available for two (subunits c and a) or

Fig. 4. Phylogenetic tree for 38 homologous a-subunits of F-type ATPases obtained using the TREE program. Abbreviations are as indicated in Table 3. Four clusters are depicted (see text). Relative branch lengths are shown adjacent to the branches. The percentage SD value calculated from the distance matrix was 5.77.

even three (subunits a, c and b) of these subunits. Fig. 7 (a and b) shows the results obtained for the c-subunits with the programs TREE and PAPA, respectively. The branching order is practically the same with either program. Further, when comparing the tree shown in Fig. 7(a) (c-subunits) with that depicted in Fig. 7(c) (a-subunits), it can be seen that the two trees are almost identical with respect to branching order although the c-subunit tree is more compact than the a-subunit tree. The former observation strongly suggests that the c- and a-subunits evolved in parallel from common ancestral proteins. Since these phylogenetic trees generally follow the phylogenies of the organisms/organelles of origin, it appears that they evolved primarily by divergence during and following vertical transmission of genetic information. The fact that branch lengths are generally longer in Fig. 7(c) than in Fig. 7(a) suggests that the rate of evolutionary divergence for the a-subunits was greater than that for the c-subunits. Interestingly, the branch positions for the *R. rubrum* subunits and the *P. modestum* subunits cluster with the mitochondrial subunits and with the chloroplast and cyanobacterial subunits, respectively, on all three trees (Fig. 7a–c). The positional differences of the *R. rubrum* subunits probably reflect experimental error (i.e. compare Fig. 7a–c).

Fig. 7(d) presents comparable data for the bacterial and chloroplastic b-subunit sequences (note that the non-homologous mitochondrial sequences were excluded because of a lack of demonstrable homology). The tree in Fig. 7(d) resembles those in Fig. 7(a, c) with the sole exception of the *P. modestum* b-subunit which clusters more closely with the Gram-negative bacterial b-subunits. This interesting positional difference may reflect a different origin of the b-subunit gene as compared with the c- and a-subunit genes of *P. modestum*.

Comparing branch lengths for proteins represented in Fig. 7(d) with the corresponding proteins represented in Fig. 7(a, c), it can tentatively be concluded that of these three subunits, the b-subunits diverged from their common evolutionary ancestor most rapidly, the c-subunits diverged most slowly, and the a-subunits diverged at an intermediate rate. The consensus sequence evaluation discussed above provides confirmatory evidence for this conclusion.

DISCUSSION

The present study serves to define the phylogenetic relationships between the three integral membrane subunits which comprise all F₀ sectors of the F-type ATPases of bacteria, chloroplasts and mitochondria. One of these subunits, the c-subunit, also exhibits homology to the corresponding subunit in an archaeal ATPase and the eukaryotic vacuolar ATPases. In the latter case, these subunits have evolved to about twice their original size due to an intragenic tandem duplication encompassing most of the gene encoding the primordial c-subunit (Mandel *et al.*, 1988).

Our studies suggest that the intragenic duplication event in the primordial c-subunit gene that gave rise to large eukaryotic V-type ATPase c-subunits (Fig. 2) and extragenic duplication of the evolving b-subunit gene that gave rise to two b-subunits in the F-type ATPases of cyanobacteria and chloroplasts (Fig. 6) occurred as early events during evolutionary history. The former event occurred about 2 billion years ago, just after archaea diverged from eukaryotes, but before the represented eukaryotes diverged from each other. The latter event may have occurred at about the same time after cyanobacteria diverged from other bacteria, but before they invaded eukaryotes to give rise to chloroplasts. Our analyses indicate that like the intragenic duplication event occurring in the evolving V-type ATPase c-subunit gene, the latter event probably occurred only once. Although chloroplasts may have evolved from cyanobacteria more than once, all of those that acquired the duplicated b-type subunit genes apparently acquired them from cyanobacteria in which this duplication event had already occurred.

Based on the phylogenetic analyses reported here and elsewhere (Gogarten *et al.*, 1992; Kibak *et al.*, 1992; Yokoyama *et al.*, 1990, 1994; Solioz & Davies, 1994; Takase *et al.*, 1994), we propose three pathways for the evolution of all related F- and V-type ATPases. We

Alignment #:	1	10	20	30	40	50	Alignment #:	51	60	70	80	90	100
Residue							Residue						
Variability:	5		8	8	8	99	Variability:	7	8				
Asp-C (22)	LFDPN	ATLPLMAIQF	LA	ITIIINLIYYPLGKIL	DERDEYIA	(63)	Asp-C (64)	NSLTASAALS	KANDLTKRYEQDLAE	SRUKAQDI	IKNAQQDAQNI	VSSK	(112)
Asp-C1 (20)	LFDLD	ATLPLMAIQF	LL	LALIINATLYPKLRAI	DGRNEIVR	(61)	Asp-C1 (62)	NNLEAQERLS	KAETLAETQEELAG	ARRQAQTI	IADAAQGRKIAEK		(110)
Ssp-B1 (20)	LFDLD	ATLPLMAIQF	LV	LVLLINAVYYPFGKVL	DDRDQFVR	(61)	Ssp-B1 (62)	GGGQDAKARLA	EVKALTAQYEQELAA	TRKQSQA	ILAEQAETAGRIAAQ		(110)
Osi-C1 (20)	LFDIN	ATLPLVAIQF	LL	LVLLINAVYISPLTII	ERREYIL	(61)	Osi-C1 (62)	NKLAKASEIL	SQANLETAQYEQELAT	VRKQAQLE	ITNSQRIKREILIE		(110)
Sol-C1 (84)	LFDPN	LTPLIIMAEF	LF	LVFALDKIYYPLGDMF	DKRDASIK	(125)	Sol-C1 (126)	EQLSGVKDT	SEVQLEEQANAVHRA	ARAEISAA	INVKKGTQLEVEAK		(174)
Ssp-B1 (01)	MFDDP	ATLPLMAIQF	VV	LVFALLIAIFYPKHNVL	DERADYIR	(42)	Ssp-B1 (43)	TNEZDAREKLA	KAKAITQYEQQITD	ARRQSQAV	IADAAQGRKIAEK		(91)
Egr-C (26)	VFETNI	INLAIVVGLT	FY	FGKVLTSLLKTRKTI	IKNILDIDEKIRSSQ	(75)	Egr-C (76)	SSLYLAEL	FEENAKAASLRNSGTF	CLKSPDI	IRSSVNEIDIRLKQSK		(125)
Osa-C (26)	ILATNL	INLVTVVGVV	IF	FGKGVLDLLDNRRQRI	LSTIRNSELARGTI	(75)	Osa-C (76)	EQLKARIRLQ	KVLEADEYRHHGYSE	IEREKANLIN	INATISISLEQLEKSK		(125)
Tae-C (26)	ILATNL	INLVTVVGVV	IF	FGKGVLDLLDNRRQRI	LSTIRNSELARGTI	(75)	Tae-C (76)	EQLKARIRLQ	KVLEADEYRHHGYSE	IEREKANLIN	INATISISLEQLEKSK		(125)
Sol-C (26)	ILATNL	INLVSVLVGVV	IF	FGKGVLDLLDNRRQRI	LSTIRNSELARGAI	(75)	Sol-C (76)	EQLKARIRLQ	KVLEADEYRHHGYSE	IEREKANLIN	INATISISLEQLEKSK		(125)
Nto-C (26)	ILATNL	INLVSVLVGVV	IF	FGKGVLDLLDNRRQRI	LSTIRNSELARGAI	(75)	Nto-C (76)	EQLKARIRLQ	KVLEADEYRHHGYSE	IEREKANLIN	INATISISLEQLEKSK		(125)
Psa-C (16)	ILATNL	INLVSVLVGVV	IF	FGKGVLDLLDNRRQRI	LSTIRNSELARGAI	(65)	Psa-C (66)	EQLKARIRLQ	KVLEADEYRHHGYSE	IEREKANLIN	INATISISLEQLEKSK		(125)
Mpo-C (26)	LELTNL	INLVGVVGLV	VY	FGKGVLSNLLNRRKIL	TLITIDQAEZYKQAT	(75)	Mpo-C (76)	DKINQARTLR	QAQKADITRINGSQ	KEKQDL	INAADEDSKRLKLESD		(115)
Ssp-B (22)	LFETNL	INLAIIIGLL	VI	YAGRGFLGNLLSRRRA	KEATREVEKLASSA	(71)	Ssp-B (72)	QILAEQKQ	LEQAQAEARIVQEAQ	QRAEVAQ	ETATQTEADLRREAA		(126)
Ssp-B (27)	FLANLN	INLAIIIGLL	VI	YAGRGFLGNLLSRRRA	KEATREVEKLASSA	(76)	Ssp-B (77)	QILAEQKQ	LEQAQAEARIVQEAQ	QRAEVAQ	ETATQTEADLRREAA		(126)
Asp-B (34)	LIDTN	INLAIIITVL	FV	FGKGVLTGKTLKTRRE	NIETATQNAEQAADAA	(83)	Asp-B (84)	QILAEQKQ	LEQAQAEARIVQEAQ	QRAEVAQ	ETATQTEADLRREAA		(126)
Asp-C (29)	FLANLN	INLAIIIGLL	VI	YAGRGFLGNLLSRRRA	KEATREVEKLASSA	(78)	Asp-C (79)	NRLLESEKQ	LEQQLVITQVLNDAE	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Osi-C (24)	ILETG	INLAIVGLV	IF	YAGRGFLGSFLGKQKTS	IVQGGQDAEGRLEEN	(73)	Osi-C (74)	RRLSEQAQ	LEQQLVITQVLNDAE	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Eco-B (08)	ILG	QAIAVFLVFLFC	HK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(56)	Eco-B (57)	LAKASATDQ	LEQAQAEARIVQEAQ	QRAEVAQ	ETATQTEADLRREAA		(126)
Val-B (08)	LLG	QAISFALFVFC	HK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(56)	Val-B (57)	LAKASATDQ	LEQAQAEARIVQEAQ	QRAEVAQ	ETATQTEADLRREAA		(126)
Tfe-B (10)	TLIVQL	LVFVILLVALL	IK	YHMGPRKRVMDORAKI	ADGLAAAEKREGEAA	(59)	Tfe-B (60)	LAKASATDQ	LEQAQAEARIVQEAQ	QRAEVAQ	ETATQTEADLRREAA		(126)
Pmo-B (15)	HF	WQINFLFILLV	IK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(63)	Pmo-B (64)	KANQEAGT	LVREAKRKAET	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Spn-B (12)	FILIT	GSFILLVIV	IK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(64)	Spn-B (65)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Sor-B (12)	FILIT	GSFILLVIV	IK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(60)	Sor-B (61)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Efa-B (19)	IIVVS	GSFILLVIV	IK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(67)	Efa-B (68)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Bme-B (19)	DLFQ	LVFVILLVALL	IK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(68)	Bme-B (69)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Tba-B (22)	TIYQL	LVFVILLVALL	IK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(71)	Tba-B (72)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Bfi-B (10)	SALYQ	LVFVILLVALL	IK	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(58)	Bfi-B (59)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Mga-B (48)	VYVAT	IALSIILILG	IF	YVWFFPILMAAIEK	QKQIADGLASAARAKDD	(97)	Mga-B (98)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Rru-B (28)	PAFVS	IAFLMVGVVYIK	AK	KLALGADGRGA	AVKALDEARILQDDAQ	(77)	Rru-B (78)	VLAQKRE	LELAGRSNEAKTI	ITPAQKVSQ	ILDKKIDVEKLGAS		(133)
Consensus:	-----I-L-----L-----L-----L-R-----I-----I-AE-----												
Alignment #:	101	110	120	130	140	150	Alignment #:	160	170	180	190	200	
Residue							Residue						
Variability:	7		8			7	Variability:	7		8			
Asp-C (113)	IKAEQ	KADQADQALMSNTYDQL	NIQEQ	QALQNLKQVDI	LSNQIQIKLGN	(160)	Asp-C (164)	IKAEQ	KADQADQALMSNTYDQL	NIQEQ	QALQNLKQVDI	LSNQIQIKLGN	(160)
Asp-C1 (111)	VAAEQ	KAQAEQAEQAGEI	EQQKQ	QALASLEQVDA	LSRQILEKLGADLVKQR	(164)	Asp-C1 (164)	VAAEQ	KAQAEQAEQAGEI	EQQKQ	QALASLEQVDA	LSRQILEKLGADLVKQR	(164)
Ssp-B1 (111)	LAEQ	KAQAEQAEQAGEI	DQKQ	VALQALDQVDA	LSBQILEKLGARA	(159)	Ssp-B1 (111)	LAEQ	KAQAEQAEQAGEI	DQKQ	VALQALDQVDA	LSBQILEKLGARA	(159)
Osi-C1 (111)	LNISQ	KIIDLNIENITDDL	LEKQ	TALNSLDTVQS	LCVQIENRLSI	(157)	Osi-C1 (111)	LNISQ	KIIDLNIENITDDL	LEKQ	TALNSLDTVQS	LCVQIENRLSI	(157)
Sol-C1 (175)	LAEGR	KKIEVELQALGSL	EQQKQ	TSLSLSQISA	LSDDIVKVLFS	(223)	Sol-C1 (175)	LAEGR	KKIEVELQALGSL	EQQKQ	TSLSLSQISA	LSDDIVKVLFS	(223)
Ssp-B1 (92)	LAEQ	RESQKQETAAQEI	EAQ	QSALSSLEQVAA	LSNQILEKLGPELIK	(143)	Ssp-B1 (92)	LAEQ	RESQKQETAAQEI	EAQ	QSALSSLEQVAA	LSNQILEKLGPELIK	(143)
Egr-C (126)	RLILR	TEKRSVRIEKLIS	QACQKAKATI	IKRLNSKIKK	ILKQKMKSLKQKPKY	(185)	Egr-C (126)	RLILR	TEKRSVRIEKLIS	QACQKAKATI	IKRLNSKIKK	ILKQKMKSLKQKPKY	(185)
Osa-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)	Osa-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)
Tae-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)	Tae-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)
Sol-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)	Sol-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)
Nto-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)	Nto-C (126)	NETLY	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIRANISILGAE	(180)
Psa-C (116)	NKTIR	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIGATITGFGMSKAKT	(173)	Psa-C (116)	NKTIR	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	TIGATITGFGMSKAKT	(173)
Mpo-C (126)	NATIR	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	MIDYHIGLLRAMESTIE	(184)	Mpo-C (126)	NATIR	FEKQRAHNVQRV	QVQALG	QALGTLNSCINTELEFR	MIDYHIGLLRAMESTIE	(184)
Ssp-B (122)	AQDVST	EQQRVLEDELRI	AVAGAL	SVETQLSQLEDAQQR	LIDRSIALT	(172)	Ssp-B (122)	AQDVST	EQQRVLEDELRI	AVAGAL	SVETQLSQLEDAQQR	LIDRSIALT	(172)
Ssp-B (127)	AQDLG	AEQVRIEALRRRI	AEQVAKAEADLR	DRINEIDTQR	LIDRSIALT	(180)	Ssp-B (127)	AQDLG	AEQVRIEALRRRI	AEQVAKAEADLR	DRINEIDTQR	LIDRSIALT	(180)
Asp-B (134)	AADLN	AEADRLAQLRV	VALALQ	KVESELDQGISDAQKT	LIDRSIALT	(188)	Asp-B (134)	AADLN	AEADRLAQLRV	VALALQ	KVESELDQGISDAQKT	LIDRSIALT	(188)
Asp-C (129)	KASIV	VAENITQKIKIT	ALALQKVSQ	LQAVDTTQAK	IIDSSIIDLRGDI	(183)	Asp-C (129)	KASIV	VAENITQKIKIT	ALALQKVSQ	LQAVDTTQAK	IIDSSIIDLRGDI	(183)
Osi-C (124)	LATFR	SKRQIFLVEKQ	QIILLVKT	VARAQQTGPKER	TALITETIKLEGDLL	(179)	Osi-C (124)	LATFR	SKRQIFLVEKQ	QIILLVKT	VARAQQTGPKER	TALITETIKLEGDLL	(179)
Eco-B (107)	QAEIE	AEKRRREELKQ	VAILAVAGAEK	ILERSIDKQAKD	ILDNITAKL	(157)	Eco-B (107)	QAEIE	AEKRRREELKQ	VAILAVAGAEK	ILERSIDKQAKD	ILDNITAKL	(157)
Val-B (107)	EAELE	AEKRRREELKQ	VAILAVAGAEK	ILERSIDKQAKD	ILDNITAKL	(157)	Val-B (107)	EAELE	AEKRRREELKQ	VAILAVAGAEK	ILERSIDKQAKD	ILDNITAKL	(157)
Tfe-B (110)	RAEID	VEINRAEVLGQ	VVLLVNGT	QILERSIDKQAKD	ILDNITAKL	(160)	Tfe-B (110)	RAEID	VEINRAEVLGQ	VVLLVNGT	QILERSIDKQAKD	ILDNITAKL	(160)
Pmo-B (114)	EVEIE	KHKEQAKRLEQ	LVETDLAV	LAERHINEKVDKIGAN	ILDQYIEVGEEL	(168)	Pmo-B (114)	EVEIE	KHKEQAKRLEQ	LVETDLAV	LAERHINEKVDKIGAN	ILDQYIEVGEEL	(168)
Spn-B (115)	NQEL	IAQKALQSVGEVAD	ITVSLAKK	ISKNLDSARKE	LIDQYIDQLGEA	(177)	Spn-B (115)	NQEL	IAQKALQSVGEVAD	ITVSLAKK	ISKNLDSARKE	LIDQYIDQLGEA	(177)
Sor-B (111)	NQEL	IAQKALQSVGEVAD	ITVSLAKK	ISKNLDSARKE	LIDQYIDQLGEA	(177)	Sor-B (111)	NQEL	IAQKALQSVGEVAD	ITVSLAKK	ISKNLDSARKE	LIDQYIDQLGEA	(177)
Efa-B (118)	QTDIM	LERDIALNSVDD	VADLSQ	IAKILNKLESPHES	LINQYIEGLSSNETR	(175)	Efa-B (118)	QTDIM	LERDIALNSVDD	VADLSQ	IAKILNKLESPHES	LINQYIEGLSSNETR	(175)
Bme-B (119)	KQET	EQQDQVAAALQ	EVASVLSL	ASKVIEKELSEQDQK	LIBEYIEVGEVDR	(173)	Bme-B (119)	KQET	EQQDQVAAALQ	EVASVLSL	ASKVIEKELSEQDQK	LIBEYIEVGEVDR	(173)
Tba-B (122)	KQET	EQQDQVAAALQ	EVASVLSL	ASKVIEKELSEQDQK	LIBEYIEVGEVDR	(173)	Tba-B (122)	KQET	EQQDQVAAALQ	EVASVLSL	ASKVIEKELSEQDQK	LIBEYIEVGEVDR	(173)
Bfi-B (102)	EIRRE	QVASALQEVAGLSV	LATVIEKELSEADQK	LVEYIEVGEVDR		(155)	Bfi-B (102)	EIRRE	QVASALQEVAGLSV	LATVIEKELSEADQK	LVEYIEVGEVDR		(155)
Mga-B (148)	HAET	IKQESILRLEGR	IVDVALEAT	STLIFQVAKEDER	LIVNELRLD	(199)	Mga-B (148)	HAET	IKQESILRLEGR	IVDVALEAT	STLIFQVAKEDER	LIVNELRLD	(199)
Rru-B (128)	VDRIA	QAEQALQVNR	AVDVAV	SAARSILMAGSLAKADQNR	LIDALADLPKLE	(183)	Rru-B (128)	VDRIA	QAEQALQVNR	AVDVAV	SAARSILMAGSLAKADQNR	LIDALADLPKLE	(183)
Consensus:	-----I-E-----V-----A-----L-----LI-----I-----L-----												

Fig. 5. Multiple alignment of the well-aligned portions of the sequences of F-type ATPase b-subunits from bacteria and chloroplasts. Abbreviations of the proteins are as indicated in Table 4. Thirty sequences are represented. The format of presentation is essentially the same as for Fig. 1, except that only bacterial and chloroplastic sequences are presented due to the lack of demonstrable homology with putative b-subunits of mitochondrial F-type ATPases.

Fig. 5. Multiple alignment of the well-aligned portions of the sequences of F-type ATPase b-subunits from bacteria and chloroplasts. Abbreviations of the proteins are as indicated in Table 4. Thirty sequences are represented. The format of presentation is essentially the same as for Fig. 1, except that only bacterial and chloroplastic sequences are presented due to the lack of demonstrable homology with putative b-subunits of mitochondrial F-type ATPases.

assume that the common precursor of these enzymes was present in the common bacterial ancestor of the three relevant kingdoms of life, the archaeal-eukaryotic kingdom, the cyanobacterial-chloroplastic kingdom, and the bacterial-mitochondrial kingdom.

Development of V-type ATPases possibly occurred in the archaeal-eukaryotic kingdom. Following extensive divergence of V-type ATPases from F-type ATPases, archaea diverged from proeukaryotes. Only in the proeukaryotic lineage, prior to the diversification of eukaryotes, did the tandem intragenic c-subunit duplication occur. These 'differentiated' enzymes may subsequently have been acquired by bacteria as a result of horizontal transmission of the entire genetic apparatus, and some of these V-type ATPases apparently gained differing cation-specificities, i.e. for Na^+ instead of H^+ (Yokoyama *et al.*, 1990, 1994; Solioz & Davies, 1994; Takase *et al.*, 1994).

Development of chloroplastic F-type ATPases occurred

in the cyanobacterial kingdom. Although the primordial cyanobacterial enzymes did not diverge markedly from those found in bacteria, the extragenic duplication event giving rise to two distinct b-subunit polypeptides apparently occurred only within this lineage. These dissimilar b-subunit polypeptides may have evolved distinct but overlapping functions, thereby generating ATPases with heterodimeric b-subunit enzyme complexes that possess functions in addition to those associated with ATPases possessing the homodimeric b-subunits. The fully conserved arginyl residue at alignment position 31 in Fig. 5 in all b-subunits except the duplicated ones argues in favour of their evolution toward a dissimilar function. When the cyanobacteria endosymbiotically invaded primitive eukaryotes, and the bacteria then degenerated and differentiated into chloroplasts, the need for two dissimilar b-subunits was retained. If future studies establish that some eukaryotic chloroplasts possess F-type ATPases with only a single b-subunit, then either one of

Fig. 6. Phylogenetic tree for 30 homologous b-subunits of F-type ATPases from bacteria and chloroplasts generated with the TREE program. Abbreviations are as indicated in Table 4. Clusters are as follows: A, primordial b-subunits of cyanobacteria and chloroplasts; B, duplicated b-subunits (indicated B1 or C1 for cyanobacterial and chloroplastic proteins, respectively); C, bacterial b-subunits other than those from cyanobacteria. The format of presentation is as indicated in Fig. 2. Relative branch lengths are shown adjacent to the branches. The percentage SD value calculated from the distance matrix was 7.7.

the two b-subunits must have been lost, or the latter chloroplasts evolved independently of the two-b-subunit-containing chloroplasts from progenitor cyanobacteria in which b-subunit gene duplication had not occurred.

Intra- or extragenic duplication frequently gives rise to two products, one of which retains its primordial function, the other of which is free to diverge to acquire a dissimilar function (Wu & Saier, 1990; Saier, 1994). In the cases of the two duplications that gave rise to more complex ATPases, that in the c-subunit gene of evolving V-type ATPases of eukaryotes, and that in the b-subunit gene that gave rise to the primordial cyanobacterial F-type ATPase, we postulate that one of these duplications evolved to primarily provide a structural role while the other evolved to primarily provide a functional role. Thus, the V1 subunit lacks the essential catalytic glutamyl (or aspartyl) residue at alignment position 88 in Fig. 1 and therefore may be unable to function in proton translocation (Fraga *et al.*, 1994; Zhang & Fillingame, 1994). These V1 c-subunit halves have diverged more rapidly from each other and from the common V1/V2 ancestral protein than have the presumably functional V2 c-subunit halves (data not shown). Interestingly, this situation differs for the duplicated b-subunits. In this latter case, the B1 and C1 b-subunits which have lost the otherwise fully conserved arginyl residue (alignment position 31 in Fig. 5) and therefore have presumably lost an essential function, have nevertheless diverged from each other and

from the common primordial sequence more *slowly* than have the B and C b-subunits (Fig. 6). These considerations lead to the postulate that while the c-subunits play a primary functional role in proton translocation, the b-subunits primarily serve a structural role in construction of the ATPase. Similar conclusions have been reached by others using quite different reasoning (Dunn, 1992; Pedersen & Amzel, 1993).

The development of mitochondrial F-type ATPase complexity probably occurred primarily in the bacterial kingdom. No recognized duplication events occurred in the genetic apparatus encoding the F_0 sector of these ATPases. However, following entry of mitochondrial precursor bacteria into endosymbiotic relationships with early eukaryotic cells, the b-subunits diverged unrecognizably from those in bacteria, possibly in order to provide new functions. At the same time, the ATPase acquired additional subunits to allow nuclear localization of some of the ATPase genetic apparatus as well as sensitivity to complex regulatory signals generated within these multiorganellar, multicellular hosts.

In the above discussion we have suggested three pathways to account for the evolution of all F- and V-type ATPases found in nature. The evidence generated by phylogenetic tree construction suggests that the three integral membrane subunits a, b and c of the F_0 portion of F-type ATPases usually evolved as a unit. In almost all cases, the phylogenies of these three protein subunits follow the phylogenies of the organisms from which they are derived. The H^+ -translocating F_0F_1 -ATPase from *R. rubrum*, a Gram-negative purple bacterium of the α -subdivision (Woese, 1987; Olsen *et al.*, 1994), possesses c- and a-subunits which consistently cluster with those of mitochondrial F-type ATPases rather than with those of the other bacteria represented in our study (Figs 2 and 4). This fact correlates with the belief that mitochondria arose from an ancestral bacterium of the α -subdivision that also gave rise to *R. rubrum*. Although results of analyses designed to ascertain the origins of mitochondria are ambiguous, the results of ribosomal RNA analyses suggest that mitochondria were derived from members of the α -subdivision of purple bacteria (Gray, 1988; Gutell *et al.*, 1994; Olsen *et al.*, 1994).

An apparent exception to the rule that the F_0 subunits evolved as a unit may be found in the *P. modestum* ATPase. The c- and a-subunits of this enzyme complex cluster more closely with those of cyanobacteria and chloroplasts than with those of other bacteria (Figs 2 and 4), but the b-subunit clusters with the bacterial F-type ATPase b-subunits, being situated closer to the Gram-negative purple bacterial cluster of these proteins than to the Gram-positive bacterial cluster of these proteins. *P. modestum* has been reported to represent a new, separate line of descent within the bacterial phyla, moderately related to the cyanobacteria (Both *et al.*, 1991). Since the b-subunit of *P. modestum* does not cluster as do the a- and c-subunits on their respective trees, the b-subunit may have been acquired by horizontal transmission from another bacterium. Alternatively, adaptation to Na^+ translocation

Fig. 7. Phylogenetic trees for (a) c-subunits (TREE program), (b) c-subunits (PAPA program), (c) a-subunits (TREE program) and (d) b-subunits (TREE program) for those organisms/organelles for which complete sequence data are available for at least two (or all three) subunits. Abbreviations for the proteins are as indicated in Tables 1, 3 and 4, respectively. Relative branch lengths are shown adjacent to the branches. The percentage SD values calculated from the distance matrix were 7.24 (a), 4.27 (b), 4.84 (c) and 5.40 (d).

may have selectively altered the b-subunit without changing the sequences of the a- or c-subunits.

Another apparent anomaly concerns the b-subunit of *M. gallisepticum*. This protein occurs on a branch of the tree shown in Fig. 6 that is separated from the cluster of Gram-positive bacterial b-subunits by the *R. rubrum* protein, and the cyanobacterial/chloroplast proteins are found on the other side of the branch bearing the *M. gallisepticum* protein. Corresponding sequenced a- and c-subunits from this organism were not available at the time this comparative study was conducted. *M. gallisepticum* proteins

would normally be expected to cluster with those of Gram-positive bacteria (Woese, 1987).

An anomalous and functionally uncharacterized protein has been sequenced from the yeast *S. cerevisiae*. This protein is homologous to c-subunits of F- and V-type ATPases. *S. cerevisiae* possesses both F- and V-type ATPases, with the expected c-type subunits found in association with these ATPases. The functionally uncharacterized c-subunit, here designated Sce-?, is the most distant member of the c-subunit family. It forms a primary branch on the c-subunit phylogenetic tree between the

other major subdivisions (Fig. 2). Because the function of this protein is not known, little can be said regarding its significance. However, based on its phylogeny (as mentioned in the Results section), we believe that it may prove to be a constituent of a yet-to-be-discovered V-type ATPase. Alternatively, the encoding gene may be a 'pseudogene' and does not encode a functional protein (Krawiec & Riley, 1990; Riley, 1993).

In conclusion, our results lead to the suggestion that the evolution of the F₀ and V₀ sectors of F-type and V-type ATPases probably occurred almost exclusively by vertical transmission of genetic information with minimal duplication and horizontal transfer of the encoding genes. The potential exceptions are cited above.

ACKNOWLEDGEMENTS

We are grateful to Mary Beth Hiller for providing expert assistance in the preparation of this manuscript. This work was supported by USPHS grants 5RO1AI 21702 and 2RO1AI 14176 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Anraku, Y., Umemoto, N., Hirata, R. & Wada, Y. (1989). Structure and function of the yeast vacuolar membrane proton ATPase. *J Bioenerg Biomembr* **21**, 589–603.
- Birkenhager, R., Hoppert, M., Deckershebestreit, G., Mayer, F. & Altendorf, K. (1995). The F₀ complex of the *Escherichia coli* ATP synthase – investigation by electron spectroscopic imaging and immunoelectron microscopy. *Eur J Biochem* **230**, 58–67.
- Both, B., Kaim, G., Wolters, J., Schleifer, K. H., Stackebrandt, E. & Ludwig, W. (1991). *Propionigenium modestum*: a separate line of descent within the bacteria. *FEMS Microbiol Lett* **62**, 53–58.
- Brusilow, W. S. A. (1993). Assembly of the *Escherichia coli* F₁F₀ ATPase, a large multimeric membrane-bound enzyme. *Mol Microbiol* **9**, 419–424.
- Cross, R. L. & Taiz, L. (1990). Gene duplication as a means for altering H⁺/ATP ratios during the evolution of F₀F₁ ATPases and synthases. *FEBS Lett* **259**, 227–229.
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983). Establishing homologies in protein sequences. *Methods Enzymol* **91**, 524–545.
- Denda, K., Konishi, J., Hajiro, K., Oshima, T., Date, T. & Yoshida, M. (1990). Structure of an ATPase operon of an acidothermophilic archaeobacterium, *Sulfolobus acidocaldarius*. *J Biol Chem* **265**, 21509–21513.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.
- Dmitriev, O., Deckers-Hebestreit, G. & Altendorf, K. (1993). ATP synthesis energized by DpNa and Dy in proteoliposomes containing the F₀F₁ ATPase from *Propionigenium modestum*. *J Biol Chem* **268**, 14776–14780.
- Doolittle, R. F. (1986). *Of Urfs and Orfs: a Primer on How to Analyze Derived Amino Acid Sequences*. Mill Valley, CA: University Science Books.
- Doolittle, R. F. & Feng, D.-F. (1990). A nearest neighbor procedure for relating progressively aligned amino acid sequences. *Methods Enzymol* **183**, 659–669.
- Dunn, S. D. (1992). The polar domain of the b subunit of *Escherichia coli* F₁F₀-ATPase forms an elongated dimer that interacts with the F₁ sector. *J Biol Chem* **267**, 7630–7636.
- Feng, D.-F. & Doolittle, R. F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol* **183**, 375–387.
- Foster, D. L. & Fillingame, R. H. (1982). Stoichiometry of subunits in the H⁺-ATPase complex of *Escherichia coli*. *J Biol Chem* **257**, 2009–2015.
- Fraga, D., Hermolin, J. & Fillingame, R. H. (1994). Transmembrane helix–helix interactions in F₀ suggested by suppressor mutations to Ala²⁴ → Asp/Asp⁶¹ → Gly mutant of ATP synthase subunit c. *J Biol Chem* **269**, 2562–2567.
- Futai, M., Noumi, T. & Maeda, M. (1989). ATP synthase (H⁺-ATPase): results by combined biochemical and molecular biological approaches. *Annu Rev Biochem* **58**, 111–136.
- Gogarten, J. P., Kibak, H., Dittich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989a). Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci USA* **86**, 6661–6665.
- Gogarten, J. P., Rausch, T., Bernasconi, P., Kibak, H. & Taiz, L. (1989b). Molecular evolution of H⁺-ATPases. I. *Methanococcus* and *Sulfolobus* are monophyletic with respect to eukaryotes and bacteria. *Z Nat Forsch Sect C J Biosci* **44**, 641–650.
- Gogarten, J. P., Starke, T., Kibak, H., Fishmann, J. & Taiz, L. (1992). Evolution and isoforms of V-ATPase subunits. *J Exp Biol* **172**, 137–147.
- Gray, M. W. (1988). Organelle origins and ribosomal RNA. *Biochem Cell Biol* **66**, 325–348.
- Gutell, R. R., Larsen, N. & Woese, C. R. (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev* **58**, 10–26.
- Hassinen, I. E. & Vuokila, P. T. (1993). Reaction of dicyclohexylcarbodiimide with mitochondrial proteins. *Biochim Biophys Acta* **1144**, 107–124.
- Hatefi, Y. (1993). ATP synthesis in mitochondria. *Eur J Biochem* **218**, 759–767.
- Kibak, H., Taiz, L., Starke, T., Bernasconi, P. & Gogarten, J. P. (1992). Evolution of structure and function of V-ATPases. *J Bioenerg Biomembr* **24**, 415–424.
- Kluge, C. & Dimroth, P. (1992). Studies on Na⁺ and H⁺ translocation through the F₀ part of the Na⁺-translocating F₁F₀ ATPase from *Propionigenium modestum*: discovery of a membrane potential dependent step. *Biochemistry* **31**, 12665–12672.
- Kluge, C. & Dimroth, P. (1993). Specific protection by Na⁺ or Li⁺ of the F₁F₀-ATPase of *Propionigenium modestum* from the reaction with dicyclohexylcarbodiimide. *J Biol Chem* **268**, 14557–14560.
- Krawiec, S. & Riley, M. (1990). Organization of the bacterial chromosome. *Microbiol Rev* **54**, 502–539.
- Krebs, M. P. & Khorana, H. G. (1993). Mechanism of light-dependent proton translocation by bacteriorhodopsin. *J Bacteriol* **175**, 1555–1560.
- Laubinger, W. & Dimroth, P. (1988). Characterization of the ATP synthase of *Propionigenium modestum* as a primary sodium pump. *Biochemistry* **27**, 7531–7537.
- Malmström, B. G. (1989). The mechanism of proton translocation in respiration and photosynthesis (the third Datta lecture). *FEBS Lett* **250**, 9–21.
- Mandel, M., Moriyama, Y., Hulmes, J. D., Pan, Y. C., Nelson, H. & Nelson, N. (1988). cDNA sequence encoding the 16-kDa proteo-

- lipid of chromaffin granules implies gene duplication in the evolution of H⁺-ATPases. *Proc Natl Acad Sci USA* **85**, 5521–5524.
- Mitchell, P. (1979).** Compartmentation and communication in living systems. Ligand conduction: a general catalytic principle in chemical, osmotic and chemiosmotic reaction systems (the ninth Sir Hans Krebs lecture). *Eur J Biochem* **95**, 1–20.
- Monticello, R. A., Angov, E. & Brusilow, W. S. A. (1992).** Effects of inducing expression of cloned genes for the F₀ proton channel of the *Escherichia coli* F₁F₀ ATPase. *J Bacteriol* **174**, 3370–3376.
- Nelson, N. (1989).** Structure, molecular genetics, and evolution of vacuolar H⁺-ATPases. *J Bioenerg Biomembr* **21**, 553–571.
- Nelson, N. (1994).** Energizing porters by proton-motive force. *J Exp Biol* **196**, 7–13.
- Nelson, N. & Taiz, L. (1989).** The evolution of H⁺-ATPases. *Trends Biochem Sci* **14**, 113–116.
- Olsen, G. J., Woese, C. R. & Overbeek, R. (1994).** The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* **176**, 1–6.
- Pao, G. M. & Saier, M. H., Jr (1994).** The N-terminal, putative, mitochondrial targeting domain of the mitochondrial genome maintenance protein (MGM1) in yeast is homologous to the bacterial ribonuclease inhibitor, barstar. *Mol Biol Evol* **11**, 964–965.
- Pearson, W. R. & Lipman, D. J. (1988).** Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Pedersen, P. L. & Amzel, L. M. (1993).** ATP synthases. Structure, reaction center, mechanism, and regulation of one of nature's most unique machines. *J Biol Chem* **268**, 9937–9940.
- Reizer, J., Reizer, A. & Saier, M. H., Jr (1994).** A functional superfamily of sodium/solute symporters. *Biochim Biophys Acta* **1197**, 133–166.
- Riley, M. (1993).** Functions of the gene products of *Escherichia coli*. *Microbiol Rev* **57**, 862–952.
- Saier, M. H., Jr (1994).** Computer-aided analyses of transport protein sequences: gleaned evidence concerning function, structure, biogenesis, and evolution. *Microbiol Rev* **58**, 71–93.
- Saier, M. H., Jr, Werner, P. & Müller, M. (1989).** Insertion of proteins into bacterial membranes: mechanism, characteristics and comparisons with the eukaryotic process. *Microbiol Rev* **53**, 333–366.
- Saier, M. H., Jr, Fagan, M. J., Hoischen, C. & Reizer, J. (1993).** Transport mechanisms. In *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 133–156. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Schneider, E. & Altendorf, K. (1984).** Subunit b of the membrane moiety (F₀) of ATP synthase (F₁F₀) from *Escherichia coli* is indispensable for H⁺ translocation and binding of the water-soluble F₁ moiety. *Proc Natl Acad Sci USA* **81**, 7279–7283.
- Schneider, E. & Altendorf, K. (1985).** All three subunits are required for the reconstitution of an active proton channel (F₀) of *Escherichia coli* ATP synthase (F₁F₀). *EMBO J* **4**, 515–518.
- Schneider, E. & Altendorf, K. (1987).** Bacterial adenosine 5'-triphosphate synthase (F₁F₀): Purification and reconstitution of F₀ complexes and biochemical and functional characterization of their subunits. *Microbiol Rev* **51**, 477–497.
- Senior, A. E. (1990).** The proton-translocating ATPase of *Escherichia coli*. *Annu Rev Biophys Biophys Chem* **19**, 7–41.
- Serrano, R. (1988).** Structure and function of proton translocating ATPase in plasma membranes of plants and fungi. *Biochim Biophys Acta* **947**, 1–28.
- Smith, D. W. (1988).** A complete, yet flexible, system for DNA/protein sequence analysis using VAX/VMS computers. *Comput Appl Biosci* **4**, 212.
- Soloz, M. & Davies, K. (1994).** Operon of vacuolar-type Na⁺-ATPase of *Enterococcus hirae*. *J Biol Chem* **269**, 9453–9459.
- Stokes, D. L. & Nakamoto, R. K. (1994).** Structures of P-type and F-type ion pumps. *Curr Opin Struct Biol* **4**, 197–203.
- Stone, D. K., Crider, B. P., Südhof, T. C. & Xie, X.-S. (1989).** Vacuolar proton pumps. *J Bioenerg Biomembr* **21**, 605–620.
- Takase, K., Kakinuma, S., Yamato, I., Konishi, K., Igarashi, K. & Kakinuma, Y. (1994).** Sequencing and characterization of the *ntp* gene cluster for vacuolar-type Na⁺-translocating ATPase of *Enterococcus hirae*. *J Biol Chem* **269**, 11037–11044.
- Van Rosmalen, M. & Saier, M. H., Jr (1993).** Structural and evolutionary relationships among two families of bacterial periplasmic chaperone proteins which function cooperatively in fimbrial assembly. *Res Microbiol* **144**, 507–527.
- Vik, S. B. & Dao, N. N. (1992).** Prediction of transmembrane topology of F₀ proteins from *Escherichia coli* F₁F₀ ATP synthase using variational and hydrophobic moment analyses. *Biochim Biophys Acta* **1140**, 199–207.
- Walker, J. E. (1994).** The regulation of catalysis in ATP synthase. *Curr Opin Struct Biol* **4**, 912–918.
- Walker, J. E., Runswick, M. J. & Saraste, M. (1982).** Subunit equivalence in *Escherichia coli* and bovine heart mitochondria F₁F₀ ATPases. *FEBS Lett* **146**, 393–396.
- Woese, C. R. (1987).** Bacterial evolution. *Microbiol Rev* **51**, 221–271.
- Wu, L.-F. & Saier, M. H., Jr (1990).** Nucleotide sequence of the *fruA* gene, encoding the fructose permease of the *Rhodobacter capsulatus* phosphotransferase system, and analyses of the deduced protein sequence. *J Bacteriol* **172**, 7167–7178.
- Yokoyama, K., Oshima, T. & Yoshida, M. (1990).** *Thermus thermophilus* membrane-associated ATPase. Indication of a bacterial V-type ATPase. *J Biol Chem* **265**, 21946–21950.
- Yokoyama, K., Akabane, Y., Ishii, N. & Yoshida, M. (1994).** Isolation of prokaryotic V₀V₁-ATPase from a thermophilic eubacterium *Thermus thermophilus*. *J Biol Chem* **269**, 12248–12253.
- Zhang, Y. & Fillingame, R. H. (1994).** Essential aspartate in subunit c of F₁F₀ ATP synthase. Effect of position 61 substitutions in helix-2 on function of Asp²⁴ in helix-1. *J Biol Chem* **269**, 5473–5479.
- Zimányi, L., Váró, G., Chang, M., Ni, B., Needleman, R. & Lanyi, J. K. (1992).** Pathways of proton release in the bacteriorhodopsin photocycle. *Biochemistry* **31**, 8535–8543.

Received 13 June 1995; revised 23 August 1995; accepted 12 September 1995.