Evolutionary appearance of H⁺-translocating pyrophosphatases

Proteins of the proton-pumping pyrophosphatase (H⁺-PPase) family are found in the vacuolar (tonoplast) membranes of higher plants, algae and protozoa, and in both bacteria and archaea (Baltscheffsky & Baltscheffsky, 1993; Baltscheffsky et al., 1999; Drozdowicz et al., 1999; Kim et al., 1994, 1995; Sarafian et al., 1992a, b). They are therefore presumed to be ancient enzymes, having arisen before the divergence of the three domains of life (Drozdowicz & Rea, 2001). The plant, algal and archaeal enzymes, of which there may be several isoforms in a single organism, probably pump one H⁺ upon hydrolysis of pyrophosphate, thereby generating a proton motive force, positive and acidic in the tonoplast lumen or negative and basic in the cytoplasm of a prokaryote (Bäumer et al., 2002; Drozdowicz et al., 2003; Moriyama et al., 2003). These enzymes, some of which are phosphate-starvation-inducible, establish a pmf of similar magnitude to that generated by the H⁺-translocating ATPases in the same membranes (Blumwald et al., 2000; Moriyama et al., 2003; Motta et al., 2004; Palma et al., 2000; Ruiz et al., 2001; Zhen et al., 1997a, b). The bacterial and archaeal proteins may catalyse fully reversible reactions, thus being able to synthesize pyrophosphate when the pmf is sufficient (Bäumer et al., 2002; Belogurov et al., 2002; Drozdowicz et al., 1999). The enzyme from Rhodospirillum rubrum is induced in response to a number of environmental stress conditions and contributes to the pmf when light intensity is insufficient to generate a pmf that can support rapid ATP synthesis (García-Contreras et al., 2004; López-Marqués et al., 2004). All H⁺-PPases require Mg²⁺, and many, but not all of those from plant vacuoles and acidic calcisomes (acidic calcium storage compartments) of protozoa, algae, slime moulds and bacteria (Seufferheld et al., 2003) require millimolar concentrations of K⁺ (Drozdowicz et al., 2000). These proton pumps have been reported to fall into two phylogenetic subfamilies (Belogurov et al., 2002). One subfamily contains a conserved cysteine and includes the known K⁺-independent H⁺-PPases, while the other has a different conserved cysteine but lacks the first cysteine and includes the known K⁺-dependent H⁺-PPases (Belogurov et al., 2002; Drozdowicz et al., 2000; Kim et al., 1995; Zhen et al., 1994). Those from respiratory and photosynthetic bacteria as well as archaea do not seem to require K⁺, but exceptions may exist (Belogurov et al., 2002; Drozdowicz et al., 1999, 2000). It is not known whether K⁺ is transported by the K⁺-dependent enzymes. The archaeon Methanosaeta concilii Göl1 encodes within its genome two H⁺-translocating pyrophosphatases, Mvp1 and Mvp2 (Bäumer et al., 2002). Mvp1 resembles the common bacterial PPases while Mvp2 resembles plant PPases. Some bacteria also have both types of H⁺-PPases (Bäumer et al., 2002).

Both eukaryotic and prokaryotic members of the H⁺-PPase family are large proteins with 15–17 putative transmembrane α-helical spanners (TMSs) (see below). The H⁺-pyrophosphatase of Streptomyces coelicolor has been shown experimentally to have a 17 TMS topology with the substrate-binding domain exposed to the cytoplasm (Mimura et al., 2004). An extra C-terminal domain of this protein, not found in most H⁺-PPase homologues, is largely hydrophilic with a single TMS. The basic structure of these PPases is therefore believed to consist of 16 TMSs with several large cytoplasmic loops containing catalytic residues and functional motifs and both the N and C termini extracytoplasm (Mimura et al., 2004; Zhen et al., 1994). Several acidic residues in the Arabidopsis H⁺-PPase have been shown to be important for function (Zhen et al., 1997a, b). All H⁺-PPases have the enzyme commission number EC 3.6.1.1, and the transporter classification number for the H⁺-PPase family is TC 3.A.10.

In previous publications we have identified internally duplicated, triplicated and quadruplicated transmembrane segments of transporters (for a review, see Saier, 2003a). The identification of these repeat elements has provided evidence for the pathways taken for the evolutionary appearance of these important integral membrane proteins. Thus, transporters appear to have arisen independently of the other protein types by intragenic duplication of DNA elements encoding hydrophobic transmembrane channel-forming peptides (Saier, 2003a, b).

In this report we analyse the sequences of full-length H⁺-PPases with 16 putative TMSs, identifying repeat elements common to this family of proteins. This information leads to the prediction of a surprisingly unique pathway taken for their early appearance. Functional and mechanistic implications are discussed.

**Microbiology Comment** provides a forum for discussion of scientific issues arising directly from papers that have been published in the pages of Microbiology. The authors of papers under discussion will be offered an opportunity to respond.

Guidelines on how to submit a Microbiology Comment article can be found in the Instructions for Authors at http://mic.sgmjournals.org

It should be noted that the Editors of Microbiology do not necessarily agree with the views expressed in Microbiology Comment.

**Charles J. Dorman, Editor-in-Chief**
$\text{H}^+$-translocating pyrophosphatases ($\text{H}^+$-PPases) from bacteria, archaea and eukaryotes in the current NCBI database. Fifty-two non-redundant, full-length, sequence-divergent $\text{H}^+$-PPases were selected for analyses (see Table S1, available with the online version of this paper at http://mic.sgmjournals.org). These proteins contain between 659 and 823 amino acyl residues. The archaeal proteins are the smallest (mean size of about 680 residues), the eukaryotic proteins are the largest (mean size of about 770 residues) and the bacterial proteins are of intermediate size with a mean size of about 720 residues. These relative size differences are in agreement with those reported previously for other protein families (Chung et al., 2001). Using the WHAT program (Zhai & Saier, 2001a), most $\text{H}^+$-PPases were predicted to have 16 TMSs, but a few had 17 with the extra predicted TMS at the C terminus (as for the Streptomyces coelicolor enzyme). Two $\text{H}^+$-PPases had 15 predicted TMSs due to the loss of TMS1 in the 16 TMS proteins, possibly due to incorrect initiation codon assignment (see Fig. 1). These protein sequences were multiply aligned and shown to be homologous throughout most of their lengths (see Fig. S1, available with the online version of this paper at http://mic.sgmjournals.org).

Mean hydropathy, amphipathicity and similarity plots (Zhai & Saier, 2001b) for the PPases analysed are shown in Fig. 1. Sixteen peaks of hydrophobicity (labelled 1–16, thick solid line, top) correspond to 16 regions of conservation (thin solid line, bottom). In comparing the hydropathy plot with the similarity plot, it can be seen that the odd-numbered loops between TMSs are in general well conserved while the even-numbered loops are much less well conserved. Thus, the hydrophilic regions between TMSs 1 and 2, 3 and 4, 5 and 6, etc., are better conserved than those between TMSs 2 and 3, 4 and 5, 6 and 7, etc. Conservation is always observed throughout and to the left of the odd-numbered TMSs, and throughout and to the right of the even-numbered TMSs.

Baltscheffsky et al. (1999) have noted that in the $\text{H}^+$-PPase of R. rubrum, there is a 45-residue segment, corresponding to loop 5, that shows some sequence similarity with a segment in loop 15. These investigators noted that an intragenic duplication event could have occurred, accounting for this similarity. However, no statistical data or evidence for the nature of such a duplication supporting this suggestion were presented.

We investigated the evolutionary origin of $\text{H}^+$-PPase segments using the GAP (Devereux et al., 1984) and IC (Zhai & Saier, 2002) programs. We first showed that TMSs 1–6 are homologous to TMSs 11–16. Thus, when TMSs 1–6 of the PPase of Chloroflexus aurantiacus was compared with TMSs 11–16 of the homologue from Rubrivivax gelatinosus using the GAP program with default settings and 500 random shuffles, the alignment shown in Fig. 2 was obtained. This alignment gave 28·9 % identity and 36·9 % similarity with a comparison score of 9·8 SD. This last value indicates that the degree of sequence identity/similarity observed is statistically significant; this degree of sequence identity could have occurred by chance with a probability of $<10^{-30}$, a value that is considered sufficient to establish homology as discussed previously (Saier, 1994).

Returning to Fig. 1, the hydropathy plots of the TMS 1–6 segment and the TMS 11–16 segment are strikingly similar. Thus, peaks 1 and 11 are distant from peaks 2 and 12, respectively, which, however, are close to peaks 3 and 13; peaks 3 and 13 are distant from peaks 4 and 14, which, however, are close to peaks 5 and 15; and peaks 5 and 15 are distant from peaks 6 and 16. Moreover, the larger cytoplasmic loops are always better conserved than the smaller extracellular loops with the single exception of the loop between TMSs 13 and 14 as noted above. Even in this last mentioned loop, there is extensive conservation to the right of TMS 13 and to the left of TMS 14. Therefore, even this cytoplasmic loop contains conserved regions near the flanking TMSs.

Further evidence for the homology of TMSs 1–6 with TMSs 11–16 resulted from the analysis of amphipathicity with the angle per residue set at 100˚as is appropriate for an $\alpha$-helix (Fig. 1, dotted line). Striking peaks of amphipathicity occur between TMSs 1 and 2 as well as TMSs 11 and 12 with none occurring between TMSs 2 and 3 or TMSs 12 and 13. Two amphipathic peaks can be seen between TMSs 3 and 4 as well as

---

**Fig. 1.** Mean hydropathy (solid line, top), amphipathicity (dashed line, top) and similarity (bottom) for 52 full-length $\text{H}^+$-translocating pyrophosphatases. The sequences were aligned using the CLUSTAL X program with default settings (Thompson et al., 1997) (see Table S1 and Fig. S1, available with the online version of this paper at http://mic.sgmjournals.org). The plots were generated with the AveHas program (Zhai & Saier, 2001b). Alignment position is present on the x axis.
TMSs 13 and 14, but not between TMSs 4 and 5 or TMSs 14 and 15. Finally, there are two clear peaks of amphipathicity between TMSs 5 and 6 as well as TMSs 15 and 16. These similarities provide additional evidence for homology and suggest that striking structural similarities between the two repeat units have been retained during their evolutionary divergence.

Analysis of the multiple alignment (Fig. S1, available with the online version of this paper at http://mic.sgmjournals.org) revealed that the two most conserved regions in the H⁺-PPase are the loops between TMSs 5 and 6 and TMSs 15 and 16. Several residues in each of the two sequences in these regions are residues that are fully conserved in all 52 proteins. Surprisingly, most of the conserved residues are not conserved between the two homologous segments. Thus, it can be concluded that loop 5–6 has diverged in sequence from loop 15–16, and that these two regions serve distinct but possibly overlapping functions.

The results summarized above demonstrated that TMSs 1–6 and 11–16 arose from a common ancestral element, probably by intragenic duplication. We then asked what the origin of the central four TMSs (TMSs 7–10) might be. An example of our analyses is presented in Fig. 3. TMSs 7–10 of the PPase from Thermoanaerobacter tengcongensis showed 25.4% identity and 37.7% similarity to TMSs 1–4 of the enzyme from Dehalococcoides ethenogenes with a comparison score of 101 SD. TMSs 7–10 and 1–4 are therefore homologous. The duplication event resulting in the presence of TMSs 7–10 may have included only TMSs 1–4 (or TMSs 11–14) but not TMSs 5 and 6 (or TMSs 15 and 16); alternatively the duplication event might have originally included all 6 TMSs, but the last two TMSs in the second repeat element were subsequently deleted early in the evolutionary process, before the speciation and extragenic duplication events that gave rise to the many current members of the H⁺-PPase family.

Fig. 2. Alignment of TMSs 1–6 from the putative H⁺-PPase of Chloroflexus aurantiacus (Cau; gi #53796873) with TMSs 11–16 from the H⁺-PPase of Rubrivivax gelatinosus (Rge; gi #47575275). The residue number of the 6 TMS segment is presented at the beginning and end of each line. Vertical bars, identities; colons, close similarities; single points, more distant similarities. The numbered rectangular bars (residue in white on a black background) show the positions of the putative TMSs, numbered as shown in Fig. 1. The GAP program (Devereux et al., 1984) was used to generate the alignment.

Fig. 3. Alignment of TMSs 1–4 of the putative H⁺-PPase from Dehalococcoides ethenogenes (Det1; gi #57234445) with TMSs 7–10 of the PPase from Thermoanaerobacter tengcongensis (Tte; gi #20806805). The convention of presentation is as described in the legend to Fig. 2. The alignment was generated using the GAP program.
These larger proteins arose by loss of 2 TMSs. The pathway could well involve triplication of a primordial H⁺-translocating inorganic pyrophosphatase: from the evolutionary backwaters into the mainstream. Trends Plant Sci 6, 206–211.


Garcia-Contreras, R., Celis, H. & Romero, I. (2004). Importance of Rhodospirillum rubrum...


DOI 10.1099/mic.0.28581-0

**A cultural divide on the use of chemostats**

The attractions of chemostat studies are indeed too often overlooked as suggested by Hoskisson & Hobbs (2005). My laboratory routinely uses continuous cultures to provide reproducible conditions for global regulation and proteomics, the over-simplifications of Hoskisson & Hobbs (2005) need to be balanced by pointing out some of the pitfalls as well.

The first obvious point is that chemostats induce a particular nutrient-limited state. Nutrient limitation results in bacterial responses not only specific for the type of limiting nutrient, but also responses that are far from uniform for a particular nutrient. The choice of different growth or dilution rates will critically influence the expression patterns obtained (Ferenci, 1999a; Harder & Dijkhuizen, 1983). For example, a glucose-limited *Escherichia coli* cell grown at a dilution rate of 0·1 h⁻¹ is near starvation and has highly induced stress responses, whereas bacteria grown at 0·6 h⁻¹ have elevated hunger responses and increased nutrient scavenging activities (Ferenci, 2001). Dilution rates in between have a varying mix of both responses. This aspect is poorly appreciated and not well integrated into many of the studies cited in Hoskisson & Hobbs, where mostly an arbitrary dilution rate is chosen.

A second point that needs elaborating is that nutrient limitation is an extremely strong selection condition for mutational changes that rapidly sweep populations. Indeed, the historical introduction misses the aim of Novick & Szilard (1950) who were more interested in the evolutionary applications of continuous culture than in the production of reproducible, steady-state bacteria (which was Monod’s aim). The conditions to establish a steady state suggested by earlier workers (multiple chemostat culture volumes to achieve a physiological ‘steady state’) are indeed sufficient for *E.coli* populations to be swept by rpoS mutations (Notley-McRobb et al., 2002). An added complication is that the type of takeover is very strain-dependent, even amongst laboratory strains of *E.coli* (King et al., 2004). So before functional genomic assays are assessed, the time-course of changes needs to be carefully controlled. The dichotomy between the desire for ‘steady state’ in functional genomics studies and the rapid selection for change needs to be appreciated.

The third point missed by Hoskisson & Hobbs (2005) is that there is actually no true steady-state in chemostats for the simple reason that the residual concentration of limiting nutrient...