UNIVERSITY OF CALIFORNIA, SAN DIEGO

P-type ATPases in the Proteobacterial, Bacteroidetes and Fusobacterial phyla

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Kris Kumar

Committee in charge:

Professor Milton Saier, Jr., Chair
Professor Stephen Baird
Professor Randy Hampton

2009

The Thesis of Kris Kumar is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____
Chair

University of California, San Diego

2009

TABLE OF CONTENTS

LIST OF FIGURES

Figures in Chapter 1:

Figures in Chapter 2:

LIST OF TABLES

Tables in Chapter 1:

Tables in Chapter 2:

Tables in Discussion:

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Milton Saier, Jr. for his help and support while completing my research. It has been invaluable to be able to learn from him and have his constant guidance during the course of my research project. In addition, I would like to acknowledge the Saier lab community for their help and assistance, in addition to their friendship.

Parts of this Thesis are being prepared for publication. The Thesis author along with the committee chairman, Dr. Milton Saier Jr., will be a co-investigators and co-authors of this paper.

ABSTRACT OF THE THESIS

P-type ATPases in the Proteobacterial, Bacteroidetes and Fusobacterial phyla

by

Kris Kumar

Master of Science in Biology

University of California, San Diego, 2009

Professor Milton Saier, Jr., Chair

P-type ATPases are a class of cation transporting proteins located within the membranes of cells, both eukaryotic and prokaryotic. While much has been elucidated regarding the nature and function of eukaryotic P-type ATPases, much less in known about the prokaryotic enzymes. This study focuses upon the analysis of these enzymes from the Proteobacterial, Bacteroidetes and Fusobacterial phyla. 47 P-type ATPases were found in the *Bacteroides*, *Flavobacterium*, and *Fusobacterium* genuses while 218 members comprised the Protoebacterial phylum. Genomic searches of organisms with fully sequenced genomes revealed many such enzymes, which were then combined with known standards of P-type ATPases from the nine functionally characterized as well as

32 functionally uncharacterized ATPases to classify them phylogenetically. Familial representation of these proteins varied, with *Bacteroides*, *Flavobacterium*, and *Fusobacterium* genuses having homologues of Families 2, 4-7, 29, 30 and 32, and the Proteobacterial phylum having homologues of Families 1-7, 25, 27-32. In addition, 16S rRNA trees were constructed; showing largely orthologous relationships among the proteins from *Bacteroides*, *Flavobacterium*, and *Fusobacterium* genuses with no evidence for horizontal gene transfer, but Proteobacteria exhibited high frequencies of horizontal gene transfer. Motif analyses were conducted in order to further classify the proteins. All proteins showed conservation at the crucial phosphorylation site required for P-type ATPase activity, and varying degrees of motif conservation for eight other motifs analyzed. Topological analyses classified the P-type ATPases according to hydropathy profiles, and a novel topological type, Type VII, was discovered.

Introduction

In most eukaryotic and prokaryotic organisms, ions are transported in and out of cells with the assistance of a class of active ion transporters named P-type ATPases. The Gibbs free energy derived ATP from hydrolysis provides the energy required to power transport of these ions across membranes. This energy allows the build up of ionic electrochemical potential gradients (Apell, 2004). The P-type ATPases are very vital to many different biological processes. Nutrient absorption, Ca^{2+} signal transduction (Axelsen, 1997), removal of toxic levels of ions within cells and generating the membrane potential needed for muscular contractions are just a few of their important functions (Kühlbrandt, 2004).

This study focuses upon the nine functionally characterized P-type ATPase families along with the 26 functionally uncharacterized P-type ATPase (FUPA) families listed within the Transport Classification Database (TCDB). Of the characterized P-type ATPase families, Family 1 ATPases transport Na^+, K^+ ions, Family 2 transports Ca^{2+}, and Family 3 enzymes catalyze H^+ transport, while Family 4 ATPases transport Mg^{2+}. Family 5 enzymes transport Cu^{2+} ions with either inwardly or outwardly directed polarity, while Family 6 P-type enzymes are responsible for the transport of Heavy Metal ions including Zn^{2+}, Cd^{2+}, Co^{2+}, Hg^{2+}, Ni^{2+}, Cu^{2+}, Ag^+ and Pb^{2+}. Family 7 transporters, or Kdp-type transporters, comprise the class of enzymes responsible for K^+ transport. Families 1-6 are all found in both eukaryotic and prokaryotic organisms. Family 8 P-type ATPases are found only in eukaryotes, and are responsible for phospholipid uptake. The last of the

characterized families, Family 9 ATPases, transport Na^+ and or K^+ and are found primarily in fungi and protozoa.

The name P-type ATPases comes from a significant feature during the enzyme reaction cycle, the formation of a phosphorylated intermediate. This phosphorylation state is thereby responsible for the designation, P-type (Axelsen, 1997).

Bacterial P-type ATPases are found in cytoplasmic membranes of these organisms and are oriented in such a way that one side of the pump faces the cytoplasm while the other faces the outside of the cell. Activity is based upon the transition between two main conformations of the P-type enzyme, E1 and E2, which are regulated by phosphorylation and dephosphorylation (Okkeri et al., 2002). Ions moving from inside the cell to the extracellular space bind deep within the transmembrane region. Once these ions are bound, phosphorylation of the enzyme is promoted. Phosphorylation is dependent on ATP, which is in turn converted to two parts, the phosphate moiety that binds the P-type ATPase and the ADP moiety that is momentarily associated with the enzyme (Gadsby, 2007). When the ion and resulting phosphate group are bound, the pump enters the E1P state in which ions cannot be accessed from either side of the ATPase. This phase of occlusion allows for greater control of P-type ATPase function so access to the binding regions varies between the two different sides of the enzyme located within the membrane. The ADP, after momentarily associating with the pump, is released, allowing the ATPase to relax into an E2P state. This causes relaxation of the pump, opening of the extracellular side of P-type ATPase and release of the ions.

The reverse reaction can occur when ions from the extracellular space enter the P-type ATPase and bind to binding sites in the E2P conformation. This binding then

dephosphorylates the enzyme, leaving the transporter in an E2 state in which the ions are occluded as discussed previously. The ATPase then goes from the occluded E2 state to the E1 conformation, relaxes, and releases the ions into the cytoplasm.

The phosphorylation state is achieved via the formation of a phosphorylated aspartyl residue within the P-type ATPase. This aspartate is located within one of the nine conserved P-type ATPase motifs, motif 5, DKTGLT (Møller et al., 1996). The aspartate in motif 5 is located within the large cytosolic domain C, which is also known as the nucleotide-binding domain. In addition to domain C, region B flanks the N-terminal part of domain C, and region J flanks the C-terminal of domain C and connects it with the C-terminal transmembrane region of the P-type ATPase. Within the nucleotide-binding region of domain C, lies the aspartyl phosphorylation site (Okerri, 2002).

While much has been discovered regarding eukaryotic P-type ATPases, the enzymes in prokaryotic species are less well studied (Bramkamp et al., 2007). This study is divided into two parts, the first examining the P-type ATPases of the phyla Bacteroidetes and Fusobacteria. Eleven species of bacteria within the closely related gram-negative genuses of *Bacteroides, Flavobacterium* and *Fusobacterium* were analyzed. *Bacteroides* are typically symbionts that occupy the intestines of mammalian species, including the clinically important *Bacteroides fragilis*. While they do play a useful role in mammals, at times they can become opportunistic pathogens. *Flavobacterium* species on the other hand are generally found within the soil and freshwater sources, and are able to cause disease within fish such as salmon and trout

(Duchaud et al., 2007). Members of the genus *Fusobacterium* are usually found within mucous membranes of humans, and generally cause diseases in humans.

The second part of this study focused upon members of the phylum Proteobacteria. Proteobacteria are further subdivided into five different classes designated by the Greek letters, α, β, γ, δ, and ε. All five classes of Proteobacteria contain a wide variety of species that can be either pathogenic or harmless; most can be free living.

In this study, phylogenetic analyses were undertaken in order to classify the bacterial P-type ATPases based on their relationships to the enzymes of known function as well as the FUPA proteins. In addition, 16S rRNAs are studied to allow identification of orthology or horizontal gene transfer of the proteins and familial relationships between species. Multiple alignments containing members of the same species are used in order to look at the nine conserved motifs and understand the conservation of a motif within a particular family. Finally, multiple alignments were used to understand the topological types and classify the families based on the known topologies of P-type ATPases. A novel topology is also identified.

Methods

Note – Methods used are applicable to both parts of this study.

Protein Sequence Collection

Representatives of P-type ATPases specific to the organisms analyzed in this study were initially compiled from Membranetransport.org (Paulsen et al., 2002). This database has membrane transport proteins for a wide variety of organisms whose genomes have undergone full sequencing. Out of the proteins found, only full-length proteins were kept, and others that did not meet this standard were discarded.

Genomic BLAST (Basic Local Alignment Search Tool) was performed on each of the organisms included in this study. BLAST searches revealed multiple P-type ATPases per organism, and proteins that showed good sequence similarity to the proteins used in the BLAST were kept. Redundant proteins and proteins that were not of full length were omitted. Redundancies were checked using BLAST2, which compared two sequences against each other for sequence similarity. Sequences that showed identity after BLAST2 were analyzed for length, and only the longer of the two sequences was kept to ensure that no vital information was discarded and lost.

Phylogenetic Anayses

Using the program ClustalX (Thompson et al., 1997), multiple alignments were created of the protein sequences (S2 and S4), the 16S rRNAs of the organisms for the *Bacteroides*, *Flavobacterium* and *Fusobacterium* and the 16SrRNAs from each genus represented in Proteobacteria, and the protein sequences along with standards of P-type

ATPases from the Transport Classification Database (TCDB) (S1 and S3). The standard proteins used in this study included both the functionally characterized P-type ATPases and the functionally uncharacterized P-type ATPases (FUPA) listed within TCDB. These proteins helped to serve as references so that familial assignment of the proteins included within this study could be determined. Phylogenetic trees were then created using the TreeView program. 16S rRNA trees were used in order to determine whether or not the proteins in this study were orthologous or had undergone horizontal gene transfer.

Hydropathy Anayses

Multiple alignments of the representatives of each family were created using ClustalX. These multiple alignments were then used in the programs AveHAS and WHAT to generate hydropathy plots of the proteins (Zhai and Saier, 2001). AveHAS is used when the number of proteins in a particular family exceeds one, as it takes an average of the hydropathy for a particular family. The WHAT program on the other hand, is used in cases in which the number of proteins for a particular family is just one. The hydropathy analyses are able to determine the hydrophobicity of various locations within the proteins or a single protein. As a result, the locations of any transmembrane segments (TMS) can be estimated.

Regions at which there existed significant N-terminal or C-terminal extensions that altered hydropathy plots were further analyzed. Residues from these extensions were isolated, and BLAST searches were performed in hopes of elucidated whether or not conserved domains existed within those regions of the P-type ATPases. This allowed for a clearer understanding of the nature and function of these particular regions.

Motif Analyses

The nine conserved motifs within P-type ATPases (Møller et al., 1996) were analyzed for all of the families in this study. By looking at the multiple alignments of a particular family using ClustalX, conservation of these motifs was determined.

Chapter 1. *Bacteroides*, *Flavobacterium* and *Fusobacterium* P-type ATPases

The fully sequenced genomes of eleven bacterial species within the closely related Bacteroidetes and Fusobacterial phyla, were analyzed for P-type ATPases. 11 organisms derived from the genuses *Bacteroides*, *Flavobacterium* and *Fusobacterium* were studied and had a total of 28, 9, and 10 enzymes respectively. The total number of P-type ATPases identified in this study was 47.

The average number of enzymes per organism was found to be around three. Of all of the organisms, *Bacteroides fragilis* NCTC 9343 and *Flavobacterium johnsoniae* UW101 had the largest number of P-type ATPases with six each (Table A1). *Flavobacterium psychrophilum* JIP02/86 and *Fusobacterium nucleatum subsp. vincentii* ATCC 49256 had the least number of P-type ATPases with three and two such enzymes respectively. Overall, variation in the number of P-type ATPases correlated to the size in mega base pairs (Mbp) of the genome (Table A1 & Figure A3).

The representation of P-type ATPases according to family proved interesting. Eleven Family 5 (Cu^{2+}) and Family 6 (Heavy Metal (HM)) were identified (Table A2). Each organism analyzed has exactly one Heavy Metal ATPase. At least one member from Family 5 is seen within the eleven organisms except in *Flavobacterium psychrophilum* JIP02/86 and *Fusobacterium nucleatum subsp. vincentii* ATCC 49256. Family 2 (Ca^{2+}) and Family 7 (K^{+}) each have eight P-type ATPases, with organisms from the genus *Fusobacterium* lacking these K^{+} transporters. Family 4 (Mg^{2+}) has only 3 members, all of which are found in species of *Bacteroides* (Table A2). In addition, functionally uncharacterized P-type ATPases from Families 29, 30 and 32 contained two, one and three enzymes respectively.

The relationship between genome size of an organism and the number of ATPases

Figure 3A shows the relationship between the genome size of the eleven organisms included in this study plotted against the number of P-type ATPases within the genome of that particular organism. A best-fit line was drawn, and approximately a linear relationship was observed (Figure A3).

Table A1 lists the organisms included in this study and provides information about the genome sizes and numbers of open reading frames recognized. It also lists the numbers of P-type ATPases that were found encoded within these genomes. The largest genome recorded is that of *Bacteroides ovatus* ATCC 8483, 6.5 Mbp, but *Flavobacterium johnsoniae* UW101 and *Bacteroides thetaiotaomicron* VPI-5482 have genomes that are almost as large (6.1 Mbp and 6.3 Mbp, respectively). All three organisms encode 5-6 P-type ATPases. The small genome of *Fusobacterium nucleatum subsp. vincentii* ATCC 49256 (2.1 Mbp) encodes just 2 P-type ATPases. This suggests a rough correlation between genome size and numbers of encoded P-type ATPases as seen in Table A1 and Figure A3.

Phylogenetic Analyses

The phylogenetic relationships between the 47 P-type ATPases are seen in Figure A1. In addition, phylogenetic relationships between the proteins in this study with standards from TCDB are shown in Figure 2A. Figure 1A shows the enzymes clustering according to familial type, with five clusters representing the characterized families and three clusters that represent the functional uncharacterized P-type ATPases (FUPAs) (Figure A1). The five characterized families cluster with proteins specific for Ca^{2+}

(3.A.3.2), Mg^{2+} (3.A.3.4), Cu^{+} (3.A.3.5), Heavy Metal (3.A.3.6), and K^{+} (Kdp, 3.A.3.7) (Figure 1B). The FUPA proteins cluster with proteins specific for FUPA29 (3.A.3.29), FUPA30 (3.A.3.30), and FUPA32 (3.A.3.32) (Figure A2). The three FUPA32 proteins cluster loosely with members from Family 6 (Figure A1). The lone FUPA30 protein clusters loosely between Family 2 and Family 4 proteins (Ca^{2+} and Mg^{2+}). FUPA29 does not cluster loosely with any of the established families of P-type ATPases.

A phylogenetic tree of the 16S rRNAs of the eleven organisms was also made to study possibly orthology or horizontal gene transfer (Figure A4). Each organism segregated with members from its respective genus. The two species from the genus *Flavobacterium* clustered loosely by the six members of the *Bacteroides* genus. Members of the *Fusobacterium* genus clustered further away from the other two genuses (Figure A4).

Ca^{2+} homologues

Eight members are homologous members of the calcium efflux family (3.A.3.2). These members range from 862-901 amino acids in length (Table A3). There are two subclusters within these proteins as seen in Figure 1A. Six of these proteins, one each from the six *Bacteroides* species, all probably orthologous, comprise the first of the two subclusters. The second includes two *Fusobacterium nucleatum* proteins from different strains showing a high degree of sequence identity as expected. No Ca^{2+}-type *flavobacterial* ATPase homologues were identified.

Mg^{2+} homologues

Loosely clustering near Family 2 are the Magnesium Family 4 homologues. Only three members of this family are seen in one cluster (Figure A1), with amino acid lengths for each protein at 883 amino acids (Table A3). These proteins (Bth1, Bov1 and Bfr1) are likely to be orthologous *Bacteroides* proteins. No Mg^{2+} systems were found in *Flavobacteria* or *Fusobacteria*. It is thus clear that phylogenetic clustering patterns are dictated by the phylogenies of the source organisms, suggesting that these organisms were subject to different evolutionary pressures. The results also suggest a lack of horizontal ATPase gene transfer between the three organismal types.

Cu^{+} and Heavy Metal homologues

The two largest families represented in these organisms are the Family 5 Copper (3.A.3.5) and Family 6 Heavy Metal (3.A.3.6) ATPase families with eleven members each. For the copper ATPases, amino acids lengths range from 735-845 amino acids (Table A3). Family 5 shows three early diverging subclusters. The first subcluster includes two orthologs from two different strains of *Fusobacterium nucleatum*. The second includes two paralogs from *Flavobacterium johnsoniae* UW101, and the third includes seven proteins from different species of *Bacteroides*. Five proteins in this last subcluster *(*Bov3, Bth3, Bca2, Bfr3 and Bvu2) are all likely to be orthologues (Figure A1).

For the heavy metal ATPases, residue lengths range from 572-663 amino acids (Table A3). Family 6 includes three primary subclusters. The first subcluster includes six proteins, each from one of the six different species of *Bacteroides* examined; the second

subcluster includes two proteins from the two different species of *Flavobacterium*, and the third subcluster includes three proteins from different species of *Fusobacterium* (Figure A1). All of these proteins are probably orthologous as their relative distances from each other are similar to those for the organismal 16S rRNAs (Figure A4). In both families, the proteins segregate according to organismal phylogeny. This fact strongly suggests that there has been essentially no horizontal transfer of the encoding genetic material between these three organismal types over a period of at least one, and more likely two billion years.

Kdp-type K^+ homologues

The Kdp-type K^+ uptake ATPases (3.A.3.7), are well represented with eight members, each from a different organism, with amino acid lengths ranging from 677 to 685 amino acids (Table 3A). They fall into two closely related subclusters. The first includes a set of six proteins from the six *Bacteroides* species, all of which are likely to be orthologous. The second cluster is represented by proteins from the two different species of *Flavobacterium* examined. These two proteins are also likely to be ortholgous to each other, and may be orthologous to the *Bacteroides* proteins as well. Fusobacteria seem to lack these ATPases.

Functionally uncharacterized family 29 (FUPA29)

Of the families with no known function, FUPA29 contains just two members (Table A2). FUPA29 family homologues (792 and 795 amino acids) are derived exclusively from *Flavobacterium* species, each protein from a different species. Members

of this family most closely resemble Cu^{+} and Heavy Metal ATPases, both in terms of topology, (Topological Type I; see below) and phylogeny (Figure A1).

Functionally uncharacterized family 30 (FUPA30)

Finally a single protein from *Flavobacterium johnsoniae* UW101, Fjo1, represents the third family of unknown function. This protein has 838 amino acyl residues (Table A3). Because it belongs to the previously identified FUPA30 family, one can assume that it is not an orphan protein, and most likely, not a pseudogene. This protein and other FUPA30 family members are very loosely associated with the large cluster of Ca^{2+} and Mg^{2+} ATPases, and exhibit the Type II topology (Figure A1).

Functionally uncharacterized family 32 (FUPA32)

FUPA32 contains three members (Table A2) with the residue number for all three proteins at 735 amino acids (Table A3). FUPA32 homologues consist of a cluster of three proteins derived from three distinct strains of *Fusobacterium nucleatum*. Like FUPA29 members, FUPA32 members also resemble Cu^{+} and Heavy Metal ATPases, both in terms of topology (Topological Type I; see below) and phylogeny (Figure A1).

Conserved Motifs

P-type ATPases have nine well-conserved motifs in their primary structure (Møller et al., 1996). These nine motifs, sequentially numbered from the N terminus to the C terminus, 1-9, are well recognizable when viewing the multiple alignment

sequences of individual P-type ATPase families. These motifs are usually found in P-type ATPases with Type I and Type II topologies.

The nine motifs are fount in three different cystolic locations, Region (domain) B, Region (domain) C and Region (domain J). Towards the N terminus, Region B is the small cytosolic loop. This region contains three motifs of high sequence similarity (Møller et al., 1996). The first motif, Motif 1 (PGD) is closest to the N terminus and is followed closely by Motif 2 (PAD). Motif 3, TGES, is very important to P-type ATPase activity. A mutation in any of these residues can lead to a reduction in ion transport, showing that structural modification can lead to inhibition of the ATPase and loss of activity (Møller et al., 1996 & Fagan and Saier, 1994). Region B has an antiparallel specific β-sheet. In addition, Region B is important for energy transduction. Mutations in the region can reduce rates of ion movement. This reduction in ion movement occurs without loss of ion binding or phosphorylation capacity (Møller et al., 1996). Motif 4, PEGL, is found in transmembrane segment 4 (TMS4). This motif can cause reduction in energy transduction to a modest extent, particularly when the well conserved proline is mutated (Møller et al., 1996).

Region C, or the large cytosolic loop, starts before the phosphorylation site (N-terminus) and ends before Region J (Møller et al., 1996). Motif 5 in this domain, DKTGTLT, contains an aspartate (D) residue that is phosphorylated. Phosphorylation of motif 5 is extremely essential for ATPases function, and its elimination leads to loss of function. Motif 6 (KGAPE) is thought to be closely associated with ATP binding, yet may not participate directly (Møller et al., 1996). Like motif 5, motif 7 (DPPR) and Motif 8 (MVTGD) are critical for phophorylation. Motif 8 is thought to have a β-α-β structure

(Møller et al., 1996). The last motif, motif 9, located within the J region is often called the hinge motif. Parts of it are well conserved, such as the internal GDG-N region, a characteristic feature of these enzymes.

Motif 1 (PGD) proved to vary dramatically, depending upon family (Table A4). The central G is fully conserved in just four of the families, but it is the dominant residue in all but one of these families. The first residue (P) shows the poorest conservation, while the last residue (D) is reasonably well conserved. Only one family, the Family 4 Mg^{2+} ATPases, has the well conserved PGD motif. All other families have substantially altered motifs 1. The P at position 1 can be substituted by a hydrophobic residue (V or I), or the strongly hydrophilic residue, K. A fully conserved K is found in Families 7, 29 and 32. In fact, in family 32, motif 1 is almost unrecognizable, being KDD.

Motif 2 (PAD) is fully conserved in only two families, the Ca^{2+} and Mg^{2+} ATPases (Table A4). However, the D is fully conserved in all proteins. In the FUPA30 protein, the P is replaced by an N, which like P, is an α-helix breaker and participates in β-turn formation with high probability. In FUPA32, this residue is replaced by a fully conserved S. Serine, like proline and asparagine, has a low propensity for α-helix formation and a high propensity for β-turn formation (Saier, 1987). In all Type I ATPases, the A is replaced by a hydrophobic residue (V or L), but in all Type II and Type III ATPases, an A can be found at this position.

Motif 3 (TGES) is fully conserved in families 4, 6 and 7. In fact, TGE is fully conserved in most families with only the fourth residue exhibiting substantial variation. Particularly in Type I ATPases, this residue can vary depending on the family. Thus, in families 5, 6, 29 and 32, this terminal residue is P, S, A and Y, respectively (Table A4).

Interestingly, the only family that does not follow this pattern of conservation of the first three residues of this motif is Family 5. Family 5 has S at the first position instead of T, while retaining the middle two residues, G and E, with good conservation. The final residue in all Type I ATPases varies from the norm and is P.

Motif 4, PEGL, is largely conserved in the Ca^{2+} ATPases, but in the other Type II ATPases represented, Family 4, the last two residues are fully conserved but differ, being ML for Family 4. Of the Type I topological types, PEGL is replaced by the largely conserved PCAL motif except for Family 32 where this motif is fully conserved as SCGI. In these enzymes this motif is known to function in substrate binding (Møller et al., 1996). The Kdp-type K^+ ATPases also exhibit a distinctive motif 4: PTTI (Table A4).

Motif 5 (DKTGTLT), the phosphorylation site, is almost fully conserved in all families, functionally characterized and uncharacterized. The only variation occurs in the next to the last residue where the L can be conservatively substituted with I. As expected the first residue, the aspartyl phosphorylation site, is fully conserved in all proteins included within this study (Table A4).

Motif 6 (KGAPE) is relatively poorly conserved. The full KGAPE motif is found only in the Ca^{2+} ATPases with the last two residues, P and E, not being well conserved (Table A4). The Mg^{2+} ATPases have a fully conserved KGAVE motif while the FUPA30 motif is dramatically different (GNSEK). FUPA29 proteins exhibit a fully conserved motif at this position, but this motif differs at every position being IKIGS. Family 5 and 6 proteins do not exhibit good conservation of this motif.

Motif 7 (DPPR) is fully conserved in only one family, FUPA32. In the Kdp-type K^+ ATPases there is full conservation, but the motif is DIIK. In two Type II families, 4

and 30, the motif is fully conserved as DPPK. The aspartyl residue in the beginning of this motif is well conserved within all families except in FUPA29, where the D is replaced with an N (Table A4).

The last two residues in motif 8 (MVTGD) are fully conserved in every ATPase examined. Furthermore, in every family, the central residue is an aliphatic hydroxyl amino acid, T or S. However, the first two hydrophobic residues, M and V, can be substituted by almost any aliphatic hydrophobic amino acid, L, I, V or M (Table A4). The last two residues, G and D are well conserved within all of the ATPases included in this study.

Finally, motif 9 (VAVTGDGVNDSPALKKADIGVAM), the hinge motif, is surprisingly well conserved (Table A4). The internal GDG-N is fully conserved, a characteristic of most P-type ATPases. Other residues within the hinge motif exhibit variable degrees of conservation (Table A4).

Topological Analyses

Members of five functionally characterized families were identified in the *Fusobacterium, Bacteroides,* and *Flavobacterium* genuses. Average hydropathy plots were derived using the AveHAS program for all families except FUPA30, which has only one member. For this protein, Fjo1, the hydropathy plot was drawn using the WHAT program. The three established topological types are Type I, Type II and Type III. Currently, Type I consists of Cu^+ (Family 5) and Heavy Metal (Family 6) ATPases. Type II consists of Na^+, K^+ (Family 1), Ca^{2+} (Family 2), H^+ (Family 3), Mg^{2+} (Family 4), Phospholipid (Family 8) and Na^+ or K^+ (Family 9). Type III topologies occur only in the

Kdp-type K^+ Family 7. Only two of the characterized families represented (Ca^{2+}; Family 2 and Mg^{2+}; Family 4) are of the type II topology (Figures A5 and A6). Two characterized families (Copper; Family 5 and Heavy Metal; Family 6) are of the Type I topology (Figures A7 and A8). And the last family (Kdp-type K^+; Family 7) is of the Type III topology, in accordance with expectation (Figure A9).

In addition to the functionally characterized families, three previously identified functionally uncharacterized families, FUPA29, FUPA30 and FUPA32, were identified as discussed above. A FUPA29 family member had previously been identified in the δ-proteobacterium, *Bdellovibrio bacteriovorus*. Two homologues in Flavobacteria were identified, one in each of the two organisms, *Flavobacterium johnsoniae* UW101 and *Flavobacterium psychrophilium* JIP02/86. No such members were identified in the *Fusobacterium* and *Bacteroides* genuses. However, NCBI-BLAST searches revealed that members of this family are also present in *Planctomycetes (Candidatus)*, and *Sphingobacteria (Cytophaga, Algoriphagus and Pediobacter)*. These proteins in general exhibit somewhat ambiguous topologies. However, most likely, they exhibit the Type I topology, although in some cases, an additional TMS, preceding peaks A and B, is predicted, called peak A' (Figure A10). Indeed, this suggestion is in agreement with the very loose clustering pattern observed for these proteins together with the Copper and Heavy Metal ATPases (Figure A1). BLAST searches revealed that FUPA29 homologues are most closely related to FUPA27 proteins, derived from α-, β- and γ-proteobacteria as well as spirochetes. Unexpectedly, the FUPA27 proteins also show ambiguous hydropathy plots with a single potential TMS sometimes predicted to precede TMSs A and B. These observations confirm a close relationship between these two families, and

suggest that FUPA27 and FUPA29 proteins are both phylogenetically and topologically similar. These homologues might also serve a similar function, possibly transporting, for example, heavy metals. FUPA29 proteins appear to exhibit in a new topological type, classified as Type VII, closely related to Type I (Figure A10).

In contrast to FUPA29 proteins, a single FUPA30 homologue is found in *Flavobacterium johnsoniae* UW101. However, homologues are also found in δ-proteobacteria, spirochetes, and as shown here, *Flavobacteria*. The *Flavobacterium johnsoniae* UW101 protein, Fjo1 exhibits a typical Type II topology as is true of other FUPA30 homologues (Figure A11). This observation is in agreement with the phylogenetic analysis shown in Figure A1 where the FUPA30 homologue clusters loosely with Ca^{2+} ATPases, and even more loosely with Mg^{2+} ATPases. FUPA30 proteins may transport Ca^{2+}, based upon topology and phylogeny.

The third uncharacterized family represented in organisms most closely related to and included within the Bacteriodetes phylum is the FUPA32 family. α-, β-, δ-, γ- and ε-Proteobacteria, actinobacteria, firmicutes and spirochetes also possess members of this large family. These proteins are phylogenetically about equally distantly related to the known Copper and Heavy Metal ATPases. Although these proteins gave variable topological predictions with three different programs, it is likely they have the Type I topology (Figure A12). Members of this family were also discovered in *Verrucomicrobia* and *Euryarchaeota*. Of all the families tabulated within the Transport Classification Database (TCDB), the FUPA32 proteins are most closely related to members of the FUPA31 family. These observations emphasize the diversity of Family 32. It includes members from many bacterial kingdoms as well as the *Euryarchaeota*.

Parts of this Thesis are being prepared for publication. The Thesis author along with the committee chairman, Dr. Milton Saier Jr., will be a co-investigators and co-authors of this paper.

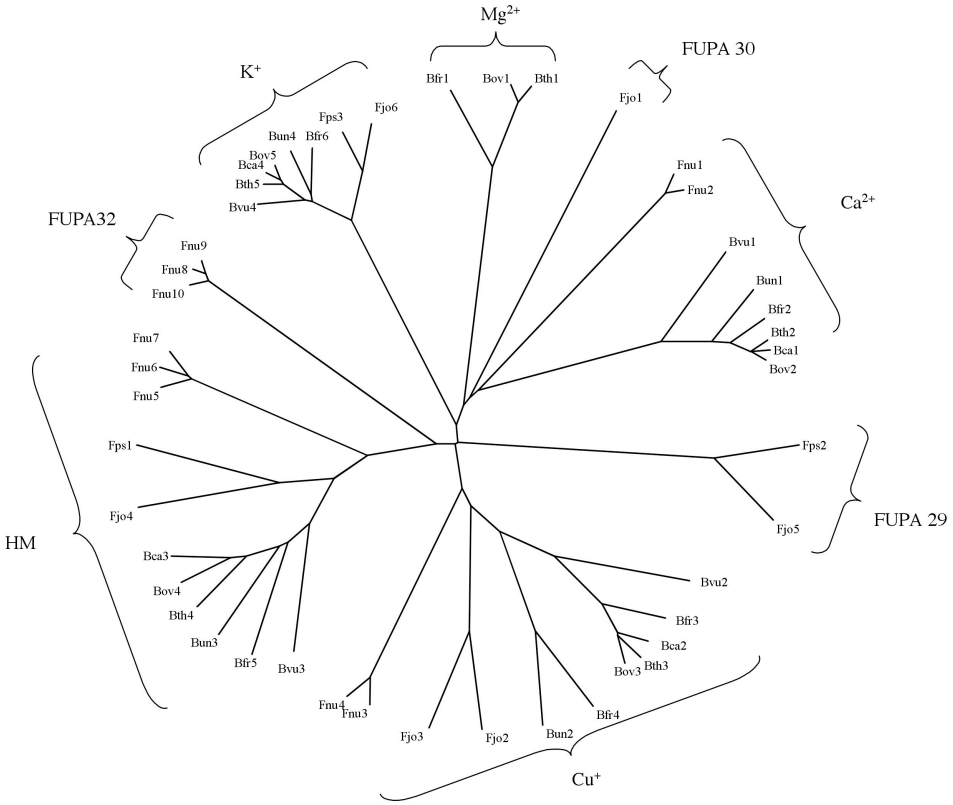

Figure A1: Phylogenetic tree made by ClustalX of 47 P-type ATPases from the eleven species of *Bacteroides, Flavobacterium* and *Fusobacterium*. Clustering is according to P-type ATPase Family. Families 2, 4-7 and FUPA29, 30 and 32 are represented.

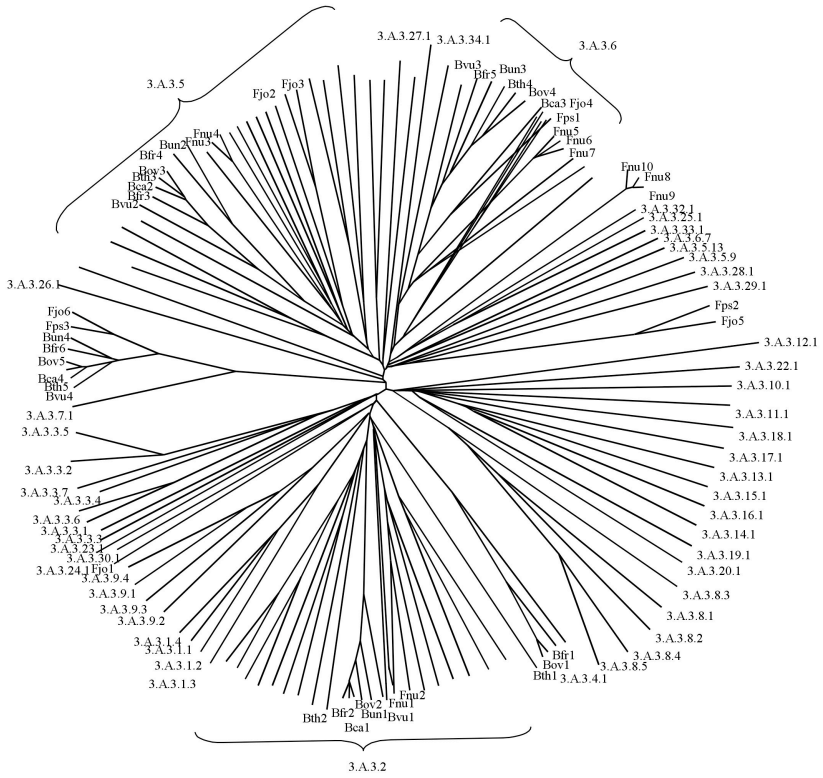

Figure A2: Phylogenetic tree made by ClustalX of 47 P-type ATPases from the eleven species of *Bacteroides, Flavobacterium* and *Fusobacterium* along with standards from the Transport Classification Database (TCDB). Clustering occurs according to P-type ATPase Family.

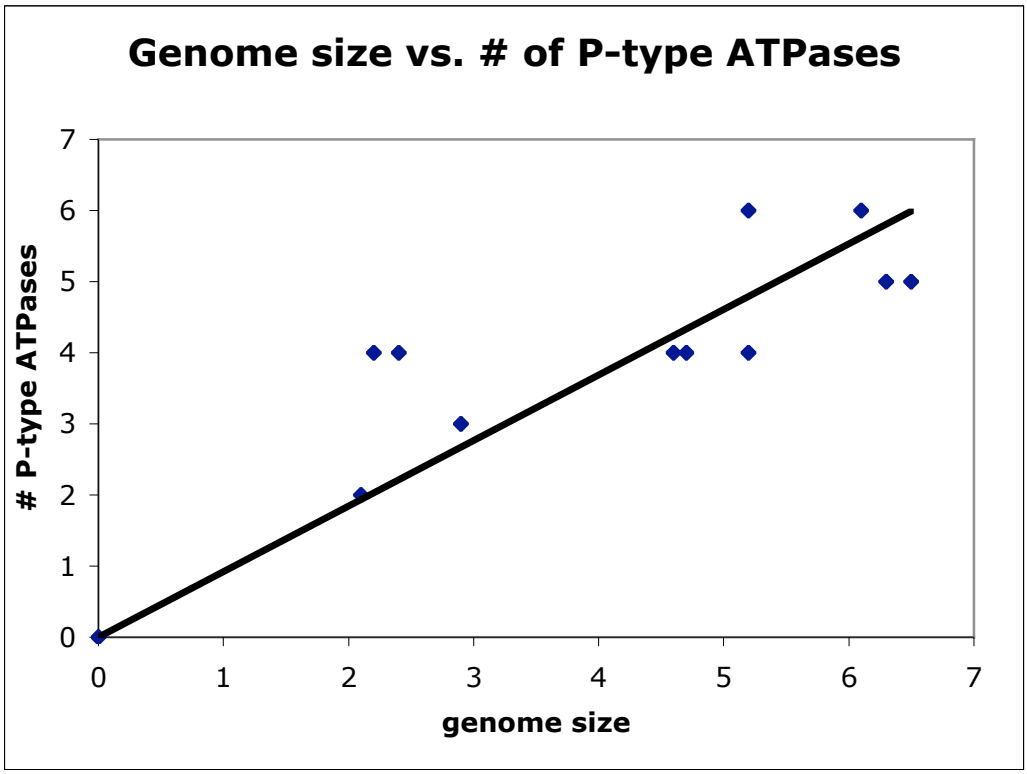

Figure A3: Genome sizes of the organisms used in this study vs. # of P-type ATPases found in that organism.

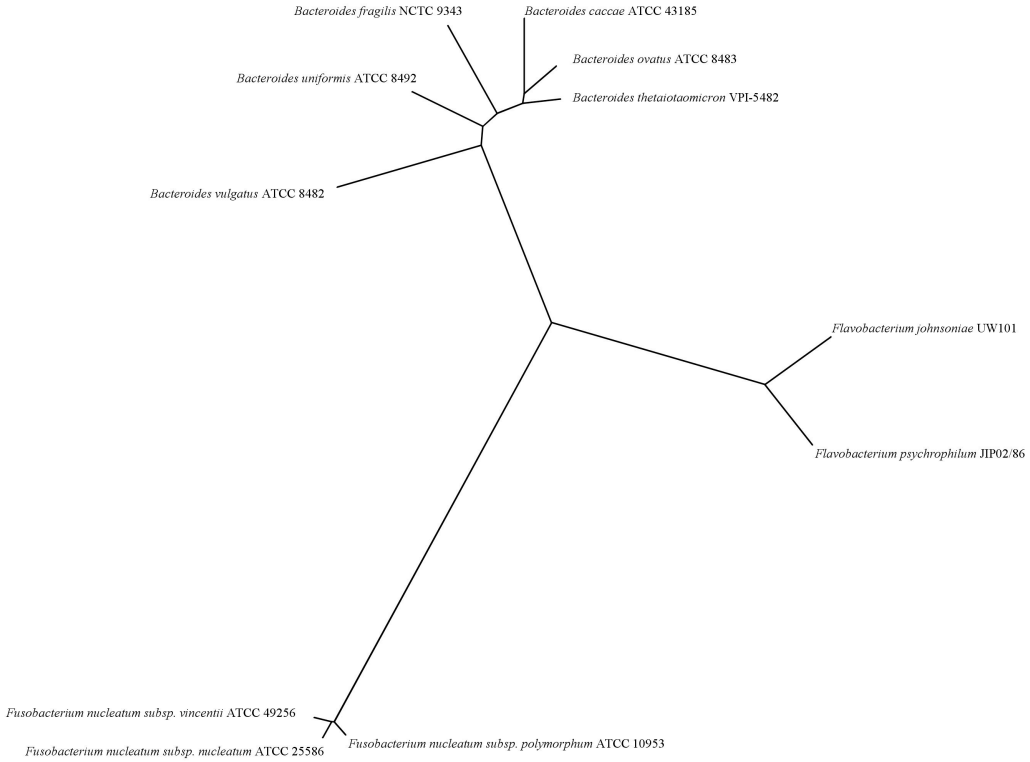

Figure A4: Tree of 16S rRNAs from all eleven organisms represented in this study. These organisms represent species from the three genuses *Bacteroides*, *Flavobacterium* and *Fusobacterium*.

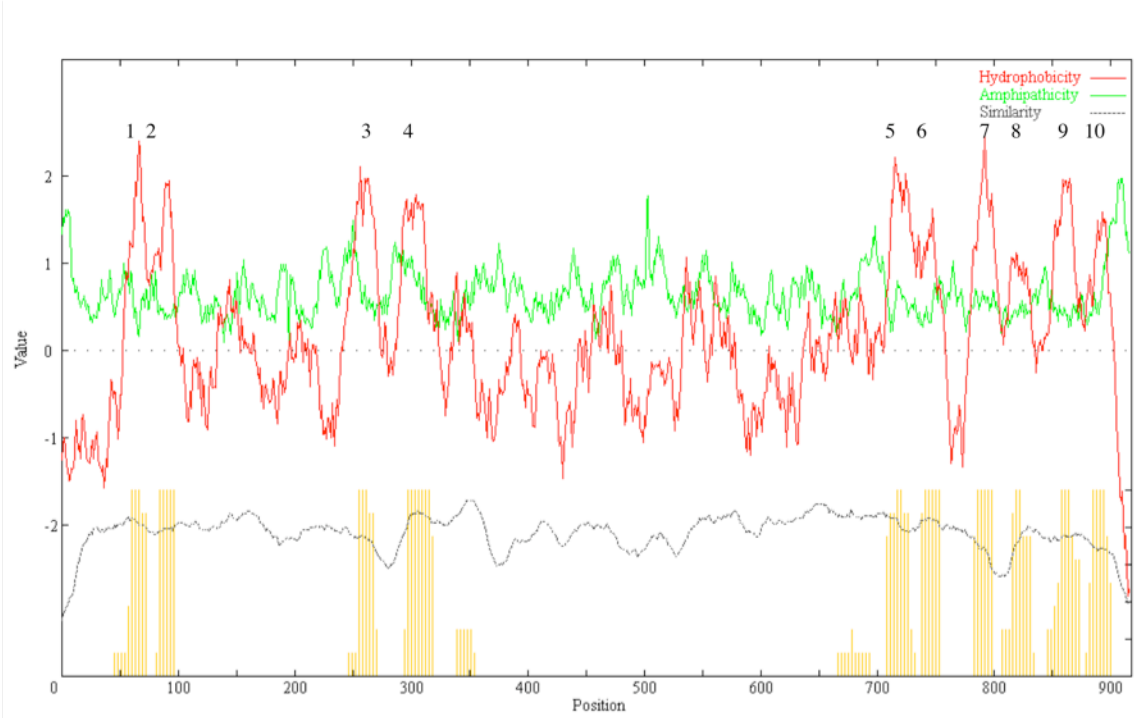

Figure A5: Topological analysis of the eight members of Family 2, Ca^{2+}, using the AveHAS program. Family 2 shows typical Type II topology.

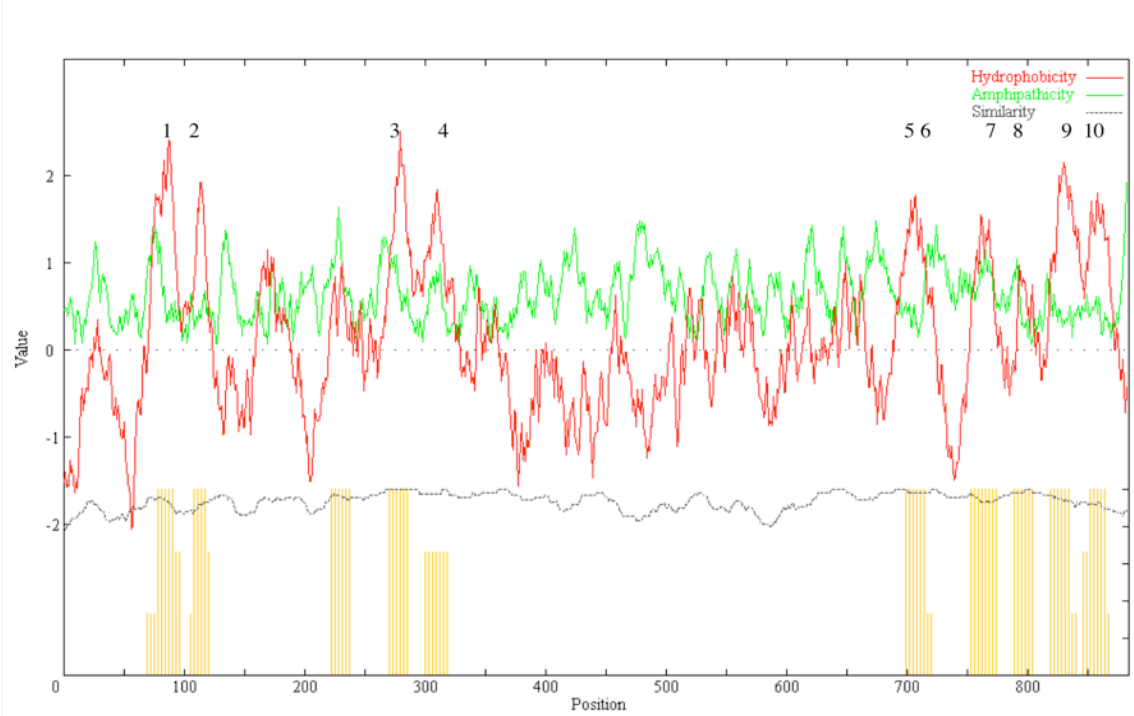

Figure A6: Topological analysis of the three members of Family 4, Mg^{2+}, using the AveHAS program. Family 4 shows typical Type II topology.

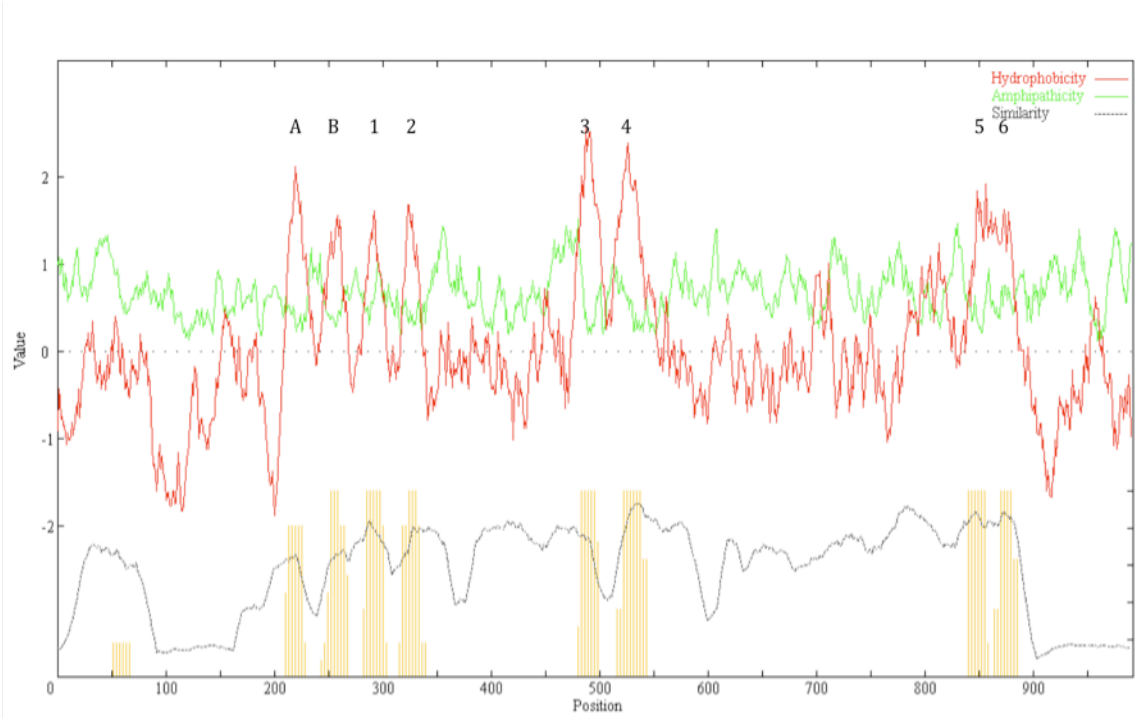

Figure A7: Topological analysis of the 11 members of Family 5, Cu^{+}, using the AveHAS program. Family 5 shows typical Type I topology.

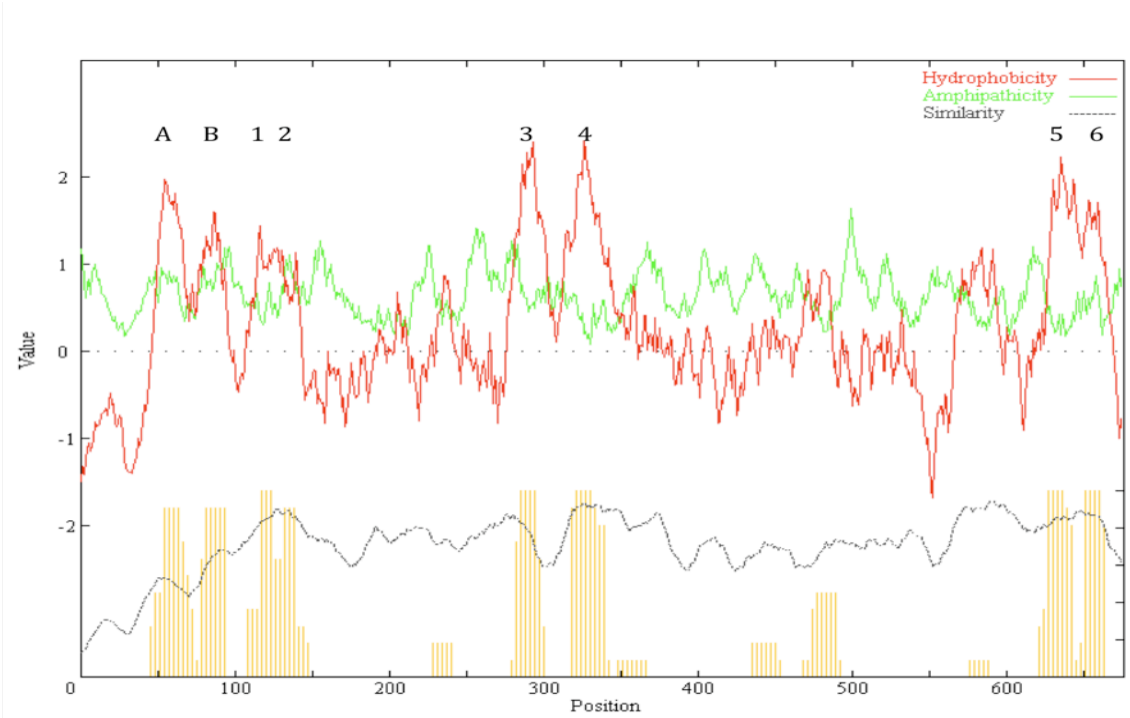

Figure A8: Topological analysis of the 11 members of Family 6, Heavy Metal, using the AveHAS program. Family 6 shows typical Type I topology.

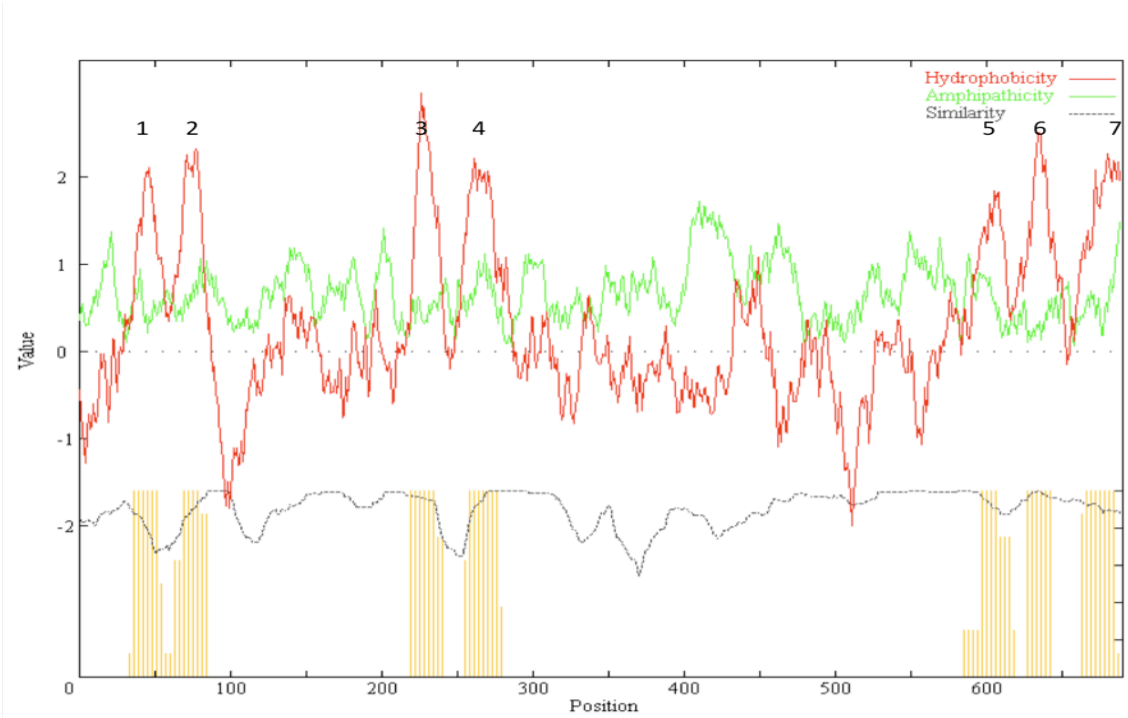

Figure A9: Topological analysis of the eight members of Family 7, K^+, using the AveHAS program. Family 7 shows typical Type III topology.

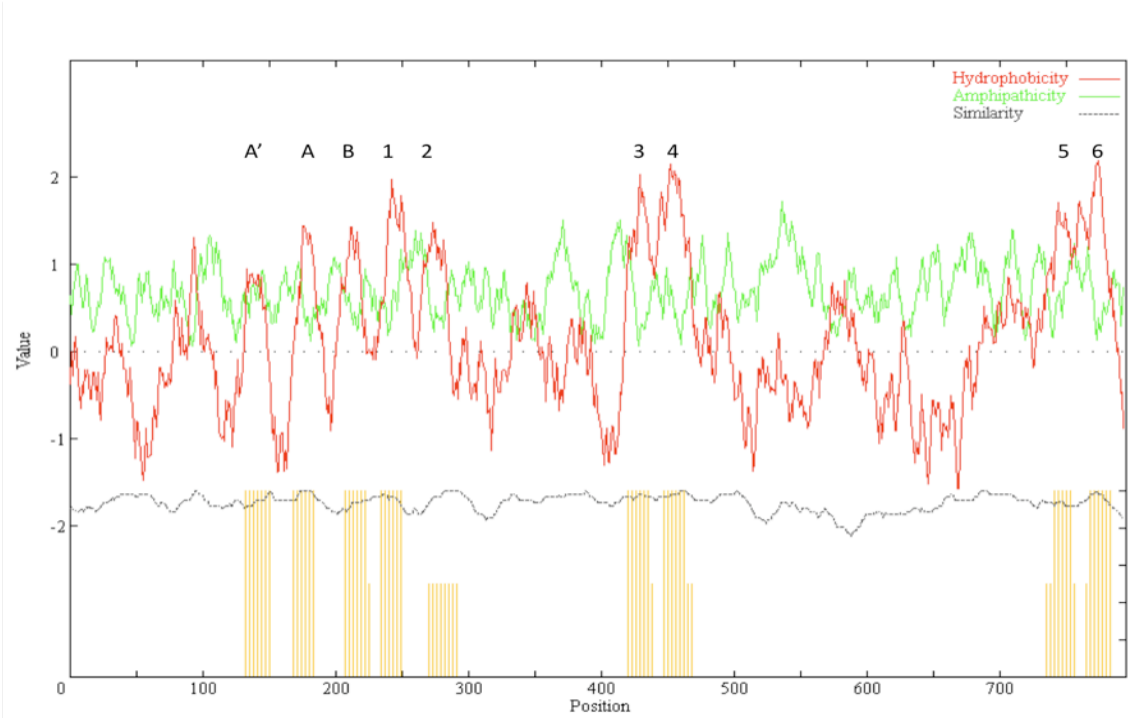

Figure A10: Topological analysis of the two members of Family 29, FUPA29, using the AveHAS program. Family 29 shows the novel Type VII topology, with A' marking the peak preceding the typical Type I topology.

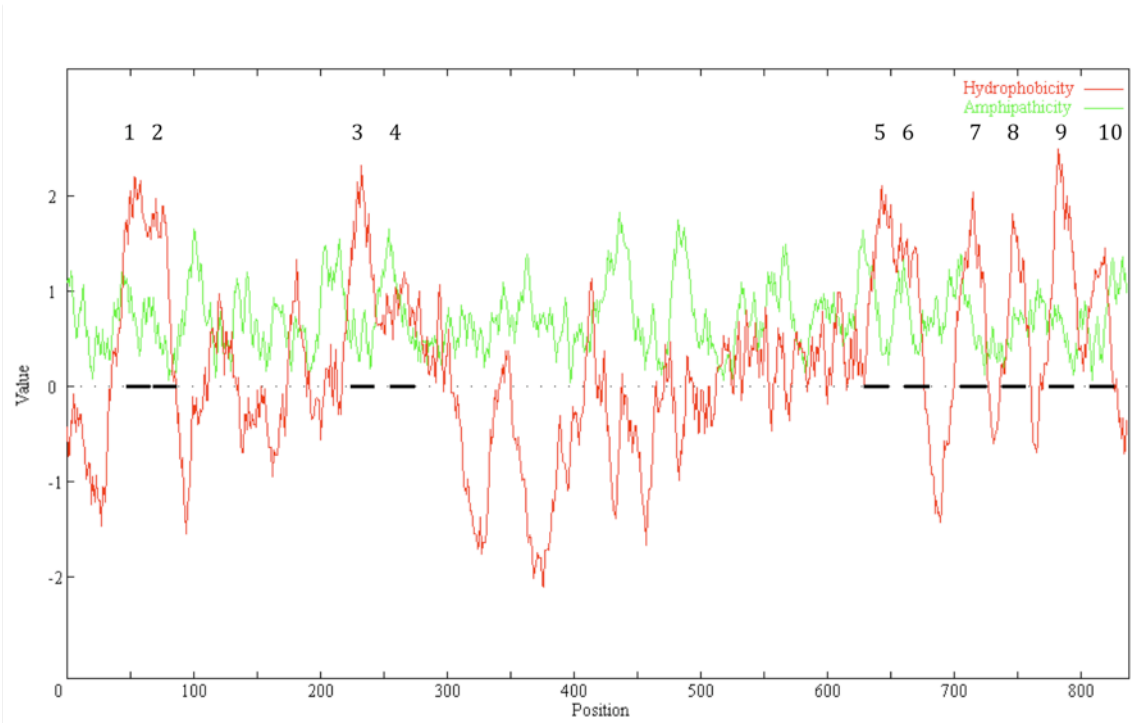

Figure A11: Topological analysis of the one member of Family 30, FUPA30, using the WHAT program. Family 30 shows the typical Type II topology.

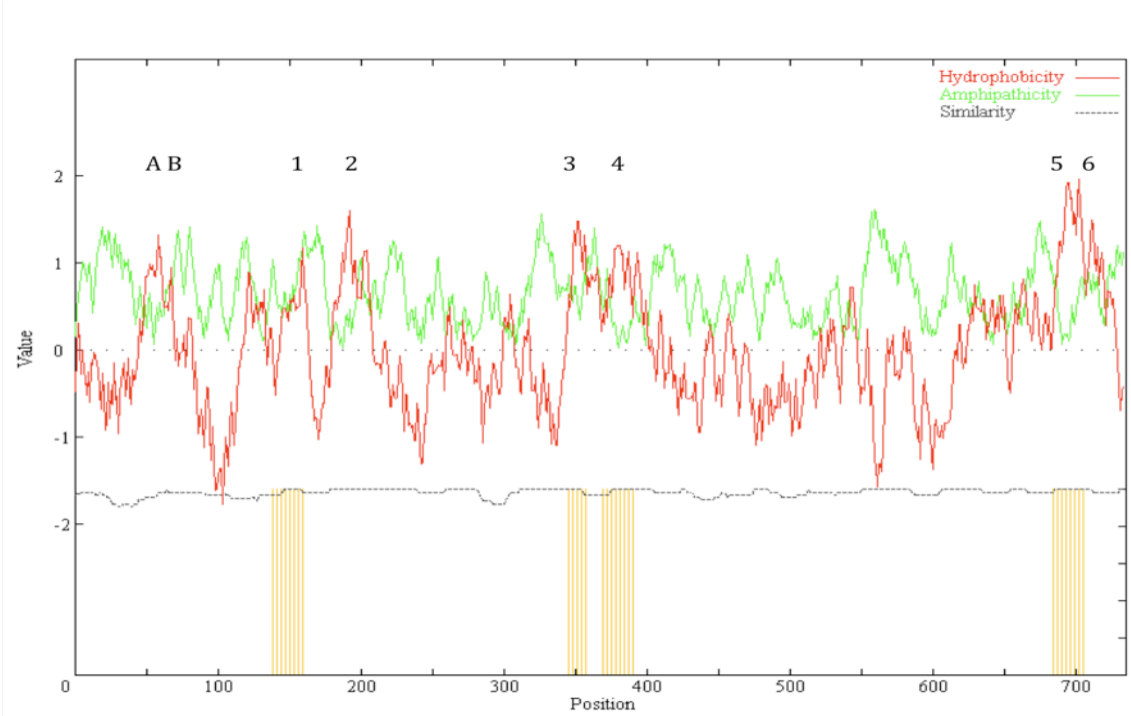

Figure A12: Topological analysis of the three members of Family 32, FUPA32, using the AveHAS program. Family 32 shows the typical Type I topology.

Table A1: Properties of eleven species of *Bacteroidetes, Flavobacteria*, and *Fusobacteria*. Genome size (column 2) and the estimated numbers of open reading frames (column 3) are also listed. Finally the numbers of P-type ATPases found in each organism are also presented in column 4.

Oraganisms	Genome Size (Mbp)	ORF Count	# of P-type ATPases
Bacteroides caccae ATCC 43185 (Bca*)*	4.6	3855	4
Bacteroides fragilis NCTC 9343 (Bfr)	5.2	4274	6
Bacteroides ovatus ATCC 8483 (Bov)	6.5	5536	5
Bacteroides thetaiotaomicron VPI-5482 (Bth)	6.3	4778	5
Bacteroides uniformis ATCC 8492 (Bun*)*	4.7	4663	4
Bacteroides vulgatus ATCC 8482 (Bvu)	5.2	4065	4
Flavobacterium johnsoniae UW101 (Fjo)	6.1	5017	6
Flavobacterium psychrophilum JIP02/86 (Fps)	2.9	2412	3
Fusobacterium nucleatum subsp. nucleatum ATCC 25586 (Fnun*)*	2.2	2067	4
Fusobacterium nucleatum subsp. polymorphum ATCC 10953 (Fnup)	2.4	2375	4
Fusobacterium nucleatum subsp. vincentii ATCC 49256 (Fnuv)	2.1	2250	2

Table A2: Number of P-type ATPases per family encoded within eleven *Bacteroides*, *Flavobacterium* and *Fusobacterium* genomes. A total of 47 P-type ATPases were identified. The number of P-type ATPases in families 2, 4-7, FUPA 29,30 and 32 are compiled in this table. The complete organismal names along with the corresponding abbreviations (in parenthesis) are provided in column 1.

Organisms	Family 2 (Ca^{2+})	Family 4 (Mg^{2+})	Family 5 (Cu^{+})	Family 6 (HM)	Family 7 (K^{+})	FUPA 29	FUPA 30	FUPA 32	Total
Bacteroides caccae ATCC 43185 (Bca)	1	0	1	1	1	0	0	0	4
Bacteroides fragilis NCTC 9343 (Bfr)	1	1	2	1	1	0	0	0	6
Bacteroides ovatus ATCC 8483 (Bov)	1	1	1	1	1	0	0	0	5
Bacteroides thetaiotaomicron VPI-5482 (Bth)	1	1	1	1	1	0	0	0	5
Bacteroides uniformis ATCC 8492 (Bun)	1	0	1	1	1	0	0	0	4
Bacteroides vulgatus ATCC 8482 (Bvu)	1	0	1	1	1	0	0	0	4
Flavobacterium johnsoniae UW101 (Fjo)	0	0	2	1	1	1	1	0	6
Flavobacterium psychrophilum JIP02/86 (Fps)	0	0	0	1	1	1	0	0	3
Fusobacterium nucleatum subsp. nucleatum ATCC 25586 (Fnun)	1	0	1	1	0	0	0	1	4
Fusobacterium nucleatum subsp. polymorphum ATCC 10953 (Fnup)	1	0	1	1	0	0	0	1	4
Fusobacterium nucleatum subsp. vincentii ATCC 49256 (Fnuv)	0	0	0	1	0	0	0	1	2
Total	8	3	11	11	8	2	1	3	47

Table A3: 47 P-type ATPases from the genuses *Bacteroides, Flavobacterium* and *Fusobacterium* included in this study, organized by family. In each family, proteins are shown in alphabetical order. Protein size in amino acids (aas), average size ± standard deviations (S.D.), and Genbank Index numbers are also reported. Phylum is also reported.

Abbreviation	Organism	Protein Size (# of aas)	Genbank Index # (GI #)	Phylum
Family 2 Ca^{2+}				
Bca1	*Bacteroides caccae* ATCC 43185	901	153806064	Bacteroidetes
Bfr2	*Bacteroides fragilis* YCH46	894	53715628	Bacteroidetes
Bov2	*Bacteroides ovatus* ATCC 8483	901	160883654	Bacteroidetes
Bth2	*Bacteroides thetaiotaomicron* VPI-5482	896	29347951	Bacteroidetes
Bun1	*Bacteroides uniformis* ATCC 8492	894	160891488	Bacteroidetes
Bvu1	*Bacteroides vulgatus* ATCC 8482	876	150003870	Bacteroidetes
Fnu1	*Fusobacterium nucleatum subsp. nucleatum* ATCC 25586	862	19704357	Fusobacteria
Fnu2	*Fusobacterium nucleatum subsp. polymorphum* ATCC 10953	862	167008204	Fusobacteria
Average protein size ± S.D.		886±17		
Family 4 Mg^{2+}				
Bfr1	*Bacteroides fragilis* YCH46	883	53711502	Bacteroidetes
Bov1	*Bacteroides ovatus* ATCC 8483	883	160882232	Bacteroidetes
Bth1	*Bacteroides thetaiotaomicron* VPI-5482	883	29346398	Bacteroidetes
Average protein size ± S.D.		883		
Family 5 Cu^{+}				
Bca2	*Bacteroides caccae* ATCC 43185	735	153806703	Bacteroidetes
Bfr3	*Bacteroides fragilis* NCTC 9343	736	60681570	Bacteroidetes
Bfr4	*Bacteroides fragilis* NCTC 9343	836	60683305	Bacteroidetes
Bov3	*Bacteroides ovatus* ATCC 8483	736	160886041	Bacteroidetes
Bth3	*Bacteroides thetaiotaomicron* VPI-5482	738	29346501	Bacteroidetes
Bun2	*Bacteroides uniformis* ATCC 8492	840	160891311	Bacteroidetes
Bvu2	*Bacteroides vulgatus* ATCC 8482	739	150005539	Bacteroidetes
Fjo2	*Flavobacterium johnsoniae* UW101	837	146297956	Bacteroidetes
Fjo3	*Flavobacterium johnsoniae* UW101	845	146300819	Bacteroidetes
Fnu3	*Fusobacterium nucleatum subsp. nucleatum* ATCC 25586	769	19703590	Fusobacteria
Fnu4	*Fusobacterium nucleatum subsp. polymorphum* ATCC 10953	769	167006957	Fusobacteria
Average protein size ± S.D.		780±49		
Family 6 Heavy Metal				
Bca3	*Bacteroides caccae* ATCC 43185	650	153806079	Bacteroidetes
Bfr5	*Bacteroides fragilis* NCTC 9343	648	60680801	Bacteroidetes
Bov4	*Bacteroides ovatus* ATCC 8483	652	160883629	Bacteroidetes
Bth4	*Bacteroides thetaiotaomicron* VPI-5482	652	29347922	Bacteroidetes
Bun3	*Bacteroides uniformis* ATCC 8492	646	160892200	Bacteroidetes
Bvu3	*Bacteroides vulgatus* ATCC 8482	572	150005019	Bacteroidetes
Fjo4	*Flavobacterium johnsoniae* UW101	663	146301088	Bacteroidetes
Fps1	*Flavobacterium psychrophilum* JIP02/86	660	150024529	Bacteroidetes
Fnu7	*Fusobacterium nucleatum subsp. nucleatum* ATCC 25586	614	19703603	Fusobacteria
Fnu6	*Fusobacterium nucleatum subsp. polymorphum* ATCC 10953	614	167006945	Fusobacteria
Fnu5	*Fusobacterium nucleatum subsp. vincentii* ATCC 49256	614	34763530	Fusobacteria
Average protein size ± S.D.		635±28		
Family 7 K^{+}				
Bca4	*Bacteroides caccae* ATCC 43185	677	153808315	Bacteroidetes

Table A3: (Continued)

Abbreviation	Organism	Protein Size (# of aas)	Genbank Index # (GI #)	Phylum
Bfr6	*Bacteroides fragilis* YCH46	682	53711870	Bacteroidetes
Bov5	*Bacteroides ovatus* ATCC 8483	677	160884983	Bacteroidetes
Bth5	*Bacteroides thetaiotaomicron* VPI-5482	677	29347834	Bacteroidetes
Bun4	*Bacteroides uniformis* ATCC 8492	682	160887723	Bacteroidetes
Bvu4	*Bacteroides vulgatus* ATCC 8482	677	150005773	Bacteroidetes
Fjo6	*Flavobacterium johnsoniae* UW101	685	146299734	Bacteroidetes
Fps3	*Flavobacterium psychrophilum* JIP02/86	677	150025739	Bacteroidetes
Average protein size ± S.D.		679±3		
FUPA29				
Fps2	*Flavobacterium psychrophilum* JIP02/86	792	150025225	Bacteroidetes
Fjo5	*Flavobacterium johnsoniae* UW101	795	146300291	Bacteroidetes
Average protein size ± S.D.		793.5±2		
FUPA30				
Fjo1	*Flavobacterium johnsoniae* UW101	838	146302203	Bacteroidetes
Average protein size ± S.D.		838		
FUPA32				
Fnu8	*Fusobacterium nucleatum subsp. nucleatum* ATCC 25586	735	19704525	Fusobacteria
Fnu9	*Fusobacterium nucleatum subsp. polymorphum* ATCC 10953	735	167008397	Fusobacteria
Fnu10	*Fusobacterium nucleatum subsp. vincentii* ATCC 49256	735	34764203	Fusobacteria
Average protein size ± S.D.		735		

Table A4: Conserved motifs for the 8 *Bacteroides, Flavobacterium* and *Fusobacterium* P-type ATPases families. The nine consensus motifs along with the conserved motifs of these families are shown. Degree of conservation is as follows: * signifies an identity, : signifies a close similarity, . signifies a distant similarity, _ signifies a lack of similarity among the residues.

Consensus	**PGD**	**PAD**	**TGES**	**PEGL**	**DKTGTLT**	**KGAPE**	**DPPR**	**MVTGD**	**VAVTGDGVNDSPALKKADIGVAM**
Family 2 Ca^{2+}									
Seq. Motif	VGD	PAD	TGEP	PEGL	DKTGTLT	KGAPE	DPIR	IVTGD	VAVTGDGTNDAPALNHAQVGLSM
Conservation	_.*	***	***_	**.*	*******	***_:	**_*	::***	_*:****.******::*::*::*
Family 4 Mg^{2+}									
Seq. Motif	PGD	PAD	TGES	PEML	DKTGTLT	KGAVE	DPPK	ILSGD	VGFLGDGINDAGALRQSDIGISV
Conservation	_**	***	****	****	*******	*****	****	*****	*******:***************
Family 5 Cu^{+}									
Seq. Motif	VGD	PVD	SGEP	PCAL	DKTGTLT	KGSYQ	DPIK	MLTGD	VAMVGDGINDSQALAQADVGIAM
Conservation	_*:	.**	.**.	**:*	*****:*	_____	*_::	:****	***.******:_***_:*:.:*:
Family 6 HM									
Seq. Motif	IGE	PLD	TGES	PCAL	DKTGTLT	EAVID	DELK	ILSGD	VAFVGDGINDAPVLALSDVGIAM
Conservation	_._	_**	****	***:	*****:*	____:	*__*	:*:**	:.*****:**:**:*_:_.****
Family 7 K^{+}									
Seq. Motif	KGD	PAD	TGES	PTTI	DKTGTIT	VDLAD	DIIK	MVTGD	VAMMGDGTNDAPALAQANVGVAM
Conservation	*.*	.:*	****	****	*******	:_*_*	****	*****	**********:*****
FUPA29									
Seq. Motif	KGD	PVD	TGEA	PCAL	DKTGTIT	IKIGS	NQYR	ILSGD	VMMVGDGLNDAGALAQSNVGISI
Conservation	**:	***	****	***:	*******	*****	*.**	:****	**********:**********
FUPA30	Only	One	Protein	No	Consensus				
FUPA32									
Seq. Motif	KDD	SVD	TGEY	SCGI	DKTGTIT	ETKIG	DPPR	LLTGD	VIMIGDGVNDAPALSYANVGVAL
Conservation	***	***	****	****	*******	***:*	****	*****	**********

Chapter 2. Proteobacterial P-type ATPases

The fully sequenced genomes of 55 organisms within the Proteobacterial phylum were analyzed for P-type ATPases. All five classes of Proteobacteria were examined (α-, β-, γ-, and δ-Proteobacteria) in this study, with a total of 48 different genuses representing 55 distinct organisms. The total number of P-type ATPases identified in all of the species was 218.

The largest amount of enzymes studied for a particular species was 12 belonging to the organism *Methylococcus capsulatus str. Bath*, a γ-Proteobacteria. All other species reported a range of total ATPases, with the lowest amount being one enzyme (Table B1). The average number of P-type ATPases was 3 enzymes per species.

Familial representation in the Proteobacterial phylum is quite extensive. Of the functionally characterized P-type ATPases, Families 1-7 are represented while Families 8 and 9 show no representation. In addition, there are a number of functionally uncharacterized ATPases that were identified. The Proteobacterial phylum showed representative members of FUPA25 and 27-32.

Distribution of P-type ATPases in the 5 Proteobacterial classes.

Table B1 summarizes the findings with respect to the distribution of P-type ATPases within the Proteobacterial phylum. This table presents the numbers of organisms analyzed, the average genome sizes, and the predicted average numbers of genes per genome for each of the 5 classes. It also presents the total numbers of the P-type ATPases in each of these 5 classes.

First examining overall totals for the 55 Proteobacteria examined, the two dominant families are Family 5 (Copper, 32%) and Family 6 (Heavy Metal, 16%). Four families are represented to about the same extent: The uncharacterized Family 27 (FUPA27) includes 12% of the total P-type ATPases; Family 7 Kdp-type K^+ ATPases include 11%; Family 2 Ca^{2+} ATPases include 10% of the total; and Family 4 Mg^{2+} ATPases include 8% of the total. All other families represented (Families 1 and 3, and all remaining FUPA families) include between 0.5% and 1.8% of the total P-type ATPases.

A single homologue of Family 1 Na^+/K^+ ATPases is found each in the γ- and δ-Proteobacteria. However Ca^{2+} ATPases (Family 2) are lacking in only the ε-Proteobacteria examined. The distributions of these enzymes vary tremendously, however, depending on the organismal class. For example, 4% of all α-Proteobacterial P-type ATPases are of this family, but 12% of the β-Proteobacteria ATPases, 10% of the γ-Proteobacteria, and 27% of the δ- Proteobacteria enzymes belong to this family.

Family 3 H^+ (or Heavy Metal; Hao et al., 1999) ATPases are found only in δ-Proteobacteria, but Family 4 Mg^{2+} ATPases are found in α-, β-, and γ-Proteobacteria (but not δ- Proteobacteria or ε-Proteobacteria) in about equal percentages (8-11%).

Family 5, 6 and 7 are found in all 5 classes of Proteobacteria. Copper ATPases are found in α-, β-, γ-, δ-Proteobacteria (27-36%) but 62% of the ε-Proteobacteria ATPases are of this type. Similarly the three classes of Proteobacteria have similar percentages of Heavy Metal ATPases (12-18%) but δ-Proteobacteria have only 4% while ε-Proteobacteria have 23%. Unlike other Proteobacteria studied, the ε- Proteobacteria have

85% of their ATPases as members of Families 5 and 6. Family 7 proteins show similar percentages (8-12%) for all 5 classes of Proteobacteria (Table B1).

The functionally uncharacterized P-type ATPases (FUPA) are not detected in all five Proteobacterial classes. For example, FUPA25 is represented in the α-, γ-, δ-Proteobacteria but not in the β- and ε-Proteobacteria. By contrast FUPA27 is found in α-, β-, γ-Proteobacteria, but not in δ- or ε-Proteobacteria and FUPA30 is represented among α-, β-, and δ-Proteobacteria. FUPA32 is represented in all classes except the γ-Proteobacteria. This last observation is surprising considering that a total of 29 species of γ-Proteobacteria were examined. All other FUPA families are found in only one class of Proteobacteria. While two or three FUPA28 or 31 members are found only in γ-Proteobacteria, a single FUPA30 homologue is found in δ-Proteobacteria.

Variations of P-type ATPase Distributions in the Various Organisms

Variation of P-type ATPase representation exists within the constituent members of the five Proteobacterial classes in this study (Table B2). The sizes of the proteins according to family, and their Genbank Index numbers can be seen in Table B3.

Family 1 and 3 ATPases were only seen in 0 to 1 copies, with the overwhelming majority of species having no members. The majority of organisms examined lack Ca^{2+} ATPases. However the numbers of these enzymes can vary from 0 to 4. Only one organism, *Methylococcus capsulatus str. Bath*, a γ-Proteobacteria, has 4 paralogues of this family. No organism has three paralogues of Family 2, and few have two. These

enzymes are particularly prevalent among the δ-Proteobacteria.

The Mg^{2+} ATPases (Family 4) are maximally present in just one or two copies per organism. They are about equally distributed among the α-, β-, γ-Proteobacteria. The Copper ATPases (Family 5) are present in almost all organisms in copies ranging from 1 to 3 per organism. The only two organisms with three such enzymes are *Methylococcus capsulatus str. Bath*, a γ-Proteobacteria, and *Sinorhizobium meliloti* 1021, an α-Proteobacteria. The Heavy Metal ATPases (Family 6) are present in 0 to 2 copies per organism. The K^+ (Family 7) are either lacking or are present in a single copy within the various organisms examined with the sole exception of *Pseudomonas fluorescens* Pf-5, which has two paralogues.

FUPA25 members are always lacking or are present in single copy in all organisms, and similarly FUPA27 members usually show the same distribution pattern. However two closely related α-Proteobacteria, *Mesorhizobium loti* MAFF303099 and *Sinorhizobium meliloti* 1021, have two such proteins. Remaining FUPA families are never represented with more than one member per organism.

Size Variation Among Proteobacterial P-type ATPases

As summarized in Table B4, each of the families, or topological types show a distinctive size range. The four characterized Type II ATPases (Families 1, 2, 3, and 4) have sizes of approximately 895 residues on average. There is little size variation between these families (Table B4).

Functionally known Type I ATPases (Families 5 and 6) have substantially smaller sizes as compared to Type II ATPases. Their average size is 767 residues as is typical. Family 6, Heavy Metal ATPases, are shorter on average than Copper ATPases by about 60 residues. This is due to the smaller number of N-terminal Heavy Metal binding domains (Yamaguchi et al. 2007) in these proteins.

Type III ATPases (Kdp-type K^+) show little size variation and are smaller than any of the Type I or Type II ATPases (689 ± 30 residues).

FUPA families generally show unexpected size variation. FUPA25 and FUPA32 family members, most closely resembling Type I, are in general shorter than Type I ATPases by 80 to 90 residues. On the other hand, FUPA28 and FUPA31 are longer by 80 to 90 residues. Of these proteins, the two FUPA28 homologues both had C-terminal hydrophilic extensions that accounted for their size differences, and the FUPA31 proteins exhibited different extensions near their C-termini. The four FUPA30 proteins, most closely resembling Type II ATPases, are about 50 residues shorter than average than the Type II enzymes. They seem to be missing an N-terminal segment normally found in these enzymes. Type VII ATPases, FUPA27 and FUPA29, are of the typical Type I size even though they may differ in topology.

Phylogenetic Analysis of Proteobacterial P-type ATPases

Two of the 218 proteins included in our study proved to be aberrant. These two proteins (which are included in Table B3) are Sen5, which we believe to be truncated, and Mca10, which is much longer than expected. The phylogenetic tree, including the

remaining 216 proteins is shown in Figure B1. Regardless of organismal source, the proteins grouped phylogenetically according to familial type. Phylogenetic relationships of these Proteobacterial enzymes with standards from TCDB are shown in Figure B2.

Figure 3B shows the phylogenetic tree for the 16s rRNAs from the 48 genuses represented in Figure 7. As can be seen, all five classes of Proteobacterial rRNAs cluster distinctly with one exception, *Xanthomonas*, which clusters loosely with the β- rRNAs. It can also be seen that the γ-Proteobacteria and δ-Proteobacteria rRNAs are more diverse in sequence than the α-, β-, and ε-Proteobacteria rRNAs. Comparison of the protein tree (Figure B1) with the rRNA tree (Figure B3) suggests that there has been extensive horizontal transfer of genes encoding P-type ATPases between Proteobacterial classes.

Na^+, K^+ homologues

The two homologues of Family 1 cluster together as seen in Figure B1. These two proteins, Tcr1 and Pca5, branch distantly from all other Proteobacterial proteins but closest to Ca^{2+} ATPases (Family 2) and FUPA30 proteins. These proteins belong to the Type II topology.

Ca^{2+} homologues

22 members comprise the Ca^{2+} P-type ATPases. Out of all of the subclusters, none are extremely tight, though all the Family 2 homologues are located next to one another (Figure B1). The proteins are loosely associated with the Family 1 Na^+/K^+ homologues

and cluster near FUPA30 proteins as well. These proteins belong to the Type II topology. When looking at the putative calcium ATPases, there is little clustering according to class suggesting that horizontal gene transfer has taken place.

H^+ homologues

Only 2 H^+ homologues are seen in Figure B1, Dps2 and Gsu6, both δ-Proteobacteria. Both proteins do not branch closely with each other. These proteins seem to have very loose clustering with Mg^{2+} homologues. These proteins belong to the Type II topology.

Mg^{2+}homologues

Family 4 enzymes consist of 18 proteins. Loose clustering occurs with the H^+ ATPases. Proteins from each class are scattered throughout the cluster of Mg^{2+} ATPases showing that there are not any orthologous proteins and horizontal gene transfer has occurred. These proteins belong to the Type II topology.

Cu^+ and Heavy Metal homologues

Topological Type I Copper (Family 5) and Heavy Metal (Family 6) homologues comprise the largest two families in this study. Copper ATPases have the most enzymes, 70, which all cluster together and span multiple subclusters according to Figure B1.

Members of this family cluster loosely near the single FUPA29 homologue and the FUPA27 proteins.

Heavy Metal P-type ATPases consist of 35 different homologues. Family 6 proteins clusters loosely with the FUPA25 enzymes. Though Family 5 and Family 6 enzymes are of the same topological type (Type II), they clearly segregate from each other as seen in Figure B1.

The Copper ATPases show the same intermixing of Proteobacterial classes although one of the γ-Proteobacteria clusters appears to lack members from the other classes. Nevertheless, even within this cluster of 15 proteins, it is apparent that they cannot all be orthologous (Figures B1 and B3). In the Heavy Metal ATPase Family, limited segregation is reflected by the fact that the β-Proteobacteria proteins cluster together as do the ε-Proteobacteria proteins. However, otherwise one sees extensive intermixing. This large amount of intermixing between classes of the Type I proteins suggests horizontal gene transfer between members.

Kdp-type K^+ homologues

Family 7 (topological Type III) proteins consist of 24 homologues in a tight cluster. The Kdp-type K^+ proteins do not seem to cluster near other families (Figure B1). In the Family 7 ATPases, a greater degree of segregation according to organismal class is observed. Proteins from each class cluster separately, except that γ-Proteobacteria proteins fall into four clusters separated by proteins from the other classes. These proteins though are not likely orthologous (Figures B1 and B3).

Functionally uncharacterized families 25, 27-32 (FUPA25, 27-32)

Four enzymes make up Family 25. These proteins branch loosely with each other and cluster loosely with a subcluster of Family 7 homologues.

FUPA27 proteins make up the largest family of uncharacterized enzymes within this study. 26 proteins compromise this family. These proteins cluster loosely with homologues of FUPA28 and Family 5, Copper ATPases. Two proteins, Lpn8 and Lpn9, make up the FUPA28 ATPases. FUPA29 consists of a single homologue Bba3. This protein clusters loosely between FUPA27 and Family 5 proteins (Figure B1). Families 27, 28 and 29 all cluster loosely with the Copper ATPases, more distantly to the Heavy Metal ATPases (Figure B1).

FUPA30 has 4 homologues, clustering near Family 1 and Family 4 proteins. FUPA31 has two homologues, Mca11 and Mca12, while FUPA32 has four homologues. FUPA31 and FUPA32 homologues are about equidistant between the Heavy Metal and Copper ATPases. These clustering patterns are in agreement with the topological analyses reported below.

A comparable intermixing of Proteobacterial classes exists within all of the functionally uncharacterized P-type ATPase families. This seems to suggest that there are not orthologous proteins, and horizontal gene transfer occurred.

Conserved Motifs

Motif 1 (PGD) shows variations for all of the families examined, both

characterized and uncharacterized. The central G residue is fully conserved in all families, with the sole exception being Family 6, Heavy Metal ATPases. In Family 6 the central G shows extremely poor conservation (Table B5). H^+ ATPases (Family 3) have the expected motif, PGD, fully conserved. In Families 31 and 32, the first residue is changed from P to the more hydrophobic residue A. The third residue of this motif, D, can be changed in only the Copper ATPases to E, though conservation within this family is poor.

The second motif, PAD, shows full conservation within Family 3, the H^+ ATPases, and Family 4, the Magnesium ATPases (Table B5). The last residue, D, shows full conservation for nearly all families analyzed. The sole exception being the Kdp-type K^+ (Family 7) ATPases, in which this residue is just semi conserved. In Family 7, an S can replace the P, though it is not well conserved (Table 5B). Serine, like proline, has a low propensity for α-helix formation and a high propensity for β-turn formation (Saier et al., 1987). The Type I Family 5 proteins can have the central A replaced by the hydrophobic amino acid V (Table B5).

Motif 3 (TGES) is fully conserved in Families 2, 3, 7 and 30 (Table B5). The characterized Type II Families 1 and 4 do not observe full conservation of this motif, though the central G and E are fully conserved. FUPA25 and 32 have full conservation of the first three residues of this motif, TGE, though variations can be seen in the final amino acid, serine. Families 5, 25, and 28 (Type I topologies) show the terminal residue S changed to P.

Motif 4, PEGL, shows full conservation in just Family 1 (Table B5). The first residue, P, has full conservation in all families studied except for Family 7 and FUPA32 homologues. Family 7 shows extremely poor conservation of this residue, and in Family 32, the proline (P) is replaced with a serine (S) that is fully conserved throughout that particular family's homologues (Table B5).

Motif 5, (DKTGTLT) the phosphorylation site, is almost fully conserved in all families, functionally characterized and uncharacterized. The first residue, the aspartyl phosphorylation site shows full conservation throughout all of the families examined (Table B5). FUPA31 proteins have the poorest conservation of this motif (though the aspartyl residue is still fully conserved) while FUPA32 proteins have the entire motif conserved. Variations in this motif occur in the Kdp-type K^+ (Type III) enzymes, where the motif is a fully conserved DKTGTIT. FUPA25 proteins have the second residue, lysine (K), changed to the hydrophobic amino acid leucine (L).

Motif 6 (KGAPE) shows extremely poor conservation throughout all of the families examined within this study (Table B5). The full motif is found only in Ca^{2+} ATPases, with the last three residues being poorly conserved. Family 3, H^+ ATPases, have a fully conserved KGAPQ, with the hydrophilic amino acid glutamine (Q) in place of the hydrophilic glutamic acid (E) (Table B5). Families 5, 6 and uncharacterized Family 27 do not exhibit good conservation between their respective members at this position.

Motif 7, DPPR, is only fully conserved in one family (Table 5B). The functionally characterized H^+ ATPases, Family 3, have this motif fully conserved. Functionally

uncharacterized Family 27 shows full conservation at this motif position, although instead of DPPR, the conserved motif for this family is DPLR. The aspartyl residue at the beginning of this motif is well conserved in all families except the Type I Copper ATPases and FUPA31 enzymes (Table B5).

Full conservation of the expected sequence of motif 8 (MVTGD) is not seen in any of the families analyzed due to subtle changes in amino acids. Type I Families 1-3 and FUPA30 homologues show MITGD with full conservation instead of MVTGD. The hydrophobic V is replaced with an amino acid of similar hydrophobicity, I (Table B5). In every family, the central residue is an aliphatic hydroxyl amino acid, T or S, except for Family 31 where this residue is replaced by the aliphatic hydrophobic amino acid I. The last two residues of this motif (GD) are fully conserved in all families examined except for FUPA28 and FUPA31 homologues.

Motif 9, VAVTGDGVNDSPALKKADIGVAM, the hinge motif is not very well conserved in most of the families analyzed (Table B5). The internal sequence GDG-N is well conserved in most of the families studied. Characterized families 2-7 and uncharacterized families 25, 27, 30 and 32 all show a good degree of conservation in this internal stretch as is typically for most P-type ATPases (Table B5). Other residues within the hinge motif exhibit varying degrees of conservation.

Topological Analyses

Functionally characterized Families 1-4 all exhibit the typical Type II topology as expected, with a few exceptions (Figures B4-B7). The Noc1 Ca^{2+} ATPase homologue

exhibits a unique 200 residue extension (Figure B5). This extension includes a single putative transmembrane segment following the last expected TMS. Assuming this to be TMS 10, the C-terminal portion of this protein should be on the outside. Through BLAST searches of this particular extension, it was determined that the C-terminal domain is a LysM peptidoglycan binding domain (Buist et al, 2008). This finding confirms its extracellular location and provides convincing evidence that peak 10 is truly a transmembrane spanner (Figure B5). This finding is thought to be the first report of a P-type ATPase that contains a LysM domain. Another putative Ca^{2+} ATPase, Mca2, proved to have a 140 residue N-terminal domain that was found at the N-termini of many other proteins. These include P-type ATPases, phosphoesterases and ABC transporters. The NCBI Conserved Domain Database did not recognize this sequence as a conserved domain. Thus it may be an unclassified domain of unknown function. FUPA30 homologues also show Type II topology (Figure B15).

In addition, topological analysis of the 70 Copper homologues showed hydrophilic extensions at the N-terminal (Figure B8). These proteins, after BLAST against TCDB known families were found to be part of the MerP domain, a mercury transport in bacteria (Yamaguchi et al., 2006). In addition, NCBI BLAST searches showed that these domains were present in other bacterial organisms as well. Of the 70 Copper homologues, the MerP domain existed in copies of 0-2 at the N-terminal.

The AveHAS plots for families 5 and 6 reveal only a single Type I topology characteristic of these families (Figure B8 and Figure B9). However both families contain extensive N-terminal domains of up to 300 residues. This is the region where the heavy

metal binding domains are often present in multiple numbers (up to six copies) are to be found (Yamaguchi et al., 2007). As expected, the Kdp-type K^+ ATPases show the uniform 7 TMS Type III topology, with no apparent exceptions (Figure B10).

Of the six remaining FUPA families (FUPA25, 27-29, 31, and 32) all exhibit the basic characteristics of the Type I topology (Figures B12-B14, B16 and B17). However, only Families 28 and 31 exhibit this topology unequivocally (Figures B13 and Figure B16). Family 27 and 29 proteins exhibit this basic topology but they possibly also have one or two TMSs preceding the first of these established TMSs (TMS A), noted as peak A' in Figures B12 and B14. This is referred to as topology as Type VII, a novel topological type. The proteins in Family 32 may lack peaks A and B and have the most ambiguous topological type of the families examined (Figure B17). Some members of Families 29 and 32 (e.g., Bba4 and Wsu4, respectively) appear to exhibit one or more N-terminal heavy metal binding domains, present in all copper and heavy metal ATPases (Yamaguchi et al., 2007). This observation suggests that these ATPases may similarly transport heavy metal ions (Figure B14 and Figure B17).

Parts of this Thesis are being prepared for publication. The Thesis author along with the committee chairman, Dr. Milton Saier Jr., will be a co-investigators and co-authors of this paper.

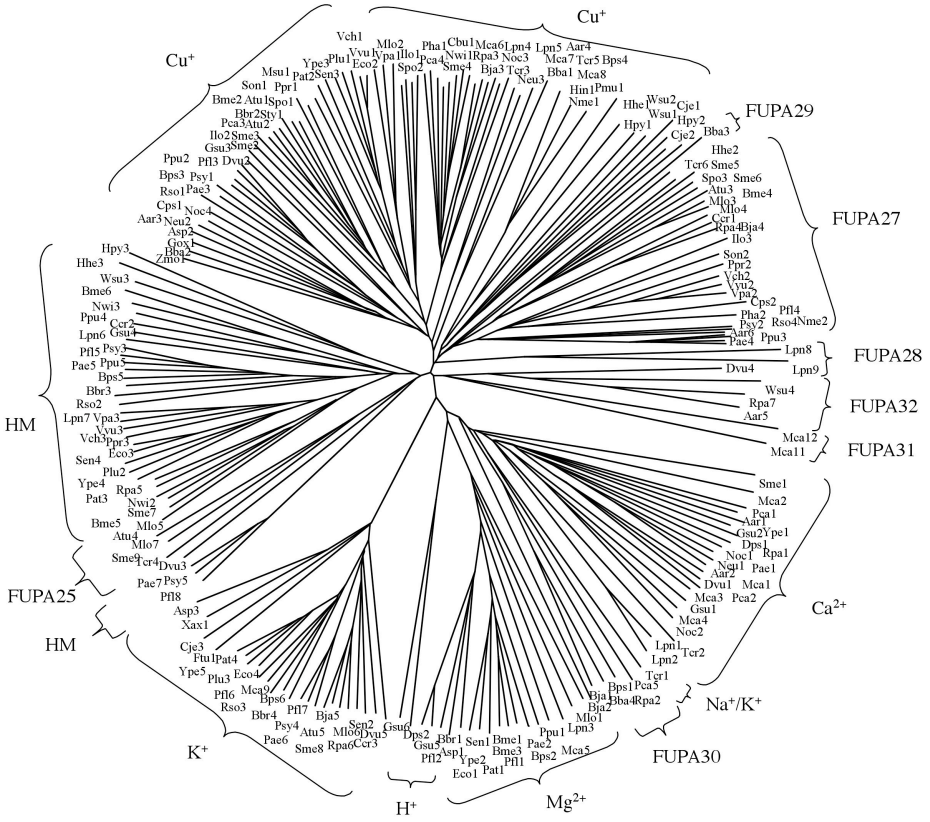

Figure B1: Phylogenetic tree made by ClustalX of 216 P-type ATPases from the Proteobacterial phylum. Clustering is according to P-type ATPase Family. Families 1-7 and FUPA25, 27-32 are represented. Tree excludes aberrant proteins Sen5 (truncated) and Mca10 (longer than expected).

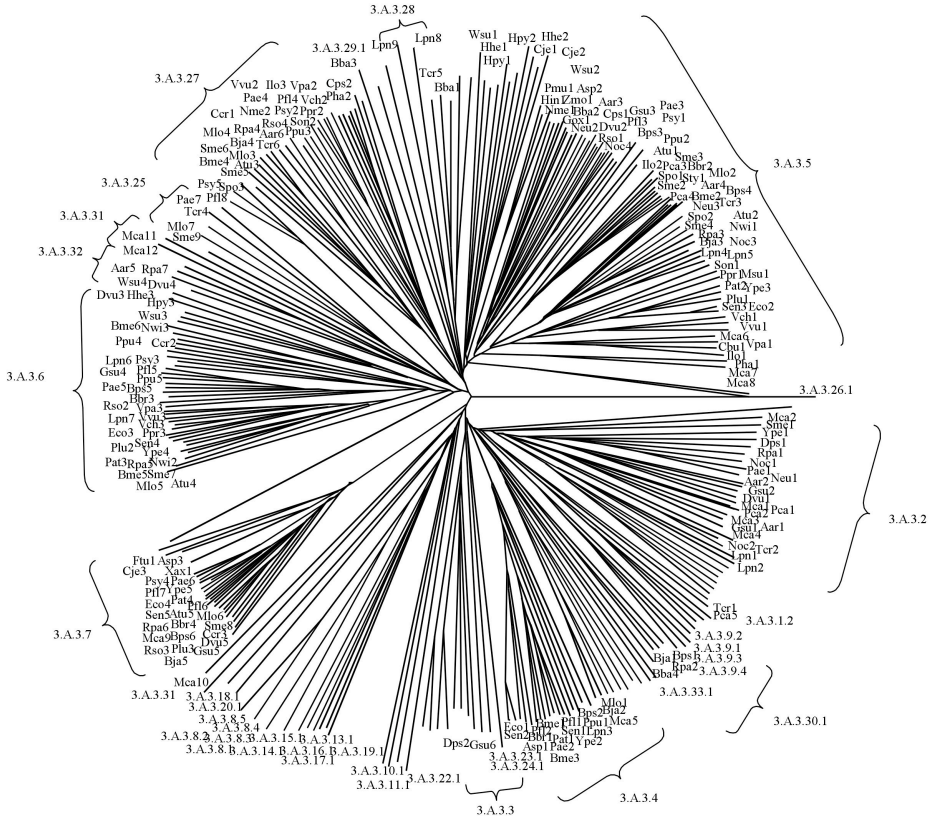

Figure B2: Phylogenetic tree made by ClustalX of 218 P-type ATPases from the Proteobacterial phylum along with standards from the Transport Classification Database (TCDB). Clustering occurs according to P-type ATPase Family.

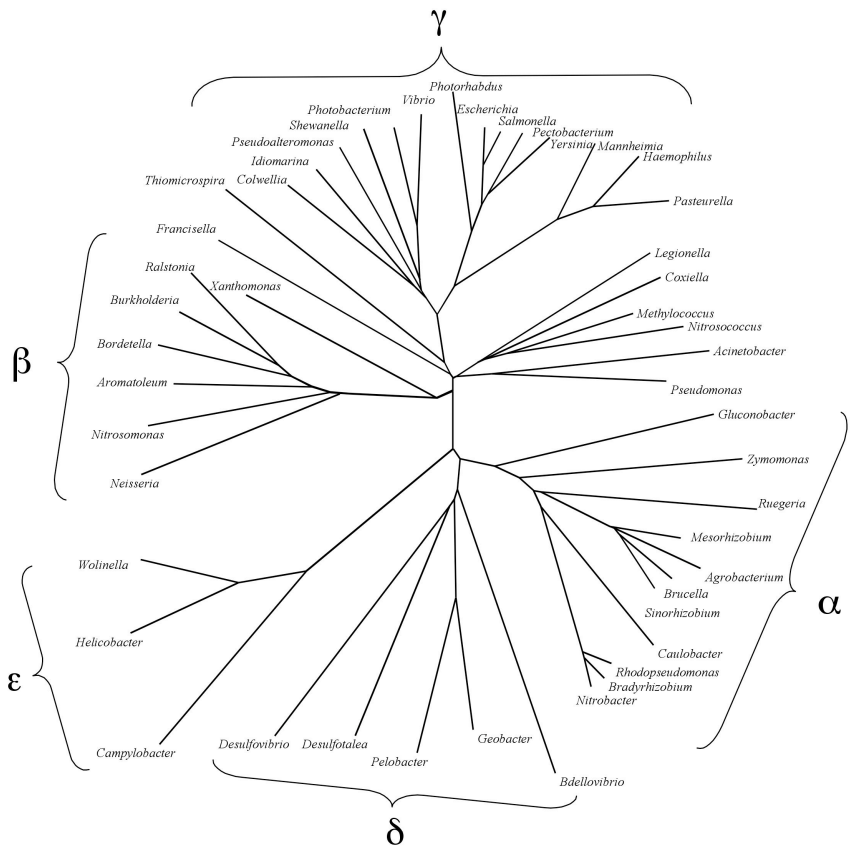

Figure B3: Tree of 16S rRNAs from all 48 genuses of Proteobacteria represented in this study. These 48 genuses represent 55 organisms and 218 proteins. The 48 Proteobacterial genuses separate according to the five Proteobacterial classes, α, β, γ, δ, ε.

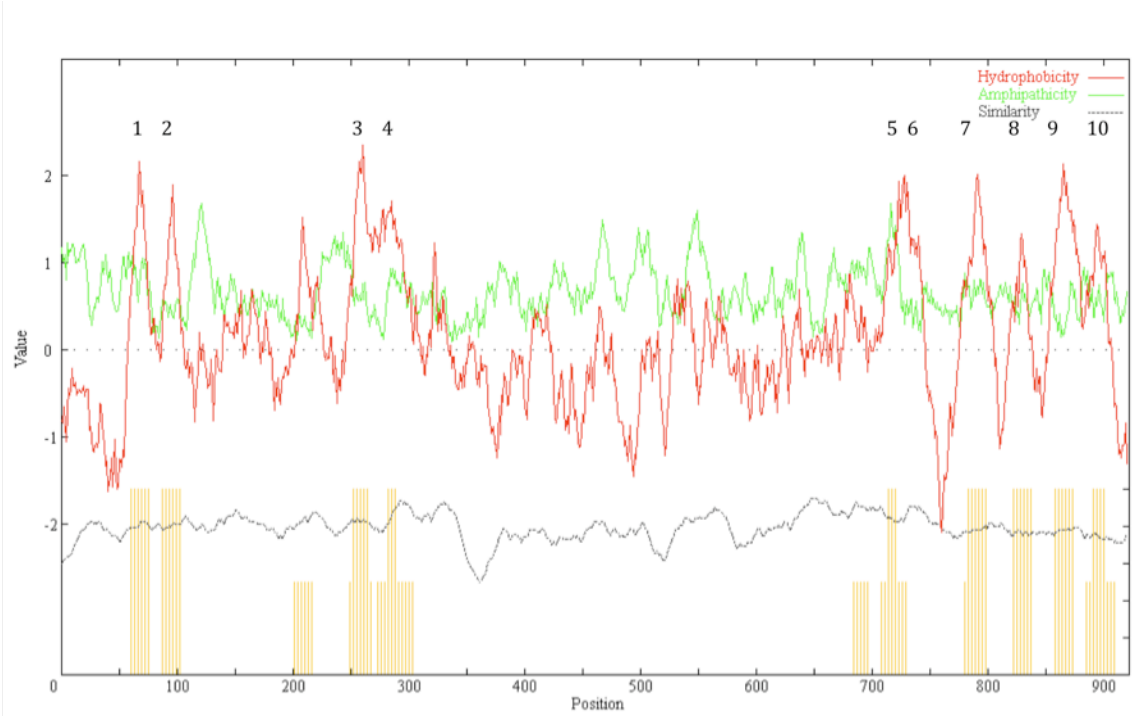

Figure B4: Topological analysis of the two members of Family 1, Na^+/K^+, using the AveHAS program. Family 1 shows typical Type II topology.

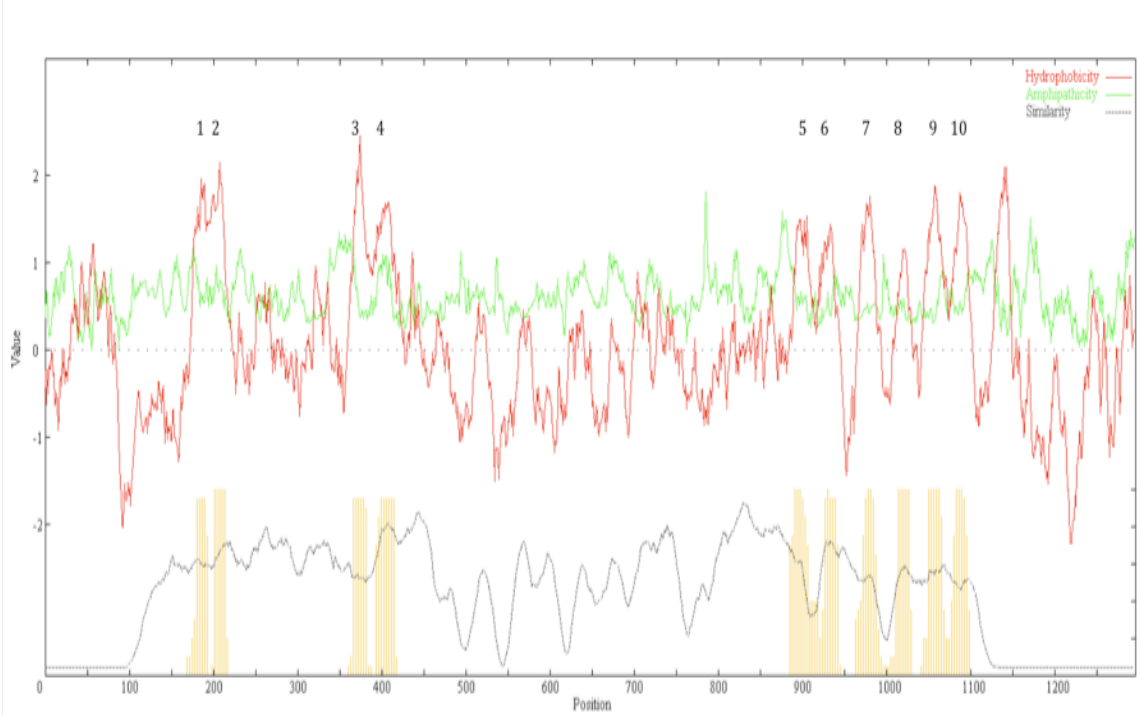

Figure B5: Topological analysis of the 22 members of Family 2, Ca^{2+}, using the AveHAS program. Family 2 shows typical Type II topology.

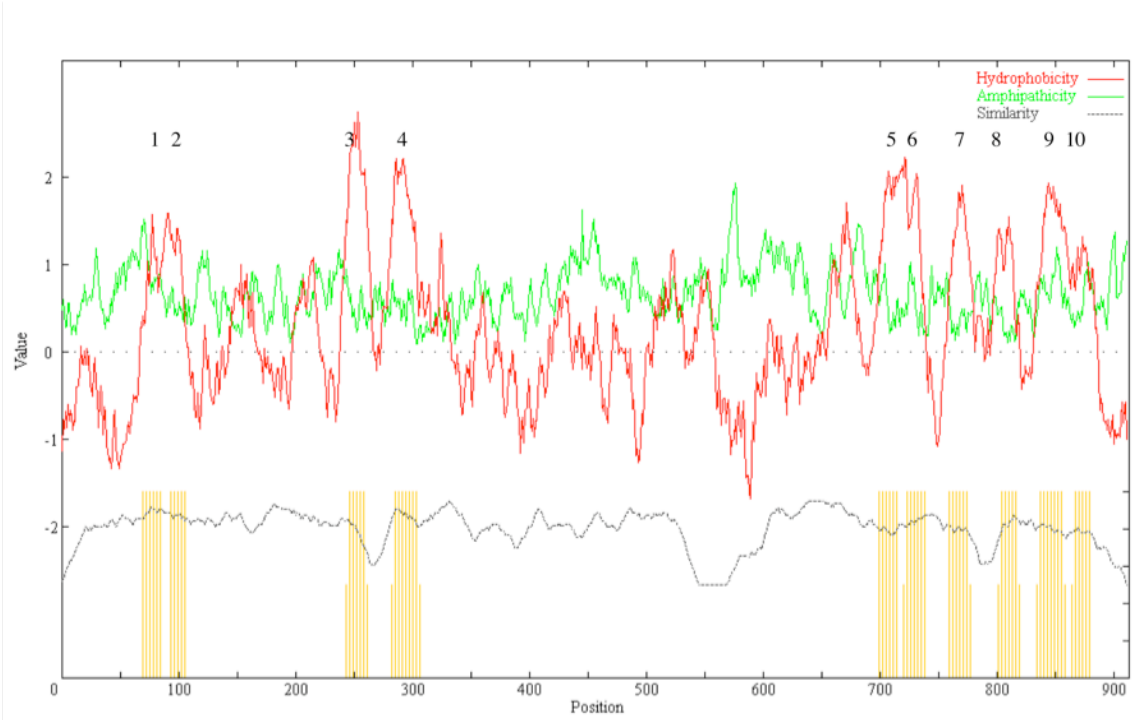

Figure B6: Topological analysis of the two members of Family 3, H^+, using the AveHAS program. Family 3 shows typical Type II topology.

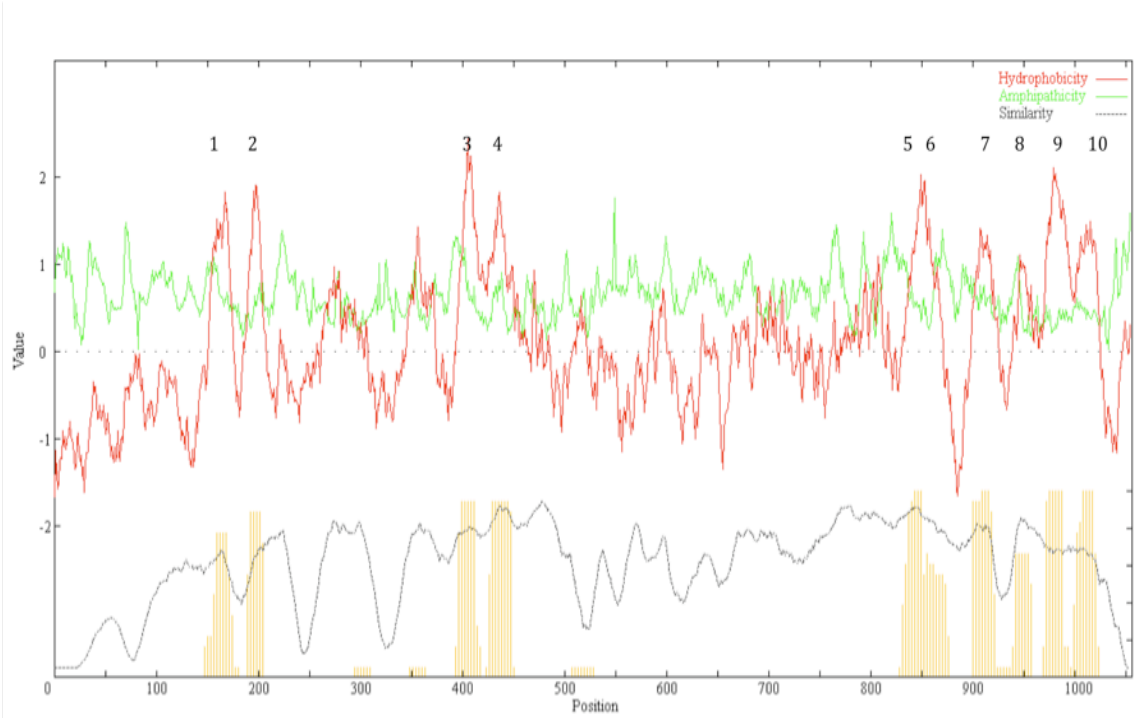

Figure B7: Topological analysis of the 18 members of Family 4, Mg^{2+}, using the AveHAS program. Family 4 shows typical Type II topology.

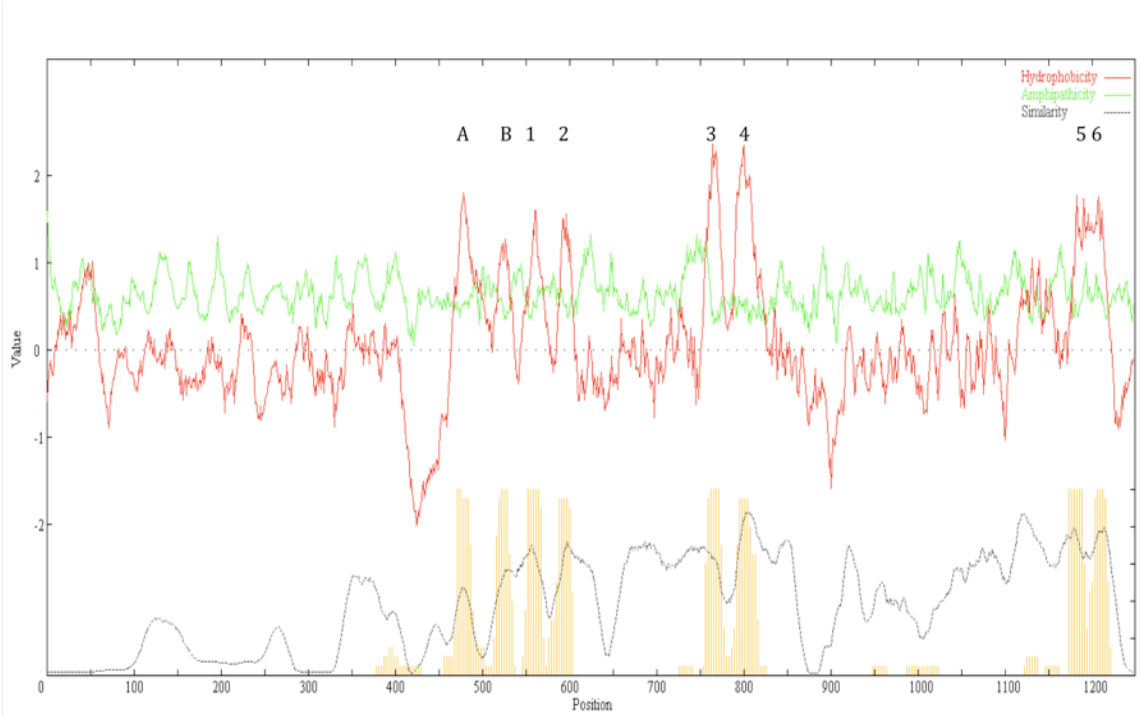

Figure B8: Topological analysis of the 70 members of Family 5, Cu^{+}, using the AveHAS program. Family 5 shows typical Type I topology.

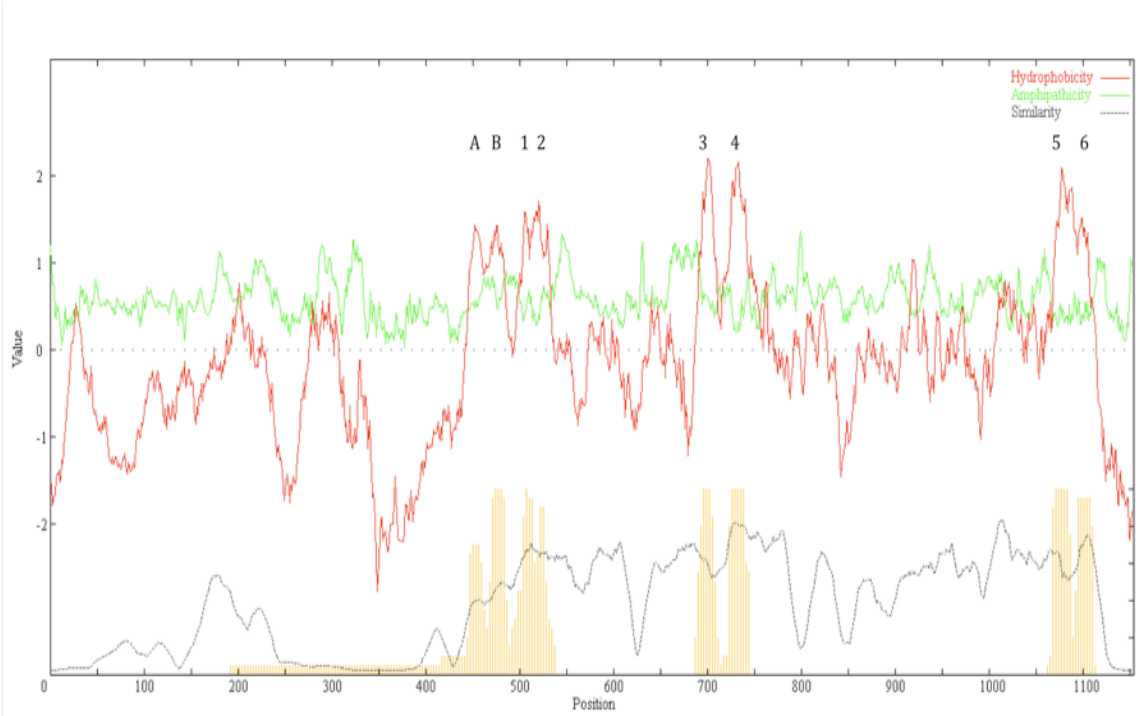

Figure B9: Topological analysis of the 35 members of Family 6, Heavy Metal, using the AveHAS program. Family 6 shows typical Type I topology.

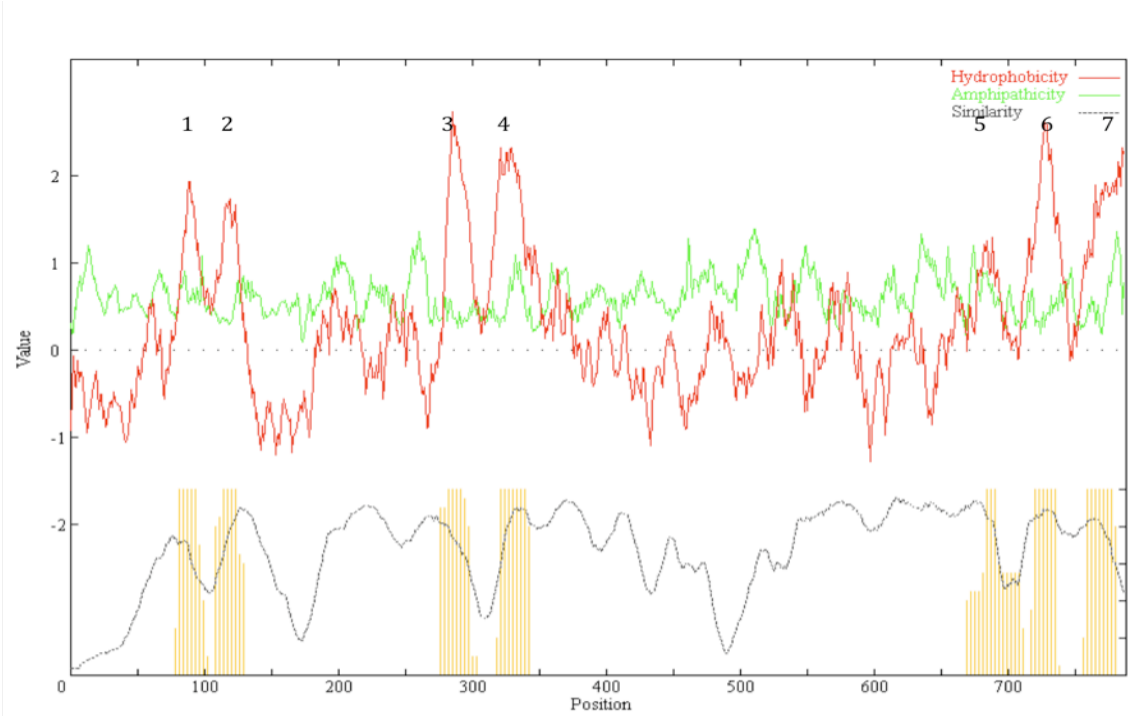

Figure B10: Topological analysis of the 25 members of Family 7, K^+, using the AveHAS program. Family 7 shows typical Type III topology.

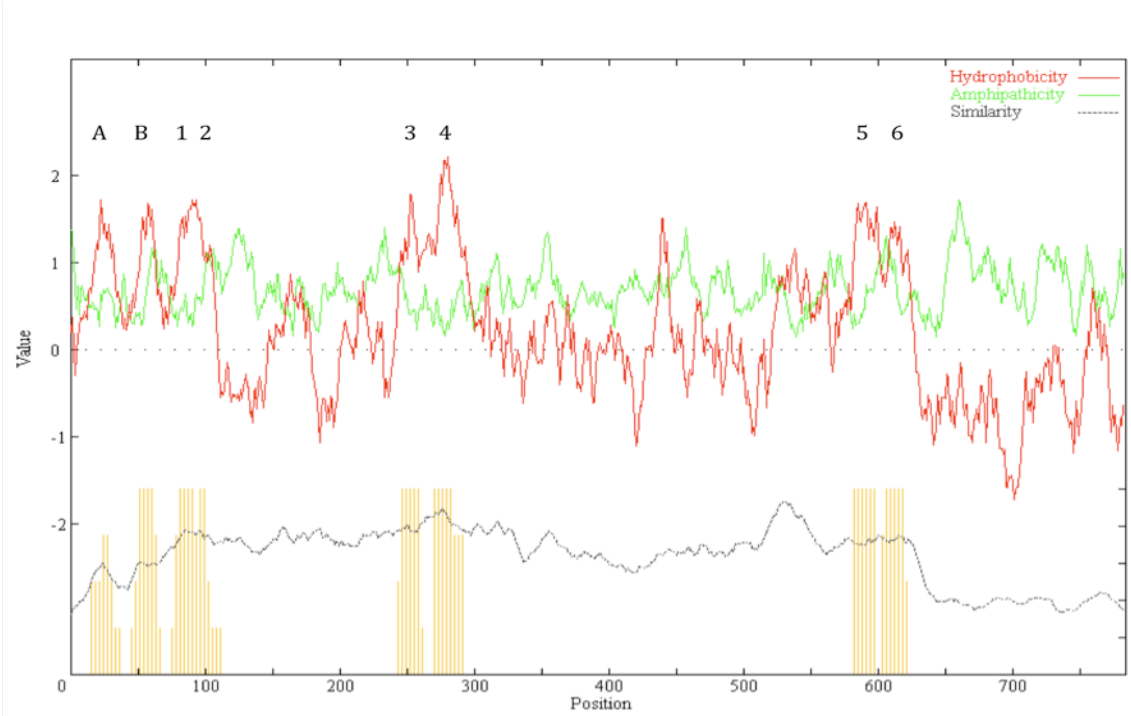

Figure B11: Topological analysis of the four members of Family 25, FUPA25, using the AveHAS program. Family 25 shows the typical Type I topology.

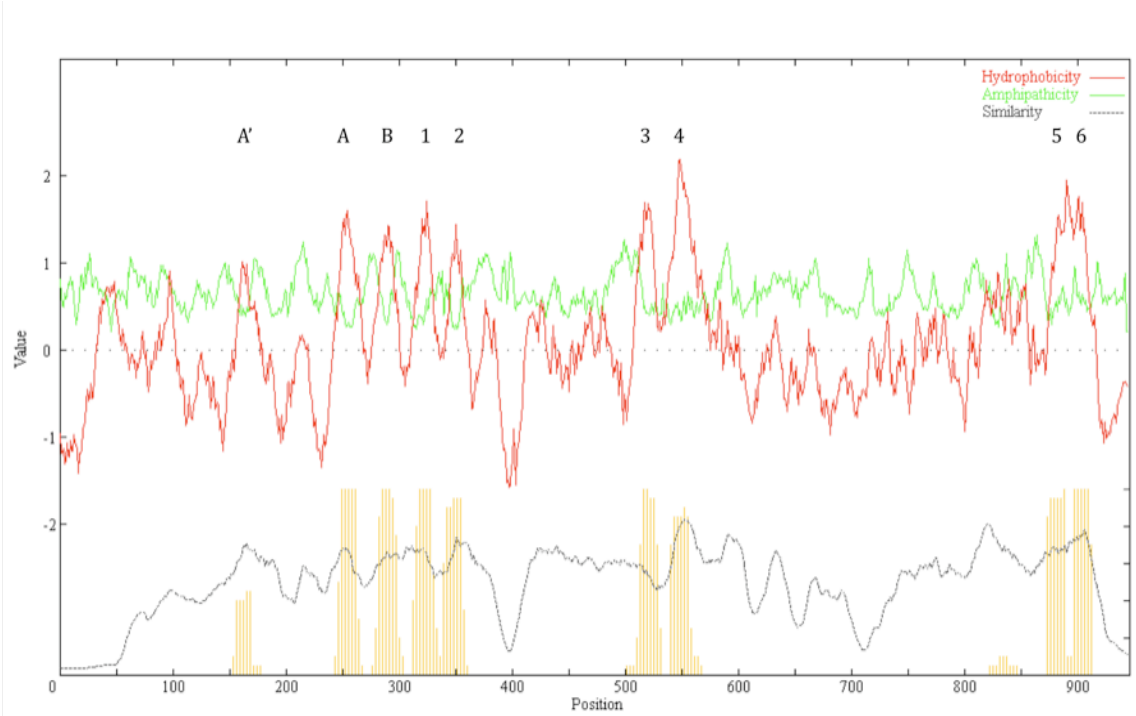

Figure B12: Topological analysis of the 26 members of Family 27, FUPA27, using the AveHAS program. Family 27 shows the novel Type VII topology, with A' marking the peak preceding the typical Type I topology.

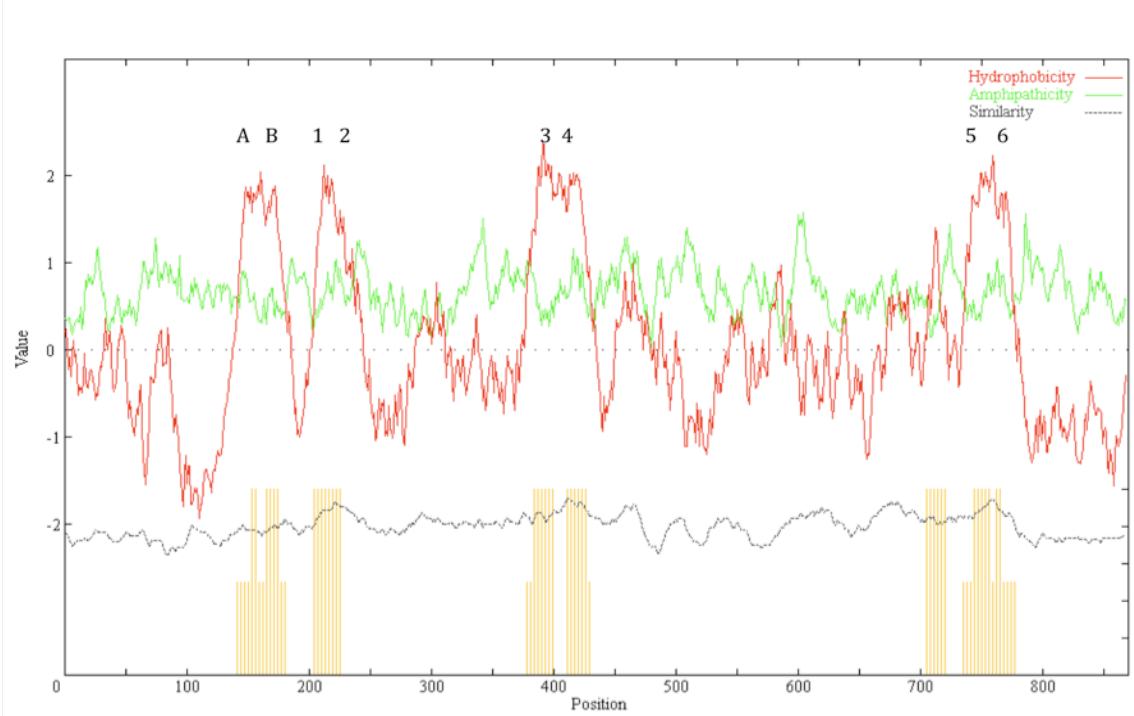

Figure B13: Topological analysis of the two members of Family 28, FUPA28, using the AveHAS program. Family 28 shows the typical Type I topology.

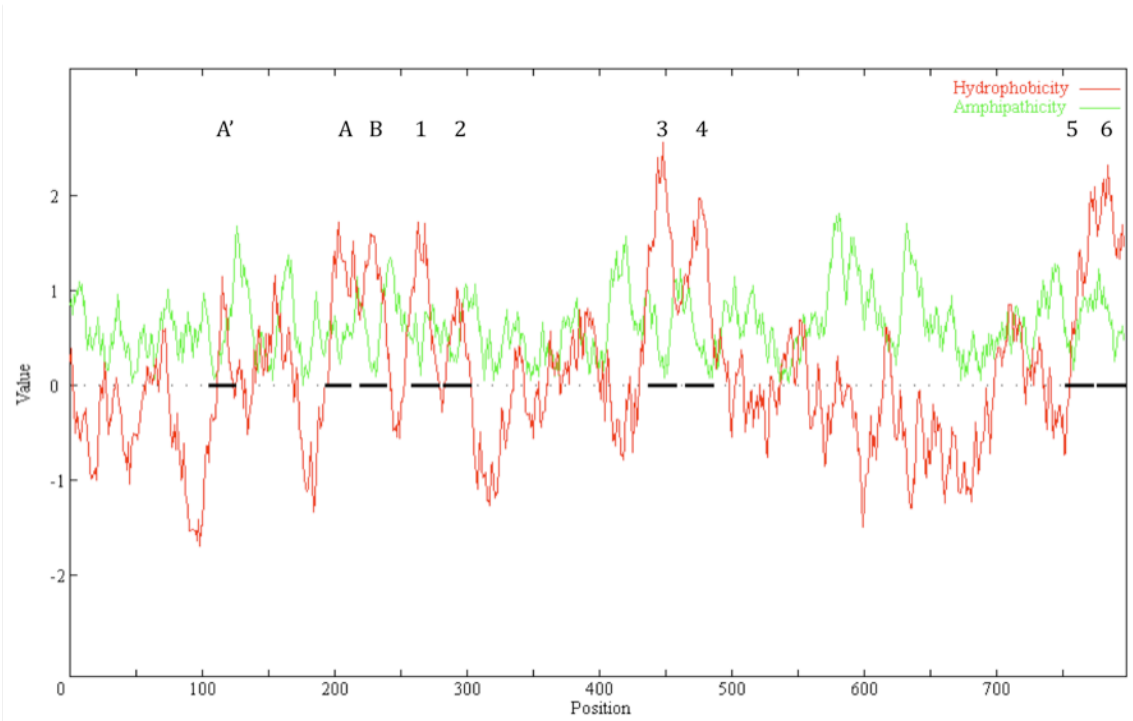

Figure B14: Topological analysis of the one member of Family 29, FUPA29, using the WHAT program. Family 29 shows the novel Type VII topology, with A' marking the peak preceding the typical Type I topology.

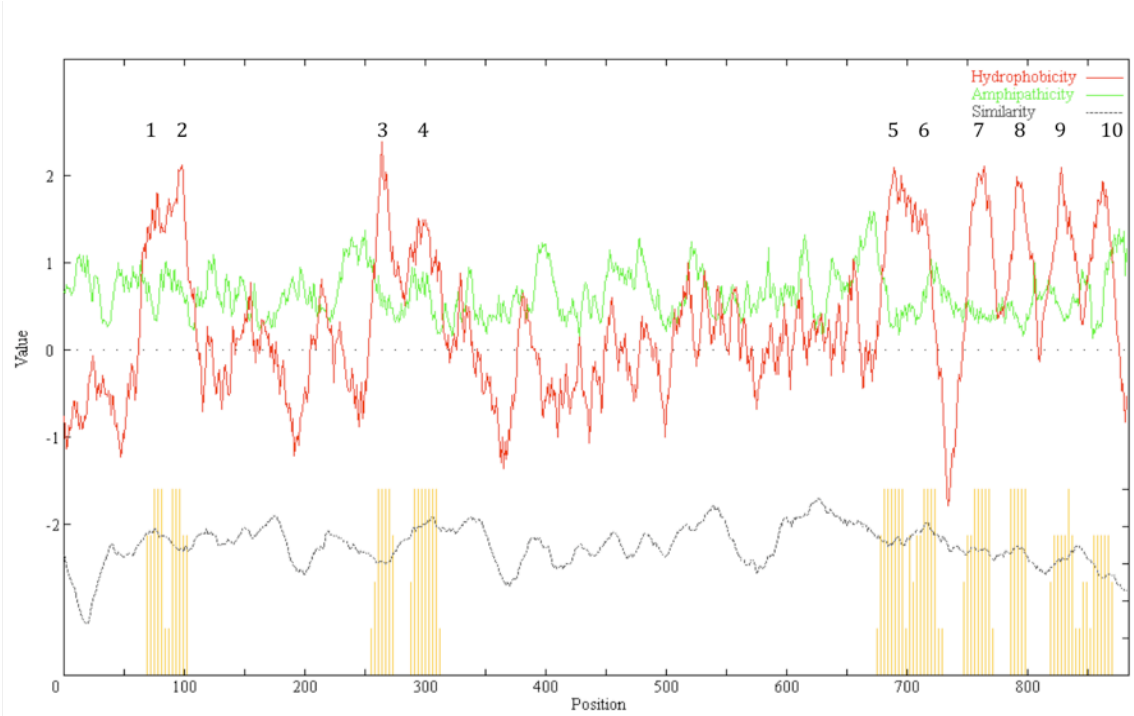

Figure B15: Topological analysis of the four members of Family 30, FUPA30, using the AveHAS program. Family 30 shows the typical Type II topology.

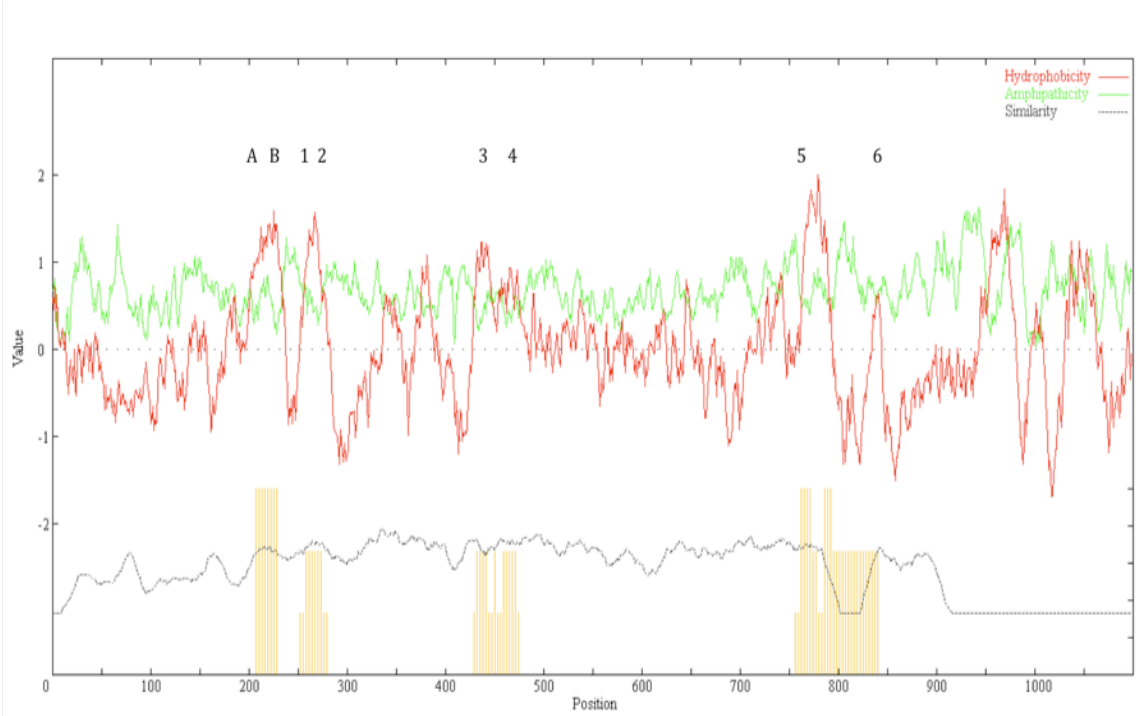

Figure B16: Topological analysis of the three members of Family 31, FUPA31, using the AveHAS program. Family 31 shows the typical Type I topology.

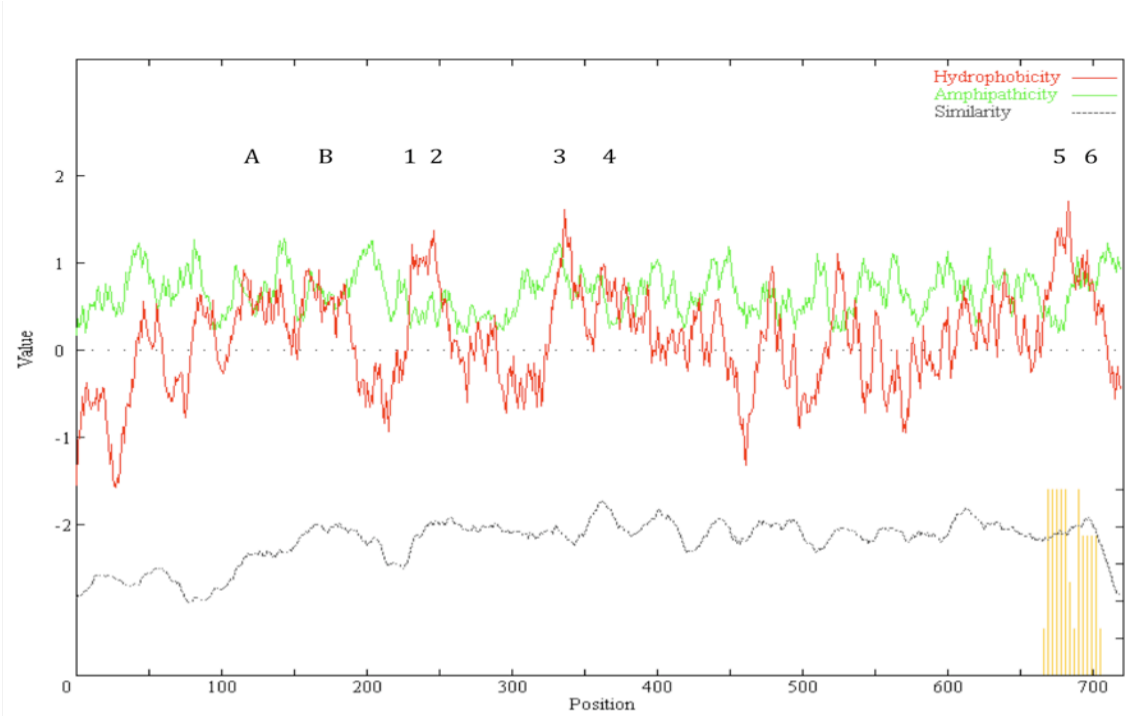

Figure B17: Topological analysis of the four members of Family 32, FUPA32, using the AveHAS program. Family 32 shows a topology similar to Type I, but is still ambiguous.

Table B1: Distribution of P-type ATPases in the five classes of proteobacteria. Number of organisms per class, average genome size per class and average ORF count are reported. Numbers of proteins for each class and the percentage of those proteins out of the total members of that particular class are reported. In addition, total numbers of proteins per family are reported along with the percentage of those proteins out of the total number of P-type ATPases analyzed.

Class	# of Organisms	Average Genome Size (Mbp)	Average ORF Count	Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	FUPA 25	FUPA 27	FUPA 28	FUPA 29	FUPA 30	FUPA 31	FUPA 32	Total
Alpha	11	5.0	4633	0/0%	2/4%	0/0%	4/8%	14/28%	9/18%	6/12%	2/4%	10/20%	0/0%	0/0%	2/4%	0/0%	1/2%	50
Beta	6	4.6	4084	0/0%	3/12%	0/0%	2/11%	9/36%	3/12%	3/12%	0/0%	3/12%	0/0%	0/0%	1/4%	0/0%	1/4%	25
Gamma	29	4.3	3715	1/0.9%	11/10%	0/0%	12/11%	33/30%	19/18%	13/12%	1/0.9%	13/12%	2/2%	0/0%	0/0%	3/3%	0/0%	108
Delta	5	3.3	3383	1/4%	6/27%	2/9%	0/0%	6/27%	1/4%	2/9%	1/4%	0/0%	0/0%	1/4%	1/4%	0/0%	1/4%	22
Epsilon	4	1.8	1768	0/0%	0/0%	0/0%	0/0%	8/62%	3/23%	1/8%	0/0%	0/0%	0/0%	0/0%	0/0%	0/0%	1/8%	13
Total	55	4.2	5350	2	22	2	18	70	35	25	4	26	2	1	4	3	4	218
Total %				0.9	10	0.9	8	32	16	11	1.8	12	0.9	0.5	1.8	1.3	1.8	

Table B2: Number of P-type ATPases per family encoded within 55 Proteobacterial genomes. A total of 218 P-type ATPases were identified. The number of P-type ATPases in Families 1-7, FUPA 25, 27-32 are compiled in this table. The complete organismal names along with the corresponding abbreviations (in parenthesis) are provided in column 1.

Organism	Genome Size	ORF Count	Class	Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	FUPA 25	FUPA 27	FUPA 28	FUPA 29	FUPA 30	FUPA 31	FUPA 32	Total
Agrobacterium tumefaciens str. C58 (Atu)	5.65	5360	α	0	0	0	0	2	1	1	0	1	0	0	0	0	0	5
Bradyrhizobium japonicum USDA 110 (Bja)	9.11	8317	α	0	0	0	1	1	0	1	0	1	0	0	1	0	0	5
Brucella melitensis 16M (Bme)	3.3	3198	α	0	0	0	2	1	2	0	0	1	0	0	0	0	0	6
Caulobacter crescentus CB15 (Ccr)	4.01	3737	α	0	0	0	0	0	1	1	0	1	0	0	0	0	0	3
Gluconobacter oxydans 621H (Gox)	2.92	2664	α	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Mesorhizobium loti MAFF303099 (Mlo)	7.59	7275	α	0	0	0	1	1	1	1	1	2	0	0	0	0	0	7
Nitrobacter winogradskyi Nb-255 (Nwi)	3.4	3122	α	0	0	0	0	1	2	0	0	0	0	0	0	0	0	3
Rhodopseudomonas palustris CGA009 (Rpa)	5.46	4841	α	0	1	0	0	1	1	1	0	1	0	0	1	0	1	7
Silicibacter pomeroyi DSS-3 (Spo)	4.4	4252	α	0	0	0	0	2	0	0	0	1	0	0	0	0	0	3
Sinorhizobium meliloti 1021 (Sme)	6.7	6205	α	0	1	0	0	3	1	1	1	2	0	0	0	0	0	9
Zymomonas mobilis subsp. mobilis ZM4 (Zmo)	2.06	1998	α	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
α				**0**	**2**	**0**	**4**	**14**	**9**	**6**	**2**	**10**	**0**	**0**	**2**	**0**	**1**	**50**

Table B2: (Continued)

Organism	Genome Size	ORF Count	Class	Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	FUPA 25	FUPA 27	FUPA 28	FUPA 29	FUPA 30	FUPA 31	FUPA 32	Total
Aromatoleum aromaticum EbN1 (Aar)	4.29	4124	β	0	2	0	0	2	0	0	0	1	0	0	0	0	1	6
Bordetella bronchiseptica RB50 (Bbr)	5.34	4994	β	0	0	0	1	1	1	1	0	0	0	0	0	0	0	4
Burkholderia pseudomallei K96243 (Bps)	7.25	5729	β	0	0	0	1	2	1	1	0	0	0	0	1	0	0	6
Neisseria meningitidis MC58 (Nme)	2.27	2079	β	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2
Nitrosomonas europaea ATCC 19718 (Neu)	2.81	2461	β	0	1	0	0	2	0	0	0	0	0	0	0	0	0	3
Ralstonia solanacearum GMI1000 (Rso)	5.81	5116	β	0	0	0	0	1	1	1	0	1	0	0	0	0	0	4
β				**0**	**3**	**0**	**2**	**9**	**3**	**3**	**0**	**3**	**0**	**0**	**1**	**0**	**1**	**25**
Acinetobacter sp. ADP1 (Asp)	3.6	3325	γ	0	0	0	1	1	0	1	0	0	0	0	0	0	0	3
Colwellia psychrerythraea 34H (Cps)	5.3	4910	γ	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2
Coxiella burnetii RSA 493 (Cbu)	2.1	2009	γ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Escherichia coli str. K-12 substr. MG1655 (Eco)	4.6	4237	γ	0	0	0	1	1	1	1	0	0	0	0	0	0	0	4
Francisella tularensis subsp. tularensis SCHU S4 (Ftu)	1.89	1603	γ	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Haemophilus influenzae Rd KW20 (Hin)	1.83	1714	γ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Idiomarina loihiensis L2TR (Ilo)	2.84	2628	γ	0	0	0	0	2	0	0	0	1	0	0	0	0	0	3

Table B2: (Continued)

Organism	Genome Size	ORF Count	Class	Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	FUPA 25	FUPA 27	FUPA 28	FUPA 29	FUPA 30	FUPA 31	FUPA 32	Total
Legionella pneumophila subsp. pneumophila str. Philadelphia 1 (Lpn)	3.4	2942	γ	0	2	0	1	2	2	0	0	0	2	0	0	0	0	9
Mannheimia succiniciproduce ns MBEL55E (Msu)	2.31	2384	γ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Methylococcus capsulatus str. Bath (Mca)	3.3	2959	γ	0	4	0	1	3	0	1	0	0	0	0	0	3	0	12
Nitrosococcus oceani ATCC 19707 (Noc)	3.52	3017	γ	0	2	0	0	2	0	0	0	0	0	0	0	0	0	4
Pasteurella multocida subsp. multocida str. Pm70 (Pmu)	2.4	2015	γ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Pectobacterium atrosepticum SCRI1043 (Pat)	5.06	4472	γ	0	0	0	1	1	1	1	0	0	0	0	0	0	0	4
Photobacterium profundum SS9 (Ppr)	6.4	5480	γ	0	0	0	0	1	1	0	0	1	0	0	0	0	0	3
Photorhabdus luminescens subsp. laumondii TTO1 (Plu)	5.69	4683	γ	0	0	0	0	1	1	1	0	0	0	0	0	0	0	3
Pseudoalteromon as haloplanktis TAC125 (Pha)	3.85	3486	γ	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2
Pseudomonas aeruginosa PAO1 (Pae)	6.3	5567	γ	0	1	0	1	1	2	1	0	1	0	0	0	0	0	7

Table B2: (Continued)

Organism	Genome Size	ORF Count	Class	Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	FUPA 25	FUPA 27	FUPA 28	FUPA 29	FUPA 30	FUPA 31	FUPA 32	Total
Pseudomonas fluorescens Pf-5 (Pfl)	6.5	6137	γ	0	0	0	2	1	2	2	0	1	0	0	0	0	0	8
Pseudomonas putida KT2440 (Ppu)	6.1	5350	γ	0	0	0	1	1	2	0	0	1	0	0	0	0	0	5
Pseudomonas syringae pv. tomato str. DC3000 (Psy)	6	5471	γ	0	0	0	0	1	2	1	0	1	0	0	0	0	0	5
Salmonella enterica subsp. enterica serovar Typhi str. CT18 (Sen)	4.8	4395	γ	0	0	0	2	1	1	1	0	0	0	0	0	0	0	5
Salmonella typhimurium LT2 (Sty)	4.8	4451	γ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Shewanella oneidensis MR-1 (Son)	5.14	4778	γ	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2
Thiomicrospira crunogena XCL-2 (Tcr)	2.43	2192	γ	1	1	0	0	2	0	0	1	1	0	0	0	0	0	6
Vibrio cholerae O1 biovar El Tor str. N16961 (Vch)	4	3835	γ	0	0	0	0	1	1	0	0	1	0	0	0	0	0	3
Vibrio parahaemolyticus RIMD 2210633 (Vpa)	5.17	4832	γ	0	0	0	0	1	1	0	0	1	0	0	0	0	0	3
Vibrio vulnificus CMCP6 (Vvu)	5.21	4537	γ	0	0	0	0	1	1	0	0	1	0	0	0	0	0	3
Xanthomonas axonopodis pv. citri str. 306 (Xax)	5.17	4312	γ	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Yersinia pestis CO92 (Ype)	4.65	4.83	γ	0	1	0	1	1	1	1	0	0	0	0	0	0	0	5
γ				**1**	**11**	**0**	**12**	**33**	**19**	**13**	**1**	**13**	**2**	**0**	**0**	**3**	**0**	**108**

Table B2: (Continued)

Organism	Genome Size	ORF Count	Class	Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	FUPA 25	FUPA 27	FUPA 28	FUPA 29	FUPA 30	FUPA 31	FUPA 32	Total
Bdellovibrio bacteriovorus HD100 (Bba)	3.7	3583	δ	0	0	0	0	2	0	0	0	0	0	1	1	0	0	4
Desulfotalea psychrophila LSv54 (Dps)	3.66	3236	δ	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2
Desulfovibrio vulgaris str. Hildenborough (Dvu)	3.2	3531	δ	0	1	0	0	1	0	1	1	0	0	0	0	0	1	5
Geobacter sulfurreducens PCA (Gsu)	2.5	3445	δ	0	2	1	0	1	1	1	0	0	0	0	0	0	0	6
Pelobacter carbinolicus DSM 2380 (Pca)	3.66	3118	δ	1	2	0	0	2	0	0	0	0	0	0	0	0	0	5
δ				**1**	**6**	**2**	**0**	**6**	**1**	**2**	**1**	**0**	**0**	**1**	**1**	**0**	**1**	**22**
Campylobacter jejuni subsp. jejuni NCTC 11168 (Cje)	1.64	1576	ε	0	0	0	0	2	0	1	0	0	0	0	0	0	0	3
Helicobacter hepaticus ATCC 51449 (Hhe)	1.8	1875	ε	0	0	0	0	2	1	0	0	0	0	0	0	0	0	3
Helicobacter pylori 26695 (Hpy)	1.66	1576	ε	0	0	0	0	2	1	0	0	0	0	0	0	0	0	3
Wolinella succinogenes DSM 1740 (Wsu)	2.11	2044	ε	0	0	0	0	2	1	0	0	0	0	0	0	0	1	4
ε				**0**	**0**	**0**	**0**	**8**	**3**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**13**
Average	4.194	5350	**Σ→**	**2**	**22**	**2**	**18**	**70**	**35**	**25**	**4**	**26**	**2**	**1**	**4**	**3**	**4**	**218**

Table B3: 218 P-type ATPases from the phylum of Proteobacteria included in this study, organized by family. In each family, proteins are shown in alphabetical order. Protein size in amino acids (aas), average size ± standard deviations (S.D.), and Genbank Index numbers are also reported. Proteobacterial class for each protein is also reported (α, β, γ, δ or ε).

Abbreviation	Organism	Protein Size (# of aas)	Genbank Index # (GI #)	Class
Family ` Na^+/K^+				
Tcr1	*Thiomicrospira crunogena* XCL-2	892	78485409	γ
Pca5	*Pelobacter carbinolicus* DSM 2380	896	77919873	δ
Average protein size ± S.D.		894±3		
Family 2 Ca^{2+}				
Sme1	*Sinorhizobium meliloti* 1021	900	16263078	α
Rpa1	*Rhodopseudomonas palustris* CGA009	914	39934334	α
Aar1	*Aromatoleum aromaticum* EbN1	897	56477784	β
Aar2	*Aromatoleum aromaticum* EbN1	911	56475557	β
Neu1	*Nitrosomonas europaea* ATCC 19718	912	30248632	β
Lpn1	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	842	52841251	γ
Lpn2	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	904	52841327	γ
Mca1	*Methylococcus capsulatus* str. Bath	919	53804203	γ
Mca2	*Methylococcus capsulatus* str. Bath	1031	53802308	γ
Mca3	*Methylococcus capsulatus* str. Bath	905	53804876	γ
Mca4	*Methylococcus capsulatus* str. Bath	884	53802673	γ
Noc1	*Nitrosococcus oceani* ATCC 19707	1082	77164900	γ
Noc2	*Nitrosococcus oceani* ATCC 19707	884	77165595	γ
Pae1	*Pseudomonas aeruginosa* PAO1	902	15596626	γ
Tcr2	*Thiomicrospira crunogena* XCL-2	898	78486262	γ
Ype1	*Yersinia pestis* CO92	908	16120780	γ
Dps1	*Desulfotalea psychrophila* LSv54	919	51244159	δ
Dvu1	*Desulfovibrio vulgaris* str. Hildenborough	917	46580402	δ
Gsu1	*Geobacter sulfurreducens* PCA	871	39996778	δ
Gsu2	*Geobacter sulfurreducens* PCA	916	39997423	δ
Pca1	*Pelobacter carbinolicus* DSM 2380	899	77918506	δ
Pca2	Pelobacter carbinolicus DSM 2380	906	77918957	δ
Average protein size ± S.D.		915±49		
Family 3 H^+				
Dps2	*Desulfotalea psychrophila* LSv54	858	51244834	δ
Gsu6	*Geobacter sulfurreducens* PCA	868	39997447	δ
Average protein size ± S.D.		863±7		
Family 4 Mg^{2+}				
Bja2	*Bradyrhizobium japonicum* USDA 110	832	27382164	α
Bme1	*Brucella melitensis* 16M	928	17988400	α
Bme3	*Brucella melitensis* 16M	928	17988400	α
Mlo1	*Mesorhizobium loti* MAFF303099	896	13475516	α
Bbr1	*Bordetella bronchiseptica* RB50	918	33600086	β
Bps2	*Burkholderia pseudomallei* K96243	928	53721996	β
Asp1	*Acinetobacter sp.* ADP1	920	50085121	γ
Eco1	*Escherichia coli* str. K-12 substr. MG1655	898	16132064	γ
Lpn3	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	855	52842590	γ
Mca5	*Methylococcus capsulatus* str. Bath	951	53804030	γ
Pae2	*Pseudomonas aeruginosa* PAO1	903	15600018	γ
Pat1	*Pectobacterium atrosepticum* SCRI1043	903	50119386	γ
Pfl1	*Pseudomonas fluorescens* Pf-5	903	70730713	γ
Pfl2	*Pseudomonas fluorescens* Pf-5	921	70731434	γ

Table B3: (Continued)

Abbreviation	Organism	Protein Size (# of aas)	Genbank Index # (GI #)	Class
Ppu1	*Pseudomonas putida* KT2440	920	26989364	γ
Sen1	*Salmonella enterica subsp. enterica serovar Typhi* str. CT18	908	16762554	γ
Sen2	*Salmonella enterica subsp. enterica serovar Typhi* str. CT18	902	16763257	γ
Ype2	*Yersinia pestis* CO92	899	16121926	γ
Average protein size ± S.D.		906±27		
Family 5 Cu^{+}				
Atu1	*Agrobacterium tumefaciens* str. C58	861	15888276	α
Atu2	*Agrobacterium tumefaciens* str. C58	836	15888531	α
Bja3	*Bradyrhizobium japonicum* USDA 110	823	27375811	α
Bme2	*Brucella melitensis* 16M	826	17988013	α
Gox1	*Gluconobacter oxydans* 621H	791	58039131	α
Mlo2	*Mesorhizobium loti* MAFF303099	839	13474443	α
Nwi1	*Nitrobacter winogradskyi* Nb-255	831	75675104	α
Rpa3	*Rhodopseudomonas palustris* CGA009	973	39934730	α
Sme2	*Sinorhizobium meliloti* 1021	826	16263001	α
Sme3	*Sinorhizobium meliloti* 1021	827	16264766	α
Sme4	*Sinorhizobium meliloti* 1021	733	16263042	α
Spo1	*Silicibacter pomeroyi* DSS-3	828	56695696	α
Spo2	*Silicibacter pomeroyi* DSS-3	785	56709147	α
Zmo1	*Zymomonas mobilis subsp. mobilis* ZM4	740	56551811	α
Aar3	*Aromatoleum aromaticum* EbN1	803	56478367	β
Aar4	*Aromatoleum aromaticum* EbN1	785	56476640	β
Bbr2	*Bordetella bronchiseptica* RB50	808	33600166	β
Bps3	*Burkholderia pseudomallei* K96243	976	53721259	β
Bps4	*Burkholderia pseudomallei* K96243	807	53717943	β
Neu2	*Nitrosomonas europaea* ATCC 19718	782	30249022	β
Neu3	*Nitrosomonas europaea* ATCC 19718	722	30249201	β
Nme1	*Neisseria meningitidis* MC58	720	15677191	β
Rso1	*Ralstonia solanacearum* GMI1000	748	17548065	β
Asp2	*Acinetobacter sp.* ADP1	802	50085487	γ
Cbu1	*Coxiella burnetii* RSA 493	742	29654798	γ
Cps1	*Colwellia psychrerythraea* 34H	791	71280081	γ
Eco2	*Escherichia coli* str. K-12 substr. MG1655	834	16128468	γ
Hin1	*Haemophilus influenzae* Rd KW20	722	16272245	γ
Ilo1	*Idiomarina loihiensis* L2TR	753	56460326	γ
Ilo2	*Idiomarina loihiensis* L2TR	749	56459705	γ
Lpn4	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	736	52841258	γ
Lpn5	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	735	52841854	γ
Mca6	*Methylococcus capsulatus* str. Bath	725	53803908	γ
Mca7	*Methylococcus capsulatus* str. Bath	779	53805105	γ
Mca8	*Methylococcus capsulatus* str. Bath	831	53804835	γ
Msu1	*Mannheimia succiniciproducens* MBEL55E	750	52424943	γ
Noc3	*Nitrosococcus oceani* ATCC 19707	724	77165022	γ
Noc4	*Nitrosococcus oceani* ATCC 19707	823	77163740	γ
Pae3	*Pseudomonas aeruginosa* PAO1	792	15599115	γ
Pat2	*Pectobacterium atrosepticum* SCRI1043	907	50120132	γ
Pfl3	*Pseudomonas fluorescens* Pf-5	798	70734212	γ
Pha1	*Pseudoalteromonas haloplanktis* TAC125	748	77359957	γ
Plu1	*Photorhabdus luminescens subsp. laumondii* TTO1	911	37527684	γ
Pmu1	*Pasteurella multocida subsp. multocida* str. Pm70	724	15603757	γ
Ppr1	*Photobacterium profundum* SS9	965	54309978	γ
Ppu2	*Pseudomonas putida* KT2440	799	26987324	γ
Psy1	*Pseudomonas syringae pv. tomato* str. DC3000	732	28867978	γ
Sen3	*Salmonella enterica subsp. enterica serovar Typhi* str. CT18	833	16759476	γ
Son1	*Shewanella oneidensis* MR-1	753	24373257	γ
Sty1	*Salmonella typhimurium* LT2	762	16763733	γ
Tcr3	*Thiomicrospira crunogena* XCL-2	776	78486448	γ
Tcr5	*Thiomicrospira crunogena* XCL-2	670	78485287	γ

Table B3: (Continued)

Abbreviation	Organism	Protein Size (# of aas)	Genbank Index # (GI #)	Class
Vch1	*Vibrio cholerae O1 biovar El Tor* str. N16961	915	15642213	γ
Vpa1	*Vibrio parahaemolyticus* RIMD 2210633	911	28897532	γ
Vvu1	*Vibrio vulnificus* CMCP6	912	27363721	γ
Ype3	*Yersinia pestis* CO92	961	16123263	γ
Bba1	*Bdellovibrio bacteriovorus* HD100	692	42523748	δ
Bba2	*Bdellovibrio bacteriovorus* HD100	724	42523682	δ
Dvu2	*Desulfovibrio vulgaris* str. Hildenborough	905	46580729	δ
Gsu3	*Geobacter sulfurreducens* PCA	797	39997547	δ
Pca3	*Pelobacter carbinolicus* DSM 2380	834	77919195	δ
Pca4	*Pelobacter carbinolicus* DSM 2380	767	77919310	δ
Cje1	*Campylobacter jejuni subsp. jejuni* NCTC 11168	699	15792485	ε
Cje2	*Campylobacter jejuni subsp. jejuni* NCTC 11168	785	15792479	ε
Hhe1	*Helicobacter hepaticus* ATCC 51449	747	32262231	ε
Hhe2	*Helicobacter hepaticus* ATCC 51449	795	32262572	ε
Hpy1	*Helicobacter pylori* 26695	745	15645686	ε
Hpy2	*Helicobacter pylori* 26695	788	15646112	ε
Wsu1	*Wolinella succinogenes* DSM 1740	732	34556850	ε
Wsu2	*Wolinella succinogenes* DSM 1740	799	34558447	ε
Average protein size ± S.D.		798±70		
Family 6 Heavy Metal				
Atu4	*Agrobacterium tumefaciens* str. C58	905	17934751	α
Bme5	*Brucella melitensis* 16M	813	17986337	α
Bme6	*Brucella melitensis* 16M	673	17988441	α
Ccr2	*Caulobacter crescentus* CB15	677	16126959	α
Mlo5	*Mesorhizobium loti* MAFF303099	749	13472245	α
Nwi2	*Nitrobacter winogradskyi* Nb-255	712	75677312	α
Nwi3	*Nitrobacter winogradskyi* Nb-255	665	75676140	α
Rpa5	*Rhodopseudomonas palustris* CGA009	709	39936323	α
Sme7	*Sinorhizobium meliloti* 1021	743	15963877	α
Bbr3	*Bordetella bronchiseptica* RB50	778	33601483	β
Bps5	*Burkholderia pseudomallei* K96243	836	53720986	β
Rso2	*Ralstonia solanacearum* GMI1000	784	17548540	β
Eco3	Escherichia coli str. K-12 substr. MG1655	732	16131341	γ
Lpn6	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	635	52841244	γ
Lpn7	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	729	52841243	γ
Pae5	*Pseudomonas aeruginosa* PAO1	740	15598886	γ
Pae7	*Pseudomonas aeruginosa* PAO1	661	15597631	γ
Pat3	*Pectobacterium atrosepticum* SCRI1043	787	50123272	γ
Pfl5	*Pseudomonas fluorescens* Pf-5	818	70733177	γ
Pfl8	*Pseudomonas fluorescens* Pf-5	634	70733474	γ
Plu2	*Photorhabdus luminescens subsp. laumondii* TTO1	761	37527955	γ
Ppr3	*Photobacterium profundum* SS9	801	54308413	γ
Ppu4	*Pseudomonas putida* KT2440	665	26986786	γ
Ppu5	*Pseudomonas putida* KT2440	750	26991815	γ
Psy3	*Pseudomonas syringae pv. tomato* str. DC3000	752	28872391	γ
Psy5	*Pseudomonas syringae pv. tomato* str. DC3000	641	28872637	γ
Sen4	*Salmonella enterica subsp. enterica serovar Typhi* str. CT18	732	16762728	γ
Vch3	*Vibrio cholerae O1 biovar El Tor* str. N16961	768	15641046	γ
Vpa3	*Vibrio parahaemolyticus* RIMD 2210633	768	28897733	γ
Vvu3	*Vibrio vulnificus* CMCP6	739	27365397	γ
Ype4	*Yersinia pestis* CO92	788	16123955	γ
Gsu4	*Geobacter sulfurreducens* PCA	713	39997245	δ
Hhe3	*Helicobacter hepaticus* ATCC 51449	695	32262134	ε
Hpy3	*Helicobacter pylori* 26695	686	15645410	ε
Wsu3	*Wolinella succinogenes* DSM 1740	707	34557497	ε
Average protein size ± S.D.		736±61		

Table B3: (Continued)

Abbreviation	Organism	Protein Size (# of aas)	Genbank Index # (GI #)	Class
Family 7 K^+				
Atu5	*Agrobacterium tumefaciens* str. C58	718	15891162	α
Bja5	*Bradyrhizobium japonicum* USDA 110	706	27381889	α
Ccr3	*Caulobacter crescentus* CB15	686	16125839	α
Mlo6	*Mesorhizobium loti* MAFF303099	697	13472737	α
Rpa6	*Rhodopseudomonas palustris* CGA009	709	39936068	α
Sme8	*Sinorhizobium meliloti* 1021	680	16263707	α
Bbr4	*Bordetella bronchiseptica* RB50	721	33602892	β
Bps6	*Burkholderia pseudomallei* K96243	686	53718808	β
Rso3	*Ralstonia solanacearum* GMI1000	744	17548100	β
Asp3	*Acinetobacter sp.* ADP1	671	50085998	γ
Eco4	*Escherichia coli* str. K-12 substr. MG1655	682	16128673	γ
Ftu1	*Francisella tularensis subsp. tularensis* SCHU S4	679	56708741	γ
Mca9	*Methylococcus capsulatus* str. Bath	690	53803781	γ
Pae6	*Pseudomonas aeruginosa* PAO1	690	15596831	γ
Pat4	*Pectobacterium atrosepticum* SCRI1043	714	50120280	γ
Pfl6	*Pseudomonas fluorescens* Pf-5	690	70731913	γ
Pfl7	*Pseudomonas fluorescens* Pf-5	694	70731647	γ
Plu3	*Photorhabdus luminescens subsp. laumondii* TTO1	688	37525373	γ
Psy4	*Pseudomonas syringae pv. tomato* str. DC3000	692	28869445	γ
Sen5	*Salmonella enterica subsp. enterica serovar Typhi* str. CT18	569	16759648	γ
Xax1	*Xanthomonas axonopodis pv. citri* str. 306	682	21241528	γ
Ype5	*Yersinia pestis* CO92	688	16122896	γ
Dvu5	*Desulfovibrio vulgaris* str. Hildenborough	678	46581738	δ
Gsu5	*Geobacter sulfurreducens* PCA	689	39997576	δ
Cje3	*Campylobacter jejuni subsp. jejuni* NCTC 11168	681	15792027	ε
Average protein size ± S.D.		689±30		
FUPA25				
Mlo7	*Mesorhizobium loti* MAFF303099	617	13475503	α
Sme9	*Sinorhizobium meliloti* 1021	746	16263085	α
Tcr4	*Thiomicrospira crunogena* XCL-2	759	78485067	γ
Dvu3	*Desulfovibrio vulgaris* str. Hildenborough	633	46581204	δ
Average protein size ± S.D.		689±74		
FUPA27				
Atu3	*Agrobacterium tumefaciens* str. C58	763	15888852	α
Bja4	*Bradyrhizobium japonicum* USDA 110	730	27377880	α
Bme4	*Brucella melitensis* 16M	752	17987852	α
Ccr1	*Caulobacter crescentus* CB15	724	16125656	α
Mlo3	*Mesorhizobium loti* MAFF303099	762	13475370	α
Mlo4	*Mesorhizobium loti* MAFF303099	762	13475529	α
Rpa4	*Rhodopseudomonas palustris* CGA009	732	39933093	α
Sme5	*Sinorhizobium meliloti* 1021	755	16262778	α
Sme6	*Sinorhizobium meliloti* 1021	757	16263112	α
Spo3	*Silicibacter pomeroyi* DSS-3	725	56698342	α
Aar6	*Aromatoleum aromaticum* EbN1	817	56478348	β
Nme2	*Neisseria meningitidis* MC58	823	15676928	β
Rso4	*Ralstonia solanacearum* GMI1000	851	17545993	β
Cps2	*Colwellia psychrerythraea* 34H	820	71279318	γ
Ilo3	*Idiomarina loihiensis* L2TR	792	56460410	γ
Pae4	*Pseudomonas aeruginosa* PAO1	811	15596746	γ
Pfl4	*Pseudomonas fluorescens* Pf-5	778	70729297	γ
Pha2	*Pseudoalteromonas haloplanktis* TAC125	791	77360787	γ
Ppr2	*Photobacterium profundum* SS9	769	54309022	γ
Ppu3	*Pseudomonas putida* KT2440	882	26990952	γ
Psy2	*Pseudomonas syringae pv. tomato* str. DC3000	823	28869200	γ
Son2	*Shewanella oneidensis* MR-1	799	24373906	γ
Tcr6	*Thiomicrospira crunogena* XCL-2	828	78486299	γ

Table B3: (Continued)

Abbreviation	Organism	Protein Size (# of aas)	Genbank Index # (GI #)	Class
Vch2	*Vibrio cholerae O1 biovar El Tor* str. N16961	790	15641448	γ
Vpa2	*Vibrio parahaemolyticus* RIMD 2210633	787	28898313	γ
Vvu2	*Vibrio vulnificus* CMCP6	789	27365909	γ
Average protein size ± S.D.		785±40		
FUPA28				
Lpn8	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	852	52840486	γ
Lpn9	*Legionella pneumophila subsp. pneumophila* str. Philadelphia 1	847	52842897	γ
Average protein size ± S.D.		849±3		
FUPA29				
Bba3	*Bdellovibrio bacteriovorus* HD100	798	42524029	δ
Average protein size ± S.D.		798		
FUPA30				
Bja1	*Bradyrhizobium japonicum* USDA 110	850	27378926	α
Rpa2	*Rhodopseudomonas palustris* CGA009	852	39935958	α
Bps1	*Burkholderia pseudomallei* K96243	837	53722128	β
Bba4	*Bdellovibrio bacteriovorus* HD100	825	42522486	δ
Average protein size ± S.D.		841±13		
FUPA31				
Mca10	*Methylococcus capsulatus* str. Bath	1068	53804058	γ
Mca11	*Methylococcus capsulatus* str. Bath	673	53802306	γ
Mca12	*Methylococcus capsulatus* str. Bath	839	53804062	γ
Average protein size ± S.D.		860±198		
FUPA32				
Rpa7	*Rhodopseudomonas palustris* CGA009	698	39935402	α
Aar5	*Aromatoleum aromaticum* EbN1	694	56475751	β
Dvu4	*Desulfovibrio vulgaris* str. Hildenborough	606	46581732	δ
Wsu4	*Wolinella succinogenes* DSM 1740	716	34557896	ε
Average protein size ± S.D.		678±49		

Table B4: Average size of P-type ATPases ± standard deviation by family. The number of proteins belonging to each family is also reported.

P-type ATPase Family (number of proteins in the family)	Average size ± Standard Deviation
Na^+/K^+ (2)	894 ± 3
Ca^{2+} (22)	915 ± 49
H^+ (2)	863 ± 7
Mg^{2+} (18)	906 ± 27
Cu^+ (70)	798 ± 70
Heavy Metal (35)	736 ± 61
K^+ (25)	689 ± 30
FUPA25 (4)	689 ± 74
FUPA27 (26)	785 ± 40
FUPA28 (2)	849 ± 3
FUPA29 (1)	798
FUPA30 (4)	841 ± 13
FUPA31 (3)	860 ± 198
FUPA32 (4)	678 ± 49

Table B5: Conserved motifs for the Proteobacterial P-type ATPases families. The nine consensus motifs along with the conserved motifs of these families are shown. Degree of conservation is as follows: * signifies an identity, : signifies a close similarity, . signifies a distant similarity, _ signifies a lack of similarity among the residues present at a particular position.

Consensus	**PGD**	**PAD**	**TGES**	**PEGL**	**DKTGTLT**	**KGAPE**	**DPPR**	**MVTGD**	**VAVTGDGVNDSPALKKADIGVAM**
Family 1 Na^+/K^+									
Seq. Motif	PGD	PAD	TGES	PEGL	DKTGTLT	KGAVE	DPAR	MITGD	VAMTGDTVNDSPALKAADIGIAM
Conservation	***	.**	.**.	****	*******	***_*	**.*	*****	**********.****_**.****
Family 2 Ca^{2+}									
Seq. Motif	PGD	PAD	TGES	PEGL	DKTGTLT	KGAPE	DPPR	MITGD	VAMTGDGVNDAPALKRADIGVAM
Conservation	.**	.:*	****	**.*	****.*	.*:__	*__:	*.***	_.:****.**.***._*_.*.::
Family 3 H^+									
Seq. Motif	PGD	PAD	TGES	PVAL	DKTGTLT	KGAPQ	DPPR	MITGD	VAMTGDGVNDAPALKKADAGIAV
Conservation	***	***	****	***.	*******	*****	****	*****	*.**********.**._**
Family 4 Mg^{2+}									
Seq. Motif	PGD	PAD	TGES	PEML	DKTGTLT	KGAVE	DPPK	VLTGD	VGFMGDGINDAPALRDADIGISV
Conservation	_**	***	.**.	**.*	*******	***_:	**__	:::**	**::******_:::_**.*.**
Family 5 Cu^+									
Seq. Motif	PGE	PVD	TGEP	PCAL	DKTGTLT	KGVAG	DPLK	MLTGD	VAMVGDGINDAPALAAADVGIAM
Conservation	_*_	__*	.**_	*_::	****_:*	_*___	.__:	:.**	____***_**._.*__:__.._.
Family 6 HM									
Seq. Motif	PGD	PAD	TGES	PCAL	DKTGTLT	HPLAQ	DTPR	MLTGD	VAMVGDGINDAPALKAASIGIAM
Conservation	___	__*	.***	***.	*****.*	*__._	*__:	:::**	____****.**.._:__._:...:
Family 7 K^+									
Seq. Motif	PGD	SVD	TGES	PTTI	DKTGTIT	KGAVD	DIVK	MITGD	VAMTGDGTNDAPALAQADVAVAM
Conservation	._*	_::	****	____	*******	**._.	*_:*	*_***	.** ***.****.*.*.*.*.. *

Table B5: (Continued)

Consensus	**PGD**	**PAD**	**TGES**	**PEGL**	**DKTGTLT**	**KGAPE**	**DPPR**	**MVTGD**	**VAVTGDGVNDSPALKKADIGVAM**
FUPA25									
Seq. Motif	PGD	PVD	TGEP	PCPL	DKTGTLT	VGQPS	DEVR	LLTGD	VMMVGDGVNDAPALAAADVGVAM
Conservation	_**	***	***.	**.*	*****.*	_*__:	*_:*	:_:**	.::****.******* _*_.**
FUPA27									
Seq. Motif	PGD	PAD	TGES	PCAL	DKTGTLT	IGQPA	DRIR	LLSGD	VLMVGDGINDAPVLAAAHVSVAM
Conservation	___	..*	.**_	***:	****.**	_____	:__:	:_:**	.___***.**_._:__*._*_:_
FUPA28									
Seq. Motif	PGE	PVD	TGAP	PCAL	DLNGTLT	IIGNK	DPLR	ICTGA	VAMVGDAANDATALAASDIGIAV
Conservation	_*_	***	_*__	**.*	*******	_**_*	****	.*.**	***.*******.*.*_ **.****
FUPA29	Only	One	Protein	No	Consensus				
FUPA30									
Seq. Motif	PGD	PAD	TGES	PEEF	DKTGTLT	KGAPE	DPLR	MITGD	VAMTGDGVNDAPALKAAHIGIAM
Conservation	_**	.**	****	***:	*****.*	**.**	**_*	*****	**.*******.*.****..****
FUPA31									
Seq. Motif	AGD	PVD	TGES	PTLV	DKTGTLT	KGAEI	DPEI	VVIDD	VALVGDGVNDSPALKKADVSVSL
Conservation	_**	*_*	_*_.	*:_	*__.___	.._.:	____	:__._	.___***__:::____.*.*:.:
FUPA32									
Seq. Motif	AGD	PVD	TGES	SCAL	DKTGTLT	KAAEA	DEVR	MLTGD	VAFVGDGINDAPALAGAHVGIAM
Conservation	_**	*.*	***.	***:	*******	__*._	*_:*	.****	:_::********** __*.*.:_:

Discussion

Eleven species of bacteria belonging to the three closely related genuses *Bacteroides*, *Flavobacterium* and *Fusobacterium*, were analyzed for P-type ATPases. In total, 47 different proteins exhibiting similarity to known standards of P-type ATPases from TCDB were isolated. Of the genuses, *Bacteroides*, *Flavobacterium* and *Fusobacterium*, they encoded in their genomes 28, 9 and 10 P-type ATPases respectively (Table A1).

The 47 enzymes from these eleven organisms were shown to have characteristics of functionally characterized P-type ATPases as well as many functionally uncharacterized P-type ATPases following phylogenetic analyses with standard proteins. Familial representation showed many proteins belonging to the characterized families of known function. Family 2 (Ca^{2+}) ATPases had 8 such homologues while Family 4 (Mg^{2+}) has only 3 such members, the latter present only in *Bacteroides* species (Table A2). Of the topological Type I families, Family 5 (Copper) and Family 6 (Heavy Metal) had 11 homologues each. Kdp-type K^+ uptake ATPases (Family 7) had eight members, one each from different organisms in Bacteroidetes. Functionally characterized Families 1, 3, 8 and 9 have no representation in the analysis of genuses *Bacteroides*, *Flavobacterium* and *Fusobacterium* reported in this study. In addition to functionally characterized enzymes, members from three families of functionally uncharacterized proteins were examined. FUPA29, 30 and 32 had two, one and three homologues, respectively (Table A2).

Of all of the organisms, *Bacteroides fragilis* NCTC 9343 and *Flavobacterium johnsoniae* UW101 were the only organisms to have two members of the Copper family of ATPases. All other species in this study had just one or zero members in this family.

It is noteworthy that Family 5 is also the family with the largest number of total proteins, 11 (Table A2). A fairly good correlation between genomic size and the number of P-type ATPases per organism was observed (Figure A3). This all seems to suggest that the organisms analyzed here have had a high degree of genetic stability during their evolutionary history.

Phylogenetic analysis of these 47 proteins was performed in order to understand familial relationships and evolutionary pressures that could have contributed to their presence. By looking at the 16S rRNAs of the organisms, it was apparent that horizontal gene transfer had not occured (Figure A4). The majority of the proteins within all of the functionally characterized families seem to be othologous.

Motif analysis between members of a particular family showed varied degrees of conservation. All of the identified proteins are expected to be functional regardless of whether they belong to functionally characterized families or not (Table A4). Motif 1 (PGD) was not very well conserved in Type I homologues (Family 5 and 6), and most families except Family 4 had the first residue changed from a P. Motif 2 (PAD) showed higher degrees of conservation within these families, with Families 2 and 4 showing good conservation. Motif 3 (TGES) shows a good degree of conservation for the first three residues, TGE, being well conserved in all families except Family 5. Motif 4 (PEGL) is well conserved within Ca^{2+} homologues, with differences in the motif occurring for the other families examined. Motif 5 (DKTGTLT), containing the phosphorylation site, shows high degrees of conservation in all the families, with full conservation at the critical aspartyl residue. Overall, motifs 6 (KGAPE) and 7 (DPPR) are not well conserved. Motif 8 (MVTGD) is fully conserved in Family 7 and shows varying degrees

of conservation in the other families. Motif 9 (VAVTGDGVNDSPALKKADIGVAM), the hinge motif, is well conserved. The internal GDG-N is fully conserved in all families as is true of many other P-type ATPases (Table A4).

Of the functionally characterized families, topologies are as expected with Families 5 and 6 having the characteristic Type I topology, Families 2 and 4 the characteristic Type II topology and Family 7 the characteristic Type III topology (Figures A5-A9). Of the FUPA homologues, FUPA30 showed the Type II topology, and FUPA32 showed the Type I topology (Figure A11 and A12).

In addition, FUPA29 showed a topology that had some features of Type I yet were vague (Figure A10). These homologues clustered near Type I proteins of Family 5 and Family 6. BLAST searches with members of these proteins, and after examining their hydropathy plots, it was suggested that members of this family have a TMS preceding TMSs A and B, noted A'. Thus a new probable topological type was identified. We have classified these homologues topological Type VII.

In the second part of this study, the phylum of Proteobacteria was analyzed for P-type ATPases. In contrast to the three genuses hat were studied previously, this study looked at all five classes of Proteobacteria, including 48 different genuses and a total of 55 different organisms. Classification of 218 P-type ATPases was performed assigning them to the functionally characterized and functionally uncharacterized families. The organism *Methylococcus capsulatus str. Bath*, a γ-Proteobacterium had the largest number of enzymes studied, 12, although on average, there were 3 enzymes per species (Table B2).

Phylogenetic analysis seen in Figures B1 and B2 showed that the phylum of Proteobacteria had members from Families 1-7 as well as homologues from FUPA25, 27-32. Family 1 (Na^+/ K^+) had two such members while Family 2 showed 22 different homologues. Family 3, H^+ ATPases had 2 members, both from δ-Proteobacteria. Family 4 Mg^{2+} homologues had 18 proteins present in α-, β-, and γ-Proteobacteria. Family 5 Copper ATPases had the largest representation. 32% of the ATPases examined were of this type. These proteins were found in all five classes of Proteobacteria. Family 6 (Heavy Metal) proteins consist of 35 different homologues while the Kdp-type K^+ (Family 7) has 25 such enzymes. Four enzymes make up Family 25 while Family 27 contains the largest number of homologues of all of the functionally uncharacterized families, 26 members. FUPA28, 29, 30, 31 and 32 contained two, one, four, three and four members, respectively.

16S rRNA analysis of the 48 genuses that included these 55 organisms showed genuses grouping according to class as expected (Figure B3). The only exception to this seemed to be the genus *Xanthomonas*, which clusters loosely with the β-Proteobacterial rRNAs. It is possible to see that the γ-Proteobacterial and δ-Proteobacterial rRNAs are more diverse in sequence than the α-, β-, and ε-Proteobacteria rRNAs (Figure B3). Comparison of the 16S rRNA phylogenetic clustering with the phylogenetic relationships between the 216 proteins (2 proteins out of 218 proved to be aberrant: Sen5 a Family 7 protein, and Mca10, a FUPA31 homologue) showed that there were few orthologous members between the proteins (Figure B1 and Figure B3). In each of the subclusters within the phylogenetic tree of the proteins, there was a high degree of intermixing between the five Proteobacterial classes. Subclusters of proteins of the same class had

16S rRNA relationships that were groundly not characteristic of orthology; rather, the genuses were far and separate. As a result, it is apparent that a high frequency of horizontal transfer of genes encoding P-type ATPases has occurred between members of the Proteobacterial phylum.

Table B5 shows the motif analysis for the phylum. Motif 1 (PGD) in Proteobacteria is not well conserved in all families examined while motif 2 (PAD) is only fully conserved in Families 3 and 4. Motif 3 shows full conservation in Families 2, 3, 7 and 30 while motif 4 has poor conservation in all families except Family 1. The fifth motif (DKTGTLT) is well conserved in most families, with only FUPA31 showing poor conservation among its homologues. The critical aspartyl residue is fully conserved in all families. Both motifs 6 and 7 show poor conservation in the majority of the families analyzed. Motif 8, MVTGD, shows subtle changes in almost all families though the last two residues, GD, are fully conserved in all families except FUPA28 and FUPA31 homologues. The hinge motif, VAVTGDGVNDSPALKKADIGVAM (motif 9) is not very well conserved overall in most families, though Families 2-7, 25, 27, 30 and 32 all show good degrees of conservation in the internal stretch GDG-N as is typical for most P-type ATPases.

Topological analyses of the families of Proteobacterial P-type ATPases were conducted. Typical results were seen of the functionally characterized enzymes, with Families 1-4 having Type II topology, Copper and Heavy Metal homologues having the Type I topology and Kdp-type K^+ ATPases showing the Type III topology (Figures B5-B10). In addition, FUPA30 proteins showed Type II topology (Figure B15). Families 25, 28, 31 and 32 all exhibit the basic Type I topology (Figures B11, B13, B16 and B17).

Families 27 and 29, on the other hand, have characteristics of Type I families but also have a TMS preceding TMS A (A'), the first of the established TMSs (Figure B12 and Figure B14). This topology represents a novel topology, Type VII. Thus, Family 27 and 29 members of the Proteobacterial analysis have topological Type VII, a type highly similar to Type I.

The Noc1 Ca^{2+} ATPase homologue exhibits a unique 200 residue extension (Figure B5), which through BLAST searches was determined to be a LysM peptidoglycan binding domain (Buist et al, 2008). This appears to be the first report of such a domain in P-type ATPases. Another Ca^{2+} ATPase, Mca2, had a 140 residue N-terminal domain that was found at the N-termini of many other proteins but was not recognized as a conserved domain in the NCBI Conserved Domain Database. This domain may be an unclassified domain of unknown function.

Topological analysis of the Copper ATPases showed extensions at the N-termini that were largely hydrophilic (Figure B8). These proteins, after BLAST searches against TCDB were found to be part of the MerP domain family. MerP plays a role in mercury transport in bacteria (Yamaguchi et al., 2006). These MerP domains exist among a large majority of the proteins of Family 5, and were identified in copies ranging from one to two at the N-termini. Existence of the MerP domain in Copper homologues as well as LysM domain in Ca^{2+} homologues shows the wide diversity of P-type ATPases in the Proteobacterial phylum.

Comparison of the Proteobacterial ATPases to those from the genuses *Bacteroides*, *Flavobacterium* and *Fusobacterium* shows some interesting differences. Firstly, the sheer size of the number of organisms analyzed in the five classes of

protoebacteria, 55, is much larger in comparison to the 11 of *Bacteroides*, *Flavobacterium* and *Fusobacterium* (Figure C1). In addition, there are 218 P-type ATPases of the Proteobacteria versus just 47 of the *Bacteroides*, *Flavobacterium* and *Fusobacterium*. Yet for each organism in both studies, the average number of proteins per organisms was three. *Bacteroides*, *Flavobacterium* and *Fusobacterium* species rarely had more than one member of a particular family per organism while in the Proteobacteria this occurred much more frequently (Figure A2 and Figure B2).

Proteobacterial P-type ATPases had members from Families 1-7 and FUPA proteins of Families 25, 27-32. *Bacteroides*, *Flavobacterium* and *Fusobacterium* had just Families 2, 4-7 and FUPA homologues of families 29, 30, and 32 (Figure C1). Average size of the proteins from the *Bacteroides*, *Flavobacterium* and *Fusobacterium* was 784 residues, while Proteobacterial proteins were larger, on average, by 23 residues (807). Proteobacterial proteins were larger when analyzed according to family than the *Bacteroides*, *Flavobacterium* and *Fusobacterium* except in FUPA32 proteins (Table C1). These enzymes were 678 residues in length in the Proteobacteria but were 735 residues in length in the *Bacteroides*, *Flavobacterium* and *Fusobacterium* (Table C1).

In addition to variations in sizes between families of the two studies, percentages of the total number of proteins were also determined (Figure C1). Ca^{2+} ATPases in *Bacteroides*, *Flavobacterium* and *Fusobacterium* represented a higher percentage of the total enzymes than in Proteobacteria (17% versus 10%). Families 4-7 in both studies had comparable percentages of the total, though Copper ATPases were 32% of the total of Proteobacterial enzymes versus 23% in *Bacteroides*, *Flavobacterium* and *Fusobacterium*. Copper ATPases made up the highest percentage of enzymes in both of the studies,

though it is important to note that in *Bacteroides*, *Flavobacterium* and *Fusobacterium*, the highest percentage of homologues was seen in both Family 5 and Family 6 (11 proteins each, 23%). In general, the FUPA proteins in both studies had small percentages of the total except for FUPA27 in Proteobacteria. FUPA27 homologues had 12% of the total in the Proteobacteria phylum (26 proteins out of 218). The highest percentage of FUPA homologues in *Bacteroides*, *Flavobacterium* and *Fusobacterium* was 6% for FUPA32 (3 proteins out of 47). This fact reveals the differences in P-type ATPase representation in these two phyla (Table C1).

Both studies also reported similar topological results for the FUPA29 homologues (Figure A10 and B14). In each case, the AveHAS plots showed a topological type similar to Type I, though there were peaks in front of the first expected TMS, TMS A. This was true of other proteins similar to FUPA29 which were isolated using BLAST as mentioned previously. The new peak is noted as TMS A'. As a result of this vagueness in the topological classification of these proteins, the novel Type VII was designated to them. Proteobacteria also had Type VII FUPA27 proteins as mentioned previously, with the TMS A' preceding the first known Type I peak, A (Figure B12).

In addition, both studies show through motif analysis that the phosphorylation site on the aspartyl residue located within motif 5 (DKTGTLT) is fully conserved in all of the ATPases examined, suggesting that they are indeed functional enzymes. In general the motifs in *Bacteroides*, *Flavobacterium* and *Fusobacterium* have a higher degree of conservation of motifs within a particular family than those of Proteobacteria (Table A4 and Table A5). Most apparent are the functionally characterized enzymes. Families 2, 4-7 in *Bacteroides*, *Flavobacterium* and *Fusobacterium* have a higher degree of conservation

between the proteins of a particular family than the corresponding members in Proteobacteria. This is in part due by the greater amount of proteins in Proteobacteria (218) versus those in the *Bacteroides*, *Flavobacterium* and *Fusobacterium*. Changes in a few residues in a motif of Proteobacteria can disrupt the conservation of the family. Because of the increase in P-type ATPases, this can occur much more frequently.

Parts of this Thesis are being prepared for publication. The Thesis author along with the committee chairman, Dr. Milton Saier Jr., will be a co-investigators and co-authors of this paper.

Table C1: Comparison of proteins between *Bacteroides, Flavobacteria, Fusobacteria* and Proteobacteria. Total number of proteins of each family is shown as well as the percentage of the total in parentheses. Average size of each of the families is reported as well, expressed as members of amino acyl residues ± standard deviation.

Family	*Bacteroides, Flavobacteria, Fusobacteria* - # and (% of total)	Average size of proteins for *Bacteroides, Flavobacteria, Fusobacteria*	Proteobacteria - # and (% of total)	Average size of proteins for Proteobacteria
Family 1	-	-	2 and (0.9%)	894 ± 3
Family 2	8 and (17%)	886 ± 17	22 and (10%)	915 ± 49
Family 3	-	-	2 and (0.9%)	863 ± 7
Family 4	3 and (6%)	883	18 and (8%)	906 ± 27
Family 5	11 and (23%)	780 ± 49	70 and (32%)	798 ± 70
Family 6	11 and (23%)	635 ± 28	35 and (16%)	736 ± 61
Family 7	8 and (17%)	679 ± 3	25 and (11%)	689 ± 30
FUPA25	-	-	4 and (1.8%)	689 ± 74
FUPA27	-	-	26 and (12%)	785 ± 40
FUPA28	-	-	2 and (0.9%)	849 ± 3
FUPA29	2 and (4%)	794 ± 2	1 and (0. 5%)	798
FUPA30	1 and (2%)	838	4 and (1.8%)	841 ± 13
FUPA31	-	-	3 and (1.3%)	860 ± 198
FUPA32	3 and (6%)	735	4 and (1.8%)	678 ± 49
Total	47	784	218	807

References

Apell, HJ. 2004. How do P-type ATPases transport ions?. Bioelectrochemistry. **63**(1-2):149-56

Axelsen, K., and M. Palmgren. 1997. Evolution of substrate specificities in the P-type ATPase Superfamily. J Mol Evol. **46**:84-101.

Bramkamp, M., K. Altendorf, and J.C. Greie. 2007. Common patterns and unique features of P-type ATPases: a comparative view on the KdpFABC complex from Escherichia coli. Mol Membr Biol. **24(5-6)**:375-86.

Buist, G., A. Steen, J. Kok, and O.P. Kuipers. LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol. **68(4)**:838-47.

Duchaud E., M. Boussaha, V. Loux, J.F. Bernardet, C. Michel, B. Kerouault, S. Mondot, P. Nicolas, R. Bossy, C. Caron, P. Bessieres, J.F. Gibrat, S. Claverol, F. Dumetz, M. Le Henaff, and A. Benmansour. 2007. Complete genome sequence of the fish pathogen Flavobacterium psychrophilum. **25(7)**:763-9.

Fagan, M.J., and M.H. Saier Jr. 1994. P-type ATPases of eukaryotes and bacteria: sequence analyses and construction of phylogenetic trees. J Mol Evol. **38(1)**:57-99.

Gadsby, D. C. 2007. Structural biology: ion pumps made crystal clear. Nature. **450(7172)**:957-9.

Hao, Z., S. Chen, and D.B. Wilson. 1999. Cloning, expression, and characterization of cadmium and manganese uptake genes from *Lactobacillus plantarum*. Appl. Environ. Microbiol. **65**: 4746-4752.

Kühlbrandt, W. 2004. Biology, structure and mechanism of P-type ATPases. Nat Rev Mol Cell Biol. **5(4)**:282-95

Lauf, P.K., T.J. McManus, M. Haas, B. Forbush 3rd, J. Duhm, P.W. Flatman, M.H. Saier Jr, and J.M. Russell. 1987. Physiology and biophysics of chloride and cation cotransport across cell membranes. Fed Proc. **46(7)**:2377-94.

Møller, J.V., B. Juul, and M. le Maire. 1996. Structural organization, ion transport, and energy transduction of P-type ATPases. Biochim Biophys Acta. **1286(1)**:1-51.

Okkeri, J., E. Bencomo, M. Pietilä, and T. Haltia. 2002. Introducing Wilson disease mutations into the zinc-transporting P-type ATPase of Escherichia coli. The mutation P634L in the 'hinge' motif (GDGXNDXP) perturbs the formation of the E2P state. Eur J Biochem. **269(5)**:1579-86.

Paulsen, I. 2002. TransportDB. membranetransport.org.

Thompson, J. D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**: 4876-4882.

Yamaguchi, A., D. Tamang, and M.H. Saier Jr. 2007. Mercury Transport in Bacteria. Water Air Soil Pollut. **182**:219-234.

Zhai, Y., and M.H. Saier Jr. 2001. A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. J Mol Mocrobiol Biotechnol. **3(4)**:501-2.