UNIVERSITY OF CALIFORNIA SAN DIEGO

The Expansion of the Major Facilitator Superfamily (MFS) to Include Novel Transporters, and the Potential Relationship between the MFS and the LysE Superfamily

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

Master of Science

in

Biology

by

Pauldeen Mikail Davejan

Committee in charge:

Professor Milton H. Saier Jr., Chair
Professor Randolph Y. Hampton, Co-Chair
Professor Nigel M. Crawford

2018

The Thesis Pauldeen Mikail Davejan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California San Diego

2018

Dedication

I would first and foremost like to dedicate this thesis to Dr. Saier. You have been a fantastic professor and guide to me throughout these years. I don't think I will meet many individuals as authentically themselves as him. I would also like to thank Arturo Medrano without his help I would have not been able to acquire the expertise that I currently have in bioinformatics. Thank you for being so patient and allowing me to hone my skills in biology and bioinformatics. Thank you to all my friends within the Saier lab for being there as a support system throughout the trying times as well as the good times spent here. I would most importantly like to thank my family and friends back home who have always been a cornerstone in my life, pushing me to pursue a higher education and better myself in my life.

Table of Contents

List of Figures

List of Tables

ABSTRACT OF THE THESIS

The Expansion of the Major Facilitator Superfamily (MFS) to Include Novel Transporters, and the Potential Relationship between the MFS and the LysE Superfamily

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

Master of Science

in

Biology

by

Pauldeen Mikail Davejan

University of California San Diego, 2018

Professor Milton H. Saier Jr., Chair
Professor Randolph Y. Hampton, Co-Chair

The Major Facilitator Superfamily (MFS) [1] is currently the largest characterized superfamily of transmembrane secondary transport proteins [2]. Its diverse members are found in

all organisms in the biosphere and function by uniport, symport, and/or antiport mechanisms. In 2012 we described a total of 74 recognized families, classified phylogenetically within the MFS, all of which included only transport proteins [3]. In this study, we assign several previously uncharacterized transport protein families in the Transporter Classification Database (TCDB; http://www.tcdb.org) to the MFS using established statistical methodologies. In addition, relationships at the superfamily level between the Major Facilitator Superfamily (MFS) and the (LysE) Superfamily have been discovered. Specifically, for the MFS having a common origin with the Resistance to Homoserine/Threonine Family (RhtB; TC# 2.A.76), Tellurium Ion Resistance Family (TerC; TC# 2.A.109), and (17) Nickel/cobalt Transporter (NicO; TC# 2.A.113) is presented. Global alignments and hydropathy plots of transport proteins were generated to assist in determining homology. The use of several sequence analysis programs to search for internal repeats, combined with existing protein structural data, provide strong evidence that the ubiquitous 12 transmembrane segment (TMS) topology arose from a 6 TMS gene duplication which in turn, arose from a 3 TMS duplication. Furthermore analysis of PFAM domains, 3D structures, provided further evidence for homology. Negative control studies were conducted between members of TCDB that are currently not known to be members of the MFS to ensure statistical significance. Positive control studies were conducted between members currently within the MFS.

Introduction

The Transporter Classification (TC #) Database is an IUBMB approved system of classification for recognized and hypothetical transport proteins [4-6]. Using functional and phylogenetic information derived from publications on transport systems, TCDB classifies transport proteins into over 1000 families[6-8]. Ongoing efforts are focused on the identification of distant relationships between transport protein families, allowing categorization of both existing and novel families into superfamilies (see the Superfamily hyperlink in TCDB)[6, 9].

The largest and most diverse superfamily of secondary carriers characterized to date is the MFS. Members of the MFS recognize their substrates stereospecifically and utilize a carrier-mediated process to catalyze transport across biological membranes[10]. MFS secondary carriers transport by (1) uniport, where single molecular entities are transported by facilitated diffusion or by potential driven processes for charged solutes, (2) symport, where two or more solutes are transported in the same direction, driven by chemiosmotic energy, in this case, the electrochemical gradient of protons, called the proton motive force (pmf), or [11] antiport, where two or more solutes are transported in opposite directions, again using chemiosmotic energy to drive the vectorial process. Most members of the MFS share a three dimensional structure that consists of two domains surrounding a central substrate binding site[12]. These transporters operate by an alternating access mechanism where the two halves of the protein move, relative to each other, like a rocker switch, mediated in part by salt bridge formation and breakage during the transport cycle [13].

MFS porters, or permeases, are known to exhibit specificity for sugars, drugs, neurotransmitters, amino acids, organic and inorganic ions, as well as many other ligands,

depending on the specific porter [14].Typical transporters of the MFS are of 400-600 amino acyl residues (aa) in length and with few exceptions, possess either a 12 or 14 trans-membrane α-helical segment (TMS) topology.

Aside from the MFS families listed under TC# 2.A.1, when the studies reported here were initiated, there were 6 additional transport protein families in TCDB with evidence supporting homology with the MFS family. These families and those reported here are characterized within the MFS superfamily in TCDB (see Superfamily hyperlink).

The Glycoside-Pentoside-Hexuronide:Cation Symporter (GPH) Family (TC # 2.A.2), as its name suggests, consists of symporters that catalyze uptake of glycosides in conjunction with a monovalent ion, usually H^+.Most of the functionally characterized proteins of this family are from bacteria, although some homologues are found in archaea and eukaryotes. Members of the GPH family are usually around 500 aas in length and possess the characteristic 12 TMS topology.

The ATP:ADP Antiporter (AAA) Family (TC # 2.A.12) contains transporters that are obligate exchange translocases with specificity frequently for ATP and ADP [15]. They function by taking up ATP into the cell in exchange for ADP, but can also transport inorganic phosphate and other phosphorylated nucleosides [16].These proteins are around 500-600 aas in length with 12 putative TMS and are most commonly found in intracellular pathogens.

The Proton-dependent Oligopeptide Transporter (POT/PTR) Family (TC # 2.A.17) contains members from animals, plants, yeasts, archaea, and bacteria, catalyzing peptide uptake. They are usually 500-600 aas in length and exhibit 12 putative TMS. It has been suggested that pairs of salt bridge interactions between the transmembrane α-helical structures work together to

provide the alternating access transport mechanism [17]. Mammalian members of this transporter family, PepT1 and PepT2, are responsible for the uptake of pharmaceutically important drug molecules such as antibiotics and antiviral agents [17].

The Reduced Folate Carrier (RFC) Family (TC # 2.A.48) includes uptake porters for folates, reduced folates, thiamine, and folate analogues including the anti-cancer drug, methotrexate. Folates, also known as vitamin B9, are essential vitamins for humans, and folate deficiency contributes to a variety of health problems [18].These RFC homologues mediate the intestinal absorption of methotrexate, pralatrexate, and dietary folates [18]. Amino acid replacement experiments have shown that the region between TMS 1 and 2 forms a substrate-binding pocket [19]. Like other MFS family porters, RFC members are typically between 500-600 aas and possess 12 putative or established TMS.

The Organo Anion Transporter (OAT) Family (TC # 2.A.60) contains proteins that catalyze facilitated transport of large amphipathic organic ions such as bromosulfobromophthalein, prostaglandins, bile acids, steroid conjugates, thyroid hormones, oligopeptides, drugs, toxins, and xenobiotics [20, 21]. Human OAT transporters play important roles in drug and metabolite transport across the blood-brain barrier and in the kidneys [22, 23]. These transporters have 12 putative transmembrane segments and are usually 600-700 aas in length. Their evolutionary histories and functional diversification have been examined [23].

Members of the Folate-Biopterin Transporter (FBT) Family (TC # 2.A.71) transport folate and biopterin across the cell membrane and are believed to function by H^+ symport [24]. Most functionally characterized members of the FBT family are from mammals and protozoa, but homologues exist in bacteria, plants, and algae [11].

In this thesis, I assign additional protein families in TCDB to the MFS superfamily. I provide evidence that these proteins all share a common evolutionary origin. In addition to the above described families, previously characterized as part of the MFS, evidence is presented that members of the following families share a common origin with recognized MFS superfamily proteins: (1) the Equilibrative Nucleoside Transporters (ENT; TC# 2.A.57), (2) Ferroportins (Fpn; TC# 2.A.100), [11] Eukaryotic Riboflavin Transporters (E-RFT; TC# 2.A.125). We also have obtained evidence that suggests homology on a superfamily level between the MFS and the LysE superfamily. Members of the latter superfamily include**,** (4) the Resistance to Homoserine/Threonine Family (RhtB; 2.A.76), (5) the Tellurium Ion Resistance Family (TerC; TC# 2.A.109), and (6) the Nickel/cobalt Transporter (NicO; TC# 2.A.113). Sufficient evidence using our statistical methodologies s suggest that members of these families, or domains within these proteins, derive from a common ancestor. The families included in this study are tabulated in Table 1-2.

Methods

Detecting Homology Between Families

We used the program areFamiliesHomologous [25] that automates the three main steps in our strategy to detect remote homology based on the transitivity property of homology [6]. The first step is to run famXpander to retrieve an exhaustive list of homologous proteins for each family based on BLAST [26] searches (E-value < 10, coverage ≥ 50%) against the National Center for Biotechnology Information (NCBI) non-redundant (NR) protein database. Retrieved sequences are then post-processed to remove redundancies using CD-HIT [27] with a redundancy cutoff of 80%. This list of non-redundant candidate homologs is assumed to be representative of the entire family.

Second, the lists of homologues generated with famXpander are then used as input to the program Protocol2 [28] to compare each family against all families in the Major Facilitator Superfamily (MFS). Protocol2 takes the FASTA files from two families and compares them using the Smith-Waterman algorithm as implemented in SSEARCH [29]. For each hit Protocol2 estimates the GSAT [28]) score based on 1000 shuffles, which refers to the number of standard deviations a given global Needleman-Wunsch score, as generated by the program Needle from the EMBOSS suite of programs [30], is from the average score obtained from the shuffled sequences. The result is a report in HTML format, containing the protein accessions, the GSAT score, and the corresponding alignment highlighting the residues in TMSs as predicted by HMMTOP [31]. Based on results from previous research [5, 7, 32] we initially considered only hits with a GSAT score ≥ 15.0. Given that MFS has a repeat unit of 6 TMS we also required that the alignments should be composed of at least 100aa and include a minimum of 5 TMS.

Third, if we label two proteins from different families as A and D, any homolog of A as B and any homolog of C as D, then Protocol2 reports the scores between C-D. For all significant hits reported in the second steps the program calculates the GSAT scores for A-B and C-D. The lowest of the three scores is considered the comparison score for any particular Protocol2 hit. The three scores are given in Table 1-3 but only the B-C comparison scores are reported for in Table 4-5.

HVORDAN, an in-house program, was used to graphically illustrate the results of protocol 2, it displays the hydropathy plots of all proteins in the transitivity path (A-B-C-D), as well as the alignment between B and C. Figures 2,3, and 5-8 illustrate some of our best hits.

Conserved Domains Between Families

All members of our control sets, as well as families inferred to be related to MFS, had their sequences scanned against the Pfam database [33]. The HHMER suite of programs [34] was used to search for domains applying a gathering threshold. We required that proteins A-B and C-D had direct Pfam hits with the characteristic domains of their respective families. To identify additional evidence supporting the relationship suggested by Protocol2 between A and D, we considered a domain as shared between B and C if there was at least 50% overlap —it is not uncommon for Pfam domains to span 2 or more TMS repeats units—of their respective domains in the Protocol2 alignment. The rationale is that if there is a significant similarity between B and C in regions that contain the characteristic domains of their respective families, then the two domains are likely to be related. This is equivalent to projecting the domain of B onto C and vice versa.

Identification of Internal Sequence Repeats

HHrepID [35], AncientRep [28], and the unpublished program tmsRepeat (https://github.com/SaierLaboratory), were used to detect the possible internal repeat unit of any protein family for which the repeat unit is not known. Using based on hidden Markov Models (HMM) HHrepId takes a single protein sequence or a multiple alignment and identifies possible sites where a internal duplications are likely to be. We considered alignments significant if they generated a P-value < 10^{-5}. Several programs are required as prerequisite to AncientRep. These include extractFamily [25], which extracts proteins sequences of specified families from TCDB in several formats, ClustalO [36], and AveHAS [37]. AncientRep receives a multiple alignment files in Clustal format as its input and allows the user to determine in what regions to search for internal repeats using as references the plots generated by AveHAS. GSAT scores are reported for candidate repeat sequences along with hydropathy plots illustrating the repeated regions. All members that are proposed to be additions of MFS show at least a score of 16.1 SD. The program tmsRepeat splits a proteins sequence into bundles of TMS according to the expected repeat unit size, and compares all non-overlapping TMS bundles with the Smith-Waterman algorithm as implemented in SSEARCH [29] (using composition statistics based 10,000 shuffles). Hits with E-value < 10^{-2} were considered significant.

Analysis of Protein 3D Structures

Deuterocol is an in-house unpublished program that was used to conduct the comparison of 3D structures among all families considered this project. The increasing availability of high-resolution 3D structures for multiple transporter families now allows analogous analyses to areFamiliesHomologous using 3D structural alignments. Deutorocol extracts all structures available in PDB for multiple families, cuts them into α-helix bundles (depending on the length

of the known or expected TMS repeat unit shared among families) and superposes the resulting structures. For our purpose we cut the α-helix bundles of 4 TMSs to 6 TMS. Results can be filtered by several criteria including RMSD values, length and coverage of each bundle in the alignments. We considered alignments with RMSD < 3.5 Å, coverage of at least 70% of the bundle, and where the helices aligned made sense in terms of the repeat unit of the families.

Selection of a Negative Control

Transmembrane segments contain low complexity hydrophobic regions that can generate statistically significant alignments, but this does not necessarily entail shared ancestry, as it may be the result of physical-chemical constraints in the membrane environment [38, 39]. As previously reported [25], we tackle this problem by using the GSAT score as a normalizing scale where the scores between known homologous (Positive Control) and non-homologous (Negative Control) proteins can be compared to infer a critical value for the GSAT score that can discriminate between the two sets. Our positive control consists of the 7 families in TCDB currently documented to be part of MFS. The negative control was comprised of a total of 28 families from TCDB that have no known relationship to the MFS and for which the evolutionary path leading to their TMS topology is either not known or not similar to MFS.

Results

Family expansion

All of the protein families within TCDB belonging to subclass 2.A are electrochemical potential-driven uniporters, symporters, and antiporters [6]. Prior to the studies reported here, the MFS included 7 recognized protein families outside of those listed under TC# 2.A.1. As discussed in the introduction, because of the current study, we have added 3 novel families not currently in a superfamily to the MFS and provided preliminary evidence suggesting that 3 families currently in a superfamily (LysE) have a potential relationship with the MFS. It should be made clear that the families within the LysE superfamily do not have nearly as strong relationship to the MFS as the novel families discussed in this thesis. Table 1 identifies all families that are discussed and provides scores expressed in standard deviations (SD) that in conjunction with other types of evidence suggests homology. Protein alignments, internal repeat analyses, domain analysis, 3D structural evidence, and a common evolutionary pathway for the appearance of members of all of these families within the MFS provided evidence suggesting homology. If a family was shown to have strong evidence across at least 3 of these criteria we would consider the family as having a high potential for homology in the MFS. Our strategy is illustrated in the flow chart depicted in figure 1. Analyses involving the ENT, FPN, E-RFT families provide strong statistical evidence for homology. We also have preliminary evidence that suggests members of the LysE superfamily and the MFS have a potential relationship seen in Table 2. The members that we examine here are the RhtB, TerC and NicO families. Beyond these families discussed in this thesis, I have also conducted research on 9 other families that provide evidence to suggest homology with the MFS, their results can be seen in Table 5.

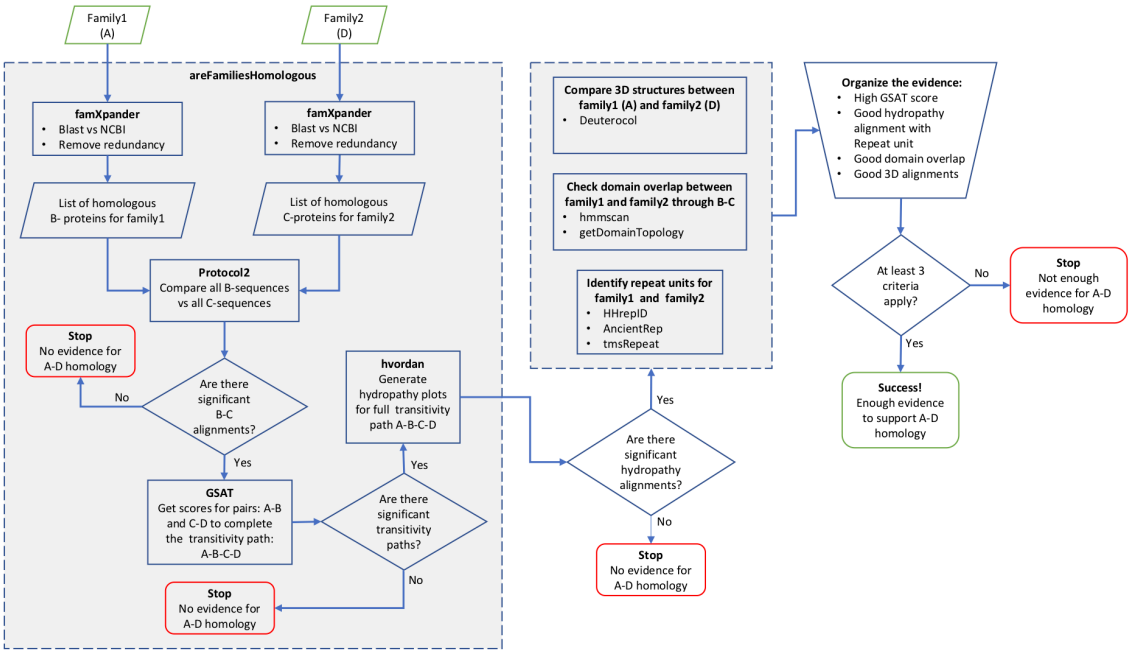

Figure 1. Flowchart of the strategy to infer distant homology between pairs of families

Our strategy is based on the Superfamily principle that takes advantage of the transitivity property of homology. Proteins A and D are considered homologous if it can be shown that there are significant structural properties shared across the path A-B-C-D. Each pair of families is first put through the pipeline areFamiliesHomologous. Proteins A (from Family1) and D (from Family2) are first BLASTed against the NCBI NR database using the program famXpander to get the lists of with candidate homologues Bs and Cs, respectively. Protocol2 then compares the lists of Bs versus Cs to identify significant B-C alignments. The output of this pipeline is the GSAT scores for the pairs A-B, B-C, and C-D, as well as the individual hydropathy plots for each protein including the alignment B-C. If a protein pair B-C had a significant score and good hydropathy alignment, we would proceed to identify the TMS repeat units and Pfam domain topologies for all proteins A-B-C-D. If available, all 3D structures associated with the families of proteins A and D were aligned with the program Deuterocol to identify significant alignments that show congruence with the repeat units for each family. The flow chart lists all programs used; see methods for their descriptions. Four lines of evidence were applied to infer homology between pairs of families: 1) the GSAT score, 2) hydropathy alignments congruent with repeat units, 3) significant overlap of characteristic Pfam domains, and 4) the comparison of 3D structures, if available, between the families of proteins A and D. We considered two families homologous if at least 3 of the 4 criteria were satisfied.

Comparison between Negative and Positive Controls

Twenty eight families with no known relationship to MFS (TC# 2.A.1) were selected for this study. There were some Protocol2 hits with high z-scores between MFS and these families, however upon further examination we saw that these scores were due to the presence of hydrophilic regions in the Protocol2 alignments. Since hydrophilic regions are not useful to study the relationships between transmembrane domains, we removed these regions and recalculated the scores considering only segments that contained TMS. The removal of hydrophilic regions dropped the GSAT score below significant levels. The highest score observed between MFS and the 28 families in the negative control set was 14.3 SD with the Betaine/Carnitine/Choline Transporter Family (TC # 2.A.15), a member of the APC superfamily (Table 3). All other families within the negative control displayed scores below 15 SD. In contrast, the lower score that connects the 7 families in our positive control was 15.1 SD (.8 SD away from 14.3). Therefore we consider any score above 15.1 SD as one evidence of homology between the proposed family and MFS. Throughout the course of our studies all scores contained within the proposed additions have shown to be above 16 SD.

The Equilibrative Nucleoside Transporter (ENT) Family (TC # 2.A.57)

Members of the ENT family are typically 350-500 aas in length and possess 11 putative TMS. ENT family members catalyze nucleoside transport and have homologues in fungi, protozoa, nematodes, and mammals. Members of the human ENT family, SLC29, are known to import drugs used in cancer, AIDS, and parasitic disease treatments [40]. ENT family members have been experimentally shown to have a topology with a cytoplasmic N-terminus and an extracellular C-terminus [41]. Site directed mutagenesis experiments provided evidence for structural commonality and common evolutionary origin between members of the MFS and ENT families, suggesting similar packing of TMSs around a solvent accessible binding site [40].

Comparing TMS 1-11 of the MFS homologue WP_056965629 (12 TMS) with TMS 1-11 of the ENT homologue KVI06040 (11 TMS) gave a comparison score of 18.0 SD (Figure 2). The domain of this MFS homologue WP_056965629 spans across TMS 1-12. The domain of the ENT homologue KVI06040 spans across TMS 4-11. Analysis of the PFAM domains of these proteins shows that nearly 100% of the characteristic domains of both these families are found within their respective Protocol2 alignment.

Another example shows that TMS 4-11 of the MFS homologue WP_025791574 (12 TMS) align well with TMS 4-11 of the ENT homologue XP_013309414 (11 TMS) with a score of 18.0 SD. The domain of the MFS homologue WP_025791574 is seen to span across TMS 1-10 and the domain of the ENT homologue XP_013309414 spans across TMS 4-11. Analysis of the PFAM domains of these proteins indicates that nearly 100% of the characteristic domains of both families are found within their respective Protocol2 alignment.

TMS 1-9 of the MFS homologue KZT53059 (12 TMS) aligned with TMS 1-9 of the ENT homologue XP_004996090 (11 TMS) with a score of 20.2 SD, however TMSs 8 and 9 of both

proteins do not align as well as TMSs 1-7. The domain of the MFS homologue KZT53059 is shown to be across TMS 1-11, and the domain of the ENT homologue XP_004996090 is found to be across TMS 4-11. Analysis of the PFAM domains of these proteins reveal that over 70% of the characteristic domain of the ENT homologue XP_004996090 was found within the Protocol2 alignment. The MFS homologue KZT53059 has over 60% overlap of its characteristic domain within the Protocol2 alignment, which contains one entire repeat unit of the MFS suggesting that these domains are indeed homologous.

The results of AncientRep for the ENT family suggest a repeat unit of 6+6 with the loss the C-Terminus TMS through an alignment of 5 TMS alignments (Table 6). Protocol2 alignments clearly support the loss of the C-terminus TMS (Figure 2). The application of these criteria gives us confidence that these two families are homologous and allows classification of the ENT family within the MFS.

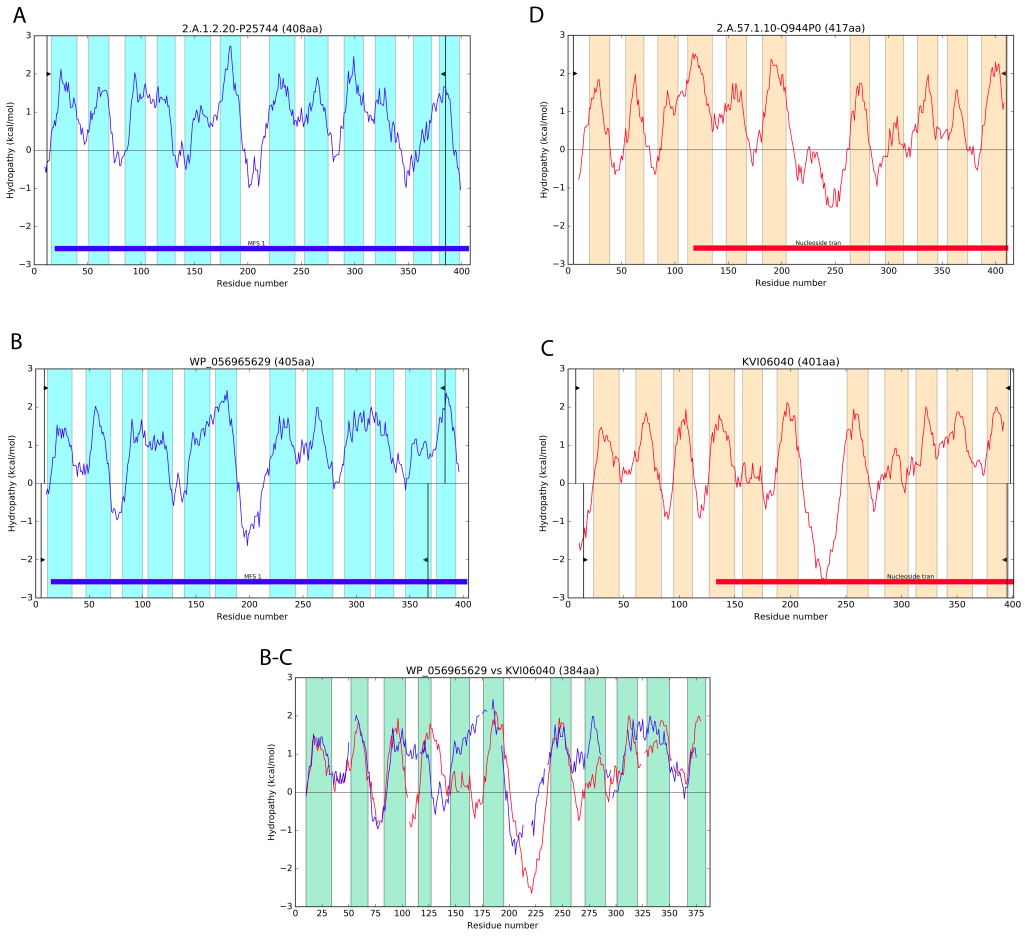

Figure 2. Hydropathy Plots and Alignments between families 2.A.1 and 2.A.57

Graphical output of the hvordan program displaying the hydropathy plots and Pfam domain content for proteins A-B-C-D that correspond to a top scoring Protocol2 alignment B-C. A. Shows the hydropathy plot of the MFS protein A (2.A.1.2.20-P25744). B. Shows the hydropathy plot of the MFS homologue B (WP_056965629). C. Shows the hydropathy plot of the ENT homologue C (KVI6040). D. Shows the hydropathy plot of the ENT protein D (2.A.57.1.10-Q944P0). B-C. Shows the Protocol2 alignment of the MFS homologue (B) and the ENT homologue (C). The blue and tan vertical bars seen in panels A, B, C, and D indicate the locations of the Trans-membrane segments (TMSs).The blue and red horizontal bars in panels A, B, C, and D indicate the location of the respective Pfam domains. The wedges seen in panels A and the top half of panel B indicate the regions of the protein A that aligned with B. The same relationship holds true for panel D and the top half of panel C. The wedges seen in the bottom half of panels B and C indicate the regions of these proteins that are involved in the Protocol2 alignment shown in panel B-C. The green vertical bars seen in panel B-C indicate the areas where the TMSs of panels B and C overlap.

The Ferroportin (Fpn) Family (TC # 2.A.100)

Proteins of the ferroportin family are required for the export of iron and manganese into the systemic circulation [42, 43]. These iron regulated transport proteins are found in the basolateral membranes of mammalian intestinal epithelial cells [44]. Ferroportins are essential for iron homeostasis; studies have shown that mice lacking these proteins die during embryonic development [45]. Members of the Fpn family are between 400 and 800 aas in length, and studies with antisera have suggested a topology of 11 TMSs, with the C-termini exposed on the cell surface [42].

Comparing TMSs 1-11 of the MFS homologue KYC90238 (12 TMS) with TMSs 1-11 of the Fpn homologue KEQ60961 protein (12 TMS) gave a comparison score of 16.5 SD (Figure 3). The domain of the MFS homologue KYC90238 spans across TMS 1-12 and the domain of the Fpn homologue KEQ60961 spans through TMS 1-12. The Pfam domains of both the Fpn homologue and the MFS homologue are contained within their Protocol2 alignments.

In another case we see TMSs 1-9 of the MFS homologue WP_046324300 (12 TMS) with TMSs 1-9 of Fpn homologue XP_002638873 (12 TMS) yielded a comparison score of 19.3 SD. The domain of the MFS homologue WP_046324300 spans across TMS 1-12 and the domain of the Fpn homologue XP_002638873 spans across TMS 1-11. In addition, this Fpn homologue also hits the characteristic MFS domain (Pfam Id PF7690) between TMS 1-9. There is nearly 70% of the domain of the Fpn homologue, but nearly 100% of its MFS domain within the protocol 2 alignment. The MFS homologue includes over 70% of its characteristic domain within the protocol 2 alignment.

Examining TMSs 2-9 of the MFS homologue KPL72370 (12 TMS) with TMSs 2-9 of the Fpn homologue CDJ86150 (12 TMS) gave a comparison scored of 18.6 SD. We see that the

MFS homologue KPL72370 spans its domain across TMS 1-11 and the Fpn homologue CDJ86150 spans its domain across TMS 1-11. The Protocol2 alignment contains over 70% of the domain of the MFS homologue and over 70% of the domain of the Fpn homologue.

Given that structures are available for both families, we performed comparisons of sets of 4-6 alpha-helix bundles to cover more than half of one 6-TMS repeat unit. Figure 4 shows our best result where the alignment of alpha-helices 3-6 of the MFS member 2.A.1.1.3 (PDB: 4QIQ) with alpha-helices 3-6 of the FPN member 2.A.100.2.1 (PDB: 5AYM) yields an RMSD value of 1.88 Å across 104 amino acyl residues. This alignment is in agreement with the basic repeat unit of MFS where 6 TMS arose from a duplication of 3 TMS. Le Gac et al. [46] constructed a 3D model of Ferroportins by homology to an MFS crystal structure (EmrD), successfully predicting critical amino acids with a generated ferroportin model.

Unfortunately we could not identify a repeat unit based on protein sequences for Fpn using the programs AncientRep, HHrepID and tmsRepeat as detailed in methods.

Altogether, these results provide enough evidence for homology between Fpn proteins and the MFS superfamily, and indicate that in the Fpn family, the C-terminal TMS was lost relative to the typical 12 TMS MFS topology to generate the characteristic 11 TMS Fpn.

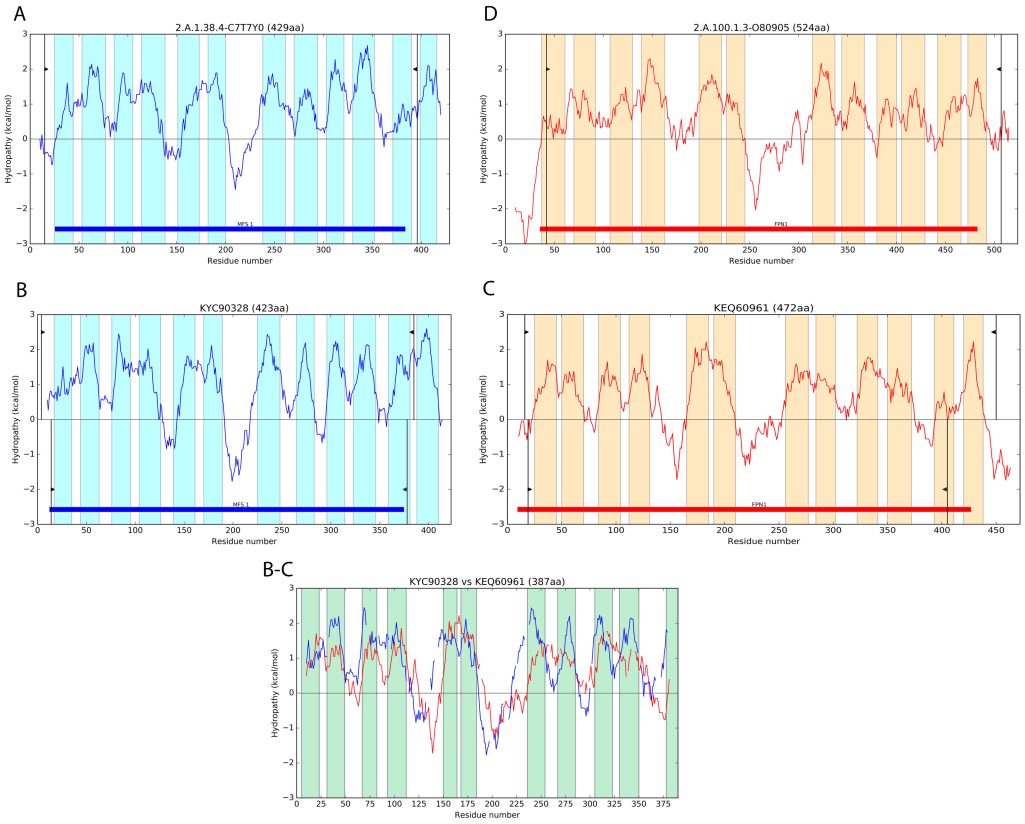

Figure 3. Hydropathy Plots and Alignments between families 2.A.1 and 2.A.100

Representation of the output of the hvordan program to graphically display top scoring candidate homologs. **A.** Shows the hydropathy plot of the MFS protein A (2.A.1.38.4-7T7Y0). **B.** Shows the hydropathy plot of the MFS homologue B (KYC90328). **C.** Shows the hydropathy plot of the Fpn homologue C (KEQ60961). **D.** Shows the hydropathy plot of the Fpn protein D (2.A.100.1.3-O80905). **B-C.** Shows the Protocol2 alignment of the MFS homologue (B) and the Fpn homologue (C). For description of the regions delimiting the wedges as well as all bars see description of figure 2.

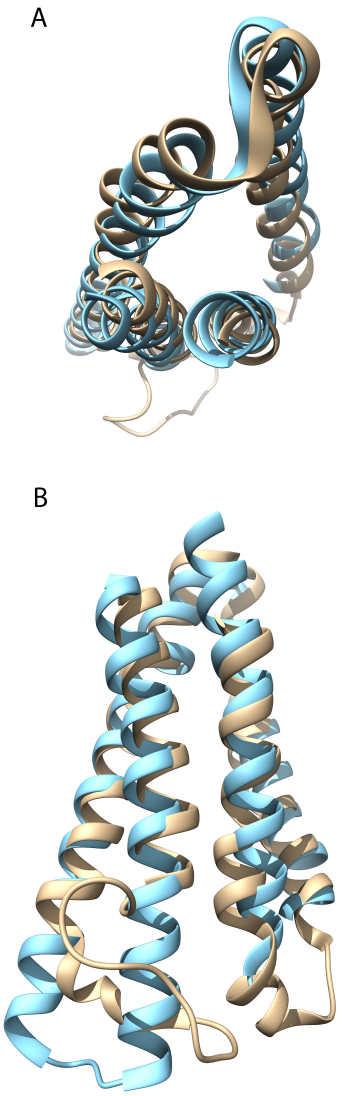

Figure 4. Structural alignment between families 2.A.1 and 2.A.100

Example 3D structural alignment of two 4-helix bundles generated by the program deuterocol. This alignment is between α-helices 3-6 of protein 2.A.1.1.3 (PDB: 4QIQ; tan color) and α-helices 3-6 of protein 2.A.100.2.1 (PDB: 5AYM; cyan color). The alignment is 104 amino-acyl residues long with a root-mean square deviation (RMSD) value of 1.88 Å. A. top down view of the alignment where the top of panel B is rotated along the X axis towards the observer. B. frontal view of the alignment.

The Eukaryotic Riboflavin Transporter (E-RFT) Family (TC # 2.A.125)

Members of the E-RFT family are typically of 450 amino acyl residues in length and possess 11 putative TMS. As their name suggests, the E-RFT family transports riboflavin (vitamin B2), which is critical for cellular redox functions. Riboflavin in the forms of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) act as cofactors in biological oxidation-reduction reactions [47]. Deficiencies in riboflavin can lead to developmental abnormalities in adolescence and is a risk factor for anemia, cancer, and cardiovascular disease [48]. Studies performed on the rat riboflavin transporter 2 (rRFT2) showed that it is inhibited by the presence of lumiflavin, FMN, and FAD [49]. These findings suggest that riboflavin transporters are able to transport various riboflavin derivatives.

In our first example it should be noted that the ENT family has been shown to be homologous to the MFS based on our previous results. Comparing TMS 1-8 of the ENT homologue, KUI64529 (11 TMS), with TMS 1-8 of the E-RFT homologue, XP_008068176 (12 TMS), yielded a comparison score of 19.3 SD (Figure 5). We see the ENT homologue KUI64529 has its domain span across TMS 4-11. The E-RFT homologue XP_008068176 shows to have its domain spanning across TMS 7-9. The Protocol2 results shows that nearly 70% of the domain of the ENT homologue is aligned and over 70% of the domain of the E-RFT homologue within the alignment. Since we have evidence for homology between the ENT family and MFS, and E-RFT is suggested to be related to ENT, by the superfamily principle we can suggest homology between MFS and E-RFT.

Examining TMSs 6-11 of MFS homologue WP_049859152 (12 TMS) we observed an alignment with TMS 6-11 of E-RFT homologue XP_798894 (11 TMS) yielding a comparison score of 17.2 SD. The domain of the MFS homologue WP_049859152 spans throughout TMSs

1-12. The domain of the E-RFT homolog XP_798894 spans across TMSs 7-9. The MFS homolog in this case contains nearly 70% of its domain within the Protocol2 alignment. This alignment includes the last TMS of its first repeat domain, and the first 5 TMSs of its second repeat domain. The E-RFT homologue has nearly 100% of its domain within the Protocol2 alignment.

TMSs 3-9 of the MFS homologue KIL83779 with TMSs 3-9 of the E-RFT homologue XP_002670990 showed to have comparison scores of 16.4 SD. The MFS homologue KIL83779 has its domain spanning across TMSs 1-12. The domain of the E-RFT homologue XP_002670990 spans across TMSs 7-9. The Protocol2 results show that over 50% of the domain of the MFS homologue is within the alignment, and 100% of the domain of the E-RFT homologue is within the alignment.

While we do have evidence suggesting homology between the MFS and E-RFT through a direct relationship, by the superfamily principle [50] we also have evidence suggesting homology between MFS and E-RFT through a relationship with the ENT family. The results also suggest that the 11 TMS E-RFT family member topology could have arisen from a loss of the C-Terminus TMS of an original 12 TMS topology. This conclusion can be reached when examining the Protocol2 alignment between the ENT homologue and the E-RFT homologue, where the N-terminus aligns between the two proteins while C-terminus does not. While the repeat unit of this family has yet to be well established we have evidence that it consists of a 3+3 topology, which is consistent with MFS. This was observed with both our programs AncientRep, and TMSRepeat where TMS 4-6 align with 7-9 in the same protein 2.A.125.1.5. Overall our data suggests that family E-RFT is part of MFS.

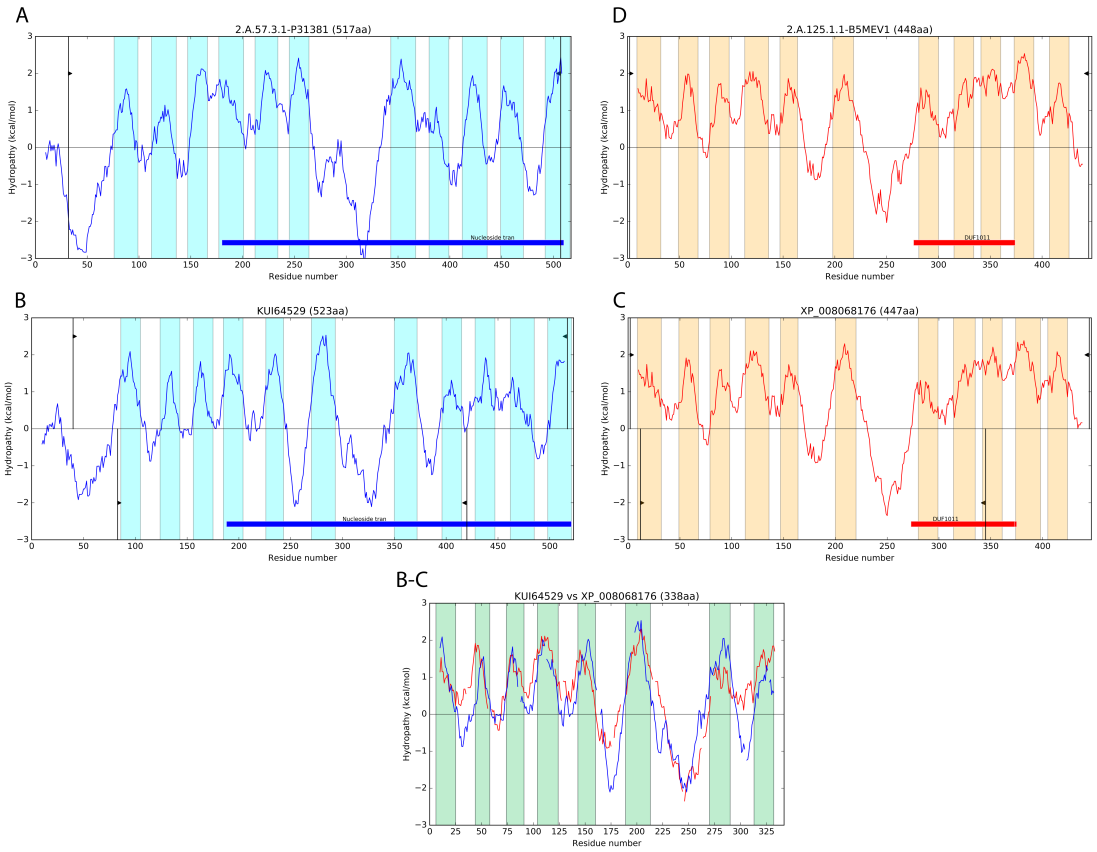

Figure 5. Hydropathy Plots and Alignments between families 2.A.57 and 2.A.125

Representation of the output of the hvordan program to graphically display top scoring candidate homologs. **A.** Shows the hydropathy plot of the ENT protein A (2.A.57.3.1-p31381). **B.** Shows the hydropathy plot of the ENT homologue B (KUI64529). **C.** Shows the hydropathy plot of the E-RFT homologue C (XP_008068176). **D.** Shows the hydropathy plot of the E-RFT protein D (2.A.125.1.1-B5MEV1). **B-C.** Shows the Protocol2 alignment of the ENT homologue (B) and the E-RFT homologue (C). For description of the regions delimiting the wedges as well as all bars see description of figure 2.

The Potential Relationship between the LysE Superfamily and the MFS

The Tellurium Ion Resistance (TerC) Family (TC # 2.A.109)

The TerC family, a member of the LysE superfamily is thought to function in tellurium ion resistance [51]. It has been documented that members of the TerC family are found in a wide range of organisms including eukaryotes, prokaryotes, and archaea [52]. There are also findings that show the TerC family to be involved in a metal sensing stress response system demonstrated in bacteria as well. It is well known that this family typically has a 7 TMS core but some members of this family are seen to have up to 9 TMSs. Members of this family tend to be between 180 and 350 residues in length.

Comparing TMS 4-9 of the MFS homologue WP_019618799 (12 TMS) with TMS 1-6 of the TerC homologue WP_049319872 (7 TMS) showed to have a comparison score of 22.3 SD (Figure 6). The domain of the MFS homologue WP_019618799 spans across TMSs 1-11. The domain of the TerC homologue WP_049319872 spans across TMSs 1-6. We observe that over 50% of the characteristic MFS domain is overlapping in the Protocol2 alignment. The TerC homologue on the other hand shows to have nearly 100% of its domain overlapping in the the Protocol2 alignment.

Comparing TMS 4-9 of the MFS homologue ERL43808 (12 TMS) with TMS 1-6 of the TerC homologue WP_013626834 we see a comparison score of 17.4 SD. The domain of this MFS homologue shows to span from TMS 5-12 of the protein. The TerC homologue on other hand shows to have its domain spanning across TMS 1-6. We see that there is 100% overlap of the domain of the TerC homologue in the Protocol2 alignment, while the pfam results for the MFS homologue demonstrate a domain overlap above 50% of the protein.

Comparison of TMS 4-10 of the MFS homologue WP_051266504 (12 TMS) to TMS 1-7 of the TerC homologue AMA61886 (7 TMS) yielded comparison score of 16.1 SD. We see that the domain of the MFS homologue WP_051266504 spans across TMS 1-11 while the domain of the TerC homologue AMA61886 spans across TMS 1-6. The overlap of the MFS homologue in the Protocol2 alignment appears to be over 50% overlap while the TerC homologue shows to have 100% overlap of its domain within the protocol 2 alignment.

AncientRep supports the previously documented TMS repeat unit of 3+3 for the TerC family [53] (Table 6). Overall the evidence we present here suggests a relationship between the TerC family and MFS, but it is not enough to claim homology. Even if TerC has the 3+3 topology, it may have undergone rearrangements of the 3 TMS precursor repeat unit relative to the MFS. Further studies must be conducted to increase the degree of confidence for homology between these two families. The other LysE superfamily members that are discussed (LysE, and NicO) in this thesis show the exact same behaviour as this family, suggesting that they all must undergo further studies.

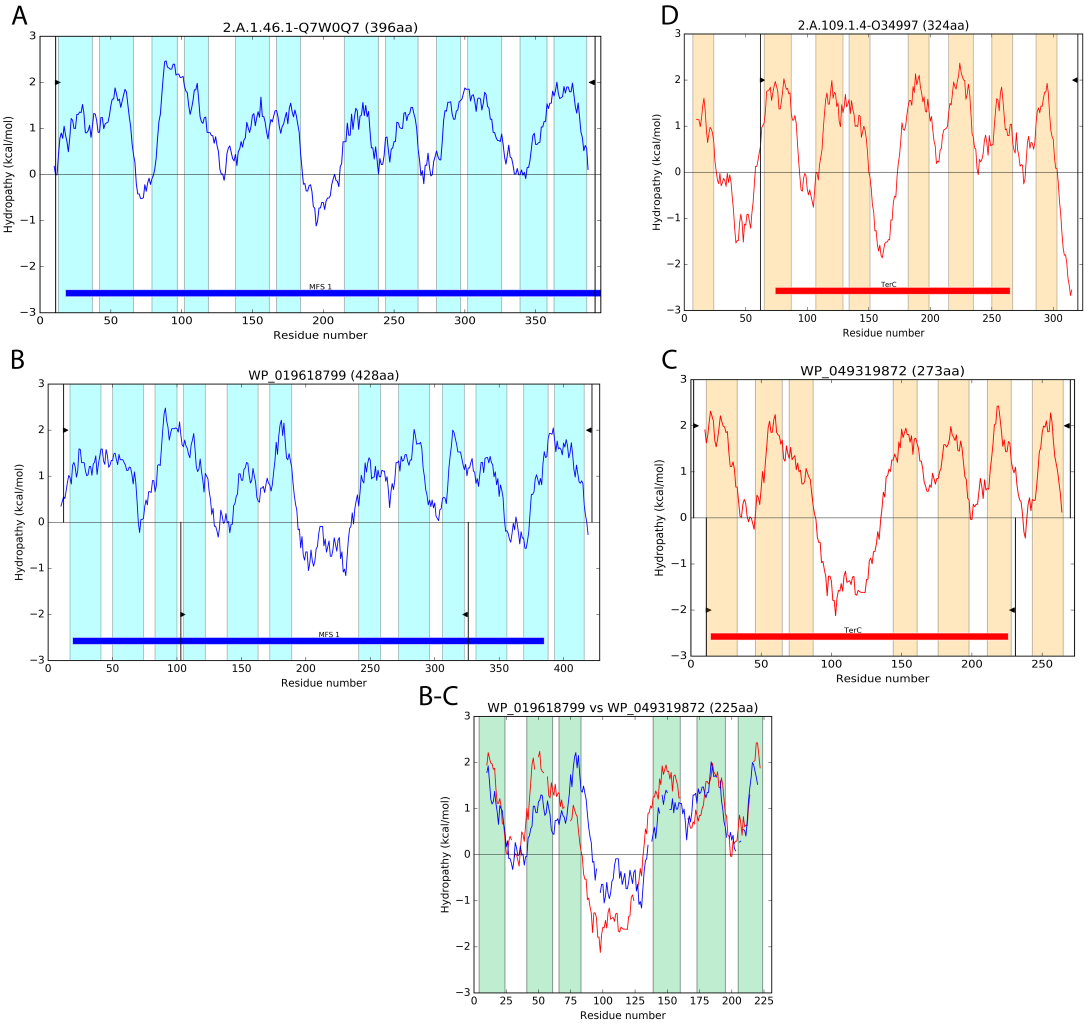

Figure 6. Hydropathy Plots and Alignments between families 2.A.1 and 2.A.109

Representation of the output of the hvordan program to graphically display top scoring candidate homologs. **A.** Shows the hydropathy plot of the MFS protein A (2.A.1.46.1-Q7W0Q7). **B.** Shows the hydropathy plot of the MFS homologue B WP_019618799). **C.** Shows the hydropathy plot of the TerC homologue C (WP_049319872). **D.** Shows the hydropathy plot of the TerC protein D (2.A.109.1.4-O34997). **B-C.** Shows the Protocol2 alignment of the MFS homologue (B) and the TerC homologue (C). For description of the regions delimiting the wedges as well as all bars see description of figure 2.

The Resistance to Homoserine/Threonine (RhtB) Family (RhtB); 2.A.76

Proteins of the RhtB family are known to be be amino acid efflux porters. While the first members of this family were known to be Homoserine and Threonine exporters, there a variety of other amino acid exporters within this family such as leucine exporters, cysteine, O-acetylserine, and azaserine exporters as annotated by TCDB. This family is seen to have homologues in archaea, as well as both Gram-negative and Gram-positive bacteria. Some proteins of this family display controlled expression under the influence of the leucine-responsive protein (Lrp) [54]. Given the involvement of the family in the export of small molecules, there have been suggestions that these exporters play a role as mediators in the secretion of signaling molecules or exist to avoid the accumulation of substrate compounds to toxic levels .[4, 42, 55]. Findings have been documented that pimT, a member of this family, is involved in the export of the quorum-sensing pimaricin inducer PI-factor (2,3-diamino-2,3-bis(hydroxymethyl)-1,4-butanediol) [56], a positively charged molecule that acts as an autoinducer in *S. natalensis*. Members of the RhtB family are proteins typically seen to be 190-240 amino acids in length. They also display a membrane topology 6 TMSs across nearly all members of the family.

Examining TMS 4-8 of the MFS homologue WP_018560854 (12 TMSs) we see that is overlaps with TMS 1-5 of the RhtB homologue WP_061785927 (6 TMS) and generates an alignment score of 16.3 SD (figure 7). The domain of the MFS homologue WP_018560854 spans across TMSs 1-8 and the domain of the LysE homologue spans across TMS 1-6. Furthermore we observe that the domain of the RhtB homologue WP_061785927 overlaps over 70% of its domain in the Protocol2 alignment while the MFS homologue WP_018560854

overlaps over 50% of its domain in this same alignment. We observe this overlap to be with half of the 2 MFS domains.

We also see TMS 5-9 of the MFS homologue WP_051569412 (12 TMS) TMS 2-6 of the RhtB homologue WP_064516356 (6 TMS) with a comparison scored of 16.3 SD. The domain of this MFS homolgoue WP_051569412 spans across TMS 1-11 and the domain of the RhtB homologue WP_064516356 spans across TMS 1-6. The domain overlap shows to be over 70% overlap of the RhtB homologue domain in the Protocol2 alignment the MFS homologue domain on the other hand has below 50% overlap of in the Protocol2 alignment. This overlap of the MFS domains spans across roughly half of one of the MFS domains.

Comparing TMS 6-9 of the MFS homologue WP_049366914 (13 TMS) with TMS 2-5 of the RhtB homologue WP_027578727 we see an alignment score of 18.0 SD. The MFS homologue WP_049366914 appears to have an N-terminal insertion suggested by its alignment with the MFS protein 2.A.1.20.4 (12 TMS). The alignment between these two spans across 12 TMSs, but the MFS homolog WP_049366914 only aligns TMS 2-13 and not TMS #1. This indicates that TMS #1 is not characteristic of MFS. The MFS homolog has its domain span across TMSs 2-11 this further confirms that TMS 1 is not characteristic of MFS. The RhtB homologue WP_027578727 has its domain span across TMSs 1-6. We see that the domain overlap of the RhtB homologue in this particular case to be over 70% overlap in the protocol 2 alignment. In the case of the MFS homologue this protein's domain overlap is nearly 50% of the proteins domain however the alignment is across 2 halves of the MFS domain.

While there is previous documentation suggesting a 3+3 repeat unit [53] for the RhtB family our studies with AncientRep (Table 6) provide evidence to further solidify this claim. The evolutionary pathway of the RhtB family shows to be in line with the evolutionary pathway of

MFS where 12 TMSs originated from the duplication of 3 TMSs to 6TMSs and the duplication of 6TMSs to 12 TMSs total. Therefore both MFS and the RhtB family show originated from a 3+3 TMS repeat unit.

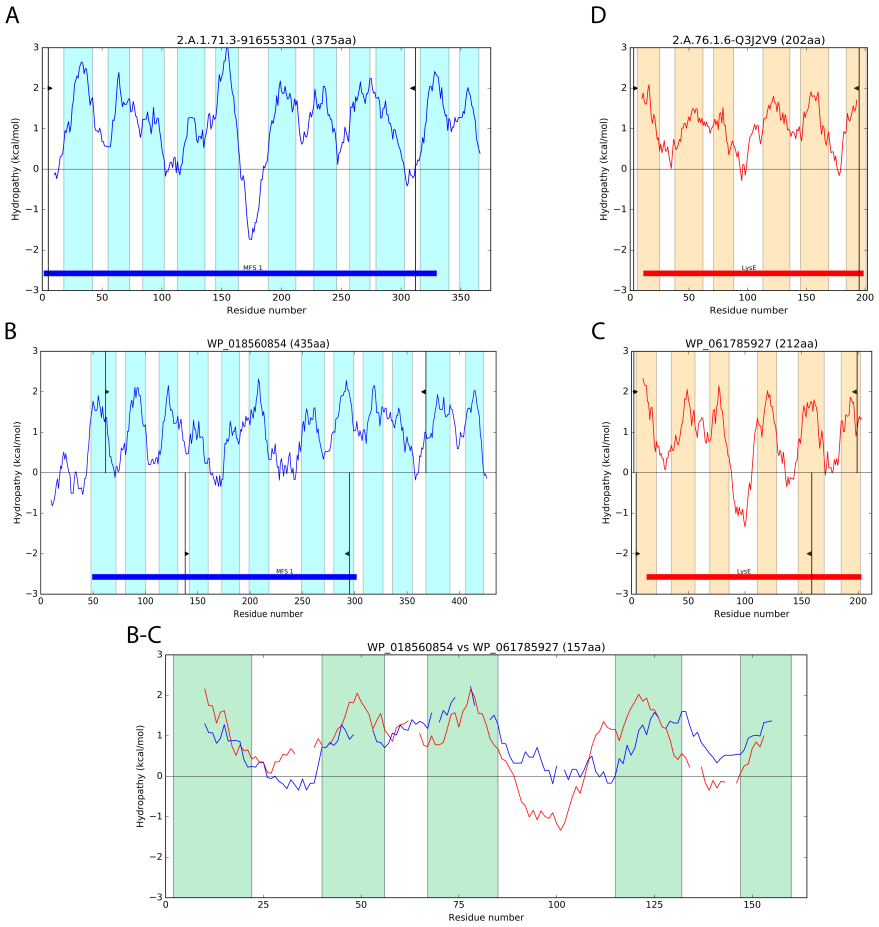

Figure 7. Hydropathy Plots and Alignments between families 2.A.1 and 2.A.76

Representation of the output of the hvordan program to graphically display top scoring candidate homologs. **A.** Shows the hydropathy plot of the MFS protein A (2.A.1.71.3-916553301). **B.** Shows the hydropathy plot of the MFS homologue B (WP_018560854). **C.** Shows the hydropathy plot of the LysE homologue C (WP_061785927). **D.** Shows the hydropathy plot of the LysE protein D (2.A.76.1.6-Q3J2V9). **B-C.** Shows the Protocol2 alignment of the MFS homologue (B) and the TerC homologue (C). For description of the regions delimiting the wedges as well as all bars see description of figure 2.

The Nickel/cobalt Transporter (NicO) Family (TC # 2.A.113)

While we do have evidence that suggests a relationship between this family and MFS these data suggest that the relationship between this family and the MFS are not as strong as the relationship between the other LysE superfamily members mentioned above. There has been preliminary research on the NicO family showing that they play a role in Nickel and Cobalt export in bacteria [57]. This transporters purpose is thought to assist in metal resistance. Members of this family are found throughout prokaryotes, eukaryotes, and Archea. While members of this family have a range of 6-8 TMSs they are comprised of a 6 TMS core and are thought to have a 3+3 repeat unit [53]. Members of the NicO family are typically comprised of 214-597 amino-acyl residues with most proteins being in the approximate range for 214-320 amino acids in length.

Comparison TMS 4-8 of the MFS homologue WP_022710359 (12 TMS) with TMS 1-5 the NicO homologue WP_052486258 (6 TMS) yields a comparison score of 16.1 SD. The domain of this MFS homologue spans across TMS 1-11 in this protein while the domain of the NicO homologue spans across TMS 1-6. Overall we see that over 70% of the domain of the NicO homologue is found in the Protocol2 alignment (Figure 8). The domain of the MFS homologue in this case shows over 50% overlap of its domain in the Protocol2 alignment. Comparison between TMS 5-8 of the MFS homologue AKH41752 (12 TMS) and TMSs 2-5 of the NicO homologue WP_057850321 (6 TMS) yielded a comparison score of 17.0 SD. The MFS homologue AKH41752 has its domain span across TMSs 1-12. The Pfam database is unable to recognize a domain for the NicO homologue, however if we project the domain of the NicO protein (TC # 2.A.113.1.5) onto the NicO homologue WP_052486258 based on the alignment between the two proteins we see that there TMS 1-5 overlap with TMS 1-5. This projection

suggests that the NicO homologue WP_052486258 has the NicO domain at least across TMSs 1-5**.** The NicO homologue shows over 70% overlap of its characteristic domain in the Protocol2 alignment. Conversely the domain of this MFS homologue shows a domain overlap nearly 50% in the Protocol2 alignment.

Comparison between TMSs 3-8 of the MFS homologue KUE99291 (12 TMS) with the NicO homologue WP_038012784 (6 TMS) yielded a comparison score of 17.2 SD.The MFS characteristic domain of this MFS homologue KUE99291 spans across TMS 1-11. The domains of the NicO homologue WP_038012784 spans across TMS 1-6. Overall we see 70% overlap of the NicO homologue domain within the Protocol2 alignment while the MFS homologue shows a domain overlap of over 50% within the Protocol 2 alignment.

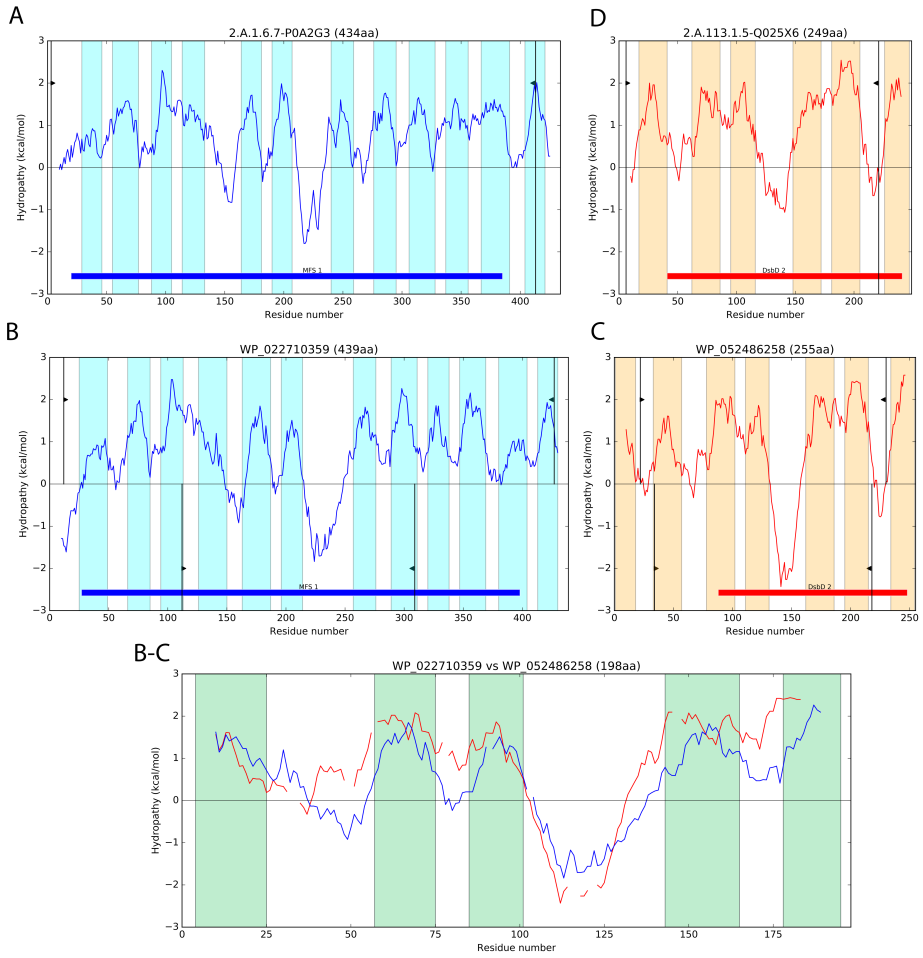

Figure 8. Hydropathy Plots and Alignments between families 2.A.1 and 2.A.113

Representation of the output of the hvordan program to graphically display top scoring candidate homologs. **A.** Shows the hydropathy plot of the MFS protein A (2.A.1.6.7-P0A2G3). **B.** Shows the hydropathy plot of the MFS homologue B (WP_018560854). **C.** Shows the hydropathy plot of the NicO homologue C (WP_052486258). **D.** Shows the hydropathy plot of the NicO protein D (2.A.113.1.5-Q025X6). **B-C.** Shows the Protocol2 alignment of the MFS homologue (B) and the NicO homologue (C). For description of the regions delimiting the wedges as well as all bars see description of figure 2

.

Discussion

In this study, we have expanded the largest family of secondary carriers currently recognized, the Major Facilitator Superfamily, MFS. Prior to this study, nearly 100 families were known to comprise the MFS. However, with the exception of a few receptors, virtually all of these proteins were known or assumed to be transport proteins. Our present efforts have allowed us to include several additional transport protein families, members of which are believed to function as secondary carriers. These include the Equilibrative Nucleoside Transporters (ENT; TC# 2.A.57) with 11 TMS, the Ferroportins (Fpn; TC# 2.A.100) with 11 TMS, and the Eukaryotic Riboflavin Transporters (E-RFT; TC# 2.A.125) with 11 TMS. All of the proteins that comprise these families were shown to have the basic 6 TMS repeat unit, giving rise to 12 TMS, although the predominant members of these three families appear to have lost a single TMS at their N- or C-termini.

Interestingly, we have found evidence that suggests a relationship between superfamilies; in this case between the MFS and the LysE superfamilyles. The RhtB, TerC and NicO families of the LysE superfamily show significant alignments that span 6 TMSs with members of MFS. However, these alignments cover the second and first halves of the two MFS 6 TMS repeat units. This is intriguing but not enough to support homology. The LysE families may share with MFS the basic construction of their repeat unit (3+3), but it is possible that they have undergone a rearrangement of the basic precursor 3 TMS repeat unit. As more 3D structures and protein sequences become available in public repositories, we will be able to study and greater detail the relationship between these two families.

The negative control provided a good reference to detect families homologous to MFS. The highest z-score seen within the negative control was 14.3 SD. This contrast well with the

lowest score of 16.4 SD obtained for the new families added to MFS. Similarly, the lowest GSAT score observed for the 3 LysE families was 16.1 SD, clearly higher than the scores observed for the negative control. Nevertheless, the GSAT score alone is not enough to claim homology. The four criteria used in this project: 1) GSAT score, 2) repeat unit congruence with good hydropathy alignments, 3) Pfam domain overlap, and 4) 3D structure alignments (when available), clearly make the inference of homology more reliable. At least two additional criteria can be incorporated in the detection of homology, that is, common sequence motifs between families and conservation of functional residues. However, for the purpose of this projects we considered the first 4 criteria strong enough to reliably infer homology.

While much is understood about the function of MFS, there is still a plethora of members that have yet to be fully understood. Over the past years there has been a significant amount of 3D structural data that has become available for members of MFS, however there are still many mechanistic and functional questions that remain unanswered. Functional predictions through conservation of genome context would allow better characterization of these families and improve our understanding of how members of these families might be related based on additional functional characteristics.

Table 1.

The proposed additions to the MFS that have been discussed throughout this thesis. All proteins that make up the relationship, their respective GSAT scores, and alignments have been noted below. It should also be noted that all members below had PFAM domains which aligned at least 5 of their 6 TMS with the characteristic MFS repeat unit. Boxes in grey beneath the comparison score column indicate the lowest comparison score between A-B, B-C, and C-D. The lowest score of this data set is 16.4 SD.

Family	Protein A	Protein B	Protein C	Protein D	Comparison score (SD)			TMS alignment
					A-B	B-C	C-D	
2.A.57 The Equilibrative Nucleoside Transporter (ENT) Family	2.A.57.1.1-Q99808	XP_004996090	KZT53059	2.A.1.14.17-O43000	60.5	20.2	52.2	1-7 w/ 1-7 (p8,9 w/ p8,9)
2.A.57 The Equilibrative Nucleoside Transporter (ENT) Family	2.A.57.1.10-Q944P0	KVI06040	WP_056965629	2.A.1.2.20-P25744	48.9	18	79.1	1-11 w/ 1-11 suggests c terminal deletion
2.A.57 The Equilibrative Nucleoside Transporter (ENT) Family	2.A.57.1.6-Q9BZD2	XP_013309414	WP_025791574	2.A.1.57.4-Q1D1G6	45.2	17.96	26.8	4-11 w/ 4-11 suggests c terminal deletion
2.A.100 The Ferroportin (Fpn) Family	2.A.100.1.4-Q9NP59	XP_002638873	WP_046324300	2.A.1.21.21-H5X1B8	57.9	20.5	32.5	1-9 w/ 1-9
2.A.100 The Ferroportin (Fpn) Family	2.A.100.1.1-Q9JHI9	CDJ86150	KPL72370	2.A.1.21.23-Q1IWX2	62.9	18.6	39.1	2-9 w/ 2-9
2.A.100 The Ferroportin (Fpn) Family	2.A.100.1.3-O8O905	KEQ60961	KYC90328	2.A.1.38.4-C7T7Y0	53.3	16.5	39	1-11 w/ 1-11

Table 1. *(continued)*

Family	Protein A	Protein B	Protein C	Protein D	Comparison score (SD)			TMS alignment
					A-B	B-C	C-D	
2.A.125 The Eukaryotic Riboflavin Transporter (E-RFT) Family	2.A.125.1.1-B5MEV1	XP_008068176	KUI64529	2.A.57.3.1-P31381	156.6	19.3	69	1-8 w/ 1-8
2.A.125 The Eukaryotic Riboflavin Transporter (E-RFT) Family	2.A.125.1.2-Q9NQ40	XP_798894	WP_049859152	2.A.1.1.26-O34718	83.3	17.2	104	6-11 w/ 6-11
2.A.125 The Eukaryotic Riboflavin Transporter (E-RFT) Family	2.A.125.1.5-D2VY17	XP_002670990	KIL83779	2.A.1.53.4-M1V8F4	422	16.4	29.8	3-9 w/ 3-9

Table 2.
The LysE superfamily members which show preliminary evidence suggesting a relationship between the LysE superfamily and the MFS. All proteins that make up the relationship, their respective GSAT scores, and alignments are seen below. It should also be noted that all members below had PFAM domains which aligned at least 5 of their 6 TMSs with the characteristic MFS repeat unit. Boxes in grey beneath the comparison score column indicate the lowest comparison score between A-B, B-C, and C-D. Lowest score of this data set is 16.1 SD.

Family	Protein A	Protein B	Protein C	Protein D	Comparison score (SD)			TMS alignment
					A-B	B-C	C-D	
2.A.76 The Resistance to Homoserine/Threonine (RhtB) Family	2.A.76.1.9-C6GYU4	WP_027578727	WP_049366914	2.A.1.20.4-Q9S3K0	33.2	18	116.8	2-5 w/ 6-9, 13 TMS MFS
2.A.76 The Resistance to Homoserine/Threonine (RhtB) Family	2.A.76.1.6-Q3J2V9	WP_061785927	WP_018560854	2.A.1.71.-916553301	18.9	16.3	29.1	1-5 w/ 4-8
2.A.76 The Resistance to Homoserine/Threonine (RhtB) Family	2.A.76.1.2-P0AG38	WP_064516356	WP_051569412	2.A.1.2.69-O34546	33.4	16.3	18.5	2-6 w/ 5-9
2.A.109 The Tellurium Ion Resistance (TerC) Family	2.A.109.1.4-O34997	WP_049319872	WP_019618799	2.A.1.46.1-Q7W0Q7	76.5	21.8	46.8	1-6 w/ 4-9
2.A.109 The Tellurium Ion Resistance (TerC) Family	2.A.109.2.2-Q7URC1	WP_013626834	ERL43808	2.A.1.24.4-Q1CWQ3	54.3	17.4	63	1-6 w/ 4-9
2.A.109 The Tellurium Ion Resistance (TerC) Family	2.A.109.1.4-O34997	AMA61886	WP_051266504	2.A.1.46.2-Q83N16	81.9	16.1	55.6	1-7 w/ 4-10

Table 2. *(continued)*

Family	Protein A	Protein B	Protein C	Protein D	Comparison score (SD)			TMS alignment
					A-B	B-C	C-D	
2.A.113 The Nickel/Cobalt Transporter (NicO) Family	2.A.113.1.9-F8C138	WP_057850321	AKH41752	2.A.1.1.92-P38055	30.3	16.95	33.7	2-5 w/ 5-8
2.A.113 The Nickel/Cobalt Transporter (NicO) Family	2.A.113.1.7-D7A5Z9	WP_038012784	KUE99291	2.A.1.2.91-E9ELL4	18.4	17.2	35	1-5 w/ 3-8
2.A.113 The Nickel/Cobalt Transporter (NicO) Family	2.A.113.1.5-Q025X6	WP_052486258	WP_022710359	2.A.1.6.7-P0A2G3	46.7	16.1	67.6	1-5 w/ 4-8

Table 3.
The positive control comparisons (PC) between members of MFS are seen below. All proteins that make up the relationship, their respective GSAT scores, and alignments are seen below. It should also be noted that all members below had PFAM domains, which aligned at least 5 of their 6 TMSs with the characteristic MFS repeat unit. Boxes in grey beneath the comparison score column indicate the lowest comparison score between A-B, B-C, and C-D. Lowest PC score was 15.1 SD.

Family	Protein A	Protein B	Protein C	Protein D	Comparison Score (SD)			B-C TMS Alignment
					A-B	B-C	C-D	
2.A.1	2.A.1.2.38-Q56RY7	WP_05125705	WP_05125705	2.A.12.1.18-Q6MDZ0	44.2	23.5	61.9	1-8 w/ 1-8
2.A.1	2.A.1.2.69-O34546	WP_019240224	WP_011716441	2.A.17.1.7-Q5KYD1	15.1	85.2	72.56	1-9 w/ 1-9
2.A.1	2.A.1.2.88-Q8TZJ0	WP_055144063	WP_015030855	2.A.2.6.3-Q8EEC4	29.7	171.9	81.5	1-12 w/ 1-12
2.A.1	2.A.1.1.33-Q8NJ22	BAP73800	CAF91078	2.A.48.1.4-Q9BZV2	267.9	23.1	119.2	2-8 w/ 2-8
2.A.1	2.A.1.49.5-Q6NMN6	XP_001962407	XP_017024685	2.A.60.1.9-Q96BD0	178.9	194.6	19.9	4-9 w/ 4-9
2.A.1	2.A.1.57.1-Q9HWG8	WP_006194690	WP_015114708	2.A.71.2.1-O68867	17.4	155.4	145.5	1-11 w/ 1-11
2.A.12	2.A.12.1.18-Q6MDZ0	WP_038610631	WP_005490832	2.A.17.1.7-Q5KYD1	45.7	17.5	68.7	2-8 W/ 2-8
2.A.12	2.A.12.1.16-Q39002	EPS64663	WP_035918481	2.A.2.3.2-P94488	296.9	17.4	47.8	4-9 w/ 4-9
2.A.12	2.A.12.1.2-Q9S6V3	CRX3903	AKM21193	2.A.60.1.10-Q9EPZ7	162.2	16.2	201.6	5-9 w/ 5-9
2.A.12	2.A.12.1.18-Q6MDZ0	WP_003458335	BAU28254	2.A.71.2.3-Q55721	52.1	15.1	17.8	2-10 w/ 2-10
2.A.17	2.A.17.4.6-H2DJV9	WP_006441187	WP_057883689	2.A.2.3.12-Q745T8	26.9	22	21.3	2-8 w/ 2-8
2.A.17	2.A.17.1.1-P0C2U2	WP_017258137	XP_006861994	2.A.48.1.4-Q9BZV2	68.5	17.4	187.9	2-6 w/ 2-6
2.A.17	2.A.17.1.7-Q5KYD1	KVV13772	XP_002866685	2.A.60.1.11-Q6ZQN7	59.2	17.1	18.6	2-6 w/ 2-6
2.A.17	2.A.17.1.7-Q5KYD1	WP_012415004	WP_018399855	2.A.71.2.1-O68867	67.3	18.8	127.6	1-11 w/ 1-11

Table 3. *(continued)*

Family	Protein A	Protein B	Protein C	Protein D	Comparison Score (SD)			B-C TMS Alignment
					A-B	B-C	C-D	
2.A.2	2.A.2.6.3-Q8EEC4	WP_038331042	XP_012826759	2.A.48.1.4-Q9BZV2	108.9	22	142.9	2-8 w/ 2-8
2.A.2	2.A.2.4.1-Q39232	XP_016200557	XP_016953445	2.A.60.1.27-Q9W269	170	17.5	248.7	3-8 w/ 3-8
2.A.2	2.A.2.5.2-Q6D188	WP_002769940	XP_004025031	2.A.71.2.1-O68867	29.2	23.3	86.6	2-10 w/ 2-10
2.A.48	2.A.48.1.4-Q9BZV2	OCT81094	XP_017571680	2.A.60.1.16-Q9H2Y9	153.4	15.3	93.8	6-9 w/ 6-9
2.A.48	2.A.48.1.4-Q9BZV2	KFB48653	XP_002455071	2.A.71.3.1-Q9XIQ7	78.5	16.1	58.1	1-10 w/ 1-10
2.A.60	2.A.60.1.9-Q96BD0	XP_009860177	KOO31281	2.A.71.2.5-W2QLK8	167.2	15.2	88.5	2-6 w/ 2-6

Table 4.
The negative control (NC) comparison between members not shown to be related to MFS. All proteins that make up the relationship, the B-C GSAT scores, and B-C alignments are seen below. The highest B-C comparison score is highlighted in grey (14.3 SD).

Family	Protein A	Protein B	Protein C	Protein D	Categorized Superfamily	B-C Comparison Score (SD)	B-C Alignment
1.A.26	1.A.26.1.2-Q5SMG8	WP_062056322	CRG86951	2.A.1.19.38-Q9C101	Tog	9.4	1 w/ 7
1.A.76	N/A	N/A	N/A	N/A	TOG	-	N/A
2.A.112	2.A.112.1.14-Q6UX68	XP_014351855	XP_007674592	2.A.1.2.33-P53389	Tog	8.8	7-9 w/ 4-6
2.A.120	2.A.120.1.1-Q8CR56	WP_012731876	KPM44295	2.A.1.1.112-Q9P3U6	APC	9.3	8-10 w/ 10-12
2.A.123	2.A.123.1.9-Q8L9J7	ABK26022	KIV79942	2.A.1.1.83-Q7SD12	Tog	14	3-7 w/ 2-6
2.A.15	2.A.15.1.3-Q4ZLW8	WP_021015360	WP_019189406	2.A.1.7.13-Q08280	APC	14.3	10-12 w/ 4-6
2.A.18	2.A.18.2.10-Q9FF99	XP_016718435	KPI42421	2.A.1.2.39-Q5JAK9	APC	13.3	2-p5 w/ 1-p4
2.A.29	2.A.29.6.1-Q00319	OBZ89534	XP_012740382	2.A.1.2.63-P38776	Mc	12.97	2-4 w/ 7-9
2.A.37	2.A.37.1.11-Q6UWJ1	KTG40702	XP_013271457	2.A.1.14.11-P53322	CPA	11.2	(nothing with 1) 1-3 w/ 2-5 (3 with nothing)
2.A.39	2.A.39.3.10-A6N844	OAP59585	CDK24061	2.A.1.14.17-O43000	APC	13.4	10-12 w/ 3-6 (4 w/ nothing)
2.A.43	2.A.43.2.3-Q12010	OAV92255	AMV20189	2.A.1.15.16-Q1DA07	TOG	13.2	4-6 w/ 5-8 (4 w/ 5,6)
2.A.45	2.A.45.2.4-Q10SY9	XP_010671838	XP_003053146	2.A.1.2.33-P53389	IT	9.3	6-9 w/ 7-11
2.A.47	2.A.47.2.2-P27514	XP_007393201	XP_008877621	2.A.1.2.61-Q2V4F9	IT	9.9	1-3 w/ 1-3

Table 4. *(continued)*

Family	Protein A	Protein B	Protein C	Protein D	Categorized Superfamily	B-C Comparison Score (SD)	B-C Alignment
2.A.6	2.A.6.2.27-Q9I6X6	WP_018151710	WP_039981502	2.A.1.3.36-C5W790	RND	10.4	1 w/ 1
2.A.61	2.A.61.1.2-P45428	WP_016458608	KXJ86048	2.A.1.2.40-P38125	IT	6.8	11 w/ 4
2.A.66	2.A.66.1.14-Q96FL8	XP_015736246	XP_003719672	2.A.1.2.77-Q8NKG7	MOP	12.2	1-3 w/ 1-3
2.A.68	2.A.68.1.1-P46133	WP_025485563	CEQ39468	2.A.1.12.2-P36035	IT	12.9	2-5 w/ 2-5
2.A.69	2.A.69.2.3-B8MZ51	XP_013282821	XP_017549438	2.A.1.4.9-Q8NCC5	BART	12.6	6-9 w/ 10-12
2.A.7	2.A.7.3.63-927580526	WP_062001407	WP_039718070	2.A.1.31.2-Q55937	DMT	3.71	2-3 w/ 6-7
2.A.70	N/A	N/A	1	N/A	CPA	-	N/A
2.A.72	2.A.72.3.8-O22881	KVI10022	SAM08437	2.A.1.1.43-A0ZXK6	APC	11.5	8-10 w/ 1-4
2.A.87	N/A	N/A	1	N/A	BART	-	N/A
2.A.94	2.A.94.1.2-B0WNM6	KRY43159	KNA23640	2.A.1.2.61-Q2V4F9	IT	8.4	No TMSs align
3.A.1	2.A.1.3.36-C5W790	WP_038137994	WP_051647262	3.A.1.105.15-P75777	ABC	6.3	1 w/ 1
3.B.1	3.B.1.1.2-Q57079	WP_051840193	XP_001389178	2.A.1.13.18-P36032	CPA	13.1	1 w/ 10
3.E.1	3.E.1.7.1-Q93WP2	AGF84748	KXG36523	2.A.1.82.1-F2CRE4	TOG	8.36	4-6 w/ 3-5
9.B.25	9.B.25.1.1-Q03327	OAL46946	XP_011680324	2.A.1.13.12-Q7RTX9	NO	7.6	1-3 w/ 7-9
9.B.34	9.B.34.1.3-D4ZZR3	SAL68543	KEQ81412	2.A.1.2.67-P53283	NO	14	2 w/ 6

Table 5.
Families that have also been studied at the same depth as the ones mentioned throughout this thesis, but are not discussed throughout this thesis. All proteins that make up the relationship, the B-C GSAT scores, and B-C alignments are seen below. PFAM Domain notes have also been indicated below.

Family	Protein A	Protein B	Protein C	Protein D	B-C Comparison Score (SD)	TMS alignment	Pfam Domains Notes
1.A.8 The Major Intrinsic Protein (MIP) Family	2.A.2.6.1-O14091	XP_016623124	CCI45070	1.A.8.9.9-F6QEC2	15	1-5 w/ 1-5	More than 70% of the PFAM domain of C is found to be aligning with B.
1.A.16 The Formate-Nitrite Transporter (FNT) Family	1.A.16.2.5-A1RWM1	WP_012751216	KZP16952	2.A.1.40.2-Q5R542	15.7	4-8 w/ 1-5	Roughly 50% of the PFAM domain of B is seen to be aligning with C in this case.
3.A.3.4 Mg^{2+} ATPase Proteins	3.A.3.4.1-P36640	WP_051580591	OAP59195	2.A.1.1.125-C4QVV9	16.5	11-14 w/ 8-11	B and C share greater than 70% of their PFAM domains in the alignment of B and C
4.D.2 The Glycosyl Transferase 2 (GT2) Family	4.D.2.4.3-P75770	AJC47630	WP_010342725	2.A.1.3.45 - E5Y3Y1	104	2-8 w/ 2-8	The alignment between B and C shows that they align well above 70% of their respective domains.

Table 5. *(continued)*

Family	Protein A	Protein B	Protein C	Protein D	B-C Comparison Score (SD)	TMS alignment	Pfam Domains Notes
4.D.2 The Glycosyl Transferase 2 (GT2) Family	2.A.60.1.16-Q9H2Y9	EHJ71138	WP_060743451	4.D.2.1.7-C9XJW8	16.1	3-5 w/ 2-4	Of the PFAM domain of C we see that roughly 50% is contained in the alignment between B and C.
4.D.2 The Glycosyl Transferase 2 (GT2) Family	2.A.2.3.4-P31435	WP_034882575	WP_009402631	4.D.2.4.3-P75770	14.8	2-7 w/ 2-7	We see that over 70% of the domain of C is found to be aligning between B and C.
5.B.2 The Eukaryotic Cytochrome b_{561} ($Cytb_{561}$) Family	5.B.2.3.2-G9NAJ2	XP_009849912	XP_008626745	2.A.1.49.3-F4IKF6	15.9	2-6 w/ 2-6	Over 70% of the Domain of B is found to be in the alignment between B and C.
9.B.57 The Conidiation and Conidial Germination Protein (CCGP) Family	9.B.57.1.2-P53584	CCE41607	CBY23915	2.A.1.19.36-Q9VCA2	15.5	8-12 w/ 2-6	The alignment between B and C share roughly 60% of both domains B and C.
9.B.104 The Rhomboid Protease Family	9.B.104.1.1-P09391	WP_011201337	XP_006957555	2.A.1.1.69-A1Z264	19	1-5 w/ 1-5	When B and C align we see that it aligns 5 TMSs very well. This alignment contains well over 70% of the PFAM domain of B when aligning with C.

Table 5. *(continued)*

Family	Protein A	Protein B	Protein C	Protein D	B-C Comparison Score (SD)	TMS alignment	Pfam Domains Notes
9.B.111 The 6 TMS Lysyl tRNA Synthetase (LysS) Family	9.B.111.1.2-O69916	WP_037604593	WP_014387355	2.A.2.6.3-Q8EEC4	15.5	4-6 w/ 4-6	Although on 3 TMS's are aligning, over 70% alignment of the domain of B is seen in the B to C alignment.
9.B.143 The 6 TMS DUF1275/PF06912 (DUF1275) Family	9.B.143.1.2-D4XPL7	WP_054410151	SCB71721	2.A.1.8.11-Q166T6	16.5	3-6 w/ 8-11	Roughly 70% of the PFAM domain of B are found to be present in the B-C alignment.

Table 6.
Table of results providing internal repeat data (if available) on all families discussed in this thesis. The analysis of the results and GSAT score is also shown below. A GSAT score above 4 in the case of thesis results is considered to be significant.

TCID	Method	Result	Implied repeat unit	GSAT z-core (SD)
2.A.57.2.9, 2.A.57.2.4	Ancient Rep	TMS 1-5 aligns with TMS 7-11	6+6 since the family underwent a C terminal Deletion	6
2.A.57.5.3, 2.A.57.3.1	Ancient Rep	TMS 1-5 aligns with TMS 7-11	6+6 family underwent C terminal Deletion	6
2.A.100.1.1	AncientREP, HHrepID, Radar	No meaningful results found	N/A	N/A
2.A.100.1.2	AncientREP, HHrepID, Radar	No meaningful results found	N/A	N/A
2.A.100.1.3	AncientREP, HHrepID, Radar	No meaningful results found	N/A	N/A
2.A.100.1.4	AncientREP, HHrepID, Radar	No meaningful results found	N/A	N/A
2.A.100.1.5	AncientREP, HHrepID, Radar	No meaningful results found	N/A	N/A
2.A.100.2.1	AncientREP, HHrepID, Radar	No meaningful results found	N/A	N/A
2.A.125.1.5, 2.A.125.1.5	TMS Repeat, Ancient Rep	TMS 4-6 aligns with TMS 7-9	Suggests 3+3	6
2.A.76.1.12, 2.A.76.3.1	AncientRep	TMS 2-3 aligns with TMS 5-6	Suggests 3+3	7
2.A.76.3.1, 2.A.1.7	AncientRep	TMS 1-2 aligns with TMS 4-5	Suggests 3+3	6
2.A.76.1.7, 2.A.76.1.10	AncientRep	TMS 1-2 aligns with TMS 4-5	Suggests 3+3	5

Table 6. *(continued)*

TCID	Method	Result	Implied repeat unit	GSAT z-core (SD)
2.A.109.1.5, 2.A.109.2.2	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	7
2.A.109.2.2, 2.A.109.1.2	AncientRep	TMS 1-3 aligns with TMS 7-9	Suggests 3+3	5
2.A.109.2.2, 2.A.109.2.2	AncientRep	TMS 1-2 aligns with TMS 4-5	Suggests 3+3	5
2.A.109.1.5, 2.A.109.1.5	AncientRep, HHRepId	TMS 1-2 aligns with TMS 4-5	Suggests 3+3	4
2.A.113.1.6, 2.A.113.1.5	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	13
2.A.113.1.3, 2.A.113.1.11	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	9
2.A.113.1.7, 2.A.113.2.3	AncientRep	TMS 2-3 aligns with TMS 5-6	Suggests 3+3	7
2.A.113.2.1, 2.A.113.2.3	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	7
2.A.113.1.6, 2.A.113.1.2	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	7
2.A.113.1.1, 2.A.113.1.11	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	6
2.A.113.1.2, 2.A.113.1.11	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	6
2.A.113.1.3, 2.A.113.2.2	AncientRep	TMS 1-3 aligns with TMS 4-6	Suggests 3+3	5

References

1.Marger MD, Saier MH, Jr. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends in biochemical sciences. 1993;18(1):13-20. Epub 1993/01/01. PubMed PMID: 8438231.

2. Pao SS, Paulsen IT, Saier MH, Jr. Major facilitator superfamily. Microbiology and molecular biology reviews : MMBR. 1998;62(1):1-34. Epub 1998/04/08. PubMed PMID: 9529885; PubMed Central PMCID: PMCPmc98904.

3. Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier MH, Jr. The major facilitator superfamily (MFS) revisited. The FEBS journal. 2012;279(11):2022-35. Epub 2012/03/31. doi: 10.1111/j.1742-4658.2012.08588.x. PubMed PMID: 22458847; PubMed Central PMCID: PMCPmc3425384.

4. Saier MH, Jr., Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic acids research. 2006;34(Database issue):D181-6. Epub 2005/12/31. doi: 10.1093/nar/gkj001. PubMed PMID: 16381841; PubMed Central PMCID: PMCPmc1334385.

5. Saier MH, Jr., Yen MR, Noto K, Tamang DG, Elkan C. The Transporter Classification Database: recent advances. Nucleic acids research. 2009;37(Database issue):D274-8. Epub 2008/11/22. doi: 10.1093/nar/gkn862. PubMed PMID: 19022853; PubMed Central PMCID: PMCPmc2686586.

6. Yee DC, Shlykov MA, Vastermark A, Reddy VS, Arora S, Sun EI, The transporter-opsin-G protein-coupled receptor (TOG) superfamily. The FEBS journal. 2013;280(22):5780-800. Epub 2013/08/29. doi: 10.1111/febs.12499. PubMed PMID: 23981446; PubMed Central PMCID: PMCPmc3832197.

7. Saier MH, Jr. Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. Microbiological reviews. 1994;58(1):71-93. Epub 1994/03/01. PubMed PMID: 8177172; PubMed Central PMCID: PMCPmc372954.

8. Saier MH, Jr. A functional-phylogenetic classification system for transmembrane solute transporters. Microbiology and molecular biology reviews : MMBR. 2000;64(2):354-411. Epub 2000/06/06. PubMed PMID: 10839820; PubMed Central PMCID: PMCPmc98997.

9. Saier MH, Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. The Transporter Classification Database (TCDB): recent advances. Nucleic acids research. 2016;44(D1):D372-9. Epub 2015/11/08. doi: 10.1093/nar/gkv1103. PubMed PMID: 26546518; PubMed Central PMCID: PMCPmc4702804.

10. Moraes TF, Reithmeier RA. Membrane transport metabolons. Biochimica et biophysica acta. 2012;1818(11):2687-706. Epub 2012/06/19. doi: 10.1016/j.bbamem.2012.06.007. PubMed PMID: 22705263.

11. Eudes A, Kunji ER, Noiriel A, Klaus SM, Vickers TJ, Beverley SM, Identification of transport-critical residues in a folate transporter from the folate-biopterin transporter (FBT) family. The Journal of biological chemistry. 2010;285(4):2867-75. Epub 2009/11/20. doi: 10.1074/jbc.M109.063651. PubMed PMID: 19923217; PubMed Central PMCID: PMCPmc2807340.

12. Yan N. Structural Biology of the Major Facilitator Superfamily Transporters. Annual review of biophysics. 2015;44:257-83. Epub 2015/06/23. doi: 10.1146/annurev-biophys-060414-033901. PubMed PMID: 26098515.

13. Zhang XC, Zhao Y, Heng J, Jiang D. Energy coupling mechanisms of MFS transporters. Protein science : a publication of the Protein Society. 2015;24(10):1560-79. Epub 2015/08/04. doi: 10.1002/pro.2759. PubMed PMID: 26234418; PubMed Central PMCID: PMCPmc4594656.

14. Saier MH, Jr., Beatty JT, Goffeau A, Harley KT, Heijne WH, Huang SC, The major facilitator superfamily. Journal of molecular microbiology and biotechnology. 1999;1(2):257-79. Epub 2000/08/16. PubMed PMID: 10943556.

15. Winkler HH, Neuhaus HE. Non-mitochondrial ATP transport. Trends in biochemical sciences. 1999;24(2):64-8. Epub 1999/03/31. PubMed PMID: 10098400.

16. Trentmann O, Jung B, Neuhaus HE, Haferkamp I. Nonmitochondrial ATP/ADP transporters accept phosphate as third substrate. The Journal of biological chemistry. 2008;283(52):36486-93. Epub 2008/11/13. doi: 10.1074/jbc.M806903200. PubMed PMID: 19001371; PubMed Central PMCID: PMCPmc2606016.

17. Newstead S. Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. Biochimica et biophysica acta. 2015;1850(3):488-99. Epub 2014/05/27. doi: 10.1016/j.bbagen.2014.05.011. PubMed PMID: 24859687; PubMed Central PMCID: PMCPmc4331665.

18. Matherly LH, Wilson MR, Hou Z. The major facilitative folate transporters solute carrier 19A1 and solute carrier 46A1: biology and role in antifolate chemotherapy of cancer. Drug metabolism and disposition: the biological fate of chemicals. 2014;42(4):632-49. Epub 2014/01/08. doi: 10.1124/dmd.113.055723. PubMed PMID: 24396145; PubMed Central PMCID: PMCPmc3965896.

19. Flintoff WF, Williams FM, Sadlish H. The region between transmembrane domains 1 and 2 of the reduced folate carrier forms part of the substrate-binding pocket. The Journal of biological chemistry. 2003;278(42):40867-76. Epub 2003/08/12. doi: 10.1074/jbc.M302102200. PubMed PMID: 12909642.

20. Hong M. Critical domains within the sequence of human organic anion transporting polypeptides. Current drug metabolism. 2014;15(3):265-70. Epub 2014/01/01. PubMed PMID: 24372098.

21. Hagenbuch B, Stieger B. The SLCO (former SLC21) superfamily of transporters. Molecular aspects of medicine. 2013;34(2-3):396-412. Epub 2013/03/20. doi: 10.1016/j.mam.2012.10.009. PubMed PMID: 23506880; PubMed Central PMCID: PMCPmc3602805.

22. Tsigelny IF, Kovalskyy D, Kouznetsova VL, Balinskyi O, Sharikov Y, Bhatnagar V, Conformational changes of the multispecific transporter organic anion transporter 1 (OAT1/SLC22A6) suggests a molecular mechanism for initial stages of drug and metabolite transport. Cell biochemistry and biophysics. 2011;61(2):251-9. Epub 2011/04/19. doi: 10.1007/s12013-011-9191-7. PubMed PMID: 21499753.

23 .Zhu C, Nigam KB, Date RC, Bush KT, Springer SA, Saier MH, Jr., Evolutionary Analysis and Classification of OATs, OCTs, OCTNs, and Other SLC22 Transporters: Structure-Function Implications and Analysis of Sequence Motifs. PloS one. 2015;10(11):e0140569. Epub 2015/11/05. doi: 10.1371/journal.pone.0140569. PubMed PMID: 26536134; PubMed Central PMCID: PMCPmc4633038.

24. Tazoe Y, Hayashi H, Tsuboi S, Shioura T, Matsuyama T, Yamada H, Reduced folate carrier 1 gene expression levels are correlated with methotrexate efficacy in Japanese patients with rheumatoid arthritis. Drug metabolism and pharmacokinetics. 2015;30(3):227-30. Epub 2015/05/25. doi: 10.1016/j.dmpk.2015.02.001. PubMed PMID: 26003891.

25. Medrano-Soto A, Moreno-Hagelsieb G, McLaughlin D, Ye ZS, Hendargo KJ, Saier MH, Jr. Bioinformatic characterization of the Anoctamin Superfamily of Ca2+-activated ion channels and lipid scramblases. PloS one. 2018;13(3):e0192851. doi: 10.1371/journal.pone.0192851. PubMed PMID: 29579047; PubMed Central PMCID: PMCPMC5868767.

26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389-402. PubMed PMID: 9254694; PubMed Central PMCID: PMCPMC146917.

27. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England). 2012;28(23):3150-2. doi: 10.1093/bioinformatics/bts565. PubMed PMID: 23060610; PubMed Central PMCID: PMCPMC3516142.

28. Reddy VS, Saier MH, Jr. BioV Suite--a collection of programs for the study of transport protein evolution. The FEBS journal. 2012;279(11):2036-46. doi: 10.1111/j.1742-4658.2012.08590.x. PubMed PMID: 22568782; PubMed Central PMCID: PMCPMC3978091.

29. Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics. 1991;11(3):635-50. PubMed PMID: 1774068.

30. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16(6):276-7. PubMed PMID: 10827456.

31. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics (Oxford, England). 2001;17(9):849-50. Epub 2001/10/09. PubMed PMID: 11590105.

32. Yen MR, Choi J, Saier Jr MH. Bioinformatic Analyses of Transmembrane Transport: Novel Software for Deducing Protein Phylogeny, Topology, and Evolution. Journal of molecular microbiology and biotechnology. 2009;17(4):163-76.

33. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, The Pfam protein families database: towards a more sustainable future. Nucleic acids research. 2016;44(D1):D279-85. doi: 10.1093/nar/gkv1344. PubMed PMID: 26673716; PubMed Central PMCID: PMCPMC4702930.

34. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7(10):e1002195. doi: 10.1371/journal.pcbi.1002195. PubMed PMID: 22039361; PubMed Central PMCID: PMCPMC3197634.

35. Alva V, Nam SZ, Soding J, Lupas AN. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic acids research. 2016;44(W1):W410-5. doi: 10.1093/nar/gkw348. PubMed PMID: 27131380; PubMed Central PMCID: PMCPMC4987908.

36. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539. doi: 10.1038/msb.2011.75. PubMed PMID: 21988835; PubMed Central PMCID: PMCPMC3261699.

37. Zhai Y, Saier MH, Jr. A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. Journal of molecular microbiology and biotechnology. 2001;3(2):285-6. Epub 2001/04/26. PubMed PMID: 11321584.

38. Wong WC, Maurer-Stroh S, Eisenhaber F. Not all transmembrane helices are born equal: Towards the extension of the sequence homology concept to membrane proteins. Biol Direct. 2011;6:57. doi: 10.1186/1745-6150-6-57. PubMed PMID: 22024092; PubMed Central PMCID: PMCPMC3217874.

39. Wong WC, Maurer-Stroh S, Eisenhaber F. More than 1,001 problems with protein domain databases: transmembrane regions, signal peptides and the issue of sequence homology. PLoS

Comput Biol. 2010;6(7):e1000867. doi: 10.1371/journal.pcbi.1000867. PubMed PMID: 20686689; PubMed Central PMCID: PMCPMC2912341.

40. Valdes R, Elferich J, Shinde U, Landfear SM. Identification of the intracellular gate for a member of the equilibrative nucleoside transporter (ENT) family. The Journal of biological chemistry. 2014;289(13):8799-809. Epub 2014/02/06. doi: 10.1074/jbc.M113.546960. PubMed PMID: 24497645; PubMed Central PMCID: PMCPmc3979384.

41. Baldwin SA, Beal PR, Yao SY, King AE, Cass CE, Young JD. The equilibrative nucleoside transporter family, SLC29. Pflugers Archiv : European journal of physiology. 2004;447(5):735-43. Epub 2003/07/03. doi: 10.1007/s00424-003-1103-2. PubMed PMID: 12838422.

42. Yeh KY, Yeh M, Glass J. Interactions between ferroportin and hephaestin in rat enterocytes are reduced after iron ingestion. Gastroenterology. 2011;141(1):292-9, 9.e1. Epub 2011/04/09. doi: 10.1053/j.gastro.2011.03.059. PubMed PMID: 21473866.

43. Madejczyk MS, Ballatori N. The iron transporter ferroportin can also function as a manganese exporter. Biochimica et biophysica acta. 2012;1818(3):651-7. Epub 2011/12/20. doi: 10.1016/j.bbamem.2011.12.002. PubMed PMID: 22178646.

44. Delaby C, Pilard N, Puy H, Canonne-Hergaux F. Sequential regulation of ferroportin expression after erythrophagocytosis in murine macrophages: early mRNA induction by haem, followed by iron-dependent protein expression. The Biochemical journal. 2008;411(1):123-31. Epub 2007/12/13. doi: 10.1042/bj20071474. PubMed PMID: 18072938.

45. Mitchell CJ, Shawki A, Ganz T, Nemeth E, Mackenzie B. Functional properties of human ferroportin, a cellular iron exporter reactive also with cobalt and zinc. American journal of physiology Cell physiology. 2014;306(5):C450-9. Epub 2013/12/07. doi: 10.1152/ajpcell.00348.2013. PubMed PMID: 24304836; PubMed Central PMCID: PMCPmc4042619.

46. Le Gac G, Ka C, Joubrel R, Gourlaouen I, Lehn P, Mornon JP, Structure-function analysis of the human ferroportin iron exporter (SLC40A1): effect of hemochromatosis type 4 disease mutations and identification of critical residues. Human mutation. 2013;34(10):1371-80. Epub 2013/06/21. doi: 10.1002/humu.22369. PubMed PMID: 23784628.

47. Yonezawa A, Masuda S, Katsura T, Inui K. Identification and functional characterization of a novel human and rat riboflavin transporter, RFT1. American journal of physiology Cell physiology. 2008;295(3):C632-41. Epub 2008/07/18. doi: 10.1152/ajpcell.00019.2008. PubMed PMID: 18632736.

48. Powers HJ. Riboflavin (vitamin B-2) and health. The American Journal of Clinical Nutrition. 2003;77(6):1352-60.

49. Yamamoto S, Inoue K, Ohta KY, Fukatsu R, Maeda JY, Yoshida Y, Identification and functional characterization of rat riboflavin transporter 2. Journal of biochemistry. 2009;145(4):437-43. Epub 2009/01/06. doi: 10.1093/jb/mvn181. PubMed PMID: 19122205.

50. Chang AB, Lin R, Keith Studley W, Tran CV, Saier MH, Jr. Phylogeny as a guide to structure and function of membrane transport proteins. Molecular membrane biology. 2004;21(3):171-81. Epub 2004/06/19. doi: 10.1080/09687680410001720830. PubMed PMID: 15204625.

51. Burian J, Tu N, Kl'ucar L, Guller L, Lloyd-Jones G, Stuchlik S, In vivo and in vitro cloning and phenotype characterization of tellurite resistance determinant conferred by plasmid pTE53 of a clinical isolate of Escherichia coli. Folia Microbiol (Praha). 1998;43(6):589-99. PubMed PMID: 10069007.

52. Anantharaman V, Iyer LM, Aravind L. Ter-dependent stress response systems: novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. Mol Biosyst. 2012;8(12):3142-65. doi: 10.1039/c2mb25239b. PubMed PMID: 23044854; PubMed Central PMCID: PMCPMC4104200.

53. Tsu BV, Saier MH, Jr. The LysE Superfamily of Transport Proteins Involved in Cell Physiology and Pathogenesis. PloS one. 2015;10(10):e0137184. doi: 10.1371/journal.pone.0137184. PubMed PMID: 26474485; PubMed Central PMCID: PMCPMC4608589.

54. Kutukova EA, Zakataeva NP, Livshits VA. [Expression of the genes encoding RhtB family proteins depends on global regulator Lrp]. Mol Biol (Mosk). 2005;39(3):374-8. PubMed PMID: 15981566.

55. Morgan JL, Strumillo J, Zimmer J. Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature. 2013;493(7431):181-6. Epub 2012/12/12. doi: 10.1038/nature11744. PubMed PMID: 23222542; PubMed Central PMCID: PMCPmc3542415.

56. Vicente CM, Santos-Aberturas J, Guerra SM, Payero TD, Martin JF, Aparicio JF. PimT, an amino acid exporter controls polyene production via secretion of the quorum sensing pimaricin-inducer PI-factor in Streptomyces natalensis. Microb Cell Fact. 2009;8:33. doi: 10.1186/1475-2859-8-33. PubMed PMID: 19505319; PubMed Central PMCID: PMCPMC2698837.

57. Marrero J, Auling G, Coto O, Nies DH. High-level resistance to cobalt and nickel but probably no transenvelope efflux: Metal resistance in the Cuban Serratia marcescens strain C-1. Microb Ecol. 2007;53(1):123-33. doi: 10.1007/s00248-006-9152-7. PubMed PMID: 17186148.