

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Bioinformatic Characterization of the Anoctamin Family of Calcium-Activated Chloride Channels and the Establishment of Homology Using Protein Sequence Analysis

### Permalink

<https://escholarship.org/uc/item/6vk83636>

### Author

McLaughlin, Daniel

### Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Bioinformatic Characterization of the Anoctamin Family of Calcium-Activated  
Chloride Channels and the Establishment of Homology Using Protein  
Sequence Analysis

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

in

Biology

by

Daniel McLaughlin

Committee in charge:

Professor Milton Saier, Chair  
Professor Mauricio Montal, Co-Chair  
Professor David Traver

2015



The Thesis of Daniel McLaughlin is approved, and it is acceptable in quality and form for publication on microfilm electronically:

---

---

Co-Chair

---

Chair

University of California, San Diego

2015

## TABLE OF CONTENTS

SIGNATURE PAGE.....	iii
TABLE OF CONTENTS .....	iv
LIST OF ABBREVIATIONS.....	v
LIST OF FIGURES.....	vi
LIST OF TABLES.....	viii
AKNOWLEDGEMENTS.....	ix
ABSTRACT OF THE THESIS .....	x
INTRODUCTION.....	1
METHODS .....	6
RESULTS.....	11
DISCUSSION.....	22
CONCLUSION .....	32
APPENDIX.....	34
REFERENCES.....	48

## LIST OF ABBREVIATIONS

ANO—Anoctamin

TMC—Transmembrane Channel-like Protein

DUF—Domain of Unknown Function

TMS—Transmembrane Segment

VIC—Voltage-gated Ion Channel

TRP—Transient Receptor Potential Channel

CPA—Monovalent Cation-Proton Antiporter

GSAT—Global Sequence Alignment Search Tool

## LIST OF FIGURES

<b>Figure 1:</b> Predicted structures for various members of the Anoctamin Superfamily .....	11
<b>Figure 2:</b> Superfamily Tree of the Anoctamin Superfamily created by SuperfamilyTree and drawn with FigTree .....	12
<b>Figure 3:</b> Average hydropathy and amphipathicity of the Anoctamin Family (1.A.17.1) with similarity and TMS prediction.....	14
<b>Figure 4:</b> Average hydropathy and amphipathicity of the TMC Family (1.A.17.4) with similarity and TMS prediction.....	15
<b>Figure 5:</b> Average hydropathy and amphipathicity of the DUF221 Family (1.A.17.5) with similarity and TMS prediction.....	16
<b>Figure 6:</b> Average hydropathy and amphipathicity of the Multi-domain Family (1.A.17.2) with similarity and TMS prediction.....	17
<b>Figure 7:</b> Average hydropathy and amphipathicity of the Uncharacterized Family (1.A.17.3) with similarity and TMS prediction. ....	18
<b>Figure 8:</b> Predicted evolutionary pathway for the appearance of members of the Anoctamin Superfamily and other families from a primordial 6 TMS VIC Family member .....	30
<b>Figure 9:</b> Average hydropathy and amphipathicity of the VIC Family (1.A.1) with similarity and TMS prediction.....	34
<b>Figure 10:</b> Average hydropathy and amphipathicity of the TRP Family (1.A.4) with similarity and TMS prediction.....	35
<b>Figure 11:</b> Average hydropathy and amphipathicity of the CPA1 Family (2.A.36) with similarity and TMS prediction.....	36
<b>Figure 12:</b> Average hydropathy and amphipathicity of the TrK Family (2.A.38) with similarity and TMS prediction.....	37
<b>Figure 13:</b> GSAT binary alignment Control: CFTR (ABC Superfamily) compared with TMC .....	38
<b>Figure 14:</b> GSAT binary alignment between a TMC and an ANO protein.....	39
<b>Figure 15:</b> GSAT binary alignment between a TMC and a VIC protein.....	40
<b>Figure 16:</b> GSAT binary alignment between a TMC and a TRP protein. ....	41
<b>Figure 17:</b> GSAT binary alignment between an ANO and a TRP protein. ....	42
<b>Figure 18:</b> GSAT binary alignment between a TMC and a CPA1 protein. ....	43
<b>Figure 19:</b> GSAT binary alignment between a TRP and a CPA1 protein.....	44

<b>Figure 20:</b> GSAT binary alignment between an ANO and a Trk protein. ....	45
<b>Figure 21:</b> GSAT binary alignment between a TRP and a Trk protein. ....	46
<b>Figure 22:</b> GSAT binary alignment between a CPA1 and a Trk protein. ....	47

## LIST OF TABLES

<b>Table 1:</b> Average length and number of TMSs for each family in the Anoctamin Superfamily .....	13
<b>Table 2:</b> GSAT comparison scores between each family in the Anoctamin Superfamily .....	19
<b>Table 3:</b> GSAT comparison scores between 1.A.17 and distant families .....	21

## ACKNOWLEDGEMENTS

I would like to acknowledge the support of Professor Milton Saier, my committee chair, as well as the members of Saier Lab.

## ABSTRACT OF THE THESIS

Bioinformatic Characterization of the Anoctamin family of Calcium-activated Chloride Channels and the Establishment of the Anoctamin Superfamily Using Protein Sequence Analysis

by

Daniel McLaughlin

Master of Science in Biology

University of California, San Diego, 2015

Professor Milton Saier, Chair  
Professor Mauricio Montal, Co-Chair

The Anoctamin Family of proteins is closely related to both Transmembrane Channel-like proteins and DUF221 proteins whose functions are largely unknown. These three families are grouped into a larger superfamily called the Anoctamin Superfamily. Here, we have analyzed the

topologies of each of these three families of proteins using Bioinformatic programs included in the BioV Suite: AveHAS, GSAT, CLUSTAL X, and SuperTree. We have constructed structural schemes for each of these families and deduced a possible function for the DUF221 Family. We suggest members are involved in NaCl tolerance. We have also found several families of proteins in the CPA, VIC, and TRP superfamilies that are evolutionarily linked to the Anoctamin Superfamily. Identification of these homologous families is important because it provides insight into possible mechanisms for the poorly characterized proteins in the Anoctamin Superfamily. The results establish for the first time that the Anoctamin Superfamily is a sub-superfamily of the largest known superfamily of transmembrane channel proteins, the Voltage-gated Ion Channel (VIC) Superfamily.

## INTRODUCTION

### Anoctamins

Anoctamins, also referred to as TMEM16 proteins, comprise a family of proteins implicated in various channelopathies. Mutations in human Anoctamin-1 (Ano1) have been found to be associated with diseases including muscular dystrophies and cerebellar ataxia (Duran and Hartzell 2011). Further, overexpression of the gene encoding Ano1 has been linked to several forms of cancer, specifically to gastrointestinal stromal tumors and squamous cell carcinomas (Duran and Hartzell 2011). Anoctamins are present in nearly every eukaryote with 10 paralogs identified in vertebrates named Ano1 through Ano10 (Pang et al. 2014). It has been shown that Ano1 and Ano2 are  $\text{Ca}^{2+}$ -activated  $\text{Cl}^-$  channels (CaCCs). However, it is unclear whether the remaining anoctamins are activated by  $\text{Ca}^{2+}$  or even anion channels (Pang et al. 2014). It has been proposed that Ano1 and Ano2 have an 8 transmembrane segment (TMS) topology with a re-entrant loop between the fifth and sixth TMSs which is proposed to be the pore region of the channel (Pang et al. 2014). The name “Anoctamin” has been given to this protein family as a result of the 8 TMS topology and the anion conductance expressed by Ano1 and Ano2 (anion = ano; 8 = oct)(Duran and Hartzell 2011).

The mechanism through which increased intracellular  $\text{Ca}^{2+}$  concentration activates chloride conductance is currently unknown. Early tests indicate that

calmodulin, a  $\text{Ca}^{2+}$  binding protein, is required for this process. It has been proposed that  $\text{Ca}^{2+}$  bound calmodulin physically interacts with the cytosolic N-terminus of Ano1 to activate the  $\text{Cl}^-$  channel (Tian et al. 2011). A recent study, however, suggested that purified Ano1 is sufficient to mediate CaCC activity. This study proposes that calmodulin is not necessary, nor are any other accessory proteins required to mediate such activity. Instead, a set of two conserved glutamate residues between TMSs 6 and 7 are believed to be responsible for Ano1 activation by  $\text{Ca}^{2+}$ . This suggests that  $\text{Ca}^{2+}$  interacts directly with Ano1 and not indirectly through accessory proteins (Terashima et al. 2013). Galletta, however, showed that anoctamins contain a series of 4-5 consecutive glutamic acid residues that are located in the region between TMSs 2 and 3. These residues may be the site of  $\text{Ca}^{2+}$  interaction. Furthermore, an arginine followed by two glutamines found in the pore loop region may play a role in  $\text{Cl}^-$  conductance. Mutations in these amino acid residues result in altered ion selectivity and can lead to cation conductance (Galletta 2009).

Alteration of the previously mentioned residues has a strong effect on the voltage dependence of the channels. A wild type Anoctamin channel under non-optimal  $\text{Ca}^{2+}$  concentration will activate upon imposition of a positive membrane potential, and deactivation occurs when the membrane potential returns to its previous value. When the  $\text{Ca}^{2+}$  concentration is at optimal levels,

the channel becomes active at negative membrane potentials (Galiotta 2009). Splice variants of Anoctamin have different levels of voltage,  $[Ca^{2+}]$  dependency, and ion selectivities (Galiotta 2009). More research is required to uncover the process of  $Ca^{2+}$  Anoctamin activation. It is predicted that changes in the membrane potential cause a conformational change in the pore region of the channel (Galiotta 2009), but it is unclear if this is similar in other members of the Anoctamin Superfamily.

## Transmembrane Channel-like Proteins

Through sequence similarity, the transmembrane channel-like proteins (TMC) have been suggested to be homologous to anoctamins (Hahn et al. 2009). TMC proteins are predicted to have an 8 TMS topology with a re-entrant loop similar to that of the anoctamins. Additionally, many conserved amino acid residues have been identified in TMSs 4-7 that correspond in position and nature to residues in the same TMSs of the anoctamins (Hahn et al. 2009). TMC homologues are primarily found in animals, although at least one homologue has been found in choanoflagellates (Hahn et al. 2009). This differs from the species diversity found for the anoctamins.

There are 8 TMC paralogs in animals named TMC1 through TMC8. Mutations in TMC1, the most studied TMC, cause deafness in both mice and humans. It has been shown that mice lacking a functional TMC1 fail to develop

working cochlear neurosensory hair cells (Labay et al. 2010). TMC1 and TMC2 expressed in these hair cells are crucial to mechanotransduction, where  $\text{Ca}^{2+}$  enters the cell in response to sound vibrations (Kim and Fettiplace 2013). Additional experiments have elucidated a possible function for TMC1. TMC1 acts as a sensor for salt chemosensation in *Caenorhabditis elegans*. TMC1 is required for behavioral avoidance in response to increased NaCl concentrations (Chatzigeorgiou et al. 2013). Furthermore, expression of *C. elegans* TMC1 in mammalian cell culture resulted in  $\text{Na}^+$ -activated cation conductance. These data suggest a possible function for TMC1 as an ionotropic receptor (Chatzigeorgiou et al. 2013). Functions of TMCs 3-8 are not well understood, although TMC 6 and 8 are implicated in the human disease, epidermodysplasia verruciformis, which involves an increased susceptibility to human papillomavirus infection (Horton and Stokes 2014).

## DUF221 Containing Proteins

Another family that has been associated with both Anoctamin and TMC families is the DUF221 Family. DUF is an acronym that stands for Domain of Unknown Function. Little is known about these proteins, and their functions have not been elucidated (Bateman et al. 2010).

## Bioinformatic Analyses

Bioinformatic programs can be used to analyze anoctamins, TMCs, and DUF221s. In the absence of crystal structures, protein sequence analysis is important because it provides an easy and cheap resource for studying channel proteins. We use sequence analysis tools to compare anoctamins to TMCs as well as a group of other families that we believe are distantly related. Furthermore, we are interested in how our predictive tools for protein topology of these families correspond to results obtained through biochemical assays (see above). In establishing evolutionary relationships between distinct families, one of the first steps is to select candidates which have similar topologies, ion selectivities, or mechanisms of action. Since there is a notable difference between the ion conductances of anoctamins and TMCs, we predict that the mechanisms of channel activation will be most strongly conserved among family members, and that these mechanisms will differ in detail for members of dissimilar families. However, members of the generalized mechanism observed for all of these families may prove to be related.

## METHODS

### Examining conserved domains within TCDB members

All members of Anoctamin, TMC, and DUF221 families recorded in TCDB were used as query sequences for a batch conserved domain (CD) search on the Conserved Domains Database, referred to as CDD (Marchler-Bauer et al. 2013, Marchler-Bauer and Bryant 2004). Locations and occurrences of conserved domains were reported for each protein in the TC family 1.A.17. As a result of the analyses reported here, within 1.A.17, anoctamin is represented by the family 1.A.17.1, TMC is represented by 1.A.17.4, and DUF221 is represented by 1.A.17.5.

### SuperfamilyTree of TCDB members

A phylogenetic tree of the TC group 1.A.17 was created using the program SuperfamilyTree (Chen et al. 2011, Yen et al. 2009, Yen et al. 2010, Lee et al. 2014). SuperfamilyTree uses tens of thousands of comparative BLAST bit scores to calculate phylogeny for related proteins that have high sequence divergence. SuperfamilyTree generates a consensus tree based on agreement of 100 trees (Chen et al. 2011, Yen et al. 2009, Yen et al. 2010, Lee et al. 2014). The phylogenetic tree created by SuperfamilyTree was visualized using the FigTree program (<http://tree.bio.ed.ac.uk/software/figtree/>).

## Finding homologs from each cluster of the generated phylogenetic tree

One representative from each cluster of the phylogenetic tree was chosen as a query in a search for homologs. Protocol1 from the BioV Suite was used to find these homologs (Reddy and Saier 2012). This program uses PSI-BLAST with 2 iterations and a cutoff of 0.7 to generate a list of homologous protein sequences. The cutoff of 0.7 was used to eliminate sequences that exhibit similarity greater than 70% to any other sequence in the list of homologs (Altschul et al. 1997). Protocol1 creates a file containing non-redundant homologous sequences in FASTA format.

## Multiple alignments of homologs and average hydrophobicity / amphipathicity / similarity plots

Using CLUSTALX, A multiple alignment for each list of homologs was created. Sequences that introduced large gaps in the alignment were removed, and the resulting alignments were used to create new phylogenetic trees. Furthermore, the lists of homologs obtained using Protocol1 were tabulated based on the results of these new trees. Average hydrophobicity plots were then created with the web program AveHAS using these multiple alignments (Zhai and Saier 2001). AveHAS plots, or average hydrophobicity,

amphipathicity and similarity plots, are used to examine conserved topologies within families. These plots were used to estimate the positions of TMSs in members of these multiple alignments (Zhai and Saier 2001).

## Protocol2 and GSAT

Protocol2 from the BioV Suite of programs was used to find similarities between two lists of homologs obtained using Protocol1 (Reddy and Saier 2012). This program forms binary alignments between all protein sequences from each of two lists of homologs. Each binary alignment shows labeled TMSs in each sequence and provides a comparison score expressed in standard deviations (SD) (Reddy and Saier 2012). The labeled TMSs were used to indicate which segments align and are conserved between two families of proteins. The top scoring binary alignments were then verified using the Global Sequence Alignment Tool (GSAT). This TCDB web program creates a binary alignment of two sequences using optimally 20,000 random shuffles to create accurate comparison scores (Reddy and Saier 2012). We consider scores greater than 14.0 SD significant if there is overlap in the transmembrane regions of two sequences. (Reddy and Saier 2012).

The Comparison score expressed in standard deviations (SD) is used as a quantitative indication of homology. In the alignment process, an initial binary alignment is created with a score based on sequence identity and gaps.

Both of the sequences are then randomized (shuffled) and realigned to provide new scores that provide background values to determine the significance of the experimental value. This approach allows for the correction of high values obtained due to a restricted amino acid composition. The GSAT score reflects the difference obtained for the original un-randomized, alignment compared to the average scores of the shuffled alignments (Reddy and Saier 2012).

### Internal duplications

HHrepID and IntraCompare were used to identify any possible internal duplications within each family of proteins to elucidate their evolutionary origins. HHrepID uses a single protein sequence to locate occurrences of internal duplication by using an HMM-HMM comparison (Biegert and Soding 2008). IntraCompare (Reddy and Saier 2012) uses a CLUSTAL X multiple alignment as an input and allows the user to select regions in the alignment to compare using an AveHAS plot as a reference (Zhai and Saier 2002).

Comparison scores between two sections of the alignment are expressed in SD (Zhai and Saier 2002). IntraCompare was used to analyze each protein in each multiple alignment.

### Identification of distant homologs

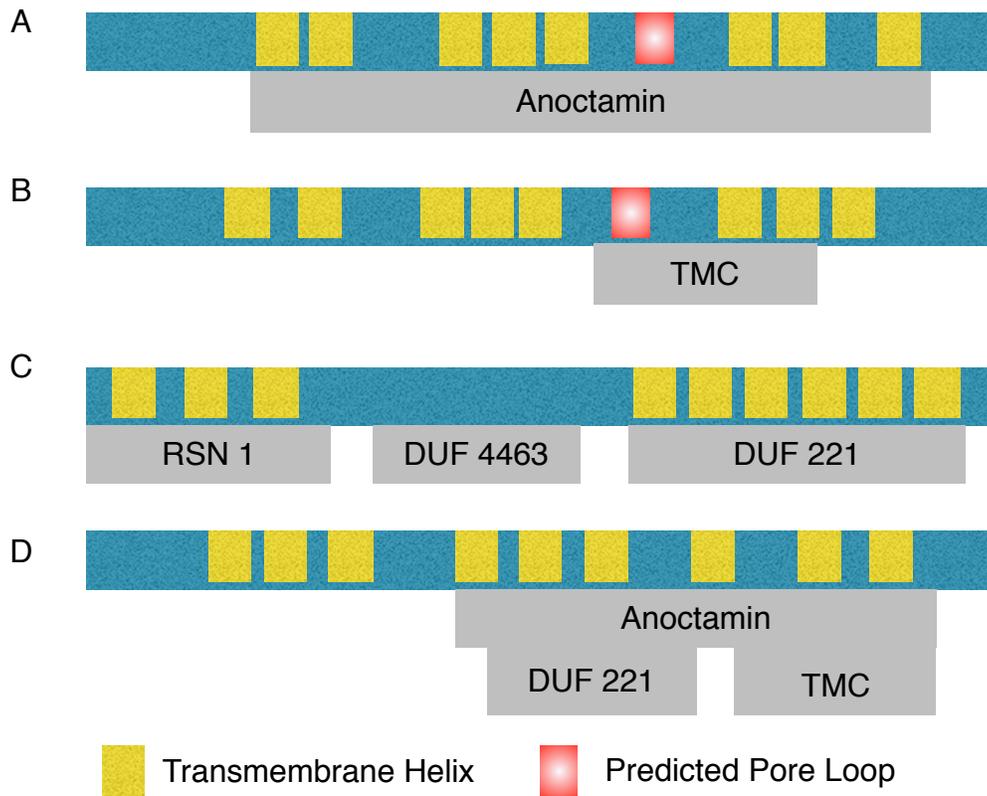
NCBI and TC PSI-BLAST searches with up to 4 iterations were used to

identify families that may be distantly related to the sequences in the Anoctamin Superfamily (TC #1.A.17). Possible candidates were used as queries for Protocol1. Then, Protocol2 was used to compare these candidate sequences to the lists of homologs created for the Anoctamin Superfamily. The top scoring binary alignments were verified using GSAT and recorded in Table 3. A negative control was used consisting of a comparison with a family that is believed to have originated via a different pathway from the other candidate families. In order to establish homology between Anoctamin families and a candidate family, the comparison scores must be three SD higher than the control (Yee et al. 2013).

## RESULTS

### Conserved Domains

The results of CDD batch searches of the Anoctamin Superfamily members have been tabulated and used to create average structural schematics for each family within the Anoctamin Superfamily (see Figure 1).



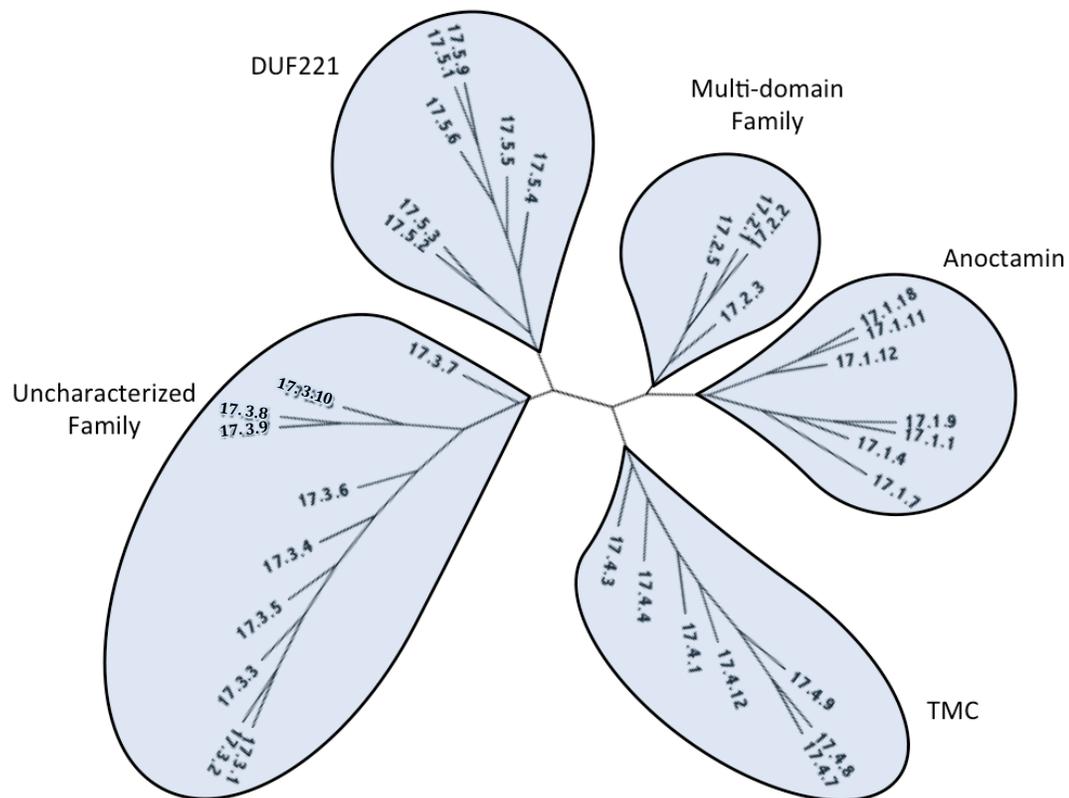
**Figure 1:** Predicted structures for various members of the Anoctamin Superfamily

In addition to the domain organizations found in the above figure, another member of the Anoctamin Superfamily was found to contain no recognized domains. The organizations depicted by C and D in the figure as

well as this extra uncharacterized member were used as a starting point in a PSI-BLAST search. Each of the new groups was expanded to contain several new proteins.

### Phylogeny of TCDB members

A phylogenetic tree of the newly expanded Anoctamin Superfamily was constructed using SuperfamilyTree.



**Figure 2:** Superfamily Tree of the Anoctamin Superfamily created by SuperfamilyTree and drawn with FigTree

In this figure, the new groups have been named Multi-domain Family and Uncharacterized Family.

## Homologs of the 1.A.17 Group

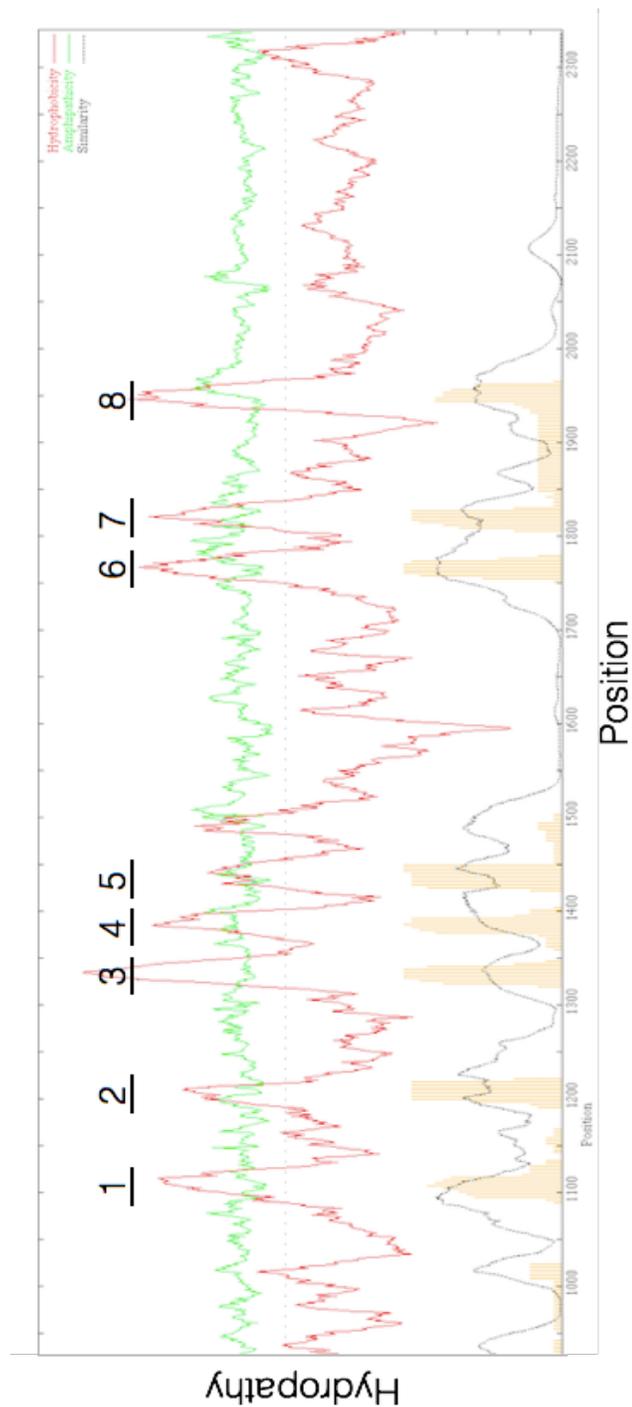
A list of homologs was compiled using a representative from each cluster of the phylogenetic tree shown in Figure 2.

**Table 1:** Average length and number of TMSs for each family in the Anoctamin Superfamily

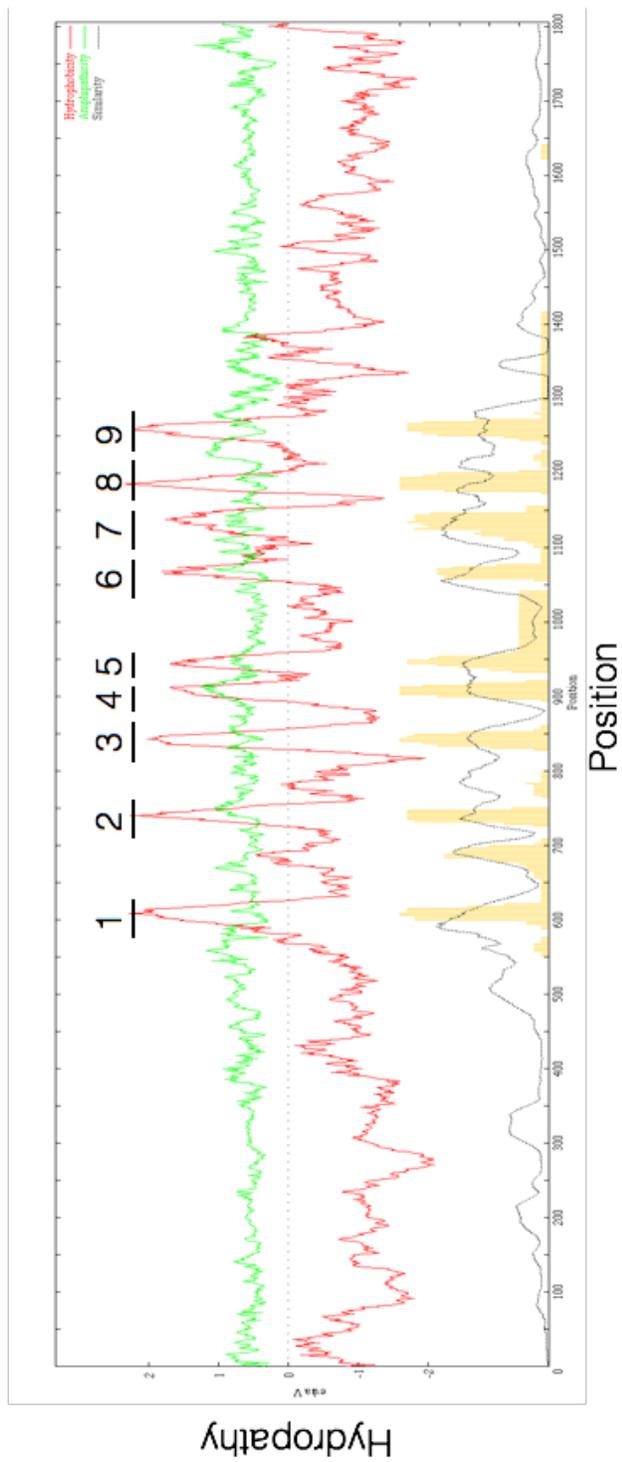
Family	Average Protein Length (a.a.)	Average number of TMSs	Organismal Groups
<b>Anoctamin</b>	897 +/- 155	8 +/- 1	Metazoa, Albunigaceae, Saprolegniaceae, Phaeophyceae, Salpingoecidae, Ichthyosporea,
<b>TMC</b>	835 +/- 143	10 +/- 1	Metazoa, Salpingoecidae, Viridiplantae, Ichthyosporea
<b>DUF221</b>	774 +/- 36	10 +/- 1	Metazoa, Viridiplantae, Fungi
<b>Multi-domain</b>	994 +/- 134	9 +/- 1	Metazoa
<b>Uncharacterized</b>	903 +/- 106	9 +/- 2	Metazoa, Viridiplantae, Fungi, Saprolegniaceae, Phaeophyceae, Pelagophyceae, Oligohymenophorea, Bacillariophyta, Spirotrichea, Eustigmatophyceae

Average hydropathy plots were drawn for each of the groups of homologs described in Table 1. These plots depict the average hydropathy value at each position on the multiple alignments created using CLUSTALX. The red lines indicate hydropathy while the green lines represent amphipathicity. Yellow bars at the bottom of the plot represent predicted TMSs while the dotted lines indicate similarity. High similarity in a region predicted to

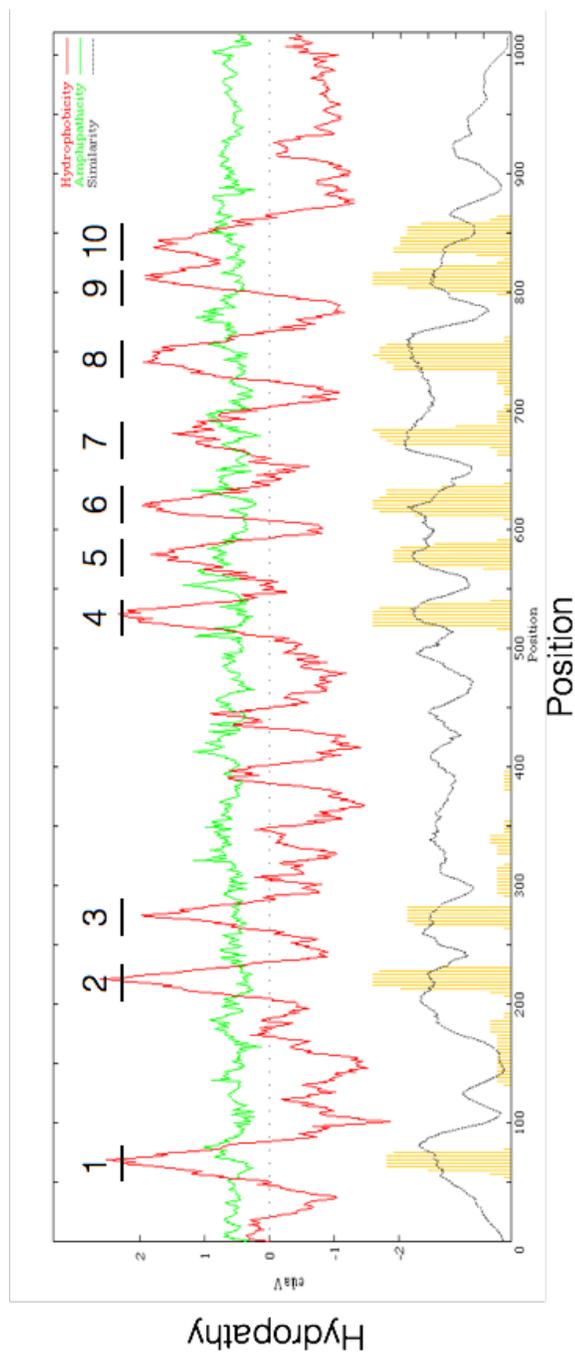
be a TMS suggests strong conservation of that TMS. Well conserved TMSs are indicated by number.



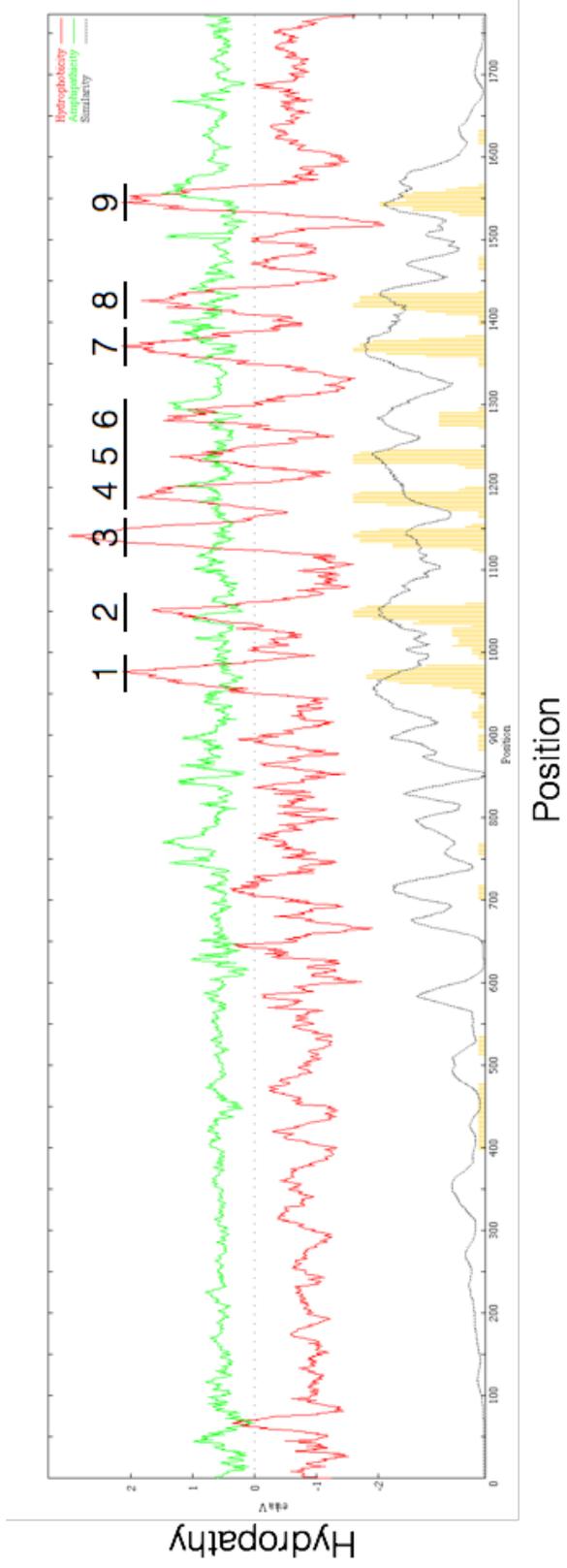
**Figure 3:** Average hydrophobicity and amphipathicity of the Anoctmamin Family (1.A.17.1) with similarity and TMS prediction.



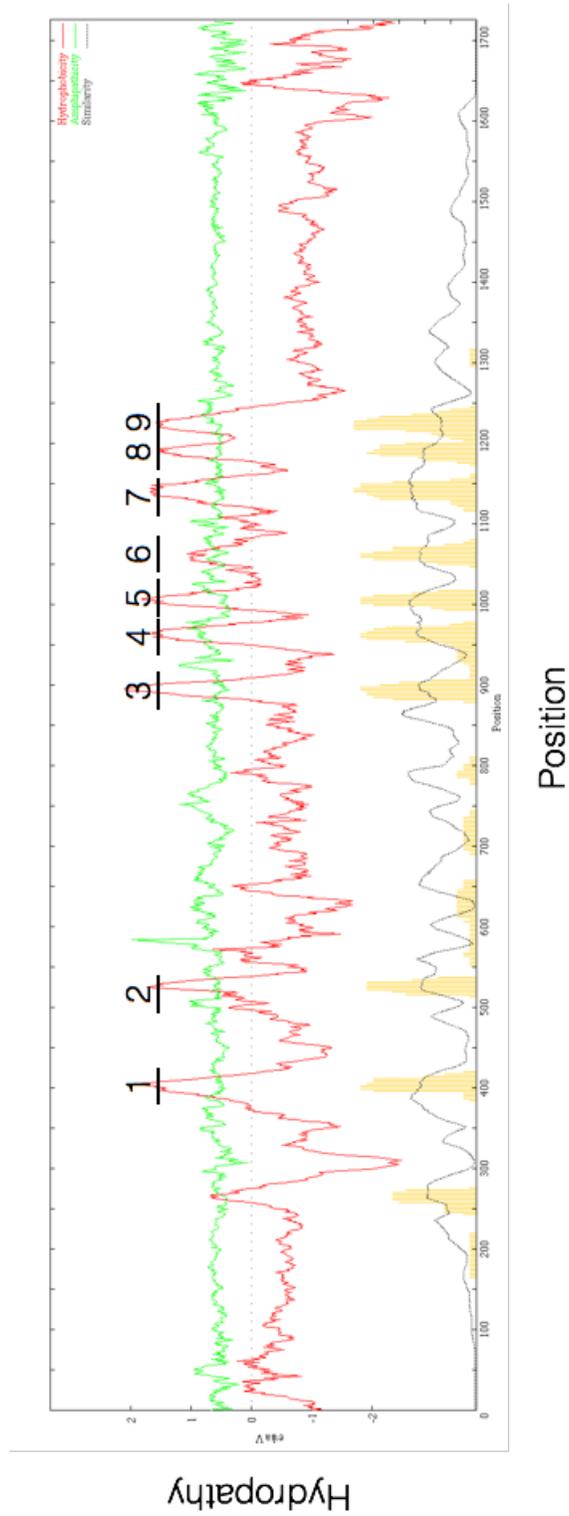
**Figure 4:** Average hydrophathy and amphipathicity of the TMC Family (1.A.17.4) with similarity and TMS prediction.



**Figure 5:** Average hydropathy and amphipathicity of the DUF221 Family (1.A.17.5) with similarity and TMS prediction.



**Figure 6:** Average hydropathy and amphipathicity of the Multi-domain Family (1.A.17.2) with similarity and TMS prediction.



**Figure 7:** Average hydrophathy and amphipathicity of the Uncharacterized Family (1.A.17.3) with similarity and TMS prediction.

## Internal Duplication

No significant data were recovered from both HHrepID and IntraCompare for any of the families in the Anoctamin Superfamily.

## Anoctamin Superfamily Comparisons

To solidify the homology of the families in the Anoctamin Superfamily, GSAT comparison scores between each family in the Anoctamin Superfamily are shown in Table 2. Significant scores appear in blue.

**Table 2:** GSAT comparison scores between each family in the Anoctamin Superfamily

	Multi-domain (1.A.17.2)	Uncharacterized (1.A.17.3)	TMC (1.A.17.4)	DUF221 (1.A.17.5)
Anoctamin (1.A.17.1)	182.4 SD	17.2 SD	25.3 SD	197.8 SD
Multi-dmoain (1.A.17.2)		19.7 SD	18.5 SD	18.3 SD
Uncharacteriz ed (1.A.17.3)			11.3 SD	14.4 SD
TMC (1.A.17.4)				14.1 SD

## Distant Families

Protocol2 was used to compare members of the Anoctamin Superfamily to possible homologs identified through use of PSI-BLAST with two iterations. Members from each family were placed in Table 3 with GSAT scores. Significant scores are highlighted. Scores within families were omitted. It

should be noted that members of DUF221, Multi-domain, and Uncharacterized families are omitted to eliminate redundancies in scores.

As a control, a member of the ABC superfamily was chosen. CFTR (TC 3.A.1.202.1) is the cystic fibrosis transmembrane regulator. It is a chloride channel found in epithelial cells as are anoctamins. It differs in that it requires ATP for activation instead of  $\text{Ca}^{2+}$  and voltage. It contains two sets of 6 TMS units with a total of 12 TMSs (Kim and Skach 2012). The highest score achieved between CFTR and any member of TC group 1.A.17 was 11 SD. The Alignments can be seen in Figures 13 through 22. A cutoff for homology of 14 SD, at least 3 SD above the control, was selected. It should be noted, however, that this is an arbitrary decision; the greater the comparison score, the better the evidence for homology.

**Table 3: GSAT comparison scores between 1.A.17 and distant families**

	1.a.17.1 (ANO)	1.a.17.4 (TMC)	1.a.1.24 (VIC)	1.a.4.1 (TRP)	1.A.4.2 (TRP)	1.a.4.4 (TRP)	2.a.36.1 (CPAT)	2.a.36.2 (CPA1)	2.a.36.4 (CPA1)	2.A.38.2 (Trk)
1.a.17.1 (ANO)		25.3 SD (5 TMSs)	11.7 SD (3 TMSs)	14.5 SD (4 TMSs)	14.5 SD (4 TMSs)	14.5 SD (4 TMSs)	13.3 SD (5 TMSs)	13.5 SD (2 TMSs)	13.5 SD (3 TMSs)	14.5 SD (2 TMSs)
1.a.17.4 (TMC)			14.7 SD (3 TMSs)	13.5 SD (4 TMSs)	16.7 SD (3 TMSs)	15.3 SD (3 TMSs)	14.1 SD (3 TMSs)	15.5 SD (4 TMSs)	17.4 SD (4 TMSs)	13.6 SD (4 TMSs)
1.a.1.24 (VIC)				12.5 SD (3 TMSs)	11.9 SD (2 TMSs)	13.2 SD (5 TMSs)	11.5 SD (3 TMSs)	11.2 SD (7 TMSs)	12.5 SD (3 TMSs)	14.4 SD (3 TMSs)
1.a.4.1 (TRP)							12.8 SD (3 TMSs)	12.3 SD (3 TMSs)	14.1 SD (3 TMSs)	14.1 SD (2 TMSs)
1.A.4.2 (TRP)							12.8 SD (3 TMSs)	12.3 SD (3 TMSs)	12.4 SD (4 TMSs)	12.8 SD (2 TMSs)
1.a.4.4 (TRP)							13.3 SD (3 TMSs)	13.5 SD (4 TMSs)	14.0 SD (4 TMSs)	11.3 SD (2 TMSs)
2.a.36.1 (CPAT)										16.0 SD (4 TMSs)
2.a.36.2 (CPAT)										11.6 SD (3 TMSs)
2.a.36.4 (CPAT)										14.0 SD (2 TMSs)
2.A.38.2 (Trk)										

## DISCUSSION

### Domains

The predicted TMS arrangements of the five families in the Anoctamin Superfamily (1.A.17) are depicted in Figure 1, showing four primary organizational patterns. In the Anoctamin Family, a single anoctamin domain was recognized that contained all 8 TMSs and the predicted pore loop. TMCs contained a single TMC domain that contains the predicted pore loop as well as the two TMSs on the C-terminus of the loop.

The DUF221 Family members contain a variety of recognized domains. First, all DUF221 proteins in 1.A.17 contained a DUF221 domain consisting of TMSs 4-9. DUF221 members are of unknown function, so it is difficult to assess the role of these proteins without examining some of the other recognized domains. A second domain found in most DUF221 proteins is DUF4463 in the hydrophilic region of the protein. Furthermore, each DUF221 protein contained a third domain recognized as RSN1 at the N-terminus, which contains TMSs 1-3. RSN1 has an unknown function, but experiments in yeast have shown that Sro7P-deficient mutants exhibit increased sensitivity to NaCl concentrations because Sro7P is responsible for localizing sodium pumps to the cell membrane to remove excess Na<sup>+</sup> from the cell. Overexpression of RSN1 has been shown to re-route these sodium pumps to the plasma membrane, restoring NaCl tolerance (Wadskog et al. 2006). Because these

domains (DUF221, DUF 4463, and RSN1) appear in nearly all DUF221 proteins, it may be deduced that the three domains function together, possibly to regulate Na<sup>+</sup> concentrations either as a cation channel or to allow other proteins to localize to the plasma membrane.

One member of this superfamily, 1.A.17.2.1, was found to contain all three domains found in anoctamin, TMC, and DUF221. Upon a BLAST search, multiple homologs to this protein were found to contain the same three domain organization. The anoctamin domain is the largest and occupies the region containing TMSs 4-9. DUF221 and TMC domains occupy smaller regions, TMSs 4-6 and TMSs 8-9, respectively. The recognized anoctamin domain in this protein is shorter than the domain recognized in other anoctamin proteins, containing 6 TMSs instead of 8. Judging from the arrangement of TMSs, it is predicted that the first 2 TMSs from a complete anoctamin domain were removed, leaving TMSs 3-8 in this protein. While the function of the proteins in this group is unknown, we predict that it acts as a channel because the anoctamin domain contains the TMSs associated with the pore region in these unique proteins. This arrangement also supports the well documented claim that Anoctamin, TMC, and DUF221 families are evolutionarily related, since there is overlap in these domains. As seen in the phylogenetic tree in Figure 2, proteins of the 1.A.17.2 group are clustered in a group apart from the adjacent Anoctamin Family. This suggests that the proteins in this group are related to

the Anoctamin Family, but divergent enough to comprise a new family. We propose the name Multi-domain Family for the group containing these proteins to reflect their unique domain organization. More research must be conducted on these proteins to reveal the details of their functions.

A final group of proteins, represented by group 1.A.17.3, did not contain a recognizable domain. Because of this, it is difficult to ascertain a function and we therefore propose the name Uncharacterized Family for this group.

## SuperfamilyTree

The phylogenetic tree presented in Figure 2 contains each family in distinct clusters. As shown in this figure, the members of the Anoctamin Superfamily form five distinct groups: Anoctamin, TMC, Uncharacterized Family, DUF221, and Multi-domain Family. The group on the tree containing the Uncharacterized Family shows the highest amount of sequence divergence.

One protein from each cluster was chosen as a query for Protocols 1 and 2. The selected proteins were 1.A.17.1.1 (ANO), 1.A.17.4.6 (TMC), 1.A.17.5.1 (DUF221), 1.A.17.2.1 (Multi-domain), and 1.A.17.3.1 (Uncharacterized).

An alignment was created for each list of homologs and was edited to remove sequences that introduced large gaps. The average number of TMSs

found in this list of anoctamin homologs was 8 as expected. In this case, the programs used to calculate topology could decipher between TMSs and predicted pore loops. This fact gave us more confidence in the accuracy of our topology prediction software. TMC, however, showed an average of 10 TMSs. Since anoctamins and TMCs are homologous, one of the extra TMSs may be a pore-lining loop and was therefore incorrectly identified as a TMS. The remaining extra TMS appears as a small hydrophobic peak between TMSs 1 and 2 in the AveHAS plot in Figure 4. This plot suggests that TMCs have a greater diversity in topology than anoctamins. While the vast majority of proteins in the Anoctamin Family are predicted to have 8 TMSs, the list of homologs for the TMC Family has a range of 8-12 TMS. The extra TMSs may be the result of partial duplications or addition of extra domains. This, however, is difficult to assess because the proteins that contain extra TMSs do not have additional recognizable domains. We predict that these proteins represent splice variants of TMCs. The most conserved TMSs (of which there are 8) are apparent in the AveHAS plot.

From the AveHAS plots for the five families, the arrangement of TMSs can be seen. The Anoctamin Family in Figure 3 shows an arrangement of 2-3-3 TMSs, with a region of hydrophobicity and amphipathicity directly following the fifth TMS. The TMC Family in Figure 4 exhibits a similar arrangement with 2-3-4 TMS configuration, with the 6<sup>th</sup> TMS expected to be a pore-lining helix.

The AveHAS plot in Figure 5 predicts a structure for DUF221s that is different from that of the anoctamins. Here, we see a 1-2-7 TMS arrangement. Based on this arrangement, as well as the data collected regarding conserved domains, DUF221 proteins may have arisen through a fusion of a DUF221 domain containing 6 TMSs similar in sequence to 6 of the TMSs in ANO with another domain containing 3 TMSs. More research is needed to define the evolutionary pathways taken for the DUF221 and anoctamin homologues.

The homologues of the Multi-domain Family shown in the plot in Figure 6 have a similar arrangement to the Anoctamin Family, although the predicted pore loop was counted as TMS 6 by the program. The similarity in TMS arrangement between the Multi-domain Family and the Anoctamin Family is substantiated by the close proximity of the two families in the superfamily tree in Figure 2. The Uncharacterized Family shown in Figure 7 shows an arrangement most similar to the arrangement of the DUF221 Family. Both contain 1-3 TMSs followed by a hydrophilic region, followed by 7 TMSs at the C-terminus. These two families also occupy adjacent locations on the superfamily tree.

Table 1 presents information about average length and species distribution for each of the five families in the Anoctamin Superfamily. The species distribution of TMC homologs is more diverse than previously thought. Originally, it was thought that TMCs were only found in mammals and

choanoflagelates (Hahn et al. 2009). In our list of homologs, we see that in addition to these two groups, TMC homologs are also present in Ichthyosporea and Viridiplantae. The Uncharacterized Family shows the greatest species diversity of the superfamily. In addition to animals and fungi, there are many species of the protozoa and viridiplantae groups.

## Homology

As seen in Table 3, several distant families of transporters and channels are suspected to be related to the 1.A.17 group. Below are the comparisons from the table with high scores.

## VIC

The Voltage-gated Ion Channel (VIC) Family consists primarily of ion selective channels with 6 TMSs as shown in Figure 9. The first four TMSs form the voltage sensor while the last two TMSs form the pore of the channel (Shimomura et al. 2011). The 1.A.1.24 group of proteins in the VIC Superfamily is composed of potassium channels. In Figure 15, a score of 14.8 SD was found in the comparison of TMC with VIC. Here, TMSs 1-3 of a TMC align well with the region containing TMSs 3-6 in the VIC protein. This corresponds with the pore region and part of the sensor region of the VIC Family member.

## TRP

The Transient Receptor Potential  $\text{Ca}^{2+}$  Channel (TRP) Family of receptors resembles the VIC proteins and is a member of the VIC Superfamily (Chang et al. 2004). They have a voltage sensor region in the first four TMSs and a pore region in the next two TMSs (TMSs 5-6). The AveHAS plot for the TRP Family in Figure 10 indicates an extra TMS in some TRP proteins at their N-termini. This region is not conserved in the VIC Superfamily. These channels transport calcium and sense various stimuli including pain, heat, and pH (Hu et al. 2011). The comparison of a TMC to a TRP in Figure 16 gave a score of 16.7 SD. TMSs 4-6 of the TMC align with TMSs 4-6 of the TRP. Furthermore, the predicted pore region of TMC is located in the same region as the pore region of the TRP channel. Additionally, the comparison of an Anoctamin with a TRP in Figure 17 gave a score of 14.5 SD. TMSs 3-6 of the TRP protein aligned with TMSs 4-7 of an Anoctamin in an arrangement similar to the TMC-TRP comparison. The evidence that Anoctamins are voltage sensitive (Galiotta 2009) supports this finding.

## CPA1

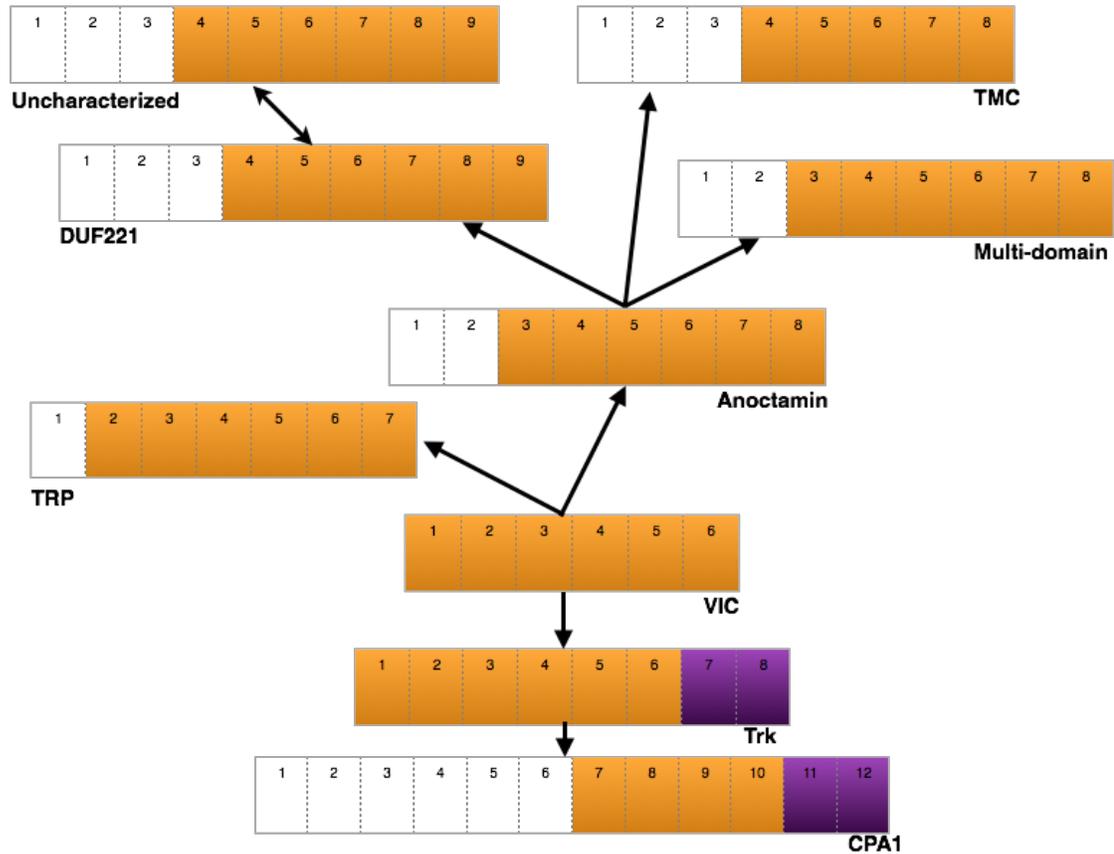
The Cation-Proton Antiporter 1 (CPA1) Family are transporters that exchange  $\text{H}^+$  and  $\text{Na}^+$ . They are predicted to have 10-12 TMSs as well as a pore region between TMSs 9 and 10 (Wells and Rao 2001). The comparison

of a TMC with a CPA1 protein in Figure 18 gave a score of 17.4 SD. TMSs 8-12 of CPA1 align with TMSs 6-9 of TMC. The pore region of CPA1 (TMSs 9-10) aligns well with the pore region of TMC (TMSs 7-8).

## Trk

The Potassium Transporter (Trk) Family of the VIC Superfamily contains an average of 8 TMSs with four pore-forming regions (Kato et al. 2007). Significant scores were found when comparing a Trk with an anoctamin, a TRP, and a CPA1. The anoctamin comparison with a Trk in Figure 20 exhibited a score of 14.5 SD. TMSs 7-9 of anoctamin (this protein has an additional N-terminal TMS) align with TMSs 7-9 of the Trk protein (This protein also has an additional N-terminal TMS). Both regions contain predicted pore regions. The comparison of a TRP with a Trk in Figure 21 exhibited a score of 14.1 SD. TMSs 5-6 of the TRP (the pore region) align with TMSs 7-8 of the Trk protein (the pore region). Finally, CPA1 aligned with Trk in Figure 22 provided a score of 16 SD. TMSs 10-13 of the CPA1 align with TMSs 5-8 of the Trk Family protein.

## Possible evolutionary pathway



**Figure 8:** Predicted evolutionary pathway for the appearance of members of the Anoctamin Superfamily and other families from a primordial 6 TMS VIC Family member

Figure 8 presents a prediction of the evolutionary pathway to the Anoctamin Superfamily containing Anoctamin, TMC, DUF221, Multi-domain, and Uncharacterized families. The pathway was estimated using information gathered from GSAT binary alignment scores, TMS overlap, and species distribution. The white boxes represent TMS additions to the N-terminus of the protein, while purple boxes represent additions to the C-terminus. Since the

VIC Superfamily contains the smallest number of TMSs and have the largest species distribution (found in all domains of life), it is predicted that it is the starting protein group in the pathway. From this group, additions at the N- and C-termini resulted in the topologies of proteins in other related families. The most conserved region found in the anoctamin-TMC group is the domain containing the pore of the channel as well as several TMSs located directly N-terminal to the pore. This is the same 6 TMS core region that is conserved in DUF221-containing proteins. We predict that this same region is conserved in most members of the VIC Superfamily as well since the TMS corresponding to this region has the highest GSAT scores. It is also likely that anoctamins have conserved the VIC sensor responsible for responding to changes in membrane potential since this region overlaps between VIC and ANO/TMC proteins.

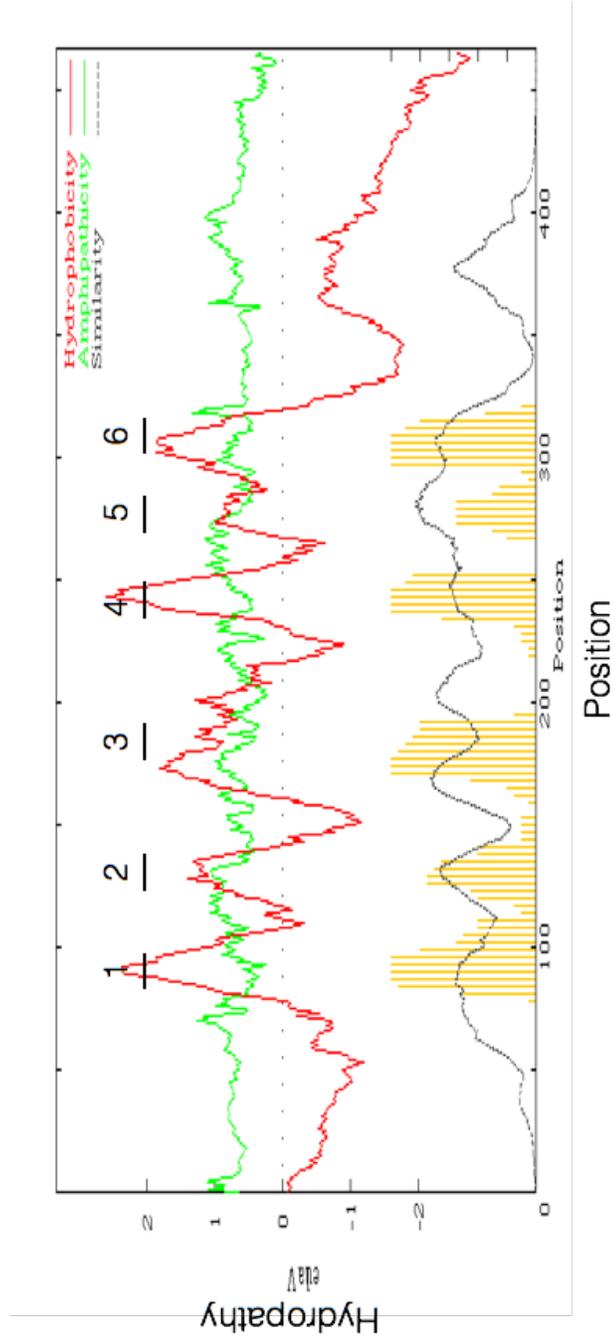
## CONCLUSION

The Anoctamin Family of proteins is closely related to both the Transmembrane Channel-like Family and DUF221 Family whose functions are largely unknown. These three families are grouped into a larger superfamily which we have called the Anoctamin Superfamily. In addition to these three families, we have found that two new families also belong to the superfamily. We named them the Multi-domain Family and the Uncharacterized Family. Here, we have analyzed the topologies of the proteins of each of these families using Bioinformatic programs included in the BioV Suite: AveHAS, GSAT, CLUSTAL X, and SuperTree. We have constructed structural schemes for each of these families and deduced a possible function for the DUF221 Family in NaCl tolerance. We have also found several families of proteins (CPA, VIC, and TRP) that are homologues to members of the Anoctamin Superfamily. These homologous families are important because they reveal the origins of and provide insight into possible mechanisms for many of the proteins in the Anoctamin Superfamily. We have found that there can be evolutionary relationships between channels and carriers with different ion selectivities. Mechanisms can be conserved even between evolutionary divergent families, as observed regarding the voltage activation seen in both VIC and Anoctamin families. When using homology approaches to find drug targets, it may be more important to consider these mechanisms instead of ion

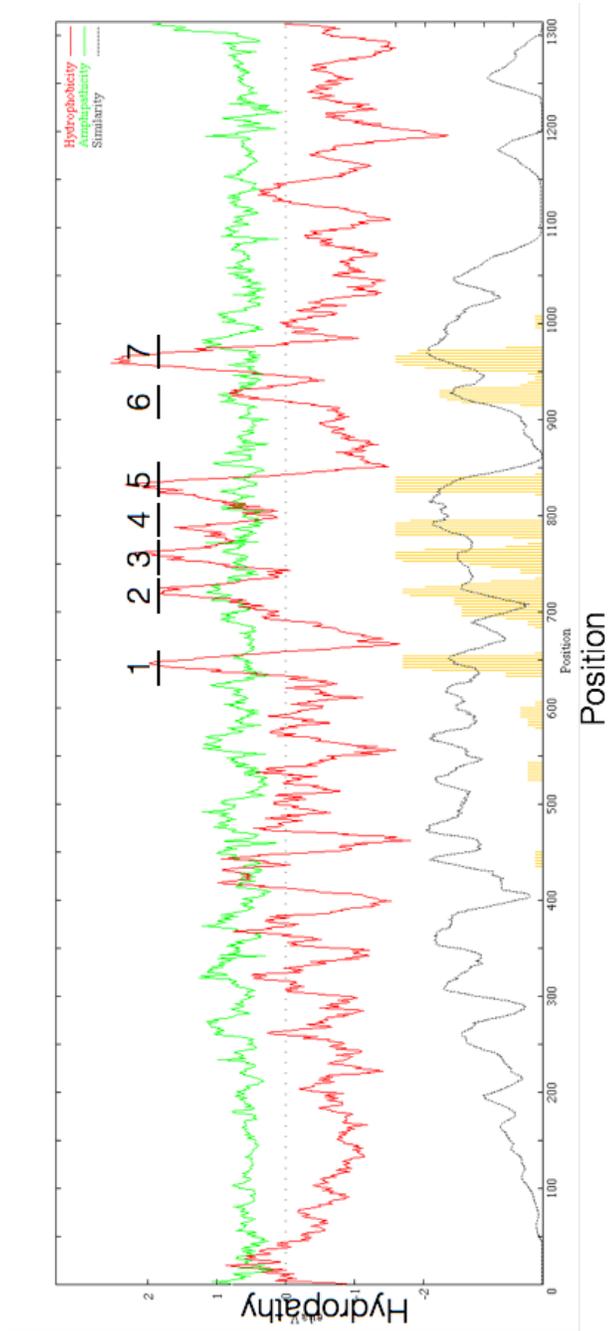
selectivity. Understanding which domains of each protein share homology should allow us to find therapies to target some proteins that are important in disease progression.

Because little information was gathered on the possibility of internal repeats within the Anoctamin Superfamily, it is important to further research this matter. This is important because it will allow us to further establish the relationships between each distant family. If we know that the families arose from a similar duplication pattern through evolution, it will provide more evidence that they are related.

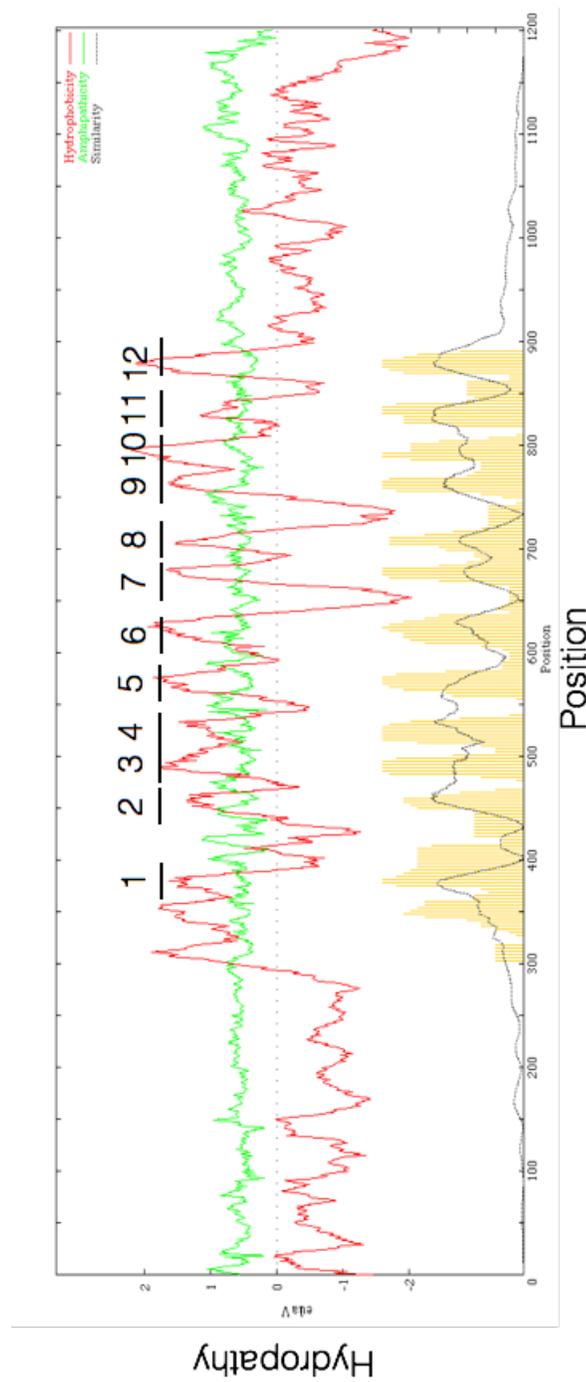
APPENDIX



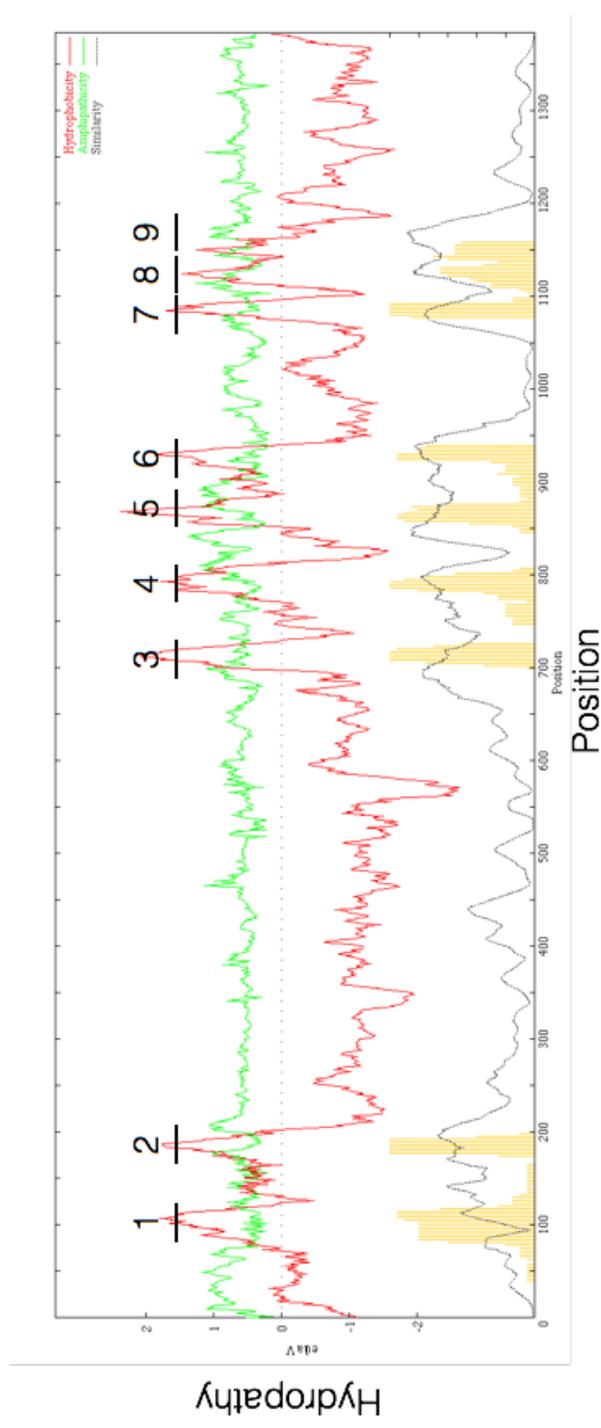
**Figure 9:** Average hydrophobicity and amphipathicity of the VIC Family (1.A.1) with similarity and TMS prediction.



**Figure 10:** Average hydropathy and amphipathicity of the TRP Family (1.A.4) with similarity and TMS prediction.



**Figure 11:** Average hydropathy and amphipathicity of the CPA1 Family (2.A.36) with similarity and TMS prediction.



**Figure 12:** Average hydrophobicity and amphipathicity of the Trk Family (2.A.38) with similarity and TMS prediction.

```

ABC (3.a.1) vs TMC (1.A.17.4)

Sequence A (ABC-CFTR) gi|513007829|
Sequence B (TMC)      gi|556725851|

# Aligned_sequences: 2
# 1: A_Sequence
# 2: B_Sequence
# Matrix: EBLOSUM62
# Gap_penalty: 8.0
# Extend_penalty: 2.0
#
# Length: 223
# Identity:      60/223 (26.9%)
# Similarity:   92/223 (41.3%)
# Gaps:         30/223 (13.5%)
# Score: 93.0
#
#=====
A_Sequence      1  MGQLLSPGSACVLIFFSGI--HGGDACRRTSFMKTKFVRVSIIVKLFFSW      48
   : ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
B_Sequence      1  LAYLLSP-LACLLLCFCGIIIRRMVKGLPQKMFLGQDY-RSPLSAKVFSW      48
   2
A_Sequence      49  ---IRPILKKGYRKRLELSDIYQIPSADSADNLSEKLEREWDRVSKKN      95
   || : || | | | | | | | | | | | | | | | | | | | | | | | |
B_Sequence      49  DFCIRG-QEAATIKRHEISNEF-----KMELEEGRHLLLLQ-      83

A_Sequence      96  PKLINALRRCFLWRFTFYGILLYLGEVTKAVQPLLLGRIIASYDPDNKEE      145
   : | | | | | | | | | | | | | | | | | | | | | | | | | | |
B_Sequence      84  -QQTRAQRACHLLTYLRVNILIGL-LVVGAI SAIFWA---TKYSQDNKEE      128
   3
A_Sequence      146 RSI AI--YLAIGLCLLFIVRTL L LHPAIFGLHHIGMQMRIAMFSLIYKKT      193
   : | | | | | | | | | | | | | | | | | | | | | | | | | | |
B_Sequence      129 SLFLLLQYLPPGVIALVNFLGPLL FVFLVQLENYPPNTEVNLTLIWCVV      177
   4
A_Sequence      194 LKLSSRVLDKISIGQLVNLSNN      216
   ||| : | | | | | | | | | | | | | | | | | | | | | | | |
B_Sequence      178 LKLASLGMFSFSLGQTVLCIGRN      200

#-----

===== FINISHED =====
Average Quality (AQ)      16.55 +/- 6.89
Standard score (Z): 11.0
Precise score (Z): 11.089

```

**Figure 13:** GSAT binary alignment Control: CFTR (ABC Superfamily) compared with TMC

```

TMC (1.A.17.4) vs ANO (1.A.17.1)

Sequence A (TMC) gi|470243840|
Sequence B (ANO) gi|67906860|

# Matrix: EBLOSUM62
# Gap_penalty: 8.0
# Extend_penalty: 2.0
#
# Length: 293
# Identity:      78/293 (26.6%)
# Similarity:   139/293 (47.4%)
# Gaps:         57/293 (19.5%)
# Score: 201.0
#
#
#=====
A_Sequence      1 RVSSRSRAFRFAVSYSVFIILIGISAICT--VY--VLKT-----VN      37
  : | | | | : | | | | : : : : | | | | | : | | : :
B_Sequence      1 KFSKRFPFRFSTSVPTIAFMILLVMAAVMTVIVYRIVVKTAIFAIDQEFIS    50
                                     4
A_Sequence     38 NYQFLANILTTSLVSAIILRVINVIYEFMCLYLTKLEGHRTWTNFRNHNT    87
  : | : : | : | : : : : : : | : : : | | | | | : : |
B_Sequence     51 SYASIIITSVTASMISLILIMILQILYERIAVWLTNLELHRTETEYEDSFT  100
                                     5
A_Sequence     88 LKLFVFKVINV--LVLYIL-----RD--KIFRDI--TDQVVQAGC    121
  : : : | : | | | | | | | | | | | | | | | | | | |
B_Sequence    101 FKMYLFAFVNYYSTSFYIAFFKGRLLPGTTPADYGRVF-GIWRQEECDPAGC  149
                                     6
A_Sequence    122 P---LIDV----GSQFF--FILLMDLTLQNIWEIVYVSGMARIGRCIGK    161
  | : : | | | | | : | | | | | | | | | | | | | |
B_Sequence    150 TQELFINIAITMCGKQFFNNFMELAMPVLMNFWR-----SRTGRKEEK    192
                                     7
A_Sequence    162 RGKKSTDSYKPEFDLAE-----EYLELLYRQFIVYLGIPYIPLVAVF    203
  | | : : : | | | : | | | | : : : : : : : : | | :
B_Sequence    193 SGKGRYEQWEQDADLADLGRGLFKEYLEMVVQFGFSTIFVAAPLAPLF    242
                                     7
A_Sequence    204 GVLNLAEFYVDKFRLLKICRRPHRVQGSMKRFLSFYLLIIAV      246
  : | | | | : | : : : | | | | | | | | : : : : |
B_Sequence    243 ALLNNLVEVRLDAYKFISQLRRP--VAKRAQDIGAWYAILVTV      283
                                     8
#-----

===== FINISHED =====
Average Quality (AQ)      17.97 +/- 7.25
Standard score (Z): 25.0
Precise score (Z): 25.259

```

**Figure 14:** GSAT binary alignment between a TMC and an ANO protein.



```

TMC (1.A.17.4) vs TRP (1.A.4.2)

Sequence A (TRP) gi|260814083|
Sequence B (TMC) gi|488557315|

# Matrix: EBLOSUM62
# Gap_penalty: 8.0
# Extend_penalty: 2.0
#
# Length: 217
# Identity:      47/217 (21.7%)
# Similarity:   94/217 (43.3%)
# Gaps:         28/217 (12.9%)
# Score: 108.0
#
#
#=====
                                4
A_Sequence      1  LLGIAIIGSWFYIIFYCRGWTLTGPFV--TMIYKMIAGDMARFGIITYCTF      48
                  |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
B_Sequence      1  LVGVNVLRL LSSMYFQC--WAVTSSNVIPHERVFKASRSDNFYMLLLLLVL      48
                                4
                                5
A_Sequence      49  IITFAPAFWIMFRDKPTEKFD-SEFGS---LMTLFQMTLGEFDYPVFNLA      94
                  ::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
B_Sequence      49  LLSLLPVAWTVLALPPS--FDCGPFSGKNRMYDVIQETI-ENDFPFPAF---      92
                                5
                                6
A_Sequence      95  EFPALAKLL-FVVFMLVPIMLLNMLIAMMSNTFQEVVDRSEREWRRQK      142
                  |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
B_Sequence      93  ---LGKIFAFIANPGLIIPAILLMVLAIYYLNSVSKSLARANAQLRK-K      137
                                6
A_Sequence      143 AKIIVSLERSFSKSSLQQFRKDYSIQLR--PDSAVHLPV-EPG--PHGLR      187
                  ::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
B_Sequence      138 IQVLREVEK--SHKSVRATARDSXGKPRGSPGSATQCPLTREGAKPHSAS      185

A_Sequence      188 ALLVIKEEAMSRACKKS      204
                  : : : | : |
B_Sequence      186 QIQGLDKRAQGPXSRS      202

#-----

===== FINISHED =====
Average Quality (AQ)      11.38 +/- 5.78
Standard score (Z): 17.0
Precise score (Z): 16.726

```

**Figure 16:** GSAT binary alignment between a TMC and a TRP protein.





```

TRP (1.A.4.4) vs CPA1 (2.A.36.4)

Sequence A (TRP) gi|573063959|
Sequence B (CPA1) gi|325109240|

# Matrix: EBLOSUM62
# Gap_penalty: 8.0
# Extend_penalty: 2.0
#
# Length: 187
# Identity:      48/187 (25.7%)
# Similarity:   79/187 (42.2%)
# Gaps:        27/187 (14.4%)
# Score: 94.0
#
#
#=====
A_Sequence      1 SKLTMSMQLDSNDE-EEDENDPTAPFFINLNALE-IAAIAD-----AKRFL      43
| :|:  :| | | | | | | | | | | | | | | | | | | | | | | | | |
B_Sequence      1 SGVTLPQKISSTLEVESGSNDPMAIFLTIGICIELIAQRMDFGPALLLLFL      50

                                     1
A_Sequence      44 SOHVVQKIITGIWSGDIIFW--DRLSAGSIKKPRFYNPQTADPFCLRV      91
| | | | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
B_Sequence      51 VQMVV-GLLVGVGVGALAVVVNRIQLGS--AGLF--PVLVSAFCLLSFG      95
                                     7

                                     2                                     3
A_Sequence      92 RYLKVYEVVFFGLFLFLYYAVLIEQNRYAITPREIMLY---VWFAAFCY      137
| | | | | :| | | | | :| :| :| :| :| :| :| :| :| :| :|
B_Sequence      96 --LSA-NLGGSG-FLAVYLSGLVLGNRRLIFQIRLYHDAIAWLA---      137
                                     8                                     9

                                     4
A_Sequence      138 DEVSEYIDAGSIFYSDVWNLFDMIMIAIGLVFAILR      174
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
B_Sequence      138 -QILMFVMLGLLSFPSRLWAVSGKALLVA AVLILVAR      173
                                     10

#-----

===== FINISHED =====
Average Quality (AQ)      14.29 +/- 5.68
Standard score (Z): 14.0
Precise score (Z): 14.039

```

**Figure 19:** GSAT binary alignment between a TRP and a CPA1 protein.

```

ANO (1.A.17.1) vs Trk (2.A.38.2)

Sequence A (ANO) gi|556514639|
Sequence B (VIC) gi|365983958|

# Matrix: EBLOSUM62
# Gap_penalty: 8.0
# Extend_penalty: 2.0
#
# Length: 232
# Identity:      63/232 (27.2%)
# Similarity:   108/232 (46.6%)
# Gaps:         49/232 (21.1%)
# Score: 103.0
#
#
#=====
A_Sequence      1 ADG--FPTLKKD---DLHTPGTEASDGKSSPNKST-----ETKITR      36
  |||  |||: | :| | | :| | | | | | | | | | | | | | | | |
B_Sequence      1 ADGQHSPITLEDNDDEELEEDEDEEESKSLASSNSSTSHAIGKERRESNTSK  50

                                     7
A_Sequence      37 AEQEEIMPKYDDATNDYLEMFI--QFGFVSMFTCVFPICGLLAFLNILE      84
  :::: : ||::| | : : : | | | | | | | | | | | | | | | |
B_Sequence      51 KKKKKNKKEEDDSSNSYIAIHLRRQLSF-DMW--FMFLGL--FIICICE      94
                                     8
                                     8
A_Sequence      85 IRSDAFKLLTNFORPFSETARGIGVWERAFEIVS-YVAIAVNIIGLIGVSG    133
  : | : | | | | | | | | | | | | | | | | | | | | | | |
B_Sequence      95 ----SGKI----QDP--EMP-AINVFSILFEIVSAYGTVGLSLGYPGTTT    133
                                     9
                                     9
A_Sequence      134 SIQALFPGLTTHQYILLL IAVEHILFALRFGLSHIVPPIPFALERSIAIL    183
  | : | | | : | : | : | : | : | : | : | : | : | : |
B_Sequence      134 SLSAQFTALSKLVII VMLVGRN-----R-GL-----PYALDRAI-ML    169
                                     10

A_Sequence      184 EHKRREALRTLE-RECRKEARRVSDENDSPDY      214
  : | | : : | : | : : : : | : | |
B_Sequence      170 PSERLERIDDIEDSKLRRQPKNIGKEDPVSGY      201

#-----

===== FINISHED =====
Average Quality (AQ)      15.06 +/- 6.08
Standard score (Z): 14.0
Precise score (Z): 14.460

```

**Figure 20:** GSAT binary alignment between an ANO and a Trk protein.

```

TRP (1.A.4.1) vs Trk (2.A.38.2)

Sequence A (TRP)  gi|313232667|
Sequence B (VIC)  gi|448114292|

# Matrix: EBLOSUM62
# Gap_penalty: 8.0
# Extend_penalty: 2.0
#
# Length: 88
# Identity:      30/88 (34.1%)
# Similarity:   44/88 (50.0%)
# Gaps:        8/88 ( 9.1%)
# Score: 87.0
#
#
#=====
A_Sequence      1 YMIYQNETSKCEEMTVNKNMYWSHGLMCISFVLYCLRLLHIFTMSHNLGP      50
  | :  : : | | | | :  | | :  | | | : : | : | :  |
B_Sequence      1 YVHGKDEVSKHEDDNTQKNFIGAHLRROLSFDLWFI-FLGLFICIAEGS      49
                                     7

A_Sequence      51 KLVMIRKMMKDI-FVLLAILAVFILSYGVVSQALLYPN      87
  | :  | | | | | :  | | | :  | | | :  | | |
B_Sequence      50 KI-----KKDINFSIFAILFEIISAYGTVGLSLGYPN      81
                                     8

#-----

===== FINISHED =====
Average Quality (AQ)      11.45 +/- 5.34
Standard score (Z): 14.0
Precise score (Z): 14.149

```

**Figure 21:** GSAT binary alignment between a TRP and a Trk protein.

CPA1 (2.a.36.1) vs Trk (2.a.38.2)

Sequence A (CPA1) gi|528290185|  
Sequence B (VIC) gi|575059220|

# Matrix: EBLOSUM62  
# Gap\_penalty: 8.0  
# Extend\_penalty: 2.0  
#

# Length: 147  
# Identity: 44/147 (29.9%)  
# Similarity: 65/147 (44.2%)  
# Gaps: 14/147 (9.5%)  
# Score: 107.0  
#  
#

```

=====
                        10                      11
A_Sequence      1  YIYLGSLSLFTE SKLEFKPLF-- IIVTVFGICAARWTAVF-- PLSRAINWF      46
                  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
B_Sequence      1  YTN T G M S L S D E S M V P F Q Q A Y P V I L V M V F L V L A G N T A F V S H G P P D N G T D W F      50
                   5                      6

                        12
A_Sequence      47  I R Y R A R S K G K E A A D E L P Y N Y Q A M L F W A G L R G A V G V A L A A G M T G A N S W A L -      95
                   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
B_Sequence      51  F - F L V L D I G N P A I D S I P V G V R - - F I D G L L Q A V A V R - A A G F A T V T L S A L A      95
                                                7

                        13
A_Sequence      96  - K A T V L V V V V L T V I I F G - G T T A R M L E I L G I R T - G V V - E E I D S D D E F D      138
                   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
B_Sequence      96  P A V K V L V V I M M Y I S V Y P I A M S V R S T N V Y E E K S L G I Y P E E D S D D E D D      142
                                                8

```

#-----

===== FINISHED =====  
Average Quality (AQ) 13.56 +/- 5.82  
Standard score (Z): 16.0  
Precise score (Z): 16.048

**Figure 22:** GSAT binary alignment between a CPA1 and a Trk protein

## REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, **25**:3389-402.
- Bateman, A., Coggill, P. & Finn, R. D. 2010. DUFs: families in search of function. *Acta crystallographica. Section F, Structural biology and crystallization communications*, **66**:1148-52.
- Biegert, A. & Soding, J. 2008. De novo identification of highly diverged protein repeats by probabilistic consistency. *Bioinformatics*, **24**:807-14.
- Chang, A. B., Lin, R., Keith Studley, W., Tran, C. V. & Saier, M. H., Jr. 2004. Phylogeny as a guide to structure and function of membrane transport proteins. *Molecular membrane biology*, **21**:171-81.
- Chatzigeorgiou, M., Bang, S., Hwang, S. W. & Schafer, W. R. 2013. tmc-1 encodes a sodium-sensitive channel required for salt chemosensation in *C. elegans*. *Nature*, **494**:95-9.
- Chen, J. S., Reddy, V., Chen, J. H., Shlykov, M. A., Zheng, W. H., Cho, J., Yen, M. R. & Saier, M. H., Jr. 2011. Phylogenetic characterization of transport protein superfamilies: superiority of SuperfamilyTree programs over those based on multiple alignments. *Journal of molecular microbiology and biotechnology*, **21**:83-96.
- Duran, C. & Hartzell, H. C. 2011. Physiological roles and diseases of Tmem16/Anoctamin proteins: are they all chloride channels? *Acta pharmacologica Sinica*, **32**:685-92.
- Galletta, L. J. 2009. The TMEM16 protein family: a new class of chloride channels? *Biophysical journal*, **97**:3047-53.
- Hahn, Y., Kim, D. S., Pastan, I. H. & Lee, B. 2009. Anoctamin and transmembrane channel-like proteins are evolutionarily related. *International journal of molecular medicine*, **24**:51-5.
- Horton, J. S. & Stokes, A. J. 2014. The transmembrane channel-like protein family and human papillomaviruses: Insights into epidermodysplasia verruciformis and progression to squamous cell carcinoma. *Oncoimmunology*, **3**:e28288.

- Hu, H., Bandell, M., Grandl, J. & Petrus, M. 2011. High-Throughput Approaches to Studying Mechanisms of TRP Channel Activation. *In*: Zhu, M. X. (ed.) TRP Channels. Boca Raton (FL).
- Kato, N., Akai, M., Zulkifli, L., Matsuda, N., Kato, Y., Goshima, S., Hazama, A., Yamagami, M., Guy, H. R. & Uozumi, N. 2007. Role of positively charged amino acids in the M2D transmembrane helix of Ktr/Trk/HKT type cation transporters. *Channels*, **1**:161-71.
- Kim, K. X. & Fettiplace, R. 2013. Developmental changes in the cochlear hair cell mechanotransducer channel and their regulation by transmembrane channel-like proteins. *The Journal of general physiology*, **141**:141-8.
- Kim, S. J. & Skach, W. R. 2012. Mechanisms of CFTR Folding at the Endoplasmic Reticulum. *Frontiers in pharmacology*, **3**:201.
- Labay, V., Weichert, R. M., Makishima, T. & Griffith, A. J. 2010. Topology of transmembrane channel-like gene 1 protein. *Biochemistry*, **49**:8592-8.
- Lee, A., Vastermark, A. & Saier, M. H., Jr. 2014. Establishing homology between mitochondrial calcium uniporters, prokaryotic magnesium channels and chlamydial IncA proteins. *Microbiology*, **160**:1679-89.
- Marchler-Bauer, A. & Bryant, S. H. 2004. CD-Search: protein domain annotations on the fly. *Nucleic acids research*, **32**:W327-31.
- Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D. & Bryant, S. H. 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic acids research*, **41**:D348-52.
- Pang, C., Yuan, H., Ren, S., Chen, Y., An, H. & Zhan, Y. 2014. TMEM16A/B associated CaCC: structural and functional insights. *Protein and peptide letters*, **21**:94-9.
- Reddy, V. S. & Saier, M. H., Jr. 2012. BioV Suite--a collection of programs for the study of transport protein evolution. *The FEBS journal*, **279**:2036-46.

- Shimomura, T., Irie, K., Nagura, H., Imai, T. & Fujiyoshi, Y. 2011. Arrangement and mobility of the voltage sensor domain in prokaryotic voltage-gated sodium channels. *The Journal of biological chemistry*, **286**:7409-17.
- Terashima, H., Picollo, A. & Accardi, A. 2013. Purified TMEM16A is sufficient to form Ca<sup>2+</sup>-activated Cl<sup>-</sup> channels. *Proceedings of the National Academy of Sciences of the United States of America*, **110**:19354-9.
- Tian, Y., Kongsuphol, P., Hug, M., Ousingsawat, J., Witzgall, R., Schreiber, R. & Kunzelmann, K. 2011. Calmodulin-dependent activation of the epithelial calcium-dependent chloride channel TMEM16A. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **25**:1058-68.
- Wadskog, I., Forsmark, A., Rossi, G., Konopka, C., Oyen, M., Goksor, M., Ronne, H., Brennwald, P. & Adler, L. 2006. The yeast tumor suppressor homologue Sro7p is required for targeting of the sodium pumping ATPase to the cell surface. *Molecular biology of the cell*, **17**:4988-5003.
- Wells, K. M. & Rao, R. 2001. The yeast Na<sup>+</sup>/H<sup>+</sup> exchanger Nhx1 is an N-linked glycoprotein. Topological implications. *The Journal of biological chemistry*, **276**:3401-7.
- Yee, D. C., Shlykov, M. A., Vastermark, A., Reddy, V. S., Arora, S., Sun, E. I. & Saier, M. H., Jr. 2013. The transporter-opsin-G protein-coupled receptor (TOG) superfamily. *The FEBS journal*, **280**:5780-800.
- Yen, M. R., Choi, J. & Saier, M. H., Jr. 2009. Bioinformatic analyses of transmembrane transport: novel software for deducing protein phylogeny, topology, and evolution. *Journal of molecular microbiology and biotechnology*, **17**:163-76.
- Yen, M. R., Chen, J. S., Marquez, J. L., Sun, E. I. & Saier, M. H. 2010. Multidrug resistance: phylogenetic characterization of superfamilies of secondary carriers that include drug exporters. *Methods in molecular biology*, **637**:47-64.
- Zhai, Y. & Saier, M. H., Jr. 2001. A web-based program for the prediction of average hydrophathy, average amphipathicity and average similarity of multiply aligned homologous proteins. *Journal of molecular microbiology and biotechnology*, **3**:285-6.

Zhai, Y. & Saier, M. H., Jr. 2002. A simple sensitive program for detecting internal repeats in sets of multiply aligned homologous proteins. *Journal of molecular microbiology and biotechnology*, **4**:375-7.