

Signature Page

Thesis approved by



Dr. Richard Hallworth,
Major Advisor



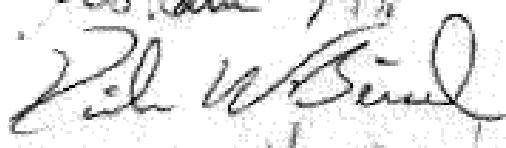
Dr. Gail Jensen
Dean of the Graduate School

Committee Members:

Dr. Hesham Ali



Dr. Kirk Beisel



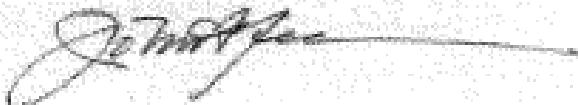
Dr. David (Zhi Zhou) He



Dr. Anthony Ricci



Dr. John Yee



Structures Involved in the Oligomerization of Prestin

By

Benjamin Currall

Submitted to the Faculty of the Graduate School of Creighton University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biomedical Sciences
Omaha, NE

May 2010

ACKNOWLEDGEMENTS

One evening, after having difficulties with some of my cloning work, I was walking through the Creighton parking lot. A professor, who I barely knew, stopped me to ask how my studies were going. We spent the next hour going over some of the problems I was having and then he invited me to come by his lab the next day so he could show me some of the tricks he used. He spent the next couple of months helping troubleshooting my experiments, despite having no obligation to do so. I would like to thank the many people who kindly assisted me throughout my doctoral especially those from outside the Biomedical Sciences Department: Dr. Venkatesh Govindarajan (GV), Cancer Center (the parking lot professor); Dr. Michael Nichols, Physics Department; Dr. John Mordeson, Math Department; and Dr. Zhao Yi (Charlie) Wang, Medical Microbiology and Immunology.

A special thanks is also due to the University of Nebraska Omaha's Bioinformatics Program who supplied both Dr. Hesham Ali and his student Kate Dempsey. Without their efforts, much of this work would not have been possible.

The professors and staff of the Biomedical Sciences Department have been especially kind and helpful during my time at Creighton University. Special thanks to my doctoral committee members who have gone through the entire process with me: Dr. Richard Hallworth (my mentor), Dr. Hesham Ali, Dr. Kirk Beisel, Dr. David He, Dr. Anthony Ricci, and Dr. Jack Yee.

And, of course, friends and family who have helped, and continue, to keep me grounded.

ABSTRACT

Hearing is an essential part of daily life for most people, yet little is known its molecular constituents. Cochlear amplification is the mechanism by which the hearing process is tuned and boosted in the inner ear. Somatic motility, a unique property of the outer hair cells in the inner ear, is a major component of cochlear amplification. Outer hair cell somatic motility is driven by the motor protein prestin, but little is known about the structure-function relationship of the motor protein prestin. This has lead to disputes over its role in cochlear amplification. **This work seeks to clarify the structure-function relationship of prestin by testing the hypothesis that the prestin protein family's function is dependent on homo-oligomerization through the STAS domain.** Förster resonance energy transfer demonstrated that homo-oligomerization occurs in several prestin homologous sequences. Subsequent sequence analysis of prestin homologous sequences revealed a model of the STAS domain, a putative protein-protein motif in the STAS domain, and two putative pore regions in the transmembrane region. Scanning cysteine mutagenesis suggested that one cysteine (C415) affects both structure and function and may have a role in disulfide bond formation. Mutation of the protein-protein motif in the STAS domain also significantly altered both structure and function, but it is unclear the role this motif plays in homo-oligomerization. These results, along with recently published structural data, were used to generate a refined model of prestin. This model postulates that the STAS domain acts as an ‘ATP-gate’ regulating prestin function. If correct, this model may help further our understanding of the structure-function relationship of prestin and its role in human hearing.

Contents

ACKNOWLEDGEMENTS	ii
ABSTRACT	iii
CHAPTER I: Introduction	1
CHAPTER II: Background	3
I. Hearing Mechanics	3
II. Cochlear Amplification	5
III. Prestin	6
A. <i>Prestin Evolution</i>	7
B. <i>Prestin Structure</i>	8
C. <i>Prestin Function</i>	11
IV. Prestin Oligomerization	14
A. <i>Homo-oligomerization and stoichiometry</i>	14
B. <i>Homologous sequence oligomerization</i>	14
C. <i>Functional implications of oligomerization</i>	15
V. Study Design	16
A. <i>Approach</i>	16
B. <i>Techniques</i>	17
CHAPTER III: Analysis of Oligomerization Conservation.....	24
I. Materials and Methods	24
A. <i>Phylogeny</i>	24
B. <i>Plasmid Constructs</i>	25
C. <i>Cell Transfection</i>	27
D. <i>FRET</i>	28
II. Results	30
A. <i>Identification of Homologous Sequences</i>	30
B. <i>Protein Synthesis and Targeting</i>	33
C. <i>FRET Efficiency of Controls</i>	37
D. <i>FRET Efficiency of Homo-Oligomerization and Hetero-Oligomerization</i>	39
III. Conclusion	41
CHAPTER IV: Identification of Oligomerization Motifs	42

I. Materials and Methods.....	42
A. <i>Sequence Curation</i>	42
B. <i>Sequence Analysis</i>	44
II. Results	44
A. <i>Higher Order Structure.....</i>	44
B. <i>Motifs</i>	68
C. <i>Primary Structure.....</i>	73
III. Conclusion.....	74
CHAPTER V: Analysis of Oligomerization Motif Disruption	76
I. Materials and Methods.....	76
A. <i>Multiple Sequence Alignment (MSA).....</i>	76
B. <i>Plasmid Constructs.....</i>	77
C. <i>Cell Transfection</i>	77
D. <i>Non-Linear Capacitance (NLC)</i>	77
E. <i>Fluorescent Resonance Energy Transfer (FRET)</i>	79
II. Results	79
A. <i>Cysteine Mutations</i>	79
B. <i>DSSG Motif.....</i>	86
III. Conclusion.....	91
CHAPTER VI: Discussion	93
I. Prestin Homo-oligomerization	93
A. <i>Analysis of Oligomerization Conservation in Prestin Homologous Sequences</i>	93
B. <i>Identification of Oligomerization Motifs.....</i>	97
C. <i>Analysis of Oligomerization Motif Disruption</i>	99
D. <i>Summary.....</i>	105
II. Transmembrane Region	105
III. STAS Domain	109
IV. Prestin Model.....	111
V. Future Work	113
A. <i>Experiments</i>	113

<i>B. Method Improvement</i>	115
APPENDIX I: Plasmid Constructs.....	118
I. Background.....	118
<i>A. Environment</i>	118
<i>B. Plasmid Structure</i>	120
<i>C. Molecular Cloning</i>	121
II. Study Plasmid Construct Design	122
<i>A. Design of Fluorescent Protein Plasmid Constructs</i>	123
<i>B. Design of Homologous Sequence-Fluorescent Protein Plasmid Constructs ..</i>	124
III. Primers Tables.....	125
<i>A. Cloning Primers</i>	125
<i>B. Mutation Primers</i>	126
<i>C. Sequencing Primers</i>	127
APPENDIX II: Förster Resonance Energy Transfer	128
I. Background	128
<i>A. Luminescence</i>	128
<i>B. Quenching</i>	132
<i>C. Förster resonance energy transfer (FRET)</i>	133
<i>D. Measuring FRET</i>	135
<i>E. Fluorescent proteins</i>	137
II. System Calibration.....	138
APPENDIX III: Bioinformatic Theory and Methods	142
I. Theory.....	142
<i>A. Analysis of Physiochemical Properties.....</i>	143
<i>B. Sequence Comparison.....</i>	148
<i>C. Advances in Sequence Analysis Theory</i>	156
II. Software, Databases, and Algorithms	157
<i>A. Software.....</i>	158
<i>B. Databases.....</i>	159
<i>C. Algorithms</i>	162
III. Curated Databases, PAIM Algorithm, and PAIM Results	167

<i>A. Curated Databases</i>	167
<i>B. Post-Alignment Identity Matrix (PAIM)</i>	174
<i>C. Detailed PAIM Results</i>	179
DETAILED METHODS	180
I. Molecular Biology	180
<i>A. Solutions</i>	181
<i>B. Protocols</i>	184
II. Cell Culture	195
<i>A. Solutions</i>	195
<i>B. Protocols</i>	196
ABBREVIATIONS	200
LITERATURE CITED	203

CHAPTER I: Introduction

Hearing loss affects quality of life and incurs large socio-economic costs. Approximately 3 of every 1,000 newborns are born with permanent hearing loss requiring lifetime care. Treatment costs for these individuals are estimated to be near \$2.1 billion (in 2003 dollars) [1, 2]. The prevalence of hearing loss increases with age; as many as 93% of older Americans have some indication of hearing loss [3-5]. Due to factors such as increased environmental hazards and an aging population, the incidence of hearing loss is increasing in the United States [6-8]. Despite the socio-economic impact of hearing loss, our basic understanding of auditory transduction remains vague.

In 2000, a protein, termed prestin, was discovered [9]. This protein was found to be essential for mammalian hearing [10, 11]. Prestin operates as a motor which oscillates at frequencies orders of magnitude greater than any previously identified protein [12]. The novel motor function of prestin is thought to be derived from the transport function seen in its related proteins in the solute carrier 26 a (Slc26a) family [13, 14]. However, little is known about prestin's structure-function relationship.

In this work, I examine one aspect of prestin structure: homo-oligomerization. Several studies have shown that prestin forms homo-oligomers, possibly as a tetramer [15-20]. It has also been shown that one prestin domain, the sulfate transporter anti-sigma factor antagonist (STAS), is able to form homo-oligomers when isolated from the rest of the protein [21]. There is much debate, however, about how prestin homo-oligomerizes and its role in motor function [20, 22]. I hypothesize that Slc26a protein function is dependent on homo-oligomerization of the STAS domain. This hypothesis depends on three premises: 1) all Slc26as form homo-oligomers, 2) homo-oligomerization is

mediated through the STAS domain, and 3) prestin motor function is dependent on oligomerization.

In the following study, three separate approaches were used to examine prestin homo-oligomerization. First, conservation of oligomerization was examined in and between several Slc26a family members. Second, the structure of Slc26a family members was examined to identify possible oligomerization sites. Third, putative homo-oligomerization sites were disrupted to determine if disruption of oligomerization correlated with disruption of function. These approaches helped clarify the structure-function relationship of prestin and lead to a better understanding of the role prestin plays in both function and dysfunction of human hearing.

CHAPTER II: Background

The human ear is capable of feats that make complex communication possible. Trained musicians can discriminate differences in frequency of less than 0.2%. Teenagers can hear frequencies from 20 Hz to 20,000 kHz. Many young adults can hear a pin drop, a sound which is about 1,000,000 times less intense than a jack hammer [23]. The source of these extraordinary abilities is just starting to be resolved as the science of the ear is advancing. Although many mysteries of ear physiology and pathology still remain, a large part of frequency and intensity processing can be explained by the intricate mechanosensory system of the mammalian inner ear.

I. Hearing Mechanics

The ear is composed of the outer, middle and inner ear (Fig. 1A). The outer ear includes the pinna and ear canal, which collects sound. The middle ear, composed of the tympanic membrane and three ossicles (malleus, incus, and stapes), transmits sound to the liquid environment of the cochlear inner ear. The resulting pressure wave causes a traveling wave across the intrinsically tuned basilar membrane (Fig 1B). Resting on the basilar membrane is the sensory apparatus, the organ of Corti, which contains both the inner and outer hair cells (IHCs and OHCs, respectively, Fig. 1C).

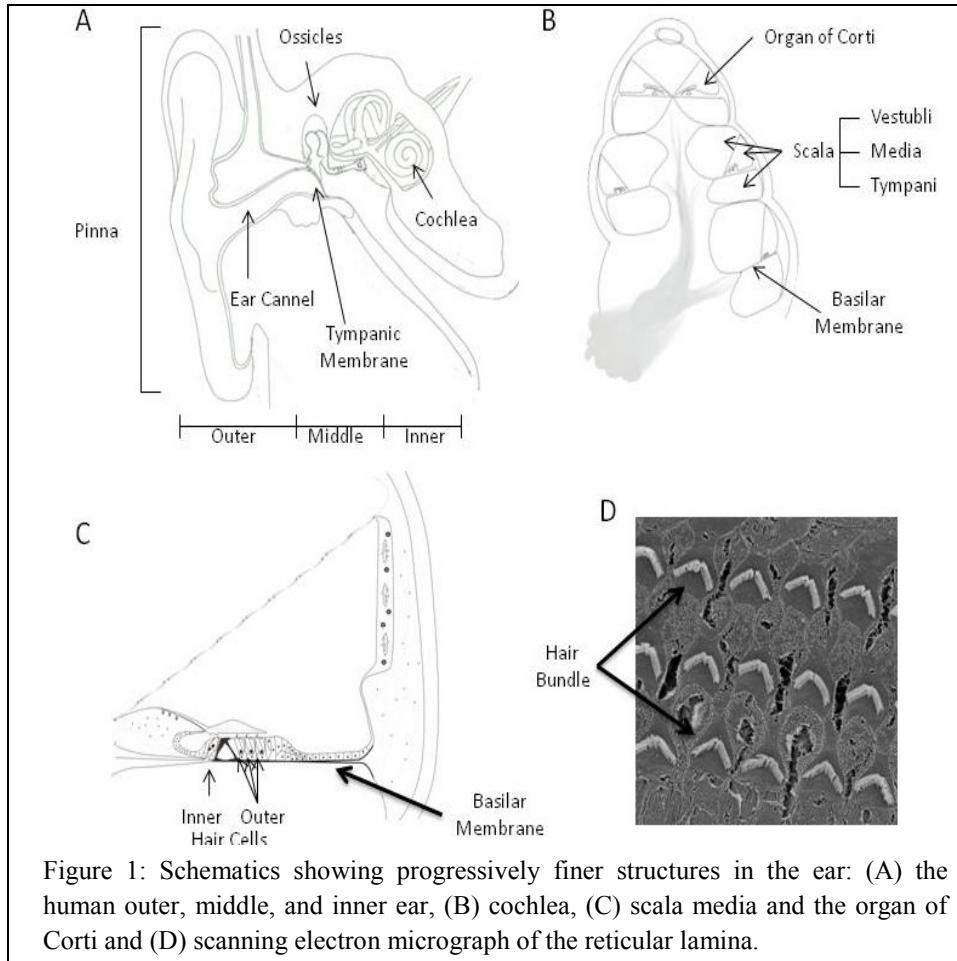


Figure 1: Schematics showing progressively finer structures in the ear: (A) the human outer, middle, and inner ear, (B) cochlea, (C) scala media and the organ of Corti and (D) scanning electron micrograph of the reticular lamina.

Cochlear hair cells convert the basilar membrane displacement into the electrical stimuli which are then conveyed to the auditory cortex. Basilar membrane movement causes a deflection of the hair bundle, composed of actin packed stereocilia, at the apex of each hair cell (Fig. 1D). Hair bundle kinofugal deflection causes opening of stereocilia mechanoelectrical transduction (MET) channels. Opening of the MET channels results in the influx of cations, which triggers depolarization of the hair cell. IHCs receive ~95% of the cochlear afferent innervation suggesting they are the major source of relaying information to the central nervous system [23]. OHCs receive mainly efferent innervation, suggesting that OHCs are not involved in direct signal transduction. Unlike IHCs, OHCs are capable of undergoing shape change, both elongation and contraction

(i.e. somatic motility), when electrically stimulated. Many hearing loss pathologies are linked OHC dysfunction [24].

II. Cochlear Amplification

In mammals, low intensity acoustic stimuli cause small movements of the basilar membrane, which can be augmented by a nonlinear active process known as cochlear amplification in mammals. Competing theories propose two different sources of cochlear amplification: 1) the somatic motility of mammalian outer hair cells and 2) force generation by the hair cell's hair bundle. Somatic motility is unique to the mammalian OHCs which elongate or shorten when hyperpolarized or depolarized, respectively. Moreover, somatic motility is possible at rates greater than 70 kHz, matching the upper frequency ranges of mammalian hearing [12]. Given these properties, OHC somatic motility is hypothesized to be the primary source of cochlear amplification [25]. Somatic motility has difficulty explaining certain phenomena associated with cochlear amplification. Theoretically, somatic motility is rate limited to approximately 1 kHz *in vivo*, well below the 160 kHz hearing upper limit that occurs in some echolocating mammals, by the electrical properties of the cell membrane [13]. Further, cochlear amplification also occurs in non-mammals despite the absence of somatic motility in their hair cells [26-28].

An alternative hypothesis suggests that the hair cell hair bundle acts as the primary source of cochlear amplification. According to this theory, re-closure of the MET can generate a force capable of causing cochlear amplification [29, 30]. This is supported by evidence that force can be supplied by the hair bundle in both non-mammalian and

mammalian species [31-33]. The debate between somatic motility and hair bundle driven cochlear amplification theories can be resolved by discerning the associated proteins.

III. Prestin

In 2000, the protein responsible for somatic motility was identified and named prestin. Prestin knockout mice showed a significant elevation of hearing threshold limits furthering the somatic motility theory of cochlear amplification [34]. OHCs from knock out (KO) mice, however, were significantly shorter than wild-type OHCs, which may have compromised the role of hair bundles in cochlear amplification. This problem was addressed by prestin knock in (KI) mice, in which the wild-type prestin was replaced by a nonfunctional prestin [11]. The OHCs of the KI were similar to wild-type except for a complete loss of somatic motility in physiological ranges. The thresholds of compound action potentials from the eighth nerve, however, were elevated by 40-60 dB in these KI mice. This suggests that prestin is, at least in part, an essential component of mammalian cochlear amplification.

However, the role prestin plays in cochlear amplification is still in question. A straight forward model would suggest that prestin driven somatic motility would occur on a cycle-by-cycle basis (i.e. OHC elongation during the falling phase and shortening during the rising phase of the BM movement would augment the intrinsic BM movement). While it has been shown that OHC somatic motility can operate at > 70 kHz *in vitro*, this may not be possible *in vivo* [12]. Somatic motility appears to be a membrane potential driven process and the rate of OHC hyper- and depolarization is limited by the cell's membrane resistance and capacitance [35-37]. This resistor-capacitor (RC) filtering

limits the voltage rate change to ~ 1 kHz in the OHC, which is well below the 160 kHz frequencies seen in echolocating mammals [37]. This is thought to result in a cycle-by-cycle frequency matching at low frequencies (referred to as the AC component) and sustained depolarization at high frequencies (referred to as the DC component) of the OHC *in vivo* (Fig. 2) [38]. The *in vivo* RC filter, and two component somatic motility, complicates the theory of prestin driven somatic motility as the source of cochlear amplification. Several competing models have arisen to take into account the features of cochlear amplification, somatic motility, and prestin function (see below) [13, 14, 30].

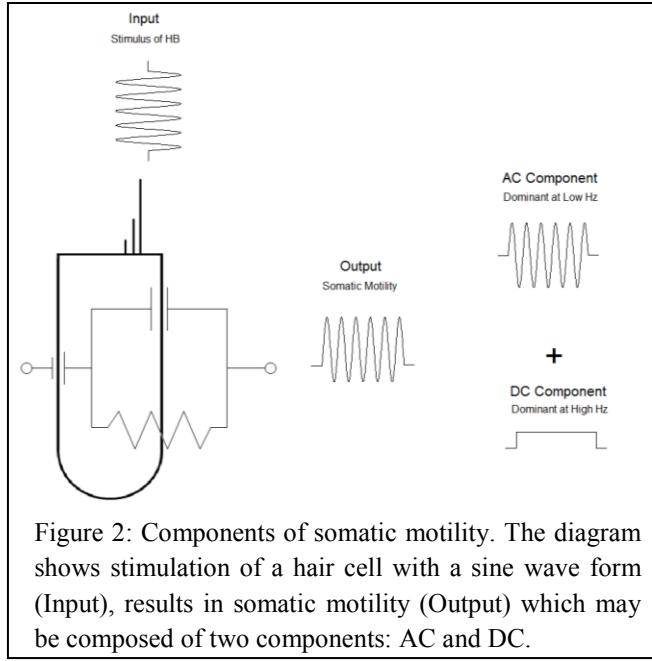


Figure 2: Components of somatic motility. The diagram shows stimulation of a hair cell with a sine wave form (Input), results in somatic motility (Output) which may be composed of two components: AC and DC.

A. Prestin Evolution

Prestin (Slc26a5) is part of the solute carrier 26 family (Slc26a family; HUGO Gene Nomenclature Committee), which in turn can be considered part of the larger SulfP family found in both eukaryotes and prokaryotes (2.A.53 as classified by Transport Classification Database, TCDB) [39]. There are 11 identified paralogs of the Slc26a

family (a1-a11), although Slc26a10 is thought to be a pseudogene. Of particular note, mammalian Slc26a5 is highly conserved when compared to other mammalian Slc26a paralogs [40, 41]. Mammalian prestin is the only Slc26a family member to act as a motor protein [28]. The other Slc26as can be classified into three groups: 1) SO₄²⁻ transporters (Slc26a1 and Slc26a2); 2) Cl⁻/HCO₃⁻ exchangers (Slc26a3, Slc26a4, non-mammalian Slc26a5, and Slc26a6); and 3) ion channels (Slc26a7 and Slc26a9). The functional roles of Slc26a8 and Slc26a11 are unknown. Interestingly, Slc26a11 is more closely related to plant (*A. thaliana*) and fruit fly (*D. melanogaster*) SulP homologous sequences than its Slc26a paralogs [42, 43]. One fruit fly paralog (CG5485), however, is found in the Johnston's organ (the hearing organ of the fruit fly) and is more closely related to prestin than other fruit fly paralogs. It is possible that the hearing related Slc26as have a common evolution-related origin separate from the other Slc26a paralogs [44].

B. Prestin Structure

The SulP protein family is defined by two distinct domains: the sulfate transporter (SulP, PF00916) domain and the sulfate transporter and anti-sigma factor antagonist (STAS, pf01740) domain (as classified by the National Center for Biotechnology Information's Conserved Domain Database and Pfam; Fig. 3A) [45]. The SulP domain is a member of the Amino Acid-Polyamine-Organocation clan (APC, CL00062; Note: the **SulP protein family** includes all proteins that have a SulP domain and a STAS domain in order from amino to carboxy terminus, such as prestin; while the **SulP domain** is a segment of the SulP protein family with a distinct and separate evolutionary path from the STAS domain). Clan members have between 10 and 14 transmembrane helices and

transport ions and/or amino acids. In the APC clan, only the SulP domain is associated with the STAS domain.

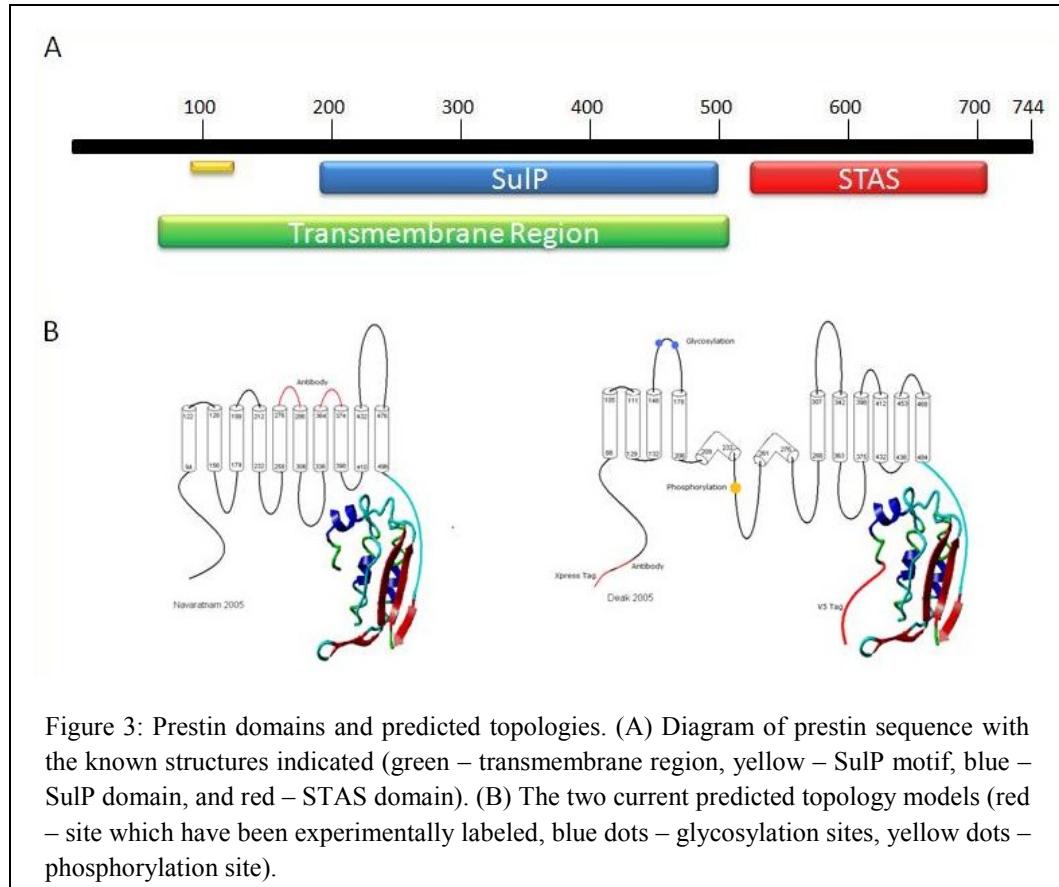


Figure 3: Prestin domains and predicted topologies. (A) Diagram of prestin sequence with the known structures indicated (green – transmembrane region, yellow – SulP motif, blue – SulP domain, and red – STAS domain). (B) The two current predicted topology models (red – site which have been experimentally labeled, blue dots – glycosylation sites, yellow dots – phosphorylation site).

The STAS domain belongs to a different clan, the ATP-Grasp clan (CL0179). The ATP-grasp clan members are ATP-dependent carboxylate-amine/thiol ligases [46]. The STAS domain, however, appears to have different function than the rest of the ATP-Grasp clan. Members of the STAS domain family, the bacterial SpoIIaa proteins, regulate sporulation through protein-protein interactions [47]. It has yet to be determined what role a sporulation factor protein domain plays in transporter proteins in mammalian species. Attempts have been made to separate the SulP domain from the STAS domain

through C-terminal truncation, but these have all resulted in a loss of membrane targeting [48-50].

In addition to the domains, a well conserved sulfate transporter motif (the SulP motif) has also been identified in the Slc26a family. The SulP motif is thought to be essential for normal function [9, 51]. The location and function of these domains and motifs in prestin are disputed [41, 49, 52].

Several groups have developed topology maps of Slc26a family members based on hydrophobicity modeling. One laboratory, led by Dr. Santos-Sacchi, has proposed a 10 transmembrane domain prestin topology (Fig. 3B, left). This model is supported by evidence of two extracellular epitopes (residues 274-290 and 359-375) [16]. Another laboratory, led by Dr. Dallos, has proposed a 12 transmembrane prestin model with two re-entrant loops (Fig. 3B, right). The Dallos model is supported by evidence of two extracellular glycosylation sites (N163, N166) and data obtained from introduced hemagglutinin tags (intracellular loop 4 and 5) [53-55]. Both the Santos-Sacchi and Dallos models agree that the amino and carboxy termini are intracellular, the SulP domain forms the majority of the transmembrane domains, and the STAS domain is intracellular. Both models also agree on an intracellular loop containing a cyclic nucleotide phosphorylation site (S238) [16, 55]. Otherwise, the two models have contradicting topologies with nearly every extracellular and intracellular loop inverted between the models (Fig. 3B). While alternate prestin topology models have also been proposed, the topology of prestin remains uncertain [51, 56]. Unfortunately, no NMR or X-ray crystallography data has been published for prestin, making it difficult to verify any model.

Of the mammalian Slc26a paralogs, mammalian prestin orthologs is the most conserved, suggesting that prestin is both important for survival in the natural environment and that normal prestin motor function is easily disrupted [40, 41]. Attempts have been made to identify residues involved in motor function, yet the vital structural components of prestin remain unknown [16, 49, 51-53, 57]. Some structural features of prestin have been described. Mammalian prestin contains 744 residues and has a molecular weight of ~86 kDa. This weight is higher than predicted based upon the number of residues and may be due to glycosylation at residues N163 and N166 [04 Matsuda]. The presence of glycosylation, however, has been disputed [16, 54]. A phosphorylation site at S238 has also been identified using phosphorylation motif software and point mutation analysis [55]. Additionally, clusters of positive and negative residues identified in the carboxy terminal have been proposed to play a role in allosterically-mediated conformational changes [52, 58]. The near-absence of high resolution prestin secondary, tertiary, and quaternary structure models makes evaluation of prestin's structure-function relationship difficult, despite data obtained from over one hundred mutations, truncations, and chimeras generated for prestin and its Slc26a homologous sequences to date [50, 53, 59-66].

C. Prestin Function

While prestin operates as a motor protein to drive somatic motility, the mechanism of this action is poorly understood. Unlike other known motor proteins, prestin does not require ATP or Ca^{2+} to generate force [67, 68]. The rate of prestin movement also excludes secondary cascades and enzymatic activity. It has been proposed that the motor

function of prestin is a modification of the transport function seen in other Slc26a proteins [13, 14].

Prestin-driven somatic motility is intimately linked to a nonlinear voltage-dependent capacitance, simply known as a nonlinear capacitance (NLC, see below) [69]. Both OHCs and prestin transfected human embryonic kidney (HEK) cells exhibit NLC. NLC is absent or severely reduced in both prestin paralogs and non-mammalian orthologs, neither of which acts as motor proteins [27, 28, 43, 56]. One study of prestin's non-conserved charged residues suggested NLC is associated with an extrinsic anion, specifically internal chloride [53]. This evidence led to the development of the partial anion transporter (PAT) function model [13, 14, 30]. The PAT model suggests that, on depolarization, Cl^- anions are partially transported through the transmembrane portion of prestin towards the extracellular surface and force the protein into a 'contracted' state. When hyperpolarized, Cl^- anions move in the opposite direction through the transmembrane portion of the prestin channel resulting in an 'extended' state (boxed region of Fig. 4). This model explains both somatic motility, OHCs lengthen when hyperpolarize and shorten when depolarized, and NLC, a charge moves in rather than through the membrane. The PAT model further illustrates that prestin motor function is a subset of transporter transition states (Fig. 4).

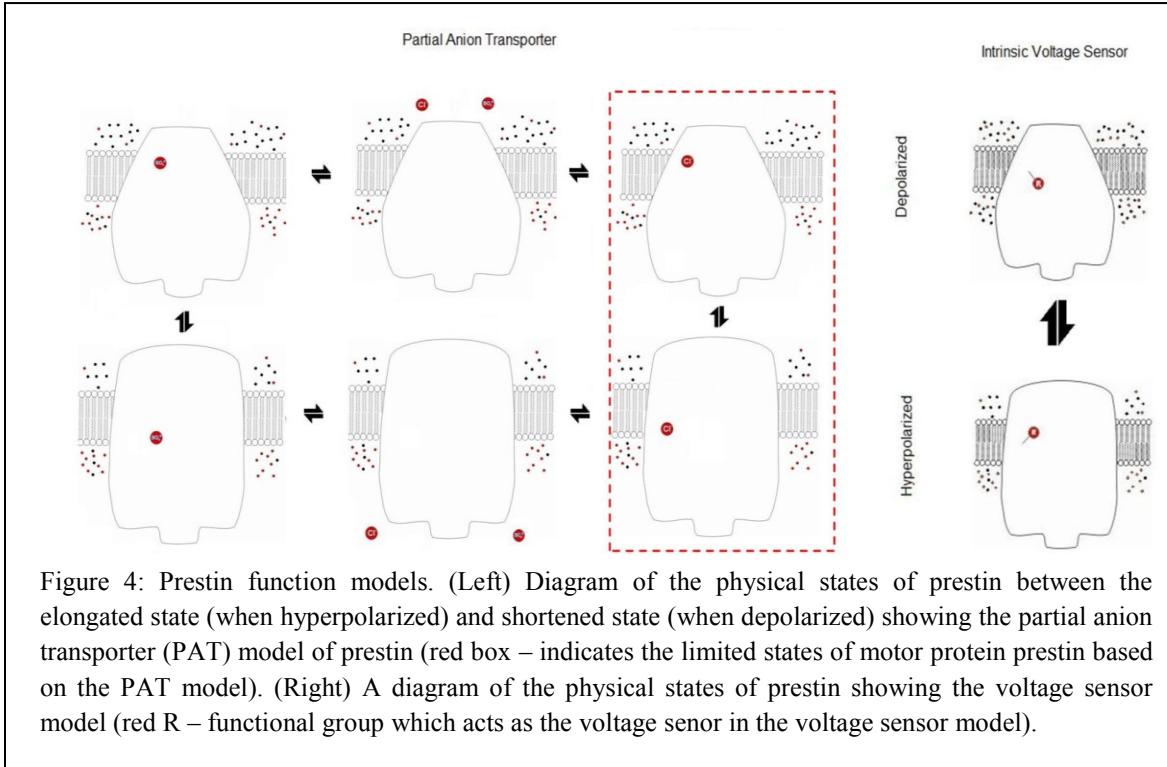


Figure 4: Prestin function models. (Left) Diagram of the physical states of prestin between the elongated state (when hyperpolarized) and shortened state (when depolarized) showing the partial anion transporter (PAT) model of prestin (red box – indicates the limited states of motor protein prestin based on the PAT model). (Right) A diagram of the physical states of prestin showing the voltage sensor model (red R – functional group which acts as the voltage senor in the voltage sensor model).

Recent evidence has challenged the PAT model. Mutating conserved charged residues alters NLC, indicating NLC may be carried by charged intrinsic residues [70]. NLC has also been shown in non-mammalian prestin where anion transport is complete, further suggesting NLC may be carried by a charged residue [71]. Another study reports that mammalian prestin transports anions and that the transport function can be separated from NLC [57]. These data suggest an alternative model, somewhat akin to the voltage sensor of the voltage-gated sodium channel, wherein an intrinsic voltage gate drives the motor function (Fig. 4). Unfortunately, without high resolution structural models of prestin, it is difficult to validate either of these competing models.

IV. Prestin Oligomerization

A. Homo-oligomerization and stoichiometry

Recently, several papers confirmed that prestin quaternary structure consists of homo-oligomers [16-20]. Both 3D electron microscopy and Western blots provided evidence that prestin forms a homo-tetramer, possibly stabilized by disulfide bonds [17, 19]. The tetramer state has been challenged by another study which provides evidence that prestin, as well as other Slc26a members, forms dimers [20]. Central to resolving the stoichiometry question is establishing the regions involved in oligomerization. The STAS domain has been shown to participate in protein-protein interactions which may be stabilized by disulfide bonds [21, 43, 48]. In addition, amino terminal truncation resulted in the loss of both NLC and oligomerization [16]. Despite these preliminary results, several questions remain unanswered. For example, it is not known if oligomerization occurs through a single region, as in voltage-gated potassium channels, or whether multiple regions are involved, as in hemoglobin tetramerization [72, 73]. It is also unclear if oligomerization is required for function [20, 22, 74].

B. Homologous sequence oligomerization

Little evidence exists on whether or not other Slc26a family members form homo-oligomers. One study uses Western blot evidence to suggest that rat prestin (Slc26a5), a zebrafish ortholog (*Danio rerio* Slc26a5), a human paralog (*Homo sapiens* SLC26A3), and a prokaryotic homologous sequence (*Pseudomonas aeruginosa* SulP) all formed dimers [20]. This study, however, is in contrast to several other studies, also including Western blot data, which demonstrate mammalian prestin forms tetramers [17]. Another

study suggested that while Slc26a4 prestin paralogs can form dimers, they more likely exist as monomers [75]. Confirming whether homo-oligomerization is conserved across Slc26a family members, or whether it is specific for prestin, will help identify both structures that are involved in prestin oligomerization as well as the functional implication of prestin oligomerization.

C. Functional implications of oligomerization

Little is known about whether oligomerization is necessary for normal prestin function. One study has suggested prestin oligomerization is dependent on cholesterol levels [76]. Furthermore, when cholesterol levels were depleted, normal NLC was disrupted. In contrast, co-transfection of different prestin mutants suggests that oligomerization may not be necessary for function, although oligomerization may have allosteric effects [20]. Two recent studies also suggest that while prestin may form homo-oligomers, the subunits function independently [22, 74]. These studies, however, did not directly examine the function of prestin monomers. Thus, it remains unclear if prestin homo-oligomerization is crucial for function, as with voltage-gated potassium channels, or plays more of a modifying role, as with chloride channels [77]. The functional implications of homo-oligomerization could be better addressed by specifically disrupting homo-oligomerization without changing other aspects of the prestin structure-function relationship.

V. Study Design

I hypothesized that **Slc26a function is dependent on homo-oligomerization mediated through the STAS domain.** This hypothesis can be broken down into three parts. First, homo-oligomerization is conserved across all Slc26a family members. Despite the scarce evidence, previous studies do suggest that homo-oligomerization is a conserved property of the Slc26a family [20]. Second, homo-oligomerization is mediated through the STAS domain. A single report suggests that the STAS domain is able to form homo-oligomers when synthesized in the absence of the rest of the protein [21]. Further, the SpoIIaa, the STAS domain ancestor, functions as a sporulation inhibitor through protein-protein interaction, a function that may be maintained through its decedents [54]. Third, homo-oligomerization is necessary for function. The current literature is in disagreement over the functional importance of homo-oligomerization, but it is interesting to note that many channels and transporters are composed of several similar subunits or repetitive domains (e.g. chloride channels, potassium channels, sodium channels, gap junctions) [20, 22, 74]. These hypotheses were addressed by three different approaches that aim to determine conservation of oligomerization, to identify structures involved in oligomerization, and to measure functional implication of oligomerization.

A. Approach

Determine Conservation of Homo-Oligomerization

The first approach attempted to determine if homo-oligomerization is conserved throughout the Slc26a family. Conservation indicates both conserved structures involved

in and functional importance of homo-oligomerization in the Slc26a family. Molecular biology and Förster's resonance energy transfer (FRET) techniques were used to determine if interaction occurred between prestin homologous sequences.

Identify Structures Involved in Oligomerization

Conservation of oligomerization would predict that structures involved in oligomerization would also be conserved. Bioinformatic techniques were employed to determine related structures and *de novo* structures based in prestin based on homology (see Appendix III). Conserved structures were cross-referenced with current literature to determine if they have a putative role in prestin homo-oligomerization.

Measure Functional Implication of Homo- Oligomerization

The third approach determined the functional importance of homo-oligomerization by disrupting putative structures associated with oligomerization. Prestin structures were disrupted using molecular biology techniques, while oligomerization and function were measured using FRET and NLC, respectively.

B. Techniques

Förster Resonance Energy Transfer (FRET)

Förster resonance energy transfer (FRET) is a phenomenon that occurs when non-radiative resonance energy is transferred, through dipole-dipole interactions, between molecules. The efficiency of that transfer is related to the quantum yield of the donor (*i.e.* the excited molecule), molar absorptivity of the acceptor (*i.e.* the excitable molecule), the

spectral overlap between donor and acceptor, and the distance between the donor and acceptor (Fig. 5). FRET efficiency is dominated by the distance between the donor and acceptor, falling off by the sixth power as the distance increases. For the majority of donor /accepter pairs FRET is no longer detectable when there is a distance greater than 10 nm between the two. FRET is an effective microscopy technique in that it is able to resolve distances less than 10 nm while conventional light microscopy is practically limited to approximately 400 nm due to diffraction limits.

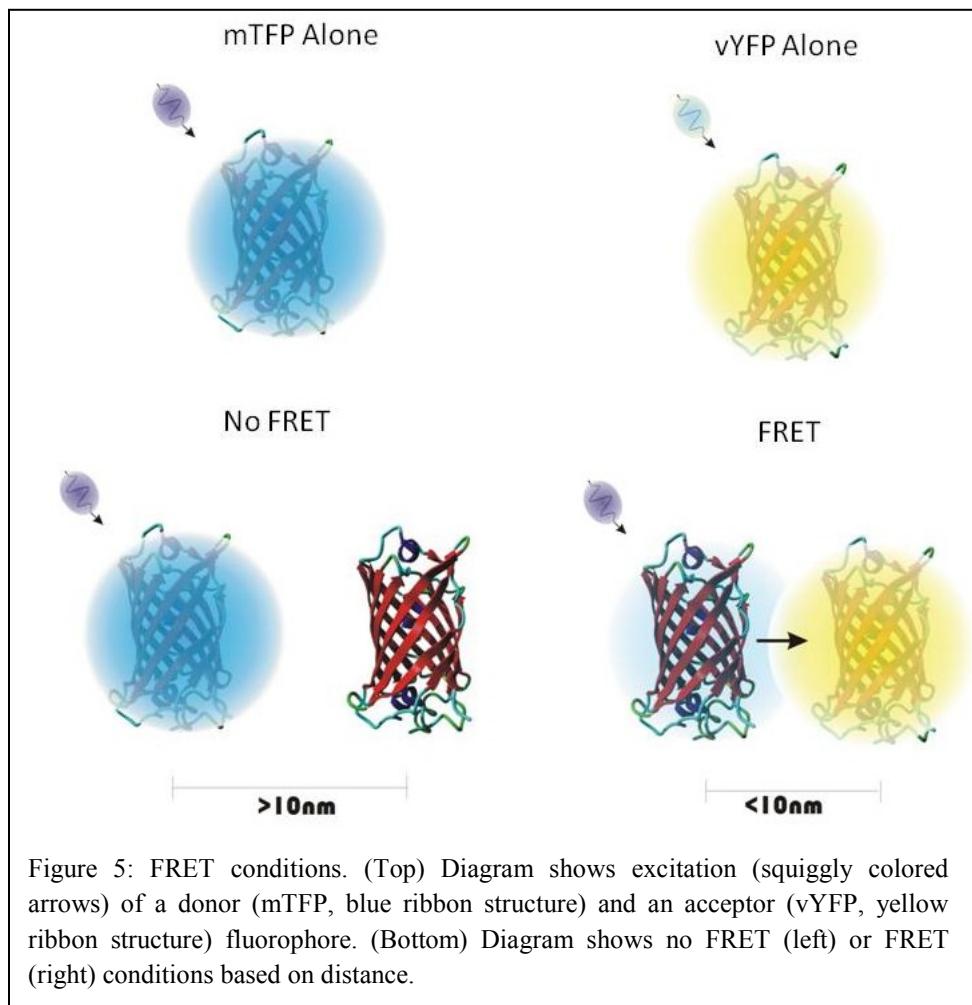


Figure 5: FRET conditions. (Top) Diagram shows excitation (squiggly colored arrows) of a donor (mTFP, blue ribbon structure) and an acceptor (vYFP, yellow ribbon structure) fluorophore. (Bottom) Diagram shows no FRET (left) or FRET (right) conditions based on distance.

FRET can be used in proteomic studies to examine protein-protein interaction or protein conformational changes. This is typically performed by adding donor and

acceptor fluorophores to proteins of interest. The FRET efficiency can then be determined by measuring changes in anisotropy, intensity, and/or lifetime changes in the fluorophores. Changes in FRET efficiency will indicate that protein-protein interactions and/or conformational changes are occurring (a more thorough description of FRET is given in Appendix II). Unlike other protein-protein interaction techniques, FRET can be measured without disrupting the cellular environment, presenting conditions more like those *in vivo*.

Sequence Analysis

Sequence analysis can be used to: 1) analyze sequence physiochemical properties, or 2) compare multiple sequences for conservation. Analysis of sequence physiochemical properties can be used to determine interaction sites (*e.g.* methylation, phosphorylation), binding (*e.g.* purine to pyridine, disulfide bonds), environment solubility (*e.g.* melting temperature, hydrophobicity), and secondary structure (*e.g.* stem-loops, α -helix) among other properties. This type of sequence analysis can also be used to suggest higher order structure-function relationships (*i.e.* *ab initio* protein structure modeling).

Sequence comparison generally uses alignment methods in which nucleotide or amino residues are aligned in columns and assayed for identity and similarity. If the sequences, or motifs/residues in the sequence, are well conserved (*i.e.* relatively high number of identical or similar residues in columns), then the sequences, or motifs/residues, are predicted to share similar functional properties. Sequence comparison can be used to: 1) compare a sequence of interest to databases of conserved sequences, or 2) compare a sequence of interest to homologous sequences. By comparing

the sequence of interest to a database, previously identified properties of a given sequence can be identified such as family membership, domain structure, motif presence, and possibly known structure-function. Comparing the sequence of interest to homologous sequences allows for relationships between proteins to be revealed, as well as identification of novel motifs and conserved residues, which may also have functional importance.

Non-Linear Capacitance (NLC)

Most cell membranes have a linear capacitance component due to the inherent capacitance of the plasma membrane (Fig. 6). The capacitance charge is stored on the plasma membrane, because the plasma membrane acts as an insulator between the electrical potential difference of the intra- and extracellular media. The plasma membrane capacitance can be described in the equation:

$$C = \frac{\epsilon A}{d},$$

where C is capacitance, ϵ is the permittivity, A is the area of the cell membrane, and d is the width of the plasma membrane. When the permittivity is constant, the capacitance is linear across membrane potentials:

$$C_{Lin} = \frac{Q}{V},$$

where C_{Lin} is the linear capacitance, Q is the ratio of charges on either side of the plasma membrane, and V is the membrane potential. In many cells, the permittivity of the cell changes at different voltages resulting in a non-linear capacitance (NLC). This can occur by redistribution or re-orientation of lipid molecules, membrane fusion of lipid vesicles,

ions moving in the membrane, and conformation changes of membrane proteins [78]. The NLC of cells can be calculated as a first order Boltzmann derivative:

$$C_{NL} = \frac{Q_{Tot}ze/kT}{\exp\left[\left(\frac{ze}{kT}\right)(V-V_{Pk})\right][1+\exp\left\{-\left(\frac{ze}{kT}\right)(V-V_{Pk})\right\}]^2},$$

where C_{NL} is the non-linear capacitance, Q_{Tot} is the total charge transferred, z is the valence of the charge transfer, e is the electron charge, k is the Boltzmann constant, T is the absolute temperature, V is the membrane potential, and V_{Pk} is the membrane voltage at half maximal charge [79]. The three variables in the equation (Q_{Tot} , z , V_{Pk}) can be correlated with physical properties of molecules. Q_{Tot} is measured as the area under the curve and calculates the total charge transferred when the molecules change confirmation state. V_{Pk} is measured as the voltage at which the maximum NLC is reached and calculates the membrane potential where the molecular are equally distributed between confirmation states. z is inversely related to the width and calculates the membrane potential range over which the change of conformational states occurs.

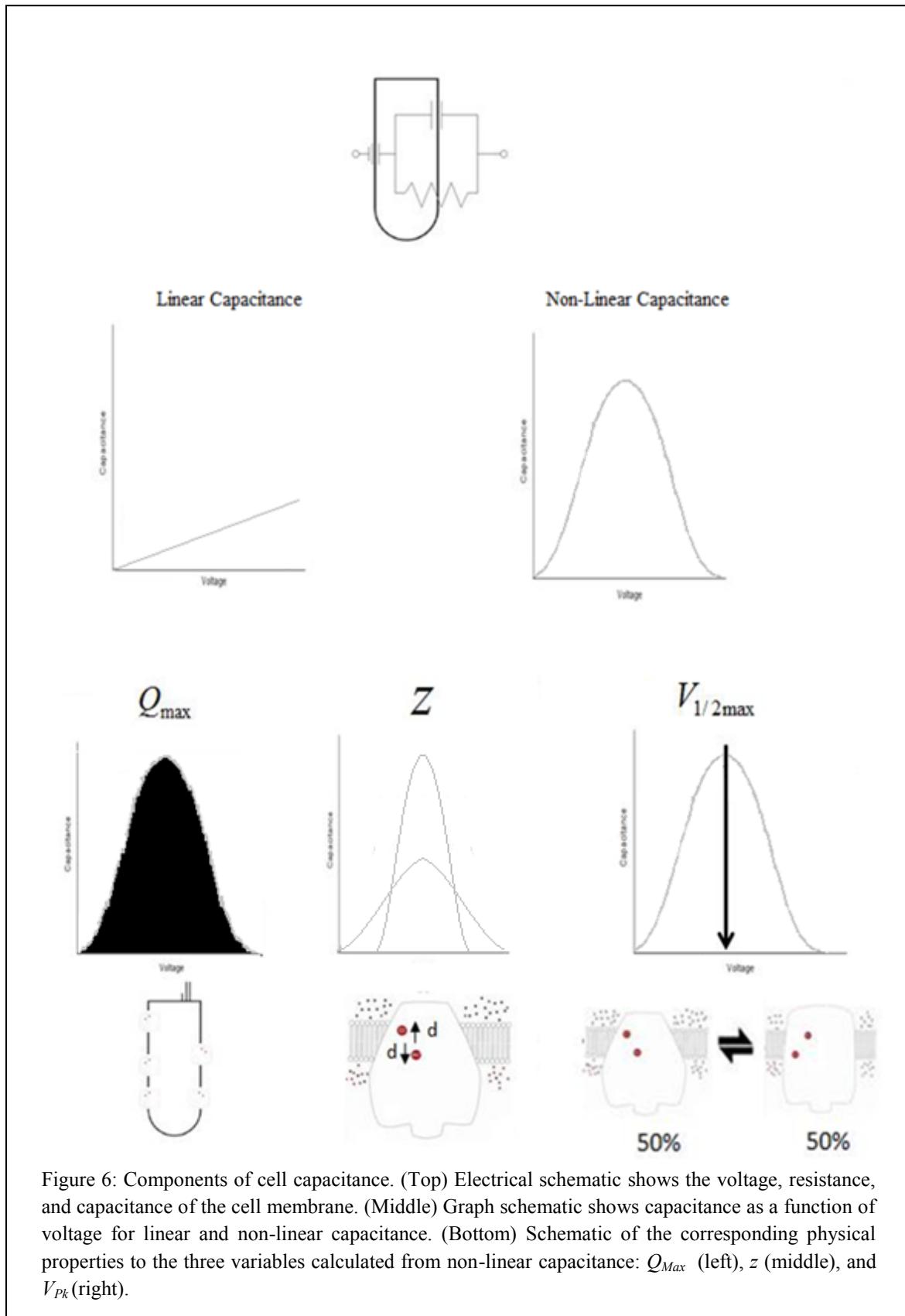


Figure 6: Components of cell capacitance. (Top) Electrical schematic shows the voltage, resistance, and capacitance of the cell membrane. (Middle) Graph schematic shows capacitance as a function of voltage for linear and non-linear capacitance. (Bottom) Schematic of the corresponding physical properties to the three variables calculated from non-linear capacitance: Q_{\max} (left), z (middle), and V_{pk} (right).

Voltage-gated channels are often associated with NLC because voltage driven confirmation changes cause movement of charged residues in the membrane. The OHC has a large NLC that is correlated with the appearance of densely packed putative prestin molecules in the OHC lateral wall. When prestin is transfected into cultured cells, these cells also display NLC [9]. Since transport currents have not been measurable with mammalian prestin, NLC has become the hallmark method for studying mammalian prestin function [69].

CHAPTER III: Analysis of Oligomerization Conservation

If oligomerization is conserved in all prestin homologous sequences, then oligomerization may be essential for function and, by extrapolation, structures responsible for oligomerization will also be conserved. Although several studies have shown that mammalian prestin homologous sequences form homo-oligomers, it is not known if more distantly-related prestin homologous sequences also form oligomers [16-18, 20]. In this chapter, the phylogenetic relationship of prestin homologous sequences was determined using bioinformatic techniques and the oligomerization of prestin homologous sequences was examined using FRET in cultured cells.

I. Materials and Methods

A. Phylogeny

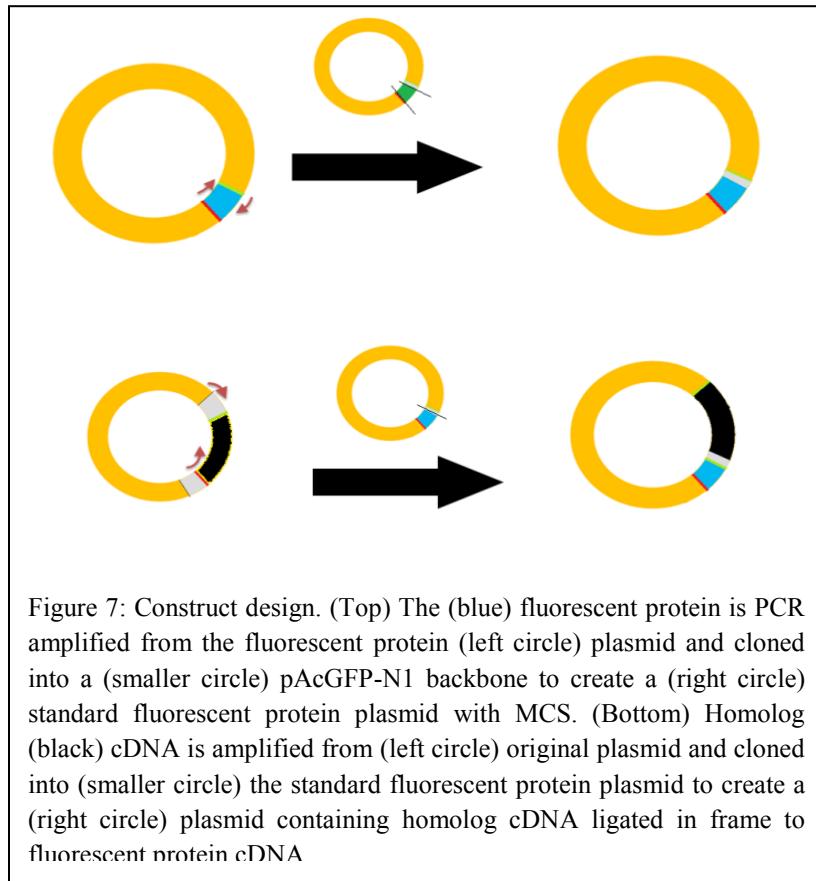
The *H. sapiens* prestin FASTA files sequence was obtained from the National Center for Biotechnology Information's (NCBI's) Entrez Gene from its online database (www.ncbi.nlm.nih.gov/gene). The *H. sapiens* prestin ortholog ‘isoform a’ (NP_945350.1) was used to retrieve prestin homologous sequences using the Basic Local Alignment Search Tool (BLAST) through the NCBI gateway (www.blast.ncbi.nlm.nih.gov). In the BLAST software, the blastp algorithm was used to search the reference sequence database across representative species including species from other kingdoms (see Appendix III for species). Sequences that did not contain both a plasma membrane SulP domain and a carboxy terminal STAS domain were removed from further analysis. FASTA files for the retrieved sequences were aligned using CLC Main Workbench’s (CLC Bio, Cambridge, MA) custom alignment algorithm with default

parameters [80]. If a sequence had a large gap or insertion (> 50 residues) compared to its nearest homologous sequence, it was rejected. Phylogeny analysis was also performed with CLC Main Workbench using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) with 100 bootstrap replicates.

B. Plasmid Constructs

FRET requires donor and acceptor fluorophores attached to putative protein-protein interacting partners. In this study, either monomeric teal fluorescent protein (mTFP) or Cerulean cyan fluorescent protein (cCFP) were used as the donor fluorophore, while Venus yellow fluorescent (vYFP) was used as the acceptor fluorophore. In all experiments, except for the link protein (see below), the mTFP protein was used, because of its superior FRET characteristics (see Appendix II) [81]. The donor pmTFP1-C plasmid construct, which contained the cDNA for mTFP was obtained from Allele Biotechnology (San Diego, CA). The acceptor pVenus plasmid construct, which contained the cDNA for the acceptor Venus yellow fluorescent protein (vYFP), was obtained from Dr. Atsushi Miyawaki (RIKEN Brain Science Institute, Saitama, Japan). Using PCR cloning, the mTFP and vYFP cDNA were cloned into the pAcGFP-N1 plasmid construct replacing the cDNA for *Aequorea coerulescens* green fluorescent protein to create pmTFP-N1, pmTFP-N2, pmTFP-N3, pvYFP-N1, pvYFP-N2, and pvYFP-N3 plasmid constructs (Fig. 7). A plasmid containing *Meriones unguiculatus* (gerbil) prestin (gP) cDNA, ligated in frame to enhanced green fluorescent protein (eGFP) cDNA, (this plasmid will be referred to as pgPG) was obtained from Dr. Peter

Dallos (Northwestern University, Chicago, Illinois). Plasmids containing chicken and zebrafish prestin (cP and zP, respectively) cDNA were obtained from Dr. Dominic Oliver



(University of Tübingen, Tübingen, Germany). Plasmids containing cDNA for the human *SLC26A2* (A2, MHS1010-9204232), *SLC26A3* (A3, MHS1010-7430037), and *SLC26A11* (A11, MHS1010-9204612) were obtained from Open Biosystems (Thermo Scientific, Huntsville, AL). The cDNA for the prestin homologous sequences were PCR cloned into both the pmTFP-N1 and pvYFP-N1 plasmid construct. Naming convention for the plasmid constructs used the following format:

pXF,

where p indicates the plasmid, X is the abbreviation of the cDNA, and F the abbreviation of the fluorescent protein (cCFP, C; mTFP, T; eGFP, G; and vYFP; Y). The pLink

positive control plasmid construct, containing cDNA for cCFP ligated in frame to cDNA for vYFP, and p38Y control, containing cDNA for SLC38A2 ligated in frame to cDNA for vYFP, were obtained from Dr. Jian Zhu (St. Jude's Research Hospital, Memphis, TN). Plasmid construct sequence was confirmed by sequencing of the inserts performed at the Creighton University Molecular Biology Core Laboratory. Proper protein synthesis was confirmed by examining both membrane targeting and fluorescent spectrum using the Zeiss LSM 510 META NLO confocal microscope located in the Creighton University Integrated Biomedical Imaging Facility (CUIBIF).

C. Cell Transfection

On reaching 80-100% confluence, plated human embryonic kidney (HEK) 293 cells were transfected with plasmid(s) using LipofectamineTM 2000 following the manufacturer's protocol (Invitrogen, Carlsbad, CA). In most cases, culture medium was replaced 6-24 hours after transfection to reduce LipofectamineTM toxicity. Approximately 48-72 hours after transfection, cells were fixed using 4% paraformaldehyde in phosphate buffered saline (PBS). After 30 minutes fixation, cells were washed with PBS and mounting medium was applied. Coverslips (#1 ½ 30 mm round, Warner Instruments, Hamden, CT) were applied and sealed using rubber cement (Elmer's Products Incorporated, Columbus, OH). Plates were stored at 4°C until imaged. If few co-transfected cells had a roughly equivalent donor and acceptor fluorophore intensity (see FRET Step 1 below), then co-transfection was performed and plasmid volumes were adjusted accordingly.

D. FRET

Transfected fixed cells were imaged using a 40x 1.4 n.a. oil immersion objective (Carl Zeiss Incorporated, Thornwood, NY), on the CUIBIF Zeiss LSM 510 META NLO confocal microscope. Cells containing donor alone and/or acceptor fluorophores were imaged with 4x digital magnification. Acceptor photobleaching lifetime imaging (apFLIM) consisted of five steps: 1) pre-bleach spectrum, 2) pre-bleach lifetime, 3) acceptor photobleaching, 4) post-bleach lifetime, and 5) post-bleach spectrum (Fig. 8). Pre-bleach spectral images were excited with 453 nm laser line and filtered using the META detector (series of 10.7 nm band pass filters) in lambda mode from 473.3 nm to 601.7 nm. Spectral images were linear unmixed into either donor or acceptor values based on previously acquired donor and acceptor spectral profiles. If a cell did not contain approximately equal intensities of donor and acceptor fluorescence, then it was rejected from further analysis. Pre-bleach lifetime images were imaged using two photon excitation with 820 nm (for cCFP) or 870 nm (for mTFP) laser wavelengths (Chameleon Ultra, Coherent Incorporated, Santa Clara, CA) and analyzed using time correlated single photon counting (TCSPC, Becker and Hickl, Berlin) system. Acceptor photobleaching was performed by 30 repetitions of excitation of a region of interest containing the cell with 545 nm laser line and band pass filtered from 565 nm to 615 nm. Post-bleach lifetimes and spectral images were analyzed after photobleaching.

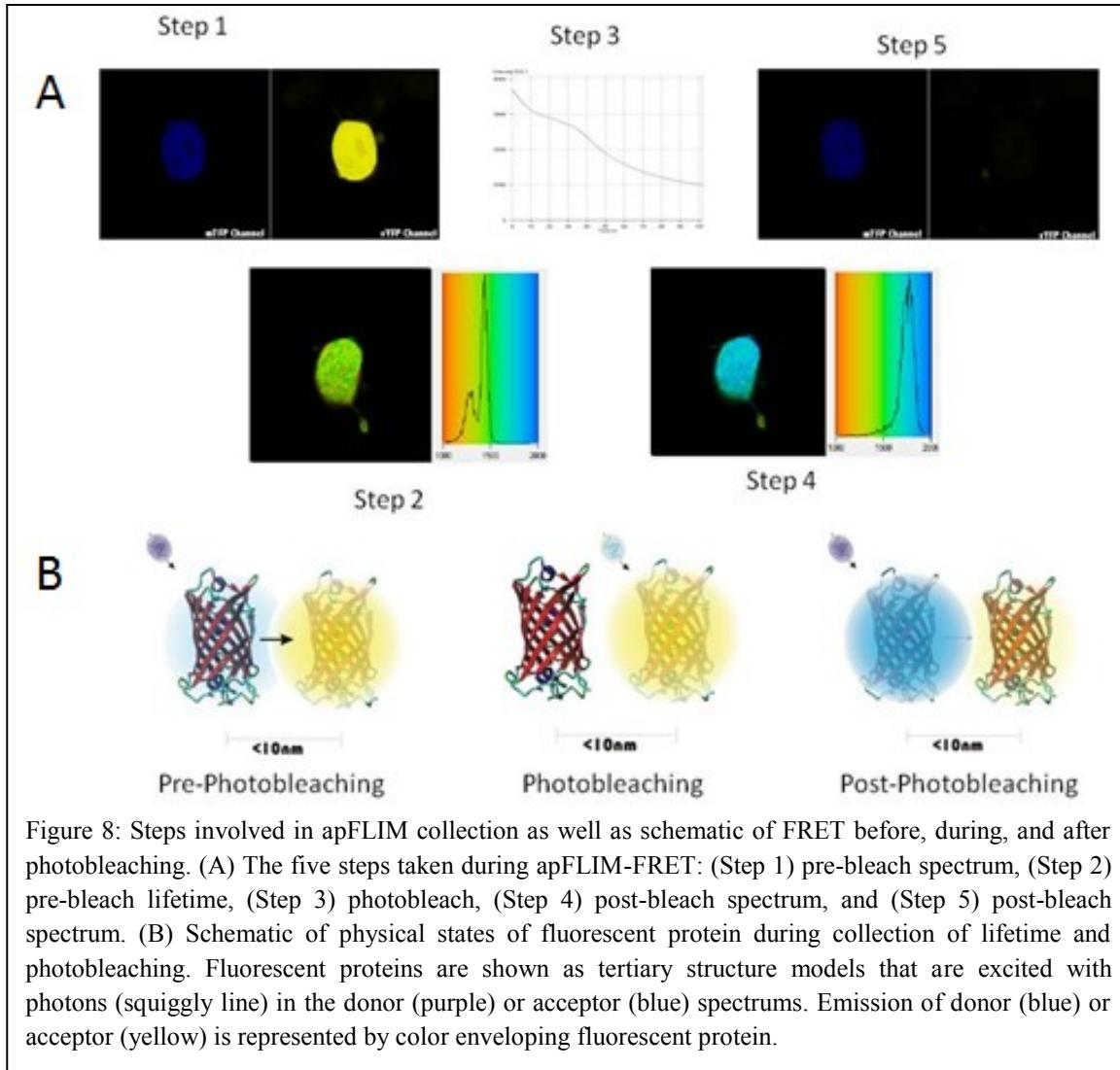


Figure 8: Steps involved in apFLIM collection as well as schematic of FRET before, during, and after photobleaching. (A) The five steps taken during apFLIM-FRET: (Step 1) pre-bleach spectrum, (Step 2) pre-bleach lifetime, (Step 3) photobleach, (Step 4) post-bleach spectrum, and (Step 5) post-bleach spectrum. (B) Schematic of physical states of fluorescent protein during collection of lifetime and photobleaching. Fluorescent proteins are shown as tertiary structure models that are excited with photons (squiggly line) in the donor (purple) or acceptor (blue) spectrums. Emission of donor (blue) or acceptor (yellow) is represented by color enveloping fluorescent protein.

Donor lifetimes were calculated using SPCImaging as either 1 (mTFP) or 2 (cCFP) lifetime components with a threshold of 25 photon counts per pixel. Color range was set with the minimum and maximum limits at 500 and 3000 ps, respectively. The histogram distribution of the decay matrix was exported into Origin and fitted to a Gaussian curve (OriginLab, Northampton, MA). Cells with an $R^2 < 0.900$ were not used for further analysis. FRET efficiency was calculated with the following equation:

$$E_{FRET} = \frac{\tau_D - \tau_{DA}}{\tau_D},$$

where E_{FRET} is FRET efficiency, τ_D is the lifetime of donor alone (after photobleaching), and τ_{DA} is the lifetime of the donor in the presence of an acceptor (before photobleaching).

II. Results

A. Identification of Homologous Sequences

Over 110 prestin homologous sequences were identified and aligned based on the filtered BLAST database searches. These included protein sequences from five major kingdoms (animals, plants, fungi, amoeba, and bacteria). The full alignment and functional classes are seen in Figure 9. As noted in Chapter II, mammalian Slc26a, animal SulP, and plant Sultr proteins appear to be evolutionarily related. Although the nomenclature of non-vertebrate chordates uses both Slc26a and SulP designations, the paralog designation of Slc26a no longer follows the same classification as vertebrate Slc26a. For clarification, this work will hereafter refer to the SulP family as the entire family of these related proteins, to Slc26as as all vertebrate proteins in the SulP family of proteins, to Sultrs as all plant proteins in the SulP family of proteins, and to SulPs as all remaining species of proteins in the SulP family of proteins (note that the *Caenorhabditis elegans* SulP1 should not be considered orthologous to either Slc26a1 or other SulP1 proteins from other species., because orthology can no longer be established using this analysis).

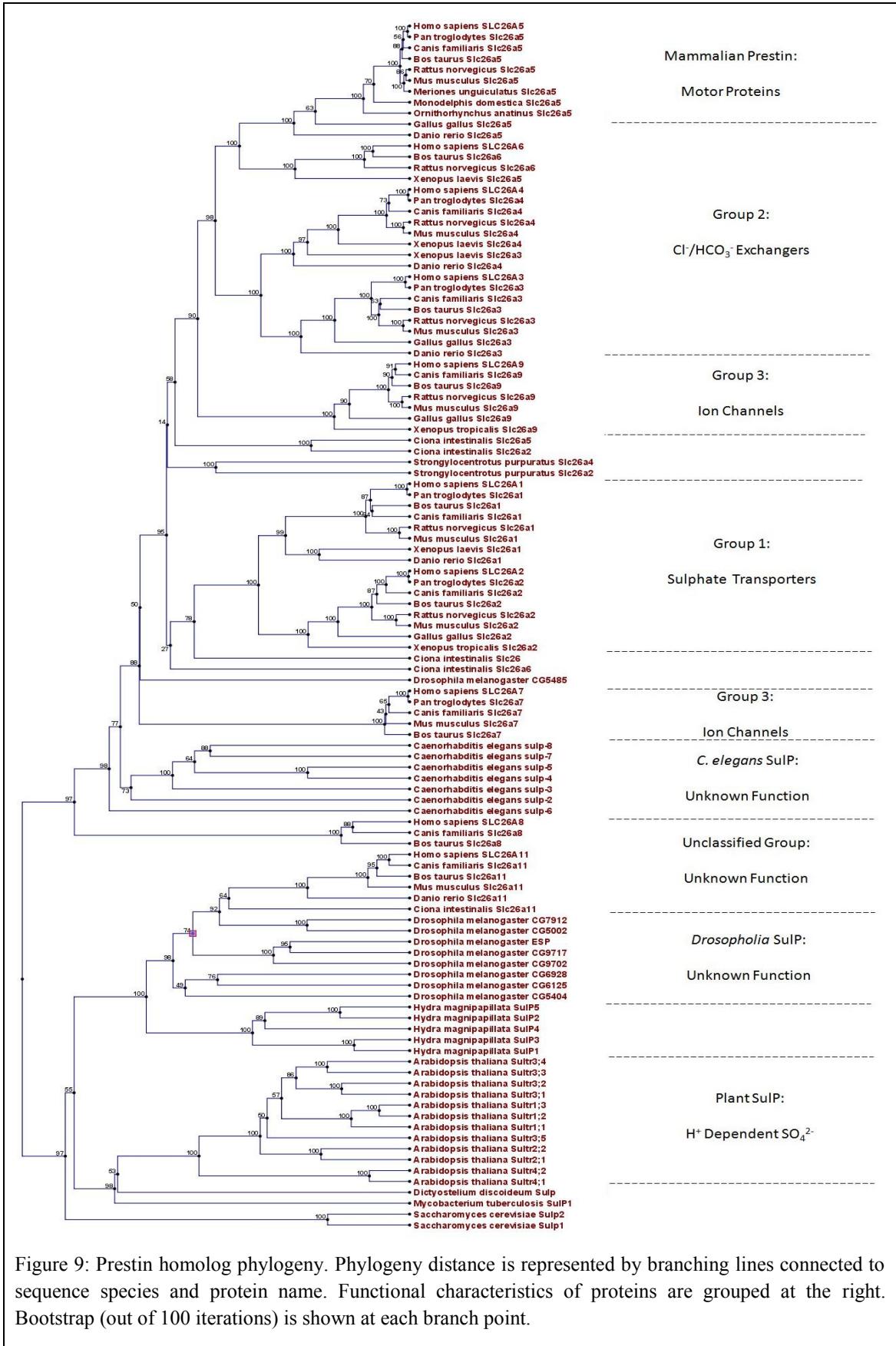


Figure 9: Prestin homolog phylogeny. Phylogeny distance is represented by branching lines connected to sequence species and protein name. Functional characteristics of proteins are grouped at the right. Bootstrap (out of 100 iterations) is shown at each branch point.

The earliest major branch occurred in an ancient chordate, with the mammalian Slc26a11 being more similar to plant homologous sequences and Slc26a1 to Slc26a9 being more similar to the rest of the chordate homologous sequences. This branching pattern is also seen with *D. melanogaster*, which has the majority of its SulP paralogs grouped to the Slc26a11 branch, while one paralog, CG5485, is grouped with the Slc26a1-9 branch. Interestingly, the CG5485 protein, which is found in the Johnston's organ, is more closely related to the Slc26a1-9 branch than to the other *D. melanogaster* paralogs. Also of note, Slc26a11 is the only mammalian SulP paralog without a significant intervening sequence (IVS, see Chapter IV for more details). Although the function of the Slc26a11 paralogs is unknown, some plant homologous sequences are dependent on H⁺ for function.

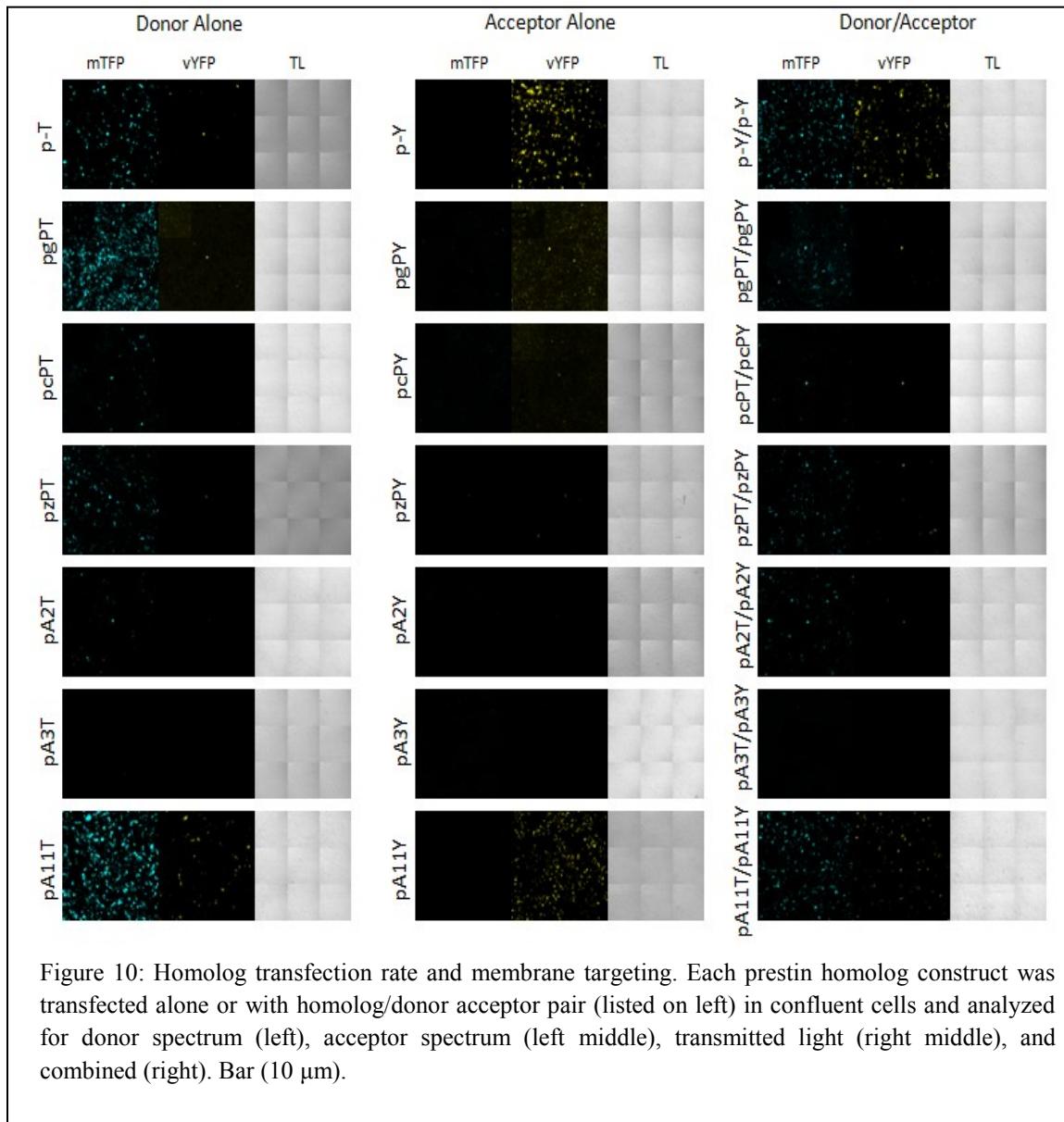
After the Slc26a11 branch, the next major branch in the SulP family includes the Slc26a8 paralogs. Currently, the function of Slc26a8 is unknown. The *C. elegans* paralogs form the next major branch, but their function is also uncharacterized. The Slc26a7 paralogs, which form the next major branch, are characterized as ion channels. The Slc26a9 paralogs, which also function as ion channels, are separated from the Slc26a7 branch by another major branch that includes the Slc26a1 and Slc26a2 paralogs which function as sulfate transporters. The last major branch of SulP proteins includes prestin (Slc25a5), Slc26a3, Slc26a4, and Slc26a6 paralogs. All of these paralogs, except for the mammalian prestin orthologs, act as of Cl⁻/HCO₃⁻ exchangers. Of note is the high sequence similarity of mammalian prestin, especially eutherian mammals, when compared to other paralogs (for functional reviews see [43, 56]).

Based on SulP phylogeny and available cDNA, a selection of proteins of interest were chosen to test homo-oligomerization in the SulP family based on SulP phylogeny. The cDNA for gerbil prestin (gP) is the most commonly used mammalian prestin in prestin physiology and was therefore used in the following experiments to represent mammalian prestin. Both chicken (cP) and zebrafish prestin (zP) are readily available non-mammalian prestin orthologs. Human SLC26A2 (A2) and SLC26A3 (A3) represented major branch points of prestin during chordate evolution. SLC26A11 (A11) is the most distant paralog and may have branched from the rest of the SulP family during or before early animal evolution. These six proteins were examined as representatives of major branches in SulP family homologous sequences. SLC38A2 (abbreviated as 38) was a readily available membrane protein that is unrelated to the SulP family. The cDNA for these proteins were ligated inframe to cDNA for donor or acceptor fluorescent proteins to create donor (pgPT, pcPT, pzPT, pA2T, pA3T, and pA11T) plasmid constructs and acceptor (pgPY, pcPY, pzPY, pA2Y, pA3Y, and pA11Y) plasmid constructs.

B. Protein Synthesis and Targeting

Not all proteins were equally synthesized from the plasmid constructs in the cells despite similar transfection protocols. Cells were transfected with equal concentration of donor, acceptor, or donor and acceptor (co-transfection noted as donor/acceptor) to analyze number of fluorescent cells, fluorescent intensity in cells, and membrane targeting (Fig. 10, Table 1). The majority of cells transfected with the p-T plasmid construct (plasmids containing only the cDNA for mTFP) were fluorescent. Cultures transfected with either pgPT or pA11T also had fluorescence in the majority of cells and

were membrane targeted rather than throughout the cytoplasm. Cellular fluorescent intensity and prevalence in cultures transfected with pcPT, pzPT, and pA2T were diminished compared to p-T, pgPT, and pA11T transfected cells. The fluorescent intensity of cells in cultures transfected with pA3T was below the limits of detection, despite correct cDNA insertion. The fluorescent intensity and prevalence in cells transfected with acceptor constructs were diminished compared to cells transfected with donor constructs in all experiments, except the p-Y transfected cells. When co-transfected, the intensity and number of fluorescent cells were further diminished relative to transfection with either donor or acceptor alone, in all plasmid constructs except p-T and p-Y plasmid constructs.



	Fluorescent Intensity	Transfection Rate	Membrane Targeted	Cell Confluency
p-T	+++	+++	No	> 90%
pgPT	++	+++	Yes	> 90%
pcPT	+	+	Yes	> 90%
pzPT	+	+	Yes	> 90%
pA2T	+	+	Yes	> 90%
pA3T	-	-	N/A	> 90%
pA11T	+++	+++	Yes	> 90%

	Fluorescent Intensity	Transfection Rate	Membrane Targeted	Cell Confluency
p-Y	+++	+++	No	> 90%
pgPY	+	++	Yes	> 90%
pcPY	+	+	Yes	> 90%
pzPY	+	+	Yes	> 90%
pA2Y	+	+	Yes	> 90%
pA3Y	-	-	N/A	> 90%
pA11Y	+++	+++	Yes	> 90%

	Fluorescent Intensity	Transfection Rate	Membrane Targeted	Cell Confluency
p-T/p-Y	++	+++	No	> 90%
pgPT/pgPY	+	++	Yes	> 90%
pcPT/pcPY	+	+	Yes	> 90%
pzPT/pzPY	+	+	Yes	> 90%
pA2T/pA2Y	+	+	Yes	> 90%
pA3T/pA3Y	-	-	N/A	> 90%
pA11T/pA11Y	++	++	Yes	> 90%

Table 1: Quantitative estimates of plasmid transfection efficiencies in HEK cells.

C. FRET Efficiency of Controls

The sensitivity of the FRET detection system was determined using multiple controls (Fig. 11). Two positive controls were used: the pLink, in which cCFP is ligated in-frame to vYFP; and pgPT/pgPY, in which gerbil prestin donor and acceptor plasmid constructs are co-transfected. pLink-transfected cells were found to have significantly greater E_{FRET} than cells transfected with the pgPT/pgPY pair (Mann-Whitney $p \leq 0.001$), likely due to the forced 1:1 donor:acceptor ratio in the pLink synthesized protein as well as the forced short distance between the donor and acceptor fluorophores.

Several negative controls were used, including donor alone transfect cells (p-T and pgPT) and donor co-transfected with a non-oligomerizing plasmid construct (p-T/p-Y and pgPT/p38Y). These negative controls all had significantly smaller FRET efficiencies than those of the positive controls (Mann-Whitney $p \leq 0.001$). This verifies the system's ability to detect difference between oligomerizing and non-oligomerizing proteins. The fluorescent proteins did not appear to interact with themselves or each other, based on the negligible FRET efficiencies in cells transfected with p-T plasmid constructs or co-transfected with p-T/p-Y plasmid constructs ($E_{FRET} = 0.92\%$ and 0.13% , respectively). Cells co-transfected with pgPT and the non-oligomerization p38Y had E_{FRET} values that were not significantly different from p-T or p-T/p-Y transfected cells, suggesting that nonspecific interactions did not significantly contribute to E_{FRET} ($E_{FRET} = 1.70\%$, Mann-Whitney > 0.05). Cells transfected with pgPT alone, however, had significantly greater FRET efficiencies than either p-T or p-T/p-Y transfected cells (Mann-Whitney test $p \leq 0.001$). This higher FRET efficiency may be due to the formation of puncta by gerbil

prestin in HEK cells, resulting in donor/donor FRET [82, 83]. The FRET efficiencies of gPT transfected cells, however, were indistinguishable from the FRET experiments of the gPT/p38Y transfected cells.

For the rest of this report, the FRET efficiencies of pgPT/pA38Y transfected cells was considered the lower threshold detection level of FRET ($E_{FRET} = 1.70\%$) and the FRET efficiency of pgPT/pgPY transfected cells was considered the normal E_{FRET} of gerbil prestin subunits ($E_{FRET} = 7.64\%$).

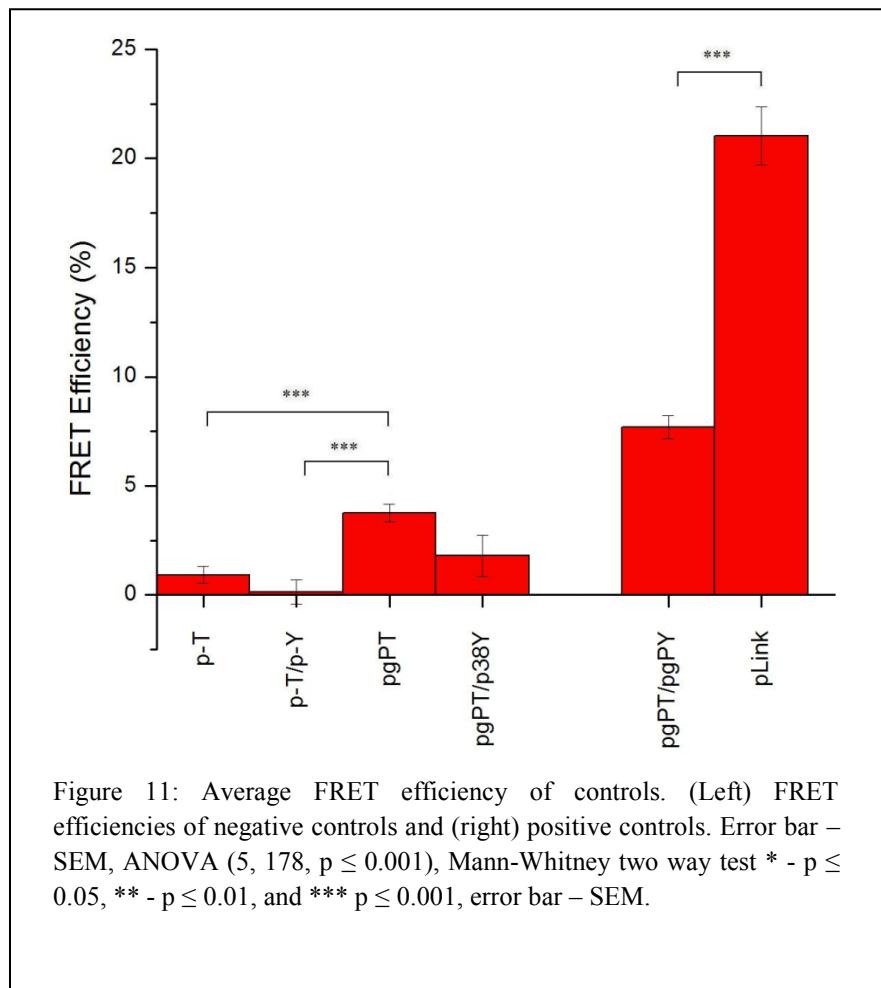
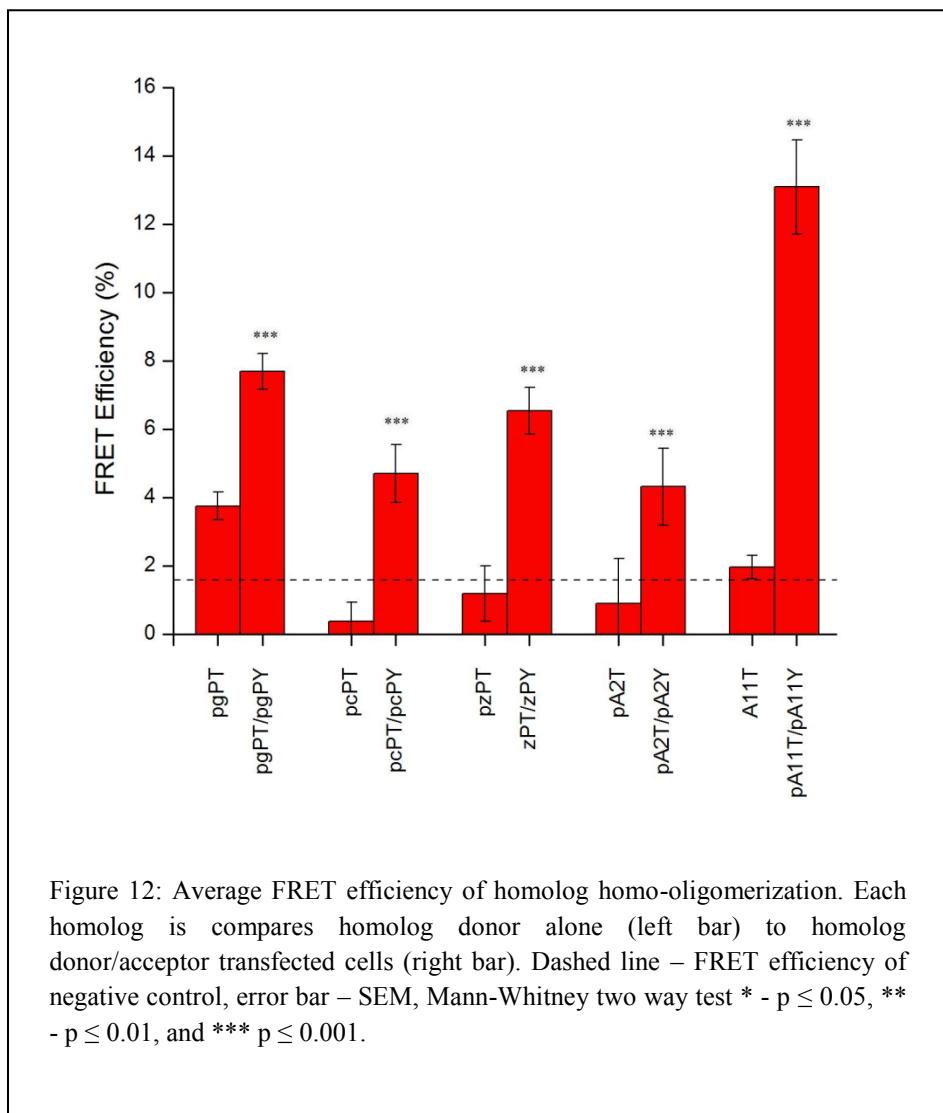


Figure 11: Average FRET efficiency of controls. (Left) FRET efficiencies of negative controls and (right) positive controls. Error bar – SEM, ANOVA (5, 178, $p \leq 0.001$), Mann-Whitney two way test * - $p \leq 0.05$, ** - $p \leq 0.01$, and *** $p \leq 0.001$, error bar – SEM.

D. FRET Efficiency of Homo-Oligomerization and Hetero-Oligomerization

FRET was used to examine homo-oligomerization in cells transfected with various prestin homologous sequences (Fig. 12). Cells were transfected with donor alone (negative bleach control) or donor and acceptor (test) plasmid constructs. Donor/acceptor co-transfected cells for all the prestin orthologs (gPres, cPres, and zPres), as well as the prestin paralogs (A2 and A11), all showed significant increases in FRET efficiency over cells transfected with donor alone or pgPT/p38Y (Student t-test $p \leq 0.001$, Dunnett's t-test $p \leq 0.05$, respectively).



The ability of each prestin homologous sequence to undergo hetero-oligomerization with gerbil prestin was also examined (Fig. 13). The donor pgPT plasmid construct was transfected along with the acceptor plasmid construct for each prestin homologous sequence. Only one co-transfection, pgPT/pA2Y, had a significantly reduced FRET from pgPT/pgPY ($E_{FRET} = 3.90\%$ and 7.69% , respectively; Dunnett's two way t-test $p \leq 0.05$). The pgPT/pA2Y E_{FRET} , however, was significantly higher than the negative control ($E_{FRET} = 3.90\%$ and 1.80% , respectively; Dunnett's one way t-test $p \leq 0.05$). The E_{FRET} observed in the remaining cells co-transfected with pgPT and given homologous sequence-coupled acceptor were not significantly different from E_{FRET} of cells co-transfected with pgPT/pgPY, but was significantly greater than the FRET efficiency of the negative control cells co-transfected with pgPT/p38Y (Dunnett's one way t-test $p \leq 0.05$).

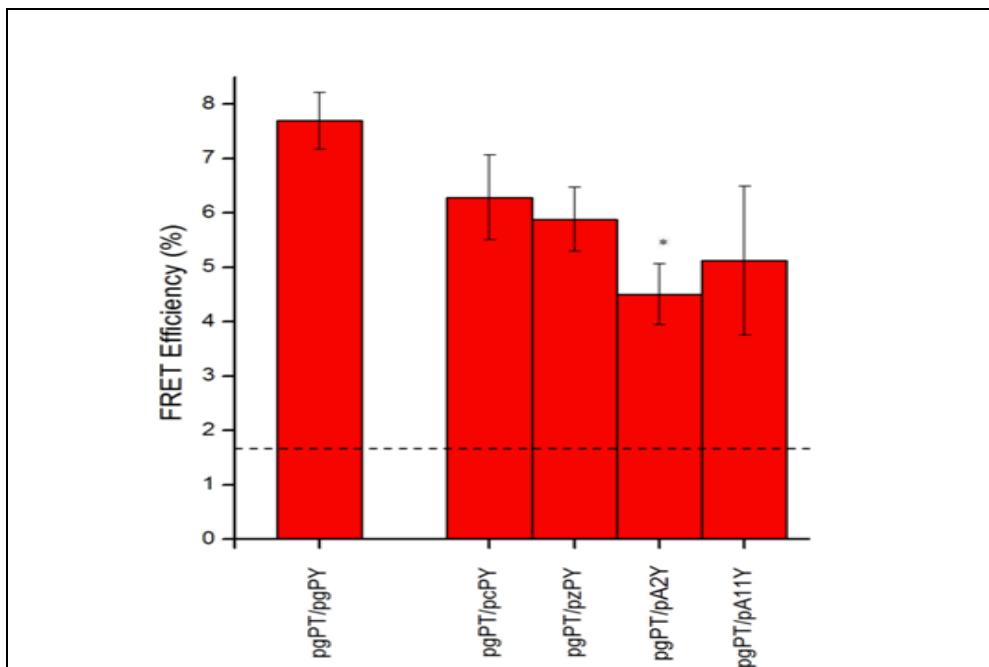


Figure 13: Figure 12: Average FRET efficiency of homolog hetero-oligomerization. FRET efficiency is compared between cells transfected with gerbil prestin donor/acceptor constructs (left) and gerbil prestin donor/homolog acceptor constructs (right). Dashed line – FRET efficiency of negative control, error bar – SEM, ANOVA (4, 78, $p \leq 0.01$), Dunnett's two way t-test * - $p \leq 0.05$, ** - $p \leq 0.01$, and *** $p \leq 0.001$.

III. Conclusion

This chapter examined the conservation of oligomerization in prestin homologous sequences. Five homologous sequences were tested for oligomerization based upon their phylogeny, function, and protein synthesis (*M. unguiculatus* Slc26a5, *G. gallus* Slc26a5, *D. rerio* Slc26a5, *H. sapiens* SLC26A2, and *H. sapiens* SLC26A11). All homologous sequences had positive indications of homo-oligomerization. In addition, all homologous sequence, except for *H. sapiens* SLC26A2, demonstrated interaction with gerbil prestin. This suggests that not only is homo-oligomerization conserved throughout the SulP family, but also that similar oligomerization motifs may be conserved.

CHAPTER IV: Identification of Oligomerization Motifs

As demonstrated in Chapter III, oligomerization is a conserved property of the SulP family. Thus, it is hypothesized that this conserved property is associated with a conserved structure. Sequence analysis was used to examine the structural elements of prestin. Prestin structure was examined on three levels: 1) higher order structures, 2) motifs, and 3) primary structure. The higher order structures of prestin were examined by first delineating the major structures (i.e. domains) in prestin. The identified prestin domains were then compared to other related domains to determine conservation of primary, secondary, and tertiary structure. Where appropriate, a template based model was created. Next, prestin motifs were examined by comparing prestin with known motifs as well as analyzing prestin homologous sequences for *de novo* motifs. Finally, the primary structure of prestin was analyzed across homologous sequences to determine the amino residue conservation. Structures conserved throughout prestin and its relatives indicate structures important for oligomerization which was then used to determine the structure-function relationship of prestin homo-oligomerization (Chapter V).

I. Materials and Methods

A. Sequence Curation

Although gerbil prestin is the most often used prestin ortholog in biological hearing research studies, the human genomic, and corresponding proteomic data is the best researched sequences in bioinformatic studies. Therefore, the initial sequence used for sequence analysis was obtained from NCBI's RefSeq for *H. sapiens* SLC26A5 isoform a protein (NP_945350.1). Since the output of any algorithm depends on the

assumptions and contents of the input, five separate sequence database types were curated to support various analyses: Type I sequences based on domain diversity, Type II sequences based on evolutionary splits, Type III sequences based on kingdoms, Type IV sequences based on non-mammalian kingdoms, and Type V sequences based on the most diverse mammalian sequences available (*i.e.* proteria, metatheria, and eutheria). Different sequence sources were used to populate these databases.

For Type I databases, FASTA files were derived from NCBI's Conserved Domain Database (CDD) [45]. For all domain families, the first 10 mFASTA files returned from the “the most diverse members” Type Selection were placed into a FASTA format flat file. FASTA file titles were modified to include accession numbers as well as species and gene name (see Appendix III for list of sequences in databases).

For the remaining databases, sequences were curated from both NCBI and Ensembl BLAST queries based on species of interest. For each paralog in a given species, the FASTA file for the isoform most similar to NP_945350.1 was kept for further analysis. FASTA files were renamed based on organism and phylogeny. Each FASTA file was submitted to NCBI's CDD and sequences which did not have both the SulP and STAS domain, in order from amino to carboxy terminal, were removed from the initial database. The remaining FASTA files were curated in CLC Workbench (v5.1) and submitted to the ‘sequence alignment’ algorithm. Any sequence with a large gap or insert (> 50 residues) compared to its closest homologous sequence was removed from the initial database. Specific sequences were then chosen to populate the remaining four databases based on evolutionary branches (Type II), kingdoms (Type III), non-mammalian kingdoms (Type IV), or diverse mammals (Type V).

B. Sequence Analysis

Multiple programs were used for sequence analysis (Table 2). The input to each program was either the prestin RefSeq (NP_945340.1) or one of the five types of curated databases listed above. For clarity, input and run-time parameters are noted with each result. Detailed descriptions of servers, software, databases, and algorithms can be found in Appendix III.

Type	Name	Description	References
Servers and Software Packages	National Center for Biotechnology Information (NCBI)	American Governmental Server	
	European Molecular Biology Laboratory - European Bioinformatic Institute (EMBL-EBI)	European Governmental Server	
	Expert Protein Analysis Systems (ExPASy)	Swiss Governmental Server	
	Bioinformatic Toolkit (BITK)	Max Plank Institute Server	06 Biegert
	CLC Main Workbench (CLC)	Commercial Software	
	Interactive Structure Based Alignment Program (STRAP)	Humboldt University (Germany) Software	01 Gille
Databases	Entrez Gene	RefSeq Genes	05 Maglott
	Ensembl	Gene and Proteins	09 Hubbard
	Entrez Protein	RefSeq Proteins	
	InterPro	Protein Signature Sequences	09 Hunter, 05 Quevillon
	Prosite	Protein Sites, Motifs, Domains, and Families	
	Entrez Conserved Domains Database (CDD)	Protein Domains	09 Marchler-Bauer
	Pfam	Protein Domains	10 Finn
	Transport Classification Database (TCDB)	Protein Families	09 Saier
Algorithms	Basic Local Alignment Search Tool (BLAST)	Sequence Comparison for Database Searches	90 Altschul
	HHpred	Sequence Comparison for Database Searches	05 Soding, 05 Soding
	Create Alignment	Sequence Comparison for Alignment	87 Feng
	ClustalW and ClustalX	Sequence Comparison for Alignment	07 Chenna
	Unweighted Pair Group Method with Arithmetic Mean (UPGMA)	Sequence Comparison for Phylogeny	73 Sneath
	2ZIP	Residue Analysis for Leucine Zippers	
	Coils	Residue Analysis for Coiled-Coils	91 Lupas
	TMPred	Transmembrane Modeling	93 Hofmann
	TOPCONS	Transmembrane Modeling	09 Bensel
	Swiss-Model	Template Based Modelling	06 Arnold, 09 Kiefer, 06 Kopp, 03 Schwede

Table 2: List of software, database, and algorithms used in this study.

II. Results

A. Higher Order Structure

Domains

The *H. sapiens* SLC26A5 prestin FASTA residue sequence (NP_945350.1) was submitted to four separate domain database searches: The National Center for Biotechnology Information's (NCBI) Conserved Domain Database (CDD), European

Molecular Biology Laboratory - European Bioinformatic Institute's (EMBL-EBI) InterPro, Welcome Trust Sanger Institute's Protein Family Database (Pfam), and the Expert Protein Analysis System's (ExPaSy) PROSITE. As stated in the Background, two different domains can be identified in SulP proteins: one domain in the transmembrane region of protein and one domain in the intracellular carboxy region of the protein.

Three of the four databases identified the domain in the transmembrane portion of the prestin protein (Fig. 14). Although NCBI's CDD does not have its own record for the family of this domain, the domain is associated with the Xanthine Uracil Permease Superfamily (XUP, CL00967). The XUP superfamily is associated with several other records including two records from Pfam (PF00916 and PF00860), two records from Clusters of Orthologous Groups of proteins (COG; COG2233 and COG2252), two records from Protein Clusters (PRK10720 and PRK11412), and two records from The Institute for Genomic Research's (TIGR) database of protein families (TIGR00801, TIGR03713) (data not shown). InterPro also indicates a similar domain to the CDD record (IPR011547), but it is based on the Pfam record PF00916. The Pfam record for this domain is PF00916 and is part of the clan for the Amino Acid-Polyamine-Organocation (APC) superfamily (CL0062). This clan includes 16 domain family members (Table 3).

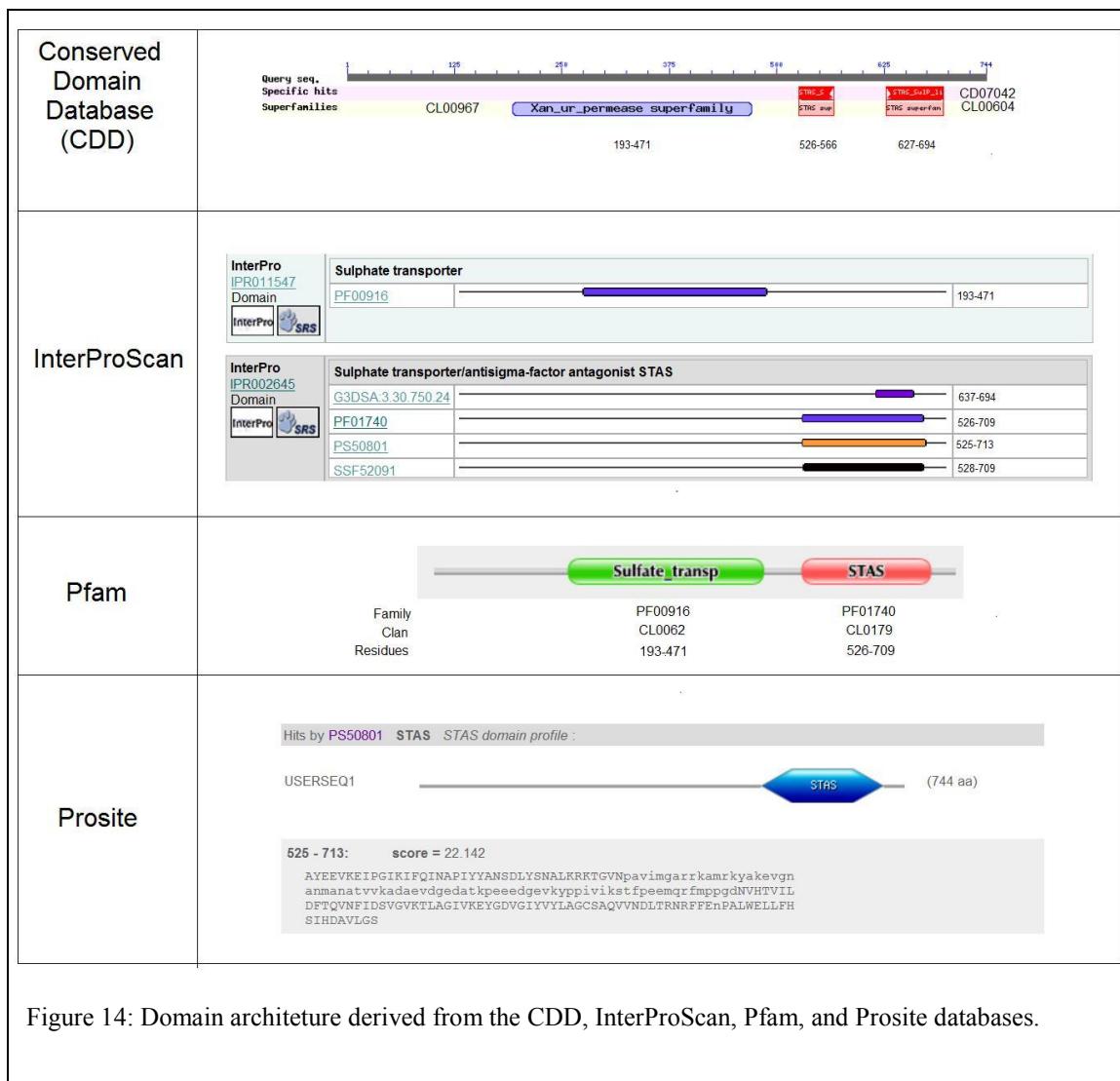


Figure 14: Domain architecture derived from the CDD, InterProScan, Pfam, and Prosite databases.

Clan	Domain Family	Abbreviation	Pfam #
Amino Acid-Polyamine-Organocation (APC) (CL0062)	Amino Acid Permease	AAP	PF00324
	Amino Acid Transporter	AAT	PF01490
	Benzoate E Transporter	BenE	PF03594
	Branched-Chain Amion Acid Transporter	BCAAT	PF05525
	Carbon Starvation Protein	CstA	PF02554
	Cobalt Transporter	ChiQ	PF02361
	HCO ₃ ⁻ Transporter	HCO3	PF00955
	Natural Resistance-Associated Macrophage	Nramp	PF01566
	Cytosine/Purine, Uracil, Thiamine, Allantoin Transporter	CPT	PF02133
	Sodium:Alanine Symporter	NAS	PF01235
	Sodium:Solute Symporter	SSF	PF00474
	Spore Germination Protein	SprG	PF03845
	SulP Transporter	SulP	PF00916
	Tripartite Tricarboxylate Transporter	TctB	PF07331
	Tryptophan/Tyrosine Permease	TTP	PF03222
	Xanthine Uracil Permease	XUP	PF00860

Table 3: List of APC clan members.

The carboxy terminal domain is noted in all four databases and is referred to as the Sulfate Transporter and Anti-Sigma Factor Antagonist (STAS) domain. CDD lists the STAS domain in prestin as a more specific STAS SulP-like Sulfate Transporter (CD07042) which is part of the STAS superfamily (CL00604). The STAS superfamily includes three CDD records (CD07041, CD07042, and CD07043), one Pfam record (PF01740), two COG records (COG1366 and COG3113), and two TIGR records (TIGR00377 and TIGR02886). CDD shows the human prestin STAS domain encompassed by residues 526-694, but with a break (*i.e.* IVS) in the domain from residue 567 to residue 626. The InterPro record for the STAS domain (IPR002645) is derived from four different sources. Gene3D (3.30.750.24), Pfam (PF01740), Prosite (PS50801), and Superfamily (SSF52091) listed the STAS domain residues as 637-694, 526-709, 525-713, and 528-709, respectively. As shown in InterPro, the Pfam record for the STAS domain extends from residue 525-709. Pfam places the STAS domain in the ATP-Grasp

clan (CL0179). This clan includes 14 Pfam clan members (Table 4). Prosite's record for the STAS domain is PS50801. In human prestin, the STAS domain is encompassed by 525-713, but also shows a break in the STAS domain (indicated by lower case letters) between residues 567-637. It is unclear from these records where the STAS domain begins, ends, and breaks.

Clan	Domain Family	Abbreviation	Pfam #
ATP-Grasp (CL0062)	ATP-Grasp 1	ATP-Grasp_1	PF02222
	ATP-Grasp 2	ATP-Grasp_2	PF08442
	ATP-Grasp 3	ATP-Grasp_3	PF02655
	Carbamoyl-Phosphate Synthetase L Chain ATP Binding	CSPase	PF02786
	D-Ala D-Ala Ligase	Dala_Dala	PF07478
	DUF1297	DUF1297	PF06973
	DUF407	DUF407	PF04174
	Phosphoribosylglycinamide Synthetase ATP-Grasp	GARS_A	PF01071
	Prokaryotic Glutathione Synthetase ATP-Grasp	GSH-S_ATP	PF02955
	Inositol 1, 3, 4-Triphosphate 5/6-Kinase	Ins134	PF05770
	RimK-Like ATP-Grasp	RimK	PF08443
	Sulfate Transporter and Anti-Sigma Factor Antagonist	STAS	PF01740
	Synapsin ATP Binding	Synapsin_C	PF02750
	Tubulin-Tyrosine Ligase	TTL	PF03133

Table 4: List of ATP-Grasp clan members.

The description of the relationships between domains, domain families, and domain clans differs from one database to the next. For the remainder of this study domain families will be defined as protein domains which have demonstrated evolutionary homology as classified in the Pfam database (designated with a 'PF' prefix). Domain clans will be defined by domain families which have implied evolutionary and/or functional homology as classified in the Pfam database (designated with a 'CL' prefix). It is possible that there is an evolutionary relationship between the prestin domain families (SulP and STAS domain family) and other family members within their respective clans (APC and ATP-Grasp domain clan). In the following sections, the primary structure of

domain family members within these clans will be analyzed to determine structure homology. Any domain families that share structure homology will be grouped as a domain superfamily. Secondary structure of clan members will also be analyzed to determine secondary structure homology. Where appropriate, homology based modeling will be used to imply tertiary structure.

SulP Domain

APC Clan Primary Structure Homology

The CDD-derived Type I database for each member of the Pfam APC clan (CL0062) was used to determine the relationship between the SulP domain family and other APC clan family members. These sequences in each database were submitted to HHpred using 8 local alignment PSI-BLAST iterations and then searched across the Pfam 24.0 database (Fig. 15). The returned e-values were tabulated, and any returned domain with an e-value > 500 was removed (see Appendix III). The results indicated the APC clan was composed of three separate groups. Group 1 includes the prestin SulP domain family (PF00916) as well as the BenE (PF03594), XUP (PF00860), and HCO3 (HCO3) domain families. Group 3 had both ChiQ (PF02361) and TctB (PF07331) domain families, which were not related to either group 1 or group 2, nor were they related to each other. Group 2 includes the rest of the APC clan members. The grouping of SulP, BenE, XUP, and HCO3 domain families suggest a structural homology and will be designated as the SulP domain superfamily. More PSI-Blast iterations or more sophisticated algorithms, however, may establish help redefine the SulP superfamily within the APC clan.

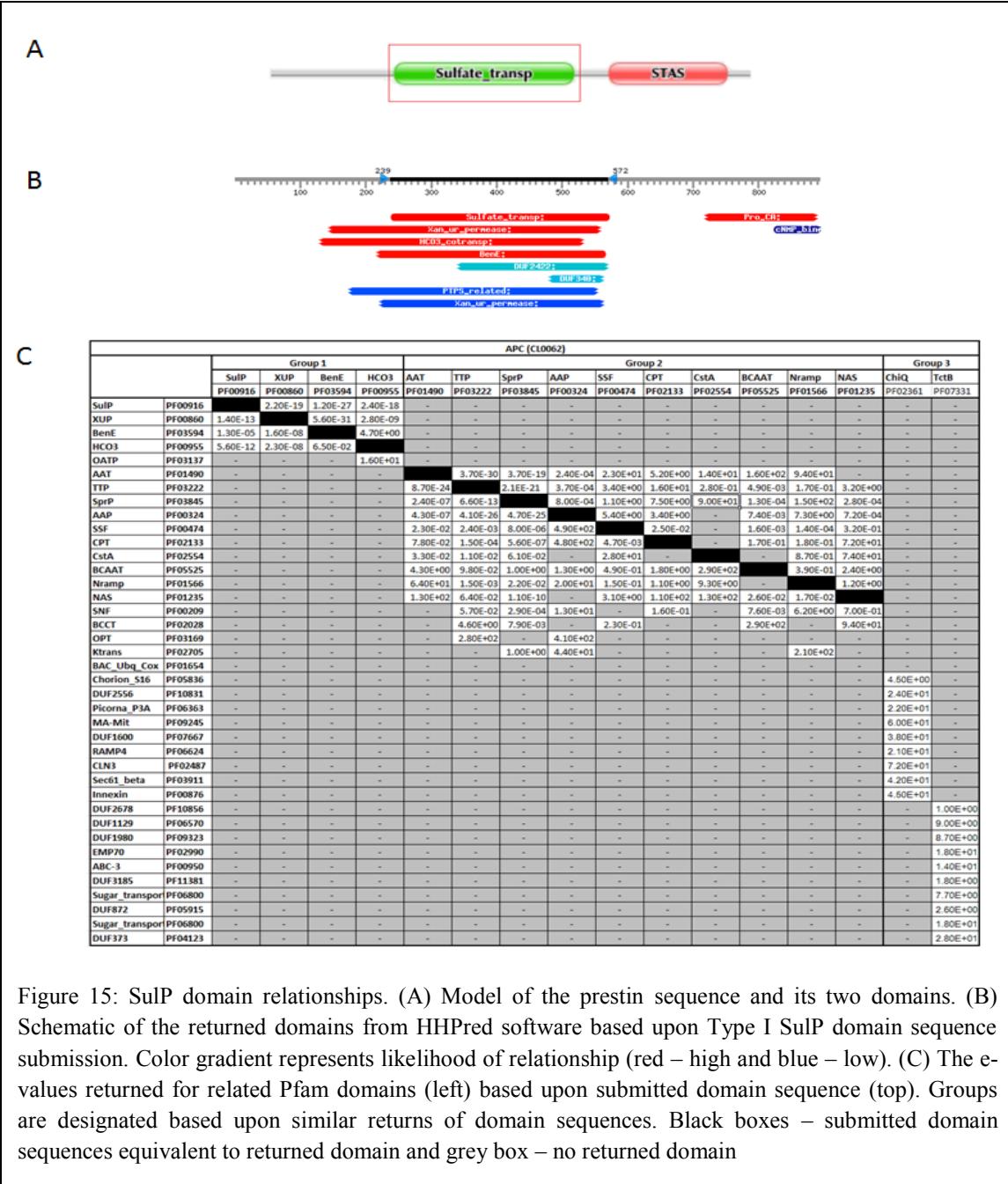


Figure 15: SulP domain relationships. (A) Model of the prestin sequence and its two domains. (B) Schematic of the returned domains from HHpred software based upon Type I SulP domain sequence submission. Color gradient represents likelihood of relationship (red – high and blue – low). (C) The e-values returned for related Pfam domains (left) based upon submitted domain sequence (top). Groups are designated based upon similar returns of domain sequences. Black boxes – submitted domain sequences equivalent to returned domain and grey box – no returned domain

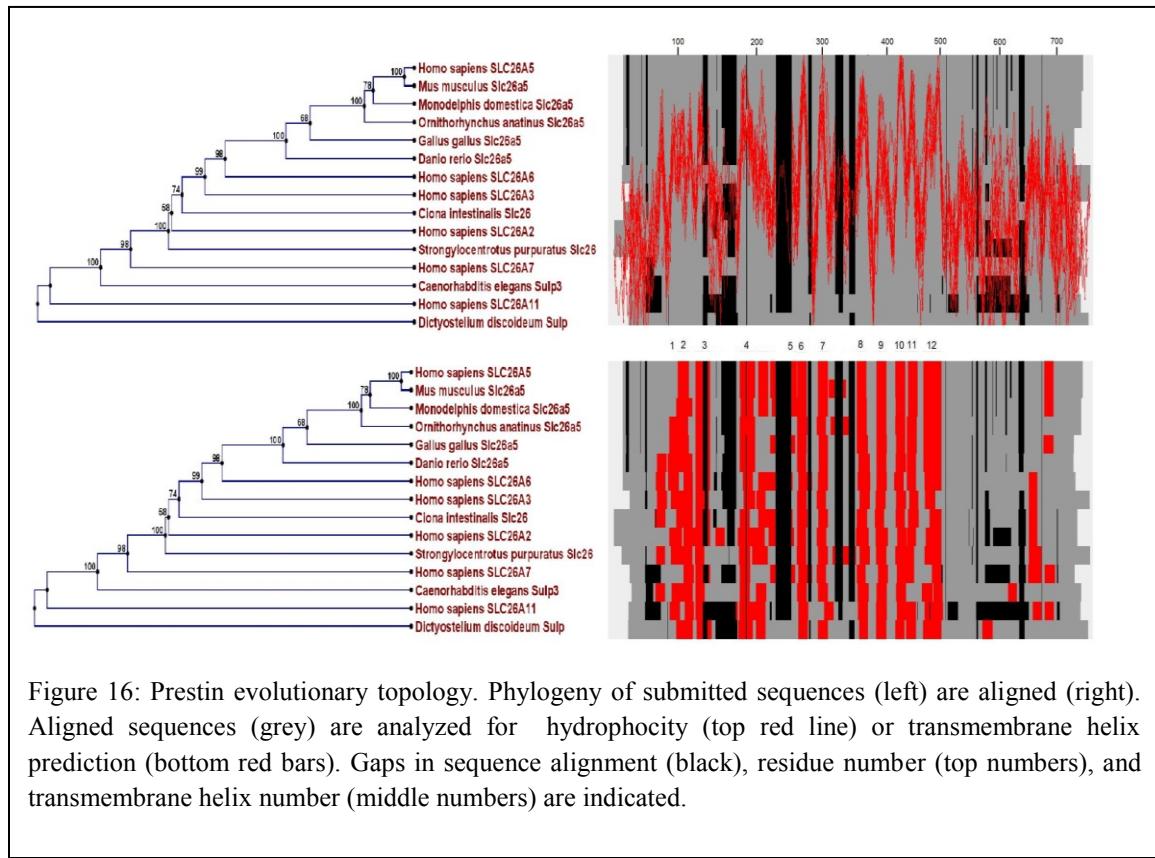
In the SulP domain superfamily, the prestin SulP domain family is most closely related to the XUP family (SulP compared to XUP e-value = 1.4e^{-13} , XUP compared to SulP e-value = 2.2e^{-19}). The SulP family showed a closer relationship to the HCO3 family than the BenE (SulP compared to HCO3 e-value = 5.6e^{-12} , SulP compared to BenE e-

value = $1.3e^{-5}$), but the XUP family was more closely related to the BenE family than the HCO3 family (XUP compared to HCO3 e-value = $2.3e^{-8}$, XUP compared to BenE e-value = $1.6e{-8}$). The HCO3 family had the weakest relationship to the BenE family (HCO3 compared to BenE e-value = 4.7). This pattern was similar for all reciprocal experiments. This suggests XUP is the most closely related domain family to SulP, followed by HCO3. BenE is the least closely related domain. Although no SulP domain superfamily members were related to other APC clan family members, HCO3 showed a weak relationship to the OatP domain family (PF03137, e-value = 16), which, according to Pfam, is part of the large major facilitator superfamily clan (MFS CL0015). The relationships between the SulP, HCO3, XUP, BenE, and OatP family domains were examined further using secondary structure analysis.

SulP Secondary Structure

Secondary structure of the sequences for SulP domain superfamily members were analyzed for conservation of structure. Since the SulP domain in prestin is contained in the plasma membrane portion of the protein, secondary structure was analyzed in terms of transmembrane helices (TMHs). TMHs of prestin homologous sequences were analyzed prior to comparison across SulP domain superfamily. Prestin homologous sequence secondary structure was compared along evolutionary lines using a Type II database. The sequences chosen for the Type II database were based on major sequence branches and functional changes from the larger phylogeny (Fig. 16, see Chapter III Fig. 9 for large phylogeny). Four mammalian orthologs were aligned, including a metatherian (opossum, *Monodelphis domestica*) and a prototherian (platypus, *Ornithorhynchus*

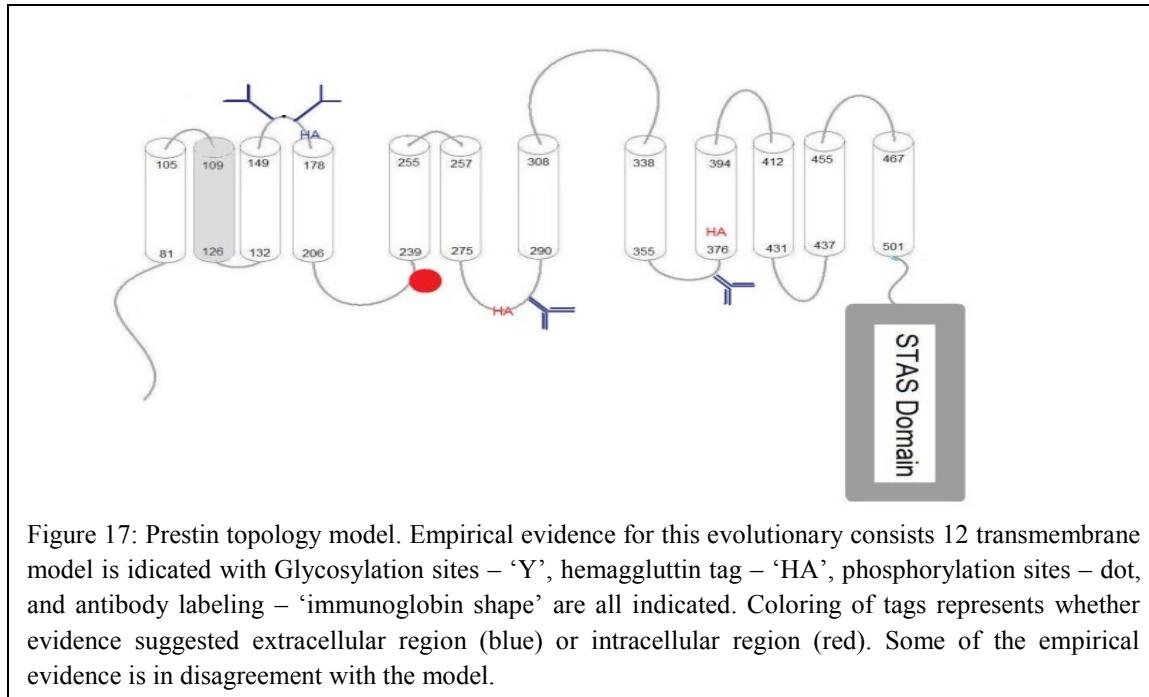
anatinus) ortholog, both of which are thought to act as motor proteins. The chicken and zebrafish prestin orthologs were also analyzed and function as $\text{Cl}^-/\text{HCO}_3^-$ exchangers. Slc26a6, Slc26a4, and Slc26a3 are the closest paralogs to prestin and, along with non-mammalian Slc26a5 proteins, function as $\text{Cl}^-/\text{HCO}_3^-$ exchangers. Slc26a2 represents prestin homologous sequences that function as SO_4^{2-} transporters. Slc26a7 represent prestin homologous sequences that function as ion transporters. The sea squirt (*Ciona intestinalis*), sea urchin (*Strongylocentrotus purpuratus*), and round worm (*C. elegans*) represent prestin homologous sequences from the invertebrate, echinoderm, and nematode animals, respectively. The Slc26a11 is the most distant prestin paralog and its function is unknown. The amoeba (*Dictyostelium discoideum*) homologous sequence is the most distantly related homologous sequence listed.



Alignment of the type II database sequence was performed using the ClustalW algorithm in the Interactive Structure-based Sequences Alignment Program (STRAP). The aligned sequences were analyzed for hydrophobicity, using the Kyte-Doolittle algorithm, and TMHs, using the TMPred algorithm, in the STRAP software (Fig. 16A and 16B, respectively). TMHs 6-9, denoted on top of TMPred TMHs results, are clearly delineated stretches of hydrophobic residues followed by stretches of hydrophilic residues, which are conserved along prestin homologous sequence evolution. TMHs 10 and 11 are close together, but have hydrophobic regions with a small, but distinct hydrophilic region in between. In the *C. elegans* sequence, however, these two TMHs are calculated as a single TMH. TMH 12 is a relatively long stretch of hydrophobic residues with a small hydrophilic stretch in the middle, a pattern seen across prestin homologous sequence evolution. It is unclear whether this region is composed of 1 or 2 TMHs or a more complex membrane spanning structure. In all homologous sequences, TMHs 1-5 have a complex hydrophobicity profile. The best defined TMH in this area is TMH 4, but TMH 5 is not clearly defined. In some species TMH 5 appears to be combined with TMH 4, but in others it is divided into two separate TMHs. The hydrophobicity profiles of residues in TMHs 4 and 5 show two hydrophobic peaks and three hydrophilic valleys, but these peaks and valleys are relatively small, possibly due to the close approximation of both hydrophilic and hydrophobic residues. TMH 1-3 are poorly delineated. The first hydrophobic peak is bordered by large hydrophilic valleys present in all prestin homologous sequences. The peak itself, however, is relatively small and is not predicted to be a TMH in most sequences. Carboxyl to this peak is a region that is generally hydrophobic with several small peaks and valleys. After TMH 3, a sharp hydrophilic

valley is indicative of a loop between TMHs. Between these two hydrophobic peaks, it is unclear whether there are 2, 3 or more TMHs in any of the prestin evolutionary homologous sequences. It is possible that this area has a complex transmembrane profile and/or could represent a pore region.

A topological model was created based on the prestin evolutionary homologous sequences using both the topology data and empirical data from the literature (Fig. 17). Empirical evidence suggesting extracellular components is shown in blue and intracellular components are shown in red (see Chapter II for details). The SulP motif is shown in TMH 2 as a grey area. N-glycosylation sites are shown as a “Y” at residue N163 and N166. A cyclic nucleotide phosphorylation site at residue S238 is shown as a red sphere. Antibody labeling, shown as a hemagglutinin-like shape, suggests that the loops between TMHs 6 and 7 as well as 8 and 9 are extracellular. This evidence, however, is contradicted by the hemagglutinin evidence, shown as “HA”, which demonstrated that the loops between TMHs 6 and 7 as well as 8 and 9 are intracellular. Despite the contradicting empirical evidence and poorly distinguished topology, a general prestin homologous sequence topology can be characterized by approximately 12 TMHs, intracellular termini, TMHs 1-6 poorly characterized, TMHs 7-9 clearly delineated, TMHs 10 and 11 closely approximated TMHs, and TMH 12 as a large or two closely approximated TMHs.



The prestin evolutionary homologous sequence topology was compared against the topology of APC Group 1 domain family members to determine if secondary structure was conserved among this group. Type I databases were obtained for the SulP, XUP, HCO3, and BenE domain families. These sequences were aligned with ClustalW and TMHs were predicted using TMPred in the STRAP software. The STAS domain and Band3 domain were removed from sequences for clarity (SulP family members and HCO3 family members, respectively). The aligned sequences for BenE, XUP, SulP, and HCO3 domains, along with predicted TMH, are shown in Figure 18.

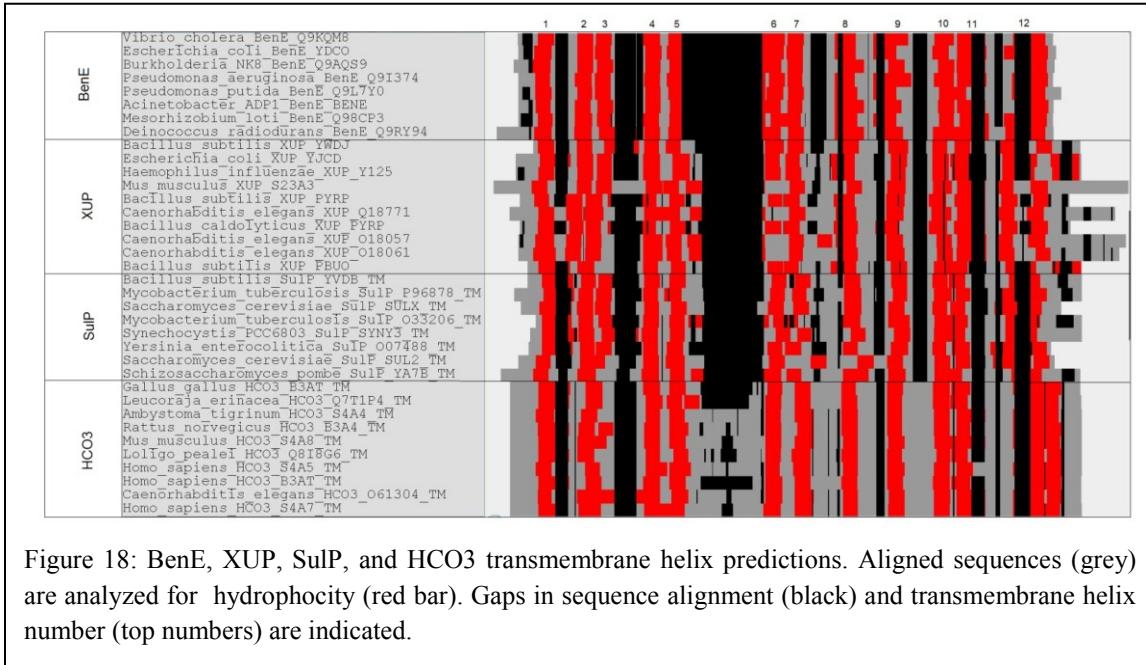


Figure 18: BenE, XUP, SulP, and HCO3 transmembrane helix predictions. Aligned sequences (grey) are analyzed for hydrophobicity (red bar). Gaps in sequence alignment (black) and transmembrane helix number (top numbers) are indicated.

The predicted TMHs in and between the domain families are well aligned and display a similar TMH pattern as the prestin homologous sequences. Approximately 12 TMHs are seen in each of the sequences. TMHs 7-9 are clearly delineated and well aligned in all sequences. TMHs 10 and 11 are well aligned and closely approximated, although in some sequences these two TMHs are predicted to be one large TMH. In SulP and BenE domain sequences, TMH 12 is predicted to be one large TMH, while in XUP and HCO3 domain sequences this same area is predicted to contain two smaller TMHs. TMHs 2-6 are not as clearly delineated as TMH 7-11, but are aligned. One confounding factor is a large insertion between TMH 5 and 6 in the HCO3 domain sequences, which may skew the alignment of the predicted TMH 5 and 6. TMH 2 and 3 are clearly delineated in most of the BenE, XUP, and SulP sequences, but is predicted to be a single TMH in most of the HCO3 sequences. TMH 1 is clearly delineated and well aligned in all sequences. Although there are differences between sequence TMH predictions in these

Group 1 domain families, the conserved general TMH pattern would suggest they have similar tertiary structure.

SulP domain family member topologies were also compared to the OatP domain family as well as other APC clan members through the CPT [PF02133; note: CPT family is also referred to as the Nucleobase:Cation Symporter 1 (NCS1) family and thought to be related to the XUP family which is also referred to as the Nucleobase:Cation Symporter 2 (NCS2) family by the Transporter Classification Database (see www.tcdb.com for details) domain family] (Fig. 19). Both OatP and CPT families have 12 transmembrane domains and intracellular termini as verified by solved structure models (see SulP Tertiary Structure section below). Neither of these two 12 TMH profiles match the SulP family 12 TMH topology profile. All 12 OatP TMHs are clearly delineated with inter-TMH loops of similar size except for one relatively large loop between TMH 9 and 10. OatP TMHs 10, 11, and 12 align with the STAS domain of SulP family members not the transmembrane SulP domain of prestin. The CPT 12 TMH profile does not fare much better in topology comparison to the SulP domain family topology. Although the TMHs are not as clearly delineated as OatP, they are better delineated than the SulP family members. In addition, the CPT TMHs 11 and 12 also align with the STAS domain of the SulP domain family members. Based on these differences in topologies, it is unlikely that SulP domain family members have structure-function relationships similar to members of either the MFS clan or the rest of the APC clan.

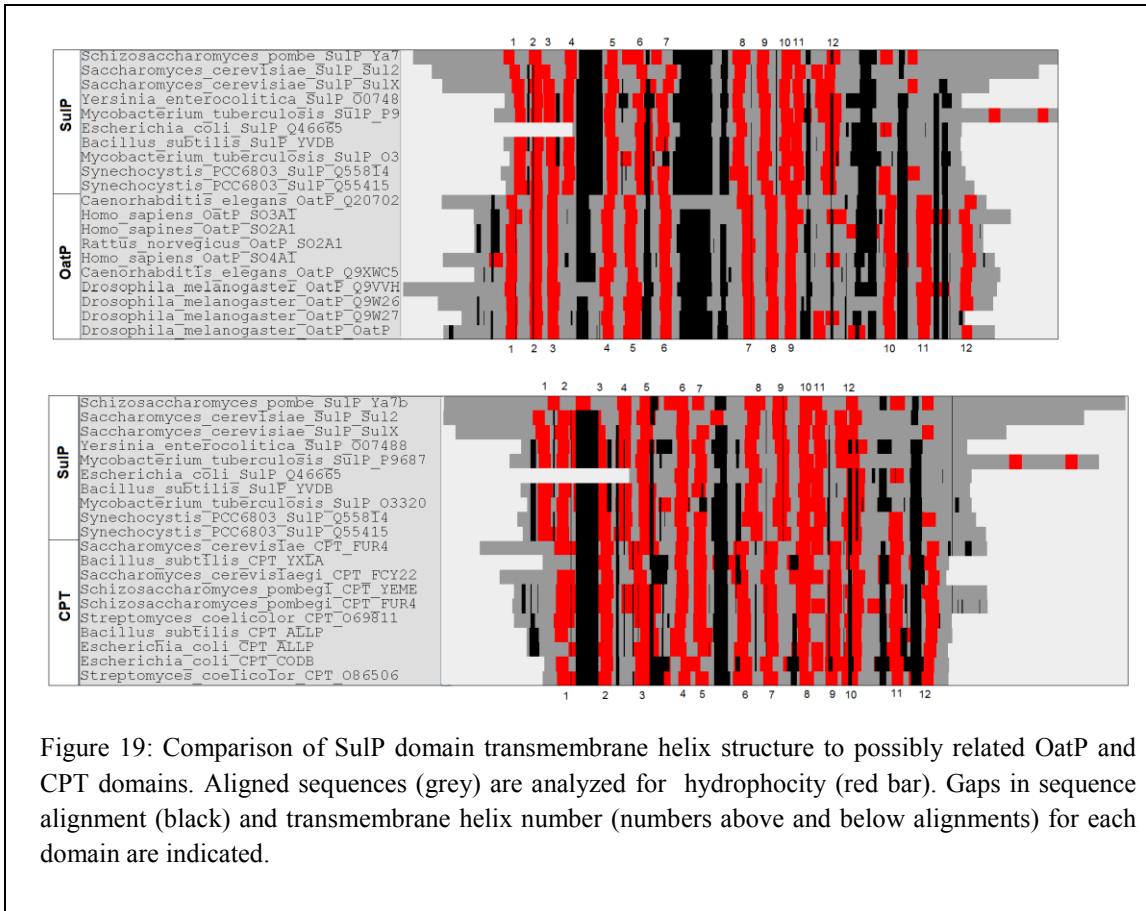


Figure 19: Comparison of SulP domain transmembrane helix structure to possibly related OatP and CPT domains. Aligned sequences (grey) are analyzed for hydrophobicity (red bar). Gaps in sequence alignment (black) and transmembrane helix number (numbers above and below alignments) for each domain are indicated.

Tertiary Structure

Their similar topology suggests that APC clan Group 1 family members may share similar tertiary structure and possibly similar structure-function relationships. A representative from each of the four SulP domain superfamily members was run in HHpred with 8 local alignment PSI-BLAST iterations and against the most current PDB database (Fig. 20). No PDBs were listed for representatives from HCO3 (*H. sapiens* SLC4A1, NP_000333.1), SulP (*H. sapiens* SLC26A5, NP_945350.1), XUP (*H. sapiens* SLC23A1, NP_005838.3), or BenE (*Vibrio cholerae* BenE, YP_001217497.1) protein families (data not shown). An OatP (*H. sapiens* SLC01A2, NP_602307.1) and CPT family representative was also analyzed (Escherichia coli CodB, YP_003220321.1). The

OatP family representative was most closely related to the PDB file 1PW4, while the CPT family representative was most closely related to the PDB file 2JLN. Due low sequence homology and secondary structure profile, the human prestin sequence was not modeled against any PDB.

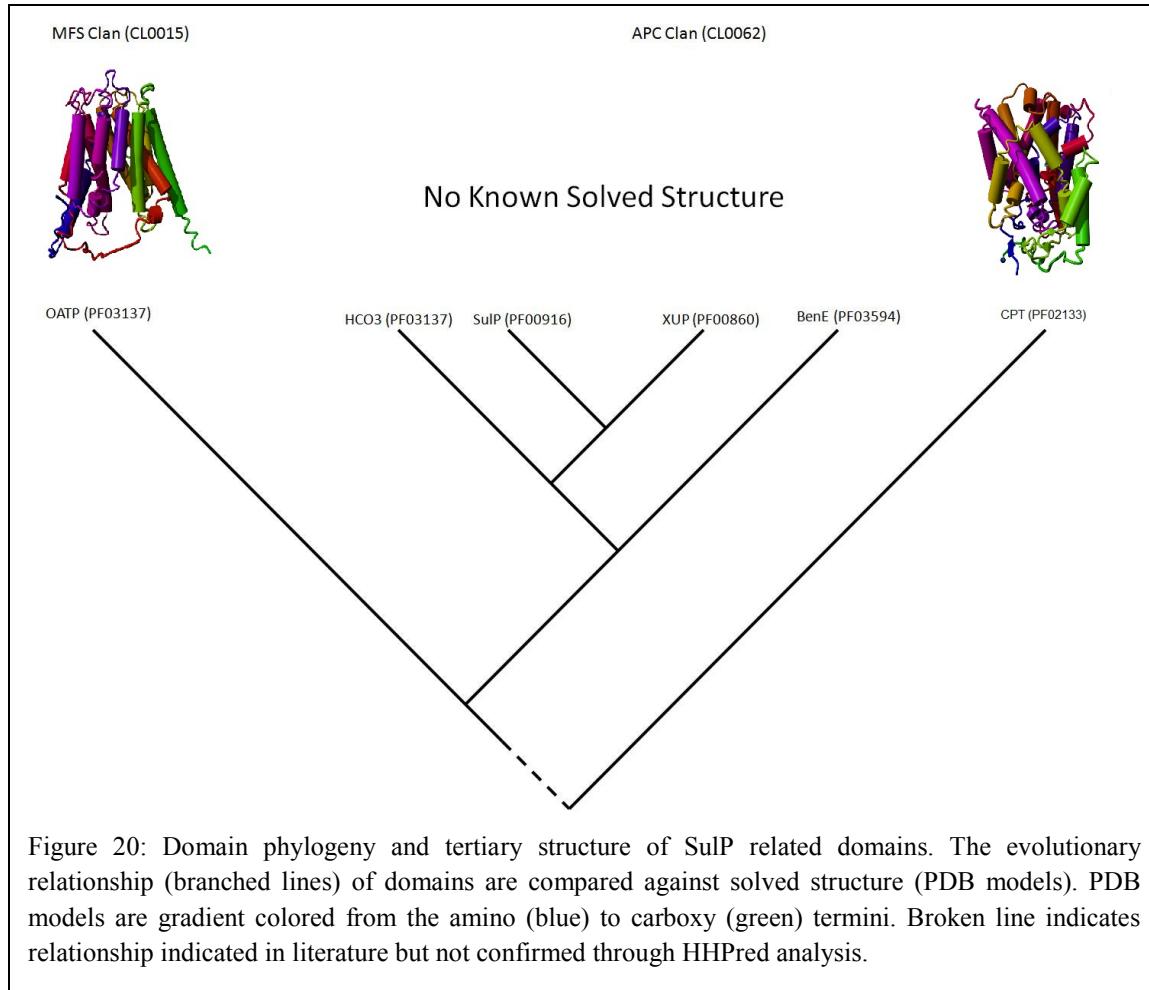


Figure 20: Domain phylogeny and tertiary structure of SulP related domains. The evolutionary relationship (branched lines) of domains are compared against solved structure (PDB models). PDB models are gradient colored from the amino (blue) to carboxy (green) termini. Broken line indicates relationship indicated in literature but not confirmed through HHpred analysis.

STAS Domain

ATP-Grasp Domain Clan Primary Structure Homology

As with the SulP family and the APC clan, the STAS domain was compared to other members of the ATP-Grasp clan (Fig. 21). The ATP-Grasp domain clan forms two

groups based on HHPred analysis. ATP-Grasp group 1 includes only the STAS domain family. ATP-Grasp clan group 2 includes the rest of the Pfam identified domain families and was termed the ATP-Grasp domain superfamily. None of the other identified STAS HHPred related Pfam domains had an e-value small enough (e-value $\leq 1.00 \times 10^{-3}$) to be considered as part of a STAS domain superfamily.

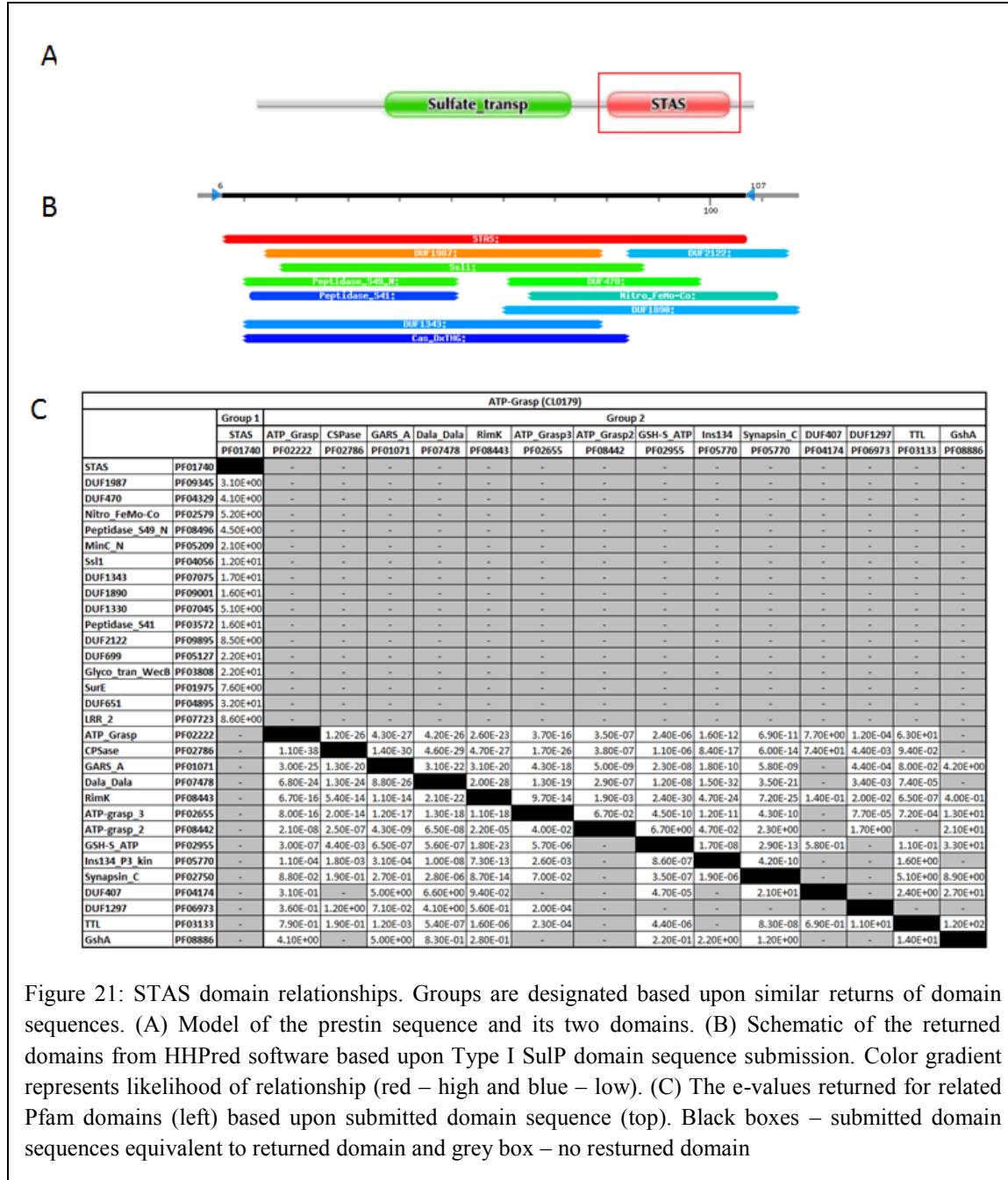


Figure 21: STAS domain relationships. Groups are designated based upon similar returns of domain sequences. (A) Model of the prestin sequence and its two domains. (B) Schematic of the returned domains from HHPred software based upon Type I SulP domain sequence submission. Color gradient represents likelihood of relationship (red – high and blue – low). (C) The e-values returned for related Pfam domains (left) based upon submitted domain sequence (top). Black boxes – submitted domain sequences equivalent to returned domain and grey box – no returned domain

STAS Secondary Structure

The STAS domain is a cytoplasmic domain, and therefore secondary structure was characterized in terms of α -helices and β -sheets. A Type II database for the STAS domain was created to examine the evolution of STAS secondary structure. In addition to the evolutionary sequences used in the SulP secondary structure analysis (see SulP domain section above), the prototypical SpoIIaa sequence (PDB 1AUZ) and the most closely related solved structure sequence (PDB 2KA5, see below) also added to the database. The sequences in the database were aligned using the ClustalW algorithm in STRAP. Sequences were then edited to more clearly define the STAS domain. The STAS amino terminal was defined by a conserved proline (P506 in human prestin) and the STAS carboxy terminal was defined by three residues past a conserved SIHDAV motif (S713 in human prestin). STAS sequence phylogeny was then analyzed using CLC Workbench (Fig. 22). Despite starting with the same sequences, the phylogeny of the STAS domain is different to the phylogeny of the whole sequence (see Fig. 17 above). The opossum (*M. domestica*) and platypus (*O. anatinus*) STAS domain sequences are more closely related to each other than to the next nearest branch point. This also holds true for human SLC26A6 and frog (*X. laevis*) Slc26a5 as well as the round worm (*C. elegans*) SulP3 and ameba (*D. discoideum*) STAS sequences. This suggests slightly different evolutionary pressures on the STAS domain than on the rest of the sequence.

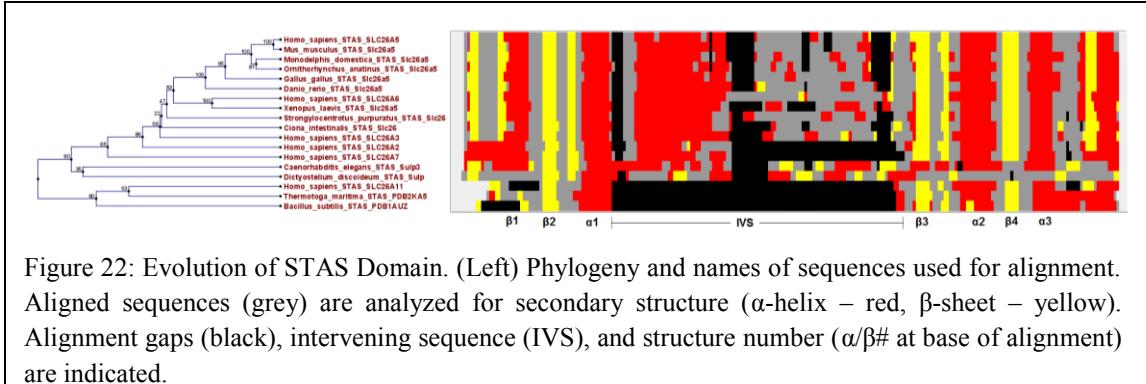


Figure 22: Evolution of STAS Domain. (Left) Phylogeny and names of sequences used for alignment. Aligned sequences (grey) are analyzed for secondary structure (α -helix – red, β -sheet – yellow). Alignment gaps (black), intervening sequence (IVS), and structure number ($\alpha/\beta\#$ at base of alignment) are indicated.

The STAS domain sequences were aligned in STRAP using the ClustalW algorithm. Secondary structure analysis was performed in STRAP using the Self-Optimized Prediction Method A (SOPMA). The α -helices and β -sheets were named according to the naming convention of 1AUZ, which has the following structure (in order from amino to carboxy end): β_1 , β_2 , α_1 , β_3 , α_2 , β_4 , α_3 . Other than the round worm (*C. elegans*) and amoeba (*D. discoideum*) sequences, β_2 through α_3 were clearly delineated and well aligned. The poor alignment of secondary structure in the round worm and amoeba may be due to distinct evolutionary paths or inaccurate cDNA for the selected representatives. STAS domains have several features that differ from the SpoIIaa sequences. Except for human SLC26A11, all the STAS domains are characterized by a large intervening sequence (IVS) between α_1 and β_3 . Also, amino terminal to α_1 , the SpoIIaa secondary structure differs dramatically from the STAS domain structure. Both SpoIIaa proteins (1AUZ, 2KA5) are characterized by at least two β -sheets and only one small α -helix in 2KA5. However, all of the selected STAS domain sequences have an additional β -sheet between β_2 and α_1 . Most of the STAS domains also have two β -sheets and a large α -helix where β_1 is located in the SpoIIaa proteins. This higher complexity in

the secondary structure of the STAS amino terminal may be related to the link between STAS and SulP domain in SulP protein family members.

STAS domain secondary structure was compared to ATP-Grasp domain superfamily secondary structure through the representative ATP-Grasp 1 domain family (PF02222, Fig. 23). Type I databases were collected for both domain families and aligned, using ClustalW algorithm, and the secondary structure was analyzed, using the SOPMA algorithm, in the STRAP software. No correlation was found between the STAS domain and the ATP-Grasp 1 domain family members in terms of secondary structure. The secondary structure in both families, however, was well conserved. This further suggests no evolutionary relationship between the STAS domain family and the rest of the ATP-Grasp domain clan members.

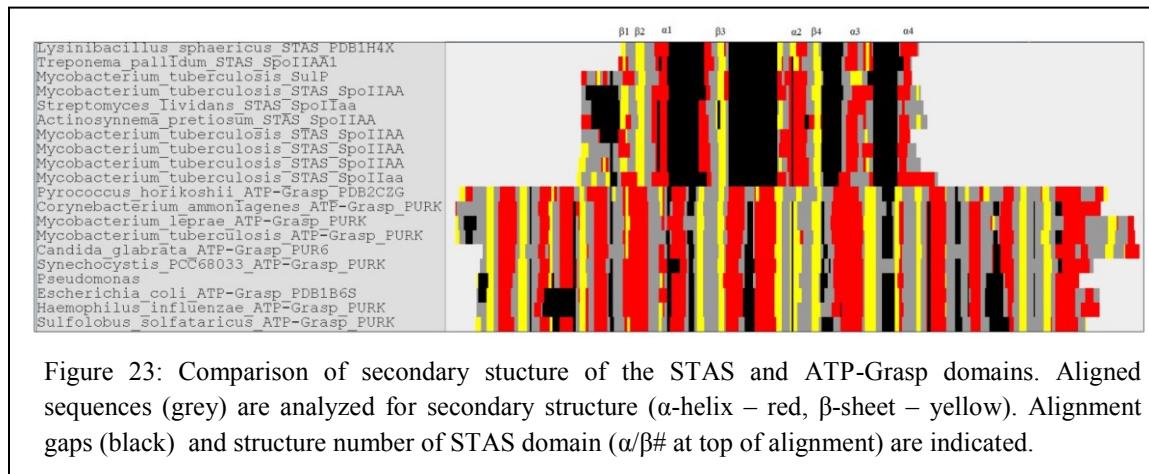


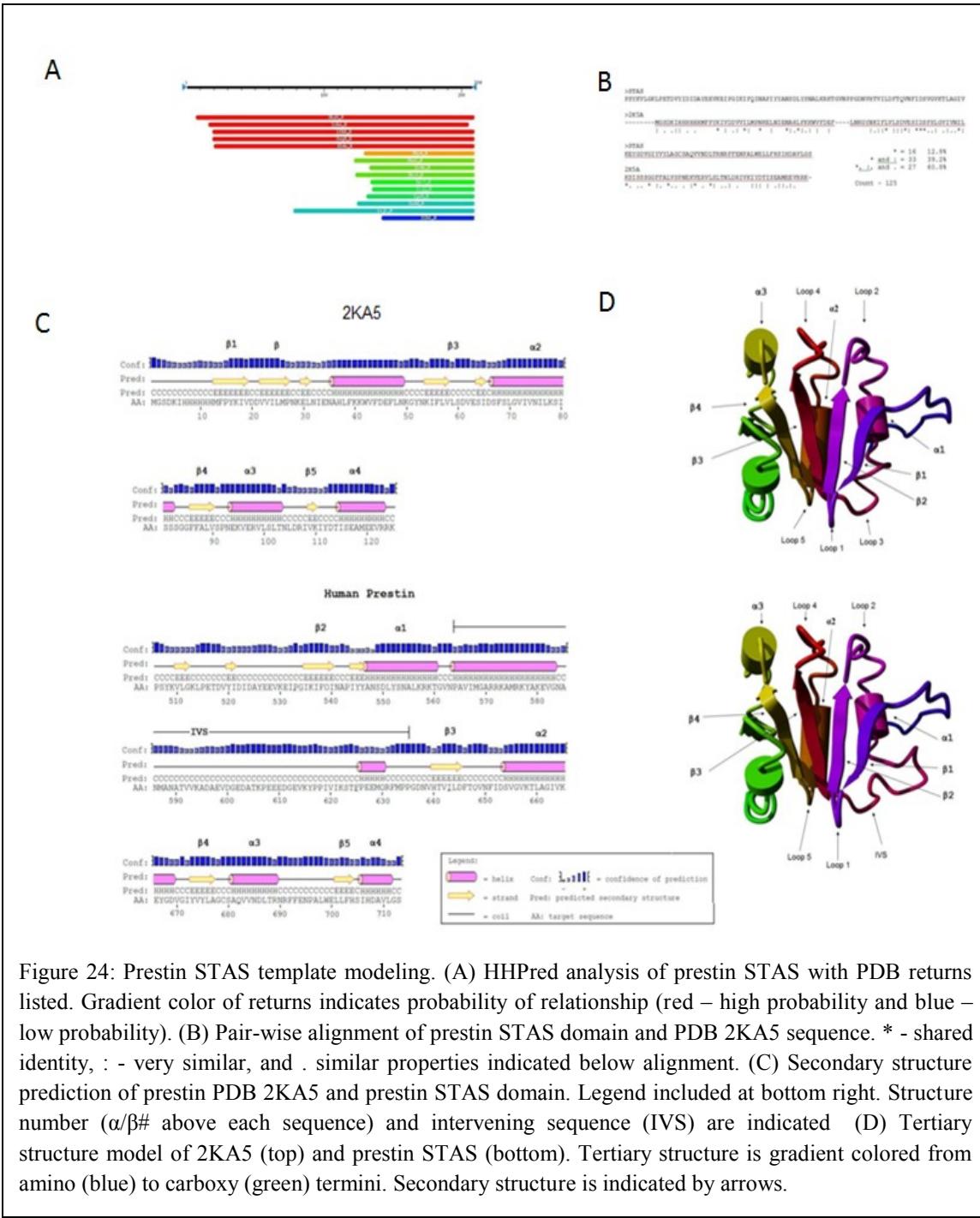
Figure 23: Comparison of secondary structure of the STAS and ATP-Grasp domains. Aligned sequences (grey) are analyzed for secondary structure (α -helix – red, β -sheet – yellow). Alignment gaps (black) and structure number of STAS domain ($\alpha/\beta\#$ at top of alignment) are indicated.

STAS Tertiary Structure

The human STAS domain sequence (as defined in the STAS Secondary Structure section above) was submitted to HHpred and searched across the most recent PDB database (Fig. 24). The PDB 2KA5 was the closest associated structure (e -value = 8.7e-21). In order to better model the STAS domain, the IVS was removed from the human

STAS domain sequence starting at residue A565, which followed a proline near the start of the IVS as identified in the previous section, and ending at residue P635, which is a proline near the end of the IVS as identified in the previous section. ClustalW alignment of the human STAS domain, with the IVS removed, and 2KA5 sequence shows only a 12.8% identity (represented as an * below alignment), which is below the twilight zone of sequence alignment. The percentage of identical and highly similar residues (:), however, was 39.2%. When the number of similar residues (.) is added, there is a 60.8% similarity between the IVS-less human STAS sequence and the 2KA5 sequence. This similarity was considered sufficient to perform homology modeling. PSIPred was performed to compare the secondary structure of the full human STAS sequence and the 2KA5 sequence. As in the previous section, α -helices and β -sheets are named in convention with the 1AUZ PDB. The β_2 , α_1 , β_3 , α_2 , β_4 and α_3 profile is seen in both sequences if the IVS is disregarded. In addition, both the human STAS and 2KA5 sequences have a β_5 and α_4 secondary structures not seen in SpoIIaa sequences. The amino β_1 , however, differs between the human STAS and 2KA5 sequences. Tertiary structure homology modeling was performed between the IVS-less human STAS domain sequence and the 2KA5 PDB structure using the Sequence Alignment method of Swiss-Model. Structural Assessment with Swiss-Model was performed on both the 2KA5 PDB structure and the predicted STAS domain structure. The original 2KA5 PDB had a G_{mean} of 0.586 (z-score = -1.256) and all residues were predicted to be in 5 Å of predicted position (see Appendix III for description of tertiary structure statistics). The 2KA5 modeled IVS-less human STAS domain had a Q_{mean} of 0.482 (z-score = -2.508) and, except for residue L1, all residues were predicted to be in 5 Å of predicted position. Most of the high energy

residue positions for the STAS homology model are seen in the carboxy and amino termini. The relatively high energy profile of the STAS model may also be due to the high energy seen in the initial 2KA5 model. Although this model of the IVS-less human STAS domain will be used as a possible arrangement of the primary and secondary structure, residue resolution should be considered ≤ 5 Å. Despite low resolution, the tertiary structure can be characterized as a β -sheet with one face bordered by α -helices (α_1 and α_2) and the other face exposed to the environment. The third α -helix is in plane with loop 2 and loop 4.



The 2KA5 modeled IVS-less human STAS domain was compared to other solved structures from representative ATP-Grasp clan members including a Synapsin ATP Binding Domain (Synapsin C, PF02750), Inositol 1, 3, 4-trisphosphate 5/6-kinase

(Ins124, PF05770), and ATP-Grasp 1 (PF02222) family members (Fig. 25). ATP-Grasp clan members tend to be much larger than the STAS domain. Also, although there is little similarity between primary and secondary structure, a similar tertiary structure is seen in all models (circled in red). Similar to the STAS domain model, a central β -sheet is bordered by α -helices on one face while the other face is exposed to the environment. In many ATP-Grasp clan members, nucleotide triphosphates are catalyzed on the exposed face of this β -sheet. This structure is often repeated in the same domain of the ATP-Grasp clan family members. It is possible that this represents convergent evolution between the STAS domain and ATP-Grasp clan members.

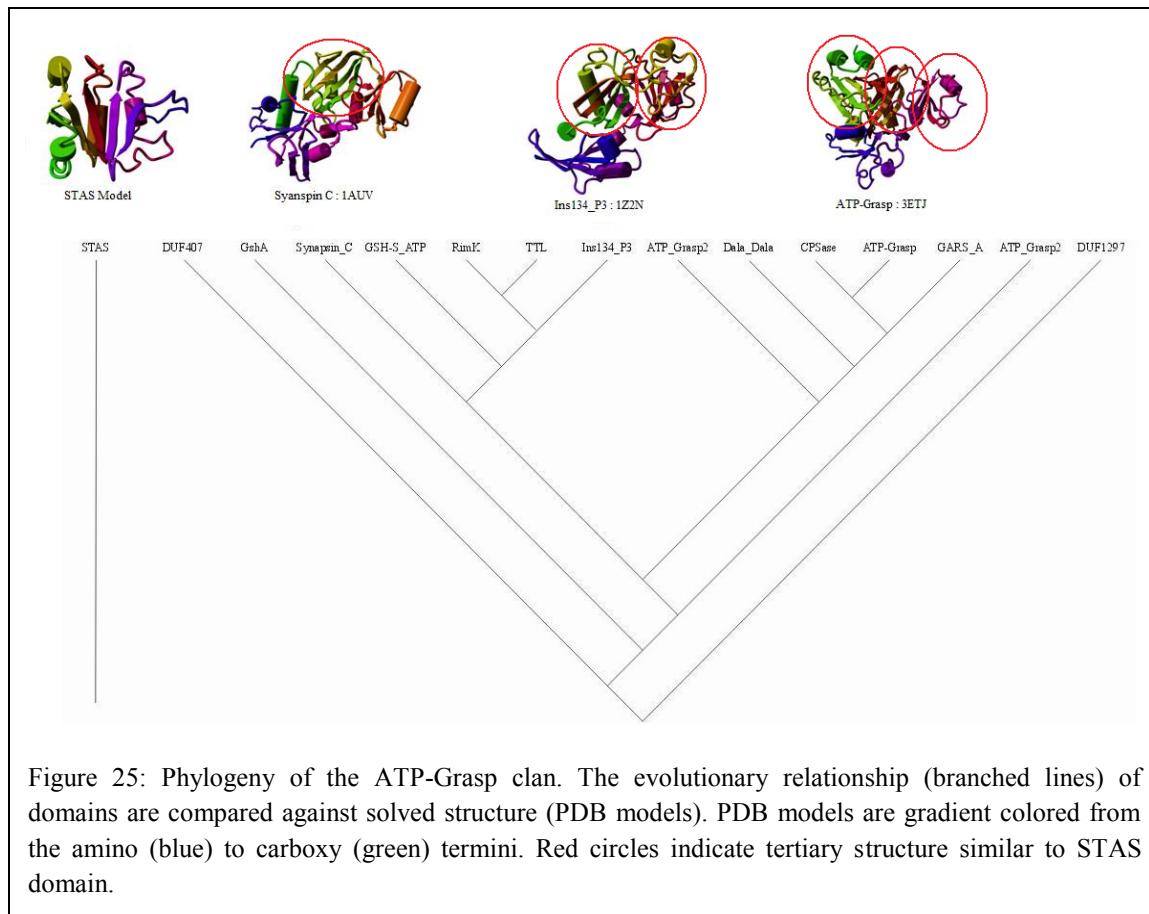


Figure 25: Phylogeny of the ATP-Grasp clan. The evolutionary relationship (branched lines) of domains are compared against solved structure (PDB models). PDB models are gradient colored from the amino (blue) to carboxy (green) termini. Red circles indicate tertiary structure similar to STAS domain.

B. Motifs

Known Motifs

Many common motifs are stored in accessible databases. The human prestin sequence (NP_945350.1) was used to query some of these databases to search for known motifs. A ScanProsite query, selecting for motifs with a high probability of occurrence, revealed several high-probability motifs (Fig. 26). Three N-glycosylation and two cAMP/cGMP-dependent phosphorylation sites have been explored in previous studies (see Chapter II). An additional 15 possible phosphorylation sites remain unexplored. Several N-myristoylation sites are also predicted, but most of these sites are unlikely to be relevant since there is no evidence of post-translational cleavage of prestin. ScanProsite also identified the SulP motif found in previous studies. InterPro also identifies the SulP motif, but InterPro's record is based on the ScanProsite database. None of these motifs are suggested to play a role in oligomerization.

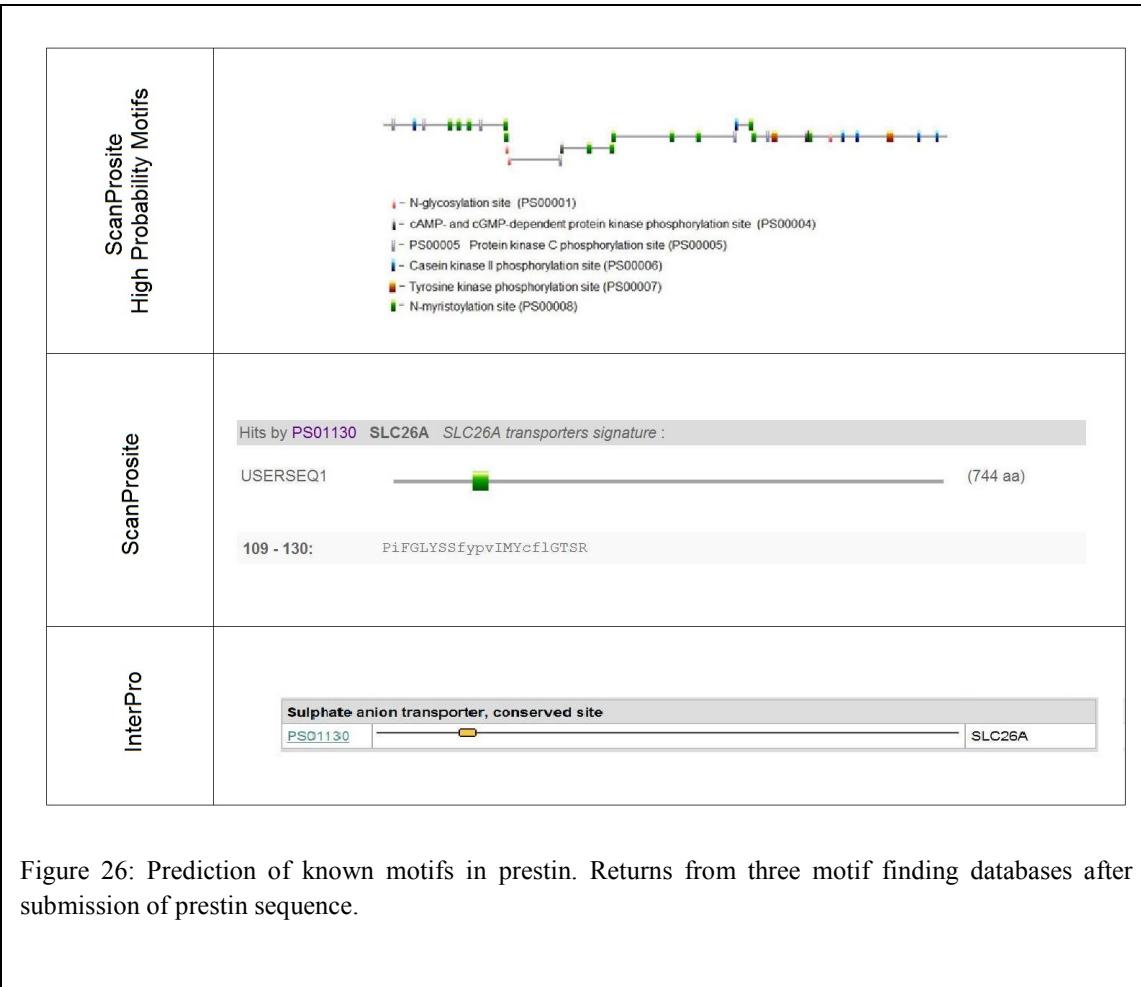
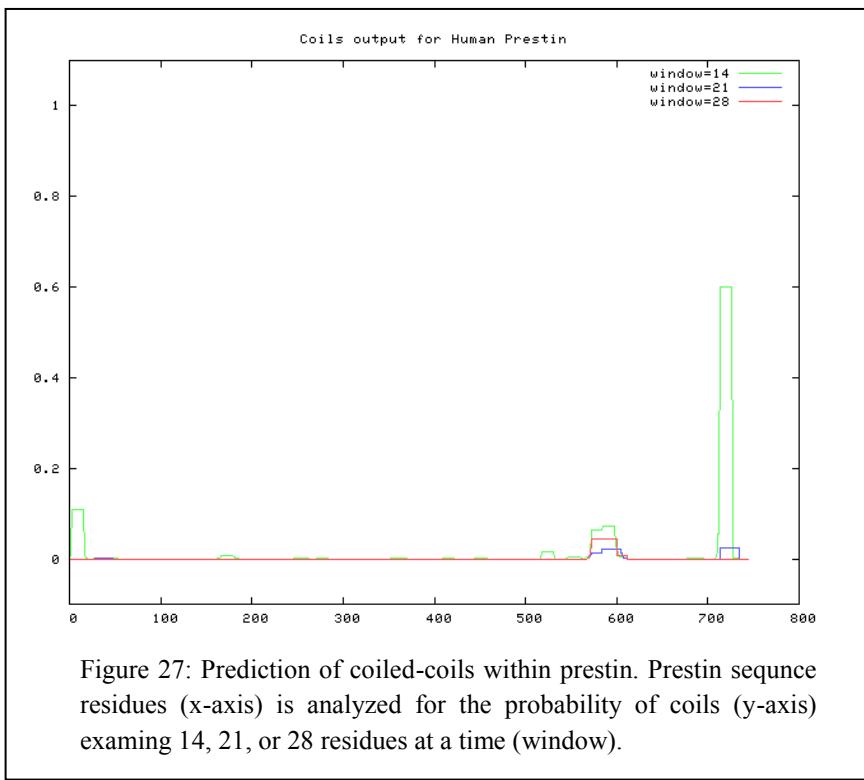


Figure 26: Prediction of known motifs in prestin. Returns from three motif finding databases after submission of prestin sequence.

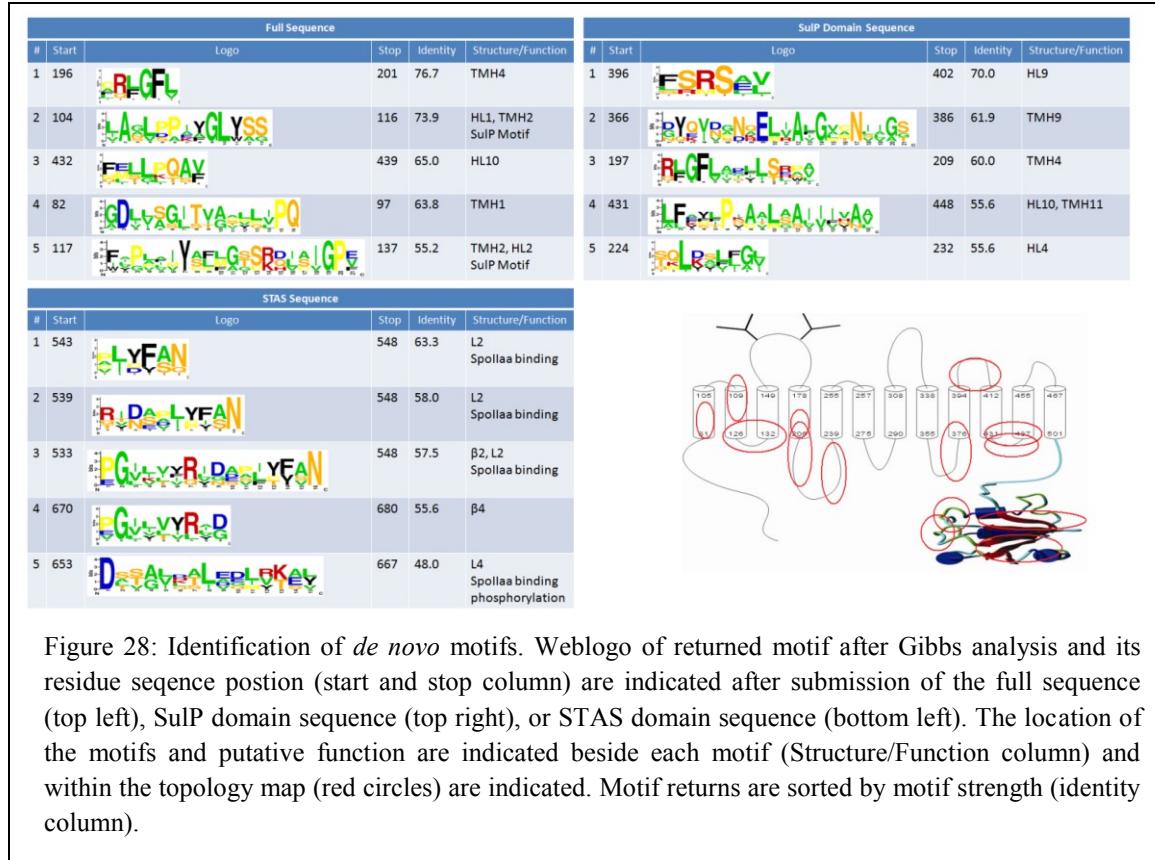
The human prestin sequence was also submitted to coiled-coil and leucine zipper prediction software (Fig. 27). The COILS software was run with the MTIDK matrix with no weighting of positions. The output suggested a 60% probability of a coiled-coil at the carboxy terminal (residue 713-726). This possible coiled-coil appears only in a 14 residue window in a region of the carboxy terminal that is not necessary for normal function (see Chapter II). The 2ZIP software also did not identify either a leucine zipper or coiled-coil (data not shown). Therefore any prestin homo-oligomerization does not depend on the well established coiled-coil or leucine zipper oligomerization motifs.



De Novo Motifs

Since no known oligomerization motif was found in previous sections, prestin was assessed for unique motifs. As suggested by the results in Chapter III, homo-oligomerization is conserved throughout the SulP family. Thus the entire SulP family was used to assay for *de novo* conserved motifs by creating a Type III database with the SulP family member representatives from each kingdom. This database was further divided into three databases based on the full length, SulP domain, and STAS domain sequences for SulP family. The SulP domains for each member of the Type III database were delineated by domain designation in the CDD database. STAS domains for each member of the Type III database were delineated as described in the STAS Secondary Structure section above. The full length, SulP domain, and STAS domain sequence databases were submitted to Gibbs Motif Sampler using recursive sampling with various motif widths (4,

8, 12, 16, and 20). Identity was analyzed for each motif by averaging the identity percentage for each residue position. Based on motif strength, the top five results for each database were compared to topology and tertiary structure model based on previous sections (Fig. 28). Gibbs Motif Sampler identified several known as well as *de novo* motifs, which are identified in the topology structure as red circles.



Gibbs Motif Sampler analysis of the full human prestin sequence identified the SulP motif along with three *de novo* motifs. The SulP motif was identified in return 2 and 5, although these two returns did not overlap. Despite being a known motif in the SulP family, the SulP motif has no known function. The *de novo* TMH 1 and 4 motifs, found in return 1 and 4, respectively, correspond to the region of poorly delineated TMHs. The

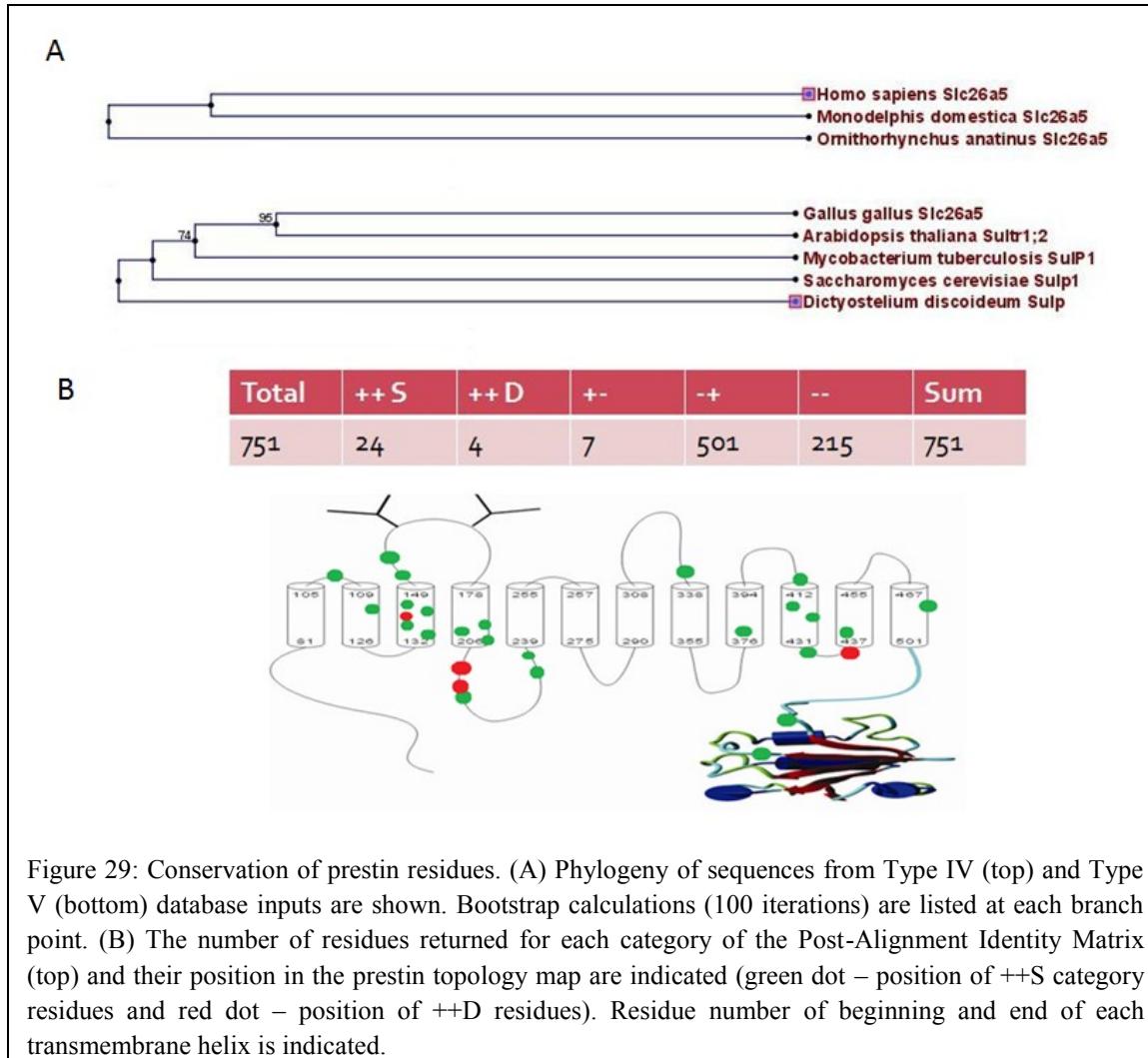
de novo motif in the helical loop 10 (HL10) is near the carboxy end of the SulP domain and is not associated with any function.

In the Gibbs analysis of the SulP domain, the *de novo* motifs had similar identity strength to both the STAS and full length sequence results, suggesting equivalent validity between the three sets of returns. Two of the returns, return 3 and 4, for the SulP domain overlapped two returns from the full length analysis, returns 1 and 3, respectively. This again suggests that TMH 4 and 10 may be important for function. A motif was also identified in the HL 4 near TMH 4 by return 5 in the SulP domain Gibbs analysis. The top two returns for the Gibbs analysis of the SulP domain corresponded to regions in and around TMH 9. There is no known function that corresponds to any of these *de novo* motifs.

The Gibbs analysis of the STAS domain reinforced some of the putative functional structures of the SpoIIaa associated domain. The first three returns identified a portion of the STAS loop 2 (L2), which is part of the protein-protein binding area of SpoIIaa and SpoIIab. Return 3 also included the β 2 helix that borders on one face of the central β -sheet, which, based on the tertiary model in the previous section, is thought to participate in ATP catalysis. The fourth return also identified part of the central β -sheet: β 4. The β 2 and β 4 motifs have very similar profiles suggesting they share similar function. The fifth return includes part of the STAS domain loop 4 (L4). L4 includes a DSSG motif, which is phosphorylated in SpoIIaa proteins. Phosphorylation of this DSSG motif inhibits sporulation in bacterial and fungal species.

C. Primary Structure

Residue conservation was also analyzed using a novel Post-Alignment Identity Matrix (PAIM) algorithm. PAIM was used to identify residues conserved in all homologous sequences as well as residues that are conserved specifically in prestin homologous sequences that function as motor proteins or transporter proteins. Type IV and Type V databases were curated and inputted into the PAIM algorithm (Fig. 29). Results were tabulated in terms of residues with conserved identity in and between both databases (++S); residues with conserved identity in both databases, but different residue between databases (++D); residues with conserved identity in non-motor prestin homologous sequences, but non-conserved identity in motor prestin homologous sequences (+-); residues with non-conserved identity in non-motor prestin homologous sequences, but conserved identity in motor prestin homologous sequences (-+), and residues with non-conserved identity in either database (--). Of human prestin's 744 residues, only 24 residue identities were conserved throughout all sequences (++S, shown as green circles). Four residue positions had conserved identity in both databases, but were different between databases (++D, shown in red). It is uncertain whether any of these residues are essential for function (for a list of these residues see Appendix III). The clustering of these residues, however, may indicate regions that are important for the structure-function relationship. These conserved residues cluster around TMHs 2-4 and TMHs 9-11.



III. Conclusion

This chapter used bioinformatic techniques to determine conserved structures. Two domains were identified in prestin: the SulP and STAS domains. The SulP domain shared distant homology with three other domains: the BenE, HCO3, and XUP domains. All of the related domains shared secondary structure, but no solved structure is available for proteins within and of these domain families. The STAS domain was not found to be related to any other domains within its clan members. However, solved structures exist within its own domain family. Secondary structure was also conserved and a template

model was made based upon other STAS domain family members. Prestin did not contain any known oligomerization motifs, but it did contain several conserved motifs and residues. These conserved motifs and residues suggest possibly two pore regions: TMH 1-4 and TMH 8-11 in the transmembrane region. The STAS domain contains a DSSG motif that may regulate protein-protein interaction through phosphorylation.

CHAPTER V: Analysis of Oligomerization Motif Disruption

Disruption of homo-oligomerization sites could help determine its functional importance. Prestin homo-oligomerization may occur through disulfide bonds and/or a protein-protein interaction site in the STAS domain. As shown in Chapter III, oligomerization is conserved throughout prestin evolution and therefore it is also expected that oligomerization sites are also conserved. The disulfide bond conservation, which occurs via cysteine residues, was examined using multiple sequence alignment (MSA) of major prestin evolutionary branches. The role disulfide bonds play in prestin homo-oligomerization was further examined by performing a cysteine scanning mutagenesis in gerbil prestin. Homo-oligomerization was examined through FRET analysis and any associated functional change was measured with NLC.

One putative homo-oligomerization site was identified: the DSSG motif of the STAS domain. In SpoIIaa proteins, phosphorylation of serine in the DSSG motif disrupts protein-protein interaction. The conservation of this motif was also examined throughout the major branches in prestin's evolution using MSA. The structural and functional importance of this putative phosphorylation site was also examined by mutation of the serine to a threonine, which can be phosphorylated, or to an alanine, which cannot be phosphorylated.

I. Materials and Methods

A. *Multiple Sequence Alignment (MSA)*

Databases for sequences were curated (see Chapter IV) and submitted to CLC Workbench. The “Create Alignment” algorithm was performed using “Gap open cost” of

10, a “Gap extension cost” of 1, and an “End gap cost” under the “Very accurate” settings [80]. Phylogeny was subsequently run through the same software using the “Create tree” algorithm based on UPGMA and 100 bootstrap iterations.

B. Plasmid Constructs

Plasmid constructs developed in the previous chapter (see Chapter III) were used here. Disruption of residues and motif were performed using the QuickChange® II XL Site Directed Mutagenesis kit according to manufacturer’s recommendations (Stratagene). Mutagenesis primer design was assisted by web-based PrimerX software (www.bioinformatics.org, primers used are listed in Appendix I).

C. Cell Transfection

Methods similar to those described in the previous chapter were used for FRET experiments (see Chapter III). For nonlinear capacitance studies, cells were grown to between 15-30% confluence and Lipofectamine™ 2000 volume was adjusted to maximize transfection rate and minimize cell loss.

D. Non-Linear Capacitance (NLC)

Approximately 24 to 48 h after transfection, cell medium was replaced with an extracellular channel blocking solution containing NaCl (120 mM), MgCl (2.0 mM), TEA (20.0 mM), CoCl (2.0 mM), dextrose (10.0 mM), and HEPES (10.0 mM), buffered to pH 7.25 and adjusted to 300 mOsm. Cells were imaged at room temperature using a 100x 1.4 numerical aperture objective with mercury excitation filtered with a FITC

spectrum dichroic on an Olympus IX-70 inverted microscope. Only cells with clearly membrane-targeted fluorophores were selected for analysis.

Electrophysiological measurements were performed using a Warner Instruments PC-501A patch clamp amplifier (Hamden, CT). Voltage commands, driven by custom software (TestPoint, C. E. C. Corp., Burlington, MA), were low passed filtered (at 5 kHz) prior to digitization (A/D-D/A KCPI 3801 Board, Keithley Instruments, Cleveland, OH). Cells were whole-cell patch clamped using pulled patch pipettes (Sutter P-97 puller, Novato, CA; glass capillaries 8250, A-M Systems, Carlsborg, WA) polished to approximately 3 μ m (Narashige MF-830 polisher, East Meadow, NY). The patch pipettes were back filled with an intracellular channel blocking solution consisting of CsCl (140 mM), EGTA (10.0 mM), HEPES (10.0 mM), MgCl (2.0 mM), and K₂ATP (2 mM). Only patch pipettes with an initial resistance between 1.5 and 5.0 M Ω were used. After establishing whole-cell patch clamp, only cells with an input resistance of at least 500 M Ω were considered for further analysis. Cells were maintained at a -70 mV holding potential.

NLC was measured using the two-sinusoid method [84]. Two sinusoidal voltage commands (195.3 Hz and 390.6 Hz), with amplitudes of 10 mV, were superimposed on voltage step command ranging from -100 mV to 60 mV (5 mV intervals). Membrane potentials were corrected off-line for series resistance. Corrected capacitance-membrane potentials were fitted by the Levenberg-Marquardt algorithm in Origin Software (OriginLab, Northampton, MA) to the Boltzmann equation:

$$C_{NL} = C_{Lin} + \left(\frac{Q_{Tot}z}{kT} \right) \left(\frac{b}{(1+b)^2} \right),$$

$$b = \exp\left(\frac{-ze}{kT}(V - V_{Pk})\right)$$

where C_{NL} is the NLC, C_{Lin} is the linear capacitance, V is the membrane potential, V_{Pk} is the membrane potential at peak NLC, e is electron charge, z is the valence of charge transferred, Q_{Tot} is the total number of elementary charges transferred, k is Boltzmann's constant, and T is absolute temperature. Only cells with a fit of $R^2 > 0.9$ were used to determine the three Boltzmann curve variable parameters (z , Q_{Tot} , and V_{Pk}). The Q_{Tot} parameter was normalized for variance in cell size by dividing Q_{Tot} by C_{Lin} . NLC for both gerbil prestin mutants and gerbil prestin wild-type plasmid construct transfected HEK cells were obtained on the same day to compensate for day-to-day system variances. The number of transfected cells with an $R^2 > 0.9$ fit was compared to the total number of transfected cells patched to determine prestin mutant loss-of-function.

E. Fluorescent Resonance Energy Transfer (FRET)

FRET efficiency was collected and calculated as described in Chapter III.

II. Results

A. Cysteine Mutations

Multiple Sequence Alignment (MSA)

Cysteine is the only residue that can form disulfide bonds. It is also a tiny, polar residue which shares similar properties with serine (tiny, polar), threonine (small, polar), glycine (tiny, nonpolar), and alanine (tiny, nonpolar; see Appendix III for residue

properties). The conservation of cystyl residues was examined along evolutionary lines using a Type II database of prestin homologous sequences curated along major evolutionary branches (Fig. 30, see Chapter IV for further details). Cystyl residue conservation can then be analyzed by identity (same residue) or similarity (residue with similar property). Only the C52 residue position has conserved identity throughout evolution. Positional gaps were present in a few sequences, but this is most likely due to poor overall conservation in the amino terminal rather than a change of properties in the residue position. The cystyl at C124, which is part of the SulP motif, has no conserved identity, even in mammalian sequences. However, at this residue position, small and polar residues are conserved throughout the Slc26a5 orthologs. In the C192 and C196 residue positions, neither identity nor similarity is conserved. The C260 residue position is occupied by small residues in mammalian sequences, but not throughout the Slc26a5 orthologs. Both C381 and C395 residue positions are characterized by small and/or polar residues throughout prestin evolution, except for a single valyl residue in the *C. elegans* SulP sequence for C381. The C415 residue position is unique. It is the only residue position in which the identity is conserved in all of the motor protein homologous sequences but none of the non-motor protein homologous sequences. The C679 residue position has conserved similarity (polar, small) only throughout the sulfate exchanger homologous sequences. The conservation of similarity may extend to more distantly related homology, but this is obscured due to a possible misalignment with the *C. intestinalis* sequence as a result of poor alignment in the carboxy terminal due to the IVS. Of all the cystyl residues, C52 is the only cystyl residue conserved throughout prestin evolution and therefore the only residue likely to participate in disulfide bonding in all

SulP proteins. C415 is the only cystyl residue that is unique to prestin motor protein homologous sequence and therefore may have particular importance for motor function.

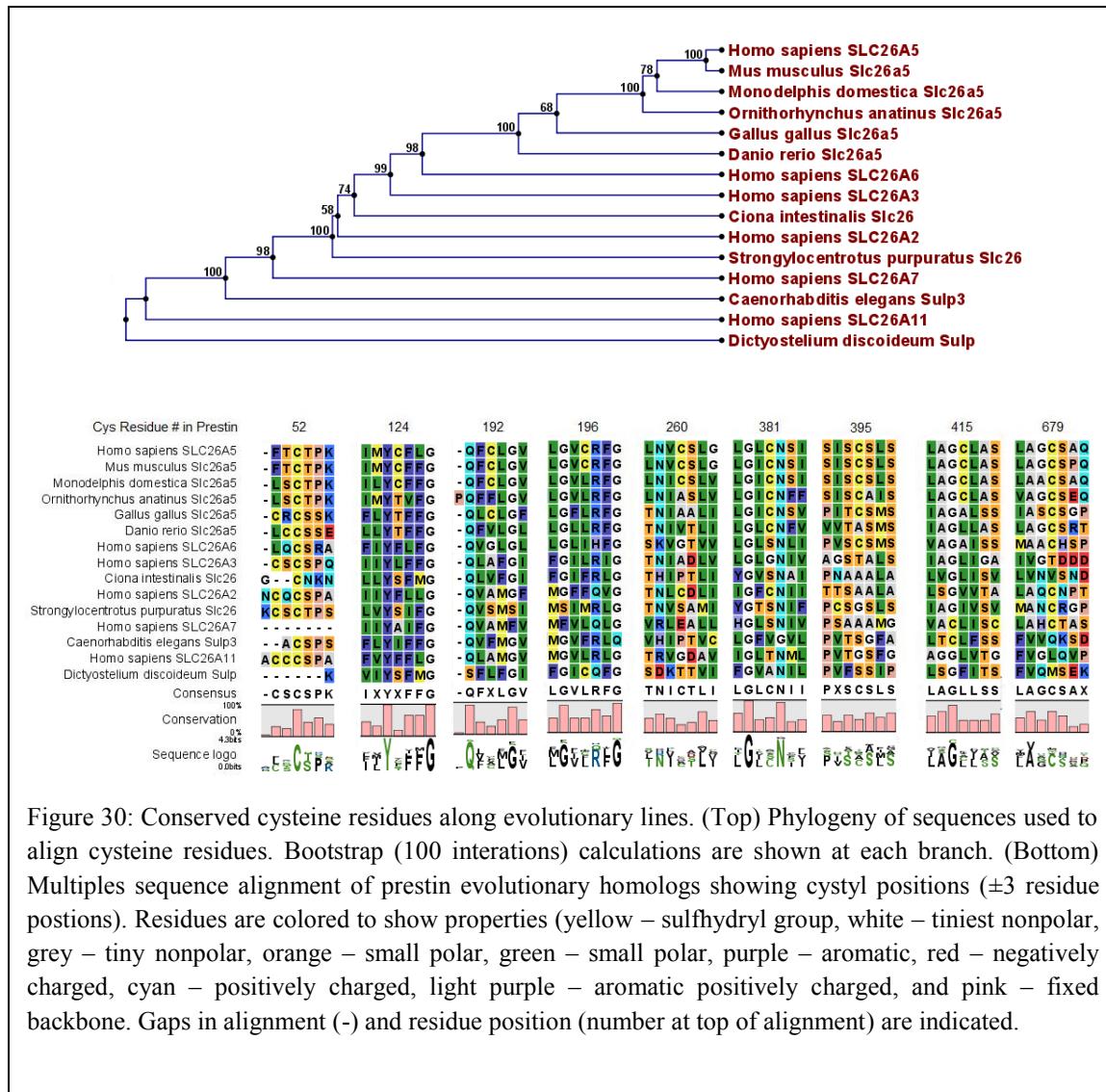


Figure 30: Conserved cysteine residues along evolutionary lines. (Top) Phylogeny of sequences used to align cysteine residues. Bootstrap (100 iterations) calculations are shown at each branch. (Bottom) Multiple sequence alignment of prestin evolutionary homologs showing cystyl positions (± 3 residue positions). Residues are colored to show properties (yellow – sulphydryl group, white – tiniest nonpolar, grey – tiny nonpolar, orange – small polar, green – small polar, purple – aromatic, red – negatively charged, cyan – positively charged, light purple – aromatic positively charged, and pink – fixed backbone). Gaps in alignment (-) and residue position (number at top of alignment) are indicated.

Nonlinear Capacitance (NLC)

Each of the nine cystyl residues were individually mutated to an alanyl residue to remove the sulfhydryl group and disrupt any disulfide bond, as well as polar bonding, at that residue position. The three NLC parameters (V_{PK} , Q_{Tot} , and z) of the cystyl residue mutations were examined (Fig. 31). The well conserved amino cysteine (C52) did not show significant change in any of three NLC parameters. Two cysteine-to-alanine mutations, however, did have significant changes in V_{PK} : C124A and C415A. Cells transfected with the pgPG C124A plasmid construct had significantly hyperpolarized V_{PK} compared to cells transfected with pgPG wild-type (-72.8 and -49.9 mV, respectively; Student's t-test $p \leq 0.001$). This suggests that more prestin C124A proteins were in the shortened state at resting membrane potential compared to prestin wild-type proteins. This effect may reflect the C124 residue position located in the putative pore region of the SulP motif in TMH 2 (see chapter IV). In Slc26a5 orthologs, the C124 residue position is conserved as a polar residue suggesting an importance for polarity at this residue position. Cells transfected with the pgPG C415A plasmid construct had significantly depolarized V_{PK} compared to pgPG wild-type (-32.1 and -53.7 mV, respectively; Student's t-test $p \leq 0.001$). In contrast to C124A mutations, this suggests more prestin C415A proteins were in the elongated state at resting membrane potential compared to prestin wild-type proteins. This effect may reflect the C415 residue position in a putative second putative pore forming region in TMH 8-11. It is also the only cystyl residue position with conserved identity exclusively in mammalian prestin orthologs.

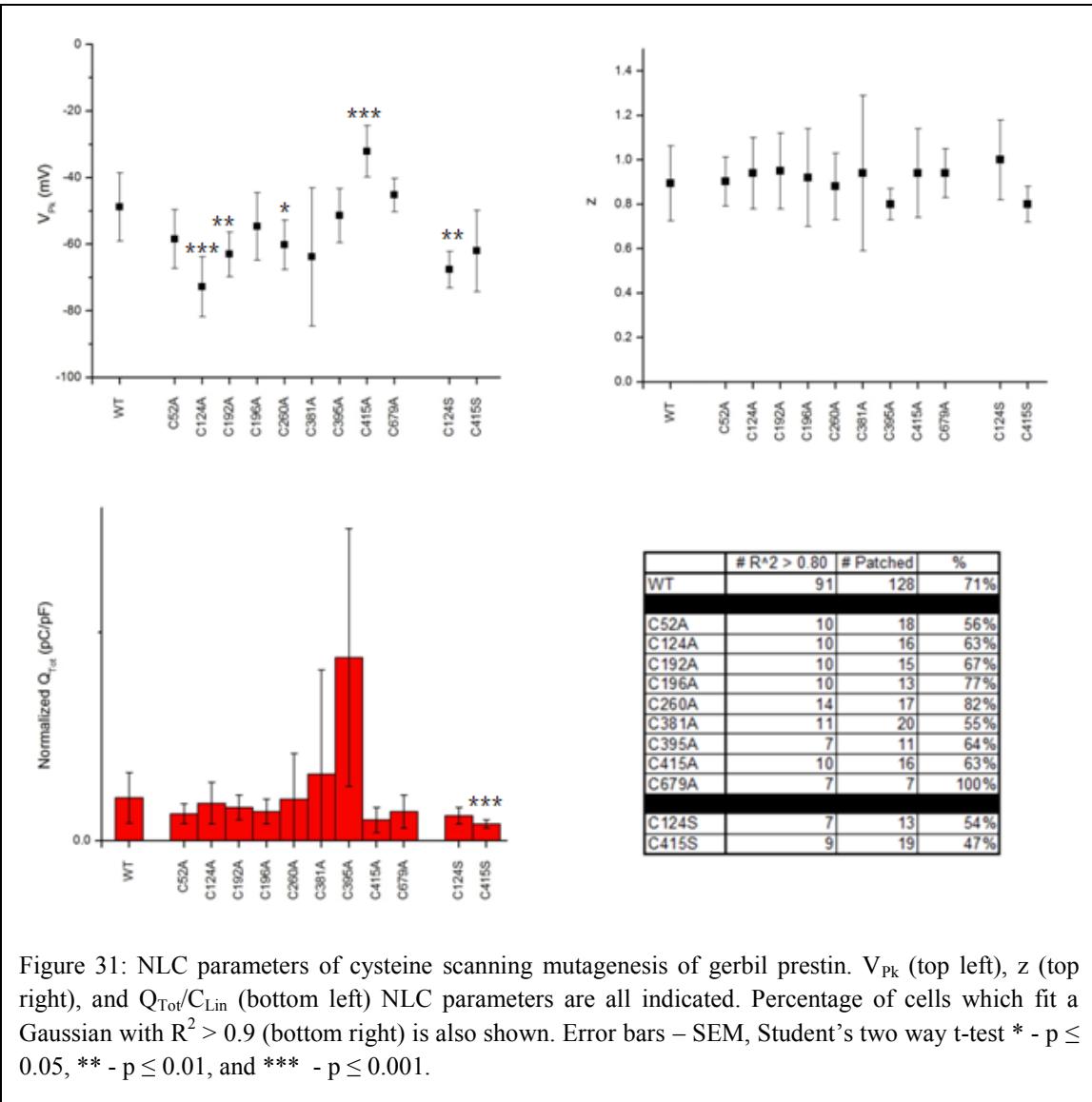


Figure 31: NLC parameters of cysteine scanning mutagenesis of gerbil prestin. V_{pk} (top left), z (top right), and $Q_{Tot}/CLin$ (bottom left) NLC parameters are all indicated. Percentage of cells which fit a Gaussian with $R^2 > 0.9$ (bottom right) is also shown. Error bars – SEM, Student's two way t-test * - $p \leq 0.05$, ** - $p \leq 0.01$, and *** - $p \leq 0.001$.

C124 and C415 were substitution-mutated to serine to examine the importance of polarity for these residue positions. Cells transfected with pgPG C124S plasmid constructs NLC parameters were not significantly different than the NLC parameters of cells transfected with pgPG C124A plasmid constructs (-67.7 and -72.8 mV, respectively; Student's t-test $p > 0.05$). This suggests that polarity at this residue position is not the only property that affects normal prestin function. Cells transfected with pgPG C415S plasmid constructs, however, did have significantly hyperpolarized V_{pk} when compared to cells transfected with pgPG C415A (-59.3 and -32.1 mV, respectively; Student's t-test $p \leq 0.001$). Moreover, NLC parameters of cells transfected with pgPG C415S were indistinguishable from cells transfected with pgPG wild-type. However, the Q_{tot}/C_{lin} of cells transfected with pgPG C415S was reduced in comparison to cells transfected with pgPG wild-type (3.07 and 8.11 fC/pF, respectively; Student's t-test $p \leq 0.001$). This may correlate to a loss in membrane insertion of C415S gerbil prestin proteins compared to wild-type gerbil prestin proteins. This suggests that the small polar properties of C415 residue are important for the normal motor protein structure-function of mammalian prestin orthologs.

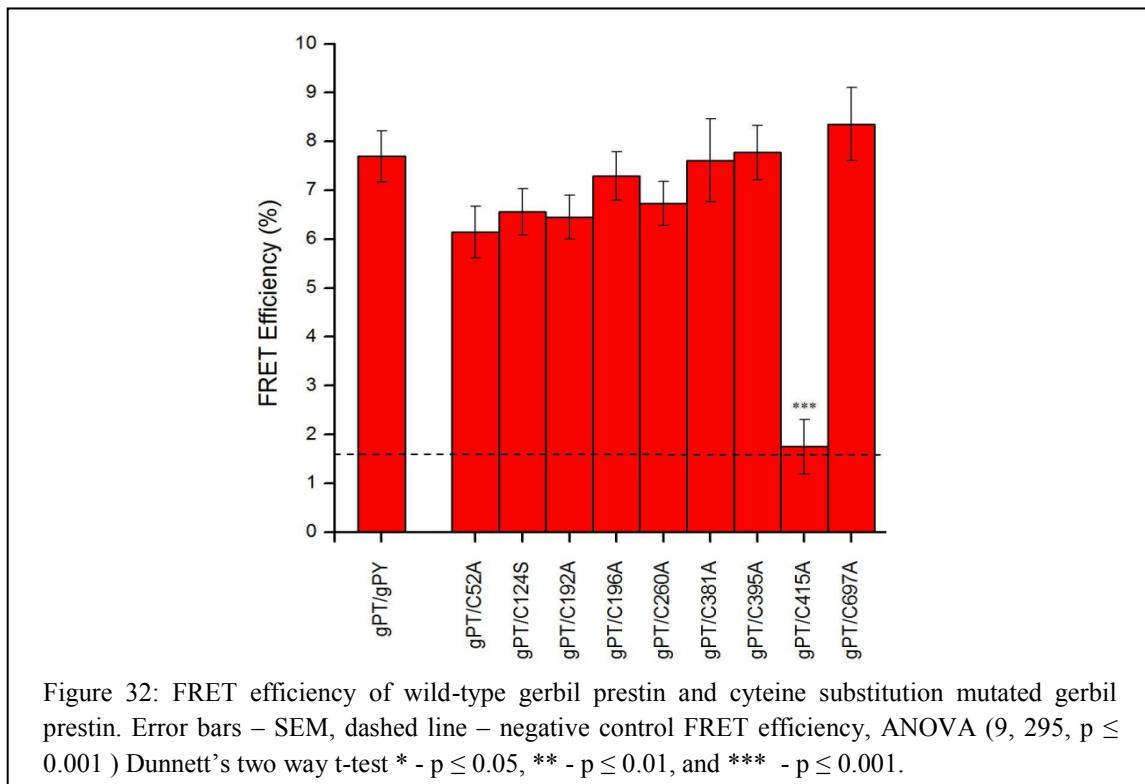
Significant, although smaller, changes were also seen with cysteine-to-alanine mutations at residues 192 and 260. Cells transfected with pgPG C192A or C260A were hyperpolarized in comparison to cells transfected with pgPG wild-type (for pgPG C192A versus wild-type: -63.0 and -52.2 mV, respectively with a Student's t-test of $p \leq 0.01$; for pgPG C260A versus wild-type: -60.2 and -51.0 mV, respectively with a Student's t-test of $p \leq 0.05$). Although significant, these small changes in V_{pk} may reflect small changes in prestin conformation than loss of disulfide bonds.

Since only cells with appreciable NLC signals were used in the calculations, the total number of cells that were successfully whole-cell patch clamped ($R_S > 500 \text{ M}\Omega$) was compared to those that were successfully clamped with a NLC that had a Gaussian fit with $R^2 > 0.08$. This allowed us to determine if a transfected cell population had diminished NLC signal. Only one transfected cell population had a lower than 50% cells with appreciable NLC (C415S). This may be due to a possible loss in membrane targeting (data not shown).

Fluorescent Resonance Energy Transfer (FRET)

If human prestin cysteines participate in disulfide bond formation, then substitution mutation of cysteines should reduce the binding efficiency between prestin subunits. This was examined by measuring FRET efficiencies in cells co-transfected with wild-type gerbil prestin donor plasmid constructs (pgPT wild-type) and cysteine substitution mutated gerbil prestin acceptor plasmid constructs (pgPY C###A/S). To avoid loss of oligomerization due to possible misfolding properties, mutations to C124 and C415 were carefully considered. pgPG C124S was used for FRET analysis because of smaller hyperpolarization of V_{PK} in transfected cells relative to pgPG C124S transfected cells. Due to possible loss of membrane insertion of C415S prestin proteins, the pgPY C415A plasmid construct was used despite depolarization of V_{PK} . FRET efficiencies were calculated for each donor/acceptor co-transfected cells (Fig. 32). Only cells co-transfected with pgPT wild-type and pgPY C415A had a significant change in E_{FRET} compared to cells co-transfected with pgPT wild-type and pgPY wild-type (2.11% and 7.64%, respectively, Dunnett's two way t-test $p \leq 0.001$). Moreover, the E_{FRET} of

cells co-transfected with pgPT wild-type and pgPY C415A were indistinguishable from the cells co-transfected with negative control of pgPT wild-type and p38Y (showed as a dotted line in figure 31; 2.11% and 1.80%, respectively, Dunnett's one way t-tests $p > 0.05$). Loss of E_{FRET} indicates a significant loss in oligomerization between C415A prestin subunits and the wild-type prestin subunits.



B. DSSG Motif

Multiple Sequence Alignment (MSA)

As with the cysteine mutant MSA, a Type II database was curated which included four mammalian prestin orthologs and several homologous sequences based on evolutionary branches (Fig. 33). In addition, two *SporIlaa* homologous sequences with solved tertiary structure (PDB 2KA5 and PDB 1AUZ) were curated into the database. To

avoid misalignment, only the STAS domain sequence was aligned (see Chapter IV for STAS domain delineation). Alignment and phylogeny were performed as described in the Cysteine Mutations section above. In the aligned DSSG motif, either the first or second residue position, after a universally conserved aspartic acid, is occupied by a phosphorylatable serine or threonine. Only three species sequences do not share this motif: *C. intestinalis*, *S. purpuratus*, and *C. elegans*. It is possible that these three sequences may be misaligned, erroneous, or that the DSSG motif in these sequence serve a purpose other than phosphorylation. A serine follows the aspartic acid located at residue S654 in the human and gerbil sequences (gerbil sequence not shown).

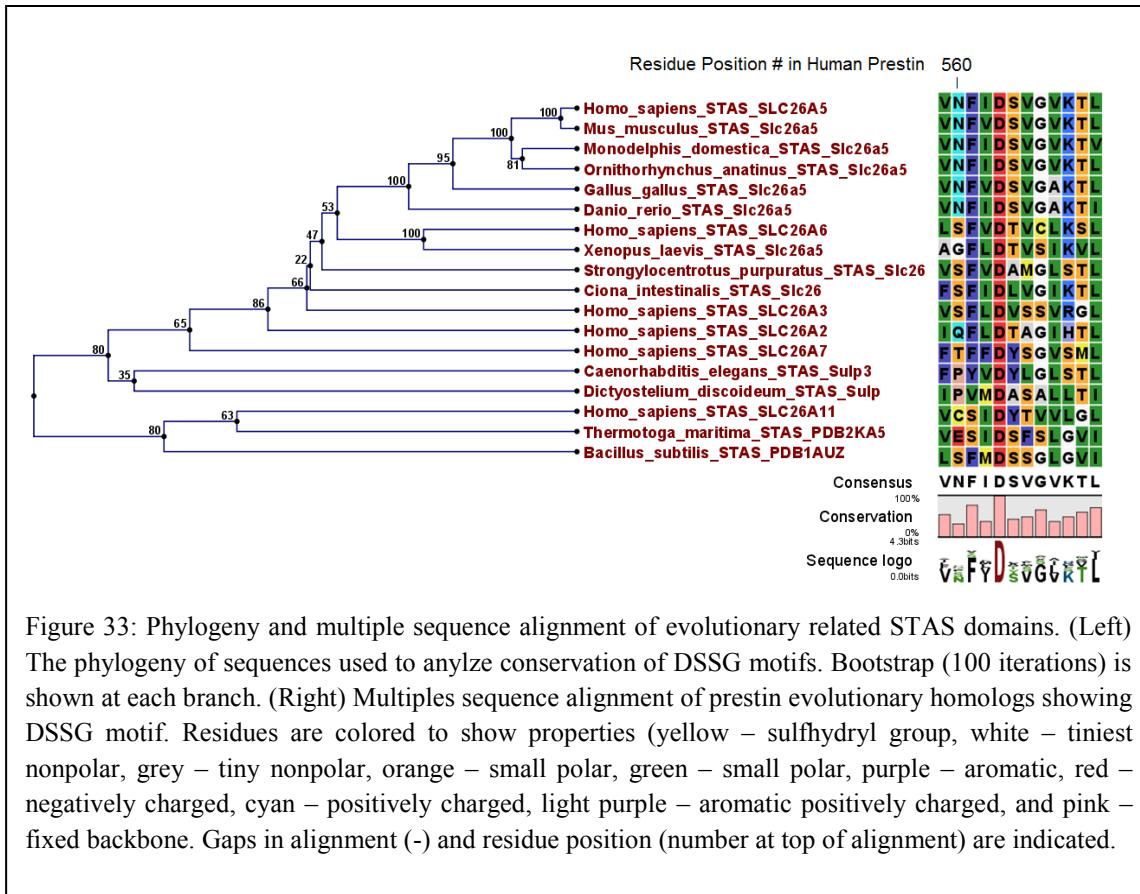


Figure 33: Phylogeny and multiple sequence alignment of evolutionary related STAS domains. (Left) The phylogeny of sequences used to analyze conservation of DSSG motifs. Bootstrap (100 iterations) is shown at each branch. (Right) Multiple sequence alignment of prestin evolutionary homologs showing DSSG motif. Residues are colored to show properties (yellow – sulphydryl group, white – tiniest nonpolar, grey – tiny nonpolar, orange – small polar, green – small polar, purple – aromatic, red – negatively charged, cyan – positively charged, light purple – aromatic positively charged, and pink – fixed backbone). Gaps in alignment (-) and residue position (number at top of alignment) are indicated.

Nonlinear Capacitance (NLC)

Based on the MSA in the previous section, the DSSG motif serine (S654) was considered a putative phosphorylation site. The gerbil S654 was substitution mutated to a threonine, which can be phosphorylated, and an alanine, which cannot be phosphorylated. The NLC parameters were calculated for cells transfected with the pgPG S654A and pgPG S654T plasmid constructs (Fig. 34). Only the V_{PK} of cells transfected with pgPG S654T was significantly different from cells transfected with either pgPG or pgPG S654A (-68.7 and -51.2 mV, respectively, Student's t-test $p \leq 0.01$). Cells transfected with pgPG S654T had a significantly hyperpolarized V_{PK} suggesting more subunits were in the shortened state at rest. The percentage of cells with a discernable NLC current, however, was dramatically less in cells transfected with pgPG S654A than in either cells transfected with pgPG S654T or pgPG wild-type. This suggests that prestin S654 is important for prestin function.

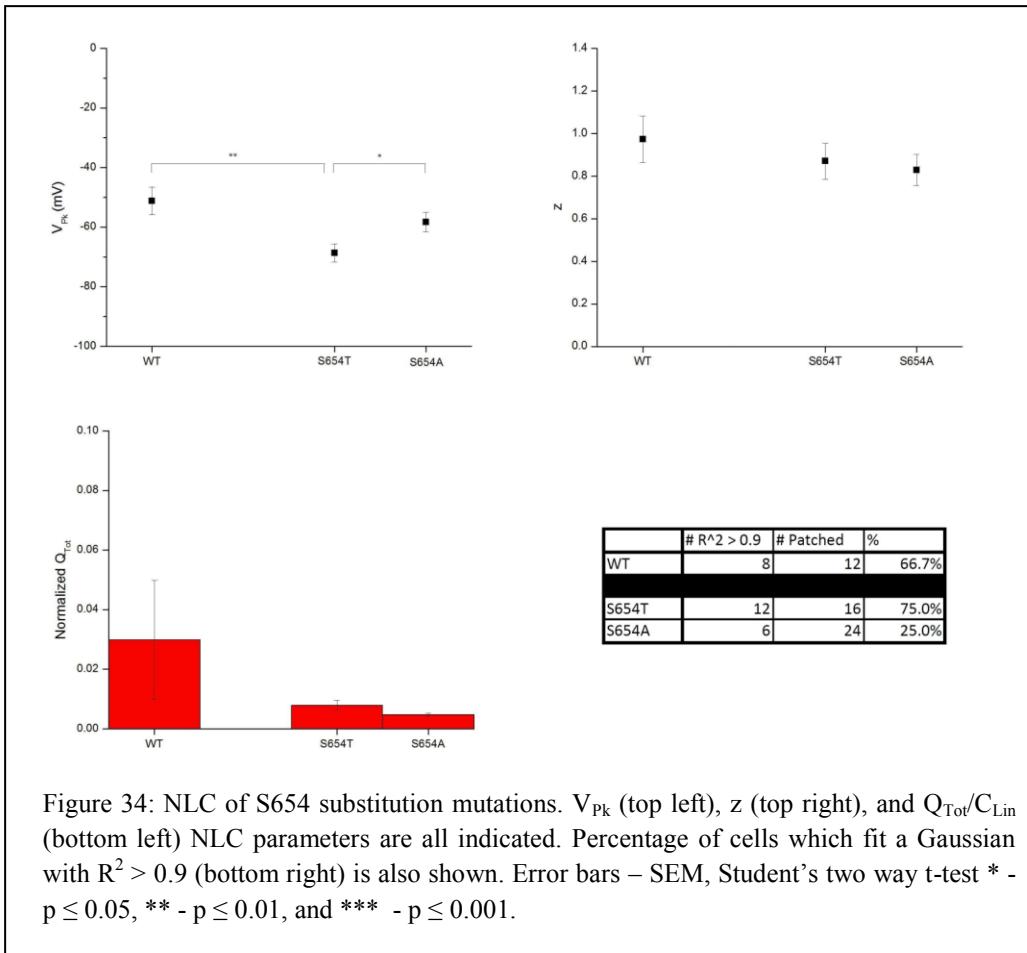
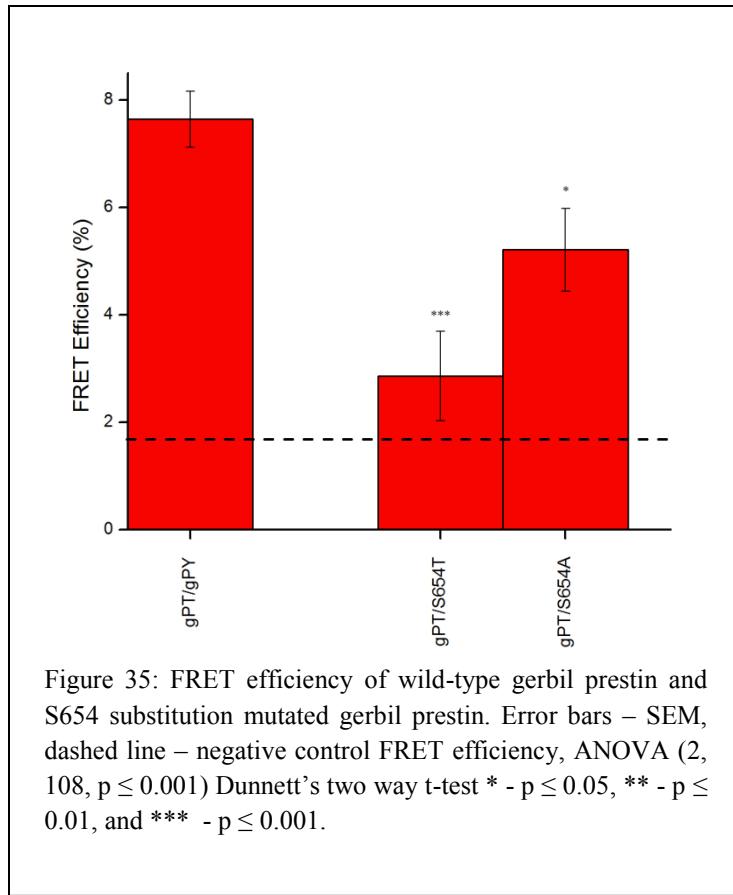


Figure 34: NLC of S654 substitution mutations. V_{pk} (top left), z (top right), and Q_{tot}/C_{Lin} (bottom left) NLC parameters are all indicated. Percentage of cells which fit a Gaussian with $R^2 > 0.9$ (bottom right) is also shown. Error bars – SEM, Student's two way t-test * - $p \leq 0.05$, ** - $p \leq 0.01$, and *** - $p \leq 0.001$.

Förster Resonance Energy Transfer (FRET)

Substitution mutations of the putative phosphorylation site were also evaluated for changes in higher order structure using FRET (Fig. 35). Cells were co-transfected with wild-type gerbil prestin donor plasmid constructs (pgPT wild-type) and S654 substitution mutated gerbil prestin acceptor constructs (pgPY S654T/A). Cells co-transfected with pgPT wild-type and pgPY S654T or pgPT wild-type and pgPY S654A had significantly reduced E_{FRET} when compared to cells co-transfected with the pgPT wild-type and pgPY wild-type (2.86%, 5.21%, and 7.69%, respectively; Dunnett's two way t-test for wild-type/S654T versus wild-type/wild-type and wild-type/S654A versus wild-type/wild-type is $p \leq 0.001$ and $p \leq 0.05$, respectively). The FRET efficiency of cells co-transfected with pgPT wild-type and pgPY S654T were also indiscernible from the FRET efficiencies of cells co-transfected with the negative control of pgPT wild-type and p38Y (2.86% and 1.80%, respectively, Dunnett's one way t-test $p > 0.05$). The FRET efficiency of cells co-transfected with pgPT wild-type and pgPY S654A, however, were significantly greater than the negative control pgPT wild-type and p38Y (5.21% and 1.80%, respectively; Dunnett's one way t-test $p \leq 0.01$). Loss of E_{FRET} indicates a significant loss in oligomerization between S654 mutated prestin subunits and wild-type prestin subunits. Surprisingly, the larger effect was seen in with the S654T prestin subunits, which should be able to phosphorylate similar to wild-type subunits.



III. Conclusion

In this chapter the role of homo-oligomerization in prestin function was tested by disrupting putative oligomerization motifs. The nine cysteine residues were mutated to remove possible disulfide bonds. Only one cysteine residue mutation, C415A, reduced FRET efficiencies to negative control levels. This mutation also significantly shifted NLC V_{pk} , but did not eliminate NLC. It is possible that C415 does stabilize oligomerization through disulfide bonds, but this disulfide bond is not essential for function. It is also possible that C415 is part of an essential pore region and, when disrupted, disrupts both normal function and oligomerization. Disruption of the putative phosphorylation site in the DSSG motif (S654A) resulted in fewer transfected cells with a

discernable NLC. The S654A mutation also significantly reduced FRET efficiencies, but these FRET efficiencies were greater than baseline. Contrary to what was expected, S654T mutations, which should be phosphorylated similar to wild-type prestin, also had significant changes in FRET efficiencies and NLC. It is uncertain what role the DSSG motif plays in prestin, but disruption of this motif causes changes to both structure and function.

CHAPTER VI: Discussion

This study examined the role homo-oligomerization plays in the structure-function relationship of prestin, a protein essential for human hearing. In this discussion, the conclusions drawn from the results in the previous chapters will be discussed. Beyond the scope of the hypothesis, however, the results in this study may be used to shed light on other aspects of prestin's structure-function relationship. Based on the information garnered from this work a novel model is suggested. In addition, ways to test hypotheses generated by this work and ways to improve the methodology used are suggested.

I. Prestin Homo-oligomerization

The hypothesis that homo-oligomerization is required for SulP family member function was examined in a three step approach. First, the ability to oligomerize was examined throughout prestin homologous sequences to examine the conservation of homo-oligomerization. Second, prestin structural components were analyzed to identify homo-oligomerization motifs. Third, putative homo-oligomerization structures were disrupted to examine the functional role of homo-oligomerization.

A. Analysis of Oligomerization Conservation in Prestin Homologous Sequences

In Chapter III, it was concluded that prestin homo-oligomerization is conserved throughout SulP evolution. The rationale for that conclusion is discussed below.

If homo-oligomerization is conserved throughout SulP evolution, then it is likely that the structures involved in oligomerization are also conserved. Homologous sequences were specifically chosen to examine the presence of homo-oligomerization

along prestin evolutionary lines. The human prestin paralog SLC26A11 appears to have diverged from the rest of the Slc26a family members as early as chordate evolution if not earlier. The SLC26A2 paralog is functionally distinct from the prestin paralogs and represents a major branch in the Slc26a1-9 lineage. The SLC26A3 paralog has a similar function to the prestin orthologs, but it is the most distantly-related paralog in that functional group. Both zebrafish and chicken prestin are orthologs, but they function as ion transporters rather than as motor proteins. Thus, examination of these homologous sequences was used in this study to provide a thorough analysis of homo-oligomerization evolution.

Not all prestin homologous sequences, however, behaved as expected in the cell expression system. Attempts to equalize donor and acceptor fluorescence intensities in co-transfection experiments, there were differences in protein synthesis between plasmid constructs. Chicken and zebrafish proteins were synthesized at lower levels than the gerbil prestin or human SLC26A11 proteins. It may not be surprising that chicken and zebrafish prestin homologous sequences may not be readily synthesized in mammalian cell lines due to differences in differences between mammalian and non-mammalian protein synthesis, interaction, regulation, and degradation. Interestingly, SLC26A2 and SLC26A3, which are human proteins, were also synthesized at lower levels than the gerbil prestin and human SLC26A11. In fact, since the SLC26A3 protein is found in the kidney, HEK cells should be ideal for SLC26A3 synthesis [56]. However, synthesis of the SLC26A3 protein was not detected. Co-factors that are necessary for SLC26A3 synthesis may be absent in this particular cell line. Notably SLC26A11, which is the most distantly related prestin paralog, was synthesized at levels similar, if not greater, than

gerbil prestin. This may be explained by the fact that SLC26A11 does not have an intervening sequence (IVS), which is seen in the rest of the chordate SulP family (see Chapter IV). The IVS is hypothesized to mediate protein-protein interactions [43]. Since SLC26A11 does not have an IVS, it, unlike other Slc26as, may not require co-factors for proper synthesis. While questions about synthesis control remain unresolved, it was possible to determine difference in FRET efficiencies in cells transfected with all homologous sequences except SLC26A3.

All of the prestin homologous sequences tested had a significant increase in FRET efficiencies when co-transfected with corresponding homologous sequence donor/acceptor pairs (*i.e.* pgPT/pgPY, pcPT/pcPY, and pA11T/pA11Y) relative to the FRET efficiency of cells transfected with homologous sequence donor alone or cells co-transfected with pgPT/p38Y. Oligomerization potential must be analyzed carefully since FRET efficiencies can be affected by at least three different factors: 1) changes in ratio of donor:acceptor protein, 2) changes in protein structure, and 3) changes in protein-protein binding efficiencies (*i.e.* oligomerization). While it is difficult to control the ratio of donor:acceptor plasmid present or donor:acceptor protein synthesis in any given cell after LipofectamineTM co-transfection, this was compensated for by adjusting the amount of plasmid transfected and selecting only cells with approximately equal donor:acceptor fluorescent intensities. This reduced the variability in FRET efficiencies due to different donor:acceptor synthesis rate. Plasmid constructs were designed to ensure the link between each homologous sequence and its fluorescent protein was identical. This limited variation in fluorophore distance from one plasmid construct to the next. Notably, some variation may be unavoidable as a result of differences in tertiary structure between

homologous sequences. By limiting the first two factors, more confidence can be placed in the relationship between changes in FRET efficiencies and oligomerization. Therefore, these results demonstrate that homo-oligomerization occurs in the prestin homologous sequences tested.

Despite the variety of homologous sequences examined, two possibilities could explain the presence of homo-oligomerization in the homologous sequences chosen: 1) homo-oligomerization is conserved from a common SulP ancestor (homo-oligomerization structures likely conserved in all SulP proteins), or 2) homo-oligomerization arose independently in different Slc26a homologous sequences (homo-oligomerization structures likely unique to different branches of SulP proteins). If homo-oligomerization structures are conserved, then it is likely that SulP homologous sequences should form oligomers with each other. When this was tested, E_{FRET} was significantly greater in cultures co-transfected with gerbil prestin donor plasmid constructs in all homologous sequences except for when co-transfected with the SLC26A2 paralog. However, the average FRET efficiencies for these cells all had significantly greater FRET efficiencies than the negative control (pgPT/p38Y). The loss in FRET efficiencies for SLC26A2 co-transfection may have been due to reduced synthesis of the SLC26A2 protein in the HEK cells compared to gerbil prestin. These FRET efficiency results strongly argue that oligomerization is occurring between homologous sequences. Thus, prestin homo-oligomerization is most likely an evolutionarily conserved property in the SulP family. Examination of conserved structures throughout the SulP family should reveal conserved structures involved in prestin homo-oligomerization.

B. Identification of Oligomerization Motifs

In Chapter IV, bioinformatic techniques were used to search for conserved structures within prestin homologous sequences. Only the DSSG motif was linked to a possible role as a protein-protein interaction motif. However, several other aspects of prestin structure were also revealed. This is discussed below.

Both the prestin SulP and STAS domains were identified and characterized. Analysis of the SulP domain of prestin suggests that this domain may be part of a larger domain superfamily within the APC domain clan. Analysis of primary structure indicates that the SulP domain family is related to the XUP, HCO3, and BenE domain families, and no other APC clan members. Analysis of hydrophobicity also suggested a secondary structure profile that matched the XUP, HCO3, and BenE secondary structure profiles, but not that of any other APC or MFS domain clan member despite similar features such as 12 putative TMHs and intracellular termini. Unfortunately, to date, none of these related domains are associated with a PDB structure, making it difficult to predict tertiary structure (see Chapter VI for more detailed comparison). The similarity in primary, secondary, and possibly tertiary structure suggests SulP, XUP, HCO3, and BenE may share similar structure-function relationships. Therefore, data on any protein with these domains may have bearing on prestin as well. Of all these related proteins, HCO3 has been the most extensively studied. Interestingly, before prestin was identified, the OHC motor protein was actually predicted to be a relative of HCO3 proteins [85]. Like prestin, HCO3 domain family members are thought to form dimers and possibly tetramers. However, the motifs regulating the HCO3 homo-oligomerization are not known and

could not be correlated to the SulP domain family [86, 87]. More empirical data are necessary to determine the possibly shared oligomerization sites in these transmembrane domains.

The structure of the STAS domain was better characterized. Although the primary and secondary structure of the STAS domain did not correlate to other members of the ATP-Grasp domain clan, the prestin STAS domain did share homology with several PDB structures including those for other STAS domains. Despite not sharing a similar primary or secondary structure with other ATP-Grasp domain clan members, homology modeling of the prestin STAS domain had similar tertiary structure to ATP-Grasp proteins. This similarity of tertiary structure may correlate to similar function, specifically nucleotide triphosphates (NTP) catalysis [88]. The tertiary structure of the model prestin STAS domain was also similar to the homologous SpoIIaa structure, which functions as an inhibitor of sporulation. SpoIIaa sporulation inhibition is disrupted by phosphorylation of the protein-protein interaction region (loop 2, loop 4, and α 3) [47]. The conservation of this SpoIIaa structure in the model prestin STAS domain indicates that phosphorylation regulated protein-protein interaction may also occur at this site in prestin. Although the protein-protein interaction site in SpoIIaa regulates interaction between SpoIIaa and SpoIIab proteins, this site may have evolved into a homo-oligomerization site in prestin proteins. Thus, this presents an attractive area to examine homo-oligomerization and is analyzed in the following chapters.

Prestin was also examined for motifs and residues that may play a role in homo-oligomerization. Coiled-coils and leucine zippers are common oligomerization motifs, but prestin did not appear to contain any of these motifs. None of the other identified

known motifs correlated to oligomerization. Thus, Gibbs Motif Sampler and our novel PAIM algorithm were employed to find *de novo* motifs and residues of interest. In the transmembrane portion of prestin, Gibbs Motif Sampler and PAIM suggested two areas of high conservation: TMH 1-4 and 8-11. It is unclear what role these conserved motifs may play in transport or motor function in SulP proteins, but the higher conservation suggests that they are important for function, possibly as pore forming regions. It is also unclear if these regions have a role in oligomerization.

In the STAS domain, Gibbs Motif Sampler and PAIM identified several interesting regions. Similar conserved motifs were found in β 2 and β 4, an area that corresponds to the putative NTP catalysis site. A DSSG motif was also identified in L4, which is part of the putative protein-protein interaction site. The DSSG motif is a well-described phosphorylation site in SulP proteins, but has not been described in prestin prior to this work [48, 50, 89, 90]. The phosphorylation of the DSSG motif results in disruption of the protein-protein interaction in SpoIIaa, and, if STAS domain homo-oligomerization occurs in the same manner removal of the phosphorylation site should stabilize homo-oligomerization [47]. If the DSSG does regulate homo-oligomerization through phosphorylation, then replacing the phosphorylatable residue with a non-phosphorylatable residue should disrupt homo-oligomerization and possibly function.

C. Analysis of Oligomerization Motif Disruption

In Chapter V, both putative disulfide bonds and the DSSG motif were disrupted in an attempt to prevent homo-oligomerization. C415 mutations reduced FRET efficiency to negative control levels and affected NLC. However, C415 is not conserved, suggesting

that if it does participate in disulfide bonding, than disulfide bonding at C415 is unique to mammalian prestin orthologs. It may be more likely that C415 is part of a motif required for normal function as suggested in Chapter IV. DSSG mutations also affected both FRET efficiencies and NLC, but the results were incongruous with what was expected for a homo-oligomerization site regulated by phosphorylation. The results and conclusions of Chapter V are discussed in more detail below.

Cysteine Mutants

If disulfide bonds are essential for maintaining conserved SulP homo-oligomerization, then the cystyl residues participating in disulfide bonding should also be conserved. However, when the only conserved cysteine (C52) was mutated to alanine no changes in either NLC or FRET efficiency were observed. This argues against a conserved role of disulfide bonds in prestin quaternary structure. Two of the cysteine-alanine substitution mutants, however, showed a significant shift in NLC V_{pk} : C124A and C415A. The previous chapter suggested C124 is in the SulP motif, and part of a possible conserved pore region. The C124 residue position itself, however, is not well conserved. In the Slc26a5 homologous sequences, the C124 residue position contains either a cystyl or a threonyl residue. Both residues are polar, although threonyl are considerably larger due to an additional methyl group. This size difference did not affect FRET efficiencies suggesting changes in function where due to specific functional characteristics rather than homo-oligomerization. It is still possible that C124 does form disulfide bonds, but if it does, disulfide bonding at this residue position is not essential for homo-oligomerization.

More likely, taking into account the previous chapter's results, C124 is an important residue in a motif essential for mammalian prestin motor function.

The C415 residue position contains a cystyl residue that is unique to mammalian prestin. When C415 is substituted with either a non-polar or polar residue, there is a significant change in NLC properties of the prestin mutation. Also, the C415A mutation was the only mutation to show a substantial loss in FRET efficiency between itself and wild-type gerbil prestin. One explanation of these results is that C415 participates in a disulfide bond. When mutated to residues other than cystyl, subunits are less efficient at forming homo-oligomers (loss of FRET efficiency), which changes the physiological characteristics of prestin (depolarization of V_{pk} and reduction in Q_{Tot}/C_{Lin}). Another possible explanation is that C415 is an essential residue in a functionally crucial pore region (C415 is in the putative pore-forming region of TMH 8-11). Misfolding due to mutation of this residue could account for the loss of FRET efficiency, loss of membrane insertion (reduction of Q_{Tot}/C_{Lin} in the C415S mutant), and changes to physiology (depolarization of V_{pk} in C415A mutants). Further analysis of this residue position may help distinguish between these two possibilities. For example, Western blot analysis in reducing and non-reducing states can be used to demonstrate the presence of a disulfide bond [17]. If the C415A mutant did not have similar disulfide bonding to wild-type prestin it would help confirm the C415's role in disulfide bonding. However, the evidence of uniqueness of C415 to mammalian prestin orthologs would suggest that, if C415 does participate in disulfide bonding, it is a unique property of prestin. It is unclear whether or not C415 participates in disulfide bonding, but mutations of C415 affect prestin structure and function.

Significant, although smaller, changes in NLC V_{pk} were also seen in cells transfected with pgPG C192A and pgPG C260A. The C192 residue position is in a putative pore. The C260 residue position is not predicted to be in an essential functional area according to the previous chapter's results. Neither residue position has discernable evolutionary pattern. It is unclear what role either of these residues plays in gerbil prestin or its homologous sequences.

Recently, two papers examined cysteine mutations [91, 92]. These studies disagreed on the presence and functional importance of disulfide bonds. McGuire *et al* argued that disulfide bonds, if present, were not essential for prestin function. They performed cysteine to serine mutations (preserving size and polarity while eliminating disulfide bond formation) on all nine prestin cystyl residues. They also found that C124 and C415 played roles in function, but they suggest that the effect of C124 mutation was caused by lack of membrane insertion and that mutations in C415 disrupted hydrophobicity. Unfortunately, although the cysteine mutations were analyzed with Western blots, these were carried out solely in reduced conditions and did not analyze the presence or absence of disulfide bonds. The McGuire report also suggested a functional role for C260, C381, and C395 as well as C192 and C196 in combination. Although we did not perform serine scanning mutagenesis on all cystyl residues, our results did show similar functional changes with C192A and C260A, although not with C381A or C395A. Functional changes in C381 and C395 may be explained by their position in TM helix 9, a putative pore forming helix, and their position in or around hydrophilic residue positions (Fig. 29, also see previous chapter). The functional importance of C260 is more elusive, but McGuire *et al.* theorized that C260 is part of a tetrameric pore.

Bai *et al.* argues that prestin is stabilized by intra-subunit disulfide bonding and that homo-oligomerization is independent of disulfide bonding. Bai *et al.* analyzed disulfide bonding by performing single and multiple cysteine-to-serine mutations. As with the results in this work and McGuire *et al.*, C415 mutations result in a significant change in NLC properties. Further, Bai *et al.* shows that NLC was lost completely in combination mutations of C415S with C192/C196S, C260S, or C381S. In addition the combination mutation of C192/C196S and C395S results in complete loss of NLC. This evidence was used to argue that C415 and either C192/C196 participate in disulfide bond formation. Although it is intriguing that NLC can be abolished by only two or three single point mutations, the argument that C415 and C192/C196 form a disulfide bond is not well-supported. When Bai *et al.*, examined combination mutation of C415A (which disrupts disulfide bonds as efficiently as C415S mutations) with C192/C196S, C260S or C381S, NLC was partly recovered. This argues that the loss of NLC in C415S combination mutations may not have been due to loss of disulfide bonds, but rather changes in other cysteine-specific properties.

It is not surprising that many of the cysteine mutations resulted in significant changes in NLC. Several other papers have also shown a small, but significant, change in NLC after mutating a variety of mammalian prestin residues [16, 53]. This may be the result of the high conservation of mammalian homologous sequences compared to its paralogs [40, 41]. This suggests motor prestin homologous sequence function does not tolerate mutations well. The most interesting result is the change of NLC in C415 mutations, which is common to all three reports. The evidence that it is the only cysteine unique to mammalian prestin and the only cysteine mutation to cause a loss in FRET

efficiency only reinforces the suggestion that C415 may be of particular importance. However, its role in disulfide bonding and/or homo-oligomerization is still unclear.

DSSG Motif

Chapter IV described a DSSG motif that correlated with a SpoIIaa phosphorylation site regulating protein-protein binding. In SpoIIaa, protein-protein binding is disrupted when this motif is phosphorylated [47]. This led to the hypothesis that the S654 residue in the gerbil prestin DSSG motif has a similar function. Alanine substitution should eliminate phosphorylation at this site and stabilize protein-protein binding, while threonine, another phosphorylatable residue, should have little effect on normal function. If this possible phosphorylation site participates in prestin homo-oligomerization, then it was expected that cells transfected with pgPG S654A mutation should stabilize homo-oligomerization (increase or no change in both FRET efficiency and NLC $Q_{\text{Tot}}/C_{\text{Lin}}$) while cells transfected with pgPG S654T should have similar properties to wild-type prestin. S654A mutation, however, did not cause a significant change in the NLC parameters, although a smaller percentage of the cells transfected with pgPG S654A had a discernable NLC compared to cells transfected with either pgPG wild-type or pgPG S654T. In addition, cells co-transfected with the pgPG wild-type donor and pgPG S654A acceptor had FRET efficiencies less than wild-type donor/acceptor pairs, although significantly greater than the negative controls. S654T mutations, which were hypothesized to have little effect on prestin structure-function relationship, caused a significant hyperpolarization in NLC V_{pk} as well as FRET efficiencies indistinguishable from the negative controls. Both the S654A and S654T

mutations effects are contrary to the initial hypothesis based on the idea that the DSSG motif participating in homo-oligomerization which is then regulated through a phosphorylation site. Despite this counter evidence, both mutations affected the structure and/or function suggesting that the residue is important for normal prestin function. A possible role for this putative phosphorylation site is discussed below.

D. Summary

Although this work was unable to conclusively state that homo-oligomerization is essential for Slc26a function, advances were made in prestin structure-function relationship. Combined with recent structural data, obtained from prestin and its relatives, several novel hypotheses are extrapolated below.

II. Transmembrane Region

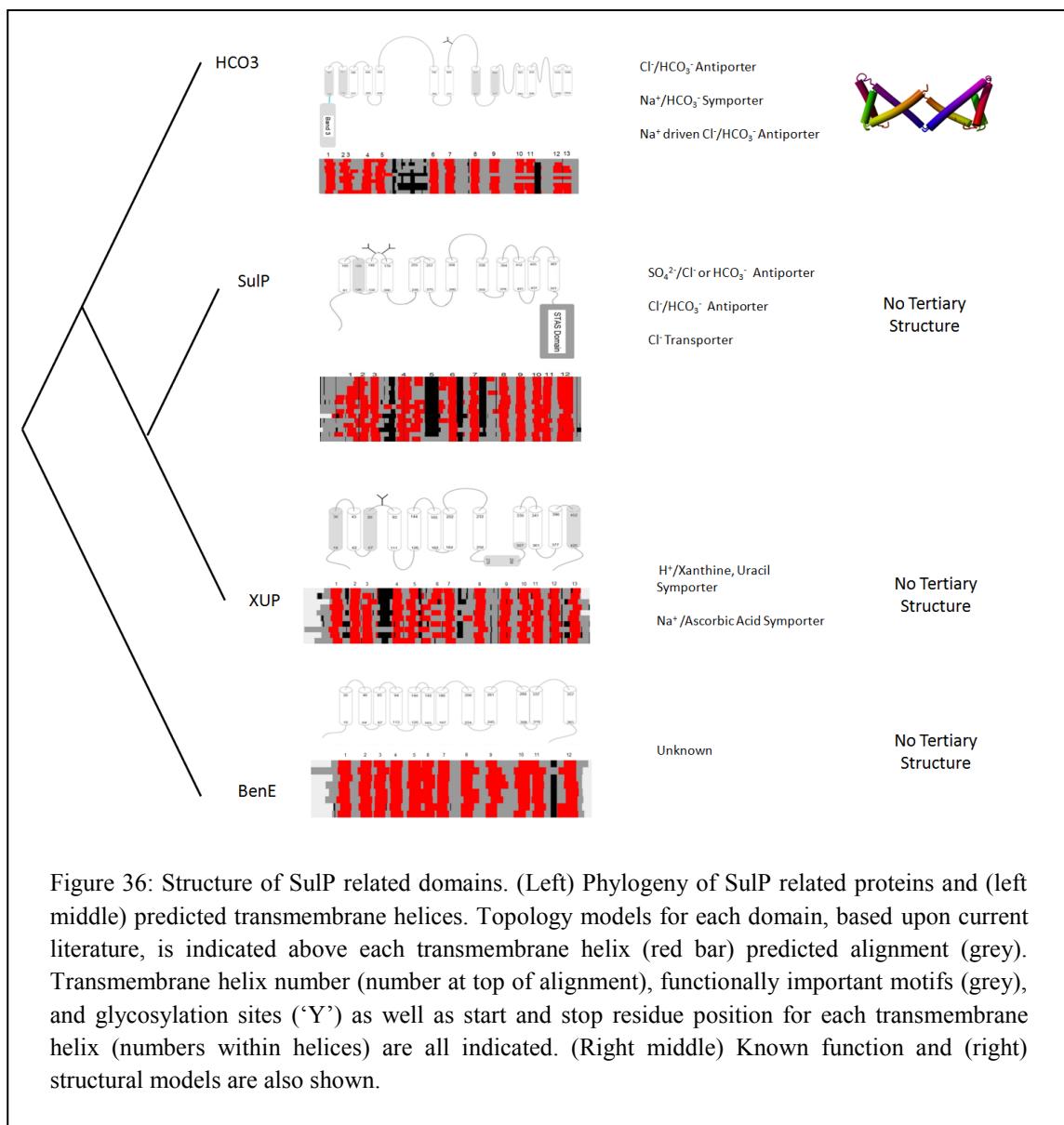
In this work, a new putative topology model for the SulP family is proposed (Chapter IV). The SulP topology has 12 TMHs with two glycosylation sites between TMH 3 and 4, a cyclic nucleotide phosphorylation site between TMH 4 and 5, and two putative pore regions in TMH 1-4 and TMH 8-11. Also shown in Chapter IV, the prestin SulP family's primary and secondary structure is related to the primary and secondary structure of the HCO₃, XUP, and BenE families (Fig. 35). These families have intracellular termini and share similar hydrophobicity profiles (~12 putative TMHs, with TMHs 1-6 poorly delineated, TMHs 7-9 clearly delineated, TMHs 10 and 11 closely approximated, and TMH 12 either one long TMH or two closely approximated helices). Based on this similarity of structure, it is likely that all members of this group share

similar functional mechanisms (e.g. many Na^+ , K^+ , and Ca^{2+} channels share a 4+2 structure and form ion channels through the interaction of 4 pore loop in the last 2 TMHs of the structure). Until now, the SulP, BenE, XUP, and HCO3 domains were studied as disparate transmembrane domains. If studied as similar domains, research of one domain's structure-function relationship can be translated into a hypothesis for another's structure-function relationship. A brief synopsis of each of these family members is given.

The least studied clan member, the BenE family, is classified as a benzoate transporter homologous sequence due to chromosomal location, but this protein family shares little homology to the other benzoate transporters [93]. It is not known if the termini in these proteins are intracellular. Based on its relationship to SulP, XUP, and HCO3 domains, a novel topology for BenE is proposed that contains 12 TMHs and intracellular termini (Fig. 36). Neither the tertiary or quaternary structure has been examined in the BenE protein family.

The XUP domain is the closest related domain to the SulP domain. The XUP family, which includes the solute carrier 23 (Slc23) family, is also referred to as either the nucleobase ascorbate transporter (NAT) family or the nucleobase cation symport 2 (NCS2) family (refer to www.TCDB.org). Proteins in the XUP family act as symporters for cations (Na^+ or H^+) and larger molecules (xanthine, uracil, ascorbic acid). The XUP family is also predicted to have 12 TMHs. However, the 12th TMH in the XUP family is split into two TMHs. The 8th TMH is predicted to be entirely in the cytoplasm to keep both termini on the intracellular side of the plasma membrane [94]. A glycosylation site is located between TMH 3 and 4 [95]. Two important motifs have been identified in the

XUP domain: one is located in TMH 1 and a second motif carboxyl to TMH 9 [94, 96]. No solved tertiary structures are associated with the XUP family. The oligomerization state of the XUP domain has not been well studied.



The HCO₃ family, which includes the solute carrier 4 (Slc4) family [also known as anion exchangers (AE), or red cell band 3] is the best studied protein family in this group. This family is associated with a large amino terminal composed of the band 3

domain (PF07565). The predicted topology for the HCO3 family include 13 TMH with a short strand of residues passing through the plasma membrane to ensure both termini are on the intercellular side of the plasma membrane [97]. A glycosylation site has been identified between TMH 7 and 8 [98]. Functional importance has also been shown for TMH 1 and 2 as well as TMH 8, 9, and 10 [98, 99]. It has been shown that HCO3 proteins form dimers and possibly tetramers [98]. A recent electron crystallography study suggests that TMHs 1 and 2 form a “V-shaped” pore similar to Clc channels. This study further suggests that TMH 8 and 9 interact with a homo-dimer to form a central pore region [100].

Due to their similarity in primary and secondary structure, it is likely that tertiary structures as well as structure-function relationships are shared in all four families. This is further supported by evidence across families that regions in TMH 1-4 and TMH 8-11 have functional importance. If data from the XUP and HCO3 family domains can be extrapolated to the SulP family, it would suggest that the SulP family members also form a dimer (and possibly a tetramer) with two pore forming regions (one ‘edge’ pore with TMH 1 and 2 as well as a ‘central’ pore with TMH 8 and 9). These two regions also contain many of the putative motifs and conserved residues identified in Chapter IV. Thus, it is purposed that, like the HCO3 domain model, prestin forms two pore regions: TMH 1-3 at the ‘edge’ of the protein and TMH 8-9 at the center of the protein. This model presents the first look at higher resolution of prestin pores in the transmembrane region.

III. STAS Domain

In Chapter IV, a homology model of the STAS domain was created based on the 2K5A SpoIIaa solved structure. This model features a β -sheet which is hypothesized to play a role in NTP catalysis. Also a protein-protein phosphorylation regulated binding motif was identified (DSSG). Mutations of the phosphorylation site, however, contradicted the hypothesis that this motif participated in homo-oligomerization. Although this motif does not participate in homo-oligomerization, several recently submitted solved structures for the STAS domain may clarify the mutational results and help define what role this motif, as well as the domain as a whole, plays in the SulP family of proteins.

Two new STAS domain crystallized structures have been reported in the Protein Database: 3MGL and 3LLO (Fig. 37). The 3MGL solved structure is the STAS domain of SulP protein from the *Vibrio cholera* (a gram negative bacterial species). It shows the quaternary structure of the STAS domain as a homo-tetramer. In this solved structure the β -sheet planes are facing one another towards the center of the protein. The DSSG motifs are pointed towards the presumptive plasma membrane portion of the protein (based on the orientation of amino termini of the solved structure). Unfortunately, no publication has been submitted along with the 3MGL solved structure. The 3LLO solved structure is derived from the partial structure of the STAS domain [505-563]GS[637-718] from rat (*Rattus norvegicus*) prestin (Slc26a5). This structure is fairly similar to the 2K5A modeled structure, but α 1 is tilted relative to α 2 and the amino terminal is more complex. The study which proposed the 3LLO structure suggests that the SpoIIaa homologous protein-protein interaction area, which contains the DSSG motif, faces the

transmembrane portion of the prestin protein [101]. Based on these two models, the DSSG motif may interact with the cytoplasmic portion of the transmembrane region of prestin and the β -sheet planes face each other. These models may help explain the role of the DSSG in prestin model (see below).

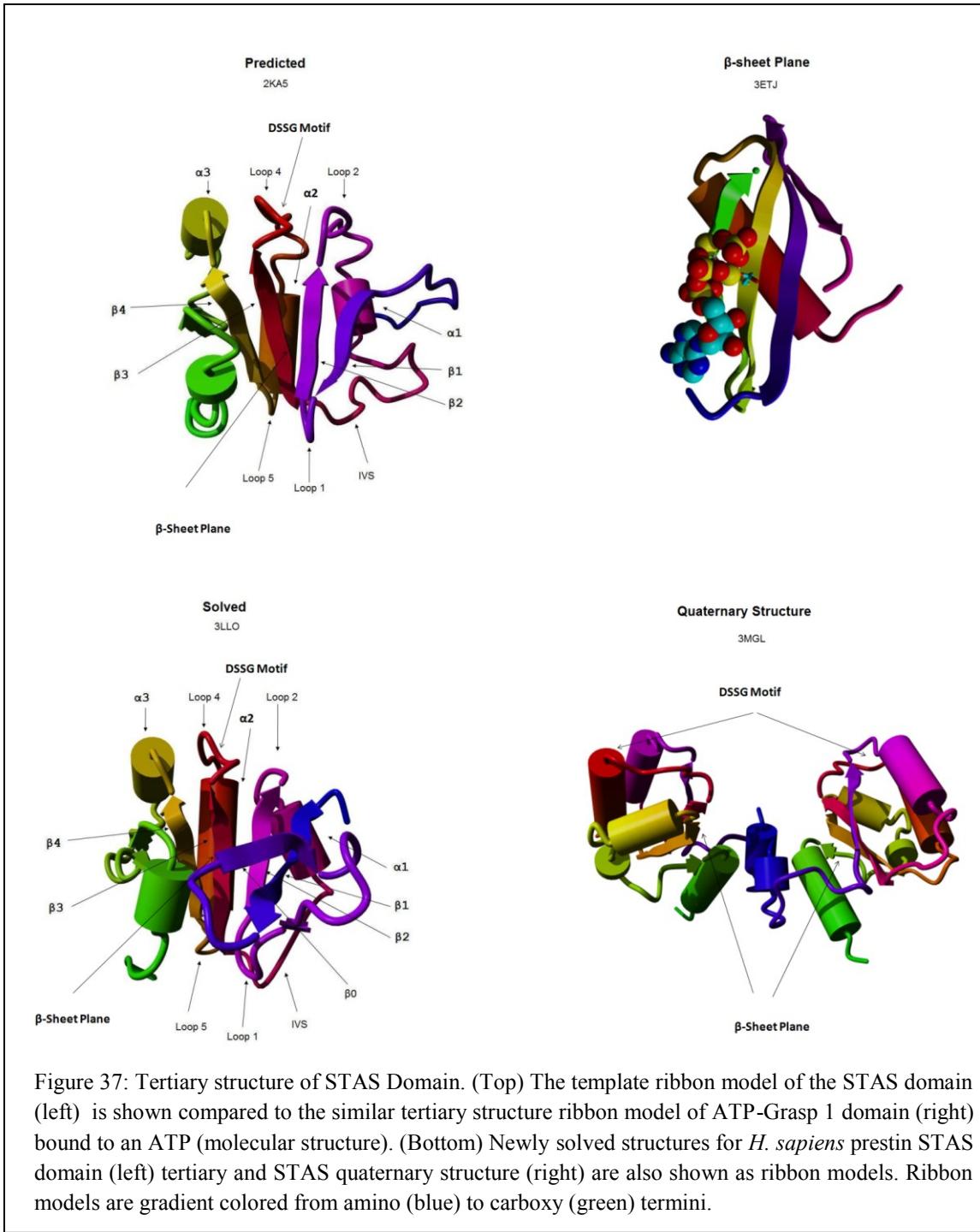
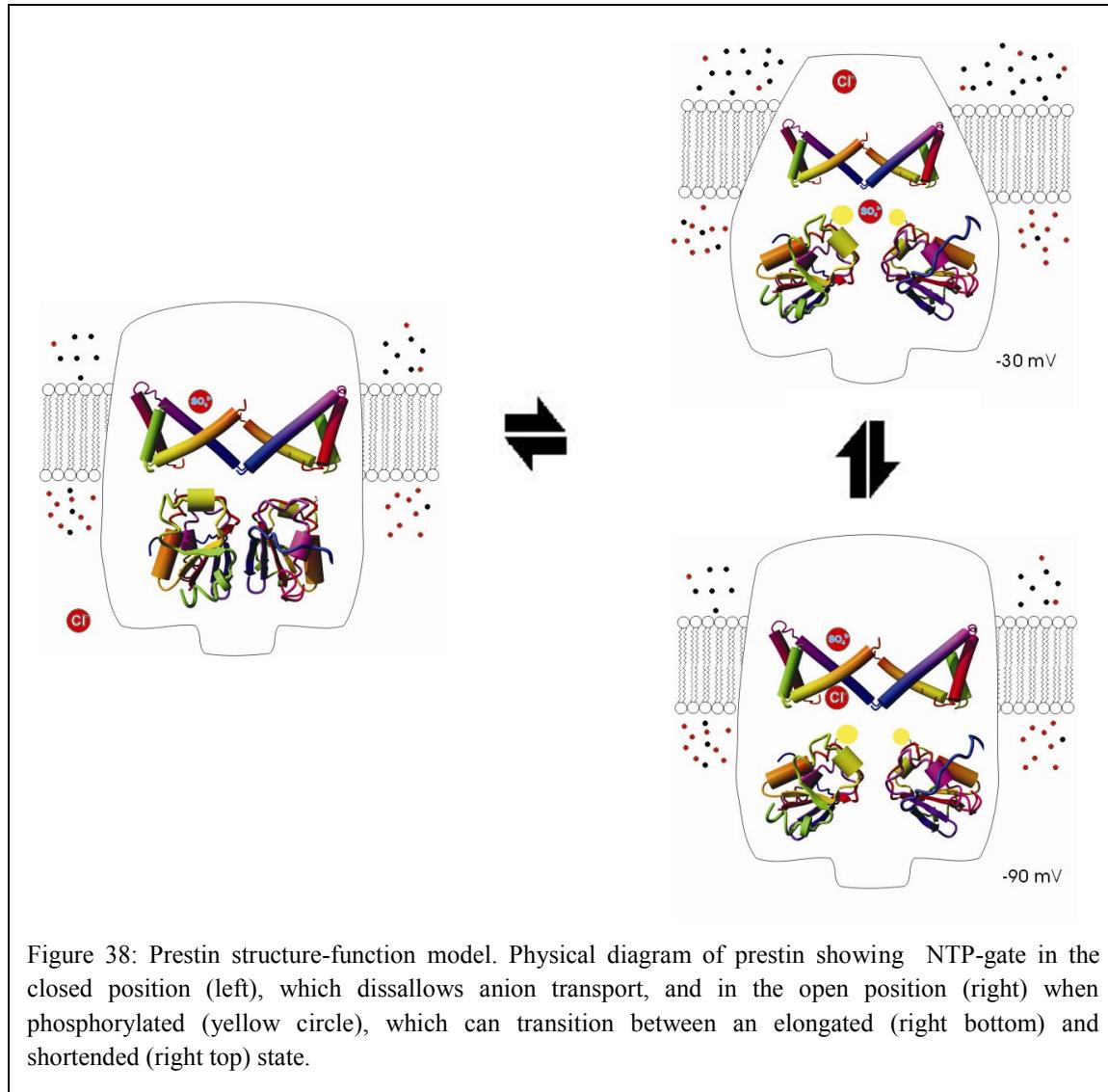


Figure 37: Tertiary structure of STAS Domain. (Top) The template ribbon model of the STAS domain (left) is shown compared to the similar tertiary structure ribbon model of ATP-Grasp 1 domain (right) bound to an ATP (molecular structure). (Bottom) Newly solved structures for *H. sapiens* prestin STAS domain (left) tertiary and STAS quaternary structure (right) are also shown as ribbon models. Ribbon models are gradient colored from amino (blue) to carboxy (green) termini.

IV. Prestin Model

A theoretical structure-function model of prestin is proposed based on evidence in this work (Fig. 38). As anions pass through the channel, the protein changes its shape between an ‘extended’ and a ‘shortened’ state (see Chapter II for further details). Anions pass through two different pore regions: a Cl⁻ pore near the protein’s exterior and a SO₄²⁻ pore near the protein’s interior. On the cytoplasmic side, the STAS domain forms an ‘NTP-gate’. When phosphorylated at the DSSG motif, the interaction between the STAS domain and the transmembrane region is disrupted. This forces the NTP-gate into an open position allowing transition from the extended to shortened state. When dephosphorylated, the DSSG motif again interacts with the transmembrane region of the protein and the NTP-gate is forced into a closed position. When in the closed position, the STAS domains’ β-sheet planes are close enough to result in NTP catalysis which phosphorylates the STAS domain forcing the NTP-gate back into the open position. This would effectively close the channel when the cell is NTP-starved, but keep it open when NTP is available.

Using this model, the results from the mutational analysis of the DSSG motif can be re-interpreted. Removal of the DSSG phosphorylation site could force the NTP-gate into a closed position resulting in fewer cells having a discernable NLC (this result is seen in the S654A mutation from Chapter V). Serine-to-threonine mutations may stabilize phosphorylation resulting in a greater percentage of NTP-gates in the open position (increased distance between carboxy terminals would explain the loss of FRET efficiency in S645T mutations from Chapter V). This presents the first suggestion that the STAS domain acts as an NTP-gate for prestin and the SulP family of proteins.



An immediate objection to the NTP-gate model is evidence that OHC fast motility is independent of ATP [67, 68]. These studies disrupted ATP production and/or added nonhydrolysable ATP analog. These perturbations of ATP concentration did not change the characteristics of fast somatic motility and suggested that somatic motility was not dependent on ATP. These experiments, however, do not necessarily argue against the NTP-gate model. In the NTP-gate model, ATP would not be required for the cycle-by-cycle elongation and shortening of prestin which drives somatic motility (which is what

was tested in the above experiments). Instead ATP would only be required in small amounts to open the NTP-gate through a phosphorylation event that may last hours to days. In fact, prestin function has already been shown to be dependent on phosphorylation of at least one other site [55]. Rather than play a direct role in elongation and shortening of the prestin molecule, the NTP-gate hypothesis suggests that the STAS domain in prestin refines the motor protein activity by opening, to allow motor function, or closing, to inhibit motor function.

V. Future Work

A. *Experiments*

The most intriguing suggestion of this work is that the STAS domain may operate as an NTP-gate. Several experiments are proposed to test this hypothesis. Prestin phosphorylation can be tested using either radio labeled phosphate or 2D gels. If prestin is phosphorylated, then confirmation of the phosphorylation site(s) can be performed through site-directed mutagenesis. Also, *ab initio* modeling of both the SpoIIaa and STAS domain in the presence of ATP may be able to predict the likelihood of ATP catalysis by this domain. Another possible test would be to analyze prestin characteristics when the STAS domain is either removed or replaced. These last types of experiments have been shown to be problematic in previous studies probably due to incorrect delineation of prestin domains [49].

In addition to the NTP-model hypothesis, the STAS domain has another interesting characteristic: the IVS. Although its function was not directly addressed in this work, the structure was examined. The function of the IVS is unknown, but part of

the structure has been shown to participate in hetero-oligomerization [102, 103]. Seeing that the IVS differs among paralogs, it implies that the IVS may serve as a paralog specific hetero-oligomerization site. Examination of the functional and interactional (*e.g.* pull-down assays, FRET analysis) implications of prestin with and without the IVS may be fruitful. cDNA for the cystic fibrosis transport receptor (CFTR), which is known to bind to prestin, has been coupled to acceptor fluorophores to permit examination of the putative role of the IVS. This work is still ongoing and is not examined in this work [104].

Sequence analysis, performed in Chapter IV of this work, identified several motifs and residues of interest. Both structural (*i.e.* FRET) and functional (*i.e.* NLC) analysis of mutations in these structural features may also prove to be fruitful. The results of these studies, however, may be difficult to interpret without a more precise picture of the structure-function relationship. In this work, it is suggested that many of these motifs and residues may participate in pore regions, but there is little structural data to confirm this hypothesis. A solved structure would dramatically help interpret these results, but none currently exist for the transmembrane region for prestin, any of its homologous sequences, or any of its clan members. Although no solved structures are available, current work by our laboratory is analyzing prestin topology using scanning cysteine accessibility methods (SCAM). We are also examining other structural features, such as homo-oligomer stoichiometry, using total internal reflection (TIRF) fluorescence. Although the resolution of this structural analysis is coarse in comparison with the resolution of x-ray or nuclear magnetic resonance (NMR) crystallography, it will greatly assist in the mutational analysis of these conserved structures.

B. Method Improvement

In this work, FRET was used to determine loss in oligomerization. FRET, however, is a sensitive indicator of distance rather than an absolute measurement of protein-protein interaction. Thus, FRET efficiency may be affected by at least three different factors: 1) the stoichiometry of the fluorophores, 2) the efficiency of oligomerization, and 3) any subsequent changes in conformation. Changes in FRET efficiency due to stoichiometry can actually be utilized in hetero-oligomerization studies to determine stoichiometry (for example see [105]), but in homo-oligomerization this can be problematic especially when dealing with transfection. In transfected cells, there is cell-to-cell variability in the stoichiometry of donor-to-acceptor plasmids. This will result in a ratio-based difference in amount of donor-to-acceptor fluorophores and the efficiency of energy transfer. In these experiments, the variability was minimized by selecting cells that had similar donor-to-acceptor intensity ratios. This problem could have been better addressed by plasmid injection in oocytes. Although early attempts at plasmid injection failed, oocyte injection remains an enticing option. Despite several improvements to the accuracy and precision of the FRET methodology (*i.e.* FLIM and acceptor photobleaching, compare result errors to [16, 106]), variances in FRET efficiencies do not discriminate between loss in oligomerization and changes in protein confirmation. Western blot analysis could help differentiate between oligomerization and confirmation based changes to FRET efficiency. For our laboratory, and many others, Western blot based experiments have been inconsistent. This is likely due to a lack of reliable antibodies and difficulties stemming from variations in extraction from cell membrane. If method refinement continues with prestin Western blots, they will

eventually be used to help clarify the FRET results. In conjunction, TIRF is being utilized to perform single molecule imaging (SMI). Once the parameters are determined, it may be possible to perform SMI FRET analysis in live cells and analyze changes in FRET alongside changes in function. These prospective experiments will more closely simulate the *in vivo* structure-function relationship of prestin.

NLC has become the ‘hallmark’ of prestin function, but problems still remain with using NLC as a functional characterization of prestin in HEK cells. Although HEK cells are robust, available and easily transfected, transfection of HEK cells results in stoichiometry variability of plasmids from cell-to-cell in the same culture. Furthermore, protein production is limited by the physical constraints of the cell. The resulting NLC signal in transfected HEK cells is relatively small compared the OHC NLC signal. This makes it difficult to accurately calculate the NLC parameter Q_{Tot} in HEK transfected cells. In addition, the number of synthesized prestin proteins, unlike OHCs, is not dependent on cell size, but instead on the number of plasmids transfected, the efficiency of protein synthesis, and the amount of time since transfection all of which will vary cell-to-cell in the same culture. Thus, variations in Q_{Tot} are further exacerbated by the inability to normalize by cell size (*i.e.* C_{Lin}). This makes it difficult to accurately interpret loss of function in cells transfected with mutated prestin plasmids. Although electrophysiological measurements from oocytes come with their own problems, plasmid injection would help mitigate cell-to-cell variances in protein production and possibly result in more accurate Q_{Tot} calculations. Again, oocyte plasmid injection offers an enticing option.

Sequence analysis is used extensively in this work, but the precision of the data is limited. There are innumerable sequence analysis algorithms, based on various assumptions, currently available to researchers. Few of these algorithms, however, use statistical methods to determine data precision. In this work, the input sequences, algorithm parameters, and output interpretation were carefully considered. However, more advanced methods need to be developed. Part of the inability to calculate precision comes from the lack of an established ‘normal distribution’ in genetic variance. We are currently developing algorithms and methods that take into account evolution, structure, and function in sequence analysis so that accuracy and precision can be measured (for further detail see Appendix III and [107]). These new approaches should help better identify and interpret the structure-function relationship of prestin.

APPENDIX I: Plasmid Constructs

This appendix will serve to explain the reasoning and methods used in construct design. First, background information on plasmids and their use in molecular cloning will be discussed. Next, the design of the constructs used in this report will be explained. Finally, a list of primers, and their purpose, will be detailed.

I. Background

Plasmids are extrachromosomal DNA that can carry genes. Plasmids can also replicate in and are transferable to a variety of cellular environments. These properties make plasmids an attractive option as a vector for protein genes. Molecular techniques readily allow insertion of DNA into plasmids permitting biologists to study proteins in a more easily controlled and cost effective environment. However, the *in vitro* environment used in many studies performed with plasmids can skew results and may result in erroneous conclusions. Therefore, plasmid environment, structure, and design must be carefully considered to prevent erroneous results (for more in-depth review of plasmids and molecular cloning see [108, 109]).

A. Environment

Often the first consideration when using a plasmid as a vector is the cellular environment needed for the desired experiment. Plasmids can be transferred to a variety of environments. In bacteria the transfer of plasmids occurs naturally through three different mechanisms: conjugation (cell-to-cell contact), transduction (transferred through virus), and transformation (taken up from extracellular environment). Deliberate

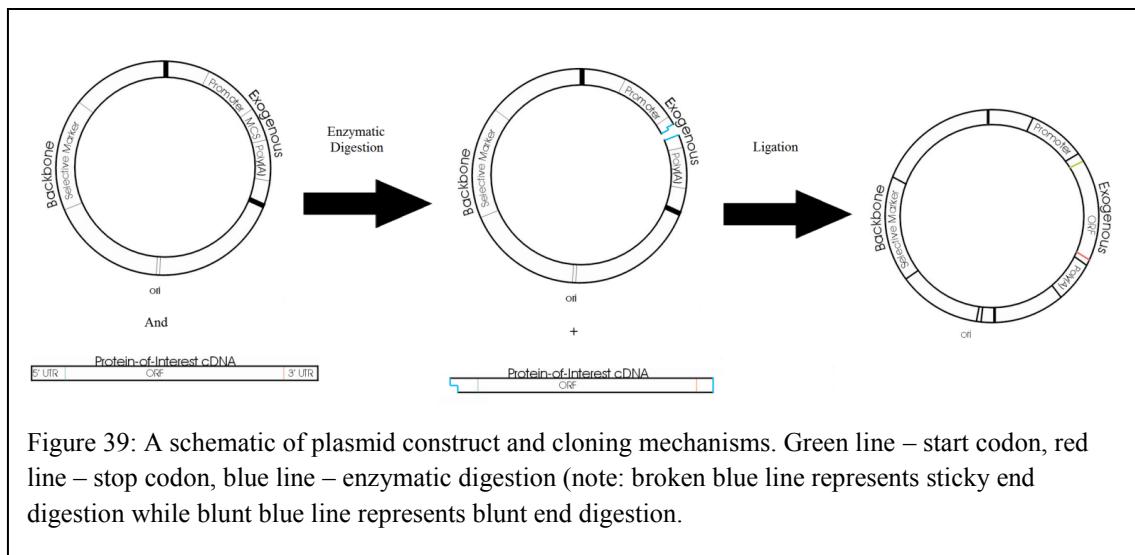
introduction of plasmids into prokaryotes usually involves transformation methods (*e.g.* heat-shock, calcium chloride, and electroporation). Deliberate introduction into eukaryotic cells, a process known as transfection, typically uses either transduction or transformation (*e.g.* cationic liposome and electroporation). The method used to transfer plasmids may depend on the desired effect.

Plasmids can be used for self-replication, mRNA expression, and/or protein synthesis. Self-replication of plasmids is driven by a prokaryotic origin of replication (*ori*) and typically performed in the prokaryotic *E. coli* due to their rapid growth a variety of available strains [110]. Plasmid self-replication can also be performed *in vitro* using polymerase chain reaction (PCR) techniques, but decreased plasmid fidelity and increased replication cost for PCR typically makes this a less attractive option. In eukaryotic cells, plasmid replication does not occur and plasmids are diluted through degradation and mitosis. Plasmids can be used, however, to insert DNA into the chromosomal DNA (stable cell lines) where the inserted DNA will be inherited during mitosis. Expression of mRNA synthesis is dependent on the promoter, which is typically a T7 or SP6 promoter for prokaryotic expression or CMV (cytomegalovirus) promoter for eukaryote expression. Synthesis of mRNA can also be performed *in vitro* to obtain uniform mRNA product. In both prokaryotes and eukaryotes, expressed mRNA will be synthesized into proteins. Prokaryotic cells can synthesize large quantities of proteins in short periods, but may result in protein misfolding. Eukaryotic cells present a more *in vivo*-like environment for eukaryotic proteins, but synthesize fewer proteins. Protein synthesis is a complex event and may require specific co-factors, interacting proteins, post-translation modification, and other mechanisms not found in every cell. Careful

consideration of cellular environment should be taken to avoid inclusion bodies and/or protein misfolding [111]. The choice of cellular environment, however, can be limited by experiment type (*e.g.* yeast two-hybrid) or available cell type (some of the most readily available are cell lines include human embryonic kidney, Chinese hamster ovarian, and HeLa cell).

B. Plasmid Structure

Artificial plasmid designed for use as a vector (*i.e.* plasmid constructs) typically includes backbone and exogenous DNA (see Fig. 39). The plasmid construct backbone contains DNA important for plasmid maintenance including the ori and selective markers. The ori is the site at which replication is initiated in bacteria and allows plasmids to be reproduced in bacteria [110]. Selective markers, which use either positive (*e.g.* X-Gal) or negative (*e.g.* antibiotic resistance) selection, help identify bacterial colonies which have been transformed with plasmid construct.



The exogenous DNA is introduced DNA important for transcription and translation of the mRNA for the protein-of-interest. Transcription is initiated by the promoter and is dependent on cellular environment. Typically T7 or SP6 promoters are used for prokaryotic expression and CMV (cytomegalovirus) promoter are used in eukaryotic expression. Multiple cloning sites (MCS), located 3' to the promoter, contain several unique restriction enzyme sites and are used to facilitate insertion of protein-of-interest cDNA. The poly(A) sequence, located 3' of the MCS, results in transcription of several adenosine nucleotides 3' to the stop codon (poly A tail), which slows down degradation of the coding sequence. An open reading frame is a sequence, after the promoter, which includes has both a start codon (AUG) and a stop codon (UAG, UAA, or UGA) in the same reading frame. The protein-of-interest cDNA, which can be ligated into the MCS, is double stranded DNA that encodes the 5' untranslated region (UTR), protein-of-interest ORF (without introns), and 3' UTR (see Appendix III). Protein tags are either inherit (*e.g.* antibody) or introduced (*e.g.* poly-his, FLAG, V5) which can be used to assist in synthesis, purification, and/or labeling. Protein tags can be 5' or 3' of the protein-of-interest cDNA, but must be within the ORF and inframe with the protein-of-interest to produce conjugated protein tags. Protein tags can also be transcribed by the same protein, but separately transcribed if an internal ribosome entry site (IRES) is used [112].

C. Molecular Cloning

After determining proper cellular environment(s) and plasmid, recombination of plasmid construct may be necessary. If the plasmid does not already contain the protein-

of-interest cDNA, then molecular cloning techniques are necessary. Before cloning the protein-of-interest cDNA into the plasmid, careful consideration must be taken to choose the correct isoform (several homologous sequences and alternate splice forms may exist for any given protein-of-interest). The protein-of-interest cDNA is often contained in an original protein construct. Also, the plasmid construct (plasmid which includes protein-of-interest cDNA) must be designed so that the desired ORF is delimitated by a Kozak sequence (Shine-Dalgarno sequence in prokaryotes) and stop codon. If bicistronic, then an IRES should be placed between the Kozak sequences for each ORF. If the desired protein-of-interest is to be conjugated to another protein and/or tag, then codons must be inframe and be within a single ORF. Many plasmid constructs contain template sequences to which aid in designing these criteria such as a MCS followed by several frame shifted stop codons and a poly(A) sequence.

An excess of protein-of-interest cDNA (insert) to initial plasmid (backbone), increases the likelihood of producing the desired plasmid construct. Excess insert can be produced through excising insert from protein-of-interest cDNA original plasmid construct or PCR cloning the insert. PCR cloning is preferential because of the ability to engineer unique restriction sites using PCR primers. The backbone must also be digested so that the appropriate ends of the excised insert can be ligated into the backbone through either blunt (paired nucleotide) and/or sticky (unpaired nucleotides) end. Backbone and insert can be combined and allowed to anneal while DNA ligase repairs nicks in plasmid construct. Correct plasmid construct sequence can be confirmed using DNA sequencing.

II. Study Plasmid Construct Design

In this study, prestin homologous sequences were conjugated to FRET pair fluorescent proteins (mTFP and vYFP). This was performed to by ligating prestin homologous sequence cDNA into plasmid constructs which contained cDNA for either mTFP or vYFP. The mTFP and vYFP plasmid constructs were engineered from the same plasmid backbone and the MCS was located the same distance from the Kozak sequence for each fluorescent protein cDNA to minimize differences in transcription and protein synthesis. The design of the fluorescent protein plasmid constructs and homologous sequence constructs is described below.

A. Design of Fluorescent Protein Plasmid Constructs

The mTFP was chosen as a donor fluorophore because of its enhanced FRET properties (see Appendix II). To facilitate cloning methods, three novel plasmid constructs were created in which the cDNA mTFP was frame shifted from the MCS by one nucleotide (pmTFPN-1, pmTFPN-2, and pmTFPN-3). The mTFP cDNA was PCR amplified to create insert using primers to add a 5' BamHI and 3' HpaI restriction enzyme cut sites (Table X). Three separate forward primers were used to insert 5' BamHI restriction enzyme site at different coding frames. The pAcGFPN-1 (PT3716-5, Clontech Laboratories, Mountain View, CA) was used as the backbone for the new pmTFPN-X plasmid constructs. The cDNA for the *Aequorea coerulescens* green fluorescent protein (AcGFP) was excised by restriction digestion using BamHI and HpaI. The three frame shifted inserts were ligated into the backbone to create pmTFP-N1, pmTFP-N2, and pmTFP-N3. The novel pvYFPN-1, pvYFPN-2, and pvYFPN-3 were created in a similar

manner using BamHI and XbaI to replace the AcGFP cDNA with frame shifted vYFP cDNA inserts (Table 4).

B. Design of Homologous Sequence-Fluorescent Protein Plasmid Constructs

In order to minimize the difference in mRNA translation and protein synthesis, the prestin homologous sequences were placed in similar plasmid constructs. The original prestin homologous sequence plasmid constructs were used to PCR amplify the prestin homologous sequence cDNA and add a 5' NheI and 3' XhoI restriction site to create the insert (Table 4). The forward primer was designed to include an ACCATG sequence to improve the Kozak sequence recognition. The reverse primer was designed so that the XhoI restriction site replaced the stop codon for each prestin homologous sequence cDNA. The pmTFP-N1 and pvYFP-N1 were used as the backbone plasmid construct to maintain reading frame between the prestin homologous sequences and fluorescent proteins. This strategy also ensured a uniform 18 amino peptide link between the prestin homologous sequences and fluorescent proteins (LELKLRILQSTVPRARDP).

III. Primers Tables

A. Cloning Primers

Cloning primers were used to build plasmid constructs for FRET experiments and are listed below:

	Forward Primer Sequence	Enzyme	Reverse Primer Sequence	Enzyme
mTFP				
pmTFP-N1	GCTGGAGGATCCAA <u>TG</u> GTGAGCAAGGG	BamHI	GCGGGGGTTAAC <u>TTA</u> CTTGTACAGCTCG	HpaI
pmTFP-N2	GCTGGAGGATCCAA <u>ATG</u> GTGAGCAAGGG	BamHI	GCGGGGGTTAAC <u>TTA</u> CTTGTACAGCTCG	HpaI
pmTFP-N3	GCTGGAGGATCCAAA <u>ATG</u> GTGAGCAAGGG	BamHI	GCGGGGGTTAAC <u>TTA</u> CTTGTACAGCTCG	HpaI
vYFP				
pvYFP-N1	ATAGGATCC <u>AA</u> TG <u>GT</u> GAGCAAGGGC	BamHI	TGAG <u>TCTAGA</u> CTTGTACAGCTCGTCC	XbaI
pvYFP-N2	ATAGGATCC <u>AA</u> AT <u>GG</u> TGAGCAAGGGC	BamHI	TGAG <u>TCTAGA</u> CTTGTACAGCTCGTCC	XbaI
pvYFP-N3	ATAGGATCC <u>ATG</u> GTGAGCAAGGGC	BamHI	TGAG <u>TCTAGA</u> CTTGTACAGCTCGTCC	XbaI
Meriones unguiculatus Slc26a5				
p _g PT	GCTAGC <u>GCTAGC</u> ACC <u>ATG</u> GATCATGCCGAAG	NheI	GCTATA <u>CTCGAG</u> CTCGGGTGTGGTGGGT	XbaI
p _g PY	GCTAGC <u>GCTAGC</u> ACC <u>ATG</u> GATCATGCCGAAG	NheI	GCTATA <u>CTCGAG</u> CTCGGGTGTGGTGGGT	XbaI
Gallus Gallus Slc26a5				
p _c PT	GCGGAC <u>GCTAGC</u> ACC <u>ATG</u> GAAGATGCTCAAG	NheI	CGTCAC <u>CTCGAG</u> GTGGTCTAAGGCAGTC	XbaI
p _c PY	GCGGAC <u>GCTAGC</u> ACC <u>ATG</u> GAAGATGCTCAAG	NheI	CGTCAC <u>CTCGAG</u> GTGGTCTAAGGCAGTC	XbaI
Danio rerio Slc26a5				
p _z PT	GC <u>ACTG</u> G <u>CCTAGC</u> ACC <u>ATG</u> GAGCACGTA <u>CT</u>	NheI	CGAG <u>CTCTCGAG</u> GTGGATGTTGGTGG	XbaI
p _z PY	GC <u>ACTG</u> G <u>CCTAGC</u> ACC <u>ATG</u> GAGCACGTA <u>CT</u>	NheI	CGAG <u>CTCTCGAG</u> GTGGATGTTGGTGG	XbaI
Homo sapiens SLC26A2				
p _A 2T	GCAGCG <u>GCTAGC</u> ACC <u>ATG</u> T <u>CTTC</u> CAGAA <u>AGT</u>	NheI	CGGCG <u>CCCTCGAG</u> AT <u>CACTA</u> ACTAAG <u>ACT</u>	XbaI
p _A 2Y	GCAGCG <u>GCTAGC</u> ACC <u>ATG</u> T <u>CTTC</u> CAGAA <u>AGT</u>	NheI	CGGCG <u>CCCTCGAG</u> AT <u>CACTA</u> ACTAAG <u>ACT</u>	XbaI
Homo sapiens SLC26A3				
p _A 3T	GCAGCG <u>GCTAGC</u> ACC <u>ATG</u> A <u>TTGAACCC</u> TTT	NheI	CGGCG <u>CCCTCGAG</u> GA <u>ATT</u> TTG <u>TTCAAC</u>	XbaI
p _A 3Y	GCAGCG <u>GCTAGC</u> ACC <u>ATG</u> A <u>TTGAACCC</u> TTT	NheI	CGGCG <u>CCCTCGAG</u> GA <u>ATT</u> TTG <u>TTCAAC</u>	XbaI
Homo sapiens SLC26A11				
p _A 11T	GC <u>ATTA</u> G <u>CTAGC</u> ACC <u>ATG</u> C <u>CTT</u> C <u>GGTG</u>	NheI	CGATA <u>ACTCGAG</u> T <u>GCCTTGAGCAGGGC</u>	XbaI
p _A 11Y	GC <u>ATTA</u> G <u>CTAGC</u> ACC <u>ATG</u> C <u>CTT</u> C <u>GGTG</u>	NheI	CGATA <u>ACTCGAG</u> T <u>GCCTTGAGCAGGGC</u>	XbaI

Table 5: List of cloning primers used in experiments.

B. Mutation Primers

Mutation primers were used to change codons and transcribed amino residues for cysteine scanning mutagenesis and DSSG mutations. They are listed below:

	<i>Forward Primer</i>	<i>Reverse Primer</i>	<i>Mutation</i>
Cysteine Scanning Mutagenesis			
C52A	GGATAAGCTGAAGCAGGCCGTTACA GCC ACTCCCCAAAAGATAAG	CAAACCTATCTTTGGGAGT GGC TGTGAACGCCCTGCTTCAGC	TGC→GCC
C124A	CCTGTTATCATGTAC GCT TTCTTGGGACCTCAGACAC	GATAGTGTCTGGAGGTCCAAAGAA AGC GTACATGATAAC	TGT→GCT
C124S	CTGTGTTATCATGTACT CCT TTCTTGGGACCTC	GAGGTCCCAAAAGAA AGA GTACATGATAACAGG	TGT→TCT
C192A	GCTCTAGGAATCATTCAGTT GCC CTAGGTGTTGCGAG	GGTTTCTGACACACTAG GCG AAACTGAATGATTCC	TGC→GCC
C196A	GTTTGCTTAGGTG GGC AGGTTGGATTGTG	CACAAATCCAACCT GCC CACACCTAGGCCAAAC	TGC→GCC
C260A	GCAGAAATGTTAAAAAACCTCAACG GCT TCCTAGGCCTGCGCC	CCATCAGGCCGACGCCCTAGGG AGC CACGTTGAGGTTTAAC	TGT→GCT
C381A	CATCGCTTGGGATA GCC AACTCCATCGGATC	GATCCGATGGAGTT GCG TATCCCCAAAGCGATG	TGC→GCC
C395A	CTTCCAGACCTTCTCATTT GCC CTCTGTCAGCAGCC	AACGGCTGGAGACAAAGGA GCG GGAAATGGAGAAGGTCTGG	TGC→GCC
C415A	GGAGGGAAAACACAGCTCGCAGGG GCC TTGGCCTCGATG	GCAGAAATCATCAGCGAGGCCA AGGC ACCTGCGAGCTGTGTTTCC	TGC→GCC
C415S	GCTCGCAGGTT CC CTTGCCCTCGCTGATGATTCT	CGAGGCCAA GG AACCTGCGAGCTGTGTTTCC	TGC→TCC
C679A	GTTGGTATTATGTGACTTAGCAGGA GCC AGCCCACAAAGTCG	GTAAGTCGACTTGTG GGC TGGCTCTGCTAAGTACACATAAAATACC	TGC→GCC
DSSG Mutations			
S654A	CAATTTATCGAC GCT GTGGAGTAAAACCTGGCTGTGAT	GGGTTTTACTCCAAC AGC GTCGATAAAATTGACTTGTG	TCT→GCT
S654T	CAATTTATCGAC ACT GTGGAGTAAAACCTGGCTGTGAT	GGGTTTTACTCCAAC AGT GTCGATAAAATTGACTTGTG	TCT→ACT

Table 6: List of mutation primers used in experiments.

C. Sequencing Primers

When plasmid constructs were manipulated (i.e. cloning or mutagenesis), the coding sequence was verified to check for proper and/or improper manipulation. Sequencing primers are listed below:

	Forward Primer	Start Nucleotide
Promoter Primer		
CMV	CGCAAATGGCGGTAGGCGTG	NA
	Forward Primer	Start Nucleotide
<i>Meriones unguiculatus</i> Slc26a5		
gPres0	TTCACATGCACTCCAAA	148
gPres1	CATGTCTGTCACCCTGCT	537
gPres1.5	CTAGAGTTCTTGCTGTG	874
gPres2	TTCCTGCTCCTGCTCG	1,179
gPres3	CGCCCCAATTACTATGC	1,623
gPres4	CCCCCAATAGTCATCAA	1,849
gPres5	AATGCCACACCCACCACA	2,206
<i>Gallus Gallus</i> Slc26a5		
GgA5A	CAGAAGTATTGTGGAG	43
GgA5B	GACAAGTGCCTGATGAAG	452
GgA5C	CTGCCAGTCTCTTCAATT	874
GgA5D	GGTGCACTCTTCAATT	1,258
GgA5E	CCAATAGTGAGTCGTACA	1,658
GgA5F	GGTGTCTGTCTGTATT	2,056
<i>Danio rerio</i> Slc26a5		
DrA5A	CCGTCGCTACACTAATG	28
DrA5B	GTCGTGATGATTGGTGGGG	425
DrA5C	CAACGAACGCTTCAAGAA	843
DrA5D	GAAATTGCAGGTCTTG	1,240
DrA5E	CCTCCATCTACTTCGAA	1,634
DrA5F	AACAGTTGACGTGAAGGT	2,043
<i>Homo sapiens</i> SLC26A2		
HsA2A	CAACATAACGTTTCACCC	22
HsA2B	TTTCTCTGGGTACCTC	453
HsA2C	CTCACCTCCCTCGGACT	823
HsA2D	GGAAATGTATGCCATTGG	1,248
HsA2E	AGTTCACTGCTTGGCTTG	1,651
HsA2F	GCAGAGATTATGAAGCCA	2,012
<i>Homo sapiens</i> SLC26A3		
HsA3A	GGCCAGTGTATTCTACAA	38
HsA3B	TCAAAAGCAGTCCAGAT	436
HsA3C	ATCAGCGCTTCAAAGACA	836
HsA3D	TTGCTGGGCTTATTGGTG	1,235
HsA3E	CATTGGTTCTTAGGCG	1,644
HsA3F	GGATCAAGGTAGATGTGT	2,027
<i>Homo sapiens</i> SLC26A11		
HsA11A	CAGTGGCTGAAGATGGAT	136
HsA11B	TACAGAACATCCCCAGGC	527
HsA11C	GACGATCTCTTACCGA	885
HsA11D	CGCTCTGGCGTGTAAAGA	1,277
HsA11E	GGTTCCAGTACTCTCTA	1,697

Table 7: List of sequencing primers used in experiments.

APPENDIX II: Förster Resonance Energy Transfer

This appendix will serve to explain the methods used in the works Förster Resonance Energy Transfer (FRET) techniques. First background of information about luminescence, quenching, FRET, lifetime, intensity, and fluorescent proteins will help explain the reasons for using the particular FRET methods (for a more in-depth review see [113]. Then, the results of the system calibration will be explained.

I. Background

A. Luminescence

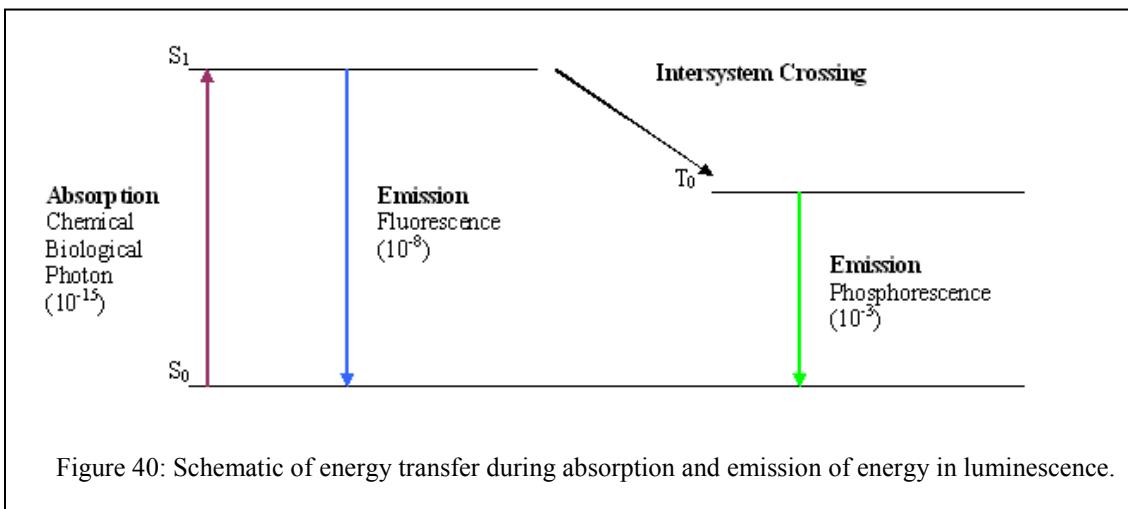
Luminescence is the radiate release of energy from an electron in an excited state in the form of a photon. Electrons can be excited to these higher energy states through a variety of means (*e.g.* chemiluminescence, bioluminescence), but for many applications, the electrons are excited with photons, often from a light source where the excitation wavelengths can be controlled (*e.g.* laser). Luminescence is generally divided into two categories: fluorescence and phosphorescence. Fluorescence occurs when fluorescent molecule's (fluorophore's) electron, which is paired to an electron with opposite spin in the ground state, decays from an excited singlet state. Phosphorescence, on the other hand, occurs when a paired electron's spin is inverted (*i.e.* becomes the same spin as its ground state pair) resulting in the decay from a forbidden triplet state, a process known as intersystem crossing. The emission rate (*i.e.* the amount of photons released per second) is intrinsic to the fluorophore, but, these two processes have different emission rates. The emission rate is usually expressed in terms of quantum yield:

$$Q = \frac{\Gamma}{\Gamma + k_{nr}},$$

where Q is the quantum yield, Γ is the emission rate and k_{nr} is the non-radiative decay rate (*i.e.* rate of energy loss through pathways other than photon release). Quantum yield is expressed as a percentage so if there is no non-radiative decay, then quantum yield is 100% and if the non-radiative decay is equal to the emission rate, then the quantum yield is 50%. Emission rate can also be expressed in terms of lifetime:

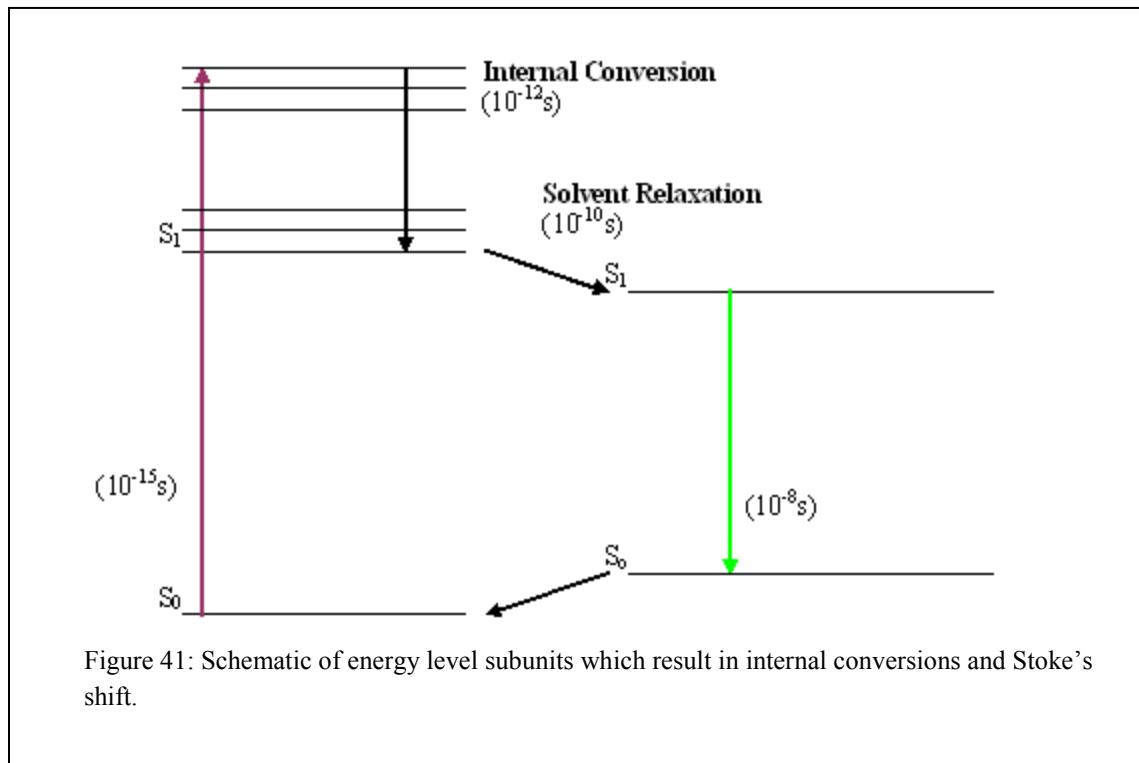
$$\tau = \frac{1}{\Gamma + k_{nr}},$$

where τ is the fluorescent lifetime. Fluorescent lifetime is a random process, but if the decay rate follows a single exponential, τ represents the time when 63% of the fluorophore's electrons have returned to ground state after excitation. The typical lifetime for fluorescence is approximately 10^{-8} (10 ns). Phosphorescence, because the electron enters the forbidden triplet state, takes much longer to decay and are typically 10^{-3} (1 ms) though they can last for up to hours. Luminescence processes can often be represented in schematic diagrams as shown in Figure 40.



Before decaying to ground state, the excited electron may undergo several other processes that result in a loss of energy (Fig. 41). Within any singlet state there are typically several vibration levels. Also, there may exist several higher order singlet states, though the energy difference between these states are often much less than between the first singlet state and the ground state. The energy from the higher energy singlet states and vibration levels, in most cases, rapidly relax ($\leq 10^{-12}$ s) to the lowest vibration level of the first singlet state losing its energy to the environment. This process is known as internal conversion.

Solvents can also change the energy level within singlet states. A fluorophore with an electron in an excited state will typically have a larger dipole. The dipole of the solvent will often rearrange to accommodate the fluorophore dipole. This rearrangement decreases the energy difference between the first singlet state and the ground state. The consequence is that more energy must be absorbed to excite the electron (shorter wavelength light) than energy that is emitted to relax the electron (longer wavelength light), a phenomenon known as Stoke's shift. Several other environmental factors may add to the spectral shift, including hydrogen bonding, charge shift, solvent polarity, probe-probe interactions and conformational changes.



Fluorescence can be disrupted in many ways. During photobleaching, the fluorophore undergo a change, typically the creation or deletion of covalent bond that results in a molecule that can no longer fluoresce. All fluorophores undergo photobleaching, but some fluorophores are more resistant to photobleaching than others. Typically, a fluorophore in the excited state has a larger dipole than the fluorophore in the ground state. This larger dipole provides the energy necessary to form or break covalent bonds. Thus, the rate of photobleaching is directly dependent upon the rate of excitation. Photobleaching can be limited by reducing the amount of excitation photons placed into the system. Of course, this will also correspond with a reduction in emission. So, excitation must be balanced with the ability to create a satisfactory image.

Photobleaching is typically a permanent process and reduces the amount of fluorophores present in a given sample reducing the overall intensity.

B. Quenching

Quenching also reduces the amount of fluorescent emission. Quenching occurs when a quenching molecule comes in contact with the fluorophore. There are two basic types of quenching: static and dynamic. Static quenching occurs when a molecular forms a complex with the fluorophore in the ground state, which results in a nonfluorescent molecule. This loss of fluorescence intensity described by:

$$\frac{F_0}{F} = 1 + K_s[Q],$$

where F_0 and F are the fluorescence intensity in the absence and presence of the quencher, respectively and $[Q]$ is the concentration of the quencher. K_s is the static quenching association constant and is given by:

$$K_s = \frac{[FQ]}{[F][Q]},$$

where $[F]$ and $[Q]$ are the concentrations of the fluorophore and quencher respectively and $[FQ]$ is the concentration of the complex. Static quenching is wholly dependent upon the concentration and ability to associate with the fluorophore. Both photobleaching and static quenching effectively reduce the amount of molecules that fluoresce, but unlike photobleaching, once the quencher is removed, the fluorophore is again able to fluoresce. Another consequence of the action of both photobleaching and static quenching is that the intensity of the system is reduced without changing the lifetime.

Dynamic, also known as collision, quenching affects both intensity and lifetime. Like static quenching, collision quenching occurs when a molecule comes into contact with the fluorophore. Unlike static quenching, this occurs when the fluorophore is in the excited state and results in a non-radiative return of the electron to the ground state. Also, unlike static quenching, the molecule does not form a complex with the fluorophore. This loss of fluorescence is described by:

$$\frac{F_0}{F} = 1 + K_D [Q],$$

where F_0 and F are the fluorescence intensity in the absence and presence of the quencher, respectively, and $[Q]$ is the concentration of the quencher. K_D is the dynamic quenching association constant and is given by:

$$K_s = k_q \tau_o$$

where k_q is the collision quenching constant and τ_o is the fluorophore lifetime in the absence of the quencher. Collision quenching is dependent upon the rate of diffusion and concentration of the molecule. Thus, the longer the fluorescent lifetime, the more time a molecule has to diffuse towards the excited fluorophore. This disparate effect on lifetime is the most common way to distinguish between static and collision quenching.

C. Förster resonance energy transfer (FRET)

Förster resonance energy transfer (FRET) can be considered a variation of collision quenching where energy is transferred through resonance rather than collision.

If one fluorophore (the donor) has a spectral emission overlap with the spectral excitation of another fluorophore (the acceptor), then the energy contained in the donor electron's excited singlet state can be non-radiatively transferred to the acceptor electron's ground singlet state through dipole-dipole resonance, but only if the fluorophores are close enough for the transfer to occur (< 10 nm). Subsequently the acceptor's photon will decay back to the ground singlet state emitting a photon in the acceptor's spectral emission range. This produces an interesting effect that while exciting the donor, the resulting emission is that of the acceptor. Since FRET is a variation of the collision quenching process, both the donor's fluorescence intensity and lifetime are reduced.

Theodor Förster first described the resonance energy transfer in 1948. The efficiency of energy transfer can be described as:

$$E = \frac{1}{\left(\frac{r}{R_0}\right)^6 + 1},$$

where r is equal to the distance between the two fluorophore and R_0 is equal to the distance where FRET is 50% efficient. R_0 is further described as:

$$R_0 = (8.8 \times 10^{-28} \kappa^2 n^{-4} Q_D J)^{1/6},$$

where κ is relative orientation between the fluorophore dipoles, n is the refractive index of the media between the fluorophore dipoles and Q_D is the quantum coefficient of the donor. J is the spectral overlap and can be further described as:

$$J = \int_0^{\infty} f_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda,$$

where f_D is the normalized donor emission spectrum and ε_A is the extinction coefficient of the acceptor. For most FRET experiments κ can be considered to be 2/3 which is the orientation that fluorophores will assume during the resonance energy transfer if allowed to freely rotate [114]. In cases where the fluorophore is not free to rotate, however, κ may become a factor in the interpretation of the results [115]. Different media with different refractive indexes (n) may also affect the FRET efficiency, but are typically minor when considering other factor such as distance and overlap between fluorophores. When choosing fluorophores, the spectral overlap, donor quantum coefficient, and acceptor extinction coefficient all play major roles. The higher the donor quantum coefficient and acceptor extinction coefficient the larger the R_0 , resulting in a better FRET efficiency given the same distance between fluorophores. More spectral overlap between the donor and acceptor fluorophores also results in a larger R_0 , but this may also cause unintended excitation of the acceptor and/or make it difficult to separate the emission spectrums of the two fluorophores.

D. Measuring FRET

Intensity

There are several ways to examine the FRET efficiency. One of the most straight forward methods used is sensitized emission FRET (seFRET). In seFRET, the donor is excited while either the donor or the donor and acceptor emission bandwidth is

monitored. If FRET is occurring, image intensity should decrease in the donor bandwidth while increasing in the acceptor bandwidth. The seFRET method, however, can result in many false positives due to emission cross contamination and auto-fluorescence and thus requires extensive controls. In spectral imaging FRET (siFRET), the entire spectrum of both the donor and acceptor emission is monitored. This allows for better characterization of the emission and proper assignment of donor and acceptor components, however, siFRET requires longer excitation times and photobleaching may complicate the interpretation of the results. Acceptor photobleaching FRET (apFRET) is often used to verify FRET in either of the above methods. In apFRET, the acceptor is photobleached by excitation with a laser tuned to the acceptor's excitation peak. Once photobleached, the acceptor's quenching properties are eliminated and the donor's emission intensity should increase. Photobleaching, however, can also complicate the interpretation due to its disturbance of the local environment such as the creation of reactive oxygen species, which can cause both quenching and photobleaching, and increase in temperature, which increases collision quenching. The FRET efficiency of these methods is calculated using the equation:

$$E = 1 - \frac{I_{DA}}{I_D},$$

where I_{DA} is the intensity of the donor in the presence of the acceptor and I_D is the intensity of the donor in the absence of the acceptor.

Fluorescent Lifetime

All of the aforementioned methods used intensity measurements to calculate FRET. Fluorescent lifetime imaging (FLIM) can also be used as a measurement of FRET and has some distinct advantages. As previously mentioned, fluorescent lifetimes are not affected by collision quenching. Fluorescent lifetimes are also unaffected by photobleaching, so long as enough photons are collected to accurately calculate the fluorescent lifetime. Furthermore, FLIM methods tend to eliminate all contamination of the acceptor emission in the donor lifetime. Thus, interpretations of changes to fluorescent lifetimes are simplified to static quenching affects. The longer an electron remains in the excited state the higher the probability its energy will be released through non-radiative methods via static quenching molecules. Thus, if FRET occurs fluorescent lifetime will decrease, giving a relatively unambiguous result. One of the greatest drawbacks of FLIM, however, is in the specialized equipment required and the associated costs. As with siFRET and seFRET, FLIM-FRET can use acceptor photobleaching to verify a FRET event. The FRET efficiencies in FLIM are calculated using the equation:

$$E = 1 - \frac{\tau_{DA}}{\tau_D},$$

where τ_{DA} is the fluorescent lifetime of the donor in the presence of the acceptor and τ_D is the fluorescent lifetime of the donor in the absence of the acceptor (for further review see [116, 117].

E. Fluorescent proteins

In order to measure *in situ* protein-protein interactions, fluorescent proteins are often used due to their relative ease of attaching to the proteins of interest. A major drawback

of using fluorescent proteins is that the chromophore is contained within a β -barrel structure with a diameter of approximately 4 nm. As a result, the closest two fluorescent protein chromophores can be is slightly greater than 4 nm, which means the highest FRET efficiency that can be expected is 40% [117]. Furthermore, placement of the fluorophore can have profound effects; for example, if one fluorescent protein is attached to an extracellular end of a protein, while the other is placed on an intracellular end of another, the chromophores will be too far separated for FRET to occur even if the proteins interact. These problems can result in a false negative.

II. System Calibration

In this work the mTFP was used as a donor and vYFP was used as an acceptor. The mTFP was chosen because it is specifically engineered to prevent dimeric interaction seen in other fluorescent proteins [118]. In addition, mTFP is brighter and more photostable than other comparable donors. The vYFP was chosen as an acceptor because of its high FRET efficiency with mTFP [81]. FLIM was combined with apFRET to take advantage of both methods. Cells were fixed to avoid changes in pH and cytotoxicity. Fixation, however, may both change the refractive index of the medium and inhibit free rotation of the fluorophore, both of which may affect R_0 . Fixation may also result in additional static quenching by fixing a quenching molecule in proximity to the fluorophore.

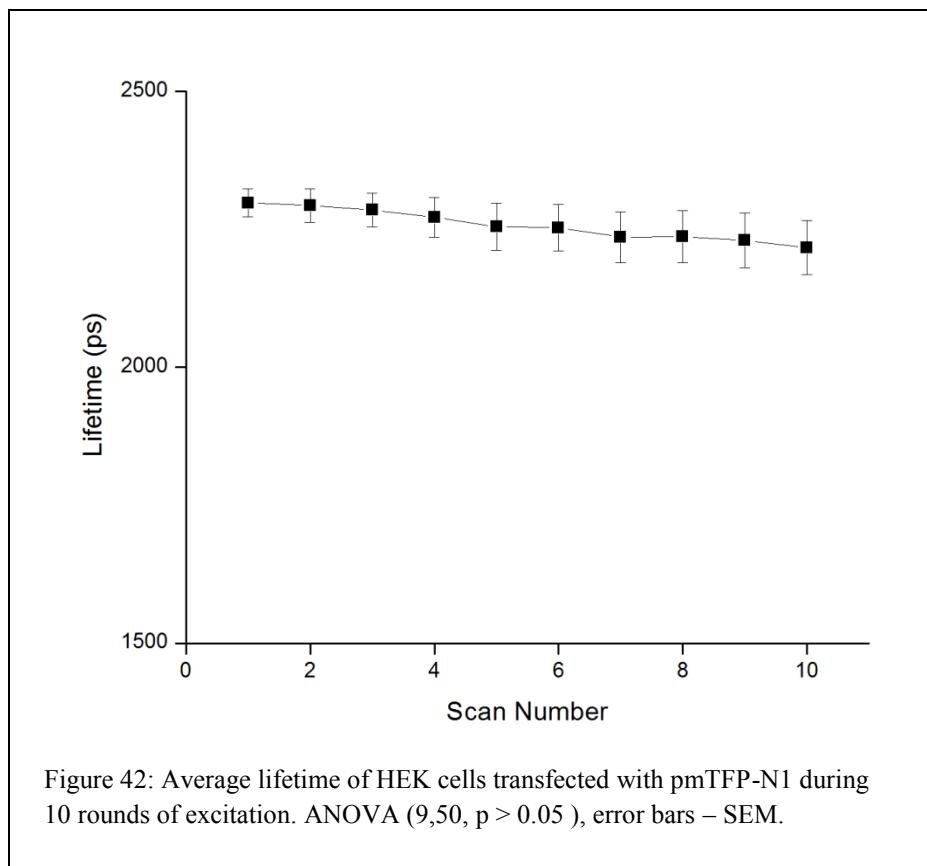
Literature suggests that mTFP fluorophore has a single lifetime (3200 ps) [118]. Our system parameters were analyzed by examining HEK cells transfected with pmTFP-N1 and fixed approximately 48 hours after transfection. HEK cells were imaged on the

CUIBIF Zeiss LSM 510 META NLO confocal microscope. Excitation of the mTFP was performed using 2 photon excitation at 870 nm with the Chameleon Ultra (Coherent Incorporated, Santa Clara, CA). The Becker and Hickl SPCImage software was used to fit the collected photon lifetimes to either single, double, or triple exponential. The χ^2 of each pixel for the different exponential fits was exported and the average χ^2 was calculated for pixels with a photon count peak greater than 125. Three separate cells were analyzed, but the average χ^2 did not improve when multiple exponentials were fit for any of the cells, suggesting mTFP lifetime could be calculated using a single exponential fit (Table 7). The average χ^2 also did not improve with multiple exponential fits after 10 rounds of excitation, suggesting that a single exponential fit for mTFP lifetime is stable during photobleaching.

1st Round Excitation			10th Round Excitation			
	Chi1	Chi2	Chi1	Chi2	Chi3	
T001						
	1.1261	1.1384	1.3618	1.1061	1.1199	1.2780
T002						
	1.0919	1.1058	1.2217	1.0979	1.1114	1.2381
T005						
	1.1037	1.1106	1.2344	1.1058	1.1044	1.1921

Table 8: List of average χ^2 when fitted with a single (Chi1), double (Chi2), or triple (Chi3) exponential on three different cells transfected with pmTFP-N1.

The stability of the mTFP lifetime was calculated, using a single exponential fit, during successive rounds of excitation. Though lifetime shortened and standard error of the mean increased with each scan, there was no significant difference between any scan (Fig. 42). This suggests that the lifetime is stable despite photobleaching that is occurring with each round.



While the mTFP lifetime is stable and fit with a single exponential, the average lifetime (2258 ps) is less than that calculated by Al *et al.* (3200 ps) [118]. This study, however, used a different lifetime acquisition system and living cells. Another study found mTFP lifetimes closer to those seen in this study (2650 ps) when using Becker and Hickl SPCImage, although their analysis was also in live cells [81]. It is possible that the difference in lifetime is due to fixation and different imaging media. The shorter lifetimes may also affect the resolution of the FRET system. However, the stability of the mTFP lifetime suggests that errors due to donor photobleaching should be minimal.

APPENDIX III: Bioinformatic Theory and Methods

This appendix will serve as an explanation and general reference to the bioinformatic tools used in this report. The basic theory of bioinformatic sequence analysis will be explained as well as some of advancements made by the author. Also provided is a more thorough description of the software, databases, and algorithms used in this report. In addition, sequences curated in the various database as well as more detailed results are given.

I. Theory

Bioinformatics serves as a marriage between computer and biological sciences. Bioinformatic techniques are used in various studies including data mining, high-throughput analysis, and genetic compilation. This work utilizes a bioinformatic technique called protein sequence analysis. Protein sequence analysis examines the amino residue sequence of proteins to determine the structure-function relationship. This analysis can be performed using two distinct approaches: 1) analysis of sequence physiochemical properties, or 2) sequence comparison. Analysis of sequence physiochemical properties can be used to determine protein post-translation modification (*e.g.* glycosylation, phosphorylation), molecular binding (*e.g.* hydrogen bonds, disulfide bonds), environment solubility (*e.g.* hydrophobicity, hydrophobicity), secondary structure (*e.g.* stem-loops, α -helix), tertiary structure, protein function (*e.g.* enzymatic activity, structural interaction), and other properties. Sequence comparison typically uses the concept of homology to determine protein structure-function relationships. Certain

aspects of sequence analysis, however, have been ignored by common algorithms and an attempt was made to correct for this throughout the work.

A. Analysis of Physiochemical Properties

Primary Structure

Proteins are composed of amino acids with 20 different types of R groups. Once linked through peptide bonding, the composition and position of the R groups and amino backbone define the function of the protein. The R groups each have their own unique properties, although several properties are shared amongst groups (Fig. 43). When analyzing the physiochemical properties of proteins, it is important to note residues with specific physiological properties. For example, several residues can be post-translationally modified including those with either amine groups, which can be glycosylated, or carboxy groups, which can be glycosylated or phosphorylated. Also of note are three amino acids with particular properties: glycine, proline, and cysteine. Glycine is the smallest amino acid and the only amino acid which is not chiral, allowing greater flexibility around glycine residues. Proline is the only amino acid in which the R group binds to the backbone amine, which creates rigidity in the backbone and often disrupts secondary structure. Cysteine is the only amino acid capable of forming disulfide bonds, which often serve to stabilized protein tertiary and quaternary structure. Less specific properties are also important to note. Hydrophilic R groups, along with the hydrophilic backbone, tend to be arranged so that they are exposed to the more energetically favorable hydrophilic intra- and extracellular environments. These polar and charged residues also typically fold into active sites, where they can destabilize

substrates, or ion channels, where charged molecules can pass. The physiochemical properties along any stretch of the sequence can also indicate sections of the protein exposure to different environments, which can be used to determine higher order structure.

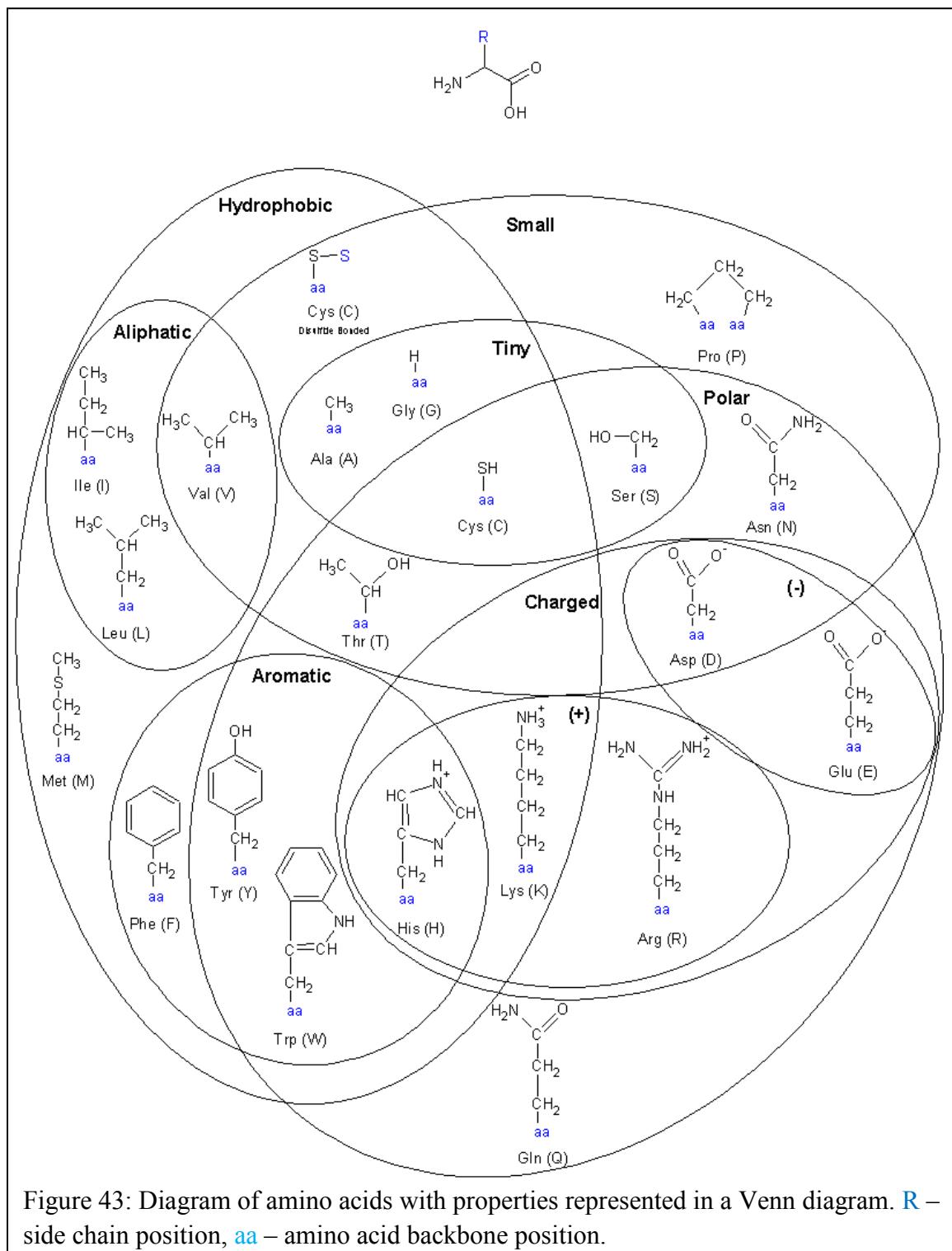


Figure 43: Diagram of amino acids with properties represented in a Venn diagram. R – side chain position, aa – amino acid backbone position.

Secondary Structure

While the primary structure of proteins can be informative, residue interaction is not simply defined by distance along primary structure. Ultimately, residue interaction is defined by the three dimension folding of the protein, which in turn is defined by the formation of the lowest energy structure [119]. This allows the prediction of both secondary and tertiary structure based on the physiochemical properties of the residues within the protein sequence. A major factor in secondary structure folding of proteins is the hydrophobicity of the surrounding environment. The secondary structure of proteins within the hydrophilic environment of the intra- or extracellular space is typically assigned to three common structures: 1) α -helix, 2) β -sheet, or 3) loop/random coils. A protein sequence's secondary structure is predicted based upon the probability and/or physiochemical properties of residues existing within each of the three structures. Secondary structure predictions can be improved by examining other factors, including homology and solvent accessibility [120]. Several different algorithms can be used to analyze protein structure, including SOPMA and PSIPred (see below). These algorithms, however, are not applicable to sections of the protein sequences that are embedded within the plasma membrane.

Proteins within the hydrophobic environment of the plasma membrane tend to form two different types of structures: β -barrels or α -helical bundles [121]. The β -barrel structure forms a cylinder within the plasma membrane composed of β -sheet like secondary structure, which allows the backbone amine and carboxy groups to interact, while hydrophobic residues tend to face the hydrophobic aliphatic fatty acid tails of the plasma membrane and hydrophobic residues face the often hydrophilic protein core.

More often, plasma membrane proteins are composed of α -helical bundles, which are made of α -helix like structures that allow the hydrophilic backbone amine and carboxy groups to interact and hydrophobic residues to face the plasma membrane. Depending on the protein, the α -helical like secondary structure can span the plasma membrane once or many times with variable loop stretches in between. The α -helical bundle proteins can typically be identified based on a relatively highly hydrophobic stretch of residues. Based upon the hydrophobic/hydrophilic profile of these proteins, the position and number of transmembrane helices can be determined. Several algorithms are used to calculate the hydrophobicity profile, such as Kyte-Doolittle, and transmembrane domains, such as TMPred (see below, www.ch.embnet.org/software/TMPRED_form.html) [122]. More recent algorithms take into account additional transmembrane information such as the positive inside rule, re-entrant loops, and sequence alignment [123-125]. The secondary structure of proteins can be used to predict not only characteristics of the protein itself, but also higher order structure.

Tertiary Structure

Predictions of tertiary structure, using the physiochemical properties of the amino residues alone, are often referred to as either *ab initio*, *de novo*, or free modeling. Molecular dynamic studies are usually employed to simulate polypeptide folding energy, typically employing force field rather than quantum mechanics calculations, in different media over time. Several software programs have been written to perform these calculations, including AMBER and CHARMM [126, 127]. While *ab initio* methods promise potential not seen in other methods (*e.g.* the only input necessary is peptide

sequence, physics based calculation of binding energies, moment-by-moment confirmation change) and are still considered the ‘Holy Grail’ of protein tertiary structure prediction, resolution, accuracy, and investment make *ab initio* modeling impractical for most laboratories. More recent programs have begun employing threading (i.e. using peptide fold database as templates) which improves the accuracy of protein modeling while reducing investment of resources [128, 129]. These methods are employed by software such as I-TASSER and RAPTOR [130, 131]. Even more accurate measurements can be made when comparing the sequence of interests to an evolutionarily closely related sequence with known structure, a process which is known as template modeling (TM, see below). TM, however, depends on the availability of solved structures in the Protein Database (PDB, www.pdb.org). The quality of these different methods is evaluated biennially by the Protein Structure Prediction Center in their Critical Assessment of protein Structure Prediction (CASP) experiments [<http://predictioncenter.org/>]. The accuracy of the modeling method can be determined by comparing a model structure to a solved structure. The precision of the model algorithm is calculating using the root mean squared deviation (RMSD) between the atoms of the superimposed modeled and solved structures. The most recent CASP (CASP9) for 2010 is currently underway.

B. Sequence Comparison

Homology

Sequence comparison generally employs sequence homology to determine protein structure-function relationship. Sequence homology, which predicts the evolution of

sequences from common ancestors, is based upon the conservation of residues between sequences. Residue conservation is typically predicted using alignment methods, in which the amino residues of one template sequence is compared to the test sequence-of-interest. The two sequences are considered homologous if greater than 20% of residues share identity (identical amino residues in the same aligned position), a point known as ‘the twilight zone’ [128]. Homologous sequences can subsequently be divided into sequences which are separated by speciation events (orthologs) or by gene duplication events (paralogs). While ortholog function is typically conserved along evolutionary lines, paralogs may often gain or lose function due to different evolutionary pressures. Several algorithms have been created to aid in the determination, retrieval, and classification of homologous sequences and are described below.

Based upon the axiom that protein structure defines function, conservation of structure should result in conservation of function. Thus, residues that are conserved within homologous sequences are predicted to be essential for protein function, while non-conserved residues are thought to be less important for function. This line of reasoning has led to the creation of several algorithms that either compare a sequence-of-interest across a database with known structure-function relationships, or compare a sequence-of-interest across several homologous sequences to determine de novo structure-function relationships (see below). However, evolution, structure, and function all influence residue conservation and careful consideration must be taken when using these algorithms (see below).

Protein sequences are garnered from several sources, including mass spectrometry, complementary DNA, and/or genome. Mass spectrometry allows for the direct sequencing of the polypeptide sequence by analyzing the mass-to-charge (m/z) ratio. In many modern mass spectrometry methods, the m/z ratio is obtained for the parent molecule (the original polypeptide) and several daughter molecules (fractions of the original polypeptide split by gas bombardment). The parent-daughter m/z profile can be compared against a database to determine the original polypeptide sequence. As the mass spectrometry technology advances, with methods such as dual mass spectrometry and quadrupole ion traps, both the measurement capabilities and the sequence accuracy will continue to improve. Protein sequences can also be derived from complementary DNA (cDNA) [132]. The cDNA for a protein-of-interest is derived by reverse transcribing the specific mRNA for the protein-of-interest, or from cDNA libraries, where the entire organisms or tissue specific mRNA is reverse transcribed (*i.e.* expressed sequence tags, ESTs). Since the cDNA is derived from mRNA, only the exons are present and the synthesized protein sequence can be determined based upon the cDNA codons. Of particular benefit, the exon-only property allows cDNA to be cloned into plasmids and expressed in *in vitro* systems where *in vivo* splicing mechanisms may not be present (see Appendix I for further details). However, researchers should be aware of possible alternative splicing (various exon/intron expressions) of mRNA for the protein-of-interest. Protein sequences can also be derived directly from the genome (www.ornl.gov/sci/techresources/Human_Genome/home.shtml). Though the pure size of genomes may make manual searches for the gene encoding the protein-of-interests impractical, advances in bioinformatics have made computer aided gene discovery

simpler. Potential genes can be predicted through such methods as open reading frame (ORF) detection, syntenic genes, and gene homology. Deriving protein sequences from genes, however, is often difficult since the exon/introns pattern is not always known. After the protein sequence is derived, the sequence itself, along with correlative data, is often curated into open source databases.

Several governmental organizations have created open source databases for the curation of genomic and proteomic information. The United States of America established the National Center for Biotechnology Information (NCBI) in 1988 (www.ncbi.nlm.nih.gov). NCBI works in conjunction with European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) and DNA Database of Japan (DDBJ) to collect and curate protein sequences (see below). Along with information about the derivation of proteins sequence, these databases also cross correlate sequences with information such as encoding gene, protein homology, and domain architecture. This information is open access to the public, but, because of the massive amount of associated data, retrieval of sequences can be problematic. One algorithm, known as the Basic Local Alignment Search Tool (BLAST), has come to dominate sequence retrieval [133].

BLAST, and variations of the algorithm, requires only the sequence of input to search across available databases. The algorithm uses local pair-wise alignment to retrieved sequences (see below). BLAST returns are evaluated with an e-value, which describes the probability due to chance that a sequence in the database has better similarity than the returned result (*e.g.* an e-value of $1e^{-3}$ would imply about 1000 similar sequences would be found in a database of $1e^6$) [133]. Several derivatives of BLAST

have been developed for purposes such as weak homology (position specific iteration, PSI-BLAST) and pattern searching (pattern hit initiated, PHI-BLAST) [134, 135]. BLAST, and its various versions, can still return an overwhelming amount of sequences. Curation of these sequences can be performed manually or in local sequence database software, such as CLC Workbench and Interactive Structure based Sequence Alignment Program (STRAP; see below for more detail). The relationship of these retrieved sequences can make a difference in analysis, so it is important to note the sequence information such as source (*e.g.* ‘raw’ sequence such as EST databases, ‘derived sequences’ from genomic databases, or ‘established’ sequences from annotated database), spliceoform (various sequences derived from alternative splicing of the exons/introns of the same gene), and species. Also, careful consideration of aims and objectives should be performed before curating any sequence database (see section C below).

Sequence Alignment

Sequence comparison is typically performed using sequence alignment. Alignment methods compare the linear arrangement of sequences and try to maximize the number of matched nucleotides or residues, while minimizing the number of mismatches or gaps between sequences. Most alignments are based upon a matrix algorithm which places sequences on either X or Y axis and assigns numerical values for matches and mismatches [136]. Algorithm can predict an alignment between the sequences by adding gaps within the sequences while maximizing matches and minimizing mismatches (Fig. 44). When aligning protein sequences, more consideration is given mismatches due to redundancy of codons. This is typically calculated with substitution matrices such as

Point Accepted Mutation (PAM), Block Substitution Matrix (BLOSUM), or hidden Markov model (hMM) matrices [132, 137]. Alignment of the entire sequence (*i.e.* global alignment) works well for closely related sequences [138]. In some cases, however, major portions of sequences (*e.g.* domains) can be gained or lost over the course of evolution and in such cases, alignment of only sections of sequences (*i.e.* local alignment) may be more fruitful [139]. While alignment of two sequences, often referred to as pair-wise sequence alignment, is used in many applications (including BLAST), sequence analysis can be improved by aligning multiple sequences (multiple sequence alignment, MSA). MSA is based on the same matrix based algorithms as pair-wise sequence alignment, but, due to the computation requirements of MSA, heuristic methods are often employed. Heuristic MSA methods typically employ phylogenetic concepts such as guide tree.

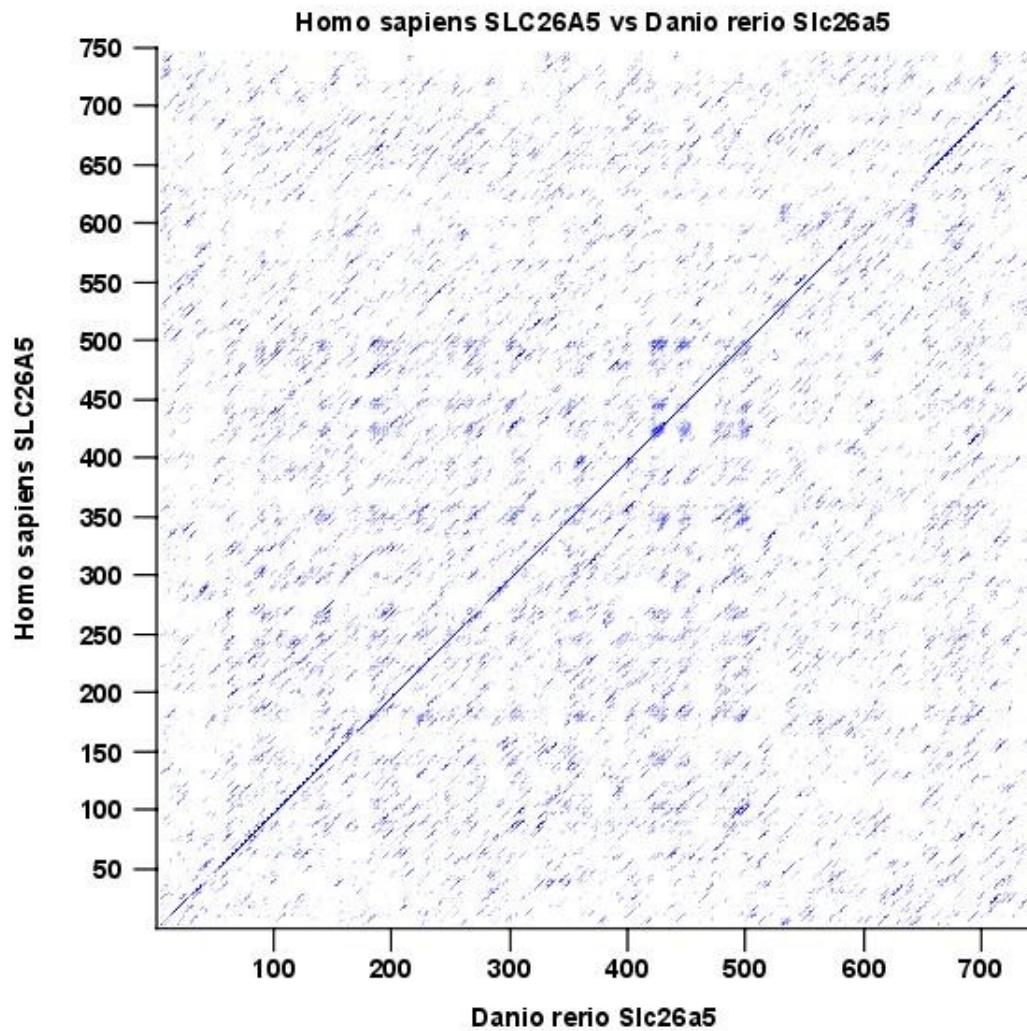


Figure 44: Dot plot of *Homo sapiens* SLC26A5 and *Danio rerio* Slc26a5 protein sequences. Blue dot indicates where residue is equivalent in both sequences.

Phylogeny

Phylogeny examines the evolutionary relationships between groups. This relationship is often represented as a phylogenetic tree, in which a branching diagram represents diversification of individuals from an evolutionarily related ancestor. In sequence analysis, the relationship between a set of sequences is typically calculated based upon pair-wise sequence alignment [136]. The sequence relationship distances are

then used to calculate various branch points based on different algorithmic assumptions (*e.g.* Unweighted Pair Group Method with Arithmetic Means, UPGMA; or Neighbor Joining; see below). This can be followed by an MSA of all sequences and recalculation of evolutionary distances. This process is often iterative to improve phylogeny, but statistical means to calculate accuracy are wanting because of lack of proper tools (*e.g.* mutational rate, generation, and evolutionary pressures are unknown as well as original common ancestors are extinct). One calculation of phylogeny accuracy, bootstrap, iterates the initial phylogenetic analysis of a subgroup of the submitted sequences several times. Bootstrap calculations then determine how many times the original phylogenetic branching is maintained for each iteration.

Motif Finding Software

Motif finding software is based upon the assumption that motifs are contained in areas of high residue conservation. In essence, motif finding software is a local sequence alignment. Small windows of a given sequence (typically 8-24 residues) are compared to several other sequences. Windows with higher rates of residue conservation are returned as putative motifs. As with all alignment algorithms, the level of conservation depends on input and so careful consideration must be taken when choosing sequences for input.

Template Based Modeling

Template based modeling (TBM) is performed by comparing the sequence-of-interest to a sequence with a solved PDB structure. After pair-wise alignment of the two sequences, the PDB sequence is replaced by the sequence-of-interests atom-by-atom

[132]. After atomic replacement, basic molecular dynamic modeling helps improve the accuracy. The template based model can then be evaluated by examining the energy in the structure. One method of evaluation is G_{mean} , which averages several energy evaluations and compares the calculated G_{mean} to other solved structures and also calculates the z-value [140].

C. Advances in Sequence Analysis Theory

A book chapter by the author suggests that there are missing elements in sequence comparison theory [107]. Essentially, sequence comparison attempts to determine conserved structures within a given set of sequences. Conserved sequences are hypothesized to be important for function. This follows from the reasoning that the structure-function relationship of proteins is conserved during evolution. Many sequence comparison algorithms, however, do not take into consideration either evolutionary distance or functional differences between sequences. When neither evolution nor function are taken into account when performing sequence comparison, results vary dramatically depending upon sequences chosen (*e.g.* what is “conserved” in one set of sequences may not be “conserved” in another set of sequences, even if using homologous sequences in both sets). Despite several articles that have echoed this observation, few sequence comparison users take evolution and function into account when performing sequence comparison. In this work, “informed” sequence comparison was used which utilized structure, function, and evolution to curated database for sequence comparison and interprets the results within those elements (Fig. 45).

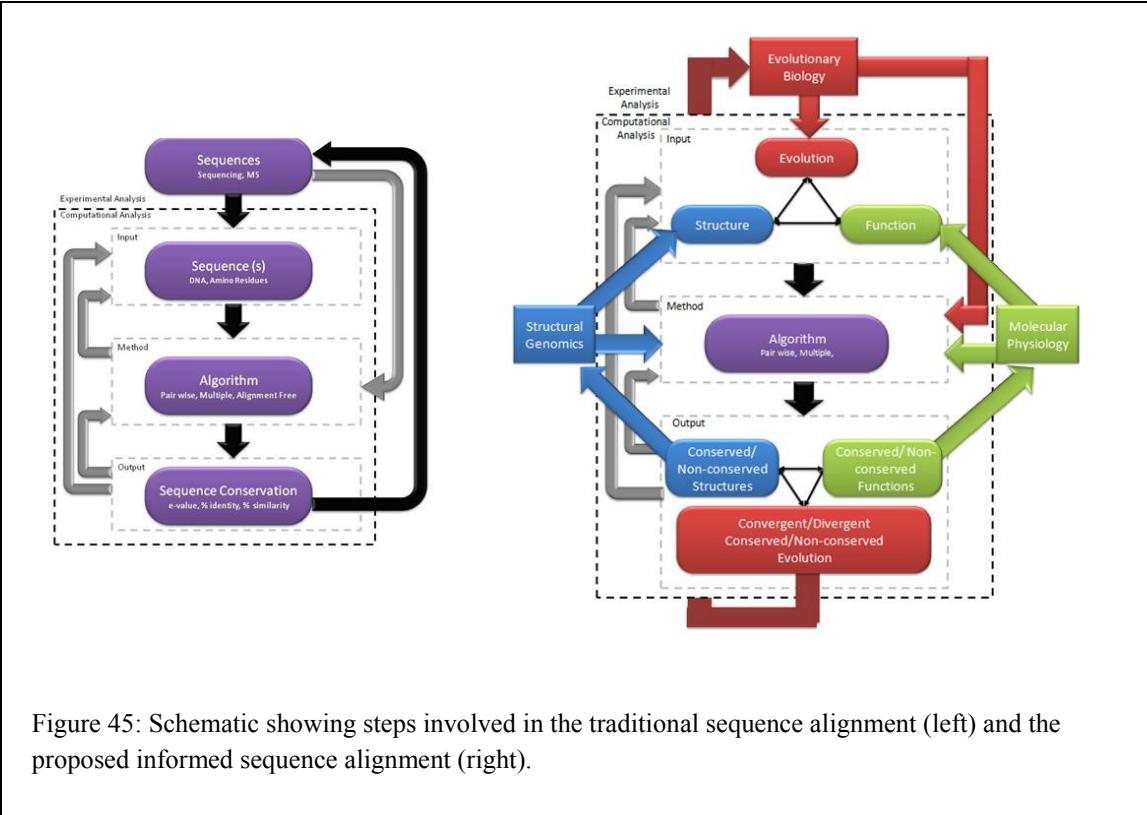


Figure 45: Schematic showing steps involved in the traditional sequence alignment (left) and the proposed informed sequence alignment (right).

II. Software, Databases, and Algorithms

There are an innumerable number of bioinformatic algorithms available for sequence analysis. These algorithms are typically used for either analysis of physiochemical properties or sequence comparison. Analysis of physiochemical properties can be used to determine *ab initio* structure-function based upon protein sequence. Sequence comparison analyzes the composition of different protein sequences to determine conserved structure-function relationships, often coalesced around gateway servers such as the National Center for Biotechnology Information (NCBI), European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), Expert Protein Analysis Systems (ExPASy) and the Bioinformatic Toolkit (BiTk). In addition, these algorithms can be integrated into software packages such as the Interactive

Structure based Sequence Alignment Program (STRAP), Vector NTI, and CLC Workbench. Several of these algorithms were used in this study.

A. Software

Yet Another Scientific Artificial Reality Application (YASARA)

YASARA aims to improve molecular models using all-atom force field algorithm for protein and nucleotide optimization *in vacuo* [141]. YASARA has expanded its applications to include homology modeling, structure assessment, and dynamic simulations. YASARA also incorporates an intuitive graphical user interface (GUI) that allows for researchers to visualize protein database (PDB) files in a variety of ways from any angle. This software was utilized in the examination of tertiary structures.

CLC Main Workbench

CLC Main Workbench is a commercially available program that integrates software for analysis of DNA, RNA, and protein. It utilizes its own custom algorithm to perform multiple sequence analysis and create phylogenetic trees [80].

Interactive Structure based Sequence Alignment Program (STRAP)

STRAP is a freely available program that integrates several proteomic tools including database curation, sequence analysis, homology modeling [142]. One of its most powerful tools is the ability to view several analysis methods and multiple aligned sequences.

B. Databases

Genomic Databases

National Center for Biotechnology Information (NCBI). In essence, the NCBI serves two complimentary roles: 1) as a repository for genetic information and 2) as a collection of tools to analyze genetic information (www.ncbi.nlm.nih.gov).

European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI). Like NCBI, the EMBL-EBI is a collection of both genetic information as well as tools used to curate and analyze this genetic information (www.ebi.ac.uk).

Curated Sequence Databases

Entrez Gene and Entrez Protein. The Entrez Gene database cultivates information about known genes from model organisms [143]. Using this database, several potential prestin homologous sequences were curated based upon prestin associated gene names and identifiers. The database was also used to confirm putative prestin homologous sequences by serving as a gateway to associated information on sequence.

Ensembl. Like NCBI, Ensembl serves both as a repository of genetic information as well as its own set of tools and databases to analyze that information. Ensembl sequence curation, however, revolve around a given set of model species [144]. In addition, many genes are extensively annotated including connections to both paralogs and orthologs.

Algorithm Database

Expert Protein Analysis System (ExPASy). ExPASy both organizes and serves as a gateway to many proteomic databases and tools [145]. It also serves as the host for several of the Swiss Institute of Bioinformatics (SIB) databases.

SWISS-MODEL. SWISS-Model is an online collection of algorithms designed to perform TBM (<http://swissmodel.expasy.org/>). SWISS-Model maintains a repository of TBM or users can create their own TBM based upon alignment of a sequence-of-interest with a solved structure [146, 147].

Max Plank Institute's (MPI's) Bioinformatic Toolkit. The MPI Bioinformatic Toolkit is an intuitive web based collection of several different programs designed to help in the analysis of protein structure-function relationship [148]. The range of tools includes programs from sequence curation to tertiary structure prediction, for which the results of one program can often be fed into another.

Domain and Motif Databases

Interpro. Interpro is a database that collects and organizes ‘signature’ motifs, domains, and families [149]. Sequences-of-interest can be scanned against these signature sequences to determine possible properties and sequence relationships.

Prosite. Like CDD, Prosite serves as a database for protein domains and families [150]. Prosite, however, is based on sequence generalized profiles. These profiles can be used to make hypothesis about important structure-function relationships within domains. In addition, sequences can be scanned across the Prosite database for known motifs as well as domains using ScanProsite. In this study, Prosite was used to delineate domains as well as examine phylogeny between distantly related prestin homologous sequences.

Conserved Domains Database (CDD). The CDD brings together data from several different databases about protein domains [45]. For each domain, several sequences are aligned to create a domain “footprint”. Domains are also annotated with putative function as well as organized into superfamilies. The CDD was used in this information to curate diverse representative sequences from domains of interest as well as to determine distant prestin homologous sequences.

Pfam. Pfam is a database of protein families based on hidden Markov Model (hMM) profiles [46]. The protein families of Pfam encompass entire proteins or domains within proteins and can be used to examine evolution and structure-function relationship. Pfam also classifies protein families into larger, more distantly related clans. This database was used to examine prestin relationship to other known proteins.

Transport Classification Database (TCDB). The TCDB is a database that organizes known transport proteins by both their sequence homology and function [151, 152]. Each family entry is accompanied by a summary of related information, relevant citations, and

curated sequences. TCDB also integrates protein families into larger superfamilies based upon BLAST searches. TCDB was used to further classify prestin related proteins and access protein information.

C. Algorithms

Basic Local Alignment Search Tool (BLAST).

BLAST uses a heuristic pair-wise alignment of short sequences to compare nucleotides or amino residue sequences [153]. This tool can be used to search for homologous sequences in the NCBI repository. In this study, two different BLAST algorithms were used: 1) blastp and 2) PSI-BLAST. Blastp is the basic BLAST algorithm and most often used to find homologous sequences. PSI-BLAST uses position specific interaction to find more distantly related sequences. In essence, this is accomplished by running several iterations of the blastp algorithm, but between each iteration, a position-specific score matrix is produced and serves as the input for the next iteration [134]. Putative prestin homologous sequences were identified by searching the protein database for matches to the prestin reference sequence (NP_945350.1) using both sets of algorithms.

ClustalW and ClustalX

The Clustal family of algorithms is one of the oldest and most used MSAs [154, 155]. Like many other MSA it uses heuristic methods that combine pair-wise alignment and guide trees to complete alignment of entire sequence dataset.

Conserved Domain Architecture Retrieval Tool (CDART).

CDART uses a RPS-BLAST (reverse-position-specific) algorithm and the CDD resources to determine domains within any given sequence [156]. This tool allows the user to visually examine both the architecture (positional relationship between domains) of the protein of interest and other proteins that may share similar architecture. In this study, CDART was used to confirm near prestin homologous sequences as well as search for other sequences containing related domains.

UPGMA

The Unweighted Pair Group Method with Arithmetic Mean is a clustering method used to develop phylogenetic trees [157].

2ZIP

2ZIP is an algorithm designed to identify leucine zipper motifs based on sequence alone [158]. It used coiled-coil finding algorithms along with leucine residue spacing to identify these motifs.

HHpred.

In order to analysis distantly related proteins, sequence of interests were submitted to the HHpred program [159, 160]. HHpred utilizes seed HMM profiles to perform pair-wise alignments and discover remote homology between proteins.

Gibbs Motif Sampling Algorithm (Gibbs)

The Gibbs algorithm was originally defined in 1993 by Lawrence *et al.* for local multiple alignment [161]. The goal of the method is to find “subtle” patterns assumed to be present in a set of varied protein or DNA sequences. It is heralded as a fast method for finding ungapped common patterns in linear time with a relatively high sensitivity. The method requires no prior information and works solely from the input set of data, highlighting two ideas: one, that no expertise or knowledge of secondary or tertiary structures is required, and two, the fact that it is of utmost importance to provide the program with intelligent input. Gaps are not used in Gibbs pattern finding; it is maintained that the need for gaps generally comes from variable loop size in secondary structures (such as in RNA loops) but the active site is conserved, resulting in short conserved patterns. Also allowed for is differential positioning of patterns among sequences, due to genomic rearrangement events. Thus the Gibbs method of pattern detection allows for the discovery of short conserved sequences in DNA or protein sequences that are varied among patterns. Gibbs is best suited for general pattern finding and in performance studies has shown comparable sensitivity and specificity on a position and site-specific level (n_{Sn} , n_{Sp} , s_{Sp}), compared to other tools [162, 163]. Gibbs does not require auxiliary data for run-time and execution under default parameter settings provide optimal results. Issues with the Gibbs detection program, however, includes the inability of the algorithm to find multiple sites within highly conserved sequences. The algorithm instead returns the same site under varying lengths or no sites at all. In addition, if sequences are highly conserved, the algorithm has a tendency to find multiple sites within the sequence that are have some similarity but are clearly two

separate sites. This occurs as a result of normal issues surrounding finding unique sites among very related sequences and the lack of an option to include auxiliary data.

Post-Alignment Identity Matrix (PAIM)

Post-Alignment Identity Matrix (PAIM) is a novel custom algorithm designed to compare conservation between two separate sets of protein sequences. The concept behind PAIM is that comparison between homologous proteins with different functions will reveal residues that are important for the homologous protein family as well as residues that are important for the different function in each protein set. The two protein sets are curated by selecting the most diverse sequences in each functional protein class. These two protein sets undergo MSA separately and then their consensus sequences are PWA against each other. In order to simplify the algorithms and, in future adaptations, take into more sequence information, a custom MSA has been created using vector analysis. Results of this analysis are then organized into five different groups: 1) conserved residue position in both sets with equivalent identity (putatively important for homologous sequence function); 2) conserved residue position in both sets with different identity (putatively important for different function in either set); 3) conserved residue position in set 1, but not set 2 (putatively important for set 1 function); 4) conserved residue position in set 2, but not set 1 (putatively important for set 2 function); and 5) residue position not conserved in either set (not important for function in either set). These results allow analysis of functionally different homologous sequences in a way that no other currently available MSA is capable of.

Uniform Scoring Method

The Uniform Scoring Method (USM) is a novel method developed for this work. It estimates a score for a motif based upon the number of sequences that contain the motif and the frequency of aligned residues within the motif. This is performed with a hash data object, which is similar to an array, but instead of defining positions by number, as in an array, the hash defines positions as keys. A key is a placeholder that points to where data described by the key is held. This is a more efficient data storage object if the entire array does not need to be accessed. For each residue position in the alignment, it is easier to count which residue is most common by inserting the residue itself as a key in the hash – thus, the hash will contain only those residues in that position. This is defined in the pseudocode below:

Let n = number of sequences in total

Let l = the length of the alignment

- (1) For $x = 0$ to l
- (2) Define $\%ResidueHash$ where key = the character of residue _{x}
- (3) For $i = 0$ to n
 - (4) Add Residue _{x,i} to $\%ResidueHash$
- (5) Define the Most Represented Residue (MRR) as the key in $\%ResidueHash$ with the largest count
- (6) $positionalResidue_x = MRR(key)$;
- (7) $positionalScore = MRR(count) / n$;
- (8) Output $positionalResidue_x$, $positionalScore_x$

This returns an identity percentage for the motif and was used with the Gibbs Motif Sampler to rank the motif returns.

III. Curated Databases, PAIM Algorithm, and PAIM Results

A. Curated Databases

Five database types were curated in Chapter IV for use in identifying prestin motifs.

Database Type II-IV are listed in the figures. Type I database sequences are listed below:

Type I Database: Sequences Based on Domain Diversity - APC Clan (II)

Name #	Name	Species	gi #
AAP PF00324	Slc12a3	<i>Mus musculus</i>	GI:81890266
	AMINO ACID PERMEASE	<i>Mycoplasma pulmonis</i>	GI:75544410
	Hypothetical protein	<i>Thermoplasma acidophilum</i>	GI:74576458
	Cationic amino acid transporter	<i>Thermoplasma acidophilum</i>	GI:74576434
	Putative amino acid permease	<i>Streptomyces coelicolor</i>	GI:81789495
	Slc12a9	<i>Homo sapiens</i>	GI:74752435
	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:75028064
	CG4357-PA, isoform A	<i>Drosophila melanogaster</i>	GI:74948526
	Slc12a1	<i>Mus musculus</i>	GI:27151793
AAT PF01490	Na-K-Cl cotransporter	<i>Bos taurus</i>	GI:75058334
	Amino acid permease	<i>Arabidopsis thaliana</i>	GI:75220733
	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:74965509
	Vacuolar amino acid transporter 7	<i>Saccharomyces cerevisiae</i>	GI:731839
	Vacuolar amino acid transporter 5	<i>Saccharomyces cerevisiae</i>	GI:110282938
	Vacuolar amino acid transporter 6	<i>Saccharomyces cerevisiae</i>	GI:731507
	SLC38A3	<i>Homo sapiens</i>	GI:52783419
	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:74957462
	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:74960303
BenE PF03594	Putative amino acid permease	<i>Caenorhabditis elegans</i>	GI:12644223
	N amino acid transport system protein	<i>Neurospora crassa</i>	GI:2507070
	Benzoate membrane transport protein	<i>Deinococcus radiodurans</i>	GI:81552054
	Inner membrane protein ydcO	<i>Escherichia coli K-12</i>	GI:20140524
	Probable transporter	<i>Pseudomonas aeruginosa</i>	GI:81541300
	BenE	<i>Pseudomonas putida</i>	GI:75416939
	CbeE protein	<i>Burkholderia sp. NK8</i>	GI:75411493
	Benzoate membrane transport protein	<i>Acinetobacter sp. ADP1</i>	GI:543886
	Mlr5064 protein	<i>Mesorhizobium loti</i>	GI:81778533
BCAAAT PF05525	Benzoate transport protein	<i>Vibrio cholerae</i>	GI:81544673
	Transport protein	<i>Deinococcus radiodurans</i>	GI:81551299
	Putative branched-chain amino acid transport protein	<i>Streptococcus pyogenes serotype M1</i>	GI:81534000
	Branched-chain amino acid permease	<i>Clostridium acetobutylicum</i>	GI:81530386
	Branched-chain amino acid transporter	<i>Clostridium perfringens</i>	GI:81768052
	Similar to branched-chain amino acid transport system carrier	<i>Staphylococcus aureus subsp. aureus Mu50</i>	GI:81782317
	Branched-chain amino acid transport system carrier	<i>Lactobacillus delbrueckii subsp. lactis</i>	GI:1705503
	Branched chain amino acid transport system II carrier protein	<i>Fusobacterium nucleatum subsp. nucleatum</i>	GI:81483792
	Branched-chain amino acid transport system carrier protein	<i>Fusobacterium nucleatum subsp. nucleatum</i>	GI:81485044
CstA PF02554	Sodium-coupled branched-chain amino acid carrier protein	<i>Clostridium perfringens</i>	GI:81768332
	Sodium-coupled branched-chain amino acid carrier protein	<i>Clostridium perfringens</i>	GI:81767852
	Carbon starvation protein A	<i>Bifidobacterium longum</i>	GI:81754067
	Carbon starvation protein A	<i>Haloquadratum walsbyi DSM 16790</i>	GI:121687016
	Carbon starvation protein CstA	<i>Haloarcula marismortui</i>	GI:74516990
	hypothetical protein	<i>Eubacterium ventriosum ATCC 27560</i>	GI:154484377
	hypothetical protein	<i>Dorea longicatena DSM 13814</i>	GI:153855727
	carbon starvation-induced protein	<i>Algoriphagus sp. PR1</i>	GI:126647224
	hypothetical protein	<i>Clostridium sp. L2-50</i>	GI:160894933
	carbon starvation protein CstA	<i>Halorubrum lacusprofundi ATCC 49239</i>	GI:222481115
	Hypothetical protein	<i>Collinsella aerofaciens ATCC 25986</i>	GI:139439480
	carbon starvation protein CstA	<i>Anaeromyxobacter sp. Fw109-5</i>	GI:153004459

Type I Database: Sequences Based on Domain Diversity - APC Clan (II)

Name #	Name	Species	gi #
ChiQ	Uncharacterized protein MG181	<i>Mycoplasma genitalium</i>	GI:1351502
	Cobalt transport system permease	<i>Synechococcus elongatus PCC 7942</i>	GI:75345165
	Slr1978 protein	<i>Synechocystis sp. PCC 6803</i>	GI:81671163
	Uncharacterized protein MG302	<i>Mycoplasma genitalium</i>	GI:1351537
	Uncharacterized protein MG302 homolog	<i>Mycoplasma pneumoniae</i>	GI:2496372
	Uncharacterized protein MJ1571	<i>Methanocaldococcus jannaschii</i>	GI:37538025
	Hypothetical protein PH0131	<i>Pyrococcus horikoshii</i>	GI:74570453
	Putative HMP/thiamine permease protein ykoC	<i>Bacillus subtilis</i>	GI:81342348
	Uncharacterized protein Mb2352c	<i>Mycobacterium bovis</i>	GI:54040526
HCO3	Putative permease ybaF	<i>Bacillus subtilis</i>	GI:239938826
	SLC4A5	<i>Homo sapiens</i>	GI:182691595
	SLC4A1	<i>Homo sapiens</i>	GI:114787
	Slc4a1	<i>Gallus gallus</i>	GI:114786
	Slc4a9	<i>Rattus norvegicus</i>	GI:29427384
	Slc4a4	<i>Ambystoma tigrinum</i>	GI:82069164
	SLC4A7	<i>Homo sapiens</i>	GI:229462789
	Slc4a8	<i>Mus musculus</i>	GI:81878258
	Na-dependent Cl/HCO3 exchanger	<i>Loligo pealei</i>	GI:74814631
Nramp	HCO3 transporter	<i>Caenorhabditis elegans</i>	GI:74764081
	Anion exchanger 3	<i>Leucoraja erinacea</i>	GI:82095714
	Mn2+/Fe2+ transporter	<i>Lactococcus lactis</i>	GI:122548924
	Putative manganese transpoter	<i>Carboxydothermus hydrogenoformans Z-2901</i>	GI:123576040
	Ethylene signaling protein	<i>Solanum lycopersicum</i>	GI:122194996
	Natural resistance-associated macrophage protein	<i>Ralstonia pickettii 12J</i>	GI:187928750
	Mn2+ and Fe2+ transporter	<i>Burkholderia dolosa AUO158</i>	GI:254254355
	natural resistance-associated macrophage protein	<i>Acidothermus cellulolyticus 11B</i>	GI:117927804
	natural resistance-associated macrophage protein	<i>Acidiphilium cryptum JF-5</i>	GI:148260407
CPT	putative manganese transport protein	<i>Bradyrhizobium sp. BTAi1</i>	GI:148258582
	natural resistance-associated macrophage protein	<i>Pyrobaculum arsenaticum DSM 13514</i>	GI:145590688
	natural resistance-associated macrophage protein	<i>Pyrobaculum calidifontis JCM 11548</i>	GI:126459810
	Uracil permease	<i>Saccharomyces cerevisiae</i>	GI:585161
	Putative purine-cytosine permease yxIA	<i>Bacillus subtilis</i>	GI:81637692
	Purine-cytosine permease FCY22	<i>Saccharomyces cerevisiae</i>	GI:74583809
	Uncharacterized permease C29B12.14c	<i>Schizosaccharomyces pombe</i>	GI:74581949
	Uracil permease	<i>Schizosaccharomyces pombe</i>	GI:19862932
	Putative integral membrane transporter	<i>Streptomyces coelicolor</i>	GI:81344201
NAS	Probable allantoin permease	<i>Bacillus subtilis</i>	GI:2492820
	Putative allantoin permease	<i>Escherichia coli K-12</i>	GI:2498121
	Cytosine permease	<i>Escherichia coli K-12</i>	GI:76364218
	Putative transmembrane transport protein	<i>Streptomyces coelicolor</i>	GI:81345396
	Na(+)-linked D-alanine glycine permease	<i>Pseudoalteromonas haloplanktis</i>	GI:399339
PF02133	Sodium/proton-dependent alanine carrier protein	<i>Bacillus sp. PS3</i>	GI:399023
	Uncharacterized transporter	<i>Haemophilus influenzae</i>	GI:1175322
	Uncharacterized transporter	<i>Haemophilus influenzae</i>	GI:1175473
	Uncharacterized transporter yaaJ	<i>Escherichia coli K-12</i>	GI:401444
	Amino-acid carrier protein AlsT	<i>Bacillus subtilis</i>	GI:2500890

Type I Database: Sequences Based on Domain Diversity - APC Clan (III)

Name #	Name	Species	gi #
SSF PF00474	Slc5a3	<i>Canis lupus familiaris</i>	GI:400353
	Slc5a4	<i>Sus scrofa</i>	GI:400322
	SLC5A1	<i>Homo sapiens</i>	GI:127803
	SLC5A2	<i>Homo sapiens</i>	GI:400337
	Sodium/pantothenate symporter	<i>Escherichia coli K-12</i>	GI:1172008
	Sodium/pantothenate symporter	<i>Haemophilus influenzae</i>	GI:1172009
	Sodium/proline symporter	<i>Haemophilus influenzae</i>	GI:1172774
	Sodium/proline symporter	<i>Escherichia coli K-12</i>	GI:131658
	Uncharacterized symporter ywcA	<i>Bacillus subtilis</i>	GI:239938811
SprG PF03845	Cation/acetate symporter ActP	<i>Escherichia coli K-12</i>	GI:418555
	Spore germination protein	<i>Bacillus halodurans</i>	GI:81786838
	BH2187 protein	<i>Bacillus halodurans</i>	GI:81786677
	Spore germination protein GerlB	<i>Bacillus cereus</i>	GI:75343855
	Spore germination protein	<i>Bacillus halodurans</i>	GI:81787255
	Spore germination protein	<i>Bacillus halodurans</i>	GI:81787142
	Spore germination protein	<i>Clostridium acetobutylicum</i>	GI:81531052
	Germination protein	<i>Bacillus megaterium</i>	GI:75340500
	Spore germination protein B2	<i>Bacillus subtilis</i>	GI:251757271
SulP PF00916	Spore germination protein A2	<i>Bacillus subtilis</i>	GI:1708033
	Spore germination protein yndE	<i>Bacillus subtilis</i>	GI:81669032
	SUL2	<i>Saccharomyces cerevisiae</i>	GI:6094366
	SULX	<i>Saccharomyces cerevisiae</i>	GI:1711596
	Hypothetical protein	<i>Schizosaccharomyces pombe</i>	GI:1175439
	Sulfate permease	<i>Yersinia enterocolitica</i>	GI:75340486
	CARBONATE DEHYDRATASE	<i>Mycobacterium tuberculosis</i>	GI:81671928
	Prs-associated putative membrane protein	<i>Escherichia coli</i>	GI:75491127
	Putative sulfate transporter	<i>Bacillus subtilis</i>	GI:81340951
TctB PF07331	Sulfate transporter	<i>Mycobacterium tuberculosis</i>	GI:81669068
	Low affinity sulfate transporter	<i>Synechocystis sp. PCC 6803</i>	GI:81672393
	Low affinity sulfate transporter	<i>Synechocystis sp. PCC 6803</i>	GI:81672150
	Probable transmembrane protein	<i>Ralstonia solanacearum</i>	GI:81506069
	Uncharacterized 16.3 kDa protein in TAR-I ttuC' 3'region	<i>Agrobacterium vitis</i>	GI:4033502
	Hypothetical protein SMB20026	<i>Sinorhizobium meliloti</i>	GI:81813863
	Blr3453 protein	<i>Bradyrhizobium japonicum</i>	GI:81737995
	hypothetical protein	<i>Agrobacterium tumefaciens str. C58</i>	GI:15887471
	hypothetical protein	<i>Agrobacterium tumefaciens str. C58</i>	GI:16119558
TTP PF03222	Tyrosine permease	<i>Pantoea agglomerans</i>	GI:2492536
	Low affinity tryptophan permease	<i>Escherichia coli K-12</i>	GI:135930
	Tryptophan-specific transport protein	<i>Shigella flexneri</i>	GI:77416582
	Low affinity tryptophan permease	<i>Proteus vulgaris</i>	GI:135931
	Putative tyrosine-specific transport protein	<i>Arabidopsis thaliana</i>	GI:75318027
	Tyrosine Transport	<i>Chlamydophila pneumoniae</i>	GI:81789764
	Tyrosine Transport	<i>Chlamydia trachomatis</i>	GI:81345323
	Tyrosine Transport	<i>Chlamydia trachomatis</i>	GI:81345326
	Tyrosine-specific transport protein	<i>Shigella flexneri</i>	GI:77416686
	Tyrosine-specific transport protein 1	<i>Haemophilus influenzae</i>	GI:1174834
	Tyrosine-specific transport protein 2	<i>Haemophilus influenzae</i>	GI:1174835
	TnaB	<i>Haemophilus influenzae</i>	GI:75425998
	Tryptophan-specific transport protein	<i>Haemophilus influenzae</i>	GI:1171061

Type I Database: Sequences Based on Domain Diversity - APC Clan (IV)

Name #	Name	Species	gi #
XUP PF00860	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:74963224
	Putative purine permease ywdJ	<i>Bacillus subtilis</i>	GI:254763349
	Slc23a3	<i>Mus musculus</i>	GI:81862839
	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:74958007
	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:74958003
	Guanine/hypoxanthine permease pbuO	<i>Bacillus subtilis</i>	GI:81637628
	Putative permease yjcD	<i>Escherichia coli K-12</i>	GI:84028043
	Putative permease	<i>Haemophilus influenzae</i>	GI:1176347
	Uracil permease	<i>Bacillus subtilis</i>	GI:239938688
	Uracil permease	<i>Bacillus caldolyticus</i>	GI:730437

Table 9: Type I database sequences used for APC clan.

Type I Database: Sequences Based on Domain Diversity - ATP-Grasp Clan (I)

Name	#	Name	Species	gi #
ATP-Grasp_1	PF02222	2CZG_B	<i>Pyrococcus horikoshii</i>	GI:88192723
		Phosphoribosylaminoimidazole carboxylase ATPase Subunit	<i>Corynebacterium ammoniagenes</i>	GI:3914498
		Phosphoribosylaminoimidazole carboxylase ATPase Subunit	<i>Mycobacterium leprae</i>	GI:1172768
		Phosphoribosylaminoimidazole carboxylase ATPase Subunit	<i>Mycobacterium tuberculosis</i>	GI:54041756
		Phosphoribosylaminoimidazole carboxylase	<i>Candida glabrata</i>	GI:51338742
		Phosphoribosylaminoimidazole carboxylase ATPase Subunit	<i>Synechocystis sp. PCC 6803</i>	GI:2500017
		Phosphoribosylaminoimidazole carboxylase ATPase Subunit	<i>Pseudomonas aeruginosa</i>	GI:12644682
		1B6S_D	<i>Escherichia coli</i>	GI:6730108
		Phosphoribosylaminoimidazole carboxylase ATPase Subunit	<i>Haemophilus influenzae</i>	GI:1172767
		Phosphoribosylaminoimidazole carboxylase ATPase Subunit	<i>Sulfolobus solfataricus</i>	GI:3914484
ATP-Grasp_2	PF08442	1EUC_B	<i>Sus scrofa</i>	GI:9955020
		Probable succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	<i>Caenorhabditis elegans</i>	GI:1710876
		Succinyl-CoA ligase [GDP-forming] subunit beta, hydrogenosomal	<i>Trichomonas vaginalis</i>	GI:1711579
		Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	<i>Saccharomyces cerevisiae</i>	GI:1711580
		Probable succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	<i>Schizosaccharomyces pombe</i>	GI:21264033
		Probable succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	<i>Caenorhabditis elegans</i>	GI:1710875
		Succinyl-CoA ligase [GDP-forming] subunit beta, hydrogenosomal	<i>Neocallimastix frontalis</i>	GI:9789854
		Succinyl-CoA ligase [ADP-forming] subunit beta-1	<i>Archaeoglobus fulgidus</i>	GI:21264526
		Succinyl-CoA ligase [ADP-forming] subunit beta	<i>Thermus thermophilus</i>	GI:135027
		Succinyl-CoA ligase [ADP-forming] subunit beta-2	<i>Streptomyces coelicolor</i>	GI:21264032
ATP-Grasp_3	PF02655	3DF7_A	<i>Archaeoglobus fulgidus</i>	GI:195927566
		Hypothetical protein	<i>Xanthobacter autotrophicus</i>	GI:75493592
		Hypothetical protein	<i>Pirellula sp.</i>	GI:81659153
		Predicted ATP-utilizing enzymes of the ATP-grasp superfamily	<i>Methanopyrus kandleri</i>	GI:74561410
		Predicted ATP-dependent carbolligase related to biotin carboxylase	<i>Methanopyrus kandleri</i>	GI:74558911
		Uncharacterized protein MJ0776	<i>Methanocaldococcus jannaschii</i>	GI:2833581
		Hypothetical protein	<i>Methanococcus maripaludis</i>	GI:74556275
		Conserved protein	<i>Methanothermobacter</i>	
		Hypothetical protein	<i>thermautotrophicus str. Delta H</i>	GI:74509527
		Conserved protein	<i>Archaeoglobus fulgidus</i>	GI:74511359
			<i>Methanosarcina mazei</i>	GI:74525986
CSPase	PF02786	2C00_B	<i>Pseudomonas aeruginosa</i>	GI:199583513
		PYR1-3	<i>Dictyostelium discoideum</i>	GI:166208504
		CAD protein	<i>Mesocricetus auratus</i>	GI:131696
		Protein ura1	<i>Schizosaccharomyces pombe</i>	GI:1172781
		Acetyl-/propionyl-coenzyme A carboxylase alpha chain	<i>Mycobacterium leprae</i>	GI:13432116
		Acetyl-/propionyl-coenzyme A carboxylase alpha chain	<i>Saccharomyces cerevisiae</i>	GI:1709946
		1BXR_G	<i>Escherichia coli</i>	GI:4929941
		Carbamoyl-phosphate synthase arginine-specific large chain	<i>Bacillus subtilis</i>	GI:254763438
		Carbamoyl-phosphate synthase pyrimidine-specific large chain	<i>Bacillus caldolyticus</i>	GI:1168772
		Carbamoyl-phosphate synthase arginine-specific large chain	<i>Trichosporon cutaneum</i>	GI:1168773
Dala_Dala	PF07478	1IOW_A	<i>Escherichia coli</i>	GI:157831486
		D-alanine--D-alanine ligase	<i>Leuconostoc mesenteroides</i>	GI:3913422
		D-alanine--D-alanine ligase A	<i>Shigella flexneri</i>	GI:62288068
		D-alanine--D-alanine ligase	<i>Enterococcus faecalis</i>	GI:30179787
		D-alanine--D-alanine ligase	<i>Synechocystis sp. PCC 6803</i>	GI:2829607
		D-alanine--D-alanine ligase	<i>Bacillus subtilis</i>	GI:2829831
		D-alanine--D-alanine ligase	<i>Borrelia burgdorferi</i>	GI:3913421
		D-alanine--D-alanine ligase	<i>Aquifex aeolicus</i>	GI:6225246
		D-alanine--D-alanine ligase	<i>Buchnera aphidicola</i>	GI:22654229
		D-alanine--D-alanine ligase	<i>Haemophilus influenzae</i>	GI:1169257

Type I Database: Sequences Based on Domain Diversity - ATP-Grasp Clan (II)

Name #	Name	Species	gi #
DUF1297 PF06973	2R7L_A	<i>Methanocaldococcus jannaschii</i>	GI:161761060
	Hypothetical protein	<i>Sulfolobus solfataricus</i>	GI:74542714
	Hypothetical conserved protein	<i>Candidatus Nitrosocaldus sp.</i>	GI:121679516
	Hypothetical protein	<i>crenarchaeote</i>	GI:74498781
	Hypothetical protein	<i>Archaeoglobus fulgidus</i>	GI:74514408
	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase-like protein	<i>Ignicoccus hospitalis KIN4/I</i>	GI:156937734
	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase-like protein	<i>Pyrobaculum islandicum DSM 4184</i>	GI:119871614
	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase-like protein	<i>Caldivirga maquilingensis IC-167</i>	GI:159040613
	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase-like protein	<i>Metallosphaera sedula DSM 5348</i>	GI:146302801
	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase-like protein	<i>Nitrosopumilus maritimus SCM1</i>	GI:161528278
DUF407 PF04174	Hypothetical protein	<i>Pseudomonas aeruginosa</i>	GI:81539667
	Hypothetical protein	<i>Pseudomonas aeruginosa</i>	GI:81541251
	Mll0717 protein	<i>Mesorhizobium loti</i>	GI:75543924
	Hypothetical protein	<i>Synechococcus sp. PCC 7002</i>	GI:75343164
	Hypothetical protein CC1602	<i>Caulobacter vibrioides</i>	GI:81535435
	Mll2876 protein	<i>Mesorhizobium loti</i>	GI:81779892
	Hypothetical protein MLCB1913.13	<i>Mycobacterium leprae</i>	GI:81556230
	Uncharacterized protein Rv2567/MT2643	<i>Mycobacterium tuberculosis</i>	GI:6226938
	Uncharacterized protein Mb2434c	<i>Mycobacterium bovis</i>	GI:54040527
	Uncharacterized protein sll0335	<i>Synechocystis sp. PCC 6803</i>	GI:2496793
	Hypothetical protein SCO6429	<i>Streptomyces coelicolor</i>	GI:81344208
	Hypothetical protein	<i>Sulfolobus solfataricus</i>	GI:74541643
	Hypothetical protein	<i>Deinococcus radiodurans</i>	GI:81550887
GARS_A PF01071	1GSO_A	Unknown	GI:6730014
	Phosphoribosylamine-glycine ligase	<i>Synechocystis sp. PCC 6803</i>	GI:3024481
	Phosphoribosylamine-glycine ligase	<i>Mycobacterium leprae</i>	GI:3024486
	Phosphoribosylamine-glycine ligase	<i>Mycobacterium tuberculosis</i>	GI:54041753
	Phosphoribosylamine-glycine ligase	<i>Bacillus subtilis</i>	GI:1316111
	Phosphoribosylamine-glycine ligase, chloroplastic	<i>Arabidopsis thaliana</i>	GI:12644306
	Phosphoribosylamine-glycine ligase	<i>Haemophilus influenzae</i>	GI:1172752
	Phosphoribosylamine-glycine ligase	<i>Methanocaldococcus jannaschii</i>	GI:3024488
	Phosphoribosylamine-glycine ligase	<i>Archaeoglobus fulgidus</i>	GI:7994661
	Hypothetical protein	<i>Caenorhabditis elegans</i>	GI:74964353
GSH_S_ATP PF02955	1GSH_A	<i>Escherichia coli B</i>	GI:157831253
	Glutathione synthetase	<i>Gloeobacter violaceus</i>	GI:39931339
	Glutathione synthetase	<i>Thermosynechococcus elongatus BP-1</i>	GI:29611754
	Glutathione synthetase	<i>Nostoc sp. PCC 7120</i>	GI:20141389
	Glutathione synthetase	<i>Synechococcus elongatus PCC 7942</i>	GI:8134483
	Glutathione synthetase	<i>Caulobacter vibrioides</i>	GI:20138320
	Glutathione synthetase	<i>Rhodopseudomonas palustris</i>	GI:47605581
	Glutathione synthetase	<i>Bartonella quintana</i>	GI:81647244
	Glutathione synthetase	<i>Brucella suis</i>	GI:29611773
	Glutathione synthetase	<i>Rhizobium tropici</i>	GI:75402492

Type I Database: Sequences Based on Domain Diversity - ATP-Grasp Clan (III)

Name	#	Name	Species	gi #
Ins134	PF05770	2Q7D_A	<i>Homo sapiens</i>	GI:151567948
		Putative inositol 1,3,4-trisphosphate 5/6-kinase	<i>Oryza sativa</i>	GI:75162282
		Inositol-tetrakisphosphate 1-kinase 1	<i>Arabidopsis thaliana</i>	GI:75202063
		Inositol-tetrakisphosphate 1-kinase 2	<i>Arabidopsis thaliana</i>	GI:83288250
RimK	PF08443	Putative inositol 1,3,4-trisphosphate 5/6-kinase	<i>Oryza sativa</i>	GI:75168629
		Ribosomal protein S6 modification protein	<i>Escherichia coli K-12</i>	GI:84027899
		273aa long hypothetical ribosomal protein S6 modification protein	<i>Pyrococcus horikoshii</i>	GI:74571626
		Uncharacterized protein MG012 homolog	<i>Mycoplasma pneumoniae</i>	GI:2496264
		Gamma-F420-2:alpha-L-glutamate ligase	<i>Methanocaldococcus jannaschii</i>	GI:62906875
STAS	PF01740	Ribosomal protein S6 modification protein	<i>Haemophilus influenzae</i>	GI:1172934
		1H4X_A	<i>Lysinibacillus sphaericus</i>	GI:15826430
		Anti-sigma F factor antagonist	<i>Treponema pallidum</i>	GI:81344737
		Uncharacterized protein Rv1364c/MT1410	<i>Mycobacterium tuberculosis</i>	GI:1723069
		ANTI-ANTI-SIGMA FACTOR RSFB	<i>Mycobacterium tuberculosis</i>	GI:81669742
		Hypothetical protein	<i>Streptomyces lividans</i>	GI:75344027
		Hypothetical protein	<i>Actinosynnema pretiosum subsp. auranticum</i>	GI:75490434
		Uncharacterized protein Rv1365c/MT1411	<i>Mycobacterium tuberculosis</i>	GI:1723070
		Anti-sigma factor antagonist	<i>Mycobacterium tuberculosis</i>	GI:81668904
		Anti-sigma factor antagonist	<i>Mycobacterium tuberculosis</i>	GI:81670091
Synapsin_C	PF02750	Hypothetical protein	<i>Mycobacterium tuberculosis</i>	GI:81342282
		1PK8_B	<i>Rattus norvegicus</i>	GI:47168546
		Synapsin-3	<i>Rattus norvegicus</i>	GI:6707725
		Synapsin-2	<i>Homo sapiens</i>	GI:223634710
TTL	PF03133	Synapsin s-syn-short	<i>Loligo pealei</i>	GI:74764332
		CG32238-PA	<i>Drosophila melanogaster</i>	GI:74871827
		CG4089-PA	<i>Drosophila melanogaster</i>	GI:74869013
		Tubulin glycyclase 3B	<i>Drosophila melanogaster</i>	GI:257096967
		Tubulin monoglycyclase TTLL3	<i>Homo sapiens</i>	GI:172046606
		Tubulin glycyclase 3A	<i>Drosophila melanogaster</i>	GI:74869876
		Putative tubulin-tyrosine ligase	<i>Leishmania major</i>	GI:74827407
		Tubulin--tyrosine ligase	<i>Bos taurus</i>	GI:586134
		Tubulin--tyrosine ligase-like protein 12	<i>Caenorhabditis elegans</i>	GI:6686247
		Tubulin--tyrosine ligase-like protein 12	<i>Homo sapiens</i>	GI:20455527
		TTL domain-containing protein C12B10.04	<i>Schizosaccharomyces pombe</i>	GI:1723548

Table 10: Type I database sequences used for ATP Grasp clan.

B. Post-Alignment Identity Matrix (PAIM)

If we group residues according to unique functionality, for example, we can define a reference group of sequences (exhibiting the unique function) together as a reference sequence set. We can group other sequences exhibiting the normal or shared function together as the reference vector sequence set. We can then do an alignment of each sequence set individually (based on identity or similarity, or ideally perform both) and determine what is 100% conserved within each group. We can represent this result with a Result Vector that contains two objects, the result in sequence form and the result in binary form. In binary form, if the position is 100% conserved, it will be marked as 1, otherwise it will be marked as a 0. In sequence form, if the position is 100% conserved, it will be marked as that residue (or the most frequent residue if examining similarity) or the most frequent residue at that position. Residues that are not 100% conserved are represented as a dash for clarity in Figure 45.

Reference Group	Reference Vector										RV(seq)	RV(bin)
	A	T	V	P	L	M	G	P	R			
A	G	V	P	L	M	K	W	R	-	-	-	-
A	P	M	P	L	M	A	C	R	1	0	0	1
A	V	L	P	L	M	V	C	R	-	-	-	0
1	0	0	1	1	1	0	0	1	-	-	-	1

A	V	Y	W	D	N	E	V	R				
A	V	D	V	D	T	N	I	R	-	-	-	-
A	V	M	G	D	L	Q	A	R	1	1	0	0
A	V	Y	K	D	I	V	T	R	-	-	-	1
1	1	0	0	1	0	0	0	1	-	-	-	-

Figure 45: Step 1 in the new method; creation of reference and test vectors.

We define the Reference Group as a set of n sequences having some structural, functional, or evolutionary difference from the set of m sequences in the Test Group.

$$RG = \begin{bmatrix} n_{i,j} & n_{i+1,j} & n_{i+2,j} \\ n_{i,j+1} & \dots & \dots \\ n_{i,j+2} & \dots & \dots \end{bmatrix}$$

Clustal alignment allows for output of a positional alignment score based on identity or a similarity matrix that can be mapped to the alignment of the Reference Group:

$$RG_{alignment} = \{score_j, score_{j+1}, score_{j+2}, \dots\}$$

Based on some threshold t_x where x is equal to either identity or similarity, and t is equal to the level of x allowable from 0 to 100, we can define the binary Reference Vector (RV) as an object containing the a 1 or a 0 if the alignment score for each position matches or exceeds the threshold t defined by the user:

$$RV_{binary} = [bin_j \ bin_{j+1} \ bin_{j+2} \dots]$$

Where:

$$bin_j = \begin{cases} 0, & score_j < t \\ 1, & score_j \geq t \end{cases}$$

For each position in the Reference Vector object we also can define a characteristic sequence Reference Vector containing the most represented residue(s) if the RV_{bin} for that position is 1, or a null value if the RV_{bin} is 0.

$$RV_{sequence} = [seq_j \ seq_{j+1} \ seq_{j+2} \dots]$$

Where:

$$seq_j = \begin{cases} \emptyset, & score_j < t \\ n_{i,j}, & score_j \geq t \end{cases}$$

The result of this is a vector object containing a sequence representation of only the residues meeting our threshold of identity or similarity and an accompanying binary vector. We create vector objects for both the reference and test sequence sets, resulting in a Reference and a Test Vector. Using the sequence and binary vectors combined, we can then align the Reference and Test Vectors using Clustal (or any other alignment technique) using the same parameters used to align the original test set, where:

$$RV_{binary} = [bin_r \ bin_{r+1} \ \dots]$$

$$TV_{binary} = [bin_t \ bin_{t+1} \ \dots]$$

$$Score = \begin{cases} + + & bin_r = bin_t \\ + - & bin_r = 1, bin_t = 0 \\ - + & bin_r = 0, bin_t = 1 \end{cases}$$

And the score for null values results in “--“ which essentially represents noise. It is necessary to further examine those scores that are a ++ conservation, as this means that for both positions in both vectors there was a conserved residue; however this is not sufficient to assume they are the same residue. Indeed, it is important to know if these residues are the same or different, which can be determined as such:

$$Score = \begin{cases} ++, Same & seq_r = seq_t \\ ++, Diff & seq_r \neq seq_t \end{cases}$$

The score of this analysis will categorize all residues into one of the five following results described below:

++, SAME: The residue in that position is 100% conserved in both vectors and is the same residue in both vectors. This is the result we can see the easiest from normal MSA.

++, DIFF: The residue in that position is 100% conserved in both vectors but is not the same residue. (Particularly interesting in the case of unique function)

+ -: The residue in that position is 100% conserved in the reference vector but not the test vector. Might indicate its importance to the unique function.

- +: The residue in that position is 100% conserved in the test vector but not the reference vector. Might indicate its importance to the shared function.

--: The residue in that position is not conserved in neither the reference vector nor the test vector. It likely has little impact on shared or unique function.

The diagram illustrates a sequence comparison step. On the left is a 5x9 grid representing a sequence alignment between two proteins. The columns are labeled A, -, -, P, L, M, -, -, R. The rows are labeled A, V, +, +, s at the top, and d at the bottom. The grid contains various symbols: A, V, +, +, s in the first row; -, -, +, +, - in the second; -, -, +, +, - in the third; -, -, +, +, - in the fourth; and R, R, +, +, s in the fifth. A diagonal line from the bottom-left (d) to the top-right (R) passes through several shaded cells. To the right of the grid is a large rectangular box containing scoring logic:

```

If RVpos(x) == 1 and RVpos(x) == 1 then
    If RVseq(x) == Rvseq(x) then ++, SAME
    If RVseq(x) != Rvseq(x) then ++, UNIQUE
If RVpos(x) == 1 and RVpos(x) == 0 then +-+
If RVpos(x) == 0 and RVpos(x) == 1 then +--
If RVpos(x) == 0 and RVpos(x) == 0 then --
  
```

Figure 46: Step 2 in our new method explaining how vector comparison is scored

This allows us to form a table of resulting important residues, based upon what function of the proteins one is trying to examine further. By reducing the signal-to-noise ratio, this provides a way to find “important residues” in a high-throughput way, or “residue classification.”

Residues	Grouped to:	Position:
A,R	++S	1
L D	++D	5
P,M	+-	4,6
V	-+	2
(na)	--	3,7,8

C. Detailed PAIM Results

The ++S and ++D PAIM results are indicated in the prestin model in Chapter IV.

The residue numerical position is indicated below.

++S		++D	
<i>H. sapiens</i>	SLC26A5 Residue	<i>H. sapiens</i>	Aligned Homolog Residue
P72		F191	V/L
S86		F278	L/I
A102		G302	A
G112		C415	V/L/I/A
Y114			
F117			
P119			
R130			
G127			
H131			
G135			
F137			
P153			
G199			
K276			
P331			
A362			
Y368			
N372			
Q373			
A377			
G379			
N382			
G407			
L440			
W462			
G487			
D655			

Table 11: List of ++S and ++D residues retrieved from PAIM analysis.

DETAILED METHODS

Laboratory specific methods are listed in this section. Molecular biology and cell culture methods are common to many biological laboratories, but have been specifically adapted to our laboratory. Standardized protocols have been developed to reflect these changes. NLC and FRET methods are explained in the relevant chapters and are not repeated here.

I. Molecular Biology

This section lists solutions and protocols used to make and verify work specific plasmid constructs.

A. Solutions

Protocol: Solutions and Media

Reagents	Preparation
0.5M EDTA	<ul style="list-style-type: none">• 7.3g EDTA (Fisher, AC11843-2500)• QS to 50mL Nanopure H₂O• pH to 8.0, Autoclave, Store at RT
1M Tris-HCl	<ul style="list-style-type: none">• 7.9g Tris-HCl (Fisher, PR-H5121)• QS to 50mL Nanopure H₂O• pH to 8.0, Autoclave, Store at RT
3M Sodium Acetate	<ul style="list-style-type: none">• 12.3g sodium acetate (anhydrous) (Fisher, S210-500)• QS to 50mL• pH to 5.2, Autoclave, Store at RT
10N NaOH	<ul style="list-style-type: none">• 20g NaOH (Fisher, BP359-500)• QS to 50mL• Store at RT
0.5N NaOH	<ul style="list-style-type: none">• 1g NaOH (Fisher, BP359-500)• QS to 50mL• Store at RT
10% SDS (w/v)	<ul style="list-style-type: none">• 5g SDS (Fisher, 03-500-509)• QS to 50mL• pH to 7.2, Store at RT
TE Buffer (Tris, EDTA)	<ul style="list-style-type: none">• 500µL 1M Tris-HCl• 100µL 0.5M EDTA• 49.6mL Nanopure H₂O <p>Store at RT</p>
EB Elution Buffer	<ul style="list-style-type: none">• 500µL 1M Tris-HCl• 49.5mL Nanopure H₂O <p>Store at RT</p>
X-Gal	<ul style="list-style-type: none">• 80mg X-Gal (BCIG) (AB-0479)• 1mL DMF (dimethylformamide) <p>Wrap in foil, Store at -20°C</p>
IPTG	<ul style="list-style-type: none">• 24mg IPTG (Isopropyl β-D-1-thiogalactopyranoside)• 1mL Nanopure H₂O <p>Filter (0.22µm), Store at -20°C</p>
Kanamycin	<ul style="list-style-type: none">• 50mg kanamycin• 1mL Nanopure H₂O <p>Filter (0.22µm), Store at -20°C</p>
Ampicillin	<ul style="list-style-type: none">• 100mg ampicillin• 1mL Nanopure H₂O <p>Filter (0.22µm), Store at -20°C</p>
Chloramphenicol	<ul style="list-style-type: none">• 20mg chloramphenicol• 1mL Nanopure H₂O <p>Filter (0.22µm), Store at -20°C</p>

**Protocol:
Solutions and Media**

Reagents	Preparation
Tetracycline	<ul style="list-style-type: none"> • 20mg tetracycline • 1mL Nanopure H₂O Filter (0.22µm), Store at -20°C
LB Media	<ul style="list-style-type: none"> • 10g Bacto-Tryptone • 5g Yeast Extract • 10g NaCl • 2mL 0.5N NaOH QS to 1L with Nanopure H ₂ O pH to 7.5, Autoclave, Store at 4°C (allow to cool before storage)
LB Agar for Selective Plates	<ul style="list-style-type: none"> • 10g Bacto-Tryptone • 5g Yeast Extract • 10g NaCl • 15g Agar • 2mL 0.5N NaOH QS to 1L with Nanopure H ₂ O pH to 7.5, Autoclave, Cool to 50°C and add 1mL sugar(s) and/or 1mL antibiotic(s), pour into plates, allow to set overnight, Store at 4°C Do not pour down sink (use biohazard waste)
50x TAE Buffer Stock (Tris, Acetic Acid, EDTA)	<ul style="list-style-type: none"> • 242g Tris base • 57.1mL glacial acetic acid • 100mL 0.5M EDTA QS to 1L with Nanopure H ₂ O pH to 8.0 Autoclave
1x TAE Buffer (Tris, Acetic Acid, EDTA)	<ul style="list-style-type: none"> • 0.3L 50x TAE Buffer • 14.7L Nanopure H₂O
100x Dye Stock	<ul style="list-style-type: none"> • 0.2085g Bromophenol Blue • 0.2085g Xylene Cyanol QS to 5mL with NAF H ₂ O
DNA LB (Loading Buffer)	<ul style="list-style-type: none"> • 40µL 100x Dye Stock • 1000µL glycerol • 400µL 50x TAE Buffer Stock • 560µL NAF H₂O Vortex well
1kb DNA Ladder	<ul style="list-style-type: none"> • 100µL 1kb DNA ladder (NEB, N3232L) • 100µL DNA LB • 800µL 1x TAE Buffer

**Protocol:
Solutions and Media**

Reagents	Preparation	
ALS #1 (Alkaline Lysis Solution)	<ul style="list-style-type: none"> • 50mM glucose • 25mM Tris-HCl • 10mM EDTA • 10uLRNase I_f • 46mL Nanopure H₂O 	<ul style="list-style-type: none"> 1.25mL of 2M glucose per 50mL 1.25mL of 1M Tris-HCl per 50mL 1mL of 0.5M EDTA per 50mL 10uLRNase I_f 46mL Nanopure H₂O
ALS #2 (Alkaline Lysis Solution)	<ul style="list-style-type: none"> • 0.2N NaOH • 1% (w/v) SDS • 8.8mL Nanopure H₂O 	<ul style="list-style-type: none"> 200µL of 10N NaOH per 10mL 1000µL of 10% SDS per 10mL 8.8mL Nanopure H₂O
ALS #3 (Alkaline Lysis Solution)	<ul style="list-style-type: none"> • 30mL of 5M KOAc (potassium acetate) per 50mL • 5.75mL of glacial acetic acid per 50mL • 14.25 Nanopure H₂O 	<ul style="list-style-type: none"> 30mL of 5M KOAc (potassium acetate) per 50mL 5.75mL of glacial acetic acid per 50mL 14.25 Nanopure H₂O

Hallworth Laboratory Protocol

Adopted From: Molecular Cloning 3rd Ed. by Sambrook & Russell 2001

Last Modified: 05/29/09

Created by: Benjamin Cumall

Page 3 of 3

B. Protocols

Protocol: DNA Cloning

Procedure:

- 1) PCR amplify PCR insert (see Protocol: Polymerase Chain Reaction).
- 2) Cleanup PCR product (see Protocol: PCR Cleanup).
- 3) Digest PCR product and 2000ng of backbone plasmid (See Protocol: DNA Enzymatic Digestion).
- 4) Run digested insert and backbone on a preparative gel (see Protocol: Agarose Gel Analysis).
- 5) Extract appropriately cut insert and backbone from agarose gel (see Protocol: Agarose Gel Extraction).
- 6) Ligate insert and backbone (see Protocol: DNA Ligation).
- 7) Transform ligation product (see Protocol: Transformation).
- 8) Mini prep 8 colonies from selective plate containing transformants grown overnight at 37°C (see Protocol: Mini Prep).
- 9) Restriction enzyme digest the mini preps to determine if cloning worked (see Protocol: DNA Enzymatic Digestion).
- 10) Submit possible successfully cloned mini preps for sequencing (see Protocol: Sequencing).
- 11) If cloning is successful, then maxi prep appropriate plasmids from mini preps (see Protocol: Maxi Prep).

Protocol: Polymerase Chain Reaction

Before Starting:

- Design primers with correct restriction sites, appropriate overlap, similar annealing temperature, and any other desired elements in order to clone into backbone plasmid (ordered through Creighton Molecular Core Lab).
- Prepare working solutions of primers (10pMoles/ μ L)
- Defrost primers, Phusion HF buffer (NEB, F-530S), and dNTP (NEB, N0447S).

Procedure:

- 1) Use chart below to setup reaction in PCR tube (Fisher, 14-230-225):

Plasmid Template Solution	HF Buffer (μ L)	1x Forward Primer (μ L)	1x Reverse Primer (μ L)	Phusion DNA Polymerase (μ L)	NAF H_2O QS to (μ L)
100ng	10	2	2	0.5	50

Volume of plasmid to be added is determined by dividing the desired amount of plasmid by the concentration of the plasmid solution.

- 2) Place PCR tube into thermocycler, close lid, and run on STD3 protocol.
- 3) Place in 4°C until ready to use.

Protocol: PCR Cleanup

Before Starting:

- Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).
- All centrifugation steps are carried out at 17,900 × g (13,000 rpm) in a conventional tabletop microcentrifuge at room temperature.
- Add 1:250 volume pH indicator I to Buffer PB (i.e., add 120 µl pH indicator I to 30 ml Buffer PB or add 600 µl pH indicator I to 150 ml Buffer PB). The yellow color of Buffer PB with pH indicator I indicates a pH of 7.5.
- Add pH indicator I to entire buffer contents. Do not add pH indicator I to buffer aliquots.
- If the purified PCR product is to be used in sensitive microarray applications, it may be beneficial to use Buffer PB without the addition of pH indicator I.

Procedure:

- 1) Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix.
- 2) Check that the color of the mixture is yellow. If the color of the mixture is orange or violet, add 10 µl of 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.
- 3) Place a QIAquick spin column in a provided 2 ml collection tube.
- 4) To bind DNA, apply the sample to the QIAquick column and centrifuge for 30s.
- 5) Discard flow-through. Place the QIAquick column back into the same tube.
- 6) To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 30s.
- 7) Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min.
IMPORTANT: Residual ethanol from Buffer PE will not be completely removed unless the flow-through is discarded before this additional centrifugation.
- 8) Place QIAquick column in a clean 1.5 ml microcentrifuge tube.
- 9) To elute DNA, add 20 µl Buffer EB to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge.
IMPORTANT: Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA.
- 10) Label and place elutant in 4°C until ready for use.

Protocol: DNA Enzymatic Digestion

Before Starting:

- Identify appropriate restriction enzyme for digestion based upon substrate of interest (www.neb.com is a useful website for this step).
- Identify appropriate buffer, heating conditions and if bovine serum albumin (BSA) is needed for the particular restriction enzyme.
- Set water bath to appropriate temperature (typically 37°C)
- Defrost buffer(s) and BSA (if necessary) by hand or in water bath.
- Place enzyme on ice.

Procedure:

- 1) Use chart below to setup reaction in PCR tube (Fisher, 14-230-225):

	Plasmid Solution	Buffer (μ L)	BSA (if used) (μ L)	Enzyme #1 (μ L)	Enzyme #2 (optional) (μ L)	NAF H_2O QS to (μ L)
For Analysis	100-500ng	1.0	0.1	0.25	0.25	10
For Cloning	500-2000ng	2.5	0.25	1.0	1.0	25

Volume of plasmid to be added is determined by dividing the desired amount of plasmid by the concentration of the plasmid solution.

Note: If more than one reaction is being set up using the same enzymes, then a master mix, containing all components except the plasmid solution can be created.

- 2) Briefly centrifuge tube (at least 1,000rpm for 1sec).
- 3) Place tube in water bath for 1 hour.
- 4) Run on agarose gel to analyze and/or separate plasmid fractions (see Protocol: Agarose Gel Analysis).

Protocol: Agarose Gel Analysis

Before Starting:

- Make TAE Buffer (see Protocol: Solutions & Media)
- Make DNA LB (see Protocol: Solutions & Media)

Preparing Agarose Gel

- 1) Mix 0.7g agarose (GenePure LE Quick Dissolve, Agarose ISC BioExpress, E-3119-125) and 100mL of TAE buffer to make 0.7% agarose gel ($X\mu\text{g}/100\text{mL}$).
Note: For smaller DNA strands make higher percentage gels, for larger DNA strands use lower percentage gels, and for gel extractions use 0.5% gels.
- 2) Heat solution while stirring until solution becomes clear.
- 3) Remove from heat and add 10 μL of ethidium bromide (Fisher, BP1302-10).
- 4) Add gel to tray (Owl Separation Systems, B2) and place appropriate comb (B2-20 for amounts $\leq 15\mu\text{L}$, B2-12 for amounts $\leq 30\mu\text{L}$) in before gel sets.

Loading Agarose Gel

- 1) Premix DNA, DNA LB, and H₂O according to table below (this can be performed on parafilm or in centrifuge tubes).

	Analytical Gel ($\geq 100\text{ng DNA}$)		Preparative Gel ($\geq 1\mu\text{g DNA}$)	
	10 μL	30 μL	10 μL	30 μL
DNA	X μL ($[\text{DNA}]/100$)	X μL ($[\text{DNA}]/100$)	X μL ($[\text{DNA}]/1000$)	X μL ($[\text{DNA}]/1000$)
LB	2 μL	2 μL	2 μL	2 μL
H ₂ O	8-X μL	28-X μL	8-X μL	28-X μL

Where [DNA] is the concentration of DNA sample in ng/ μL .

Note: The amount of DNA suggested is the minimum for procedures, more DNA can be used at investigator's discretion.

- 2) Load 10 μL of 1kb DNA ladder or premix into appropriate wells.
- 3) Close lid and run for 60 minutes (or longer for better separation) at 60mV (or higher for faster runs).
Note: Gel can be run in 4°C cold room or fridge to reduce diffusion.
- 4) Image with dark box analyzer.

Protocol: Agarose Gel Extraction

Before Starting:

- The yellow color of Buffer QG indicates a pH ≤ 7.5 .
- Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).
- All centrifugation steps are carried out at $17,900 \times g$ (13,000 rpm) in a conventional table-top microcentrifuge at room temperature.

Procedure:

- 1) Excise the DNA fragment from the agarose gel with a clean, sharp scalpel. Minimize the size of the gel slice by removing extra agarose.
- 2) Add 600 μ L of Buffer QG.
- 3) Incubate at 50°C for 10 min (or until the gel slice has completely dissolved) inverting every 2-3 minutes.
- 4) After the gel slice has dissolved completely, check that the color of the mixture is yellow. If the color of the mixture is orange or violet, add 10 μ L of 3 M sodium acetate until mixture is yellow.
- 5) Add 200 μ L of isopropanol to the sample and mix.
- 6) Place a QIAquick spin column in a provided 2 ml collection tube.
- 7) To bind DNA, apply the sample to the QIAquick column, and centrifuge for 1 min.
- 8) Discard flow-through and place QIAquick column back in the same collection tube.
- 9) Add 0.5 ml of Buffer QG to QIAquick column and centrifuge for 1 min.
- 10) To wash, add 0.75 ml of Buffer PE to QIAquick column and centrifuge for 1 min.
- 11) Discard the flow-through and centrifuge the QIAquick column for an additional 1 min.
- 12) Place QIAquick column into a clean 1.5 ml microcentrifuge tube.
- 13) To elute DNA, add 30 μ L of Buffer EB to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge for 1 min.
IMPORTANT: Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA.
- 14) Label and place elutant in 4°C until ready for use.

Protocol: Ligation

Before Starting:

- Defrost T4 DNA Ligase buffer (NEB, M0202S) by hand or in water bath.
- Place T4 DNA Ligase (NEB, M0202S) on ice.

Procedure:

- 1) Use chart below to setup reaction in PCR tube (Fisher, 14-230-225):

Insert Plasmid Solution	Backbone Plasmid Solution	Buffer (μ L)	Ligase (μ L)	NAF H_2O QS to (μ L)
300-3000ng	100-500ng	2.0	1.0	20

Volume of plasmid to be added is determined by dividing the desired amount of plasmid by the concentration of the plasmid solution.

- 2) Briefly centrifuge tube (at least 1,000rpm for 1sec).
- 3) Incubate at room temperature for at least 1 hour.
- 4) Run on agarose gel to analyze (see Protocol: Agarose Gel Analysis).
- 5) Transform successful ligation (see Protocol: Transformation).

Protocol: Transformation

Before Starting:

- Make appropriate selective plates (see Protocol: Solutions & Media).
- Place appropriate selective plate in incubator at 37°C.
- Place SOC medium at room temperature.
- Turn heat bath to 42°C.
- If necessary, separate *E. coli* stock into 50µL aliquots into flat-bottomed, thin-walled centrifuge tubes (Fisher, 05-408-138).

Procedure:

- 1) Defrost chemically competent *E. coli* on ice (~5-15min).
Note: High efficiency 10 β *E. coli* (NEB, C3019H) should be used for transforming recovered, ligated or mutated products, while sub-cloning efficiency 5 α *E. coli* (NEB, C2988J) should be used for mini or maxi prep products. For non-methylated plasmid DNA use *dam/dcm* *E. coli* (NEB, C2925I)
- 2) Add at least 50ng of vector into vial of chemically competent *E. coli* and mix gently (do not use more than 5µL of plasmid solution).
- 3) Incubate on ice for 30min
Note: Reducing incubation time reduces transformation efficiency.
- 4) Heat shock cells for 30sec at 42°C without shaking.
Note: Both time and temp are crucial for maximizing transformation efficiency.
- 5) Immediately transfer and incubate on ice for 5min.
- 6) Add 500µL of room temperature SOC medium.
- 7) Cap tube tightly and shake horizontally (200 rpm) at 37°C for 1 hour.
Note: Incubating for less than 1 hour will reduce efficiency, while heating for greater than 1 hour will result in homologous colonies.
- 8) Spread 250µL on pre-warmed selective plate.

**Protocol:
Mini Prep
Plasmid Extraction**

Before Starting:

- Make ALS solutions (see Protocol: Solutions & Media)
- Make LB media (see Protocol: Solutions & Media)
- Make EB (see Protocol: Solutions & Media)
- Thaw antibiotic(s) (see Protocol: Solutions & Media)
- Grow transformed culture until colonies are visible (see Protocol: Transformation)

Procedure:

- 1) Mix 3mL of LB media with 3 μ L of appropriate antibiotic (based upon plasmid resistance) in culture tube.
- 2) Pick individual colony from culture plate with sterile toothpick and place in culture tube.
- 3) Grow cultures at 37°C overnight at 200rpm.
- 4) Transfer 1.5mL of each culture to microfuge tubes.
- 5) Pellet cells at 4°C for 2min at 13,000rpm.
- 6) Remove supernate and invert tubes on paper towel.
- 7) Add 100 μ L of ALS #1 and vortex until pellet is in suspension.
- 8) Add 200 μ L of ALS #2 (made fresh) and mix by inversion (gently so as not to shear DNA).
- 9) Add 150 μ L of ALS #3 and mix by inversion (gently so as not to shear DNA).
- 10) Incubate on ice for 5 min.
- 11) Pellet cell debris at 4°C for 5min at 13,000rpm.
- 12) Transfer supernate to a sterile microfuge tube.
- 13) Add 900 μ L of 100% EtOH (Fisher,) to microfuge tube and mix by inversion.
- 14) Incubate at room temperature for 2min.
- 15) Pellet DNA at 4°C for 2min at 13,000rpm.
- 16) Remove supernate and invert microfuge tubes on paper towel.
- 17) Allow pellet to dry until it becomes transparent (~10min).
- 18) Add 30 μ L of EB and flick until pellet is in suspension.

Protocol: Sequencing

Before Starting:

- Design primers with approximately 15 matching nucleotides for forward sequencing or 15 reverse complementary nucleotides for reverse sequencing (primers can be ordered through Creighton's Molecular Biology Core Facility).
- Dilute working primers to 1pMole/ μ L.
- Defrost working primers.

Procedure:

- 1) Nanodrop plasmid solutions to determine concentration.
- 2) Add enough plasmid to clean microfuge tube and dilute to concentration $\geq 250\text{ng}/\mu\text{L}$ for at least 2 runs ($15\mu\text{L}$ per sequence).
- 3) Add enough working primer to clean centrifuge tube for at least 2 replications ($10\mu\text{L}$ per sequence).
- 4) Fill out "Request Sequencing" form.
- 5) Place samples and form in/on refrigerator in Molecular Biology Core Facility (pass code: 02680)

**Protocol:
Maxi Prep
Plasmid Purification**

Before Starting:

- Add the provided RNase A solution to Buffer P1 before use. Use one vial of RNase A (centrifuge briefly before use) per bottle of Buffer P1, to give a final concentration of 100 µg/ml.
- Add the provided LyseBlue reagent to Buffer P1 and mix before use.
- Check Buffer P2 for SDS precipitation due to low storage temperatures. If necessary, dissolve the SDS by warming to 37°C.
- Pre-chill Buffer P3 at 4°C.

Procedure:

- 1) Add either toothpick or 100µL of culture from mini prep to 1L of LB containing 1mL of appropriate antibody (see Protocol: Solutions & Media).
- 2) Harvest the bacterial cells by centrifugation at 6000 x g for 15 min at 4°C.
Note: If you wish to stop the protocol and continue later, freeze the cell pellets at -20°C.
- 3) Resuspend the bacterial pellet in 10 mL Buffer P1 by vortexing.
- 4) Add 10 mL Buffer P2, mix thoroughly by vigorously inverting the sealed tube 4–6 times, and incubate at room temperature (15–25°C) for 5 min.
- 5) Add 10 mL of chilled Buffer P3, mix immediately and thoroughly by vigorously inverting 4–6 times, and incubate on ice for 20 min.
- 6) Centrifuge at 6,000rpm for 30 min at 4°C.
- 7) Equilibrate QIAGEN-tip 500 by applying 10 mL Buffer QBT.
- 8) Apply the supernatant from step 6 to the QIAGEN-tip by pouring through a cell strainer (Fisher, 08-771-1).
- 9) Wash the QIAGEN-tip with 2 x 30 mL Buffer QC.
- 10) Place QIAGEN-tip into clean 50mL container.
- 11) Elute DNA with 15 mL Buffer QF.
- 12) Precipitate DNA by adding 10.5 mL (0.7 volumes) room-temperature isopropanol to the eluted DNA.
- 13) Mix and centrifuge immediately at 6,000rpm for 30 min at 4°C.
- 14) Carefully decant the supernatant careful not to disturb pellet.
- 15) Wash DNA pellet with 5 mL of room-temperature 70% ethanol
- 16) Centrifuge at 6,000rpm for 10 min.
- 17) Carefully decant the supernatant without disturbing the pellet.
- 18) Air-dry the pellet for 5–10 min, and redissolve the DNA in a suitable EB.

II. Cell Culture

This section includes solutions and protocols used in laboratory specific cell cultures

A. Solutions

Protocol: Solutions and Media for Cell Culture	
Reagents	Preparation
Culture Media	<ul style="list-style-type: none">• 5 mL Fetal bovine serum (FBS) (Invitrogen, 26140-079)• 45 mL of Dulbecco Modified Essential Medium (DMEM)(Invitrogen, 11885-092) Combine in aseptic conditions Store at 4°C
Cryopreservation Media	<ul style="list-style-type: none">• 9 mL Fetal bovine serum (FBS) (Invitrogen, 26140-079)• 1 mL Dimethyl sulfoxide (DMSO) (Fisher, 67-68-5) QS to 50mL Nanopure H ₂ O pH to 8.0, Autoclave, Store at RT
Phosphate Buffered Solution (PBS)	<ul style="list-style-type: none">• 45.6 g NaH₂PO₄·H₂O (dissolve in 1 L distilled H₂O) (Fisher, S369-1)• 166.04 g Na₂HP0₄ anhydrous (dissolve in 2 L distilled H₂O) (Fisher, S374-3)• 1.5 g Thimerosal (Sigma, T5125) Combine solutions and QS to 15 L before adding thimerosal Store at RT
4% Paraformaldehyde in PBS	<ul style="list-style-type: none">• 4 g Paraformaldehyde (Sigma, P6148-500G)• 100 mL PBS Heat until dissolved, filter pH to 7.4, Store at 4°C
Mounting Media	<ul style="list-style-type: none">• 5 mL Glycerol (Fisher, 800687)• 5 mL PBS Combine and stir Store at 4°C

B. Protocols

Protocol: Cell Passage

Before Starting:

Make Culture Medium (DMEM w/10% FBS)
Prewarm Culture Medium
Prewarm 0.25% Trypsin

Passage Cell

- 1) Check flask for confluency (cells should be between 70-90% confluent).
- 2) Aspirate medium from flask.
- 3) Add 2.0mL trypsin and place the flask into incubator until all cells detach when flask is tapped (~1 minute).
- 4) Add 3.0mL of DMEM w/10% FBS to flask.
- 5) Place medium containing suspended cells and place in 15mL conical tube and centrifuge for 3.5min at 850 rpm.
- 6) Remove supernate and resuspend in 1mL medium
- 7) Load resuspended cells into flask containing 5mL of medium to obtain appropriate confluency (e.g. if flask was originally 100% confluent, loading 300 μ L of resuspend cells will result in 30% confluency) and place back in incubator. Note – original flask can be reused if rinsed at least once with 5mL of medium.
- 8) The remaining resuspended cells can be used for plating or other experiments.

**Protocol:
Cell Culture - Cryopreservation**

Before Starting:

- Culture five (5) 75mm culture flasks until ~90% confluent
- Make 10mL Freezing Medium: FBS*, Qualified (Invitrogen: 26140-079) with 10% (v/v) DMSO
- Make 50mL Culture Medium: DMEM, low glucose (Invitrogen: 11885-092) with 10% (v/v) FBS*, Qualified (Invitrogen: 26140-079)
- Fill Cryo-Safe (Fisher: 03-410-495) with 100% isopropanol and chill to 4°C in refrigerator overnight.
- Pre-warm Culture Medium (warm) to 37°C
- Pre-warm 0.25% Trypsin to 37°C
- Pre-chill Culture Medium (cool) to 4°C
- Pre-chill Freezing Medium to 4°C
- Pre-chill Centrifuge to 4°C
- Label and Pre-chill Cryotubes (Fisher: 05-669-64) to 4°C

Cryopreserve Cells

- 1) Check flasks for confluency (cells should be near 90% confluent).
- 2) Aspirate media from flasks.
- 3) Add 5.0 mL trypsin into flasks.
- 4) Place the flasks into incubator until all cells detach when flasks are gently tapped (~1 minute).
- 5) Add 10.0 mL of Culture Medium (warm) to flasks.
- 6) Triturate cells in flasks by pipetting media several times.
- 7) Place media, containing suspended cells, into separate 15 mL conical tube.
- 8) Centrifuge 15 mL conical tubes for 4min at 850 rpm cooled to 4°C.
- 9) From centrifuged 15 mL conical tubes, remove supernate carefully to avoid disturbing pellets.
- 10) Resuspend cells in Culture Medium (cool) by trituration to achieve cell density level of 2×10^6 cells/mL.
- 11) Add 0.5 mL of cell suspension to Cryotubes.
- 12) Add 0.5 mL Freezing Medium to Cryotubes containing cell suspension.
- 13) Label and load Cryotubes into pre-chilled Cryo-Safe.
- 14) Place Cryo-Safe in -70°C overnight.
- 15) Transfer cryotubes to long term storage.

* - FBS should be separated into 50mL aliquots and frozen to avoid freeze thaw cycles and contamination.

Hallworth Laboratory Protocol

Adopted From: Fundamentals in Cell Culture (BCACC), Methods in Molecular Biology, vol 1368 (Springer), and Cell Culture Manual 3rd edition (Sigma)
Created by: Benjamin Curall Modified: 11/01/10

Protocol: Plasmid Transfection

Before Starting:

- Aliquot Opti-MEM®I (Invitrogen 31985-062) in 15 mL conical tubes and store at 4°C.
- Plate cells on 35 x 10 mm (BD Falcon 353801) at least 48 hours prior to transfection (plates should be ≥ 80% confluent at time of transfection; note: if using glass dishes, then pre-coat with collagen).
- Pre-warm Opti-MEM®I to room temperature.

Procedure:

- 1) Dilute 4 µg DNA¹ in Opti-MEM®I to a final volume of 50 µL in 1.5 mL centrifuge tubes for each transfection.²
- 2) Dilute 20 µL Lipofectamine™ (Invitrogen 11668-019) in Opti-MEM®I to a final volume of 50 µL in 1.5 mL centrifuge tubes for each transfection to make Lipofectamine™ master mix.^{3,4,5}
- 3) Incubate tubes at room temperature for ~5 minutes.
- 4) Add 50 µL of Lipofectamine™ master mix into each DNA diluted centrifuge tube to make transfection solution.
- 5) Incubate transfection solution room temperature for ≥ 20 minutes.
- 6) Add transfection solution to plated cells.
- 7) Optional – replace cell culture medium 6-24 hours after transfection.

Cell protein expression is at maximum 24-72 hours after transfection.

¹ To maximize transfection rate, a ratio of ≥ 8 µg : 20 µL DNA : Lipofectamine™ should be used.

² To simplify method, especially with multiple transfections, the same volume of DNA is used in each tube as long as there is ≥ 8 µg DNA for every 20 µL of Lipofectamine™.

³ If performing several transfections, the same centrifuge tube should be used to make a Lipofectamine™ master mix to avoid pipetting errors.

⁴ Lipofectamine™ is toxic to cells and should be adjusted to each new Lipofectamine™ stock to maximize transfection rate while minimizing cell death.

⁵ The amount of Lipofectamine™ may need to be adjusted to cell confluency to maximize transfection rate while minimizing cell death.

**Protocol:
Cell Culture - Fixation**

Before Starting:

- Make 4% PFA
- Make Mounting Medium (MM)

Cryopreserve Cells

- 1) Remove culture media from culture plates.
- 2) Gently add enough 4% PFA to cover cells (~ 1 mL in 35 mm culture dish) to avoid removing cells.
- 3) Let stand for 30 minutes at room temperature.
- 4) Rinse 3 times with PBS (~ 2 mL PBS per rinse).
- 5) Add 200 μ L (for 35 mm dish) of MM.
- 6) Place 30 mm round coverslip (Warner Instruments, 64-1499) over cells avoiding air bubbles.
- 7) Remove excess MM with tissue (Kimwipe, Fisher Scientific 06666A)
- 8) Seal edges of coverslip with rubber cement (Elmer's).
- 9) Store in 4°C and minimize light exposure until ready to image.

ABBREVIATIONS

38 – SLC38A2

A2 – SLC26A2

A3 – SLC26A3

A11 – SLC26A11

AC – Alternating Current

APC – Amino Acid-Polyamine-Organocation

apFLIM – Acceptor Photobleaching FLIM

apFRET – Acceptor Photobleaching FRET

BenE – Benzoate E Transporter

BiTk – Bioinformatic Toolkit

BLAST – Basic Local Alignment Search Tool

BLOSUM – Block Substitution Matrix

BM – Basilar Membrane

CASP – Critical Assessment of Protein Structure Prediction

CDART – Conserved Domain Architecture Tool

CDD – Conserved Domain Database

cDNA – Complimentary DNA

CFTR – Cystic Fibrosis Transport Receptor

COG – Clusters of Orthologous Groups

cP – Chicken Prestin

CPT – Cytosine/Purine, Uracil, Thiamine, Allantoin Permease

CUIBIF – Creighton University Integrated Biomedical Imaging Facility

DC – Direct Current

EBI - European Bioinformatic Institute

eGFP – Enhance Green Fluorescent Protein

EMBL - European Molecular Biology Laboratory

ExPASy – Expert Protein Analysis System

FLIM – Fluorescent Lifetime Imaging

FRET – Förster Resonance Energy Transfer

gP – Gerbil Prestin

HCO3 – HCO_3^- Transporter

HEK – Human Embryonic Kidney

hMM – Hidden Markov Model

HUGO – Human Genome Organization

IHC – Inner Hair Cells

IRES – Internal Ribosome Entry Site
IVS – Intervening Sequence

KO – Knock Out
KI – Knock In

MCS – Multiple Cloning Site
MET – Mechanical Electrical Transduction Channel
MFS – Major Facilitator Superfamily
MPI – Max Plank Institute
MSA – Multiple Sequence Alignment
mTFP – Monomeric Teal Fluorescent Protein

NCBI – National Center for Biotechnology Information
NCS – Nucleobase:Cation Symporter
NLC – Non-Linear Capacitance
NMR – Nuclear Magnetic Resonance
NTP – Nucleotide Triphosphate

OHC – Outer Hair Cells
ORF – Open Reading Frame

PAIM – Post-Alignment Identity Matrix
PAM – Point Accepted Mutation
PAT – Partial Anion Transfer
PCR – Polymerase Chain Reaction
PDB – Protein Database
Pfam – Protein Family Database
PSI – Position Specific Iteration

RC – Resistor Capacitor
RefSeq – Reference Sequence

SCAM – Scanning Cysteine Accessibility Method
seFRET – Sensitized Emission FRET
siFRET – Spectral Imaging FRET
Slc26a – Solute Carrier 26 a protein family
Slc38a – Solute Carrier 38 a protein family
SMI – Single Molecule Imaging
STAS –Sulfate Transporter Anti-Sigma Factor Antagonist
STRAP – Interactive Structure based Sequence Alignment Program
SulP – Sulfate Transporter

TBM – Template Based Modeling
TCDB – Transport Classification Database
TCSPC – Time Correlated Single Photon Counting

TIGR – The Institute for Genomic Research

TIRF – Total Internal Reflection

TMH – Transmembrane Helices

UPGMA – Unweighted Pair Group Method with Arithmetic Mean

USM – Uniform Scoring Method

vYFP – Venus Yellow Fluorescent Protein

XUP – Xanthine Uracil Permease

YASARA - Yet Another Scientific Artificial Reality Application

zP – Zebrafish Prestin

LITERATURE CITED

1. Mohr, P.E., Feldman, J.J., Dunbar, J.L., McConkey-Robbins, A., Niparko, J.K., Rittenhouse, R.K., and Skinner, M.W. (2000). The societal costs of severe to profound hearing loss in the United States. *International journal of technology assessment in health care* *16*, 1120-1135.
2. Honeycutt, A., Dunlap, L., Chen, H., and al Homsi, G. (2004). Economic Costs Associated with Mental Retardation, Cerebral Palsy, Hearing Loss, and Vision Impairment --- United States, 2003. In MMWR, Volume 53. pp. 57-59.
3. (2008). Universal screening for hearing loss in newborns: US Preventive Services Task Force recommendation statement. *Pediatrics* *122*, 143-148.
4. Niskar, A.S., Kieszak, S.M., Holmes, A.E., Esteban, E., Rubin, C., and Brody, D.J. (2001). Estimated prevalence of noise-induced hearing threshold shifts among children 6 to 19 years of age: the Third National Health and Nutrition Examination Survey, 1988-1994, United States. *Pediatrics* *108*, 40-43.
5. Agrawal, Y., Platz, E.A., and Niparko, J.K. (2008). Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999-2004. *Archives of internal medicine* *168*, 1522-1530.
6. Daniel, E. (2007). Noise and hearing loss: a review. *The Journal of school health* *77*, 225-231.
7. Fligor, B.J., and Cox, L.C. (2004). Output levels of commercially available portable compact disc players and the potential risk to hearing. *Ear and hearing* *25*, 513-527.
8. McMahon, C.M., Kifley, A., Rochtchina, E., Newall, P., and Mitchell, P. (2008). The contribution of family history to hearing loss in an older population. *Ear and hearing* *29*, 578-584.
9. Zheng, J., Shen, W., He, D.Z., Long, K.B., Madison, L.D., and Dallos, P. (2000). Prestin is the motor protein of cochlear outer hair cells. *Nature* *405*, 149-155.
10. Liberman, M.C., Gao, J., He, D.Z., Wu, X., Jia, S., and Zuo, J. (2002). Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. *Nature* *419*, 300-304.
11. Dallos, P., Wu, X., Cheatham, M.A., Gao, J., Zheng, J., Anderson, C.T., Jia, S., Wang, X., Cheng, W.H., Sengupta, S., et al. (2008). Prestin-based outer hair cell motility is necessary for mammalian cochlear amplification. *Neuron* *58*, 333-339.
12. Frank, G., Hemmert, W., and Gummer, A.W. (1999). Limiting dynamics of high-frequency electromechanical transduction of outer hair cells. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 4420-4425.
13. Dallos, P., Zheng, J., and Cheatham, M.A. (2006). Prestin and the cochlear amplifier. *The Journal of physiology* *576*, 37-42.
14. Ashmore, J. (2008). Cochlear outer hair cell motility. *Physiological reviews* *88*, 173-210.
15. Kalinec, F., Holley, M.C., Iwasa, K.H., Lim, D.J., and Kachar, B. (1992). A membrane-based force generation mechanism in auditory sensory cells. *Proceedings of the National Academy of Sciences of the United States of America* *89*, 8671-8675.

16. Navaratnam, D., Bai, J.P., Samaranayake, H., and Santos-Sacchi, J. (2005). N-terminal-mediated homomultimerization of prestin, the outer hair cell motor protein. *Biophysical journal* *89*, 3345-3352.
17. Zheng, J., Du, G.G., Anderson, C.T., Keller, J.P., Orem, A., Dallos, P., and Cheatham, M. (2006). Analysis of the oligomeric structure of the motor protein prestin. *The Journal of biological chemistry* *281*, 19916-19924.
18. Wu, X., Currall, B., Yamashita, T., Parker, L.L., Hallworth, R., and Zuo, J. (2007). Prestin-prestin and prestin-GLUT5 interactions in HEK293T cells. *Developmental neurobiology* *67*, 483-497.
19. Mio, K., Kubo, Y., Ogura, T., Yamamoto, T., Arisaka, F., and Sato, C. (2008). The motor protein prestin is a bullet-shaped molecule with inner cavities. *The Journal of biological chemistry* *283*, 1137-1145.
20. Detro-Dassen, S., Schanzler, M., Lauks, H., Martin, I., zu Berstenhorst, S.M., Nothmann, D., Torres-Salazar, D., Hidalgo, P., Schmalzing, G., and Fahlke, C. (2008). Conserved dimeric subunit stoichiometry of SLC26 multifunctional anion exchangers. *The Journal of biological chemistry* *283*, 4177-4188.
21. Pasqualetto, E., Seydel, A., Pellini, A., and Battistutta, R. (2008). Expression, purification and characterisation of the C-terminal STAS domain of the SLC26 anion transporter prestin. *Protein expression and purification* *58*, 249-256.
22. Wang, X., Yang, S., Jia, S., and He, D.Z. (2010). Prestin forms oligomer with four mechanically independent subunits. *Brain Res* *1333*, 28-35.
23. Dallos, P., Popper, A., and Fay, R. (1996). *The Cochlea*, (New York: Springer-Verlag).
24. Oghalai, J.S. (2004). The cochlear amplifier: augmentation of the traveling wave within the inner ear. *Current opinion in otolaryngology & head and neck surgery* *12*, 431-438.
25. Dallos, P. (2008). Cochlear amplification, outer hair cells and prestin. *Current opinion in neurobiology* *18*, 370-376.
26. Manley, G.A. (2001). Evidence for an active process and a cochlear amplifier in nonmammals. *Journal of neurophysiology* *86*, 541-549.
27. He, D.Z., Beisel, K.W., Chen, L., Ding, D.L., Jia, S., Fritzsch, B., and Salvi, R. (2003). Chick hair cells do not exhibit voltage-dependent somatic motility. *The Journal of physiology* *546*, 511-520.
28. Schaechinger, T.J., and Oliver, D. (2007). Nonmammalian orthologs of prestin (SLC26A5) are electrogenic divalent/chloride anion exchangers. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 7693-7698.
29. Hudspeth, A.J. (2008). Making an effort to listen: mechanical amplification in the ear. *Neuron* *59*, 530-545.
30. Fettiplace, R. (2006). Active hair bundle movements in auditory hair cells. *The Journal of physiology* *576*, 29-36.
31. Kennedy, H.J., Evans, M.G., Crawford, A.C., and Fettiplace, R. (2006). Depolarization of cochlear outer hair cells evokes active hair bundle motion by two mechanisms. *J Neurosci* *26*, 2757-2766.
32. Fettiplace, R., Ricci, A.J., and Hackney, C.M. (2001). Clues to the cochlear amplifier from the turtle ear. *Trends in neurosciences* *24*, 169-175.

33. LeMasurier, M., and Gillespie, P.G. (2005). Hair-cell mechanotransduction and cochlear amplification. *Neuron* *48*, 403-415.
34. Cheatham, M.A., Huynh, K.H., Gao, J., Zuo, J., and Dallos, P. (2004). Cochlear function in Prestin knockout mice. *The Journal of physiology* *560*, 821-830.
35. Ashmore, J.F. (1987). A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier. *The Journal of physiology* *388*, 323-347.
36. Santos-Sacchi, J., and Dilger, J.P. (1988). Whole cell currents and mechanical responses of isolated outer hair cells. *Hearing research* *35*, 143-150.
37. Housley, G.D., and Ashmore, J.F. (1992). Ionic currents of outer hair cells isolated from the guinea-pig cochlea. *The Journal of physiology* *448*, 73-98.
38. Dallos, P. (1985). Response characteristics of mammalian cochlear hair cells. *J Neurosci* *5*, 1591-1608.
39. Busch, W., and Saier, M.H., Jr. (2002). The transporter classification (TC) system, 2002. *Critical reviews in biochemistry and molecular biology* *37*, 287-337.
40. Franchini, L.F., and Elgoyen, A.B. (2006). Adaptive evolution in mammalian proteins involved in cochlear outer hair cell electromotility. *Molecular phylogenetics and evolution* *41*, 622-635.
41. Okoruwa, O.E., Weston, M.D., Sanjeevi, D.C., Millemon, A.R., Fritzsch, B., Hallworth, R., and Beisel, K.W. (2008). Evolutionary insights into the unique electromotility motor of mammalian outer hair cells. *Evolution & development* *10*, 300-315.
42. Smith, F.W., Ealing, P.M., Hawkesford, M.J., and Clarkson, D.T. (1995). Plant members of a family of sulfate transporters reveal functional subtypes. *Proceedings of the National Academy of Sciences of the United States of America* *92*, 9373-9377.
43. Dorwart, M.R., Shcheynikov, N., Yang, D., and Muallem, S. (2008). The solute carrier 26 family of proteins in epithelial ion transport. *Physiology (Bethesda, Md)* *23*, 104-114.
44. Weber, T., Gopfert, M.C., Winter, H., Zimmermann, U., Kohler, H., Meier, A., Hendrich, O., Rohbock, K., Robert, D., and Knipper, M. (2003). Expression of prestin-homologous solute carrier (SLC26) in auditory organs of nonmammalian vertebrates and insects. *Proceedings of the National Academy of Sciences of the United States of America* *100*, 7690-7695.
45. Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., et al. (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic acids research* *37*, D205-210.
46. Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. *Nucleic acids research* *38*, D211-222.
47. Masuda, S., Murakami, K.S., Wang, S., Anders Olson, C., Donigian, J., Leon, F., Darst, S.A., and Campbell, E.A. (2004). Crystal structures of the ADP and ATP bound forms of the *Bacillus* anti-sigma factor SpoIIAB in complex with the anti-anti-sigma SpoIIAA. *Journal of molecular biology* *340*, 941-956.

48. Rouached, H., Berthomieu, P., El Kassis, E., Cathala, N., Catherinot, V., Labesse, G., Davidian, J.C., and Fourcroy, P. (2005). Structural and functional analysis of the C-terminal STAS (sulfate transporter and anti-sigma antagonist) domain of the *Arabidopsis thaliana* sulfate transporter SULTR1.2. *The Journal of biological chemistry* *280*, 15976-15983.
49. Zheng, J., Du, G.G., Matsuda, K., Orem, A., Aguinaga, S., Deak, L., Navarrete, E., Madison, L.D., and Dallos, P. (2005). The C-terminus of prestin influences nonlinear capacitance and plasma membrane targeting. *Journal of cell science* *118*, 2987-2996.
50. Dorwart, M.R., Shcheynikov, N., Baker, J.M., Forman-Kay, J.D., Muallem, S., and Thomas, P.J. (2008). Congenital chloride-losing diarrhea causing mutations in the STAS domain result in misfolding and mistrafficking of SLC26A3. *The Journal of biological chemistry* *283*, 8711-8722.
51. Rajagopalan, L., Patel, N., Madabushi, S., Goddard, J.A., Anjan, V., Lin, F., Shope, C., Farrell, B., Lichtarge, O., Davidson, A.L., et al. (2006). Essential helix interactions in the anion transporter domain of prestin revealed by evolutionary trace analysis. *J Neurosci* *26*, 12727-12734.
52. Bai, J.P., Navaratnam, D., Samaranayake, H., and Santos-Sacchi, J. (2006). En block C-terminal charge cluster reversals in prestin (SLC26A5): effects on voltage-dependent electromechanical activity. *Neuroscience letters* *404*, 270-275.
53. Oliver, D., He, D.Z., Klocker, N., Ludwig, J., Schulte, U., Waldegger, S., Ruppertsberg, J.P., Dallos, P., and Fakler, B. (2001). Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein. *Science (New York, N.Y.)* *292*, 2340-2343.
54. Matsuda, K., Zheng, J., Du, G.G., Klocker, N., Madison, L.D., and Dallos, P. (2004). N-linked glycosylation sites of the motor protein prestin: effects on membrane targeting and electrophysiological function. *Journal of neurochemistry* *89*, 928-938.
55. Deak, L., Zheng, J., Orem, A., Du, G.G., Aguinaga, S., Matsuda, K., and Dallos, P. (2005). Effects of cyclic nucleotides on the function of prestin. *The Journal of physiology* *563*, 483-496.
56. Mount, D.B., and Romero, M.F. (2004). The SLC26 gene family of multifunctional anion exchangers. *Pflugers Arch* *447*, 710-721.
57. Bai, J.P., Surguchev, A., Montoya, S., Aronson, P.S., Santos-Sacchi, J., and Navaratnam, D. (2009). Prestin's anion transport and voltage-sensing capabilities are independent. *Biophysical journal* *96*, 3179-3186.
58. Adler, H.J., Belyantseva, I.A., Merritt, R.C., Jr., Frolenkov, G.I., Dougherty, G.W., and Kachar, B. (2003). Expression of prestin, a membrane motor protein, in the mammalian auditory and vestibular periphery. *Hearing research* *184*, 27-40.
59. Karniski, L.P. (2001). Mutations in the diastrophic dysplasia sulfate transporter (DTDST) gene: correlation between sulfate transport activity and chondrodysplasia phenotype. *Human molecular genetics* *10*, 1485-1490.
60. Hoglund, P., Sormaalaa, M., Haila, S., Socha, J., Rajaram, U., Scheurlen, W., Sinaasappel, M., de Jonge, H., Holmberg, C., Yoshikawa, H., et al. (2001). Identification of seven novel mutations including the first two genomic

- rearrangements in SLC26A3 mutated in congenital chloride diarrhea. Human mutation 18, 233-242.
61. Lopez-Bigas, N., Melchionda, S., de Cid, R., Grifa, A., Zelante, L., Govea, N., Arbones, M.L., Gasparini, P., and Estivill, X. (2001). Identification of five new mutations of PDS/SLC26A4 in Mediterranean families with hearing impairment. Human mutation 18, 548.
 62. Taylor, J.P., Metcalfe, R.A., Watson, P.F., Weetman, A.P., and Trembath, R.C. (2002). Mutations of the PDS gene, encoding pendrin, are associated with protein mislocalization and loss of iodide efflux: implications for thyroid dysfunction in Pendred syndrome. The Journal of clinical endocrinology and metabolism 87, 1778-1784.
 63. Makitie, O., Savarirayan, R., Bonafe, L., Robertson, S., Susic, M., Superti-Furga, A., and Cole, W.G. (2003). Autosomal recessive multiple epiphyseal dysplasia with homozygosity for C653S in the DTDST gene: double-layer patella as a reliable sign. Am J Med Genet A 122A, 187-192.
 64. Makela, S., Eklund, R., Lahdetie, J., Mikkola, M., Hovatta, O., and Kere, J. (2005). Mutational analysis of the human SLC26A8 gene: exclusion as a candidate for male infertility due to primary spermatogenic failure. Molecular human reproduction 11, 129-132.
 65. Howitt, S.M. (2005). The role of cysteine residues in the sulphate transporter, SHST1: construction of a functional cysteine-less transporter. Biochimica et biophysica acta 1669, 95-100.
 66. Madden, C., Halsted, M., Meinzen-Derr, J., Bardo, D., Boston, M., Arjmand, E., Nishimura, C., Yang, T., Benton, C., Das, V., et al. (2007). The influence of mutations in the SLC26A4 gene on the temporal bone in a population with enlarged vestibular aqueduct. Archives of otolaryngology--head & neck surgery 133, 162-168.
 67. Kachar, B., Brownell, W.E., Altschuler, R., and Fex, J. (1986). Electrokinetic shape changes of cochlear outer hair cells. Nature 322, 365-368.
 68. Holley, M.C., and Ashmore, J.F. (1988). On the mechanism of a high-frequency force generator in outer hair cells isolated from the guinea pig cochlea. Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character 232, 413-429.
 69. Dallos, P., and Fakler, B. (2002). Prestin, a new type of motor protein. Nature reviews 3, 104-111.
 70. Rybalchenko, V., and Santos-Sacchi, J. (2008). Anion control of voltage sensing by the motor protein prestin in outer hair cells. Biophysical journal 95, 4439-4447.
 71. Albert, J.T., Winter, H., Schaechinger, T.J., Weber, T., Wang, X., He, D.Z., Hendrich, O., Geisler, H.S., Zimmermann, U., Oelmann, K., et al. (2007). Voltage-sensitive prestin orthologue expressed in zebrafish hair cells. The Journal of physiology 580, 451-461.
 72. Haitin, Y., and Attali, B. (2008). The C-terminus of Kv7 channels: a multifunctional module. The Journal of physiology 586, 1803-1810.

73. Ackers, G.K., and Holt, J.M. (2006). Asymmetric cooperativity in a symmetric tetramer: human hemoglobin. *The Journal of biological chemistry* *281*, 11441-11443.
74. He, D.Z., Jia, S., Sato, T., Zuo, J., Andrade, L.R., Riordan, G.P., and Kachar, B. (2010). Changes in plasma membrane structure and electromotile properties in prestin deficient outer hair cells. *Cytoskeleton (Hoboken, N.J)* *67*, 43-55.
75. Porra, V., Bernier-Valentin, F., Trouttet-Masson, S., Berger-Dutrieux, N., Peix, J.L., Perrin, A., Selmi-Ruby, S., and Rousset, B. (2002). Characterization and semiquantitative analyses of pendrin expressed in normal and tumoral human thyroid tissues. *The Journal of clinical endocrinology and metabolism* *87*, 1700-1707.
76. Organ, L.E., and Raphael, R.M. (2009). Lipid lateral mobility in cochlear outer hair cells: regional differences and regulation by cholesterol. *J Assoc Res Otolaryngol* *10*, 383-396.
77. Accardi, A., and Picollo, A. (2010). CLC channels and transporters: proteins with borderline personalities. *Biochimica et biophysica acta* *1798*, 1457-1464.
78. Armstrong, C.M., and Bezanilla, F. (1974). Charge movement associated with the opening and closing of the activation gates of the Na channels. *The Journal of general physiology* *63*, 533-552.
79. Santos-Sacchi, J. (1991). Reversible inhibition of voltage-dependent outer hair cell motility and capacitance. *J Neurosci* *11*, 3096-3110.
80. Feng, D.F., and Doolittle, R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of molecular evolution* *25*, 351-360.
81. Day, R.N., Booker, C.F., and Periasamy, A. (2008). Characterization of an improved donor fluorescent protein for Forster resonance energy transfer microscopy. *Journal of biomedical optics* *13*, 031203.
82. Rajagopalan, L., Greeson, J.N., Xia, A., Liu, H., Sturm, A., Raphael, R.M., Davidson, A.L., Oghalai, J.S., Pereira, F.A., and Brownell, W.E. (2007). Tuning of the outer hair cell motor by membrane cholesterol. *The Journal of biological chemistry* *282*, 36659-36670.
83. Koushik, S.V., and Vogel, S.S. (2008). Energy migration alters the fluorescence lifetime of Cerulean: implications for fluorescence lifetime imaging Forster resonance energy transfer measurements. *Journal of biomedical optics* *13*, 031204.
84. Kakehata, S., and Santos-Sacchi, J. (1996). Effects of salicylate and lanthanides on outer hair cell motility and associated gating charge. *J Neurosci* *16*, 4881-4889.
85. Lim, D.J., and Kalinec, F. (1998). Cell and molecular basis of hearing. *Kidney Int Suppl* *65*, S104-113.
86. Rettig, M.P., Orendorff, C.J., Campanella, E., and Low, P.S. (2001). Effect of pH on the self-association of erythrocyte band 3 in situ. *Biochimica et biophysica acta* *1515*, 72-81.
87. Perlman, D.F., and Goldstein, L. (2004). The anion exchanger as an osmolyte channel in the skate erythrocyte. *Neurochemical research* *29*, 9-15.

88. Thoden, J.B., Holden, H.M., and Firestone, S.M. (2008). Structural analysis of the active site geometry of N5-carboxyaminoimidazole ribonucleotide synthetase from Escherichia coli. *Biochemistry* *47*, 13346-13353.
89. Aravind, L., and Koonin, E.V. (2000). The STAS domain - a link between anion transporters and antisigma-factor antagonists. *Curr Biol* *10*, R53-55.
90. Shibagaki, N., and Grossman, A.R. (2006). The role of the STAS domain in the function and biogenesis of a sulfate transporter as probed by random mutagenesis. *The Journal of biological chemistry* *281*, 22964-22973.
91. McGuire, R.M., Liu, H., Pereira, F.A., and Raphael, R.M. (2010). Cysteine mutagenesis reveals transmembrane residues associated with charge translocation in prestin. *The Journal of biological chemistry* *285*, 3103-3113.
92. Bai, J.P., Surguchev, A., Bian, S., Song, L., Santos-Sacchi, J., and Navaratnam, D. (2010). Combinatorial cysteine mutagenesis reveals a critical intramonomer role for cysteines in prestin voltage sensing. *Biophysical journal* *99*, 85-94.
93. Kim, D., Kim, S.W., Choi, K.Y., Lee, J.S., and Kim, E. (2008). Molecular cloning and functional characterization of the genes encoding benzoate and p-hydroxybenzoate degradation by the halophilic Chromohalobacter sp. strain HS-2. *FEMS microbiology letters* *280*, 235-241.
94. Koukaki, M., Vlanti, A., Goudela, S., Pantazopoulou, A., Gioule, H., Tournaviti, S., and Diallinas, G. (2005). The nucleobase-ascorbate transporter (NAT) signature motif in UapA defines the function of the purine translocation pathway. *Journal of molecular biology* *350*, 499-513.
95. Wang, Y., Mackenzie, B., Tsukaguchi, H., Weremowicz, S., Morton, C.C., and Hediger, M.A. (2000). Human vitamin C (L-ascorbic acid) transporter SVCT1. *Biochemical and biophysical research communications* *267*, 488-494.
96. Gournas, C., Papageorgiou, I., and Diallinas, G. (2008). The nucleobase-ascorbate transporter (NAT) family: genomics, evolution, structure-function relationships and physiological role. *Molecular bioSystems* *4*, 404-416.
97. Pushkin, A., and Kurtz, I. (2006). SLC4 base (HCO₃⁻, CO₃²⁻) transporters: classification, function, structure, genetic diseases, and knockout models. *Am J Physiol Renal Physiol* *290*, F580-599.
98. Alper, S.L. (2009). Molecular physiology and genetics of Na⁺-independent SLC4 anion exchangers. *The Journal of experimental biology* *212*, 1672-1683.
99. Ellory, J.C., Guizouarn, H., Borgese, F., Bruce, L.J., Wilkins, R.J., and Stewart, G.W. (2009). Review. Leaky Cl⁻-HCO₃⁻ exchangers: cation fluxes via modified AE1. *Philosophical transactions of the Royal Society of London* *364*, 189-194.
100. Yamaguchi, T., Ikeda, Y., Abe, Y., Kuma, H., Kang, D., Hamasaki, N., and Hirai, T. (2010). Structure of the membrane domain of human erythrocyte anion exchanger 1 revealed by electron crystallography. *Journal of molecular biology* *397*, 179-189.
101. Pasqualetto, E., Aiello, R., Gesiot, L., Bonetto, G., Bellanda, M., and Battistutta, R. (2010). Structure of the cytosolic portion of the motor protein prestin and functional role of the STAS domain in SLC26/SulP anion transporters. *Journal of molecular biology* *400*, 448-462.

102. Alvarez, B.V., Vilas, G.L., and Casey, J.R. (2005). Metabolon disruption: a mechanism that regulates bicarbonate transport. *The EMBO journal* *24*, 2499-2511.
103. Dorwart, M.R., Shcheynikov, N., Wang, Y., Stippec, S., and Muallem, S. (2007). SLC26A9 is a Cl(-) channel regulated by the WNK kinases. *The Journal of physiology* *584*, 333-345.
104. Homma, K., Miller, K.K., Anderson, C.T., Sengupta, S., Du, G.G., Aguinaga, S., Cheatham, M., Dallos, P., and Zheng, J. (2010). Interaction between CFTR and prestin (SLC26A5). *Biochimica et biophysica acta* *1798*, 1029-1040.
105. Zheng, J., Madison, L.D., Oliver, D., Fakler, B., and Dallos, P. (2002). Prestin, the motor protein of outer hair cells. *Audiology & neuro-otology* *7*, 9-12.
106. Greeson, J.N., Organ, L.E., Pereira, F.A., and Raphael, R.M. (2006). Assessment of prestin self-association using fluorescence resonance energy transfer. *Brain Res* *1091*, 140-150.
107. Liu, L.A. ed. (In Press). *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications*.
108. Sambrook, R., and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual*, 3 Edition, (Cold Spring Harbor: Cold Spring Harbor Press).
109. Wong, D.W.S. (2006). *The ABCs of Gene Cloning*. 2nd Edition. (New York: Springer Science).
110. Mott, M.L., and Berger, J.M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol* *5*, 343-354.
111. Sabate, R., de Groot, N.S., and Ventura, S. (2010). Protein folding and aggregation in bacteria. *Cell Mol Life Sci* *67*, 2695-2715.
112. Komar, A.A., and Hatzoglou, M. (2005). Internal ribosome entry sites in cellular mRNAs: mystery of their existence. *The Journal of biological chemistry* *280*, 23425-23428.
113. Lakowicz, J. (2006). *Principles of Fluorescence Spectroscopy*, 3rd Edition, (New York: Springer Science).
114. Patterson, G.H., Piston, D.W., and Barisas, B.G. (2000). Forster distances between green fluorescent protein pairs. *Analytical biochemistry* *284*, 438-440.
115. VanBeek, D.B., Zwier, M.C., Shorb, J.M., and Krueger, B.P. (2007). Fretting about FRET: correlation between kappa and R. *Biophysical journal* *92*, 4168-4178.
116. Hallworth, R., Currall, B., Nichols, M.G., Wu, X., and Zuo, J. (2006). Studying inner ear protein-protein interactions using FRET and FLIM. *Brain Res* *1091*, 122-131.
117. Piston, D.W., and Kremers, G.J. (2007). Fluorescent protein FRET: the good, the bad and the ugly. *Trends in biochemical sciences* *32*, 407-414.
118. Ai, H.W., Henderson, J.N., Remington, S.J., and Campbell, R.E. (2006). Directed evolution of a monomeric, bright and photostable version of *Clavularia cyan* fluorescent protein: structural characterization and applications in fluorescence imaging. *The Biochemical journal* *400*, 531-540.
119. Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science (New York, N.Y)* *181*, 223-230.

120. Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* 292, 195-202.
121. Rigden, D.J. ed. (2009). From Protein Structure to Function with Bioinformatics (New York: Springer Science).
122. Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydrophobic character of a protein. *Journal of molecular biology* 157, 105-132.
123. Krishtalik, L.I., and Cramer, W.A. (1995). On the physical basis for the cis-positive rule describing protein orientation in biological membranes. *FEBS letters* 369, 140-143.
124. Yan, C., and Luo, J. (2010). An analysis of reentrant loops. *The protein journal* 29, 350-354.
125. Bernsel, A., Viklund, H., Hennerdal, A., and Elofsson, A. (2009). TOPCONS: consensus prediction of membrane protein topology. *Nucleic acids research* 37, W465-468.
126. Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. *Journal of computational chemistry* 26, 1668-1688.
127. Brooks, B.R., Brooks, C.L., 3rd, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). CHARMM: the biomolecular simulation program. *Journal of computational chemistry* 30, 1545-1614.
128. Bowie, J.U., Luthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science* (New York, N.Y 253, 164-170.
129. Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature* 358, 86-89.
130. Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* 5, 725-738.
131. Xu, J., Li, M., Kim, D., and Xu, Y. (2003). RAPTOR: optimal protein threading by linear programming. *Journal of bioinformatics and computational biology* 1, 95-117.
132. Edwards, D., Stajich, J., and Hansen, D. eds. (2009). Bioinformatics: Tools and Applications (New York, NY: Springer).
133. Altschul, S.F., and Lipman, D.J. (1990). Protein database searches for multiple alignments. *Proceedings of the National Academy of Sciences of the United States of America* 87, 5509-5513.
134. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25, 3389-3402.
135. Zhang, Z., Schaffer, A.A., Miller, W., Madden, T.L., Lipman, D.J., Koonin, E.V., and Altschul, S.F. (1998). Protein sequence similarity searches using patterns as seeds. *Nucleic acids research* 26, 3986-3990.
136. Rosenberg, M.S. ed. (2009). Sequence Alignment (Los Angeles, CA: University of California Press).

137. Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* *89*, 10915-10919.
138. Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology* *48*, 443-453.
139. Sadler, J.R., Waterman, M.S., and Smith, T.F. (1983). Regulatory pattern identification in nucleic acid sequences. *Nucleic acids research* *11*, 2221-2231.
140. Benkert, P., Tosatto, S.C., and Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* *71*, 261-277.
141. Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins* *47*, 393-402.
142. Gille, C., and Frommel, C. (2001). STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* (Oxford, England) *17*, 377-378.
143. Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. (2005). Entrez Gene: gene-centered information at NCBI. *Nucleic acids research* *33*, D54-58.
144. Hubbard, T.J., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., et al. (2009). Ensembl 2009. *Nucleic acids research* *37*, D690-697.
145. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* *31*, 3784-3788.
146. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* (Oxford, England) *22*, 195-201.
147. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic acids research* *37*, D387-392.
148. Biegert, A., Mayer, C., Remmert, M., Soding, J., and Lupas, A.N. (2006). The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic acids research* *34*, W335-339.
149. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). InterPro: the integrative protein signature database. *Nucleic acids research* *37*, D211-215.
150. Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B.A., de Castro, E., Lachaize, C., Langendijk-Genevaux, P.S., and Sigrist, C.J. (2008). The 20 years of PROSITE. *Nucleic acids research* *36*, D245-249.
151. Saier, M.H., Jr., Tran, C.V., and Barabote, R.D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic acids research* *34*, D181-186.
152. Saier, M.H., Jr., Yen, M.R., Noto, K., Tamang, D.G., and Elkan, C. (2009). The Transporter Classification Database: recent advances. *Nucleic acids research* *37*, D274-278.

153. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic acids research* *36*, W5-9.
154. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* *22*, 4673-4680.
155. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* (Oxford, England) *23*, 2947-2948.
156. Geer, L.Y., Domrachev, M., Lipman, D.J., and Bryant, S.H. (2002). CDART: protein homology by domain architecture. *Genome research* *12*, 1619-1623.
157. Sneath, P.H.A. (1973). Numerical taxonomy : the principles and practice of numerical classification (San Francisco: W. H. Freeman).
158. Bornberg-Bauer, E., Rivals, E., and Vingron, M. (1998). Computational approaches to identify leucine zippers. *Nucleic acids research* *26*, 2740-2746.
159. Soding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* (Oxford, England) *21*, 951-960.
160. Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* *33*, W244-248.
161. Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., and Wootton, J.C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* (New York, N.Y) *262*, 208-214.
162. Quest, D., Dempsey, K., Shafullah, M., Bastola, D., and Ali, H. (2008). MTAP: the motif tool assessment platform. *BMC bioinformatics* *9 Suppl 9*, S6.
163. Tompa, M., Li, N., Bailey, T.L., Church, G.M., De Moor, B., Eskin, E., Favorov, A.V., Frith, M.C., Fu, Y., Kent, W.J., et al. (2005). Assessing computational tools for the discovery of transcription factor binding sites. *Nature biotechnology* *23*, 137-144.