MOTIF IDENTIFICATION IN THE MOTOR PROTEIN PRESTIN

Kathryn Dempsey*, Dhundy Bastola, and Hesham Ali

Departmen of Pathology and Microbiology, University of Nebraska Medical Center
Omaha, NE 68198,U.S.A.
**Email: kmdempsey@unmc.edu*

Benjamin Currall and Richard Hallworth

Department of Biosciences, Creighton University
Omaha, NE 68178, U.S.A.

One potential contribution of bioinformatics is the use of in silico methods to more rapidly identify potential structure-function relationships of newly discovered proteins. One such class of proteins is the solute carrier family 26 (Slc26) family of proteins. These proteins play a central role in anion transport and are associated with diseases such as Pendred syndrome and dystrophic dysplasia. Of particular interest is the mammalian Slc26a5 (prestin) protein, which, unlike other members of the family, acts as a unique voltage-dependent motor protein that functions at rates of greater than 70 kHz. In this study, prestin homologs were identified and their sulfate transporter and antisigma factor antagonist (STAS) superfamily domains were analyzed through both Gibbs and MEME motif identification software. The results of both software packages identified a DSVG motif within the STAS domain. Multiple executions of our analyses were performed for consistency and confirmed the DSVG motif as a strong signal in the STAS domain of our input. This domain is found to align with the solved structure of bacterial spoIIaa protein using three-dimensional modeling techniques. The DSVG motif within the STAS domain aligns with a key spoIIaa phosphorylation site essential for normal function. Our analysis elucidated the structure-function relationship of prestin protein and this approach could be readily extrapolated to other domains and protein families. Future plans include development of a software pipeline that allows for identification of de novo signals that would suggest evolutionary pressure on prestin, related orthologs, and SLC family members.

1. INTRODUCTION

An intriguing aspect of proteomics is how conserved structures can be used in variety of ways to create unique function. This makes identifying conserved structures with small variances essential to understanding unique functions of homologous proteins. Bioinformatics continues to offer innovative *in silico* algorithms that will elucidate important domains, motifs, and residues among related sequences. In this study, we present an approach for determining short, specific motifs within protein structures. Application of traditional and novel computational techniques enables us to combine alignment, phylogeny, secondary structure prediction, pattern search, and 3-D visualization to predict potentially important motifs in protein structure. With the ability to examine these key residues at multiple levels of protein structures, we take steps toward finding critical conserved motifs and deducing unique structure-function relationships within protein families.

A newly discovered class of proteins is the solute carrier family 26 (Slc26) family. Slc26 proteins can be found in a variety of organs ranging from the brain to the kidney. Almost all participate in monovalent anion transport usually involving chloride, bicarbonate, sulfate, and/or oxalate. All family members contain both a xanathine uracil permease (XUP) superfamily domain, a sulphate transporter, and anti-sigma factor antagonist (STAS) superfamily domain. While the STAS superfamily domain shares homology with the well described bacterial sporylation factor (SpoIIaa), little is known about the structure-function relationship of the XUP superfamily domain.

There are currently 11 identified human SLC26 genes (a1-a11). However, one gene, SLC26A10, is believed to be a pseudogene. Several diseases have been correlated to mutations within this family of proteins including Pendred syndrome (SLC26A4),

*Corresponding author

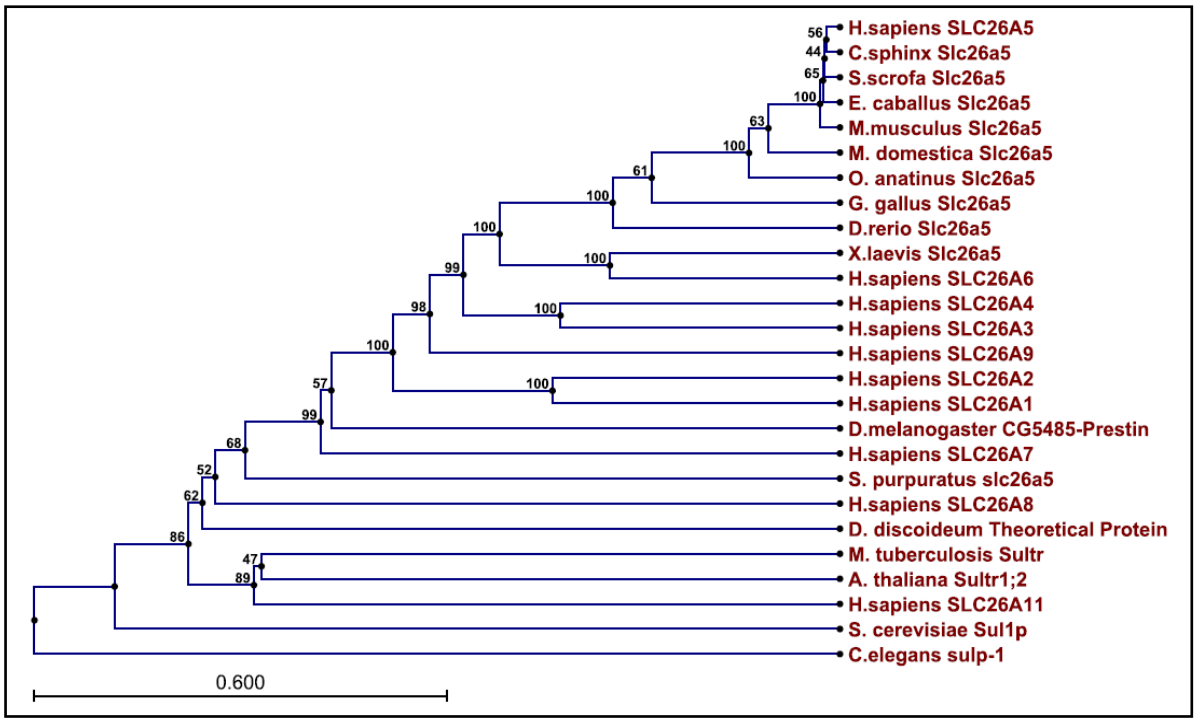

Figure 1: Phylogenetic tree of selected SLC26 protein homologs. Bootstraps run under 100 replicates are represented at each node. The human SLC26A5 is placed as the top node. Scale bar represents recent change in residues.

congenital chloride-losing diarrhea (SLC26A3), and chrondodysplasia (SLC26A2).[1] One member of this family, however, seems to have lost its ability to transport anions and instead operates as a unique form of a motor protein. Slc26a5 (prestin) is found in the membrane of the inner ear outer hair cell (OHC). When the OHC is depolarized, the cell shrinks, and when hyperpolarized, the cell elongates, in a process known as somatic motility. Somatic motility has been shown to operate at rates of greater than 70kHz and play a role in the amplification of sound input within the inner ear.[2] This theory is strengthened by the demonstration that knocking out of Slc26a5 results in loss of hearing in mice.[3] Even more intriguing, this motor function of Slc26a5 proteins occurs only in mammals, while the non-mammalian Slc26a5 functions as an anion transporter.[4] This unique function of mammalian prestin presents an opportunity to analyze both shared and unique structural components of the Slc26 family to help deduce the structure-function relationship.

2. DATA PREPARATION

In this study, we begin analyzing shared domains, motifs, and residues throughout the entire Slc26 family. Slc26 homologs were identified using the local alignment (BLAST) and using the naming conventions in the Ensembl Genome Browser. Sequences that did not contain both a XUP and STAS superfamily domain, in that order, were rejected. From the generated list of sequences, 25 SLC26A5 homologs were chosen for further analysis, which included 9 human paralogs (SLC26A1, SLC26A2, SLC26A3, SLC26A4, SLC26A6, SLC26A7, SLC26A8, SLC26A9, and SLC26A11) and 16 orthologs (*C. sphinx* Slc26a5, *S. scrofa* Slc26a5, *E. caballus* Slc26a5, *M. musculus* Slc26a5, *M. domestica* Slc26a5, *O. anatinus* Slc26a5, *G. gallus* Slc26a5, *D. rerio* Slc26a5, *X. laevis* Slc26a5, *S. purpuratus* Slc26a5, *D. melanogaster* CG5485, *C. elegans* Sulp-1, *S. cerevisiea* Sul1p, *A. thaliana* Sultr1;2, *M. tuberculosis* Sultr, and *D. discoideum* theoretical protein). These sequences were then structurally aligned and phylogenetically mapped using the 'Create Alignment' and 'Create Tree' applications of CLC Workbench (Fig. 1). Other than the known superfamily domains of XUP and STAS, no conserved domain or motif was identified.

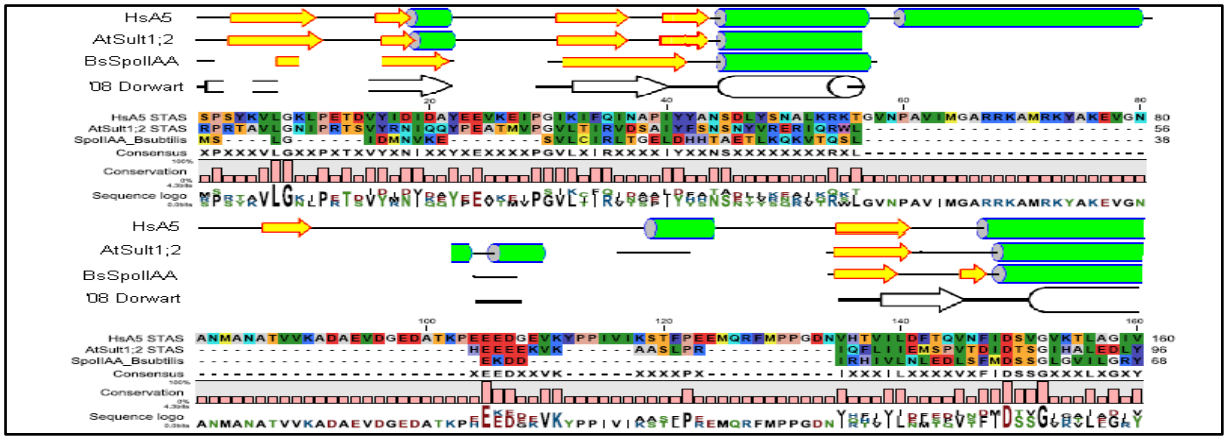

Figure 2: The PSIPred results for comparison of human SLC26A5, A. thaliana, and spoIIaa

The STAS domains of these sequences were considered for further analysis due to homology with another protein with a known structure-function relationship. Though the nature of the STAS domain is not well defined, the role of its homolog SpoIIAA is well known. Relating the STAS domain to the solved crystal structure of SpoIIAA can help elucidate function.[5] The surface of the SpoIIAA that surrounds S57 is participates in protein-protein interactions as a phosphorylation site.[6] Though there exists a well conserved loop around the S57 in the STAS region, it is not certain whether the serine/threonine is still phosphorylated in the Slc26 family. The STAS domain has also been shown to bind to other proteins[6] as well as itself.[7] Several mutations have been performed in the SpoIIAA and STAS domains resulting in a wide variety of affects, yet it is still difficult to define the role of STAS domain in Slc26 family

In chordates, there is a variable intervening sequence (IVS) within the STAS domain which must be taken into consideration. To determine the beginning and end of the STAS domain, PSIPRED[8] was used to predict secondary structure of the spoIIaa and compared to the nuclear magnetic resonance (NMR) solved spoIIaa structure (1AUZ). The PSIPRED secondary structure predictions for spoIIaa and *Arabidopsis thaliana* Sultr1;2, which does not contain an IVS region, were used to define the STAS domain in chordate Slc26s as shown in Figure 2. The newly delineated STAS domains were used for alignment and pattern detection.

3. PATTERN DETECTION

Motif detection programs are used to find patterns in protein or DNA sequences such as transcription factor binding sites or conserved signals in related or co-expressed genes. Two programs, Gibbs Motif Sampler and MEME, were chosen for our analysis based upon their popularity in literature and ability to search protein sequences. These programs are also able to find conserved patterns based on primary protein sequences without requiring extraneous information from the user.

Gibbs Motif Sampler[9] uses a recursive sampling algorithm to find patterns in user input. We ran the Gibbs Motif Sampler in recursive mode over our dataset, assuming 0+ sites per sequence submitted as set by the default. Multiple runs were performed searching for motifs of varying lengths (12, 14, 18, 22, and 24) and each motif returned was required to be found in at least 10 sequences. The Gibbs returns top scoring motifs based on expectation value. Duplicates were omitted from the final analysis.

We used another motif detection program, MEME,[10] which uses an expectation maximization algorithm. We ran MEME over our dataset with the default parameters. Multiple runs were performed searching for motifs of varying lengths (12, 14, 18, 22, and 24). MEME scores motifs based on a Bayesian threshold of confidence where top scoring results are reported with their respective confidence score. Duplicates were omitted from final analysis.

The ClustalWv1.8[11] alignment program was used to more precisely align the STAS sequence dataset based on the default Gonnet PAM 250 matrix. Upon visual evaluation of alignment results, as shown in Figure 3, it was noted that one residue, an aspartic acid (D), was 100% conserved throughout all STAS data. This observation allowed further insight into a possible pattern for prestin and homologs that could confer some importance with regard to conserved structure-function relationships.

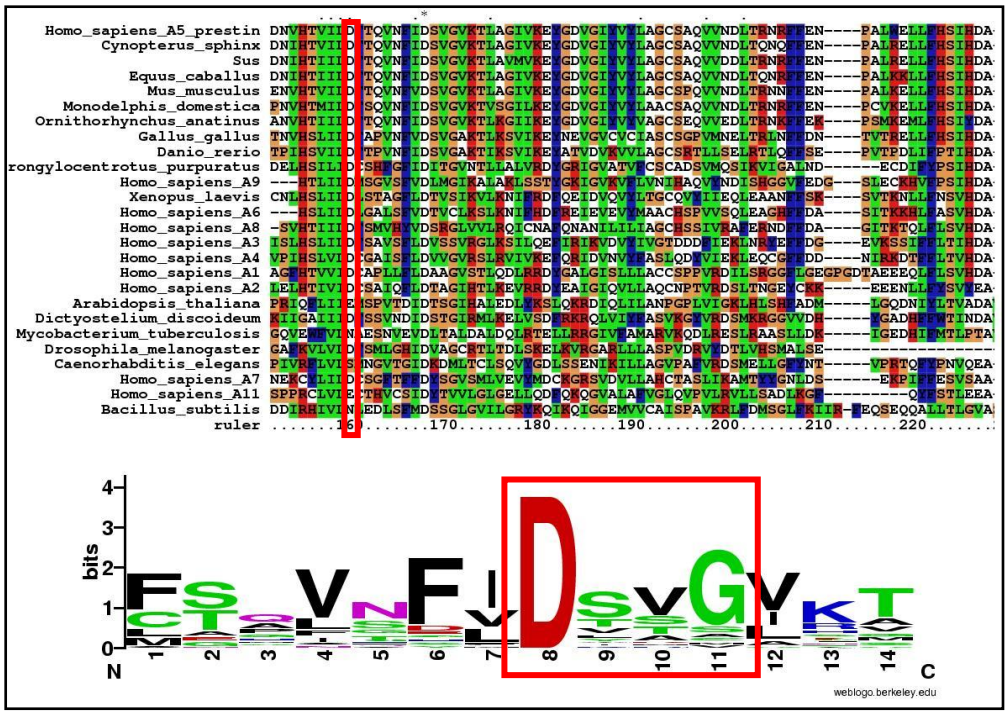

Figure 3: WebLogo Representation of Alignment Results. An aspartic acid (D) found to be 100% conserved is highlighted.

Visual evaluation of the top 5 results of each motif detection program combined with results of Clustal alignment revealed a conserved DS[V|S]G motif. In order to determine the possible functional relevance of the motif, the human SLC26A5 was submitted to the online 3Djigsaw database. The entire STAS domain was correctly identified by the 3D-Jjigsaw[12] (Pfam01740.13, identity: 100%, e-value: 3e-67) and aligned with the Protein Database (PDB) structure of a bacterial SpoIIaa (1AUZ). The PDB structures for both 1AUZ and STAS domain were aligned using the YASARA software and the DSVG motif was identified. The STAS DSVG motif aligned with a known SpoIIaa phosphorylation site (DSSG) important for protein-protein interaction.

4. DISCUSSION

This project has revealed a potentially significant motif that we expect to have a role in the Slc26 structure-function relationship. We employed alignment and motif detection to identify a conserved motif, DSVG. Identification of motifs makes it possible to conduct further mutagenesis experiments to determine impact on function. As observed in the SpoIIaa, the Slc26 DSVG motif may contain an important phosphorylation site that regulates homo- or heteromeric protein binding. Future work will include extending this analysis to the XUP superfamily domain, though this may prove more difficult due to the lack of previously solved structures. In this work, we have begun the initial stages to examine the shared structure-function relationship between homologs. In the future, we hope to extend these studies to not only analyze shared structure-function relationship, but also the unique function found within particular proteins.

Acknowledgments

This project was funded by the NIH grant number P20 RR016469 from the INBRE Program of the National Center for Research Resources, NSF-EPSCoR EPS-0346476, and State of Nebraska LB692.

References

1. Mount DB, Romero MF. The SLC26 gene family of multifunctional anion exchangers *Eur J Physiol* 2004; **447**:710–721.
2. Ashmore J. Cochlear outer hair cell motility. *Physiol Rev* 2008; **88**:173-210.
3. Cheatham MA, Huynh KH, Gao J, Zuo J, Dallos P. Cochlear function in prestin knockout mice. J Physiol 2004; 560.3:821-830.
4. Schaechinger TJ, Oliver D. Nonmammalian orthologs of prestin (SLC26A5) are electrogenic divalent/chloride anion exchangers. PNAS 2007; **104**: 7693–7698.
5. Aravind L, Koonin EV. The STAS domain – a link between anion transporters and antisigma-factor antagonists. *Curr Biol* 2000; **10(2)**:R53-5.
6. Dorwart MR, Scheynikov N, Wang Y, Stipped S, Muallem S. SLC26A9 is a Cl(-) channel

regulated by the WNK kinases. *J Physiol* 2007; **584(Pt 1)**:333-45.

7. Paqualletto E, Seydel A, Pellini A, Battistutta R. Expression, purification, and characterization of the C-terminal STAS domain of the SLC26 anion transporter prestin. *Protein Expr Purif* 2008. **58(2)**:249-56.
8. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS. & Jones DT. Protein structure prediction servers at University College London. *Nucl. Acids Res.* 2005; 33(Web Server issue): W36-38.
9. Thompson W, Rouchka EC, and Lawrence CE. Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucleic Acids Res 2003*. **31**(13):3580-3585
10. Bailey TL, Williams N, Misleh C, and Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research 2006. **34**:*W369-W373
11. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23(21)**:2947-2948
12. Bates PA, Kelley LA, MacCallum RM, and Sternberg MJE. Enhancement of Protein Modelling by Human Intervention in Applying the Automatic Programs 3D-JIGSAW and 3D-PSSM. *Proteins: Structure, Function and Genetics* 2001. **5**:39-46