Research Article

Evolutionary Influenced Interaction Pattern as Indicator for the Investigation of Natural Variants Causing Nephrogenic Diabetes Insipidus

Steffen Grunert and Dirk Labudde

Hochschule Mittweida, University of Applied Sciences, Technikumplatz 17, 09648 Mittweida, Germany

Correspondence should be addressed to Steffen Grunert; sgrunert@hs-mittweida.de

The importance of short membrane sequence motifs has been shown in many works and emphasizes the related sequence motif analysis. Together with specific transmembrane helix-helix interactions, the analysis of interacting sequence parts is helpful for understanding the process during membrane protein folding and in retaining the three-dimensional fold. Here we present a simple high-throughput analysis method for deriving mutational information of interacting sequence parts. Applied on aquaporin water channel proteins, our approach supports the analysis of mutational variants within different interacting subsequences and finally the investigation of natural variants which cause diseases like, for example, nephrogenic diabetes insipidus. In this work we demonstrate a simple method for massive membrane protein data analysis. As shown, the presented in silico analyses provide information about interacting sequence parts which are constrained by protein evolution. We present a simple graphical visualization medium for the representation of evolutionary influenced interaction pattern pairs (EIPPs) adapted to mutagen investigations of aquaporin-2, a protein whose mutants are involved in the rare endocrine disorder known as nephrogenic diabetes insipidus, and membrane proteins in general. Furthermore, we present a new method to derive new evolutionary variations within EIPPs which can be used for further mutagen laboratory investigations.

1. Introduction

Integral membrane proteins are coded by 20–30% of all open reading frames of known genomes [1–3]. As elements in accomplishing numerous molecular processes, that is, signal transduction and passive and active transport of an extensive number of chemical compounds and ions, mutations in genes coding for membrane proteins are often linked to diseases [4]. Despite their biological importance, relatively little is known about folding, functional mechanics, and synthesis of membrane proteins [1]. This is due to experimentally costly and complex procedures, since membrane proteins are difficult to handle in lab experiments [5]. To understand correspondence between genetic mutations and the effects on protein mechanics, the development of novel theoretical approaches is highly demanded. In our work we demonstrate a high-throughput in silico approach for the investigation

of the influences of genetic variations within interacting sequence part in membrane proteins, which are directly linked to nephrogenic diabetes insipidus. Nephrogenic diabetes insipidus (NDI) is a disorder which can be acquired as a side effect of surpassing drug taking or which is caused by inherited genetic mutations. Autosomal recessive and dominant inherited NDI are linked to mutations in genes encoding the integral membrane aquaporin-2 water channel [6, 7]. X-linked inheritable NDI is caused by mutations in the gene encoding the AVP type-2 receptor membrane protein (V2R) [8, 9]. In the general population, inherited NDI shows a low prevalence of one case per 20,000–30,000 people [10–12]. Aquaporin-2 water channels and V2R are essential elements in the water reabsorption through the apical cell membrane. This water composes the main part of preurine, a product that results from ultrafiltration in the kidney. The process of water reabsorption from the preurine

is essential to ensure the body's fluid balance and is realised by membrane-integrated aquaporin-2 water channels. The insertion of aquaporin-2 into the human kidney cell membrane is triggered by the antidiuretic hormone, which is also referred to as arginine vasopressin (AVP). The AVP blood concentration is regulated by the controlled release of AVP in the pituitary gland which is adapted according to the body's fluid balance. In the process, the binding of AVP to V2R leads to the activation of the receptor. In this state, V2R is able to interact with the guanine nucleotide-binding G(s) subunit alpha [13, 14]. Subsequently, the activation of adenylcyclase 6 takes place, leading to cAMP synthesis and increase of cAMP concentration in the cell plasma [15, 16]. By means of protein kinase A, cAMP triggers the phosphorylation of aquaporin-2 molecules which are stored in cytoplasmic vesicles that have bound to the endoplasmic reticulum. The phosphorylation induces the translocation and fusion of the cytoplasmic vesicles into the plasma membrane and finally leads to the insertion of aquaporin-2 molecules into the apical membrane [17]. Inactive mutants of V2R and aquaporin-2 cause a reduced water reabsorption in the kidneys [18]. Consequences are the typical symptoms of NDI, for example, sensorineural deafness, urinary tract anatomy, ataxia, peripheral neuropathy, mental retardation, psychiatric illness, a daily output of 15–20 L highly dilute (<100 mOsmol/kg) urine (polyuria), and compensatory excessive liquid intake [18–20]. In newborn infants, NDI is characterized by dehydration symptoms, irritability, and poor feeding as well as poor weight gain. A schematic illustration of these molecular coherences is given in Figure 1. The direct inspection of the aquaporin-2 gene as well as the V2 receptor gene (AVPR2) has become accomplishable in clinical practice [21] for differential NDI diagnosis and has been substituting dehydration testing over the last years [18].

2. Materials and Methods

As the first step, we want to realise a task which is involved in the prediction of homologue sequence parts within transmembrane α-helices. This means that aquaporin specific evolutionary interaction pattern pairs (EIPPs) were generated as described in current work of [22]. In this work, Grunert and Labudde show that the combination of interaction information and sequence motifs with evolutionary variation can be used for 3D structure prediction. They obtained key information from homologue sequences to separate and predict membrane protein structures in the context of interacting pattern and their evolutionary variation. Patterns as motif representatives are investigated for evolutionary covariation. Here, a motif has been described in previous work of [23] and can be written in a generalized, regular expression-like form of XYn, where X and Y correspond to amino acids separated by $n - 1$ highly variable positions. Interaction information contributes to detecting interacting pattern with evolutionary background. This means that evolutionary variation at pattern positions was marked as X. Here, different mutation types like that described in [22] may apply at specific X-position. Subsequently, in this work recently published proteins with PDB-Ids: 4nef, 4oj2 were used to transfer family

specific EIPPs to these aquaporin-2 representative proteins. For mention, the protein structure (PDB-Id: 4nef, 4oj2) was published by Frick et al. [24] and Vahedi-Faridi et al. [25] Beyond, both protein structures were considered as unknown structures at time of EIPP generation caused by missing Pfam entries. This led to no consideration of both proteins by EIPPs generation. Aquaporin specific EIPPs were derived from known structures of the corresponding PF00230 family. After obtaining of EIPPs, they were employed to generate interaction block schemes (IBSs). Here we try to illustrate that IBSs are useful graphical visualisation media to represent different interacting patterns which distinguish evolutionary. More specifically, we are able to show if a mutation within a pattern has influence on the evolutionary variability of the interacting counterpart. Eventually, IBSs can be used to support the understanding of the three-dimensional fold for the respective interaction partners and the whole protein structure. Moreover, transmembrane helical information was derived from PDBTM [26] for the proteins to be investigated (PDB-Ids: 4nef, 4oj2). Afterwards, EIPPs were applied on helices and sequence similarity of the incurred interacting ranges compared to known structures of the other family members was calculated. For further investigation, mutations occurring in nephrogenic diabetes insipidus patients were aggregated from recently published works [27–34] and registered on sequences of proteins to be investigated. These natural variants of NDI can also be obtained from UniProt (http://www.uniprot.org/uniprot/P41181). Finally, IBSs were applied to similar sequence parts which include NDI mutational effects.

2.1. Evolutionary Variations within EIPPs. In this section we describe a method to derive variation at X-positions from evolutionary sequence record. To realise this task, the full unknown seed structure dataset (9641 proteins) of the representative family (PF00230) was derived from Pfam database [35]. Transmembrane helical information was obtained using TMHMM Server v.2.0 [36]. A variety of methods have been developed to predict structural features from sequence, such as α-helical membrane-spanning helices and extra/intracellular domains. Basically, TMHMM performs a prediction of intra/extracellular regions and integral membrane helices based on sequence. Beside perresidue predictions TMHMM also lists underlying perresidue assignment probabilities as an indicator of prediction uncertainty. Consequently, helical information was used to apply our derived EIPPs at unknown structures. Here, Xpositions were investigated in a closer way when both existing EIPP counterparts were registered in different helices. For the detecting of new evolutionary variations, the amino acid occupancy from unknown structure information was used to compare with amino acids of known structures at specific X-positions. At last, new amino acids at variable X-position were registered and added. Finally, with this method we are able to extend evolutionary information within interacting sequence parts which can be used for further mutational or
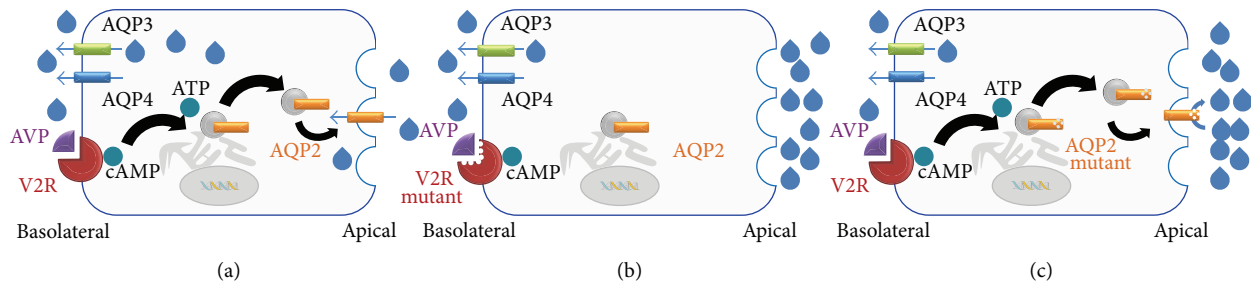
(a)   (b)   (c)

FIGURE 1: (a) In normally regulated water absorption in kidney cells, the antidiuretic hormone arginine vasopressin (AVP) is released in the pituitary gland, binds to the V2 receptor (V2R), and subsequently induces a series of phosphorylation reactions which lead to the insertion of aquaporin-2 water channels in the apical membrane that allow water molecules to pass the membrane. (b) Genetic mutations in the gene encoding V2R lead to reduced binding affinity and protein stability in V2R. Dysfunctional V2R mutants cause a significantly reduced amount of inserted aquaporin-2 proteins and thus decrease the water flux through the apical membrane. On the other hand, dysfunctional aquaporin-2 mutants decrease the water reabsorption as well (see (c)). Reduced water reabsorption is directly linked to an increased output of highly diluted urine (polyuria) and excessive drinking (polydipsia) which are the most severe symptoms observable in nephrogenic diabetes insipidus patients [12, 18, 20].

TABLE 1: Structural similar helical ranges of two given aquaporin-2 representatives in relation to aquaporin family members (PF00230). Similarity values describe what percentages are consistent helical ranges which are covered by EIPPs.

PDB-Id: 4nef		PDB-Id: 4oj2	
Helix	Similarity	Helix	Similarity
1	100%	1	100%
2	100%	2	100%
3	79.1%	3	79.3%
4	100%	4	100%
5	94.4%	5	95.2%
6	100%	6	100%
7	100%	7	100%

general investigations of membrane proteins or in this case specifically aquaporin water channel proteins.

3. Results and Discussion

Our structure prediction shows, if an unknown structure tends to a family affiliation, family specific EIPPs have to resurface on the protein sequence. Here, EIPPs were derived from known crystal structures of the aquaporin family (PF00230) and marked to α-helical structure of recently published aquaporin-2 representative proteins (PDB-Ids: 4nef, 4oj2). As mentioned before, aquaporin-2 representatives have not been considered in the EIPP generation process and make them to transparent unknown structures. Similarity results are shown in Figures 2 and 3 and confirm the already enlightened family affiliation and predicted structures. This means in all TM-helices EIPPs could be found and cover the helical range with up to 100% as listed in Table 1.

Here our prediction results explain the mightiness of EIPPs. On the one hand they provide useful and powerful information to predict α-helical structures within the transmembrane environment of homologue membrane proteins. On the other hand, we are able to describe selected interacting
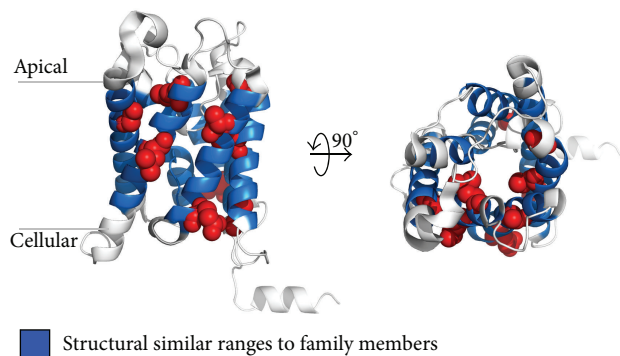

■ Structural similar ranges to family members

FIGURE 2: Structural colouring of helical ranges with high similarity to aquaporin family members (PF00230). Side and top-down view of the aquaporin-2 representative protein (PDB-Id: 4nef). Blue coloured cartoon residue ranges are present. These consist of family specific EIPPs which were found in known structures of the given family (PF00230). Red coloured spheres illustrate natural variants derived from UniProt (http://www.uniprot.org/uniprot/P41181) of the protein sequence caused by nephrogenic diabetes insipidus. All figures were rendered with PyMOL.

areas which are constrained by evolution. To evaluate this assumption, different IBSs were generated and applied to highly conserved interacting sequence parts which were derived from Pfam HMM-logos [35]. One IBS example is shown in Figure 4 which illustrates two interacting patterns. These patterns are part of the aquaporin-2 representative protein with PDB-Id: 4nef. Within an X-positioned pattern, we are able to register evolutionarily designed variable positions. Our IBSs additionally show that an interaction with another pattern takes place. In this work, the goal was not to show which pattern position is involved in spatial interaction but rather to show that two patterns build an interacting block. Figure 4 shows examples of variable positions, which can be occupied by different natural variants. With our IBSs, we can show that an interaction between two blocks is given, when the respective positions are occupied by the
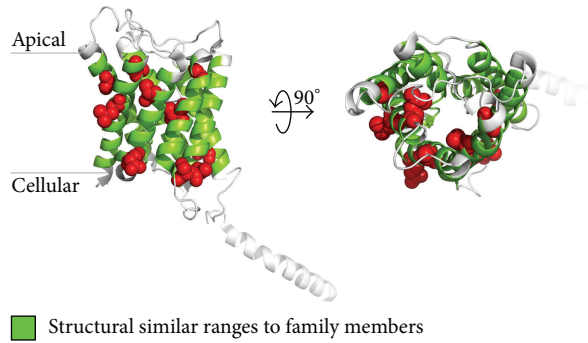
FIGURE 3: Structural colouring of helical ranges with high similarity to aquaporins (PF00230). Side and top-down view of the aquaporin-2 representative protein (PDB-Id: 4oj2). Green coloured cartoon residue ranges are present. These consist of family specific EIPPs which were found in known structures of the given family (PF00230). Red coloured spheres illustrate natural variants derived from UniProt (http://www.uniprot.org/uniprot/P41181) of the protein sequence caused by nephrogenic diabetes insipidus. All figures were rendered with PyMOL.
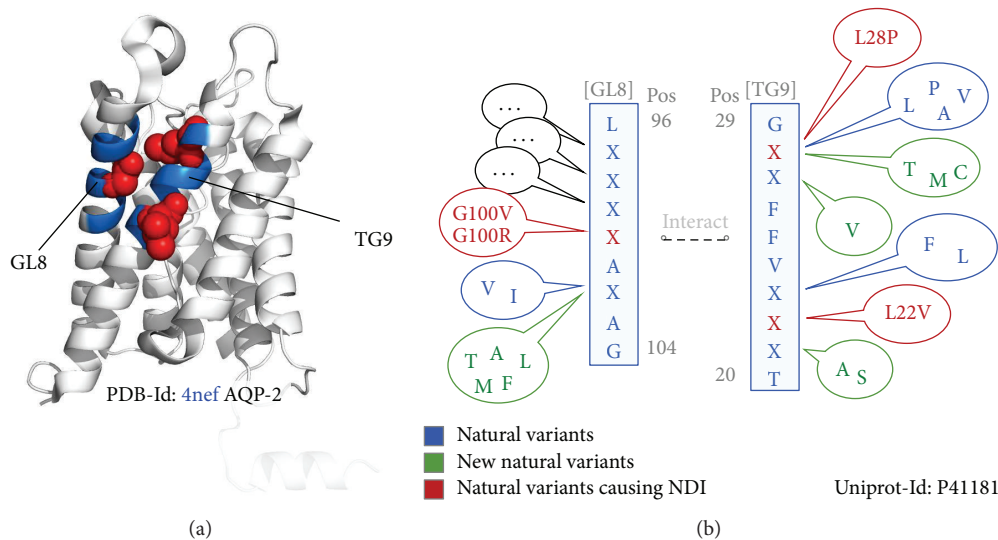

FIGURE 4: Examples of two spatially interacting subsequences. (a) More specifically, GL8 motif representative pattern (left) interacts with TG9 (right). Both patterns are coloured in blue. Red spheres representing natural variants causing nephrogenic diabetes insipidus (NDI). If we look at these patterns, it is not important to know which pattern position is involved in a spatial interaction. It is more important, that these two pattern build an interacting block. (b) The corresponding interaction block scheme (IBS), with the interacting pattern in blue letters. Red letters illustrating positions causing NDI. All patterns have variable positions. These positions are marked with X. Different coloured bubbles are present and address an X-position with possible natural variants. These are examples of possible occupations, which can be realized by natural variants derived from known structures (blue), natural variants causing NDI (red), or new natural variants derived from unknown structures (green).

amino acids. This implies that a TG9-GL8 interaction is given with Phe23-Ala101 or Phe26 and Ala101 or in the case of NDI caused by Leu28Pro [28], a TG9-GL8 interaction is given with Leu28 and Val102 (red coloured amino acid occupation) at specific positions. Here, IBSs give us a quick overview, that within or across a block evolutionary changes take place. That a destabilizing amino acid substitution is compensated by another position over the evolutionary time scale has been explained in detail in previous work of Morcos et al. [37]. Mutational variation information at specific X-positions can close this gap. This leads to further results in our work, the detection of new X-positioned

variations caused by evolution. Here, many variations of new possible amino acid substitutions within different sequence pattern could be obtained. As one example, the TG9 motif representative pattern TLIXVFFXXG is given. For this, X-positions can be occupied with the following amino acids: X3F, X3L, X7G, X8V, X7A, X8P, and X8L, which lead to a final regular expression similar to PROSITE [38–41] pattern syntax TLI[F,L]VFF[G,A][V,A,P,L]G. The evaluation of the VG5 submotif representative pattern VFFGXG shows the flexibility of evolution. Referring to natural variant causing NDI L28P, the fourth position (starting from zero) can be occupied with the following amino acids: X4L, X4P. Here, our

method has spawned new amino acids for this position. X4T, X4M, and X4C are able to complete this block which leads to final regular expression VFFG[V,A,P,L,T,M,C]G. Ultimately, this shows a greater variability outside of VG5. For further tasks in genomics or proteomics like protein modelling or mutational investigations, new amino acid replacements can be included. This expands the view on what structural mechanisms could also be possible to realise the three-dimensional fold within the respective sequence part and finally to ensure the protein function.

4. Conclusion

In the present work, we have applied a new approach for extracting short, spatially interacting amino acid sequence parts, so-called evolutionary interaction pattern pairs (EEIPs), from known structures of membrane proteins, more specifically aquaporin water channel proteins. Based on EIPPs, structure similarity of recently published aquaporin-2 representative proteins was determined and this in silico analysis confirms the aquaporin family affiliation. EIPPs were obtained and employed to generate interaction block schemes of highly conserved sequence parts annotated with natural variants caused by diabetes insipidus. Newly amino acid variations have been discovered. In our further works we will prove the relevance. In conclusion, it is a fact that disease patterns play an important role in membrane proteins but currently few involved structures are available. Different works have shown mutations on a membrane protein sequence influencing disease patterns. Besides, mutations are used in the diagnosis of biomarkers. However, the application of interaction block schemes can lead to better indicators and this in silico analysis can support laboratory mutagen investigations.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contribution

Steffen Grunert and Dirk Labudde participated in the design of this study. Steffen Grunert designed all methods and performed the implementation. Steffen Grunert evaluated the results. Dirk Labudde provided valuable consultation on structural biology and procedural steps. Steffen Grunert and Dirk Labudde wrote the paper. All authors read and approved the final paper.

Acknowledgments

References

[1] A. Marsico, D. Labudde, T. Sapra, D. J. Muller, and M. Schroeder, "A novel pattern recognition algorithm to classify membrane protein unfolding pathways with high-throughput single-molecule force spectroscopy," *Bioinformatics*, vol. 23, no. 2, pp. e231–e236, 2007.

[2] G. C. Brito and D. W. Andrews, "Removing bias against membrane proteins in interaction networks," *BMC Systems Biology*, vol. 5, article 169, 2011.

[3] S. Tan, T. T. Hwee, and M. C. M. Chung, "Membrane proteins and membrane proteomics," *Proteomics*, vol. 8, no. 19, pp. 3924–3932, 2008.

[4] M. Luckey, *Membrane Structural Biology: With Biochemical and Biophysical Foundations*, Cambridge University Press, 2008.

[5] P. G. Sadowski, A. J. Groen, P. Dupree, and K. S. Lilley, "Subcellular localization of membrane proteins," *Proteomics*, vol. 8, no. 19, pp. 3991–4011, 2008.

[6] P. M. T. Deen, M. A. J. Verdijk, N. V. A. M. Knoers et al., "Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine," *Science*, vol. 264, no. 5155, pp. 92–95, 1994.

[7] S. M. Mulders, D. G. Bichet, J. P. L. Rijss et al., "An aquaporin-2 water channel mutant which causes autosomal dominant nephrogenic diabetes insipidus is retained in the golgi complex," *The Journal of Clinical Investigation*, vol. 102, no. 1, pp. 57–66, 1998.

[8] A. M. W. van den Ouweland, J. C. F. M. Dreesen, M. Verdijk et al., "Mutations in the vasopressin type 2 receptor gene (AVPR2) associated with nephrogenic diabetes insipidus," *Nature Genetics*, vol. 2, no. 2, pp. 99–102, 1992.

[9] W. Rosenthal, A. Seibold, A. Antaramian et al., "Molecular identification of the gene responsible for congenital nephrogenic diabetes insipidus," *Nature*, vol. 359, no. 6392, pp. 233–235, 1992.

[10] S. Ananthakrishnan, "Diabetes insipidus in pregnancy: etiology, eva luation, and management," *Endocrine Practice*, vol. 15, no. 4, pp. 377–382, 2009.

[11] R. Krysiak, I. Kobielusz-Gembala, and B. Okopien, "Recurrent pregnancy-induced diabetes insipidus in a woman with hemochromatosis," *Endocrine Journal*, vol. 57, no. 12, pp. 1023–1028, 2010.

[12] G. L. Robertson, "Diabetes insipidus," *Endocrinology & Metabolism Clinics of North America*, vol. 24, no. 3, pp. 549–572, 1995.

[13] N. Wettschureck and S. Offermanns, "Mammalian G proteins and their cell type specific functions," *Physiological Reviews*, vol. 85, no. 4, pp. 1159–1204, 2005.

[14] G. Milligan and E. Kostenis, "Heterotrimeric G-proteins: a short history," *British Journal of Pharmacology*, vol. 147, no. 1, pp. S46–S55, 2006.

[15] N. Defer, M. Best-Belpomme, and J. Hanoune, "Tissue specificity and physiological relevance of various isoforms of adenylyl cyclase," *American Journal of Physiology: Renal Physiology*, vol. 279, no. 3, pp. F400–F416, 2000.

[16] J. Hanoune, Y. Pouille, E. Tzavara et al., "Adenylyl cyclases: structure, regulation and function in an enzyme superfamily," *Molecular and Cellular Endocrinology*, vol. 128, no. 1-2, pp. 179–194, 1997.

[17] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[18] E. L. Los, P. M. T. Deen, and J. H. Robben, "Potential of nonpeptide (ant)agonists to rescue vasopressin v2 receptor mutants for the treatment of X-linked nephrogenic diabetes insipidus," *Journal of Neuroendocrinology*, vol. 22, no. 5, pp. 393–399, 2010.

[19] T. M. Strom, K. Hörtnagel, S. Hofmann et al., "Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein," *Human Molecular Genetics*, vol. 7, no. 13, pp. 2021–2028, 1998.

[20] M. Birnbaumer, "V2R structure and diabetes insipidus," *Receptors and Channels*, vol. 8, no. 1, pp. 51–56, 2002.

[21] T. M. Fujiwara and D. G. Bichet, "Molecular biology of hereditary diabetes insipidus," *Journal of the American Society of Nephrology*, vol. 16, no. 10, pp. 2836–2846, 2005.

[22] S. Grunert and D. Labudde, "The observation of evolutionary interaction pattern pairs in membrane proteins," Submitted to *BMC Structural Biology*.

[23] Y. Liu, D. M. Engelman, and M. Gerstein, "Genomic analysis of membrane protein families: abundance and conserved motifs," *Genome biology*, vol. 3, no. 10, 2002.

[24] A. Frick, U. K. Eriksson, F. de Mattia et al., "X-ray structure of human aquaporin 2 and its implications for nephrogenic diabetes insipidus and trafficking," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 17, pp. 6305–6310, 2014.

[25] A. Vahedi-Faridi, D. Lodowski, A. Schenk et al., The structure of aquaporin, Journal, To be Published.

[26] D. Kozma, I. Simon, and G. E. Tusnády, "PDBTM: protein data bank of transmembrane proteins after 8 years," *Nucleic Acids Research*, vol. 41, no. 1, pp. D524–D529, 2013.

[27] M. C. Canfield, B. K. Tamarappoo, A. M. Moses, A. S. Verkman, and E. J. Holtzman, "Identification and characterization of aquaporin-2 water channel mutations causing nephrogenic diabetes insipidus with partial vasopressin response," *Human Molecular Genetics*, vol. 6, no. 11, pp. 1865–1871, 1997.

[28] N. Marr, D. G. Bichet, S. Hoefs et al., "Cell-biologic and functional analyses of five new Aquaporin-2 missense mutations that cause recessive nephrogenic diabetes insipidus," *Journal of the American Society of Nephrology*, vol. 13, no. 9, pp. 2267–2277, 2002.

[29] S.-H. Lin, D. G. Bichet, S. Sasaki et al., "Two novel aquaporin-2 mutations responsible for congenital nephrogenic diabetes insipidus in Chinese families," *The Journal of Clinical Endocrinology & Metabolism*, vol. 87, no. 6, pp. 2694–2700, 2002.

[30] P. Carroll, H. Al-Mojalli, A. Al-Abbad et al., "Novel mutations underlying nephrogenic diabetes insipidus in Arab families," *Genetics in Medicine*, vol. 8, no. 7, pp. 443–447, 2006.

[31] S. M. Mulders, N. V. A. M. Knoers, A. F. van Lieburg et al., "New mutations in the aqp2 gene in nephrogenic diabetes insipidus resulting in functional but misrouted water channels," *Journal of the American Society of Nephrology*, vol. 8, no. 2, pp. 242–248, 1997.

[32] R. Vargas-Poussou, L. Forestier, M. D. Dautzenberg, P. Niaudet, M. Déchaux, and C. Antignac, "Mutations in the vasopressin v2 receptor and aquaporin-2 genes in 12 families with congenital nephrogenic diabetes insipidus," *Journal of the American Society of Nephrology*, vol. 8, no. 12, pp. 1855–1862, 1997.

[33] K. Goji, M. Kuwahara, Y. Gu, M. Matsuo, F. Marumo, and S. Sasaki, "Novel mutations in aquaporin-2 gene in female siblings with nephrogenic diabetes insipidus: evidence of disrupted

[34] water channel function," *The Journal of Clinical Endocrinology & Metabolism*, vol. 83, no. 9, pp. 3205–3209, 1998.

[34] M. Kuwahara, "Aquaporin-2, a vasopressin-sensitive water channel, and nephrogenic diabetes insipidus," *Internal Medicine*, vol. 37, no. 2, pp. 215–217, 1998.

[35] M. Punta, P. C. Coggill, R. Y. Eberhardt et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 40, no. 1, pp. D290–D301, 2012.

[36] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *Journal of Molecular Biology*, vol. 305, no. 3, pp. 567–580, 2001.

[37] F. Morcos, A. Pagnani, B. Lunt et al., "Direct-coupling analysis of residue coevolution captures native contacts across many protein families," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 49, pp. E1293–E1301, 2011.

[38] C. J. A. Sigrist, E. De Castro, L. Cerutti et al., "New and continuing developments at PROSITE," *Nucleic Acids Research*, vol. 41, no. 1, pp. D344–D347, 2013.

[39] C. J. A. Sigrist, L. Cerutti, N. Hulo et al., "Prosite: a documented database using patterns and profiles as motif descriptors," *Briefings in bioinformatics*, vol. 3, no. 3, pp. 265–274, 2002.

[40] E. de Castro, C. J. A. Sigrist, A. Gattiker et al., "Scanprosite: detection of prosite signature matches and prorule-associated functional and structural residues in proteins," *Nucleic Acids Research*, vol. 34, supplement 2, pp. W362–W365, 2006.

[41] C. J. A. Sigrist, E. de Castro, P. S. Langendijk-Genevaux, V. Le Saux, A. Bairoch, and N. Hulo, "ProRule: a new database containing functional and structural information on PROSITE profiles," *Bioinformatics*, vol. 21, no. 21, pp. 4060–4066, 2005.