Functional Analyses of Membrane Protein Mutants involved in Nephrogenic Diabetes insipidus: An Energy-based Approach

Florian Heinke and Dirk Labudde
Department of MNI
University of Applied Sciences Mittweida, Germany

1 Introduction

Integral membrane proteins are coded by 20 – 30% of all open reading frames of known genomes (Marsico et al., 2007; Brito & Andrews, 2011; Tan et al., 2008). As elements in accomplishing numerous molecular processes, i.e. signal transduction, 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 (Luckey, 2008). Despite their biological importance, relatively little is known about folding, functional mechanics and synthesis of membrane proteins (Marsico et al., 2007). This is due to experimentally costly and complex procedures, since membrane proteins are difficult to handle in lab experiments (Sadowski et al., 2008). To understand correspondences between genetic mutations and the effects on protein mechanics, the development of novel theoretical approaches is highly demanded. In our work we demonstrate a theoretical approach to discuss the influences of genetic mutations 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 (Deen et al., 1994; Mulders et al., 1998). X-linked inheritable NDI is caused by mutations in the gene encoding the AVP type-2 receptor membrane protein (V2R) (van den Ouweland et al., 1992; Rosenthal et al., 1992). In the general population, inherited NDI shows a low prevalence of one case per 20,000 – 30,000 people (Ananthakrishnan, 2009; Krysiak et al., 2010; Robertson, 1995). 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 pro-urine; a product that results from ultra-filtration in the kidney. The process of water reabsorption from the pro-urine 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 argenine 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 (Wettschureck & Offermanns, 2005; Milligan & Kostenis, 2006). Subsequently, the activation of adenyl cyclase 6 takes place, leading to cAMP synthesis and increase of cAMP concentration in the cell plasma (Defer et al., 2000; Hanoune et al., 1997). 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 (Kanehisa & Goto, 2000).

Inactive mutants of V2R and aquaporin-2 cause a reduced water reabsorption in the kidneys (Los et al., 2010). Consequences are the typical symptoms of NDI, e.g. 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 (Los et al., 2010; Strom et al., 1998; Birnbaumer, 2002). In newborn infants, NDI is characterised by dehydration symptoms, irritability, 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 (Fujiwara & Bichet, 2005) for differential

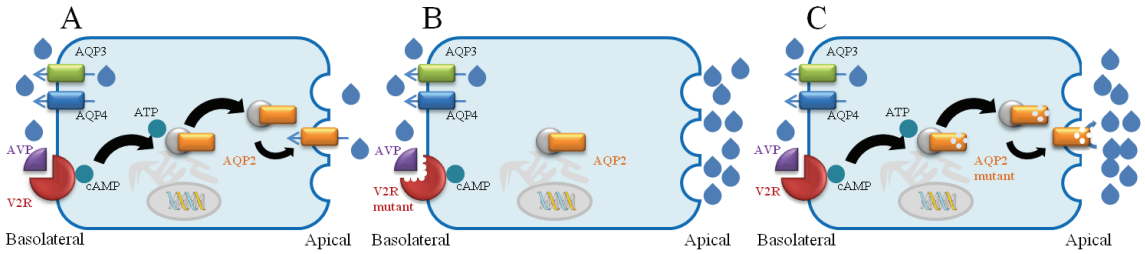

Figure 1: A: In normally regulated water absorption in kidney cells, the antidiuretic hormone arginine vasopressin (AVP) is released in the pituitary gland, binds to the V2 receptor (V2R), and subsequently induces a series of phosphorylation reactions which lead to the insertion of aquaporin-2 water 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 dilute urine (polyuria) and excessive drinking (polydipsia) which are the most severe symptoms observable in nephrogenic diabetes insipidus patients (Los et al., 2010; Robertson, 1995; Birnbaumer, 2002).

NDI diagnosis and has been substituting dehydration testing over the last years (Los et al., 2010).

The analyses elucidated in this work focus on the stability of aquaporin-2 and stability discrepancies which can be observed in aquaporin-2 mutant structures. Further discussions deal with V2R and the shifts in stability induced by interactions with AVP. The key methodology utilised in the investigations given in this chapter is based on the calculation and comparison of so-called protein energy profiles. Protein energy profiles are representations of structure stability, whereas physico-chemical properties and spatial information are abstracted to a sequence of fuzzy numbers by utilising a coarse-grained energy model. This approach provides in general the opportunity to inspect spatial and conformational modifications as consequences of protein-environment interactions (Heinke & Labudde, 2012). By means of pairwise energy profile comparisons, energetic discrepancies become observable; whereby information of effects in function and activity, which, for example, might have resulted from mutations, can be gained. More detailed discussions of energy profile-based methods are given in the following section.

Besides the strategies presented in this chapter, there are numerous techniques that are well described in the literature and which can be applied to address similar problems. However, these methods rely on structure and/or sequence data. With no experimental structure data available for the investigated key proteins, such techniques become limiting. To circumvent these restrictions, this book chapter is focused on the application of basic concepts of statistical physics to derived structure models and known protein structures. From this, insights are gained about the possible correspondences of observed mutations and protein structure destabilisation. The methods for calculating and comparing protein energy profiles as well as an energy profile database are freely available at http://bioservices.hs-mittweida.de/Epros (Heinke et al., 2013).

2 Protein Energy Profiles as an Investigation Methodology for Protein Functionality and Stability

2.1 Theory of Protein Energy Profiles

To investigate the influences of mutations on protein function and stability, a coarse-grained knowledge-based energy model has been implemented. Like in many coarse-grained energy models, smoothing of physical information is achieved by reducing system complexity (Zhang et al., 2004). In particular, coarse-grained energies are derived in our model from statistics and concepts of statistical physics as well as by applying a straight-forward residue contact function. The basic concepts which have been applied in this model are discussed in (Sippl, 1993) and (Tanaka & Scheraga, 1975; Tanaka & Scheraga, 1976). Basically, the energy of a residue is approximated by applying Boltzmann principles to amino acid-wise observations made from a dataset of experimental protein structures to the observed residue and the residue it is interacting with. Here, the dataset for statistics generation has consisted of 380 non-redundant α-helical transmembrane protein structures that had been obtained from the Protein Data Bank of Transmembrane Proteins (PDBTM) (Tusnady et al., 2004; Tusnady et al., 2005). Subsequently, statistics have been derived by counting the number of occurrences of each amino acid i in which the observed residues are found to be exposed at the surface ($n_{i,out}$) or found to be buried ($n_{i,in}$) in the protein structure (Dressel et al., 2007). Additionally, the topological state s of the observed residue is taken into account. If i is located inside the membrane, s is assigned as TM, with s = nTM otherwise. Topological data had been obtained from the PDBTM database. C_β atoms (or C_α-atoms in cases when observing glycine) are declared as spatial residue-representative points. Using these statistics, the energy of residue i can be approximated:

$$e_i = -\ln\left(\frac{n_{i,in,s}}{n_{i,out,s}}\right) + k_i \ , \tag{1}$$

with $k_i = 0$ if s = nTM, or

$$k_i = -\ln\left(\frac{n_{i,\mathrm{TM}}}{n_{i,\mathrm{nTM}}}\right) \tag{2}$$

otherwise. To approximate the total energy E_i^*, all potentials of all interacting residues are taken into account (Dressel et al., 2007; Heinke & Labudde, 2012):

$$E_i^* = \sum_{\forall j | j \neq i} f(i,j)\,(e_i + e_j) \ , \tag{3}$$

with

$$f(i,j) = \begin{cases} 1, & \|i-j\| < 8\text{Å} \\ 0, & \text{else} \end{cases} \tag{4}$$

The sequence of all E_i^* of a given protein structure corresponds to the protein energy profile. An energy profile can be interpreted as a physiochemical and structure-specific representation, since spatial and chemical information are included in the computation. One can address that energy profile-based approaches can be realised from energy profiles derived from more sensitive, all-atom and physics-based methods (for instance see (Mrozek et al., 2006)), i.e. molecular dynamics (MD) techniques. However, the application of such techniques to membrane proteins is in general difficult to handle computationally, since the structure

has to be embedded in a lipid bilayer, which increases the total number of atoms and thus system complexity. Additionally, simulations over long time scales (≈10ns – 100ns) are required to draw meaningful conclusions (Luckey, 2008). In the model applied in this study, the effects of the membrane bilayer are modelled by means of term 2. Note that energy values computed by this model are given in arbitrary unit entities. In Figure 2, the energy profile of the modelled aquaporin-2 structure is plotted (see Figure 2C) and mapped on the protein structure (see Figure 2A and B) using a rainbow colouring scheme representing the corresponding energy values. As shown, residues with low energy values (blue) occur mainly in membrane spanning helices. In contradistinction to this observation, residues which are located in extra or intra-cellular protein regions show increased energy values, as depicted by dark yellow and green colouring. Accordingly, membrane spanning α-helices rest stabilised in the hydrophobic membrane environment. This general observation correlates with previously published data derived from experiments (for examples see (Fleming & Engelman, 2001; Finger et al., 2006; Luckey, 2008)).

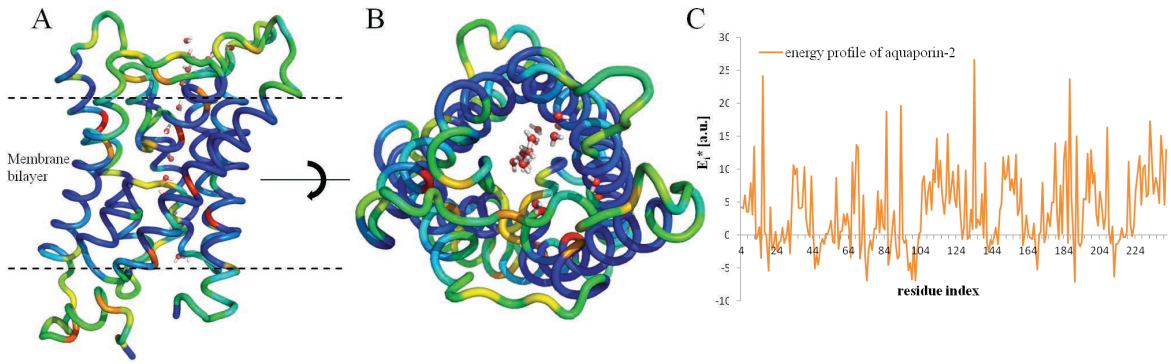

Figure 2: A and B: Energy values computed by the coarse-grained energy model utilised in this work are mapped onto the structure model of aquaporin-2. Obviously, residues with low energy values occur mainly in membrane spanning helices - an observation which is in agreement with experimental data (Fleming & Engelman, 2001; Finger et al., 2006; Luckey, 2008). Water molecules passing the pore are shown as spheres. C: The plotted energy profile of aquaporin-2.

Furthermore, a methodology for computing pairwise and multiple energy profile alignments was implemented. By that, energetic shifts can be analysed and global energy profile distances can be derived (Heinke & Labudde, 2012). Energy profile distances (dScores) can be used as input for hierarchical clustering methods, such as UPGMA (Unweighted Pair Group Method with Arithmetic Mean) (Sokal & Michener, 1958) or Neighbor-Joining (NJ) (Saitou & Nei, 1987). Similar to the approach discussed in the work of Eisenberg et al (Bowie et al., 1991) and Kozielski et al (Mrozek et al., 2006; Mrozek et al., 2007; Mrozek et al., 2009), energy profiles can be aligned by means of dynamic programming. Therefore, an energy-energy scoring function was implemented. It is derived by distances between power-equal intervals of the gaussian integral of the energy distribution. For scoring two energy values, each energy value is assigned to its interval in the gaussian integral. The distance between both integrals corresponds to the pairwise energy score. This scoring is used for aligning two given energy profiles A and B by alignment algorithms, f.e. the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970) or the Smith-Waterman algorithm (Smith & Waterman, 1981). The estimation of alignment significance is provided by normalising the resulting score x_r by taking into account the best possible score x_{opt} and the average permutation score $\bar{x}_p$. The latter is

derived by permuting and realigning the given energy profiles iteratively. As discussed in (Gusfield, 1993; Higgins et al., 1996), this normalised score is referred to as distance score (dScore) and is defined as:

$$\mathrm{dScore}\left(x_{\mathrm{r}}\right) = -\log\left(\frac{x_{\mathrm{r}} - \overline{x}_{\mathrm{p}}}{x_{\mathrm{opt}} - \overline{x}_{\mathrm{p}}}\right) \tag{5}$$

with

$$x_{\mathrm{opt}}\left(A,B\right) = \frac{\delta\left(|A| + |B|\right)}{2}. \tag{6}$$

Here, δ denotes the best possible pairwise energy score. In general, significant energy profile alignments correspond to dScores of less than 2.5 dits. The alignment of two identical energy profiles corresponds to a dScore of 0 dits (Heinke & Labudde, 2012).

2.2 Correspondences of Energy, Function and Structure

To investigate correlations between coarse-grained energies computed by our model and energies derived by MD, 220 non-redundant globular protein structures (sequence identity < 25%) had been obtained from the Protein Data Bank (PDB) (Berman et al., 2000) and have been analysed by the TINKER molecular dynamics software suite (Ponder, 2001). By this analysis, all-atom energies have been computed and investigated for correlations with energies derived by our coarse-grained energy model. As shown in Figure 3A, total binding energies calculated by TINKER (E_{FG}) correlate very well with the sums of all energies computed by our model (E_{CG}). Thus, our coarse-grained energy model can be used to draw physically and biologically meaningful conclusions concerning residue stability and destabilising effects of point mutations.

Furthermore, sequence, structural and functional correspondences to pairwise energy profile distances were investigated. For this purpose, 2,700 non-redundant globular protein structures and their corresponding GO-term annotations had been obtained from the PDB and UniProt (Apweiler et al., 2004), respectively. Sequence identities and structural similarities have been recorded. To investigate functional correspondences, GO-term annotations (Ashburner et al., 2000) have been compared semantically utilising the G-SESAME web server (Du et al., 2009). As depicted in Figure 3B, sequence identities (seqId), structural similarities ($-\log(\text{p-value})$, calculated by FATCAT (Ye & Godzik, 2003)), functional similarity (semantic GO-term annotation similarity, illustrated by a blue-to-red colouring scheme) correlate strongly with energy profile distances (dScores). From this observation, it can be deduced that energy profiles yield sequence, structural and functional information as proposed. Additionally, these correspondences can be transferred to α-helical membrane proteins as well. Thus, energy profile differences correspond to functional and structural divergences and can be analysed in detail. According to this, dScores can be applied as a measure of structural stability.

3 Protein Stability of Aquaporin-2

3.1 Description of Aquaporin-2

Aquaporins belong to a family of related water channels widely present in nature. They provide pores with high water permeability and consist of four identical subunits that form a tetramer complex after insertion into the membrane bilayer (Pollard & Earnshaw, 2007). In human tissues, 12 different isoforms are expressed but only aquaporin-2, -3 and -4 are present in the principal collecting duct cells in the membrane,

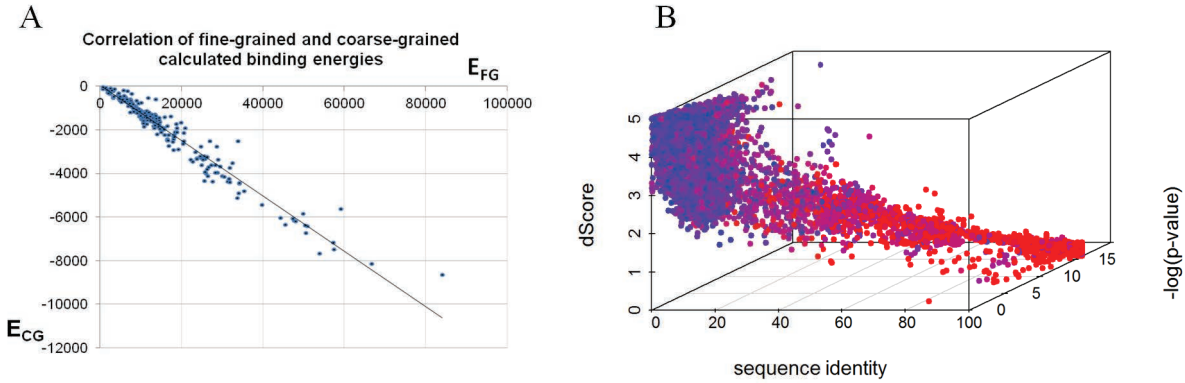

Figure 3: A: The total binding energies of 220 non-redundant globular protein structures (sequence identity < 25%) were calculated by means of the TINKER molecular dynamics software suite (Ponder, 2001) and plotted against the sum of all E_i^* computed by our coarse-grained energy model. A linear correlation to all-atom binding energies was found. B: Scatter plot of sequence identities, structural similarities ($-\log$(p-value)) and energy profile distance (dScore). Additionally, dots are coloured according to the semantic similarity of the two GO-term annotations, as a representation of functional correspondences, of both proteins. Blue coloured alignments indicate no detectable functional similarity between both proteins, whereas red colouring points to two identical GO-term sets. As shown, dScores correspond to functional, structural and sequence similarity simultaneously. Thus, energy profile differences correspond to functional and structural divergences and can be analysed in detail.

whereas only aquaporin-2 is linked to NDI (King et al., 2004). Each aquaporin subunit consists of a bundle of six membrane spanning helices (helices H1-H6) and two long, distinct loop regions of both holding a short α−helix located close to the membrane surface (helices HE, HB). As discussed in the literature, two highly conserved Asn-Pro-Ala-motifs are located in both helices HE and HB, facing each other in opposite α-helical direction (Chen et al., 2006). It is proposed, that this characteristic structural feature induces a bipolar electric field which is, besides Cys 189 (residue numbering according to aquaporin-1, PDB-Id 1fqy), mainly responsible for proton selectivity in aquaporins. It has been shown that residues Phe 56, His 180 and Arg 195 induce a secondary free energy barrier located at the extracellular site of the protein which contributes to selectivity as well. An attenuation of the secondary energy barrier and, thus, reduced selectivity have been observed in Arg 195 mutants by means of MD simulations (de Groot et al., 2003; Chakrabarti et al., 2004a; Chakrabarti et al., 2004b; Ilan et al., 2004). Interestingly, the residues located in the interior of the aquaporin water channels show hydrophobic properties which, as proposed, increase water permeability. The structure of aquaporin-1 is depicted in Figure 4 with helices H1-H6 and HB with HE highlighted by orange and red colouring, respectively (see Figure 4A). In this figure, the residues mainly involved in water transport are depicted in detail in B. However, the three-dimensional structure of aquaporin-2 has not been determined experimentally yet, but, because of the strong homology in this family, it is proposed but neither experimentally nor theoretically proven that the general aforementioned aquaporin characteristics apply to aquaporin-2 as well.

Single-molecule force spectroscopy (SMFS) is one general way for investigating molecular stability and interactions experimentally. It is demonstrated in the following sections that the energy profile-based

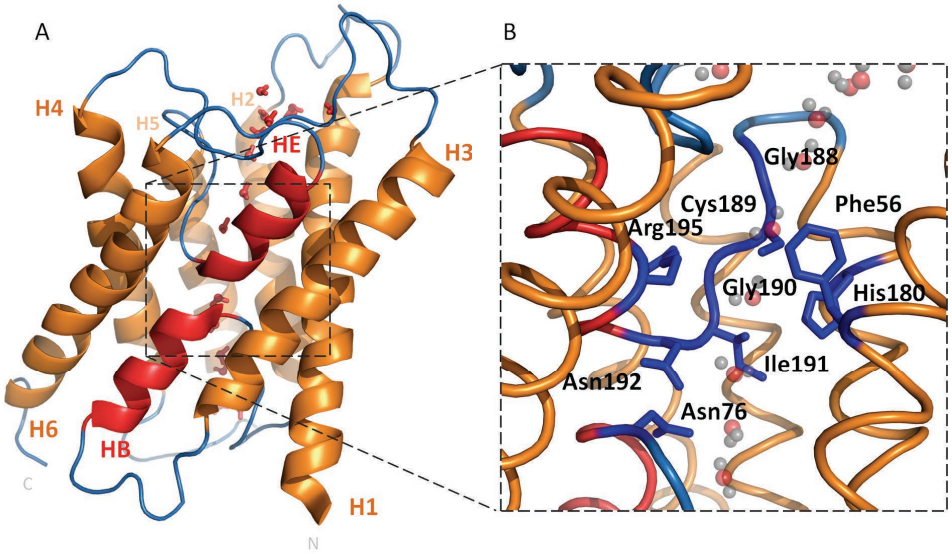

Figure 4: Homology studies of the aquaporin family suggest that all aquaporins share common structural features. It is shown, that aquaporins consist of six membrane spanning helices (H1-H6, highlighted in gold in A) and two short membrane loops, each embedding a single short helix (HB and HE, highlighted in red in A). In each helix HB and HE, a highly conserved Asn-Pro-Ala motif is present. In the folded structure, both motifs face each other in opposite direction and establish a biploar electric field which is proposed to be responsible for water permeability (de Groot et al., 2003; Chakrabarti et al., 2004a; Chakrabarti et al., 2004b; Ilan et al., 2004). B: The residues involved in water transport are highlighted as sticks (PDB-Id: 1fqy, residue numbering is given according to aquaporin-1).

approach can be applied to transfer information derived from SMFS data to structural and stabilising features in proteins, e.g. aquaporin-1 and aquaporin-2, which confirms the aforementioned hypothesis.

3.2 Theoretical Analysis of Protein Stability of Aquapoprin-2

Single-molecule force spectroscopy (SMFS) has been introduced as a valuable approach to investigate the stability and stabilising effects in molecules, e.g. intra- and intermolecular forces. There have been numerous studies which discuss the use of SMFS as a method for analysing stabilising forces, probing energy landscapes and measuring so-called unfolding events by unfolding the membrane protein of interest in a controlled manner (Müller & Engel, 1999; Müller et al., 1999; Seelert et al., 2003; Janovjak et al., 2004; Janshoff et al., 2000). To this day, aquaporin-1 is the only member of the aquaporin family that has been investigated by SMFS (Möller et al., 2003). However, SMFS is still an error-prone, computationally- and resource-demanding methodology. Additionally, understanding and accommodating experimental observations to structural data by means of a generalised model has not been achieved yet.

As an approach for investigating protein stability in aquaporins, the energy profiles of aquaporin-1,-2,-3,-4, and -5 have been computed and aligned for deriving all pairwise dScores. SMFS data of aquaporin-1 has been gathered from literature and studied for correlations with energy profile and structure data of aquaporin-1 and aquaporin-2. However, as elucidated earlier, energy profile-based analyses require spatial

information of the protein of interest. To this end, the three-dimensional structures of aquaporin-1 (PDB-Id: 1fqy), aquaporin-4 (PDB-Id: 3gd8), aquaporin-5 (PDB-Id: 3d9s) have been obtained from the PDB. Theoretical structure models of aquaporin-2 and -3 have been retrieved from the ModBase database (Pieper et al., 2011). The aquaporin-2 model has been produced by utilising comparative modelling and the protein structure of aquaporin-5 serving as modeling template (PDB-ID 3d9s). With both proteins sharing a sequence identity of 68%, the resulting structure model has been found to be reliable in quality. Quality re-evaluation has been performed by means of the protein structure analysis tool VADAR (version 1.8 (Willard et al., 2003)). One measure of quality given by VADAR is the quality index which summarises side chain misfoldings, stereo-chemical clashes and insufficient atom packing for each residue. Residues reported with a quality index <4 are considered to be poorly modelled. The quality index plot of aquaporin-2 is shown in Figure 5. As illustrated, most residues in the aquaporin-2 model are found to be modelled reliably and only few spots of rather poor modelling quality are reported. Analogue to the aquaporin-2 model, the structure model of aquaporin-3 has been generated by means of comparative modelling. However, only the aquaporin-homolog *E. coli* glycerol facilitator (PDB-Id: 1ldf) has been found to be the best matching experimental structure which is acceptable for comparative modelling (43% sequence identity). Subsequent comparative modelling has resulted to an aquaporin-3 model with average reliability (data not shown).

In the process of analyses, the corresponding energy profiles have been computed and aligned in a pairwise manner as elucidated earlier. From this, all pairwise dScores have been derived and used for hierarchical clustering by means of UPGMA. By that, a dScore-distance tree has been generated (see Figure 6). Note, that the longest branch in this tree (0.78) indicates a dScore of 1.56 dits. This dScore corresponds to the distance between the energy profile of aquaporin-3 and the other investigated aquaporins. With the range of highly significant dScores being defined with 0 – 2.5 dits, this observed distance indicates a good energy profile agreement between all investigated structures. Hence, it can be hypothesised that stabilising features and characteristics are conserved in the aquaporin family.

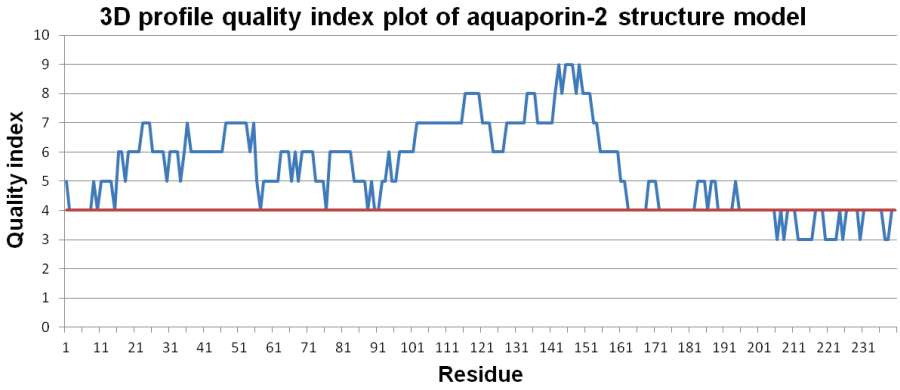

Figure 5: In energy profile-based analyses structure data of the protein of interest is required. In this study, a structure model of aquaporin-2 has been retrieved from the ModBase database (Pieper et al., 2011). Model quality has been assessed by utilising VADAR (Willard et al., 2003). One measure of quality given by VADAR is the residue quality index which reports side chain misfolding, stereo-chemical overlaps and insufficient atom packing. Residues showing a quality index <4 (red line) are assessed as poorly modelled. In this case, the obtained aquaporin-2 model is found to be of good modelling quality.

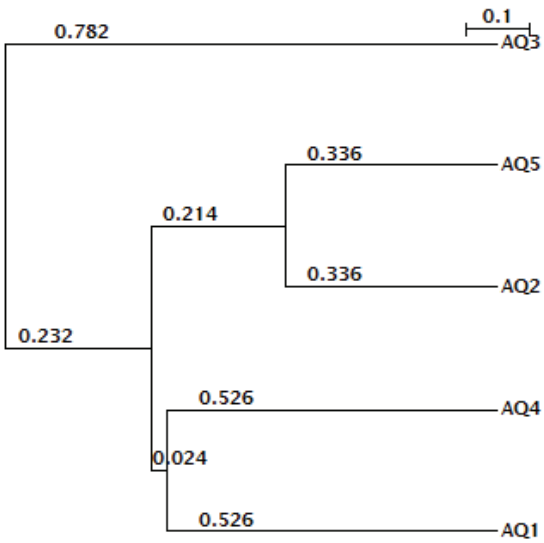

Figure 6: The three-dimensional structures of aquaporin-1, -4 and -5 had been retrieved from the Protein Data Bank and structure models of aquaporin-2 and -3 had been generated. From the pairwise energy profile comparisons of these structures, dScores have been computed and used for generating a dScore-distance tree by means of Unweighted Pair Group Method with Arithmetic Mean (UPGMA) as depicted in this figure. The branch lengths given in this tree are in the range of significant energy profile similarity (dScore < 2.5, corresponding to a branch length of 1.25), indicating strong energy profile commonalities. Since energy profile similarity correlates with common stabilising and functional features in protein structures, it can be proposed that this observation can be applied to the investigated aquaporins as well. We postulate that energetic and, thus, stabilising characteristics are highly conserved in aquaporins. This allows the energy profile-based investigation of destabilising effects in aquaporin-2 mutants which lead to NDI.

To strengthen this hypothesis, we have investigated SMFS data derived from different α-helical membrane proteins concerning possible correlations to energy profiles. We found, that unfolding events can be identified on the level of energy profiles by taking into account the residue-residue contact information that can be computed from the given protein structure. As discussed in the previous section, a contact between two residues i and j is assumed according to Equation 4. The information needed for further investigations is generated by computing the sum of contacts c_i for each residue i. In the following argumentations the sequence of residue-residue contact sums $(c_1, c_2, \ldots, c_i, \ldots, c_n)$ is referred to as the residue-residue contact profile (RRCP). In general, an RRCP is a representation of spatial information. The unfolding events and stabilising effects, that can be detected by utilising SMFS, are manifested in spatial and physo-chemical properties. Thus, taking into account the spatial information yield by RRCPs as well as the information yield in energy profiles, these properties can be identified.

Regarding to aquaporin-2, it can be observed that the RRCPs of aquaporin-1 and aquaporin-2 share common features at a high degree of similarity. With respect to known SMFS and energy profile data as well as the structural similarity between both proteins (p-value = $8.5e^{-12}$, RMSD = 1.45 Å), it can be postulated that SMFS data derived from aquaporin-1 can be directly transferred to aquaporin-2, which emphasises the homology of protein stability in both aquaporins. Experiments carried out on aquaporin-2 might very likely

result to similar observations. According to this and the results given in the previous paragraph, stabilising features are conserved in the aquaporin family. To depict these agreements, the RRCPs and energy profiles of aquaporin-2 and aquaporin-1 as well as SMFS data of aquaporin-1 (e.g. observed SMFS peaks) are illustrated in Figure 7A. By transferring RRCP, energy profile and SMFS data of aquaporin-1 to aquaporin-2, the locations of SMFS unfolding events in aquaporin-2 can be predicted (represented by green bars in Figure 7B).

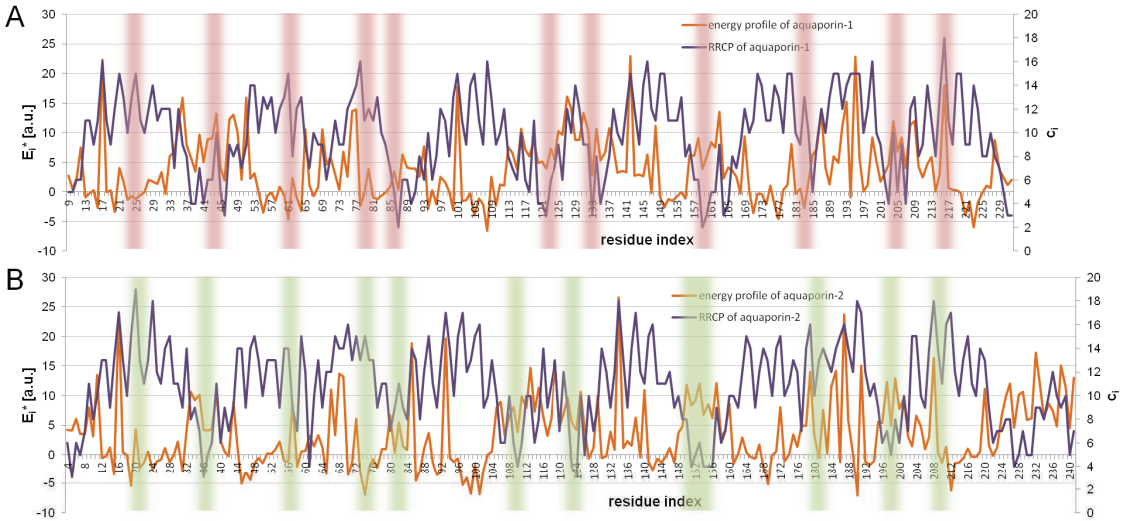

Figure 7: As given by energy profile distances and shown in Figure 6, energy profile characteristics are highly similar in the investigated aquaporin structures and, thus, conserved in the aquaporin family. In this Figure, the energy profile and residue-residue contact profile (RRCP) of aquaporin-1 (see A) and aquaporin-2 (see B) are illustrated. A residue-residue contact is assumed, if $f(i,j) = 1$ (see Equation 4). The sum of all contacts of residue i corresponds to the residue-residue contact number c_i. An RRCP corresponds to the sequence $(c_1, \ldots, c_i, \ldots, c_n)$. We investigated diverse α-helical membrane protein structures with data derived from single-molecule force spectroscopy (SMFS) - an approach for measuring and probing protein energy landscapes and stabilising characteristics (Müller & Engel, 1999; Müller et al., 1999). Correlations have been found in unfolding events (peaks) derived by SMFS experiments to energy profile and RRCP features. The positions of detected SMFS peaks in aquaporin-1 are highlighted by red bars in A. The energy profiles of both aquaporins as well as their RRCPs share common features. Hence, the positions of SMFS peaks in aquaporin-2 are very likely similar located and distinct as observed in aquaporin-1. Predicted SMFS peaks in aquaporin-2 are highlighted by green bars in B

3.3 Protein Stability of Aquaporin-2 Mutants linked to NDI

Over the last decades, numerous aquaporin-2 mutants have been identified in NDI patients. In our analysis, we focused on seven mutations which differ in activity and the phenotype observed in NDI patients and experiments. First, the mutations D150E and L22V+C181W lead to a reduced water flux through the apical

membrane (Guyon et al., 2009; Canfield et al., 1997). In contrast, the mutants T125M+G175R and G196D are experimentally affirmed to prevent water permeability totally (Goji et al., 1998; Guyon et al., 2009). In addition, mutations in the gene encoding aquaporin-2 have been experimentally characterised that cause protein misfolding and prevent routing of the protein to the plasma membrane. The mutants A147T and T126M have been chosen to represent this group of mutants (Mulders et al., 1997). Finally, N68S has been included in the energy profile-based analysis. This mutation is located in the first conserved Asn-Pro-Ala motif. Although it is proven that the N68S mutant does not show water permeability, it is not clear whether the reduced water flux is caused by a disrupted water pore or by protein misfolding and, thus, impaired transport out of the endoplasmic reticulum (Mulders et al., 1997).

To investigate potential correspondences, the mutant structures had been generated using comparative modelling with the aquaporin-5 structure as modelling template. The energy profiles had been computed, aligned and, from this, dScores have been derived. The dScore-distance tree obtained by means of UPGMA (see Figure 8) shows the obtained hierarchical clustering. It can be seen, that mutants leading to an impaired transport are clustered in a single group integrating the N68S mutant as well. According to this, it can be predicted that the bipolar electric field established by α-helices HB and HE is not present in N68S mutants. This affects aquaporin-2 folding significantly in such a way that the native conformation cannot be attained. This leaves N68S mutants in a misfolded and not-active state resting in the endoplasmic reticulum. Furthermore, the mutants D150E and G196D form a separate cluster distant from the correctly folded but (partially) dysfunctional L22V+C181W and T125M+G175R mutants. This indicates the independence of occurring effects induced by these mutations. The high similarity of the energy profiles of D150E mutant and G196D mutant points to close correspondences in the mutated water transport mechanism. However, the observed energetic divergences lead to the significantly different, experimentally confirmed phenotypes. According to our hypothesis, this conclusions are analogue to the mutants L22V+C181W and T125M+G175R, where, as indicated by a branch length of 0.02, the energy profile distance and, thus, energetic divergences are much greater than in the cluster holding the D150E and G196D mutants.

Thus, the observed energy profile distances are in good agreement with experimental data and correspond to water transport activity in aquaporin-2 mutants as well as overall protein stability, since mutants leading to protein misfolding can be distinguished on the level of energy profiles. To validate these correlations, pairwise structure alignments have been computed using the FATCAT algorithm (Ye & Godzik, 2003). With the p-values (all p-values ≈ 0, indicating almost identical C_α trace matches) obtained from these calculations no meaningful clustering could be generated. Hence, the observed clustering is the result of the similar destabilisations and energy discrepancies described by the coarse-grained energy model, and is not caused by structure similarities.

4 Investigation of Protein Stability in V2 Vasopressin Receptor

4.1 Description of V2 Vasopressin Receptor

The V2 vasopressin receptor (V2R) belongs to the class A of the so-called G-protein-coupled receptors. Like most members in this class, V2R consists of seven transmembrane α-helices (Barberis et al., 1998). The residues involved in binding the agonist anti-diuretic hormone arginine-vasopressine (AVP) are basically located in α-helices H2–H5, spanning the four sequence regions 88–96, 119–127, 284–291 and 311–317

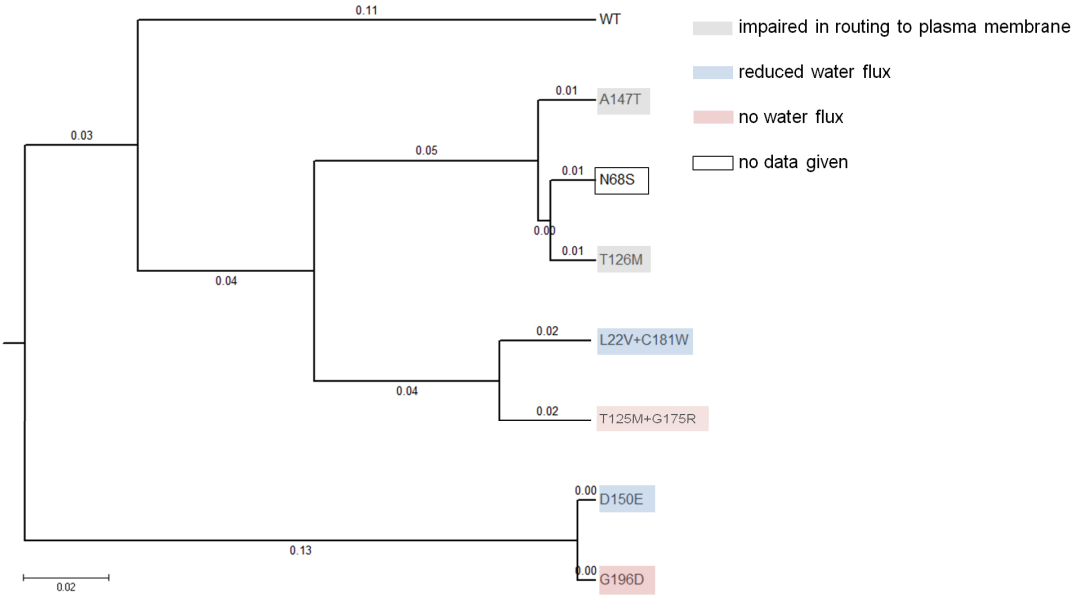

Figure 8: Structure models of seven aquaporin-2 mutants and corresponding energy profiles were generated. Pairwise dScores were computed and used for hierarchical clustering by means of UPGMA. As shown, groupings obtained by clustering correspond to phenotypes and activity observed in experiments. The energy profile of the N68S mutant is arranged to the energy profiles of mutants A147T and T126M. Both mutants are proven to be responsible for incorrect folding, leaving the resulting dysfunctional protein resting in the endoplasmic reticulum. By that, no membrane insertion takes place and water reabsorption is not established leading to symptoms of NDI. From the clustering of both mutants with N68S, it can be predicted that protein structure misfolding is the most likely cause for the loss of functionality in N68S mutants. The clades of (partially) dysfunctional mutants D150E, G196D and L22V+C181W, T125M+G175R, respectively, show relatively long energy profile distances. This indicates that in each clade destabilising effects are highly similar with differences leading to the differing phenotypes observed in experiments.

(Slusarz et al., 2006). Once AVP has bound to V2R, the structure of V2R is passing through multiple allosteric rearrangements. In the subsequent active state, V2R is capable of interacting with cytosolic G-protein activating adenyl-cyclase. As a result of this interaction, a cascade of multiple phosporylation events is taking place, whereas, at last, aquaporin-2 is activated, translocated and finally integrated in the apical membrane (Los et al., 2010; Robben et al., 2006).

For the energy profile-based investigation of bound and unbound states of V2R, a structure model had to be computed since, to this day, no experimentally determined structure is listed in the public structure databases and, because of the limitations of comparative modelling, no reliable model had been generated also. Thus, extensive structure modelling needed to be carried out. By means of the I-TASSER modelling pipeline (Roy et al., 2010) and NAMD2 (Phillips et al., 2005) as the subsequent program of energy minimization and model-reliability assessment, a structure model of V2R has been generated. In further MD simulations, the root mean square fluctuation of the C_α-trace of the V2R structure model has been found to be 2.7 Å which confirmed the good modelling quality that had been additionally evaluated by VADAR (Willard et al., 2003). Furthermore, the docking of AVP to V2R has been simulated utilising the Molecular Docking Server (Bikadi & Hazai, 2009). By this, a structure model of V2R bound to AVP has been generated.

4.2 Analysis of Energy Profiles of V2 Vasopressin Receptor in bound and unbound State

From the two resulting structure models, both energy profiles have been generated and compared directly. The most significant energetic shifts ΔE_i^* (defined as $\Delta E_i^* = E_{i,\text{unbound}}^* - E_{i,\text{bound}}^*$) correspond very well to the sequence regions responsible for AVP binding (see Figure 9). Additionally, a large number of mutants has been identified which are located in these sequences regions and which negatively affect AVP binding. Well-described examples are A84D (Albertazzi et al., 2000), I130F (Pasel et al., 2000; Robben et al., 2005) and P322S (Morin et al., 1998; Vargas-Poussou et al., 1997). Especially Asp 85 shows the greatest energetic shift, e.g. an energetic increase of 8.8 a.u. during binding. As a hydrophilic and polar residue, aspartic acid is more often observed in extra- and intracellular regions of membrane proteins than in the rather hydrophobic environment of membrane spanning helices or membrane-associated regions in general. With respect to the model applied in this study, the value k_{Asp} described by equation 2 is found to be >1 for aspartic acids which are located in membrane-spanning segments. Thus, interactions with membrane-located aspartic acids are approximated to cause destabilising influences on α-helical membrane proteins in general. With focus on the bound and unbound states of V2R, the energetic increase of Asp 85 is caused by interactions to AVP. According to our hypothesis, Asp 85 and the destabilization during and after coupling to AVP are main triggers for the structural rearrangements of the receptor. Previously published results from experimental studies and reports on NDI patients substantiate these conclusions (Sadeghi et al., 1997; Kalenga et al., 2002; Rocha et al., 1999). These studies indicate the responsibility of Asp 85 in coupling to guanine nucleotide-binding G(s) subunit alpha. As reported, D85N mutants show a 20-fold decrease in coupling efficiency. Thus, the application of the energy profile-based approach as a method for investigating protein structure stability is substantiated by these experimental findings.

However, further energy profile-based inspections are limited due to restrictions in modelling large-scale molecular rearrangements.

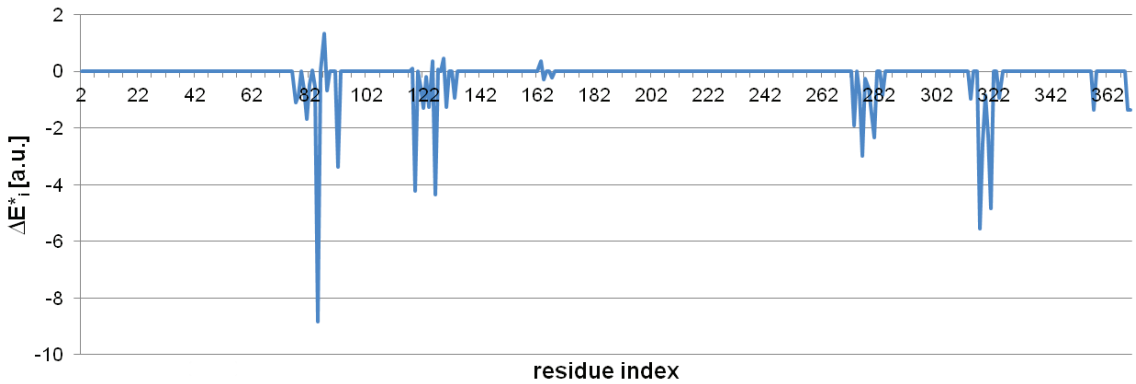

Figure 9: The generation of a human V2R structure model had been carried out by means of molecular modelling and MD. To investigate energetic alternations induced by hormone binding, the docking of arginine vasopressin (AVP) to the structure model has been performed. The superposition of the energy profiles (with $\Delta E_i^* = E_{i,\text{unbound}}^* - E_{i,\text{bound}}^*$) of V2R in bound and unbound state to AVP shows four distinct regions with varying energy values. These regions correspond to the residues mainly involved in AVP binding. The most distinct energetic shift is found at residue Asp 85. From this observation, it can be postulated that the destabilization of the polar and hydrophilic Asp 85 plays a main role in the structural rearrangements observed in V2R after binding to AVP. Previously published data substantiate this conclusion (Sadeghi et al., 1997; Kalenga et al., 2002; Rocha et al., 1999).

5 Conclusion

In this study membrane proteins and mutants which are involved in nephrogenic diabates insipidus have been investigated on the basis of theoretical assumptions. By this, a coarse-grained energy model has been applied which allows the calculation of so-called energy profiles from protein structure data. As shown, energy profiles can be compared and aligned to investigate discrepancies in protein structure, function and stability. In the cases of the analysed proteins and mutants, experimental observations have been substantiated by employing these approaches. Furthermore, predictions concerning effects on protein stability in protein mutants could be made by including experimental data.

Similar to the procedures elucidated in this work, analogue biological questions might be addressed in general in the same manner. Thus, efficient high-throughput in silico techniques might be established that permit the comparison to experimental data and draw valuable conclusions concerning the stability and possible stability variations of proteins of interest.

6 Acknowledgements

The authors would like to thank the Free State of Saxony and the University of Applied Sciences Mittweida for funding. Additionally, very special thanks go to Daniel Stockmann, Steffen Grunert and Michael Spranger for their support, motivation and programming.

References

Albertazzi, E. et al. (2000). Nephrogenic diabetes insipidus: functional analysis of new AVPR2 mutations identified in italian families. *J Am Soc Nephrol*, 11(6), 1033–1043.

Ananthakrishnan, S. (2009). Diabetes insipidus in pregnancy: etiology, evaluation, and management. *Endocr Pract*, 15(4), 377–382.

Apweiler, R. et al. (2004). Uniprot: the universal protein knowledgebase. *Nucleic Acids Res*, 32(Database issue), D115–D119.

Ashburner, M. et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1), 25–29.

Barberis, C., Mouillac, B., & Durroux, T. (1998). Structural bases of vasopressin/oxytocin receptor function. *J Endocrinol*, 156(2), 223–229.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res*, 28(1), 235–242.

Bikadi, Z. & Hazai, E. (2009). Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J Cheminform*, 1, 15.

Birnbaumer, M. (2002). V2R structure and diabetes insipidus. *Receptors Channels*, 8(1), 51–56.

Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016), 164–170.

Brito, G. C. & Andrews, D. W. (2011). Removing bias against membrane proteins in interaction networks. *BMC Syst Biol*, 5, 169.

Canfield, M. C., Tamarappoo, B. K., Moses, A. M., Verkman, A. S., & Holtzman, E. J. (1997). Identification and characterization of aquaporin-2 water channel mutations causing nephrogenic diabetes insipidus with partial vasopressin response. *Hum Mol Genet*, 6(11), 1865–1871.

Chakrabarti, N., Roux, B., & Pomès, R. (2004a). Structural determinants of proton blockage in aquaporins. *J Mol Biol*, 343(2), 493–510.

Chakrabarti, N., Tajkhorshid, E., Roux, B., & Pomès, R. (2004b). Molecular basis of proton blockage in aquaporins. *Structure*, 12(1), 65–74.

Chen, H., Wu, Y., & Voth, G. A. (2006). Origins of proton transport behavior from selectivity domain mutations of the aquaporin-1 channel. *Biophys J*, 90(10), L73–L75.

de Groot, B. L., Frigato, T., Helms, V., & Grubmüller, H. (2003). The mechanism of proton exclusion in the aquaporin-1 water channel. *J Mol Biol*, 333(2), 279–293.

Deen, P. M., Verdijk, M. A., Knoers, N. V., Wieringa, B., Monnens, L. A., van Os, C. H., & van Oost, B. A. (1994). Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science*, 264(5155), 92–95.

Defer, N., Best-Belpomme, M., & Hanoune, J. (2000). Tissue specificity and physiological relevance of various isoforms of adenylyl cyclase. *Am J Physiol Renal Physiol*, 279(3), F400–F416.

Dressel, F., Marsico, A., Tuukkanen, A., Schroeder, M., & Labudde, D. (2007). Understanding of SMFS barriers by means of energy profiles. In *Proceedings of German Conference on Bioinformatics* (pp. 90–99).

Du, Z., Li, L., Chen, C. F., Yu, P. S., & Wang, J. Z. (2009). G-SESAME: web tools for GO-term-based gene similarity analysis and knowledge discovery. *Nucleic Acids Res.*, 37, W345–349.

Finger, C., Volkmer, T., Prodöhl, A., Otzen, D. E., Engelman, D. M., & Schneider, D. (2006). The stability of transmembrane helix interactions measured in a biological membrane. *J Mol Biol*, 358(5), 1221–1228.

Fleming, K. G. & Engelman, D. M. (2001). Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc Natl Acad Sci U S A*, 98(25), 14340–14344.

Fujiwara, T. M. & Bichet, D. G. (2005). Molecular biology of hereditary diabetes insipidus. *J Am Soc Nephrol*, 16(10), 2836–2846.

Goji, K., Kuwahara, M., Gu, Y., Matsuo, M., Marumo, F., & Sasaki, S. (1998). Novel mutations in aquaporin-2 gene in female siblings with nephrogenic diabetes insipidus: evidence of disrupted water channel function. *J Clin Endocrinol Metab*, 83(9), 3205–3209.

Gusfield, D. (1993). Efficient methods for multiple sequence alignment with guaranteed error bounds. *Bull Math Biol*, 55(1), 141–154.

Guyon, C., Lussier, Y., Bissonnette, P., Leduc-Nadeau, A., Lonergan, M., Arthus, M.-F., Perez, R. B., Tiulpakov, A., Lapointe, J.-Y., & Bichet, D. G. (2009). Characterization of D150E and G196D aquaporin-2 mutations responsible for nephrogenic diabetes insipidus: importance of a mild phenotype. *Am J Physiol Renal Physiol*, 297(2), F489–F498.

Hanoune, J., Pouille, Y., Tzavara, E., Shen, T., Lipskaya, L., Miyamoto, N., Suzuki, Y., & Defer, N. (1997). Adenylyl cyclases: structure, regulation and function in an enzyme superfamily. *Mol Cell Endocrinol*, 128(1-2), 179–194.

Heinke, F. & Labudde, D. (2012). Membrane protein stability analyses by means of protein energy profiles in case of nephrogenic diabetes insipidus. *Comput Math Methods Med*, 2012, 790281.

Heinke, F., Schildbach, S., Stockman, D., & Labudde, D. (2013). eProS - A Database and Toolbox for Investigating Protein Sequence-Structure-Function Relationships through Energy Profiles. *Nucleic Acids Res*, Annual Database Issue *(in press)*.

Higgins, D. G., Thompson, J. D., & Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol*, 266, 383–402.

Ilan, B., Tajkhorshid, E., Schulten, K., & Voth, G. A. (2004). The mechanism of proton exclusion in aquaporin channels. *Proteins*, 55(2), 223–228.

Janovjak, H., Struckmeier, J., Hubain, M., Kedrov, A., Kessler, M., & Müller, D. J. (2004). Probing the energy landscape of the membrane protein bacteriorhodopsin. *Structure*, 12(5), 871–879.

Janshoff, Neitzert, Oberdörfer, & Fuchs (2000). Force spectroscopy of molecular systems-single molecule spectroscopy of polymers and biomolecules. *Angew Chem Int Ed Engl*, 39(18), 3212–3237.

Kalenga, K., Persu, A., Goffin, E., Lavenne-Pardonge, E., van Cangh, P. J., Bichet, D. G., & Devuyst, O. (2002). Intrafamilial phenotype variability in nephrogenic diabetes insipidus. *Am J Kidney Dis*, 39(4), 737–743.

Kanehisa, M. & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1), 27–30.

King, L. S., Kozono, D., & Agre, P. (2004). From structure to disease: the evolving tale of aquaporin biology. *Nat Rev Mol Cell Biol*, 5(9), 687–698.

Krysiak, R., Kobielusz-Gembala, I., & Okopien, B. (2010). Recurrent pregnancy-induced diabetes insipidus in a woman with hemochromatosis. *Endocr J*, 57(12), 1023–1028.

Los, E. L., Deen, P. M. T., & Robben, J. H. (2010). Potential of nonpeptide (ant)agonists to rescue vasopressin V2 receptor mutants for the treatment of X-linked nephrogenic diabetes insipidus. *J Neuroendocrinol*, 22(5), 393–399.

Luckey, M. (2008). *Membrane Structural Biology - With Biochemical and Biophysical Foundation*. Cambridge University Press.

Marsico, A., Labudde, D., Sapra, T., Muller, D. J., & Schroeder, M. (2007). A novel pattern recognition algorithm to classify membrane protein unfolding pathways with high-throughput single-molecule force spectroscopy. *Bioinformatics*, 23(2), e231–e236.

Milligan, G. & Kostenis, E. (2006). Heterotrimeric G-proteins: a short history. *Br J Pharmacol*, 147 Suppl 1, S46–S55.

Möller, C., Fotiadis, D., Suda, K., Engel, A., Kessler, M., & Müller, D. J. (2003). Determining molecular forces that stabilize human aquaporin-1. *J Struct Biol*, 142(3), 369–378.

Morin, D., Ala, Y., Sabatier, N., Cotte, N., Hendy, G., Vargas, R., Dechaux, M., Antignac, C., Hibert, M., Bichet, D., & Barberis, C. (1998). Functional study of two V2 vasopressin mutant receptors related to NDI. P322S and P322H. *Adv Exp Med Biol*, 449, 391–393.

Mrozek, D., Malysiak, B., & Kozielski, S. (2006). EAST: Energy Alignment Search Tool. In L. Wang, L. Jiao, G. Shi, X. Li, & J. Liu (Eds.), *Fuzzy Systems and Knowledge Discovery*, volume 4223 of *Lecture Notes in Computer Science* (pp. 696–705).: Springer Berlin / Heidelberg.

Mrozek, D., Malysiak, B., & Kozielski, S. (2007). An optimal alignment of proteins energy characteristics with crisp and fuzzy similarity awards. In *FUZZ-IEEE'07* (pp. 1–6).

Mrozek, D., Malysiak-Mrozek, B., & Kozielski, S. (2009). Alignment of protein structure energy patterns represented as sequences of fuzzy numbers. In *Fuzzy Information Processing Society, 2009. NAFIPS 2009. Annual Meeting of the North American*.

Mulders, S. M. et al. (1998). An aquaporin-2 water channel mutant which causes autosomal dominant nephrogenic diabetes insipidus is retained in the Golgi complex. *J Clin Invest*, 102(1), 57–66.

Mulders, S. M., Knoers, N. V., Van Lieburg, A. F., Monnens, L. A., Leumann, E., Wühl, E., Schober, E., Rijss, J. P., Van Os, C. H., & Deen, P. M. (1997). New mutations in the AQP2 gene in nephrogenic diabetes insipidus resulting in functional but misrouted water channels. *J Am Soc Nephrol*, 8(2), 242–248.

Müller, D. J. & Engel, A. (1999). Voltage and pH-induced channel closure of porin OmpF visualized by atomic force microscopy. *J Mol Biol*, 285(4), 1347–1351.

Müller, D. J., Sass, H. J., Müller, S. A., Büldt, G., & Engel, A. (1999). Surface structures of native bacteriorhodopsin depend on the molecular packing arrangement in the membrane. *J Mol Biol*, 285(5), 1903–1909.

Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48(3), 443–453.

Pasel, K., Schulz, A., Timmermann, K., Linnemann, K., Hoeltzenbein, M., Jääskeläinen, J., Grüters, A., Filler, G., & Schöneberg, T. (2000). Functional characterization of the molecular defects causing nephrogenic diabetes insipidus in eight families. *J Clin Endocrinol Metab*, 85(4), 1703–1710.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem*, 26(16), 1781–1802.

Pieper, U. et al. (2011). ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res*, 39(Database issue), D465–D474.

Pollard, D. & Earnshaw, W. (2007). *Cell Biology*. Springer Verlag Berlin Heidelberg.

Ponder, J. (2001). *TINKER – Software Tools for Molecular Design*. Technical report, Dept. of Biochemistry and Molecular Biophysics, Washington University, School of Medicine, St. Louis.

Robben, J. H., Knoers, N. V. A. M., & Deen, P. M. T. (2005). Characterization of vasopressin v2 receptor mutants in nephrogenic diabetes insipidus in a polarized cell model. *Am J Physiol Renal Physiol*, 289(2), F265–F272.

Robben, J. H., Knoers, N. V. A. M., & Deen, P. M. T. (2006). Cell biological aspects of the vasopressin type-2 receptor and aquaporin 2 water channel in nephrogenic diabetes insipidus. *Am J Physiol Renal Physiol*, 291(2), F257–F270.

Robertson, G. L. (1995). Diabetes insipidus. *Endocrinol Metab Clin North Am*, 24(3), 549–572.

Rocha, J. L., Friedman, E., Boson, W., Moreira, A., Figueiredo, B., Liberman, B., de Lacerda, L., Sandrini, R., Graf, H., Martins, S., Puñales, M. K., & De Marco, L. (1999). Molecular analyses of the vasopressin type 2 receptor and aquaporin-2 genes in brazilian kindreds with nephrogenic diabetes insipidus. *Hum Mutat*, 14(3), 233–239.

Rosenthal, W., Seibold, A., Antaramian, A., Lonergan, M., Arthus, M. F., Hendy, G. N., Birnbaumer, M., & Bichet, D. G. (1992). Molecular identification of the gene responsible for congenital nephrogenic diabetes insipidus. *Nature*, 359(6392), 233–235.

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 5(4), 725–738.

Sadeghi, H., Robertson, G. L., Bichet, D. G., Innamorati, G., & Birnbaumer, M. (1997). Biochemical basis of partial nephrogenic diabetes insipidus phenotypes. *Mol Endocrinol*, 11(12), 1806–1813.

Sadowski, P. G., Groen, A. J., Dupree, P., & Lilley, K. S. (2008). Sub-cellular localization of membrane proteins. *Proteomics*, 8(19), 3991–4011.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4(4), 406–425.

Seelert, H., Dencher, N. A., & Müller, D. J. (2003). Fourteen protomers compose the oligomer III of the proton-rotor in spinach chloroplast ATP synthase. *J Mol Biol*, 333(2), 337–344.

Sippl, M. J. (1993). Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J Comput Aided Mol Des*, 7(4), 473–501.

Slusarz, M. J., Giełdoń, A., Slusarz, R., & Ciarkowski, J. (2006). Analysis of interactions responsible for vasopressin binding to human neurohypophyseal hormone receptors-molecular dynamics study of the activated receptor-vasopressin-g(alpha) systems. *J Pept Sci*, 12(3), 180–189.

Smith, T. F. & Waterman, M. S. (1981). Identification of common molecular subsequences. *J Mol Biol*, 147(1), 195–197.

Sokal, R. & Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38, 1409–1438.

Strom, T. M., Hörtnagel, K., Hofmann, S., Gekeler, F., Scharfe, C., Rabl, W., Gerbitz, K. D., & Meitinger, T. (1998). Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. *Hum Mol Genet*, 7(13), 2021–2028.

Tan, S., Tan, H. T., & Chung, M. C. M. (2008). Membrane proteins and membrane proteomics. *Proteomics*, 8(19), 3924–3932.

Tanaka, S. & Scheraga, H. A. (1975). Model of protein folding: inclusion of short-, medium-, and long-range interactions. *Proc Natl Acad Sci USA*, 72(10), 3802–3806.

Tanaka, S. & Scheraga, H. A. (1976). Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*, 9(6), 945–950.

Tusnady, G. E., Dosztanyi, Z., & Simon, I. (2004). Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics*, 20(17), 2964–2972.

Tusnady, G. E., Dosztanyi, Z., & Simon, I. (2005). PDBTM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res*, 33(Database issue), D275–D278.

van den Ouweland, A. M., Dreesen, J. C., Verdijk, M., Knoers, N. V., Monnens, L. A., Rocchi, M., & van Oost, B. A. (1992). Mutations in the vasopressin type 2 receptor gene (AVPR2) associated with nephrogenic diabetes insipidus. *Nat Genet*, 2(2), 99–102.

Vargas-Poussou, R., Forestier, L., Dautzenberg, M. D., Niaudet, P., Déchaux, M., & Antignac, C. (1997). Mutations in the vasopressin V2 receptor and aquaporin-2 genes in 12 families with congenital nephrogenic diabetes insipidus. *J Am Soc Nephrol*, 8(12), 1855–1862.

Wettschureck, N. & Offermanns, S. (2005). Mammalian g proteins and their cell type specific functions. *Physiol Rev*, 85(4), 1159–1204.

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D., & Wishart, D. S. (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*, 31(13), 3316–3319.

Ye, Y. & Godzik, A. (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, 19 Suppl 2, ii246–ii255.

Zhang, C., Liu, S., Zhou, H., & Zhou, Y. (2004). The dependence of all-atom statistical potentials on structural training database. *Biophys J*, 86(6), 3349–3358.