# The Hidden Intricacies of Aquaporins: Remarkable Details in a Common Structural Scaffold

*Nikolaus Gössweiner-Mohr, Christine Siligan, Kristyna Pluhackova,\* Linnea Umlandt, Sabina Koefler, Natasha Trajkovska, and Andreas Horner\**

Evolution turned aquaporins (AQPs) into the most efficient facilitators of passive water flow through cell membranes at no expense of solute discrimination. In spite of a plethora of solved AQP structures, many structural details remain hidden. Here, by combining extensive sequence- and structural-based analysis of a unique set of 20 non-redundant high-resolution structures and molecular dynamics simulations of four representatives, key aspects of AQP stability, gating, selectivity, pore geometry, and oligomerization, with a potential impact on channel functionality, are identified. The general view of AQPs possessing a continuous open water pore is challenged and it is depicted that AQPs' selectivity is not exclusively shaped by pore-lining residues but also by the relative arrangement of transmembrane helices. Moreover, this analysis reveals that hydrophobic interactions constitute the main determinant of protein thermal stability. Finally, a numbering scheme of the conserved AQP scaffold is established, facilitating direct comparison of, for example, disease-causing mutations and prediction of potential structural consequences. Additionally, the results pave the way for the design of optimized AQP water channels to be utilized in biotechnological applications.

## 1. Introduction

Aquaporins (AQPs), part of a larger family of major intrinsic proteins, are one of the best-studied protein families with currently 20 non-redundant high-resolution structures (≤3.70 Å) solved. Since their first discovery in 1992 by Peter Agre and coworkers,[1,2] thirteen different types of AQPs (AQP0-12) were discovered in mammals. The narrow AQP pores combine enormous permeability, conducting water in a single-file manner close to the diffusion limit of water in bulk, with exceptional selectivity.[3] A subset of AQPs, the aquaglyceroporins (AQGPs), paralogs of AQPs, are also able to conduct glycerol and other small neutral solutes.[4,5] Bacteria also express members of AQPs and AQGPs, generally functioning with one copy of each paralog but, some lacking both. Unicellular eukaryotes and fungi follow a similar pattern, with a clear division between AQPs or AQGPs and a heterogeneous distribution in the number of copies of each paralog in the different genera.[6] So far, no archaea has been found that possesses both paralogs concurrently. Plants exhibit the highest AQP diversity, with five main subfamilies (plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), nodulin-26 like intrinsic proteins (NIPs), small basic intrinsic proteins (SIPs), and X intrinsic proteins (XIPs)), which are each further divided into subgroups.[6] Furthermore, in primitive plant species, two additional subfamilies, GLPF-like intrinsic proteins (GIPs) and hybrid intrinsic proteins (HIPs), have been found.[7] However, the full diversity of AQ(G)Ps is still not represented by the numerous high-resolution structures, as exemplified by only three plant AQP structures and none of the unorthodox AQ(G)Ps (represented by AQP11 and 12 in mammals).

Their critical involvement in cellular water homeostasis and great selectivity renders AQ(G)Ps important in several key areas: i) AQ(G)Ps are expressed in a wide variety of tissues throughout the mammalian body, where they play a role in an extensive range of physiological functions,[5] including water/salt homeostasis, exocrine fluid secretion and epidermal hydration. Due to their important tasks throughout the body, AQ(G)Ps are involved in various human diseases, including glaucoma, cancer, epilepsy, and obesity.[8,9] Mutations in their primary sequence cause genetic diseases like nephrogenic diabetes insipidus, congenital cataracts and keratoderma.[10] Therefore, AQ(G)Ps represent potential drug targets.[9,11,12] Moreover, ii) AQ(G)Ps fulfill pivotal functions in plants, where they also participate in the regulation of cellular water homeostasis,[13] thus steering transpiration sensitivity to soil drying as well as to

N. Gössweiner-Mohr, C. Siligan, L. Umlandt, S. Koefler, N. Trajkovska, A. Horner
Institute of Biophysics
Johannes Kepler University Linz
Gruberstr. 40, Linz 4020, Austria
E-mail: andreas.horner@jku.at
K. Pluhackova
Stuttgart Center for Simulation Science
University of Stuttgart
Cluster of Excellence EXC 2075, Universitätsstr. 32
70569 Stuttgart, Germany
E-mail: kristyna.pluhackova@simtech.uni-stuttgart.de

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smll.202202056.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

NANO · MICRO

**small**

www.small-journal.com

high atmospheric vapor pressure deficit.[14] Hence, they represent the perfect target to address abiotic stresses, like drought, through genetic engineering.[15] Drought is one of the major threats to agricultural production worldwide and therefore, research efforts have focused on the development of drought-resistant crops in the past years. To reach this goal, it is important to explore the underlying molecular mechanisms as well as the interrelated pathways and signaling networks through which AQ(G)Ps induce drought tolerance.[15] Additionally, iii) due to their structural stability and easy handling, AQPs are candidates for building blocks of next-generation filter membranes[16–27] and are already used as a benchmark for newly designed artificial water channels in terms of permeability and selectivity.[28] Thus, concepts of solute and solvent flux through narrow membrane channels can be directly transferred to this challenging and emerging field of material science[29–36] and used to optimize their performance.

AQ(G)Ps exhibit two constriction sites, the aromatic/arginine selectivity filter (ar/R), with its aromatic residue(s) and the positively charged arginine, and the dual asparagine-proline-alanine (NPA) motive, with the structurally opposing half-helices creating a positive dipole moment, both of which are involved in proton rejection.[37–39] Progress in our understanding of mechanisms underlying AQ(G)P functionality was mainly achieved by structural investigations,[4,40–56] sequence comparisons,[7] non-quantitative in vitro assays,[37] yeast complementation assays[37,57] and in silico studies.[38,58,59] It was generally believed that the amino acid composition of the ar/R-region determines the differences in the specificity of AQPs and AQGPs, primarily by affecting the pore size.[37,58,59] However, several studies pointed out that solute specificity cannot be explained by differences in pore size alone.[58,60,61] One of the latest studies emphasized that substrate discrimination in water channels depends on a complex interplay between solute, pore size as well as polarity, and that channel pore size determines exclusion properties but not solute selectivity.[62] Besides water, selected AQ(G)Ps are shown to facilitate a wide variety of neutral molecules across lipid bilayers,[63] like hydrogen peroxide,[57,64] metalloids,[65,66] monocarboxylic acid species,[67] ammonia,[56] and lactate/lactic acid.[68,69] In individual cases, the substrate selectivity was linked to a specific residue, for example, His63 in helix 2 of $At$Tip2;1 is crucial for ammonia specificity,[56] or substitution of Ile254 with Met at the periplasmic end of helix 6 in $Hv$PIP2;3 impairs transport of $CO_2$, while having no impact on water permeability.[70] Alternatively, the presence of Cys residues lining the water pore was suggested to be connected to the ability to transport $H_2O_2$.[71] However, the general molecular determinants enabling the permeability of single substrates across their highly conserved tertiary and quaternary structure remain largely elusive. This may partly be due to the lack of quantitative methods to estimate neutral solute permeabilities.

Quantitative unitary water permeabilities ($p_f$) of AQPs, summarized in **Figure 1**, are rare and span almost one order of magnitude. Currently, diverse methods are used to obtain $p_f$, delivering sometimes more and sometimes less coherent results.[72–76] Only recently, driven by methodological advancements,[3,74,76–78] the major determinants of single-file transport of water through membrane channels were found to be the
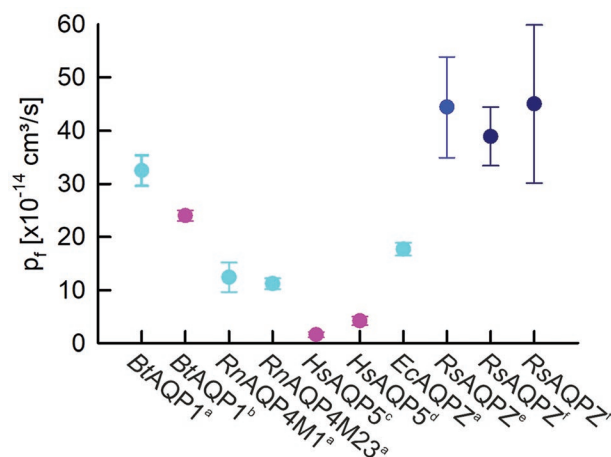


**Figure 1.** Published $p_f$ values of tetrameric AQPs at diverse temperatures. $p_f$ values are included if they are based on accurate channel counting and $p_f$ estimation. a) Stopped-flow light scattering and fluorescence correlation spectroscopy (FCS) counting in *Escherichia coli* polar lipid extract (PLE) at 5 °C.[3,79] b) Micropipette aspiration technique and FCS at room temperature (RT).[72] c) Polarized cell monolayer with FCS at RT.[82] d) Single cell swelling using laser scanning microscopy (LSM) and FCS at RT.[82] e) Stopped-flow light scattering and FCS counting in a 4:1 phosphatidylcholine/ phosphatidylserine (PC/PS) mixture at 10 °C. f) Stopped-flow light scattering and FCS counting in PMOXA-PDMS-PMOXA at 15 °C. Values are temperature-color coded (5 °C – cyan, 10 °C – blue, 15 °C – dark blue, RT – pink).

number of H-bonds water molecules form with pore-lining residues,[3] positive charges at the pore mouth,[79] channel gating,[75] and potentially the shape of the vestibule.[80] Thereby, high $p_f$ is intricately linked to a low activation energy of water transport.[81]

Although many excellent papers and reviews on AQ(G)Ps have been published, covering their evolution,[6,83,84] potential roles in disease,[8–12] AQ(G)P gating,[83,84,86–88] biotechnological applications[27,90–92] and their protein-protein interactions,[93,94] to name a few, a thorough analysis elucidating their subtle structural differences and determining the molecular basis for solute selectivity, protein gating, and protein stability, is missing so far. Even recent structural papers and reviews analyze and compare only small sub-sets of AQ(G)P structures, which does not educe patterns within or between AQ(G)P groups. Here, we aim to fill this gap by analyzing the complete non-redundant set of 20 currently available high-resolution AQ(G)P structures from human and mammals, bacteria, plants, and fungi available in the PDB database, with a resolution between 0.88 and 3.7 Å (**Table 1**). The multitude of AQ(G)P structures offers the unique possibility to study the structural diversity of a class/family of proteins, providing an extensive overview of the differences in scaffold and amino acid distribution, with a strong focus on protomer/tetramer stability, pore geometry, and the channel functionality. This examination brings us one step further in the desire to understand the molecular architecture/construction plan of AQ(G)Ps and membrane channels in general. In this respect, AQ(G)Ps can be seen as a model case of how to conduct similar analysis on other protein families, e.g. urea channels,[95,96] in the future. Moreover, our results presented herein lay the ground to predict the potential

**Table 1.** List of 20 AQ(G)P structures used in this study. A non-redundant list of 16 AQP and 4 AQGP (grey) structures with resolution better than 3.7 Å, excluding low resolution and lower resolved redundant structures as well as mutant proteins. MmAQPM,[40] AfAQPM, AtAQPZ, EcAQPZ,[41] EcGLPF,[4] KpAQY1,[42] PfAQP,[43] OaAQP0,[44] BtAQP0,[45] HsAQP1,[46] BtAQP1,[47] HsAQP2,[48] HsAQP4,[49] RnAQP4,[50] HsAQP5,[51] HsAQP7,[52] HsAQP10,[53] SoPIP2;1,[54] AtPIP2;4,[55] AtTIP2;1.[56] A full list of currently 60 AQ(G)P structure entries available in the PDB databank can be found in Table S1, Supporting Information.

| Order of life | Organism | Short name | PDB | Reso. [Å] | No. residues total/PDB/trunc |
|---|---|---|---|---|---|
| Archaea | Methanothermobacter marburgensis | MmAQPM | 2F2B | 1.68 | 246/1-245/7-241 |
| | Archaeoglobus fulgidus | AfAQPM | 3NE2 | 3 | 246/2-245/8-241 |
| Bacteria | Agrobacterium tumefaciens | AtAQPZ | 3LLQ | 2.01 | 256/1-227/4-222 |
| | Escherichia coli | EcAQPZ | 1RC2 | 2.5 | 231/1-231/4-224 |
| | Escherichia coli | EcGLPF | 1FX8 | 2.2 | 281/6-259/10-252 |
| Protozoa/Yeast | Komagataella phaffii | KpAQY1 | 3ZOJ | 0.88 | 279/11-273/47-261 |
| | Plasmodium falciparum | PfAQP | 3C02 | 2.05 | 258/8-249/12-239 |
| Mammals | Ovis aries | OaAQP0 | 2B6O | 1.9 | 263/5-239/12-220 |
| | Bos taurus | BtAQP0 | 1YMG | 2.24 | 263/6-239/12-220 |
| | Homo sapiens | HsAQP1 | 4CSK | 3.28 | 292/3-235/13-228 |
| | Bos taurus | BtAQP1 | 1J4N | 1.8 | 271/1-247/13-230 |
| | Homo sapiens | HsAQP2 | 4NEF | 2.75 | 242/2-240/12-220 |
| | Homo sapiens | HsAQP4 | 3GD8 | 1.8 | 223/32-254/37-249 |
| | Rattus norvegicus | RnAQP4 | 2ZZ9 | 2.8 | 301/30-253/37-249 |
| | Homo sapiens | HsAQP5 | 3D9S | 2 | 266/1-245/13-222 |
| | Homo sapiens | HsAQP7 | 6QZI | 1.9 | 247/33-279/36-273 |
| | Homo sapiens | HsAQP10 | 6F7H | 2.3 | 301/18-268/23-261 |
| Plants | Spinacia oleracea | SoPIP2;1 | 1Z98 | 2.1 | 281/24-274/40-261 |
| | Arabidopsis thaliana | AtPIP2;4 | 6QIM | 3.7 | 259/32-275/40-267 |
| | Arabidopsis thaliana | AtTIP2;1 | 5I32 | 1.18 | 275/1-238/20-233 |

structural and functional effects of disease-causing mutations in AQ(G)Ps.

## 2. Results and Discussion

### 2.1. AQ(G)Ps Exhibit a Common Scaffold

AQ(G)Ps are homo-tetrameric membrane channels, in which the functional pores reside within each of the four protomers. As an exception, formation of PIP hetero-tetramers in plants[97–99] may represent a novel mechanism to adjust water transport across the plasma membrane.[100] Assemblies of multiple tetramers into orthogonal arrays of particles have been found for an isoform of AQP4,[101–103] with suggested but unproven functional implications on water permeability,[104,105] cell–cell adhesion,[49,101,106] and AQP4 polarization to astrocyte end-feet.[107,108] Transport of ions and gas molecules through the central fifth pore is debated, with missing proof of biological significance.[73,109,110] Generally, each AQ(G)P protomer consists of six transmembrane helices (H) and two half-helices (HH), connected by loops (L). The two half-helices meet at the membrane center and form one out of two selectivity filters within the single-file pore. The opposite dipole moments of the half helices[37–39] with the highly conserved dual NPA motifs at their ends flip water molecules in the middle of the single-file region.[111] The ar/R selectivity filter accommodates the highly conserved positively charged arginine and two aromatic residues, together constituting a major checkpoint for solute discrimination[37] in the narrowest part of the pore. Thereby, a combination of desolvation and electrostatic barriers disrupt proton conduction via the Grotthuss mechanism,[109] with the NPA region imposing the major electrostatic barrier inside the pore according to MD simulations.[112–115] These fundamental AQ(G)P features, the 3D arrangement of α-helices in the protomer, as well as the oligomeric assembly are depicted in **Figure 2**A–C, with bacterial EcAQPZ serving as the archetypical template due to its minimal sequence length, high structural stability, and selectivity for water. The AQ(G)P topology and fold illustrate that the N- and C-termini, as well as loops L2, L3, and L5 are oriented towards the cytoplasm whereas loops L1, L4, L6, and L7 are oriented towards the periplasm. Helices H6, HH1, HH2, and H3 are facing toward the lipid membrane and H1, H2, H5, and H4 form most of the protein–protein interface.

Alignment of all 20 non-redundant AQ(G)P structures revealed a perfectly conserved transmembrane protein scaffold (Figure 2D) with variations in the loops, N- and C-termini. Therefore, we defined a common scaffold of all AQ(G)P structures shown in Figure 2D for further structural considerations. Regions showing some variability in transmembrane helix and loop length are excluded from most of the structural analysis further on. The structural alignment does not only visualize

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com
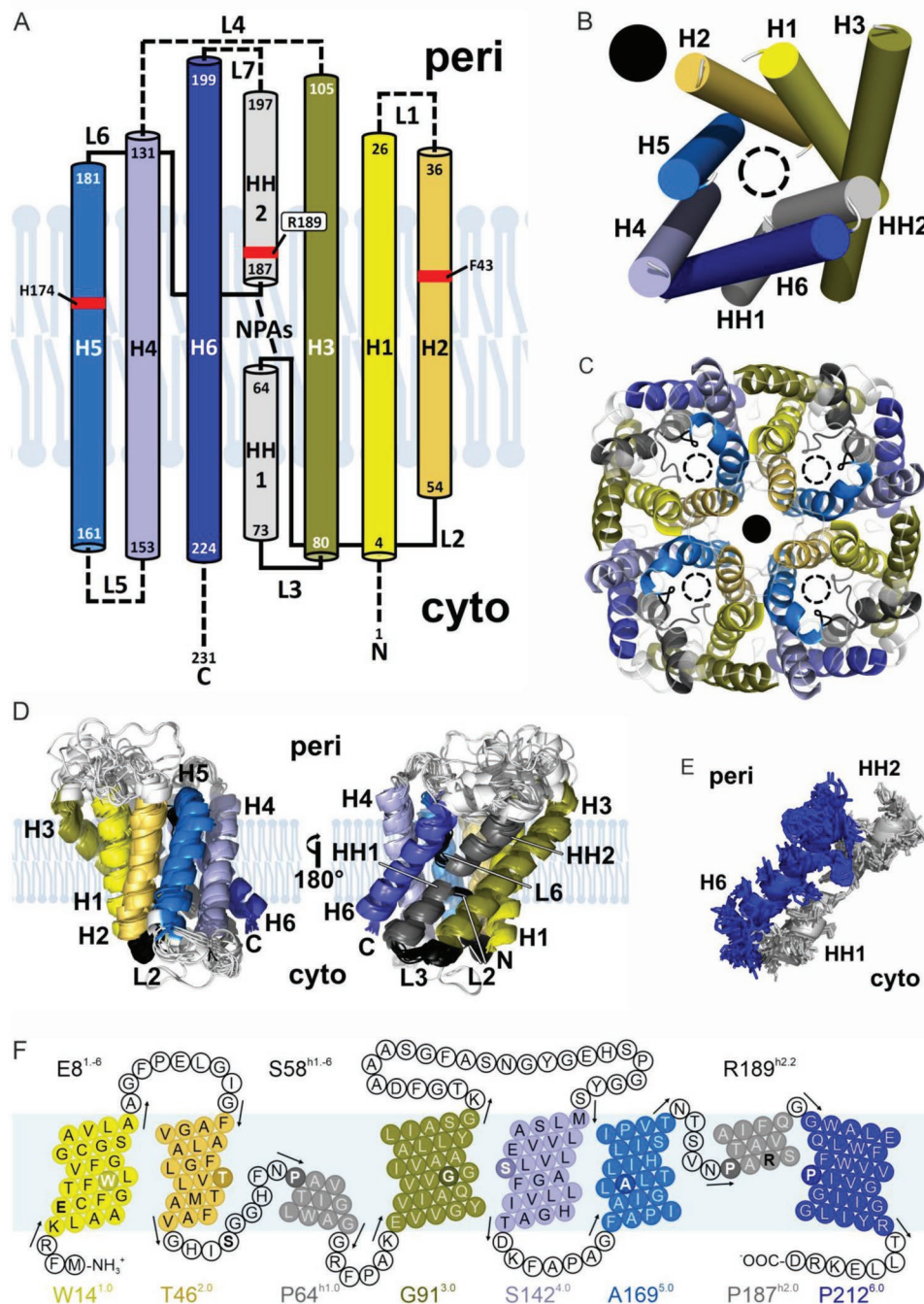
**Figure 2.** The conserved AQ(G)P scaffold. A) Topology map of *Ec*AQPZ. The NPA regions as well as residues of the ar/R selectivity filter (red boxes) are highlighted. Variable loops and termini are depicted with dashed lines. B) Transmembrane helix arrangement of the *Ec*AQPZ protomer (periplasmic view) showing the color-coding of the transmembrane helices as used throughout the manuscript. The water pore is indicated as a black dashed lined circle and the central pore in the tetrameric AQP as a black filled circle. C) Tetrameric assembly of AQ(G)P protomers (periplasmic view of the *Ec*AQPZ structure). Pore positions are indicated as in (C). D) Alignment of the 20 nonredundant AQ(G)P structures using PyMOL reveals a very conserved AQ(G)P scaffold. Loop regions in white are more mobile and variable. The rest of the structure after omitting these mobile loop regions is called "core" throughout the manuscript. E) Representative structural conservation along single amino acid positions of H6, HH1, and HH2. F) Generalized numbering scheme illustrated on a snake-plot representation of *Ec*AQPZ. White bold letters in darker circles represent the residue at the center of the membrane of the corresponding helix, serving as a reference residue in the numbering scheme and are listed at the bottom of each helix. Black bold letters illustrate three examples of highly-conserved residues, that is, Glu8[1,-6],(2) Ser58[h1,-6],(2) R189[h2.2] (*).

the highly conserved α-helical assembly but also the perfectly structurally conserved loops L2, L3, and L6, connecting H2

with HH1, HH1 with H3, and H5 with HH2, respectively. A closer look at the protein backbone (Figure 2E) shows that it is

possible to assign each amino acid in the common scaffold to a respective 3D position within the structure, further supporting the notion of a universal AQ(G)P fold. Moreover, the precisely conserved $C_\alpha$ positions of individual amino acids of all AQ(G)Ps analyzed in this work will confer confidence to expand the analysis of amino acid distributions—including their chemical properties—at secondary or tertiary structural interfaces to a multitude of AQ(G)P sequences with yet unresolved structures in the future.

### 2.2. Numbering Scheme for AQ(G)Ps

The universal AQ(G)P fold with its conserved $C_\alpha$ positions allows us to introduce a numbering scheme for AQ(G)Ps similar to the Ballesteros–Weinstein numbering of G protein-coupled receptors (GPCRs) which was already introduced in 1995.[116] This will provide a framework to relate site-specific properties to the sequences of structurally yet unresolved AQ(G)Ps facilitating structural and functional comparisons and predictions of, for example, disease-causing mutations in mammalian AQ(G)Ps. The numbering scheme is informative of the amino acid (AA) present at that position, the real AA number in a particular AQ(G)P, and the relative position of AAs. All these AA position-specific pieces of information are condensed in identifiers, derived as follows: the AA is labeled by a standard one or three letter code and its real AA number, then the AA identifier in superscript starts with the helix number (1-6), for example, 1, for helix H1. To avoid confusion HH1 is denoted as, h1, and HH2 as, h2. This helix identifier is separated by a point from the position relative to a reference residue which was chosen to be the residue in the center of the lipid bilayer of the corresponding helix (see Figure 2F). That reference residue is assigned the number 0. For example, in H1 it is a tryptophan in position 14 in $Ec$AQPZ, whose identifier would be 1.0, that is, $\text{Trp14}^{1.0}$. A glycine residue located five AA after $\text{Trp14}^{1.0}$ will be $\text{Gly19}^{1.5}$, the glutamate acid residue six positions earlier in the sequence is $\text{Glu8}^{1.-6}$. This general numbering scheme is illustrated in Figure 2F on the snake-plot representation of $Ec$AQPZ and used throughout the manuscript to relate AA positions of different AQ(G)Ps to each other. Mutations are identified in this numbering scheme in the usual manner, with the wild-type residue followed by the mutant AA. For example, $\text{E8K}^{1.-6}$ defines the Lys mutation of the wild-type Glu in helix 1, residue 8 in the $Ec$AQPZ sequence, in the position 1.-6. The identification scheme allows for a systematic comparison of mutations done in different AQ(G)Ps at the same location. Moreover, the assignment of the reference residue to the mid of the bilayer allows for rapid comparison of the helix length on the periplasmic or cytoplasmic side. Figure 2F illustrates the selected reference AA in each helix with the corresponding identifiers according to $Ec$AQPZ as well as three representative identifiers of conserved AAs. According to our numbering scheme, AAs in loops can principally have two identifiers, for example, $\text{Ser58}^{h1.-6} = \text{Ser58}^{2.12}$. We suggest using the identifier related to the next C-terminal helix, for example, $\text{Ser58}^{h1.-6}$ in our example, although both define uniquely the same position. This has the advantage that L2 and L6 are affiliated with the identifiers h1 and h2 of the neighboring half helices HH1

and HH2, respectively. However, it might also be reasonable to number N-terminal residues of L1, L4, L5, and L7 according to their adjacent N-terminal helix, as this emphasizes the structural vicinity and conservation. Clearly, AQP identifiers depict a unique position within the AQ(G)P scaffold if the residue is located within the AQ(G)P core only. Identifiers for AAs located in L1, L4, L6, L7, the N-terminus, and the C-terminus are only relative in nature, as the exact position may vary among AQ(G)Ps due to differing lengths and folds of the corresponding structural element.

### 2.3. Structural Features and Basic Differences of AQ(G)Ps

Despite a common scaffold, AQ(G)Ps are not identical, showing distinct differences. A sequence alignment (Figure S1, Supporting Information) illustrates that all AQ(G)Ps exhibit NPA motives at the ends of HH1 and HH2 except $Pf$AQP and $Hs$AQP7, which possess NLA and NAA motifs at the end of HH1, respectively, and NPS instead of NPA at the end of HH2. While $\text{Arg189}^{h2.2}$ (in $Ec$AQPZ), as part of the ar/R selectivity filter, is perfectly conserved among all AQ(G)Ps, $\text{Phe43}^{2.-3}$, and $\text{His174}^{5.5}$ slightly vary. $\text{Phe43}^{2.-3}$ in $Ec$AQPZ corresponds to a Trp in $Ec$GLPF and $Pf$AQP and a His in $At$Tip2;1. In $Hs$AQP10, $\text{Gly61}^{2.-3}$ at this position provides space for $\text{Phe58}^{2.-6}$, reaching into the same space from one $\alpha$-helical turn above (three residues earlier, toward the N-terminus, and periplasm). The second aromatic residue, $\text{His174}^{5.5}$ ($Ec$AQPZ) is replaced by an Ile in $At$Tip2;1 and AQPMs or by Gly in all AQGPs. In the glycerol conducting AQGPs $Ec$GLPF, $Pf$AQP, $Hs$AQP7, and $Hs$AQP10, $\text{His174}^{5.5}$ is functionally substituted by Phe, Phe, Tyr, and Ile, located in L6, 9 residues further to the C-terminus (identifier h2.-4), respectively. This illustrates quite plainly the difference between AQGPs, AQPs, AQPMs, and of the ammonia permeable $At$Tip2;1 concerning their solute selectivity. Our classification of AQ(G)Ps into these four groups for further analysis throughout the manuscript is confirmed by the sequence similarity map (Figure S2, Supporting Information) and the polygenetic tree of the 20 candidate sequences (Figure S3, Supporting Information). Overall, the sequence similarity of truncated AQ(G)Ps, quantified by estimation of identity scores, varies between 58.3% for $At$PIP2.4 and $Hs$AQP10 and 86.0% for $Hs$AQP2 and $Hs$AQP5. Similarly, denominated AQPs from different organisms vary between 86.5% and 86.9% for AQPZ and AQPM to 98.9% and 98.1% for AQP0 and AQP4, respectively (Figure S20, Supporting Information).

Focusing on the length of helices and loops, shown in Figure S4 and S5 Supporting Information, respectively, following observations were made. While the TM helix length of HH1 is constant in all investigated AQ(G)P structures, HH2 is longer for AQGPs and AQPMs compared to AQPs. No clear pattern can be observed regarding the length of H2, H4, and H5. H6 is shorter in most mammalian AQPs except $Hs$AQP7 and H3 is strikingly longer in AQGPs towards the periplasmic side. The length of H1 is generally most diverse, being longest in plant AQPs and $Kp$AQY1. The lengths of L2, L3, and L6 are perfectly conserved, showing 9, 6, and 5 residues, respectively, whereas L5 is longer in plant PIPs and slightly longer in AQGPs. The length of L1 is most diverse among

all loops, with AQPMs having by far the longest loop with 25 amino acids compared to only six residues in both AQP0. L4 is much longer in AQGPs (*Ec*GLPF, *Hs*AQP7, *Hs*AQP10), slightly longer in *At*AQPZ, *Ec*AQPZ, and *Pf*AQP, and shortest in *At*TIP2;1 and *Kp*AQY1. L7 is longest in AQGPs and slightly elongated in AQPMs and PIPs. In addition to these AQ(G)P specific variations in helix and loop length, there are obvious differences between AQGPs and AQPs in the length of H3 and HH2 as well as the corresponding periplasmic loops L4 and L7, connecting H3-H4 and HH2-H6, respectively. Both the loop and N- and C-termini diversity suggests regulatory functions via the interaction with peripheral proteins.[93] In this respect, archaeal and bacterial AQ(G)Ps as well as *Hs*AQP7 exhibit the shortest termini by far (Figure S5, Supporting Information). Having a closer look at the loops apparent polarity confirms the positive-inside rule,[117] with basic residues being markedly more abundant on the cytoplasmic side (Figures S6–S8, Supporting Information). Nevertheless, the extend and distribution of surface electrostatics varies greatly between the 20 AQ(G)P structures (Figure S8, Supporting Information), in compliance with the possibility to serve as binding platform for peripheral proteins.

### 2.4. Disparities in Pore Geometry and Radius Determine Solute Permeability

Classical AQPs exhibit a long, narrow single-file region with the narrowest part formed by the ar/R selectivity filter. This can be seen by a channel radius of 0.8-1.8 Å for the blue colored exemplary pore radii in **Figure 3**A and the long narrow (green surface) pore surface in Figure 3B for *Ec*AQPZ. In case of *Kp*AQY1, Tyr31[1,-26], part of a putative gating mechanism, occludes the pore on the cytoplasmic side of the channel visualized in Figure S9, Supporting Information. Mutational studies in combination with MD simulations imply that gating by Tyr31[1,-26] may be regulated by a combination of phosphorylation and mechanosensitivity in *Kp*AQY1.[118] AQGPs (red lines in Figure 3A) show a much larger pore radius throughout most of the pore, as is apparent for the respective pore surface representations of *Ec*GLPF and *Hs*AQP10 in Figure 3B. AQPMs show an intermediate pore radius, with a broad pore geometry at the cytoplasmic side similar to AQGPs, yet a narrow ar/R filter at the periplasmic end of the channel, similar to AQPs. This mixed shape of the pore correlates well with AQPMs solute selectivity and the evolutionary analysis (shown in Figure S3, Supporting Information). In detail, i) AQPM has been observed to transport glycerol, albeit at lower rates as compared to *Ec*GLPF.[119] ii) Moreover, a glycerol molecule was bound inside the pore in the crystal structure of *Mm*AQPM.[40] iii) Finally, the evolutionary analysis revealed that AQPMs are an intermediate between AQPs and AQGPs.[6,120,121] However, the biological relevance of the observed increased permeability to larger neutral solutes is questionable in case of the primary host of *Mm*AQPM, *Methanothermobacter marburgensis*, as it relies on $CO_2$ as its sole carbon source,[119] and thus does not require a dedicated glycerol facilitator for survival. The hyperthermophilic sulphate-reducing archaeon *Archaeoglobus fulgidus* additionally encodes a glycerol facilitator *Af*GLPF,[122] utilizing the glycerol derivative diglycerol phosphate as an osmolyte under high-salt conditions.[123,124]

### 2.5. Solute-Specialized AQ(G)Ps Differ Subtly in their Scaffold Structure

AQGPs exhibit a wider pore geometry than AQPs in order to accommodate and conduct neutral solutes like glycerol (Figure 3A). However, as mutational studies of *Ec*AQPZ engineered with three signature amino acids of *Ec*GLPF (F43W[2,-3]/H174C[5.5]/T183F[h2,-4]) suggest,[60] the wider pore geometry and solute selectivity of AQGPs is not solely achieved by a different set of pore lining residues. This proposes that in addition to pore lining residues, different relative positions of one or more α-helices within each protomer to each other could play an equally important role in shaping the NPA region and the ar/R selectivity filter. To locate these subtle structural differences, we first analyzed the RMSD of all AQP structures to each other (Figure S11, Supporting Information). The RMSD plot roughly confirms the polygenetic tree (Figure S3, Supporting Information) and the sequence similarity plot (Figure S2, Supporting Information), clearly grouping AQGPs or mammalian AQPs together. The analysis revealed that the mammalian AQP1 is very much different from the rest of mammalian AQPs, except for AQP0 and AQY1, as well as from bacterial and archaeal AQPs. Plant PIPs are different from *At*TIP2;1 and show a high RMSD when compared to basically all other AQ(G)Ps. The analysis also revealed distinct and unexpected differences between similar proteins from different organisms, as for AQPMs. However, general differences between AQPs and AQGPs could not be established at this level of comparing RMSD of full AQ(G)P structures.

For a more detailed analysis, we next concentrated on the variance of single $C_\alpha$ positions. A comparison between all 20 structures (Figure S12, Supporting Information), with *Ec*AQPZ as a reference, revealed that the periplasmic side is generally more diverse as compared to the core and the cytoplasmic side. Helices H2 and H5 in AQGPs deviate from the rest of the AQPs. *At*TIP2;1 differs on the cytoplasmic side of H2 similar to the AQGPs. AQPMs and PIPs are rather similar to AQPs as compared to AQGPs. On the other hand, the position of the NPA motive including vicinal amino acids and the ar/R selectivity filter are universally conserved.

In order to gain information about the relative directional deviation of single α-helices to each other, we had a closer look on helix-helix distances in five layers along the membrane normal (Figure S13, Supporting Information). Thereby, layer position 1 is located at the periplasmic side, layer position 3 in the middle of the membrane and layer position 5 at the cytoplasmic side (Figure 3C). To find differences between AQPs, AQGPs, and AQPMs, we further averaged the respective helix–helix distances into these three groups (Figure S14, Supporting Information). Visualization of the results reveals subtle differences in the relative α-helix positions within AQGPs as compared to AQPs for each of the five layers (Figure S15A, Supporting Information). Whereas the cytoplasmic vestibule (layer 5) seems to be generally wider, in the center of the membrane, at the height of the periplasmic NPA motive (layer 3), the major
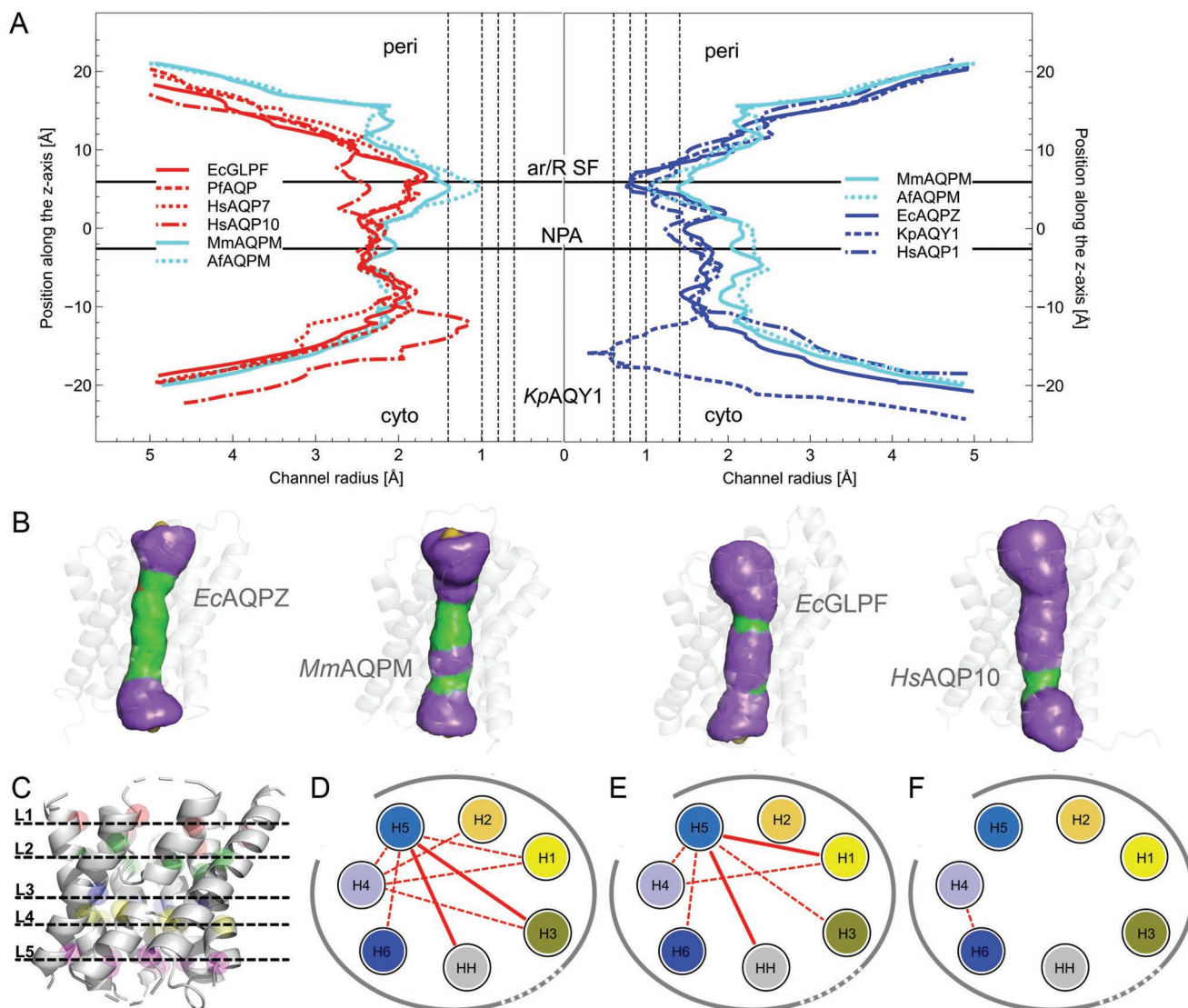
**Figure 3.** Pore geometries and scaffold differences of AQ(G)Ps. A) Pore radii as estimated with the program HOLE.[125,126] AQGPs are indicated in red, 3 representative AQPs are labeled in blue and the two AQPM structures in cyan. The vertical lines guiding the eye highlight pore radii of 0.6, 0.8, 1.0, and 1.4 Å. B) Exemplary pore surfaces of EcAQPZ, MmAQPM, EcGLPF, and HsAQP10 generated by HOLE.[125,126] Protomer structures are depicted in a transparent side view (periplasm up/cytoplasm down). The color scheme according to the program HOLE is the following: Green areas represent pore radii suitable for a single water molecule only, red colored pore surfaces are too narrow to let water molecules pass, and purple regions can accommodate more than one water molecule per cross section. Pore radii and pore geometries of our full set of non-redundant AQP structures can be found in Figures S9,S10, Supporting Information. C) Positions of the 5 layers (shown for EcAQPZ) used to determine helix–helix distances along the channel (shown in Figure S14, Supporting Information). D) Averaged helix–helix distance differences of AQGPs to AQPs, E) AQGPs to AQPMs and F) AQPMs to AQPs. Increase of distance by 0.5–1 Å is shown by a dashed red line and above 1 Å by a solid red line.

difference is an increased distance between H5 and helices H3, H4, as well as H6 by >1 Å. An average over all five layers (Figure S16, Supporting Information) illustrates the major differences in the relative backbone arrangements between AQPs and AQGPs. That is, H5 is shifted towards the central pore of the tetramer, leading to a larger distance to mainly H3 and HH1. Moreover, the distance of H4 to H1, H2, H3, and H5 is increased. Together with the enlarged distance between H4 and H5, these results point to a larger spread of the AQGP helices compared to AQPs (Figure 3D). The increased distance between H4 and H5 could be further verified by measuring their center

of mass distances (Figure S17, Supporting Information). Furthermore, inter-protomer spacing characterized by center of mass distances between (i) contacting helices H2:H5, (ii) single protomers, and (iii) the corner half-helices HH (shown in Figure S17, Supporting Information) revealed both that the distance between protomers of AQGPs is larger by 1.1 Å, and the overall extension of tetrameric AQGPs is larger by 2.6 Å. The differences in the interhelix distances between AQGPs and AQPMs are similar to those of AQGPs and AQPs (Figure 3E). AQPMs and AQPs on the other hand are rather similar in their averaged helix-helix distances (Figure 3F). However, a more

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

detailed view on the individual layers in Figure S15, Supporting Information, shows that the cytoplasmic vestibule of AQPMs is slightly tighter, with H1 in closer contact to H2, and the periplasmic vestibule slightly wider, with H3 oriented outwards the channel center.

The overall increased pore geometry as well as a wider scaffold of AQGPs compared to AQPs strengthens the notion that the selectivity filter is also shaped by the backbone (distances of helices to each other) and not solely by the corresponding side chains. As AQPMs exhibit a similar scaffold as compared to AQPs (Figure 3F) but a comparable pore geometry to AQGPs with low glycerol permeability (Figure 3A), we speculate that it is possible to reach glycerol selectivity with an AQP backbone, yet at low permeability. High glycerol permeabilities require an adaptation of the relative helix positions to each other.

## 2.6. Water is Guided through the AQ(G)P Pore Along a Line of H-Bond Forming Residues

The mobility of pore water in AQ(G)Ps was suggested to be governed by the number of H-bonds, $N_H$, water molecules may form with pore lining residues.[3] The mobility increases in a logarithmic dependence with lower $N_H$, in line with the multiplicity of binding options at higher $N_H$ densities. Except for the ar/R selectivity filter with a multitude of H-bond forming options for single-file water molecules, water is guided through the cytoplasmic half of the channel by forming H-bonds with residues on one side of the channel only, likely in order to stabilize the orientation of the water molecules in the pore and to allow for rapid water flow (Figure S18, Supporting Information). As depicted in **Figure 4** and Figure S19, Supporting Information, most of these interactions occur with backbone oxygens (residues labeled a, b, c, d, i, j, k, l, and m) as weak H-bond acceptors, whereas only five residues form H-bonds to water via their side chains (labeled e, f, g, h, m). All five of those residues are involved in the formation of the selectivity filters. Two out of these residues are the conserved Asns of the two NPA motifs. The remaining three residues belong to the ar/R selectivity filter: $Arg^{h2.2}$ in position f, a prevalently conserved $His^{5.5}$ ($His174^{5.5}$ in $Ec$AQPZ, position g), and position m. The last position is filled with a $Phe^{2.-3}$ in most of the AQ(G)P structures except for $Ec$GLPF, $Pf$AQP, and $At$TIP2;1, which exhibit $Trp^{2.-3}$, $Trp^{2.-3}$ and $His^{2.-3}$, respectively. Only $Trp^{2.-3}$ and $His^{2.-3}$ are able to contribute to H-bond formation by their side chains. Overall, this greatly limits the variability in the number of H-bonds single-file water molecules can form with pore lining residues in position g and m. Generally, nitrogens are thought to build stronger H-bonds than oxygen atoms,[127] with the strength of H-bonds formed with water having a pronounced effect on the H-bond dynamics[128] and inducing a moderate slowdown in the water H-bond exchange dynamics due to an excluded volume effect,[128] similar to that of hydrophobic groups. Such H-bonds are in AQ(G)Ps formed with side chains of the five above mentioned amino acids (Figure 4A labels e, f, g, h, m). In $Ec$AQPZ, the side chain of $Asn182^{h2.-5}$ (which is a Gly in all other 19 structures) is also capable of H-bond formation. All other H-bond donor groups are backbone oxygens, likely designed by nature for increased water permeability.

Detailed analysis of the average number of H-bonds of pore lining residues with pore water molecules in our MD simulations revealed interesting differences (Figure S20). In detail, on average 16.3, 14.1, 17.1, and 16.9 H-bonds were observed for $Ec$AQPZ, $Bt$AQP1, $Hs$AQP4, and $Ec$GLPF, respectively. As the rest of the narrow pore is decorated by hydrophobic residues only, we speculate whether substitution of those hydrophobic residues for AAs capable of forming H-bonds would disturb rapid sliding of water molecules through the narrow pore as well as interfere with high permeabilities of AQ(G)Ps.
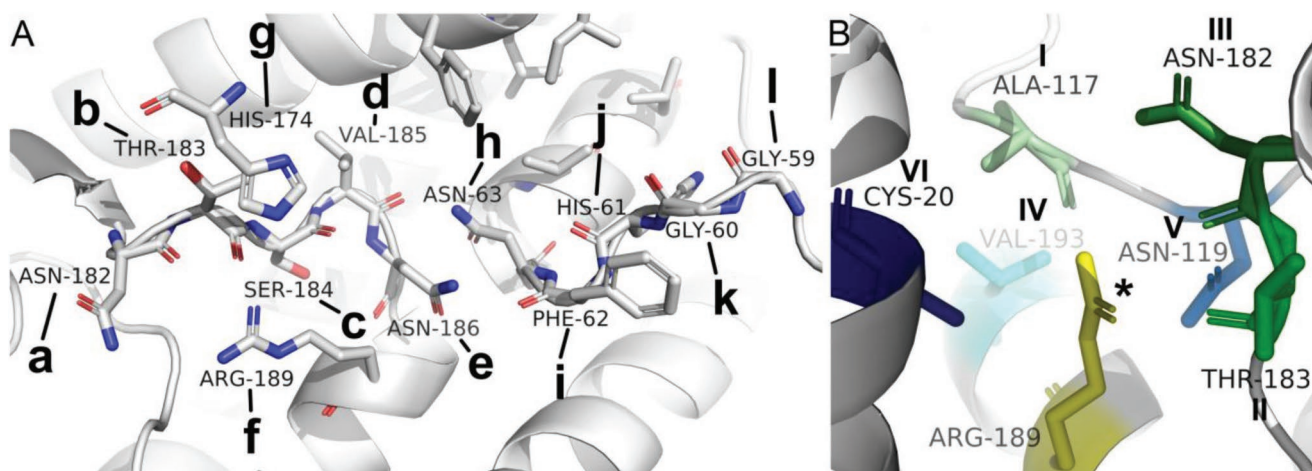
## 2.7. AQ(G)P Gating

Even though AQ(G)Ps are commonly seen as constantly open passive facilitators of water and other neutral solutes, there is a plethora of mechanisms reported influencing their transport capabilities. Water passage can be modulated by movement of side chains of pore lining residues into the pore[54,86,87,89,129] or even by large scale rearrangements of structural elements.[108] A similar process involving two periplasmic loops is also thought to gate the proton gated inner membrane urea channel of *Helicobacter pylori*, $Hp$UreI.[130] In AQ(G)Ps, structurally caused permeability modulations have been proposed to be triggered by pH,[88] divalent cation binding,[131] phosphorylation,[132] mechanical stress,[118] and protein binding[133] and can act at different positions along the pore.

### 2.7.1. Arg in the ar/R filter Region

Numerous indications from in silico[59,134–137] and in vitro[41,138,139] studies point to a conserved and highly flexible $Arg^{h2.2}$ in the ar/R selectivity filter able to block water passage through the single-file water pore of AQPs. This could provide AQPs with a putative transmembrane voltage or lipid asymmetry dependent regulation mechanism. However, gating by the conserved $Arg^{h2.2}$ is still under debate,[140] with a clear experimental proof missing. Herein, we analysed the hydrogen bond network, stabilizing the position of the side chain of $Arg^{h2.2}$ in the channel pore in the 20 high-resolution structures and of four selected targets also by MD simulations, in order to gain a view on its dynamics. In our structural analysis, six neighboring residues were determined which form at least one contact with $Arg^{h2.2}$ in one of the 20 structures (Figure S21, Supporting Information). For illustration, those seven residues (including $Arg^{h2.2}$) were mapped onto the $Ec$AQPZ structure (Figure 4B). It is important to note that no AQ(G)P structure showed interaction with all of those seven residues at once. The nitrogen atoms on the $Arg189^{h2.2}$ side chain may form H-bonds to the backbone oxygens of following residues: $Arg^{h2.2}$ itself (labeled *), the position of $Ala117^{4.-25}$ (I) in L4 and the positions of $Asn182^{h2.-5}$ (III) and $Thr183^{h2.-4}$ (II) in L6. Furthermore, we discovered $Ser^{h2.6}$ in the position of $Val193^{h2.6}$ (IV) in HH2, $Asn^{4.-23}$ in position of $Asn119^{4.-23}$ (V) in L4 and $Glu^{1.6}$ in the position of $Cys20^{1.6}$ (VI) in H1 as potential side chain H-bond interaction sites with the respective $Arg^{h2.2}$ in certain AQ(G)Ps. The presence of $Asn^{4.-23}$ does not necessarily imply formation of an H-bond to $Arg^{h2.2}$. Fascinatingly, the overall amount of

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**NANO · MICRO
small**

www.small-journal.com

**Figure 4.** Line of H-bond donors and acceptors guide water through the AQ(G)P pore. A) Residues in our template *Ec*AQPZ potentially forming H-bonds to passing water molecules according to the sub-Ångström structure of *Kp*AQY1, which for the first time directly visualized H-bonds of crystalized single-file water molecules in the pore with pore lining residues. Through visual inspection of potentially solvent exposed residues in the single-file pore of *Ec*AQPZ, we additionally included Phe62[h1.-2] and Val185[h2.-2]. Indeed, our MD simulations have shown that residues in these two positions are not only capable to form H-bonds with pore waters in *Ec*AQPZ, yet also in *Bt*AQP1, *Hs*AQP4 and *Ec*GLPF (see Figure S20, Supporting Information). Position m (not shown) is only relevant in *Ec*GLPF, *At*TIP2;1 and *Pf*AQP as all other 17 AQ(G)Ps exhibit a Phe[2.-3] in this position (Figures S19, Supporting Information). Residues are labeled from left (periplasm) to right (cytoplasm). For a cytoplasmic and periplasmic view, see Figure S18, Supporting Information. B) Hydrogen bonding network stabilizing the Arg[h2.2] side chain in the ar/R filter. Potential interaction partners observed in the 20 nonredundant AQ(G)P structures are highlighted in the *Ec*AQPZ structure and colored based on the classification of all individual positions (Figure S21, Supporting Information). H-bonding with Arg[h2.2] itself and amino acids I, II, and III occurs with the backbone, whereas amino acids IV, V, and VI interact via their side chains.
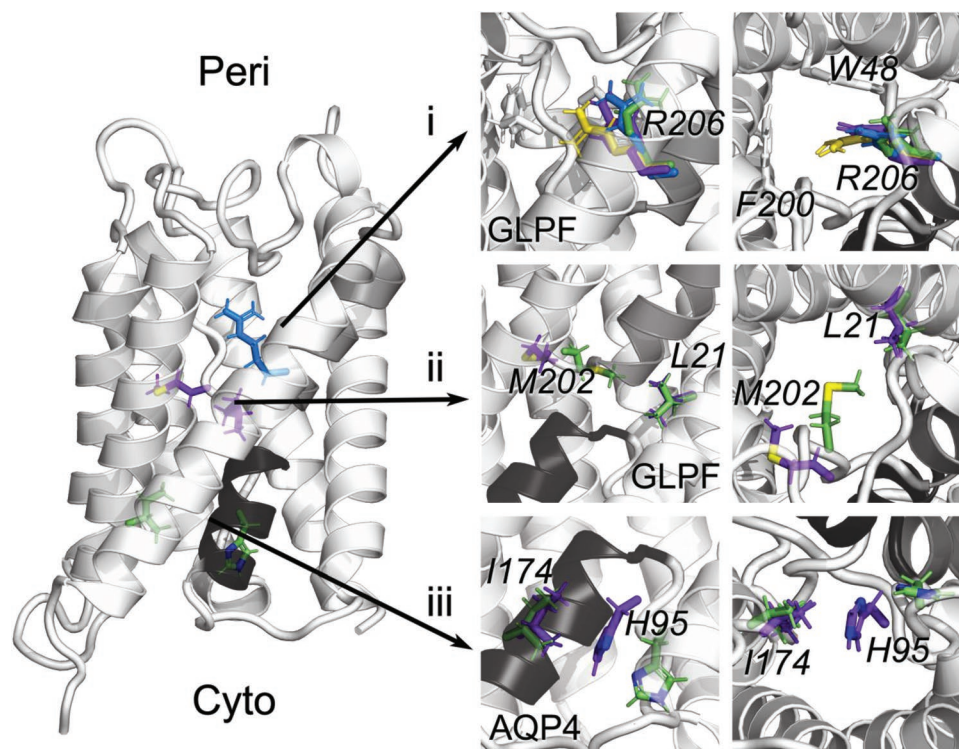
H-bonds stabilizing the respective Arg[h2.2] varies widely, with a minimum of one H-bond in *At*Tip2;1, two H-bonds in *Ec*GLPF, *Hs*AQP7, and *Hs*AQP10, five H-bonds in *Pf*AQP and *At*AQPZ, and a maximum of six potential H-bonds counted in *Hs*AQP5.

To verify the structurally found differences in H-bond stabilization of the conserved Arg[h2.2] and to thereby be able to explain and predict potential Arg[h2.2] gating in vivo, we conducted MD simulations of selected AQ(G)Ps. Indeed, all atom MD simulations of *Bt*AQP1, *Hs*AQP4, *Ec*AQPZ, and *Ec*GLPF confirmed the Arg[h2.2] H-bond interaction partners found in our structural analysis (Figure S22, Supporting Information), with the exception that Arg189[h2.2] in *Ec*AQPZ does also form an H-bond with itself for about 50% of the simulation time and the H-bond to Thr183[h2.-4] is almost never observed in MD simulations. Also, the probability for Arg197[h2.2] to form an H-bond to Ser201[h2.6] in *Bt*AQP1 amounted to only 20%. In addition, it got obvious that the H-bond probability between individual AQP protomers was rather homogeneous for *Bt*AQP1 and *Ec*AQPZ, indicating a relatively stable conformation, while the H-bond network in *Hs*AQP4 was different in chain D compared to the other protomers, hinting at a less stable H-bond network and higher Arg[h2.2] flexibility. Strikingly, in *Ec*GLPF, the situation was completely different, exhibiting a very unstable and disperse H-bond network of neighboring residues with Arg[h2.2]. This finding is underpinned with on average $0.83 \pm 0.12$, $1.84 \pm 0.11$, $2.31 \pm 0.04$, and $2.56 \pm 0.04$ (the error denotes SEM, $n = 4$ chains) H-bonds with neighboring aminoacids in *Ec*GLPF, *Hs*AQP4, *Ec*AQPZ, and *Bt*AQP1, respectively. We further visualized the flexibility of Arg[h2.2] in *Ec*GLPF and chain D of *Hs*AQP4 by plotting the angle of the respective arginines in the pore. As it can be seen in Figure S23, Supporting Information, Arg[h2.2] exhibits indeed larger flexibility. Moreover, the orienta-

tion of Arg[h2.2] in the pore modulates water passage though the respective pores (Figure S24, Supporting Information). While Arg[h2.2] leads to an open pore at a dihedral angle of around 300° and 100°, the pore is closed at an intermediate dihedral angle of 180° (**Figure 5**, subset i). Hence, we conclude that the flexibility of Arg[h2.2] in the pore varies significantly between AQPs, having a severe impact on the rate of water passage through the respective channel. The number of H-bonds found in AQP structures stabilizing Arg[h2.2] may serve to predict this tendency in silico and potentially also in vivo. We speculate that Arg[h2.2] in *Hs*AQP7, *Hs*AQP10, and *At*TIP2;1 are highly flexible as in the case of *Ec*GLPF, but Arg[h2.2] in AQPMs, *So*PIP2;1, and *Rn*AQP4 may be similarly stable as compared to *Hs*AQP4. *Hs*AQP5, *Kp*AQY1, *At*AQPZ, and *Pf*AQP seem to have an even more refined H-bond network comprising one or two additional H-bonding partners.

### 2.7.2. Dynamic Constriction Regions Identified by MD Simulations

In addition, our MD simulation unraveled two more locations where pore lining residues modulate water flux by correlating the positions of the pore lining residues with the number of water passage events through *Ec*GLPF, *Hs*AQP4, *Ec*AQPZ, and *Bt*AQP1 over time (Figure S24, Supporting Information). The reduction of water permeability is thereby caused by (temporary) translation of the respective side chains into the channel pore, namely the conserved His[h1.-3] at the cytoplasmic end of the water pore region (Figure 5, subset iii) and Met[h2.-2] directly preceding the second NPA motif (Figure 5, subset ii). Met[h2.-2] is conserved in 8 out of 20 here investigated AQ(G)Ps, all other structures show hydrophobic Ile, Leu, or Val in the respective

**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

NANO · MICRO
**small**
www.small-journal.com

**Figure 5.** AQ(G)P gating in MD simulations. MD simulations of *Bt*AQP1, *Hs*AQP4, *Ec*AQPZ, and *Ec*GLPF have revealed three possible dynamic gating sites by residues moving into the pore lumen. Gating site (i) is the conserved Arg$^{h2.2}$ in the ar/R selectivity filter. Arg$^{h2.2}$ can occupy four different positions, described here by the dihedral angle $C_\alpha C_\beta C_\gamma C_\zeta$. The angle of 320° (colored green) represents an open state in the position of Arg$^{h2.2}$, similar to that observed in most crystal structures. Arg$^{h2.2}$ in *Bt*AQP1 and *Ec*AQPZ only rarely leaves this position. The angle of ≈270° (colored blue) represents another open state, however, with the guanidyl group slightly more protruding to the pore and thus slowing down the water passage. While the pore is blocked by Arg$^{h2.2}$ at the angle of ≈200° (colored yellow), further reduction of the dihedral angle to ≈100° (colored purple) results in pore opening again. The second restriction site (ii) is characterized by large scale movement of Met$^{h2.-2}$ into the pore lumen, here shown for Met202$^{h2.-2}$ in *Ec*GLPF and quantified by measuring the minimum distance of Met202$^{h2.-2}$ to Leu21$^{1.1}$. At short distances (<0.4 nm) the pore is blocked (shown in green), at larger distances (colored purple) the pore is open. The third constriction site is the highly conserved His$^{h1.-3}$ in the cytoplasmic pore lumen (iii). The pore closure is visualized for *Hs*AQP4 and measured by estimating the minimal distance between the His95$^{h1.-3}$ and Ile174$^{4.7}$. Small distances (shown in purple) result from flipping of the histidine into the pore and lead to its closure. At long distances (His95$^{h1.-3}$ and Ile174$^{4.7}$ shown as green sticks) the pore is open. The middle panels show side views on the constriction sites and the right panels show views from the periplasmic (i and ii) and cytoplasmic (iii) side into the pore.

position, which anchor loop L6 well in the hydrophobic surroundings. The backbone of this residue is stabilized by a hydrogen bond with the conserved Glu$^{4.-4}$ (Figure 9, label 10) in H4.

Met202$^{h2.-2}$ positioning in the *Ec*GLPF pore, measured by a minimal distance between Met202$^{h2.-2}$ and Leu21$^{1.1}$, strongly modulates water passage in our simulation (Figure 5ii, Figure S24, Supporting Information). While a distance of around 0.7 nm allocates an open pore, distances of around 0.3–0.4 nm correspond to a closed pore. Met212$^{h2.-2}$ of *Hs*AQP4 in the same position does not exhibit gating behavior over the course of our simulations. The slightly reduced average minimal distance of 0.6 nm between Met212$^{h2.-2}$ and Phe48$^{1.1}$ in *Hs*AQP4 as compared to *Ec*GLPF is in agreement with a smaller pore radius at the methionine position (Figure S16, Supporting Information, position along the pore of about 0 Å corresponding to the membrane center). In order to elucidate the reason for this difference, we analyzed the H-bonds stabilizing the backbone of these methionines (Figure S23, Supporting Information). Indeed, the H-bonding of the conserved Glu$^{4.-4}$ (Glu163$^{4.-4}$ and Glu152$^{4.-4}$ in *Hs*AQP4 and *Ec*GLPF, respectively) is much

less frequent in *Ec*GLPF, compared to *Hs*AQP4. Especially, in chain B of *Ec*GLPF, exhibiting the largest closure probability by Met202$^{h2.-2}$ (Figure S24, Supporting Information), Met202$^{h2.-2}$ is H-bonded to Glu152$^{4.-4}$ only for 10% of the simulation time (Figure S23, Supporting Information). This observation confirms our supposition that the movement of this methionine to the pore lumen is interconnected with deficient anchoring of its backbone to the rest of the protein. Recently, methionine in the hydrophobic core of the spidroin protein was shown to bestow the protein with exceptional dynamics, enabling it to adjust its shape and thus to customize its function.[141] Unfortunately, the literature evidence about the role of methionine in AQ(G)Ps is scarce. Met212$^{h2.-2}$ in *Hs*AQP4 was suggested to directly interact with His201$^{5.5}$ in the ar/R filter and Met209$^{h2.-2}$ in *Hs*AQP8 was suggested to be involved in protection mechanism of cells against damage by reactive oxygen species.[71] For the first time, our results here point to the water-flux mediating role of Met$^{h2.-2}$ positioned in the vicinity of the NPA and ar/R motifs inviting further experimental and simulational investigations.

Gating by the conserved His$^{h1.-3}$ at the cytoplasmic vestibul was first suggested by Janosi et al. for *Hs*AQP5,[142] followed by

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**NANO · MICRO
small**

www.small-journal.com

Alberga et al. for *Hs*AQP4,[143] and soon afterward it was shown to be further influenced by pH.[144] While pH gating will be discussed in the next section, here, we concentrate on His[h1.-3] gating in a single protonation state. Our MD simulations show that besides *Hs*AQP4 also *Bt*AQP1 and *Ec*AQPZ, unlike *Ec*GLPF, can be gated by the movement of the cytoplasmic His[h1.-3] into the pore lumen (Figure S24, Supporting Information). Our analysis of H-bond patterns between the cytoplasmic His[h1.-3] and various interaction partners has revealed consistent stabilisation of His[h1.-3] by the neighboring amino acids. In detail, the most stable anchoring occured in *Ec*GLPFs (His66[h1.-3] is H-bonded by Glu14[1.-6] and Thr72[h1.3]), rather stable anchoring was observed for His76[h1.-3] by Glu17[1.-6] in *Bt*AQP1 and for His61[h1.-3] by Glu8[1.-6] and Thr67[h1.3] in *Ec*AQPZ, while no systematic stabilisation took place in *Hs*AQP4 (Figure S23, Supporting Information).

### 2.7.3. pH Gating

Numerous in vivo, in vitro, and in silico studies indicate pH gating of AQ(G)Ps and pin-point this effect to specific His residues as pH gates in the respective structures.[53,54,88,131,142–155] Analyzing these publications, we were able to structurally separate five regions/positions in AQ(G)Ps where protonation/deprotonation of His residues was shown to modulate AQ(G)P function. We were overwhelmed by the fact that such a small and "simple" protein is proposed to be gated by pH via so many different ways, even though the molecular mechanisms of AQ(G)P modulation remains elusive in most of the referenced cases. One study on *Hs*AQP3 in lung cells indicated an additional His[2.-13] (putative sixth gating site) in L1 at the periplasmic side,[149] which we skipped in our analysis due to the fact that there is no high-resolution structure for AQP3 available, yet. **Figure 6** illustrates all 5 potential pH regulation sites in AQPs and AQGPs found in the literature (colored yellow and in green colors), mapped onto the *Ec*AQPZ structure as well as three additional His positions conserved in AQPs (shown in different blue tones). The latter histidines are found neither in any AQGP, nor in any AQPM. Only two pH gating sites have been structurally resolved so far, namely the pore exclusion of PIPs induced by His[5.-12] in L5 (bright green, E)[54,88] and a different side chain orientation in *Hs*AQP10 of the conserved His[h1.-3] at the cytoplasmic end of the single-file pore (yellow, A).[53] The
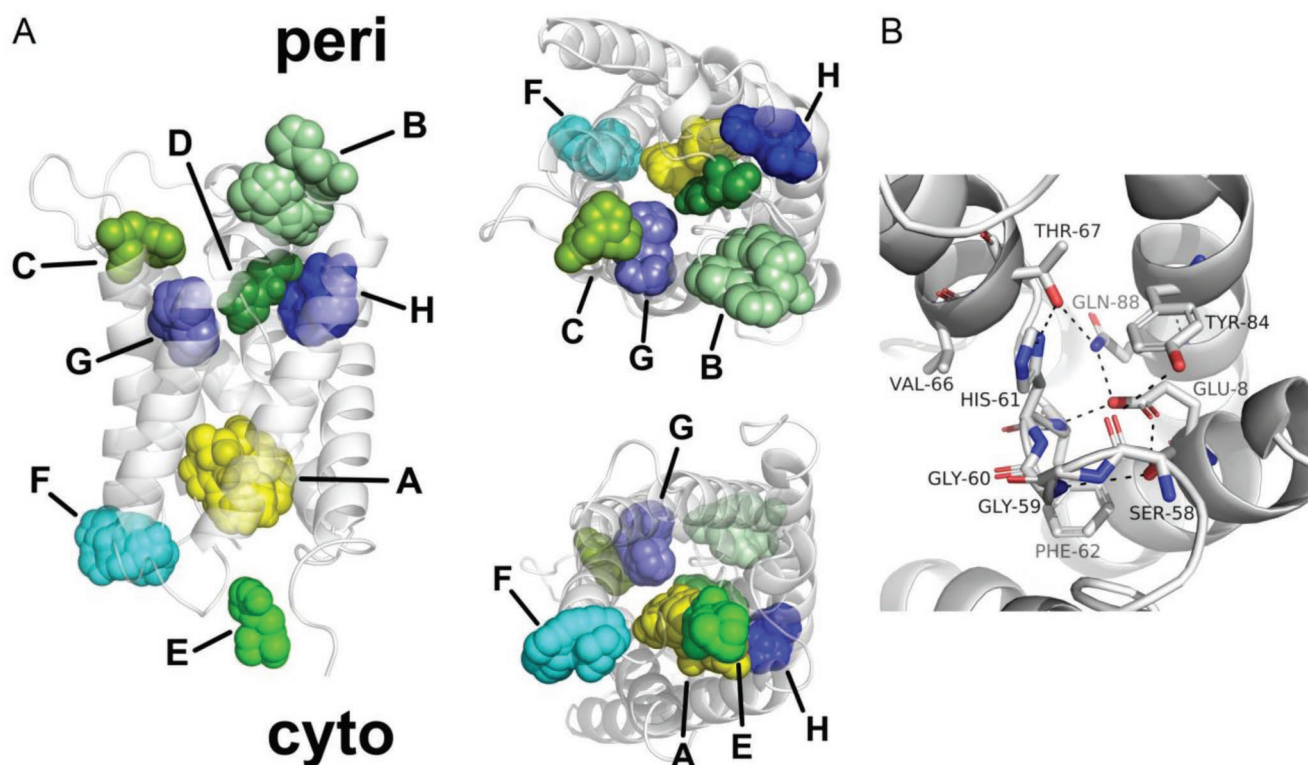


**Figure 6.** Possible pH gating sites and conserved His residues in AQ(G)Ps. A) His positions among AQ(G)P structures potentially involved in pH gating shown in *Ec*AQPZ. Color code according to the classification of individual positions listed in Figures S25 and S27, Supporting Information. Please note, that in no AQ(G)P all here highlighted histidines are present at once. Occurrence of the respective His is 19, 5, 2, 1, 2, 8, 13, and 10 out of 20 for the positions A–H. Position A, His[h1.-3], at the cytoplasmic side of the pore is conserved in all AQ(G)Ps except in *Kp*AQY1. Position B, His[4.-14] or His[4.-16], in L4 at the periplasmic side appears in *Ec*GLPF, *Oa*AQP0, *Bt*AQP0, *Hs*AQP4, and *Rn*AQP4. Positions C, His[2.-11], in the periplasmic pore in L1, D, His[4.-18], in L4 in the periplasmic pore and E, His[5.-12], in L5 at the cytoplasmic mouth contain His only in AQP0s, *At*Tip2;1 and PIPs, respectively. Position F, His[h1.-8], in the cytoplasmic vestibule is conserved in *Ec*AQPZ, *Oa*AQP0, *Bt*AQP0, *Hs*AQP1, *Bt*AQP1, *Hs*AQP2, *Hs*AQP4, and *Rn*AQP4. Position G, His[5.5], is a part of the ar/R selectivity filter in all AQPs except AQGPs, AQPMs and *At*Tip2;1. Position H, His[6.-6], in H6 at the periplasmic side is conserved in all classical AQPs and *At*Tip2;1 except for PIPs. In addition, periplasmic and cytoplasmic views are depicted. B) Hydrogen bond network stabilizing the position of His61[h1.-3] (position A, yellow) in L2 (*Ec*AQPZ). Hydrogen bonds connecting Glu8[1.-6] (2), Ser58[h1.-6], Gly60[h1.-4], Phe62[h1.-2], His61[h1.-3] (A), Thr67[h1.3] (5), Tyr84[3.-7] (6), and Gln88[3.-3] (7) in sticks are indicated with dashed lines.

other positions in green color were located solely by computational and functional studies. In detail, position A, His[h1.-3], was located using MD simulations and experiments with $Hs$AQP5[142] and $Hs$AQP4.[143,144] Position D, His[4.-18], was identified in grapevine $Vv$TIP2;1[154] using yeast cell assays and in $At$TIP5;1[147] in oocytes. Position C, His[2.-11], is important in AQP0[145,146] as shown by pH dependent water measurements with oocytes. Position B, His[4.-14], was located in $Bt$AQP0 and $Rn$AQP4 using oocytes[146] in $Hs$AQP3[149] in lung cells and in $Hs$AQP7 using yeast cells and MD simulations.[153] Figure S25, Supporting Information, gives an overview of the number of His residues in the respective AQP structure and the occurrence in the respective potential pH gating site. A potential pH gating site which deserves detailed inspection is the pore lining His[h1.-3] at the cytoplasmic entrance (within L2) labeled A and colored in yellow in Figure 6A, conserved in all investigated AQ(G)Ps except for $Kp$AQY1. MD simulations of $Hs$AQP5[142] and $Hs$AQP4,[143,144] experimental studies on $Hs$AQP4[144] and $Hs$AQP10[53] and structural investigations on $Hs$AQP10[53] revealed this His residue as a pH gate. However, the function of several other AQPs carrying His in this position, including AQP1, were reported not to depend on pH. In an attempt to distinguish whether these differences result from inconclusive measurements or from differences in the local surroundings of the histidines, we have analyzed the H-bonds networks stabilizing His[h1.-3] in the pore (Figure 6B). However, the amino acids involved in this network are largely conserved throughout the analyzed AQP structures (Figure S26, Supporting Information). An H-bond network analysis similar to the one performed for the conserved Arg[h2.2] in the selectivity filter revealed that His[h1.-3] in position A is stabilized in 15 out of 20 structures at its side chain and in 19 out of 20 cases also at its backbone. However, we found no obvious correlation of the number of H-bonds stabilizing His[h1.-3] itself or L2, housing His[h1.-3], with AQ(G)Ps reported to be gated by pH at this position. Another explanation might be the variability in pKa values between 2.07 and 5.14 found using the program propka[156,157] (Figure S27, Supporting Information). Hence, shifted pKa values of the respective His residues may be responsible for the differences in pH gating. However, considering the limited accuracy of the pKa predictions and the limited number of functional studies, no obvious correlation could be described either.

Interestingly, even though our structural analysis of the H-bond networks did not reveal any obvious indication why His[h1.-3] may gate water flux through some AQPs but not others (Figure S24, Supporting Information), we found differences in His[h1.-3] flexibility in our MD simulations. Whereas His[h1.-3] formed stable H-bonds in $Bt$AQP1, $Ec$AQPZ, and $Ec$GLPF, it was very mobile in $Hs$AQP4, hardly forming any H-bonds with neighboring residues (Figure S23, Supporting Information). $Bt$AQP1, $Ec$AQPZ, and $Ec$GLPF showed significant H-bond probabilities with the conserved Glu[1.-6] (position 2) in H1 and the conserved Thr[h1.3] six positions towards the C-terminus located in HH1 (position 5 in Figure 9). In addition, all three AQ(G)Ps formed H-bonds with the conserved Ser[h1.-6] in L2 three residues towards the N-terminus (position 2 in Figure 9), although only in the case of $Bt$AQP1 to a significant extend. Partly in line with these in silico results, our structural analysis revealed that the side chain of the conserved His[h1.-3] forms

either an H-bond with itself as in the case of $Bt$AQP1, the mentioned Thr[h1.3] six positions towards the C-terminus located in HH1 (Figure S26, Supporting Information) or its backbone with the conserved Glu[1.-6] in H1. However, interactions with Ser[h1.-6] in L2 are not obvious from the high-resolution structures. Intriguingly, a recent in silico study suggested the role of water coordination around His[h1.-3] and interactions with the conserved Glu[1.-6] in H1 in the gating behavior of $Hs$AQP10.[158] Similarly, we found major differences in the interaction with Glu[1.-6], as outlined above and in the average number of H-bonds the His forms with pore water molecules (Figure S20, Supporting Information). His95[h1.-3] of $Hs$AQP4 forms at neutral pH on average $0.7 \pm 0.5$ H-bonds with other pore lining amino acids, with $0.3 \pm 0.3$ H-bonds to Glu[1.-6] in H1 as calculated from the data presented in Figure S23, Supporting Information—the errors denote SEM, $n = 4$ chains. On the contrary, His76[h1.-3] in $Bt$AQP1, His61[h1.-3] in $Ec$AQPZ, and His66[h1.-3] in $Ec$GLPF form $1.5 \pm 0.1$, $2.0 \pm 0.1$, and $1.9 \pm 0.04$ H-bonds with other amino acids and $1.1 \pm 0.04$, $1.1 \pm 0.03$, and $1.0 \pm 0.02$ H-bonds particularly with Glu[1.-6], respectively. Vice versa, His95[h1.-3] of $Hs$AQP4 formed on average $3.2 \pm 0.2$ H-bonds with pore lining water molecules while His76[h1.-3] in $Bt$AQP1, His61[h1.-3] in $Ec$AQPZ, and His66[h1.-3] in $Ec$GLPF formed only $2.5 \pm 0.1$, $2.6 \pm 0.2$, and $2.2 \pm 0.2$, respectively (Figure S20, Supporting Information). Hence, as our analysis of the crystal structures does not match the results from MD simulations, we suggest that His flexibility and pH gating reported in previous research may be due to differences in the vicinal water dynamics impinging on the stabilizing H-bonds of His[h1.-3].

## 2.8. Roles and Stability of the Tetrameric Fold of AQ(G)Ps

The AQ(G)P tetramers are stabilized via hydrophobic interactions at the protomer-protomer interface (**Figure 7**A). However, our analysis of all AQ(G)P interfaces revealed large differences in the shape and pattern of the hydrophobic protein-protein interfaces (Figure S28, Supporting Information) compared to the rather uniform hydrophobic belt on the protein-lipid surface surrounding the tetrameric assembly (Figure S29, Supporting Information). The central interfaces of the here studied AQ(G)Ps exhibit almost perfect surface complementarity with a differently pronounced dent at the center of the membrane toward the lipid bilayer. Additional stabilization is ensured by H-bonds and salt bridges. These are often necessary to imply specificity of the interaction interface.[159] Salt-bridges at the protomer-protomer interface of the core proteins (i.e., proteins with truncated N- and C-terminus and without L1, L4, L5, and L7) are rare: only four AQ(G)Ps are stabilized by a salt-bridge in this region, and only $Af$AQPM and $Ec$AQPZ exhibit multiple salt-bridges, that is, 3 and 2, respectively. Also, the full structures are stabilized by salt-bridges only in 8 out of 20 AQ(G)Ps. The number of hydrogen bonds at the protomer-protomer interface varies between 4 for $At$Tip2;1 to 15 for $Kp$AQY1. The situation changes when considering only the truncated protein versions (missing the N- and C-termini), with 13 remaining H-bonds for $Hs$AQP7 and $Hs$AQP10 and only 3 H-bonds stabilizing the tetramer for $At$Tip2;1 and $Kp$AQY1 (Table S2, Supporting Information). While most of the formed H-bonds are
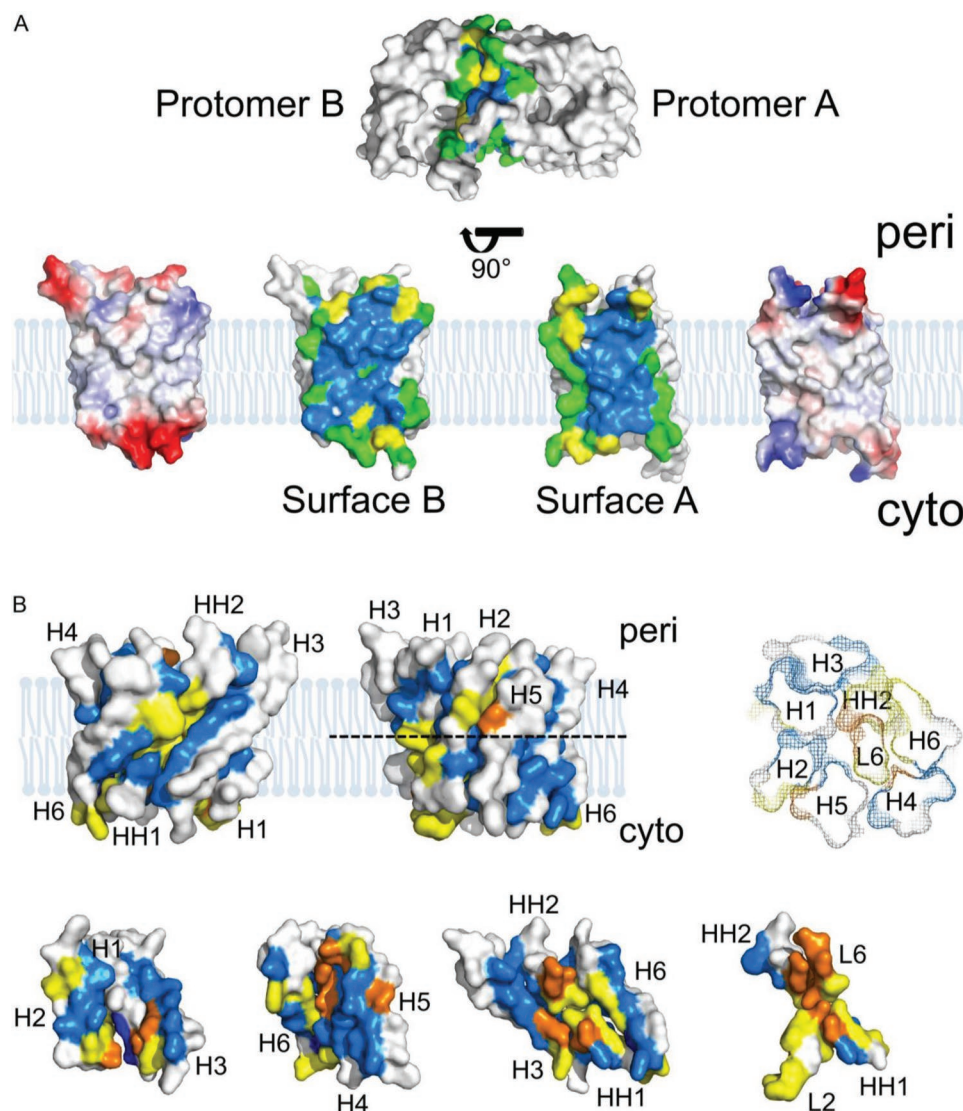
**Figure 7.** Exemplary inter- and intra-protomer interactions of *Ec*AQPZ. A) Hydrophobic residues that are involved in protomer-protomer surface interactions are colored in green (low degree of buried surface) and blue (high degree of buried surface). Residues involved in hydrogen bond formation are highlighted in yellow. The pictures at the very left and very right in the bottom row show net surface charge representations as estimated by PyMOL: white (no charge – hydrophobic), red (negative charge) and blue (positive charge) coloring. B) Residues involved in hydrophobic interactions are colored in light blue (single interaction) or dark blue (multiple interactions). Residues forming hydrogen bonds are labeled in yellow (single hydrogen bond) or orange (multiple hydrogen bonds formed). The projection of a horizontal cut within the *Ec*AQPZ protomer (dashed line) is shown in mesh representation at the top-right and the helix numbers are indicated (periplasmic view). Representative interaction surfaces among helices are shown in detail at the bottom.

also found within the truncated version of the AQ(G)Ps, some remarkable exceptions exist as illustrated by *Kp*AQY1 (out of 15 H-bonds only 3 are located at the truncated scaffold) and for *So*PIP2.1 (only 4 out of 10 H-bonds are found at the truncated proteins). The number of salt bridges varies between 0 (for most of the mammalian AQ(G)Ps) and 5 (for *Ec*AQPZ) in the full structures and 0 (for 16 structures) and 3 (for *Af*AQPM) for the truncated structures. Given those numbers, *At*Tip2;1 exhibits the lowest density of H-bonds per 1000 Å$^2$ with only 2.3 and *Hs*AQP10 the highest with 9.5 for the full-length protein and 1.9 and 11 for the truncated protein respectively. The average AQ(G)P interface in our analysis has $5.2 \pm 1.6$ H-bonds per 1000 Å$^2$

and $0.4 \pm 0.7$ salt bridges per 1000 Å$^2$. The truncated AQ(G)Ps exhibit on average $5.0 \pm 2.2$ H-bonds per 1000 Å$^2$ and $0.2 \pm 0.5$ salt bridges per 1000 Å$^2$. Both values are at the lower end or even below published values for soluble protein interfaces of 5–10 hydrogen bonds per 1000 Å$^{2[160,161]}$ and around one salt bridge per 1000 Å$^2$.[161] A detailed analysis of amino acids involved in the interaction at the protomer-protomer interface (Figure S30, Supporting Information) revealed that H1, H2, the periplasmic end of H3, and the cytoplasmic end of H5 of one protomer interact with the periplasmic end of H2, H4, H5, and the cytoplasmic end of H6 of the neighboring protomer. In addition, the N- and C-terminus as well as L1, the N-terminal region of L4,

L5 and the C-terminal side of L4 are involved in protomer contacts. While the interaction pattern is very diverse in the loops, it is more defined in the transmembrane region, with largely similar patterns but varying individual amino acids. This leads to an amino acid specific interaction strength via the number of hydrogen bonds, salt bridges, and the surface area involved in hydrophobic interactions (Figure S30, Supporting Information).

Our analysis clearly shows that hydrogen bonds and salt bridges almost exclusively occur at the protein-aqueous phase interface (Figure 7A, Figure S28, Supporting Information). As there are hardly any methods available unveiling the relative contribution of each amino acid to the overall stability of the tetramer or the single protomers, we estimated buried surface areas[162] and interaction free energies ($\Delta G$) of the respective interfaces using PDBePISA.[163] Overall, $\Delta G$ is lower for mammalian AQPs as compared to archaea, bacteria, protozoa or yeast (Table S2, Supporting Information, Figure S31, Supporting Information). With an overall hydrophobic protein thickness of all AQ(Q)P structures of 30.2 ± 1.0 nm as extracted from the OPM database[164] we did not find any dependence of $\Delta G$ on the potential membrane thickness (data not shown). Compared to the linear relation of protomer-protomer $\Delta G$s on the buried surface area of the respective interface (Figure S31, Supporting Information), there is no clear trend between inter- and intra-protomer $\Delta G$s (Figure S32, Supporting Information). Concomitant is no obvious trend for intra-protomer $\Delta G$s on the AQ(G)P classes/subfamilies (Figure S32, Supporting Information).

Here, we would like to note that depending on the method used to calculate these $\Delta G$ values, absolute values may vary. Consequently, using PDBePISA only allows us to rank relative differences between AQ(G)P structures. PDBePISA comprises any kind of contact interaction, including hydrogen bonds, into the $\Delta G$ calculation. However, as it assumes that potential hydrogen bonding partners become satisfied by hydrogen bonds to water upon dissociation/unfolding, this may significantly lower the contribution of hydrogen bonds on the overall $\Delta G$s of membrane proteins. As PDBePISA is designed for soluble protein complexes, neglecting the contribution and presence of the lipid bilayer during, for example, oligomer disruption into single protomers, this may lead to an underestimation of the contribution of H-bonds and salt bridges to the overall $\Delta G$ of membrane proteins. Experimental and theoretical studies suggest $E_{hb} \approx 2$–10 kcal mol$^{-1}$,[165] whereas the remaining effect of decreased entropy of solvent due to the loss of mobility by bound molecules leads to an overall estimation of about $E_{hb} \approx 0.6$–1.5 kcal mol$^{-1}$ only.[166,167] Limited experimental data on the stabilization effect of salt bridges suggest that free energy contribution of a salt bridge is close to that of a hydrogen bond, amounting to $E_{sb} \approx 0.9$–1.25 kcal mol$^{-1}$.[168,169] A disulfide bond may contribute up to 2–8 kcal mol$^{-1}$,[170,171] yet, at lower occurrence. Hence, especially at the protomer–protomer interface of transmembrane proteins, hydrogen bonds might appear as major contributors into $\Delta G$. This is a very interesting finding given the fact that AQPs exhibit on average less H-bonds per 1000 Å$^2$ protomer–protomer interface than soluble proteins.

The residue specific analysis of intra-protomer helix–helix interactions reveals the key importance of a strong H-bond network anchoring L2 (connecting H2 and HH1), L6 (con-

necting H5 with HH2), and neighboring residues of the respective HH (Figure S33, Supporting Information). These stabilizing H-bond networks, discussed in detail in the next chapter, ensure the desired selectivity and proper functionality of the protein, keeping the ar/R filter and NPA region in place. Moreover, L2 and L6 are inevitable in forming H-bonds by their backbone oxygens with pore water molecules, something not possible with a helical structure, where all backbone oxygens and amides are required to stabilize the secondary structure. Hence, structurally stable selectivity filters as well as pore lining (loop) residues are inevitable for proper function of AQ(G)Ps. As can be seen in Figure 7B, the rest of the helices surrounding the selectivity filter region are stabilized mainly by hydrophobic interactions (blue) with single H-bonds stabilizing the ends of the $\alpha$-helices. The two intra–protomer helix–helix interfaces with the most noticeable differences between AQPs and AQGPs are H1–H3 and H2–H5. In detail, AQGPs and AQPMs exhibit on average 4.3 and 4.0 stabilizing H-bonds between H1 and H3 and 2.5 and 2.0 H-bonds between H2 and H5 compared to 0.5 and 0.6 H-bonds in AQPs, respectively.

Thermostability assays with reconstituted *Rn*AQP4, *Ec*GLPF, *Mm*AQPM, and *Ec*AQPZ in *E. coli* total lipid extract revealed that these proteins turned inactive at 70, 90, 90, and 100 °C, respectively.[119] While *Ec*AQPZ revealed the steepest temperature dependence, *Ec*GLPF exhibited the flattest, which is reflected by the estimated thermal denaturation values ($T_m$) in **Figure 8**. A linear fit to the data of $\Delta G$s (Table S2, Supporting Information) versus $T_m$ values for these four AQ(G)Ps shows a linear correlation of high credibility (Figure 8). This is surprising as PDBePISA $\Delta G$ values are thought to be a good relative approximation for hydrophobic interactions, underestimating the contribution of H-bonds or salt bridges in membrane proteins. Since neither the H-bonds nor salt bridges (8/0, 5/0, 8/1, and 6/5) nor the size of the protomer-protomer interaction interfaces (1488.6, 1545.3, 1763.4, and 1706.3 Å$^2$)
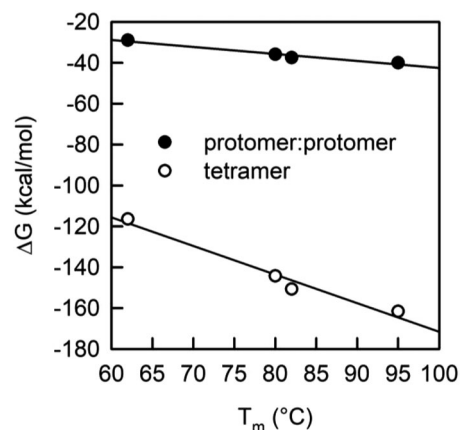


**Figure 8.** AQ(G)P thermostability. Thermal denaturation values ($T_m$) in the order of *Rn*AQP4 < *Ec*GLPF < *Mm*AQPM < *Ec*AQPZ[119] are compared to the $\Delta G$s calculated by PDBePISA. The data is fitted with a linear regression resulting in $k$-values of −0.34 and −1.4 for the protomer-protomer (full circles) and tetramer (empty circles) case. Protomer-protomer $\Delta G$s of −29, −35.9, −37.5, and −40.0 kcal mol$^{-1}$ and tetramer $\Delta G$s of −116.4, −144.3, −150.7, and −161.6 kcal mol$^{-1}$ for *Rn*AQP4, *Ec*GLPF, *Mm*AQPM, and *Ec*AQPZ, respectively, were used according to Table S2, Supporting Information.

yield a similar correlation, we are led to the conclusion that hydrophobic interactions are the main determinant of thermal denaturation of AQ(G)Ps, with only a marginal contribution of H-bonds and salt bridges. The latter two are, as mentioned above, likely responsible for the specifity of the protomer–protomer interactions instead.

### 2.9. Roles of Highly Conserved Residues

In addition to the conserved Arg[h2.2] in the ar/R selectivity filter including some of its interaction partners as well as the conserved His[h1.-3] which we discussed in detail in Sections 2.7.1. and 2.7.3. and the well-known NPA motifs, our sequence alignment (Figure S1, Supporting Information) reveals additional highly conserved amino acids. For illustration purposes, we color coded these positions in our AQ(G)P scaffold (**Figure 9**). A closer look, shown in detail in Figure S34, Supporting Information, discloses their diverse roles in stabilizing the AQ(G)Ps. The NPA motifs (orange, +) connect the half helices HH1 and HH2 with the inner pore loops L2 and L6, respectively. The extended H-bond network connects the NPA motifs to the transmembrane helical bundle of AQ(G)Ps. The inner pore loops L2 and L6 are stably anchored to the transmembrane helices by residues highlighted in magenta, in detail by Glu8[1.-6] (2), Thr67[h1.3] (5), Glu138[4.-4] (10), Thr181[5.12] (13), and Gln88[3.-3] (7). Moreover, Gly59[h1.-5] (4) shapes L2 via a loop internal interaction. Conserved Glys (labeled in dark blue) in the middle of helices ensure close H:H contact and include Gly91[3.0] (8), Gly167[5.-2] (12), and Gly215[6.3] (16). These Glys are parts of G(A/S)XX XG(A), P(G/S/A)XX XG and G(A/V)XX XGXX XG(A/F) motifs similar to the well-known glycine zippers.[172] Lipid interactions or larger spacer residues are labeled in cyan or grey (conserved in AQGPs, only) and include lipid-binding Arg3[1.-11] (11), spacers Tyr84[3.-7] (6), Phe/Tyr208[6.-4] (in case of AQGPs Pro[6.-4]) (20), and Tyr223[6.11] (17), as well as membrane anchors[173] Trp206[6.-6] (14), Phe98[3.2] (18), and Phe149[4.-7] (19). Moreover, Pro212[6.0] (15) and neighboring hydrophobic AAs are part of the double glycine zipper motif in H6 enabling close interaction with H4 and HH2. Other conserved positions are labeled in green: Gly21[1.7] (3) responsible for a kink in H1 which is, however, only obvious in seven structures (*Ec*GLPF, *Ec*AQPZ, *Af*AQPM, *Mm*AQPM, *At*AQPZ, *At*TIP2;1, *Bt*AQP1). Ser130[4.-12] (9) at the crossover between L4 and H4 stabilizes the oligomeric assembly via a backbone interaction to H3 of the neighboring protomer in *Ec*AQPZ, and Ser142[4.0] (11) (Thr[4.0] in all structures except of *Ec*AQPZ) in H4 stabilizing the connection with H6, keeping the cytoplasmic side of these helices in place.

Taken together, AQ(G)Ps contain multiple conserved residues whose art, localization, and H-bonding networks hint to very diverse roles, some of which are open for experimental validation.

### 2.10. Lipid Interaction Sites with Negatively Charged Lipids are Conserved

*Ec*AQPZ is significantly stabilized by cardiolipin (CL) as shown by mass spectrometry.[174] In addition, it was discovered that
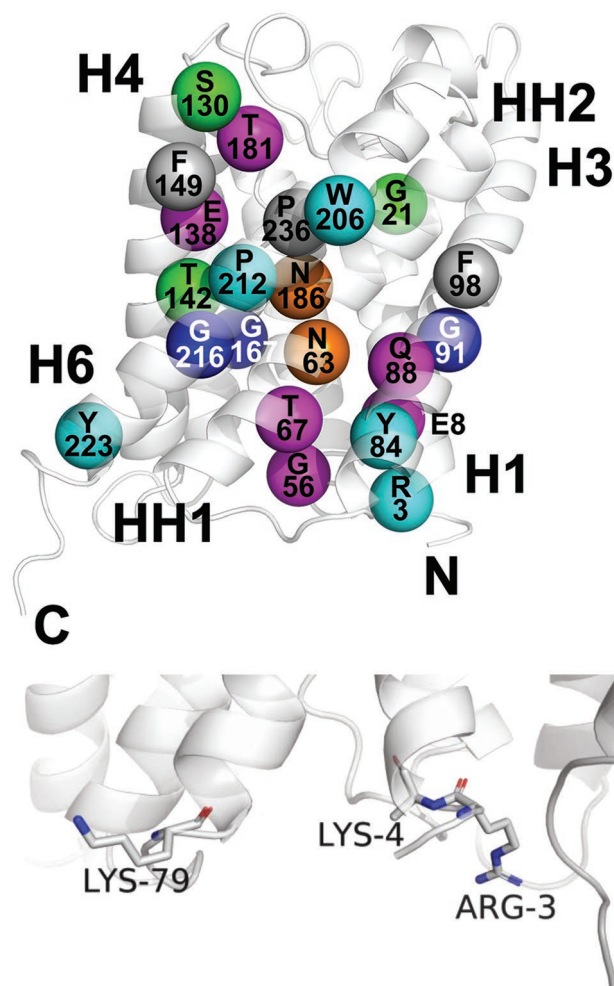


**Figure 9.** Location of conserved amino acid residues in the AQ(G)P scaffold. 21 locations are delineated in the sequence alignment in Figure S1, Supporting Information: Arg3[1.-11] (1, 65% conserved and 95% charge similarity), Glu8[1.-6] (2, 100% conserved), Gly21[1.7] (3, 85% conserved), Gly56[h1.-5] (4, 100% conserved), Thr67[h1.3] (5, 85% conserved), Tyr84[3.-7] (6, 90% conserved), Gln88[3.-3] (7, 100% conserved), Gly91[3.0] (8, 90% conserved), Ser130[4.-12] (9, 40% conserved and 90% type similarity), Glu138[4.-4] (10, 95% conserved), Thr142[4.0] (11, 95% conserved), Gly167[5.-2] (12, 100% conserved), Thr181[5.12] (13, 85% conserved), Trp206[6.-6] (14, 85% conserved), Pro212[6.0] (15, 100% conserved), Gly216[6.3] (16, 100% conserved), Tyr223[6.11] (17, 80% conserved and 100% steric similarity), Asn63[h1.-1] and Asn186[h2.-1] as part of the NPA motifs (+, 100% conserved), Phe98[3.2] (18, 100% conserved in AQGPs), and Phe149[4.-7] (19, 100% conserved in AQGPs), Pro236[6.-4] (20, 100% conserved in AQGPs). Snapshots of the H-bonding networks of these conserved residues are depicted in Figure S34, Supporting Information. Conservation is calculated as probability occurrence of the very same AA in all 20 high-resolution structures. Lipid interactions or larger spacer residues are labeled in cyan, the NPA motive in orange, residues fixing HHs and loops to HHs in pink, Glys and Alas for close H:H contact in dark blue, and other special positions in green. Bottom: Arg/Lys[1] at the groove between two protomers (light and dark grey) represent a conserved lipid interaction site. Shown are the Arg3[1.-11] and Lys4[1.-10] at the cytoplasmic end of H1 and Lys79[3.-12] at the cytoplasmic end of H3 of *Ec*AQPZ.

CL binds at the contact sites of protomers in the tetrameric assembly.[175] Investigating CL interactions utilizing a multitude of *E. coli* inner membrane proteins in silico identified a typical CL binding site to harbor two to three basic residues in close

proximity, at least one polar residue, and one or more aromatic residues slightly deeper within the membrane.[176] Similar to *Ec*AQPZ, CL was also found to bind preferentially to this cytoplasmic crevice at the contact of two neighboring *Bt*AQP1 protomers, as evident from MD simulations using a native-like model of *E. coli* polar lipid extract.[177] In the latter study, the rest of the cytoplasmic protein surface was covered with negatively charged phosphatidylglycerols, which were attracted by a number of positively charged residues (Lys7[1.-16], Lys8[1.-15], Arg12[1.-11], Arg95[3.-11], Arg243[6.25], and Lys245[6.27]), similarly to CL. Our analysis here reveals that the above-mentioned position of positively charged residues at the contact site between AQP protomers is conserved. Figure 9 illustrates the relative orientation of positive residues Arg3[1.-11] and Lys4[1.-10] at the cytoplasmic end of H1 and Lys79[3.-12] at the cytoplasmic end of H3 of *Ec*AQPZ. Both locations are equipped with positive charges in all AQPs under investigation (Figure S1, Supporting Information). Positively charged residues have been shown before to specifically target proteins to negatively charged surfaces of diverse membranes[178] and to control the orientation of the protein upon insertion into the lipid membrane.[179] The above mentioned conserved positively charged residues at the mouth of the crevice between two AQ(G)P protomers hint to the fact that AQ(G)Ps might additionally use the high abundance of negatively charged lipids at the cytoplasmic side of the membranes to stabilize their tetrameric structure. Yet, the absolute impact of specific lipid interactions of an AQ(G)P embedded in a native lipid bilayer on AQ(G)P stability has not yet been addressed.

## 2.11. The Central Pore

The biological significance of AQ(G)P tetramerization is still elusive. In general, oligomerization might lead to structural and/or proteolytic stability, functional diversity, regulatory mechanisms, and formation of binding cavities.[180] One side-effect of AQ(G)P tetramerization is the formation of a potential central pore. Several studies suggest ion[181–184] and gas[185] transport through the central, potential pore. As the topic is highly debated,[73,109,110] yet biological significance is still missing, we have conducted an analysis of the geometrical and chemical properties of these potentially channel forming central pores. In comparison to the protomeric pores, most of the profiles of the central channel reveal constrictions of ≤0.65 Å radius, which is geometrically too narrow for water molecules to pass (Figure S35, Supporting Information).[186] Notable exceptions are *Bt*AQP0, *Bt*AQP1, *Hs*AQP1, and *Hs*AQP2 with 0.7 Å, *Mm*AQPM and *At*Tip2;1 with 0.9 Å, *Hs*AQP4 and *Rn*AQP4 with 0.9-1 Å, and *Ec*GlpF with 1 Å. Figure S36, Supporting Information, illustrates the plethora of different pore geometries found for our structural dataset. Compared to the protomeric pore geometries, the shapes are much more diverse and less conserved. The considerable different pore lengths are partly caused by significant differences in the length of the cyto-/periplasmic loops connecting the transmembrane helices. *Pf*AQP is completely closed at the cytoplasmic side.

However, the hydrophobic characteristic of the inner surface of the central pores is conserved among all analyzed AQ(G)Ps:

Except for a ring of charged residues at the periplasmic side of the channel in case of *Ec*GLPF, *Pf*AQP, and mammalian AQP1s, the narrow pore is solely decorated by hydrophobic side chains (Figure S37, Supporting Information), which renders potential water passage through the central pore illusive. This presumption could be confirmed by analyzing water permeation events using MD simulations, where 0.99% (11 vs 1106), 0.96% (16 vs 1663), 2.49% (9 vs 361), and 0.43% (6 vs 1135) of the water molecules passed the central pore as compared to the 4 protomer pores within a simulation time of 500 ns for *Bt*AQP1, *Ec*AQPZ, *Hs*AQP4, and *Ec*GLPF, respectively. In contrast, narrow (2 Å diameter) carbon nanotubes (CNTs) are thought to be highly water permeable despite their hydrophobic nature.[187] This discrepancy may partly be due to different pore surfaces, which are homogeneous in CNTs and highly inhomogeneous in AQPs. In addition, CNTs are decorated with benzenes compared to methyl groups in AQPs. Moreover, a ring of negatively charged residues may serve as a putative divalent cation binding spot as found for Glu43[2.-8] in *Ec*GLPF,[188] further obstructing the pore for other solutes (Figure S37). An alanine mutant (Glu43Ala[2.-8]) suggested the involvement of the respective $Mg^{2+}$ binding site in the overall stability of *Ec*GLPF, shifting the oligomeric state dramatically towards the monomer.[189] Stabilization of the tetramer in the presence of $Mg^{2+}$ ions was also reported by Borgnia et al.[190] However, the effect could not be confirmed in experiments with EDTA as a chelator.[191] In addition to Glu43[2.-8], the crystal structure of *Ec*GLPF localized a second $Mg^{2+}$ next to Trp42[2.-9].[4] Glu43[2.-8] might also be involved in the binding of specific lipids in vivo.[4,51] In wider central channels, that is, the hexameric *Hp*UreI, pore obstruction by lipids is a common phenomenon.[95] In the yeast AQP, *Kp*AQY1, another finding suggests chloride ions to be bound in the central channel next to Trp5.-9.[118] Generally, structural comparison with narrow ion channels reveals that their ion selective pores are not hydrophobic but decorated with coun-tercharges in respect to the transported ion permeating the channel.[192,193] Thus, the hydrophobic central pore of AQ(G)Ps is unlikely to efficiently transport water or small ions even if the narrow constriction would theoretically provide enough space.

## 3. Conclusions and Outlook

With the overall structure of AQ(G)Ps being well established, including a sub-Ångström structure of *Kp*AQY1,[42] recent structures combined with experiments and simulations[52,53,55] have provided great insight into the transport, selectivity, and regulation mechanisms of individual AQPs. Here, we took a new approach, analyzing all 20 non-redundant high-resolution structures deposited in the protein database to unravel structural peculiarities of AQ(G)Ps hidden by the analysis of single AQP structures or limited subsets.

In principle, AQ(G)Ps constitute simple proteins, with an equal fold of six transmembrane helices and two half-helices, nevertheless being regulated by multiple molecular mechanisms, bestowing AQ(G)Ps diverse selectivities, and hence making them capable of fulfilling manifold functions. First, we want to elaborate on the question how it is possible that seemingly highly similar proteins exhibit such a wide variance in $p_f$

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

values, as shown in Figure 1. The major determinants of single-file transport of water through narrow membrane channels were suggested to be the number of H-bonds water molecules may form with pore lining residues,[3] channel gating by pore lining residues,[75] positive charges at the pore mouth potentially reducing the dehydration penalty,[79] and possibly the shape of the entry/exit vestibules.[80] Moreover, the geometry of the pore as well as structural changes due to lipid interactions may depict additional determinants. The characteristic hourglass shape of AQPs was found to be the optimum for a hydrodynamic dissipation process, maximizing channel permeability.[80] However, as it can be seen from Figure 3, Figures S8 and S9, Supporting Information, the shape and angle of the hourglass shaped pore entrances are very conserved among AQ(G)Ps, thus leaving hardly any room to explain permeability differences among AQ(G)Ps. Similarly, the overall pore geometry (width) could potentially explain $p_f$ differences between AQPs and AQGPs but hardly among AQPs, as crucial differences in pore geometry within this group are missing (Figure S10, Supporting Information). The effect of positive charges at the pore entrance potentially reducing the dehydration penalty is thought to be minor.[79] The latter estimate expanded to the current set of 20 AQ(G)P structures (Figure S8, Supporting Information) reveals comparable differences in the number and distribution of positive AAs next to the pore entrances. Similarly, our structural analysis suggests that the potential number of H-bonds water molecules may form with pore lining residues is hardly variable. H-bonds formed to single-file water molecules are either formed via side chains of conserved residues in the NPA motives or in the ar/R filter or via backbone oxygens of L2 and L6. Interestingly, our MD simulations depict significant variations in the overall number of H-bonds single-file water molecules form with these conserved pore lining residues among different AQ(G)Ps, which would principally enable a certain variability in $p_f$. What is left are gating effects by pore lining residues (e.g., the phenolic barriers drastically reducing the permeability of AQP0[45]) or lipid specific effects.

In line with these insights, our structural analysis and MD simulations clearly show that AQ(G)Ps do not exhibit a universally open pore. In contrast, water flow through the single-file pores is modulated by pore lining residues as sporadically already mentioned in literature. Even though the physiological significance of water flux modulation via the conserved His at the cytoplasmic side of the single-file region can be envisioned similar to pH regulated human adipose glycerol flux through HsAQP10,[53] the picture is less obvious regarding gating of the conserved Arg^h2.2 of the ar/R selectivity filter. Theoretically the position of the Arg^h2.2 sidechain in the pore could be regulated via the transmembrane potential, lipid asymmetry, or binding of highly charged soluble proteins to the pore entrance, the latter two impinging on the membrane potential. So far, such an effect on water permeability through HsAQP1 and HsAQP4 was only seen in silico at unphysiologically high membrane potentials of >±0.5 V.[137] However, high transmembrane potential can lead to electroporation of the membrane[194,195] or to denaturation of transmembrane proteins.[196] The importance of side chain fluctuations of Met^h2.-2 positioned in the direct vicinity of the NPA and ar/R motives awaits clarification. Are they a result of the imperfection of the respective AQP or part

of an elaborate regulation mechanism? This is a question which is not easy to answer. However, what we could show for the dataset analysed herein by comparing the H-bond network stabilizing the Arg^h2.2, it is possible to predict its mobility in silico with a clear impact on water passage through the pores (Figures S21–S24, Supporting Information).

Which structural effects specific lipid interactions have on AQ(G)P functionality and how such direct structural regulatory mechanisms would look like, remains an open challenge. Nonetheless, as it is known that specific binding of lipids stabilizes the oligomeric assembly,[162,197] we speculate that stabilization of the oligomeric assembly has an impact on the flexibility of pore lining residues. This could change the probability of pore lining residues to reside within the channel pore obstructing it for a certain percentage of time as well as potentially influence its selectivity. Hence, whereas the maximum water permeability of AQ(G)Ps is defined by the pore characteristics discussed above, we speculate that the net flux through an AQ(G)P is defined by channel gating via flexible pore lining residues. Examples could be a strongly modulated ribitol transport capability of EcGLPF by negatively charged lipids[198] and an increased EcAQPZ water permeability after CL binding.[174]

Substrate discrimination in water channels is thought to depend on a complex interplay between the solute, pore size, and polarity, with the pore size determining the exclusion properties but not solute selectivity.[62] In accordance, the pore size of AQGPs is larger throughout the pore compared to AQPs and AQPMs (Figure 3, Figure S10, Supporting Information). However, so far, it was not clear if this larger pore size is solely constituted by pore lining residues or if the AQGP scaffold is distinctly different from that of AQPs. Figures S15 and S17, Supporting Information, visualize that this is indeed the case. AQGPs do not only exhibit a larger protomer-protomer distance by 1.1 Å and an overall increased dimension of the tetrameric assembly by 2.6 Å along the protomer-lipid interface, but also the relative distances of helices in the protomer are increased. Thereby, H5 localization further away from the protomer pore goes hand in hand with an increased H4:H5 distance. The latter is possible due to an evolutionary groove at the H4:H5 interface, with perfectly parallel oriented helices within the protomeric arrangement (Figure S38, Supporting Information). In contrast, all other helix-helix interactions are much more refined, with the helices tilted to each other forming helix-helix cross overs with highly conserved Gly and Ala residues enabling close helix-helix contacts. Concerning AQPMs, we showed that despite having a broader pore than AQPs, yet a rather similar scaffold size, it is possible to reach glycerol selectivity with an AQP scaffold but at the expense of efficient permeability. To achieve both, a special set of pore lining residues and an adequately expanded scaffold as in AQGPs seem mandatory.

Our analysis reveals internal H-bond networks stabilizing both HH and adjacent loops L2 and L6, which harbor most of the H-bonds donating and accepting residues in the single-file region of AQ(G)Ps. Also, the H-bond network shapes the structure and the dynamics of the ar/R filter and thus its selectivity (Figure 7B, Figure S33, Supporting Information). Together with the strong conservation of the involved residues (Figure 1, Figure S34, Supporting Information) this highlights the importance of a stable single-file region for AQ(G)P selectivity and

**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

**NANO · MICRO
small**
www.small-journal.com

permeability. We speculate that the tetrameric arrangement is inevitable to ensure this structural necessity as monomeric *Ec*AQPZ[199] and *Ec*GLPF[200–202] exhibited reduced activities in terms of water and ribitol flux, respectively. Monomeric *Ec*GLPF was also less resistant to proteolysis in *E. coli*,[189] with a significantly stabilized tetrameric assembly in vivo.[191] Additional stabilization via specific lipid interactions within the complex natural lipid composition[177] may have evolved in parallel to inter- and intra-protomer evolution. In any case, optimization of tetramer stability is a tradeoff with protein aggregation and protein folding, as protomer stability in the lipid bilayer is a prerequisite for AQ(G)P folding into the membrane and consequent tetramerization. This protomer pre-insertion into the membrane may explain why AQ(G)P protomers exhibit hardly any polar residues at the protomer-protomer interface within the membrane (Figure S28, Supporting Information). Instead, the oligomeric assembly is mainly stabilized via hydrophobic interactions within this region, supported by lipid interactions, H-bonds, and salt bridges. The latter located mostly at the lipid bilayer to aqueous phase interfacial regions. Even though the individual contributions to tetramer stability still stay elusive, the here discovered correlation of the thermal denaturation temperatures for four AQ(G)Ps with their ΔGs calculated by PDBePISA imply the crucial role of hydrophobic interactions for the stability of the AQ(G)P fold.

AQ(G)Ps are seen as potential building blocks of next generation filter membranes.[16–27] Furthermore, they are treated as the "holy grail" of native water conducting pores, due to their superior performance in respect to their high permeability and great selectivity. Hence AQ(G)Ps serve as templates for a whole community designing artificial water channels.[29–36] Desalination and distillation are among the most resource intensive industrial processes.[203] Therefore, an enormous effort is spent to develop filter membranes with increased efficiency. As a first step, functional *Ec*AQPZ was already successfully incorporated into polymer vesicles and membranes,[204] with already some successful commercialization of Aquaporin Inside desalination membranes from Aquaporin A/S.[205] Still, this idea is afflicted with the prejudice that AQ(G)Ps are not stable enough[16] and might degrade in the presence of harsh conditions used in the process of industrial membrane formation. This may be even more relevant when using other AQ(G)Ps, which are inherently less stable than *Ec*AQPZ, to expand the range of applications from pure desalination or water filtration to biotechnological applications like small molecule recovery. Our rigorous structural analysis is a first step towards the goal to create bioinspired AQ(G)P variants with optimized stability and tuned selectivity for next generation biomimetic separation membranes. An ideal water channel in this respect shall combine high permeability, perfect selectivity, and exceptional structural stability, as to withstand adverse physical and chemical conditions. Altering the amino acid sequence of AQ(G)Ps in order to optimize the key aspects mentioned above, requires comprehensive understanding of the underlying structural and functional features. Consequently, this work can be used to choose the most appropriate scaffold in terms of AQ(G)P function and stability and combine interaction hotspots, purely hydrophobic in nature or specific via H-bonds or salt-bridges, found in other AQ(G)Ps into the chosen scaffold.

## 4. Experimental Section

*Analysis and Preparations of the AQ(G)P Structures*: The PDB database was surveyed for all currently available AQP and AQGP structures. The resulting comprehensive list, stating the respective resolution, method of structure elucidation as well as year of submission and reference to the original publication, if available, can be found in Table S1, Supporting Information.

For the analysis presented in this paper, a non-redundant selection of AQ(G)P structures was generated, by exclusion of identical, marginally mutated, or substrate containing structures of the same organism. The targets were chosen based on their apparent resolution (the higher the better) and the absence of mutations (wild type structures were preferred). The resulting list contained 20 structures: two AQPM, a subclass of archaeal AQPs first discovered in *Methanothermobacter marburgis*, three bacterial AQ(G)Ps, two structures originating from protozoa or yeast, three plant AQPs, and ten mammalian AQ(G)P structures. The corresponding amino acid sequences were aligned in Jalview (version 2.10.5)[206] utilizing the ClustalOmega algorithm with default settings and subsequently used for the visualization of shared AQ(G)P features.

To obtain root mean square deviation (RMSD) values for structure alignments of wildtype AQ(G)Ps, the standard alignment function of PyMOL (Schrödinger, version 2.3.2) was used. *Ec*AQPZ (PDB: 1RC2) served as a reference for most of the following analysis due to its high resolution and most compact structure, with minimal loops, and N- and C-termini. Furthermore, it showed an astonishing stability[119] rendering *Ec*AQPZ a promising candidate for biotechnological applications.[204,207] N- and C-termini of all AQ(G)Ps were removed according to the remaining quality of the alignment. The resulting truncated models were named the AQ(G)P "core". Helix and loop lengths, as well as the respective amino acid composition of the chains were analyzed in PyMOL. In the core models loops 1, 4, 5, and 7 were omitted due to their high flexibility/variability. The $C_\alpha$ RMSD values of the core structures relative to *Ec*AQPZ (PDB: 1RC2) were calculated using the PyMOL align command, with outlier rejection turned off. Next to RMSD estimation for each $C_\alpha$ also RMSDs of $C_\alpha$ in five distinct horizontal plains along the z-axis of the AQ(G)Ps were estimated.

The PDBePISA server[163,208] was used to calculate interaction areas and the gain in solvation energy upon interface formation, $\Delta Gs$, between AQ(G)P protomers as well as internally, between individual helices and loops. Furthermore, detailed analysis of individual interactions derived from PDBePISA provided information about amino acids involved in hydrophobic interactions (buried area percentage of interfacing residues), hydrogen bonds, and salt bridges. The detailed interaction profiles were registered to a template sequence alignment of the 20 AQ(G)P structures and thus gave rise to detailed depictions of the inter- and intra-molecular interactions within these structures.

The shape of AQ(G)P protomer channels as well as the shape of the central pore in AQ(G)P tetramers were analyzed with the program HOLE[126] utilizing the xplor algorithm and providing coordinates of a single water molecule in the center of the aligned AQ(G)P structures to serve as a start point for the channel traces.

The pKa values of histidine residues were evaluated using the program propka (version 3.4.0).[156,157]

All structure figures were prepared in PyMOL.[209] Graphs were plotted using the programs VEUSZ (Jeremy Sanders, version 2.1.1) and R.[210]

Evolutionary analysis was conducted in MEGA11.[211] The phylogenetic tree with the highest log likelihood (−4884.76) is shown in Figure S3, Supporting Information. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. This analysis involved 20 amino acid sequences. There was a total of 168 positions in the final dataset.

*MD Simulations of AQ(G)Ps*: MD simulations were performed using the GROMACS package version 2016.3 and 2018.6.[212] In each simulation system a tetrameric AQ(G)P was embedded in an *E. coli* polar lipid

extract membrane model "Avanti"[213] containing 14 different lipid types by our sequential multiscaling methodology[214] as described in ref. [213]. The membrane consisting of 384 lipids, was solvated by ≈29 000 water molecules and counter Na$^+$ ions. The tetrameric $Bt$AQP1, $Hs$AQP4, $Ec$GLPF, and $Ec$AQPZ were prepared based on the crystal structures 1J4N,[47] 3GD8,[49] 1FX8,[4] and 2ABM,[139] respectively. Each chain of $Bt$AQP1 thus contained the resolved residues M1-S249, the shortened C-terminus was thereby capped by an amine group. Each $Hs$AQP4 chain consisted of residues Q32-P254 and each $Ec$GLPF chain was built of T6-E267, thereby the loop residues P260-E267 were added to the crystal structure by Modeller 9.[215] In both, $Hs$AQP4 and $Ec$GLPF the shortened N-termini carried an NH$_2$ group and the shortened C-termini were capped by an amine group. Each $Ec$AQPZ chain contained all $Ec$AQPZ residues, that is, M1-D231 and thus both termini were charged. The histidine in the ar/R filter of all three AQPs was protonated on N$^\delta$,[216] due to its higher water permeability. All other histidines were protonated on N$^\epsilon$. All titratable amino acids were protonated according to their preferred protonation state at pH 7. The 500 ns long production MD simulations were performed at 296 K, using the CHARMM36m force field[217–219] and TIP4p water model.[220] The trajectories were written out every 2 ps. For more details see ref. [213].

The estimation of the number of water molecules permeating the AQ(G)P pores was performed by g_flux[221] through a cylinder with a radius of 0.75 nm and a length of 2 nm. Analysis of hydrogen bonds, side chain angles, and distances between residues succeeded by standard GROMACS analysis tools.

*Statistical Analysis*: Pre-processing of the data is explained in the "Analysis and Preparations of the AQ(G)P Structures" and "MD Simulations of AQ(G)Ps" sections. No outliers were rejected during our analysis. $At$Tip2;1 was not grouped with AQPs, AQGPs or AQPMs due to its NH$_3$ permeability.

If not stated otherwise, the analysis of the properties from MD simulations was done separately for each chain, that is, $n = 4$, and the standard errors of the means are given. Sample sizes for our structural comparisons were $n = 11$ for AQP structures ($Kp$AQY1, $Oa$AQP0, $Bt$AQP0, $Hs$AQP1, $Bt$AQP1, $Hs$AQP2, $Hs$AQP4, $Rn$AQP4, $Hs$AQP5, $So$PIP2;1, and $At$PIP2;1), $n = 4$ for AQGP structures ($Ec$GLPF, $Pf$AQP, $Hs$AQPZ, and $Hs$AQP10), and $n = 2$ for AQPM structures ($Mm$AQPM and $Af$AQPM). The errors denoted standard deviations of the mean.

The significance of the statistical difference was estimated for properties extracted from MD simulations due to the underlaying normal distributions of the compared properties. In detail, Figure S20, Supporting Information, shows an average number of H-bonds formed between pore-residues and pore water molecules. The significance of the differences between different AQ(G)Ps was estimated using an unpaired one-sided $t$-test in Microsoft Excel. The obtained $p$-values were characterized as strongly significant ($p < 0.01$) and weakly significant ($0.01 \leq p \leq 0.05$).

Calculations were performed in Microsoft Excel.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

N.G.-M. and C.S. contributed equally to this work. N.G-M., C.S., L.U., S.K.N.T., and A.H. analyzed AQ(G)P structures. K.P. performed and analyzed the molecular dynamics simulations. N.G-M., C.S., K.P. and A.H. conceptualized the study. N.G-M., C.S., K.P. and A.H. interpreted the results. N.G-M., C.S., K.P. and A.H. wrote the manuscript. All authors approved the manuscript.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Keywords

[1] G. M. Preston, T. P. Carroll, W. B. Guggino, P. Agre, *Science* **1992**, *256*, 385.

[2] M. L. Zeidel, S. V. Ambudkar, B. L. Smith, P. Agre, *Biochemistry* **1992**, *31*, 7436.

[3] A. Horner, F. Zocher, J. Preiner, N. Ollinger, C. Siligan, S. A. Akimov, P. Pohl, *Sci. Adv.* **2015**, *1*, e1400083.

[4] D. Fu, A. Libson, L. J. Miercke, C. Weitzman, P. Nollert, J. Krucinski, R. M. Stroud, *Science* **2000**, *290*, 481.

[5] L. S. King, D. Kozono, P. Agre, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 687.

[6] F. Abascal, I. Irisarri, R. Zardoya, *Biochim. Biophys. Acta Gen. Sub.* **2014**, *1840*, 1468.

[7] R. K. Deshmukh, J. Vivancos, G. Ramakrishnan, V. Guérin, G. Carpentier, H. Sonah, C. Labbé, P. Isenring, F. J. Belzile, R. R. Bélanger, *Plant J.* **2015**, *83*, 489.

[8] A. S. Verkman, *Annu. Rev. Med.* **2012**, *63*, 303.

[9] M. C. Papadopoulos, S. Saadoun, *Biochim. Biophys. Acta* **2015**, *1848*, 2576.

[10] A. K. Azad, T. Raihan, J. Ahmed, A. Hakim, T. H. Emon, P. A. Chowdhury, *Front. Genet.* **2021**, *12*, 654865.

[11] A. S. Verkman, M. O. Anderson, M. C. Papadopoulos, *Nat. Rev. Drug Discovery* **2014**, *13*, 259.

[12] M. C. Papadopoulos, A. S. Verkman, *Nat. Rev. Neurosci.* **2013**, *14*, 265.

[13] C. Maurel, L. Verdoucq, D. T. Luu, V. Santoni, *Annu. Rev. Plant Biol.* **2008**, *59*, 595.

[14] A. Shekoofa, T. Sinclair, *Cells* **2018**, *7*, 123.

[15] S. M. Zargar, P. Nagar, R. Deshmukh, M. Nazir, A. A. Wani, K. Z. Masoodi, G. K. Agrawal, R. Rakwale, *J. Proteomics* **2017**, *169*, 233.

ADVANCED
SCIENCE NEWS

www.advancedsciencenews.com

NANO · MICRO
small

www.small-journal.com

[16] J. To, J. Torres, *Membranes* **2015**, *5*, 352.

[17] W. Song, C. Lang, Y.-x Shen, M. Kumar, *Annu. Rev. Mater. Res.* **2018**, *48*, 57.

[18] T. Ren, M. Erbakan, Y. Shen, E. Barbieri, P. Saboe, H. Feroz, H. Yan, S. McCuskey, J. F. Hall, A. B. Schantz, G. C. Bazan, P. J. Butler, M. Grzelakowski, M. Kuma, *Adv. Biosyst.* **2017**, *1*, 1700053.

[19] A. Abdelrasoul, H. Doan, A. Lohi, C.-H. Cheng, *Polym. Sci., Ser. A* **2018**, *60*, 429.

[20] A. Fuwad, H. Ryu, N. Malmstadt, S. M. Kim, T.-J. Jeon, *Desalination* **2019**, *458*, 97.

[21] X. S. Li, R. Wang, C. Tang, A. Vararattanavech, Y. Zhao, J. Torres, T. Fane, *Colloids Surf., B* **2012**, *94*, 333.

[22] C. Tang, Z. Wang, I. Petrinić, A. G. Fane, C. Hélix-Nielsen, *Desalination* **2015**, *368*, 89.

[23] S. Qi, R. Wang, G. K. M. Chaitra, J. Torres, X. Hu, A. G. Fane, *J. Membr. Sci.* **2016**, *508*, 94.

[24] Y. Li, S. Qi, M. Tian, W. Widjajanti, R. Wang, *Desalination* **2019**, *467*, 103.

[25] P. Wagh, I. C. Escobar, *Environ. Prog. Sustainable Energy* **2019**, *38*, e13215.

[26] Z. Liang, Y. Yun, M. Wang, G. Liu, P. Lu, W. Yang, C. Li, *RSC Adv.* **2019**, *9*, 10715.

[27] M. C. Martinez-Ballesta, P. Garcia-Ibanez, L. Yepes-Molina, J. J. Rios, M. Carvajal, *Cells* **2018**, *7*, 179.

[28] W. Song, M. Kumar, *Curr. Opin. Chem. Eng.* **2019**, *25*, 9.

[29] Y. X. Shen, W. Song, D. R. Barden, T. Ren, C. Lang, H. Feroz, C. B. Henderson, P. O. Saboe, D. Tsai, H. Yan, P. J. Butler, G. C. Bazan, W. A. Phillip, R. J. Hickey, P. S. Cremer, H. Vashisth, M. Kumar, *Nat. Commun.* **2018**, *9*, 2294.

[30] W. Song, Y. M. Tu, H. Oh, L. Samineni, M. Kumar, *Langmuir* **2019**, *35*, 589.

[31] Y. X. Shen, W. Si, M. Erbakan, M. Kumar, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 9810.

[32] I. Kocsis, Z. Sun, Y. M. Legrand, M. Barboiu, *npj Clean Water* **2018**, *1*, 13.

[33] S. Murail, T. Vasiliu, A. Neamtu, M. Barboiu, F. Sterpone, M. Baaden, *Faraday Discuss.* **2018**, *209*, 125.

[34] M. Barboiu, A. Gilles, *Acc. Chem. Res.* **2013**, *46*, 2814.

[35] M. Barboiu, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 11674.

[36] Y. Le Duc, M. Michau, A. Gilles, V. Gence, Y.-M. Legrand, A. van der Lee, S. Tingry, M. Barboiu, *Angew. Chem., Int. Ed.* **2011**, *50*, 11366.

[37] E. Beitz, B. Wu, L. M. Holm, J. E. Schultz, T. Zeuthen, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 269.

[38] B. L. de Groot, H. Grubmuller, *Curr. Opin. Struct. Biol.* **2005**, *15*, 176.

[39] H. Li, H. Chen, C. Steinbronn, B. Wu, E. Beitz, T. Zeuthen, G. A. Voth, *J. Mol. Biol.* **2011**, *407*, 607.

[40] J. K. Lee, D. Kozono, J. Remis, Y. Kitagawa, P. Agre, R. M. Stroud, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18932.

[41] D. F. Savage, P. F. Egea, Y. Robles-Colmenares, J. D. O'Connell, R. M. Stroud, *PLoS Biol.* **2003**, *1*, 334.

[42] U. K. Eriksson, G. Fischer, R. Friemann, G. Enkavi, E. Tajkhorshid, R. Neutze, *Science* **2013**, *340*, 1346.

[43] Z. E. Newby, J. O'connell Iii, Y. Robles-Colmenares, S. Khademi, L. J. Miercke, R. M. Stroud, *Nat. Struct. Mol. Biol.* **2008**, *15*, 619.

[44] T. Gonen, Y. Cheng, P. Sliz, Y. Hiroaki, Y. Fujiyoshi, S. C. Harrison, T. Walz, *Nature* **2005**, *438*, 633.

[45] W. E. Harries, D. Akhavan, L. J. Miercke, S. Khademi, R. M. Stroud, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14045.

[46] D. R. Carrillo, J. T. Y. Ying, D. Darwis, C. H. Soon, T. Cornvik, J. Torres, J. Lescar, *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2014**, *70*, 1657.

[47] H. X. Sui, B. G. Han, J. K. Lee, P. Walian, B. K. Jap, *Nature* **2001**, *414*, 872.

[48] A. Frick, U. K. Eriksson, F. de Mattia, S. Törnroth-Horsefield, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6305.

[49] J. D. Ho, R. Yeh, A. Sandstrom, I. Chorny, W. E. C. Harries, R. A. Robbins, L. J. W. Miercke, R. M. Stroud, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7437.

[50] K. Tani, T. Mitsuma, Y. Hiroaki, A. Kamegawa, K. Nishikawa, Y. Tanimura, Y. Fujiyoshi, *J. Mol. Biol.* **2009**, *389*, 694.

[51] R. Horsefield, K. Nordén, M. Fellert, A. Backmark, S. Törnroth-Horsefield, A. C. Terwisscha Van Scheltinga, J. Kvassman, P. Kjellbom, U. Johanson, R. Neutze, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13327.

[52] S. W. de Mare, R. Venskutonyte, S. Eltschkner, B. L. de Groot, K. Lindkvist-Petersson, *Structure* **2020**, *28*, 215.

[53] K. Gotfryd, A. Filipa Mósca, J. W. Missel, S. Friis Truelsen, K. Wang, M. Spulber, S. Krabbe, C. Hélix-Nielsen, U. Laforenza, G. Soveral, P. A. Pedersen, P. Gourdon, *Nat. Commun.* **2018**, *9*, 4749.

[54] S. Törnroth-Horsefield, Yi Wang, K. Hedfalk, U. Johanson, M. Karlsson, E. Tajkhorshid, R. Neutze, P. Kjellbom, *Nature* **2006**, *439*, 688.

[55] H. Wang, S. Schoebel, F. Schmitz, H. Dong, K. Hedfalk, *Biochim. Biophys. Acta Biomembr.* **2020**, *1862*, 183065.

[56] A. Kirscht, S. Shreyas Kaptan, G. P. Bienert, F. Chaumont, P. Nissen, B. L. de Groot, P. Kjellbom, P. Gourdon, U. Johanson, *PLoS Biol.* **2016**, *14*, e1002411.

[57] G. P. Bienert, A. L. B. Møller, K. A. Kristiansen, A. Schulz, I. M. Møller, J. K. Schjoerring, T. P. Jahn, *J. Biol. Chem.* **2007**, *282*, 1183.

[58] J. S. Hub, B. L. de Groot, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1198.

[59] Y. Wang, K. Schulten, E. Tajkhorshid, *Structure* **2005**, *13*, 1107.

[60] D. F. Savage, J. D. O'Connell, 3rd, L. J. Miercke, J. Finer-Moore, R. M. Stroud, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 17164.

[61] R. Oliva, G. Calamita, J. M. Thornton, M. Pellegrini-Calace, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4135.

[62] P. Kitchen, M. M. Salman, S. U. Pickel, J. Jennings, S. Törnroth-Horsefield, M. T. Conner, R. M. Bill, A. C. Conner, *Sci. Rep.* **2019**, *9*, 20369.

[63] B. Wu, E. Beitz, *Cell. Mol. Life Sci.* **2007**, *64*, 2413.

[64] G. P. Bienert, F. Chaumont, *Biochim. Biophys. Acta* **2014**, *1840*, 1596.

[65] R. Mukhopadhyay, E. Beitz, *Adv. Exp. Med. Biol.* **2010**, *679*, 57.

[66] G. P. Bienert, M. D. Schussler, T. P. Jahn, *Trends Biochem. Sci.* **2008**, *33*, 20.

[67] M. Rothert, D. Ronfeldt, E. Beitz, *J. Biol. Chem.* **2017**, *292*, 9358.

[68] J. D. R. Schmidt, P. Walloch, B. Hoger, E. Beitz, *Biochimie* **2021**, *188*, 7.

[69] G. P. Bienert, B. Desguin, F. Chaumont, P. Hols, *Biochem. J.* **2013**, *454*, 559.

[70] I. C. Mori, J. Rhee, M. Shibasaka, S. Sasano, T. Kaneko, T. Horie, M. Katsuhara, *Plant Cell Physiol.* **2014**, *55*, 251.

[71] V. Montiel, R. Bella, L. Y. M. Michel, H. Esfahani, D. De Mulder, E. L. Robinson, J.-P. Deglasse, M. Tiburcy, P. H. Chow, J.-C. Jonas, P. Gilon, B. Steinhorn, T. Michel, C. Beauloye, L. Bertrand, C. Farah, F. Dei Zotti, H. Debaix, C. Bouzin, D. Brusa, S. Horman, J.-L. Vanoverschelde, O. Bergmann, D. Gilis, M. Rooman, A. Ghigo, S. Geninatti-Crich, A. Yool, W. H. Zimmermann, H. L. Roderick, et al., *Sci. Transl. Med.* **2020**, *12*, eaay2176.

[72] D. Boytsov, C. Hannesschlaeger, A. Horner, C. Siligan, P. Pohl, *Biotechnol. J.* **2020**, *15*, 1900450.

[73] C. Hannesschlaeger, A. Horner, P. Pohl, *Chem. Rev.* **2019**, *119*, 5922.

[74] C. Hannesschlager, T. Barta, C. Siligan, A. Horner, *Sci. Rep.* **2018**, *8*, 8516.

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

[75] A. Horner, P. Pohl, *Faraday Discuss.* **2018**, *209*, 9.

[76] J. Wachlmayr, C. Hannesschlaeger, A. Speletz, T. Barta, A. Eckerstorfer, C. Siligan, A. Horner, *Nanoscale Adv.* **2021**, *4*, 58.

[77] T. Hoomann, N. Jahnke, A. Horner, S. Keller, P. Pohl, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10842.

[78] D. G. Knyazev, A. Lents, E. Krause, N. Ollinger, C. Siligan, D. Papinski, L. Winter, A. Horner, P. Pohl, *J. Biol. Chem.* **2013**, *288*, 17941.

[79] A. Horner, C. Siligan, A. Cornean, P. Pohl, *Faraday Discuss.* **2018**, *209*, 55.

[80] S. Gravelle, L. Joly, F. Detcheverry, L. Bocquet, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16367.

[81] A. Horner, P. Pohl, *Science* **2018**, *359*.

[82] L. Erokhova, A. Horner, P. Kugler, P. Pohl, *J. Biol. Chem.* **2011**, *286*, 39926.

[83] R. Zardoya, *Biol. Cell* **2005**, *97*, 397.

[84] R. N. Finn, J. Cerda, *Biol. Bull.* **2015**, *229*, 6.

[85] S. Kreida, S. Toernroth-Horsefield, *Curr. Opin. Struct. Biol.* **2015**, *33*, 126.

[86] K. Hedfalk, S. Törnroth-Horsefield, M. Nyblom, U. Johanson, P. Kjellbom, R. Neutze, *Curr. Opin. Struct. Biol.* **2006**, *16*, 447.

[87] S. Tornroth-Horsefield, K. Hedfalk, G. Fischer, K. Lindkvist-Petersson, R. Neutze, *FEBS Lett.* **2010**, *584*, 2580.

[88] A. Frick, M. Jarva, S. Tornroth-Horsefield, *FEBS Lett.* **2013**, *587*, 989.

[89] R. Sachdeva, B. Singh, *Prog. Biophys. Mol. Biol.* **2014**, *114*, 69.

[90] J. P. Bezerra-Neto, F. C. de Araújo, J. R. C. Ferreira-Neto, M. D. da Silva, V. Pandolfi, F. F. Aburjaile, T. Sakamoto, R. L. de Oliveira Silva, E. A. Kido, L. L. B. Amorim, J. M. Ortega, A. M. Benko-Iseppon, *Curr. Protein Pept. Sci.* **2019**, *20*, 368.

[91] M. del Carmen Martinez-Ballesta, M. Carvajal, *Plant Sci.* **2014**, *217*, 71.

[92] P. Gena, M. Pellegrini-Calace, A. Biasco, M. Svelto, G. Calamita, *Food Biophys.* **2011**, *6*, 241.

[93] J. V. Roche, S. Tornroth-Horsefield, *Int. J. Mol. Sci.* **2017**, *18*, 2255.

[94] J. Sjöhamn, K. Hedfalk, *Biochim. Biophys. Acta-Gen. Sub.* **2014**, *1840*, 1614.

[95] D. Strugatsky, R. McNulty, K. Munson, C.-K. Chen, S. M. Soltis, G. Sachs, H. Luecke, *Nature* **2013**, *493*, 255.

[96] Y. Cui, K. Zhou, D. Strugatsky, Yi Wen, G. Sachs, Z. H. Zhou, K. Munson, *Sci. Adv.* **2019**, *5*, eaav8423.

[97] K. Fetter, V. Van Wilder, M. Moshelion, F. Chaumont, *Plant Cell* **2004**, *16*, 215.

[98] M. C. Berny, D. Gilis, M. Rooman, F. Chaumont, *Mol. Plant* **2016**, *9*, 986.

[99] R. B. Heinen, G. P. Bienert, D. Cohen, A. S. Chevalier, N. Uehlein, C. Hachez, R. Kaldenhoff, D. Le Thiec, F. Chaumont, *Plant Mol. Biol.* **2014**, *86*, 335.

[100] C. Jozefkowicz, L. Sigaut, F. Scochera, G. Soto, N. Ayub, L. I. Pietrasanta, G. Amodeo, F. L González Flecha, K. Alleva, *Biophys. J.* **2016**, *110*, 1312.

[101] Y. Hiroaki, K. Tani, A. Kamegawa, N. Gyobu, K. Nishikawa, H. Suzuki, T. Walz, S. Sasaki, K. Mitsuoka, K. Kimura, A. Mizoguchi, Y. Fujiyoshi, *J. Mol. Biol.* **2006**, *355*, 628.

[102] B. Yang, D. Brown, A. S. Verkman, *J. Biol. Chem.* **1996**, *271*, 4577.

[103] A. Rossi, T. J. Moritz, J. Ratelade, A. S. Verkman, *J. Cell Sci.* **2012**, *125*, 4405.

[104] C. Silberstein, R. Bouley, Y. Huang, P. Fang, N. Pastor-Soler, D. Brown, A. N. Van Hoek, *Am. J. Physiol. Renal. Physiol.* **2004**, *287*, F501.

[105] R. A. Fenton, H. B. Moeller, M. Zelenina, M. T. Snaebjornsson, T. Holen, N. Macaulay, *Cell. Mol. Life Sci.* **2010**, *67*, 829.

[106] H. Zhang, A. S. Verkman, *J. Mol. Biol.* **2008**, *382*, 1136.

[107] D. S. Frydenlund, A. Bhardwaj, T. Otsuka, M. N. Mylonakou, T. Yasumura, K. G. V. Davidson, E. Zeynalov, Ø. Skare, P. Laake, F.-M. Haug, J. E. Rash, P. Agre, O. P. Ottersen, M. Amiry-Moghaddam, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13532.

[108] S. Noell, P. Fallier-Becker, U. Deutsch, A. F. Mack, H. Wolburg, *Cell Tissue Res.* **2009**, *337*, 185.

[109] S. M. Saparov, D. Kozono, U. Rothe, P. Agre, P. Pohl, *J. Biol. Chem.* **2001**, *276*, 31515.

[110] S. P. Tsunoda, B. Wiesner, D. Lorenz, W. Rosenthal, P. Pohl, *J. Biol. Chem.* **2004**, *279*, 11364.

[111] E. Tajkhorshid, P. Nollert, M. Ø. Jensen, L. J. W. Miercke, J. O'Connell, R. M. Stroud, K. Schulten, *Science* **2002**, *296*, 525.

[112] N. Chakrabarti, B. Roux, R. Pomes, *J. Mol. Biol.* **2004**, *343*, 493.

[113] H. Chen, Y. Wu, G. A. Voth, *Biophys. J.* **2006**, *90*, L73.

[114] J. S. Hub, H. Grubmuller, B. L. de Groot, *Aquaporins*, Springer, Berlin **2009**, p. 57.

[115] A. Burykin, A. Warshel, *FEBS Lett.* **2004**, *570*, 41.

[116] J. A. Ballesteros, H. Weinstein, *Methods Neurosc.* **1995**, *25*, 366.

[117] G. von Heijne, *J. Mol. Biol.* **1992**, *225*, 487.

[118] G. Fischer, U. Kosinska-Eriksson, C. Aponte-Santamaría, M. Palmgren, C. Geijer, K. Hedfalk, S. Hohmann, B. L. de Groot, R. Neutze, K. Lindkvist-Petersson, *PLoS Biol.* **2009**, *7*, e1000130.

[119] D. Kozono, X. Ding, I. Iwasaki, X. Meng, Y. Kamagata, P. Agre, Y. Kitagawa, *J. Biol. Chem.* **2003**, *278*, 10649.

[120] K. Ishibashi, Y. Morishita, Y. Tanaka, *Aquaporins* (Ed.: B. Yang), Springer, Netherlands, Dordrecht, **2017**, pp 35.

[121] R. Zardoya, X. Ding, Y. Kitagawa, M. J. Chrispeels, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14893.

[122] H.-P. Klenk, R. A. Clayton, J.-F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C. I. Reich, L. K. McNeil, J. H. Badger, A. Glodek, L. Zhou, et al., *Nature* **1997**, *390*, 364.

[123] P. Lamosa, A. Burke, R. Peist, R. Huber, M - Y. Liu, G. Silva, C. Rodrigues-Pousada, J. Legall, C. Maycock, H. Santos, *Appl. Environ. Microbiol.* **2000**, *66*, 1974.

[124] L. O. Martins, R. Huber, H. Huber, K. O. Stetter, M. S. Da Costa, H. Santos, *Appl. Environ. Microbiol.* **1997**, *63*, 896.

[125] O. S. Smart, J. M. Goodfellow, B. A. Wallace, *Biophys. J.* **1993**, *65*, 2455.

[126] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, M. S. Sansom, *J. Mol. Graph* **1996**, *14*, 354.

[127] H.-J. Böhm, S. Brode, U. Hesse, G. Klebe, *Chem. – A Eur. J.* **1996**, *2*, 1509.

[128] F. Sterpone, G. Stirnemann, J. T. Hynes, D. Laage, *J. Phys. Chem. B* **2010**, *114*, 2083.

[129] N. Smolin, B. Li, D. A. Beck, V. Daggett, *Biophys. J.* **2008**, *95*, 1089.

[130] R. McNulty, J. P. Ulmschneider, H. Luecke, M. B. Ulmschneider, *Nat. Commun.* **2013**, *4*, 2900.

[131] L. Verdoucq, A. Grondin, C. Maurel, *Biochem. J.* **2008**, *415*, 409.

[132] C. Rodrigues, A. Filipa Mósca, A. P. Martins, T. Nobre, C. Prista, F. Antunes, A. C. Gasparovic, G. Soveral, *Int. J. Mol. Sci.* **2016**, *17*, 2090.

[133] S. L. Reichow, D. M. Clemens, J. A. Freites, K. L. Németh-Cahalan, M. Heyden, D. J. Tobias, J. E. Hall, T. Gonen, *Nat. Struct. Mol. Biol.* **2013**, *20*, 1085.

[134] L. Xin, H. Su, C. H. Nielsen, C. Tang, J. Torres, Y. Mu, *Biochim. Biophys. Acta* **2011**, *1808*, 1581.

[135] F. Zhu, E. Tajkhorshid, K. Schulten, *FEBS Lett.* **2001**, *504*, 212.

[136] R. Mom, B. Muries, P. Benoit, J. Robert-Paganin, S. Réty, J.-S. Venisse, A. Padua, P. Label, D. Auguin, *FEBS Lett.* **2021**, *595*, 41.

[137] J. S. Hub, C. Aponte-Santamaria, H. Grubmuller, B. L. de Groot, *Biophys. J.* **2010**, *99*, L97.

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

[138] K. Murata, K. Mitsuoka, T. Hirai, T. Walz, P. Agre, J. B. Heymann, A. Engel, Y. Fujiyoshi, *Nature* **2000**, *407*, 599.

[139] J. Jiang, B. V. Daniels, D. Fu, *J. Biol. Chem.* **2006**, *281*, 454.

[140] Y. Zhao, H. Xie, L. Wang, Y. Shen, W. Chen, B. Song, Z. Zhang, A. Zheng, Q. Lin, R. Fu, J. Wang, J. Yang, *J. Am. Chem. Soc.* **2018**, *140*, 7885.

[141] J. C. Heiby, B. Goretzki, C. M. Johnson, U. A. Hellmich, H. Neuweiler, *Nat. Commun.* **2019**, *10*, 4378.

[142] L. Janosi, M. Ceccarelli, *PLoS One* **2013**, *8*, e59897.

[143] D. Alberga, O. Nicolotti, G. Lattanzi, G. P. Nicchia, A. Frigeri, F. Pisani, V. Benfenati, G. F. Mangiatordi, *Biochim. Biophys. Acta* **2014**, *1838*, 3052.

[144] S. Kaptan, M. Assentoft, H. P. Schneider, R. A. Fenton, J. W. Deitmer, N. Macaulay, B. L. De Groot, *Structure* **2015**, *23*, 2309.

[145] K. L. Nemeth-Cahalan, J. E. Hall, *J. Biol. Chem.* **2000**, *275*, 6777.

[146] K. L. Nemeth-Cahalan, K. Kalman, J. E. Hall, *J. Gen. Physiol.* **2004**, *123*, 573.

[147] G. Soto, R. Fox, N. Ayub, K. Alleva, F. Guaimas, E. J. Erijman, A. Mazzella, G. Amodeo, J. Muschietti, *Plant J.* **2010**, *64*, 1038.

[148] T. Zeuthen, D. A. Klaerke, *J. Biol. Chem.* **1999**, *274*, 21631.

[149] M. Zelenina, A. A. Bondar, S. Zelenin, A. Aperia, *J. Biol. Chem.* **2003**, *278*, 30037.

[150] A. de Almeida, A. P. Martins, A. F. Mósca, H. J. Wijma, C. Prista, G. Soveral, A. Casini, *Mol. BioSyst.* **2016**, *12*, 1564.

[151] C. Tournaire-Roux, M. Sutka, H. Javot, E. Gout, P. Gerbeau, D.-T. Luu, R. Bligny, C. Maurel, *Nature* **2003**, *425*, 393.

[152] M. Yasui, A. Hazama, T.-H. Kwon, S. Nielsen, W. B. Guggino, P. Agre, *Nature* **1999**, *402*, 184.

[153] A. Mósca, A. De Almeida, D. Wragg, A. Martins, F. Sabir, S. Leoni, T. Moura, C. Prista, A. Casini, G. Soveral, *Cells* **2018**, *7*, 207.

[154] L. Leitão, C. Prista, T. F. Moura, M. C. Loureiro-Dias, G. Soveral, *PLoS One* **2012**, *7*, e33219.

[155] M. Fischer, R. Kaldenhoff, *J. Biol. Chem.* **2008**, *283*, 33889.

[156] M. H. Olsson, C. R. Sondergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525.

[157] C. R. Søndergaard, M. H. Olsson, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 2284.

[158] S. F. Truelsen, J. W. Missel, K. Gotfryd, P. A. Pedersen, P. Gourdon, K. Lindorff-Larsen, C. Hélix-Nielsen, *Biochim. Biophys. Acta Biomembr.* **2021**, *1864*, 183809.

[159] T. Steinbrecher, S. Prock, J. Reichert, P. Wadhwani, B. Zimpfer, J. Bürck, M. Berditsch, M. Elstner, A. S. Ulrich, *Biophys. J.* **2012**, *103*, 1460.

[160] R. P. Bahadur, P. Chakrabarti, F. Rodier, J. Janin, *Proteins* **2003**, *53*, 708.

[161] D. Xu, C. J. Tsai, R. Nussinov, *Protein Eng.* **1997**, *10*, 999.

[162] K. Gupta, J. A. C. Donlan, J. T. S. Hopper, P. Uzdavinys, M. Landreh, W. B. Struwe, D. Drew, A. J. Baldwin, P. J. Stansfeld, C. V. Robinson, *Nature* **2017**, *541*, 421.

[163] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774.

[164] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, H. I. Mosberg, *Bioinformatics* **2006**, *22*, 623.

[165] I. K. McDonald, J. M. Thornton, *J. Mol. Biol.* **1994**, *238*, 777.

[166] C. N. Pace, B. A. Shirley, M. McNutt, K. Gajiwala, *FASEB J.* **1996**, *10*, 75.

[167] A. R. Fersht, *Trends Biochem. Sci.* **1987**, *12*, 301.

[168] A. Horovitz, L. Serrano, B. Avron, M. Bycroft, A. R. Fersht, *J. Mol. Biol.* **1990**, *216*, 1031.

[169] M. Akke, S. Forsen, *Proteins: Struct., Funct., Genet.* **1990**, *8*, 23.

[170] S. F. Betz, *Protein Sci.* **1993**, *2*, 1551.

[171] J. Clarke, A. R. Fersht, *Biochemistry* **1993**, *32*, 4322.

[172] S. Kim, T.-J. Jeon, A. Oberai, D. Yang, J. J. Schmidt, J. U. Bowie, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14278.

[173] A. J. Situ, S.-M. Kang, B. B. Frey, W. An, C. Kim, T. S. Ulmer, *J. Phys. Chem. B* **2018**, *122*, 1185.

[174] A. Laganowsky, E. Reading, T. M. Allison, M. B. Ulmschneider, M. T. Degiacomi, A. J. Baldwin, C. V. Robinson, *Nature* **2014**, *510*, 172.

[175] V. Schmidt, M. Sidore, C. Bechara, J. P. Duneau, J. N. Sturgis, *Biochim. Biophys. Acta Biomembr.* **2019**, *1861*, 431.

[176] R. A. Corey, W. Song, A. L. Duncan, T. B. Ansell, M. S. P. Sansom, P. J. Stansfeld, *Sci. Adv.* **2021**, *7*, eabh2217.

[177] K. Pluhackova, A. Horner, *BMC Biol.* **2021**, *19*, 4.

[178] S. G. Garg, S. B. Gould, *Trends Cell Biol.* **2016**, *26*, 894.

[179] G. von Heijne, *Nature* **1989**, *341*, 456.

[180] J. Neumann, N. Klein, D. E. Otzen, D. Schneider, *Arch. Biochem. Biophys.* **2014**, *564*, 281.

[181] A. J. Yool, A. M. Weinstein, *N. Physiol. Sci.* **2002**, *17*, 68.

[182] A. J. Yool, E. M. Campbell, *Mol. Aspects Med.* **2012**, *33*, 553.

[183] S. D. Tyerman, S. A. McGaughey, J. Qiu, A. J. Yool, C. S. Byrt, *Annu. Rev. Plant Biol.* **2021**, *72*, 703.

[184] M. Kourghi, J. V. Pei, M. L. De Ieso, S. Nourmohammadi, P. H. Chow, A. J. Yool, *Clin. Exp. Pharmacol. Physiol.* **2018**, *45*, 401.

[185] Y. Wang, J. Cohen, W. F. Boron, K. Schulten, E. Tajkhorshid, *J. Struct. Biol.* **2007**, *157*, 534.

[186] G. Portella, B. L. de Groot, *Biophys. J.* **2009**, *96*, 925.

[187] A. Berezhkovskii, G. Hummer, *Phys. Rev. Lett.* **2002**, *89*, 064503.

[188] R. M. Stroud, P. Nollert, L. Miercke, *Membr. Proteins* **2003**, *63*, 291.

[189] F. Cymer, D. Schneider, *Biochemistry* **2010**, *49*, 279.

[190] M. J. Borgnia, P. Agre, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2888.

[191] J. J. Galka, S. J. Baturin, D. M. Manley, A. J. Kehler, J. D. O'neil, *Biochemistry* **2008**, *47*, 3513.

[192] D. A. Doyle, João M Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, R. Mackinnon, *Science* **1998**, *280*, 69.

[193] J. Payandeh, T. Scheuer, N. Zheng, W. A. Catterall, *Nature* **2011**, *475*, 353.

[194] E. Neumann, K. Rosenheck, *J. Membr. Biol.* **1972**, *10*, 279.

[195] S. A. Kirsch, R. A. Bockmann, *Biochim. Biophys. Acta* **2016**, *1858*, 2266.

[196] F. Huang, Z. Fang, J. Mast, W. Chen, *Bioelectromagnetics* **2013**, *34*, 253.

[197] R. Patel, S. M. Smith, C. Robinson, *Biochim. Biophys. Acta–Mol. Cell Res.* **2014**, *1843*, 1620.

[198] N. Klein, N. Hellmann, D. Schneider, *Biophys. J.* **2015**, *109*, 722.

[199] V. Schmidt, J. N. Sturgis, *ACS Omega* **2017**, *2*, 3017.

[200] F. Cymer, D. Schneider, *Biol. Chem.* **2012**, *393*, 1215.

[201] M. Trefz, R. Keller, M. Vogt, D. Schneider, *Biochim. Biophys. Acta* **2017**, *1860*, 88.

[202] N. Klein, M. Trefz, D. Schneider, *Int. J. Mol. Sci.* **2019**, *20*, 927.

[203] D. S. Sholl, R. P. Lively, *Nature* **2016**, *532*, 435.

[204] M. Kumar, M. Grzelakowski, J. Zilles, M. Clark, W. Meier, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20719.

[205] L. Xia, M. F. Andersen, C. Hélix-Nielsen, J. R. McCutcheon, *Ind. Eng. Chem. Res.* **2017**, *56*, 11919.

[206] A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, *Bioinformatics* **2009**, *25*, 1189.

[207] M. Kumar, J. E. Habel, Y. X. Shen, W. P. Meier, T. Walz, *J. Am. Chem. Soc.* **2012**, *134*, 18631.

[208] E. Krissinel, *J. Comput. Chem.* **2010**, *31*, 133.

[209] W. L. DeLano, *CCP4 Newsl. Protein Crytallogr* **2002**, *40*, 82.

[210] R. Core Team, *R: A Language and Environment for Statistical Computing*, http://www.R-project.org (accessed: December 2020).

[211] K. Tamura, G. Stecher, S. Kumar, *Mol. Biol. Evol.* **2021**, *38*, 3022.

[212] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1–2*, 19.

[213] K. H. Pluhackova, A. Horner, *BMC Biol.* **2021**, *19*, 4.

[214] K. Pluhackova, T. A. Wassenaar, R. A. Bockmann, *Methods Mol. Biol.* **2013**, *1033*, 85.

[215] B. Webb, A. Sali, *Curr. Protoc. Bioinf.* **2016**, *54*, 561.

[216] G. Hu, L. Qi, X. Dou, J. Wang, *Mol. Simul.* **2013**, *39*, 261.

[217] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, E. Lindahl, *J. Chem. Theory Comput.* **2010**, *6*, 459.

[218] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, A. D. MacKerell Jr., *Nat. Methods* **2017**, *14*, 71.

[219] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. Mackerell, R. W. Pastor, *J. Phys. Chem. B* **2010**, *114*, 7830.

[220] W. L. Jorgensen, J. D. Madura, *Mol. Phys.* **1985**, *56*, 1381.

[221] O. Beckstein, M. S. P. Sansom, *Phys. Biol.* **2004**, *1*, 42.