Structural Basis for Silicic Acid Uptake by Higher Plants

Bert van den Berg [1]*, **Conrado Pedebos** [2,5], **Jani R. Bolla** [3,4], **Carol V. Robinson** [3,4], **Arnaud Baslé** [1] **and Syma Khalid** [2,5]*

1 - Biosciences Institute, The Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH, UK
2 - School of Chemistry, University of Southampton, Southampton SO17 1BJ, UK
3 - Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, UK
4 - The Kavli Institute for Nanoscience Discovery, South Parks Road, Oxford OX1 3QU, UK
5 - The Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Correspondence to Bert van den Berg and Syma Khalid: School of Chemistry, University of Southampton, Southampton SO17 1BJ, UK (S. Khalid). Biosciences Institute, The Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH, UK (B. van den Berg). *bert.van-den-berg@ncl.ac.uk (B. van den Berg), s.khalid@soton.ac.uk (S. Khalid), @ConradoPedebos* 🐦 *(C. Pedebos), ProfSyK* 🐦 *(S. Khalid)*
https://doi.org/10.1016/j.jmb.2021.167226
Edited by Daniel L. Minor

Abstract

Many of the world's most important food crops such as rice, barley and maize accumulate silicon (Si) to high levels, resulting in better plant growth and crop yields. The first step in Si accumulation is the uptake of silicic acid by the roots, a process mediated by the structurally uncharacterised NIP subfamily of aquaporins, also named metalloid porins. Here, we present the X-ray crystal structure of the archetypal NIP family member from *Oryza sativa* (OsNIP2;1). The OsNIP2;1 channel is closed in the crystal structure by the cytoplasmic loop D, which is known to regulate channel opening in classical plant aquaporins. The structure further reveals a novel, five-residue extracellular selectivity filter with a large diameter. Unbiased molecular dynamics simulations show a rapid opening of the channel and visualise how silicic acid interacts with the selectivity filter prior to transmembrane diffusion. Our results will enable detailed structure–function studies of metalloid porins, including the basis of their substrate selectivity.

Introduction

Many higher plants, and members of the *Poaceae* (grasses) in particular, accumulate Si to high levels (up to 10% w/w for rice).[1] Si is generally not yet considered an essential plant element, but a high Si content provides resistance to abiotic and biotic stresses, improves the light-interception ability by plants in a community, and minimises transpiration losses.[1–3] Silicic acid (H_4SiO_4, $pK_a = 9.25$) is the naturally occurring bioavailable form of Si. At pH values in most soils, it is a polar but neutral molecule and soluble to concentrations of ~2 mM.[4] In shoots and leaves, accumulating silicic acid is spontaneously transformed to solid amorphous silica (SiO_2–nH_2O) called silica bodies, which are deposited mainly in the cell walls of different tissues and generate structural and mechanical stability.[1,2] The facilitated diffusion of Si and other metalloids such as boron and arsenic across bilayers is mediated by members of the NIP (<u>N</u>odulin26 <u>I</u>ntrinsic <u>P</u>rotein) subfamily of aquaporins,[5] also termed metalloid porins.[2] NIPs occur not only in roots but in most plant tissues, and can be divided into three functional groups, NIP-I, NIP-II and NIP-III, based on the composition of the four-residue selectivity fil-

ter (SF) or aromatic-arginine (ar/R) region at the extracellular mouth of the channel.[2] NIP-I members have the most stringent substrate selectivity and most appear to only transport water, glycerol and arsenous acid (H_3AsO); however, boric acid and lactic acid have recently also been shown to be transported by NIP-I proteins.[6,7] NIP-II members additionally transport boric acid, while members of the NIP-III subgroup have the broadest selectivity and also transport silicic acid. Based on the pre-

dicted predominance of small residues at the SF (typically GSGR), it has been proposed that NIP-III channels may have the widest SF,[2] which would explain why they transport the widest range of substrates.

Lsi1 (low silicon rice 1), caused by the absence of *Oryza sativa* NIP2;1 (OsNIP2;1), is a rice mutant defective in silicic acid uptake with various pest-sensitive phenotypes, and has a grain yield of only 10% compared to wild type rice.[8] OsNIP2;1 has
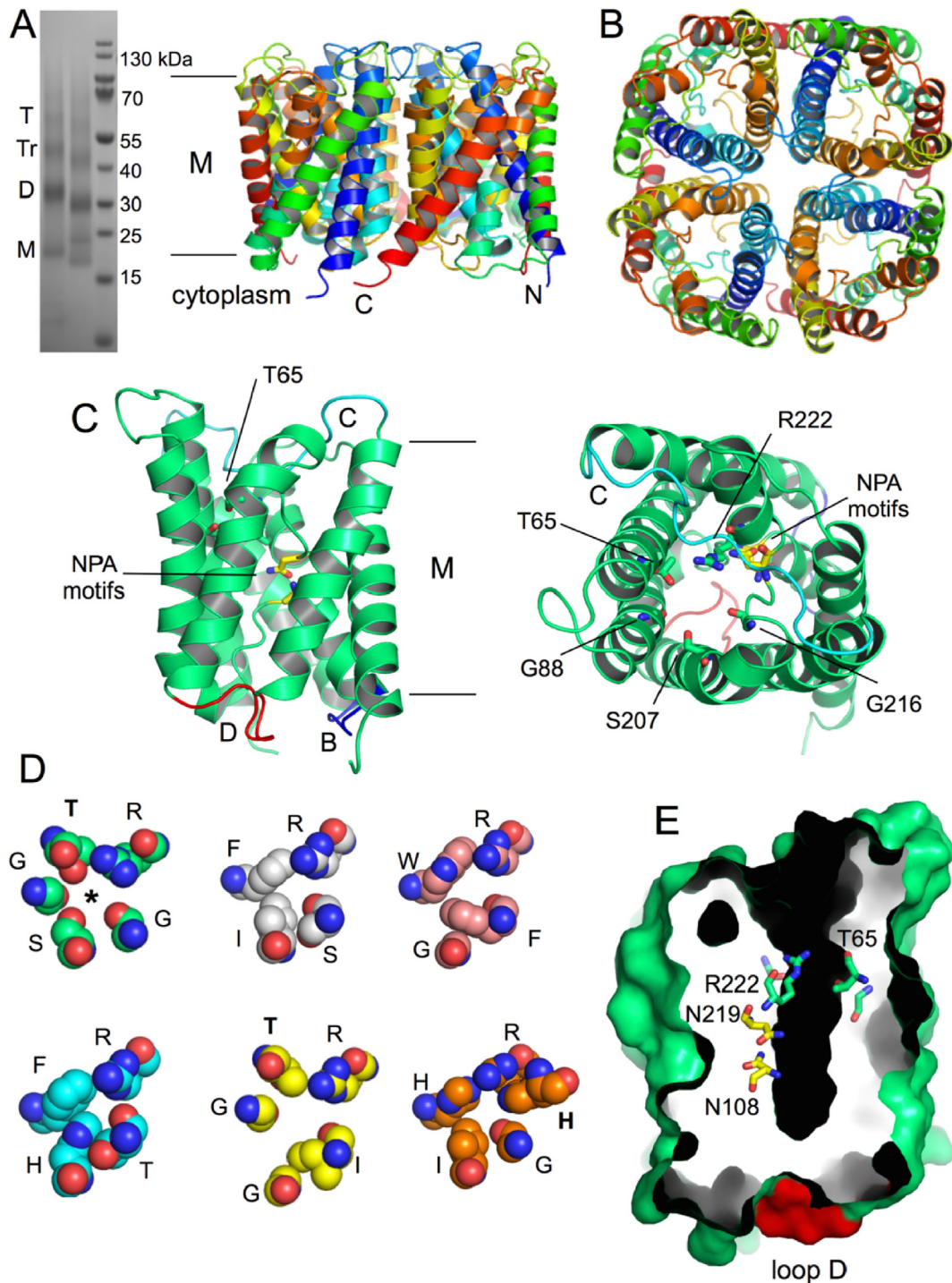
been characterised extensively using *Xenopus* oocytes and was found to transport silicic acid efficiently, but boric acid and glycerol very poorly, indicating substrate selectivity.[9] OsNIP2;1 is highly expressed on the distal side of plasma membranes of root cells, and functions together with the active silicon efflux transporter Lsi2, localised in the proximal membranes of root cells, to drive unidirectional silicic acid transport towards the xylem.[10,11] Phylogenetically, NIP family members cluster together with bacterial and archaeal AqpN proteins in arsenic resistance (*ars*) operons, suggesting NIP channels may have evolved from arsenous acid efflux proteins.[12] Indeed, OsNIP2,1 also transports arsenous acid efficiently,[13] making it a major cause of toxic arsenic accumulation in the rice grain. In contrast to classical aquaporins (AQPS) and aquaglyceroporins (AQGPs),[5,14] and despite considerable interest in members of this subfamily,[2,8] no structural information is yet available for any NIP family member, and the basis for metalloid selectivity remains therefore unclear. To enable such studies in the future, we report here the X-ray crystal structure and exploratory molecular dynamics simulations of OsNIP2,1.

Results and discussion

We expressed OsNIP2;1 to high levels (∼1 mg/liter) in *Saccharomyces cerevisiae*, followed by purification in various detergents. The protein behaves well in decyl maltoside (DM) and decyl maltose neopentyl glycol (DMng) (Supplementary Figure 1(A)). Well-diffracting crystals could only be obtained from a truncated form of OsNIP2;1 generated via limited trypsinolysis, which was shown to comprise residues 38–264 by proteomic analysis. The truncated form is a tetramer in solution, as confirmed by native mass spectrometry (Supplementary Figure 1(B) and (C)). The X-ray crystal structure was determined using data to 2.6 Å resolution, via molecular replacement with archaeal AqpM (Supplementary Table 1). A low-resolution structure (3.8 Å) was also obtained for the full-length (FL) protein, and is

identical to that of the truncated protein, suggesting the poorly conserved N- and C-terminal ∼35 residues that are invisible in the FL structure are likely disordered. As expected from sequence similarity (Supplementary Figure 2), OsNIP2;1 displays the typical aquaporin fold, with a tetrameric assembly in which each protomer has six transmembrane helices and two half-helices in cytoplasmic loop B and extracellular loop E, which together form a seventh pseudo-transmembrane segment with the important NPA motifs in the centre of the bilayer (Figure 1(B) and Supplementary Figure 2).

Inspection of the extracellular mouth of the channel reveals that the "GSGR" SF includes a fifth residue, Thr65 (Figure 1(C) and (D)). In most other AQP and AQGP structures, the SF has only four residues due to the presence of an aromatic residue at the first position (*e.g.* FISR in AqpM), which occludes the conserved glycine or threonine that corresponds to Thr65 in OsNIP2;1 (Figure 1 (D)). Thus, our structure suggests that plant silicic acid transporters have a novel, 5-residue SF (TGSGR) that is substantially wider than the 4-residue SF of conventional AQPs (Figure 1(D)). Intriguingly, a similar, but as yet unrecognised 5-residue SF (TGGIR) is present in human AQP10,[15] a known AQGP which was recently shown to transport silicic acid at levels that may be physiologically relevant.[16] However, in contrast to AQP10, OsNip2;1 transports glycerol very poorly,[8] and we propose that this is due to the lack of a large hydrophobic SF residue (*e.g.* Trp and Phe in GlpF; Ile in AQP10), which is known to interact with the backbone of the amphipathic glycerol molecule.[17] Another AQP with an extended SF is TIP2,1 from *Arabidopsis thaliana* (AtTIP2;1).[18] In AtTIP2;1, an extra histidine residue from loop C (His131) lines the pore and pushes the absolutely conserved SF Arg to the side, causing it to form a hydrogen bond with the other SF histidine (His63) (Figure 1(D)). In other aquaporins, the equivalent residue is either a Thr or Asn (Supplementary Figure 2), the sidechains of which interact with the sidechain of the SF Arg residue but are too short to

Figure 1. X-ray crystal structure of OsNIP2;1. (A) SDS-PAGE gel of purified OsNIP2;1 before (left lane) and after (middle lane) limited trypsinolysis. Molecular markers have been loaded in the right lane. The likely oligomeric states of OsNIP2;1 have been indicated (monomer to tetramer). (B) Cartoon overview of the OsNIP2;1 tetramer, rainbow coloured by chain from the N terminus (blue) to the C terminus (red). The approximate membrane (M) boundaries are shown. Left, view from the membrane plane; right, top view from the extracellular side. (C) Monomer cartoon viewed from the membrane plane (left panel) and from the extracellular side. Selectivity filter (SF) residues and the asparagine residues of the two NPA motifs are shown as sticks. Cytoplasmic loops B and D are coloured dark blue and red, respectively. Extracellular loop C is coloured cyan. (D) Comparison of SFs of aquaporins. Shown are (top row) OsNIP2;1 (green carbons), *M. marburgensis* AqpM (gray), *E. coli* GlpF (pink), and (bottom row) SoPIP2;1 (cyan), *H. sapiens* AQP10 (yellow) and A. thaliana TIP2;1 (orange). The fifth SF residue (T or H) is shown in bold. The OsNIP2;1 channel is indicated with an asterisk. (E) Membrane plane slab surface view showing the cytoplasmic block of the channel caused by loop D (red). SF residues and the asparagines of the two NPA motifs are shown as sticks.

contribute to the SF itself. Thus, while the SFs of OsNIP2;1 and AtTIP2;1 both contain five residues rather than the canonical four, the different positions and properties of the fifth residue (Thr in OsNIP2;1, His in AtTIP2;1) likely affect pore properties in different ways. However, it is clear, at least from the static crystal structure, that the OsNIP2;1 SF is unique in having four oxygen atoms lining the channel, providing multiple hydrogen bond donor and acceptor groups for the four hydroxyl groups of the translocating silicic acid molecule.

As in other plant aquaporins such as SoPIP2;1, the sidechain of the SF Arg residue forms a hydrogen bond with the carbonyl of a glycine in the long extracellular loop C that lines the external vestibule of the channel (Gly155 in OsNIP2;1, Gly151 in SoPIP2;1). This interaction likely constrains the Arg sidechain in a position that allows substrate translocation (Supplementary Figure 3). Interestingly, a recent study suggested that in Si transporters, a precise spacing of 108 residues between the NPA motifs is crucial for Si selectivity.[19] While appearing baffling at first, our structure provides an explanation for these data. The residue affected by the insertions or deletions made by the authors corresponds to Gly155 in OsNIP2,1. Disrupting this interaction might cause increased mobility of the Arg side chain and/or occlusion of the channel by loop C, precluding efficient substrate passage. Thus, rather than depending on a precise number of amino acids between the NPA motifs, efficient silicic acid transport requires the SF Arg side chain to interact with the backbone of a glycine residue in loop C.

On the cytoplasmic side, the OsNIP2;1 channel is completely closed by loop D (residues [185]ATDTRA[191]; Figure 1(E)), suggesting protein-based regulation of transport activity. The closed channel of OsNIP2;1 is perhaps surprising, because silicic acid is not known to be toxic and no phosphorylation of OsNIP2;1 has yet been observed. Moreover, in plants, the "overaccumulation" of silicic acid and its deposition as silica bodies is precisely the function of the transporter. On the other hand, OsNIP2;1 expression is downregulated upon continuous silicic acid exposure in some rice cultivars,[8] and loop D is conserved in NIPs, suggesting functional importance. Classical water-specific plant aquaporins such as PIP2,1 from spinach (SoPIP2;1) utilise (de)phosphorylation and protonation of a channel exit-lining histidine residue to protect the plant from drought stress and floods respectively, conditions which require the channel to close.[20] No evidence for phosphorylation was detected for yeast-expressed FL and truncated OsNIP2;1 by mass spectrometry analyses (Supplementary Figure 1) and inspection of the electron density, suggesting that channel opening might require phosphorylation of, for example, Thr186 or Thr188 in loop D. To test this notion and to assess the significance of the

closed OsNIP2;1 channel observed in the crystal, we performed equilibrium molecular dynamics (MD) simulations on the tetrameric assembly embedded in a POPC bilayer (Supplementary Table 2). Strikingly, by taking the distance between Phe121 in loop B and Arg189 in loop D as a proxy for channel opening, three independent 0.5 μs simulations show a pronounced shift (up to ~15 Å for the Arg189 side chain) of loop D that opens the channel (Figure 2(A) and (B)). Most (10 out of 12) monomers showed either a partial or complete channel opening (Figure 2(B) and Supplementary Figure 4(A)), suggesting the closed conformation of the channel was selected by the crystallisation process. HOLE profiles show that even the closed channels in the MD simulations are more open than the crystal structure (Figure 2(C)). Interestingly, and as expected from the fact that OsNIP2;1 has been shown experimentally to transport water,[9] the channels fill with water on a picosecond time scale during system equilibration. Calculation of bidirectional water flux gives a reasonable correlation with the open/closed state of the channel. In OsNIP2,1$_{WATER1}$ (Figure 2(B)), for example, the tetrameric flux corresponds to 7.1 waters/ns, taken over the entire 500 ns of the simulation. However, the mostly open channel C (yellow) accounts for 2.8 waters/ns, and the mostly closed channel D (red) for 0.9 waters/ns. As a comparison, the glycerol-specific AQP7,[21] which also transports water very efficiently, yields a slightly higher flux of 12 waters/ns for the tetramer. The efficient water transport by OsNIP2,1 suggests that there might indeed be a requirement for channel closure, not to prevent silicic acid accumulation, but to avoid deleterious consequences of water stress. There are many candidate residues for channel gating, including phosphorylation of threonine(s) in loop D or serine(s) in the non-resolved C terminus (Supplementary Figure 2), or protonation of ionisable histidine residue(s), analogous to SoPIP2;1.[20] Given their considerable differences, comparison of the SoPIP2,1 and OsNIP2;1 structures does not give any obvious clues (Supplementary Figure 5). Another candidate mechanism to generate/stabilise a closed channel might be protonation of Asp187 in loop D, which interacts with Arg119 in loop B and Arg189 in loop D. To test this possibility, we performed MD simulations with protonated Asp187 in the same way as for wild type. As shown in Supplementary Figure 4(B), the channels also open readily, indicating that Asp187 is not likely to cause OsNIP2;1 gating. Clearly, establishing whether gating occurs in NIP channels and elucidation of the gating mechanism remains an important goal for future work.

We next generated a system with silicic acid on both sides of the bilayer and performed unbiased MD simulations. Initial setups with 0.1 and 0.5 M silicic acid did not give any translocation events in 250 ns simulations, and we therefore used 1 M
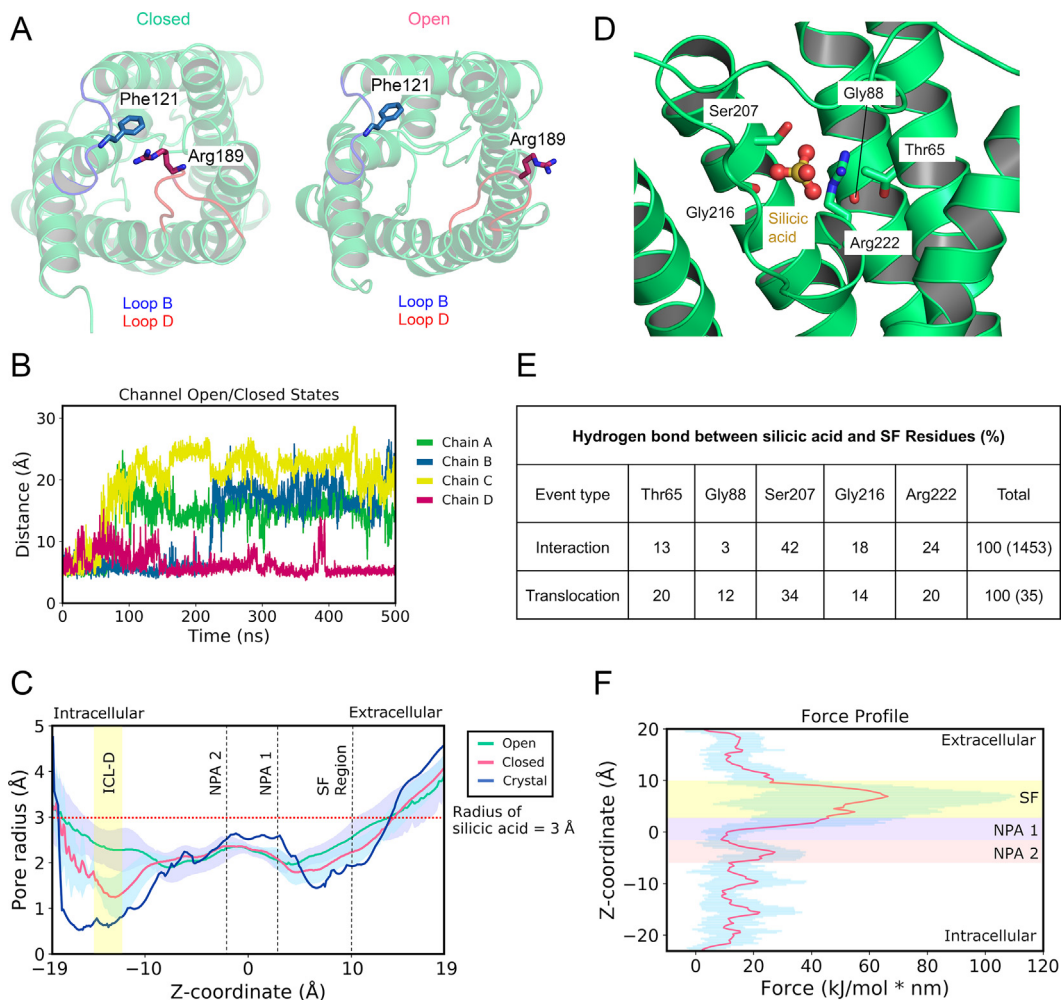
Figure 2. Molecular dynamics simulations of OsNIP2;1. (A) Comparison of the closed (left panel) and open channel structures. (B) Channel opening in the tetramer vs. simulation time. (C) HOLE profiles for the X-ray crystal structure and open and closed states from the MD simulations. The radius of the silicic acid molecule is indicated for reference. (D) Simulation snapshot of a silicic acid molecule in the SF. (E) Statistics for hydrogen bonds between silicic acid and SF residues for all interaction events and translocations only. (F) Steered MD force profile for silicic acid translocation. Average (red) value and standard deviations (blue) are shown for ten simulations.

silicic acid. In total, 3 μs of simulations were performed with silicic acid, yielding a total of three uptake events (defined as movement from the extracellular side into the space corresponding to the cytoplasm; Supplementary Movie). The translocating silicic acid molecules form hydrogen bonds with all residues of the SF including Thr65 (Figure 2(D) and (E)), but Ser207 appears to be especially important. Only a small fraction of interaction events at the SF result in translocation, suggesting that passage through the SF constitutes a thermodynamic or kinetic energetic barrier. Interestingly, the diameter of silicic acid (6 ± 0.5 Å) is larger than that of most of the channel, as determined via HOLE. This apparent discrepancy is due to the fact that HOLE calculates diameters based on the largest sphere that fits the channel. Given that both the channel

and silicic acid are non-spherical, this leaves enough space for permeation, but it does not allow co-permeation of silicic acid hydration waters. As expected from its symmetry, the translocating silicic acid molecule has no fixed orientation and rotates during its passage through the channel (Figure 3). At the NPA region, the hydroxyl hydrogens of the silicic acid point away from the asparagine side chains, analogous to the reorientation of water in classical aquaporins.

In addition to three spontaneous uptake events, we also observed three export events, with silicic acid moving from the cytoplasmic side to the extracellular side. In these cases, prior to entering the channel, the silicic acid interacts extensively with residues located at the channel entrance, especially His106, Arg119 and Asp187 (Supplementary Figure 6). Interactions with the
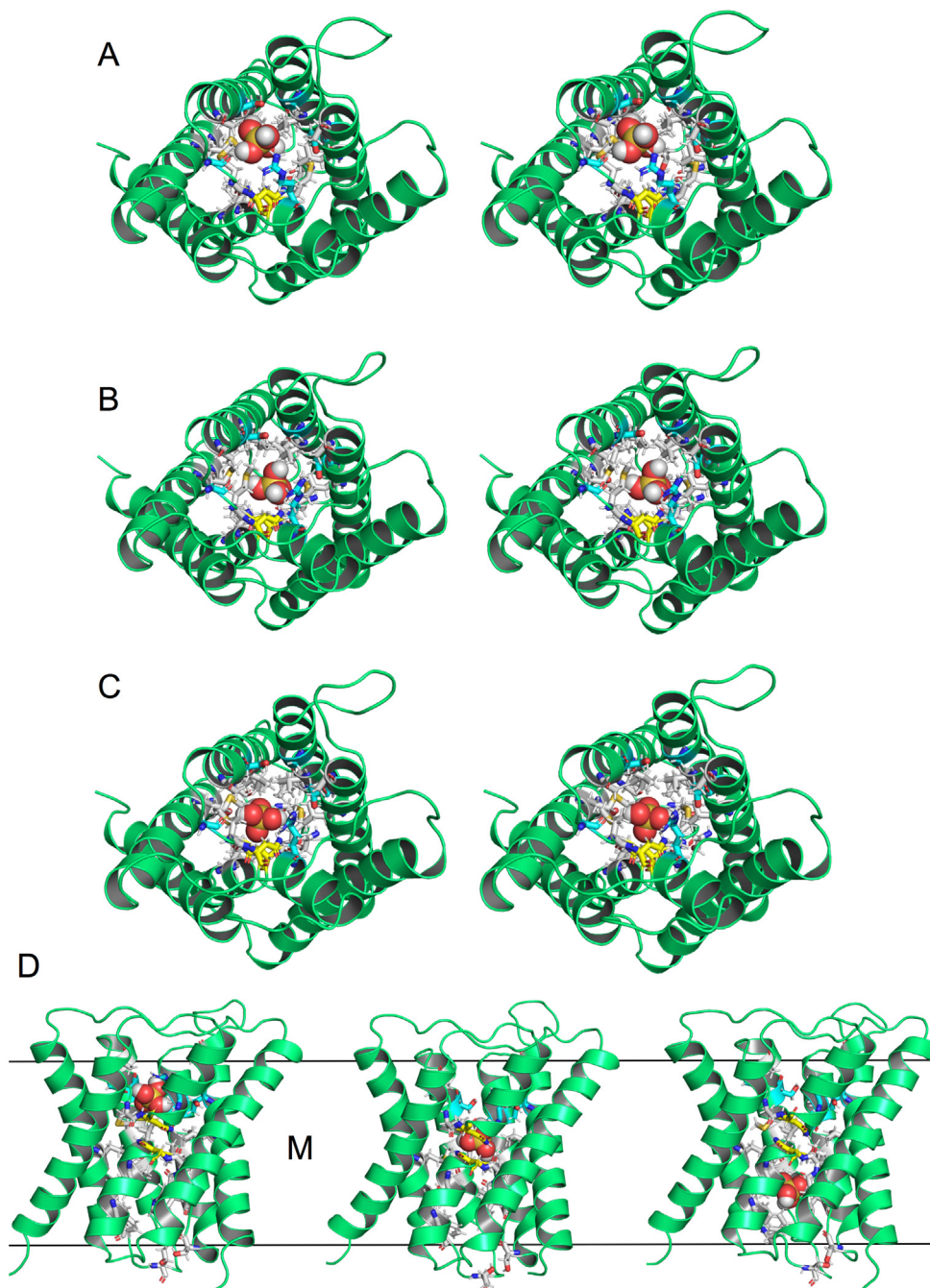
Figure 3. Silicic acid translocation through OsNIP2;1. **(**A–C) Stereo view snapshots viewed from the extracellular side of silicic acid interacting with the SF (A), NPA motifs (B) and with His106 in the exit site (C). (D) Views from the plane of the membrane, with the membrane boundaries indicated. SF residues are coloured cyan, Asn108 and Asn219 of the NPA motifs are yellow, and other channel-lining residues are grey. The silicic acid molecule is shown as a space-filling model.

same residues are observed during uptake, after silicic acid exits the channel proper and prior to diffusion away from the protein. Thus, although the sample size is small, the similar number of silicic acid uptake and export events suggests that, in the absence of a gradient across the bilayer, silicic acid transport would be bidirectional.

To increase the number of translocation events without the need for excessively long simulations, steered MD simulations were performed to provide further information about the location of translocation barriers. The SF region combined with the NPA1 motif (z ∼ 10–0 Å) provides the largest barrier for the silicic acid molecule, with a

large deviation which indicates that the size of the barrier is dependent on the conformational arrangement of the amino acids in these regions. A second, smaller, barrier exists for the NPA2 domain ($z \sim -2$ to -6 Å), which the silicic acid molecule seems to cross relatively easily in the simulations. Residence times for the silicic acid in each of these regions were 16.0 ns ± 1.6 for the SF, 9.9 ns ± 2.5 for NPA1, and 8.4 ns ± 1.3 for NPA2 (Methods). These values correspond reasonably well to those obtained from the spontaneous permeation events (SF, 18.8 ns ± 13. 5; NPA1, 8.3 ns ± 4.7; NPA2, 3.5 ns ± 2.7), and confirm that the SF presents the largest permeation barrier for silicic acid translocation.

Si is a very important element for plants,[1] and accumulating evidence suggests that it has health benefits for humans as well.[3] Our results provide a platform that will lead to an improved understanding of silicic acid uptake by plants that may be utilised for Si biofortification of important crops. Within the wider context of metalloid transport, the study of NIP family members at the atomic level will generate insights into metalloid selectivity, that, together with *in vitro* and *in planta* studies, may enable rational manipulation of, *e.g.*, boron levels in barley and arsenic accumulation in the rice grain.

Materials and methods

OsNIP2;1 expression and purification

The gene sequence encoding for full-length *osnip2;1* was obtained by gene synthesis (Eurofins genomics; Supplementary Figure 7) and optimised for expression in *Saccharomyces cerevisiae*. Notwithstanding their prediction as cytoplasmic residues, three putative N-terminal N-glycosylation sites were removed via replacement with glutamines (N4Q/N13Q/N26Q). The resulting gene, encoding a hexa-histidine sequence at the C terminus for purification, was cloned via BamHI and XhoI restriction sites into the 83v vector[22] digested with the same enzymes. The plasmid was moved into the yeast W303 Δ*pep4* expression strain via the lithium acetate method. Transformants were selected on SCD -His plates (ForMedium), incubated at 30 °C.

For expression, cells were grown in shaker flasks at 30 °C for ~20–24 hrs in synthetic minimal medium lacking histidine and with 1% (w/v) glucose to a typical A_{600} of 6–8. Cells were subsequently spun down for 15 mins at 4200 rpm and resuspended in YP medium containing 1.5% (w/v) galactose, followed by another 16–20 hrs growth at 30 °C/225 rpm, and harvested by centrifugation for 20 mins at 4200 rpm. Final A_{600} values typically reached 18–20. Cells were resuspended in TSB buffer (20 mM Tris, 300 mM NaCl, pH 8) in the presence of 5 mM EDTA and lysed by 1–2 passes through a cell disrupter operated at 35–37 kpsi (TS-Series 0.75 kW;

Constant Systems). Membranes were collected from the suspension by centrifugation at 200,000g for 90 mins (45Ti rotor; Beckman). Membrane protein extraction was performed by homogenisation in TSB with a 1:1 (w/w) mixture of dodecyl-β-D-maltoside and decyl-β-D-maltoside (DDM/DM) followed by stirring at 4 °C for 1 hr or overnight. Depending on the amount of processed cells, one or two protease inhibitor tablets were added at this stage (Complete EDTA-free protease cocktail; Sigma). Typically, 1 g (1% w/v) of total detergent was used for membranes from 2 liters of cells. The membrane extract was centrifuged for 35 mins at 200,000g and the supernatant loaded onto a 10 ml Nickel column (Chelating Sepharose; GE Healthcare) equilibrated in TSB with 0.2% DDM pH 8. The column was washed with 15 column volumes buffer containing 30 mM imidazole and eluted in 3 column volumes with 200 mM imidazole. The protein was purified to homogeneity by gel filtration chromatography (Superdex 200 16/600) in 10 mM Hepes, 100 mM NaCl, 0.05% DDM pH 7–7.5. For polishing and detergent exchange, a second gel filtration column was performed using various detergents. The final yield of OsNIP2;1 was about 1 mg per 4 liters of culture with $A_{600} \sim$ 18–20. Proteins were concentrated to ~10–15 mg/ml using 100 kD cutoff centrifugal devices (Millipore), flash-frozen and stored at −80 °C prior to use.

Crystallisation and structure determination

Crystallisation screening trials by sitting drop vapor diffusion were set up at 4 °C and 20 °C using in-house screens and the MemGold, MemGold2, MemChannel and MemTrans screens (Molecular Dimensions) with a Mosquito crystallisation robot (TTP Labtech). Crystals were harvested directly from the initial trials or optimised by hanging drop vapor diffusion using larger drops (typically 2–3 μl total volume). Crystals diffracting beyond 5 Å were obtained only with 0.05% decyl-maltose neopentyl glycol (DMng), with the best crystal showing useable diffraction to just beyond 4 Å (C2 space group). However, the diffraction was anisotropic, and molecular replacement solutions could not be obtained. To improve diffraction, limited proteolysis trials were performed at 4 °C with trypsin and chymotrypsin, using a 100-fold excess (w/w) of OsNIP2;1 over protease. The digestions with trypsin showed removal of ~5–10 kD from the protein based on SDS-PAGE. Following a large-scale digest (~5 mg), the truncated protein was subjected to SEC in DMng as described above (Supplementary Figure 1(A)), with the addition of 10% glycerol to the buffer. Native mass spectrometry showed a molecular mass for the monomer of 23,960 Da, indicating the likely removal of residues 1–37 from the N terminus and

residues 265–304 from the C terminus (predicted molecular mass 23964.1 Da). After initial screening as described above, diffracting, block-shaped crystals with various morphologies in space group P1 were obtained following optimisation of the MemGold H11 hit condition (1 mM CdCl$_2$, 30 mM MgCl$_2$.6H$_2$O, 0.1 M MES pH6.5, 30% PEG400), by slightly varying the PEG 400 concentration (28–32% w/v). Due to severe anisotropy in most crystals, a number of crystals had to be screened in order to obtain a moderately anisotropic dataset that allowed successful structure solution. Datasets (360°) were collected at beamline I24 at the Diamond Light Source and were autoprocessed via XDS within Autoproc[23] and STARANISO[24]. Useful phases were obtained via molecular replacement (MR) with Phaser,[25] using as search model the tetramer of the AqpM aquaporin from *Methanothermobacter marburgensis* (PDB ID: 2EVU), which has 34% sequence identity to OsNIP2;1. The asymmetric unit (AU) contains two OsNIP2;1 tetramers, corresponding to a solvent content of ~65% (Matthews volume 3.5 Å^3/Da). The two tetramers within the AU pack via their cytoplasmic faces, in part via bridging cadmium ions present in the crystallisation mixture. However, it is not clear whether the cadmium ions play a critical role in lattice formation, given that the C2 crystals for the full-length protein, obtained in the absence of cadmium, pack in a very similar manner. The initial model was improved via iterative cycles of Autobuilding within Phenix,[26] manual building within Coot,[27] and refinement via Phenix. The data for refinement were cut off at 3.0 Å, since higher resolution cutoffs led to unstable refinements with high clash scores and many rotamer and Ramachandran outliers (>3%). Structure validation was carried out with MolProbity.[28] The data collection and refinement statistics are summarised in Supplementary Table 1.

Native mass spectrometry

Prior to MS analysis, the protein sample was buffer exchanged into 200 mM ammonium acetate pH 8.0 and 0.05% (w/v) LDAO, using a Biospin-6 (BioRad) column and introduced directly into the mass spectrometer using gold-coated capillary needles (prepared in-house). Data were collected on a Q-Exactive UHMR mass spectrometer (ThermoFisher). The instrument parameters were as follows: capillary voltage 1.2 kV, quadrupole selection from 1,000 to 20,000 *m/z* range, S-lens RF 100%, collisional activation in the HCD cell 100 V, trapping gas pressure setting kept at 7.5, temperature 200 °C, resolution of the instrument 12500. The noise level was set at 3 rather than the default value of 4.64. No in-source dissociation was applied. Data were analysed using Xcalibur 4.2 (Thermo Scientific).

Proteomics

For protein identification, proteins were digested with both trypsin and chymotrypsin and the resultant peptides were loaded onto a reverse phase C18 trap column (Acclaim PepMap 100, 75 μm × 2 cm, nano viper, C18, 3 μm, 100 Å, ThermoFisher, Waltham, MA, U.S.A) using an Ultimate 3000 and washed with 50 μL of solvent A at 10 μl/min. The desalted peptides were then separated using a 15 cm pre-packed reverse phase analytical column (Acclaim PepMap 100, 75 μm × 15 cm, C18, 3 μm, 100 Å, ThermoFisher, Waltham, MA, U.S.A) using a 45 min linear gradient from 5% to 40% solvent C (80% acetonitrile, 20% water, 0.1% formic acid) at a flow rate of 300 nl/min. The separated peptides were electrosprayed into an Orbitrap Eclipse Tribrid mass spectrometry system in the positive ion mode using data-dependent acquisition with a 3 s cycle time. Precursors and products were detected in the Orbitrap analyzer at a resolving power of 60,000 and 30,000 (@ *m/z* 200), respectively. Precursor signals with an intensity > 1.0×10^{-4} and charge state between 2 and 7 were isolated with the quadrupole using a 0.7 *m/z* isolation window (0.5 *m/z* offset) and subjected to MS/MS fragmentation using higher-energy collision induced dissociation (30% relative fragmentation energy). MS/MS scans were collected at an AGC setting of 1.0×10^4 or a maximum fill time of 100 ms and precursors within 10 ppm were dynamically excluded for 30 s. Data were searched against the Oryza sativa subsp. japonica (Rice) proteome using ProteinProspector (v6.2.2) with the following search parameters: trypsin/chymotrypsin digestion; fixed modification was set to carbamidomethyl (C); variable modifications set as oxidation (M), acetylated protein N terminus, and phosphorylation (STY).

Molecular Dynamics Simulations

Systems were constructed by submitting the crystallographic structure of OsNIP2;1 to the CHARMM-GUI server[29,30] and using the Membrane Builder feature.[31] The aquaporin was inserted in a membrane composed of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) phospholipids and with explicit water solvation. A number of counterions were added to the simulation to neutralize charges, with an extra salt concentration of 0.15 M of potassium chloride ions for all simulations. Concentrations of 0.1, 0.5, and 1 M of silicic acid were tested, aiming to increase the number of translocation events observed (uptake/export). Silicic acid molecules were randomly inserted in the simulation box by making use of the *gmx insert-molecules* tool. Silicic acid molecule topology was constructed using pre-existing silicate parameters[32] from the CHARMM36m force field.[33] Atomic partial charges for silicic acid were obtained using quantum

mechanics (QM) calculations of Hirshfeld charges[34] in Gaussian16 software.[35] For this, an optimization of the system was performed at the MP2 level[36–40] with a 6-311G** basis set.[41,42]

Molecular dynamics simulations were performed using the GROMACS simulation suite (version 2020.4)[43] along with CHARMM36m force field[33,44] and TIP3P water model.[45] Initially, equilibration simulations were run employing NVT and NPT ensembles with position restraints in the protein and phosphate atoms of the phospholipids, lasting for 1 and 50 ns, respectively. Subsequentially, production simulations in NPT ensemble ran for either 500 ns or 1 microsecond. Simulations were performed at both 310 K and 330 K (to enhance the occurrence of silicic acid translocation events), and the velocity rescale thermostat[46] with a coupling constant of $\tau = 0.1$ was applied to keep a constant temperature. The Parrinello-Rahman barostat[47] with a time constant of 2 ps was used to maintain pressure semi-isotropically at 1 atm. Long-range electrostatics were treated by the particle mesh Ewald method.[48] Covalent bonds were constrained by the LINCS algorithm,[49,50] allowing an integration step of 2 fs. Values for long-range electrostatics and van der Waals cut-offs were set to 1.2 nm. Starting velocities were modified at the beginning of each different replicate to improve conformational sampling. Ten steered MD simulations were performed for 40 ns to estimate the force required for one silicic acid molecule to translocate through the channel. To achieve this, a harmonic spring constant with a force of 800 kJ mol^{-1} nm^{-2} was attached to the Si atom, which was then pulled through the z-axis at a constant velocity of 0.1 nm ns^{-1}. Different initial starting velocities were employed for each of the ten independent runs. Final forces were computed and an average force profile and standard deviation were obtained for the pulling coordinate.

Molecules were manipulated, visualized, and analysed utilizing VMD[51] and Pymol[52] software. The HOLE software[53] was utilized for the calculation of pore radius analyses. Distance and angle cut-offs to count hydrogen bonds between atoms were 3.5 Å and 20°, respectively. A translocation event was counted once a silicic acid molecule moved from one side of the simulation box, through the inside of the protein channel, and exited on the opposite side. Residence time analysis was performed by calculating the minimum distance between silicic acid and the atoms of each domain. The molecule was considered to be interacting with the domain if the distance was less than 4 Å. Once this distance was inside the cut-off, the number of subsequent frames that it remained below the cut-off was calculated and converted to the corresponding amount of time (in ns). The HOLE profiles from the open and closed states were obtained by dividing the trajectories into 0–200 and 300–500 ns portions. A total of 12 portions were used (6 for each

state) in the analysis, using data from different channels, followed by calculation of the average profile and standard deviation. For the closed state, portions were chosen if distance values between R189 and F121 were below 10 Å: chain B (OsNIP2,1$_{WATER1}$, chains A, B, and C (OsNIP2;1$_{WATER2}$), chain A (OsNIP2;1$_{WATER3}$). For the open state, portions were chosen when distance values were above 10 Å: chains A, B, and C (OsNIP2;1$_{WATER1}$), chain B (OsNIP2;1$_{WATER2}$), chain A and C (OsNIP2;1$_{WATER3}$).

CRediT authorship contribution statement

Bert van den Berg: Conceptualization, Investigation, Visualization, Writing– original draft, Writing – review & editing. **Conrado Pedebos:** Investigation, Visualization. **Jani R. Bolla:** Investigation, Visualization. **Carol V. Robinson:** Supervision. **Arnaud Baslé:** Resources. **Syma Khalid:** Supervision, Writing – review & editing.

Acknowledgements

Author contributions

MD simulations were performed by CP, supervised by SK. JRB performed the mass spectrometry experiments. AB collected X-ray crystallography data. BvdB conceived the project, cloned, expressed and purified OsNIP2;1, and determined the crystal structures. The manuscript was written by BvdB with input from SK and CP.

Competing interests

The authors declare no competing interests.

Data availability

The data supporting the findings of this study, including simulation snapshots and parameters

are available from the corresponding authors upon reasonable request. Coordinates and structure factors have been deposited in the Protein Data Bank with accession code 7NL4 [http://doi.org/10.2210/pdb7NL4/pdb].

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167226.

References

1. Epstein, E., (1999). Silicon. *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, **50**, 641–664.
2. Pommerrenig, B., Diehn, T.A., Bienert, G.P., (2015). Metalloido-porins: Essentiality of Nodulin 26-like intrinsic proteins in metalloid transport. *Plant Sci.*, **238**, 212–227.
3. Farooq, M.A., (2015). Dietz KJ (2015) Silicon as versatile player in plant and human biology: overlooked and poorly understood. *Front. Plant Sci.*, **6**, 994–1017.
4. Exley, C., Guerriero, G., Lopez, X., (2019). Silicic acid: The omniscient molecule. *Sci. Total Environ.*, **665**, 432–437.
5. King, L.S., Kozono, D., (2004). Agre P (2004) From structure to disease: the evolving tale of aquaporin biology. *Nature Rev. Mol. Cell Biol.*, **5**, 687–698.
6. Diehn, T.A., Bienert, M.D., Pommerrenig, L.Z., Spitzer, C., Bernhardt, N., Fuge, J., Bieber, A., Richet, N., et al., (2019). Boron demanding tissues of *Brassica napus* express specific sets of functional Nodulin26-like Intrinsic Proteins and BOR1 transporters. *Plant J.*, **100**, 68–82.
7. Beamer, Z.G., Routray, P., Choi, W.-G., Spangler, M.K., Lokdarshi, A., Roberts, D.M., (2021). Aquaporin family lactic acid channel NIP2;1 promotes plant survival under low oxygen stress in Arabidopsis. *Plant Physiol.*,. https://doi.org/10.1093/plphys/kiab196.
8. Ma, J.F., Tamai, K., Yamaji, N., Mitani, N., Konishi, S., Katsuhara, M., Ishiguro, M., Murata, Y., et al., (2006). A silicon transporter in rice. *Nature*, **440**, 688–691.
9. Mitani, N., Yamaji, N., Ma, J.F., (2008). Characterization of substrate specificity of a rice silicon transporter, Lsi1. *Pflugers Arch.*, **456**, 679–686.
10. Ma, J.F., Yamaji, N., Mitani, N., Tamai, K., Konishi, S., Fujiwara, T., Katsuhara, M., Yano, M., (2007). An efflux transporter of silicon in rice. *Nature*, **448**, 209–212.
11. Ma, J.F., Yamaji, N., (2008). Functions and transport of silicon in plants. *Cell. Mol. Life Sci.*, **65**, 3049–3057.
12. Pommerrenig, B., Diehn, T.A., Bernhardt, N., Bienert, M.D., Mitani-Ueno, N., Fuge, J., Bieber, A., Spitzer, C., et al., (2020). Functional evolution of nodulin 26-like intrinsic proteins: from bacterial arsenic detoxification to plant nutrient transport. *New Phytol.*, **225**, 1383–1396.
13. Ma, J.F., Yamaji, N., Mitani, N., Xu, X.Y., Su, Y.H., McGrath, S.P., Zhao, F.J., (2008). Transporters of arsenite in rice and their role in arsenic accumulation in rice grain. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 9931–9935.
14. Kreida, S., Törnroth-Horsefield, S., (2015). Structural insights into aquaporin selectivity and regulation. *Curr. Opin. Struct. Biol.*, **33**, 126–134.
15. Gotfryd, K., Mósca, A.F., Missel, J.W., Truelsen, S.F., Wang, K., Spulber, M., Krabbe, S., Hélix-Nielsen, C., et al., (2018). Human adipose glycerol flux is regulated by a pH gate in AQP10. *Nature Commun.*, **9**, 4749–4759.
16. Garneau, A.P., Carpentier, G.A., Marcoux, A.A., Frenette-Cotton, R., Simard, C.F., Rémus-Borel, W., Caron, L., Jacob-Wagner, M., et al., (2015). Isenring P Aquaporins Mediate Silicon Transport in Humans. *PLoS ONE*, **10**, e0136149
17. Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J., Stroud, R.M., (2004). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, **290**, 481–486.
18. Kirscht, A., Kaptan, S.S., Bienert, G.P., Chaumont, F., Nissen, P., de Groot, B.L., Kjellbom, B., Gourdon, P., et al., (2016). Crystal Structure of an Ammonia-Permeable Aquaporin. *PLoS Biol.*, **14**, 1002411.
19. Deshmukh, R.K., Vivancos, J., Ramakrishnan, G., Guérin, V., Carpentier, G., Sonah, H., Labbé, C., Isenring, P., et al., (2015). A precise spacing between the NPA domains of aquaporins is essential for silicon permeability in plants. *Plant J.*, **83**, 489–500.
20. Törnroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R., Kjellbom, P., (2006). Structural mechanism of plant aquaporin gating. *Nature*, **439**, 688–694.
21. de Maré, S.W., Venskutonytė, R., Eltschkner, S., de Groot, B.L., Lindkvist-Petersson, K., (2020). Structural Basis for Glycerol Efflux and Selectivity of Human Aquaporin 7. *Structure*, **28**, 215–222.
22. Hays, F.A., Roe-Zurz, Z., Stroud, R.M., (2010). Overexpression and purification of integral membrane proteins in yeast. *Methods Enzymol.*, **470**, 695–707.
23. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., Bricogne, G., (2011). Data processing and analysis with the autoPROC toolbox. *Acta Cryst.*, **D67**, 293–302.
24. Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., (2018). Bricogne G STARANISO. Global Phasing Ltd, Cambridge, United Kingdom.
25. McCoy, A.J. et al, (2007). Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
26. Adams, P.D. et al, (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
27. Emsley, P., Cowtan, K., (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

28. Chen, V.B. et al, (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

29. Jo, S., Kim, T., Iyer, V.G., Im, W., (2008). CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.*, **29**, 1859–1865.

30. Lee, J. et al, (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.*, **12**, 405–413.

31. Wu, E.L. et al, (2014). CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *J. Comput. Chem.*, **35**, 1997–2004.

32. Lopes, P.E.M., Murashov, V., Tazi, M., Demchuk, E., MacKerell Jr, A., (2006). Development of an Empirical Force Field for Silica. Application to the Quartz-Water Interface. *J. Phys. Chem. B*, **110**, 2782–2792.

33. Huang, J. et al, (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, **14**, 71–73.

34. Hirshfeld, F., (1977). Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta*, **44**, 129–138.

35. Frisch, M.J. et al, (2016). Gaussian 16, Revision C.01. Gaussian, Inc., Wallingford CT.

36. Head-Gordon, M., Pople, J.A., Frisch, M.J., (1988). MP2 energy evaluation by direct methods. *Chem. Phys. Lett.*, **153**, 503–506.

37. Saebø, S., Almlöf, J., (1989). Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chem. Phys. Lett.*, **154**, 83–89.

38. Frisch, M.J., Head-Gordon, M., Pople, J.A., (1990). Direct MP2 gradient method. *Chem. Phys. Lett.*, **166**, 275–280.

39. Frisch, M.J., Head-Gordon, M., Pople, J.A., (1990). Semi-direct algorithms for the MP2 energy and gradient. *Chem. Phys. Lett.*, **166**, 281–289.

40. Head-Gordon, M., Head-Gordon, T., (1994). Analytic MP2 Frequencies Without Fifth Order Storage: Theory and Application to Bifurcated Hydrogen Bonds in the Water Hexamer. *Chem. Phys. Lett.*, **220**, 122–128.

41. McLean, A.D., Chandler, G.S., (1980). Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11-18. *J. Chem. Phys.*, **72**, 5639–5648.

42. Francl, M.M. et al, (1982). Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.*, **77**, 3654–3665.

43. Abraham, M.J. et al, (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, 19–25.

44. Klauda, J.B. et al, (2010). Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B*, **114**, 7830–7843.

45. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

46. Bussi, G., Donadio, D., Parrinello, M., (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101

47. Parrinello, M., Rahman, A., (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182.

48. Darden, T., York, D., Pedersen, L., (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

49. Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., (1997). LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, **18**, 1463–1472.

50. Hess, B., (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.*, **4**, 116–122.

51. Humphrey, W., Dalke, A., (1996). VMD: visual molecular dynamics. *J. Mol. Graph.*, **15**, 33–38.

52. The PyMOL Molecular Graphics System, Version 2.3 Schrödinger, LLC.

53. Smart, O.S., Neduvelil, J.G., Wang, X., Wallace, B.A., Sansom, M.S.P., (1996). HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.*, **14**, 354–360.