BBA - Biomembranes 1865 (2023) 184056

Contents lists available at ScienceDirect

BBA - Biomembranes

journal homepage: www.elsevier.com/locate/bbamem

Computational modelling of diatom silicic acid transporters predicts a conserved fold with implications for their function and evolution

Michael J. Knight [a], Benjamin J. Hardy [a], Glen L. Wheeler [b], Paul Curnow [a,*]

[a] School of Biochemistry, University of Bristol, UK
[b] Marine Biological Association of the UK, UK

A B S T R A C T

Diatoms are an important group of algae that can produce intricate silicified cell walls (frustules). The complex process of silicification involves a set of enigmatic integral membrane proteins that are thought to actively transport the soluble precursor of biosilica, dissolved silicic acid. Full-length silicic acid transporters are found widely across the diatoms while homologous shorter proteins have now been identified in a range of other organisms. It has been suggested that modern silicic acid transporters arose from the union of such partial sequences. Here, we present a computational study of the silicic acid transporters and related transporter-like sequences to help understand the structure, function and evolution of this class of membrane protein. The AlphaFold software predicts that all of the protein sequences studied here share a common fold in the membrane domain which is entirely different from the predicted folds of non-homologous silicic acid transporters from plants. Substrate docking reveals how conserved polar residues could interact with silicic acid at a central solvent-accessible binding site, consistent with an alternating access mechanism of transport. The structural conservation between these proteins supports a model where modern silicon transporters evolved from smaller ancestral proteins by gene fusion.

1. Introduction

Silica is an important biomaterial found in diverse organisms including plants, sponges, and diatoms. The process by which this biomineral forms within the living cell – known generally as biosilicification – has not been precisely defined at the molecular level, and remains of interest as a remarkable aspect of cellular biology and as the inspiration for novel man-made materials [1–3].

Diatoms are unicellular algae that play an important role in global geochemical cycles [4,5]. Diatoms are generally distinguished by an outer cell wall, or frustule, that is almost entirely formed of silica with a minor organic component [6]. Frustule mineralization is an obligatory part of the lifecycle of many (but not all) of the thousands of extant diatom species [7]. A cadre of biomolecules have now been identified that seem to be dedicated to frustule formation [4,8–17], and understanding the molecular basis of silicification remains an active area of study [18].

One enigmatic family of diatom proteins are the silicic acid transporters (SITs) that reside within the plasma membrane underneath the silicified frustule. These integral membrane proteins are thought to act as permeases which carry out the active transport of the dissolved and bioavailable form of silica, monomeric silicic acid ($Si(OH)_4$). Active transport of silicic acid was proposed decades ago because Si-starved diatoms show saturable, hyperbolic uptake kinetics suggestive of a protein carrier (reviewed in [7,19]). The silicon requirements of the diatom cell, and the concentrative uptake of silicic acid against a concentration gradient that precedes frustule synthesis, are also consistent with an active transport mechanism [20] although this has been debated [21,22]. Multiple studies attest that the SITs are intimately linked to silicon metabolism in the diatom [23–28], and that they appear to transport silicic acid in model systems [15,29,30]. However, further detailed characterisation of these proteins has proven difficult. The SITs are poorly-expressed in recombinant cells and, when they can be produced, are incompatible with many of the detergent-based and detergent-free methods that have been developed for studying membrane proteins [30].

SITs have subsequently been found in a range of other silicifying eukaryotes - most notably the choanoflagellates and the silicifying haptophyte, *Prymnesium neolepis* [31] - suggesting that they play a common role in Si transport in these organisms. All SITs possess a similar organisation of two pseudo-repeat domains of five predicted transmembrane helices (5TM), wherein each domain contains a pair of the highly conserved GXQ motifs [32]. An additional family of proteins has been identified that exhibit similarity to the individual 5TM domains of the SITs. These SIT-like proteins (SITLs) are found in a range of eukaryote lineages including foraminifera, dinoflagellates, radiolarians

* Corresponding author.
E-mail address: p.curnow@bristol.ac.uk (P. Curnow).

https://doi.org/10.1016/j.bbamem.2022.184056
Received 12 July 2022; Received in revised form 16 September 2022; Accepted 20 September 2022
Available online 30 September 2022

and some metazoans. Although many of these organisms are not extensively silicified, they are known to produce some silicified structures or to utilise Si as part of other (non-siliceous) biomineralization processes [31,32]. The SITLs have not yet been characterised experimentally, but their similarity to the SITs suggests that SITs may have been formed by duplication and fusion of an ancestral SITL protein. Many other membrane transporters exhibit pseudo-repeat domains, suggesting that duplication and fusion is a common process in their evolution, although there are relatively few examples where the evolved and ancestral forms are both found in eukaryotes [33].

A high-resolution three-dimensional structure of the SITs would provide important insights into their biological function. For example, few other proteins are known that specifically engage with monomeric silicic acid via precise protein-substrate interactions, with perhaps the only clear example being the silicic acid channel from plants [34]. How then might the SITs capture and release this substrate? The SITs appear to be coupled to the electrochemical sodium gradient [15,30,35], but how is this bioenergetic coupling achieved? And what is the role of the GXQ sequence motifs that are absolutely conserved across all known SITs [32,36]? A further question arises over the SITLs. If such 'half-SITs' were the evolutionary predecessors of the modern SITs, what is the exact structural relationship between the SITLs and the SITs? Unfortunately experimental structures that could address these questions have not been forthcoming and, to our best knowledge, are probably not imminent given the difficulties in preparing the SITs for biophysical studies.

The AlphaFold software has recently emerged as a transformative new technology in molecular modelling. AlphaFold uses machine learning to produce computational models of protein structures that are in unprecedented agreement with experimental structures [37]. This software has now been applied in a high-throughput automated pipeline to predict the structures of multiple proteomes, and these are publicly available in an online database (https://alphafold.ebi.ac.uk). These predicted structures include silicic acid transport proteins that are unrelated to the diatom SITs and have different modes of action, namely the plant aquaporin-like Lsi1 and plant efflux transporters Lsi2 and Lsi3. However, at the time of writing the AlphaFold database does not include any diatom SITs.

Here, we use AlphaFold to predict the structure of several full-length diatom SITs as well as homologous SITLs from eukaryotic and prokaryotic sources. The results suggest a strong conservation of a novel fold within the transmembrane domain of the SITs. Using these models for computational substrate docking reveals a plausible binding mode for silicic acid where it is coordinated by highly conserved polar residues within a solvent accessible cleft, consistent with the activity of a membrane permease. AlphaFold also predicts that the same SIT transmembrane fold can be formed by homodimerisation of SITL proteins, supporting the idea that modern SITs evolved from gene duplication and fusion of SITLs [32]. We hope that making these models available to the community will stimulate further studies of these proteins.

2. Materials and methods

Sequences for full-length diatom silicon transporters and SITL sequences were submitted to either of two online Google Colaboratory notebook implementations of AlphaFold 2.1.0 [37,38] available at (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb) or (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb); the results were indistinguishable whichever implementation was used. Either the monomer or multimer models were used with relaxation, prokaryotic sequences were designated as such where this option was available, and no templates were specified. Molecular docking was performed with MOE, specifying the conserved glutamines as being involved in substrate binding. Initial placement of silicic acid was performed using the Triangle Matcher algorithm and refined poses were scored using the GBVI/WSA model. Protein placement into a DOPC bilayer potential was performed with the PPM server [39] (https://opm.phar.umich.edu/ppm_server). Coiled-coil interactions were analysed with Socket 2 [40] (http://coiledcoils.chm.bris.ac.uk/socket2/home.html). Symmetry calculations were performed on the membrane domain of PtSIT1 with *SymD* [41] implemented at the *icn3D* server (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html). PyMOL [42] was used to display model structures and for between-model RMSD calculations. Colouring models by pLDDT were performed with the PyMOL extension *coloraf* (https://github.com/cbalbin-bio/pymol-color-alphafold) by S. Bliven and C. Balbin.

The following protein sequences were used for modelling: Full-length diatom SITs from *Cylindrotheca fusiformis* (CfSIT1, AAD13804.1), *Nitzschia alba* (NaSIT, ABB81817.1), *Phaeodactylum tricornutum* (PtSIT1, GenBank ID: XP_002183269.1), *Skeletonema costatum* (ScSIT, ABB81824.1), *Thalassiosira oceanica* (ToSIT, EJK72537.1), *Thalassiosira pseudonana* SIT 1 (TpSIT1, ABB81825.1), *Thalassiosira pseudonana* SIT 3 (TpSIT3, ABB81827.1). SITL proteins from coccolithophores *Coccolithus braarudii* (CbSITL) and *Calcidiscus leptoporus* (CleL), annelid worm *Capitella teleta* (CteL, ELT99670.1), Dictyochophyceae *Florenciella parvula* (FpaL, CAMPEP_0119469700), copepod *Calanus finmarchicus* (CfiL, GAXK01179749.1), bacterium *Rhodococcus opacus* (RopL, WP_043827568.1) and cyanobacterium *Synechococcus* sp. KORDI-100 (SkoL, WP_038543010.1). All sequences and model PDB files are provided as Supplementary data.

Systems containing PtSIT1 and the CbSITL dimer inserted into 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayers were assembled using the Membrane Builder application [43] of CHARMM-GUI [44,45]. Systems were built with explicit water solvation and with an approximate physiological salt concentration of 150 mM sodium chloride, with sufficient counterions to neutralise charges. Where required, silicic acid was inserted at 10 mM with the *gmx insert-molecules* GROMACS application using a water-replacement method. The molecular topology and partial charges for silicic acid were those developed by van der Berg et al. [34]. Systems were parameterised using GROMACS 2020.2 [46], and topology and parameter files converted to Amber format using *ParmEd*. Molecular dynamics simulations were performed using the Amber20 simulation suite [47] on the BlueCrystal (Phase 4) University of Bristol high-performance computer, using the CHARMM36m forcefield [48] with TIP3P water [49]. Water molecules were kept rigid using SETTLE [50], and hydrogens constrained using SHAKE. Systems were first minimised by 5000 cycles of steepest-descent minimisation with restraints applied to non-hydrogen protein atoms, followed by 5000 steps with restraints to protein C-alpha atoms, and 5000 steps with no restraints. PtSIT1 systems were heated to 303.15 K using Langevin dynamics over 100 ps with a friction constant of 1 ps^{-1} and integration step of 2 fs. CbSITL systems were heated first with a shortened integration step of 0.5 fs over 12.5 ps, followed by 50 ns with a 1 fs integration time. Positional restraints were applied to protein non-hydrogen atoms for all heating steps. The Berendsen barostat [51] was then applied for 500 ps to maintain semi-isotropic pressure at 1 Bar with a pressure relaxation time of 2 ps, with positional restraints applied to protein non-hydrogen atoms. Systems were equilibrated for a further 1 ns at constant temperature and pressure with positional restraints applied to protein C-alpha atoms only, utilising GPU acceleration via the CUDA implementation of *pmemd*. For simulations of PtSIT1 with docked silicic acid, positional restraints were applied to ligand silicon and oxygen atoms during the first two rounds of minimisation and all equilibration steps. Production simulations were run for 500 ns (25,000 frames) in triplicate, giving a total simulation time of 1.5 μs per system, and visualised in VMD [52]. Backbone RMSF and RMSD were determined with CPPTRAJ [53].

3. Results

3.1. A conserved fold for diatom SITs

The full-length amino acid sequences of seven diatom silicon transporters were submitted to AlphaFold for computational modelling. An overlay of the top-ranked models for each protein is shown in Fig. 1. All models converged upon the same basic fold in the probable membrane domain, with local Cα RMSD ≤ 2.5 Å over at least 375 atoms against CfSIT1. The AlphaFold confidence in predicting this membrane fold was considered 'very high' (pLDDT > 90), with lower confidence in flanking soluble regions. This was confirmed by repeat modelling runs of some individual proteins which gave consistent results for the high-confidence membrane regions, but showed variability in the structure of the low-confidence soluble domains. We thus focus our further analysis below on the SIT membrane domain.

One of the diatom SITs, PtSIT1, was taken as a representative model for closer inspection, but the observations below can be applied to all of the models. Fig. 2 shows the insertion of PtSIT1 into a lipid bilayer using the PPM server [39]. This was in excellent agreement with bioinformatic analysis of the primary sequence and prior experimental data [29,30,36]; all of the sections of the protein expected to form membrane-embedded alpha-helices are found as such within the AlphaFold model. The overall fold of the protein comprised a central core of six transmembrane helices (TM 2–4 and 7–9). A DALI search [54] against the AlphaFold database and the PDB suggested that the core bundle could have some structural homology to members of the Major Facilitator Superfamily of solute transporters. TM helices 1 and 6 sit at the outside of the core bundle, lying at a slight angle akin to a 'sash' across the main body of the protein. Helices 5 and 10 do not stretch across the membrane, but instead lie laterally in the membrane plane as 'belts' that surround the core bundle at the inner edge of the cytoplasmic leaflet. They are unwound in the middle of each helical segment. The presence of such lateral helices has been previously observed in the experimental structures of some other permeases, for example the magnesium transporter CorB (PDB ID: 7M1T) and the betaine transporter BetP (4DOJ). The membrane-embedded portion of the model was determined to have a C2 symmetry axis when viewed from the top or bottom of the membrane (*SymD* Z-score 28.5). It thus appears that the SITs adopt a well-conserved and novel membrane fold, which is very different from the experimental structures and AlphaFold predictions for non-homologous silicon transporters from plants ([34] and https://alphafold.ebi.ac.uk).

3.2. Substrate docking simulations

The AlphaFold model of PtSIT1 was used for docking simulations of the presumed natural ligand, $Si(OH)_4$. The top-scoring pose from this simulation is shown in Fig. 3. Silicic acid appears to access a central binding site in the protein core via a narrow solvent-exposed cavity that that is only accessible at the outer side of the membrane. Such a binding site would be consistent with a rocker-switch alternating access mechanism, whereby global protein motions alternately expose this central substrate-binding site to either face of the membrane [55,56]. The structural model here appears to represent PtSIT1 in an outward-facing conformation. At the base of this solvent-accessible cleft lie a series of polar amino acid residues, including all of the glutamines from the four characteristic GXQ motifs and Q153 from the XQXXXQX motif in TM 4

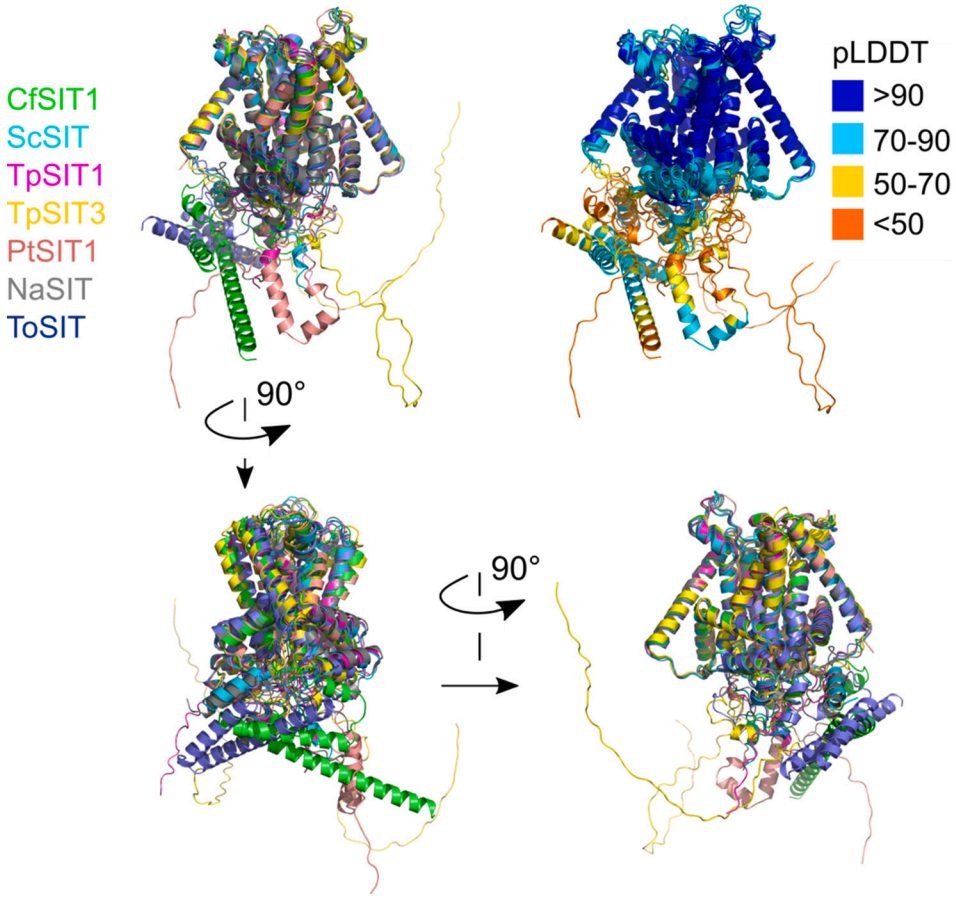

Fig. 1. AlphaFold models of diatom SITs. Superposition of seven diatom SIT structures predicted by AlphaFold, using PyMOL command *extra_fit* against template CfSIT1. The fold in the membrane domain, which is predicted with high confidence, closely overlaps between all models (Cα RMSD < 2.5 Å). The soluble domains are more variable with lower model confidence. Models are coloured according to the legend.

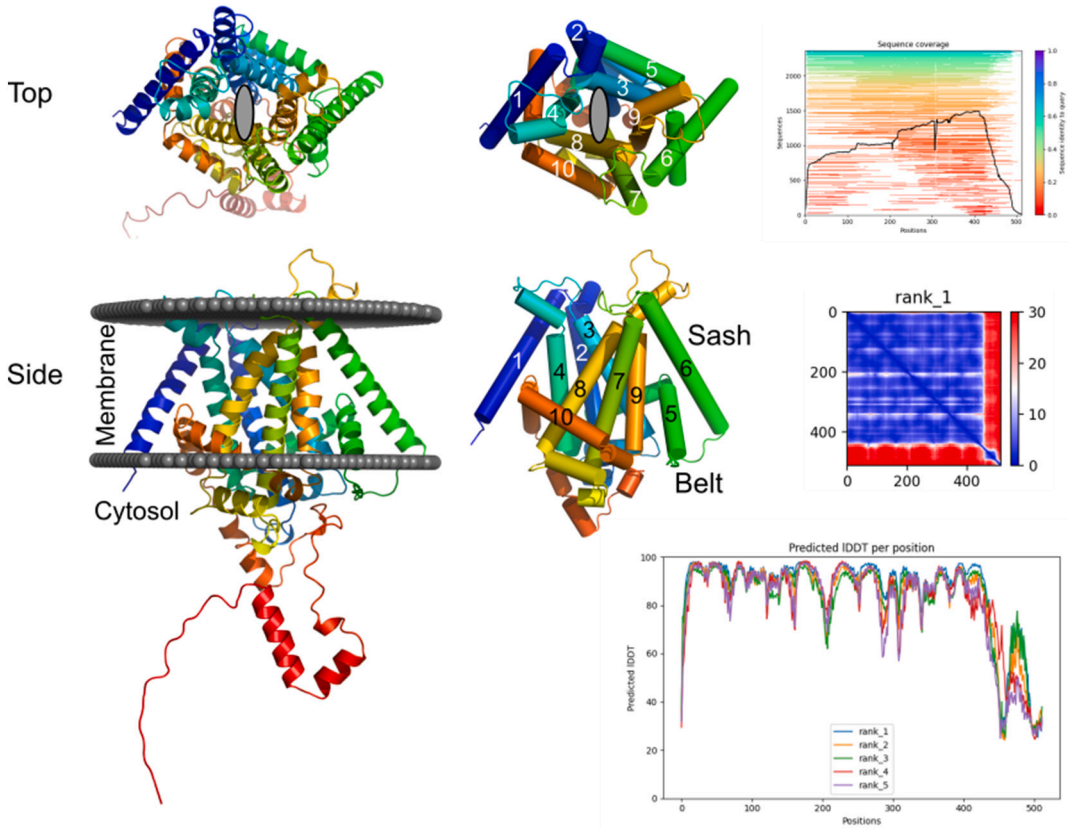

Fig. 2. Modelled structure of PtSIT1. The peptide chain of PtSIT1 is coloured as a rainbow from blue (N-terminus) to red (C-terminus), with membrane helices numbered from 1 to 10 according to their sequence order. The low-confidence C-terminal region is removed from the cylinder representation for display purposes. Two-fold rotational symmetry (C2) within the structure is apparent from the *top view*. A *side view* shows the position of the transmembrane domain within the boundaries of a bilayer membrane (grey spheres) by the PPM server. $\Delta G_{transfer}$ −82.9 kcal/mol, hydrophobic thickness 31.2 ± 0.9 Å, tilt angle 6 ± 1 Å. The lateral 'belt' helices lying parallel to the membrane can be seen in dark orange and dark green with the transmembrane 'sash' helices in blue and light green. The associated AlphaFold output data for this model are shown on the right.

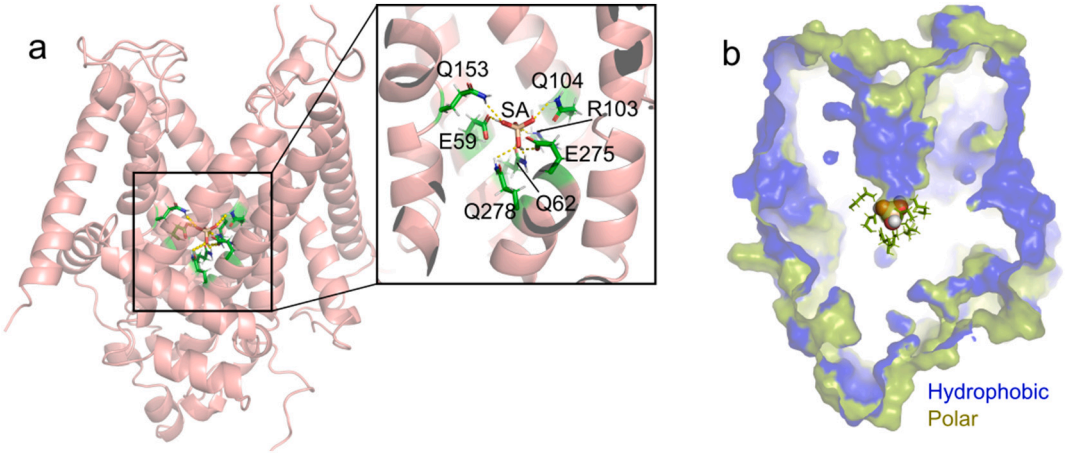

Fig. 3. Silicic acid docking to PtSIT1. (a) Monomeric silicic acid (SA) interacts with multiple polar sidechains in a narrow solvent-accessible cavity at the protein centre. Hydrogen bond interactions are shown as dashed yellow lines in the inset, with the protein clipped at the front of the image for presentation. Q62, Q104 and Q278 occur in the sequence context of highly conserved GXQ motifs. (b) Surface representation of the same image, showing the cluster of polar residues (as sticks) surrounding SA (as spheres) at the base of the substrate pocket. The image is clipped from the front to show a slice through the protein. Hydrophobic and polar surface residues are shown as blue and green, respectively.

[32,36]. These are oriented so that the Q sidechains face into the putative binding site and can engage in hydrogen bonding with silicic acid. Other conserved residues, with both negative and positive charge, were also able to take part in hydrogen bonding interactions with the substrate (E59, R103, E275). Other ligand poses, which are provided as Supplementary information, were very similar but did show some minor differences in ligand orientation. This suggests that $Si(OH)_4$ does not have a single dominant binding mode but instead can accept and donate adventitious hydrogen bonds with multiple polar groups within the binding site. This same solvent-accessible binding site was observed in

several other of the different SIT protein models, but not in all of them; some models were redolent of a more 'occluded' state where the opening to the cavity was closed off.

3.3. Structural homology between SITS and CbSITL

AlphaFold was also used to predict the structure of several SITL proteins. This group of proteins have previously been suggested to be 'half-SITs', featuring just five transmembrane helices [31,57]. We began by submitting the sequence of the SIT-like protein known as CbSITL from the coccolithophore *C. braarudii* (formerly *C. pelagicus*) [31] to AlphaFold. This was used for two independent simulation runs in either monomer or multimer mode.

Remarkably, the structure of monomeric CbSITL was predicted to be very similar to part of the full-length diatom SITs (Fig. 4a). When run in multimer mode, AlphaFold predicted that two monomers would assemble into a domain-swapped dimer that essentially reconstituted the membrane fold predicted for the diatom SITs (Fig. 4b,c). The two cytoplasmic helices of CbSITL packed against each other via knobs-into-holes interactions that could support dimerisation, but the homodimer assembly is more likely driven by the close shape complementarity (the accurate fitting-together) of the surfaces of the two protomers. In support of this, the soluble coiled-coil section did not persist in molecular dynamics simulations (see below) whereas the membrane assembly was essentially unchanged. In the dimer conformation of CbSITL the conserved GXQ motifs were in similar positions as for the full-length diatom SITs, pointing into the centre of the bundle.

3.4. Other SITLs share the SIT fold

Six other SITL homologues from bacteria, annelid worm and copepod were also modelled by AlphaFold and the resulting models are shown in Fig. 5. All models retained the same SIT-like membrane fold as CbSITL (RMSD < 1.5 Å) but showed substantial variability in their cytoplasmic domains. The only exception to this structural conservation was the protein from *Synechococcus* which was missing the N-terminal 'sash' helix.

3.5. Molecular dynamics of PtSIT1

The AlphaFold models of PtSIT1 and CbSITL were used for unbiased molecular dynamics simulations in POPC bilayers (Fig. 6 and Supplementary information). The structural features of the transmembrane region persisted in every simulation, rapidly reaching a stable RMSD < 3 Å from the starting state and with low RMSF for the transmembrane helices. The soluble domains were more dynamic as expected. In the case of PtSIT1 we performed six independent simulation runs of 500 ns (3 μs total) of the docked structure shown in Fig. 3. In three cases no additional silicic acid was included in the simulation, while in the other three runs silicic acid was introduced to the solvent at 10 mM; this is well above normal environmental concentrations of silicic acid, which are expected to be in the micromolar range, but probably close to intracellular concentrations [58] and represents only 5 molecules of silicic acid per simulation. In three of these six simulations the substrate remained present at the binding site throughout, albeit with some dynamical motions and rearrangements of the precise silanol-sidechain hydrogen bonding. In the other three simulations silicic acid was displaced into the solvent within 100 ns. There was no obvious difference between simulations with or without silicic acid. Once fully displaced into bulk solvent silicic acid continuously sampled the outer regions of the substrate cleft but did not re-bind. This is consistent with a dissociable protein-ligand interaction of moderate affinity, characteristic of a solute transporter. Water was able to enter the substrate binding site in these simulations, and the cavity became slightly extended further into the protein but did not form a continuous channel (Fig. 6c). No structural transitions were observed that could be interpreted as a significant part of the transport cycle, but such major changes were not expected to be observed on the timescale of these unguided simulations [59]. Silicic acid did not enter the hydrophobic bilayer core or spontaneously traverse the lipid bilayer, but instead was primarily

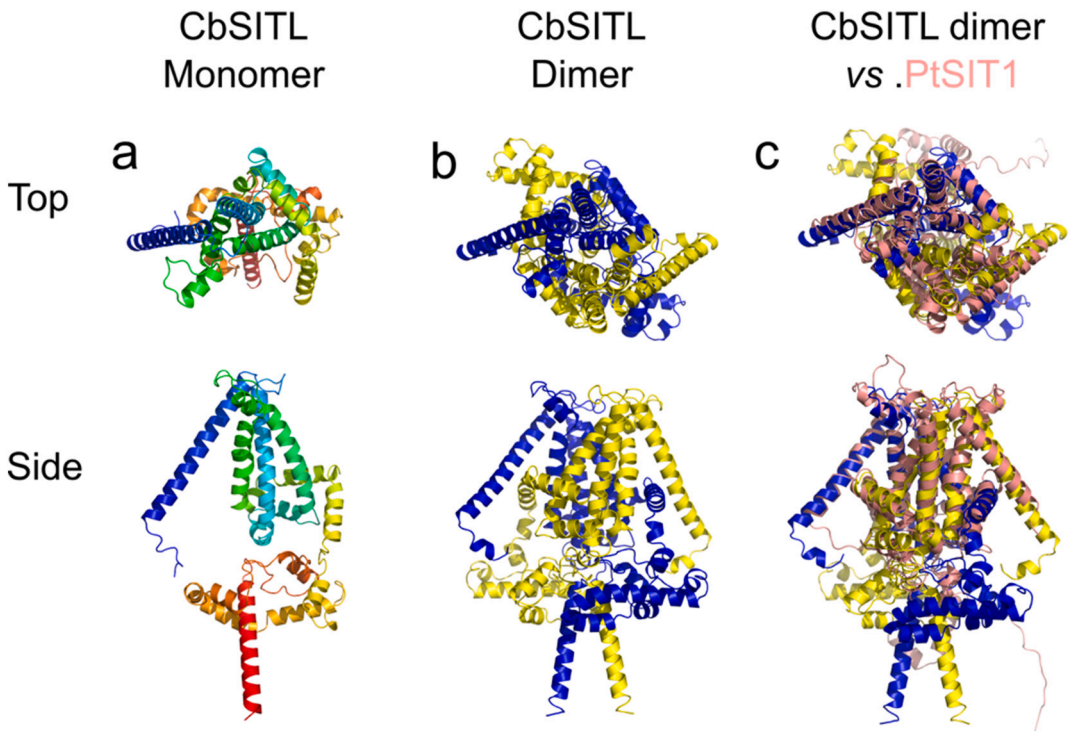

Fig. 4. AlphaFold prediction of the SIT-like protein from *C. braarudii* (CbSITL). (a) CbSITL monomer in rainbow colour scheme where the N-terminal is blue. (b) CbSITL can be successfully modelled ab initio as a strand-swapped dimer; one monomer is coloured yellow, the other blue. (c) The CbSITL dimer shares the core SIT fold, illustrated by overlay with the representative diatom protein PtSIT1 (salmon). RMSD = 3.5 Å over 2041 atoms within the membrane domain.

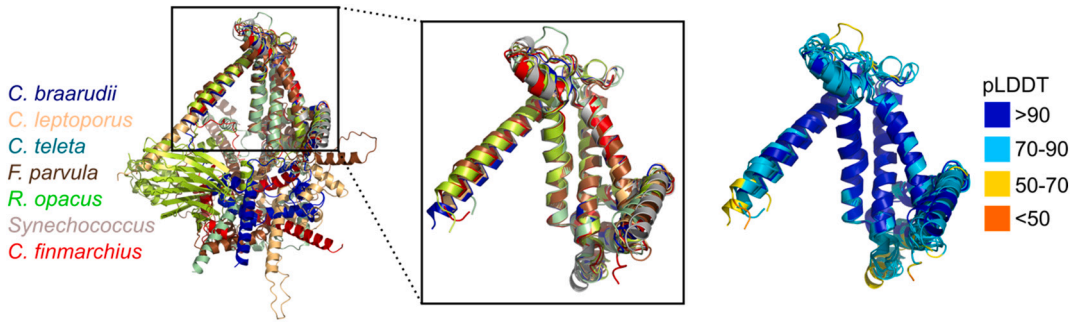

Fig. 5. The core SIT fold is retained among SIT-like proteins. Homologous SITL sequences were submitted to AlphaFold independently. The resulting models show the same membrane fold as CbSITL but have considerable diversity in their soluble domains. The overlay of the conserved membrane fold is displayed on the right, with AlphaFold confidence shown. All folds have RMSD < 1.5 Å against CbSITL over at least 1350 atoms. Models are coloured according to the legend.

associated with the lipid headgroup region.

Comparable simulations of the CbSITL dimer provided similar results (Fig. 6). There was limited fluctuation in the transmembrane domain during the timecourse of the simulation and no major structural transitions were observed. However, a major difference to PtSIT1 was that the central core of CbSITL was not accessible to water during simulation (Fig. 6c, d); the structure of CbSITL thus appears to be in a closed or occluded conformation on the simulation timescale. In parts of the simulation sodium could be seen to interact with, or 'plug', the top of the CbSITL bundle. The implications of this are unclear. As discussed above, the cytoplasmic coiled-coil predicted in the original CbSITL model did not persist during these simulations.

Independent simulations repeatedly showed occasional and transient sodium occupation of a surface region at the cytoplasmic-facing base of the 'sash' TM6 and the 'belt' TM7 in both PtSIT1 and the equivalent region of dimeric CbSITL. These spontaneous interactions typically persisted for about 3–10 ns of simulation time and could have some relevance to sodium cotransport.

4. Discussion

This study reports the predicted three-dimensional structures of a family of membrane proteins that remain largely resistant to detailed experimental analysis. Clearly, there is a degree of uncertainty around such models and their precise details ought to be interpreted with caution until they can be confirmed. Nonetheless, they may be useful for the community in the continued absence of experimental data and do account for several prior observations on the SITs.

The computational models are in close agreement with bioinformatic analyses [32,36] and low-resolution experimental data [29,30] which consistently suggest that the diatom SITs comprise bundles of 10 transmembrane alpha-helices. The AlphaFold models feature 8 authentic transmembrane helices and two 'belt' helices that do not cross the membrane, but instead lie partially submerged in the headgroup region parallel to the membrane plane. Another previous observation is the absolute sequence conservation of GXQ and XQXXXQX motifs in SITs and SITLs [32,36]; the persistence of these motifs through evolution implies that they play an important role in the SITs and SITLs, and mutations to individual GXQ motifs were shown to diminish silicic acid binding and transport in vitro [30]. In the model structures several of these conserved glutamines (as well as other conserved residues) line a solvent-accessible pocket which could plausibly coordinate silicic acid via side-chain hydrogen bonding.

The binding mode discussed here suggests a redundancy in protein-substrate interactions with implications for substrate selectivity. Presumably almost any small, polar compound could enter the substrate cleft and be sequestered by sidechain hydrogen bonding. This would be consistent with experimental data showing that soluble germanic acid ($Ge(OH)_4$) is a competitive inhibitor of silicic acid transport, a phenomenon that has been exploited for studies of Si transport in diatoms (e.g. [19,60]) and plants [61]. Perhaps the relatively high abundance of, and low competition for, silicic acid means that greater selectivity is not required, and the size of the substrate cleft alone provides sufficient discrimination. The redundancy in protein-substrate interactions implied by the models does seem at odds with the absolute conservation of GXQ sequence motifs and other hydrogen-bonding residues [32,36]. If at least seven residues can engage with the substrate (Fig. 3) then one might expect a modest effect, if any, when substituting any single one of them. Yet there is clearly strong selective pressure to preserve these interactions, and mutating just one GXQ motif appears to have a substantial impact on protein function [30]. Further work will be required to resolve this paradox.

A recent study reported the crystal structure and molecular dynamics simulation of the aquaporin-like silicic acid channel from rice [34]. Although this gated channel facilitates diffusion, rather than the secondary-active transport associated with the SITs, there are some parallels with the current study. The rice silicic acid channel features multiple hydrogen-bonding interactions between silanol hydroxyls and amino acid side-chains, which are especially prevalent at the selectivity filter region. Very few instances of silicic acid passage through the open channel were observed in unbiased simulations, even at 1 M silicic acid, and protein-silanol hydrogen bonding appears to be at least part of the channel gating mechanism. This supports the general notion that specific protein-silanol hydrogen bonds, as observed in our simulations here, can be a key factor in silicic acid transport.

AlphaFold predicts a striking structural similarity between SITs and SITLs. Specifically, it appears that SITL dimers closely resemble the full SIT membrane fold. This supports the notion of SITLs as evolutionary precursors that could be thought of as 'half-SITs'. It might be that the full SIT fold, arising from duplication and fusion of genes encoding these two half-domains [32], is required for efficient transport or electrochemical coupling as has been suggested for other membrane transporters [33]. Modern transport proteins can feature either topological inversion of these respective half-domains, or can have constituent domains that retain the same transmembrane orientation. The latter case is predicted here for the SITs.

5. Conclusions

This study uses computational tools to investigate membrane proteins thought to act as diatom silicic acid transporters, as well as their sequence homologues from other organisms. To our best knowledge the AlphaFold models provided here mark the first detailed insight into the possible three-dimensional structure of the SIT protein family. The models predict that membrane-embedded regions of the SITs and related SITLs have close structural similarity, implying a universal transmembrane fold that is shared by the wider grouping of SIT/SITL proteins. The models are further used to investigate a potential binding

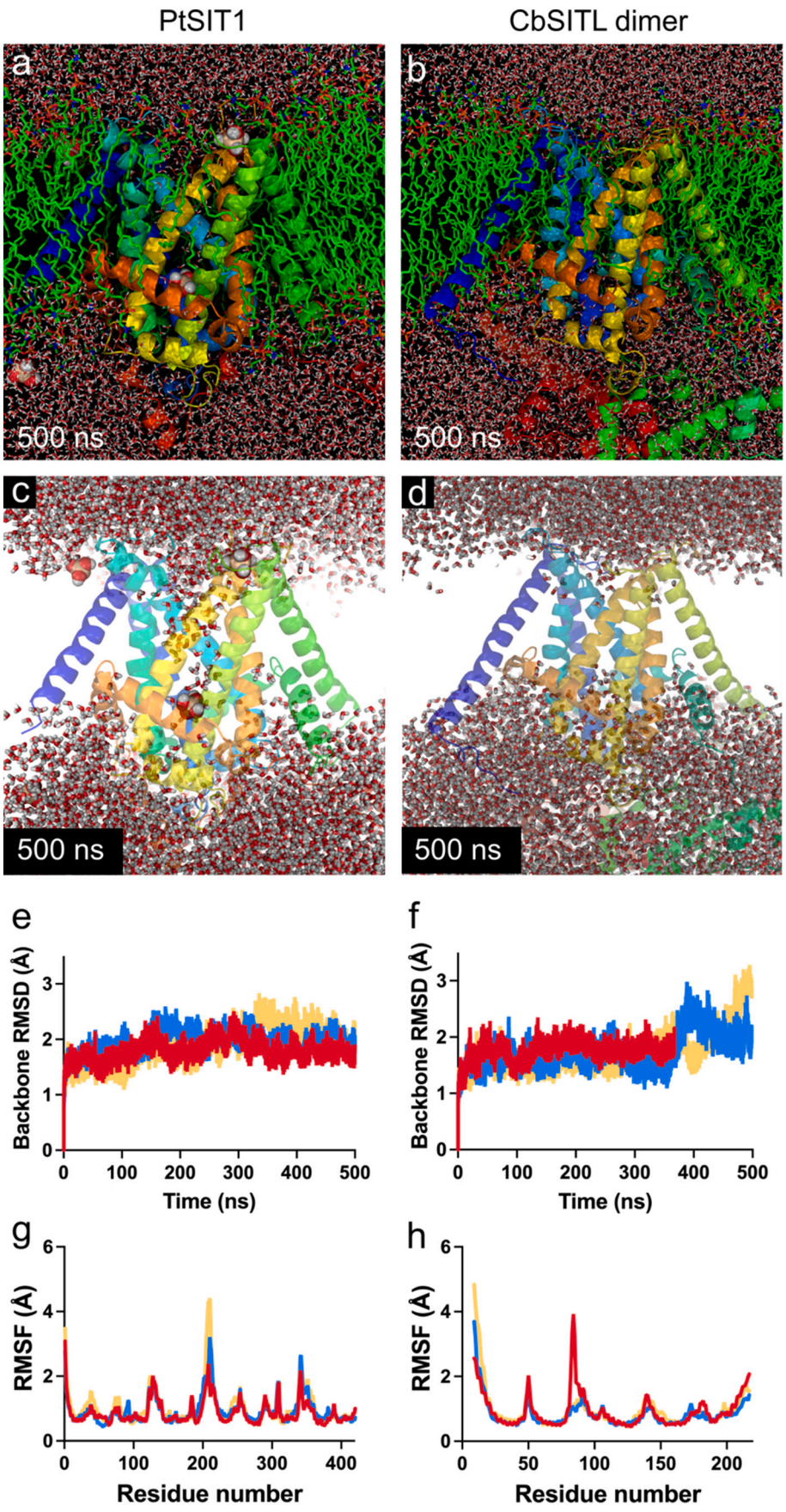

Fig. 6. Molecular dynamics simulations. (a) PtSIT1 and (b) CbSITL dimer after 500 ns in a POPC lipid bilayer. Movies for the entire trajectory and associated structure files are provided as Supplementary material. (c) The apparent substrate binding site of PtSIT1 is accessible to water, but the protein never forms a continuous water channel. (d) The CbSITL dimer remains closed to solvent. (e–f) Trajectory RMSD and (g–h) trajectory RMSF plots demonstrate the relative stability of the models during simulation. For both proteins the C-terminal regions are far more dynamic, as expected, and are not shown for presentation purposes. One monomer of the CbSITL dimer is analysed.

mode of silicic acid, namely sidechain hydrogen bonding in a solvent-accessible active site. The sidechains involved in these interactions include glutamine residues that occur in the context of highly-conserved sequence motifs, and this apparent functional role could account for their conservation. We hope that this work will encourage further research into these unusual proteins, including experimental testing of the ideas proposed here.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbamem.2022.184056.

CRediT authorship contribution statement

MJK, BJH, GW and PC conceived experiments. MJK, BJH and PC designed and performed experiments. MJK, BJH, GW and PC interpreted experimental data. PC wrote the manuscript with input from all authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

BJH is supported by a studentship from the EPSRC/BBSRC Syn-BioCDT (EP/L016494/1).

References

[1] N. Kröger, Prescribing diatom morphology: toward genetic engineering of biological nanomaterials, Curr. Opin. Chem. Biol. 11 (2007) 662–669.
[2] N. Kröger, N. Poulsen, Diatoms: from cell wall biogenesis to nanotechnology, Annu. Rev. Genet. 42 (2008) 83–107.
[3] A. Reid, F. Buchanan, M. Julius, P.J. Walsh, A review on diatom biosilicification and their adaptive ability to uptake other metals into their frustules for potential application in bone repair, J. Mater. Chem. B 9 (6728–37) (2021).
[4] B. Tesson, M. Hildebrand, Extensive and intimate association of the cytoskeleton with forming silica in diatoms: control over patterning on the meso- and micro-scale, PloS One 5 (12) (2010), e14300.
[5] E.V. Armbrust, The life of diatoms in the world's oceans, Nature 459 (2009) 185–192.
[6] F.E. Round, R.M. Crawford, D.G. Mann, The Diatoms: Biology And Morphology of the Genera, Cambridge University Press, 1990.
[7] M. Hildebrand, Diatoms, biomineralization processes, and genomics, Chem. Rev. 108 (2008) 4855–4874.
[8] N. Kröger, C. Bergsdorf, M. Sumper, Frustulins: domain conservation in a protein family associated with diatom cell walls, Eur. J. Biochem. 239 (1996) 259–264.
[9] N. Kröger, R. Deutzmann, C. Bergsdorf, M. Sumper, Species-specific polyamines from diatoms control silica morphology, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 14133–14138.
[10] N. Kröger, R. Deutzmann, M. Sumper, Polycationic peptides from diatom biosilica that direct silica nanosphere formation, Science 286 (1999) 1129–1132.
[11] N. Kröger, R. Deutzmann, M. Sumper, Silica precipitating peptides from diatoms. The chemical structure of silaffin-A from Cylindrotheca fusiformis, J. Biol. Chem. 276 (28) (2001) 26066–26070.
[12] N. Kröger, G. Lehmann, R. Rachel, M. Sumper, Characterization of a 200-kDa diatom protein that is specifically associated with a silica-based substructure of the cell wall, Eur. J. Biochem. 250 (1997) 99–105.
[13] A. Scheffel, N. Poulsen, S. Shian, N. Kröger, Nanopatterned protein microrings from a diatom that direct silica morphogenesis, Proc. Natl. Acad. Sci. U. S. A. 108 (8) (2011) 3175–3180.
[14] B. Tesson, M. Hildebrand, Characterization and localization of insoluble organic matrices associated with diatom cell walls: insight into their roles during cell wall formation, PLOS One 8 (4) (2013), e61675.
[15] M. Hildebrand, B.E. Volcani, W. Gassmann, J.I. Schroeder, A gene family of silicon transporters, Nature 385 (1997) 688–689.
[16] A. Kotzsch, P. Gröger, D. Pawolski, P.H.H. Bomans, N.A.J.M. Sommerdijk, M. Schlierf, N. Kröger, Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization, BMC Biol. 15 (2017) 65.
[17] A.W. Skeffington, M. Gentzel, A. Ohara, A. Milentyev, C. Heintze, L. Bottcher, S. Gorlich, A. Shevchenko, N. Poulsen, N. Kroger, Shedding light on silica biomineralization by comparative analysis of the silica-associated proteomes from three diatom species, Plant J. 110 (6) (2022) 1700–1716.
[18] E.D. Tommasi, J. Gielis, A. Rogato, Diatom frustule morphogenesis and function: a multidicsiplinary survey, Mar. Genomics 35 (2017) 1–18.
[19] K. Thamatrakoln, M. Hildebrand, Silicon uptake in diatoms revisited: a model for saturable and non-saturable uptake kinetics and the role of silicon transporters, Plant Physiol. 146 (2008) 1397–1407.
[20] K. Thamatrakoln, A.B. Kustka, When to say when: can excessive drinking explain silicon uptake in diatoms? BioEssays 31 (2009) 322–327.
[21] H.J. Brasser, H.J. van der Strate, W.W.C. Gieskes, G.C. Krijger, E.G. Vrieling, H. T. Wolterbeek, Compartmental analysis suggests macropinocytosis at the onset of diatom valve formation, SILICON 4 (2012) 39–49.
[22] E.G. Vrieling, Q. Sun, M. Tian, N.A.J.M. Sommerdijk, Salinity-dependent diatom biosilicification implies an important role of external ionic strength, Proc. Natl. Acad. Sci. U. S. A. 104 (25) (2007) 10441–10446.
[23] K. Thamatrakoln, M. Hildebrand, Analysis of Thalassiosira pseudonana silicon transporters indicates distinct regulatory levels and transport activity through the cell cycle, Eukaryot. Cell 6 (2) (2007) 271–279.
[24] R.P. Shrestha, B. Tesson, T. Norden-Krichmar, S. Federowicz, M. Hildebrand, A. E. Allen, Whole transcriptome analysis of the silicon response of the diatom Thalassiosira pseudonana, BMC Genomics 13 (2012) 499.
[25] R.P. Shrestha, M. Hildebrand, Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses, Eukaryot. Cell 14 (1) (2015) 29–40.
[26] M. Hildebrand, D.R. Higgins, K. Busser, B.E. Volcani, Silicon-responsive cDNA clones isolated from the marine diatom Cylindrotheca fusiformis, Gene 132 (2) (1993) 213–218.
[27] M. Hildebrand, K. Dahlin, B.E. Volcani, Characterization of a silicon transporter gene family in Cylindrotheca fusiformis: sequences, expression analysis, and identification of homologs in other diatoms, Mol. Gen. Genet. 260 (1998) 480–486.
[28] T. Mock, M.P. Samanta, V. Iverson, C. Berthiaume, M. Robison, K. Holtermann, C. Durkin, S.S. BonDurant, K. Richmond, M. Rodesch, T. Kallas, E.L. Huttlin, F. Cerrina, M.R. Sussman, E.V. Armbrust, Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon biogenesis, Proc. Natl. Acad. Sci. U. S. A. 105 (5) (2008) 1579–1584.
[29] P. Curnow, L. Senior, M.J. Knight, K. Thamatrakoln, M. Hildebrand, P.J. Booth, Expression, purification and reconstitution of a diatom silicon transporter, Biochemistry 51 (2012) 3776–3785.
[30] M.J. Knight, L. Senior, B. Nancolas, S. Ratcliffe, P. Curnow, Direct evidence of the molecular basis for biological silicon transport, Nat. Commun. 7 (2016) 11926.
[31] G.M. Durak, A.R. Taylor, C.E. Walker, I. Probert, C. de Vargas, S. Audic, D. Schroeder, C. Brownlee, G.L. Wheeler, A role for diatom-like silicon transporters in calcifying coccolithophores, Nat. Commun. 7 (2016) 10543.
[32] A.O. Marron, S. Ratcliffe, G.L. Wheeler, R.E. Goldstein, N. King, F. Not, C. de Vargas, D.J. Richter, The evolution of silicon transport in eukaryotes, Mol. Biol. Evol. 33 (12) (2016) 3226–3248.
[33] R. Keller, C. Ziegler, D. Schneider, When two turn into one: evolution of membrane transporters from two half modules, Biol. Chem. 395 (12) (2014) 1379–1388.
[34] B. van den Berg, C. Pedebos, J.R. Bolla, C.V. Robinson, A. Baslé, S. Khalid, Structural basis for silicic acid uptake by higher plants, J. Mol. Biol. 433 (21) (2021), 167227.
[35] P. Bhattacharya, B.E. Volcani, Sodium-dependent silicate transport in the apochlorotic marine diatom Nitzschia alba, Proc. Natl. Acad. Sci. U. S. A. 77 (11) (1980) 6386–6390.
[36] K. Thamatrakoln, A.J. Alverson, M. Hildebrand, Comparative sequence analysis of diatom silicon transporters: towards a mechanistic model of silicon transport, J. Phycol. 42 (2006) 822–834.
[37] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.
[38] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold - making protein folding accessible to all, biorxiv, 2021.
[39] M.A. Lomize, I.D. Pogozheva, H. Joo, H.I. Mosberg, A.L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes, Nucleic Acids Res. 40 (2012) D370–D376.
[40] P. Kumar, D.N. Woolfson, Socket2: a program for locating, visualizing and analyzing coiled-coil interfaces in protein structures, Bioinformatics 37 (23) (2021) 4575–4577.
[41] C. Kim, J. Basner, B. Lee, Detecting internally symmetric protein structures, BMC Bioinf. 11 (2010) 303.
[42] W.L. DeLano, The PyMOL molecular graphics system. http://www.pymol.org, 2002.
[43] E.L. Wu, X. Cheng, S. Jo, H. Rui, K.C. Song, E.M. Davila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R.M. Venable, J.B. Klauda, W. Im, CHARMM-GUI membrane builder toward realistic biological membrane simulations, J. Comput. Chem. 35 (27) (2014) 1997–2004.
[44] J. Lee, X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks 3rd, A. D. MacKerell Jr., J.B. Klauda, W. Im, CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field, J. Chem. Theory Comput. 12 (1) (2016) 405–413.
[45] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, J. Comput. Chem. 29 (11) (2008) 1859–1865.

[46] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multilevel parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25.
[47] D.A. Case, T.E. Cheatham 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (16) (2005) 1668–1688.
[48] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmuller, A.D. MacKerell Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins, Nat. Methods 14 (1) (2017) 71–73.
[49] W.L. Jorgensen, J. Chandrasekar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (2) (1998).
[50] S. Miyamoto, P.A. Kollman, SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, J. Comput. Chem. 13 (8) (1992) 952–962.
[51] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. Dinola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (8) (1984) 3684–3690.
[52] W. Humphrey, A. Dalke, K. Schulten, VMD - visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[53] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theory Comput. 9 (7) (2013) 3084–3095.
[54] L. Holm, DALI and the persistence of protein shape, Protein Sci. 29 (2020) 128–140.
[55] R.M. Ryan, R.J. Vandenberg, Elevating the alternating-access model, Nat. Struct. Mol. Biol. 23 (2016) 187–189.
[56] D. Drew, O. Boudker, Shared molecular mechanisms of membrane transporters, Annu. Rev. Biochem. 85 (2016) 543–572.
[57] C.A. Durkin, J.A. Koester, S.J. Bender, E.V. Armbrust, The evolution of silicon transporters in diatoms, J. Phycol. 52 (5) (2016) 716–731.
[58] S. Kumar, K. Rechav, I. Kaplan-Ashiri, A. Gal, Imaging and quantifying homeostatic levels of intracellular silicon in diatoms, Sci. Adv. 6 (42) (2020).
[59] T. Jiang, P.C. Wen, N. Trebesch, Z. Zhao, S. Pant, K. Kapoor, M. Shekhar, E. Tajkhorshid, Computational dissection of membrane transport at a microscopic level, Trends Biochem. Sci. 45 (3) (2020) 202–216.
[60] F. Azam, B.B. Hemmingsen, B.E. Volcani, Germanium incorporation into the silica of diatom cell walls, Arch. Mikrobiol. 92 (1973) 11–20.
[61] M. Nikolic, N. Nikolic, Y. Liang, E.A. Kirkby, V. Römheld, Germanium-68 as an adequate tracer for silicon transport in plants.Characterization of silicon uptake in different crop species, Plant Physiol. 143 (1) (2007) 495–503.