PNAS RESEARCH ARTICLE | BIOCHEMISTRY CHEMISTRY

The molecular basis for pore pattern morphogenesis in diatom silica

Christoph Heintze[a], Iaroslav Babenko[a,b], Jirina Zackova Suchanova[a], Alastair Skeffington[c], Benjamin M. Friedrich[b,d], and Nils Kröger[a,b,e,1]

Edited by Joanna Aizenberg, Harvard University, Cambridge, MA; received July 5, 2022; accepted October 13, 2022

Biomineral-forming organisms produce inorganic materials with complex, genetically encoded morphologies that are unmatched by current synthetic chemistry. It is poorly understood which genes are involved in biomineral morphogenesis and how the encoded proteins guide this process. We addressed these questions using diatoms, which are paradigms for the self-assembly of hierarchically meso- and macroporous silica under mild reaction conditions. Proteomics analysis of the intracellular organelle for silica biosynthesis led to the identification of new biomineralization proteins. Three of these, coined dAnk1-3, contain a common protein–protein interaction domain (ankyrin repeats), indicating a role in coordinating assembly of the silica biomineralization machinery. Knocking out individual *dank* genes led to aberrations in silica biogenesis that are consistent with liquid–liquid phase separation as underlying mechanism for pore pattern morphogenesis. Our work provides an unprecedented path for the synthesis of tailored mesoporous silica materials using synthetic biology.

biomineralization | mesoporous silica | silica deposition vesicle | phase separation | ankyrin-repeat domain

Numerous organisms produce inorganic materials with amazingly complex morphologies and extraordinary properties in a process termed biomineralization. Prominent examples include the single-domain magnetite nanocrystals of bacteria that act as sensitive magnetic field sensors (1), the nacreous calcium carbonate layers of mollusks with exceptionally high fracture resistance (2), and the hierarchically porous, silica cell walls of diatoms with intriguing photonic properties (3). A fundamental understanding how genetically encoded machineries are capable of establishing physical and chemical forces that drive morphogenesis of such intricate mineral structures is currently lacking. Therefore, unveiling the mechanisms of biomineralization holds the promise of gaining advanced capabilities to synthesize minerals with tailored properties using environmentally benign processes.

Diatoms are single-celled microalgae that produce silica-based cell walls characterized by species-specific, hierarchical patterns of meso- and macropores, ribs, tubes, and spines among others (4,5). A diatom cell wall is composed of two interlocking halves each consisting of a plate- or dome-shaped valve and an array of ring-shaped girdle bands. Each valve and each girdle band are produced intracellularly in a dedicated membrane-bound compartment termed the silica deposition vesicle (SDV) (6). The valves are more intricately patterned than the girdle bands and thus have generally been the focus of studies on silica morphogenesis. Previously, various models were put forward addressing the mechanism of valve silica morphogenesis in diatoms. The models differ depending on whether they require an organic template or not. The template-free models assume diffusion-limited aggregation (7, 8) or stochastic aggregation in combination with lateral inhibition (9) of silica precursors inside the SDV. The template-dependent models hypothesize that biomolecular assemblies in the SDV lumen, in the SDV membrane, or in the surrounding cytoplasm direct silica morphogenesis (6, 10–13). Only fairly recently gene inactivation studies in the model diatom *Thalassiosira pseudonana* identified the first proteins (SAP1, SAP3, Sin1, Thaps3_21880) involved in diatom silica morphogenesis (14–16). However, a mechanistic understanding of their functions has not yet been achieved. Furthermore, the silica architecture was largely intact in the mutants indicating that additional proteins and possibly other components are involved in the morphogenesis process. In the past, the discovery of proteins involved in silica morphogenesis was severely hampered by the lack of a method for isolating SDVs. Therefore, biochemical analyses of diatom cell walls (17–20) and transcriptomics analyses of synchronized diatom cell cultures have been pursued (21, 22). However, both techniques are prone to reveal many (if not mainly) proteins that are unrelated to silica morphogenesis. To address this problem, we pursued here a direct approach for identifying silica morphogenic proteins by establishing a method for isolating SDVs from *T. pseudonana* and performing proteomics characterization of the SDVs. Through subsequent analysis of the silica development in loss-of-function mutants,

Significance

Inorganic materials harboring patterns of nanoscale pores are important for many applications including catalysis, sensing, and optoelectronics. However, methods for their eco-friendly technological production are lacking. Contrarily, microalgae like diatoms biosynthesize mesoporous silica cell walls under physiological conditions, but so far their pore patterns could not be tailored for desired applications. Here we discovered that biosynthesis of the pore patterns in diatom silica is largely controlled by only three proteins, dAnks1-3. Through knockout of the dAnk encoding genes, we provide the first proof-of-principle demonstration for the production of silica with engineered nanoscale pore patterns through a renewable process. This capability also opens the door for investigating the yet unknown function of porous silica for diatom physiology.

Author contributions: C.H. and N.K. designed research; C.H. and I.B. performed research; I.B., J.Z.S., and B.M.F. contributed new reagents/analytic tools; C.H., I.B., A.S., and N.K. analyzed data; and C.H., I.B., and N.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence may be addressed. Email: nils.kroeger@tu-dresden.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2211549119/-/DCSupplemental.

Published December 2, 2022.

we identified new silica morphogenic proteins and gained insight into the mechanism for pore pattern formation in diatom silica.

Results

Identification of Valve SDV Proteins. We aimed to isolate SDVs by consecutive centrifugation steps in different high-density media. Our efforts were focused on valve SDVs, because they can be specifically labeled in synchronized diatom cells using the fluorescent dye PDMPO allowing for their straightforward identification in a cell lysate using epifluorescence microscopy (23). After centrifugation through a 90% Percoll® containing cushion, valve SDVs were present in two fractions. The upper fraction, PF1, contained valve SDVs with diameters <3 µm and homogeneous PDMPO staining patterns (*SI Appendix*, Fig. S1*A*, yellow arrow). In the lower fraction, PF2, cell wall fragments (*SI Appendix*, Fig. S1*B*, black arrow) were present along with valve SDVs that possessed strongly PDMPO fluorescent rims whose diameters were larger than that of valve SDVs in fraction PF1 (*SI Appendix*, Fig. S1*B*, yellow arrow). Based on the previous observations (23), the sizes of the SDVs and their PDMPO fluorescence patterns indicated that PF1 contained valve SDVs at early to middle stages of development, whereas the valve SDVs in PF2 were at late stages of development. Since a key goal was to isolate SDVs free of cell walls we only processed PF1 further. Fluorescence microscopy indicated the presence of plastid fragments in PF1 (*SI Appendix*, Fig. S1*A*, white arrow) suggesting that also membranes from organelles other than plastids might be present. To further purify the valve SDVs, they were centrifuged twice through buffers with increasing sucrose concentration (first 42.5% then 62.5%). In both cases, the valve SDVs accumulated near the bottom of the tube, and plastid fragments remained mainly in the supernatant (*SI Appendix*, Fig. S1 *C* and *D*). Nevertheless, the final valve SDV fraction, SF2, still contained some plastid fragments (*SI Appendix*, Fig. S1*D*, white arrow).

Proteomics analysis (two biological replicates each with two technical replicates) revealed 1,132 putative valve SDV proteins that were identified in two independent SDV preparations by at least two peptide hits. Based on the annotation in the Universal Protein Resource (UniProt) database, 754 proteins shared similarities to proteins with known functions and 378 proteins possessed no known or suspected function ("predicted" or "uncharacterized" protein; Dataset S1). To narrow down the pool of candidate SDV proteins for further analysis, we focused on those proteins, whose mRNA expression was upregulated during valve SDV synthesis. For this, we used the data from a previous transcriptomics study on the expression of 6,688 genes (total predicted protein-coding genes in a synchronized *T. pseudonana* cultures (22). The expression profile of the *sin1* gene, which encodes for an SDV protein (19, 22), was used as a reference. In our analysis, only the gene regulation during the Si shift-up experiment was considered, which is different to the previous study on *sin1*-like gene expressions that also included gene regulation during Si shift-down (22). The spectral clustering algorithm was implemented to partition the transcriptomics dataset into 11 clusters based on their expression profiles during the Si shift-up (Fig. 1*A* and *SI Appendix*, Fig. S2), and the *sin1* cluster contained 402 genes (Dataset S2). Only 46 of the genes in the *sin1* cluster encode putative valve SDV proteins that were identified in the proteomics analysis (Fig. 1 *B* and *C* and Dataset S3). After removing seven apparent contaminants (1 mitochondrial carrier, 3 plastid proteins, 1 histone, 1 UMP-CMP kinase, 1 endoplasmic reticulum ATPase), the remaining 39 proteins were termed the Valve SDV proteome v1.0 (VSP1.0; *SI Appendix*, Table S1).

BLASTp searches of the VSP1.0 against the NCBI database revealed 20 proteins with predicted functions, including the SDV protein Sin1 and seven subunits of the H^+-ATPase that was previously shown to be located in SDVs and involved in valve biogenesis (24). One protein is a predicted Na^+/H^+ antiporter that might be required for homeostasis of pH and Na^+, which are both critical factors in the chemistry of silica formation (25). Six proteins were predicted to be involved in vesicle transport (3 clathrin coat proteins, 1 Arf, 1 Rab, 1 v-SNARE) and may be required for SDV biogenesis. The identified silicon transporter was previously localized in the plasma membrane, but no information was provided whether it is also located in the SDV membrane (26). The predicted heat shock protein might be involved modulating the properties of SDV membrane analogous to the influence of heat shock proteins on the membrane fluidity of thylakoid membranes in *Synechocystis* sp. PCC 6803 (27). The presence of actin is consistent with the previously observed association of microfilaments with valve SDVs (28, 29). Two kinases are present that might catalyze the phosphorylation of silaffins and silacidins, which are abundant silica-associated phosphoproteins in diatoms (17, 30). The predicted S-adenosylmethionine (SAM) synthetase might be required for biosynthesis of the polyamine-type modifications of silaffins (17) and the silica-associated long-chain polyamines (17, 31), which both contain propylamine residues hypothesized to be derived from SAM (32).

From the 19 VSP1.0 proteins with unknown functions, nine were previously annotated as "SiMat7-like proteins" based on their domain structure, which consists of a N terminal ER transit peptide, a large luminal domain with eight conserved cysteine residues, a single transmembrane helix, and a short cytosolic domain at the C terminus (*SI Appendix*, Table S1; note that SiMat7 is the former name of the SDV protein Sin1) (22). Five of the SiMat7-like proteins and an additional four VSP1.0 proteins are encoded by genes of the "SiMat7 gene cluster," which includes all genes that are downregulated during silicic acid starvation and upregulated after silicic acid replenishment with an mRNA expression peak during valve formation (*SI Appendix*, Table S1) (22). Three members of the SiMat7 gene cluster (Tp23225, Tp5147, Tp21058) encode proteins predicted to be located in the cytosol, because they lack transmembrane domains and organelle targeting sequences. They also share the presence of ankyrin repeat domains (*SI Appendix*, Fig. S3), which are widely occurring protein–protein interaction motifs involved in numerous cellular functions including protein interactions on the cytosolic surface of membranes (33). Interestingly, transcriptomic studies in the diatom *Seminavis robusta* also implicated ankyrin repeat domain-bearing proteins in cell wall formation (34), which suggest that these types of proteins might be general components of the diatom machinery for silica biogenesis. We performed a phylogenetic analysis of the three ankyrin repeat domain bearing VSP1.0 proteins (Tp23225, Tp5147, and Tp21058) using the Marine Microbial Eukaryotic Transcriptome Sequencing Project database (35), as well as the PLAZA diatoms database (36). This revealed that Tp23225 is restricted to centric diatoms (Mediophyceae), whereas Tp5147 and Tp21058 are very common and possibly ubiquitous among diatoms, with a phylogenetic distribution similar to Sin1 (*SI Appendix*, Fig. S4). Tp5147 has additional putative homologues in a few other silica forming protists such as the chrysophyte *Paraphysomonas bandaiensis* and in non-silicifying groups such as the Prymnesiophytes (Haptophyta). Some BLAST hits outside of the diatoms may be due to diatoms being prey for certain protists (e.g., the ciliate, *Tiarina fusa*) or due to tertiary endosymbioses such as in *Krytoperidinium foliaceum* (Myzozoa), which has a second nucleus of diatom origin (37). To highlight the presence of

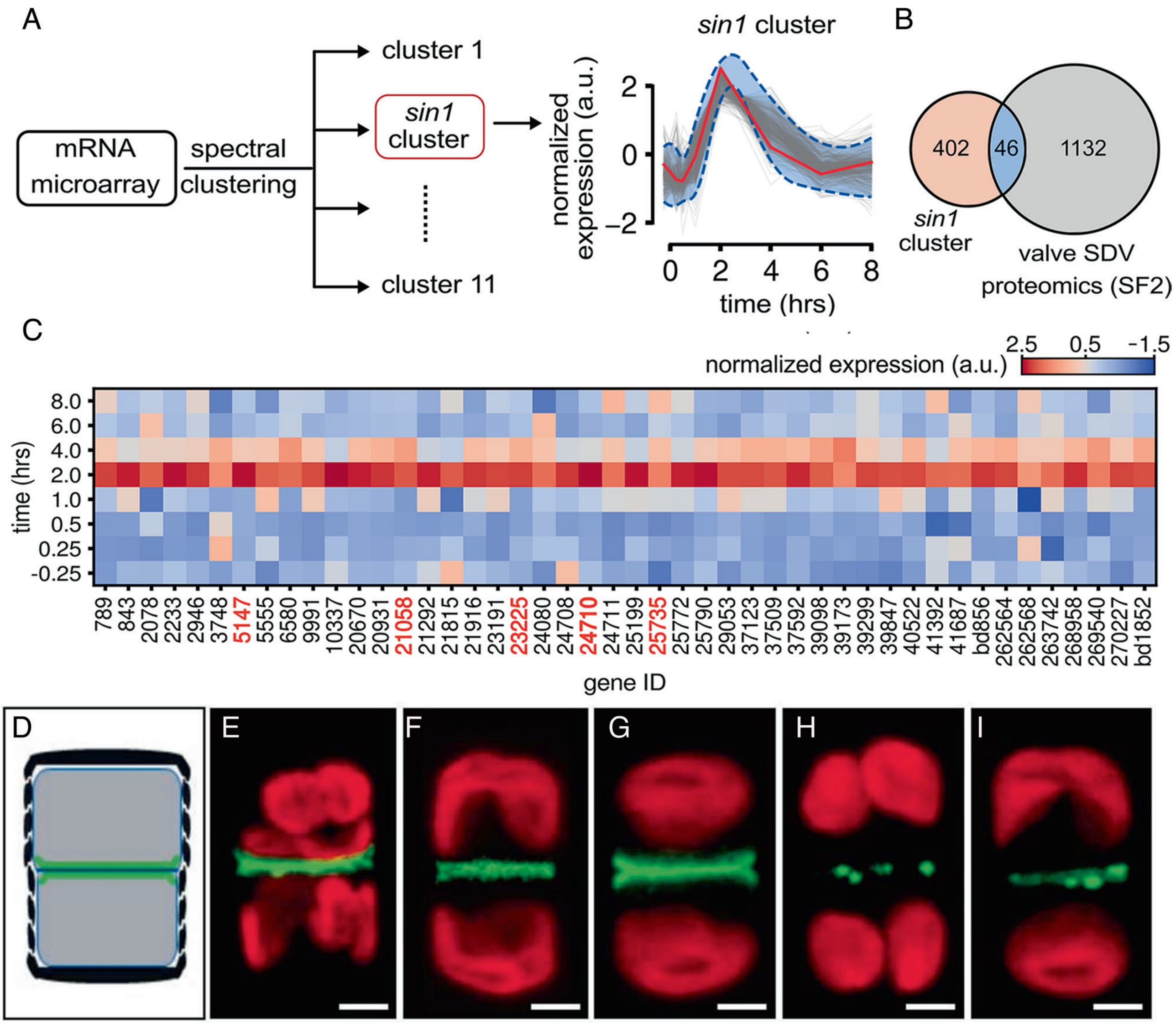

Fig. 1. Identification of valve SDV proteins. (*A*) Scheme of the cluster analysis workflow that established in a published transcriptomics dataset (22) 11 different gene expression profiles. The expression patterns in the *sin1* cluster is shown: red line—*sin1* profile, grey lines—profiles of the other 401 genes, blue area—cluster average within two standard deviations. (*B*) The Euler diagram shows the overlap between proteins identified in the valve SDV fraction (Dataset S1) and genes of the *sin1* cluster (Dataset S2). (*C*) Heat map of mRNA expression of *sin1* cluster genes that encode proteins identified in the valve SDV fraction. Gene IDs of *sin1* (24710), Tp25735, and *dank1-3* (Tp23225, Tp5147, Tp21058) are highlighted in red. The y-axis indicates time before (negative) and after addition of silicic acid, which triggers synchronous cell cycle progression. Valve SDV development occurs within hours 2.5–4.5 (23). (*D*) Scheme of a dividing diatom cell in cross section: black—cell wall, blue—plasma membrane, grey—protoplasm, green—valve SDV (other compartments not shown). (*E-I*) Confocal fluorescence microscopy images of individual dividing cells expressing the indicated proteins as GFP fusions (z-projection of optical slices through mid-cell region): green—GFP, red—chloroplast autofluorescence. (Scale bars: 2 μm.)

the ankyrin repeats ("Ank"), we named these VSP1.0 proteins dAnk1 (Tp23225), dAnk2 (Tp5147), and dAnk3 (Tp21058) with "d" indicating their prevalence in diatoms. The remaining six VSP1.0 proteins have not been implicated in silica biogenesis and have never been studied before.

To start investigating whether the VSP1.0 proteins are associated with valve SDVs in vivo, we selected for GFP tagging the predicted cytosolic proteins dAnk1-3 and the predicted transmembrane protein Tp25735 (N-terminal ER transit peptide, three transmembrane domains; *SI Appendix*, Fig. S3). In dividing cells of *T. pseudonana,* each of the GFP fusion proteins was present at the site of the developing valve SDVs in the mid-cell region matching the location of Sin1-GFP (Fig. 1 *D*–*I* and *SI Appendix*, Fig. S5). While the GFP fusions of Tp25735 and dAnk1 were homogeneously distributed across the expanding valve SDVs (Fig. 1 *F* and *G* and *SI Appendix*, Fig. S5 *A* and *B*), dAnk2-GFP and dAnk3-GFP displayed dotted patterns (Fig. 1 *H* and *I* and *SI Appendix*, Fig. S5 *C* and *D*). Altogether, the localization experiments strongly support the assumption that Tp25735 and dAnk1-3 are bona fide valve SDV proteins.

Functional Analysis of dAnk Proteins In Vivo. Given the intracellular locations and sequence characteristics of dAnk1-3, we hypothesized that they are part of a protein–protein interaction network at the cytoplasmic surface of the valve SDV membrane that might be required for SDV function. To test this hypothesis, we generated independent knockout (KO) strains of *T. pseudonana* each lacking an individual *dank* gene using the CRISPR/Cas9 method previously established for this organism (15, 38). For generating the KO strains, three plasmids were designed each encoding Cas9 and three or four guide RNAs (gRNA) specific to one *dank* gene (*SI Appendix*, Fig. S6). The KO plasmids were introduced into *T. pseudonana* wild type by biolistic particle bombardment. Sequencing the products from genomic PCRs specific for each targeted gene was used to identify independent transformant strains in which the targeted *dank* gene contained a frameshift mutation (*SI Appendix*, Fig. S7). In all cases, the disrupted *dank* genes encoded rather short polypeptides with largely altered amino acid sequences (*SI Appendix*, Fig. S8). Electron microscopy analyses of the isolated biosilica from each KO strain revealed differences to the wild-type morphology in the valves (Fig. 2 and *SI Appendix*, Figs. S9–S12). The valve of *T. pseudonana* has a circular shape and contains a network of ridges composed of branched radial ribs and bridges between them (Fig. 2 *A*, *E*, *I*, and *N*; see *SI Appendix*, Fig. S9 for details). The space delineated by the ridge network is filled with a layer of silica, called cribrum plate, which contains numerous circular openings 28 ±4.4 nm in diameter (Fig. 2*E*), called cribrum pores. The circular shape of the valve and the ridge network appeared to be unchanged in all KO strains (Fig. 2 *B*–*D* and *SI Appendix*, Figs. S10 *A*–*C*, S11 *A*–*C*, and S12 *A*–*C*). In the *dank1KO* strains, the pattern of cribrum pores appeared to be different with the pores positioned predominantly in the center between two neighboring ribs (Fig. 2*F* red arrowheads and *SI Appendix*, Fig. S10 *D*–*F*) rather than close to the ribs as in the wild type (Fig. 2*E* red arrowheads and *SI Appendix*, Fig. S9*C*). In the *dank2KO* strains and the *dank3KO* strains, the pore densities appeared to be enhanced and reduced, respectively (Fig. 2 *G* and *H* and *SI Appendix*, Figs. S11

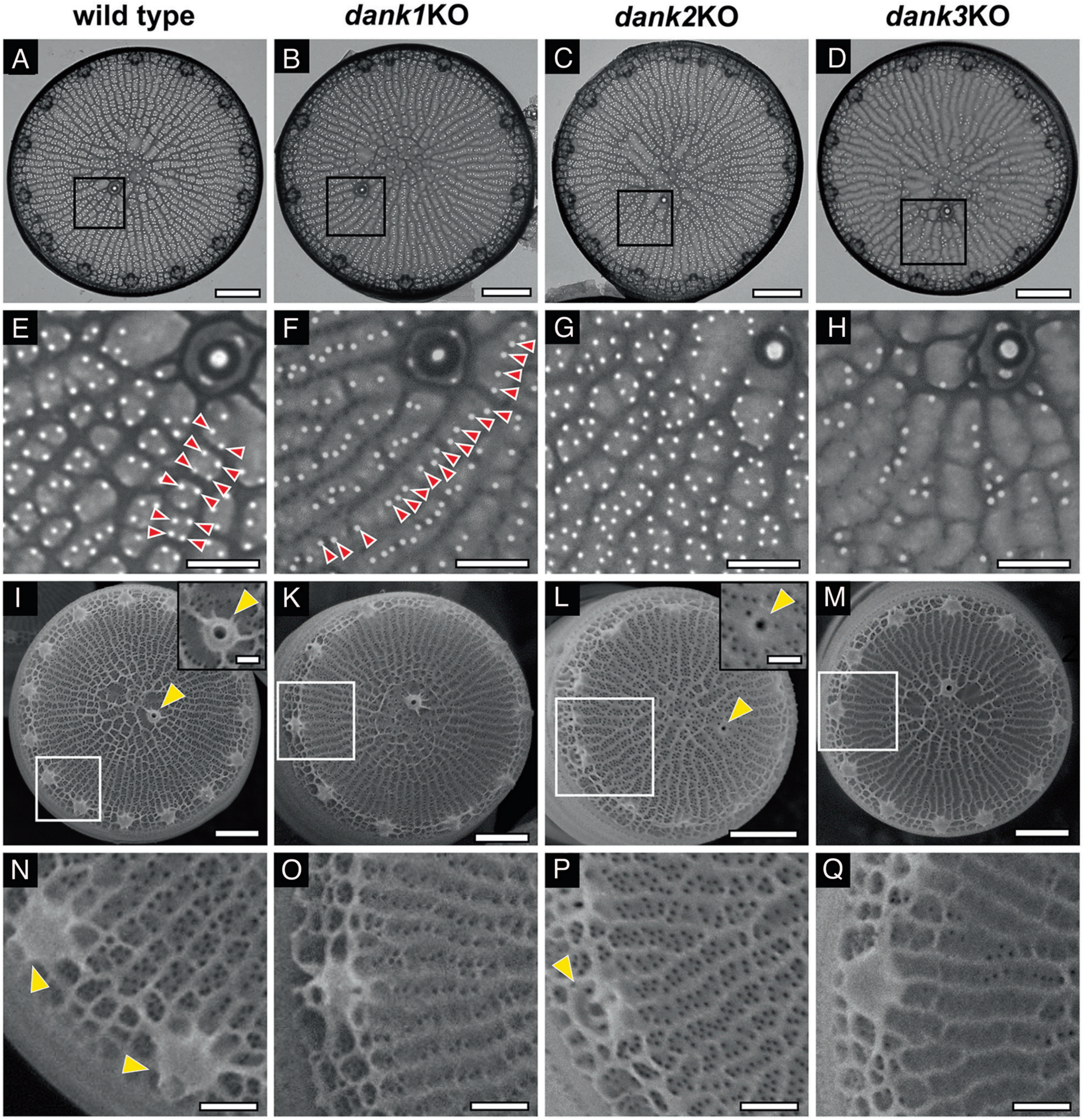

Fig. 2. Valve morphologies in *T. pseudonana* wild type and *dank* knockout mutants. (*A*–*D*) TEM and (*I*–*M*) SEM images of entire valves. (*E*–*H*, *N*–*Q*) Details corresponding to the boxed areas in *A*–*D* and *I*–*M*, respectively. Red arrowheads point to cribrum pores and yellow arrowheads to the fultoportulae. The inset in *L* is a zoom-in of the region around a central fultoportula (yellow arrowhead). (Scale bars: 1 µm in *A*–*D* and *I*–*M*, 300 nm in *E*–*H* and *N*–*Q*, 200 nm in *L*, *Inset*.)

D–F and S12 *D–F*). Scanning electron microscopy (SEM) analysis revealed that dAnk2 but neither dAnk1 nor dAnk3 is also involved in fultoportula morphogenesis. Fultoportulae are specialized pore systems that are located near the center and the rim of the valve (*SI Appendix*, Fig. S9) and are required for the secretion of chitin fibers (39). At the distal valve surface, a fultoportula is covered by a funnel-shaped tube (Fig. 2 *I*, *K*, and *M* and *SI Appendix*, Figs. S9 *B* and *D*, S10*G*, and S12*G*) (20), but in *dank2KO* mutants, the distal tubes are missing in all fultoportulae (Fig. 2 *L* and *P* and *SI Appendix*, Fig. S11 *G–M*).

To evaluate the abundance and patterns of cribrum pores in wild-type and *dank*KO strains, we developed an image analysis algorithm for cribrum pore recognition, allowing the quantitative description of observed valve morphology in a semiautonomous fashion (Fig. 3*A*). To describe the pore pattern arrangement in wild type and mutants, we introduced two morphological features: pore density (PD) and pore distribution function (PDF) (Fig. 3*A*). The quantification of PD (Fig. 3*B*) confirmed the statistically significant decrease of cribrum pore densities on average ($P = 3.14 \cdot 10^{-14}$, independent two-sample *t* test) in the *dank1KO* (65.3 ± 24.0 μm^{-2}) and the *dank3KO* (48.8 ± 13.9 μm^{2}) strains ($P = 8.27 \cdot 10^{-25}$, independent two-sample *t* test), respectively, compared to wild type (105.4 ± 22.3 μm^{-2}) (Fig. 3*B*). In contrast, the cribrum PD in the *dank2KO* strains (147.2 ± 29.4 μm^{-2}) was ~40% increased ($P = 1.32 \cdot 10^{-11}$, independent two-sample *t* test, Fig. 3*B*). Further analysis revealed that in all *dank* single KO strains the diameters of the cribrum pores were significantly reduced by 10 to 20% compared to the average pore diameter of 28.3 nm ±4.4 nm in wild type (*SI Appendix*, Table S2). To obtain information about the geometrical arrangement of the pores, we used a PDF measurement. The PDF is a

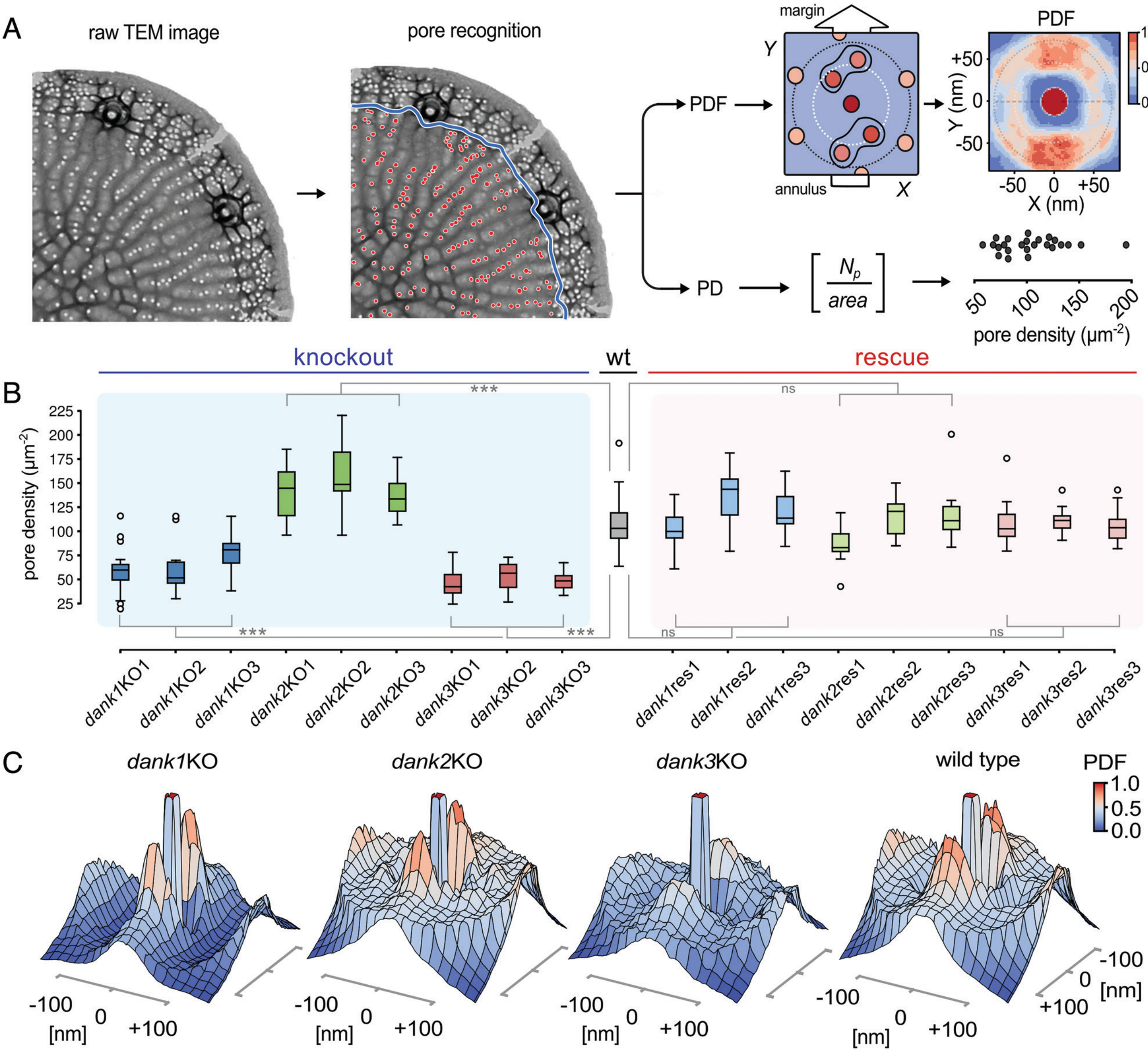

Fig. 3. Analysis of the pore patterns in valve silica from wild type and *dank* knockout mutants. (*A*) Recognition of cribrum pores from a TEM image of an individual valve. The girdle band margin is cropped from the analysis and only pores in the area underneath the blue line were analyzed. The recognition algorithm enables measurement of the PD and the PDF. (*B*) Statistical analysis of the pore densities from wild type, *dank* knockout mutants, and *dank* rescue strains: central line—median value, box edges—25th and 75th percentiles, whiskers—minimum and maximum values, white circles—sample outliers (independent two-sample *t* test ***$P < 0.001$). Differences that were not statistically significant are denoted "ns." The number of analyzed valves was 50 for the wild type and between 13 and 19 for each genetically modified strain. (*C*) Averaged PDF functions for the *dank* knockout strains within a domain of 300 nm × 300 nm. The central peak marks the position of the reference pore, while the two satellite peaks characterize the relative position of the closest neighbors. The color scale indicates normalized probabilities of finding a neighboring pore.

probability distribution function of finding the closet adjacent pore with respect to a reference pore position, and was inspired by the pair distribution function, a standard tool in condensed matter physics (40–42). By default, the centered peak of the PDF marks the position of the reference pore (Fig. 3*C*). The presence of two satellite peaks reveals a preferred relative positioning of nearby pores (Fig. 3*C*). The PDF is calculated in such a way that the orientation of satellite peaks reflects the positioning of pores in rows parallel to the radially aligned ribs, while the distance of the satellite peaks from the central peak characterizes the typical distance between neighboring pores (60 ± 18 nm) in these rows (Fig. 3*C*). The characteristic configuration of a central peak flanked by two satellite peaks is particularly pronounced in the *dank1KO* mutants (Fig. 3*C* and *SI Appendix*, Fig. S13). This reflects the observation that the *dank1KO* strain predominantly features only a single row of pores between neighboring ribs. In wild type, *dank2KO*, and *dank3KO* strains, the row of pores corresponding to the peaks is flanked by a second row of pores either left or right of the first row. Averaging over the entire strain yields intermediate values of the PDFs left and right of the peaks, in contrast to the "trenches" visible in the PDF for *dank1KO* (Fig. 3*C*). In the *dank3KO* mutant, the height of the secondary peaks is reduced (Fig. 3*C*). This is consistent with a configuration where pores are still arranged with equidistant spacing along rows, but some positions in these rows are skipped (as reflected by the lower PD in the *dank3KO* strain; Fig. 3*B*). Thus, the PDF analysis confirmed that the patterns of cribrum pores are significantly different in *dank1KO* and *dank3KO* strains, which is indicative of different functions in pore formation. We also generated double KO strains, termed *dank1/3dKO*, which lacked both the *dank1* and the *dank3* gene (*SI Appendix*, Figs. S14 and S15). The valves from three independent *dank1/3dKO* strains were largely devoid of cribrum pores (15.1 ± 11.7 μm^{-2}) exhibiting only 14% of the PD of wild-type valves (*SI Appendix*, Fig. S16). Altogether, the data from pore analysis of the single- and double-KO strains strongly suggest that the dAnk proteins are involved in both the formation and patterning of cribrum pores in *T. pseudonana*.

To corroborate that the observed morphological phenotypes in the KO mutants were generated by the inactivation of the targeted *dank* genes rather than off-target effects, we attempted to restore the wild-type phenotype by introducing an intact copy of the *dank* gene into the corresponding KO mutant. The introduced *dank* copy was mutated to lack the recognition sites for the guide RNAs without altering the encoded amino acid sequence (*SI Appendix*, Fig. S17) and was fused in frame to the *egfp* gene to facilitate the identification of transformant strains that express the reintroduced *dank* gene. Transformant strains that exhibited GFP fluorescence were termed "rescue clone" and denoted *dankXresN* with *X* and *N* specifying the introduced *dank* gene and the number of the independent clone, respectively. Confocal fluorescence microscopy confirmed proper sorting of each dAnk-GFP fusion protein to the valve SDVs in the cells of the rescue strains (*SI Appendix*, Fig. S18). Analysis by our automated cribrum pore recognition algorithm revealed that the pore densities (Fig. 3*B*) and the PDFs in valves from the rescue clones (*SI Appendix*, Fig. S19) very closely matched the characteristics of wild-type valves (Fig. 3 *B* and *C*). In the *dank2res* strains, the distal tubes of the fultoportulae were also restored (*SI Appendix*, Fig. S20). Altogether, these data clearly confirmed that the lack of the dAnk proteins caused the altered valve morphotypes in the *dank* KO strains.

To gain insight into the mechanism by which dAnk proteins influence the morphogenesis of the cribrum pore patterns, we performed TEM analysis of nascent valve SDVs in different developmental stages. As described previously for the wild type, the stages of valve development were determined by the progression of characteristic morphological features (23). In the early stage of valve formation, the space between the radially growing silica ribs is completely silica free (Fig. 4*A*). The mid-stage is characterized by the onset of silica deposition in the inter-rib space generating a corrugated, wave-like pattern along the ribs (Fig. 4 *B* and *C*). Each trough of this wave pattern is the origin of a cribrum pore, as it remains silica free during silica deposition in the inter-rib spaces (Fig. 4*C* and *SI Appendix*, Fig. S21). Eventually, in the mature valve, the space between the ribs is completely filled with a porous layer of silica (i.e., the cribrum plate) with occasional silica bridges between neighboring ribs (Fig. 4*D*).

The cribrum plates are punctuated by double rows of pores, where each row runs adjacent to a rib (Fig. 4*D*). The observed development of the pore pattern is consistent with a liquid-liquid phase separation process that generates nanodroplets in the inter-rib spaces. The nanodroplets might be mainly composed of long-chain polyamines that are abundant components in diatom biosilica (6) We hypothesize that the droplets align along the silica ribs due to electrostatic interactions between the negatively surface of the rib silica (Fig. 4*A'* black lines) and a presumed polycationic surface charge on each nanodroplet (Fig. 4*A'*, orange circles). Electrostatic repulsion between the nanodroplets would prevent droplet fusion and lead to their equidistant positioning along each side of the rib (Fig. 4*A'*). As the ribs expand laterally, the inter-rib spaces become increasingly filled with silica except where the nanodroplets are located (Fig. 4 *B'* and *C'*). This scenario would perfectly explain the wave-like silica patterns at the rib boundaries during intermediate stages of cribrum plate development (Fig. 4*B*). In areas where the distance between neighboring ribs is larger than 120 nm, pores are also present in the central region of the cribrum plates (23). These pores may be templated by excess nanodroplets that do not fit along the ribs and become electrostatically repelled by the rib-attached nanodroplets thus accumulating in the center. We noticed that the density of developing cribrum pores in immature valves was about twice as high than in mature valves (*SI Appendix*, Fig. S22). This suggests, according to our model, that a substantial fraction of the nanodroplets disassembles at later stages while silica deposition is still occurring. As a consequence, such initial pores become filled in with silica. This hypothesis is consistent with the observation that mature valves contain many round spots that are less electron dense than the surrounding cribrum plate silica and are located in positions where cribrum pores would be expected (Fig. 4 *D* and *D'* red arrowheads).

Using the nanodroplet-based model, it is possible to explain the altered cribrum pore patterns in the valves of *dank* KO mutants. In developing valves lacking *dank1* the widening of the ribs appears to proceed faster than the development of the nanodroplets (Fig. 4 *E*, *E'*, *F*, and *F'*). Therefore, by the time the droplets appear, the silica-free space between the ribs has narrowed down so much that it can accommodate only a single row of nanodroplets (Fig. 4 *F* and *F'*). As a result, single rows of cribrum pores, each roughly equidistant from the neighboring ribs, emerge as a dominant feature in the valve silica of *dank1KO* mutants (Fig. 4*G* orange arrowheads, Fig. 4*G'* and *SI Appendix*, Fig. S21). Also in mature valves of *dank1KO* mutant spots with reduced electron densities were observed (Fig. 4*H* red arrowheads) indicating that occasional disassembly of nanodroplets is also occurring in these mutants (Fig. 4*H'*). TEM analysis of valve development in *dank3KO* mutants provided further support that nanodroplet stability is an important factor in pore pattern morphogenesis. The early stages of valve development in these mutants are similar to wild type with wave-like silica boundaries developing at the onset of lateral expansion of the ribs (Fig. 4 *I* and *K*). However, these wave-like patterns fail to develop continuously

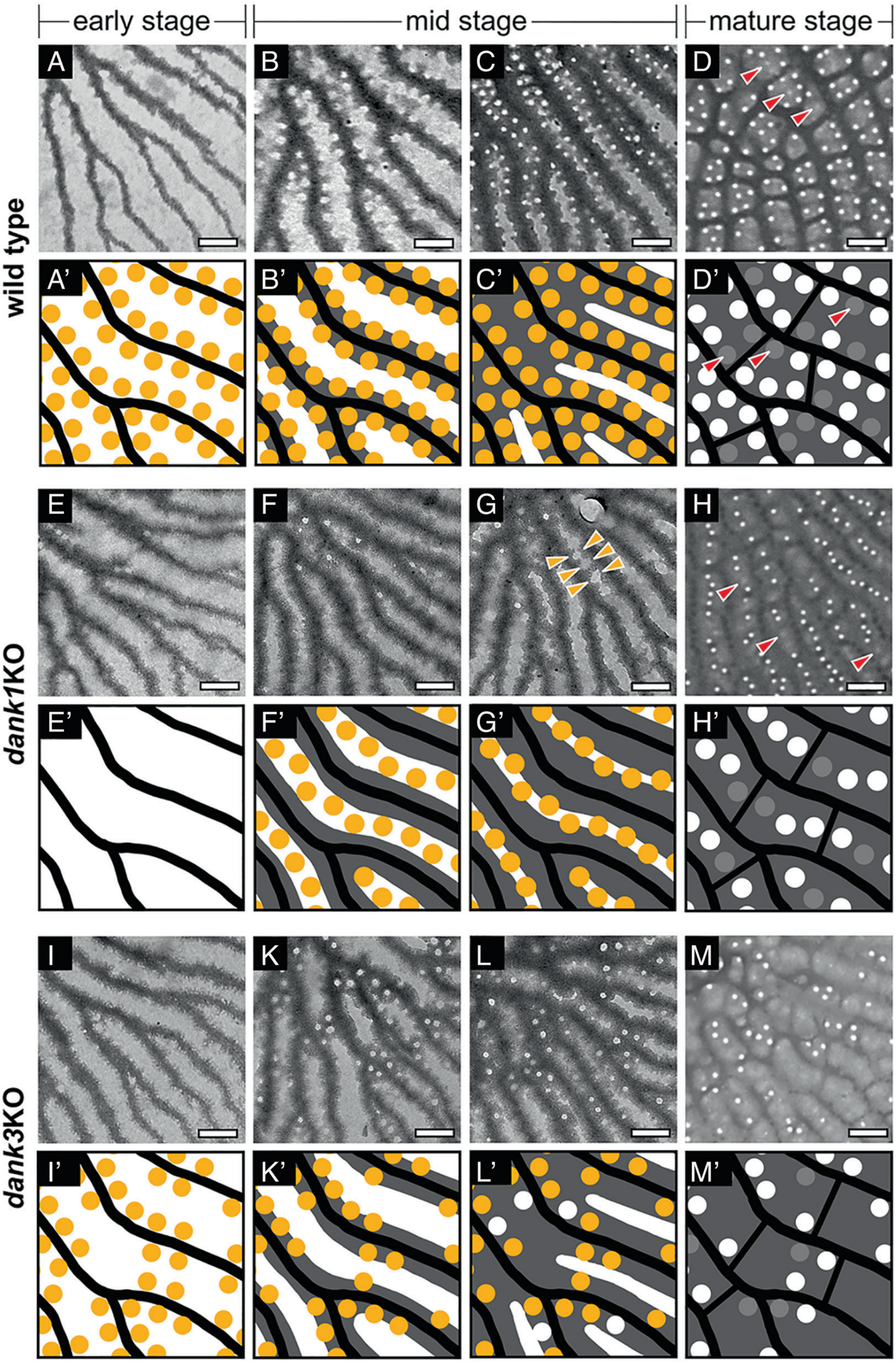

Fig. 4. Morphogenesis of porous silica patterns in valves from *T. pseudonana* wild type, *dank1KO*, and *dank3KO* strains. TEM images show representative sub areas of nascent valve silica from (*A–D*) wild type, (*E–H*) *dank1KO*, and (*I–M*) *dank3KO* in different developmental stages proceeding from *Left* to *Right*. (Scale bars: 200 nm.) Based on these TEM data and assuming liquid-liquid phase separation to occur in the SDV (11,12), we propose models for cribrum pore formation in valves of (*A'–D'*) wild type, (*E'–H'*) *dank1KO*, and (*I'–M'*) *dank3KO* strains. Black—silica ribs, orange—organic nanodroplets, grey—cribrum plate silica. Red arrowheads point to less electron dense spots assumed to originate from unstable cribrum pores that were filled in with silica. Yellow arrowheads indicate forming cribrum pores.

along the ribs (Fig. 4*K*). We hypothesize that this is caused by a decreased stability of the nanodroplets leading to their accelerated disassembly compared to the nanodroplets in the wild type (Fig. 4 *I'* and *K'*). As silica deposition in inter-rib spaces progresses, the disassembly of a significant amount of nanodroplets continues, resulting in pore-free regions, which are substantially larger than in wild type (Fig. 4 *L*, *L'*, *M*, and *M'* and *SI Appendix*, Fig. S21). The strongly enhanced PD in *dank2KO* mutants (see Fig. 3*B*) can also be explained by the phase separation model by assuming that the stability of the nanodroplets is increased compared to wild type. As a consequence, fewer nanodroplets would disassemble during valve development resulting in a higher number of pores in the mature valve silica.

Discussion

In the present work, we performed the first cell wall-free enrichment and proteomics analysis of SDVs, which led to the identification of new silica biomineralization proteins. These include dAnk1-3, which are involved in the morphogenesis of pore patterns in biosilica. dAnk-like proteins appear to be largely restricted to the diatoms and widespread, if not ubiquitously present in this group of organisms. Functional analysis of dAnk1-3 in the model diatom *T. pseudonana* provided support for a LLPS-based model for pore pattern morphogenesis. In our model, the cribrum pore patterns are templated by nanodroplets whose biogenesis is strongly influenced by dAnks. dAnk1 is involved in timing the onset of nanodroplet biogenesis relative to the onset of silica deposition in the inter-rib spaces (see Fig. 4). dAnk2 and dAnk3 are mutual antagonists that influence the lifetime of the nanodroplets as outlined below. In vivo evidence for the existence of nanodroplets is currently missing. However, in vitro studies have demonstrated that silica-associated biomolecules from diatoms (silaffin proteins, long-chain polyamines) and synthetic analogs thereof undergo spontaneous LLPS in vitro under biologically relevant conditions (43–46).

The association of dAnks with developing valves and the absence of currently known organelle targeting motifs in their polypeptide sequences, predicts dAnks to be located on the cytoplasmic surface of valve SDVs (Fig. 5*A*). In this position, dAnks can only indirectly influence silica morphogenesis, which occurs exclusively in the lumen of the SDV membrane (Fig. 5*A*). As detailed in the following, we propose that dAnks influence biogenesis of the silica pore pattern through interactions with specific SDV transmembrane proteins, termed dAnk receptors (Fig. 5*A*), that remain to be identified.

According to our model, dAnk1 binds to the cytosolic domain of a transmembrane protein (dAnk1 receptor) involved in nucleating liquid-liquid phase separation (LLPS) throughout the SDV lumen thereby generating the nanodroplets required for pore pattern formation (Fig. 5*B*). The sequence of dAnk1 (and the other two danks) does not indicate a propensity for phase separation, and thus the molecular cause for LLPS in the SDV is currently speculative. LLPS might be triggered by changes in the physicochemical conditions inside the SDV that promote the aggregation of biomacromolecules. Such changes could include (among others) shifts in pH, or in the concentration and composition of metal ions (e.g., Na^+, K^+, Ca^{2+}), or in the supply of the silica precursor. The flow of these components across the SDV membrane is likely under control of transporter proteins (ion

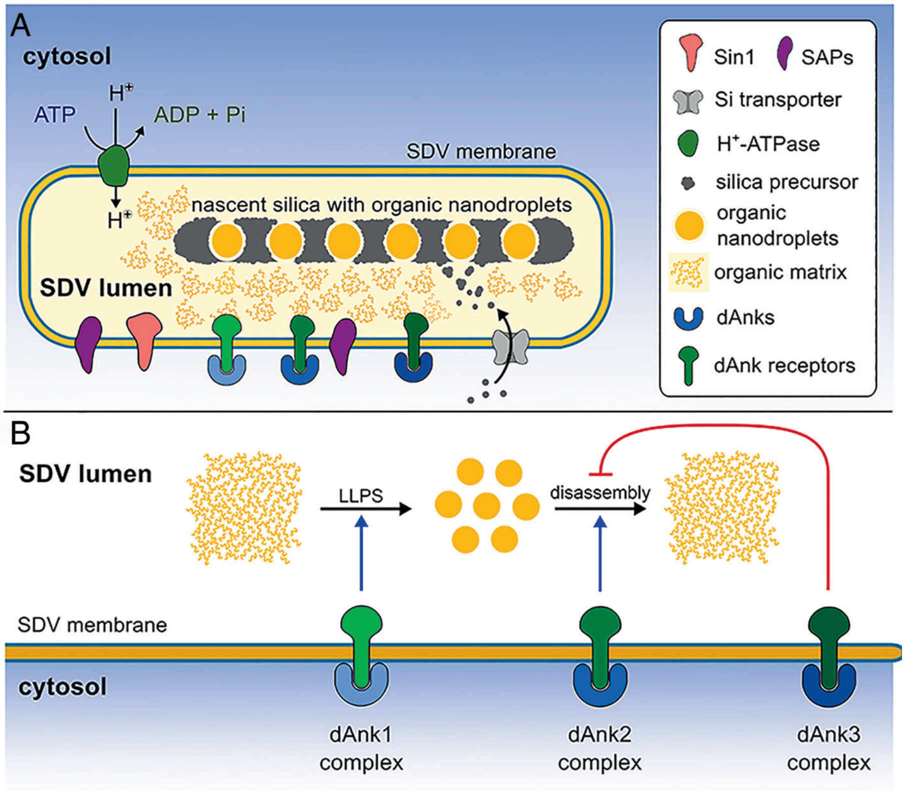

Fig. 5. Putative model for dAnk-controlled liquid-liquid phase separation (LLPS) in the SDV (components are not drawn to scale). (*A*) Current model of an SDV. The lumen is assumed to contain an organic matrix composed of strongly charged, soluble biomacromolecules (e.g., zwitterionic silaffins, polyanionic silacidins, polycationic LCPA) that are prone to undergo phase separation (11,30,43). The SDV membrane contains transmembrane proteins Sin1, SAP1, SAP3 that might interact with LCPA and silaffins in the SDV lumen (14,19). A V-type H^+-ATPase establishes and acidic pH in the SDV lumen (24). The chemical structure of the silica precursor molecules and the Si-transporter that catalyzes their import into the SDV are yet unknown. dAnks associate with the cytoplasmic surface of the SDV by binding to yet unknown transmembrane proteins (coined dAnk receptors) resulting in a dAnk complex. (*B*) The binding of dAnk1 (light blue symbol) to its receptor (light green symbol) triggers liquid–liquid phase separation in the SDV lumen. Stability of the nanodroplets is controlled by an antagonism between dAnk2 (blue symbol) and dAnk3 (dark blue symbol) complexes with their specific receptors. While dAnk2 complexes promote the disassembly of nanodroplets, dAnk3 complexes have a stabilizing effect.

channels, ion pumps), and corresponding candidates are present in the valve SDV proteome (see *SI Appendix,* Table S1). The dAnk1 complex might be involved in regulating such transporters. Alternatively, the dAnk1 complex itself might be capable of self-aggregation and aggregation with other biomolecules in the SDV lumen. The binding of dAnk1 to its receptor might trigger a conformational change resulting in the exposure of a large disordered domain of the receptor on the luminal side of the SDV membrane. Disordered domains are characteristic features of proteins that undergo LLPS (47). Even in the absence of dAnk1, the dAnk1 receptor might have an inherent ability to trigger LLPS but with significantly reduced efficiency. This would explain the presence of a reduced number of pores in the valves of *dank1KO* mutants (Figs. 2*F*, 3*B*, and 4*H*). Our model proposes further that the stability of the nanodroplets is under control of an antagonism between dAnk2 and dAnk3. The complex of dAnk2 with its receptor would promote disassembly of the nanodroplets, whereas they would be stabilized by the dAnk3 complex. These assumptions are consistent with the enhanced pore density in the absence of dAnk2 (*dank2KO* mutants) and the reduced pore density in the *dank3KO* mutants (Fig. 3*B*). Our data do not provide information on the mechanism by which the competition between dAnk2 and dAnk3 may influence the stability of the nanodroplets.

Recently, LLPS has been identified as a biological mechanism to organize many intracellular processes (47) Based on in vitro mineralization studies, it was proposed that LLPS might also play a role in biomineralization processes, but in vivo evidence is lacking so far (48). A LLPS-based model for morphogenesis of the hexagonally arranged pores in the silica of diatoms from the genus *Coscinodiscus* has been previously proposed (11). This model also assumes long-chain polyamine-bearing droplets as templates for pore formation. It further suggests that silica precipitation at the surface of the nanodroplets incorporates a fraction of the polyamines causing shrinkage and finally the disassembly of the nanodroplets (11). In the former model, pore pattern morphogenesis is solely guided by close packing of the droplets, while the SDV membrane has an entirely passive role as compartment boundary. In contrast, the results of the present study demonstrate that the SDV membrane has to play an active role in pore pattern morphogenesis by transducing the morphogenic actions of dAnks from its cytoplasmic surface into the SDV lumen. To date, the interaction partners of dAnk1-3 are unknown. However, we noted that *sin1KO* mutants exhibited a ~35% increase in pore density (Fig. 3*B*) similar to the *dank2KO* mutants (49), which would be consistent with Sin1 being part of the dAnk2 complex that destabilizes the nanodroplets. The VSP1.0 provides several more candidates for transmembrane proteins [in addition to SAP1 and SAP3 (14,15)] that might interact with dAnk2 or the other two dAnks (*SI Appendix*, Table S1). The discovery and functional characterization of dAnks provides a significant step forward in our understanding of silica pore pattern formation in diatoms. Together with the models for the dAnks' modes of action presented here, we provide an experimentally testable hypothesis how molecular interactions at the boundary of the biomineralization compartment can control mineral morphogenesis at scales of hundreds of nanometers and beyond using a LLPS process. LLPS-independent models for silica morphogenesis in diatoms hypothesized a key role for cytoskeleton fibers or other cytoplasmic organelles (e.g., "spacer vesicles") in controlling SDV development (10, 13). In these models, dAnks could play a role in providing a molecular link between proteins associated with cytoskeleton fibers or "spacer vesicles" and SDV transmembrane proteins required for SDV expansion or silica deposition. However, these models do not provide a mechanistic path for the biogenesis of pore patterns like those of the cribrum plate of *T. pseudonana* and related species.

To date, the biological functions of porous silica in diatoms have remained speculative including protection against hard UV light, acquisition of photosynthetically relevant radiation, and nutrient uptake (3, 50). Being able to predictably alter the pore patterns in silica through genetic manipulation of dAnks provides unprecedented opportunities for testing such hypotheses. From a materials perspective, diatom silica belongs to the class of meso- and macroporous materials, which have an enormously wide range of applications including photonics, catalysis, sensorics, and drug delivery (51–54) The properties of these materials critically depend on the sizes, spacing, and arrangement of the pores. Much progress has been made on developing synthetic methodologies that achieve exquisite control over the porosity of these materials, yet they generally involve environmentally hazardous reagents and energy-intensive processes. In contrast, diatoms produce meso- and macroporous silica through an environmentally benign process that is renewable and carbon neutral, and proceeds with an exponential production rate. The present work provides the first proof-of-principle demonstration that the biological production of silica materials with tailored mesoporous patterns should be feasible. This widely opens the door to an eco-friendly, biotechnological production of functional materials using genetically encoded minerals.

Materials and Methods

Chemicals and Enzymes. The enzymes for molecular genetics and phenylmethylsulfonyl fluoride (PMSF) were purchased from Thermo Fisher Scientific. Percoll®, sucrose, dithiothreitol, and iodacetamide were obtained from Sigma-Aldrich. PDMPO was purchased from Aat Bioquest. The antibiotics blasticidin S hydrochloride and nourseothricin sulfate were obtained from InvivoGen and Jena Bioscience. All oligonucleotides were purchased from Eurofins Genomics.

Culturing and Transformation of *T. pseudonana*. The wild type and all transformant strains of *T. pseudonana* (Husted) Hasle et Heimdahl clone CCMP1335 were maintained in artificial seawater medium at 18 °C and 5,000 lx in a 12-h/12-h day-night cycle, according to the recently published protocol (23). Transgenic *T. pseudonana* strains were generated by particle bombardment as described in *SI Appendix*, Supplementary Text.

Enrichment of Valve SDVs. All steps were performed at 4 °C. The density gradient buffer (DGB) that was used for all solutions was composed of 50 mM HEPES-NaOH pH 7.5, 100 mM NaCl, 50 mM sucrose and 100 µM PMSF. A 300-mL *T. pseudonana* cell culture was synchronized by two consecutive silicon starvation–replenishment cycles, PDMPO labeled, and lysed as described previously 3 h after silicic acid addition (i.e., when the proportion of cells with valve SDVs was highest) (23). The cell lysate was diluted to 30 mL with DGB, laid on top of a preformed Percoll® gradient (90 mL of 80% Percoll® in DGB were centrifuged at 30,000 g for 1 h in a 50.2Ti (Beckman-Coulter) fixed angle rotor for gradient formation) and centrifuged for 1 h at 2,000 g in an A-4-81 rotor (Eppendorf). After centrifugation, 40 mL of the gradient was removed from the top. The visible green fraction was collected (15 mL), diluted to 30 mL with DGB and subsequently centrifuged again through a freshly prepared Percoll® gradient, as described above. The visible green fraction (PF1, 15 mL) was collected, diluted to 25 mL with DGB, laid on top of a sucrose cushion (30 mL of 42.5% sucrose in DGB) and centrifuged at 5,000 g for 2 h in a SW-32Ti rotor. After centrifugation, 40 mL of the volume was removed from the top and the remaining cushion (15 mL) was collected, diluted to 25 mL with DGB, laid on top of a second sucrose cushion of higher density (30 mL of 62.5% sucrose in DGB) and centrifuged at 5,000 g for 2 h in a SW-32Ti rotor. After centrifugation, 40 mL including the upper fraction (discernible by the green color) was discarded. The remaining cushion volume (SF2, 15 mL) was collected, diluted to 60 mL with DGB and centrifuged at 5,000 g for 16 h in a SW-32Ti rotor. The supernatant was discarded, and the pelleted SF2 was stored at −20 °C until further use.

Proteomics Analysis of Valve SDVs. The proteomics analysis of SF2 was done with two biological replicates each with two technical replicates. The pelleted SF2 fraction was resuspended in 100 µL of 2% SDS solution and incubated at 37 °C for 20 min. The sample was centrifuged for 1 h at 5,000 g in a fixed angle rotor, 50 ng BSA was added to the supernatant as internal standard for sample quantity, and the sample was stored at room temperature overnight. For gel electrophoresis, 48 µL [equivalent to 34 µg, quantified by the BCA method (55)] of sample was mixed with 12 µL 5× SDS sample buffer and heated at 85 °C for 2 min. The sample was separated on a 10% Bis-Tris SDS gel (two lanes per sample) at 180 V for 6 min resulting in separation distance of approximately 2 cm. Proteins were visualized by staining with Coomassie Brilliant Blue. Each lane was sliced into four areas of approximately 5 mm length. After in-gel digestion, the peptides were extracted from the gel slices, fractions were combined and analyzed by nanoLC-MS/MS. Digestion of proteins in-gel and the extraction, and the analysis of the peptides by LC-MS/MS was performed according to procedures described previously with slight modifications (56,57). In brief, gel pieces were destained and proteins were reduced with dithiothreitol and alkylated with iodacetamide. Digestion was performed with trypsin at a final enzyme concentration of 10 ng·µL^{-1} overnight in 10 mM NH_4HCO_3. Peptides were extracted, subsequently dried, and stored until analysis at −20 °C. For LC-MS/MS analyses, the peptides were recovered in 5 µL 30% formic acid, diluted with 20 µL of water and 5 µL was injected. NanoLC-MS/MS analyses were performed with a Q-Exactive HF mass spectrometer hyphenated to nanoflow LC system (Dionex3000 RSLC). Peptides were separated in a linear gradient of 0.1% aqueous formic acid for 120 min (eluent A) and 0.1% formic acid in 60% acetonitrile (eluent B), and the mass spectrometer was operated in data-dependent acquisition mode (TopN 10). Raw files were loaded into the Progenesis QIP V4.2 software (Nonlinear Dynamics) for peak picking. Peptide and protein identification was performed with Mascot V2.8 (Matrixscience) using the UniProt database UP000001449 (02/2022). Only protein hits with at least two significant peptide hits in every replicate were considered for further analysis. The proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD033943.

Phylogenetic Analysis. The phylogenetic distribution of proteins was carried out using the MMETSP transcriptome data (35) in particular the "_clean.fasta" files from which contaminants have been removed (software: https://github.com/kolecko007/mmetsp_cleanup, data: www.imicrobe.us/#/projects/104) (58). tBLASTn searches were carried out with Seg-based filtering of low complexity regions turned on, with default parameters. For further computational details see *SI Appendix*, Supplementary Text. tBLASTn searches were carried out with default parameters and Seg-based filtering of low complexity regions turned on. A custom Perl scrip that performs the searches and generate output tables suitable for plotting is available at https://github.com/skeffington/PhyloBlast_diatoms.

Fluorescence Microscopy. Epifluorescence images of subcellular fractions were acquired with a 63× oil objective on an Axiovert 200 inverted microscope equipped with a FITC (excitation 450 to 490 nm, emission 515 nm (longpass), Zeiss) and PDMPO (excitation 365 to 395 nm, emission 520 to 550 nm) filter set.

For live cell confocal microscopy, a 10 µL aliquot of a cell culture with a density of 5.0·10^5 cells·mL^{-1} was spotted onto a coverslip and overlaid with an agarose slice (1% in ASW medium). For image acquisition, a LSM780 inverted confocal microscope with a Plan-Apochromat 63× (1.4) oil DIC M27 objective (Zeiss) was used. The GFP and chlorophyll fluorescence were excited with a 488 nm laser line (power set to 1%), a MBS 488 beam splitter, and a 32-channel GaAsP spectral detector. To separate the GFP signal from chlorophyll fluorescence, two distinct channels were used [GFP (491 to 535 nm); chlorophyll (655 to 721 nm)].

Biosilica Preparation for Electron Microscopy. The preparation of diatom cell walls for electron microscopy was done according to the previously published method (15) with minor modifications. Briefly, a wild type or transformant strain of *T. pseudonana* was grown until a cell density of 10^6 mL^{-1}. For biosilica isolation, 300 mL of the cell culture were centrifuged at 2,500 g for 3 min. The cells were resuspended in 13 mL of 2% SDS and 100 mM EDTA (pH 8) and incubated at 60 °C for 1 h under constant shaking. The sample was pelleted at 2,500 g for 3 min and the pellet was washed three times with 13 mL of 10 mM EDTA (pH 8) by centrifugation (2,500 g for 3 min) and resuspension cycles. Subsequently, the sample was resuspended in 1 mL of 100% acetone, pelleted at 2,500 g for 3 min and washed four times with 13 mL of H_2O by centrifugation (2,500 g for 3 min) and resuspension cycles. Finally, the biosilica was resuspended in 1.5 mL of water and kept at 4 °C. For SEM, 1 mL of the biosilica was stepwise transferred from H_2O to 100% ethanol by centrifugation (2,500 g for 3 min) and resuspension cycles with incrementally increasing ethanol concentrations (20%, 40%, 60%, 80%, and 100%). The dehydrated biosilica was critical point dried and subsequently immobilized on a carbon pad that was attached to an aluminum stub. The sample was sputter coated with platinum and argon as the process gas at 40 mA for 40 s. The scanning electron micrographs were taken at an acceleration voltage of 5 kV and a lower secondary electron detector (JEOL JSM 7500F field emission SEM). For transmission electron microscopy, 500 µL of the biosilica was sonicated in an ultrasonic homogenizer using an MS72 sonotrode tip (0.1 kJ for 5 s). An 8 µL aliquot of the sonicated sample was spotted on a Formvar-coated gold finder grid for 30 min. The remaining liquid was removed with a piece of filter paper and the grid was washed with H_2O for 5 min. For drying, the majority of the H_2O was removed with filter paper and the grid was air-dried overnight. The transmission electron micrographs were taken at an acceleration voltage of 80 kV (JEOL JEM-1400 TEM).

Correlative Fluorescence and Electron Microscopy of Valve SDVs. The synchronization of the cells and the PDMPO labeling of the valve SDVs was performed as described previously (23). Briefly, cells were grown to a density of 5.0·10^5 mL^{-1} and subsequently transferred into Si-free ASW for 16 h (12 h dark, 4 h light). After the starvation period, Na_2SiO_3 was added to the culture at a final concentration of 200 µM, and the cells were incubated in constant light for 9 h. The cells were again subjected to a silicon starvation period for 16 h (12 h dark, 4 h light). Na_2SiO_3 was added to the cell culture at a final concentration of 200 µM. After 3 h, PDMPO (Aat Bioquest) was added to a final concentration of 1 µM for 10 min. For lysis, the cells were resupended in lysis buffer (50 mM Hepes·NaOH pH 7.5, 150 mM NaCl, 50 mM Sucrose, 100 µM PMSF) and vortexed in the presence of nitric acid-cleaned glass beads.

Valve SDVs were imaged with electron microscopy according to the previously published method (23). Briefly, the cell lysate was immobilized on a Formvar-coated gold finder grid (Science Service). The position of valve SDVs on the grid was identified by their PDMPO fluorescence. The grid was washed three times with water and air-dried overnight. The electron micrographs were taken at an acceleration voltage of 80 kV (JEOL JEM-1400 TEM). The locations of the valve SDVs were determined based on the corresponding fluorescence images.

Spectral Clustering. The previously annotated *sin1* cluster (orignally termed SiMat7 gene cluster) is based on the log2-transformed fold change in gene expression during Si shift-down and Si shift-up periods (22). Here, only the Si shift-up signal intensities were considered. For the partitioning of the mRNA microarray, we implemented the spectral clustering segmentation technique (59–61) by using the scikit-learn (v0.21.3)) python package (62). The expression profiles were normalized with the z-score method and the affinity matrix built using the radial basis function kernel ($\gamma = 1$). As a primary input parameter, the optimal number of clusters was determined based on the Davies–Bouldin index (DBI) criterion (63), where its lowest value indicates a better data partitioning. For the present transcriptomic dataset, the local minimum of DBI = 1.33 was observed at N = 11 (*SI Appendix*, Fig. S23*A*). Additionally, the number of *sin1*-associated members was analyzed as a function of clusters numbers (*SI Appendix*, Fig. S23*B*). For the latter, the stationary phase was observed after passing N = 11.

Analysis of Pore Densities and Pore Patterns. TEM images of valves were analyzed in a semiautomated fashion using a custom-made pore recognition algorithm. The python code for image processing was partially adapted from the skimage tutorial repository using python 3.7.4 and various packages [scipy, numpy, pandas and scikit-image (64)]. For mature cell walls, TEM images the margin of the valve was cropped out manually, and only pores within the cropped region were accounted for. We used automatic multilevel thresholding (Yen's method) to automatically determine a binary mask of putative pores. The pores segmentation was achieved by applying the Euclidean distance transform algorithm (EDT) to the obtained mask, the position of each pore was then determined as the local maximum of the EDT. Subsequently, a watersheding filter was implemented that allowed to determine the outline of each pore. The pore diameter was determined through the obtained pore circumference.

For the analysis of pores in nascent biosilica, a slightly different strategy was applied. In order to identify developing pores (that commonly are not completely engulfed by silica), the ridge detection filter was implemented. Determining the positions and shapes of identified pores was achieved with the same algorithm as for mature cell walls. The margin of the valve as well as the central annulus was excluded from the area of interest, and only pores between these boundaries were considered (*SI Appendix*, Fig. S22). Additionally, only the pores within a diameter size range of 15 to 40 nm were included in the subsequent analysis. To compute the PD, the N_p within the specified region of the valve was divided by the area of that region, $PD = N_p/area$. To compute the PDF, we first selected regions of interest (ROIs) of size 300 nm × 300 nm from the corrected binary mask of pores, centered at each recognized pore. Each ROI was then rotated such that the radial direction (i.e., annulus to valve margin) points along the new y-axis. All rotated ROIs were then averaged over all identified pores yielding the PDF for an individual valve.

Molecular Genetic Methods. The verification of the gene models for *dank1-3*, and *tp25735* by RACE PCR as well as the construction of *gfp* fusion genes and KO plasmids are described in *SI Appendix*, Supplementary Text.

Statistical Analysis. The arithmetic mean and the standard deviation across multiple biological replicas were used as measures of the sample average and spread. The number of replicates for each strain including the wild type is included in the respective figure legends, where appropriate. Unless otherwise noted, all statistical analyses were performed using the Python SciPy library (2.7). Comparisons between two groups were conducted using the independent two-sample *t* test while assuming equal variance. The comparison between two groups was performed with an equal sample size corresponding to the minimal sample size of the two groups. The obtained *P* values were used to determine the statistical significance between different clones of the same KO as well as different strains (different KOs and the wild type). While comparing the produced KOs with the wild type a value of $P < 0.05$ was considered statistically significant (the corresponding *P* values are mentioned in the *Results* section). Where appropriate, asterisks are used to denote the significance of differences with gray lines drawn between the compared groups. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Differences that were not statistically significant are denoted "ns." Unless otherwise noted, all figures were generated in Matplotlib 3.5 and Seaborn 0.11.

Data, Materials, and Software Availability. All proteomics data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD033943 (65). All data on the analysis of the pores in valve silica required to evaluate the conclusions are present in the Main Text and/or the *SI Appendix*. The raw data including images and the implemented code for pore analysis are available from Figshare LLC (66).

ACKNOWLEDGMENTS. We are grateful to Nicole Poulsen (TU Dresden) and André Scheffel (TU Dresden) for critically reading the manuscript. We wish to thank Jennifer Klemm (TU Dresden) and Anne Jaczkowski (TU Dresden) for help with diatom transformation, and Ariane Kühn (TU Dresden) and Tim Schell-Dieckel (TU Dresden) for assistance in PCR screening of *T. pseudonana* mutants. We thank Marc Gentzel and Kristin Eismann from the Molecular Analysis/Mass Spectrometry Facility (CMCB, TU Dresden) for LC-MS/MS and proteomics analysis. Expert support was provided by the Light Microscopy Facility and the Electron Microscopy Facility of the Center for Cellular and Molecular Bioengineering (TU Dresden). Portions of the paper were developed from the thesis of C.H. The work was supported by the Deutsche Forschungsgemeinschaft (DFG) through grant KR1853/8-2 (to N.K.) in the framework of Research Unit 2038 (NANOMEE), and through a "Physics of Life" Starting Grant under Germany's Excellence Strategy – EXC-2068 – 390729961– Cluster of Excellence Physics of Life of TU Dresden (to N.K. and B.M.F.). B.M.F. acknowledges support by the DFG through a Heisenberg grant (FR3429/4-1). The Molecular Analysis/Mass Spectrometry Facility was supported by grants from the German Federal Ministry of Education and Research (BMBF) program "Unternehmen Region" (grants #03Z2ES1, #03Z22EB1), the German Research Foundation (INST 269/731-1 FUGG) and the European Regional Development Fund (ERDF/EFRE) (Contract # 100232736).

Author affiliations: [a]B CUBE, Center for Molecular and Cellular Bioengineering, TU Dresden, 01307 Dresden Germany; [b]Cluster of Excellence Physics of Life, Technische Universität Dresden, 01062 Dresden, Germany; [c]Camborne School of Mines & Environment and Sustainability Institute, University of Exeter, Cornwall TR10 9FE, UK; [d]Center of Advancing Electronics Dresden, Technische Universität Dresden, 01062 Dresden, Germany; and [e]Faculty of Chemistry and Food Chemistry, Technische Universität Dresden, 01062 Dresden, Germany

1. D. Faivre, D. Schüler, Magnetotactic bacteria and magnetosomes. *Chem. Rev.* **108**, 4875–4898 (2008).
2. J. Sun, B. Bhushan, Hierarchical structure and mechanical properties of nacre: a review. *RSC Adv.* **2**, 7617–7632.
3. J. W. Goessling, Y. Su, M. Kühl, M. Ellegaard, "Frustule Photonics and light harvesting strategies in diatoms" in *Diatom Morphogenesis*, V. Annenkov, J. Seckbach, R. Gordon, Eds. (Scrivener Publishing, Beverly, 2022), pp. 269–300.
4. F. E. Round, R. M. Crawford, D. G. Mann, *The Diatoms: Biology and Morphology of the Genera* (Cambridge University Press, Cambridge, UK, 1990).
5. B. E. Volcani, "Cell wall formation in diatoms: Morphogenesis and biochemistry" in *Silicon and Siliceous Structures in Biological Systems*, T. L. Simpson, B. E. Volcani, Eds. (Springer, New York, NY, 1981), pp. 157–200.
6. N. Kröger, N. Poulsen, Diatoms-from cell wall biogenesis to nanotechnology. *Annu. Rev. Genet.* **42**, 83–107 (2008).
7. R. Gordon, R. W. Drum, The chemical basis of diatom morphogenesis. *Int. Rev. Cytol.* **150**, 243–372 (1994).
8. J. Parkinson, Y. Brechet, R. Gordon, Centric diatom morphogenesis: A model based on a DLA algorithm investigating the potential role of microtubules. *Biochim. Biophys. Acta* **1452**, 89–102 (1999).
9. L. Willis, E. J. Cox, T. Duke, A simple probabilistic model of submicroscopic diatom morphogenesis. *J. R. Soc. Interface* **10**, 1–9 (2013).
10. A. M. M. Schmid, "Wall morphogenesis in *Coscinodiscus wailesii* Gran et Angst. II. Cytoplasmic events of valve morphogenesis" in *Proceedings of the 8th International Diatom Symposium*, M. Ricard, Ed. (Koenigsteinpage, Koeltz Scientific Books, 1986), pp. 293–322.
11. M. Sumper, A phase separation model for the nanopatterning of diatom biosilica. *Science* **295**, 2430–2433 (2002).
12. L. Lenoci, P. J. Camp, Diatom structures templated by phase-separated fluids. *Langmuir* **24**, 217–223 (2008).
13. D. H. Robinson, C. W. Sullivan, How do diatoms make silicon biominerals? *Trends Biochem. Sci.* **12**, 151–154 (1987).
14. B. Tesson, S. J. L. Lerch, M. Hildebrand, Characterization of a new protein family associated with the silica deposition vesicle membrane enables genetic manipulation of diatom silica. *Sci. Rep.* **7**, 1–13 (2017).
15. S. Görlich, D. Pawolski, I. Zlotnikov, N. Kröger, Control of biosilica morphology and mechanical performance by the conserved diatom gene Silicanin-1. *Commun. Biol.* **2**, 1–8 (2019).
16. A. A. Trofimov *et al.*, Deep data analytics for genetic engineering of diatoms linking genotype to phenotype via machine learning. *npj Comput. Mater.* **5**, 1–8 (2019).
17. N. Poulsen, N. Kröger, Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* **279**, 42993–42999 (2004).
18. A. Kotzsch *et al.*, Biochemical composition and assembly of biosilica-associated insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* **291**, 4982–4997 (2016).
19. A. Kotzsch *et al.*, Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization. *BMC Biol.* **15**, 1–16 (2017).
20. A. W. Skeffington, Shedding light on silica biomineralization by comparative analysis of the silica-associated proteomes from three diatom species. *Plant J.* **110**, 1700–1716, 10.1111/tpj.15765.
21. R. P. Shrestha *et al.*, Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* **13**, 1–16 (2012).
22. T. Brembu, M. S. Chauton, P. Winge, A. M. Bones, O. Vadstein, Dynamic responses to silicon in *Thalasiossira pseudonana* - identification, characterisation and classification of signature genes and their corresponding protein motifs. *Sci. Rep.* **7**, 1–14 (2017).
23. C. Heintze *et al.*, An intimate view into the silica deposition vesicles of diatoms. *BMC Mater.* **2**, 1–15 (2020).
24. D. P. Yee, M. Hildebrand, M. Tresguerres, Dynamic subcellular translocation of V-type H^+-ATPase is essential for biomineralization of the diatom silica cell wall. *New Phytol.* **225**, 2411–2422 (2019).
25. R. K. Iler, *The Chemistry of Silica* (John Wiley & Sons, New York, NY, 1979).
26. R. P. Shrestha, M. Hildebrand, Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryot. Cell* **14**, 29–40 (2015).
27. N. M. Tsvetkova *et al.*, Small heat-shock proteins regulate membrane lipid polymorphism. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13504–13509 (2002).
28. J. D. Pickett-Heaps, D. H. Tippit, J. A. Andreozzi, Cell division in the pennate diatom *Pinnularia*. IV. Valve morphogenesis. *Biol. Cell.* **35**, 199–206 (1979).
29. 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**, 1–13 (2010).
30. S. Wenzl, R. Hett, P. Richthammer, M. Sumper, Silacidins: Highly acidic phosphopeptides from diatom shells assist in silica precipitation in vitro. *Angew. Chem. Int. Ed. Engl.* **47**, 1729–1732 (2008).
31. M. Sumper, E. Brunner, G. Lehmann, Biomineralization in diatoms: Characterization of novel polyamines associated with silica. *FEBS Lett.* **579**, 3765–3769 (2005).
32. A. J. Michael, Molecular machines encoded by bacterially-derived multi-domain gene fusions that potentially synthesize, N-methylate and transfer long chain polyamines in diatoms. *FEBS Lett.* **585**, 2627–2634 (2011).
33. L. K. Mosavi, T. J. Cammett, D. C. Desrosiers, Z. Peng, The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**, 1435–1448 (2004).
34. G. Bilcke *et al.*, Diurnal transcript profiling of the diatom *Seminavis robusta* reveals adaptations to a benthic lifestyle. *Plant J.* **107**, 315–336 (2021).
35. P. J. Keeling *et al.*, The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).

36. C. M. Osuna-Cruz *et al.*, The *Seminavis robusta* genome provides insights into the evolutionary adaptations of benthic diatoms. *Nat. Commun.* **11**, 1–13 (2020).
37. J. M. Chesnick, W. H. C. F. Kooistra, U. Wellbrock, L. K. Medlin, Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *J. Eukaryot. Microbiol.* **44**, 314–320 (1997).
38. A. Hopes, V. Nekrasov, S. Kamoun, T. Mock, Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* **12**, 1–12 (2016).
39. W. Herth, The site of beta-chitin fibril formation in centric diatoms. II. The chitin-forming cytoplasmic structures. *J. Ultrastruct. Res.* **68**, 16–27 (1979).
40. A. S. Masadeh, Total scattering atomic pair distribution function: New methodology for nanostructure determination. *J. Exp. Nanosci.* **11**, 951–974 (2016).
41. P. Bordet, Application of the pair distribution function analysis for the study of cultural heritage materials. *Comptes Rendus Phys.* **19**, 561–574 (2018).
42. D. A. Keen, Total scattering and the pair distribution function in crystallography. *Crystallogr. Rev.* **26**, 143–201 (2020).
43. N. Poulsen, M. Sumper, N. Kröger, Biosilica formation in diatoms: Characterization of native silaffin-2 and its role in silica morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12075–12080 (2003).
44. M. Sumper, E. Brunner, Learning from diatoms: Nature's tools for the production of nanostructured silica. *Adv. Funct. Mater.* **16**, 17–26 (2006).
45. A. Bernecker *et al.*, Tailored synthetic polyamines for controlled biomimetic silica formation. *J. Am. Chem. Soc.* **132**, 1023–1031 (2010).
46. N. Kröger, M. Sumper, Silica-formation in diatoms: The function of long-chain polyamines and silaffins. *J. Mater. Chem.* **14**, 2059–2065 (2004).
47. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
48. D. Qin, Z. He, P. Li, S. Zhang, Liquid-liquid phase separation in nucleation process of biomineralization. *Front. Chem.* **10**, 1–11 (2022).
49. S. Horvát *et al.*, Computational analysis of the effects of nitrogen source and *sin1* knockout on biosilica morphology in the model diatom *Thalassiosira pseudonana*. *Discov. Mater.* **1**, 8 (2021).
50. G. Rosengarten, J. W. Herringer, "Interactions of diatoms with their fluid environment" in *Diatom Nanotechnology*, D. Losic, Ed. (Royal Society of Chemistry, Cambridge, UK, 2018), pp. 14–54.
51. N. Sharma *et al.*, Diatoms biotechnology: Various industrial applications for a greener tomorrow. *Front. Mar. Sci.* **8**, 636613 (2021).
52. C. Gérardin, J. Reboul, M. Bonne, B. Lebeau, Ecodesign of ordered mesoporous silica materials. *Chem. Soc. Rev.* **42**, 4217–4255 (2013).
53. Z. A. Alothman, A review: Fundamental aspects of silicate mesoporous materials. *Materials* **5**, 2874–2902 (2012).
54. S. Jarmolińska, A. Feliczak-Guzik, I. Nowak, Synthesis, characterization and use of mesoporous silicas of the following types SBA-1, SBA-2, HMM-1 and HMM-2. *Materials* **13**, 1–33 (2020).
55. P. K. Smith *et al.*, Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85 (1985).
56. A. Vasilj, M. Gentzel, E. Ueberham, R. Gebhardt, A. Shevchenko, Tissue proteomics by one-dimensional gel electrophoresis combined with label-free protein quantification. *J. Proteome Res.* **11**, 3680–3689 (2012).
57. A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
58. M. Van Vlierberghe, A. Di Franco, H. Philippe, D. Baurain, Decontamination, pooling and dereplication of the 678 samples of the Marine Microbial Eukaryote Transcriptome Sequencing Project. *BMC Res. Notes* **14**, 1–4 (2021).
59. A. Y. Ng, M. I. Jordan, Y. Weiss, "On spectral clustering: Analysis and an algorithm" in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, NIPS01* (MIT Press, Cambridge, MA, 2001), pp. 849–856.
60. N. Speer, C. Spieth, A. Zell, "Spectral clustering gene ontology terms to group genes by function" in *International Workshop on Algorithms in Bioinformatics* (Springer, Berlin, Heidelberg, 2005), pp. 1–12.
61. A. J. Chin, A. Mirzal, H. Haron, Spectral clustering on gene expression profile to identify cancer types or subtypes. *J. Teknol.* **76**, 289–297 (2015).
62. F. Pedregosa *et al.*, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
63. D. L. Davies, D. W. Bouldin, A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 224–227 (1979).
64. S. Van Der Walt *et al.*, Scikit-image: Image processing in python. *PeerJ* **2**, e453 (2014).
65. M. Gentzel, N. Kröger, Proteomics of valve silica deposition vesicles from the diatom Thalassiosira pseudonana. PRIDE Proteomics Identifications Database. https://www.ebi.ac.uk/pride/archive/projects/PXD033943/private (Accessed May 5, 2022).
66. I. Babenko, Diatom pore pattern analysis. Figshare LLC. https://doi.org/10.6084/m9.figshare.20801881 (Accessed November 22, 2022).