Biominal-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 biominal 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 biominaleralization 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 biominaleralization 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.

Numerous organisms produce inorganic materials with amazingly complex morphologies and extraordinary properties in a process termed biominaleralization. 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 of 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 biominaleralization 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 cytoplasmic 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 a direct approach for identifying silica morphogenetic 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.
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. S1A, yellow arrow). In the lower fraction, PF2, cell wall fragments (SI Appendix, Fig. S1B, black arrow) were present along with valve SDVs that possessed strongly PDMPO fluorescent rings whose diameters were larger than that of valve SDVs in fraction PF1 (SI Appendix, Fig. S1B, 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. S1A, 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. S1D, 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 culture (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. 1A and S1 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. 1B 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 silacinids, which are abundant silica-associated phosphoproteins in diatoms (17, 30). The predicted S-adenosylmethylionine (SAM) synthetase might be required for biosynthesis of the polya mine-type modifications of silaffins (17) and the silica-associated long-chain polyanimes (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-silifying 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 Kryptoperidinium foliaceum (Myyzoa), which has a second nucleus of diatom origin (37). To highlight the presence of
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](https://doi.org/10.1073/pnas.2211549119)). In all cases, the disrupted *dank* genes encoded rather short polypeptides with largely altered amino acid sequences ([SI Appendix, Fig. S8](https://doi.org/10.1073/pnas.2211549119)). 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](https://doi.org/10.1073/pnas.2211549119)). 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](https://doi.org/10.1073/pnas.2211549119) 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](https://doi.org/10.1073/pnas.2211549119)). 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](https://doi.org/10.1073/pnas.2211549119)) rather than close to the ribs as in the wild type ([Fig. 2](#) *E* red arrowheads and [SI Appendix, Fig. S9](https://doi.org/10.1073/pnas.2211549119)). 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 A–C](https://doi.org/10.1073/pnas.2211549119) and [S12 A–C](https://doi.org/10.1073/pnas.2211549119)).

![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 J, K, and M and SI Appendix, Figs. S9 B and D, S10G, and S12G) (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 dankKO 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. 3A). The quantification of PD (Fig. 3B) confirmed the statistically significant decrease of cribrum pore densities on average (P = 3.14·10⁻¹⁴, independent two-sample t test) in the dank1KO (65.3 ± 24.0 µm⁻²) and the dank3KO (48.8 ± 13.9 µm²) strains (P = 8.27·10⁻²⁵, independent two-sample t test), respectively, compared to wild type (105.4 ± 22.3 µm⁻²) (Fig. 3B). In contrast, the cribrum PD in the dank2KO strains (147.2 ± 29.4 µm²) was ~40% increased (P = 1.32·10⁻¹¹, independent two-sample t test, Fig. 3B). 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.](https://www.pnas.org/content/119/49/e2211549119/Fig3a.png)

(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 closest 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. 3C). The presence of two satellite peaks reveals a preferred relative positioning of nearby pores (Fig. 3C). 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. 3C). The characteristic configuration of a central peak flanked by two satellite peaks is particularly pronounced in the dank1KO mutants (Fig. 3C 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 PDF’s left and right of the peaks, in contrast to the “trenches” visible in the PDF for dank1KO (Fig. 3C). In the dank3KO mutant, the height of the secondary peaks is reduced (Fig. 3C). 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. 3B). 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. 3B) and the PDFs in valves from the rescue clones (SI Appendix, Fig. S19) very closely matched the characteristics of wild-type valves (Fig. 3B and C).

In the dank2res strains, the distal tubes of the fultopoorthae 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. 4A). 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. 4B 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. 4C 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. 4D).

The cribrum plates are punctuated by double rows of pores, where each row runs adjacent to a rib (Fig. 4D). 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. 4A black lines) and a presumed polycationic surface charge on each nanodroplet (Fig. 4A, orange circles). Electrostatic repulsion between the nanodroplets would prevent droplet fusion and lead to their equidistant positioning along each side of the rib (Fig. 4A). As the ribs expand laterally, the inter-rib spaces become increasingly filled with silica except where the nanodroplets are located (Fig. 4B C). This scenario would perfectly explain the wave-like silica patterns at the rib boundaries during intermediate stages of cribrum plate development (Fig. 4B). 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 as in mature valves (SI Appendix, Fig. S22). This suggests, according to our model, that a substantial fraction of the nanodroplets disassembles at earlier 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. 4D 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. 4E F and F). Therefore, by the time the drop-}

Downloaded from https://www.pnas.org by UNIVERSITY OF STIRLING on June 27, 2023 from IP address 139.153.209.199.
along the ribs (Fig. 4K). 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. 4I 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. 4L, L', M, and M' and SI Appendix, Fig. S21). The strongly enhanced PD in dank2KO mutants (see Fig. 3B) 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. 5A). In this position, dAnks can only indirectly influence silica morphogenesis, which occurs exclusively in the lumen of the SDV membrane (Fig. 5A). 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. 5A), 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. 5B). 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⁺, Ca²⁺), 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 transporters).
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. 2F, 3B, and 4H). 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. 3B). 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. 3B) 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, 30). 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, sensors, 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 blastidicin 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 & Heimdahl clone CCMIP1335 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 stavarization–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 (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 supernant was discarded, and the pellet 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 µg 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 resuspended with diethiothreitol and alkylated with iodoacetamide. Digestion was performed with trypsin at a final enzyme concentration of 10 ng·µL⁻¹ overnight in 10 mM NH₄HCO₃. 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 QI/V2.4 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 MMTESP transcriptome data (35) in particular the _* clean.fasta_ files from which contaminants have been removed (software: https://github.com/koelko007/mmetesp_cleanup; data: www.imicrobe.us/#/projects/104) (58). iBLAST 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. iBLAST searches were carried out with default parameters and Seg-based filtering of low complexity regions turned on. A custom Perl script 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 5.0×10⁵ 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) and PDMPO (excitation 365 to 395 nm, emission 520 to 550 nm) filter set. For transmission electron microscopy, 500 µL of the biosilica was sonicated in an ultrasonic homogenizer using an M572 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 H₂O for 5 min. For drying, the majority of the H₂O 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 PDAMPO labeling of the valve SDVs was performed as described previously (23). Briefly, cells were grown to a density of 5.0×10⁵ mL⁻¹ and subsequently transferred into Si-free ASW for 16 h (12 h dark, 4 h light). After the starvation period, Na₂SiO₃ 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 then incubated into a silicon starvation period for 16 h (12 h dark, 4 h light). Na₂SiO₃ 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 resuspended 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 lysis was immobilized on a Formvar-coated gold finder grid (Science Service). The position of valve SDVs on the grid was identified by their PDAMPO 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 (originally 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 (γ = 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. S23A). Additionally, the number of sin1-associated members was analyzed as a function of clusters numbers (SI Appendix, Fig. S23B). 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 skinmage 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 watershedding 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, within the specified region of the valve was divided by the area of that region, PD = N/area. To compute the PDF, we first selected regions of interest (ROIs) of size 300 nm x 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 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).

Author affiliations: aB CUBE, Center for Molecular and Cellular Bioengineering, TU Dresden, 01307 Dresden Germany; bCluster of Excellence Physics of Life of TU Dresden (to N.K. and B.M.F.). B.M.F. acknowledges support 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 #0322ES1, #0322EB1), the German Research Foundation (INST 269/731-1 FUGG) and the European Regional Development Fund (ERDF/EFRE) (Contract #100232736).