

The molecular basis for pore pattern morphogenesis in diatom silica

Christoph Heintze

TU Dresden

Iaroslav Babenko

TU Dresden

Jirina Zackova Suchanova

TU Dresden

Alastair Skeffington

University of Exeter

Benjamin Friedrich

TU Dresden

Nils Kroeger (✉ nils.kroeger@tu-dresden.de)

TU Dresden <https://orcid.org/0000-0002-8115-4129>

Article

Keywords:

Posted Date: May 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1662246/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 **The molecular basis for pore pattern morphogenesis in diatom**
2 **silica**

3

4 Christoph Heintze^a, Iaroslav Babenko^b, Jirina Zackova Suchanova^a, Alastair
5 Skeffington^c, Benjamin M. Friedrich^{b,d}, Nils Kröger^{a,b,e}

6

7 ^a B CUBE, Center for Molecular and Cellular Bioengineering, TU Dresden, 01307 Dres-
8 den, Germany.

9 ^b Cluster of Excellence Physics of Life, TU Dresden, 01062 Dresden, Germany.

10 ^c Camborne School of Mines & Environment and Sustainability Institute, University of
11 Exeter, Cornwall TR10 9FE, UK

12 ^d Center of Advancing Electronics Dresden, TU Dresden, 01062 Dresden, Germany.

13 ^e Faculty of Chemistry and Food Chemistry, TU Dresden, 01062 Dresden, Germany.

14

15 **Email:** nils.kroeger@tu-dresden.de

16

17 **Author Contributions:** NK, CH, IB and JZS designed research; CH performed experi-
18 ments; all authors analyzed and interpreted data; IB and BMF contributed new data anal-
19 ysis tools; NK, CH and IB wrote the paper.

20

21 **Competing Interest Statement:** The authors declare no competing interest.

22

23

24

25

26

27

28

29

30

31

32

33 **Abstract**

34 Biomineral forming organisms produce inorganic materials with complex, genetically en-
35 coded morphologies that are unmatched by current synthetic chemistry. It is poorly under-
36 stood which genes are involved in biomineral morphogenesis and how the encoded pro-
37 teins guide this process. We addressed these questions using diatoms, which are para-
38 digms for the self-assembly of hierarchically meso- and macroporous silica under mild
39 reaction conditions. Proteomics analysis of the intracellular organelle for silica biosynthe-
40 sis led to identification of new biomineralization proteins. Three of these, dAnk1-3, are
41 largely specific to diatoms and contain a common protein-protein interaction domain
42 (ankyrin repeats), indicating a role in coordinating assembly of the silica biomineralization
43 machinery. Knocking out individual *dank* genes led to aberrations in silica biogenesis that
44 are consistent with liquid-liquid phase separation as underlying mechanism for pore pat-
45 tern morphogenesis. Our work provides an unprecedented path for the synthesis of tai-
46 lored meso- and macroporous silicas using Synthetic Biology.

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64 **Introduction**

65 Numerous organisms produce inorganic materials with amazingly complex morphologies
66 and extraordinary properties in a process termed biomineralization. Prominent examples
67 include the single domain magnetite nanocrystals of bacteria that act as sensitive mag-
68 netic field sensors¹, the nacreous calcium carbonate layers of mollusks with exceptionally
69 high fracture resistance², and the hierarchically porous, silica cell walls of diatoms with
70 intriguing photonic properties³⁻⁵. A fundamental understanding how genetically encoded
71 machineries are capable of establishing physical and chemical forces that drive morpho-
72 genesis of such intricate mineral structures is currently lacking. Therefore, unveiling the
73 mechanisms of biomineralization holds the promise of gaining advanced capabilities to
74 synthesize minerals with tailored properties using environmentally benign processes.

75 Diatoms are single-celled microalgae that produce silica-based cell walls charac-
76 terized by species-specific, hierarchical patterns of meso- and macropores, ribs, tubes,
77 and spines among others^{6,7}. A diatom cell wall is composed of two interlocking halves
78 each consisting of a plate- or dome-shaped valve and an array of ring-shaped girdle
79 bands. Each valve and each girdle band is produced intracellularly in a dedicated mem-
80 brane bound compartment termed the silica deposition vesicle (SDV)⁸. The valves are
81 more intricately patterned than the girdle bands, and thus have generally been the focus
82 of studies on silica morphogenesis. Previously, various models were put forward address-
83 ing the mechanism of valve silica morphogenesis in diatoms. The models differ depending
84 on whether they require an organic template or not. The template-free models assume
85 diffusion-limited aggregation^{9,10} or stochastic aggregation in combination with lateral inhi-
86 bition¹¹ of silica precursors inside the SDV. The template dependent models hypothesize
87 that biomolecular assemblies in the SDV lumen, in the SDV membrane, or in the surround-
88 ing cytoplasm direct silica morphogenesis^{8,12-15}. Only fairly recently gene inactivation stud-
89 ies in the model diatom *Thalassiosira pseudonana* identified the first proteins (SAP1,
90 SAP3, Sin1, Thaps3_21880) involved in diatom silica morphogenesis¹⁶⁻¹⁸. However, a
91 mechanistic understanding of their functions has not yet been achieved. Furthermore, the
92 silica morphology was largely intact in the mutants indicating that additional proteins and
93 possibly other components are involved in the morphogenesis process. In the past, the
94 discovery of proteins involved in silica morphogenesis was severely hampered by the lack
95 of a method for isolating SDVs. Therefore, biochemical analyses of diatom cell walls¹⁹⁻²²

96 and transcriptomics analyses of synchronized diatom cell cultures has been pursued^{23,24}.
97 However, both techniques are prone to reveal many (if not mainly) proteins that are unre-
98 lated to silica morphogenesis. To address this problem, we pursued here a direct approach
99 for identifying silica morphogenic proteins by establishing a method for isolating SDVs
100 from *T. pseudonana*, and perform proteomics characterization of the SDVs. Through sub-
101 sequent analysis of the silica development in loss-of-function mutants, we identified new
102 silica morphogenic proteins and gained insight into the mechanism for pore pattern for-
103 mation in diatom silica.

104

105

106 **Results**

107 ***Identification of valve SDV proteins***

108 We aimed to isolate SDVs by consecutive centrifugation steps in different high-density
109 media. Our efforts were focused on valve SDVs, because they can be specifically fluores-
110 cently labeled in synchronized diatom cells using the dye PDMPO allowing for their
111 straightforward identification in a cell lysate using epifluorescence microscopy²⁵. After cen-
112 trifugation through a 90% Percoll[®] containing cushion, valve SDVs were present in two
113 fractions. The upper fraction, PF1, contained valve SDVs with diameters < 3 μm and ho-
114 mogeneous PDMPO staining patterns (Supplementary Fig. 1A, yellow arrow). In the lower
115 fraction, PF2, cell wall fragments (Supplementary Fig. 1B, black arrow) were present along
116 with valve SDVs that possessed strongly PDMPO fluorescent rims whose diameters were
117 larger than that of valve SDVs in fraction PF1 (Supplementary Fig. 1B, yellow arrow).
118 Based on previous observations²⁵ the sizes of the SDVs and their PDMPO fluorescence
119 patterns indicated that PF1 contained valve SDVs at early to middle stages of develop-
120 ment, whereas the valve SDVs in PF2 were at late stages of development. Since a key
121 goal was to isolate SDVs free of cell walls we only processed PF1 further. Fluorescence
122 microscopy indicated the presence of plastid fragments in PF1 (Supplementary Fig. 1A,
123 white arrow) suggesting that also membranes from organelles other than plastids might
124 be present. To further purify the valve SDVs, they were centrifuged twice through buffers
125 with increasing sucrose concentration (first 42.5% then 62.5%). In both cases, the valve
126 SDVs accumulated near the bottom of the tube, and plastid fragments remained mainly in
127 the supernatant (Supplementary Fig. 1C, D). Nevertheless, the final valve SDV fraction,

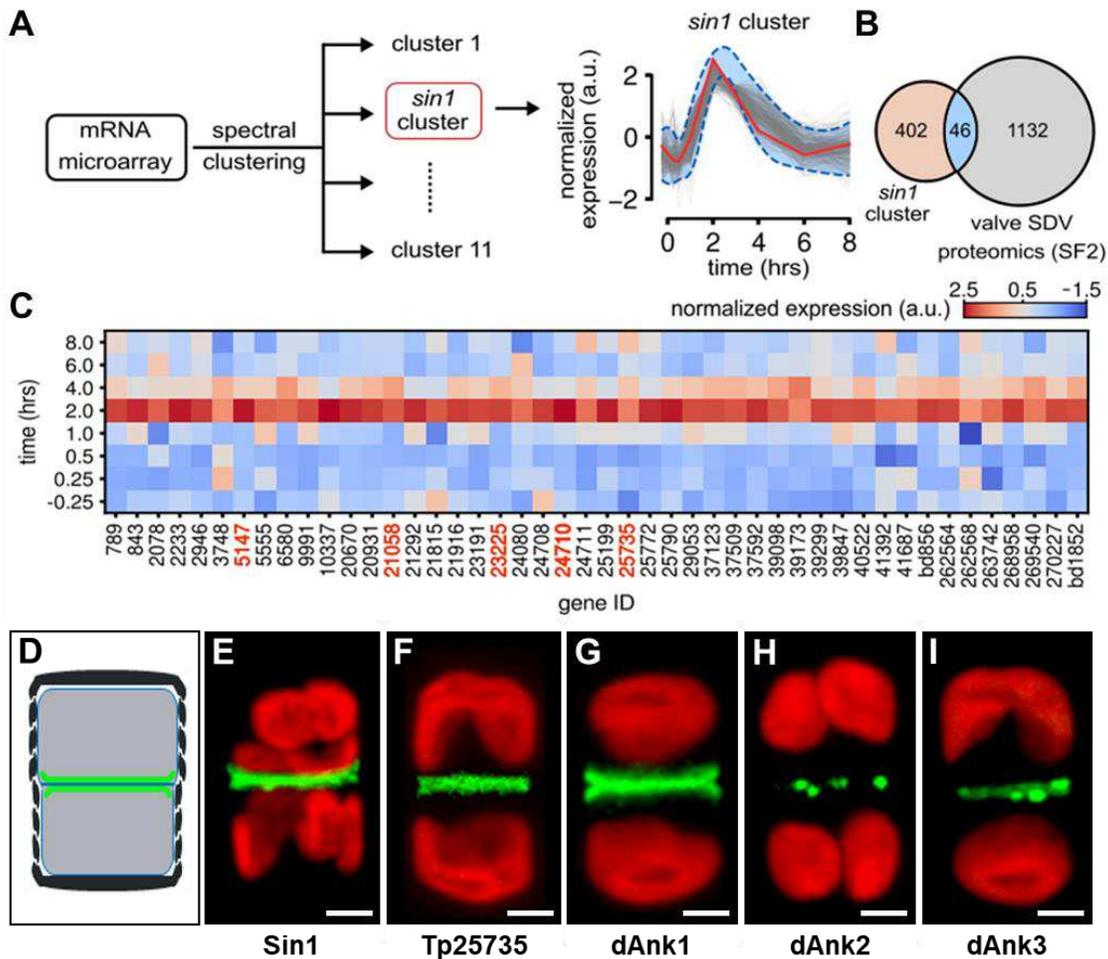
128 SF2, still contained some plastid fragments (Supplementary Fig. 1D, white arrow).

129 Proteomics analysis (two biological replicates each with two technical replicates)
130 revealed 1132 putative valve SDV proteins that were identified in two independent SDV
131 preparations by at least two peptide hits. Based on the annotation in the Universal Protein
132 Resource (UniProt) data 754 proteins shared similarities to proteins with known functions
133 and 378 proteins possessed no known or suspected function ('predicted' or 'uncharacter-
134 ized' protein; Supplementary Data 1). To narrow down the pool of candidate SDV proteins
135 for further analysis, we focused on those proteins, whose mRNA expression was upregu-
136 lated during valve SDV synthesis. For this we used the data from a previous transcriptom-
137 ics study on the expression of 6688 genes (total predicted protein coding genes in a syn-
138 chronized *T. pseudonana* cultures²⁴. The expression profile of the *sin1* gene, which en-
139 codes for an SDV protein^{21,24}, was used as a reference. In our analysis only the gene
140 regulation during the silicon replenishment period was considered, which is different to the
141 previous study on *sin1*-like gene expressions²⁴. A spectral clustering algorithm partitioned
142 the transcriptomics dataset into eleven clusters based on their expression profiles during
143 silicon replenishment (Fig. 1A, Supplementary Fig. 2), and the *sin1* cluster contained 402
144 genes (Supplementary Data 2). Only 46 of the genes in the *sin1* cluster encode putative
145 valve SDV proteins that were identified in the proteomics analysis (Fig. 1B, C, Supple-
146 mentary Data 3). After removing seven apparent contaminants (1 mitochondrial carrier, 3
147 plastid proteins, 1 histone, 1 UMP-CMP kinase, 1 endoplasmic reticulum ATPase), the
148 remaining 39 proteins were termed the Valve SDV proteome v1.0 (VSP1.0; Supplemen-
149 tary Table 1).

150 BLASTp searches of the VSP1.0 against the NCBI database revealed 20 proteins
151 with predicted functions, including the SDV protein Sin1 and seven subunits of the
152 H⁺-ATPase that was previously shown to be located in SDVs and involved in valve bio-
153 genesis²⁶. One protein is a predicted Na⁺/H⁺ antiporter that might be required for homeo-
154 stasis of pH and Na⁺, which are both critical factors in the chemistry of silica formation²⁷.
155 Six proteins were predicted to be involved in vesicle transport (3 clathrin coat proteins, 1
156 Arf, 1 Rab, 1 v-SNARE) and may be required for SDV biogenesis. The identified silicon
157 transporter was previously localized in the plasma membrane, but no information was pro-
158 vided whether it is also located in the SDV membrane²⁸. The predicted heat shock protein
159 might be involved modulating the properties of SDV membrane analogous to the influence
160 of heat shock proteins on the membrane fluidity of thylakoid membranes in *Synechocystis*

161 sp. PCC 6803²⁹. The presence of actin is consistent with the previously observed associ-
162 ation of microfilaments with valve SDVs^{30,31}. Two kinases are present that might catalyze
163 the phosphorylation of silaffins and silacidins, which are abundant silica associated phos-
164 phoproteins in diatoms^{19,32}. The predicted S-adenosylmethionine (SAM) synthetase might
165 be required for biosynthesis of the polyamine-type modifications of silaffins¹⁹ and the silica
166 associated long-chain polyamines^{19,33}, which both contain propylamine residues hypothe-
167 sized to be derived from SAM³⁴.

168 From the 19 VSP1.0 proteins with unknown functions, nine were previously anno-
169 tated as “SiMat7-like proteins” based on their domain structure, which consists of a N
170 terminal ER transit peptide, a large luminal domain with eight conserved cysteine residues,
171 a single transmembrane helix, and a short cytosolic domain at the C-terminus (Supple-
172 mentary Table 1; note that SiMat7 is the former name of the SDV protein Sin1)²⁴. Five of
173 the SiMat7-like proteins and an additional four VSP1.0 proteins are encoded by genes of
174 the “SiMat7 gene cluster”, which includes all genes that are downregulated during silicic
175 acid starvation and upregulated after silicic acid replenishment with an mRNA expression
176 peak during valve formation (Supplementary Table 1)²⁴. Three members of the SiMat7
177 gene cluster (Tp23225, Tp5147, Tp21058) encode proteins predicted to be located in the
178 cytosol, because they lack transmembrane domains and organelle targeting sequences.
179 They also share the presence of ankyrin repeat domains (Supplementary Fig. 3), which
180 are widely occurring protein-protein interaction motifs involved in numerous cellular func-
181 tions including protein interactions on the cytosolic surface of membranes³⁵. Interestingly,
182 transcriptomic studies in the diatom *Seminavis robusta* also implicated ankyrin repeat do-
183 main bearing proteins in cell wall formation³⁶, which suggest that these types of proteins
184 might be general components of the diatom machinery for silica biogenesis. We performed
185 a phylogenetic analysis of the three ankyrin repeat domain bearing VSP1.0 proteins
186 (Tp23225, Tp5147, Tp21058) using the Marine Microbial Eukaryotic Transcriptome Se-
187 quencing Project data base⁶¹, as well as the PLAZA diatoms data base⁶². This revealed
188 that Tp23225 is restricted to centric diatoms (Mediophyceae), whereas Tp5147 and
189 Tp21058 are very common and possibly ubiquitous among diatoms, with a phylogenetic
190 distribution similar to Sin1 (Supplementary Figure S4). Tp5147 has additional putative
191 homologues in a few other silica forming protists such as the chrysophyte *Paraphysomonas*
192 *bandaiensis* and in non-silicifying groups such as the Prymnesiophytes (Haptophyta).



193

194 **Fig. 1 Identification of valve SDV proteins.** (A) Scheme of the cluster analysis workflow
 195 that established eleven different gene expression profiles in the transcriptomics dataset
 196 from Brembu and co-workers (ref. 24). The graph shows the expression patterns in the
 197 *sin1* cluster. The *sin1* profile is highlighted in red, grey lines show the expression profiles
 198 of the other 401 genes, and the blue shaded area indicates the cluster average within 2
 199 standard deviations. (B) The Euler diagram shows the overlap between the proteins en-
 200 riched in the putative valve SDV proteins (Supplementary Data 1) and the genes of the
 201 *sin1* cluster (Supplementary Data 2). (C) Heat map of the expression patterns of the genes
 202 encoding the 46 proteins of the Valve SDV proteome v1.0 (VSP1.0). The gene expression
 203 patterns of *sin1* (gene ID: 24710), Tp25735, *dank1*, -2 and -3 (Tp23225, Tp5147,
 204 Tp21058) are highlighted in red. The y-axis indicates the time points before and after ad-
 205 dition of silicic acid, which triggers synchronous progression through the cell cycle. Valve
 206 SDV development starts 2.5 hours after addition of silicic acid²⁵ and is completed within
 207 90-120 min²¹. (D) Scheme of a dividing diatom cell in cross section during valve biogene-
 208 sis. Intracellular compartments other than valve SDVs are omitted for clarity. Black - cell
 209 wall, blue - plasma membrane, grey - protoplasm, green - valve SDV. (E-I) Confocal fluo-
 210 rescence microscopy images of individual dividing cells expressing the indicated proteins
 211 as GFP fusions (z-projection of optical slices through the mid cell region). The green color
 212 indicates GFP fluorescence and the red color results from chlorophyll autofluorescence of
 213 the chloroplasts. Scale bars: 2 μ m.

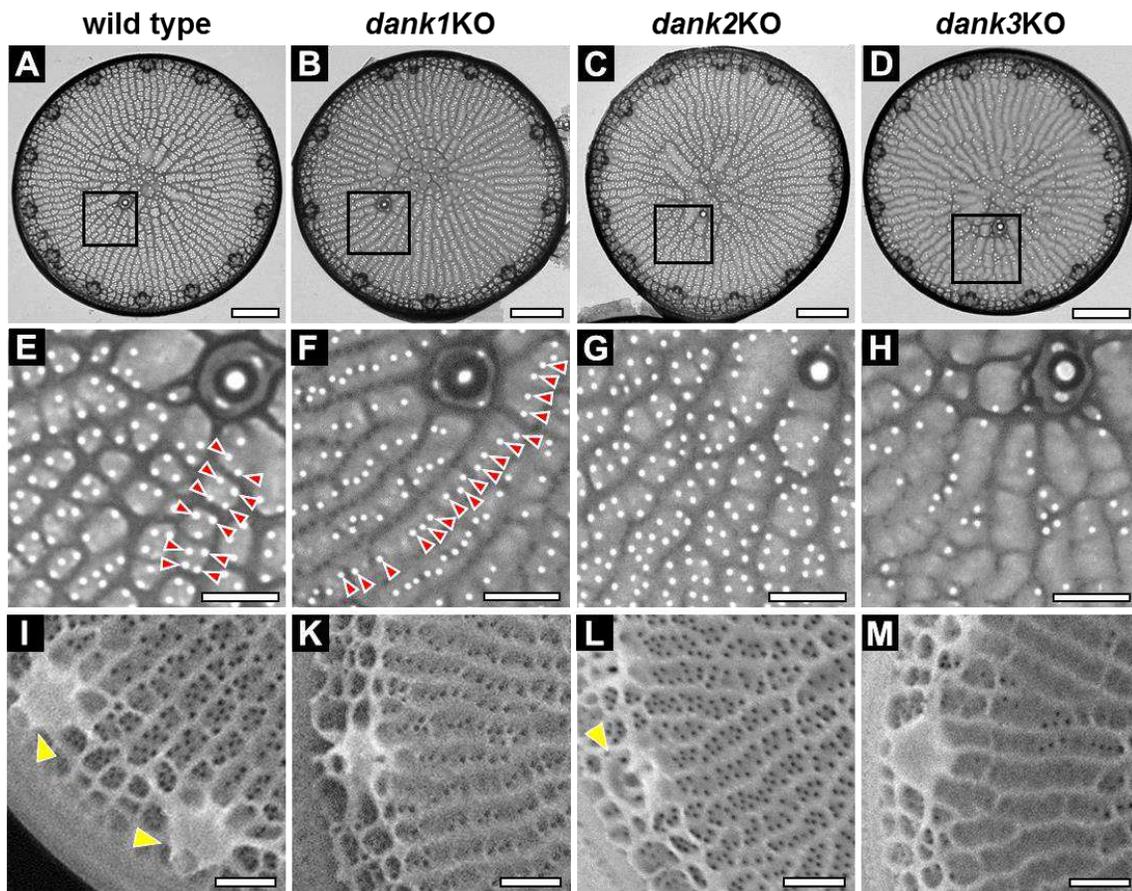
214 Some BLAST hits outside of the diatoms may be due to diatoms being prey for certain
215 protists (e.g. the ciliate, *Tiarina fusa*) or due to tertiary endosymbioses such as in *Kryto-*
216 *peridinium foliaceum* (Myzozoa), which has a second nucleus of diatom origin³⁷. To high-
217 light the presence of the ankyrin repeats ('Ank'), we named these VSP1.0 proteins dAnk1
218 (Tp23225), dAnk2 (Tp5147), and dAnk3 (Tp21058) with 'd' indicating their prevalence in
219 diatoms. The remaining six VSP1.0 proteins have not been implicated in silica biogenesis
220 and have never been studied before.

221 To start investigating whether the VSP1.0 proteins are associated with valve SDVs
222 *in vivo*, we selected for GFP tagging the predicted cytosolic proteins dAnk1-3 and the
223 predicted transmembrane protein Tp25735 (N-terminal ER transit peptide, three trans-
224 membrane domains; Supplementary Fig. 3). In dividing cells of *T. pseudonana* each of the
225 GFP fusion proteins was present at the site of the developing valve SDVs in the mid cell
226 region matching the location of Sin1-GFP (Fig. 1D-H, Supplementary Fig. 5). While the
227 GFP fusions of Tp25735 and dAnk1 were homogeneously distributed across the expand-
228 ing valve SDVs (Fig. 1E, F, Supplementary Fig. 5A, B), dAnk2-GFP and dAnk3-GFP dis-
229 played dotted patterns (Fig. 1G, H, Supplementary Fig. 5C, D). Altogether, the localization
230 experiments strongly support the assumption that Tp25735 and dAnk1-3 are *bona fide*
231 valve SDV proteins.

232

233 **Functional analysis of dAnk proteins in vivo**

234 Given the intracellular locations and sequence characteristics of dAnk1-3, we hypothe-
235 sized that they are part of a protein-protein interaction network at the cytoplasmic surface
236 of the valve SDV membrane that might be required for SDV function. To test this hypoth-
237 esis, we generated independent knockout (KO) strains of *T. pseudonana* each lacking an
238 individual *dank* gene using the CRISPR/Cas9 method previously established for this or-
239 ganism^{17,38}. For generating the KO strains, three plasmids were designed each encoding
240 Cas9 and three or four guide RNAs (gRNA) specific to one *dank* gene (Supplementary
241 Fig. 6). The KO plasmids were introduced into *T. pseudonana* wild type by biolistic particle
242 bombardment. Sequencing the products from genomic PCRs specific for each targeted
243 gene was used to identify independent transformant strains in which the targeted *dank*
244 gene contained a frameshift mutation (Supplementary Fig. 7). In all cases the disrupted
245 *dank* genes encoded rather short polypeptides with largely altered amino acid sequences
246 (Supplementary Fig. 8). Electron microscopy analyses of the isolated biosilica from each



247

248 **Fig. 2 Valve morphologies in *T. pseudonana* wild type and *dank* knockout mutants.**
 249 (A-D) TEM images of entire valves, and (E-H) details corresponding to the boxed areas in
 250 (A-D). Red arrowheads point to cribrum pores. (I-M) SEM images of details of valves.
 251 Yellow arrowheads point to the distal tubes of fultoportulae. Scale bars: 1 μm (A-D), 300
 252 nm (E-M).

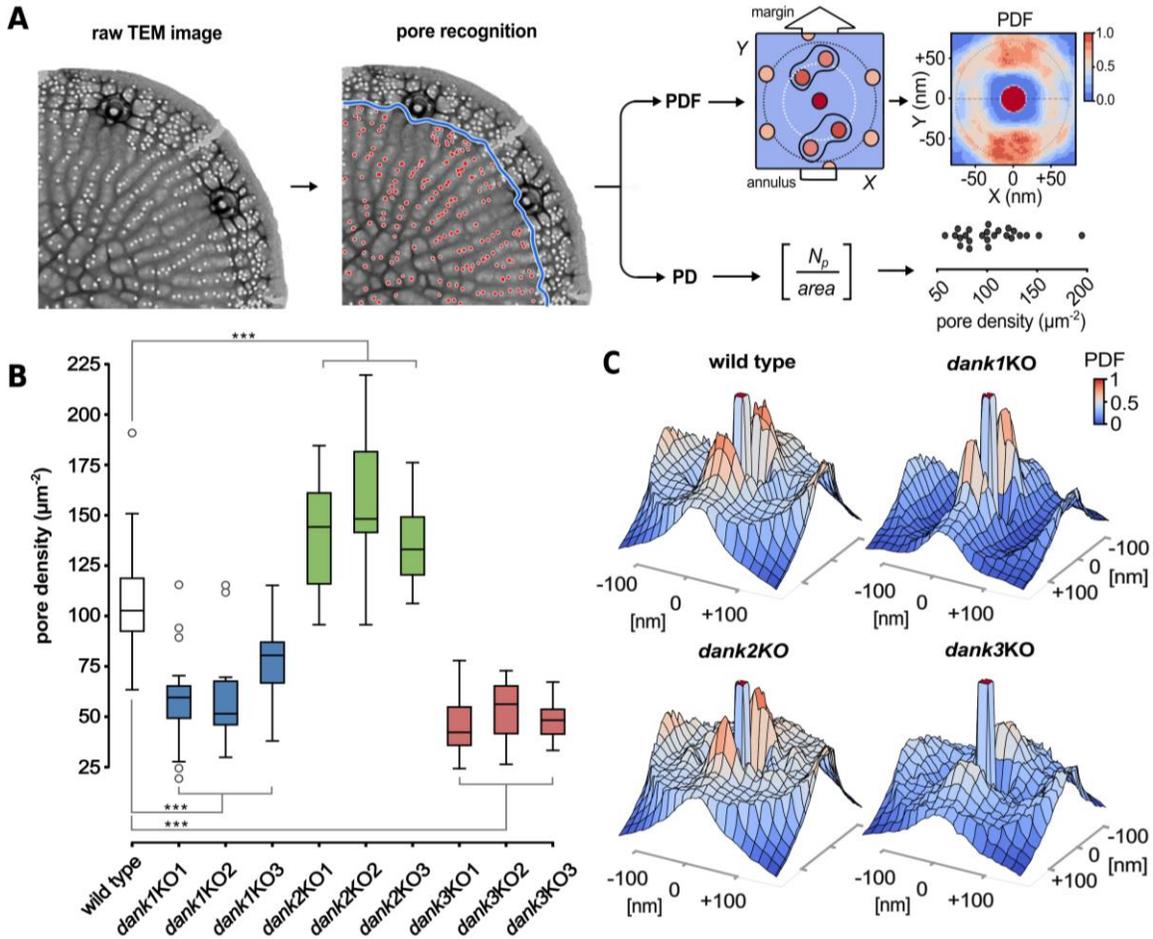
253

254

255 knockout strain revealed differences to the wild type morphology in the valves (Fig. 2,
 256 Supplementary Fig. 9-12). The valve of *T. pseudonana* has a circular shape and contains
 257 a network of ridges composed of branched radial ribs and bridges between them (Fig. 2A,
 258 E, I; see Supplementary Fig. 9 for details). The space delineated by the ridge network is
 259 filled with a layer of silica, called cribrum plate, which contains numerous circular openings
 260 26 ± 4 nm in diameter (Fig. 2E), called cribrum pores. The circular shape of the valve and
 261 the ridge network appeared to be unchanged in all knockout strains (Fig. 2B-D, Supple-
 262 mentary 10A-C, 11A-C, 12A-C). In the *dank1KO* strains the pattern of cribrum pores ap-
 263 peared to be different with the pores positioned predominantly in the center between two
 264 neighboring ribs (Fig. 2F red arrowheads, Supplementary Fig. 10D-F) rather than close to

265 the ribs as in the wild type (Fig. 2E red arrowheads, Supplementary Figure 9C). In the
266 *dankKO2* strains and the *dank3KO* strains the pore densities appeared to be enhanced
267 and reduced, respectively (Fig. 2G, H, Supplementary Fig. 11D-F, 12D-F). Scanning elec-
268 tron microscopy (SEM) analysis revealed that dAnk2 but neither dAnk1 nor dAnk3 is also
269 involved in fultoportula morphogenesis. Fultoportulae are specialized pore systems that
270 are predominantly located near the rim of the valve and required for the secretion of chitin
271 fibers³⁹. At the distal surface a fultoportula is covered by a funnel shaped tube (Fig. 2I, K,
272 M, Supplementary Fig. 9B, D, 10G, 12G)²², and in *dank2KO* mutants the distal tubes are
273 missing in all fultoportulae (Fig. 2L, Supplementary Fig. 11G-M).

274 To evaluate the abundance and patterns of cribrum pores in wild type and *dankKO*
275 strains, we developed an image analysis algorithm for cribrum pore recognition, allowing
276 the quantitative description of observed valve morphology in a semi-autonomous fashion
277 (Fig. 3A). To describe the pore pattern arrangement in wild type and mutants, we intro-
278 duced two morphological features: pore density (PD) and pore distribution function (PDF)
279 (Fig. 3A). The quantification of PD (Figure 3B) confirmed the statistically significant de-
280 crease of cribrum pore densities on average ($p = 3.14 \cdot 10^{-14}$, independent two-sample t-
281 test) in the *dank1KO* ($65.3 \pm 24.0 \mu\text{m}^{-2}$) and the *dank3KO* ($48.8 \pm 13.9 \mu\text{m}^{-2}$) strains ($p =$
282 $8.27 \cdot 10^{-25}$, independent two-sample t-test), respectively, compared to wild type ($105.4 \pm$
283 $22.3 \mu\text{m}^{-2}$) (Fig. 3B). In contrast, the cribrum pore density in the *dank2KO* strains
284 ($147.2 \pm 29.4 \mu\text{m}^{-2}$) was ~40% increased ($p = 1.32 \cdot 10^{-11}$, independent two-sample t-test,
285 Fig. 3B). The PDF is a probability distribution function of finding the closet adjacent pore
286 with respect to a reference pore position, and was inspired by the pair distribution function,
287 a standard tool in condensed matter physics⁴⁰⁻⁴². By default, the centered peak of the PDF
288 marks the position of the reference pore (Fig. 3C). The presence of two satellite peaks
289 reveals a preferred relative positioning of nearby pores (Fig. 3C). The PDF is calculated
290 in such a way that the orientation of satellite peaks reflects the positioning of pores in rows
291 parallel to the radially aligned ribs, while the distance of the satellite peaks from the central
292 peak characterizes the typical distance between neighboring pores ($60 \pm 18 \text{ nm}$) in these
293 rows (Fig. 3C). The characteristic configuration of a central peak flanked by two satellite
294 peaks is particularly pronounced in the *dank1KO* mutants (Fig 3C, Supplementary Fig.
295 13B). This reflects the observation that the *dank1KO* strain predominantly features only a
296 single row of pores between neighbouring ribs. In wildtype, *dAnk2KO* and *dAnk3KO*
297 strains the row of pores corresponding to the peaks is flanked by a second row of pores



299 **Fig. 3 Analysis of the pore patterns in valve silica from wild type and *dank* knockout**
 300 **mutants. (A)** Recognition of cribrum pores from a TEM image of an individual valve. The
 301 girdle band margin is cropped from the analysis and only pores in the area underneath the
 302 blue line are analysed. The recognition algorithm enables measurement of the pore den-
 303 sity (PD) and the pore distribution function (PDF). **(B)** Statistical analysis of the pore den-
 304 sities from wild type and the *dank* knockout strains. Data shown in B as: central line, me-
 305 dian value; box edges, 25th and 75th percentiles; whiskers, minimum and maximum val-
 306 ues; white circles sample outliers. Number of analysed valves for each strain are 50,18,
 307 14, 18, 19, 19, 12, 20, 17, 13 listed in order of the x-axis labeling (independent two-sample
 308 t-test ***p < 0.001). **(C)** Averaged PDF functions for the *dank* knockout strains within a
 309 domain of 300 nm x 300 nm. The central peak marks the position of the reference pore,
 310 while the two satellite peaks characterize the relative position of the closest neighbors.
 311 The colourbar indicates normalized probabilities of finding a neighbouring pore.

312

313

314 either left or right of the first row. After averaging, this gives intermediate values of the
 315 PDFs left and right of the peaks, in contrast to the “trenches” visible in the PDF for
 316 *dAnk1KO* (Fig. 3C). In the *dank3KO* mutant, the height of the secondary peaks is reduced

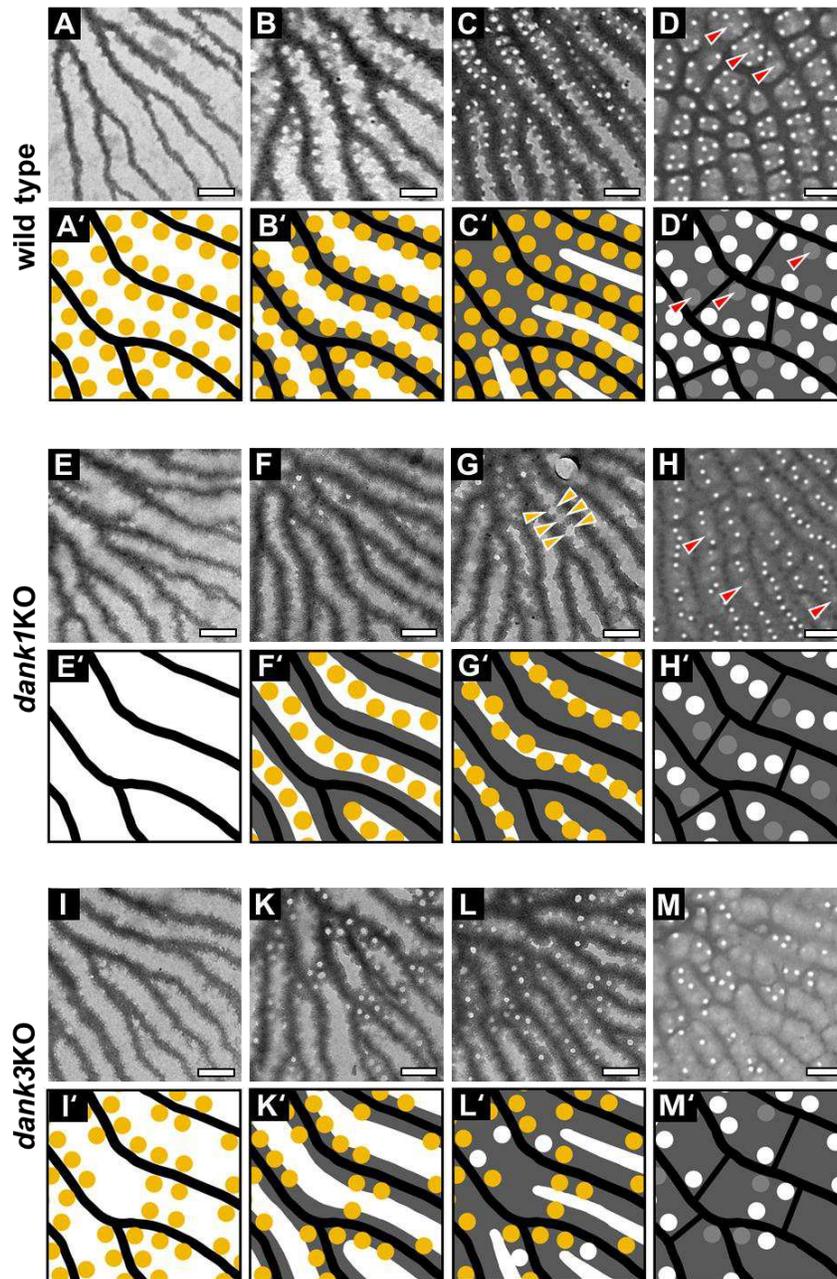
317 (Fig. 3C. This is consistent with a configuration where pores are still arranged with equi-
318 distant spacing along rows, but some positions in these rows are skipped (as reflected by
319 the lower pore density in the *dAnk3*KO strain; Fig. 3B). Thus, the PDF analysis confirmed
320 that the patterns of cribrum pores are significantly altered in *dank1*KO and *dank3*KO
321 strains. We also generated double knockout strains, termed *dank1/3dKO*, which lacked
322 both the *dank1* and the *dank3* gene (Supplementary Fig. 14, 15). The valves from three
323 independent *dank1/3dKO* strains were largely devoid of cribrum pores ($15.1 \pm 11.7 \mu\text{m}^2$)
324 exhibiting only 14% of the pore density of wild type valves (Supplementary Fig. 16). Alto-
325 gether, the data from pore analysis of the single and double knockout strains strongly
326 suggest that the dAnk proteins are involved in both the formation and patterning of cribrum
327 pores in *T. pseudonana*.

328 To corroborate that the observed morphological phenotypes in the knockout mu-
329 tants were generated by inactivation of the targeted *dank* genes rather than off-target ef-
330 fects, we attempted to restore the wild type phenotype by introducing an intact copy of the
331 *dank* gene into the corresponding knockout mutant. The introduced *dank* copy was mu-
332 tated to lack the recognition sites for the guide RNAs without altering the encoded amino
333 acid sequence (Supplementary Fig. 17) and was fused in frame to the *egfp* gene to facili-
334 tate the identification of transformant strains that express the re-introduced *dank* gene.
335 Transformant strains that exhibited GFP fluorescence were termed “rescue clone” and
336 denoted *dankXresN* with *X* and *N* specifying the introduced *dank* gene and the number of
337 the independent clone, respectively. Confocal fluorescence microscopy confirmed proper
338 sorting of each dAnk-GFP fusion protein to the valve SDVs in the cells of the rescue strains
339 (Supplementary Fig. 18). Analysis by our automated cribrum pore recognition algorithm
340 revealed that the pore densities (Supplementary Fig. 19A) and the PDFs in valves from
341 the rescue clones (Fig. 19B) very closely matched the characteristics of wild type valves
342 (Fig. 3B, C). In the *dank2res* strains the distal tubes of the fultuportulae were also restored
343 (Supplementary Fig. 20). Altogether, these data clearly confirmed that the lack of the dAnk
344 proteins caused the altered valve morphotypes in the *dank* knockout strains.

345 To gain insight into the mechanism by which dAnk proteins influence the morpho-
346 genesis of the cribrum pore patterns, we performed TEM analysis of nascent valve SDVs.
347 As described previously for the wild type²⁵, radially oriented, branched silica ribs develop
348 first (Fig. 4A). As ribs become wider, expanding laterally into the inter-rib space, their
349 boundaries attained a corrugated, wave-like shape (Fig. 4B). Each trough of this wave

350 pattern is the origin of a cribrum pore as it remains silica free during silica deposition in
351 the inter-rib spaces (Fig. 4C). Eventually, in the mature valve, the space between the ribs
352 is completely filled with a layer of silica (i.e. the cribrum plate) with occasional silica bridges
353 between neighboring ribs. The cribrum plates are punctuated by double rows of pores,
354 where each row runs adjacent to a rib (Fig. 4D). The observed development of the pore
355 pattern is consistent with a liquid-liquid phase separation process that generates
356 nanodroplets in the inter-rib spaces. We hypothesize that the droplets align along the silica
357 ribs due to electrostatic interactions between the negatively surface of the rib silica (Fig.
358 4A' black lines) and a presumed polycationic surface charge on each nanodroplet (Fig.
359 4A', yellow circles). Electrostatic repulsion between the nanodroplets would lead to their
360 equidistant positioning along each side of the rib (Fig. 4A'). Subsequently, the ribs expand
361 laterally and the inter-rib spaces becomes increasingly filled with silica except where the
362 nanodroplets are located (Fig. 4B'-C'). This scenario would perfectly explain the wave-like
363 silica patterns at the rib boundaries during intermediate stages of cribrum plate develop-
364 ment (Fig. 4B). We noticed that the density of developing cribrum pores in immature valves
365 was about twice as high than in mature valves (Supplementary Fig. 21). This suggest,
366 according to our model, that a substantial fraction of the nanodroplets disintegrates at later
367 stages while silica deposition is still occurring. As a consequence such initial pores be-
368 come filled in with silica. This hypothesis is consistent with the observation that mature
369 valves contain many round spots that are less electron dense than the surrounding crib-
370 rum plate silica and are located in positions where cribrum pores would be expected (Fig.
371 4D, D' red arrowheads).

372 Using the nanodroplet-based model it is possible to explain the altered cribrum
373 pore patterns in the valves of *dank* knockout mutants. In developing valves lacking *dank1*
374 the widening of the ribs appears to proceed faster than the development of the
375 nanodroplets (Fig. 4E, E', F, F'). Therefore, by the time the droplets appear, the silica-free
376 space between the ribs has narrowed down so much that it can accommodate only a
377 single row of nanodroplets (Fig. 4F, F'). As a result, single rows of cribrum pores, each



378

379 **Fig. 4 Morphogenesis of porous silica patterns in valves from *T. pseudonana* wild**
 380 **type, *dank1KO*, *dank2KO*, and *dank3KO* strains.** TEM images show representative sub
 381 areas of nascent valve silica from (A-D) wild type, (E-H) *dank1KO*, and (I-M) *dank3KO* in
 382 different developmental stages proceeding from left to right. Scale bars: 300 nm. Based
 383 on these TEM data and assuming liquid-liquid phase separation to occur in the SDV^{13,14},
 384 we propose models for cribrum pore formation in valves from (A'-D') wild type, (E'-H')
 385 *dank1KO*, and (I'-M') *dank3KO* strains. Black - silica ribs, orange - organic nanodroplets,
 386 grey - cribrum plate silica. Red arrowheads point to less electron dense spots assumed to
 387 originate from unstable cribrum pores that were filled in with silica. Yellow arrowheads
 388 indicate forming cribrum pores.

389 roughly equidistant from the neighboring ribs, emerge as a dominant feature in the valve
390 silica of *dank1KO* mutants (Fig. 4G yellow arrowheads, 4G'). Also in mature valves of
391 *dank1KO* mutants spots with reduced electron densities were observed (Fig. 4H red ar-
392 rowheads) indicating that occasional disintegration of nanodroplets is also occurring in
393 these mutants (Fig. 4H'). TEM analysis of valve development in *dank3KO* mutants pro-
394 vided further support that nanodroplet stability is an important factor in pore pattern mor-
395 phogenesis. The early stages of valve development in these mutants are similar to wild
396 type with wave-like silica boundaries developing at the onset of lateral expansion of the
397 ribs (Fig. 4I, K). However, these wave-like patterns fail to develop continuously along the
398 ribs (Fig. 4K). We hypothesize that this is caused by a decreased stability of the
399 nanodroplets leading to their accelerated disintegration compared to the nanodroplets in
400 the wild type (Fig. 4I', K'). As silica deposition in inter-rib spaces progresses, the disinte-
401 gration of a significant amount of nanodroplets continues, resulting in pore-free regions,
402 which are substantially larger than in wild type (Fig. 4L, L', M, M'). The strongly enhanced
403 pore density in *dank2KO* mutants (see Fig. 3B) can also be explained by the phase-sep-
404 aration model by assuming that the stability of the nanodroplets is increased compared to
405 wild type. As a consequence, less nanodroplets would disintegrate during valve develop-
406 ment resulting in a higher number of pores in the mature valve silica.

407

408

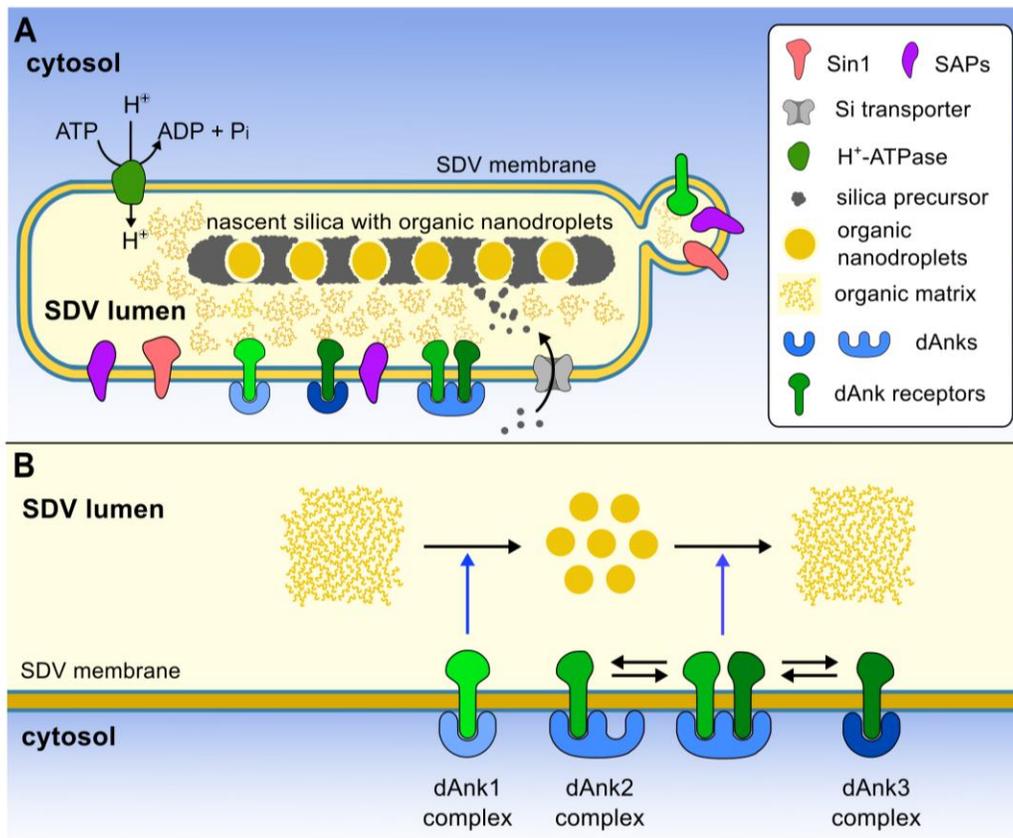
409 **Discussion**

410 In the present work, we performed the first call wall-free enrichment and proteomics anal-
411 ysis of SDVs, which led to the identification of new silica biomineralization proteins. These
412 include dAnk1-3, which are involved in the morphogenesis of pore patterns in biosilica.
413 dAnk-like proteins appear to be largely restricted to the diatoms and widespread, if not
414 ubiquitously present in this group of organisms. Functional analysis of dAnk1-3 in the
415 model diatom *T. pseudonana* provided support for a liquid-liquid phase separation based
416 model for pore pattern morphogenesis. In our model, the cribrum pore patterns are tem-
417 plated by nanodroplets whose biogenesis is strongly influenced by dAnks. dAnk1 is in-
418 volved timing the onset of nanodroplet biogenesis relative to the onset of silica deposition
419 in the inter-rib spaces (see Fig. 4). dAnk2 and dAnk3 are mutual antagonists that influence
420 the lifetime of the nanodroplets as outlined below. *In vivo* evidence for the existence of
421 nanodroplets is currently missing. However, *in vitro* studies have demonstrated that silica

422 associated biomolecules from diatoms (silaffin proteins, long-chain polyamines) and syn-
423 thetic analogies thereof undergo spontaneous liquid-liquid phase separation *in vitro* under
424 biologically relevant conditions^{43–46}.

425 The association of dAnks with developing valves and the absence of currently
426 known organelle targeting motifs in their polypeptide sequences, predicts dAnks to be lo-
427 cated on the cytoplasmic surface of valve SDVs (Fig. 5A). In this position, dAnks can only
428 indirectly influence silica morphogenesis, which occurs exclusively in the lumen of the
429 SDV membrane (Fig. 5A). As detailed in the following, we propose here that dAnks influ-
430 ence biogenesis of the silica pore patterns through interactions with specific SDV trans-
431 membrane proteins (Fig. 5A).

432 According to our model, dAnk1 binds to the cytosolic domain of a transmembrane
433 protein (dAnk1 receptor) involved in nucleating liquid-liquid phase separation (LLPS)
434 throughout the SDV lumen thereby the nanodroplets that generate the pore pattern in
435 valves (Fig. 5B). The immediate trigger for nucleating LLPS might be the import of a drop-
436 let component, a pH shift, or other changes in the physico-chemical conditions in the SDV
437 lumen that enhance the aggregation competence organic matrix components to form con-
438 densates. Even in the absence of dAnk1, the dAnk1 receptor might have an inherent ability
439 to trigger LLPS but with significantly reduced efficiency. This would explain the presence
440 of a reduced number of pores in the valves of *dank1KO* mutants (Fig. 3B, Fig. 4E-H). Our
441 model proposes further that the nanodroplet formation can be reversed by heteromeric
442 clusters of transmembrane proteins generated by interaction with dAnk2 (Fig. 3B). The
443 binding of dAnk2 to two different transmembrane proteins is supported by the presence of
444 two ankyrin repeat domains, while dAnk3 and dAnk1 have only one such protein-protein
445 interaction domain (Supplement Fig. 3). In *dank2KO* mutants such protein clusters would
446 not be able to form resulting in a stabilization (i.e. reduced destabilization) of the
447 nanodroplets compared to wild type leading to a higher pore density in the valves, which
448 was observed in our experiments (Fig. 3B). We hypothesize that in wild type dAnk3 com-
449 petes with dAnk2 for binding to a subset of the same transmembrane proteins thereby
450 reducing the number of heteromeric protein clusters that are able to destruct nanodroplets
451 (Fig. 5B). According to this scenario, nanodroplet stability would be determined by the
452 position of the equilibrium between the dAnk2 complex and the dAnk3 complex (Fig. 5B).
453 We hypothesize that competition between dAnk2 and dAnk3 occurs only within confined
454 regions of the SDV surface thereby generating “hot spots” for nanodroplet destruction. It



455

456 **Fig. 5 Model for dAnk-controlled phase separation in the SDV** (components are not
 457 drawn to scale). **(A)** Current model of an SDV. Biogenesis of the SDV is assumed to occur
 458 via the fusion of transport vesicles that provide membrane material and organic compo-
 459 nents for the SDV lumen³⁰. The SDV lumen is assumed to contain an organic matrix
 460 composed of strongly charged biomacromolecules (e.g., zwitterionic silaffins, polyanionic sila-
 461 cidins, polycationic LCPA) that are prone to undergo phase separation^{13,32,43}. The SDV
 462 membrane contains transmembrane proteins Sin1, SAP1 and SAP3 that might interact
 463 with LCPA and silaffins in the SDV lumen^{16,21}. A V-type H⁺-ATPase establishes and acidic
 464 pH in the SDV lumen²⁶. The chemical structure of the silica precursor molecules and the
 465 Si-transporter that catalyzes their import into the SDV are yet unknown. **(B)** dAnk1 (light
 466 blue symbol) may bind to a SDV transmembrane protein (light green symbol) enhancing
 467 its ability to promote liquid-liquid phase separation in the SDV lumen. dAnk2 (blue symbol)
 468 induces the formation of clusters of heteromeric transmembrane proteins (green and dark
 469 green symbols) that are restricted locally and promote disintegration of nanodroplets in
 470 many local “hot spots” in the SDV. dAnk3 (dark blue symbol) competes with dAnk2 for
 471 binding to a subset of the same transmembrane proteins (dark green symbol) thereby
 472 reducing the number of heteromeric protein clusters capable of destructing nanodroplets.

473

474

475 is unknown how the locally restricted competition between dAnk2 and dAnk3 for the same
 476 receptor molecule is achieved, but it is consistent with the spot-like distribution of both

477 proteins at the SDV (Fig. 1H, I, Supplement Fig. 5D, E, Supplement Fig. 18D-G). In con-
478 trast, dAnk1 appears to be homogeneously distributed at the SDV, which is consistent with
479 its presumed role in generating LLPS throughout the SDV lumen (Fig. 1G, Supplement
480 Fig. 5 C, Supplement Fig. 18B, C).

481 A droplet-based model for morphogenesis of the hexagonally arranged pores in
482 the silica of diatoms from the genus *Coscinodiscus* has been previously proposed¹³. How-
483 ever, the model assumed that pore pattern morphogenesis is solely guided by close pack-
484 ing of the droplets while the SDV membrane has an entirely passive role as compartment
485 boundary. In contrast, the results of the present study demonstrate that the SDV mem-
486 brane has to play an active role in pore pattern morphogenesis by transducing the mor-
487 phogenic actions of dAnks from its cytoplasmic surface into the SDV lumen. To date, the
488 interaction partners of dAnk1-3 are unknown. We noted that *sin1KO* mutants exhibited a
489 ~35% increase in cribrum pore density similar to the *dank2KO* mutants⁴⁷, which would be
490 consistent with Sin1 being part of a dAnk2 promoted cluster of SDV transmembrane pro-
491 teins that destabilizes the nanodroplets (Fig. 3C). The VSP1.0 provides several more can-
492 didates for transmembrane proteins (in addition to SAP1 and SAP3¹⁶) that might interact
493 with dAnk2 or the other two dAnks (Supplementary Table 1). The discovery and functional
494 characterization of dAnks provides a significant step forward in our understanding of silica
495 pore pattern formation in diatoms. Together with the models for the dAnks' modes of action
496 presented here, we provide an experimentally testable hypothesis how molecular interac-
497 tions at the boundary of the biomineralization compartment can control mineral morpho-
498 genesis at scales of hundreds of nanometers and beyond using a LLPS process.

499 To date, the biological functions of porous silica in diatoms have remained specu-
500 lative including protection against hard UV light, acquisition of photosynthetically relevant
501 radiation, and nutrient uptake^{3,4,63}. Being able to predictably alter the pore patterns in silica
502 through genetic manipulation of dAnks provides unprecedented opportunities for testing
503 such hypotheses. From a materials perspective, diatom silica belongs to the class of
504 meso- and macroporous materials, which have an enormously wide range of applications
505 including photonics, catalysis, sensorics, and drug delivery⁴⁸⁻⁵¹. The properties of these
506 materials critically depend on the sizes, spacing and arrangement of the pores. Much pro-
507 gress has been made on developing synthetic methodologies that achieve exquisite con-
508 trol over the porosity of these materials, yet they generally involve environmentally haz-
509 ardous reagents and energy intensive processes. In contrast, diatoms produce meso- and

510 macroporous silica through an environmentally benign process that is renewable and car-
511 bon neutral, and proceeds with an exponential production rate. The present work provides
512 the first proof-of-principle demonstration that the biological production of silica materials
513 with tailored mesoporous patterns should be feasible. This widely opens the door to an
514 eco-friendly, biotechnological production of functional materials using genetically encoded
515 minerals.

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540 **Materials and Methods**

541

542 *Chemicals and enzymes*

543 The enzymes for molecular genetics and phenylmethylsulfonyl fluoride (PMSF) were pur-
544 chased from Thermo Fisher Scientific. Percoll[®], sucrose, dithiothreitol and iodacetamide
545 were obtained from Sigma-Aldrich. PDMPO was purchased from Aat Bioquest. The anti-
546 biotics blasticidin S hydrochloride and nourseothricin sulfate were obtained from Invi-
547 voGen and Jena Bioscience. All oligonucleotides were purchased from Eurofins Ge-
548 nomics.

549

550 *Culturing and transformation of *T. pseudonana**

551 The wild type and all transformant strains of *T. pseudonana* (Husted) Hasle et Heimdahl
552 clone CCMP1335 were maintained in artificial seawater medium at 18 °C and 5000 lx in a
553 12 hours/12 hours day-night cycle, according to the recently published protocol²⁵.

554 Transgenic *T. pseudonana* strains were generated by particle bombardment as
555 described in Supporting Information.

556

557 *Enrichment of valve SDVs*

558 All steps were performed at 4 °C. The density gradient buffer (DGB) that was used for all
559 solutions was composed of 50 mM HEPES-NaOH pH 7.5, 100 mM NaCl, 50 mM sucrose
560 and 100 µM PMSF. A 300 mL *T. pseudonana* cell culture was synchronized by two con-
561 secutive silicon starvation-replenishment cycles, PDMPO labeled, and lysed as described
562 previously three hours after silicic acid addition (i.e. when the proportion of cells with valve
563 SDVs was highest)²⁵. The cell lysate was diluted to 30 mL with DGB, laid on top of a
564 preformed Percoll[®] gradient (90 mL of 80% Percoll[®] in DGB were centrifuged at 30000g
565 for 1 hour in a 50.2Ti (Beckman-Coulter) fixed angle rotor for gradient formation) and cen-
566 trifuged for 1 hour at 2000g in an A-4-81 rotor (Eppendorf). After centrifugation, 40 mL of
567 the gradient were removed from the top. The visible green fraction was collected (15 mL),
568 diluted to 30 mL with DGB and subsequently centrifuged again through a freshly prepared
569 Percoll[®] gradient, as described above. The visible green fraction (PF1, 15 mL) was col-
570 lected, diluted to 25 mL with DGB, laid on top of a sucrose cushion (30 mL of 42.5% su-

571 crose in DGB) and centrifuged at 5000g for 2 hours in a SW-32Ti rotor. After centrifuga-
572 tion, 40 mL of the volume were removed from the top and the remaining cushion (15 mL)
573 was collected, diluted to 25 mL with DGB, laid on top of a second sucrose cushion of
574 higher density (30 mL of 62.5% sucrose in DGB) and centrifuged at 5000g for 2 hours in
575 a SW-32Ti rotor. After centrifugation, 40 mL including the upper fraction (discernible by
576 the green color) were discarded. The remaining cushion volume (SF2, 15 mL) was col-
577 lected, diluted to 60 mL with DGB and centrifuged at 5000g for 16 hours in a SW-32Ti
578 rotor. The supernatant was discarded, and the pelleted SF2 was stored at -20 °C until
579 further use.

580

581 *Proteomics analysis of valve SDVs*

582 The proteomics analysis of SF2 was done with two biological replicates each with two
583 technical replicates. The pelleted SF2 fraction was resuspended in 100 µL of 2% SDS
584 solution and incubated at 37 °C for 20 min. The sample was centrifuged for 1 hour at
585 5000g in a fixed angle rotor, 50 ng BSA were added to the supernatant as internal standard
586 for sample quantity, and the sample was stored at room temperature overnight. For gel
587 electrophoresis 48 µL (equivalent to 34 µg, quantified by the BCA method⁶⁴) of sample
588 were mixed with 12 µL 5x SDS sample buffer and heated at 85 °C for 2 min. The sample
589 was separated on a 10% Bis-Tris SDS gel (two lanes per sample) at 180 V for 6 min re-
590 sulting in separation distance of approximately 2 cm. Proteins were visualized by staining
591 with Coomassie Brilliant Blue. Each lane was sliced into four areas of approximately 5 mm
592 length. After in-gel digestion, the peptides were extracted from the gel slices, fractions
593 were combined and analyzed by nanoLC-MS/MS. Digestion of proteins in-gel and the ex-
594 traction and analysis of the peptides by GeLC-MS/MS was performed according to proce-
595 dures described previously with slight modifications^{52,53}. In brief, gel pieces were des-
596 tained and proteins were reduced with dithiothreitol and alkylated with iodacetamide. Di-
597 gestion was performed with trypsin at a final enzyme concentration of 10 ng·µL⁻¹ overnight
598 in 10mM NH₄HCO₃. Peptides were extracted, subsequently dried and stored until analysis
599 at -20 °C. For LC-MS/MS analyses the peptides were recovered in 5 µL 30% formic acid,
600 diluted with 20 µL of water and 5 µL were injected. NanoLC-MS/MS analyses were per-
601 formed with a Q-Exactive HF mass spectrometer hyphenated to nanoflow LC system (Di-
602 onex3000 RSLC). Peptides were separated in a linear gradient of 0.1% aqueous formic
603 acid for 120 min (eluent A) and 0.1% formic acid in 60% acetonitrile (eluent B) and the

604 mass spectrometer was operated in data-dependent acquisition mode (DDA, TopN 10).
605 Raw files were loaded into the Progenesis QIP V4.2 software (Nonlinear Dynamics) for
606 peak picking. Peptide and protein identification was performed with Mascot V2.8 (Matrix-
607 science) using the UniProt database UP000001449 (02/2022). Only protein hits with at
608 least two significant peptide hits in every replicate were considered for further analysis.
609 The proteomics data have been deposited at the following link [https://sharing.biotec.tu-](https://sharing.biotec.tu-dresden.de/index.php/s/q5tpHESHukcV6P6)
610 [dresden.de/index.php/s/q5tpHESHukcV6P6](https://sharing.biotec.tu-dresden.de/index.php/s/q5tpHESHukcV6P6).

611

612 *Phylogenetic analysis*

613 The phylogenetic distribution of proteins was carried out using the MMETSP transcriptome
614 data⁶¹ in particular the ‘_clean.fasta’ files from which contaminants have been removed
615 (software: https://github.com/kolecko007/mmetsp_cleanup, data: [www.imicrobe.us/#/pro-](http://www.imicrobe.us/#/projects/104)
616 [jects/104](http://www.imicrobe.us/#/projects/104))⁶⁵. tBLASTn searches were carried out with Seg-based filtering of low complexity
617 regions turned on, with default parameters. For further computational details see Methods
618 S3.tBLASTn searches were carried out with default parameters and Seg based filtering of
619 low complexity regions turned on. A custom Perl scrip that performs the searches and
620 generate output tables suitable for plotting is available at [https://github.com/skeffing-](https://github.com/skeffington/PhyloBlast_diatoms)
621 [ton/PhyloBlast_diatoms](https://github.com/skeffington/PhyloBlast_diatoms)

622

623 *Fluorescence microscopy*

624 Epifluorescence images of subcellular fractions were acquired with a 63x oil objective on
625 an Axiovert 200 inverted microscope equipped with a FITC (excitation 450-490 nm, emis-
626 sion 515 nm (longpass), Zeiss) and PDMPO (excitation 365-395 nm, emission 520-
627 550 nm) filter set.

628 For live cell confocal microscopy a 10 µl aliquot of a cell culture with a density of
629 $5.0 \cdot 10^5$ cells·mL⁻¹ was spotted onto a coverslip and overlaid with an agarose slice (1% in
630 ASW medium). For image acquisition, a LSM780 inverted confocal microscope with a
631 Plan-Apochromat 63x (1.4) oil DIC M27 objective (Zeiss) was used. The GFP and chloro-
632 phyll fluorescence were excited with a 488 nm laser line (power set to 1%), a MBS 488
633 beam splitter, and a 32-channel GaAsP spectral detector. To separate the GFP signal
634 from chlorophyll fluorescence two distinct channels were used [GFP (491-535 nm); chlo-
635 rophyll (655-721 nm)].

636

637 *Biosilica preparation for electron microscopy*

638 The preparation of diatom cell walls for electron microscopy was done according to the
639 previously published method² with minor modifications. Briefly, a wild type or transformant
640 strain of *T. pseudonana* was grown until a cell density of 10^6 mL⁻¹. For biosilica isolation,
641 300 mL of the cell culture were centrifuged at 2500g for 3 min. The cells were resuspended
642 in 13 mL of 2% SDS and 100 mM EDTA (pH 8) and incubated at 60 °C for 1 hour under
643 constant shaking. The sample was pelleted at 2500g for 3 min and the pellet was washed
644 three times with 13 mL of 10 mM EDTA (pH 8) by centrifugation (2500g for 3 min) and
645 resuspension cycles. Subsequently, the sample was resuspended in 1 mL of 100% ace-
646 tone, pelleted at 2500g for 3 min and washed four times with 13 mL of H₂O by centrifuga-
647 tion (2500g for 3 min) and resuspension cycles. Finally, the biosilica was resuspended in
648 1.5 mL of water and kept at 4 °C. For scanning electron microscopy, 1 mL of the biosilica
649 was stepwise transferred from H₂O to 100% ethanol by centrifugation (2500g for 3 min)
650 and resuspension cycles with incrementally increasing ethanol concentrations (20%, 40%,
651 60%, 80% and 100%). The dehydrated biosilica was critical point dried and subsequently
652 immobilized on a carbon pad that was attached to an aluminium stub. The sample was
653 sputter coated with platinum and argon as the process gas at 40 mA for 40 sec. The scan-
654 ning electron micrographs were taken at an acceleration voltage of 5 kV and a lower sec-
655 ondary electron detector (JEOL JSM 7500F field emission SEM). For transmission elec-
656 tron microscopy, 500 µL of the biosilica was sonicated in an ultrasonic homogenizer using
657 an MS72 sonotrode tip (0.1 kJ for 5 sec). An 8 µL aliquot of the sonicated sample was
658 spotted on a Formvar-coated gold finder grid for 30 min. The remaining liquid was re-
659 moved with a piece of filter paper and the grid was washed with H₂O for 5 min. For drying,
660 the majority of the H₂O was removed with filter paper and the grid was air-dried overnight.
661 The transmission electron micrographs were taken at an acceleration voltage of 80 kV
662 (JEOL JEM-1400 TEM).

663

664 *Correlative fluorescence and electron microscopy of valve SDVs*

665 The synchronization of the cells and the PDMPO labeling of the valve SDVs was per-
666 formed as described previously²⁵. Briefly, cells were grown to a density of $5.0 \cdot 10^5$ mL⁻¹ and
667 subsequently transferred into Si-free ASW for 16 hours (12 hours dark, 4 hours light). After
668 the starvation period, Na₂SiO₃ was added to the culture at a final concentration of 200 µM,
669 and the cells were incubated in constant light for 9 hours. The cells were again subjected

670 to a silicon starvation period for 16 hours (12 hours dark, 4 hours light). Na_2SiO_3 was
671 added to the cell culture at a final concentration of 200 μM . After 3 hours PDMPO (Aat
672 Bioquest) was added to a final concentration of 1 μM for 10 min. For lysis, the cells were
673 resuspended in lysis buffer (50 mM Hepes-NaOH pH 7.5, 150 mM NaCl, 50 mM Sucrose,
674 100 μM PMSF) and vortexed in the presence of nitric acid cleaned glass beads.

675 Valve SDVs were imaged with electron microscopy according to the previously
676 published method²⁵. Briefly, the cell lysate was immobilized on a Formvar-coated gold
677 finder grid (Science Service). The position of valve SDVs on the grid were identified by
678 their PDMPO fluorescence. The grid was washed 3 times with water and air dried over-
679 night. The electron micrographs were taken at an acceleration voltage of 80 kV (JEOL
680 JEM-1400 TEM). The locations of the valve SDVs were determined based on the corre-
681 sponding fluorescence images.

682

683 *Spectral clustering*

684 The previously annotated *sin1* cluster (originally termed SiMat7 gene cluster) is based on
685 the log₂-transformed fold change in gene expression during Si shift-down and Si shift-up
686 periods²⁴. Here, only the Si shift-up signal intensities were considered. For the partitioning
687 of the mRNA microarray we implemented the spectral clustering segmentation tech-
688 nique⁵⁴⁻⁵⁶ by using the scikit-learn (v0.21.3) python package⁵⁷. The expression profiles
689 were normalized with the z-score method and the affinity matrix built using the radial basis
690 function kernel ($\gamma = 1$). As a primary input parameter, the optimal number of clusters was
691 determined based on the Davies-Bouldin index (DBI) criterion⁵⁸, where its lowest value
692 indicates a better data partitioning. For the present transcriptomic dataset the local mini-
693 mum of DBI = 1.33 was observed at N = 11 (Supplementary Fig. 22A). Additionally, the
694 number of *sin1*-associated members was analyzed as a function of clusters numbers
695 (Supplementary Fig. 22B). For the latter, the stationary phase was observed after passing
696 N = 11.

697

698 *Analysis of pore densities and pore patterns*

699 TEM images of valves were analysed in a semi-automated fashion using a custom-made
700 pore recognition algorithm. The python code for image processing was partially adapted
701 from the skimage tutorial repository⁵⁹ using python 3.7.4 and various packages (scipy,
702 numpy, pandas and scikit-image⁶⁰. For mature cell walls TEM images the margin of the

703 valve was cropped out manually, and only pores within the cropped region were accounted
704 for. We used automatic multilevel thresholding (Yen's method) to automatically determine
705 a binary mask of putative pores. The pores segmentation was achieved by applying the
706 Euclidean distance transform algorithm (EDT) to the obtained mask, the position of each
707 pore was then determined as the local maximum of the EDT. Subsequently, a water-
708 sheding filter was implemented that allowed to determine the outline of each pore. The
709 pore diameter was determined through the obtained pore circumference.

710 For the analysis of pores in nascent biosilica, a slightly different strategy was ap-
711 plied. In order to identify developing pores (that commonly are not completely engulfed by
712 silica), the ridge detection filter was implement. Determining the positions and shapes of
713 identified pores was achieved with the same algorithm as for mature cell walls. The margin
714 of the valve as well as the central annulus were excluded from the area of interest, and
715 only pores between these boundaries were considered (Supporting Fig. 19). Additionally,
716 only the pores within a diameter size range of 15 to 40 nm were included in the subsequent
717 analysis.

718 To compute the pore density (PD), the number of pores (N_p) within the specified
719 region of the valve was divided by the area of that region, $PD = N_p / \text{area}$.

720 To compute the pore distribution function (PDF), we first selected regions-of-inter-
721 est (ROIs) of size 300 nm x 300 nm from the corrected binary mask of pores, centered at
722 each recognized pore. Each ROI was then rotated such that the radial direction (i.e., an-
723 nulus to valve margin) points along the new y-axis. All rotated ROIs were then averaged
724 over all identified pores yielding the PDF for an individual valve.

725

726 *Molecular genetic methods*

727 The verification of the gene models for *dank1-3*, and *tp25735* by RACE PCR as well as
728 the construction of *gfp* fusion genes and knockout plasmids are described in Supplemen-
729 tary information.

730

731

732

733

734

735

736 **Data availability**

737 All mass spectrometry raw data generated in this study were deposited under the link
738 <https://sharing.biotec.tu-dresden.de/index.php/s/q5tpHESHukcV6P6>

739

740 **Acknowledgments**

741 We are grateful to Nicole Poulsen (TU Dresden) and André Scheffel (TU Dresden) for
742 critically reading the manuscript. We wish to thank Jennifer Klemm (TU Dresden) and
743 Anne Jaczkowski (TU Dresden) for help with diatom transformation, and Ariane Kühn (TU
744 Dresden) and Tim Schell-Dieckel (TU Dresden) for assistance in PCR screening of *T.*
745 *pseudonana* mutants. We thank Marc Gentzel and Kristin Eismann from the Molecular
746 Analysis/Mass Spectrometry Facility (CMCB, TU Dresden) for LC-MS/MS and proteomics
747 analysis. Expert support was provided by the Light Microscopy Facility and the Electron
748 Microscopy Facility of the Center for Cellular and Molecular Bioengineering (TU Dresden).
749 The work was supported by the Deutsche Forschungsgemeinschaft (DFG) through grant
750 KR1853/8-2 (to NK) in the framework of Research Unit 2038 (NANOMEE), and through a
751 “Physics of Life” Starting Grant under Germany’s Excellence Strategy – EXC-2068 –
752 390729961– Cluster of Excellence Physics of Life of TU Dresden (to NK and BMF). BMF
753 acknowledges support by the DFG through a Heisenberg grant (FR3429/4-1). The
754 Molecular Analysis/Mass Spectrometry Facility was supported by grants from the German
755 Federal Ministry of Education and Research (BMBF) program 'Unternehmen Region'
756 (grants #03Z2ES1, #03Z22EB1), the German Research Foundation (INST 269/731-1
757 FUGG) and the European Regional Development Fund (ERDF/EFRE) (Contract #
758 100232736).

759

760

761

762

763

764

765

766

767

768

769 **References**

- 770 1. Faivre, D. & Schüler, D. Magnetotactic bacteria and magnetosomes. *Chem. Rev.*
771 **108**, 4875–4898 (2008).
- 772 2. Sun, J. & Bhushan, B. Hierarchical structure and mechanical properties of nacre: a
773 review. *RSC Adv.* **2**, 7617–7632 (2012).
- 774 3. De Tommasi, E. *et al.* UV-shielding and wavelength conversion by centric diatom
775 nanopatterned frustules. *Sci. Rep.* **8**, 1–14 (2018).
- 776 4. Goessling, J. W. *et al.* Structure-based optics of centric diatom frustules: modulation
777 of the in vivo light field for efficient diatom photosynthesis. *New Phytol.* **219**, 122–
778 134 (2018).
- 779 5. Fuhrmann, T., Landwehr, S., El Rharbi-Kucki, M. & Sumper, M. Diatoms as living
780 photonic crystals. *Appl. Phys. B Lasers Opt.* **78**, 257–260 (2004).
- 781 6. Round, F. E., Crawford, R. M. & Mann, D. G. *The Diatoms: biology and morphology*
782 *of the genera. Cambridge University Press, Cambridge* (1990).
- 783 7. Volcani, B. E. *Cell Wall Formation in diatoms: morphogenesis and biochemistry. In*
784 *Silicon and siliceous structures in biological systems, Springer-Verlag, Berlin*
785 *Heidelberg* (1981).
- 786 8. Kröger, N. & Poulsen, N. Diatoms-from cell wall biogenesis to nanotechnology.
787 *Annu. Rev. Genet.* **42**, 83–107 (2008).
- 788 9. Gordon, R. & Drum, R. W. The chemical basis of diatom morphogenesis. *Int. Rev.*
789 *Cytol.* **150**, 243–372 (1994).
- 790 10. Parkinson, J., Brechet, Y. & Gordon, R. Centric diatom morphogenesis: a model
791 based on a DLA algorithm investigating the potential role of microtubules. *Biochim.*
792 *Biophys. Acta* **1452**, 89–102 (1999).
- 793 11. Willis, L., Cox, E. J. & Duke, T. A simple probabilistic model of submicroscopic
794 diatom morphogenesis. *J. R. Soc. Interface* **10**, 1–9 (2013).
- 795 12. Schmid, A. M. M. *Wall morphogenesis in Coscinodiscus wailesii Gran et Angst. II.*
796 *Cytoplasmic events of valve morphogenesis. In Proceedings of the 8th International*
797 *Diatom Symposium, O. Koeltz, Königstein* (1986).
- 798 13. Sumper, M. A Phase separation model for the nanopatterning of diatom biosilica.
799 *Science* **295**, 2430–2433 (2002).
- 800 14. Lenoci, L. & Camp, P. J. Diatom structures templated by phase-separated fluids.

- 801 *Langmuir* **24**, 217–223 (2008).
- 802 15. Robinson, D. H. & Sullivan, C. W. How do diatoms make silicon biominerals?
803 *Trends Biochem. Sci.* **12**, 151–154 (1987).
- 804 16. Tesson, B., Lerch, S. J. L. & Hildebrand, M. Characterization of a new protein family
805 associated with the silica deposition vesicle membrane enables genetic
806 manipulation of diatom silica. *Sci. Rep.* **7**, 1–13 (2017).
- 807 17. Görlich, S., Pawolski, D., Zlotnikov, I. & Kröger, N. Control of biosilica morphology
808 and mechanical performance by the conserved diatom gene Silicanin-1. *Commun.*
809 *Biol.* **2**, 1–8 (2019).
- 810 18. Trofimov, A. A. *et al.* Deep data analytics for genetic engineering of diatoms linking
811 genotype to phenotype via machine learning. *npj Comput. Mater.* **5**, 1–8 (2019).
- 812 19. Poulsen, N. & Kröger, N. Silica morphogenesis by alternative processing of silaffins
813 in the diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* **279**, 42993–42999 (2004).
- 814 20. Kotzsch, A. *et al.* Biochemical composition and assembly of biosilica-associated
815 insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *J. Biol.*
816 *Chem.* **291**, 4982–4997 (2016).
- 817 21. Kotzsch, A. *et al.* Silicanin-1 is a conserved diatom membrane protein involved in
818 silica biomineralization. *BMC Biol.* **15**, 1–16 (2017).
- 819 22. Skeffington, A. W. *et al.* Shedding light on silica biomineralization by comparative
820 analysis of the silica-associated proteomes from three diatom species. *Plant J.*
821 doi:10.1111/tpj.15765.
- 822 23. Shrestha, R. P. *et al.* Whole transcriptome analysis of the silicon response of the
823 diatom *Thalassiosira pseudonana*. *BMC Genomics* **13**, 1–16 (2012).
- 824 24. Brembu, T., Chauton, M. S., Winge, P., Bones, A. M. & Vadstein, O. Dynamic
825 responses to silicon in *Thalassiosira pseudonana* - Identification, characterisation
826 and classification of signature genes and their corresponding protein motifs. *Sci.*
827 *Rep.* **7**, 1–14 (2017).
- 828 25. Heintze, C. *et al.* An intimate view into the silica deposition vesicles of diatoms.
829 *BMC Mater.* **2**, 1–15 (2020).
- 830 26. Yee, D. P., Hildebrand, M. & Tresguerres, M. Dynamic subcellular translocation of
831 V-type H⁺-ATPase is essential for biomineralization of the diatom silica cell wall.
832 *New Phytol.* **225**, 2411–2422 (2019).
- 833 27. Iler, R. K. *The chemistry of silica. Wiley-VCH, New York* (1979).

- 834 28. Shrestha, R. P. & Hildebrand, M. Evidence for a regulatory role of diatom silicon
835 transporters in cellular silicon responses. *Eukaryot. Cell* **14**, 29–40 (2015).
- 836 29. Tsvetkova, N. M. *et al.* Small heat-shock proteins regulate membrane lipid
837 polymorphism. *Proc. Natl. Acad. Sci. USA*. **99**, 13504–13509 (2002).
- 838 30. Pickett-Heaps, J. D., Tippit, D. H. & Andreozzi, J. A. Cell division in the pennate
839 diatom *Pinnularia*. IV. Valve morphogenesis. *Biol. Cell*. **35**, 199–206 (1979).
- 840 31. Tesson, B. & Hildebrand, M. Extensive and intimate association of the cytoskeleton
841 with forming silica in diatoms : control over patterning on the meso- and micro-scale.
842 *PLoS One* **5**, 1–13 (2010).
- 843 32. Wenzl, S., Hett, R., Richthammer, P. & Sumper, M. Silacidins: highly acidic
844 phosphopeptides from diatom shells assist in silica precipitation in vitro. *Angew.*
845 *Chem. Int. Ed.* **47**, 1729–1732 (2008).
- 846 33. Sumper, M., Brunner, E. & Lehmann, G. Biomineralization in diatoms:
847 characterization of novel polyamines associated with silica. *FEBS Lett.* **579**, 3765–
848 3769 (2005).
- 849 34. Michael, A. J. Molecular machines encoded by bacterially-derived multi-domain
850 gene fusions that potentially synthesize, N-methylate and transfer long chain
851 polyamines in diatoms. *FEBS Lett.* **585**, 2627–2634 (2011).
- 852 35. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. The ankyrin repeat as
853 molecular architecture for protein recognition. *Protein Sci.* **13**, 1435–1448 (2004).
- 854 36. Bilcke, G. *et al.* Diurnal transcript profiling of the diatom *Seminais robusta* reveals
855 adaptations to a benthic lifestyle. *Plant J.* **107**, 315–336 (2021).
- 856 37. Chesnick, J. M., Kooistra, W. H. C. F., Wellbrock, U. & Medlin, L. K. Ribosomal
857 RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts
858 of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta).
859 *J. Eukaryot. Microbiol.* **44**, 314–320 (1997).
- 860 38. Hopes, A., Nekrasov, V., Kamoun, S. & Mock, T. Editing of the urease gene by
861 CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* **12**, 1–12
862 (2016).
- 863 39. Herth, W. The site of beta-chitin fibril formation in centric diatoms. II. The chitin-
864 forming cytoplasmic structures. *J. Ultrastruct. Res.* **68**, 16–27 (1979).
- 865 40. Masadeh, A. S. Total scattering atomic pair distribution function: new methodology
866 for nanostructure determination. *J. Exp. Nanosci.* **11**, 951–974 (2016).

- 867 41. Bordet, P. Application of the pair distribution function analysis for the study of
868 cultural heritage materials. *Comptes Rendus Phys.* **19**, (2018).
- 869 42. Keen, D. A. Total scattering and the pair distribution function in crystallography.
870 *Crystallography Reviews* vol. 26 (2020).
- 871 43. Poulsen, N., Sumper, M. & Kröger, N. Biosilica formation in diatoms:
872 characterization of native silaffin-2 and its role in silica morphogenesis. *Proc. Natl.*
873 *Acad. Sci. USA.* **100**, 12075–12080 (2003).
- 874 44. Sumper, M. & Brunner, E. Learning from diatoms: nature's tools for the production
875 of nanostructured silica. *Adv. Funct. Mater.* **16**, 17–26 (2006).
- 876 45. Bernecker, A. *et al.* Tailored synthetic polyamines for controlled biomimetic silica
877 formation. *J. Am. Chem. Soc.* **132**, 1023–1031 (2010).
- 878 46. Kröger, N. & Sumper, M. *The molecular basis of diatom biosilica formation. In*
879 *Biom mineralization: progress in biology, molecular biology and application.* Wiley-
880 *VCH, Weinheim* (2004).
- 881 47. Horvát, S. *et al.* Computational analysis of the effects of nitrogen source and *sin1*
882 knockout on biosilica morphology in the model diatom *Thalassiosira pseudonana*.
883 *Discov. Mater.* **1**, (2021).
- 884 48. Sharma, N. *et al.* Diatoms biotechnology: various industrial applications for a
885 greener tomorrow. *Front. Mar. Sci.* **8**, (2021).
- 886 49. Gérardin, C., Reboul, J., Bonne, M. & Lebeau, B. Ecodesign of ordered
887 mesoporous silica materials. *Chem. Soc. Rev.* **42**, 4217–4255 (2013).
- 888 50. Alothman, Z. A. A review: Fundamental aspects of silicate mesoporous materials.
889 *Materials* **5**, 2874–2902 (2012).
- 890 51. Jarmolińska, S., Feliczak-Guzik, A. & Nowak, I. Synthesis, characterization and use
891 of mesoporous silicas of the following types SBA-1, SBA-2, HMM-1 and HMM-2.
892 *Materials* **13**, 1–33 (2020).
- 893 52. Vasilj, A., Gentzel, M., Ueberham, E., Gebhardt, R. & Shevchenko, A. Tissue
894 proteomics by one-dimensional gel electrophoresis combined with label-free
895 protein quantification. *J. Proteome Res.* **11**, 3680–3689 (2012).
- 896 53. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing
897 of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858
898 (1996).
- 899 54. Ng, A. Y., Jordan, M. I. & Weiss, Y. *On spectral clustering: analysis and an*

- 900 *algorithm. In Advances in neural information processing systems, MIT Press,*
901 *Cambridge (2001).*
- 902 55. Speer, N., Spieth, C. & Zell, A. *Spectral clustering gene ontology terms to group*
903 *genes by function. In Algorithms in bioinformatics, Springer-Verlag, Berlin*
904 *Heidelberg (2005).*
- 905 56. Chin, A. J., Mirzal, A. & Haron, H. Spectral clustering on gene expression profile to
906 identify cancer types or subtypes. *J. Teknol.* **76**, 289–297 (2015).
- 907 57. Pedregosa, F. *et al.* Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*
908 **12**, 2825–2830 (2011).
- 909 58. Davies, D. L. & Bouldin, D. W. A Cluster separation measure. *IEEE Trans. Pattern*
910 *Anal. Mach. Intell.* **1**, 224–227 (1979).
- 911 59. Van Der Walt, S. & Nunez-Iglesias, J. Scikit-image: Tutorials. *GitHub*
912 <https://github.com/scikit-image/skimage-tutorials> (2019).
- 913 60. Van Der Walt, S. *et al.* Scikit-image: image processing in Python. *PeerJ* **2**, (2014).
- 914 61. Keeling, P.J., *et al.* The marine microbial eukaryote transcriptome sequencing pro-
915 ject (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans
916 through transcriptome sequencing. *PLoS Biology* **12**, e1001889 (2014).
- 917 62. Osuna-Cruz, C. M., *et al.* The *Seminavis robusta* genome provides insights into the
918 evolutionary adaptations of benthic diatoms. *Nat. Commun.* **11**, 3320 (2020).
- 919 63. Rosengarten, G. & Herringer J. W. *Interactions of diatoms with their fluid environ-*
920 *ment. In Diatom nanotechnology. The Royal Society of Chemistry, London (2018).*
- 921 64. Smith, P. K., *et al.* Measurement of protein using bicinchoninic acid. *Anal. Biochem.*
922 **150**, 76-85 (1985).
- 923 65. Van Vlierberghe, M., Di Franco, A., Philippe, H. & Baurain, D. Decontamination,
924 pooling and dereplication of the 678 samples of the marine microbial eukaryote
925 transcriptome sequencing project. *BMC Res. Notes* **14**, (2021).
- 926

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData3.xlsx](#)
- [SupplementdANKfinal2.pdf](#)
- [SupplementaryData1.xlsx](#)
- [SupplementaryData2.xlsx](#)