Supporting Information

Title: Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization

Authors: Alexander Kotzsch, Philip Gröger, Damian Pawolski, Paul H. H. Bomans, Nico A. J. M. Sommerdijk, Michael Schlierf, Nils Kröger
Supporting Material and Methods

Cloning, expression and purification of rSin1\textsuperscript{lum}. The DNA sequence encoding for amino acids 25-383 of Sin1 (Uniprot ID B8CBQ8) was amplified from \textit{T. pseudonana} cDNA using the sense primer 5’-TTACCATATGACACCAAAAAAGGCGC (Ndel site underlined) and the antisense primer 5’-ATAGGGATCCCTAGTGGTGTTGGATGTTGACCAGTATGTCGAACCTCC (BamH\textit{I} site underlined, His\textsubscript{6} tag in italics, stop codon in bold; Fig. S2b). The PCR product was incorporated into expression vector pJ404 (T5 promoter, ampicillin resistance, IPTG inducible; DNA2.0) via the \textit{NdeI} and \textit{BamH\textit{I}} restriction sites yielding the plasmid pJ404-rSin1\textsubscript{lum}, which was introduced into chemically competent \textit{E. coli} DH5\textalpha{} cells. For protein expression, 300 mL of LB medium containing 100 µg·mL\textsuperscript{-1} ampicillin were inoculated with a DH5\textalpha{} transformant clone carrying the pJ404-rSin1\textsubscript{lum} plasmid, and grown in a shaker-incubator at 37°C and 180 rpm overnight. The following day, 6 L of LB medium containing 100 µg·mL\textsuperscript{-1} ampicillin was inoculated with 240 mL of the overnight culture and grown at 37°C and 180 rpm. At an OD (at 600 nm) of ~0.6 the culture was supplemented with 1 mM IPTG and further incubated for 3 hours. Cells were harvested by centrifugation (6,000 xg, 30 min, 4°C). The cell pellet was washed with 120 mL 1% NaCl, weighed, and stored at -80°C.

Cell lysis was achieved by resuspending frozen cells (~ 28 g) in 200 mL of lysis buffer (50 mM Hepes-NaOH pH 7.5, 0.5 M NaCl) supplemented with 1 mM PMSF and 3 tablets EDTA-free protease inhibitor (Pierce), and processed using an emulsifier (Emulsiflex-C3, Avestin) tempered to 4 °C. The ice-cold cell suspension was passed through the emulsifier for a total of four times applying a pressure of 15,000-20,000 psi. After the last round of lysis, the emulsifier was washed with 20 mL lysis buffer. The lysate and wash fraction were centrifuged (12,000 xg, 20 min, 4°C). As rSin1\textsuperscript{lum} was present in inclusion bodies, the following conditions were used to solubilize and renature rSin1\textsubscript{lum}. The inclusion bodies were washed three times with 4.5 volumes of lysis buffer containing 2 M urea and centrifuged (12,000 xg, 20 min, 16°C). The resulting pellet was carefully resuspended in 6 volumes lysis buffer containing 8 M urea and stirred at room temperature for 16 hours. The solution was then centrifuged (12,000 xg, 20 min, 16°C) and the supernatant was subjected to immobilized metal affinity chromatography using 5 mL HisPur Ni-NTA resin (Thermo Scientific). The resin was equilibrated in lysis buffer, subsequently mixed with the rSin1\textsuperscript{lum} containing solution, and incubated at room temperature for 2 hours under constant mixing. The suspension was poured into a glass column that was equipped with a frit (Sigma Aldrich), the resin was washed twice with four column volumes of washing buffer 1 (50 mM Hepes-NaOH pH 7.5, 0.5 M NaCl, 8 M urea), and three times with two column volumes washing buffer 2 (50 mM Hepes-NaOH pH 7.5, 0.5 M NaCl, 2 M urea, 25 mM Imidazole). Elution of rSin1\textsuperscript{lum} was achieved by rinsing the columns with 6 column volumes elution buffer (50 mM Hepes-NaOH pH 7.5, 0.5 M NaCl, 2 M urea, 250 mM Imidazole). To allow for refolding of rSin1\textsuperscript{lum} and formation of the native disulfide bonds the combined elution fractions were dialyzed twice against 2 L renaturation buffer 1.
(20 mM Tris-HCl, 300 mM NaCl, 5% glycerol, 2.5 mM reduced glutathione, 0.25 mM oxidized glutathione, pH 8.5) and twice against 2 L renaturation buffer 2 (20 mM Tris-HCl, 300 mM NaCl, 5% glycerol, pH 8.5) at 4°C for 24 hours in each dialysis buffer. Finally, the protein solution was dialyzed against 2 L MQ buffer 1 (20 mM Tris-HCl pH 8.5) at 4°C for 24 hours.

The dialysate was subjected to anion-exchange chromatography using a MonoQ 5/50 column (GE Healthcare) equilibrated in MQ buffer 1. The dialysate was loaded onto the column at a flow rate of 0.5 mL min⁻¹, the column was washed with MQ buffer 1 for 30 min at the same flow rate, and then subjected to a linear gradient (48 min, 0.5 mL min⁻¹ flow rate) from 100% MQ buffer 1 to 50% MQ buffer 2 (20 mM Tris-HCl pH 8.5, 2 M NaCl) followed by a 7 min linear gradient to 100% MQ buffer 2. Elution of rSin1lum occurred between 20-25% MQ buffer 2. Fractions were analyzed by SDS-PAGE and Coomassie Blue staining, and those containing rSin1lum were combined.

As a final purification step the rSin1lum containing Mono Q fractions were pooled and loaded onto a Superdex 75 10/300 (GE Healthcare) size-exclusion column and chromatographed in SEC buffer (20 mM Tris-HCl pH 8.5, 0.3 M NaCl) at a flow rate of 0.5 mL min⁻¹. Superdex fractions containing pure rSin1lum (analyzed by SDS-PAGE) were combined and dialyzed against storage buffer (10 mM sodium phosphate pH 7.7) using regenerated cellulose dialysis tubing (10 kDa cutoff; SpectraPor, Spectrum Labs). Aliquots were flash-frozen and stored at -80°C (note: for cryo-TEM studies rSin1lum was used directly after purification without freezing). The concentration of rSin1lum was determined by measuring the absorbance at 280 nm using the calculated molar extinction coefficient of 42,385 M⁻¹ cm⁻¹.

### Cloning and expression of rSin1-SP

The DNA sequence encoding for amino acids 25-426 of Sin1 (Uniprot ID B8CBQ8) was amplified from *T. pseudonana* cDNA using the sense primer 5'-TTACCATATGAACACACAAAAAGGCCGC (NdeI site underlined) and the antisense primer 5'-ATAAAGGATCCCTTAGTGATGATTGATGGTGAGCCATGGCACCACCTGTC (BamHI site underlined, His₆ tag in italics, stop codon in bold; Fig. S2a). The PCR product was incorporated into expression vector pJ404 (T5 promoter, ampicillin resistance, IPTG inducible; DNA2.0) via the NdeI and BamHI restriction sites yielding the plasmid pJ404-rSin1-SP, which was introduced into chemically competent *E. coli* DH5α cells. For protein expression, 5 mL of LB medium containing 100 µg∙mL⁻¹ ampicillin were inoculated with a DH5α transformant clone carrying the pJ404-rSin1-SP plasmid, and grown in a shaker-incubator at 37°C and 180 rpm overnight. The following day 25 ml of LB medium containing 100 µg∙mL⁻¹ ampicillin was inoculated with 0.25 mL of the overnight culture and grown at 37°C and 180 rpm. At an OD (at 600 nm) of ~0.9 the culture was supplemented with 1 mM IPTG and further incubated for 3 hours. Cells were harvested by centrifugation (6,000 xg, 10 min, 4°C). The cell pellet was washed with 20 mL 1% NaCl and stored in aliquots at -80°C. For SDS-PAGE one aliquot of cells was resuspended in 100 µL sample loading buffer (SDS-PAGE) and incubated at 95°C for 10 min. The resulting mixture was diluted 50-
fold with sample loading buffer and 15 µl of the diluted rSin1-SP cell lysate were used as control in Western blots (see Material and Methods section in the main text).

**Expression of GFP-tagged Sin1 in T. pseudonana.** The start and stop codons of Sin1 were confirmed by 5' and 3'RACE PCR (Fig. S10). Total mRNA was isolated from *T. pseudonana* using oligo-dT-functionalized magnetic beads, and cDNA synthesis was performed as described previously (1). For amplification of the 3'-end, nested PCRs were performed using the gene-specific sense primer 5'-TCTTGGAAAGT TTTGACCTTG and the antisense primer 5'-GGCCACGCCTCGACTAGTAC(T)17 for the first PCR, and the gene-specific sense primer 5'-ATATGTCCATGTGACCC and the antisense primer 5'-GGCCACGCCTCGACTAGTAC for the second PCR. To amplify the 5'-end, two nested PCRs were performed using sense primer 5'-GGCCACGCCTCGACTAGTAC and the gene specific antisense primer 5'-GAATTTGACAGAGTATCCGCTG for the first PCR, and sense primer 5'-GGCCACGCCTCGACTAGTAC and gene specific antisense primer 5'-ATCGTACGCATCCTCGAG for the second PCR. All PCR products were ligated to the pJet1.2 vector (Thermo Scientific) and sequenced.

For C-terminal GFP-tagging, the Sin1 terminator region (673 bp downstream of the stop codon) was amplified from genomic DNA using the sense primer 5'-ATGGCTGCGGCCTTAAAGTCTCGTTTGAGAG3' and the antisense primer 5'-AAGAGGGTTAACGTGAGGGTTGTTTCTGAAAG3' (NotI site underlined and HpaI site in italics). The resulting PCR product was digested with NotI and HpaI and introduced into the NotI and HpaI sites of pTPNR-GFP<sub>HpaI</sub>/fcpNat(-NotI) (2) generating pPnr-GFP-TSin1/fcpNAT(-NotI). The promoter region (1013 bp upstream of the start codon) and protein coding region of Sin1 were amplified from genomic DNA using the sense primer 5'-TCGTACTTGGCGCCATACCTCTGAGGCTC-3' and the antisense primer 5'-AAACGAGGTGACAGGGTGTTCGTAAG3' (ApaI site underlined, KpnI site in italics) and introduced into the ApaI and KpnI sites of pPnr-GFP-TSin1/fcpNAT(-NotI) generating the final expression plasmid pPSin1-Sin1-GFP<sub>C</sub>-TSin1/fcpNAT(-NotI). In this plasmid Sin1 is expressed with a C-terminal GFP-tag (Sin1-GFP<sub>C</sub>) under control of the endogenous Sin1 regulatory sequences.

For N-terminal GFP-tagging, the GFP gene was inserted directly downstream of the RRL motif of Sin1. The Sin1 coding region downstream of the RRL motif (covering amino acids 48-426, “Part2”) including the terminator region were amplified from genomic DNA using the sense primer 5'-TCGTACTTGGCGCCATACCTCTGAGGCTC-3' and the antisense primer 5'-AAGAGGGTTAACGTGAGGGTGTTCGTAAG3' (NotI site underlined and HpaI site in italics). The resulting PCR product was digested with NotI and HpaI and introduced into the NotI and HpaI sites of pTPNR-GFP<sub>HpaI</sub>/fcpNat(-NotI) (3) generating pPnr-Sin1<sup>Part2</sup>-GFP<sub>C</sub>-TSin1/fcpNAT(-NotI). The region of the Sin1 gene upstream of the RRL motif and including the promoter region was amplified from genomic DNA using the sense primer 5'-TCGTACTTGGCGCCATACCTCTGAGGCTC-3' and the antisense primer 5'-GCCATCGGTACCACATCCTCG
AGGCGACG-3’ (Apal site underlined, KpnI site in italics) and introduced into the Apal and KpnI sites of pPnR-Sin1\textsuperscript{Part2}-GFP\textsubscript{C}.TSin1/fcpNAT(-NotI) generating the final expression plasmid pPSin1-Sin1\textsuperscript{Part1}.GFP\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT. This plasmid encodes a Sin1 fusion protein with the GFP-tag immediately following the RXL domain (Sin1-GFP\textsuperscript{N}), which is under expression control of the endogenous Sin1 regulatory sequences. Transformation of _T. pseudonana_ and selection of transformants on nourseothricin containing agar plates was performed as described previously (3).

**Expression of double-tagged Sin1 in _T. pseudonana_.** For N-terminal tagging of Sin1 with mTurquoise2 (abbreviated mT2) (4), the GFP gene in plasmid pPSin1-Sin1\textsuperscript{Part1}.GFP\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT (see above) was replaced by the mT2 DNA using the KpnI and NotI restriction sites flanking the GFP gene. The mT2 DNA sequence was amplified using the sense primer 5’-AGGATGGTACCGGCAGAATGGTGAGCAAGGC CG and the antisense primer 5’-TACGCAAGCGCCGTCCCTTGACGCTCGTCC (KpnI site underlined, NotI site in italics) and introduced into the KpnI and NotI sites of plasmid pPSin1-Sin1\textsuperscript{Part1}.GFP\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT generating pPSin1-Sin1\textsuperscript{Part1}.mT2\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT. Subsequently, the restriction sites for AflII and Ncol located in the Sin1 promoter region of pPSin1-Sin1\textsuperscript{Part1}.mT2\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT were removed by overlap extension PCR using the primer pairs 5’-TACA TACATAGTTAACCTATGTTG (sense) and 5’- TAGGAACCTTAACCTATGTTAG (antisense) (former AflII site underlined, mutated base in bold) and 5’-TGGGTGCTT GAGCAAGGCAGAATGGTGAGCAAGGC CG and the antisense primer 5’-TACGCAAGCGCCGTCCCTTGACGCTCGTCC (KpnI site underlined, NotI site in italics) and introduced into the KpnI and NotI sites of plasmid pPSin1-Sin1\textsuperscript{Part1}.GFP\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT generating pPSin1-Sin1\textsuperscript{Part1}.mT2\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT. Subsequently, the restriction sites for AflII and Ncol located in the Sin1 promoter region of pPSin1-Sin1\textsuperscript{Part1}.mT2\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT were removed by overlap extension PCR using the primer pairs 5’-TACA TACATAGTTAACCTATGTTG (sense) and 5’- TAGGAACCTTAACCTATGTTAG (antisense) (former AflII site underlined, mutated base in bold) and 5’-TGGGTGCTT GAGCAAGGCAGAATGGTGAGCAAGGC CG and the antisense primer 5’-ATTATACTAAGCTTACTCTAGTCA CAGCTCG (Ncol site underlined, AflII site in italics, stop codon in bold) and introduced into the Ncol and AflII sites located 3’ of the Sin1\textsuperscript{Part2} DNA sequence in pPSin1-Sin1\textsuperscript{Part1}.mT2\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT, thereby generating the plasmid pPSin1-Sin1\textsuperscript{Part1}.mT2\textsuperscript{N}.Sin1\textsuperscript{Part2}.Venus\textsuperscript{C}.TSin1/fcpNAT. To restore the original AflII and Ncol sites in the promoter region, the PSin1-Sin1\textsuperscript{Part1} sequence in this plasmid was then replaced with the equivalent sequence of plasmid pPSin1-Sin1\textsuperscript{Part1}.GFP\textsuperscript{N}.Sin1\textsuperscript{Part2}.TSin1/fcpNAT (see above) using the Apal and KpnI sites flanking this region. In this plasmid Sin1 is expressed with a N-terminal mT2-tag and a C-terminal Venus-tag (Sin1-mT2\textsuperscript{N}-GFP\textsuperscript{C}) under control of the endogenous Sin1 regulatory sequences. Transformation of _T. pseudonana_ and selection of transformants on nourseothricin containing agar plates was performed as described previously (3).

**Antibody accessibility experiments.** The binding of anti-rSin1 and anti-GFP antibodies to biosilica and organic matrices from Sin1-GFP\textsuperscript{N} expressing _T. pseudonana_ was quantified as follows. Biosilica was isolated by SDS/EDTA extraction as described previously (2), and fragmented by sonication with a MS72 sonotrode tip (Bandelin) by applying a total energy of 1.612 kJ over 80 s. The insoluble organic matrix material was prepared by incubating the biosilica with 10 M NH\textsubscript{4}F (adjusted to pH 4.5 with HCl) for 1 h at room temperature, followed by washing twice with H\textsubscript{2}O.
through centrifugation (10 min, 10,000 xg) and resuspension. The biosilica and the insoluble organic matrix were resuspended separately in blocking solution (Roti-ImmunoBlock (Carl Roth) supplemented with 0.05% (v/v) Tween 20 (Merck Millipore)) and immobilized on poly-L-lysine-coated coverslips by incubation for 1 hour at room temperature. Unbound material was removed from the coverslips by washing with blocking solution. The coverslips were overlaid with anti-rSin1 antiserum or preimmune serum (each serum was used at 1:1000 dilution) in blocking solution for 1 h. After washing 4x 5 min in TBS (50 mM Tris-HCl pH 7.5, 150 mM NaCl) containing 0.05% (v/v) Tween 20, the coverslips were overlaid with Alexafluor647-conjugated anti-rabbit IgG antibodies (0.67 µg mL⁻¹) in blocking solution for 1 hour. The coverslips were washed as described above, followed by two 5 min washes with TBS. GFP- and Alexafluor647-bearing samples were visualized using epi-illumination with a 488-nm laser and a 647-nm laser at 10 mW and 1 mW, respectively, and respective filter sets (laser bandpass (475/35, 628/40), dichroic longpass (H 488 LPXR, H 643 LPXR), and emission bandpass (525/45, 700/75). The recorded fluorescence intensities were adjusted to prevent saturation of the detector. GFP and Alexafluor647 z-stage image series were subsequently acquired with the NIS-Elements software (Nikon) using an EM CCD camera (Ixon Ultra 897, Andor) mounted on an inverted fluorescence microscope (NSTORM, Nikon) equipped with a 100 oil objective (CFI TIRF Apochromat, numerical aperture 1.49, WD 0.12 mm, Nikon) and an autofocus system (Nikon) at an exposure time of 300 ms (GFP) and 70 ms (Alexafluor647) and 4 frames/µm. For the defined regions of interest, maximum projections displaying the highest value of each pixel in all frames of the z-stacks were created using the software NIS-Elements. Alexafluor647 fluorescence intensity values were normalized by dividing through the corresponding GFP fluorescence intensity values from the same region of interest, resulting in relative fluorescence intensities (RFIs). The degree of accessibility of organic microrings in the biosilica was calculated through dividing the RFI from biosilica by the RFI from microrings.

**Immunodetection of GFP in biosilica and organic microrings.** Biosilica and organic microrings were isolated from wild-type cells and from transformant cells expressing Sin1-GFP\textsuperscript{N} and Sin1-GFP\textsuperscript{C} using a published protocol (2). Immunolabeling was performed as described under “Antibody accessibility experiments” (see above) using a polyclonal anti-GFP antibody (Clontech; final concentration: 15 µg ml⁻¹) as primary antibody, and an Alexafluor647-conjugated anti-rabbit IgG (Thermo Fisher Scientific; final concentration: 2 µg ml⁻¹) as secondary antibody. Fluorescence imaging (GFP, Alexafluor647) was performed as described above.
Fig. S1. Sequence conservation of silicanins in diatoms and other protists. Sequences were aligned using ClustalW (6). The alignment was adjusted manually to highlight the conservation of cysteine residues (red asterisks). The degree of sequence identity among all sequences is indicated by the intensity of the background color. A residue with a dark blue background is conserved in all species whereas a lighter blue background indicates that the residue is conserved only in a subset of species. Organisms: Cyclotella cryptica (Ccr), Thalassiosira oceanica (Toc), Thalassiosira rotula (Tro), Fragilariopsis cylindrus (Fcy), Pseudo-nitzschia australis (Pau), Staurosira complex (Sco), Rhizochromulina marina (Rma), Tiarina fusa (Tfu).
Fig. S2. DNA and amino acid sequences of (a) rSin1-SP and (b) rSin1_lum.
Fig. S3. Membrane association of Sin1. Total membranes from *T. pseudonana* were extracted with lysis buffer (pH 7.5) or carbonate buffer (pH 11.5). Membrane suspensions were incubated on ice and centrifuged (100,000 xg, 1 hour). Pellet (P), supernatant (S), and total membranes before extraction (M) were probed for the presence of (a) Sin1, (b) PsbD (integral transmembrane protein with 5 transmembrane helices), and (c) AtpB (peripheral transmembrane protein) by Western blot analysis with polyclonal antibodies. The faint band between 15-25 kDa in the M and P lanes of (a) could either be a degradation product of Sin1 or an unrelated protein that incidentally cross-reacts with the anti-Sin1 antibody.
**Fig. S4.** Accessibility of Sin1 in biosilica and organic matrices. The experiment was conducted according to a previously published method (2). Biosilica and silica-free, insoluble organic matrices were isolated from a transformant strain expressing Sin1-GFP\textsuperscript{N}, and subjected to immunolabeling using anti-rSin1 as primary antibody and an Alexafluor647-labeled secondary antibody. GFP and Alexafluor647 fluorescence was quantified in individual biosilica and organic matrix particles. As a control the same experiment was performed with preimmune serum as the primary antibody. The BF column shows bright field microscopy images and the GFP and AF647 columns show corresponding epifluorescence microscopy images of the same objects in the GFP and Alexafluor647 channels, respectively. The Overlay column shows the overlays of the images from the other three columns. The white rectangles and circles show the regions of interest in which the GFP and Alexafluor647 fluorescence intensities were quantified. Scale bars: 5 µm.
Fig. S5. Localization of a double-tagged Sin1 fusion protein (Sin1-mT2\textsuperscript{N}-Venus\textsuperscript{C}) in *T. pseudonana*. Live cells, biosilica, and biosilica-associated organic matrix from the transformant strain expressing Sin1-mT2\textsuperscript{N}-Venus\textsuperscript{C} (mT2 = mTurquoise2 inserted between the RXL domain and the NQ-rich domain; Venus fused to the C-terminus of Sin1). The fusion protein was expressed under control of the endogenous Sin1 promoter and terminator sequence. Magenta indicates the fluorescence of mTurquoise2 and yellow the fluorescence of Venus. The panels in (a) show confocal fluorescence images (all z-stacks) of identical individual cells in girdle view during interphase and cell division in the Venus channel (top), the mT2 channel (middle), and the overlay of both channels (bottom). For orientation chlorophyll autofluorescence (red color) is shown in the top and middle row, but was omitted in the overlay images. (b) Isolated biosilica particles (central z-stack) and organic matrix material (all z-stacks) were imaged by confocal fluorescence microscopy in the Venus channel (top) and the mT2 channel (middle), and by brightfield microscopy (bottom). Scale bars for all images: 2 µm.
Fig. S6. Anti-GFP immunolabeling of biosilica and organic matrices isolated from *T. pseudonana* transformant cells expressing Sin1-GFP\(^C\) or Sin1-GFP\(^N\). Biosilica and silica-free, insoluble organic matrices were subjected to immunolabeling using anti-GFP as primary antibody and an Alexafluor647-labeled secondary antibody. The BF column shows bright field microscopy images. The GFP and AF647 columns show corresponding epifluorescence microscopy images of the same objects in the GFP and Alexafluor647 channels, respectively. The overlay column shows the overlays of the images from the other three columns. Scale bars: 5 µm.
**Fig. S7.** Time-lapse imaging of a live cell around the time of nuclear division. Selected images from time-lapse confocal fluorescence microscopy of a Sin1-GFP$^\text{C}$ expressing cells stained with the DNA binding dye Hoechst 34580 are shown (see Movie S2). The time above the images relates to the peak of the GFP fluorescence just prior to valve exocytosis which is set as $t = 0$ min. GFP fluorescence is shown in green and Hoechst 34580 fluorescence in magenta. The images demonstrate that nuclear division was completed between -98.8 min and -95.0 min before the onset of valve SDV biogenesis between -87.4 and -83.6 min. This time frame is in good agreement with observations on Sin1-GFP$^\text{C}$ expressing cells that were labeled with PDMPO (see Fig. 4). All images are z-projections of all planes. Scale bars: 2 µm.
**Fig. S8.** Time-dependent quantitative analysis of region-specific GFP fluorescence during girdle band formation in a Sin1-GFP\(^C\) expressing cell. Images were recorded in 3.5 min intervals, and the fluorescence intensities in different regions of the cell were determined. From each frame z-projections were generated combining all nine z-planes. The schematic shows the delineations of the cellular regions that were analyzed. The line coloring in the graphs corresponds to the coloring of the cellular regions in the schematic. The frame with maximum GFP fluorescence in the mid-cell region (trace not shown) was defined as t = 0 min.
Fig. S9. Biochemical analysis of rSin1\textsuperscript{lum}. (a) Coomassie stained SDS-PAGE loaded with molecular mass standard proteins (lane 1) and 3 µg purified rSin1\textsuperscript{lum} (lane 2). (b) Circular dichroism spectra of rSin1\textsuperscript{lum} in 10 mM sodium phosphate buffer pH 7.7 at 25 °C (black line) and at 93 °C (red line).

5’-end of Sin1
1 GGGGGGGGGGGGGGGGAACACCAACTCCACCTGCAGTAATTCTGCTACACCCACCCAA
61 TACACCACAAAAACGAGCACCAGACATTGAAGTTGCTGACTCCCCGCCAATCTCTCCTCGGCCACCCG
121 ACGGCAACCCCTCTGCCCCCAAGCAACCCGCAACCACCCAAAGGCCGCTACGCAGCC
181 CCCCTCATGCGCGGAGGCCACCCCCTCCGCTCGGCTGAGATGCGTAGAT

3’-end of Sin1
1 ATATGCTGCCATGTTGACCAGCAGTGTGACTAAAGGTCGGCAAGGTGCTATTAGCAGACA
61 GGTTGCTGCCATGCTTAAAGCTCTCGTGGAGATTAAGCCATTTGTCTGTGTGCTGTGA
121 CGTGAGAGAGGATGGAGGGATGCTGATGCTCTATGATAATGTTGTTGTTGATGAAAG
181 CAACACATCATCTGCTGATGTGTTTGAAGCTTATATAAAATTTGTGATTAGTCATTAAA
241AAAAAAAAAAAAAAAA

Fig. S10. DNA sequences obtained from 5’- and 3’-RACE PCR analysis of Sin1 mRNA. Start and stop codons are highlighted in red, polyG- and polyA-tails are underlined.
**Supporting Tables**

**Table S1.** Sin1 homologues identified through tBLASTn searches. The data were grouped into non-diatoms, centric diatoms (i.e. cells with radial symmetry), and pennate diatoms (i.e. cells with bilateral symmetry). Multiple entries for the same species result from redundancies in the database (e.g. one species sequenced multiple times, multiple assemblies for one species).

| Non-diatom | MMETSP entry | BLAST hit | BLAST E-value | BLAST Score |
|------------|---------------|-----------|---------------|-------------|
| Rhizochromulina_marina_CCMP1243_SRR1300381_MMETSP1173 | c10071_g1_id len=1425 path=[1403:0-1424] | 1e-148 | 434 |
| Tiarina_fusa_LIS_SRR1296771_MMETSP0042_alt_Tiarina_fusa | c9607_g1_id len=1452 path=[870:492 86:493-1451] | 3e-115 | 350 |
| Blepharisma_japonicum_Stock-R1072_SRR12994660_MMETSP1395 | c13816_g1_id len=517 path=[1:0-516] | 2e-46 | 160 |
| Minchinia_chitonis__SRR1300215_MMETSP0186 | c716_g1_id len=329 path=[1:0-328] | 9e-29 | 104 |
| Lessardia_elongata_SMCC-104_SRR1300468_MMETSP1147 | c406_g1_id len=229 path=[207:0-228] | 2e-20 | 85.5 |
| Madagascaria_erithrochlooids_CCMP3234_SRR1300528_MMETSP1450 | c20414_g1_id | 2e-19 | 82.4 |
| Oxyrrhis_marina_LB1974_SRR130047 | c23020_g1_id len=349 path=[327:0-348] | 2e-21 | 84.0 |
| Aureococcus_anophagefferens_CCMP1850_SRR1300280_MMETSP0916 | | 2e-21 | 89.7 |
| uncultured_eukaryote_CCMP2293_MMETSP0989 | c3166_g1_id len=230 path=[208:0-229] | 6e-19 | 82.4 |
| uncultured_eukaryote_CCMP2293_SRR1300241_MMETSP0986 | c53690_g1_id len=204 path=[1:0-203] | 1e-18 | 81.3 |
| uncultured_eukaryote_CCMP2293_SRR1300242_MMETSP0988 | c56015_g1_id len=222 path=[1:0-221] | 1e-18 | 81.3 |
| Frasinodema_coloniale_CCMP1413_SRR1298662_MMETSP0406_2 | c6180_g1_id len=261 path=[1:0-260] | 8e-15 | 69.7 |
| Synchocystis_puiforielis_CCMP30302_SRR1300531_MMETSP1452 | c7938_g1_id len=215 path=[1:0-214] | 1e-13 | 65.9 |
| Durinskia_baltica_CSIRO-CS-38_SRR1296839_MMETSP0116_2 | c103628_g1_id len=437 path=[100:436] | 6e-11 | 62.0 |
| uncultured_eukaryote_CCMP2135_SRR1300533_MMETSP1467 | c8734_g1_id len=207 path=[1:0-206] | 1e-10 | 57.8 |
| Durinskia_baltica_CSIRO-CS-38_SRR1296941_MMETSP0117_2 | c5595_g1_id len=1389 path=[1370:0-1389] | 6e-10 | 61.6 |
| uncultured_eukaryote_CCMP20111_SRR1300525_MMETSP1446 | c8608_g1_id len=282 path=[1:0-281] | 0.080 | 31.6 |
| Emiliana_huxleyi_379_SRR1300232_MMETSP0997 | c25492_g1_id len=305 path=[283:0-304] | 0.081 | 32.7 |
| Emiliana_huxleyi_379_SRR1300230_MMETSP0995 | c2469_g1_id len=227 path=[1180:0-12 976:12-226] | 0.084 | 32.0 |

| Pennate diatoms | MMETSP entry | BLAST hit | BLAST E-value | BLAST Score |
|-----------------|---------------|-----------|---------------|-------------|
| Pseudo-nitzschia_australis_10249-10-AB_SRR1296848_MMETSP0140_2 | c5425_g1_id len=1494 path=[1:0-1493] | 5e-109 | 333 |
| Pseudo-nitzschia_australis_10249-10-AB_SRR1296849_MMETSP0141_2 | c19344_g1_id len=1492 path=[1:0-1491] | 5e-109 | 333 |
| Pseudo-nitzschia_australis_10249-10-AB_SRR1296847_MMETSP1139_2 | c15567_g1_id len=1505 path=[110:0-1504] | 6e-109 | 333 |
| Pseudo-nitzschia_australis_10249-10-AB_SRR1296850_MMETSP0142_2 | c303_g1_id len=1511 path=[1488:0-1479 3003:1480-1510] | 6e-109 | 333 |
| Thalassionema_frauenfeldii_CMP-1798_SRR1296708_MMETSP0786 | c4228_g1_id len=1363 path=[1365:0-1362] | 4e-108 | 330 |
| Scientific Name | Accession | Length | Path Start | Path End |
|-----------------|-----------|--------|------------|----------|
| Staurosira_CCMF2646_SRR1300449_MMETSP1361 | g1_i1_len=1479 path=[1:0-1478] | 1e+07 | 330 |
| Pseudo-nitzschia_haltii_UNIC101_SRR1300452_MMETSP1423 | g1_i1_len=1613 path=[1591:0-1612] | 6e+06 | 327 |
| Thalassionema_nitzschioides_L26-B_SRR1296923_MMETSP0158 | g1_i1_len=1470 path=[1448:0-1469] | 2e+05 | 324 |
| Thalassionema_nitzschioides_L26-B_SRR1296922_MMETSP0156 | g1_i1_len=1534 path=[110:0-1533] | 4e+05 | 324 |
| Pseudo-nitzschia_fraudulenta_WA17_SRR1294335_MMETSP0852 | g7920_g1_i1_len=1470 path=[1:0-1469] | 8e+04 | 320 |
| Pseudo-nitzschia_fraudulenta_WA17_SRR1296745_MMETSP0850 | g9223_g1_i1_len=1534 path=[1512:0-1533] | 1e+03 | 320 |
| Pseudo-nitzschia_fraudulenta_WA17_SRR1296747_MMETSP0853 | g12755_g1_i1_len=1524 path=[110:0-1523] | 1e+03 | 320 |
| Pseudo-nitzschia_fraudulenta_WA17_SRR1296746_MMETSP0851 | g13654_g1_i1_len=1530 path=[110:0-1529] | 4e+03 | 320 |
| Tryblionella_compressa_CCMF561_SRR1296829_MMETSP0744_alt_Nitzschia_punctata | g8101_g1_i1_len=1556 path=[1768:0-898 2664:899-919 240:920-1555] | 4e-03 | 319 |
| Tryblionella_compressa_CCMF561_SRR1296830_MMETSP0745_alt_Nitzschia_punctata | g4801_g1_i1_len=1583 path=[1561:0-925 2487:926-946 2508:947-1582] | 4e-03 | 319 |
| Tryblionella_compressa_CCMF561_SRR1296832_MMETSP0747_alt_Nitzschia_punctata | g4841_g1_i1_len=1575 path=[1693:0-648 2339:649-669 98:670-1574] | 4e-03 | 319 |
| Tryblionella_compressa_CCMF561_SRR1296831_MMETSP0746_alt_Nitzschia_punctata | g6786_g1_i1_len=1590 path=[1:0-919 921:920-940 942:941-972 974:... | 5e-03 | 319 |
| Cyclophora_tenuis_ECT3854_SRR1296719_MMETSP0397 | g7879_g1_i1_len=2065 path=[53:0-2064] | 1e+02 | 322 |
| Stauroneis_constricta_CCMF120_SRR1300408_MMETSP1352 | g5024_g1_i1_len=1558 path=[3163:0-541 268:542-617 83440!:618-97... | 2e+02 | 317 |
| Amphora_coffeiformis_CCM127_SRR1296713_MMETSP0316 | g5384_g1_i1_len=1282 path=[1:0-1281] | 3e+02 | 313 |
| Amphora_coffeiformis_CCM127_SRR1296713_MMETSP0318 | g11081_g1_i1_len=1308 path=[12860:1307] | 5e+02 | 313 |
| Amphora_coffeiformis_CCM127_SRR1296713_MMETSP0316 | g11557_g1_i1_len=1329 path=[13070:1328] | 8e+02 | 313 |
| Pseudo-nitzschia_pungens_cf.-circularata_SRR1300393_MMETSP1060 | g979_g1_i1_len=1557 path=[15850:1556] | 1e+00 | 313 |
| Asterionellosis_glacialis_CCM1581_SRR1300451_MMETSP1394 | g4365_g1_i1_len=1606 path=[2126:0-611 2737:612-631 505:632-659... | 4e+00 | 311 |
| Thalassiosiria_antarctica_L6-D1_SRR1296920_MMETSP0152 | g6096_g1_i1_len=1453 path=[83:0-80 164:81-123 189:124-154 220:... | 8e+00 | 309 |
| Stratiatella_unipunctata_CCMF2910_SRR1296706_MMETSP0800 | g11594_g1_i3_len=1480 path=[19750:0-145 6769:146-210 5183:211-25... | 4e+00 | 308 |
| Fragilariopsis_kerguelensis_L2-C3_SRR1296820_MMETSP0909 | g27603_g1_i1_len=1547 path=[1:0-1546] | 3e-07 | 304 |
| Fragilariopsis_kerguelensis_L2-C3_SRR1296819_MMETSP0906 | g14276_g1_i1_len=1606 path=[110:0-1605] | 5e-07 | 304 |
| Fragilariopsis_kerguelensis_L2-C5_SRR1296912_MMETSP0736 | g3453_g1_i1_len=1603 path=[1631:0-1602] | 8e-07 | 303 |
| Fragilariopsis_kerguelensis_L2-C5_SRR1296910_MMETSP0734 | g6008_g1_i1_len=1636 path=[1:0-1635] | 1e-06 | 303 |
| Fragilariopsis_kerguelensis_L2-C5_SRR1296909_MMETSP0733 | g6598_g1_i1_len=1692 path=[1670:0-1691] | 2e-06 | 303 |
| Fragilariopsis_kerguelensis_L2-C5_SRR1296907_MMETSP0732 | g7654_g1_i1_len=1713 path=[1641:0-1712] | 3e-06 | 303 |
| Nitzschia_RCC80_SRR1294404_MMETSP0014_2 | g13380_g1_i2_len=1956 path=[19340:0-254 2189:255-350 3252:351-58... | 2e-05 | 304 |
| Arthopyra_septentrionalis_CCMF2074_SRR1300527_MMETSP1449 | g18253_g1_i1_len=1760 path=[17380:1759] | 3e-05 | 300 |
| Amphiprora_paludosa_CCM125_SRR1300396_MMETSP1065 | g9243_g1_i1_len=1734 path=[17120:1733] | 3e-04 | 298 |
| Pseudo-nitzschia_arenensis_B593_SRR1296720_MMETSP0329 | g8221_g1_i1_len=1600 path=[1764:0-1523 3262:1524-1599] | 5e-04 | 296 |
| Eucampia_antarctica_CCMF1452_SRR1300514_MMETSP1437 | g4037_g1_i1_len=1491 path=[14690:1490] | 3e-02 | 290 |
| Cylindrotheca_closterium_FMMC-8-181_SRR1294408_MMETSP0017_2 | g8162_g1_i1_len=1586 path=[1630:0-924 2554:925-933 8340!:934-14... | 2e-01 | 290 |
### Thalassiosira minuscula CCMP1093 SRR1296955 MMETSP0739
- c6964_g1_i1 length=1414 path=[1:0-1413] 3e-91 286

### Thalassiosira minuscula CCMP1093 SRR1296954 MMETSP0738
- c14966_g1_i1 length=1366 path=[1:0-1365] 7e-90 282

### Thalassiosira rotula CCMP3096 SRR1296877 MMETSP0725
- c17542_g1_i1 length=1379 path=[1:0-1378] 1e-89 282

### Amphipora CCMP467 SRR1296825 MMETSP0724
- c15192_g1_i1 length=1416 path=[1:0-1415] 3e-86 273

### Cyclotella meneghiniana CCMP 338 SRR1300522 MMETSP1443
- c2622_g1_i1 length=1591 path=[1:0-1525 3214-1526-1537 89:1538-1... 2e-84 270

### Entomoneis sp. CCMP2396 SRR1300522 MMETSP1443
- c29705_g1_i1 length=985 path=[1:0-984] 6e-71 229

### Crassostrea_gigas UNCC1205 SRR1296989 MMETSP0911
- c4529_g1_i1 length=991 path=[1:0-237 5098:238-990] 3e-62 207

### Fragilariopsis kerguelensis L2-C5 SRR1296911 MMETSP0735
- c3821_g1_i1 length=1041 path=[1:0-1040] 5e-50 175

### Fragilariopsis cylindrus
- JGI CCMP02:Project 16035:Fracy1_assembly_scaffolds.fasta.gz 2e-88

| Centric diatoms | MMETSP entry | BLAST hit | BLAST E-value | BLAST Score |
|-----------------|--------------|-----------|---------------|-------------|
| Thalassiosira FW SRR1303092 MMETSP0159 | c670_g1_i1 length=1529 path=[1507:0-1528] | 0.0 | 676 |
| Thalassiosira rotula GSO102 SRR1296988 MMETSP0010 | c5549_g1_i1 length=1504 path=[1790:0-566 2357:567-591 2382:592-640... | 5e-168 | 485 |
| Thalassiosira gravis GMP1401 SRR1296986 MMETSP0493 | c4626_g1_i1 length=1462 path=[1440:0-1001 2442:1002-1461] | 2e-167 | 483 |
| Thalassiosira rotula GSO102 SRR1296989 MMETSP0911 | c5827_g1_i1 length=1459 path=[1864:0-176 8144:1177-669 2533:670-9... | 2e-166 | 479 |
| Cyclotella meneghiniana CCMP 338 SRR1303092 MMETSP0157 | c3696_g1_i1 length=1587 path=[5230:0-272 795:273-288 88:289-381 900... | 1e-162 | 473 |
| Thalassiosira minuscula CCMP1093 SRR1296956 MMETSP0740 | c6880_g1_i1 length=1458 path=[2550:0-23 279:24-195 450:196-204 459... | 2e-162 | 469 |
| Thalassiosira minuscula CCMP1093 SRR1296877 MMETSP0404_2 | c10827_g1_i2 length=1585 path=[4777:0-518 5295:519-542 568:543-777... | 3e-162 | 471 |
| Thalassiosira minuscula CCMP1093 SRR1296953 MMETSP0737 | c13590_g1_i1 length=1478 path=[269:0-70 340:71-609 120:610-611 122... | 8e-159 | 461 |
| Thalassiosira antarctica CCMP982 SRR1296921 MMETSP0904 | c16916_g1_i1 length=1436 path=[110:1435] | 3e-158 | 459 |
| Thalassiosira antarctica CCMP982 SRR1296962 MMETSP0903 | c4936_g1_i1 length=1510 path=[1581:0-315 1899:316-339 8192:136:340... | 3e-158 | 460 |
| Thalassiosira rotula GSO102 SRR1300436 MMETSP0013 | c6127_g1_i1 length=1574 path=[1551:0-178 1730:179-309 1861:310-384... | 4e-158 | 459 |
| Thalassiosira antarctica CCMP982 SRR1296965 MMETSP0905 | c7374_g1_i1 length=1477 path=[2:0-270 273:271-310 313:311-603 606... | 7e-158 | 459 |
| Thalassiosira minuscula CCMP1093 SRR1296954 MMETSP0738 | c9912_g1_i1 length=1391 path=[1570:0-386 303:387-432 590:433-704 836... | 1e-156 | 454 |
| Thalassiosira antarctica CCMP982 SRR1296959 MMETSP0902 | c4083_g1_i1 length=1477 path=[1602:0-293 3127:294-317 1923:318-654... | 2e-156 | 455 |
| Thalassiosira rotula GSO102 SRR1296990 MMETSP0912 | c6080_g1_i1 length=1552 path=[18720:466 2338:467-541 82000:1542-2... | 3e-156 | 455 |
| Detonula confervacea CCMP 353 SRR1300391 MMETSP0158 | c6268_g1_i1 length=1572 path=[1715:0-54 1770:55-260 104:261-268 81... | 7e-153 | 447 |
| Thalassiosira minuscula CCMP1093 SRR1296955 MMETSP0739 | c9223_g1_i2 length=1442 path=[197:0-149 347:150-194 3740:195-203 6... | 4e-152 | 444 |
| Strain Name | Accession Number | MMETSP ID | Length | Path Range | GC Content |
|-------------|------------------|-----------|--------|------------|-------------|
| Thalassiosira rotula | CCMP3096_SRR1296755_MMETSP0403 | c11264_g1_i1 len=1835 path=[2737:0-12 855:13-577 2:578-601 3325... | 4a-151 | 446 |
| Thalassiosira weissflogii | CCMP1010_SRR1294465_MMETSP1411 | c12439_g1_i1 len=1637 path=[16150:-1636] | 1a-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1296880_MMETSP0900_2 | c10375_g1_i1 len=1619 path=[110:-1618] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1296881_MMETSP0901_2 | c8755_g1_i1 len=1634 path=[1664:0-385 2050:386-1633] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300477_MMETSP1406 | c20482_g1_i1 len=1643 path=[110:-1642] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300478_MMETSP1407 | c12720_g1_i1 len=1637 path=[110:-1636] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300479_MMETSP1408 | c1927_g1_i1 len=1618 path=[110:-1617] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300480_MMETSP1409 | c12305_g1_i1 len=1634 path=[110:-1633] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300481_MMETSP1410 | c15370_g1_i1 len=1620 path=[1598:-1619] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300486_MMETSP1416 | c2291_g1_i1 len=1639 path=[110:-1611 3256:1612-1638] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300490_MMETSP1420 | c17283_g1_i1 len=1639 path=[110:-1638] | 1e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1296879_MMETSP0899_2 | c18420_g1_i1 len=1656 path=[16340:-1655] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300476_MMETSP1405 | c15474_g1_i1 len=1652 path=[110:-1651] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300482_MMETSP1412 | c20390_g1_i1 len=1639 path=[16170:-1638] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300485_MMETSP1415 | c11879_g1_i1 len=1656 path=[16340:-1655] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300487_MMETSP1417 | c1307_g1_i1 len=1658 path=[110:-1631 3297:1632-1657] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300488_MMETSP1418 | c12210_g1_i1 len=1639 path=[110:-1638] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300491_MMETSP1422 | c20429_g1_i1 len=1646 path=[16240:-1645] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1336_SRR1296931_MMETSP0878 | c19888_g1_i1 len=1663 path=[16410:-1662] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1336_SRR1296932_MMETSP0879 | TR5313-c0_g1_i1 len=1641 path=[16190:-1640] [-1, 1619, -2] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1336_SRR1296933_MMETSP0880 | c14442_g1_i1 len=1646 path=[110:-1645] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1336_SRR1296934_MMETSP0881 | c17309_g1_i1 len=1677 path=[110:-1667] | 2e-147 | 434 |
| Thalassiosira weissflogii | CCMP1010_SRR1300489_MMETSP1419 | c17354_g1_i1 len=1686 path=[16640:-1685] | 3e-147 | 434 |
| Thalassiosira punctigera | Tpunc2005C2_SRR1300398_MMETSP1067 | c12626_g1_i4 len=1613 path=[110:-137 7420:138-144 146:145-178 18... | 1e-140 | 417 |
| Thalassiosira gravida | CMpl4cl1_SRR1296877_MMETSP049 | c6828_g1_i1 len=1094 path=[10720:-1093] | 7e-131 | 385 |
| Skeletonema menzelii | CCMP793_SRR1296401_MMETSP0640 | c26649_g1_i1 len=1549 path=[17250:-1548] | 4e-130 | 388 |
| Skeletonema menzelii | CCMP793_SRR1296808_MMETSP663 | c4528_g1_i1 len=1546 path=[8910:-1545] | 5e-130 | 388 |
| Skeletonema marinoi | UBC1201_SRR1300462_MMETSP1428 | c6290_g1_i1 len=1525 path=[1505:0-36 1542:37-56 815628:57-1332] | 2e-128 | 384 |
| Skeletonema gretheae | CCMP-1804_SRR1296994_MMETSP57598_alt_Skeletonema_grethea | c7696_g1_i1 len=1635 path=[19110:-185 377:186-253 445:254-637 36... | 5e-128 | 384 |
| Skeletonema japonicum | CCMP2506_SRR1296995_MMETSP0593 | c8543_g1_i1 len=1782 path=[61308:689 1475:689-844 261:885-1060 ... | 5e-126 | 388 |
| Skeletonema marinoi | skelA_SRR1300407_MMETSP0918 | c7636_g1_i1 len=1489 path=[17290:-1328 3855:1329-1359 75:1360-1... | 1e-125 | 376 |
| Skeletonema marinoi | skelA_SRR1300409_MMETSP0920 | c8151_g1_i1 len=1508 path=[16910:-1508 328:159-184 142:185-1507] | 1e-125 | 376 |
| Skeletonema dohrni | skelB_SRR1296855_MMETSP0563 | c8488_g1_i2 len=1496 path=[28780:970 3832:971-1331 1315:1332-1... | 7e-125 | 374 |
| Skeletonema dohrni | skelB_SRR1296795_MMETSP0562 | c8746_g1_i1 len=1532 path=[14980:-921 2420:922-924 2423:925-969... | 9e-125 | 374 |
| Species                                | Barcode IDs                                                                 |
|----------------------------------------|-----------------------------------------------------------------------------|
| Chaetoceros _affinis_CCMP159_SRR1294453_MMETSP0088 | c2123_g1_i1 len=1489 path=[1:0-1497]                                       |
| Chaetoceros _affinis_CCMP159_SRR1294455_MMETSP0088 | c252_g1_i1 len=1559 path=[1106:0-1172]                                     |
| Chaetoceros_neogracile_CCMP1317_SRR1296836_MMETSP0754 | c6870_g1_i2 len=1558 path=[1702:0-1000]                                   |
| Chaetoceros_neogracile_CCMP1317_SRR1296992_MMETSP0492 | c18166_g1_i1 len=904 path=[1:0-903]                                       |
| Thalassiosira_oceanica_CCMP1005_SRR1300227_MMETSP0971 | c5857_g1_i1 len=992 path=[1020:0-991]                                     |
| Thalassiosira_oceanica_CCMP1005_SRR1300226_MMETSP0971 | c6409_g1_i1 len=990 path=[1076:0-850]                                       |
| Thalassiosira_oceanica_CCMP1005_SRR1294466_MMETSP1421 | c16187_g1_i1 len=1218 path=[1196:0-1217]                                   |
| Thalassiosira_oceanica_CCMP1005_SRR1294467_MMETSP1421 | c26648_g1_i1 len=999 path=[1:0-998]                                       |
| Thalassiosira_weissflogii_CCMP1010_SRR1296878_MMETSP0970 | c2963_g1_i1 len=1221 path=[1199:0-1220]                                   |
| Thalassiosira_weissflogii_CCMP1010_SRR1300483_MMETSP1413 | c15798_g1_i1 len=1128 path=[1106:0-1127]                                   |
| Thalassiosira_weissflogii_CCMP1010_SRR1300484_MMETSP1414 | c5118_g1_i1 len=1446 path=[1424:0-839]                                     |
| Thalassiosira_oceanica_CCMP1005_SRR1294424_MMETSP0973 | c8068_g1_i1 len=1401 path=[2521:0-1400]                                   |
| Chaetoceros_neogracile_CCMP1317_SRR1296636_MMETSP0754 | c6870_g1_i2 len=1558 path=[1702:0-535]                                     |
| Thalassiosira_gravida_CCMP14clt_SRR1296992_MMETSP0492 | c18166_g1_i1 len=904 path=[1:0-903]                                       |
| Chaetoceros_neogracile_CCMP1317_SRR1296835_MMETSP0753 | c5265_g1_i1 len=1555 path=[530:0-1554]                                   |
| Chaetoceros_neogracile_CCMP1317_SRR1296834_MMETSP0752 | c6349_g1_i2 len=1547 path=[616:0-1080]                                     |
| Chaetoceros_neogracile_CCMP1317_SRR1296833_MMETSP0751 | c266_g1_i1 len=1628 path=[1856:0-1627]                                   |
| Chaetoceros_sp._UNC1202_SRR1300463_MMETSP1429 | c6725_g1_i1 len=1495 path=[1:0-1494]                                       |
| Astrosyne_radiata_13vi08_1A_SRR1296718_MMETSP0418 | c5904_g1_i1 len=1547 path=[830:0-377]                                     |
| Chaetoceros_neogracile_RCC1993_SRR1303737_MMETSP1336 | c4325_g1_i1 len=1474 path=[1490:0-1473]                                   |
| Chaetoceros_affinis_CCMP159_SRR1294453_MMETSP0088 | c2123_g1_i1 len=1489 path=[1:0-1489]                                       |
| Chaetoceros_affinis_CCMP159_SRR1294455_MMETSP0088 | c252_g1_i1 len=1559 path=[1:0-650]                                        |
| Chaetoceros_affinis_CCMP159_SRR1294446_MMETSP0088 | c252_g1_i1 len=1559 path=[1:0-650]                                        |
| Chaetoceros_affinis_CCMP159_SRR1294447_MMETSP0088 | c3567_g1_i1 len=1604 path=[1:0-1603]                                       |
| Chaetoceros_affinis_CCMP159_SRR1294448_MMETSP0088 | c1716_g1_i1 len=1597 path=[1:0-1596]                                       |
| Chaetoceros_affinis_CCMP159_SRR1294450_MMETSP0088 | c1726_g1_i1 len=1597 path=[1:0-1596]                                       |
| Chaetoceros_affinis_CCMP159_SRR1294452_MMETSP0088 | c7398_g1_i1 len=1590 path=[1568:0-1589]                                   |
| Chaetoceros_affinis_CCMP159_SRR1294453_MMETSP0088 | c3654_g1_i1 len=1462 path=[1621:0-1461]                                   |
| Thalassiosira_antarctica_L6-D1_SRR1296921_MMETSP154 | c6354_g1_i1 len=1462 path=[1621:0-1461]                                   |
| Thalassiosira_antarctica_L6_D1_SRR1294449_MMETSP0154 | c3654_g1_i1 len=1462 path=[1621:0-1461]                                   |
| Thalassiosira_affinis_CCMP159_SRR12944474_MMETSP0088 | c4630_g1_i1 len=1613 path=[570:0-1062]                                     |
| Thalassiosira_affinis_CCMP159_SRR1294475_MMETSP0088 | c622_g1_i1 len=1724 path=[1:0-1279]                                        |

### Notes
- The table lists various species with their respective barcode IDs.
- The lengths and paths indicate the specific sequence details.\n- The numbers in parentheses following each length indicate the path.
Table S2. Sequence identities between *T. pseudonana* (Tps) Sin1 and Sin2, and between homologues from other diatoms and non-diatom organisms. *Cyclotella cryptica* (Ccr), *Thalassiosira oceanica* (Toc), *Thalassiosira rotula* (Tro); ii) pennate diatoms: *Fragilariopsis cylindrus* (Fcy), *Pseudo-nitzschia australis* (Pau), *Staurosira complex* (Sco); iii) non-diatom organisms *Rhizochromulina marina* (Rma), *Tiarina fusa* (Tfu).

|                | Centric diatoms | Pennate diatoms | Non-diatoms |
|----------------|-----------------|-----------------|-------------|
| **Organism**   | Tps  | Ccr  | Toc  | Tro  | Fcy  | Pau  | Sco  | Rma  | Tfu  |
| **AA number**  | 424  | 407  | 421  | 419  | 434  | 415  | 429  | 435  | 420  |
| **Sequence identity (%)** | 55   | 66   | 66   | 66   | 47   | 46   | 46   | 62   | 49   |
Table S3. Quantification of the extractability of Sin1, PsbD and AtpB from *T. pseudonanana* membranes. Western blots of the respective experiments (see Fig. S2) were analyzed using the software Image Lab 5.2.1 (Biorad). The exposure times were chosen to ensure that chemiluminescence intensities of all protein bands were (i.e. not saturated within the dynamic range of the camera. Data of three independent experiments were averaged. The membrane extractability (E) in % was calculated according to the following equation

\[ E = \frac{I_S}{I_S+I_P} \times 100. \]

in which \( I_S \) is the intensity of the protein band in the extract and \( I_P \) the intensity of the protein band in the extracted membrane.

| Extraction buffer          | Sin1 E (%) | AtpB E (%) | PsbD E (%) |
|----------------------------|------------|------------|------------|
| Lysis buffer pH 7.5        | 0          | 39 ± 10    | 0          |
| Carbonate buffer pH 11.5   | 56 ± 3     | 97 ± 4     | 1 ± 0      |

Table S4. Quantification of the accessibility of Sin1 in biosilica. The experiment was conducted according to a previously published method (2). RFI = relative fluorescence intensity (the standard error of the mean is provided), n = number of particles analyzed. Biosilica and insoluble organic matrices were isolated from *T. pseudonanana* cells expressing Sin1-GFP\(^N\), and subjected to immunolabeling using anti-Sin1 as primary antibody and an Alexafluor647-labeled secondary antibody. Fluorescence intensities were determined using epifluorescence microscopy (see Fig. S3). Alexafluor647 fluorescence intensity in individual biosilica particles served as a quantitative measure for accessibility of Sin1 for the antibody. The same immunolabeling experiment was performed with the organic matrix and the resulting Alexafluor647 fluorescence intensity served as reference value for maximum microring accessibility under the assay conditions. For each analyzed individual object (i.e. biosilica particle, organic matrix particle) the ratio of the fluorescence intensities (RFI) of Alexafluor647 and GFP was calculated thereby normalizing the immunolabeling intensity to the amount of antigen that was present in each object. Therefore, the ratio of the RFI for biosilica (RFI\(_{BS}\)) and the RFI for the organic matrix (RFI\(_{OM}\)) indicates the fraction of biosilica-associated Sin1 that is accessible to the antibody molecules.

| Sample          | RFI         | RFI\(_{BS}\) / RFI\(_{OM}\) |
|-----------------|-------------|----------------------------|
| **BS**          |             |                           |
| Girdleband biosilica | 6.66 ±0.85 (n=33) | 0.19                      |
| Valve biosilica  | 1.21 ±0.11 (n=43) | 0.03                      |
| **OM**          |             |                           |
| Organic matrices | 35.19 ±3.28 (n=36) | n. a.                     |
Table S5. Secondary structure analysis of rSin1\textsuperscript{lum}. CD spectra (see Fig. S9b) were analysed using the DiChroWeb software (7, 8) and the CDSSTR method (reference data set 7 from reference 7), and compared to the amino acid sequence-based secondary structure predicted by the PsiPred webserver (8).

|                | Temperature (°C) | α helix (%) | β sheet (%) | Turns (%) | Disordered (%) |
|----------------|------------------|-------------|-------------|-----------|---------------|
| Predicted      | -                | 28          | 14          | -         | 58            |
| Experiment     | 25               | 18          | 22          | 18        | 43            |
|                | 93               | 9           | 16          | 13        | 62            |

Table S6. Determination of free sulfhydryl groups in rSin1\textsuperscript{lum}. The concentration of free thiol groups [SH] at three different protein concentrations [rSin1\textsuperscript{lum}] was determined using cysteine as a standard. The ratio of free thiol groups that were detected per rSin1\textsuperscript{lum} molecule was less than 1, which indicated that in the vast majority of rSin1\textsuperscript{lum} molecules all 18 Cysteine residues were engaged in disulfide bonds.

| [rSin1\textsuperscript{lum}] (µM) | [SH] detected (µM) | [SH] : [rSin1\textsuperscript{lum}] |
|-----------------------------------|--------------------|-------------------------------------|
| 2.78                              | 0.9                | 0.32                                |
| 5.56                              | 1.8                | 0.32                                |
| 8.33                              | 3.1                | 0.37                                |
**Supporting Movies**

**Movie S1.** Time-lapse imaging of a Sin1-GFP\(^C\) expressing live cell around the time of valve biogenesis. Bright-field (left), GFP fluorescence (second from left), and PDMP fluorescence (third from left) images were acquired in 3.5 min intervals. Movies showing GFP fluorescence, PDMP fluorescence, and an overlay of both (right) are z-projections of all planes. The bright-field images represent the central z plane. The indicated times relate to the maximum of PDMP fluorescence during valve biogenesis (t = 0 min).

**Movie S2.** Time-lapse imaging of a live-cell around the time of nuclear division. Bright-field (left), GFP fluorescence (second from left), and Hoechst 34580 fluorescence (third from left) images were acquired in 3.8 min intervals. Movies showing GFP fluorescence, Hoechst 34580 fluorescence, and an overlay of both (right) are z-projections of all planes. The bright-field images represent the central z plane. The indicated times relate to the maximum of GFP fluorescence during valve biogenesis (t = 0 min).

**Movie S3.** Time-lapse imaging of a live cell around the time of girdle band biogenesis. Bright-field (left), GFP fluorescence (second from left), and PDMP fluorescence (third from left) images were acquired in 3.5 min intervals. Movies showing GFP fluorescence, PDMP fluorescence, and an overlay of both (right) are z-projections of all planes. The bright-field images represent the central z plane. The indicated times relate to the maximum of PDMP fluorescence during valve biogenesis (t = 0 min).

References
1. Poulsen N., Kröger N. Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana*. *J Biol Chem* **279**, 42993–42999 (2004).

2. Kotzsch A., Pawolski D., Milentyev A., Shevchenko A., Scheffel A., Poulsen N., Shevchenko A., Kröger N. Biochemical composition and assembly of biosilica-associated insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *J Biol Chem* **291**, 4982-4997 (2016).

3. Poulsen N., Chesley P.M., Kröger N. Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). *J Phycol* **42**, 1059–1065 (2006).

4. Goedhart J., von Stetten D., Noirclerc-Savoye M., Lelimousin M., Joosen L., Hink M.A., van Weeren L., Gadella Jr. T.W.J., Royant A. (2012) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **3**, 751.

5. Nagai T., Ibata K., Park E.S., Kubota M., Mikoshiba K., Miyawaki A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87-90.

6. Thompson J.D., Higgins D.G., Gibson T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680 (1994).

7. Sreerama N., Woody R.W. Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Anal Biochem* **287**, 252-260 (2000).

8. Jones D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202 (1999).