Grass silica mineralizer (GSM1) protein precipitates silica in sorghum silica cells

Santosh Kumar[1,a], Nurit Adiram-Filiba[2], Shula Blum[1], Javier Arturo Sanchez-Lopez[3], Oren Tzfadia[4], Ayelet Omid[5], Hanne Volpin[5], Yael Heifetz[3], Gil Goobes[2] and Rivka Elbaum[1]

(S.K. ORCID ID: 0000-0001-8261-5620)

[1]Robert H Smith Faculty of Agriculture, Food and Environment, Robert H Smith Institute of Plant Sciences and Genetics in Agriculture, Hebrew University of Jerusalem, Rehovot 7610001, Israel; [2]Department of Chemistry, Bar-Ilan University, Ramat Gan 5290002, Israel; [3]Department of Entomology, Robert H Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot 7610001, Israel; [4]Bioinformatics and Systems Biology, VIB/Ghent University, Gent B-9052, Belgium; [5]Danziger Innovations Limited, Mishmar Hashiva 5029700, Israel

[a]Present address: Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel

Address for correspondence:
Santosh Kumar
Tel: +972-544378508
Email: skg.research@gmail.com

Rivka Elbaum
Tel: +972-8-9489335
Email: rivka.elbaum@mail.huji.ac.il

Summary

- Silicon is absorbed by plant roots as silicic acid. In shoot tissues, silicic acid mineralizes as silica by completely unknown mechanisms. Most prominently, leaf epidermal cells called silica cells deposit silica in most of their volume.
- Using bioinformatics tools, we identified a previously uncharacterized protein in sorghum (*Sorghum bicolor*), which we named as Grass Silica Mineralizer (GSM1). We characterized the expression profile and demonstrated its activity *in vitro* and *in vivo*.
- GSM1 is a basic protein with seven repeat units rich in proline, lysine and glutamic acid. We found GSM1 expression in immature leaf and inflorescence tissues. In the active silicification zone of sorghum leaf, GSM1 was localized to the cytoplasm or near cell boundaries of silica cells. GSM1 was packed in vesicles and secreted to the paramural space. A short peptide, repeating five times in the protein precipitated silica *in vitro* at a biologically relevant silicic acid concentration. Raman and NMR spectroscopies showed that the peptide attached the silica through lysine amine groups, forming a mineral-peptide open structure. Transient overexpression of GSM1 in sorghum resulted in ectopic silica deposition in all leaf epidermal cell types.
- Our results show that GSM1 precipitates silica in sorghum silica cells.

Keywords: active silicification zone (ASZ), biomineralization, Grass Silica Mineralizer (GSM1), silica cell, silicification, silicon, *Sorghum bicolor*

Introduction

Silicon is a major constituent of plant tissues, forming 0.2 to 10% dry weight. While the uptake mechanisms of this mineral were elucidated in the last decade (Ma et al., 2007, 2006, Yamaji et al., 2008), the processes that govern its deposition are completely unknown. Silicon is available to plants as mono-silicic acid [$Si(OH)_4$] whose concentration in soil solution usually varies between 0.1 to 0.6 mM (Epstein, 1994). More than 90% of silicic acid taken up by roots is loaded into the xylem and moves with water inside the plant body to reach shoots. As plants lose water through evapo-transpiration, the silicic acid solution gets concentrated and polymerises as solid biogenic silica ($SiO_2 \cdot nH_2O$). Models of deposition suggest either spontaneous formation as a result of evapo-transpirational loss of water, or a tightly controlled process (Kumar *et al.*, 2017b). Grasses are well known for their high silica content, reaching up to 10% of their dry weight (Hodson *et al.*, 2005). The most prominent sites of silica deposition in grasses are the leaf epidermal cells, abaxial epidermal cells of the inflorescence bracts (glumes and lemma), and the walls of root endodermal cells. The silicification mechanism across the cell types is not uniform (Kumar *et al.*, 2017b). One of the cell types most frequently silicified in grass leaves is silica cells (Motomura *et al.*, 2000; Kumar & Elbaum, 2018). Silica cells are specialized epidermal cells occurring mostly as silica-cork cell pairs, both above and below the leaf longitudinal veins (Kaufman *et al.*, 1985), on internode epidermis (Kaufman *et al.*, 1969) and on the abaxial epidermis of glumes (Hodson *et al.*, 1985). In mature cells, almost the entire cell volume is filled with solid silica (Hodson *et al.*, 1985). Silica deposition in silica cells is independent of transpiration (Kumar *et al.*, 2017a) and does not occur in dead cells (Markovich *et al.*, 2015; Kumar *et al.*, 2017a; Kumar & Elbaum, 2018). Silica structure of *Triticum durum* silica cells has continuously distributed organic matter with the N/C ratio indicative of amino acids (Alexandre *et al.*, 2015). This suggests protein(s) as the templating organic matrix for the silicification process in silica cells. On a similar note, dissolution of silica from *Phalaris* lemma trichomes yields amino acid composition rich in proline-lysine and proline-glutamic acid (Harrison, 1996), suggesting protein(s) rich in these residues might be templating silica in one or the other cell-types in the inflorescence bracts.

Silicification is widespread among living beings from unicellular microbes to highly evolved multicellular organisms (Perry, 2003). Several bio-silica associated proteins have been reported, for example, silaffins (Kröger *et al.*, 1999; Poulsen & Kröger, 2004), silacidins (Wenzl *et al.*, 2008) and silicanin-1 (Kotzsch *et al.*, 2017) from diatoms; and silicateins

(Shimizu *et al.*, 1998) and glassin (Shimizu *et al.*, 2015) from sponges. Among plants, a short peptide derived from an inducible proline-rich protein precipitates silica *in vitro*. The protein, involved in systemic acquired resistance in cucumber, may precipitate silica locally at attempted fungal penetration site (Kauss *et al.*, 2003). However, all the above-mentioned protein groups do not share sequence homology.

To study silica deposition in silica cells, we used sorghum (*Sorghum bicolor*), a member of the grass (Poaceae) family. Sorghum is categorized as active silica accumulator (Hodson *et al.*, 2005; Coskun *et al.*, 2018). In grasses, leaves appear successively (Skinner & Nelson, 1995). Even within an individual leaf, there is a gradient of cell maturation. The leaf epidermal cell division zone is confined to the base of the leaf and the newly divided cells displace the maturing cells away from the leaf base. Hence, leaf epidermal cells that are close to the leaf apex are older than the cells close to the base (Skinner & Nelson, 1995). We earlier found that silica cell silicification is confined to elongating leaves, in a well-defined active silicification zone (ASZ) (Kumar *et al.*, 2017a; Kumar & Elbaum, 2018). The mineralization initiates in the paramural space of viable silica cells, producing a thick silica cell wall, and restricting the cytoplasmic space to smaller and smaller volumes (Kumar *et al.*, 2017a; Kumar & Elbaum, 2018). This model suggests that (1) the live cells secrete some biological factor(s), possibly proteinaceous in nature that catalyse the formation of silica; (2) any protein involved in silica cell silicification would show a transcriptional correlation to the silicic acid transporters. Here, we report on a previously uncharacterized protein that is expressed in silica cells and inflorescence and precipitates silica in them. Hence, we named this protein as Grass Silica Mineralizer (GSM1).

Materials and methods

Plant material, growth conditions and tissue nomenclature

Seeds of *Sorghum bicolor* (L.) Moench (line BTX-623) were surface sterilized and grown in soil as reported previously (Kumar & Elbaum, 2018). Unless indicated otherwise, we used sorghum seedlings of about 2 weeks of age for our studies. Immature, silicifying leaves (L_{IS}) in our present study is analogous to leaf-2, as reported in our earlier studies (Kumar & Elbaum, 2017, 2018; Kumar *et al.*, 2017a). The immature leaf was cut into five equal segments. The middle segment is most active in terms of silica cell silicification (Kumar *et al.*, 2017a; Kumar & Elbaum, 2018) and was named active silicification zone (ASZ). The segment just older than ASZ (towards leaf tip) was named ASZ+1, while the segment just

younger than ASZ (towards leaf base) was ASZ-1. The youngest segment (at leaf base) was named ASZ-2. Mature leaves were cut into 10 equal segments and only the eighth segment from the leaf-base was used for all the experiments.

Initial screening for the silicification factor

In order to find candidate silicification regulating factors, we first analysed publicly available data sets which measure the influence of silicon treatments on gene expression in wheat (GEO NCBI dataset GSE12936) (Chain *et al.*, 2009). Once differentially expressed genes in response to silicon treatment were identified, their probe sequences were extracted and used to BLAST against the sorghum genome. The BLAST yielded a list of 18 sorghum orthologues genes (e-value > 100 was used as orthology cutoff) (Supporting Information Table S1). The primary sequence of all the initially screened proteins were analysed for their theoretical pI value using ProtParam tool, ExPASy (https://web.expasy.org/protparam/) and the presence of predicted signal peptide (SignalP server). Proteins with predicted pI value less than 7.0 or predictably lacking a signal peptide were not considered. Proteins were further screened for the presence of internal repeat units, abundance of histidine, glutamic acid or serine, glycine-rich motifs, and frequency of proline-lysine and proline-glutamic acid residues.

Expression of GSM1 in different tissues

Equal amount of crude protein from each of root, immature silicifying leaves, mature leaves and immature inflorescence tissues were gel-separated under denaturing conditions and blotted onto nitrocellulose membrane. After blocking, the membrane was incubated with purified rabbit polyclonal antibody against peptide-1 sequence (Supporting Information Method S1 for antibody generation). The membrane was washed and incubated with anti-rabbit IgG mouse monoclonal antibody conjugated with Horseradish peroxidase. Membrane was scanned for one second after chemiluminescence development. For details of the Western hybridization protocol, see Method S2.

To check transcription of *GSM1* in immature leaf, mature leaf and immature inflorescence, RNA was isolated from these tissues and cDNA was synthesized. Equal volume of cDNA was PCR amplified using the primers specific to detect the transcript level of *GSM1* and ubiquitin-conjugating enzyme (Sb09g023560) (Shakoor *et al.*, 2014; Markovich *et al.*, 2015) as internal control (See Method S3 for detailed protocol and Table S2 for primer sequences).

Transcript abundance of *GSM1* in immature leaf pieces

Three independent sorghum seedlings with immature leaf length between 13-16 cm were used in the study. We measured relative transcript abundance of *GSM1* in ASZ-2, ASZ-1, ASZ, ASZ+1 and the youngest mature leaf. The youngest mature leaves were used as a measure of background transcription level of *GSM1* as silica cells in these regions are dead and silicification is no longer taking place in them (Kumar & Elbaum, 2018). *GSM1* transcript level in ASZ+1 was arbitrarily assumed to be 1 and the relative transcript abundance using ubiquitin-conjugating enzyme (UbCE) as internal control was calculated according to the formula of Livak and Schmittgen (2001). For comparison, relative transcript level of another housekeeping gene RNA recognition motif-containing (RRM) protein (Sb07g027950) (Shakoor *et al.*, 2014) is also reported with respect to UbCE (Supporting Information Fig. S1). For further details of the experiment, please see Method S3.

Immunolocalization of GSM1 in sorghum leaves

Small leaf pieces from ASZ and mature leaves were fixed and incubated with anti-peptide-1 polyclonal antibody. Tissues were washed and incubated with goat anti-rabbit IgG tagged with Alexa flour 488. The tissues were also stained with propidium iodide to stain the cell wall and observed under Leica SP8 inverted laser scanning confocal microscope (Wetzlar, Germany). As a control to the immunolocalization experiment, we used (i) pre-immune serum in place of anti-peptide-1 antibody, or (ii) omitted either primary, or secondary or both the antibodies in the reaction keeping the rest of the protocol unchanged. Further details of the protocol can be found in the Method S4.

Transient overexpression of GSM1 in tobacco (*Nicotiana benthamiana* Domin)

We used modified pBIN19 plasmid to overexpress GSM1 translationally fused with C-terminal GFP driven by a single CaMV35S promoter (pBIN-GSM1-GFP) (Fig. S2). The same construct lacking GSM1, but with GFP (pBIN-GFP) served as control. We immobilized the plasmids into *Agrobacterium tumefaciens* (strain EHA105) and infiltrated *Nicotiana benthamiana* source leaves on the abaxial side with slight modifications from the reported protocol (Li, 2011). Our resuspension solution consisted of 50 mM MES buffer, 2 mM NaH_2PO_4 supplemented with 200 µM acetosyringone. After 48 hours, leaf pieces were observed under laser scanning confocal microscope (excitation: 488 nm; emission filter range: 500-550 nm). Please see Method S5 for the details of the construct preparation.

Silica precipitation using Peptide-1 for observation under scanning electron microscope (SEM)

We freshly prepared 1 M silicic acid solution by mixing 150 µl of tetramethyl orthosilicate to 850 µl of 1 mM HCl for four minutes under gentle stirring. Five µl of 1 M silicic acid solution was mixed with 50 µl of peptide solution (1.5 or 2.0 mg ml^{-1} in 0.1 M potassium phosphate buffer, pH 7.0) and incubated for 5 minutes under gentle shaking at room temperature. In this way, the final silicic acid and peptide concentration in the reaction volume became 90.9 mM and 1.36 mg ml^{-1} (or 1.81 if started with 2.0 mg ml^{-1}), respectively. The precipitate was collected by centrifugation at 14000g for 5 minutes. The supernatant was thrown, and the pellet was washed thrice with 1 ml H_2O. The pellet was suspended in H_2O and 1 µl of the suspension was smeared on carbon tape, air dried and coated with iridium. The coated sample was observed in Jeol JSM IT100 (Peabody, MA, USA) scanning electron microscope (SEM).

For silica precipitation at 5 mM silicic acid, 55 µl of freshly prepared 1 M silicic acid solution was diluted in 945 µl of 0.1 M potassium phosphate buffer (pH 7.0) to form 55 mM of silicic acid. Ten µl of 55 mM silicic acid was mixed with 100 µl of Peptide-1 solution (1.5 mg ml^{-1} in 0.1 M potassium phosphate buffer, pH 7.0) and incubated for 30 min under gentle shaking condition. Thus, the final silicic acid and peptide concentration in the reaction volume became 5 mM and 1.36 mg ml^{-1}, respectively. The same reaction without Peptide-1 served as control. The precipitate was collected by centrifugation and examined in the SEM as above.

Silica precipitation and characterization by solid state NMR spectroscopy

Silica precipitation for the NMR experiment was carried out by incubating 30 mg of peptide (1.5 mg ml^{-1} in 0.1 M potassium phosphate buffer, pH 7.0) with 2 ml of freshly prepared 1 M silicic acid solution (final silicic acid and peptide concentration in the reaction volume 90.9 mM and 1.36 mg ml^{-1}, respectively) for 30 min under gentle shaking condition. The precipitate with Peptide-1 and Peptide-2 were collected by centrifugation at 5000 X g for 5 minutes and washed thrice by H_2O. After the first centrifugation step, nothing was visible in the Peptide-3 tube and hence the solution was not discarded, but the centrifugation steps were carried out along with the washing steps of Peptides -1 and -2 without further adding H_2O. Peptide-3 formed silica gel after these steps. The silica gel was washed thrice with H_2O. All the precipitates were dried at 60 ºC and the precipitate with Peptide-1 was subjected to NMR analysis. Solid state NMR spectra were recorded on a Bruker 11.7 T Avance III spectrometer (Billerica, MS, USA) equipped with a 4 mm VTN CPMAS probe at a spinning rate of 10 kHz. The ^{29}Si cross polarization experiment was done with 8192 scans and recycle delay of 6

sec and the ^{29}Si direct excitation experiment was recorded with 2560 scans and recycle delay of 60 sec. Analysis of the ^{29}Si spectrum was done using the DMFIT program (Massiot *et al.*, 2002). ^{13}C cross polarization was done with 2048 scans using recycle delay of 5 sec.

Raman spectroscopic observation of Peptide-1 and silica precipitated with Peptide-1

Lyophilized peptide solution was reconstituted in double distilled water (50 mg ml^{-1}) and 1 µl of this solution was put on a steel slide and dried at 37 ºC for 30 minutes before measuring the Raman spectra. Silica was precipitated following the protocol that we used for silica precipitation for the SEM examination. After washing thrice with double distilled water, we dried the precipitate at 60 ºC until the precipitate attained a constant dried weight. The dried powder was put on a steel slide and the Raman measurement was recorded. Raman spectra of the samples were collected using InVia microspectrometer (Renishaw, New Mills, UK) equipped with polarized 532 nm laser (45 mW max. intensity, 4 µm^2 beam) excitation under a 50X air objective lens (NA=0.75). Spectra analysis was done in WiRE3 (Renishaw), including smoothing, background subtraction, and peak picking.

Transient overexpression of GSM1 in sorghum

We created a construct exploiting the maize dwarf mosaic virus (MDMV) genome (Fig. S3; https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2016125143). We fused the expression cassette containing a CaMV35S promoter driving the expression of GSM1 and *nos*-terminator in between *Age*I and *Apa*I restriction sites to the MDMV genome. This plasmid (MDMV-GSM1) served as the overexpression plasmid, whereas the *GSM1* gene replaced by *GUS* was the control plasmid (MDMV-GUS). Plasmids were bombarded on one-week old sorghum seedlings and the seedlings were planted into soil. Leaves from the plants showing symptoms were observed in the SEM. The viral infection was confirmed using PCR and sequencing. Further details on the protocols of this experiment can be found in the Method S6.

Accession number: The nucleotide sequence of *Sorghum bicolor* GSM1 has been submitted to NCBI with the GenBank accession number- MH558953. GenBank accession number of the pBIN19 plasmid is U09365.1. The MDMV-GUS plasmid that we used in the current study is covered by a published patent number WO2016125143 (https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2016125143).

Results

Screening for putative silicification protein(s)

We screened publicly available wheat microarray data (Chain *et al.*, 2009) for transcripts which under silicon treatment showed significantly altered expression in correlation with the known silicon transporters. We shortlisted 18 genes and looked for their orthologs in sorghum for further analysis (Table S1). By screening the primary sequence of these sorghum proteins for sequence or pattern similarity with known silica mineralizing proteins (Kröger *et al.*, 1999; Kauss *et al.*, 2003; Poulsen & Kröger, 2004; Wenzl *et al.*, 2008; Shimizu *et al.*, 2015), we identified a protein in sorghum (Sb01g025970) that had seven repeat units rich in lysine, similar to silaffins from *Cylindrotheca fusiformis*, a diatom species (Kröger *et al.*, 1999); histidine-aspartic acid rich regions similar to glassin from a marine sponge, *Euplectella* (Shimizu *et al.*, 2015), and several proline residues as in a proline-rich protein (PRP1) from cucumber (Kauss *et al.*, 2003). We functionally characterized this protein and named it Grass Silica Mineralizer (GSM1).

Molecular architecture of Sorghum GSM1

Sorghum GSM1 comprises of 524 amino acids with a predicted N-terminal signal sequence (Fig. 1; Fig. S4). The signal sequence is predictably cleaved between amino acid positions 24 and 25, suggesting this protein is co-translationally imported into the endoplasmic reticulum and cleaved by signal peptidase. Further sequence characterization using TargetP predicted GSM1 to be a secretory protein with high probability (Fig. S5). Blast-based homology search revealed that GSM1 is a single copy gene in sorghum and belongs to a family of proline-rich proteins with unknown function. GSM1 is rich in proline, lysine and histidine (Table S3) and is a basic protein with a theoretical pI value of 9.28. Structure predictions suggest that about 80% of the protein is random coil and 14% is alpha helix (Fig. S6), forming intrinsically disordered tertiary structure (Fig. S7). GSM1 does not show sequence similarity with any protein known to be involved in bio-mineralization. However, GSM1 shares the following features with other silica precipitating proteins (Kröger *et al.*, 1999; Kauss *et al.*, 2003; Shimizu *et al.*, 2015). It has repeat units (R1 to R7) rich in proline, lysine, and glutamic acid (P, K, E-rich domain) with the consensus sequence KKPXPXKPKPXPKPXPXPX (Fig. 1). Near a wide range of physiological pH, close to half of this domain is positively charged (lysine-rich) and the rest of the domain is negatively charged (aspartic acid-rich). Five of the repeat units (except R1 and R3) have a histidine and aspartic acid (H, D) rich domain with the

consensus sequence DXFHKXHDYHXXXXHFH immediately preceding the P, K, E-rich domain. This aspartic acid rich domain is negatively charged under physiological pH. In five repeats (except R2 and R4) there is a third domain following the P, K, E-rich domain, rich in proline, threonine and tyrosine (P,T,Y-rich) with the consensus sequence

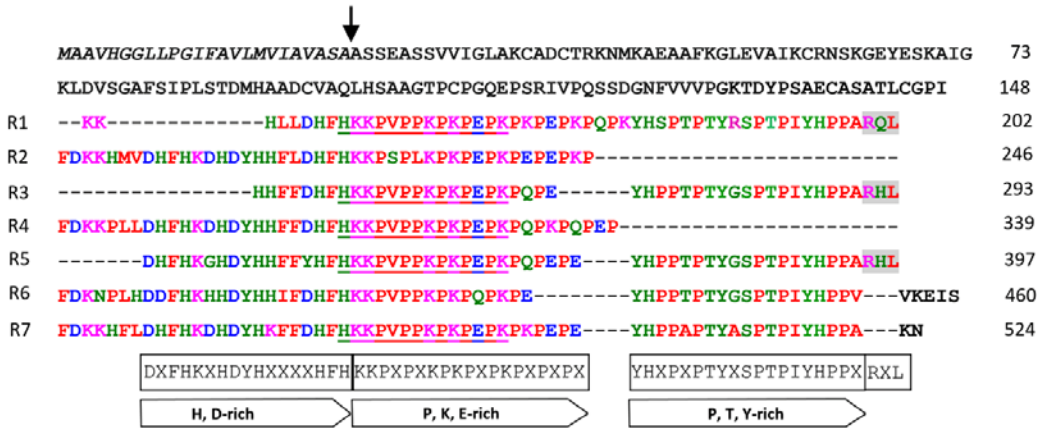

Fig. 1 Primary sequence of *Sorghum bicolor* GSM1, showing the N-terminal signal sequence and the seven repeats (R1-R7). The predicted signal sequence is shown in italics. The arrow denotes the predicted signal sequence cleavage site (See Fig. S4). All the repeats (R1-R7) have a proline, lysine and glutamic acid rich (P, K, E-rich) domain forming a zwitterionic center. In addition, a histidine-aspartic acid rich (H, D-rich) domain, negatively charged near physiological pH precedes the P, K, E-rich domain in repeats 2, 4, 5, 6 and 7. Whereas, a proline, threonine and tyrosine rich (P, T, Y-rich) domain follows the P, K, E-rich domain in repeats 1, 3, 5, 6 and 7. At the end of repeats 1, 3 and 5, there is an RXL domain (shaded) which serves as cleavage site for unknown proteases in many bio-silica associated proteins in diatoms. The underlined sequence is Peptide-1 which appears five times in the primary sequence and was used for raising antibodies and for silica precipitation assays. The consensus sequence for the domains is given in boxes below each domain, where X denotes any amino acid.

YHXPXPTYXSPTPIYHPPX. Before the start of the repeat units (amino acids 1-148), GSM1 is rich in alanine, serine, glycine and valine (Table S4). At the end of R1, R3, and R5, we found an RXL domain. This domain acts as recognition site for unknown proteases that cleave at the C-terminus of the leucine residue of the RXL domain in many bio-silica associated proteins in diatoms (Kröger *et al.*, 1999; Wenzl *et al.*, 2008; Scheffel *et al.*, 2011; Kotzsch *et al.*, 2017).

Expression pattern and subcellular localization of GSM1

To investigate the expression profiles of GSM1, we raised an antibody against Peptide-1, which appears 5 times in the P, K, E-rich domain of GSM1 (Fig. 1). Using Western hybridization, we detected several bands in immature leaves and inflorescence. GSM1 expression was not detected in roots and mature leaves (Fig. 2a). In correlation with the protein expression profile, we detected RNA transcription of *GSM1* in immature leaves and inflorescences, but not in mature leaves (Fig. 2b). RNA transcript profile of *GSM1* along immature, silicifying leaf tissues showed that *GSM1* is strongly transcribed in the leaf base,

reaching the highest levels just before the active silicification zone (ASZ-1). The transcription levels dropped by a factor of about 13 in the ASZ, and further reduced in older tissues of immature leaves (Fig. 2c). In the youngest mature leaf tissue, we found that *GSM1* is transcribed at a basic, low level constituting the background transcription level in these leaves. Background transcription level was about 1/22,000 that of the maximal transcript in ASZ-1.

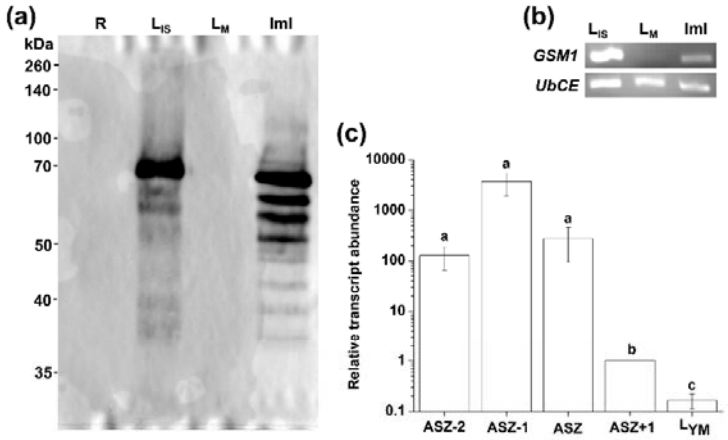

Fig. 2 Expression profile of *Sorghum bicolor* GSM1. (a) Western blot of crude protein extract from roots (R), immature silicifying leaves (L_{IS}), mature leaves (L_M), and immature inflorescences (Iml), detected by an antibody against Peptide-1. GSM1 was expressed in developing leaves and inflorescence, but not in roots and mature leaves. Multiple sized bands in the inflorescence may indicate that GSM1 is processed differently in immature leaf and inflorescence. (b) RNA transcription of *GSM1* in immature silicifying leaves (L_{IS}), mature leaves (L_M), and immature inflorescences (Iml). In accordance with protein translation, transcription was detected only in the immature tissues. *UbCE* (ubiquitin-conjugating enzyme) was used as internal control gene. (c) RNA transcript profile of *GSM1* along immature silicifying leaf tissues. Maximal transcript was found in ASZ-1, which lies just below the Active Silicification Zone (ASZ). Error bars indicate standard error (n=3). Error bar in ASZ+1 is not visible because it is very small.

To further correlate GSM1 with silica deposition, we tested its distribution in leaf epidermal cells using anti-Peptide-1 antibody. In the ASZ, the antibody bound to silica cells (Fig. 3a,b). We noticed the antibody signal in either the cytoplasmic volume or near the cell periphery of silica cells (Fig. 3a,b, Video S1). Further image analysis suggested that the cytoplasmic immunofluorescence signal originated from vesicles (Fig. 3c). In mature leaves, where silica cells are silicified and dead, the anti-Peptide-1 antibody hybridized only to the boundary of silica cells (Fig. 3d,e). No vesicles were identified by image analysis (Fig. 3f). Immunolocalization with the (i) pre-immune serum, or (ii) with the reactions lacking any one or both the antibodies as control did not fluoresce (Fig. S8).

Presence of N-terminal signal peptide in GSM1 predicted it to be secretory in nature (Fig. 1, Fig. S4, S5). Hence, to verify this prediction, we made translational fusion of green fluorescent protein (GFP) to the C-terminal of full length GSM1 (Fig S2) and transiently overexpressed it in tobacco. Similar to our observations in the silica cells of the ASZ, the

cells expressing the fusion protein fluoresced in discrete packets distributed throughout the cytoplasm. In addition, we could identify packets fusing to the cell membrane as well as diffused fluorescence at cell boundaries (Fig. 4a,b, Video S2). Whereas in control tobacco

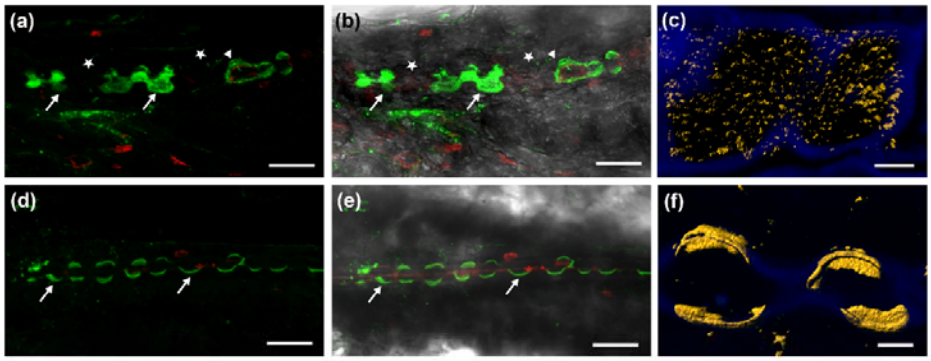

Fig. 3 Immunolocalization of *Sorghum bicolor* GSM1 in the active silicification zone (ASZ) of young leaves (a-c), and near the tip of fully silicified mature leaves (d-f). In sorghum leaves, silica cells usually pair with cork cells (marked with star) and exist as long chains of alternating silica and cork cells; but near the leaf tip, cork cells are absent and silica cells fuse to form long polylobate bodies. (a) Immunolocalization (green fluorescence) shows that GSM1 is localized to silica cells of sorghum leaf active silicification zone (ASZ), inside the cytoplasm (arrows) or near the cell periphery (arrowhead). (b) Fluorescence image in (a) merged with the corresponding bright field image. (c) The anti-peptide-1 antibody fluorescence was processed to select for punctuated regions (pseudocoloured yellow) in the ASZ. The GSM1 appears in packets in the cytoplasm and cell boundary. (d) The anti-peptide-1 antibody (green fluorescence) binds to the edges of silicified silica cells (arrows) in mature leaves. (e) Fluorescence image in (d) merged with the corresponding bright field image. (f) Processed fluorescence of mature leaf image demonstrates that peptide-1 (pseudocoloured yellow) is embedded inside the polymerized silica. Red fluorescence is from propidium iodide, blue is background autofluorescence of the cell walls. Bar 25 μm in panels (a), (b), (d) and (e); and 4 μm in panels (c) and (f).

plants that overexpressed GFP without GSM1, the fluorescence was uniform in the cytoplasm and nucleus, and lower between cells (Fig. 4c). Thus, our results show that native GSM1 is packed in vesicles inside silica cells and suggest that silica cells release GSM1 packed inside the vesicles out of the cell membrane to the paramural space. Since the apoplasm of grass leaves is super-saturated with silicic acid, secretion of the GSM1 to the apoplast may lead to immediate silica precipitation.

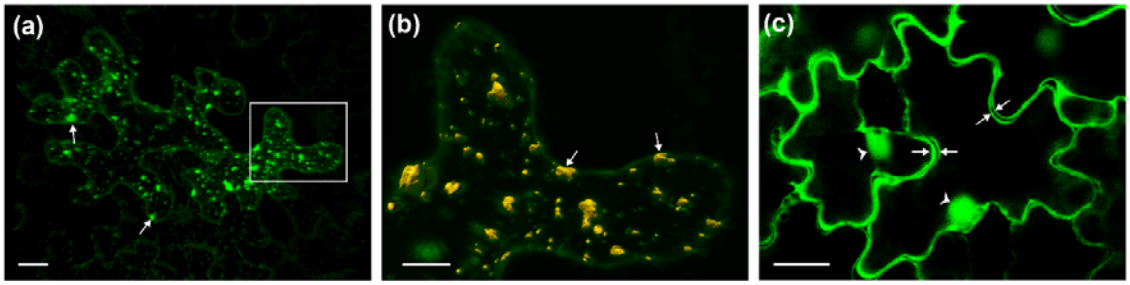

Fig. 4 Fluorescence confocal microscopy showing the secretory nature of *Sorghum bicolor* GSM1. (a) GSM1 fused with GFP was transiently overexpressed in tobacco leaf using *Agrobacterium tumefaciens*. The green fluorescence, marking the location of GSM, was found in packets inside the cytoplasm (arrows), while diffused green fluorescence can also be seen along the margin of the cell. White rectangle marks a region enlarged in panel (b). (b) Segmenting the green fluorescence to diffuse (green) and punctate (pseudocoloured yellow) regions shows packets fusing to the cell membrane (arrows) as well as diffused fluorescence at cell boundaries. (c) Control plants expressing GFP without GSM1 showing green uniform fluorescence of the cytoplasm and nucleus (arrowheads). Arrows indicate the cytoplasm of adjacent cells, demonstrating the low fluorescence between cells. Scale bars in panels (a) and (c) are

Silica *in vitro* precipitation using peptides derived from GSM1 sequence

When metastable (90.90 mM) silicic acid solution was added to the Peptide-1 solution (1.82 or 1.36 mg/ml) at neutral pH, the reaction mixture became cloudy within seconds, and very fine, white sand-like powder precipitated. Similar powder also formed when a peptide of overall same amino acid composition but with scrambled sequence was added to the silicic acid solution (KEKPVKPPKKHPPP; Peptide-2). In contrast, a mutant peptide, where the lysine residues in Peptide-1 were replaced by alanine (HAAPVPPAPAPEPA; Peptide-3), did not precipitate powder-like silica under similar conditions. In contrast, silicic acid solution with peptide-3 formed gel-like material when the reaction mixture was centrifuged (Fig. 5a). Scanning electron micrographs of the powder sediment, forming within 5 minutes with Peptide-1 and Peptide-2 revealed spheres of about 500 nm in diameter (Fig. 5b). Extending the reaction time to 30 min did not change the particle size. Silica was also precipitated by Peptide-1 (1.36 mg/ml) at silicic acid concentration of 5 mM (pH=7.0), which is equivalent to the concentration found in grass leaves (Casey *et al.*, 2003). The precipitation was invisible to

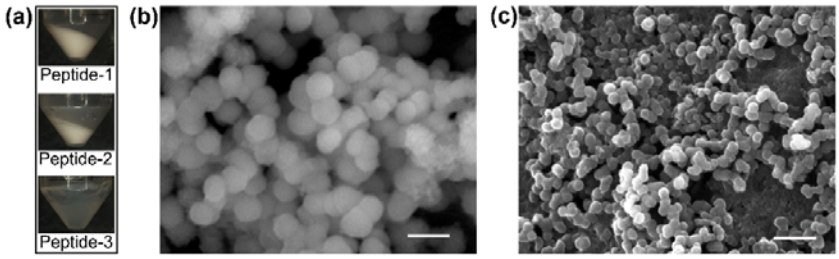

Fig. 5 Silica precipitation by peptides derived from *Sorghum bicolor* GSM1 sequence. (a) Sand-like powder sediment produced from metastable (90.9 mM) silicic acid solution with Peptide-1 and Peptide-2, while a gel-like material formed after several rounds of centrifugation of silicic acid solution with Peptide-3. (b) Scanning electron micrograph of the powder sediment formed within 5 minutes with Peptide-1 at 90.9 mM silicic acid. (c) Scanning electron micrograph of the silica precipitate formed by Peptide-1 at 5 mM silicic acid. Bar, 1 µm.

the naked eyes even after 30 minutes of incubation. However, SEM examination of the precipitate showed that there indeed was silica precipitation and the diameter of an individual silica nanosphere was about 250 nm (Fig. 5c). In the control solution (without peptide) at 5 mM silicic acid, precipitation was not observed by SEM.

Raman and NMR spectroscopic characterization of Peptide-1 – silica precipitate

The precipitated silica using metastable silicic acid and Peptide-1 was characterized by Raman and magic angle spinning (MAS) solid-state nuclear magnetic resonance (ss-NMR) spectroscopies. Raman spectroscopy of the sediment showed that the mineral is silica, recognized by the peaks at 489 cm^{-1} and 997 cm^{-1} (Aksan *et al.*, 2012). These peaks are

missing in the Raman spectrum of Peptide-1 (Fig. 6a). The bulk Si-O-Si broad peak is not shifted by much from literature values (492 cm^{-1}, (Aksan *et al.*, 2012)). However, the surface silanols are shifted by more than 10 cm^{-1} (988 cm^{-1}) (Aksan *et al.*, 2012), or 973 cm^{-1} (Gierlinger *et al.*, 2008), supporting surface modifications of the mineral. A shift in the broad peak of Peptide-1 at 1680 cm^{-1} to lower frequencies indicates a change in the peptide backbone. Such shift may indicate that a random coil structure in the pure peptide is more α-helix-like in the sediment (Anthony, 1982). In the C-H region (2850-2990 cm^{-1}) the 2979 cm^{-1} is reduced in the silica precipitate, indicating changes in the lysine CH_2- group close to the functional amine. A new shoulder at 2850 cm^{-1} could result from a symmetric stretch in CH_2- groups (Howell *et al.*, 1999). The sharp peak at 1431 cm^{-1} is assigned to deformations of NH_3^+ on the lysine functional group (Aliaga *et al.*, 2009). The broadening and shifting of this peak to higher energies (1456 cm^{-1}) indicates an interaction between the silica and the lysine functional group. The interaction may occur through hydroxyl and water groups, inhibiting some NH_3^+ bonds deformations and thus broadening the peak. The peak at 838

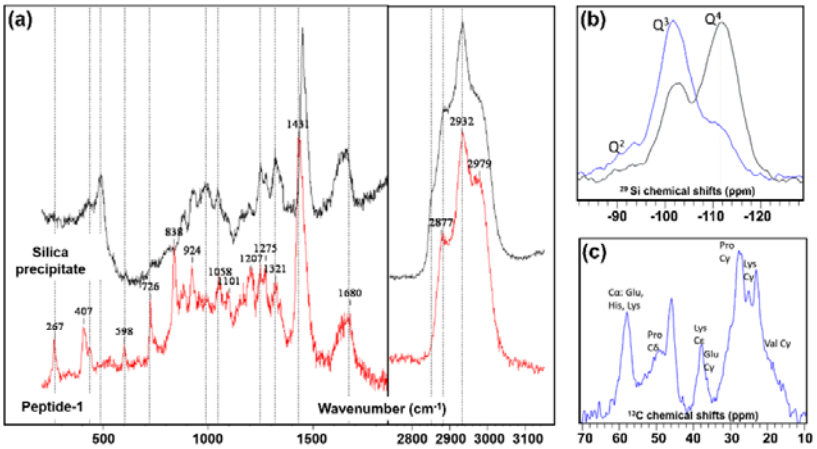

Fig. 6 Characterization of the peptide-silica precipitate using Raman and NMR spectroscopy. (a) Raman spectra of Peptide-1 (red line) and the precipitate formed with Peptide-1 (black line). The spectra show that the mineral is silica, with altered surface, and that the peptide is bound to the mineral through the lysine aliphatic amine and proline ring residues and the -COO^- C-terminal group. (b) Magic angle spinning (MAS) solid-state nuclear magnetic resonance (NMR) of ^{29}Si showing peaks of Si-$(OSi)_2(OH)_2$ (Q2), Si-$(OSi)_3(OH)$ (Q3), and Si-$(OSi)_4$ (Q4). Direct polarization (black line) samples all the Si atoms in the sample, while ^1H-cross polarization (blue line) samples Si atoms in proximity to protons. (c) ^1H-^{13}C cross polarization spectrum (aliphatic region) showing the amino acid sidechain carbons shifted to a high field by about 2 ppm. This shift reflects the shielding effect of the silica.

cm^{-1} may belong to a C-C stretch in the proline ring (Stewart & Fredericks, 1999a). This vibration disappears in the mineral precipitate, suggesting a structural change in prolines possibly embedded within the silica. Vibrations at 924, 726, and 598 cm^{-1} are assigned to groups of -COO^- at the peptide C-terminal (Stewart & Fredericks, 1999b). These groups seem to participate in the structure of the mineral, as their vibrations are missing in the precipitate.

^{29}Si NMR peak position varies according to the number of groups of -OSi and -OH bound to the central Si atom, and the Qn band is defined as Si-$(OSi)_nOH_{4-n}$ (Engelhardt & Michel, 1987). Direct polarization measurements of ^{29}Si showed the typical silicon species Q4, Q3 and Q2 at chemical shifts of -111.4 ppm, -102.1 ppm and -91.8 ppm, respectively (Fig. 6b). The relative intensities of the Q4:Q3:Q2 peaks in the spectrum were 62:33:5, indicating that there are about 5 bulk (Q4) Si atoms for every 3 surface (Q3+Q2) atoms. To examine further the silicon atoms at the surface we excited them through near-by hydrogen atoms, by cross polarization measurement (Fig. 6b). In addition to surface Q3 and Q2 signals, we detected some Q4 siloxane species. These Si atoms were excited by protons from nearby silanols and water, indicating their closeness to the surface of the silica particle.

Confirmation that the peptide was complexed with the silica was given through ^{1}H-^{13}C cross polarization measurements (Fig. 6c). The spectrum showed narrow lines (full width at half maximum (FWHM) was 282 Hz), as compared with a diatom peptide complexed with silica (FWHM of 489 Hz) (Geiger *et al.*, 2016). The narrower lines of Peptide-1 complexed in silica indicated that the structure of Peptide-1 is more uniform than that of the diatom peptide. The relatively narrow lines allowed us to identify many of the sidechain amino acid carbons (Fig. 6c). The carbon lines are shifted to a high- field by about 2 ppm, reflecting the shielding effect of the silica. The absence of the lines from aromatic carbons of histidine sidechain may suggest that they are either involved in averaging motions or experiencing large chemical shift dispersion because of a broad range of interactions with the silica surface. In either case, their lines became too broad and could not be detected.

Functional characterization of GSM1 *in planta*

To elucidate the role of GSM1 *in planta*, we transiently overexpressed it in sorghum, using a construct derived from maize dwarf mosaic virus (Fig. S3). Compared with the wild type plants, control sorghum plants infected with virus lacking the GSM1 sequence showed infection lesions and viral symptoms but no unusual silica deposition (Fig. 7a-f, S9). In contrast, the GSM1 overexpressing plants had ectopic silica deposition in patches, especially close to the veins and where viral symptoms were observed (Fig. 7g-i). Silica was also deposited in cells which are not usually silicified in sorghum leaves, like guard cells, cork cells and long cells (Fig. 7j-l). Scanning the surface of the cells by SEM in the back-scattered electron mode suggested the silicification intensity at the surface of these cells to be of a level similar to that in silica cells. Energy-dispersive X-Ray spectroscopic signal for silicon from the heavily silicified patches in the GSM1 overexpressing leaves was much higher than in the

wild type and control leaves. For energy-dispersive X-ray spectroscopic maps of other elements and their spectra, see (Fig. S10-S13).

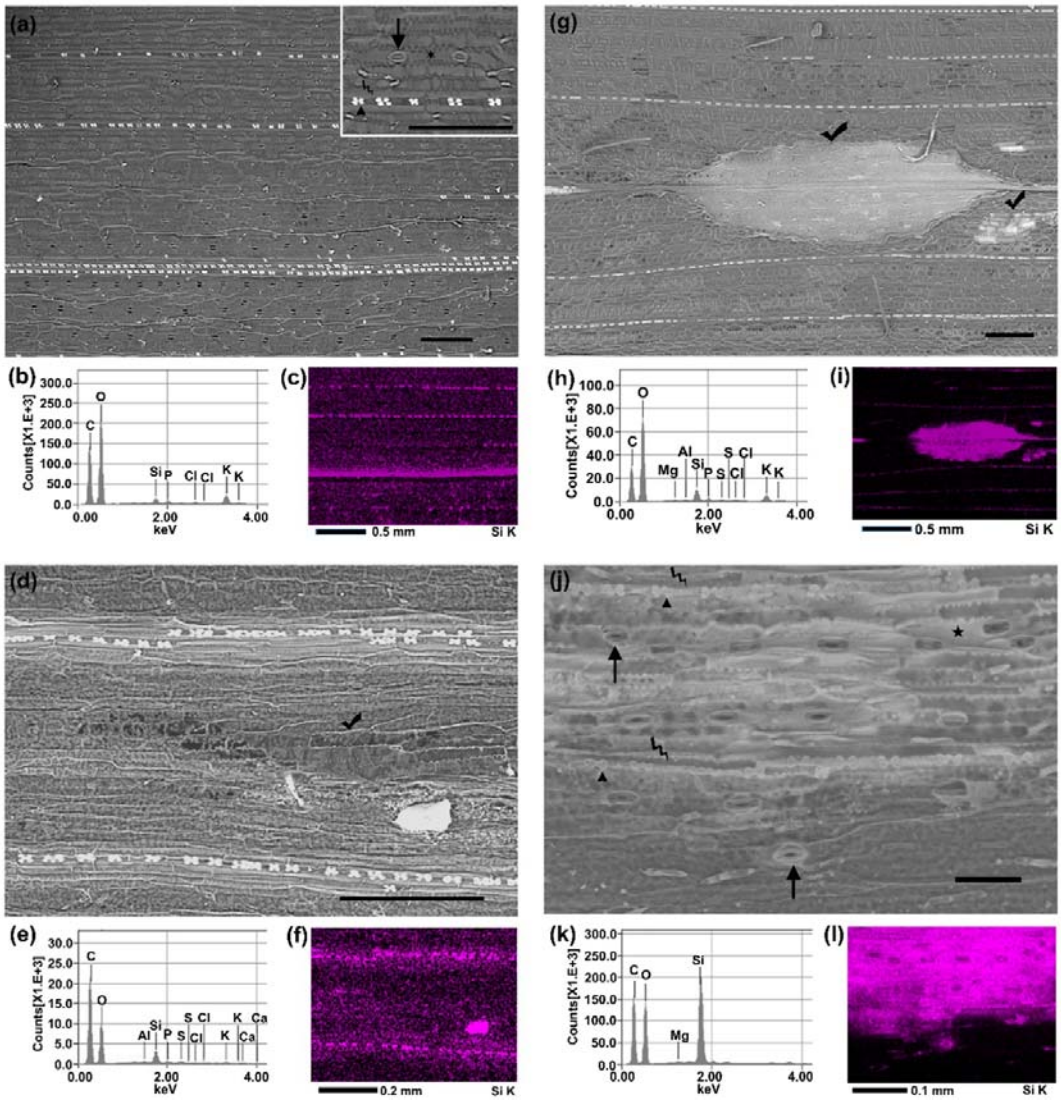

Fig. 7 Overexpression of *Sorghum bicolor* GSM1 in sorghum. Few silica cells are indicated by arrowheads, stomata by arrows, cork cells by thunder sign, long cells by stars and viral lesions by tick mark. (a,d,g,j) Scanning electron microscope (SEM) image in back-scattered electron (BSE) mode. Silicified cells appear brighter compared with the background. (b,e,h,k) Energy-dispersive X-ray spectra (EDS) of the corresponding SEM image showing the comparative elemental composition of the scanned area. (c,f,i,l) EDS map for Si-signal form the corresponding SEM image. Maps are pseudocoloured. Black colour means no signal and increasing intensity of the colour denotes higher density of Si atoms. (a) Wild type sorghum mature leaf scanned by SEM under low magnification showing the general silicification pattern in a mature leaf. Inset: epidermal cell types are indicated in a higher magnification image. (b) EDS of the image in (a). (c) EDS map for Si-signal of the image in (a). (d) SEM image of control plants showing silicification only at usual locations. Dust particle with high Si content is seen on the bottom right. (f) EDS map for Si of the image in (d). (g) SEM image of ectopic silica deposition in a viral lesion in GSM1 overexpressing plant. (h) EDS of the image in (g). (i) EDS map for Si of the image in (g). (j) SEM image of GSM1 overexpressing plants showing high intensity silica deposition in cells that do not usually silicify in wild type plants. All epidermal cell types can be seen silicified. (k) EDS of the image in (j). (l) EDS map for Si of the image in (j). Scale bars represent 200 µm in panels (a; d and g); and 50 µm in panel (j).

Discussion

Studies in plants suggest that certain genetic loci control silicification at specific sites (Dorweiler & Doebley, 1997; Piperno *et al.*, 2002; Peleg *et al.*, 2010), but no direct interaction for a protein precipitating biogenic silica was found. Our work shows that silica cells express and export GSM1 to the apoplast (Fig. 3 and 4) timed with silicification in the paramural space (Kumar *et al.*, 2017a; Kumar & Elbaum, 2018). It is obvious that silica deposition depends on the presence of silicic acid. Since GSM1 overexpression caused deposition in all epidermis cell types (Fig. 7), we can conclude that silicic acid local concentration there was high enough to support silica formation. Furthermore, GSM1 is sufficient to cause precipitation in those cells. Silica precipitation *in vitro* by Peptide-1 was observed at the concentration of 5 mM silicic acid (Fig. 5c), in the range of concentrations usually found in grass leaves (Casey *et al.*, 2003). This suggests that GSM1 in its native form is active at this silicic acid concentration. In addition, it may be post-translationally modified to precipitate silica even if silicic acid is below saturation.

Some indication for GSM1 post-translational modifications may be the time gap between its highest transcription level in the ASZ-1 tissue (Fig. 2), and the highest silicification activity in the ASZ (Kumar *et al.*, 2017a; Kumar & Elbaum, 2018). One such possible modification may be glycosylation (Elbaum *et al.*, 2009). Processed GSM1 molecules are then packed inside vesicles (Lawton, 1980) and stored in the cytoplasm for later release (Alberts *et al.*, 2002) at the time of silicification (Fig. 3). Modifications may be tissue and species specific. The difference in the size of the Western hybridization signal between immature leaf and inflorescence that we observed may be attributed to differential processing of GSM1 in these two tissues (Fig. 2a). GSM1 has three RXL domains in the primary structure (Fig. 1). RXL domains are proteolytically cleaved in many diatom biosilica associated proteins (Kröger *et al.*, 1999; Wenzl *et al.*, 2008; Scheffel *et al.*, 2011; Kotzsch *et al.*, 2016, 2017). It would be interesting to study if GSM1 undergoes alternative processing, similar to *Thalassiosira pseudonana* (a diatom species) silaffin (Poulsen & Kröger, 2004).

GSM1 is not expressed in roots where silica cells do not form. Furthermore, GSM1 is not expressed in mature leaf tissues, in correlation with the absence of viable, active silica cells (Kumar & Elbaum, 2018). This suggests that silica deposition in live silica cells is dependent on GSM1, while silica deposition in cell walls and possibly other locations are governed by other means. Our results showed that GSM1 is also expressed in immature inflorescence, consistent with the earlier transcriptomic data showing that *GSM1* is highly transcribed in the inflorescence tissues in sorghum, both before and after inflorescence emergence from the flag

leaf (Table S5 of Davidson et al. 2012). Glume abaxial epidermal cells silicify before inflorescence emergence, while macro-hairs and the long cells on the abaxial epidermis of lemma, which lacks stomata on the abaxial surface silicify after panicle emergence (Hodson *et al.*, 1984; Hodson, 2016; Kumar *et al.*, 2017b). In the absence of data pertaining to the timing of inflorescence bract silicification in sorghum, it is hard to determine in which cell-type of the inflorescence bracts, silica is precipitated by GSM1.

We did not find GSM1 expression in the protein extract from mature leaves (Fig. 2a). Hence, binding of anti-Peptide-1 antibody in the boundary of dead silica cells of mature leaves suggested that GSM1 templated the silica precipitation in silica cells and got trapped inside the silica structure (Fig 3). Silicanin-1, a biosilica associated protein from *T. pseudonana* gets embedded inside biosilica structure and is accessible to anti-silicanin-1 antibody (Kotzsch *et al.*, 2017). Similarly, anti-Peptide-1 antibody may also have access to the epitope remains on the surface of the deposited silica in silica cells. Our *In-vitro* experiments support such entrapment, with significant shifts in the NMR and Raman peaks, indicating an intimate contact between the peptide and the mineral (Fig. 6). The NMR ^{2}H-^{29}Si cross-polarization measurements suggest that the mineral is very open, with many fully coordinated Si atoms (Q_4) close to its surface (Fig. 6b). Such a structure may exist in the native plant silica and allow movement of solutes into and through the mineral. Diffusion of silicic acid through the mineral is required in order for silica cells to fully silicify, because the mineralization front is at the paramural space (Kumar & Elbaum, 2017, 2018; Kumar *et al.*, 2017a).

Over expression and localization studies of GSM1 show that this protein is involved in silicification in sorghum leaf silica cells. GSM1 has unique amino acid composition, charge distribution and probably post-translational modifications necessary for its activity. Soon after cell division, silica cells reach their final size and shape and start their preparation for silicification. GSM1 is transcribed, translated, and post-translationally modified, packed inside vesicles and stored in the cytoplasm until the cell is ready to silicify. The vesicles fuse to the membrane and release their content in the paramural space to come in contact with silicic acid. This immediately precipitates open-structured silica that allows the diffusion of more silicic acid from the apoplastic space. The rapid formation of the mineral explains the difficulty associated with finding silicification in an intermediate state. The expression pattern, localization and modifications of GSM1 in the inflorescence bracts need to be studied in detail.

Acknowledgements

S.K. was a recipient of post-doctoral fellowship from Planning and Budgeting Committee (PBC), Council of Higher Education, Israel. This research was funded by the Israel Science Foundation grant 534/14. Partial support to the Centre for Scientific Imaging, Hebrew University (Rehovot campus) by a grant from USAID (AID-ASHA-G 1400005) to the American Friends of The Hebrew University and The Robert H Smith Faculty of Agriculture, Food and Environment is gratefully acknowledged. The authors thank Prof. Shmuel Wolf for his generous gift of pBIN-GFP plasmid.

Author Contributions

S.K. and R.E. planned and designed the research; S.K. identified GSM1 from the list of putative silicification related proteins and performed most of the experiments; N.A.-F. did the NMR spectroscopy and analysed the data with G.G.; S.B. performed parts of the experiments; S.K. and J.S. performed immunolocalization experiment under the supervision of Y.H.; O.T. analysed the microarray data and prepared the list of Si responsive genes in sorghum; A.O. prepared the MDMV construct under the supervision of H.V.; S.K. and R.E. analyzed the results and wrote the paper. All the authors commented and approved the final version of the manuscript for publication.

References

Aksan A, Hubel A, Reategui E. **2012**. Silica-based composite ocular device and methods.

Alberts B, Johnson A, Lewin J, Raff M, Roberts K, Walter P. **2002**. *Molecular Biology of the Cell*. New York: Garland Science.

Alexandre A, Basile-Doelsch I, Delhaye T, Borshneck D, Mazur JC, Reyerson P, Santos GM. **2015**. New highlights of phytolith structure and occluded carbon location: 3-D X-ray microscopy and NanoSIMS results. *Biogeosciences* **12**: 863–873.

Aliaga AE, Osorio-Roman I, Garrido C, Leyton P, Cárcamo J, Clavijo E, Gómez-Jeria JS, F. GD, Campos-Vallette MM. **2009**. Surface enhanced Raman scattering study of l-lysine. *Vibrational Spectroscopy* **50**: 131–135.

Anthony TT. **1982**. *Raman spectroscopy in biology: principles and applications*. Wiley.

Casey WH, Kinrade SD, Knight CTG, Rains DW, Epstein E. **2003**. Aqueous silicate complexes in wheat, *Triticum aestivum* L. *Plant, Cell and Environment* **27**: 51–54.

Chain F, Côté-Beaulieu C, Belzile F, Menzies JG, Bélanger RR. **2009**. A comprehensive transcriptomic analysis of the effect of silicon on wheat plants under control and pathogen stress conditions. *Molecular Plant-Microbe Interactions* **22**: 1323–1330.

Coskun D, Deshmukh R, Sonah H, Menzies JG, Reynolds O, Ma JF, Kronzucker HJ, Bélanger RR. **2018**. The controversies of silicon's role in plant biology. *New Phytologist*.

Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu S-H, Jiang N, Robin Buell C. **2012**. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *The Plant journal* **71**: 492–502.

Dorweiler JE, Doebley J. **1997**. Developmental analysis of *Teosinte glume architecture1*: A key locus in the evolution of maize (Poaceae). *American Journal of Botany* **84**: 1313–1322.

Elbaum R, Melamed-Bessudo C, Tuross N, Levy AA, Weiner S. **2009**. New methods to isolate organic materials from silicified phytoliths reveal fragmented glycoproteins but no DNA. *Quaternary International* **193**: 11–19.

Engelhardt G, Michel D. **1987**. *High resolution solid state NMR of silicates and zeolites*. Chichester: John Wiley & Sons.

Epstein E. **1994**. The anomaly of silicon in plant biology. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 11–17.

Geiger Y, Gottlieb HE, Akbey Ü, Oschkinat H, Goobes G. **2016**. Studying the conformation of a silaffin-derived pentalysine peptide embedded in bioinspired silica using solution and dynamic nuclear polarization magic-angle spinning NMR. *Journal of the American Chemical Society* **138**: 5561–5567.

Gierlinger N, Sapei L, Paris O. **2008**. Insights into the chemical composition of *Equisetum hyemale* by high resolution Raman imaging. *Planta* **227**: 969–80.

Harrison CC. **1996**. Evidence for intramineral macromolecules containing protein from plant silicas. *Phytochemistry* **41**: 37–42.

Hodson MJ. **2016**. The development of phytoliths in plants and its influence on their chemistry and isotopic composition. Implications for palaeoecology and archaeology. *Journal of Archaeological Science* **68**: 62–69.

Hodson MJ, Sangster AG, Parry DW. **1984**. An ultrastructural study on the development of silicified tissues in the lemma of *Phalaris canariensis* L. *Proceedings of the Royal Society of London B* **222**: 413–425.

Hodson MJ, Sangster AG, Parry DW. **1985**. An ultrastructural study on the developmental phases and silicification of the glumes of *Phalaris canariensis* L. *Annals of Botany* **55**: 649–665.

Hodson MJ, White PJ, Mead A, Broadley MR. **2005**. Phylogenetic variation in the silicon composition of plants. *Annals of Botany* **96**: 1027–1046.

Howell NK, Arteaga G, Nakai S, Li-Chan ECY. **1999**. Raman spectral analysis in the C−H stretching region of proteins and amino acids for investigation of hydrophobic interactions. *Journal of Agricultural and Food Chemistry* **47**: 924–933.

Kaufman PB, Bigelow WC, Petering LB, Drogosz FB. **1969**. Silica in developing epidermal cells of avena internodes: electron microprobe analysis. *Science* **166**: 1015–1017.

Kaufman PB, Dayanandan P, Franklin CI. **1985**. Structure and function of silica bodies in the epidermal system of grass shoots. *Annals of Botany* **55**: 487–507.

Kauss H, Seehaus K, Franke R, Gilbert S, Dietrich RA, Kröger N. **2003**. Silica deposition

by a strongly cationic proline-rich protein from systemically resistant cucumber plants. *The Plant Journal* **33**: 87–95.

Kotzsch A, Gröger P, Pawolski D, Bomans PHH, Sommerdijk NAJM, Schlierf M, Kröger N. **2017**. Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization. *BMC Biology* **15**: 65.

Kotzsch A, Pawolski D, Milentyev A, Shevchenko A, Scheffel A, Poulsen N, Shevchenko A, Kröger N. **2016**. Biochemical composition and assembly of biosilica-associated insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *Journal of Biological Chemistry* **291**: 4982–4997.

Kröger N, Deutzmann R, Sumper M. **1999**. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* **286**: 1129–1132.

Kumar S, Elbaum R. **2017**. Estimation of silica cell silicification level in grass leaves using in situ charring method. *Bio-protocol* **7**: e2607.

Kumar S, Elbaum R. **2018**. Interplay between silica deposition and viability during the life span of sorghum silica cells. *New Phytologist* **217**: 1137–1145.

Kumar S, Milstein Y, Brami Y, Elbaum M, Elbaum R. **2017a**. Mechanism of silica deposition in sorghum silica cells. *New Phytologist* **213**: 791–798.

Kumar S, Soukup M, Elbaum R. **2017b**. Silicification in grasses: variation between different cell types. *Frontiers in Plant Science* **8**: 438.

Lawton JR. **1980**. Observations on the structure of epidermal cells, particularly the cork and silica cells, from the flowering stem internode of *Lolium temulentum* L. (Gramineae). *Botanical Journal of the Linnean Society* **80**: 161–177.

Li X. **2011**. Infiltration of *Nicotiana benthamiana* protocol for transient expression via *Agrobacterium*. *Bio-protocol* **1**: e95.

Livak KJ, Schmittgen TD. **2001**. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**: 402–408.

Ma JF, Tamai K, Yamaji N, Mitani N, Konishi S, Katsuhara M, Ishiguro M, Murata Y, Yano M. **2006**. A silicon transporter in rice. *Nature* **440**: 688–691.

Ma JF, Yamaji N, Mitani N, Tamai K, Konishi S, Fujiwara T, Katsuhara M, Yano M. **2007**. An efflux transporter of silicon in rice. *Nature* **448**: 209–212.

Markovich O, Kumar S, Cohen D, Addadi S, Fridman E, Elbaum R. **2015**. Silicification in leaves of sorghum mutant with low silicon accumulation. *Silicon* **doi:** 10.1007/s12633-015-9348-x.

Massiot D, Fayon F, Capron M, King I, Le Calvé S, Alonso B, Durand J-O, Bujoli B, Gan Z, Hoatson G. **2002**. Modelling one- and two-dimensional solid-state NMR spectra. *Magnetic Resonance in Chemistry* **40**: 70–76.

Motomura H, Fujii T, Suzuki M. **2000**. Distribution of silicified cells in the leaf blades of *Pleioblastus chino* (Franchet et Savatier) Makino (Bambusoideae). *Annals of Botany* **85**: 751–757.

Peleg Z, Saranga Y, Fahima T, Aharoni A, Elbaum R. **2010**. Genetic control over silica deposition in wheat awns. *Physiologia Plantarum* **140**: 10–20.

Perry CC. **2003**. Silicification: The processes by which organisms capture and mineralize silica. *Reviews in Mineralogy and Geochemistry* **54**: 291–327.

Piperno DR, Holst I, Wessel-Beaver L, Andres TC. **2002**. Evidence for the control of phytolith formation in Cucurbita fruits by the hard rind (Hr) genetic locus: Archaeological and ecological implications. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 10923–10928.

Poulsen N, Kröger N. **2004**. Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana*. *The Journal of Biological Chemistry* **279**: 42993–42999.

Scheffel A, Poulsen N, Shian S, Kröger N. **2011**. Nanopatterned protein microrings from a diatom that direct silica morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 3175–3180.

Shakoor N, Nair R, Crasta O, Morris G, Feltus A, Kresovich S. **2014**. A *Sorghum bicolor* expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums. *BMC plant biology* **14**: 35.

Shimizu K, Amano T, Bari MR, Weaver JC, Arima J, Mori N. **2015**. Glassin, a histidine-rich protein from the siliceous skeletal system of the marine sponge *Euplectella*, directs silica polycondensation. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 11449–11454.

Shimizu K, Cha J, Stucky GD, Morse DE. **1998**. Silicatein α: Cathepsin L-like protein in sponge biosilica. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 6234–6238.

Skinner RH, Nelson CJ. **1995**. Elongation of the grass leaf and its relationship to the phyllochron. *Crop Science* **35**: 4–10.

Stewart S, Fredericks PM. **1999a**. Surface-enhanced Raman spectroscopy of amino acids adsorbed on an electrochemically prepared silver surface. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **55**: 1641–1660.

Stewart S, Fredericks PM. **1999b**. Surface-enhanced Raman spectroscopy of amino acids adsorbed on an electrochemically prepared silver surface. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **55**: 1641–1660.

Wenzl S, Hett R, Richthammer P, Sumper M. **2008**. Silacidins: Highly acidic phosphopeptides from diatom shells assist in silica precipitation in vitro. *Angewandte Chemie* **47**: 1729–1732.

Yamaji N, Mitatni N, Ma JF. **2008**. A transporter regulating silicon distribution in rice shoots. *The Plant Cell* **20**: 1381–1389.

Supplementary Information

Fig. S1 Transcript level of *RRM* in relation to *UbCE* as internal control gene, showing that the transcript level of the two house-keeping genes do not change significantly in the tested tissues.

Fig. S2 pBIN-GSM1-GFP plasmid used for GSM1-GFP fusion protein transient overexpression in tobacco (*Nicotiana benthamiana*).

Fig. S3 Maize dwarf mosaic virus (MDMV) derived plasmid (MDMV-GSM1) used to transiently overexpress *Sorghum bicolor* GSM1 in sorghum.

Fig. S4 Signal peptide prediction in GSM1 using SignalP 4.1 server.

Fig. S5 TargetP predicts GSM1 to be secretory in nature.

Fig. S6 Secondary structure prediction of GSM1 using the program GOR4.

Fig. S7 Intrinsically disordered tertiary structure of GSM1 predicted using IUPred.

Fig. S8 Immunolocalization control reactions using pre-immune serum; or lacking either one or both the antibodies.

Fig. S9 Maize dwarf mosaic virus (MDMV) infected sorghum plants showing symptoms.

Fig. S10 Energy-dispersive X-ray spectrum and maps of different elements for a wild type sorghum mature leaf.

Fig. S11 Energy-dispersive X-ray spectrum and maps of different elements for a sorghum control plant overexpressing GUS in place of GSM1.

Fig. S12 Energy-dispersive X-ray spectrum and maps of different elements from a GSM1 overexpressing sorghum leaf.

Fig. S13 Energy-dispersive X-ray spectrum and maps of different elements from a heavily silicified patch outside viral lesions of a GSM1 overexpressing sorghum plant.

Table S1 List of the genes identified that show significantly differential transcription upon silicon treatment verses the non-treated plants.

Table S2 List of primers used in the present study.

Table S3 Amino acid composition and percentege of different amino acids in the full length of GSM1 (amino acids 1-524).

Table S4 Amino acid composition and percentege of different amino acids in GSM1 before the seven repeats (amino acids 1-148).

Method S1 Generation of anti-peptide-1 antibody.

Method S2 Western hybridization to detect GSM1 expression in different tissues.

Method S3 Quantitative real time PCR measurement of *GSM1* transcript level.

Method S4 Immunolocalization of GSM1 in sorghum leaves.

Method S5 Construct preparation for the overexpression of GSM1 in tobacco.

Method S6 Construct preparation and the overexpression of GSM1 in sorghum.

Video S1 Confocal microscopy video clip showing GSM1 localization in the cytoplasmic space and near the cell boundary of silica cells.

Video S2 Confocal microscopy video clip showing the vesicles packed GSM1 fusing to the cell membrane.