The EMBO Journal vol.13 no.19 pp.4676–4683, 1994

A new calcium binding glycoprotein family constitutes a major diatom cell wall component

Nils Kröger[1], Christian Bergsdorf and Manfred Sumper

Lehrstuhl für Biochemie I, Universität Regensburg, 93040 Regensburg, Germany
[1]Corresponding author

Communicated by W.Tanner

Diatoms possess silica-based cell walls with species-specific structures and ornamentations. Silica deposition in diatoms offers a model to study the processes involved in biomineralization. A new wall is produced in a specialized vesicle (silica deposition vesicle, SDV) and secreted. Thus proteins involved in wall biogenesis may remain associated with the mature cell wall. Here it is demonstrated that EDTA treatment removes most of the proteins present in mature cell walls of the marine diatom *Cylindrotheca fusiformis*. A main fraction consists of four related glycoproteins with a molecular mass of approximately 75 kDa. These glycoproteins were purified to homogeneity. They consist of repeats of Ca^{2+} binding domains separated by polypeptide stretches containing hydroxyproline. The proteins in the EDTA extract aggregate and precipitate in the presence of Ca^{2+}. Immunological studies detected related proteins in the cell wall of the freshwater diatom *Navicula pelliculosa*, indicating that these proteins represent a new family of proteins that are involved in the biogenesis of diatom cell walls.

Key words: calcium binding glycoprotein/*Cylindrotheca fusiformis*/diatom cell wall/silica deposition vesicle

Introduction

Diatoms are encased by two siliceous valves which together, along with organic components, comprise the cell wall of these organisms. The outstanding feature of diatoms is the intriguing species-specific design and ornamentation of their silica-based cell walls. The two parts of the cell wall are termed the hypotheca and epitheca, which overlap. Each theca is tightly enclosed by a layer of non-siliceous material (organic casing; Volcani, 1978) and is composed of a valve and several silica strips (girdle bands) which run laterally along the long axis of the cell. It is the girdle band region in which the thecae overlap. The schematic drawing (Figure 1A) shows a cross section through a typical diatom cell wall. In the case of the marine diatom *Cylindrotheca fusiformis* used in this study, the valve area between the raphe (a specialized region including a longitudinal slit through the valve) and the girdle bands is totally unsilicified, thus consisting of purely organic material (Figure 1B).

New valves are produced after cell division and cytokinesis of the mother protoplast. The resulting daughter protoplasts are trapped within the cell wall of the mother cell. If silicate is available, each daughter protoplast forms a new valve. These are formed within the protoplast in a specialized organelle, the silica deposition vesicle (SDV; Volcani, 1981). The SDV membrane (silicalemma) lies closely beneath the plasma membrane and extends during the process of silicate deposition (Volcani, 1981). Finally, the SDV spans the whole area beneath the plasma membrane, containing a new valve. The new valve is transported out of the protoplast and the daughter cells separate, each possessing a maternal epitheca and a newly formed hypotheca.

Up to now, very little has been known about the membrane processes and the proteins involved in valve

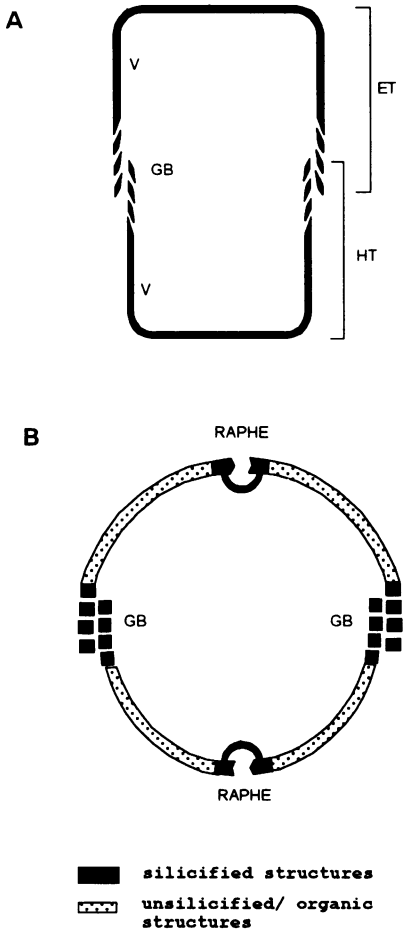

Fig. 1. Schematic drawing of a diatom cell wall. A diatom cell wall in section (**A**), consisting of two thecae: an epitheca (ET) and a hypotheca (HT). The thecae consist of the valve (V) and the girdle bands (GB). (**B**) A cross section through a cell wall of *C.fusiformis*. For a detailed description of the diatom cell wall structure see Volcani (1981) and Reimann *et al.* (1965).

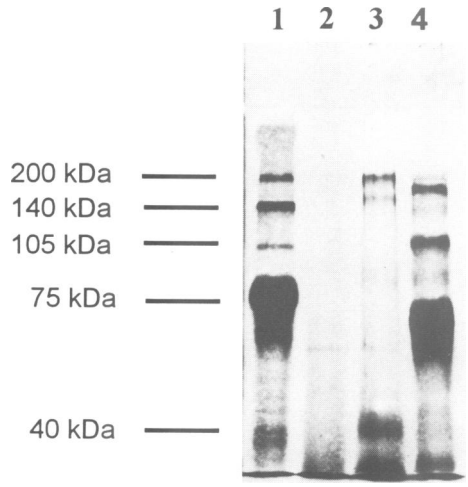

Fig. 2. Sequential extraction of *C.fusiformis* cell walls. Isolated cell walls of 10^7 cells were subjected to successive treatment with EDTA (lane 1), SDS (lane 2), HF (lane 3) or HF and EDTA (lane 4). The complete material of each extract was subjected to 8% SDS−PAGE and silver stained.

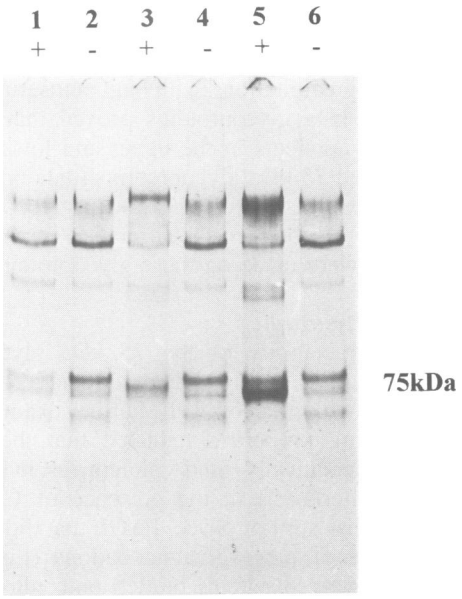

Fig. 3. ^{14}C-labelling of EDTA-extractable cell wall components. Synchronized and silicon-starved cells were labelled with $NaH^{14}CO_3$ for 2 h at various times after addition of silicate as described in the text. Equal fractions of EDTA-extracted cell wall material were analysed by 8% SDS−PAGE and subsequent autoradiography. −, silicon starved cells; +, silicon-starved cells that were resupplied with silicate. Lanes 1 and 2, labelling period 2–4 h after silicate addition to +Si cells; lanes 3 and 4, 4–6 h; lanes 5 and 6, 6–8 h.

formation. The species-specific architecture of the silica structures clearly indicates a genetic determination. It is possible that proteins found in the mature diatom cell wall are components (and structural determinants) of the valve-forming machinery within the SDV. Although amino acid analysis of purified diatom valves has revealed the presence of proteins (Hecky *et al.*, 1973), no diatom cell wall protein has been isolated so far.

Here we present the protein composition of the cell wall of the diatom *C.fusiformis*. As mentioned above, the unique feature of this particular valve is an unsilicified area between the lateral girdle bands and the silicified raphe region. The silica parts of the cell wall are tightly enclosed by organic material which seems to be continuous with the organic material in the unsilicified wall area (Reimann *et al.*, 1965). Thus, *C.fusiformis* offers the possibility of distinguishing between silica-associated and silica-independent organic cell wall components.

We have characterized a set of related *C.fusiformis* cell wall glycoproteins. These glycoproteins exhibit unique structural features and biochemical properties and are members of a new protein family found in diatoms.

Results

Sequential extraction of C.fusiformis cell walls

Diatom cells were completely lysed by vigorous shaking with glass beads. Due to the high density of biogenic silica (r = 2.1 kg/l), the diatom cell wall can be easily removed from a cell lysate by low speed centrifugation. Isolated cell walls were sequentially extracted with EDTA, followed by 1% SDS at 95°C and finally treated with liquid HF to dissolve the silica. Extracts were analysed for the presence of proteins by SDS−PAGE and subsequent silver staining (Figure 2). On lanes loaded with the SDS fractions, almost no bands are visible, indicating the virtual absence of adhering membranes (Figure 2, lane 2). The majority of cell wall proteins appear in the EDTA extract, with major bands at 200, 140, 105 and 75 kDa (Figure 2, lane 1). The HF extract reveals two high molecular mass components at 200 and 155 kDa, as well as two low molecular mass bands at 30 and 40 kDa (Figure 2, lane 4). For comparison of the proteins in EDTA and HF extracts, the EDTA fraction was in addition treated with HF. Each of the proteins in the EDTA extract showed a mobility shift following this treatment, demonstrating glycosylation, since HF specifically cleaves *O*-glycosidic bonds (Mort and Lamport, 1977). As a result, all of the polypeptides in the EDTA and HF extracts migrated with different mobilities on SDS−PAGE, suggesting that the polypeptides of both extracts are different.

^{14}C-Labelling of cell wall components during valve formation

If a diatom culture is incubated in growth medium without any added silicon, cells stop growing within 12 h. The synchronous formation of new hyothecae can be induced by resupplying the medium with silicate (Paul and Volcani, 1976). Within 8 h after addition of silicate, the daughter cells separate, each provided with a complete wall. Changes in composition of the cell wall during valve formation were followed by ^{14}C-labelling of silicate-induced cells.

A silicon-starved culture of *C.fusiformis* was divided into two aliquots. One culture was left under silicon starvation (−Si cells), the other was resupplied with silicate (+Si cells). At 2 h intervals an equal amount of cells was removed from each culture and pulse-labelled for 2 h by adding $NaH^{14}CO_3$. Subsequently, cell walls were prepared and extracted with EDTA. The EDTA extract was analysed by SDS−PAGE and autoradiography (Figure 3).

The −Si cells exhibit constant labelling in the upper 75 kDa region, but only weak labelling in the lower

75 kDa region. The label in the lower 75 kDa region markedly increases during valve formation of the +Si cells. Immunoprecipitation with an antiserum raised against the 75 kDa glycoproteins proved that both the ^{14}C-labelled components in the upper and lower 75 kDa region are indeed 75 kDa glycoproteins (data not shown).

Since the accumulation of the 75 kDa glycoproteins in the cell wall corresponds to the formation of new valves, further studies were made on these glycoproteins.

Calcium affinity studies

The EDTA-extractability of the 75 kDa glycoproteins suggests divalent cation-mediated association of these glycoproteins within the cell wall. Further examination of their ion binding properties revealed that the 75 kDa glycoproteins specifically bind calcium, as indicated by independent criteria: (i) in the presence of $CaCl_2$ they showed a mobility shift on SDS–PAGE that did not occur in the presence of other divalent cations (Figure 4A); (ii) the 75 kDa glycoproteins blotted onto nitrocellulose membrane bound $^{45}Ca^{2+}$ despite the presence of a 10^6 molar excess of Mg^{2+} (Figure 4B, lane 3). The observed intensity of $^{45}Ca^{2+}$ labelling of the 75 kDa glycoproteins was comparable with that of calmodulin (Figure 4B, lanes 1, 3 and 4), whereas equal amounts of BSA and transferrin (Figure 4B, lane 5) did not bind $^{45}Ca^{2+}$ at all under the same conditions. Calcium binding activity is observed in whole cell lysates of *C.fusiformis* (Figure 4B, lane 2), with the most intense signal residing in the 75 kDa region. There are only two additional $^{45}Ca^{2+}$ binding components detectable in the whole cell lysate, which correspond to the 200 kDa and 140 kDa glycoproteins of the EDTA cell wall extract (Figure 4B, lane 1).

Both the Ca^{2+}-specific mobility shift in SDS–PAGE and $^{45}Ca^{2+}$ binding on nitrocellulose membranes are diagnostic of Ca^{2+} binding proteins (Cartaud *et al.*, 1980; Maruyama *et al.*, 1984).

The components present in the EDTA cell wall extract form a precipitate in the presence of Ca^{2+} (Figure 4C, lanes 2 and 5). Precipitation does not occur in the presence of other divalent cations (Figure 4C, lanes 3 and 4). The purified 75 kDa glycoproteins no longer precipitate in the presence of Ca^{2+} (Figure 4A, lane 2), indicating that one or more additional components are involved in this reaction.

Analysis of neutral sugars

Synchronized and silicon-starved cells of *C.fusiformis* were pulse-labelled for 2 h with $H^{14}CO_3^-$ 6 h after the addition of silicate. The EDTA extract of cell walls was subjected to SDS–PAGE and the 75 kDa region was excised from the gel and eluted. Neutral sugars were analysed as their alditol acetates by radio gas chromatography.

Rhamnose, galactose, xylose, glucose, mannose and an unidentified sugar component in the ratio 5:4:4:2:1:4 were found to constitute the neutral sugar components of the 75 kDa glycoproteins.

Purification of the 75 kDa glycoproteins

An EDTA extract of cell walls was subjected to gel filtration through a Superdex 200 column. The fractions containing the 75 kDa glycoproteins were then loaded on a MonoQ column at pH 5 and eluted with a linear NaCl gradient. The elution profile (Figure 5A) reveals separation of the 75 kDa glycoproteins into three fractions: 75Kα (peak 1), 75Kβ (peak 2) and 75Kγ (peak 3) and 75Kδ (peak 3). Whereas 75Kα, 75Kβ and 75Kδ show about the same mobilities on SDS–PAGE, 75Kγ has a significantly lower mobility (Figure 5B). This is due to differences in

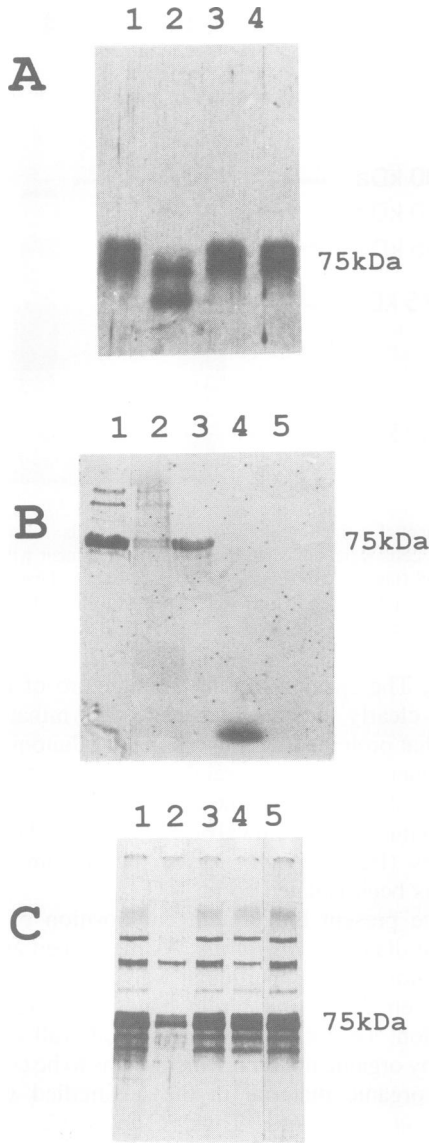

Fig. 4. Ca^{2+} binding assays. (**A**) Ca^{2+}-dependent mobility-shift of 75 kDa glycoproteins on SDS–PAGE. 75 kDa glycoproteins (1 μg) were incubated in the presence of 10 mM solutions of the indicated salt, subjected to 8% SDS–PAGE and silver stained. SDS–PAGE time was twice as long compared with normal SDS–PAGE conditions. Lane 1, EDTA, pH 8; lane 2, $CaCl_2$; lane 3, $MgCl_2$; lane 4, $ZnCl_2$. (**B**) Autoradiogram of a nitrocellulose membrane that was overlaid with $^{45}Ca^{2+}$-containing buffer at pH 6 and subsequently washed. Lane 1, EDTA cell wall extract (10 μg protein); lane 2, whole cell lysate (50 μg protein); lane 3, 5 μg 75 kDa glycoproteins; lane 4, 5 μg calmodulin; lane 5, 5 μg BSA (66 kDa) + 5 μg transferrin (78 kDa). 12% SDS–PAGE. (**C**) Ca^{2+}-dependent precipitation of EDTA-extractable cell wall components. To EDTA cell wall extracts, each containing 2 μg protein in 10 mM Tris–HCl, pH 7.5, 500 mM NaCl buffer, the indicated salt was added to a final concentration of 10 mM. After 30 min incubation at 20°C each sample was centrifuged for 5 min at 14 000 *g*. Supernatants (lanes 1–4) and the pellet (lane 5) were subjected to 8% SDS–PAGE and silver stained. Lane 1, no salt added; lanes 2 and 5, $CaCl_2$; lane 3, $MgCl_2$; lane 4, $ZnCl_2$.

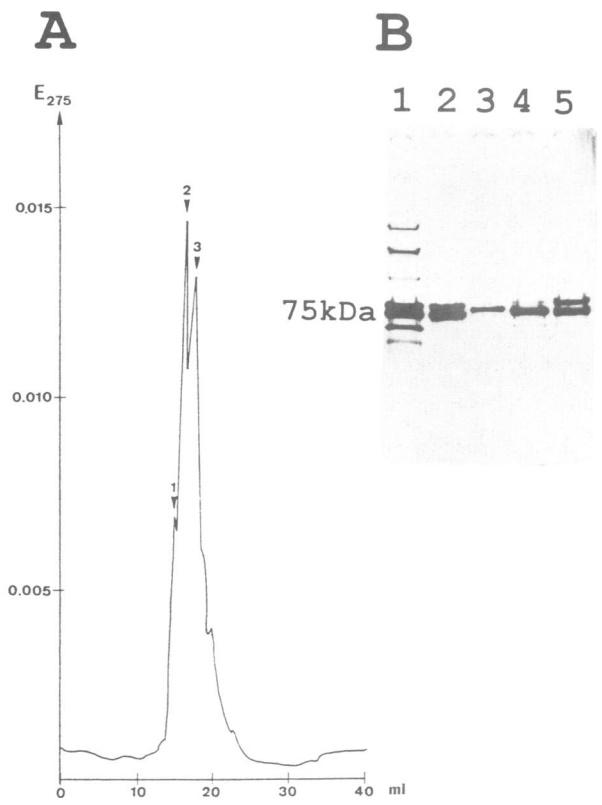

Fig. 5. Purification of 75 kDa glycoproteins. (**A**) Chromatography on MonoQ run at pH 5. Approximately 500 μg of 75 kDa glycoproteins were loaded onto the column. Elution was performed with a linear NaCl gradient (0–500 mM NaCl over 30 min, flow rate 1 ml/min, E_{275} = extinction at 275 nm). (**B**) Silver stained SDS–PAGE showing the purification steps. Equal portions of 75 kDa glycoprotein-containing fractions were loaded onto the gel. Lane 1, EDTA extract of purified cell walls; lane 2, 'fraction 60–65 ml' of Superdex 200 eluate; lane 3, peak 1 of MonoQ eluate (five times concentrated compared with all other fractions); lane 4, peak 2 of MonoQ eluate; lane 5, peak 3 of MonoQ eluate. 8% SDS–PAGE.

the polypeptide backbone of the 75Kγ glycoprotein, since deglycosylated 75Kγ has an apparent molecular weight of 68 kDa in SDS–PAGE, whereas 75Kα, 75Kβ and 75Kδ migrate at 65 kDa (data not shown). Final purification of the 75 kDa glycoproteins was achieved by native PAGE of the corresponding MonoQ eluates, the bands of the individual glycoproteins being excised and subsequently eluted into buffer.

Protein chemistry studies

The N-termini of the glycoproteins 75Kβ, 75Kγ and 75Kδ were analysed by automated Edman degradation. While the N-terminus of 75Kγ was blocked, 75Kβ and 75Kδ showed the same N-terminal sequence:

75Kβ: X A E I P I V K V
75Kδ: X A E I P I V K

The purified glycoproteins 75Kγ, 75Kδ and 75Kβ were separately digested with endoproteinase LysC and the peptides separated by reversed phase HPLC. The individual 75 kDa species showed very similar peptide patterns; in the case of 75Kδ and 75Kβ they were virtually identical. The material from well-separated peaks were subjected to amino acid sequence analysis on an automated gas phase sequencer. The amino acid sequence data obtained are summarized in Table I. The peptides of 75Kβ and 75Kγ are strikingly homologous, or even identical to each other. Remarkably, peptide L6 contains hydroxyproline, a well known constituent of cell wall proteins of algae and higher plants (Showalter, 1993).

Table I. Peptide sequences of 75 kDa glycoproteins

Peptide	Amino acid sequence
LysC peptides of 75Kβ	
L1	V G E E P A T P L G R X E G D X D K
L2	X F Q R D G G E S V P G X S G G S X D
L3	X F Q R T E N F K
L4	Q A E I P I V K
L5	X S D X G S G L K
L6	E N *P* S X *P P* X X *P* A *P* A X T T T S P G X L
LysC peptides of 75Kγ	
L7	V G E E P A T P L G R X E G D X X K (100% identity to L1)
L8	M G E E P A T P L G X X E (85% identity to L1)
L9	X Y Q R T E N F K (88% identity to L3)
L10	X S D C G S G L K (88% identity to L4)
Tryptic peptides of 75 kDa glycoprotein mixture	
T1	V G A N P G V P L E R
T2	A T D Y D F P L G L X E G D C D D
T3	A V P G X S G G A Q D S S F F D Y X V R
T4	M G E E P A T P L G X
T5	V G E E P A T P L G R
T6	X D Y X I P K
T7	T D Y X I V K

P = hydroxyproline; X = animo acid unknown.

The sequence information obtained from the LysC peptides of the purified 75 kDa glycoproteins did not allow us to design appropriate oligonucleotide primers for PCR. Therefore, we performed a tryptic digestion of the 75 kDa glycoprotein mixture. Separation and sequencing of the resulting peptides was as described above. The sequences of the tryptic peptides are summarized in Table I.

Amplification of a cDNA sequence by polymerase chain reaction (PCR)

The amino acid sequence of the tryptic peptide T3 (Table I) was used to synthesize degenerated oligonucleotide primers. The sense primer 5′-GCN GTN CCN GG-3′ was designed from amino acid positions 1–4. The antisense primer 5′-GTA RTC RAA RAA-3′ was designed from amino acid positions 14–17. The latter primer was used to reverse transcribe total RNA of *C.fusiformis*. The resulting cDNA was amplified by PCR using both the sense and antisense primers. Thirty cycles of amplification produced a cDNA fragment of 51 bp in length, which was cloned and sequenced. The deduced amino acid sequence matched the sequence of peptide T3, with cysteine at amino acid position 5. Sequence information from this 51 bp fragment was used to collect the complete cDNA sequence by the RACE–PCR technique (Frohman *et al.*, 1988). Cloning of the PCR product covering the 5′ region revealed an open reading frame matching the N-terminal sequences of 75Kβ and 75Kδ. The PCR product covering the 3′ region was 1375 bp in length, revealing an open reading frame that encoded 12 out of the 17

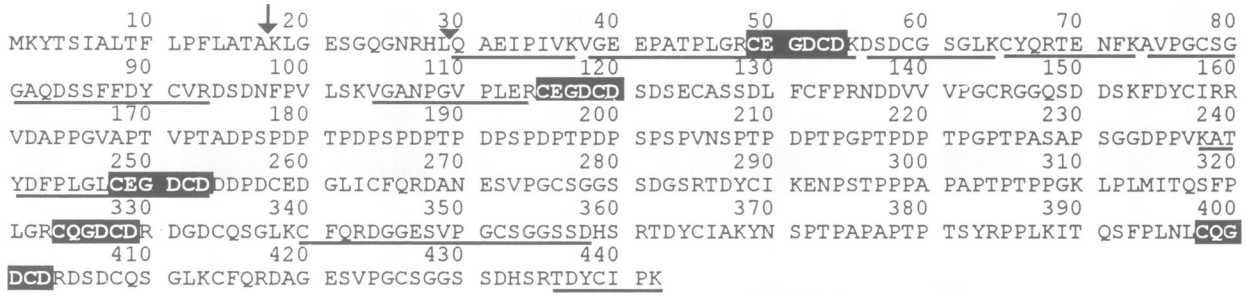

Fig. 6. Amino acid sequence of p75K. The putative signal peptidase cleavage side is marked with an arrow. The start of the mature polypeptide as derived from N-terminal amino acid sequence analysis is marked with an arrowhead. Amino acid sequences derived from isolated peptides are underlined. The hexapeptide sequence CE/QGDCD occurring once in each ACR domain is highlighted.

peptides obtained by proteolytic digestion of the 75 kDa glycoproteins (Table I). In order to rule out the possibility that the products of 5′ RACE−PCR and 3′ RACE−PCR were derived from different, but highly homologous genes, we performed three nested 5′ RACE−PCR reactions using primers that correspond to sequences near the 3′ end of the gene. The products of the second and third PCR reactions were cloned and sequenced. Both match the sequence of the previously cloned 5′ end. As a result, the products of the RACE−PCR reactions in both directions were derived from the same cDNA molecule, which we denoted as 75KcDNA.

Primary structure of the 75KcDNA-encoded polypeptide (p75K)

75KcDNA reveals an open reading frame coding for 442 amino acids (Figure 6). Amino acid positions 30–38 are identical to the N-termini of 75Kβ and 75Kδ, so that amino acids 1–29 are likely to act as a leader peptide(s) necessary to direct the protein into the cell wall. The first 17 amino acids of this pre-sequence represent a typical signal sequence allowing for the import of a nascent polypeptide into the lumen of the endoplasmatic reticulum. It contains a positively charged N-terminal end (K at amino acid position 2), followed by a stretch of uncharged and hydrophobic amino acids (amino acid positions 3–14). The tripeptide sequence A T A (amino acid positions 15–17) shows the typical motif for a signal peptide cleavage site (Perlman and Halvorsson, 1983). Consequently, amino acids 18–29 are likely to encode an additional, as yet unknown, pre-sequence. It contains three charged residues, a histidine residue and is mainly hydrophilic.

The main feature of the 75KcDNA-encoded polypeptide (p75K) is its highly repetitive structure. The hexapeptide motif CE/QGDCD occurs five times and is evenly distributed within the polypeptide. In each case the hexapeptide is embedded in a domain consisting of 53–56 amino acid residues that is rich in aspartate, glutamate and cysteine. All five domains [1 (38–93) to 5 (389–442)] share a high degree of sequence identity (43–87%) to each other (see Figure 7A). With respect to these features, we named them acidic cysteine-rich domains (ACR domains). ACR domains 2 (104–159) and 3 (239–291), ACR domains 3 and 4 (315–368) and ACR domains 4 and 5 (389–442) are connected by polyproline stretches containing proline in almost every other amino acid position. Figure 7B shows a schematic diagram of the polypeptide structure.

p75K is very acidic, with a predicted pI of 4. This property is in good agreement with the nature of the 75 kDa glycoproteins, since these firmly bind to MonoQ anion exchange matrix at pH 5. The deduced polypeptide includes all peptides obtained from the 75 kDa glycoprotein fractions except peptides T4, T7, L3, L6 and L8. However, there are sequences in p75K highly homologous to these missing peptides. For instance, the hydroxyproline-containing peptide L6 is highly homologous to the amino acid region 292–311 (*P* = hydroxyproline; X = amino acid unknown; − = space)

peptide L6: E N *P* S X *P P* X X *P* A *P* A X T T T S P G X L
aa 292–311: E N P S T P P − − P A P A P T P T P P G K L

This fact clearly indicates that the 75 kDa glycoprotein fraction represents a family of closely related polypeptides. p75K is representative of this family, as it contains the essence of all the peptide information available.

The deglycosylated polypeptides of the 75 kDa glycoproteins have apparent molecular masses on SDS−PAGE of 65 and 68 kDa respectively. These values are significantly larger than the predicted molecular weight of 46 kDa for p75K. This discrepancy was resolved by expression of a truncated form of 75KcDNA encoding amino acid positions 30–442 (the mature polypeptide) in *Escherichia coli*. This recombinant polypeptide shows the same mobility on SDS−PAGE as the deglycosylated forms of 75Kα, 75Kβ and 75Kδ (data not shown), indicating that the highly acidic polypeptides exhibit irregular electrophoretic mobilities.

Immunological studies

Polyclonal antisera against the 75 kDa glycoproteins were obtained by immunizing three rabbits with 75Kα, 75Kγ and 75Kδ respectively. As expected, the antisera cross-react with each of the purified 75 kDa glycoproteins and with all other glycoproteins of the EDTA cell wall extract, as shown by Western blot analysis (Figure 8A, lane 2). This cross-reactivity remains even after HF treatment of the EDTA extract (Figure 8A, lane 3). Therefore, the polypeptide backbones of all EDTA extractable cell wall glycoproteins share similar epitopes, again confirming the existence of a protein family. Unfortunately, the high degree of cross-reactivity among these glycoprotein species excluded studies to localize individual species in the cell wall.

The antisera raised against the 75 kDa glycoproteins of

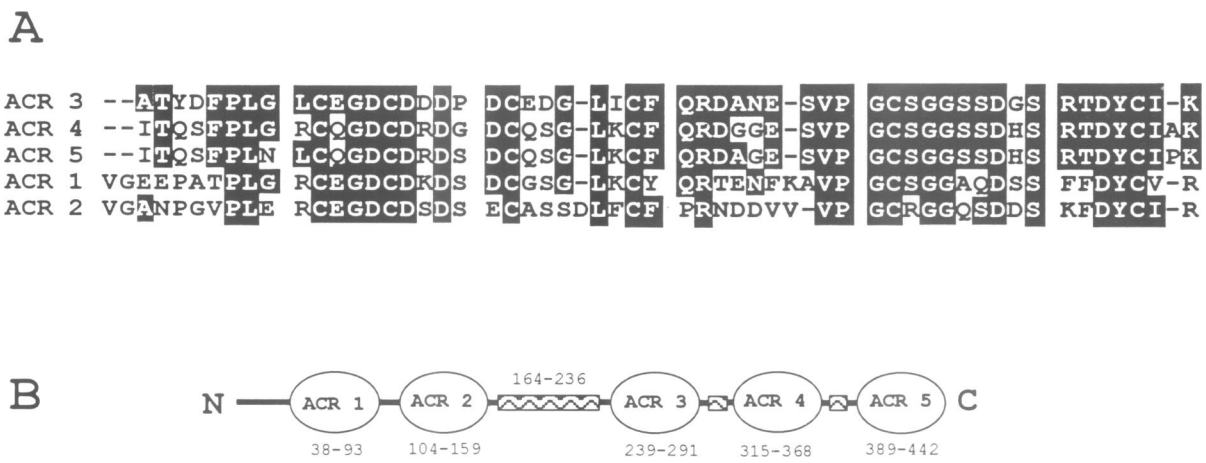

Fig. 7. Structure of the 75KcDNA-encoded polypeptide. (**A**) Alignment and comparison of the ACR domains. Only identities with the central domain (ACR 3, top line) are shown. (**B**) Schematic representation of the polypeptide structure. Hatched boxes indicate the polyproline streches. Amino acid positions of the polypeptide segments are specified.

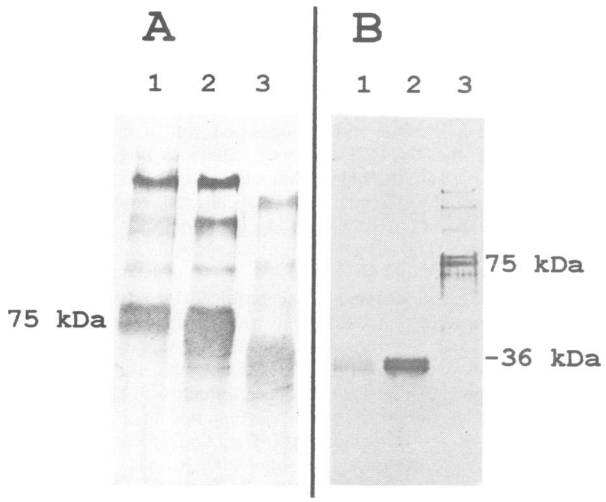

Fig. 8. Western blot analysis. (**A**) A 1:1000 dilution of the antiserum raised against 75Kγ was used. Lane 1, whole cell lysate of *C.fusiformis* (10 μg protein); lane 2, EDTA cell wall extract of *C.fusiformis* (1 μg protein); lane 3, HF-treated EDTA cell wall extract of *C.fusiformis* (1 μg protein). 8% SDS−PAGE. (**B**) A 1:200 dilution of the antiserum raised against 75Kγ was used. Lane 1, whole cell lysate of *N.pelliculosa* (10 μg protein); lane 2, EDTA cell wall extract of *N.pelliculosa* (0.5 μg protein); lane 3, EDTA cell wall extract of *C.fusiformis* (0.1 μg protein). 10% SDS−PAGE.

C.fusiformis specifically detect two proteins in an EDTA extract of purified cell walls from the freshwater diatom *Navicula pelliculosa* (Figure 8B, lane 2). These proteins have an apparent molecular mass of 36 and 34 kDa respectively. A tryptic peptide (Np.1) derived from the purified 36 kDa protein confirms the close relationship to the 75 kDa glycoproteins. Amino acids 1–15 of this peptide are strikingly homologous (60% amino acid sequence identity) to amino acids 345–359 of p75K.

Np.1 1–15:	X G T E S V P G X S G S G S X
p75K 344–359:	D G G E S V P G C S G G S S D

It is therefore apparent that the 36 kDa cell wall protein of *N.pelliculosa* is a member of the same protein family.

Discussion

Sequential extraction of *C.fusiformis* cell walls reveals two ways in which proteins are incorporated into the valve. One way seems to work in a calcium-dependent manner, since the EDTA-extractable 75 kDa glycoproteins specifically bind calcium. The second way is sensitive to HF treatment and thus may involve intimate association of the proteins with the silica parts of the cell wall. The HF-soluble proteins are putative components of an organic matrix within the silica, as is postulated for the cell wall of the diatom *Cyclotella meneghiniana* by Swift and Wheeler (1992).

Due to the striking homology of their polypeptides and their immunological similarity, the 75 kDa glycoproteins, as well as the remaining EDTA-extractable cell wall glycoproteins, would seem to be members of a protein family. The occurence of a related protein in the EDTA cell wall extract of the diatom *N.pelliculosa* indicates that these proteins are general components of diatom cell walls and not species-specific molecules connected to the unique valve architecture of *C.fusiformis*. In contrast to this diatom, *N.pelliculosa* possesses a highly silicified cell wall that is much more typical of diatoms. Thus it may be possible that differences in valve architecture of different diatom species are reflected in different biochemical properties of the members of this protein family.

p75K may represent the general structure of the members of this protein family. It appears to be composed of only two structural elements: (i) an ACR domain of ~55 bp in length containing the CE/QGDCD hexapeptide motif and (ii) a polyproline domain of variable length. Both domains may also be present in the primary structures of all members of this protein family, but may vary with regard to repeat number and arrangement. This is suggested by the modular structure of p75K, as shown in Figure 7B.

A common property of the members of this protein family is their affinity for calcium (see Figure 4B, lane1). We propose that the calcium binding sites reside within the ACR domains. These domains are similar to the EF hand motifs of calmodulin and troponin c (Strynadaka and James, 1989).

Amino acid position	1 5 10
ACR consensus sequence	C E G D C D X D X D C X S
EF hand consensus sequence (B = D or N)	X D X B X B X X X X X X E

The N-terminal regions of both consensus sequences contains acidic amino acid residues (mostly aspartate) in every other amino acid position. The amino acid residues at positions 2, 4, 6, 8 and 13 of the EF hand consensus sequence provide oxygen ligands to the Ca^{2+} ion. The glutamate residue at amino acid position 13 is conserved in all known EF hand motifs. In the ACR consensus sequence this position is occupied by a serine residue. As a consequence, this sequence may not be regarded as an EF hand motif. Nevertheless, it is plausible to ascribe Ca^{2+} binding competence to the ACR consensus sequence, since it contains potential oxygen ligands to the Ca^{2+} ion at the appropriate positions.

A striking feature of p75K is the sequence of 29 amino acids preceding the mature polypeptide. We speculate that this pre-sequence is necessary for targetting the 75 kDa glycoproteins to the cell wall. The putative signal sequence encoded by amino acids 1–17 may mediate transfer of the polypeptide into the lumen of the endoplasmatic reticulum. After transport to the Golgi apparatus and glycosylation, the second part of the pre-sequence (amino acids 18–29) may be responsible for sorting the 75 kDa glycoproteins to the SDV, the site of cell wall biogenesis. This model implies that the SDV is connected to the Golgi apparatus via transport vesicles. In order to test the hypothesis of consecutive sorting, we are currently investigating whether the immature 75 kDa glycoproteins are processed in two successive steps *in vivo*. Although the connection between the Golgi apparatus and SDV has yet to be elucidated, there is evidence that development of the SDV depends on glycosylation events in the endoplasmatic reticulum and Golgi apparatus (Lee and Li, 1992).

Interestingly, acidic and calcium binding proteins are known to be involved in the biomineralization of calcified structures, such as the $CaCO_3$ shells of molluscs and foraminifera, as well as teeth and bone, which are made of hydroxyapatite, $Ca_5(PO_4)_3OH$ (Weiner, 1985). It has been shown that these proteins are potent regulators of crystal growth by inhibiting the crystallization process (Wheeler *et al.*, 1988; Weiner and Addadi, 1991). In contrast to the crystalline structure of calcified tissues, the cell walls of diatoms contain amorphous, hydrated SiO_2 and lack Ca^{2+} as a component of the mineral. In this respect it is astonishing that similar types of molecules, namely calcium binding acidic glycoproteins, are recruited for the formation of the mineralized structure. This similarity may be coincidental and the 75 kDa glycoproteins could serve solely as structural components that are incorporated into the cell wall via calcium bridges. On the other hand, analogously to the function of the proteins in calcified tissues, the 75 kDa glycoproteins could define the molecular constraints for silicate polymerization in the SDV. In this model, silicon could not be deposited at sites where the acidic glycoproteins are located, which would result in a species-specific patterning of the silica structures. Since the 75 kDa glycoproteins, together with the other EDTA-extractable cell wall glycoproteins, precipitate in the presence of Ca^{2+} *in vitro* (see Figure 4C, lanes 2 and 5), silica deposition would be inhibited by a macromolecular framework of acidic glycoproteins that are non-covalently cross-linked by calcium bridges.

Materials and methods

Culture conditions

An axenic culture of *C.fusiformis*, kindly provided by B.E.Volcani (Scripps Institute for Oceanography, UCSD, La Jolla, CA), was grown in ASW medium at 20°C and 5000–10 000 lux (Darley and Volcani, 1969).

Synchronization of cell wall formation was achieved as follows. Cell cultures ($5–7.5\times10^5$ cells/ml) were successively subjected to a 16 h dark/6 h light/16 h dark rhythm. After the final dark period the cells were centrifuged for 10 min at 2500 *g*, washed three times with silicate-free ASW medium (ASW−Si), resuspended in ASW−Si medium to 5×10^5 cells/ml and cultured for 24 h in the light. Silicate was then added to a final concentration of 500 µM to initiate cell wall formation.

Purification of the 75 kDa glycoproteins

Harvested cells (approximately 15 g wet weight) were washed and resuspended in ten portions of 10 ml 1 mM $CaCl_2$. To each portion about 15 g glass beads (0.3 mm in diameter) were added and cell lysis was achieved by shaking each portion on a Heidolph LAX2000 vortexer at maximum speed for 6 min. Lysed cells were then centrifuged at 600 *g* for 5 min. The pellet containing the cell walls was resuspended in 1 mM $CaCl_2$ and centrifuged again at 600 *g* for 5 min. This procedure was repeated until the supernatant was colourless. The raw cell walls were pooled, resuspended in four portions of 10 ml 1 mM $CaCl_2$ and then sonicated using a Branson sonifier equipped with a microtip at half maximum power for 45 s. Cell walls were again centrifuged and washed as described above until the supernatant was colourless. The final pellets consisted of pure cell walls (modification of cell wall preparation method previously described by Coombs and Volcani, 1968).

For preparative EDTA extraction, purified cell walls were resuspended in four portions of 15 ml 100 mM EDTA, pH 8, and shaken at 4°C for 12–24 h. The suspensions were then centrifuged for 10 min at 3000 g. The supernatants were kept and the cell walls were pooled, resuspended in 50 ml H_2O and centrifuged as above. The supernatants of both centrifugation steps contained the EDTA-extractable cell wall proteins and thus were pooled. The yield was 5 mg of cell wall protein.

The EDTA cell wall extract was concentrated to 2 ml on Diaflo membrane YM 10 (10 kDa cut-off) using an Amicon concentrator. The concentrate was loaded on a Superdex 200 HiLoad 16/60 column (Pharmacia) using 20 mM Tris−HCl, pH 8, 100 mM NaCl as the running buffer at a flow rate of 1 ml/min. Fractions containing the 75 kDa glycoproteins were pooled and loaded on a MonoQ HR 5/5 column (Pharmacia) which was equilibrated in 20 mM sodium phosphate, pH 5, (buffer A). After washing the column with more than two column volumes of buffer A, elution was achieved by application of a linear chloride gradient (0–500 mM NaCl in buffer A over 30 min) at a flow rate of 1 ml/min. The 75 kDa glycoproteins elute between 200 and 250 mM NaCl (75Kα, 200 mM NaCl; 75Kβ, 220 mM NaCl; 75Kγ and 75Kδ, 245 mM NaCl). Each 75 kDa glycoprotein fraction was subjected to an individual native PAGE [discontinous electrophoresis according to Laemmli (1970), but without SDS and mercaptoethanol in the buffer]. Each 75 kDa glycoprotein was cut out of the gel and eluted by diffusion into 10 mM Tris−HCl, pH 7.5, 1 mM NaCl buffer. The yield was 5 µg 75Kα, 100 µg 75Kβ, 100 µg 75Kδ and 25 µg 75Kγ.

Pulse-labelling

A normally grown *C.fusiformis* culture (400 ml) was silicon starved as described under culture conditions. At the end of the starvation period the cell density was $5–7.5\times10^5$ cells/ml. The culture was then divided into two 200 ml portions. To one portion, 1 ml 100 mM Na_2SiO_3 was added (+Si cells) and the pH was quickly adjusted to 7.8 by adding 2 M HCl. No addition was made to the other 200 ml portion (−Si cells). At various times, 50 ml aliquots of each portion were removed and washed three times with ASW or ASW−Si medium. Finally, washed cells were resuspended in 1 ml of the corresponding medium, 40 µCi $NaH^{14}CO_3$ were added to each of the cultures and the cultures were incubated for 2 h on a shaker at 10 000 lux. At the end of the pulse period, cells were washed three times with ASW or ASW−Si medium, harvested and stored at −20°C. The preparation of cell walls and EDTA extracts were performed as described above.

Calcium affinity studies

Experiments were done with a mixture of all four 75 kDa glycoprotein species. Mobility shift experiments were performed as described in Cartaud *et al.* (1980). Studies of $^{45}Ca^{2+}$ binding to 75 kDa glycoproteins blotted onto nitrocellulose membranes were done according to Maruyama *et al.* (1984).

Sugar analysis

Synchronized *C.fusiformis* cells (50 ml, 5×10^6 cells/ml) were pulse-labelled with 0.4 mCi $NaH^{14}CO_3$ 6–8 h after addition of silicate. EDTA extraction of cell walls was performed as described above and the extracts subjected to SDS–PAGE. Radiolabelled 75 kDa glycoproteins were excised from the gel and eluted by diffusion. Neutral sugar composition was determined by radio gas chromatography of the alditol acetates according to Wenzl and Sumper (1981).

Proteolytic digestion and separation of peptides

An EDTA cell wall extract containing about 100 µg 75 kDa glycoprotein mixture was subjected to SDS–PAGE. The gel was stained with Coomassie blue and destained overnight, the region of 75 kDa was excised and the gel slice was cut into small pieces. Gel pieces were treated as follows: 2×1 h 30% methanol, 7% acetic acid; 3×1 h 50% methanol, 10% acetic acid; 2×1 h 90% ethanol. Gel pieces were dried *in vacuo*, subsequently soaked in 1 ml 10 µg trypsin/ml 0.2 M NH_4HCO_3 and incubated at 37°C overnight. The suspension was centrifuged and the supernatant was kept. After three subsequent washing steps (2× 0.2 M NH_4CO_3, 1× 0.2 M NH_4CO_3, 50% acetonitrile), the supernatant of the first centrifugation step and the wash eluates were pooled, filtered and dried *in vacuo*. The lyophilysate containing the peptides was dissolved in 6 M guanidinium hydrochloride and stored until separation was performed.

Seventy-five kDa glycoproteins 75Kβ, 75Kγ and 75Kδ (50 µg each), purified by SDS–PAGE, were separately dissolved in 200 µl 50 mM Tris–HCl, pH 8, and digested with 2 µg endoproteinase LysC at 37°C overnight. Peptides were separated by reversed phase HPLC on a Vydac 218 TP 5µm column (MZ Analysentechnik, Mainz, Germany) as described by Ertl *et al.* (1989).

Amplification of 75 K cDNA by PCR

The amino acid sequence information of the peptide AVPGXSGG-AQDSSFFDYXVR was used to synthesize the antisense oligonucleotide primer 5′-GTA RTC RAA RAA-3′ (corresponding to amino acid sequence FFDY). This primer was used to reverse transcribe RNA. The sense primer 5′-GCN GTN CCN GG-3′ (corresponding to amino acid sequence AVPG of the same peptide) was used to amplifiy the resulting cDNA by PCR. Forty cycles of amplification generated a cDNA fragment of 51 bp in length that was cloned into the *Sma*I site of pUC18 by blunt-end ligation.

RNA was extracted from the diatom cells by the method of Kirk and Kirk (1985). Reverse transcription was performed in a final volume of 20 µl containing 50 mM Tris–HCl, pH 8.3, 40 mM KCl, 6 mM $MgCl_2$, 1 mM dithiothreitol, 20 U RNAguard (Pharmacia), 400 pmol antisense oligonucleotide, 1 mM each dNTP, 1 µg RNA and 200 U MoMuLV reverse transcriptase. Incubation was at 37°C for 60 min. After addition of 80 ml 1× PCR buffer containing 1200 pmol sense and 400 pmol antisense oligonucleotide and 2.5 U Taq polymerase, thermal cycling (Perkin-Elmer Cycler 9600) was initiated (40 cycles: denaturation at 94°C for 15 s, annealing at 40°C for 20 s and extension at 72°C for 5 s).

The sequences of the remaining 5′ and 3′ stretches were established by the RACE–PCR technique (Frohman *et al.*, 1988). PCR products were blunt-end ligated into the *Sma*I site of pUC18 and sequenced. DNA sequencing was performed by the dideoxy method (Sanger *et al.*, 1977) using T7 DNA polymerase (Pharmacia). Synthetic oligonucleotides were used to sequence the cDNA in both directions.

Expression of mature 75K polypeptide in E.coli

A cDNA fragment encoding the mature 75K polypeptide (amino acid positions 30–442) was generated, cloned into the *Nde*I site of pET11a and expressed in *E.coli* following the method of Studier *et al.* (1990).

Acknowledgements

We wish to thank Dr R.Deutzmann for sequencing peptides. We are indebted to Dr J.Ortiz for critically reading the manuscript. We gratefully acknowledge invaluable support of our work by Prof. B.E.Volcani. In addition, Prof. B.E.Volcani provided us with samples of purified cell wall from *C.fusiformis* and *N.pelliculosa* for comparison with our cell wall preparations.

References

Cartaud,A., Ozon,R., Walsh,M.P., Haiech,J. and Demaille,J.G. (1980) *J. Biol. Chem.*, **255**, 9404–9408.
Coombs,J. and Volcani,B.E. (1968) *Planta (Berl.)*, **82**, 280–292.
Darley,W.M. and Volcani,B.E. (1969) *Exp. Cell Res.*, **58**, 334–342.
Ertl,H., Mengele,R., Wenzl,S., Engel,J. and Sumper,M. (1989) *J. Cell Biol.*, **109**, 3493–3501.
Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 8998–9002.
Hecky,R.E., Mopper,K., Kilham,P. and Degens,T.E. (1973) *Mar. Biol.*, **19**, 323–331.
Kirk,M.M. and Kirk,D.L. (1985) *Cell*, **41**, 419–428.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Lee,M. and Li,C.W. (1992) *Bot. Bull. Acad. Sin.*, **33**, 317–325.
Maruyama,K., Mikawa,T. and Ebashi,S. (1984) *J. Biochem.*, **95**, 511–519.
Mort,A.J. and Lamport,D.T.A. (1977) *Anal. Biochem.*, **82**, 289–309.
Paul,J.S. and Volcani,B.E. (1976) *Arch. Microbiol*, **110**, 247–252.
Perlman,D. and Halvorsson,H.O. (1983) *J. Mol. Biol.*, **167**, 391–409.
Reimann,B.E.F., Lewin,J.C. and Volcani,B.E. (1965) *J. Cell Biol.*, **24**, 39–55.
Sanger,F.S., Nicklein,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
Showalter,A.M. (1993) *Plant Cell*, **5**, 9–23.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
Strynadaka,N.C.J. and James,M.G.N. (1989) *Annu. Rev. Biochem.*, **58**, 951–998.
Swift,D.M. and Wheeler,A.P. (1992) *J. Phycol.*, **28**, 202–209.
Volcani,B.E. (1978) In Bendz,G. and Lindquist,I. (eds), *Biochemistry of Silicon and Related Problems*. Plenum Publishing, New York, NY, pp.177–204.
Volcani,B.E. (1981) In Simpson,T.L. and Volcani,B.E. (eds), *Silicon and Siliceous Structures in Biological Systems*. Springer-Verlag, Berlin, Germany, pp. 157–200.
Weiner,S. (1985) *CRC Crit. Rev. Biochem.*, **20**, 365–408.
Weiner,S. and Addadi L. (1991) *Trends Biochem. Sci.*, **16**, 252–256.
Wenzl,S. and Sumper,M. (1981) *Proc. Natl Acad. Sci. USA*,**78**, 3716–3720.
Wheeler,A.P. Rusenko,K.W., Swift,D.M. and Sikes,C.S. (1988) *Mar. Biol.*, **98**, 71–80.

Received on May 30, 1994; revised on July 14, 1994