⌘ Author's Choice

© 2013 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

Glycoproteome of Elongating Cotton Fiber Cells*[S]

Saravanan Kumar‡, Krishan Kumar‡, Pankaj Pandey‡, Vijayalakshmi Rajamani‡, Kethireddy Venkata Padmalatha§, Gurusamy Dhandapani§, Mogilicherla Kanakachari§, Sadhu Leelavathi‡, Polumetla Ananda Kumar§, and Vanga Siva Reddy‡¶

Cotton ovule epidermal cell differentiation into long fibers primarily depends on wall-oriented processes such as loosening, elongation, remodeling, and maturation. Such processes are governed by cell wall bound structural proteins and interacting carbohydrate active enzymes. Glycosylation plays a major role in the structural, functional, and localization aspects of the cell wall and extracellular destined proteins. Elucidating the glycoproteome of fiber cells would reflect its wall composition as well as compartmental requirement, which must be system specific. Following complementary proteomic approaches, we have identified 334 unique proteins comprising structural and regulatory families. Glycopeptide-based enrichment followed by deglycosylation with PNGase F and A revealed 92 unique peptides containing 106 formerly *N*-linked glycosylated sites from 67 unique proteins. Our results showed that structural proteins like arabinogalactans and carbohydrate active enzymes were relatively more abundant and showed stage- and isoform-specific expression patterns in the differentiating fiber cell. Furthermore, our data also revealed the presence of heterogeneous and novel forms of structural and regulatory glycoproteins. Comparative analysis with other plant glycoproteomes highlighted the unique composition of the fiber glycoproteome. The present study provides the first insight into the identity, abundance, diversity, and composition of the glycoproteome within single celled cotton fibers. The elucidated composition also indirectly provides clues about unicellular compartmental requirements underlying single cell differentiation. ***Molecular & Cellular Proteomics 12: 10.1074/mcp.M113.030726, 3677–3689, 2013.***

Cotton fibers are single-cell epidermal seed trichomes that undergo major developmental changes involving overlapping stages of growth including initiation, elongation, and maturation (1, 2). Matured fibers contain ~95% cellulose and 1.8% protein (3), whereas elongating fibers contain 22% protein in its primary cell wall (4). The relatively increased protein content and its consistency throughout the elongation phase correlates with its stage-specific compartmental requirement (5). Elongation is the most active and vigorous phase, during which the cell extends between 2 to 6 cm in length at a rate of >2 mm/day (1, 2). Increase in length involve the expansive deformation of the cell wall, including loosening, expansion, and remodeling. These processes collectively determine the cell wall's yielding properties and are governed by the cellulose microfibril-matrix network and associated factors, such as wall bound structural proteins and interacting enzymes (6). These proteins play crucial roles in the elongation and maturation of numerous fiber cells on the ovule surface in a synchronized fashion (7).

Earlier efforts to understand fiber cell differentiation showed the stage-specific expression of genes encoding cell wall enzymes (8), implicating their probable role in cell elongation and post elongation events (9). Furthermore, experimental data from other plant systems highlight the roles of carbohydrate active enzymes (CAZymes)[1], such as xyloglucan endotransglycosylases/hydrolases (XETs/XTHs) (10, 11), glucanases (10, 12), glycosyl transferases (GTs) (13, 14), and pectin methyl esterases (PMEs) (15, 16), in the wall modification occurring during cell development. Most of the earlier mentioned functions were suggested based on transcriptase, molecular biology or biochemical tools. Transcript level information does not reflect the structure, function or abundance of their gene products. In addition to CAZymes, genes encoding structural proteins, such as arabinogalactans (AGPs) and fasciclin-like arabinogalactans (FLAs), have been shown to play crucial roles in fiber development (17). AGPs are also known to act as signaling molecules, modulators of cell wall mechanics, pectin plasticizers (18), and stimulators of XET activity (19) and are also involved in pattern formation (20). Despite their diverse roles (21), experimental evidence concerning the het-

From the ‡Plant Transformation Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India; §National Research Centre on Plant Biotechnology (NRCPB), IARI, New Delhi, India

⌘ *Author's Choice*—Final version full access.
Received May 5, 2013, and in revised form, September 4, 2013
Published, MCP Papers in Press, September 9, 2013, DOI 10.1074/mcp.M113.030726

[1] The abbreviations used are: CAZymes, Carbohydrate active enzymes; Con A, Concanavalin A; LAC, Lectin affinity chromatography; AGPs, Arabinogalactan proteins; FLA,- Fasciclin-like arabinogalactan proteins; GPAs, Glycosylphosphatidylinositol anchored proteins; XETs/XTHs, Xyloglucan endotransglycosylases/hydrolases; PNGase, Peptide N-glycoamidase; MWCO, Molecular weight cut off, FASP, Filter aided sample preparation; SpC, Spectral counts.

erogeneity and abundance that governs the roles of the AGP family members is still emerging.

Proteomic studies employed so far to understand fiber development display an overview of key metabolic events based on the expression pattern of high abundant and detectable proteins from whole fiber extracts (5, 13, 22). However, no insight into the cell wall bound enzymes and structural proteins have been gathered through proteomic approaches. The majority of such CAZymes and structural proteins are known to be secreted, destined for the cell wall and *N*-linked glycosylated in plants (23). The glycosylation status of such proteins leads to an extended or altered conformation, which in turn is essential for crosslinking to the cell wall matrix and the strengthening of the cell wall (24). Therefore, the glycoproteome indirectly represents the proteome composition of plant cell walls and may also reflect the system specific functional properties of the wall (23, 25). Exploring the glycoproteome of cotton fibers may also provide interesting clues about the single cell compartmental makeup and the probable role of these proteins during development. To our knowledge, such studies have not been performed in cotton fibers. In this context, we have characterized the glycoproteome of the cotton fiber employing lectin affinity chromatography (LAC) followed by protein identification using complementary proteomic approaches. Our study provides evidence for the identity, abundance, heterogeneity, and novel forms of glycoproteins including cell wall destined AGPs, FLAs, and CAZymes. Comparative analysis with known plant glycoproteome data sets highlighted the unique compositional makeup of the fiber glycoproteome. Further validation using quantitative real time PCR (qRT-PCR) of the glycoprotein encoding genes revealed their stage and isoform specific expression profiles, suggesting these genes may play a regulated role in the developmental process.

EXPERIMENTAL PROCEDURES

Plant Materials—Cotton plants (*Gossypium hirsutum* cv. Coker 310) were grown in a climate controlled greenhouse. Bolls were excised from the plants during the elongation stages (5–15 days post anthesis, dpa), and fibers were carefully removed from the ovule, frozen immediately in liquid nitrogen, and stored until use.

Protein Extraction—Cotton fibers were ground into a fine powder in liquid nitrogen, along with 10% polyvinyl polypyrrolidone (PVPP) and 10% silicon dioxide (SiO_2), in a prechilled mortar and were suspended in extraction buffer containing 25 mM Tris (pH 7.5), 0.2 M $CaCl_2$, 0.5 M NaCl, 20 mM β-mercaptoethanol (β-Me), and 1× Proteinase inhibitor mixture (Roche). Extraction was performed for 2 h with constant shaking and intermittent vortexing, followed by ultrasonication at 35% amplitude for 10 min with a pulse interval of 5 s in ice-cold conditions. The sample extracts were then centrifuged at 10,000 × *g* for 20 min, and the supernatant was separated from the pellet. Three volumes of extraction buffer were added to the pellet, and the extraction was repeated. The supernatants were pooled, filtered, dialyzed overnight, and lyophilized prior to use.

Glycoprotein Capture by Lectin Affinity Chromatography—Lyophilized samples were solubilized in buffer containing 20 mM Tris (pH 7.5), 0.5 M NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, and 1 mM $MgCl_2$ and were subjected to lectin affinity chromatography (LAC) in a manually packed column containing 2 ml of concanavalin A (Con A) Sepharose resin (Sigma) (26). The bound proteins were eluted in three steps, each containing 3 column volumes (CVs) of buffer containing 0.5 M methyl α-D mannopyranoside (step I), followed by 1 M methyl α-D mannopyranoside (step II) and 1 M glucose (step III), respectively. Eluant fractions were pooled, and the buffer was exchanged and concentrated with 20 mM Tris (pH 7.5) using Amicon 10 kDa (MWCO) centrifugal filters (Vivascience, Germany).

One Dimensional (1D) and Two Dimensional (2D) SDS-PAGE—The protein samples that were enriched using LAC were subjected to 12% SDS-PAGE separation (27) in replicates. The gels were either stained with Coomassie Blue, periodic acid-Schiff (PAS) or β-glucosyl Yariv stain to visualize the proteins, glycoproteins or arabinogalactan patterns, respectively. An aliquot of the protein sample was subjected to two-dimensional gel electrophoresis (2D-SDS-PAGE) as described previously (28). The gels were stained using a silver staining procedure to visualize the spots and were stored in 1% acetic acid at 4 °C until further use.

Gel Phase Digestion and Gel Free (Solution Phase) Digestion—The glycoprotein samples that were resolved by 12% 1D-PAGE gels were excised into 0.5 mm gel slices (18 slices) from the high to low molecular weight regions. The bands from 1D-PAGE and the spots from 2D-PAGE were subjected to in-gel trypsin digestion as described by Shevchenko *et al.* (29) with minor modifications, which are described in the supplemental Methods. Solution phase glycoprotein samples were subjected to trypsin proteolysis using the filter aided sample preparation (FASP) method as previously described (30), for the glycopeptide capture and gel-free 2D LC-MALDI TOF/TOF approach. The tryptic peptides were lyophilized and stored at −80 °C prior to use.

Glycopeptide Capture—Glycopeptide capture, deglycosylation and protein identification were performed on three independent replicate samples as described by Kaji *et al.* (31). Database search parameters for the deglycosylated peptide identification are described in the supplemental Methods. The following criteria were used for the identification of glycopeptides: (1) the significance threshold was set to $p < 0.02$; (2) the expectancy cut off was set to 0.05; and (3) individual ion scores (>45) that indicated identities were only considered for identification (false discovery rate (FDR) <1%). Furthermore, the peptide was considered formerly glycosylated only if the deamidated asparagine (N) was followed by X-S/T (any amino acid except proline - serine/threonine). Additionally, only those peptides that were observed in the three replicate sample injections were reported as formerly glycosylated peptides in this current study.

Data Analysis—Data analysis was performed as shown in Fig. 1*B*. Briefly, mascot generic format (MGF) files were extracted from the individual samples and exported into PEAKS Studio (version 6.0) (32). Spectral files were subjected to protein identification through multiple search engines against the aforementioned databases. The following data analysis parameters were used: (1) a FDR <1%; (2) at least one unique peptide; and (3) a protein probability of >90%. The identified proteins were exported into BLAST2GO platform (version 2.0) (www.blast2go.com/b2ghome) (33) for Gene Ontology (GO) annotation, protein motif prediction, and pathway mapping. Potential N-terminal signal peptides were predicted using the SignalP 4.0 server (www.cbs.dtu.dk/services/SignalP) (34); integral transmembrane domains were predicted using TMHMM-2.0 (www.cbs.dtu.dk/services/TMHMM), whereas potential *N*-linked glycosylation sites were predicted using Net N-Glyc (www.cbs.dtu.dk/services/NetNGlyc), and glycosylphosphatidylinositol (GPI)-anchored proteins were predicted using the Big-PI plant predictor tool (www.mendel.imp.ac.at/gpi/plant_server.html) (35). Proteins were classified into different CAZyme families using the CAZymes Analysis Toolkit (CAT) (www.mothra.ornl.

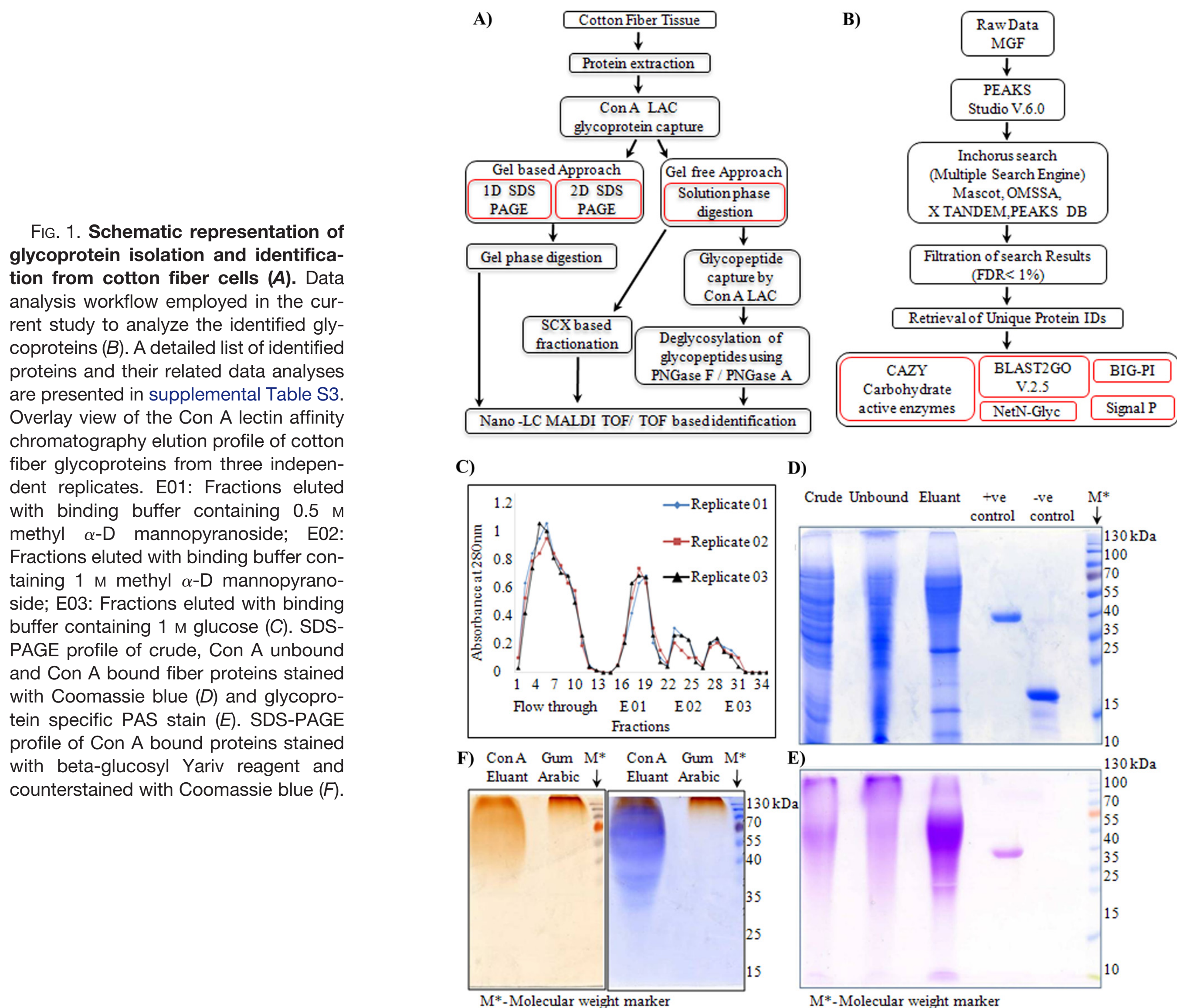


FIG. 1. **Schematic representation of glycoprotein isolation and identification from cotton fiber cells (*A*).** Data analysis workflow employed in the current study to analyze the identified glycoproteins (*B*). A detailed list of identified proteins and their related data analyses are presented in supplemental Table S3. Overlay view of the Con A lectin affinity chromatography elution profile of cotton fiber glycoproteins from three independent replicates. E01: Fractions eluted with binding buffer containing 0.5 M methyl α-D mannopyranoside; E02: Fractions eluted with binding buffer containing 1 M methyl α-D mannopyranoside; E03: Fractions eluted with binding buffer containing 1 M glucose (*C*). SDS-PAGE profile of crude, Con A unbound and Con A bound fiber proteins stained with Coomassie blue (*D*) and glycoprotein specific PAS stain (*E*). SDS-PAGE profile of Con A bound proteins stained with beta-glucosyl Yariv reagent and counterstained with Coomassie blue (*F*).

gov) (36). Leucine rich repeat sequences were predicted using the LLR finder tool (www.lrrfinder.com). Peroxidases were analyzed and classified using the Peroxibase database (www.peroxibase.toulouse.inra.fr/peroxiscan.php) (37). Spectral counting-based semiquantitative analysis, as mentioned in the supplemental Methods, was used to calculate the relative abundance of the proteins in the current data set (38).

Proteomic Data Set—The raw data from the mass spectrometry experiments have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (39) and can be found using the data set identifier PXD000178. Annotated MS/MS spectrums corresponding to proteins identified from 2D-PAGE and the glycopeptide approaches are shown in supplemental Data S1 and S2.

RESULTS

Glycoproteome of Cotton Fibers—Cotton fiber glycoproteins were enriched using Con A LAC and were identified using four independent approaches (Fig. 1*A*, 1*B* and Table I). Altogether, 334 unique proteins with ≥1 unique peptide were identified with an FDR <1%. A total of 92 proteins were identified by at least two approaches, including nine proteins that were identified by all four approaches (Fig. 2*A*). The molecular weight distribution of the identified proteins from the 1D-PAGE analysis is depicted in Fig. 3*A*. Fifty eight protein spots were identified using 2D-PAGE (Fig. 4*A*, 4*B* and supplemental Table S1), and eight of these proteins showed isoform (chain)-like patterns. Amino acid sequence variations of the protein isoforms were confirmed using tandem mass spectrometry (MS/MS) and are highlighted in Table II and in supplemental Data S1. Furthermore, we report the presence of 92 formerly *N*-linked glycosylated peptides, containing 106 glycosylation sites, from 67 unique proteins (supplemental Table S2 and supplemental Data S2). The protein identification details from the independent approaches are listed in

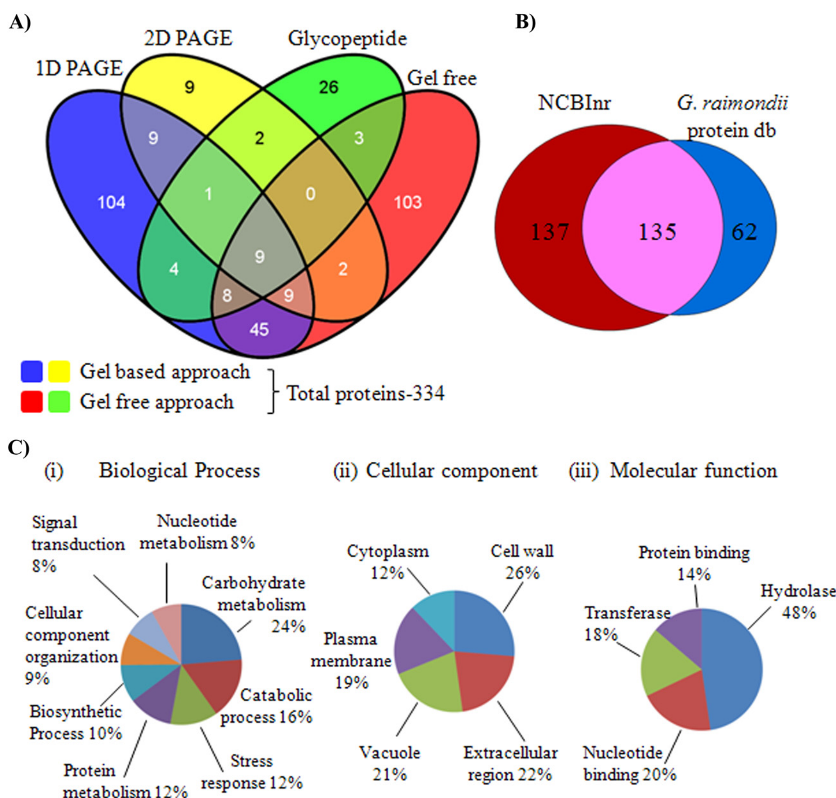
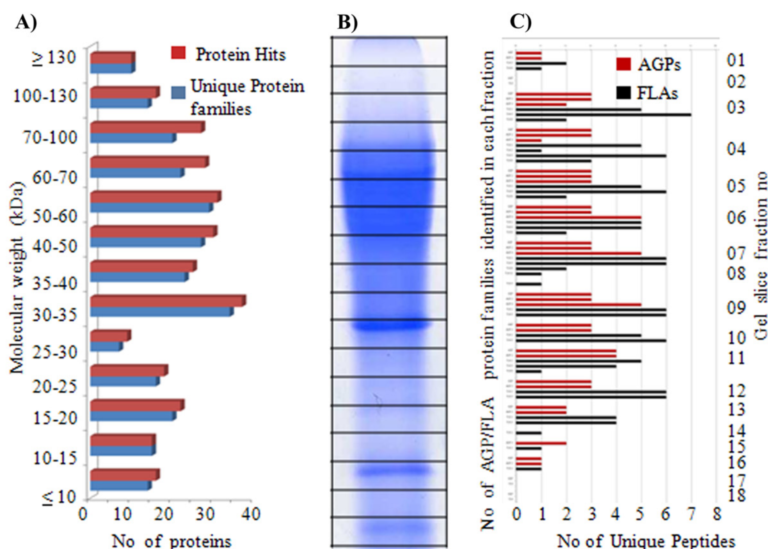
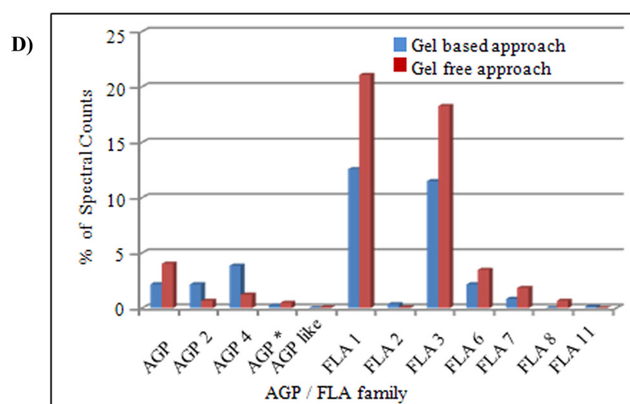


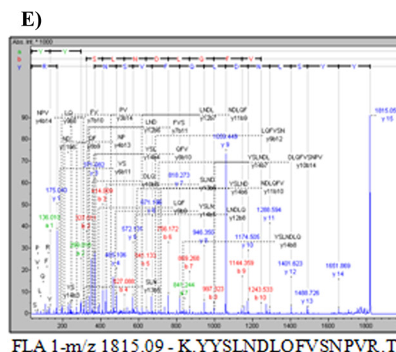
FIG. 2. **Diagrammatic view representing the overall output achieved from the gel based and gel free approaches (*A*).** Diagrammatic view representing the number of proteins identified from the two independent database search analyses (NCBInr and *G. raimondii* protein databases) (*B*). Gene Ontology (GO) based annotation and classification of the cotton fiber glycoproteome into the (i) Biological process, (ii) Cellular component and (iii) Molecular function categories (*C*).

Table I and supplemental Table S3. *In silico* analysis identified 286 unique proteins with potential *N*-linked glycosylation sites accounting for >85% of the glycoproteome. Furthermore, 46%, 13%, and 5% of the identified proteins were predicted to have signal peptide cleavage sites, transmembrane domain regions and GPI anchor sites, respectively (supplemental Table S3). GO analysis annotated 67% of the proteins to be either cell wall or extracellular region destined, and 48% of the proteins were annotated as hydrolases, of which 26% were found to be involved in carbohydrate metabolism (Fig. 2*C*). Spectral counting-based semiquantitative analysis revealed the relative abundance of redundant (repeatedly detected) AGPs and FLAs, as well as nonredundant CAZymes (Fig. 5*A*). Based on these observations, the cotton fiber glycoproteome can be classified into three categories: structural proteins, enzymatic proteins and proteins with other or unknown functions. Furthermore, we observed that the fiber glycoproteome is composed of 5% AGPs and FLAs, which account for 43% of the spectral counts (SpC), 38% CAZymes, which contribute 31% of the SpC and 57% other proteins, which account for 26% of the SpC (Fig. 5*A*, supplemental Table S4).

Structural Proteins Comprising AGPs and FLAs Display Abundance and Heterogeneity—In the present study, we identified five unique members of the AGP family and eight unique members of the FLA family (supplemental Tables S3 and S5). Using 1D-PAGE followed by LC-MALDI analysis, we observed AGP or FLA specific peptides in 15 out of the 18 gel slices excised from different molecular weight regions (Figs. 3*B* and 3*C*). The theoretical molecular weights of the identified proteins were ~25 kDa, where as their observed weights varied from < 25 to >130 kDa as determined by 1D-PAGE (Figs. 3*B*, 3*C* and supplemental Table S5). The AGP epitope specific β-Yariv staining pattern at regions >70 kDa on the 1D-PAGE gel (Fig. 1*F*) suggested the presence of dominant post-translational modifications (PTMs) that might contribute to >60% of their observed molecular weight. In addition, 10 distinct spots from various pIs and molecular weights were identified as AGPs and FLAs following 2D-PAGE analysis (Figs. 4*A*, 4*B* and supplemental Table S1). Amino acid variations observed in their fasciclin (FAS) domains are highlighted in supplemental Fig. S2 and S3. We also observed a novel, unknown peptide sequence (*m/z* 1576.61) homologous to the fasciclin region of the FLA family (Fig. 6*A*). Both the AGPs and FLAs harbored 3 to 4 potential *N*-linked glycosylation sites within the FAS domain (supplemental Figs. S2, S3), among which we identified 1 to 3 glycosylation sites per unique member (supplemental Table S2 and supplemental Data S2). Independent deglycosylation reactions with PNGase F and A (peptide N-glycoamidase) showed that FLA6 harbored two different types of core *N*-linked glycans (supplemental Fig. S5*A*). Our study further revealed that the FLAs were comparatively more abundant than the AGPs in cotton fiber. Among these proteins, FLA1 and FLA3 together constituted 30% of the identified spectral counts (Fig. 3*D*). The conserved glycosylation site (N_{147}VT) in AGP4 and FLA6 harbored different core *N*-linked glycans as determined by the deglycosylation







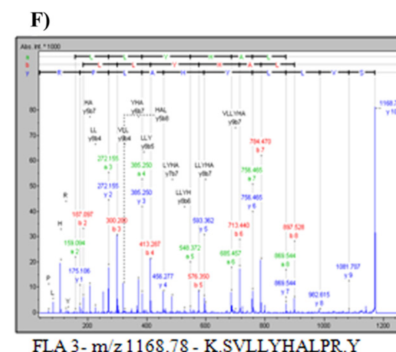


Fig. 3. **Bar graph depicting the distribution of the identified proteins across various molecular weight regions of the 1D-SDS-PAGE gel (*A*).** Fig. showing the number and position of gel slices excised from the various molecular weight regions (*B*). Bar graph depicting the number of AGP and FLA specific unique peptides observed from the gel slices excised from various molecular weight regions of the 1D-PAGE gel (*C*). Graph representing the relative abundance (% of spectral counts) of the AGP and FLA members in the gel based and gel free approaches (*D*). Representative annotated mass spectrum maps corresponding to the most abundant peptide species from FLA1 (*E*) and FLA3 (*F*).

reactions. Multiple sequence alignment followed by *in silico* analysis highlighted that all the identified family members contained conserved domain regions and sequence properties like >30% PAST and AGP-like modules (supplemental Table S5). The AGP module was longer, in terms of amino acid residues, in the AGPs than those of the FLAs (supplemental Fig. S2 and S3). Based on their domain organization and sequence specific properties, we designated them as chimeric AGPs (supplemental Figs. S2and S3 and supplemental Table S5) (40). Further gene expression analysis revealed stage and isoform specific expression patterns of certain identified members (AGP, FLA11, and FLA15) of this protein family (supplemental Fig. S6).

CAZymes are the Major Players Among the Fiber Localized Enzymatic Glycoproteins—CAZymes accounted for 38% of the fiber glycoproteome and 55% of the identified enzymes (Fig. 5*A*). Further classification and analysis showed that glycosyl hydrolases (GHs) were relatively more abundant, accounting for 59% of the CAZymes, followed by 10% carbohydrate esterases (CEs) and 5% glycosyl transferases (GTs). Carbohydrate binding modules (CBMs) accounted for 26% of the CAZymes, and the majority of them were catalytically

TABLE I

Unique proteins identified from independent technical approaches. NCBInr, NCBI nonredundant protein database. G.raimondii protein db - Gossypium raimondii protein database (refer to supplemental Methods)

Sl No	Technical approach	No of slices/ spots/fractions	Unique peptides (PSMs)		Unique protein families		Proteins unique to approach
			NCBInr	*G.raimondii* protein db	NCBInr	*G.raimondii* protein db	
01	1D SDS PAGE[a]	18 gel slices	459 (989)	1600 (2873)	149	231	104
02	2D PAGE (pI ranges: 3–10NL & 4–7L)[b]	58 spots	145 peptides (58 proteins)		41		9
03	Gel Free[c]	20 SCX fractions	461 (1575)	6004 (8891)	155	164	103
04	Glycopeptide (PNGaseF & PNGase A)	5 SCX fractions	56[d] peptides	36[d] peptides	39	28	26

[a] Refer to Fig. 3.
[b] Refer to Figs. 4*A* (4–7 Linear) and 4*B* (3–10 Non-Linear).
[c] Refer to supplemental Fig. S1.
[d] Formerly N-linked glycosylated peptides, refer to supplemental Data S2.

TABLE II

List of glycoprotein isoforms identified from 2D-PAGE gels. For spot number refer Figs. 4A and 4B. Amino acid variations confirmed by MS/MS based sequencing are shown in bold

Sl No	Spot No	Accession No	Protein ID	*m/z* $(M+H)^+$	MS/MS of discriminating peptide sequences
01	05	gi 225441645	Glucan endo-1,3-beta-glucosidase 8	979.57	R.F**YN**GL**LPR**.L
02	07	gi 356533037	Lysosomal beta glucosidase-like	2805.50	K.F**TM**GL**FEN**P**L**AD**T**SLVN**E**LGSQEHR.D
03	09	gi 33391721	Beta-D-glucosidase	2865.53	K.F**VM**GL**FEN**P**M**AD**N**SLVN**Q**LGSQEHR.E + 2 Oxidation (M)
04	10	gi 30841338	Arabinogalactan protein	3017.76	K.V**E**LVQFHIVPTYLTSSQFQTISNPLR.T
05	11	gi 150416579	Arabinogalactan protein 2	3016.81	K.V**Q**LVQFHIVPTYLTSSQFQTISNPLR.T
06	12	gi 310722811	Vacuolar invertase 1	1478.66	K.**I**PVLDDENYNMR.V
07	13	gi 310722811	Vacuolar invertase 1	1614.68	R.VLVDHSVVESFG**G**GGR.T
08	14	gi 229597364	Vacuolar invertase	1464.68	K.**V**PVLDDENYNMR.V
				1686.71	R.VLVDHSVVESFG**E**GGR.T
09	15	gi 310722811	Vacuolar invertase 1	1494.77	K.**I**PVLDDENYNMR.V+ Oxidation (M)
				1614.90	R.VLVDHSVVESFG**G**GGR.T
10	23	gi 157273646	Fasciclin-like Arabinogalactan protein 6	2986.81	K.**L**QLV**QF**H**IL**P**T**LMSTSQFQTASNPLR.T
11	24	gi 606942	Unknown (FLA-3)	1943.17	K.V**T**SAVH**TSK**PVA**V**YQ**I**DK.V
12	25	gi 150416583	Fasciclin-like arabinogalactan protein 1	1308.81	K.**V**QLV**LY**H**VI**P**K**.Y
				1815.05	K.**Y**YSL**N**D**LQF**VSNPVR.T
13	26	gi 157273638	Fasciclin-like arabinogalactan protein 2	1104.57	K.**F**YSL**A**D**FNK**.L
				1916.04	K.V**S**SAVH**STD**PVA**I**YQ**V**DK.V
14	41	gi 291042515	Germin-like protein subfamily 2	1798.88	R.IDY**K**PGG**L**NPPHTHPR.A
15	44	gi 196122046	Germin-like protein 1	1741.80	R.IDY**A**PGG**I**NPPHTHPR.A
16	51	gi 357601486	Purple acid phosphatase 1 (PAP 1)	1174.56	R.THA**YFG**W**H**R.N
17	52	gi 296082127	Unnamed protein product (PAP)	1231.58	R.THA**FYH**W**N**R.N
18	53	gi 224120334	Predicted protein (PAP)	1498.70	R.F**R**DPQP**D**YSAFR.E
19	54	gi 40217506	Acid phosphatase	1469.71	K.F**L**DPQP**E**YSAFR.E
20	55	gi 224061929	Multicopper oxidase	1315.75	R.QYLGQQ**F**YLR.V
21	56	gi 356508933	L-ascorbate oxidase homolog	1281.76	R.QYLGQQ**L**YLR.V
22	57	gi 357496611	Lamin-like protein	1877.93	K.HFYNGDWL**F**FVYDR.N
23	58	gi 319433443	Copper binding protein 4	1894.83	K.HFYNGDWL**Y**FVYDR.N

inactive toward carbohydrate substrates (Fig. 4*C*). These enzymes displayed heterogeneity in molecular weight and pI, including the presence of different isoforms; however, they were relatively less abundant and redundant compared with AGPs and FLAs.

Sixty-three unique proteins were identified as glycosyl hydrolases and were classified into 12 families. Among these proteins, GH3 constituted 25% of the GHs, followed by GH35 (15%), and other families (Fig. 4*C*(iv)). Pathway mapping and literature surveys suggested that these hydrolases catalyze similar reactions involving a diverse class of glycoconjugates. *In silico* based screening of the GHs showed the presence of noncatalytic domains apart from hydrolase specific domain regions (supplemental Table S3). To highlight a few, members of the GH16, GH32 and GH35 families had lectin-like domains, whereas the GH17 family harbored CBM43 and X8 domains. Additionally, 2D-PAGE analysis revealed four different forms of xylosidase and three different forms of glucosidase enzymes (GH3) (Table II, Figs. 4*A* and 4*B*). Genes encoding selected members of these enzyme families were predominantly expressed during the elongation phase (Fig. 5*C* & supplemental Fig. S6). Three unique members of xylo-

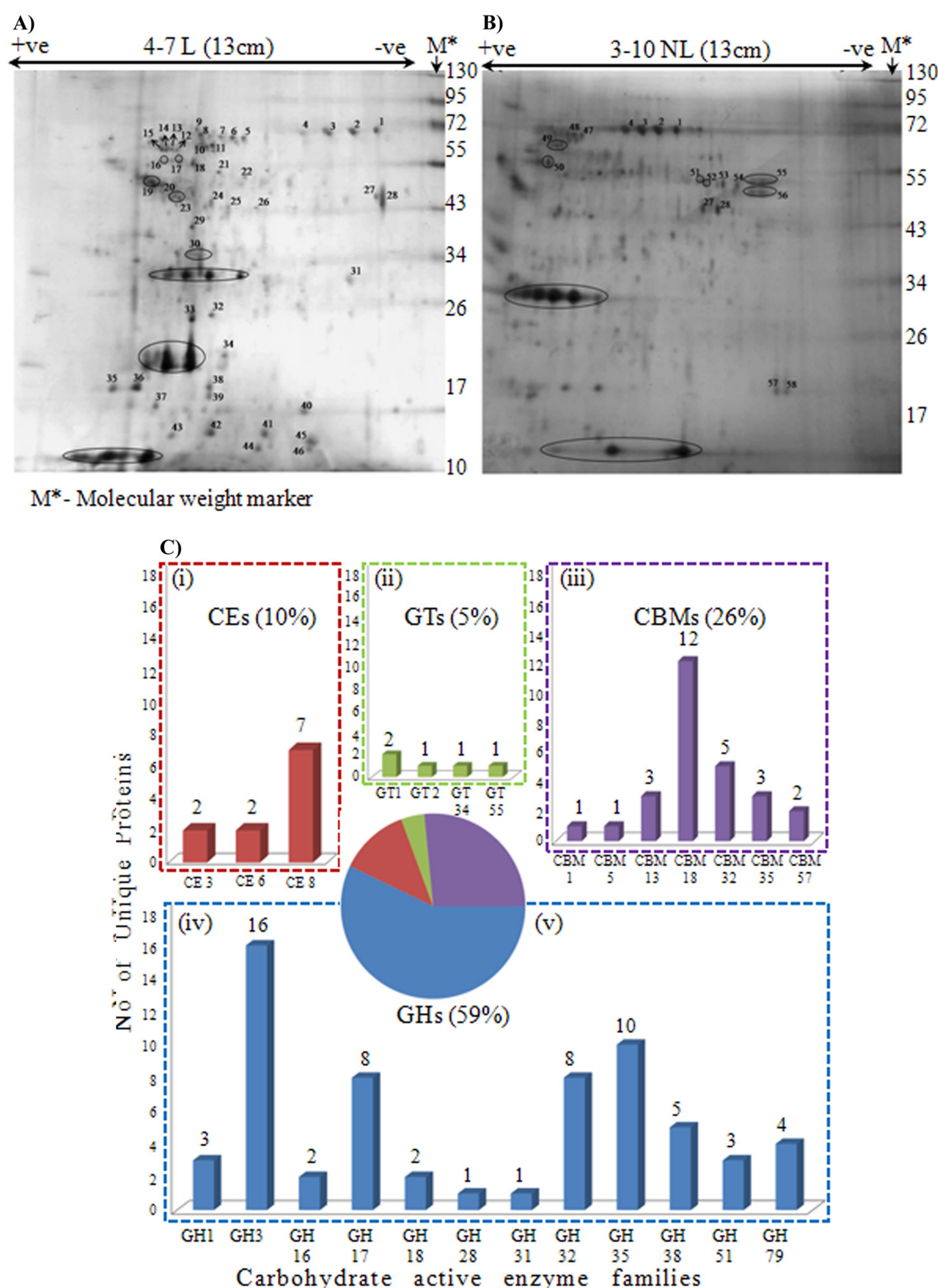
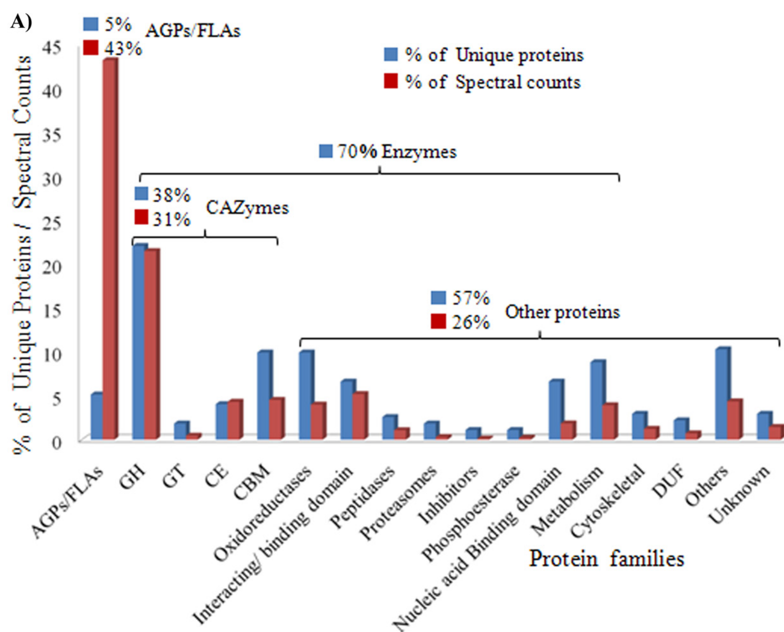


FIG. 4. **Silver stained 2D-PAGE profile of the Con A bound fiber glycoproteins in the 4–7 (linear) (*A*) and 3–10 (non-linear) (*B*) pI ranges.** *C*, Diagrammatic view representing the classification and distribution of the identified CAZyme family members.

glucan active enzymes of the GH16 family (gi 155966597) were identified, and their genes were found to be consistently expressed throughout the elongation and postelongation phases (5–25 dpa) (Fig. 5*C* and supplemental Fig. S6). Approximately 10 unique members of the glucan endo-1, 3-beta-D-glucosidase (E.C. 3.2.1.39) and cellulase (beta 1–3 glucanase, E.C. 3.2.1.4) of the GH17 family were identified; three of these proteins had predicted GPI anchor sites, whereas one had a transmembrane domain (supplemental Table S3). The GH18 family included the phosphoinositide-specific phopholipase C (PI-PLC) protein family that targets phospholipids and GPI anchored proteins (GAPs) like AGPs. Invertase and fructokinase members of the GH32 family were observed in the >70 kDa, 35 kDa, and <20 kDa regions. Ten unique members of invertases were identified; only three among them had secretory signals, and these proteins could be classified as extracellular, cytosolic or vacuolar forms (supplemental Table S3). The vacuolar forms were relatively more abundant than their extracellular counterparts, and they also exhibited variants by 2D-PAGE analysis (Table II and



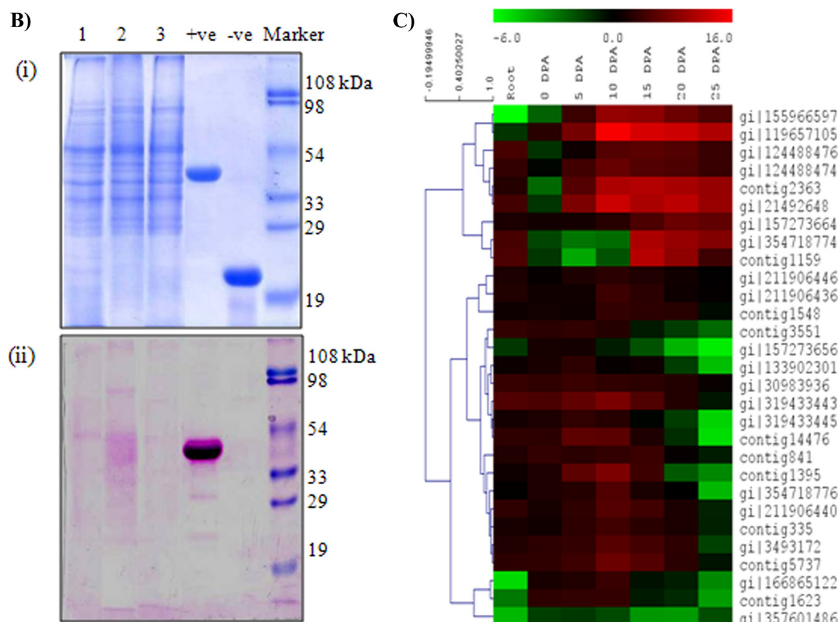
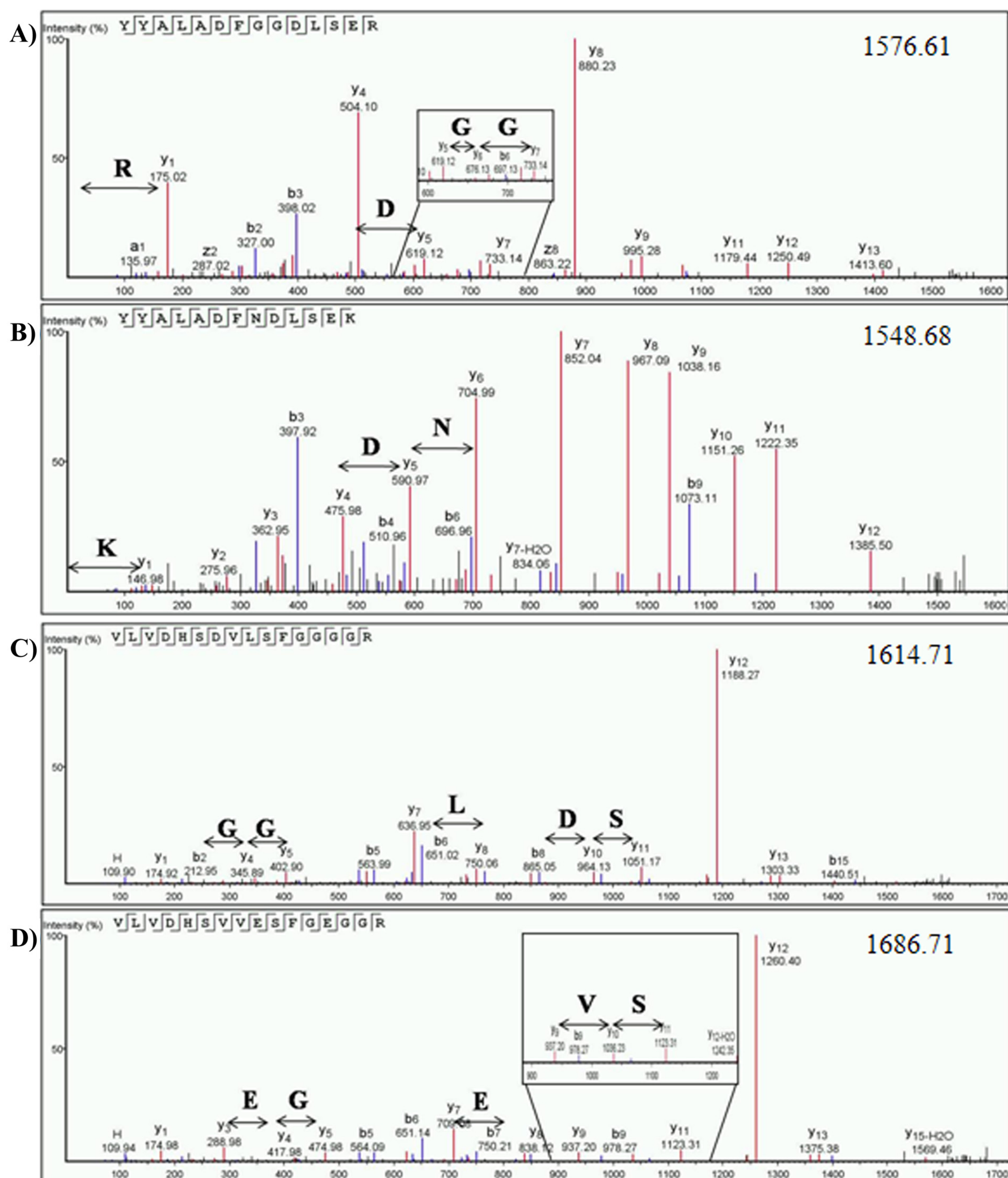


FIG. 5. ***A,* Percentage distribution of the protein families in the cotton fiber glycoproteome.** *B,* SDS-PAGE profile of cotton fiber proteins from the (lane 1) early elongation (5 dpa), (lane 2) elongation (10–15 dpa) and (lane 3) maturation (>20 dpa) stages stained with Coomassie blue (i) and with the glycoprotein specific PAS stain (ii). *C,* Quantitative real-time PCR based expression profiles of glycoprotein encoding genes under different fiber developmental stages and root relative to expression in the leaf. The heat map shows the mRNA expression profile based on hierarchical clustering of various glycoprotein encoding genes in the root and the different fiber developmental stages 0, 5, 10, 15, 20, and 25 days post anthesis (dpa) compared with the expression in leaf tissue. The colored bar at the top represents the scale for the log2 fold change in expression (supplemental Fig. S6). Protein IDs corresponding to the gi and contig numbers are listed in supplemental Table S8.

supplemental Table S1). In addition, we also observed novel/unknown peptide regions homologous to the invertase (*m/z* 1614.71, Fig. 6*C*) and fructokinase protein family (*m/z* 1617.85, supplemental Fig. S4*A*). The GH31, GH35 and GH38 families included the *N*-linked glycan processing enzymes, and among them galactosidases of GH35 were relatively more in number and in abundance. The beta-glucuronidase and heparanase-like proteins (EC: 3.2.1.31) of the GH79 family consist of proteins involved in glucuronate interconversion.

Identified members of carbohydrate esterases (CEs) were further classified into the CE3, CE6, and CE8 classes, with the latter being relatively abundant (Fig. 4*C*(i), supplemental Table S3). CE3 was comprised of phopholipase C domain-containing proteins and three isoforms of MAP3K-like protein kinases. CE6 included polygalacturonase inhibitor proteins that were found to contain two different core N-glycans (supplemental Fig. S5*D*), and the CE8 classes included pectinesterase and pectin methyl esterases (PME) (E.C.3.1.1.11). Genes encoding the different forms of PMEs showed differential stage specific expression patterns; for example, PME4 was majorly expressed during early elongation (5–10 dpa), whereas PME5 was highly expressed during the late elongation stages (15–20 dpa) (supplemental Fig. S6).

Glycosyl transferases comprising GT1, GT2, GT4, GT34, and GT55 form the third major class of the identified CAZymes (Fig. 4*C*(ii), supplemental Table S3). In addition, the



a) Novel/Unknown m/z 1576.61 YYALADFGGDLSER

b) gi|606942 m/z 1548.68 YYALADF-NDLSEK

c) Novel/Unknown m/z 1614.71 VLVDHSDVLSFGGGGR

d) gi|229597364 m/z 1686.71 VLVDHSVVESFGEGGR

FIG. 6. **Annotated line spectra corresponding to the unknown/novel peptide sequences homologous to arabinogalactans (*A*) and invertases (*C*) identified in the current study.** Line spectra of known peptide sequences homologous to the novel sequences of arabinogalactans (*B*) and invertases (*D*).

identified members of endo-xyloglucan transferase of the GH16 family and the fructosyl transferase of the GH32 family are also known to perform glycosyl transferase-like functions in plants (41, 42).

Carbohydrate binding modules form an associated class of enzymes that is comprised of 27 unique proteins classified into eight families (Fig. 4*C*(iii), supplemental Table S3). Among them, CBM43 observed within GH17 members was classified as catalytically active on carbohydrate substrates. CBM18 family is comprised of oxidoreductases and oxidases and constituted >40% of the identified CBMs. In addition, glucose-methanol-choline (GMC) oxidoreductase of the CBM1 family, protease of the CBM5 family, heat shock proteins (HSPs) of the CBM13 family, purple acid phosphatase (PAP) of the CBM32 family, domain of unknown function 1680 (DUF) of the CBM35 family and lectin like domain containing protein kinases of the CBM57 family were also identified, contributing to the catalytically noncarbohydrate active fiber localized enzymes. Among them, monocopper oxidase of the CBM18 family was found to contain different *N*-linked glycans (supplemental Fig. S5*B*), and PAP of CBM32 showed isoform variants (Table II).

Non-CAZymes Play Regulatory Roles in Fiber Cell Elongation—Non-CAZymes included non-CBM oxidoreductases, proteases, and proteins with interacting domains. Oxidoreductases, which included 40 unique proteins, accounted for 14.7% of the fiber glycoproteome. Among these proteins, 12 proteins were earlier classified under CBM18, and the remaining 28 proteins can be grouped into the reductase, disulfide isomerase, peroxidase, and copper binding protein families (Fig. 5*A*). Each of these protein groups has associated domains, such as NAD/FAD binding domains in reductases, thioredoxin-like fold/domains in disulfide isomerases and cupredoxin domains in copper binding oxidase-like proteins (supplemental Table S3). The peroxidases identified in the current study were classified as class III (secretory class) using *in silico* analysis (37). Class III peroxidases are known to be involved in wall loosening of cells that undergo growth through elongation rather than division. Proteases, proteasomes and protease inhibitors accounted for ~6% of the glycoproteome (Fig. 5*A*). The protease family included aspartic and serine carboxypeptidase, whereas the proteasomes were majorly comprised of B-type subunit family members, with one unique member containing an armadillo-like fold belonging to extra cellular matrix 29 (ECM) family. Approximately 7% of the fiber glycoproteome, constituting 18 unique proteins, had interacting or binding domain(s). These proteins included lectin like domain containing calreticulin, EF hand domain containing calmodulin, and cupin domain containing germins. Four different forms of calreticulin were identified in the current study (supplemental Table S1). In addition, we also observed proteins involved in nucleic acid, carbohydrate, and lipid metabolism. Approximately 2% of the identified proteins were classified as unknown and they contained DUF-like domains. In plants, certain DUF like domains are predicted to have cell wall binding (DUF642) and glycosyl transferase-like roles (43). Approximately 3% of the identified proteins were predicted to have no known functional domains (Fig. 5*A*).

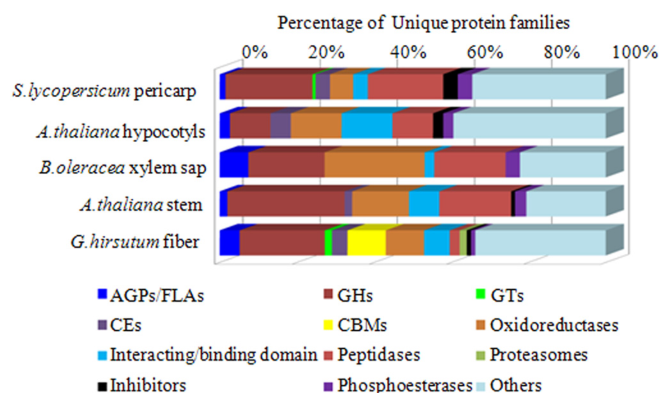


FIG. 7. **Bar graph depicting the comparative analysis of the protein families identified in the Con A bound fiber glycoproteome with other published Con A based plant glycoproteome data sets.**

Comparative Analysis of Plant Glycoproteomes Highlights the Unique Compositional Status of Cotton Fiber—The overall composition of the fiber glycoproteome was found to be similar to other known plant glycoproteome data sets (*Solanum*, *Arabidopsis*, and *Brassica* sp.) (supplemental Table S6) (26, 44, 45, 46). However, major compositional variations among the protein families were observed (Fig. 7). To highlight a few, the percentage of AGPs and FLAs in the fiber glycoproteome was comparable only with *Brassica oleracea* xylem sap, while AGPs and FLAs were relatively underrepresented (< 3 proteins) in the other data sets (Fig. 7, supplemental Table S6). CBM containing enzymes that are catalytically inactive toward carbohydrate substrates were absent from the other data sets, suggesting that this is a unique feature of the cotton fiber glycoproteome. Polygalacturonases (pectinase) of the GH28 family, oxidoreductases, proteases, and proteins with interaction and binding domains were relatively low (<10%) in the fiber glycoproteome. Additionally, the protein inhibitor class included only protease inhibitors, whereas CAZyme inhibitor was relatively low in fiber (<1%) (Fig. 7, supplemental Tables S6 and S7).

DISCUSSION

The development of cotton seed epidermal cells into long fibers has been widely studied, and various factors, such as structural proteins, CAZymes, and transcription factors, have been examined for the individual roles that they play in the differentiation process. However, systems level identification and characterization of various protein components is still lacking, and this type of analysis will help researchers better understand their roles in development. Wall yielding properties majorly regulate the differentiation events leading to the fiber's length and strength. Wall-destined proteins are known

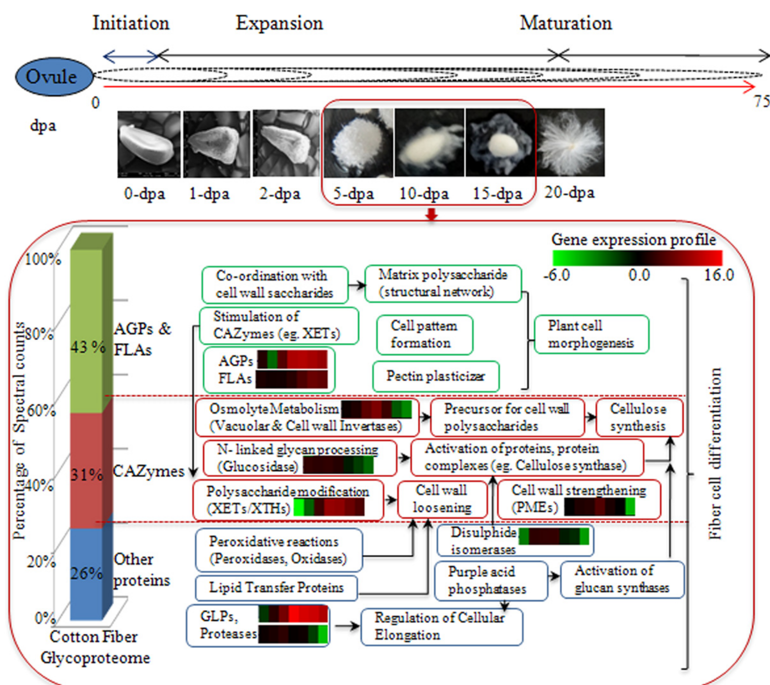


FIG. 8. **Diagrammatic representation of fiber cell undergoing overlapping stages of development (*A*).** Scanning electron microscopic images of 0, 1 and 2 dpa cotton ovules with fiber initials followed by photographic images of 5, 10, 15 and 20 dpa cotton ovules along with fiber. Fiber developmental stages utilized in the current study to identify the glycoproteome are enclosed with in red box (*B*). Bar graph depicting the distribution of the major classes of glycoproteins in cotton fiber glycoproteome (*C*). Schematic representation of the structural and functional roles of the identified glycoproteins along with their gene expression profile during fiber cell differentiation (*D*).

to be glycosylated in plants (23), and they might in turn contribute to such yielding properties. Therefore, glycoproteome approaches can be employed to study the wall-destined proteins in plants (23). Identification and analysis of the cotton fiber glycoproteome revealed that the majority of the proteins were either destined for the cell wall or were extracellular in nature. Further, our results suggested that the structural proteins were relatively abundant, compared with the CAZymes and other enzymes (Fig. 5*A*). In this study, by employing complementary proteomic approaches, we have been able to resolve the heterogeneity of the fiber glycoproteome. Briefly, a 1D-PAGE based approach revealed the presence of unique and identical peptides corresponding to same protein family from different molecular weight ranges, whereas 2D-PAGE analysis showed protein isoforms with minor amino acid variations. In addition, independent deglycosylation reactions using PNGase F and A showed overlapping and unique peptides, suggesting differences in the core *N*-linked glycans attached to these protein molecules. Together, all of these approaches suggest the presence of both protein isoforms and glycoforms in cotton fiber. The presence of the different forms of the same protein highlights the cellular requirement for proteins to perform similar functions at different developmental phases, localizations, and under different physiological conditions (Figs. 8*A*, 8*B*, 8*C* and 8*D*).

The abundance and distribution of the identified glycoproteins provided clues about the cellular makeup of the cotton fiber (47). The GHs were relatively more abundant and diverse compared with other CAZymes. Further analysis revealed that most of the GHs and non-CAZymes harbored noncatalytic carbohydrate binding modules (CBMs) or interacting domains (lectin) of broad specificity. Comparative analysis with previously reported plant glycoproteomes (26, 44, 45, 46) revealed that enzymes (CAZymes and non-CAZymes) containing noncatalytic carbohydrate binding modules were observed only in cotton fiber, which further highlights the unique feature of the fiber localized glycosylated enzymes. In a cell such as a cotton fiber, which is rich in structural and nonstructural carbohydrates, the presence of CBMs and carbohydrate interacting domains in enzymes would be advantageous as it might modulate the protein's activity by increasing or stabilizing these enzymes in close proximity to its substrate (48). On the other hand, we also observed CAZymes and other enzymes devoid of such interacting or binding domains. In addition to these features, these enzymes also had isoform variants. Such diverse and discriminating features among similar enzymes depict the cellular requirement for substrate hydrolysis, remodeling, and grafting in elongating cotton fibers (Fig. 8*D*).

Structural proteins have been proposed to have functional properties that might contribute to the dynamic status of the cell wall (49). In mammals, proteoglycans are known to be abundant molecules in the extracellular matrix, and they act as biological lubricants and stabilizers of cellular integrity (50). However, data concerning the abundance, heterogeneity and associated functional roles of such molecules in plant systems are still emerging. In order to withstand the diverse processes occurring during elongation, the fiber cell wall needs to contain responsive structural molecules. Plant AGPs and FLAs are reported to play major and diverse roles in cell development (51); however, these proteins were less represented in the non-fiber glycoproteome datasets (Fig. 7, supplemental Table S6). In the current study, we observed redun-

dant and abundant members of the AGP and FLA families, suggesting they might play major roles as structural molecules in cotton fiber cells. Heterogeneity in their distribution across various molecular weights and pIs were observed, and this may correspond to variations in amino acid sequences and PTMs. Abundance could be the major determinant of the functional parameter for such molecules. The glycan component of these proteins might play a major role during extension by acting as molecular cushions. Cell wall localized AGP signals could also stimulate enzymes such as XETs, as demonstrated in *in vitro* experiments (20). Our earlier studies showed that transcripts encoding these arabinogalactans and CAZymes were majorly down-regulated in a lintless mutant (52). Additionally, these protein encoding transcripts were highly down-regulated during conditions of drought stress (53), suggesting they play major roles in fiber development and related conditions.

In conclusion, we have made a major attempt to characterize the cotton fiber glycoproteome and have revealed cell wall destined structural and enzymatic proteins. Our comprehensive analysis identified the presence, abundance and heterogeneity of fiber localized glycoprotein families such as AGPs, CAZymes and other glycoproteins. Such structural proteins and enzyme isoforms might play non-redundant roles throughout fiber development (Figs. 8*C* and 8*D*). The diverse and heterogeneous features of the identified glycoproteins displays the tetraploid nature of *G. hirsutum* contributed by the A and D parental genomes. GO based functional annotation showed that the fiber glycoproteome followed a distribution pattern similar to other plant glycoproteomes, but there were compositional variations unique to cotton fiber. The identification of particular protein families and their abundance shown in this study reflect the major determinants of the structural and functional parameters governing the wall yielding properties of the tetraploid cotton fiber.

Acknowledgments—We wish to thank Ranjana Pathak and Israr Ahmad for their assistance during documentation.

* This work was supported by the Indian Council of Agricultural Research (ICAR) under the National Agricultural Innovation Project (NAIP), Component-4, Department of Biotechnology (DBT), Government of India, and International Centre for Genetic Engineering and Biotechnology, New Delhi. SK is the recipient of Research associateship from Indian Council of Agricultural Research, India. KK and PP are the recipients of Junior Research Fellowships from Department of Biotechnology and University Grant Commission, India respectively.

[S] This article contains supplemental Methods, Tables S1 to S9, Figs. S1 to S6, and Data files S1 and S2.

¶ To whom correspondence should be addressed: Plant Transformation Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi 110067, India. Tel.: +91 11 2674 1358; E-mail: vsreddy@icgeb.res.in.

Author Contributions: Conceived and designed the experiments: SK, PAK and VSR. Performed the experiments: SK, VR, KK, SL, PP, KVP, GD and MK. Data analysis: SK, KK, PP. Wrote the manuscript: SK, VSR.

REFERENCES

1. Kim, H. J., and Triplett, B. A. (2001) Cotton fiber growth in planta and in vitro. models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* **127,** 1361–1366
2. Qin, Y. M., and Zhu, Y. X. (2011) How cotton fibers elongate: a tale of linear cell-growth mode. *Curr. Opin. Plant Biol*. **14,** 106–111
3. Wakelyn, P. J., and French, A. D. (2007) Cotton fiber chemistry and technology. chapter 3, CRC Press, Taylor and Francis Group
4. Singh, B., Avci U., Eichler Inwood, S. E., Grimson, M. J., Landgraf, J., Mohnen, D., Sørensen, I., Wilkerson, C. G., Willats, W. G., and Haigler, C. H. (2009) A specialized outer layer of the primary cell wall joins elongating cotton fibers into tissue-like bundles. *Plant Physiol.* **150,** 684–699
5. Yang, Y. W., Bian, S. M., Yao, Y., and Liu, J. Y. (2008) Comparative proteomic analysis provides new insights into the fiber elongating process in cotton. *J. Proteome Res.* **7,** 4623–4637
6. Masuda, Y. (1990) Auxin-induced cell elongation and cell wall changes. *Bot. Mag.* **103,** 345–370
7. Bowman, D. T., Van Esbroeck, G. A., Van't Hof, J., and Jividen, G. M. (2001) Ovule fiber cell numbers in modern upland cottons. *J. Cotton Sci.* **5,** 81–83
8. Shimizu, Y., Aotsuka, S., Hasegawa, O., Kawada, T., Sakuno, T., Sakai, F., and Hayashi, T. (1997) Changes in levels of mRNAs for cell wall-related enzymes in growing cotton fiber cells. *Plant Cell Physiol.* **38,** 375–378
9. Al-Ghazi, Y., Bourot, S., Arioli, T., Dennis, E. S., and Llewellyn, D. J. (2009) Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. *Plant Cell Physiol.* **50,** 1364–1381
10. Darley, C. P., Forrester, A. M., and McQueen-Mason, S. J. (2001) The molecular basis of plant cell wall extension. *Plant Mol Biol.* **47,** 179–195
11. Nishitani, K. (1997) The role of endoxyloglucan transferase in the organization of plant cell walls. *Int. Rev. Cytol.* **173,** 157–206
12. Labavitch, J. M., and Ray, P. M. (1974) Turnover of cell wall polysaccharides in elongating pea stem segments. *Plant Physiol.* **53,** 669–673
13. Pang, C. Y., Wang, H., Pang, Y., Xu, C., Jiao, Y., Qin, Y. M., Western, T. L., Yu, S. X., and Zhu, Y. X. (2010) Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and arabidopsis root hair elongation. *Mol. Cell. Proteomics* **9,** 2019–2033
14. Willats, W. G., McCartney, L., Mackie, W., and Knox, J. P. (2001) Pectin: cell biology and prospects for functional analysis. *Plant Mol Biol.* **47,** 9–27
15. Pelloux, J., Rustérucci, C., and Mellerowicz, E. J. (2007) New insights into pectin methylesterase structure and function. *Trends Plant Sci.* **12,** 267–277
16. Moustacas, A. M., Nari, J., Borel, M., Noat, G., and Ricard, J. (1991) Pectin methylesterase, metal ions and plant cell-wall extension. *Biochem. J.* **279,** 351–354
17. Li, Y., Liu, D., Tu, L., Zhang, X., Wang, L., Zhu, L., Tan, J., and Deng, F. (2010) Suppression of GhAGP4 gene expression repressed the initiation and elongation of cotton fiber. *Plant Cell Rep.* **2,** 193–202
18. Lamport, D. T., Kieliszewski, M. J., and Showalter, A. M. (2005) Stress upregulates periplasmic arabinogalactan-proteins. *Plant Biosys.* **139,** 60–64
19. Takeda, T., and Fry, S. C. (2004) Control of xyloglucan endotransglucosylase activity by salts and anionic polymers. *Planta* **219,** 722–732
20. Seifert, G. J., and Roberts, K. (2007) The biology of arabinogalactan proteins. *Ann. Rev. Plant Biol.* **58,** 137–161
21. Ellis, M., Egelund, J., Schultz, C. J., and Bacic, A. (2010) Arabinogalactan-proteins: key regulators at the cell surface? *Plant Physiol.* **153,** 403–419
22. Liu, K., Han, M., Zhang, C., Yao, L., Sun, J., and Zhang, T. (2012) Comparative proteomic analysis reveals the mechanisms governing cotton fiber differentiation and initiation *J. Proteomics* **75,** 845–856
23. Ruiz-May, E., Kim, S. J., Brandizzi, F., and Rose, J. K. (2012) The secreted plant N-glycoproteome and associated secretory pathways. *Front. Plant Sci.* **3,** 117
24. Stafstrom, J. P., and Staehelin, L. A. (1986) The role of carbohydrate in maintaining extensin in an extended conformation. *Plant Physiol.* **81,** 242–246
25. Ruiz-May, E., Thannhauser, T. W., Zhang, S., and Rose, J. K. (2012) Analytical technologies for identification and characterization of the plant N-glycoproteome. *Front. Plant Sci.* **3,** 150
26. Catalá, C., Howe, K. J., Hucko, S., Rose, J. K., and Thannhauser, T. W. (2011) Towards characterization of the glycoproteome of tomato (Sola-

num lycopersicum) fruit using Concanavalin A lectin affinity chromatography and LC-MALDI-MS/MS analysis. *Proteomics* **11,** 1530–1544
27. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685
28. Agrawal, P., Kumar, S., and Das, H. R. (2010) Mass spectrometric characterization of isoform variants of peanut (Arachis hypogaea) stem lectin (SL-I). *J. Proteomics* **73,** 1573–1586
29. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols* **1,** 2856–2860
30. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
31. Kaji, H., Yamauchi, Y., Takahashi, N., and Isobe, T. (2006) Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site–specific stable isotope tagging. *Nat. Protocols* **1,** 3019–3027
32. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., and Ma, B. (2012) PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **11,** M111.010587
33. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–3676
34. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8,** 785–786
35. Eisenhaber, B., Wildpaner, M., Schultz, C. J., Borner, G. H., Dupree, P., and Eisenhaber, F. (2003) Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. *Plant Physiol.* **133,** 1691–1701
36. Park, B. H., Karpinets, T. V., Syed, M. H., Leuze, M. R., and Uberbacher, E. C. (2010) CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20,** 1574–1584
37. Bakalovic, N., Passardi, F., Ioannidis, V., Cosio, C., Penel, C., Falquet, L., and Dunand, C. (2006) PeroxiBase: A class III plant peroxidase database. *Phytochemistry* **67,** 534–539
38. Bräutigam, A., Hoffmann-Benning, S., and Weber, A. P. (2008) Comparative proteomics of chloroplast envelopes from C_3 and C_4 plants reveals specific adaptations of the plastid envelope to C_4 photosynthesis and candidate proteins required for maintaining C_4 metabolite fluxes. *Plant Physiol.* **148,** 568–579
39. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., and Hermjakob, H. (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41,** D1063–D1069
40. Showalter, A. M., Keppler, B., Lichtenberg, J., Gu, D., and Welch, L. R. (2010) A Bioinformatics approach to the Identification, Classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol.* **153,** 485–513
41. Nishitani, K., and Tominaga, R. (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. *J. Biol. Chem.* **267,** 21058–21064
42. Nishitani, K. (1995) Endo-xyloglucan transferase, a new class of transferase involved in cell wall construction. *J. Plant Res.* **108,** 137–148
43. Hansen, S. F., Harholt, J., Oikawa, A., and Scheller, H. V. (2012) Plant glycosyltransferases beyond CAZy: a perspective on DUF families. *Front. Plant Sci.* **3,** 59
44. Minic, Z., Jamet, E., Négroni, L., Arsene der Garabedian, P., Zivy, M., and Jouanin, L. (2007) A sub-proteome of Arabidopsis thaliana mature stems trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. *J. Exp Bot.* **58,** 2503–2512
45. Ligat, L., Lauber, E., Albenne, C., San Clemente, H., Valot, B., Zivy, M., Pont-Lezica, R., Arlat, M., and Jamet, E. (2011) Analysis of the xylem sap proteome of Brassica oleracea reveals a high content in secreted proteins. *Proteomics.* **11,** 1798–1813
46. Zhang, Y., Giboulot, A., Zivy, M., Valot, B., Jamet, E., and Albenne, C. (2011) Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry.* **72,** 1109–1123
47. Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J., and Frishman, D. (2008) Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **9,** 102–119
48. Cuskin, F., Flint, J. E., Gloster, T. M., Morland, C., Baslé, A., Henrissat, B., Coutinho, P. M., Strazzulli, A., Solovyova, A. S., Davies, G. J., and Gilbert, H. J. (2012) How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 20889–20894
49. Keller, B. (1993) Structural cell wall proteins. *Plant Physiol.* **101,** 1127–1130
50. Seror, J., Merkher, Y., Kampf, N., Collinson, L., Day, A. J., Maroudas, A., and Klein, J. (2011) Articular cartilage proteoglycans as boundary lubricants: structure and frictional interaction of surface-attached hyaluronan and hyaluronan–aggrecan complexes. *Biomacromolecules* **12,** 3432–3443
51. Knudsen, J. S., Bacic, A., and Clarke, A. E. (1998) Hydroxyproline-rich plant Glycoproteins. *Phytochemistry* **47,** 483–497
52. Padmalatha, K. V., Dhandapani, G., Kanakachari, M., Kumar, S., Dass, A., Patil, D. P., Rajamani, V., Kumar, K., Pathak, R., Rawat, B., Leelavathi, S., Reddy, P. S., Jain, N., Powar, K. N., Hiremath, V., Katageri, I. S., Reddy, M. K., Solanke, A. U., Reddy, V. S., and Kumar, P. A. (2012) Genome-wide transcriptomic analysis of cotton under drought stress reveal significant down-regulation of genes and pathways involved in fibre elongation and up-regulation of defense responsive genes. *Plant Mol. Biol.* **78,** 223–246
53. Padmalatha, K. V., Patil, D. P., Kumar, K., Dhandapani, G., Kanakachari, M., Phanindra, M. L., Kumar, S., Mohan, T. C., Jain, N., Prakash, A. H., Vamadevaiah, H., Katageri, I. S., Leelavathi, S., Reddy, M. K., Kumar, P. A., and Reddy, V. S. (2012) Functional genomics of fuzzless-lintless mutant of Gossypium hirsutum L. cv. MCU5 reveal key genes and pathways involved in cotton fibre initiation and elongation. *BMC Genomics* **13,** 624–639