Identifytion of N-linked Glycoproteins in Silkworm Serum Using Con A Lectin Affinity Chromatography and Mass Spectrometry

Zhaoming Dong,1,2,3* Lin Ye,1,2,3 Yan Zhang,1,2,3 Zhiyong Chen,1,2,3 Benchi Li,1,2,3 Tao Zhang,1,2,3 and Ping Zhao1,2,3,4

1 State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China, 2 Biological Science Research Center, Southwest University, Chongqing 400715, China, 3 Chongqing Key Laboratory of Sericultural Science, Southwest University, Chongqing 400715, China, and 4 Corresponding author, e-mail: zhaop@swu.edu.cn

Subject Editor: Amr Mohamed

Received 15 March 2021; Editorial decision 19 July 2021

Abstract

Glycosylation is one of the most common post-translational modifications to occur during protein biosynthesis, but remains poorly understood in insects. In this study, we collected serum proteins from two silkworm developmental stages, namely day 7 of the fifth instar larval stage and day 2 of the pupal stage. Results of SDS–PAGE and periodic acid-Schiff staining revealed that most serum proteins with high abundance were putative glycoproteins. LC-MS/MS identified 149 larval and 303 pupal serum proteins in the Con A lectin-enriched fractions. GO analysis revealed that many serum proteins were involved in the proteolysis and carbohydrate metabolic process. 82 N-linked glycoproteins with at least one glycosylation site were identified. N-Linked glycosylation occurred at the sequon, Asn-X-Ser/Thr, and the proportions of Ser and Thr glycosylation at the hydroxy position were found 39.6% and 60.3%, respectively. The N-glycan structures found in serum glycoproteins were mainly Man2FucGlcNAc2 (67.9%). Since storage protein 1 and transferrin had a relatively high abundance in the serum and could be significantly enriched by Con A lectin, their glycosylation was analyzed in detail. Glycoside hydrolases, serine proteases and serpins were found to form three interacting glycoprotein networks using the website STRING. This study provides important clues for the understanding of the function of N-linked glycosylation in metabolism, immunity, and metamorphosis.

Graphical Abstract

Key words: insect, N-glycan, glycosylation, glycoprotein, hemolymph

Glycosylation is one of the most common post-translational modifications (PTM) and plays an important role in different biological processes (Bertozzi and Kiessling 2001). It determines the localization, activity, and function of proteins and thus regulates various cellular processes (Kaji et al. 2003), including cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction, and endocytosis. Protein glycosylation mainly involves N-glycans and O-glycans. N-glycans are linked to...

© The Author(s) 2021. Published by Oxford University Press on behalf of Entomological Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
asparagine residues, whereas O-glycans are attached to serines and threonines (Schachter 2000).

Research on mammalian glycosylation has contributed to the diagnoses and treatment of various diseases. An increasing number of diseases related to glycosylation have been discovered, which are primarily severe syndromes (Ohtsubo and Marth 2006). Since glycoproteins comprise a major part of the serum proteome, and serum is the most commonly used fluid for diagnostics, researchers explored serum glycoproteins through lectin affinity chromatography and identified a large number of glycoproteins (Yang and Hancock 2004, Madera et al. 2007). Many of the serum glycoproteins were identified as transferrins, protease inhibitors, apolipoproteins, ceruloplasmins, and complement factors.

Insect glycoproteins have N-glycans and O-glycans with core structures similar to those produced by other eukaryotes, but they also contain some unique glycans that mediate specific functions (Walski et al. 2017, Scheys et al. 2018, Li et al. 2020). Insect glycoproteins fail to acquire antennae of the N-acetyllactosamine sugar or peripheral sugars and thus have few complex N-glycans (Chung et al. 2017). The major N-glycan structure found in insect glycoproteins is the paucimannose-type Man₃(Fuc)GlcNAc₂ (Yehuda and Padler-Karavani 2020).

In Drosophila, 138, 185, and 148 glycoproteins were identified by using Galanthus nivalis agglutinin (GNA), Nicotiana tabacum lectin (Nictaba), and Rhizoctoni solani agglutinin (RSA). GNA has a high specificity for paucimannose and highmannose N-glycans (Fouquaret et al. 2009), and Nictaba recognizes the Man,GlcNAc₂ core of N-glycans (Lanno et al. 2006), while RSA shows a high affinity for some mucin type O-glycans and terminal Gal,GalNAc residues on complex N-glycans (Blixt et al. 2004). Gene ontology (GO) analysis indicated that Drosophila glycoproteins were mainly involved in cell–cell adhesion, metabolic process, cell respiration, glycolysis, and tricarboxylic acid cycle (Vandenborre et al. 2010). In honeybees, 25 N-glycosylated royal jelly proteins with 53 N-glycosylation sites were identified, most of which were related to metabolic activities and health improvement (Zhang et al. 2014a). Some researchers identified 130, 49, 87, 118, and 218 putative glycoproteins from flour (Yang and Schachter 2000), honey bee (Apis mellifera), fruit fly (Drosophila melanogaster), and threonines (Schachter 2000). The maximum loss of 27%. The maximum ion injection times for the survey scan included days 1/3/5/7 of the fifth instar, day 2 of wandering, days 1/4/7 of the pupal stage, and day 1 of the moth stage. In each time points, four tubes of hemolymph were collected. Serum was collected by centrifuging the hemolymph at 4,000 g for 20 min at 4°C.

**Electrophoresis and Staining**

Two microliters of serum protein was separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 8% SDS-polyacrylamide gel. Following SDS-PAGE, separated proteins were stained with coomassie blue. For periodic acid-Schiff staining, protein gel was soaked in 7.5% (v/v) acetic acid for 30 min and then transferred in 0.2% (w/v) periodic acid for 1 hr at 4°C. After removing the periodic acid solution, the gel was incubated with the Schiff reagent for 1 hr at 4°C in the dark. The gel was soaked in 7.5% acetic acid for 1 hr and subsequently stored in water.

**Processing of Protein Samples From Polyacrylamide Gel**

Protein bands were excised from the polyacrylamide gel, placed in centrifuge tubes, and incubated with 100 μl decolorization solution (15 mM potassium ferric chloride and 50 mM sodium thiosulfate). After removing the decolorization solution, excised gel bands were washed three times with Milli-Q water and then incubated with 100% acetonitrile. The acetonitrile was removed when the gel appeared milky. The gel was then allowed to dry for 20 min, following which the gel bands were incubated with trypsin (10 μg/ml) at 37°C for 24 hr. Tryptic peptides were extracted from gels, by soaking in 25 μl of 30% acetonitrile/5% trifluoroacetate for 60 min, concentrated, and resuspended in 0.1% formic acid.

**Con A Affinity Chromatography**

A Con A Sepharose 4B (GE Healthcare, USA) column (1 ml) was equilibrated with 5 ml binding buffer (20 mM Tris–HCl, 150 mM NaCl, 1 mM MgCl₂, 1 mM CaCl₂, 1 mM MnCl₂, pH 7.4). Serum (0.4 ml) from day 7 of the fourth larval instar and day 2 of the pupal stage was diluted five times with binding buffer and then loaded onto the Con A column. After washing the column with 10 ml binding buffer, Con A-bound proteins were eluted with 5 ml 300 mM α-D-mannose in the binding buffer. Con A affinity chromatography was performed in four biological replicates.

**Liquid Chromatography-Tandem Mass Spectrometry**

Con A-enriched proteins were reduced with 10 mM dithiothreitol (DTT) for 150 min at 37°C and then alkylated with 50 mM iodoacetamide (IAA) for 40 min in the dark. After washing twice with 50 mM NH₄HCO₃, proteins were incubated with trypsin (1 μg/50 μg protein) overnight at 37°C in 150 μl of 50 mM NH₄HCO₃. Tryptic peptides were concentrated and resuspended in 0.1% formic acid.

Tryptic peptides were separated using the Thermo Fisher Scientific EASY-nLC 1000 system and EASY-Spray column (C18, 2 μm, 100 Å, 75 μm x 50 cm) with a 2–100% acetonitrile gradient in 0.1% formic acid over 180 min at a flow rate of 250 nl/min. The separated peptides were analyzed using a Thermo Scientific Q Exactive mass spectrometer operating in data-dependent mode. Up to 10 of the most abundant isotopic patterns with charge ≥2 from an initial survey scan were automatically selected for fragmentation by higher energy collisional dissociation with normalized collision energies of 27%. The maximum ion injection times for the survey scan...
and the MS/MS scans were 20 and 60 ms, respectively, and the ion target value for both scan modes was set to 1E6. An 18-s dynamic exclusion of previously sequenced ions was applied.

**Protein Identification and Quantification**

The resulting raw MS data were analyzed with the MaxQuant software (version 1.3.0.1) (Cox and Mann 2008). MaxQuant searches were executed against an integrated silkworm proteome database containing protein sequences from silkDB, Genbank, and Uniprot. Peptide searches were performed with the Andromeda search algorithms (Cox et al. 2011). The search parameters for protein identification specified the initial precursor and fragment mass tolerances of 6 and 20 ppm, respectively. Carbamidomethylation of cysteine was set as a fixed modification, and N-terminal protein acetylation and methionine oxidation were set as variable modifications. The minimal peptide length was set to six amino acids, and up to two miscleavages were allowed. The false discovery rate was set to 0.01 for both peptides and proteins. All common contaminants and reverse hits were removed. A minimum of one unique peptide was required for an identified protein. For comparison of protein abundance, we used the intensity-based absolute quantification (iBAQ) algorithm in MaxQuant. To compare the protein abundance in the Con A-binding fraction with that of total serum protein, an unpaired t-test was used with a significance level set at P < 0.05. The heat maps of protein abundance were generated using HemI (Heatmap Illustrator, version 1.0.3.3) (Deng et al. 2014).

**Glycoprotein Identification**

The raw MS data were searched with the Byonic software (version 2.13.17) against the Con A-enriched serum protein database (Bern et al. 2012). N-glycosylation was searched for with the following search parameters: precursor mass tolerance = 10 ppm, fragmentation type = QTOF/HCD, fragment mass tolerance = 0.02 Da; a maximum of two missed cleavages was allowed; modifications: Carbamidomethyl/+57.021464 @ C (fixed), Oxidation/+15.994915 @ N, Q (common1), Acetyl/+42.010565 @ Protein NTerm (common1), insect N-glycan (Glycan modifications). Data with a Byonic score ≥200, Delta Mod ≥10, |Log Prob| ≥3, number of unique peptides >1, protein false discovery rate (FDR) <1% was filtered. Furthermore, a manual inspection was performed for the quality of each glycopeptide.

**Bioinformatics Analysis**

The GO annotation was performed using the Singular Enrichment Analysis tool in the website AgriGO (Tian et al. 2017). The enrichment value of each GO term was defined as the ratio of protein number in the input list to protein number in the reference. The protein interaction network model was constructed by using the website STRING (Szklarczyk et al. 2011). The small interaction networks with less than five proteins were excluded. The domains and functions of proteins were annotated automatically by STRING.

**Results**

**Visualization of Glycoproteins in the Silkworm Serum**

Silkworm serum proteins were collected from nine developmental stages, including five larval stages, three pupal stages, and one adult stage, and were then separated with SDS-PAGE (Fig. 1). To visualize the serum glycoproteins, protein gels were stained with periodic acid-Schiff staining (Fig. 1A) and coomassie blue (Fig. 1B). Most proteins were stained with both reagents, including proteins with a molecular weight of >170 kDa (protein band 1), 170 kDa (protein band 2), 130 kDa (protein band 3), 90 kDa (protein band 4), 80 kDa (protein band 5), 50 kDa (protein band 6), 40 kDa (protein band 7), 30 kDa (protein band 8) (Fig. 1). Gel-based mass spectrometry identification revealed that each protein band is not a single protein, but contains 6–24 proteins (Supp Table 1 [online only]). However, the major proteins in each band occupy a very high abundance from 87.3% to 99.1% (Table 1). The major proteins in bands 1–8 are as follows: apolipophorin-I (93.5% in protein band 1), vitellogenin (98.4% in protein band 2), inter-alpha-trypsin inhibitor (97.9% in protein band 3), storage protein SP1 (93.2% in protein band 4), storage protein SP2 (98.6%), and Serpin9 & Serpin1 (99.1%).

**Fig. 1.** Electrophoretic analysis of serum proteins. (A) Periodic acid-Schiff staining following SDS–PAGE of serum protein. (B) Coomassie blue staining following SDS–PAGE of serum proteins. Two microliters of serum protein was run on an 8% polyacrylamide gel. M, protein marker; V1/V3/V5/V7/W2/P1/P4/P7/M1, day 1 of moth. The horizontal lines connected by numbers indicate protein bands. The proteins from each band were identified using a Q Exactive mass spectrometer and MaxQuant software. The abundance proportion of the major proteins in each band was calculated according to their iBAQ intensity (Table 1 and Supp Table 1 [online only]).
Identification of the major silkworm hemolymph proteins

Table 1. Identification of the major silkworm hemolymph proteins

| Band No. | Annotated namea | Uniprot ID | Genebank ID | Peptide number | Sequence coverage [%] | MW [kDa] | Relative abundanceb |
|----------|-----------------|------------|-------------|---------------|-----------------------|----------|---------------------|
| Band 1   | Apolipophorin-I | tr|G1UIS8| gi|1200716789 | 226 | 62.5% | 369 | 93.5% |
| Band 2   | Vitelligenin    | sp|Q27309| – | 79 | 54 | 203 | 98.4% |
| Band 3   | Inter-alpha-trypsin inhibitor heavy chain | – | gi|1200719391 | 27 | 27.1 | 99 | 97.9% |
| Band 4   | Storage protein 1 | sp|P09179| – | 89 | 73.4 | 87 | 93.2% |
| Band 5   | Storage protein 2 | – | gi|1174445 | 76 | 72.3 | 83 | 98.6% |
| Band 6   | Imaginal disk growth factor | – | gi|1200731398 | 21 | 42.2 | 49 | 87.3% |
| Band 7   | Serpin9         | sp|Q03383| – | 15 | 37 | 45 | 62.6% |
| Band 7   | Serpin1         | tr|C7ASM9| – | 26 | 64.3 | 43 | 36.5% |
| Band 8   | BmLP1           | sp|P09334| – | 26 | 77 | 30 | 39.2% |
| Band 8   | BmLP2           | tr|H9B444| – | 25 | 59.8 | 30 | 16.1% |
| Band 8   | BmLP4           | sp|Q08081| – | 20 | 62.4 | 30 | 14.7% |
| Band 8   | BmLP3           | sp|Q0802| – | 26 | 68.4 | 30 | 14.5% |
| Band 8   | BmLP9           | tr|H9J4G0| – | 33 | 57.6 | 44 | 11.8% |
| Band 9   | Apolipophorin-III | tr|H9JU96| – | 14 | 48.2 | 28 | 95.0% |

b“BmLP” are nomenclature of B. mori 30K proteins reported by Zhang et al. (2012).

Relative abundance was calculated according to the iBAQ intensity of each protein in Supp Table 1 (online only).

SP2 (98.6% in protein band 5), imaginal disk factor (87.3% in protein band 6), serpins (serpin9 and serpin1) (99.1% in protein band 7), and 30K proteins (Bmlp1, Bmlp2, Bmlp3, Bmlp4, and Bmlp9) (96.4% in protein band 8) (Fig. 1, Table 1 and Supp Table 1 [online only]), which are putative glycoproteins. Protein band 9 was identified as apolipophorin-III (95.0% in Band 9) (Fig. 1, Table 1 and Supp Table 1 [online only]), which could not be stained with periodic acid-Schiff staining and thus may not be a glycoprotein (Fig. 1).

Shotgun Proteomics Analysis of Con A Lectin-Enriched Proteins in the Silkworm Serum

The Con A lectin affinity column was used to enrich N-linked glycoproteins from silkworm serum at two developmental stages, day 7 of the larval fifth instar (V7) and day 2 of the pupal stage (P2). LC-MS/MS was employed to identify Con A lectin-enriched serum proteins. Combining the results of four biological replicates, we identified 202 larval serum proteins (V7 T), 149 lectin-enriched larval serum proteins (V7 L), 268 pupal serum proteins (P2 T), and 303 lectin-enriched pupal serum proteins (V7 T) (Fig. 2A Supp Table 2 [online only]). A total of 389 serum proteins were identified by LC-MS/MS, which were classified into six categories based on annotated molecular function according to previous reports (Dong et al. 2016, Chen et al. 2019), including enzyme (135), immune response (84), binding and transport (63), extracellular matrix (22), unknown function (50), and other function (35). By comparing protein numbers between serum proteins and lectin-enriched serum proteins, we found that more enzymes and proteins of unknown function were observed in lectin-enriched pupal serum proteins (Fig. 2B).

We estimated protein abundance using LC−MS/MS and the iBAQ algorithm and found that enzymes and proteins of unknown function showed higher abundance in lectin-enriched larval and pupal serum when compared to normal serum (Fig. 2C). Some proteins were highly abundant in serum but were not enriched by Con A lectin (Fig. 2D), including storage protein 1 (Genebank ID: gi|1335609), imaginal disk growth factor (Genebank ID: gi|152061158), and transferrin (SilkDB ID: BGIBMGA011424).

GO Analysis of Con A Lectin-Enriched Proteins in the Silkworm Serum

The GO annotation of silkworm serum proteins at two developmental stages was predicted by using the Singular Enrichment Analysis tool in AgriGO. In the molecular function, oxygen transporter activity (GO:0005344) is the most important GO term in the total serum proteins (Fig. 3A and B), and not found in the lectin-enriched serum proteins, indicating that most of the proteins with oxygen transporter activity may not be glycoproteins. The lectin-enriched larval and pupal serum proteins share similar molecular function GO terms, including peptidase/protease inhibitor activity (GO:0004866, GO:0004867, GO:0030414, GO:0004857), peptidase/protease activity (GO:0004252, GO:0008236, GO:0017171), and glycoside hydrolase activity (GO:0005454, GO:0016798). In the biological process, proteolysis (GO:0005975) and carbohydrate metabolic process (GO:0009056) are two important GO terms for lectin-enriched larval and pupal serum proteins (Fig. 3C and D). The GO terms of the biological process are highly correlated with the GO terms of molecular function. For instance, the carbohydrate metabolic process is related to the glycoside hydrolase activity, while the proteolysis is associated with the peptidase/protease activity and peptidase/protease inhibitor activity. Interestingly, the adhesion GO terms (GO:0007155 and GO:0022610) are unique for lectin-enriched larval serum proteins (Fig. 3C and D). In the cellular component, the extracellular region GO terms (GO:00053576, GO:0044421) and extracellular matrix GO terms (GO:00053578, GO:0030102) indicated that serum proteins are secreted out of the cell to play roles (Fig. 3E and F).

Identification of Silkworm Serum Proteins That Were Significantly Enriched by Con A Lectin

To determine which proteins were significantly enriched by Con A lectin, we calculated the ratio of protein abundance as
(enriched/unenriched) > 2 and considered P < 0.05 to be statistically significant. We found that a total of 139 glycoproteins were significantly enriched by Con A lectin, including 35 enzymes, 28 immune-related proteins, 25 proteins of unknown function, 17 binding and transport proteins, six extracellular matrix proteins, and eight other proteins (Fig. 4 and Supp Table 2 [online only]). The enzymes mainly contain glycoside hydrolases, esterases, and peptide/proteases, and the immune-related proteins are composed of immunoglobulins, serine proteases, serine protease inhibitors, and lectins.

Identification of N-linked Glycoproteins in the Silkworm Serum

Through the Byonic software search, we identified 82 N-linked glycoproteins with at least one glycosylation site, 63 of which were significantly enriched by Con A lectin (Fig. 5A and Supp Table 3 [online only]).

One to four glycosylation sites were found in each N-linked glycoprotein. We found that N-Linked glycosylation occurred at the sequon, Asn-X-Ser/Thr (Fig. 5B, Supp Table 4, and Supp Fig. 1 [online only]). The proportions of Trp and Pro were found to be 14.2%, 10.3%, 9.4%, 9.4%, and 9.4%, respectively, whereas Tyr, Cys, His, and Met accounted for 0.9%, 1.8%, 1.8%, and 1.8%, respectively. No Trp or Pro were identified at the X position.

The N-glycan structures found on serum glycoproteins were mainly HexNAc(2)Hex(2)Fuc(1) (67.9%), HexNAc(2)Hex(2) (10.4%), HexNAc(2)Hex(3) (6.6%), and HexNAc(2)Hex(3)Fuc(1) (5.7%) (Fig. 5C and Supp Table 4 [online only]). Since the N-glycan structures in insect glycoproteins are mainly HexNAc(2)Hex(2)Fuc(1) (67.9%), HexNAc(2)Hex(2)Fuc(1) (6.6%), and Man3FucGlcNAc2 (5.7%), which is consistent with previously reported paucimannose structure in the insects (Yehuda and Padler-Karavani 2020).

The N-glycan structures found on serum glycoproteins were mainly HexNAc(2)Hex(2)Fuc(1) (67.9%), HexNAc(2)Hex(2) (10.4%), HexNAc(2)Hex(3) (6.6%), and HexNAc(2)Hex(3)Fuc(1) (5.7%) (Fig. 5C and Supp Table 4 [online only]). Since the N-glycan structures in insect glycoproteins are mainly HexNAc(2)Hex(2)Fuc(1) (67.9%), HexNAc(2)Hex(2)Fuc(1) (6.6%), and Man3FucGlcNAc2 (5.7%), which is consistent with previously reported paucimannose structure in the insects (Yehuda and Padler-Karavani 2020).

Identification of N-linked Glycoproteins in the Silkworm Serum

Through the Byonic software search, we identified 82 N-linked glycoproteins with at least one glycosylation site, 63 of which were significantly enriched by Con A lectin (Fig. 5A and Supp Table 3 [online only]).

One to four glycosylation sites were found in each N-linked glycoprotein. We found that N-Linked glycosylation occurred at the sequon, Asn-X-Ser/Thr (Fig. 5B, Supp Table 4, and Supp Fig. 1 [online only]), which was consistent with previous reports (Gavel and von Heijne 1990, Shakin-Eshleman et al. 1996). The proportions of Ser and Thr at the hydroxy position were found to be 39.6% and 60.3%, respectively. The proportions of Leu, Gly, Ile, Ala, and Ser in the X position were found to be 14.2%, 10.3%, 9.4%, 9.4%, and 9.4%, respectively, whereas Tyr, Cys, His, and Met accounted for 0.9%, 1.8%, 1.8%, and 1.8%, respectively. No Trp or Pro were identified at the X position.

The N-glycan structures found on serum glycoproteins were mainly HexNAc(2)Hex(2)Fuc(1) (67.9%), HexNAc(2)Hex(2) (10.4%), HexNAc(2)Hex(3) (6.6%), and HexNAc(2)Hex(3)Fuc(1) (5.7%) (Fig. 5C and Supp Table 4 [online only]). Since the N-glycan structures in insect glycoproteins are mainly HexNAc(2)Hex(2)Fuc(1) (67.9%), HexNAc(2)Hex(2)Fuc(1) (6.6%), and Man3FucGlcNAc2 (5.7%), which is consistent with previously reported paucimannose structure in the insects (Yehuda and Padler-Karavani 2020).

N-linked Glycosylation of Storage Protein 1 and Transferrin

Among Con A lectin-enriched glycoproteins, storage protein 1 (Genebank ID: gi|1335609), imaginal disk growth factor (Genebank ID: gi|1335609), and transferrin (SilkDB ID: BGIBMGA011424) had the highest abundance (Fig. 2D). Thus, we investigated the N-linked glycosylation of these three proteins and found 0, 1, and 2 N-Linked glycosylation sites on imaginal disk growth factor, transferrin, and storage protein 1, respectively (Supp Table 4 [online only]).

Collision-induced fragmentation of the doubly- and triply-charged ions revealed the presence of five different high-mannose type N-glycans (Man$_n$GlcNAc$_m$) on the N318 residue in tryptic peptide 317-$\text{ANYTEVIER}$-325 of transferrin. Fig. 6A shows an example of an MS/MS spectrum for the triply-charged precursor (m/z 1864.634), which was identified to be the Man$_4$GlcNAc$_2$ glycoform
at the N318 site. The MS/MS spectrum provided direct evidence for the high mannose-type glycosylation linked to the N318 site. A summary of all N-linked sites and the heterogeneous glycoforms for transferrin identified in this study is provided in Supp Table 4.

For storage protein 1, eight different glycoforms (Man$_{0-4}$GlcNAc$_{1-2}$Fuc$_{0-1}$) were identified on the N494 residue of the tryptic peptide 465-ITTDELVTFVDEYDMDISNAMYLDATEM QNK-495. Fig. 6B shows an example of an MS/MS spectrum for the quadruply-charged precursor (m/z 1038.375), which was identified to be the Man$_{3}$GlcNAc$_{2}$Fuc$_{1}$ glycoform at the N494 site. A summary of all N-linked sites and heterogeneous glycoforms identified for storage protein 1 in this study is given in Supp Table 4 [online only].

Interaction Network of N-linked Glycoproteins

The 139 ConA lectin-binding proteins and 82 N-linked glycoproteins were used to construct a protein interaction network model

Fig. 3. Gene Ontology analysis of total serum proteins and Con A lectin-enriched serum proteins. GO terms enrichment analysis was performed using AgriGO. GO terms were divided into three parts: molecular function (A, B), biological process (C, D), and cellular component (E, F). GO terms are labeled using the GO number, term definition, and enrichment value. GO terms of total serum proteins from day 7 of the fifth larval instar (V7 T) and that from day 2 of pupae (P2 T) are showed as blue bars, and GO terms of lectin-enriched serum proteins from day 7 of the fifth larval instar (V7 L) and that from day 2 of pupae (P2 L) are red bars.
### Table A: Protein ID and Name

| Protein ID | Protein Name |
|------------|--------------|
| BGIBMGA007199 | Hydrodase |
| BGIBMGA012267 | Alpha-mannosidase |
| BGIBMGA003581 | Ubiquitin-protein ligase |
| BGIBMGA009885 | Beta-lactalbumin |
| BGIBMGA012647 | 5 -nucleotidease |
| BGIBMGA001387 | Aldo-keto reductase |
| BGIBMGA009352 | Antithrombin |
| BGIBMGA010071 | Proenkephalin |
| BGIBMGA001330 | Protein take-3-albumin |
| BGIBMGA005809 | Zinc transporter foli-1 |
| BGIBMGA001315 | Thrombospondin |
| BGIBMGA009487 | Glucosidase |
| BGIBMGA008231 | Divinylglycol }

### Table B: Protein ID and Name

| Protein ID | Protein Name |
|------------|--------------|
| BGIBMGA010546 | Serine protease CLIP8 |
| BGIBMGA009317 | Carbamoyl esterase |
| BGIBMGA000133 | Osiris 9 |
| BGIBMGA012616 | Arachidonic acid 1-lyase |
| BGIBMGA000144 | Uncharacterized protein |
| BGIBMGA010040 | Uncharacterized protein LOC10174936 |
| BGIBMGA002845 | Ester hydrolyase c11or8 homolog |

### Table C: Protein ID and Name

| Protein ID | Protein Name |
|------------|--------------|
| BGIBMGA010813 | Serine ocase |
| BGIBMGA007767 | Glycophosphodiester phosphodiesterase |
| BGIBMGA010044 | Uncharacterized protein |
| BGIBMGA010040 | Uncharacterized protein LOC10174936 |
| BGIBMGA000146 | Beta-fructofuranosidase BmSuc1 |
| BGIBMGA014407 | Serine protease BmSP24CLIP2 |
| BGIBMGA012239 | Basement membrane-specific heparan sulfate proteoglycan |
| BGIBMGA013358 | Proteasome subunit |
| BGIBMGA009331 | Indole-3-indoleamide hydrolase |
| BGIBMGA010194 | Beta-N-acetylglucosaminidase |
| BGIBMGA013356 | Protein takeout-like |
| BGIBMGA000079 | Osiris 9 |
| BGIBMGA003126 | Carboxylic ester hydrolase |
| BGIBMGA000064 | C-type lectin 10 |
| BGIBMGA010863 | Ferritin light chain |
| BGIBMGA000065 | C-type lectin 11 |
| BGIBMGA000069 | Ferritin light chain |
| BGIBMGA003126 | Carboxylic ester hydrolase |
| BGIBMGA010863 | Ferritin light chain |
| BGIBMGA000065 | C-type lectin 11 |

### Fig. 4. Con A lectin-binding ability of proteins from silkworm serum. A total of 139 proteins were significantly enriched by Con A lectin according to the following conditions: the ratio of protein abundance >2 and P < 0.05. Con A lectin-binding ability was quantified according to the iBAQ intensity of the Con A lectin-binding protein divided by the iBAQ intensity of the extracted protein. This heat map was generated using Heml. Colors indicate the protein category: enzymes, green; immune response proteins, red; binding and transport proteins, blue; proteins of unknown function, light grey; extracellular matrix protein, orange; other, dark grey.
by using the website STRING. The model indicated that 10 glyco-
side hydrolases formed an interaction network, while four serpins
interacted with 10 serine proteases from another network (Fig. 7).
Among them, six glycoside hydrolases and four serine proteases
were not only significantly enriched by ConA lectin, but also proved as
glycopeptide-containing proteins, whereas four glycoside hydrolases,
three serpins, and six serine proteases were ConA lectin-enriched pro-
teins but lack of glycopeptide evidence. There were six CLIP-domain
serine proteases in the network, including BGIBMGA005173 (CLIP
2), BGIBMGA010546 (CLIP 8), BGIBMGA010306 (CLIP 7/12),
BGIBMGA009551 (CLIP 17), BGIBMGA005380 (CLIP 25),
BGIBMGA014407 (CLIP 26), which may help to amplify the initial
immune recognition signal, and are negatively regulated by serpins
(Christophides et al. 2002).

Discussion

Currently, many researches on insect glycosylation focused on the
foreign recombinant glycoproteins expressed in insect cells (Marchal
et al. 2001, Rendic et al. 2008, Chung et al. 2017). The data about
the glycosylation of the insects' own proteins was mainly obtained
from studies with fruit fly, D. melanogaster (Vandenborre et al.
2010, Baycin-Hizal et al. 2011). In this study, we collected silkworm
serum proteins, enriched the N-linked glycoproteins by Con A lectin
chromatography, and identified them with LC-MS/MS. The find-
ings of this work provided details on the characterization of amino
acid site occupancy and glycan composition in silkworm serum
glycoproteins.

Using periodic acid-Schiff staining, nearly all the high-abundance
proteins in silkworm serum, such as apolipophorin-I, vitellogenin,
storage proteins, 30K proteins, serpins, and inter-α-trypsin
inhibitor, were found to be glycoproteins. Previous studies suggested
that apolipophorin-I, vitellogenin, storage proteins, and 30K pro-
teins play roles in nutrient storage (Zhang et al. 2014b), including
providing sources of amino acids and lipid transportation, whereas
serpins and the inter-α-trypsin inhibitor act as protease inhibitors
to maintain a stable body fluid environment (Zhao et al. 2012).

Using the Con A lectin affinity column to enrich N-linked glyco-
proteins, we found that storage protein 1, transferrin, and imaginal
disk growth factor were significantly enriched by Con A lectin,
whereas storage protein 2 and 30K proteins LP1 and LP2 could
not be enriched. 30K proteins have a sugar-binding CTD domain
and were reported to bind glucans (Ujita et al. 2002, Ujita et al.
2005, Yang et al. 2011). 30K proteins are sugar-binding proteins
and may not be N-linked glycoproteins, and thus could not be en-
riched by Con A lectin. Storage proteins are also called arylphorins,
used primarily as a source of aromatic amino acids for protein syn-
thesis during metamorphosis. In the tobacco worm, the storage
protein arylphorin is a hexameric glycoprotein and contains two
subunits, A1 and A2. Both subunits are glycosylated and have mo-
lecular weights of 77 kDa and 72 kDa, respectively. NMR spec-
troscopy of arylphorin glycopeptides revealed a Man\textsubscript{6}GlcNAc\textsubscript{2}
The results of comparative glycoproteomics among five insect species (Vandenborre et al. 2011). Interestingly, more glycosylated hydrolases were identified in the silkworm than in other insects (Vandenborre et al. 2011), and more glycosylated hydrolase were found in the pupae than in the larvae. Furthermore, the metabolic processes are found to be mainly related to the carbohydrates and proteins, whereas the substrates of hydrolases are also carbohydrates and proteins.

By interaction network analysis, we found that these protein hydrolases in the silkworm serum are serine proteases, most of which have CLIP domains. CLIP-domain serine proteases rapidly amplify the immune signal after pathogens are recognized, and finally lead to the formation of melanin and production of antimicrobial peptides (Zhang et al. 2014b, Chen et al. 2019). By using Edman degradation,
two N-linked glycosylation sites were observed in a silkworm CLIP serine protease, the prophenoloxidase-activating enzyme (PPAE) (Satoh et al. 1999). Serpins regulate immune response via inhibiting CLIP-domain serine proteases, and were enriched by ConA lectin in this study. The glycosylation of immune-related proteins may enhance protein stabilization, mediate interaction with pathogens, and modulate immune responses (Dwek 1996, Spiro 2002, Lis and Sharon 2005). More immune-related glycoproteins were found in the pupal hemolymph than in the larval hemolymph, may reflecting that pupae need a more stable and efficient immune response system than larvae.

Among glycoside hydrolases, three of which are α-mannosidases, including BGIBMGA002426, BGIBMGA005142, and BGIBMGA012267. These mannosidases showed higher abundance in the pupal stage than in the larval stage, and may be involved in trimming of high-mannose N-glycans to form paucimannose N-glycans (Walski et al. 2016). In T. castaneum, RNAi of α-mannosidases caused pupa malformation and prevented adult eclosion, and thus demonstrated that α-mannosidases involved in the N-glycan processing are crucial for insect metamorphosis (Walski et al. 2016).

**Conclusions**

In this study, 139 silkworm serum proteins were identified from Con A lectin-enriched fractions, 63 of which were proved as N-linked glycoproteins with at least one glycosylation site. The N-linked glycosylation sites and N-glycan structures were analyzed in detail. The GO analysis and interaction network revealed that N-linked glycosylation is important for glycoside hydrolases, serine proteases, and serpins. This study provides important clues for understanding the roles of N-linked glycosylation in insect immunity and metamorphosis.

**Supplementary Data**

Supplementary data are available at Journal of Insect Science online.

**Acknowledgments**

This work was supported by the Chongqing Research Program of Basic Research and Frontier Technology (cstc2019jcyj-msxmX0272), and the Natural Science Foundation of Chongqing (cstc2020jcyj-cxttX0001).

**Author Contributions**

Conceptualization, Z.D. and P.Z.; methodology, Z.D. and P.Z.; software, Z.D. and T.Z.; validation, Y.Z., Z.C., and B.L.; formal analysis, Z.D.; investigation, Z.D. and L.Y.; writing—original draft, Z. D.; writing—review and editing, P.Z.; supervision, P.Z.; project administration, P.Z.; funding acquisition, Z.D. and P.Z.

**References Cited**

Baycin-Hizal, D., Y. Tian, I. Akan, E. Jacobson, D. Clark, J. Chu, K. Palter, H. Zhang, and M. J. Betenbaugh. 2011. GlycoFly: a database of Drosophila N-linked glycoproteins identified using SPEG-MS techniques. J. Proteome Res. 10: 2777–2784.

Bern, M., Y. J. Kil, and C. Becker. 2012. Byonic: advanced peptide and protein identification software. Curr. Protoc. Bioinform. Chapter 13: Unit13.20.

Bertozzi, C. R., and L. L. Kiesling. 2001. Chemical glycoecology. Science. 291: 2357–2364.

Blixt, O., S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calaresu, J. Stevens, et al. 2004. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc. Natl. Acad. Sci. U. S. A. 101: 17033–17038.
Chen, S., Z. Dong, X. Ren, D. Zhao, Y. Zhang, M. Tang, J. Han, L. Ye, and P. Zhao. 2019. Proteomic identification of immune-related silkworm proteins involved in the response to bacterial infection. J. Insect Sci. 19: 13.

Christophides, G. K., E. Zdobnov, C. Barillas-Mury, E. Birney, S. Blandin, C. Blass, P. T. Brey, F. H. Collins, A. Danielli, G. Dimopoulos, et al. 2002. Immunity-related genes and gene families in Anopheles gambiae. Science. 298: 159–165.

Chung, C. Y., N. I. Majewksa, Q. Wang, J. T. Paul, and M. J. Benetbaugh. 2017. SnapShot: N-glycosylation processing pathways across kingdoms. Cell. 171: 258–258.e1.

Cox, J., and M. Mann. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26: 1367–1372.

Cox, J., N. Neuhauer, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 10: 1794–1805.

Deng, W., Y. Wang, Z. Liu, H. Cheng, and Y. Xue. 2014. HemI: a toolkit for illustrating heatmaps. PLoS One. 9: e111988.

Dong, Z., P. Zhao, Y. Zhang, Q. Song, X. Zhang, P. Guo, D. Wang, and Q. Xia. 2016. Analysis of proteome dynamics inside the silk gland lumen of Bombyx mori. Sci. Rep. 6: 21158.

Dwek, R. A. 1996. Glycobiology: toward understanding the function of sugars. Chem. Rev. 96: 683–720.

Fouquart, E., D. F. Smith, W. J. Peumans, P. Proost, J. Balzarini, S. N. Savvides, and E. J. Damme. 2009. Related lectins from snowdrop and maize differ in their carbohydrate-binding specificity. Biochem. Biophys. Res. Commun. 380: 260–265.

Gavel, Y., and G. von Heijne. 1990. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. Protein Eng. 3: 433–442.

Kaji, H., H. Saito, Y. Yamauchi, T. Shinkawa, M. Taoka, J. Hirabayashi, K. Kaisai, N. Takahashi, and T. Isole. 2003. Lectin affinity capture, isothe-coded tagging and mass spectrometry to identify N-linked glycoproteins. Nat. Biotechnol. 21: 667–672.

Lannoos, N., W. J. Peumans, E. V. Pamul, R. Alvarez, T. C. Xiong, G. Hause, C. Maziars, and E. J. Van Damme. 2006. Localization and in vitro binding studies suggest that the cytoplasmic/nuclear tobacco lectin can interact in situ with high-mannose and complex N-glycans. FEBS Lett. 580: 6329–6337.

Li, W., K. De Schutter, E. J. M. Van Damme, and G. Smagghe. 2020. Synthesis and biological roles of N-glycans in insects. Glycoconjug. J. 37: 47–56.

Lis, H., and N. Sharon. 2005. Protein glycosylation: structural and functional aspects. FEBS J. 218: 1–27.

Madera, M., Y. Mechref, I. Klouckova, and M. V. Novotny. 2007. High-sensitivity isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins from insect cells: scaled up or not? Bioil. Chem. 382: 151–159.

Ohtsubo, K., and J. D. Marth. 2006. Glycosylation in cellular mechanisms of health and disease. Cell. 126: 855–867.

Rendic, D., I. B. H. Wilson, and K. Paschinger. 2008. The glycosylation capacity of insect cells. Croat. Chem. Acta. 81: 7–21.

Ryan, R. O., D. R. Anderson, W. J. Grimes, and J. H. Law. 1985. Aplyrhin from Manduca sexta: carbohydrate structure and immunological studies. Arch. Biochem. Biophys. 243: 115–124.

Sato, H., A. Horii, M. Ochiai, and M. Ashida. 1999. Prophenoloxidase-activating enzyme of the silkworm, Bombyx mori: purification, characterization, and CDNA cloning. J. Biol. Chem. 274: 7441–7453.

Schachter, H. 2000. The joys of HexNaC. the synthesis and function of N- and O-glycan branches. Glycoconjug. J. 17: 465–483.

Scheys, F., E. J. M. Van Damme, K. De Schutter, A. Staes, K. Gevaert, and G. Smagghe. 2018. Evolutionarily conserved and species-specific glycoproteins in the N-glycoproteomes of diverse insect species. Insect Biochem. Mol. Biol. 100: 22–29.

Shakin-Eshelman, S. H., S. L. Spitalnik, and L. Kasturi. 1996. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. J. Biol. Chem. 271: 6363–6366.

Spiro, R. G. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology. 12: 43R–56R.

Szklarzcyk, D., A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Mingeuez, T. Doerks, M. Stark, J. Muller, P. Bork, et al. 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39: D561-D568.

Tian, Y., T. Liu, H. Yan, Q. You, Y. Yi, Z. Wu, X. Zuo, and Z. Su. 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community. Nucleic Acides Res. 45: 122–129.

Ujita, M., A. Kimura, D. Nishino, E. Yokoyama, Y. Banno, H. Fujii, and A. Haras. 2002. Specific binding of silkworm Bombyx mori 30-kDa lipopolysaccharides to carbohydrates containing glucose. Biosci. Biotechnol. Biochem. 66: 2264–2266.

Ujita, M., Y. Katsuno, I. Kawachi, Y. Ueno, Y. Banno, H. Fujii, and A. Haras. 2005. Glucan-binding activity of silkworm 30-kDa apolipoprotein and its involvement in defense against fungal infection. Biosci. Biotechnol. Biochem. 69: 1178–1185.

Vandenborre, G., E. J. Van Damme, B. Ghesquière, G. Mensaert, M. Hamshou, R. N. Rao, K. Gevaert, and G. Smagghe. 2010. Glycosylation signatures in Drosophila: fishing with lectins. J. Proteome Res. 9: 3235–3242.

Vandenborre, G., G. Smagbhe, B. Ghesquière, G. Mensaert, R. Nagender Rao, K. Gevaert, and E. J. Van Damme. 2011. Diversity in protein glycosylation among insect species. PLoS One. 6: e16682.

Walski, T., E. J. Van Damme, N. Smargiasso, O. Christiania, O. De Pauw, and G. Smagghe. 2016. Protein N-glycosylation and N-glycan trimming are required for postembryonic development of the pest beetle Tribolium casteum. Sci. Rep. 6: 35151.

Walski, T., K. De Schutter, E. J. M. Van Damme, and G. Smagghe. 2017. Diversity and functions of protein glycosylation in insects. Insect Biochem. Mol. Biol. 83: 21–34.

Yang, Z., and W. S. Hancock. 2004. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. J. Chromatogr. A. 1053: 79–88.

Yang, J. P., X. X. Ma, Y. X. He, W. F. Li, Y. Kang, R. Bao, Y. Chen, and C. Z. Zhou. 2011. Crystal structure of the 30 K protein from the silkworm Bombyx mori reveals a new member of the β-trefoil superfamily. J. Struct. Biol. 175: 97–103.

Yehuda, S., and V. Padler-Karavani. 2020. Glycosylated biotherapeutics: immunological effects of N-glycolylneuraminic acid. Front. Immunol. 11: 21.

Zhang, L., B. Han, R. Li, X. Lu, A. Nie, L. Guo, Y. Fang, M. Feng, and J. Li. 2014a. Comprehensive identification of novel proteins and N-glycosylation sites in royal jelly, BMC Genomics. 15: 135.

Zhang, Y., Z. Dong, D. Wang, Y. Wu, Q. Song, P. Gu, P. Zhao, and Q. Xia. 2014b. Proteomics of larval hemolymph in Bombyx mori reveals various nutrient-storage and immunity-related proteins. Amino Acids. 46: 1021–1031.

Zhou, P., Z. Dong, J. Duan, G. Wang, L. Wang, Y. Li, Z. Xiang, and Q. Xia. 2012. Genome-wide identification and immune response analysis of serine protease inhibitor genes in the silkworm, Bombyx mori. PLoS One. 7: e31168.

Zhou, Q., and H. Qiu. 2019. The mechanistic impact of N-glycosylation on stability, pharmacokinetics, and immunogenicity of therapeutic proteins. J. Pharm. Sci. 108: 1366–1377.