Amino Acids (2010) 39:751–761
DOI 10.1007/s00726-010-0517-3

ORIGINAL ARTICLE

Shotgun strategy-based proteome profiling analysis on the head of silkworm *Bombyx mori*

Jianying Li · S. Hossein Hosseini Moghaddam · Xiang Chen · Ming Chen · Boxiong Zhong

Received: 13 August 2009 / Accepted: 5 February 2010 / Published online: 3 March 2010

Abstract Insect head is comprised of important sensory systems to communicate with internal and external environment and endocrine organs such as brain and corpus allatum to regulate insect growth and development. To comprehensively understand how all these components act and interact within the head, it is necessary to investigate their molecular basis at protein level. Here, the spectra of peptides digested from silkworm larval heads were obtained from liquid chromatography tandem mass spectrometry (LC–MS/MS) and were analyzed by bioinformatics methods. Totally, 539 proteins with a low false discovery rate (FDR) were identified by searching against an in-house database with SEQUEST and X!Tandem algorithms followed by trans-proteomic pipeline (TPP) validation. Forty-three proteins had the theoretical isoelectric point (p*I*) greater than 10 which were too difficult to separate by two-dimensional gel electrophoresis (2-DE). Four chemosensory proteins, one odorant-binding protein, two diapause-related proteins, and a lot of cuticle proteins, interestingly including pupal cuticle proteins were identified. The proteins involved in nervous system development, stress response, apoptosis and so forth were related to the physiological status of head. Pathway analysis revealed that many proteins were highly homologous with the human proteins which involved in human neurodegenerative disease pathways, probably implying a symptom of the forthcoming metamorphosis of silkworm. These data and the analysis methods were expected to be of benefit to the proteomics research of silkworm and other insects.

Keywords *Bombyx mori* · Insect head · Proteomics · LTQ-Orbitrap · Gene Ontology · Pathway

Introduction

Insects have been in existence for at least 400 million years and constitute nearly 80% of species on our planet, providing us a large amount of desirable material to insight into the mechanisms of biological evolution and diversity (Grimaldi and Engel 2005). In the timeless evolutionary processes, they acquired marvelous adaptation with environmental changes. High genetic diversity, complex physiological systems, special effects on plant and animal survival and production, and so forth provide us reasonable field to do precious experiments on molecular bases of physiological mechanisms. The head of insect, including the olfactory, visual, and gustatory organs, plays the most important role in receiving environmental stimuli, and processing this information in brain and other nervous organs. Moreover, it comprises important endocrine organs, such as the brain, suboesophageal ganglion, corpus allatum, and corpus cardiacum, which are crucial for most of insect physiological processes such as growth, reproduction, diapause, and metamorphosis.

Electronic supplementary material The online version of this article (doi:10.1007/s00726-010-0517-3) contains supplementary material, which is available to authorized users.

J. Li · S. H. Hosseini Moghaddam · B. Zhong (✉)
College of Animal Sciences, Zhejiang University, Hangzhou 310029, People's Republic of China
e-mail: bxzhong@zju.edu.cn

S. H. Hosseini Moghaddam
Agriculture Faculty, University of Guilan, Rasht, Iran

X. Chen · M. Chen
College of Life Sciences, Zhejiang University, Hangzhou 310058, People's Republic of China

It is well known that the growth and developmental processes of insect are mainly regulated by the homeostasis of juvenile hormone and ecdysteroids which are biosynthesized and secreted from corpus allatum and prothoracic glands, respectively. The secretion of prothoracicotropic hormone (PTTH) from the brain–corpora cardiaca–corpora allata complex early in the fifth instar of silkworm triggers a cascade of endocrine events, resulting in the initiation of larval–pupal differentiation (Gu and Chow 2005). Moreover, it also regulates the biosynthesis of diapause hormone in suboesophageal ganglion of female pupa which is a primary endocrine factor to elicit embryonic diapause in her progeny (Wei et al. 2005; Yamashita 1996). Furthermore, environmental conditions, mostly photoperiod and temperature, affect on the endocrine system. For example, the melatonin, whose synthesis is mainly controlled by arylalkylamine *N*-acetyltransferase in vertebrates, was found in the silkworm head and was presumed to function as a neurochemical mediator of photoperiodic control of developmental events such as molting, eclosion, and diapause (Itoh et al. 1995).

In many insects, olfactory sensilla exists in the two pairs of olfactory organs of head, the antennae, and the maxillary palps. For instance, the head of silkworm contains numerous sensory cells in antenna and mandible which are crucial for searching and taking food by perceiving the smell and taste stimuli. The larval antenna of silkworm has 13 sensilla and about 52 sensory neurons in its distal portion. At the prepupal stage, cell proliferation was ceased and the antennal imaginal disc everted to form a large pupal antenna. The epidermis of the larval antenna forms the distal portion of the pupal antenna (Waku 1991).

The gene expression profile of silkworm head has been characterized based on microarray analysis (Xia et al. 2007). However, there is no strongly strict linear relationship between genes and their translational products, implying that the transcriptome analysis is insufficient for description of biological systems (Pandey and Mann 2000). On the other hand, proteomics focuses on the gene products, which are the active agents in cells. For a long time, two-dimensional gel electrophoresis (2-DE) combined with mass spectrometry (MS) has been frequently used in insect proteomic research (Zhou et al. 2008; Hosseini Moghaddam et al. 2008). But the liquid chromatography tandem mass spectrometry (LC–MS/MS) based shotgun strategy becomes relevant approach in proteomics for large-scale characterization of proteome profile. We have identified thousands of proteins using this method (Li et al. 2009a, b). Although the shotgun strategy has been widely used in human disease research (Adachi et al. 2006; de Souza et al. 2006), its applications on insect studies are still limited up to date (Heller et al. 2005; Kalume et al. 2005a; Li et al. 2009a, b). For shotgun approaches, there are mainly two prefractionation strategies for proteins prior to their proteolysis, chromatography, and MS/MS analyses to reduce complexity and increase the yield of protein identifications. The SDS-PAGE and multidimensional chromatographic approaches have been broadly applied, while they both have their own strengths and weaknesses. For example, the SDS-PAGE followed by in-gel digestion is more efficient than digestion in-solution (Adachi et al. 2006) despite that it has some defects in separation of very low- and high-molecular-weight (M_w) proteins.

The rapid development of proteomics and bioinformatics tools, especially the emergence of high-performance mass spectrometer such as the LTQ-Orbitrap, greatly promotes the credibility of results. In the present study, we analyzed the proteome of silkworm head at the fifth day of the last instar (3 days before the prepupal period) utilizing the hybrid LTQ-Orbitrap mass spectrometer. The obtained raw data sets were searched with TurboSEQUEST and X!Tandem against an in-house database followed by further validation using trans-proteomic pipeline (TPP) (Keller et al. 2005). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out to investigate the internal relationships of the identified proteins. We expect the strategy and data will improve our understanding of the molecular basis of biological regulation in the head, and contribute to the genome annotation of silkworm and, other insects as well.

Materials and methods

Sample preparation and electrophoresis

Polyvoltine silkworm strain P50 was reared on the fresh mulberry leaves under a 12-h-light/12-h-dark photoperiod at 26 ± 1°C with 70–85% relative humidity. The heads and brains were collected on the fifth day of the last instar. The sample preparation and 1D SDS-PAGE separation were carried out according to our described methods (Li et al. 2009a, b). Protein concentration of the sample was determined by 2-D Quant Kit (Amersham Biosciences, Piscataway, NJ, USA) according to the product specification. The electrophorized gel was stained with Coomassie Brilliant Blue R250 (CBB-R250, Sigma, St. Louis, MO, USA). The 2-DE was performed as we described (Zhou et al. 2008). The visualized protein spots were analyzed with ImageMaster 2D Platinum 6.0 software.

LC–MS/MS analysis

The gel lane of 1D SDS-PAGE was cut into 12 bands according to the deepness of Coomassie staining (Fig. 1). Each band was diced into small pieces ($\sim$1 mm^2). They were then subjected to in-gel tryptic digestion processes

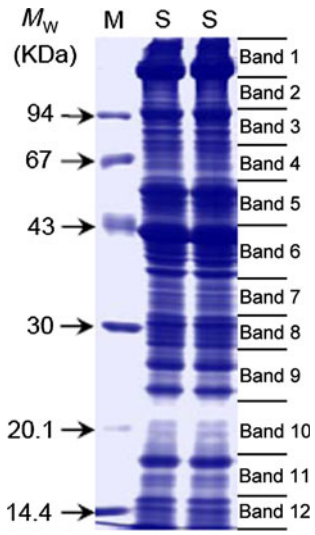

Fig. 1 One-dimensional SDS-PAGE pattern of the silkworm head proteins. *M* marker; *S* sample

(Shevchenko et al. 2006; Li et al. 2009b). The evaporated extracts of digested peptides were resuspended with 0.1% methanoic acid (Sigma) for LC–MS/MS analysis. The digested peptide mixtures were separated by the Ettan MDLC nanoflow/capillary LC system (GE Healthcare, Pittsburgh, PA, USA) equipped with a trapping column [PepMap C18, 300-μm i.d. × 5 mm, 3 μm, 100 Å (P/N 160454), Sunnyvale, CA, USA], and a nanocolumn [PepMap C18, 75-μm i.d. × 15 cm, 3 μm, 100 Å (P/N 160321), Sunnyvale, CA]. A constant flow rate of 300 nl/min was used for the nanocolumn with a 50-min solvent B (84% acetonitrile, 0.1% methanoic acid in water) gradient from 5 to 60%, and then from 60 to 95% in 10 min. The column was equilibrated with solvent A (0.1% formic acid in water) for 10 min before the next loading.

The separated peptides were detected using LTQ-Orbitrap (Thermo Finnigan, Bremen, Germany) with a nanospray configuration. The mass spectrometer was operated in the data-dependent mode with Xcalibur software version 2.0 (Thermo Electron, San Jose, CA, USA). Collision-induced dissociation (CID) was conducted with an isolation width of 2 Da, normalized collision energy of 35% for MS/MS, activation q of 0.25 and an activation time of 30 ms. Automated gain control (AGC) target ion values were set on default mode for MS, and MS/MS analyses. The five most intense ions were isolated and fragmented in linear ion trap with the dynamic exclusion settings: repeat count 2, repeat duration 30 s, exclusion duration 180 s. The precursor ion scan MS spectra (m/z 300–1,600) were acquired in the Orbitrap with the resolution $R = 60{,}000$ at m/z 400 with the number of accumulated ions being 1×10^6. The resulting fragment ions were recorded in the Orbitrap with the resolution $R = 15{,}000$ at m/z 400. Mass accuracy was determined by an external calibration 1 day prior the performing date.

Database search and bioinformatics analysis

The queried spectra were searched against the in-house database we constructed before (Li et al. 2009b). The data set consists of 43,557 proteins from predicted silkworm proteome database (PreD), NCBInr databases of *Bombyx* (BmD) and *Drosophila* (DmD). The searching was carried out with the TurboSEQUEST (Bioworks version 3.2, Thermo Electron) and X!Tandem, separately. The parameters for X!Tandem search were the same with that for SEQUEST. The monoisotopic mass tolerances of precursor ions and fragmentation ions were set to 1.5 and 1.0 Da, respectively. Trypsin enzyme was used for digestion, partially cleaved at either end of protein sequence, and two miscleavage sites were allowed. Static (carbamidomethyl) modification on cysteine (+57.0215 Da), and variable modifications (oxidation) on methionine (+15.9949 Da) were set for all searches. The target-decoy database searching (Elias and Gygi 2007) was carried out under the same parameters to evaluate the false discovery rate (FDR) of the results. Figure 2 shows the workflow of the identification processes. The identified peptides from database (original and target-decoy database) search were subjected to TPP (v4.0 JETSTREAM rev 2) for further validation. The final results protXML files from SEQUEST and X!Tandem were compared and integrated with the Computational Proteomics Analysis System (CPAS, LabKey Server 8.2) (Rauch et al. 2006). The results were exported in an .xls file.

The sequences of identified proteins from PreD were subjected to BLASTP querying against the UniProt Knowledgebase (Swiss-Prot + TrEMBL) and the new version of predicted silkworm proteome sequences (http://silkworm.swu.edu.cn/silkdb/) (The International Silkworm Genome Consortium 2008). The signatures of all identified proteins were queried against InterPro member databases by InterProScan searching (http://www.ebi.ac.uk/InterProScan/) (Mulder and Apweiler 2007; Hunter et al. 2009). The matched terms were analyzed with the WEGO [Web Gene Ontology Annotation Plot (http://wego.genomics.org.cn/cgi-bin/wego/)] as described by Ye et al. (2006). The Enzyme Commission (EC) numbers (if available) of the identified proteins were obtained by searching against the KEGG GENES (http://blast.genome.jp/). The EC numbers of matched proteins ($E \leq e$-15) were searched against KEGG reference pathway database (http://www.genome.jp/kegg/tool/search_pathway.html).

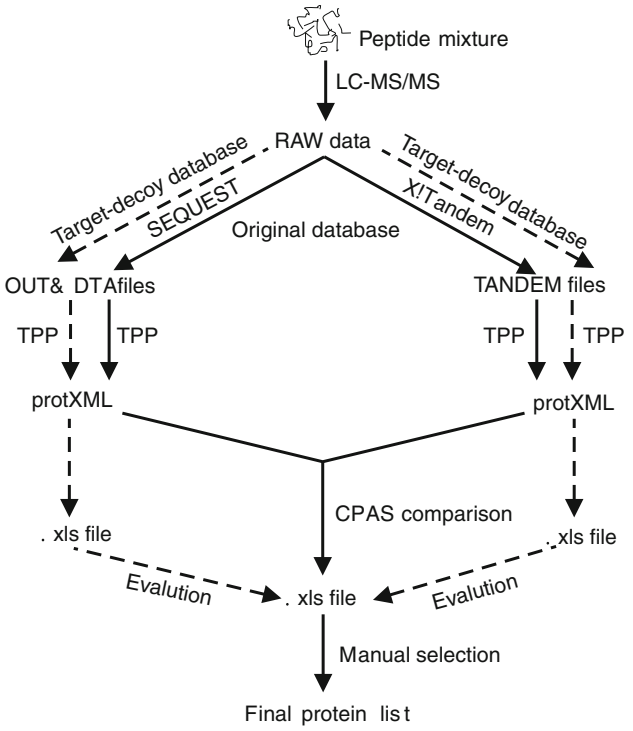

Fig. 2 Workflow of identification and evaluation of the data generated from LC–MS/MS. The raw data were subjected to searching against the target-decoy database, and original database using SEQUEST and X!Tandem under the same parameters, respectively. The resulting data were further validated by TPP. The results of original database with the two engines were compared with CPAS and then exported in EXCEL. The FDR of the results were evaluated with the data from target-decoy database. The integrated data were manually selected to decrease the redundancy

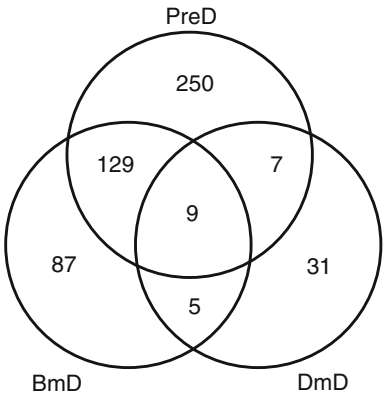

Fig. 3 Comparison of the number of identified protein entries from the three parts of database. The overlap shows the entries with the proteins from two or three sources: *PreD* predicted proteome from silkworm genome sequences; *BmD* the protein database of *Bombyx* which includes *B. mori* and *B. mandarina*; *DmD* the protein database of *D. melanogaster*. The protein sequences in BmD and DmD were all from NCBInr

Results

Post-processing of the raw data from LC–MS/MS

The 12 raw data sets from LC–MS/MS were subjected to database search using SEQUEST and X!Tandem. To evaluate the certainty of the results under different filter parameters, we computed the FDR of the identifications through target-decoy database searching followed by TPP validation with three thresholds of 0.75 (P1), 0.8 (P2) and 0.9 (P3) for the peptide probability. The thresholds for protein probability were all set at 0.9. In fact, the results of using the three criteria were different. In case that the probability of peptide increased from 0.75 to 0.9, the FDR of peptides decreased about 1.5% while there was no obvious change on FDR of proteins. With the aim of better identification with reasonable FDR, we selected the parameter P2 for original database searching and analysis that followed.

After searching against the original database and the following validation processes, totally 498 protein entries (groups) were identified (Fig. 3). Among those, 250 (50.2%), 67 (13.5%) and 31 (6.2%) entries were specifically from PreD, BmD and DmD, respectively. Due to the similarity of proteins in different database, those proteins with a part of homologous sequences were mostly assigned into the same entry by ProteinProphet. The common entries of PreD and BmD (129 entries) were far more than that of PreD and DmD (7 entries) or BmD and DmD (5 entries). For the proteins from different sources in the same entry, we could not unambiguously determine which protein or proteins were real identifications. Therefore, they were manually selected according to the higher priority of BmD, DmD, and PreD, and totally 145 and 15 proteins were deleted from PreD and DmD lists, respectively.

Using SEQUEST and X!Tandem followed by TPP validation and CPAS comparison and integration, we identified 2,235 unique peptides (supplementary Table S1), and 539 proteins with 508 overlaps between SEQUEST and X!Tandem (supplementary Table S2). By this method, about 3% (18/521, 13/526) more proteins were identified compared with using SEQUEST or X!Tandem exclusively. Furthermore, about 50% of the identified proteins from the two algorithms had the peptide sequence coverage more than 10% (Fig. 4). Among the identified proteins, 84% (454/539) were theoretically distributed in the range of p*I* 3–10, M_w 5–110 KDa. Moreover, there were 43 proteins with the p*I* more than 10 which were substantially difficult to separate by 2-DE.

Proteome profiles of silkworm head compared with brain

Compared with the proteome profile of silkworm brain described in our previous study (Li et al. 2009a), 369 proteins were common such as 14-3-3 zeta, antennal

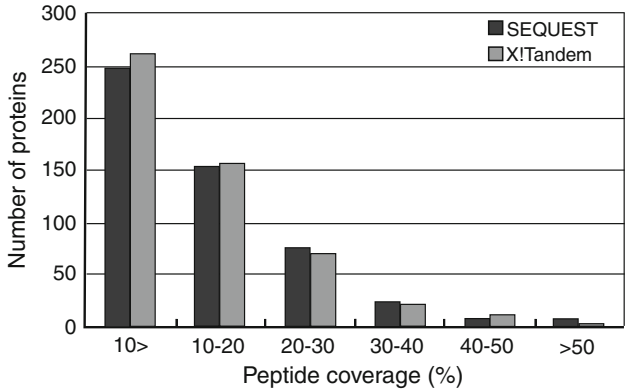

Fig. 4 Sequence coverage distribution of the identified proteins resulted from SEQUEST and X!Tandem searches. About 50% of proteins were identified with the peptide sequence coverage more than 10%

binding protein, cytochrome P450 CYP4G25, and MLE protein. Many proteins were specific for head such as most of muscle and cuticle associated proteins, as well as some proteins functionally important, such as heat shock protein 1, glutathione *S*-transferase 10, sex-specific storage-protein SP1, imaginal disk growth factor, and translationally controlled tumor protein. To provide an intuitive view on the differential expression profiles, the 2-DE proteome patterns of silkworm head and brain were presented (Fig. S1) and were analyzed with the ImageMaster 2D software. Approximately 67% protein spots of head were automatically matched with that of brain. However, 21 matched spots of head with large volumes had higher average intensity ratio, ten times more than brain.

High expression proteins

There were a number of proteins associated with muscle and cuticle (Table 1, supplementary Table S2). All the muscle-associated structural proteins including actins, myosin, tubulins, paramyosin, tropomyosin, troponin, myofilin, and Bm kettin were identified. Moreover, most of them were identified with a lot of unique peptides indicating their abundant expression. Insect cuticle is composed of proteins and chitin. Many insect cuticular proteins have a 35–36 amino acid motif known as the Rebers and Riddiford Consensus (R&R Consensus) (Rebers and Riddiford 1988). The extensive conservation of this region led to the suggestion that it functions to bind chitin. In the present study, we found nine cuticle proteins with annotation, such as larval cuticle protein 1, larval cuticle proteins LCP-17, -22, -30, and pupal cuticle protein (PCP). In the results of BLASTP against the UniProt Knowledgebase using PreD sequences, there were ten cuticle proteins. However, these proteins were assigned into different protein entries showing the big difference of their sequences.

Chemoreception proteins

Insect chemoreception systems play important roles in detecting environmental stimuli and mainly distributed in the head. In the present study, we identified several sensorium-associated proteins, including four chemosensory proteins (CSPs) and an antennal binding protein [odorant-binding protein (OBP)]. Some proteins from PreD were also involved by GO functional category analysis according to their sequences. For example, the Bmb003671 and Bmb012807 were related to olfactory behavior and sensory perception of chemical stimulus, respectively.

Apoptosis related proteins

Apoptosis is the process of programmed cell death (PCD) involving a series of biochemical events leading to cell morphology and death. Many tissues undergo PCD during insect metamorphosis. Especially atrophy of many tissues, such as the anterior silk glands (ASGs), happens at the late stage of silkworm larva. The cell death-regulatory protein GRIM19, translationally controlled tumor protein, CathD, and so on were identified in this study. However, since the apoptosis is normal cellular process, introducing which protein(s) trigger the tissue degradation needs further investigation.

Miscellaneous proteins

Except for the proteins mentioned above, there were a lot of proteins related to nervous system development, stress response, and so forth (Table 1). Apart from the known 14-3-3 protein family which is related to nervous development, many proteins from PreD may be involved in dendrite development, dendrite morphogenesis, neurotransmitter secretion, and axon guidance according to their GO annotations. The insect possesses complex stress response systems for self-protection against the environmental stresses and maintaining their normal development. The proteins involved in response to various stressing agents such as heat shock, inflammatory, oxidant, chemical, and osmotic pressure protect the cells from destabilizing. We also identified several other proteins belonging to this category, such as the time interval measuring enzyme-esterase A4 (TIME-EA4) which showed response to metal ion, and the glutathione *S*-transferase sigma responding to oxidative stress.

Functional annotation and classification

Gene Ontology signatures of 502 out of the 539 identified proteins were available. They were classified into cellular component, molecular function, and biological process according to the GO hierarchy using WEGO (Fig. 5). In the

Table 1 Partial list of the identified proteins in the head of silkworm larva

Protein accessions	Protein description	UniProt accession[a]	Unique peptide hits[b]	Theor. p*I*/M_w (KDa)	Biological process[c]	Molecular function[c]
Nervous system development associated proteins						
Bmb007857		B7XFU6	5	5.38/173.19	Neurotransmitter secretion	
Bmb019518		Q00801	5	6.97/33.70	Neurotransmitter secretion	SNAP receptor activity
Bmb019690		B3MYW1	7	4.92/155.05	Asymmetric neuroblast division	
gil114050901	14-3-3 zeta	Q2F637	4	4.9/28.17	Activation of tryptophan 5-monooxygenase activity	Diacylglycerol-activated phospholipid-dependent protein kinase C inhibitor activity
gil148298752[d]	14-3-3 epsilon protein	Q1HPT4	3	4.66/29.67	DNA damage checkpoint	Diacylglycerol-activated phospholipid-dependent protein kinase C inhibitor activity
gil153792023	MLE protein	A5JPM0	2	6.16/144.66	Axon extension	ATP-dependent helicase activity
Sensorium-associated proteins						
Bmb003671			3	8.91/22.25	Olfactory behavior	
Bmb012807		B8ZWK6	9	4.99/33.21	Sensory perception of chemical stimulus	Odorant binding
gil112983042	Chemosensory protein 9	Q3LB98	3	6.29/14.73	Metamorphosis	
gil112983052	Chemosensory protein 11	Q3LB95	1	4.97/13.52	Determination of anterior/posterior axis, embryo	Diacylglycerol binding
gil112983058	Chemosensory protein 5	Q3LBA2	2	6.73/14.65	Metamorphosis	
gil112983094	Chemosensory protein CSP1	Q8MMK7	3	5.21/14.55	Metamorphosis	
gil164448664	Antennal binding protein	Q2F5L4	1	6.71/15.49	Sensory perception of chemical stimulus	Odorant binding
Musculature proteins						
gil114052470	Muscular protein 20	Q1HPM4	4	8.7/20.13	Cell adhesion	Actin binding
gil148298826[d]	Myosin light chain 2	B5AK67	9	4.67/22.04		
gil168823429[d]	Bm kettin	Q8T103	33	5.93/544.51	Locomotion	Actin binding
gil169234928[d]	Titin1	Q8T102	2	5.06/369.03	Locomotion	Actin binding
gil195963325[d]	Paramyosin	B3VTP0	37	5.43/102.75	Mesoderm development	Motor activity
Cuticle proteins						
gil112983302	Larval cuticle protein 1	Q1XGY6	3	4.43/12.29		Structural constituent of chitin-based cuticle
gil112983982[d]	LCP18	O96052	2	4.38/11.02		Structural constituent of chitin-based cuticle
gil112984038[d]	LCP22	C0H6N0	4	8.59/18.85		Structural constituent of chitin-based cuticle
gil112984040[a]	LCP17	O02387	3	5.06/15.27		Structural constituent of chitin-based cuticle
gil187423907[d]	Pupal cuticle protein	C0H6I8	3	5.95/26.32		Structural constituent of chitin-based cuticle
Stress response						
Bmb037587[d]				4.72/6.76		Chemokine activity
gil112983667[d]	Prophenoloxidase subunit 1	Q27451	3	6.25/78.79	Defense response	Monophenol monooxygenase activity
gil112983802	Mn superoxide dismutase	Q65Y02	2	8.84/24.23	Age-dependent response to oxidative stress	

Table 1 continued

Protein accessions	Protein description	UniProt accession[a]	Unique peptide hits[b]	Theor. p*I*/M_w (KDa)	Biological process[c]	Molecular function[c]
gil116175238	Time interval measuring enzyme-esterase A4	Q08J22	1	6.12/18.25	Response to metal ion	Superoxide dismutase activity
gil160333678[d]	Glutathione *S*-transferase sigma	Q5CCJ4	4	5.85/23.34	Response to oxidative stress	Glutathione peroxidase activity
gil163838668[d]	Prophenoloxidase-2s	Q9BLG6	4	5.67/80.18	Defense response	Monophenol monooxygenase activity
Apoptosis						
gil112982880[d]	Translationally controlled tumor protein	A2TK65	3	4.66/19.86	Positive regulation of cell size	Guanyl-nucleotide exchange factor activity
gil112983370	Transport protein Sec61 alpha subunit	Q19AA9	2	8.5/51.93	Cell death	Protein transporter activity
gil112983576	CathD	Q4U5S3	1	6.25/41.62	Autophagic cell death	Aspartic-type endopeptidase activity
gil112983782[d]	Cyclophilin A	Q56C53	4	7.74/17.93	Autophagic cell death	Cyclin-dependent protein kinase regulator activity
gil114051728	Cell death-regulatory protein GRIM19	Q2F619	2	9.26/17.99	Mitochondrial electron transport, NADH to ubiquinone	NADH dehydrogenase activity
Protein binding						
Bmb026245		Q967V9	6	5.68/45.75	Negative regulation of proteolysis	Hormone binding
gil112983280	Hsc70/Hsp90-organizing protein HOP	Q5CCL7	1	6.14/62.14		Protein binding
gil112983414	Heat shock protein hsp21.4	Q5R1P5	3	5.79/21.40		Protein binding
gil112983556	90-kDa heat shock protein	Q9BLC5	6	4.99/82.42	Positive regulation of nitric oxide biosynthetic process	Nitric-oxide synthase regulator activity
gil120444903	Chaperonin	A1YM11	3	5.4/59.17	Protein folding	ATPase activity, coupled
gil148298768[d]	Heat shock protein 1	Q1HPN2	2	6.33/22.56	Embryonic development	
Protein metabolism						
Bmb019850			1	9/32.12		Estradiol 17-beta-dehydrogenase activity
gil112982735	Ribosomal protein P0	Q5UAU1	2	5.69/34.19	DNA repair	DNA-(apurinic or apyrimidinic site) lyase activity
gil118918433	Cytochrome *c* oxidase polypeptide IV	Q1HQ98	4	9.2/20.51	Mitochondrial electron transport, cytochrome *c* to oxygen	Cytochrome-*c* oxidase activity
gil156119320[d]	Low molecular lipoprotein 30 K precursor	P09335	9	6.83/30.03	Autophagy	Phosphoinositide binding
gil163838680[d]	Cytochrome P450 CYP4G25	Q1HPS6	3	8.74/63.29	Steroid biosynthetic process	Electron carrier activity
gil169234936[a]	Sex-specific storage-protein SP1	P22700	14	6.78/87.26		Nutrient reservoir activity

[a] The UniProt accessions were obtained by BLASTP against the UniProt Knowledgebase with the protein sequences in our database

[b] Average peptide number for those proteins identified by both SEQUEST and X!Tandem

[c] only two GO categories have showed

[d] The specific proteins identified in the silkworm head which are not among the proteins identified in the brain

subcategory of cell part, there were 44 important membrane proteins for signal conduction. However, no protein was found in synapse of head. For the molecular function, majority of the annotated proteins possessed the binding (197) and catalytic (186) activities. Furthermore, there were three proteins showing odorant binding. As for the

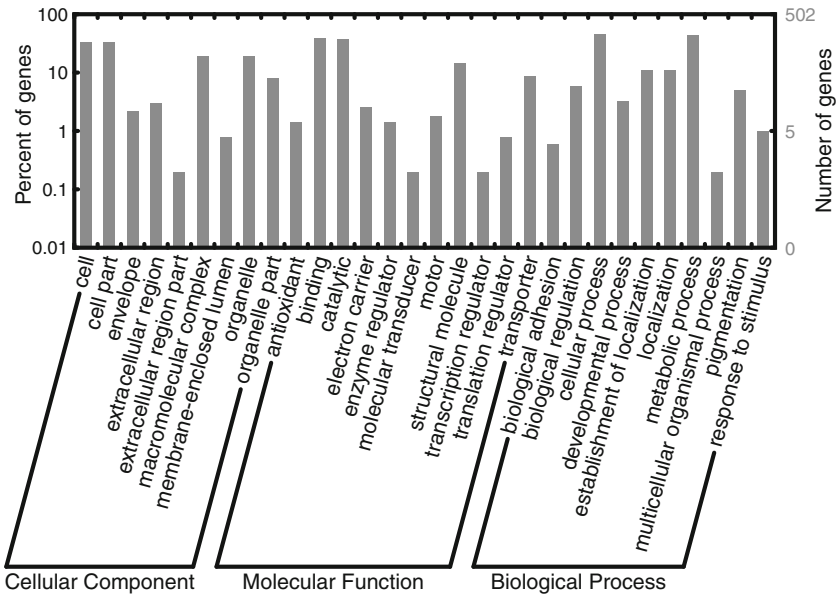

Fig. 5 GO categories of the identified proteins in the silkworm head. The identified proteins were classified into cellular component, molecular function, and biological process by WEGO according to their GO signatures. The number of genes denotes that of proteins with GO annotations

developmental process (16 proteins), there were 15 proteins taking part in the anatomical structure organization, development, and morphogenesis.

Among the identified proteins, 272 with EC annotations were involved in 73 KEGG pathways. More than half of these pathways were related to the metabolism, of which carbohydrate and amino acid metabolisms were the most active. Besides, there were five pathways related to neurodegenerative diseases. The pathways containing the most ECs were related to human neural diseases, metabolism, oxidative phosphorylation, insulin signaling, etc.

Discussion

Shotgun proteomics based on LC–MS/MS provides a highly sensitive, high-throughput method for large scale characterization of the protein components in cells, tissues and organs. However, there are still some limitations that hamper the application of this method. One of the challenges is the quality evaluation of the results (Elias and Gygi 2007). The FDR evaluation of peptides and proteins is a recommended and widely used approach (Elias and Gygi 2007). In the present study, we could get considerable proteins with FDR of 0.92% under parameter P2. The low FDR showed the qualified results and bioinformatics analysis.

Since particular species database has not been fully developed, database search remains the bottleneck for many shotgun proteomic experiments. Although thousands of proteins have been identified in human tissues (Adachi et al. 2006; de Souza et al. 2006), the identifications in insects are much less (Heller et al. 2005; Kalume et al. 2005a). In this study, the composite database considerably improved the number of identifications. Except of proteins from PreD, there were 103 entries from BmD and DmD implying that the predicted proteome from genome sequences could not cover all the translational products, and perhaps some of the genome sequences or their translations were incorrect. By using the proteomic data, it is possible to correct or supplement the gene sequences and screen new genes for genome annotation (Kalume et al. 2005b). Furthermore, using two searching algorithms SEQUEST and X!Tandem with different scoring methods increased the number and confidence of the identifications.

Taraszka et al. (2005) identified 780 proteins with less than 2,000 unique peptides from adult head of *Drosohpila*. Here, we totally identified 539 proteins with 2,235 unique peptides by using the two algorithms. These results showed that although the proteins in the silkworm database were far less than that of fruit fly, employed strategy could get considerable identifications. To decrease redundancy of the results, the proteins in the same entry were manually selected. There were 465 (86.3%) single protein hits in the final protein list and 74 proteins were included in 30 entries, which ensured the low redundancy of our results.

The identified proteins in the head were much less than that in brain (Li et al. 2009a). Some proteins such as the prothoracicotropic hormone, corazonin preprohormone, and arylalkylamine *N*-acetyltransferase were only found in the brain. It was likely that the highly expressed musculature and cuticle proteins in the head greatly decreased the chance for detection of low abundant proteins. However, we increased the chance of protein identification by slicing of gel lane into 12 pieces which were twice of the previous

experiment protocol (Li et al. 2009a). It also suggested that the less complex of sample, for example, a single tissue or organ, will be apt to detect proteins with lower expression.

Microarray-based gene expression profile of silkworm head was characterized and 78 highly expressed genes were found which might be head-specific genes (Xia et al. 2007). We compared these genes with our identifications and found only Bmb015918 (putative cuticular protein) and Bmb027605 (mitochondrial solute carrier) were common. Without considering the different developmental stage, it may be due to the discrepancy between gene expression level and protein level. Although they claimed that the OBP and CSP were also included by Blastx searching, there was no match within silkworm database. In the present study, we identified four CSPs and one OBP from BmD. Moreover, they had been identified in brain, suboesophageal ganglion, and prothoracic glands (Li et al. 2009a). These proteins are important members of insect chemoreception systems. They can detect wide range of soluble chemicals which are important for finding and assessing desirable food sources, in addition to recognizing mates and oviposition sites (Dahanukar et al. 2005). CSPs are a class of small soluble proteins expressed broadly in tissues including the antennae, head, epithelium, testes, and ovaries and so on (Gong et al. 2007). Another special chemosensory protein family is the OBP which shares none sequence similarity with CSPs. The OBPs and CSPs are supposed to be involved in carrying the chemical messages from the environment to the chemosensory receptors. These chemosensory proteins widely spread in endocrine system and may be involved in the endocrine changes. Olfactory and chemosensory neurons project their dendrites to a large lymphatic cavity where soluble binding proteins were present at high concentrations (Gong et al. 2007). Foret and Maleszka (2006) found the expression of *obp13* at high level in late larval and throughout pupal stages of honey bee. The OBP was highly expressed in the silkworm head which seems to be consistent with the previous reports. Moreover, GO analysis showed that three out of the four CSPs were involved in the process of insect metamorphosis (Table 1).

There were a lot of proteins related to musculature such as the myosin and actin among the identified proteins (Supplemental Table 1). Assuming that the more detected unique peptide reveals the higher expression level, five proteins including myosin (61 hits), actin A1 (44 hits), paramyosin (37 hits), Bm kettin (28 hits), and tropomyosin (10 hits) were the most high-abundance proteins. Actually, there was a large amount of musculature widely spread in labium, mandible, antennae, and head cavity. These high-abundance proteins reduced the chance of detection for the proteins in low expression and inevitably decreased the number of total identifications. In addition, interestingly we found nine known cuticle proteins, including the PCP. This protein accumulates in the newly synthesized pupal cuticle of silkworm. The biosynthesis of PCP is regulated at transcriptional level during post-embryonic development and the expression of the PCP gene may be triggered by the molting hormone 20-hydroxyecdysone (20E) (Nakato et al. 1992). These results implied that PCP was synthesized in the silkworm at least 3 days before prepupal stage.

A new cytochrome P450 gene, *CYP4G25*, was reported to be associated with diapause of first instar larvae of the wild silkmoth *Antheraea yamamai* (Yang et al. 2008). However, we detected this protein at the late stage of silkworm larva suggesting its possible role in insect development. We also found another diapause-related protein, TIME-EA4 which is a 156-residue protein belongs to the copper–zinc superoxide dismutase (SOD) family with a high degree of sequence homology and SOD activity (Isobe et al. 2006). However, unlike most members of this family, EA4 is a naturally glycosylated protein with a sugar chain on Asn22. Some researchers supposed that it has a specific role for measurement of the duration of cold exposure required for diapause development (Isobe et al. 2006; Kai et al. 1995; Tani et al. 2001).

In addition, we found two members of the 14-3-3 protein family. The 14-3-3 proteins are small, highly conserved acidic molecules abundant in the central nervous system (CNS) and were initially described as activators of neurotransmitter synthesis (Boston et al. 1982; Berg et al. 2003). Two isoforms 14-3-3 zeta and epsilon proteins were expressed in the silkworm head, which was consistent with their important functions in neurons. Moreover, the fifth day of the last instar of silkworm (the time of tissue sampling) is just 3 days ahead of the prepupal stage when many tissues begin to degenerate. Because of the important role of 14-3-3 proteins in the processes of apoptosis and neurodegeneration of mammal (Berg et al. 2003), they were likely to take part in degeneration of nervous system of silkworm.

Nano-LC/MS/MS technology in the present study identified a large number of important proteins besides those mentioned above. Translationally controlled tumor protein (TCTP) is a highly conserved protein upregulated in various tumours. It plays a role in regulating Rheb (Ras homologue enriched in brain), a protein controlling growth and differentiation. Reducing *Drosophila* TCTP (dTCTP) levels will reduce its cell size, cell number, and organ size (Hsu et al. 2007). Cell death-regulatory protein GRIM-19 (gene associated with retinoid-IFN-induced mortality 19) was originally identified as a nuclear protein with apoptotic nature in interferon (IFN)- and all-*trans*-retinoic acid (RA)-induced tumor cells. Huang et al. (2004) demonstrated that GRIM-19 is a functional component of mitochondrial NADH:ubiquinone oxidoreductase (complex I) and essential

for early embryonic development. Elimination of *GRIM-19* destroys the assembly and electron transfer activity of the complex I in the mitochondrial respiratory chain. The *maleless* (*mle*) gene is one of known regulatory loci required for increased transcription (dosage compensation) of X-linked genes in *Drosophila melanogaster* males, while the *mle* in females is prevented by *Sex-lethal* (*Sxl*) gene from binding to the two X chromosomes (Gorman et al. 1993). *Annexin IX* (*ANX IX*) has been identified as a 20E-inducible gene in silkworm ASGs, which undergo PCD at the late larval stage. In fact, the head includes a part of ASGs which are derived from mandible. Kaneko et al. (2006) demonstrated a possible involvement of ANX IX in the determination of PCD timing by delaying or suppressing the response to the increase of hemolymph ecdysteroid concentration during the prepupal period.

The availability of genome sequences of human and model organisms has provided a good opportunity to investigate the conservation of genes responsible for heritable diseases in humans (Bier 2005). Among our identified proteins, there were many proteins with EC annotations highly homologous with human proteins involved in the pathways of two most common neurodegenerative disorder diseases, including Alzheimer's disease (AD) and Parkinson's disease (PD) (supplementary Fig. S2). The known mutations *APP* and *SNCA* in the human can lead to AD and PD, respectively. Besides, the homologous gene mutations in flies have also been well studied for these diseases. For instance, the unique *D. melanogaster* gene that is homologous with human *PSEN* can be mutated to cause early onset of AD (Sherrington et al. 1995; Bier 2005). In the present results, the products of homologous genes of *APP*, *SNCA*, and *PSEN* were all detected. Due to broad spectrum of genes that are related to human disease genes, the *D. melanogaster* has been already proven to be a powerful model to analyze the gene function of disease processes such as developmental disorders, neurological diseases, and cancer (Bier 2005; Greenspan and Dierick 2004). Considering many genes were homologous with the genes related to human neurodegenerative disorder diseases, the silkworm could be potentially used as another model in human diseases research.

In conclusion, the shotgun LC–MS/MS is a powerful method for large-scale proteome profiling. However, the bioinformatics analysis is indispensable for accurate explanation of MS data and better understanding of the basic relationships of proteome components. The identified functionally important proteins were involved in perception and conduction of environmental stimuli, regulation of growth, and development, self-protection against stresses, etc., exhibiting the plentifulness of the complex systems in the silkworm head and also the molecular interaction in their biological pathways. Many proteins were highly homologous to human proteins which involved in the pathways of human neurodegenerative diseases. It might reflect the internal physiological status of the head to some extent due to their roles in forthcoming metamorphosis. Further research can be performed, for example, to discover how the environmental stimuli affect the CNS and lead to endocrine changes.

Acknowledgments This work was supported by the National Basic Research Program of China (Grant No. 2005CB121003), National Hi-Tech Research and Development Program of China (Grant No. 2006AA10A118), and Program for New Century Excellent Talents in University. We are thankful to the 985-Institute of Agrobiology and Environmental Sciences of Zhejiang University, for providing convenience for our experiment. We are also grateful to Mr. Wei Fan and Jisheng Li for kind help in our data analysis.

Conflict of interest statement The authors declare that they have no conflict of interest.

References

Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. Genome Biol 7:R80

Berg D, Holzmann C, Riess O (2003) 14-3-3 Proteins in the nervous system. Nat Rev Neurosci 4:752–762

Bier E (2005) *Drosophila*, the golden bug, emerges as a tool for human genetics. Nat Rev Genet 6:9–23

Boston PF, Jackson P, Thompson RJ (1982) Human 14-3-3 protein: radioimmunoassay, tissue distribution, and cerebrospinal fluid levels in patients with neurological disorders. J Neurochem 38:1475–1482

Dahanukar A, Hallem EA, Carlson JR (2005) Insect chemoreception. Curr Opin Neurobiol 15:423–430

de Souza GA, Godoy LM, Mann M (2006) Identification of 491 proteins in the tear fluid proteome reveals a large number of proteases and protease inhibitors. Genome Biol 7:R72

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4:207–214

Foret S, Maleszka R (2006) Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). Genome Res 16:1401–1413

Gong DP, Zhang HJ, Zhao P, Lin Y, Xia QY, Xiang ZH (2007) Identification and expression pattern of the chemosensory protein gene family in the silkworm, *Bombyx mori*. Insect Biochem Mol Biol 37:266–277

Gorman M, Kuroda MI, Baker BS (1993) Regulation of the sex-specific binding of the maleless dosage compensation protein to the male X chromosome in *Drosophila*. Cell 72:39–49

Greenspan RJ, Dierick HA (2004) 'Am not I a fly like thee?'—From genes in fruit flies to behavior in humans. Hum Mol Genet 13:R267–R273

Grimaldi DA, Engel MS (2005) Evolution of the insects. Cambridge University Press, London

Gu SH, Chow YS (2005) Analysis of ecdysteroidogenic activity of the prothoracic glands during the last larval instar of the silkworm, *Bombyx mori*. Arch Insect Biochem Physiol 58:17–26

Heller M, Ye M, Michel PE, Morier P, Stalder D, Jünger MA, Aebersold R, Reymond F, Rossier JS (2005) Added value for tandem mass spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. J Proteome Res 4:2273–2282

Hosseini Moghaddam SH, Du X, Li J, Cao JR, Zhong BX, Chen YY (2008) Proteome analysis on differentially expressed proteins of the fat body of two silkworm breeds, *Bombyx mori*, exposed to heat shock exposure. Biotechnol Bioprocess Eng 13:624–631

Hsu YC, Chern JJ, Cai Y, Liu M, Choi KW (2007) *Drosophila* TCTP is essential for growth and proliferation through regulation of dRheb GTPase. Nature 445:785–788

Huang G, Lu H, Hao A, Ng DC, Ponnia S, Guo K, Lufei C, Zeng Q, Cao X (2004) GRIM-19, a cell death regulatory protein, is essential for assembly and function of mitochondrial complex I. Mol Cell Biol 24:8447–8456

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2009) InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211–D215

International Silkworm Genome Consortium (2008) The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochem Mol Biol 38:1036–1045

Isobe M, Kai H, Kurahashi T, Suwan S, Pitchayawasin-Thapphasaraphong S, Franz T, Tani N, Higashi K, Nishida H (2006) The molecular mechanism of the termination of insect diapause: Part 1. A timer protein, TIME-EA4, in the diapause eggs of the silkworm *Bombyx mori* is a metallo-glycoprotein. Chembiochem 7:1590–1598

Itoh MT, Hattori A, Nomura T, Sumi Y, Suzuki T (1995) Melatonin and arylalkylamine *N*-acetyltransferase activity in the silkworm, *Bombyx mori*. Mol Cell Endocrinol 115:59–64

Kai H, Kotani Y, Miao Y, Azuma M (1995) Time interval measuring enzyme for resumption of embryonic development in the silkworm, *Bombyx mori*. J Insect Physiol 41:905–910

Kalume DE, Okulate M, Zhong J, Reddy R, Suresh S, Deshpande N, Kumar N, Pandey A (2005a) A proteomic analysis of salivary glands of female *Anopheles gambiae* mosquito. Proteomics 5:3765–3777

Kalume DE, Peri S, Reddy R, Zhong J, Okulate M, Kumar N, Pandey A (2005b) Genome annotation of *Anopheles gambiae* using mass spectrometry-derived data. BMC Genomics 6:128

Kaneko Y, Takaki K, Iwami M, Sakurai S (2006) Developmental profile of annexin IX and its possible role in programmed cell death of the *Bombyx mori* anterior silk gland. Zool Sci 23:533–542

Keller A, Eng J, Zhang N, Li XJ, Aebersold R (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol 1:2005.0017

Li JY, Chen X, Fan W, Hosseini Moghaddam SH, Chen M, Zhou ZH, Yang HJ, Chen JE, Zhong BX (2009a) Proteomic and bioinformatic analysis on endocrine organs of domesticated silkworm, *Bombyx mori* L. for a comprehensive understanding of their roles and relations. J Proteome Res 8:2620–2632

Li JY, Chen X, Hosseini Moghaddam SH, Chen M, Wei H, Zhong BX (2009b) Shotgun proteomics approach to characterizing the embryonic proteome of silkworm, *Bombyx mori*, at labrum appearance stage. Insect Mol Biol 18:649–660

Mulder N, Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol 396:59–70

Nakato H, Izumi S, Tomino S (1992) Structure and expression of gene coding for a pupal cuticle protein of *Bombyx mori*. Biochim Biophys Acta 1132:161–167

Pandey A, Mann M (2000) Proteomics to study genes and genomes. Nature 405:837–846

Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang R, Faca V, Gafken P, Zhang H, Whiteaker J, States D, Hanash S, Paulovich A, McIntosh MW (2006) Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. J Proteome Res 5:112–121

Rebers JF, Riddiford LM (1988) Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. J Mol Biol 203:411–423

Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, Tsuda T, Mar L, Foncin JF, Bruni AC, Montesi MP, Sorbi S, Rainero I, Pinessi L, Nee L, Chumakov I, Pollen D, Brookes A, Sanseau P, Polinsky RJ, Wasco W, Da Silva HAR, Haines JL, Pericak-Vance MA, Tanzi RE, Roses AD, Fraser PE, Rommens JM, St George-Hyslop PH (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375:754–760

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1:2856–2860

Tani N, Kamada G, Ochiai K, Isobe M, Suwan S, Kai H (2001) Carbohydrate moiety of time-interval measuring enzyme regulates time measurement through its interaction with time-holding peptide PIN. J Biochem 129:221–227

Taraszka JA, Kurulugama R, Sowell RA, Valentine SJ, Koeniger SL, Arnold RJ, Miller DF, Kaufman TC, Clemmer DE (2005) Mapping the proteome of *Drosophila melanogaster*: analysis of embryos and adult heads by LC-IMS-MS methods. J Proteome Res 4:1223–1237

Waku Y (1991) Developmental changes of the antenna and its neurons in the silkworm, *Bombyx mori*, with special regard to larval–pupal transformation. J Morphol 207:253–271

Wei ZJ, Zhang QR, Kang L, Xu WH, Denlinger DL (2005) Molecular characterization and expression of prothoracicotropic hormone during development and pupal diapause in the cotton bollworm, *Helicoverpa armigera*. J Insect Physiol 51:691–700

Xia Q, Cheng D, Duan J, Wang G, Cheng T, Zha X, Liu C, Zhao P, Dai F, Zhang Z, He N, Zhang L, Xiang Z (2007) Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*. Genome Biol 8:R162

Yamashita O (1996) Diapause hormone of the silkworm, *Bombyx mori*: structure, gene expression and function. J Insect Physiol 42:669–679

Yang P, Tanaka H, Kuwano E, Suzuki K (2008) A novel cytochrome P450 gene (CYP4G25) of the silkmoth *Antheraea yamamai*: Cloning and expression pattern in pharate first instar larvae in relation to diapause. J Insect Physiol 54:636–643

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34:W293–W297

Zhou ZH, Yang HJ, Chen M, Lou CF, Zhang YZ, Chen KP, Wang Y, Yu ML, Yu F, Li JY, Zhong BX (2008) Comparative proteomic analysis between the domesticated silkworm (*Bombyx mori*) reared on fresh mulberry leaves and on artificial diet. J Proteome Res 7:5103–5111