online data

Bioscience Reports

Identification of the proteome of the midgut of silkworm, *Bombyx mori* L., by multidimensional liquid chromatography (MDLC) LTQ-Orbitrap MS

Hui-Peng YAO, Xing-Wei XIANG, Lin CHEN, Ai-Qin GUO, Fang-Qing HE, Li-Pan LAN, Xing-Meng LU and Xiao-Feng WU[1]

College of Animal Science, Zhejiang University, Hangzhou 310029, People's Republic of China

Synopsis

The midgut is the digestive apparatus of the silkworm and its proteome was studied by using nano-LC (liquid chromatography) electrospray ionization MS/MS (tandem MS). MS data were analysed by using X!Tandem searching software using different parameters and validated by using the Poisson model. A total of 90 proteins were identified and 79 proteins were described for the first time. Among the new proteins, (i) 22 proteins were closely related to the digestive function of the midgut, including 11 proteins of digestive enzymes secreted by the epithelium, eight proteins of intestine wall muscle and mechanical digestion and three proteins of peritrophic membrane that could prevent the epithelium from being mechanically rubbed; (ii) 44 proteins were involved in metabolism of substance and energy; and (iii) 11 proteins were associated with signal transduction, substance transport and cell skeleton.

Key words: LTQ-Orbitrap, midgut, Poisson model, proteomics, silkworm (*Bombyx mori* L.), X!Tandem

www.bioscirep.org

INTRODUCTION

The mulberry silkworm, *Bombyx mori*, domesticated for over 5000 years from the wild progenitor *Bombyx mandarina* [1,2], is an important economical insect for silk production. As a typically representative organism of Lepidoptera, it is gradually being recognized as the model organism for insects, and would play an increasingly important role in entomology and molecular biology. Moreover, with the development of modern biotechnology, it has become an ideal bioreactor for producing recombinant proteins [3–5].

The midgut is the digestive apparatus in *B. mori* larvae, and it plays an important role in food digestion and nutrition assimilation for its growth, development and silk production. All the big molecular substances such as proteins and carbohydrates from its natural food, mulberry leaves, must be digested into small molecular compounds and subsequently assimilated by midgut epithelial cells. The larval midgut is also the first natural barrier to resist foreign pathogens. During the larval stage, there are five instars, among which the 5th instar is the most important

stage as the larva takes large amount of feed (approx. 85% of the total of whole larval stage), and accumulates most of nutrition and energy for its remaining life, including pupa and moth. The larval 5th instar is also a key period of transition from the larva to pupae. Therefore analysis of the protein composition of the midgut of the 5th instar will definitely help us to understand the larval activities involved in nutrition digestion, assimilation as well as the defence mechanism. Recently, the completion of the genetic physical map and draft sequence for the silkworm genomes [6,7] and the construction of the genome-wide microarray [8] provided a solid foundation for basic research of the silkworm at the proteomics level.

Developments in MS have dramatically increased the ability to analyse complex proteomes in depth. It is shown that LTQ-Orbitrap MS, which has recently been developed, is better than the LTQ-FT [9] to identify protein spots. X!Tandem is an open source proteomics software that could be used to find the best sequence model for a given MS/MS (tandem MS) [10]. In the present study, we analysed the silkworm midgut proteins by using LTQ-Orbitrap, and each MS data were analysed by using three kinds of parameters before being validated by the Poisson model

[10–12]. We have reported here the identification of 90 proteins of the silkworm midgut, among which 79 proteins were reported for the first time.

MATERIALS AND METHODS

Sample preparation and analysis of nano-LC (liquid chromatography) MS/MS

The *B. mori* strain P50 was provided by the Silkworm and Mulberry Station, Zhejiang University. It was reared with the mulberry leaves at 25°C. The midguts of the second day of the 5th instar were dissected, washed with 0.75% NaCl solution and rapidly stored at –80°C. Frozen tissues (220 mg) were disrupted by using a Sample Grinding kit (Amersham Biosciences) with 1 ml of lysis solution (8 M urea, 2 M thiourea and 4% CHAPS). The supernate was collected and transferred to another centrifuge tube for further clean-up using a 2-D Clean-Up kit (Amersham Biosciences). The extracted protein concentration was measured by the Bradford method [12a]. The sample was subsequently digested with trypsin, desalted and dried by using a vacuum evaporator. Tryptic peptides were resuspended in 0.1% formic acid and loaded on to an Ettan MDLC system (GE Healthcare) connected to an LTQ-Orbitrap MS (Thermo Electron Corporation) three times.

MS

Separation of tryptic peptide mixtures was achieved by nanoscale reverse-phase HPLC, in combination with online LTQ-Orbitrap. For the HPLC separation, a nano-MDLC (nano-multidimensional LC) system (Ettan MDLC; GE Healthcare) was used, employing a linear gradient of 5–45% buffer B (95% acetonitrile, 5% water and 0.1% formic acid) over 60 min. The column system consisted of a trap (0.5 mm×2 mm) and a separation column (Magic C18 AQ, 3 μm, 200 Å, 0.2 mm×150 mm), both purchased from Michrom. While column 1, trap 1 was running, column 2, trap 2 was equilibrated with buffer A (95% water, 5% acetonitrile and 0.1% formic acid) to allow continuous running of the sample through two columns.

LTQ-Orbitrap

The mass spectrometer was operated in the data-dependent mode to automatically switch between Orbitrap MS and Orbitrap MS/MS (MS^2) acquired. Survey full scan MS spectra (from m/z 200 to 2000) were acquired in the Orbitrap with resolution $R = 60\,000$ at m/z 400. The most intense ions (up to five, depending on signal intensity) were sequentially isolated for fragmentation; ions were recorded in the Orbitrap with resolution $R = 15\,000$ at m/z 400.

For accurate mass measurements, the lock mass option was enabled in both the MS and the MS/MS mode and the PCM (polydimethylcyclosiloxane) ions generated in the electrospray process from ambient air [protonated $(Si(CH_3)_2O)_6$; $m/z = 445.120025$] were used for internal recalibration in real time. For a single SIM (selected ion monitoring) scan injection of the lock mass into the C-trap, the lock mass 'ion gain' was set at 10% of the target value of the full mass spectrum. When calibrating in the MS/MS mode, the ion at 429.088735 (PCM with neutral methane loss) was used instead for recalibration.

Target ions already selected for MS/MS were dynamically excluded for 180 s. General MS conditions were: electrospray voltage, 1.8 kV, no sheath and auxiliary gas flow; ion transfer tube temperature, 200°C; collision gas pressure, 1.3 mTorr; normalized collision energy, 35% for MS, ion selection threshold was 500 counts for MS^2. An activation q-value of 0.25 and an activation time of 30 ms were applied for MS^2 acquisitions.

Protein identification using X!Tandem and high-confidence proteins selection through the Poisson model

MS data were analysed using X!Tandem open source package [10,13]. Raw data acquired by the MS above were submitted to X!Tandem along with a FASTA formatted file of the protein database containing protein data of silkworm (http://silkworm.genomics.org.cn/jsp/data.jsp) [8]. Search was performed using three different parameters named parameter 1, parameter 2 and parameter 3 for the best model, and the Poisson model was used to assess the likelihood of our false identification [11–13]. The frequency of false peptide matches, μ, was estimated by using the Poisson model satisfying the constraints that the number of false matches predicted cannot exceed the number of total matches observed [12]. Thereby μ was calculated through the following format: the total number of observed peptide matches divided by the total number of amino acids in the search database. The mean number of matches, λ, expected at random for a protein of length L is μL. The protein length-specific probability, P_{rand}, that M or more matches was observed is:

$$P_{rand} = \sum_{i=M}^{\infty} \frac{\lambda^i}{i!} \exp(-\lambda)$$

Moreover, the expected number of matches (E) is:

$$E = N_{db} P_{rand}$$

where N_{db} is the number of sequences in the database. The confidence, C, that we have identified the sequence from which the spectral data were derived and not one of the E false positives is:

$$C = \frac{1}{1 + E}$$

High-confidence ORFs (open reading frames) are defined as a confidence score of at least 0.5 based on this Poisson model.

Protein annotation

The output were obtained through X!Tandem and the Poisson model. Cross-referenced UniGene database IDs (see Supplementary Table S1 at http://www.bioscirep.org/bsr/029/bsr0290363add.htm) were used to interrogate corresponding EST (expressed sequence tag) expression profiles. Our protein hits were categorized according to the GO (gene ontology) annotation using Wego software (http://wego.genomics.org.cn/cgi-bin/wego/index.pl) and IterProScan (http://www.ebi.ac.uk/IterProScan/) annotation. The putative proteins were annotated by using pblastp and their pI and molecular mass values were analysed.

RESULTS

MS reproducibility and searching model analysis with different parameters

The sample was analysed by using LTQ-Orbitrap MS three times (midgut 1, midgut 2 and midgut 3) and subsequently MS data were respectively searched against the silkworm protein database predicted by the genome with X!Tandem of parameters 1–3 with 0.1 as the maximum valid expectation value. Data acquired above (Supplementary Table S1) were screened with the Poisson model with a probability score of 0.5 as the cut-off score, and 90 proteins were obtained (see Supplementary Table S2 at http://www.bioscirep.org/bsr/029/bsr0290363add.htm). Among these proteins, 79 proteins had a probability score of greater than 0.9, and 11 proteins had a probability score between 0.6 and 0.9 with good MS/MS spectra. These results showed that 89 proteins were identified with three or more peptides and one protein was identified with two peptides (see Table 4 and Supplementary Table S2), so our results here are very conservative. We compared three searching methods (three kinds of parameters) for seeking the best model and the results are shown in Tables 1–3. As a result, the model of parameter 1 was considered to be the best model in the results of midgut 1 and midgut 2, but the model of parameter 3 was the best in the results of midgut 3.

Protein annotation

Based on the annotated information (Figure 1 and see Supplementary Table S3 at http://www.bioscirep.org/bsr/029/bsr0290363add.htm), the proteins are classified into four categories: (i) proteins associated with special tissue (the protein related to the midgut digest), (ii) the proteins associated with substance and energy metabolism, (iii) other functional proteins and (iv) unnamed proteins (Table 4).

A total of 90 proteins were compared with midgut proteins found previously and it was found that 79 proteins were identified from the midgut for the first time. Of the 79 proteins, 44 proteins are associated with metabolism, almost accounting for 50% (Figure 2), and as is shown in Figure 2, the largest group is the category of the proteins associated with protein metabolism, accounting for 59% in the proteins associated with metabolism

Table 1 Comparison of three parameters in three experiments
Ratio 1 is the number of proteins present in all three parameters divided by the total number of proteins. Ratio 2 is the number of proteins present in two or three parameters divided by the total number of proteins.

Sample and parameter tested	Number of identified proteins	Ratio 1	Ratio 2
Midgut 1			
Parameter 1	41	0.731707	0.97561
Parameter 2	49	0.612245	0.816327
Parameter 3	45	0.666667	0.866667
Midgut 2			
Parameter 1	52	0.596154	0.884615
Parameter 2	51	0.607843	0.882353
Parameter 3	43	0.72093	0.883721
Midgut 3			
Parameter 1	41	0.804878	0.926829
Parameter 2	45	0.733333	0.844444
Parameter 3	42	0.785714	0.880952

Table 2 Comparison of three parameters in nine results
The ratios indicate the percentage of the number of the same proteins appearing in the various tests ($\geqslant 5$).

Sample	Parameter 1	Parameter 2	Parameter 3
Midgut 1	0.756	0.6734	0.7555
Midgut 2	0.7115	0.7647	0.7674
Midgut 3	0.8292	0.7333	0.7857

Table 3 Comparison of three parameters in nine results
The ratio indicates the percentage of the number of the same proteins appearing in nine tests.

Sample	Parameter 1	Parameter 2	Parameter 3
Midgut 1	0.4634	0.4081	0.4444
Midgut 2	0.3653	0.3921	0.4418
Midgut 3	0.4878	0.4222	0.4523

(Figure 3). This group includes the proteins associated with protein synthesis and folding (Table 4).

These proteins demonstrated diverse characteristics, including varying pI and molecular mass values (Table 4). It is shown that 43% of the proteins identified in the present study have a pI greater than 9 (Figure 4A) and 38% have a molecular mass > 10 000 and < 30 000 (Figure 4B).

Tissue distribution

Cross-referencing the known EST distribution (available from the UniGene database) with our identified proteins demonstrated a total of ten different tissue types that were identified with expression of at least one mRNA. The tissue distribution of every protein ID identified in the present paper is given in Supplementary Table S4 (http://www.bioscirep.org/bsr/029/bsr0290363add.htm) and illustrated in Figure 5.
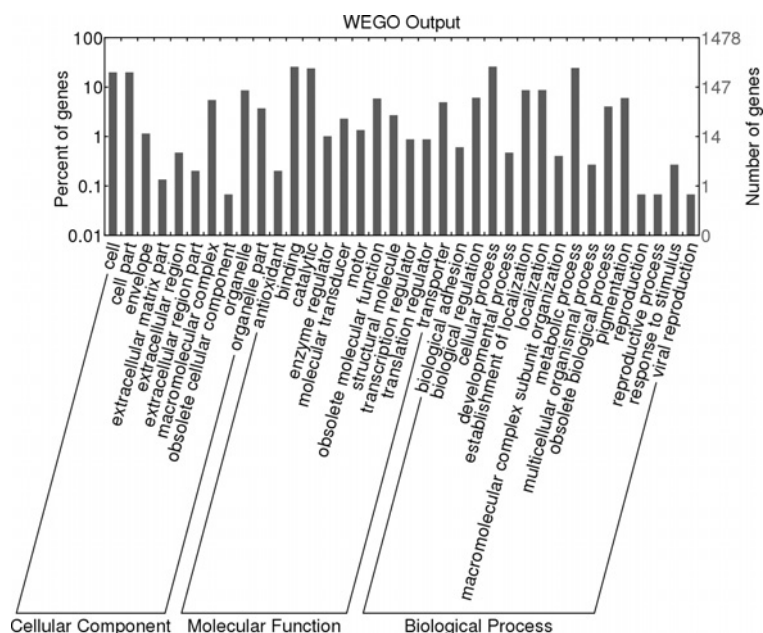
Figure 1   GO categories of the identified protein of the midgut from the larval silkworm by using Wego software

DISCUSSION

In previous papers about the midgut proteome, the methods used for identifying proteins had some disadvantages. In the studies of Zhang et al. [14] and Kajiwara et al. [15] about the midgut, the protein database used was not the silkworm protein database [14,15]. It is known that theoretical and experimental mass comparison is the basis of MS, so using the inappropriate protein database is the main reason why the results are not stringent. The findings of Hou et al. [16] showed that 32 proteins could be found in the silkworm midgut two-dimensional electrophoresis image, but there are only 26 different proteins, due to the presence of duplicate proteins. In addition, Hou et al. [16] did not show mass tolerance, and the mass tolerance selected in the other two studies associated with the midgut proteome is too large [14–16]. For example, the mass tolerances of the enzymatic peptides are 1 and 0.5 Da respectively in Zhang et al. [14] and Kajiwara et al. [15] (our maximum value of the mass tolerances of the fragment in three parameters is 0.2 Da as shown in Supplementary Table S1). Compared with the former study, our identification method is very strict and the result is satisfactory. In our study, the Shotgun proteomics methods were used for protein identification, and midgut proteins were proteolysed to peptides and subsequently subjected to direct MS analysis via CID (collision-induced dissociation) of single peptides. Two-dimensional LC separation prior to MS/MS analysis was employed. X!Tandem and the Poisson model were used to analyse MS and to acquire proteins information. This method has many advantages, e.g. it can identify a large number of proteins relatively quickly and directly analyse membrane proteins. However, this approach has some disadvantages in the random nature of ion selection for CID. The poor reproducibility in picking the same peptides for CID in replicates of the MS/MS course impacts directly on the reproducibility of protein identification [17]. Another factor impacting on it is the different algorithms, as each of them has its own strengths and weaknesses [18]. In order to achieve the best results, the experiments of CID were repeated three times. MS data were searched against the database by using X!Tandem with three parameters. Finally, we found better parameters (parameters 1 and 3). The number of peptide hits is only indicative of the confidence of the identified proteins [19]. In the previous studies, the proteins were identified by only two or more peptides [20,21], but our study showed that it was not a rigorous statistical approach. In the process of validating the results by using the Poisson model, most of the proteins with two peptide hits or peptides were rejected and only Bmb033352 was validated (Table 4).

We identified a total of 90 putative proteins. The number of new proteins is 79, accounting for 87%. Compared with 26 proteins found by Hou et al. [16], the proteins newly found refer to almost all the cellular biological functions. From Table 4, we know that the proteins associated with special tissue of the midgut are found in new proteins. For example, 11 new digestive enzymes are from the epithelium, eight new proteins relating to mechanical digestion are from muscle tissue and three new proteins are from the peritrophic membrane. Besides special tissue proteins, there are 44 proteins involved in metabolism of substance and energy and there are 11 proteins relating to signal transduction, substance transport and cell skeleton in newly found proteins. The proteins' detailed biological functions are illustrated in the following passages.

Table 4 List of proteins identified in the midgut
n.d., not determined.

Function/ classification	Protein identification	Description	Hits	Confidence	Previously described?	Molecular mass (kDa)/pI
Proteins associated with special tissue						
I. Musculature	Bmb012212	Actin	102	1	Yes	41.8/5.22
	Bmb012614	Actin	95	1	No	41.8/5.30
	Bmb021176	Arginine kinase	10	1	Yes	45.5/6.25
	Bmb024911	SM22/calponin	7	1	No	26.0/8.94
	Bmb015063	Myosin heavy chain (mhc)	22	1	No	26.5/6.64
	Bmb015062	Myosin head (motor domain)	12	1	No	134.7/5.63
	Bmb014764	Myosin tail	9	1	No	90.0/9.45
	Bmb015061	Paramyosin/tropomyosin	62	1	No	69.9/5.25
	Bmb014765	Paramyosin/tropomyosin A	4	0.996300937	No	75.5/5.61
	Bmb002610	Paramyosin/tropomyosin A	11	1	No	40.3/9.67
	Bmb004766	Tropomyosin	3	0.877134707	Yes	34.0/4.88
II. Membrane	Bmb008243	Laminin B	6	0.999979686	No	145.8/4.70
	Bmb017923	Type 4 procollagen	6	0.999640674	No	183.0/8.76
	Bmb003222	Chitin-based larval cuticle	9	0.999999999	No	111.7/5.31
III. Epithelium	Bmb003747	Serine protease	9	1	No	29.7/10.15
	Bmb018754	Serine protease	10	1	No	31.4/9.71
	Bmb002795	Membrane alanine aminopeptidase	4	0.896029435	No	108.8/5.75
	Bmb020698	Insulinase-like	3	0.901816622	No	48.8/6.15
	Bmb028877	3-Hydroxyisobutyrate dehydrogenase	5	0.999981755	No	44.8/8.91
	Bmb004877	α-Glucosidase	4	0.999958355	No	23.6/4.79
	Bmb025436	Glycoside hydrolase	3	0.909014635	No	46.2/4.80
	Bmb015353	Glycoside hydrolase	4	0.997609999	No	51.2/4.69
	Bmb008243	Low-density lipoprotein receptor	6	0.999979686	No	145.8/4.70
	Bmb000670	Sterol binding	4	0.999277202	No	43.1/6.88
	Bmb009635	Cellular retinaldehyde binding/α-tocopherol transport	3	0.825121612	No	39.8/6.06
Protein metabolism						
(i) Translation	Bmb010338	Ribosomal protein S4E	5	0.999995366	No	29.7/10.32
	Bmb030045	Ribosomal protein S5	3	0.967224192	No	30.2/9.99
	Bmb006711	Ribosomal protein S8E	4	0.999715636	No	28.5/10.60
	Bmb033626	Ribosomal protein S10b	6	1	No	20.6/9.65
	Bmb025378	Ribosomal protein S19/S15	4	0.999916644	No	17.0/10.37
	Bmb036419	Ribosomal protein S25	3	0.996994178	No	13.2/10.32
	Bmb026270	Ribosomal protein L2	4	0.999901408	No	27.9/10.76
	Bmb032730	Ribosomal protein L3	5	0.999996736	No	45.6/10.15
	Bmb020135	Ribosomal protein L4/L1e	6	0.999999972	No	50.2/10.63
	Bmb001465	Ribosomal protein L5	5	0.9999997	No	22.3/10.18
	Bmb022410	Ribosomal protein L10E	4	0.999901408	No	29.5/9.84
	Bmb023836	Ribosomal protein L11	5	0.999999927	No	17.6/9.58
	Bmb035628	Ribosomal protein L13e	4	0.999862714	No	25.1/11.37
	Bmb031937	Ribosomal protein L14	3	0.993320121	No	19.7/10.83
	Bmb030760	Ribosomal protein L22/L17	3	0.995073244	No	13.6/10.37
	Bmb014252	Ribosomal protein L14b/L23e	5	0.999999919	No	15.1/10.41
	Bmb024069	Ribosomal protein L23	4	0.999085709	No	27.9/11.12
	Bmb039472	Ribosomal protein L7Ae/L30e/S12e/Gadd45	3	0.977194492	No	29.7/10.44
	Bmb037600	Ribosomal protein L35Ae	3	0.998821274	No	10.2/10.88
	Bmb000284	60S acidic ribosomal protein	3	0.997361267	No	11.5/4.18

Table 4 (Continued)

| Function/classification	Protein identification	Description	Hits	Confidence	Previously described?	Molecular mass (kDa)/pI
	Bmb012979	Elongation factor EF-1 or G	5	0.999814361	No	97.3/6.23
	Bmb015183	Elongation factor EF-1 or G	11	1	No	50.4/9.21
(ii) Protein folding	Bmb022619	PDI	5	0.999986524	No	56.1/4.70
	Bmb010862	Peptidylprolyl cis–trans isomerase	5	0.999999945	No	18.0/7.74
	Bmb032503	PDI	5	0.999951562	No	23.3/9.51
	Bmb036625	Hsp90	4	0.991290169	No	72.4/6.36
	Bmb035865	Hsp60	3	0.8115414	Yes	61.1/5.51
	Bmb011274	HSC-70 (heat-shock cognate 70)-3	8	1	Yes	68.1/5.13
Nucleic acid/carbohydrate/lipid metabolism	Bmb004885	Nuclear receptor co-activator 7	3	0.624680364	No	53.9/4.80
	Bmb007243	Mitotic cell cycle G_2/M transition DNA damage checkpoint	3	0.722881223	No	53.1/7.22
	Bmb012751	Histone H3	5	0.999999979	No	15.3/11.27
	Bmb028559	Histone H4	3	0.99940904	No	6.3/12.01
	Bmb027946	Lactate/malate dehydrogenase	7	1	No	24.9/9.41
	Bmb020504	Lactate/malate dehydrogenase	3	0.947052423	No	37.9/6.15
	Bmb006175	GAPDH	7	1	No	35.4/7.70
	Bmb005299	PDH (acetyl-transferring)	4	0.999621762	Yes	23.4/5.36
	Bmb019850	Glucose/ribitol dehydrogenase	4	0.999819923	No	32.1/9.00
	Bmb007149	Citrate synthase	6	0.99999995	No	50.7/8.83
	Bmb025249	Glutamic-oxaloacetic transaminase	4	0.999948826	No	22.5/9.41
	Bmb033769	Enoyl-CoA hydratase/isomerase	7	1	No	38.6/8.59
	Bmb011880	Acetyl-CoA C-acyltransferase	3	0.818890182	No	37.9/6.67
	Bmb039014	FAMeT	3	0.991876073	No	16.8/5.20
Energy metabolism	Bmb033352	ATP synthase E	2	0.691524332	No	9.6/9.74
	Bmb027800	Proton-transporting ATP synthase complex (sensu Eukarya)	4	0.999996332	No	11.2/9.46
	Bmb011628	ATP synthase	30	1	Yes	54.9/5.32
	Bmb008836	ATP synthase	33	1	Yes	64.7/5.19
	Bmb016644	ATP synthase	35	1	Yes	55.0/5.26
	Bmb005542	ATP synthase	53	1	No	59.7/9.21
	Bmb016550	Cytochrome P450 4d2	3	0.989484668	No	14.3/9.25
	Bmb004448	Cytochrome c	5	0.999988323	No	55.5/8.60
	Bmb039215	Cytochrome c oxidase	4	0.999984576	Yes	17.0/5.66
Signal transduction/transport/skeleton protein	Bmb028514	Adenine nucleotide translocator 1	10	1	No	32.9/9.88
	Bmb040999	Glutamyl aminopeptidase	3	0.988729417	No	17.3/5.26
	Bmb004994	Non-membrane spanning protein tyrosine kinase	4	0.996735368	No	66.9/9.07
	Bmb020823	Serine/threonine protein kinase	3	0.602339333	No	91.1/8.42
	Bmb000655	Tau (microtubule associated complex)	4	0.965994786	No	71.1/9.93
	Bmb008789	β-Tubulin	29	1	No	48.2/4.85
	Bmb020953	β-Tubulin	6	0.999999975	No	48.3/5.46
	Bmb004930	α-Tubulin	10	1	No	51.0/5.06
	Bmb003475	β-Tubulin	10	1	No	47.5/4.72
	Bmb019690	Trol leak terribly reduced optic lobes	10	1	No	155.0/4.92
	Bmb000857	Transferrin	9	1	No	62.1/5.13
	Bmb013324	Mitochondrial outer membrane (porin)	7	1	Yes	19.7/6.83
Unannotated	Bmb006157	n.d.	3	0.999481132	No	7.4/11.72
	Bmb024309	n.d.	3	0.845164736	No	35.1/6.42
	Bmb003928	n.d.	3	0.893318237	No	43.2/10.09
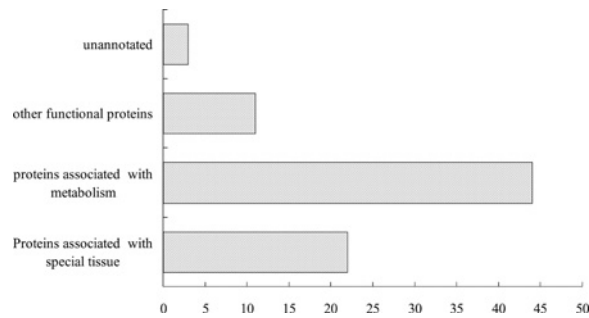
Figure 2 Functional representation of the proteins newly identified in the *B. mori* midgut
A total of 79 proteins identified were divided into four groups that are summarized in Table 4, which provides a brief overview of the *B. mori* larval midgut proteome.
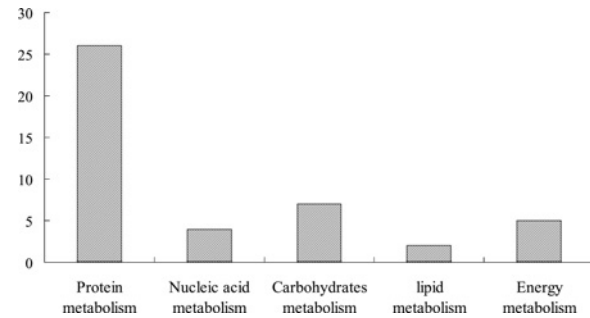

Figure 3 Functional categorization of the newly identified protein associated with metabolism
The *y*-axis shows the number of proteins in each category.

The digestive process of the midgut of the silkworm includes mechanical digestion and chemical digestion through muscular action and digestive enzyme. Peristalsis of the musculature of the midgut, such as muscle contraction, helps the digestion. Muscle contraction is caused by sliding between the thick filament and the thin filament of the myofibril. According to the paper by Scott and Jeffrey [21], myosin and paramyosin are the most abundant thick filament proteins in invertebrates. Myosin exists as a hexamer of two heavy chains, two alkali light chains and two regulatory light chains. The proteins related to the heavy chain are divided into the N-terminal globular head and the C-terminal coiled-coil tail. Myosin heavy-chain proteins identified in the present study include Bm015063, Bm015062, Bmb014764, Bmb015061, Bmb014765 and Bmb002610; among them, Bmb014764, Bmb015061 and Bmb002610 contain myosin tail, and Bmb015062 contains
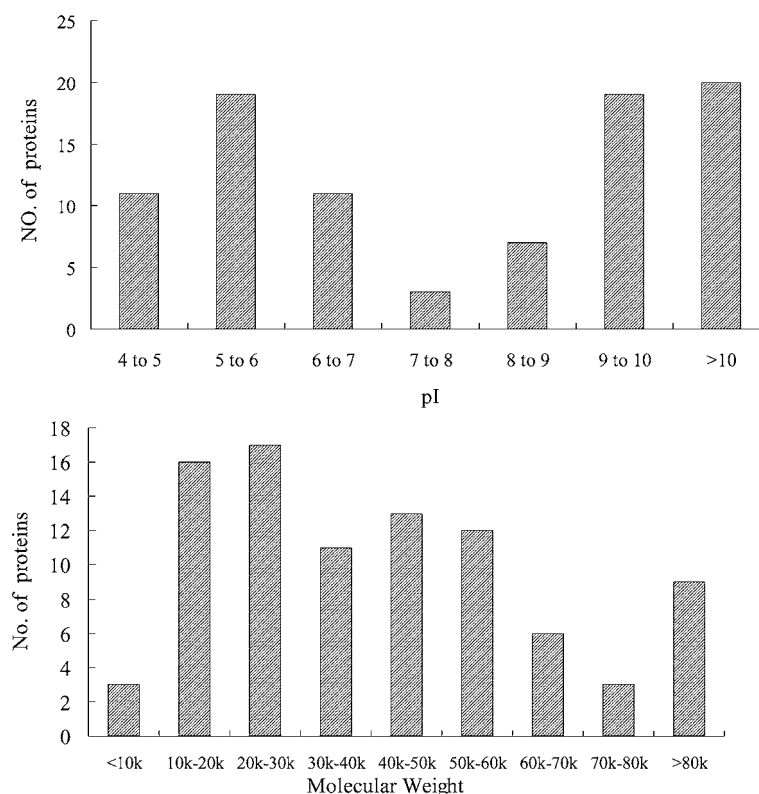

Figure 4   The pI and molecular-mass distribution of newly identified midgut proteins
(A) The pI distribution of newly identified midgut proteins. (B) The molecular mass distribution of newly identified midgut proteins.
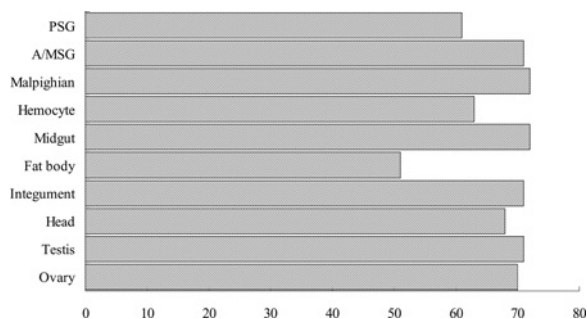
Figure 5 Pattern of EST tissue distribution of protein hits
The Unigene database was used to identify EST tissue distribution for every protein hit reported. The top ten expressed profiles are shown. The x-axis shows the number of protein hits expressed in each tissue. A/MSG, anterior/median silk gland; PSG, posterior silk gland.

myosin globular head, offering energy for muscle contraction. In addition, Bmb002610 containing a myosin tail also supplies energy according to Interpro annotation (Supplementary Table S3). Bmb014765 is a kind of paramyosin that plays an unexpected role in myoblast fusion, myofibril assembly and muscle contraction [21]. The thin filaments are primarily composed of actin (Bmb012212 and Bmb012614) and tropomyosin (Bm004766). Other than these proteins, Bmb024911 (calponin) is a type of calcium-binding protein that can inhibit the ATPase activity of myosin in smooth muscle and can be regulated by phosphorylation of a protein kinase [21]. Bmb021176 is suggested by Interpro annotation (Supplementary Table S3) to be a type of arginine kinase that is bound to actin of the thin filament.

The digestive enzyme in the midgut contains ectoenzyme, present in cavum intestinale, and endoenzyme, mainly present in microvilli of the columnar cell. Most of the enzymes identified in our analysis belong to endoenzyme except chymotrypsin-like (Bmb003747 and Bmb018754) and insulinase-like (Bmb020698). The three kinds of enzymes belong to serine proteases; among them; chymotrypsin-like prefers to cleave on the carbonyl side of aromatic residues, such as phenylalanine and tyrosine, and insulinase-like cleaves peptides on the carbonyl side of the basic amino acids, arginine or lysine [22]. Chymotrypsin-like also has strong antiviral activity against BmNPV [23]. However, ectoenzyme or proteinases in cavum intestinale cannot drastically catalyse the protein to amino acids and it needs endoenzyme. For example, Bmb002795 (membrane alanine aminopeptidase) is a kind of aminopeptidase that catalyses the removal of single amino acids in the N-terminus of small peptides and thereby plays a role in their final digestion [24] and Bmb028877 (3-hydroxyisobutyrate dehydrogenase) is one of the important enzymes related to valine metabolism [25]. Carbohydrates identified here contain α-glucosidase (Bmb004877), which has similar enzymatic activity to γ-amylase and participates in glycoprotein processing in the endoplasmic reticulum [26], and glycoside hydrolases (Bmb025436 and Bmb015353), which are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates, or between

a carbohydrate moiety [27]. The digestive enzymes associated with lipid metabolism identified here are mainly apolipoproteins; for example, Bmb000670 has sterol carrier activity and Bmb009635 can transport α-tocopherol. Bmb004877 participates in the process of regulation of lipid metabolism and is involved in neurodevelopment, and transports nutrients and vitamins [28]. It is in accord with apparent characters. The larvae midgut is not the main tissue synthesizing lipid or digesting lipid completely, and the lipid component that silkworm assimilates is transported to the fat tissue.

Peritrophic membrane can protect the midgut epithelium from mechanical rubbing from food digestion and infection from pathogens [29–31]. It is made up of protein (47%), chitin (47%) and polysaccharide (5%). The proteins associated with peritrophic membrane include Bmb008223 (laminin B), Bmb017923 (type IV procollagen) and Bmb003222 (chitin-based larval cuticle). In addition, Bmb008243 has many functional domains, so it can also be found in the epithelium (Table 4).

We identified 19 ribosomal proteins and two elongation factors (Table 4; Figure 2), which are the main proteins responsible for protein translation. The large subunits of ribosomal protein are composed of a globular, surfaced-exposed domain with long finger-like projections that extend into the rRNA core to stabilize its structure and they interact with multiple RNA elements, too [32,33]. Further, we identified 13 large units of ribosomal proteins including Bmb026270, Bmb032730, Bmb020135, Bmb001465, Bmb022410, Bmb023836, Bmb035628b, Bmb031937, Bmb030760, Bmb014252, Bmb024069, Bmb039472 and Bmb037600. Protein S4 (Bmb10338) initiates assembly of small unit and is located at junctions of five and four RNA helices respectively [32,33]. The proteins of the small unit we identified include Bmb030045, Bmb006711, Bmb033626, Bmb025378 and Bmb036419.

Newly synthesized proteins can be folded with the help of a folding helper such as chaperonin and isomerase. Chaperone proteins can help some proteins folding during synthesis and refolding during denaturing stress, thereby preventing protein misfolding and aggregation. Hsps (heat-shock proteins) are a family of molecular chaperone proteins that promote protein folding, assembly, translocation and secretion of newly synthesized polypeptides. They also participate in the removal or repairing of denatured proteins acting as molecular chaperons. Different Hsps target different proteins. Hsp70 (Bmb011274) binds to nascent proteins and helps them to reach its final native state. Hsp90 (Bmb035865 and Bmb036625), however, does not act in nascent protein folding. Instead, it targets signal transduction proteins such as steroid hormone receptors and signalling kinases [34]. Besides the chaperone's proteins, the other folding helper is isomerases including PDI (protein disulfide-isomerase) and peptidylprolyl *cis–trans* isomerase. PDI (Bmb022619 and Bmb032503) accelerates the process of forming the proper disulfide bond and switches its conformation from dimer to tetramer in its functions as a foldase [35]. Peptidylprolyl *cis–trans* isomerase (Bmb010862) catalyses peptidyl-proline circumgyration of 180 angles to accelerate the folding process and exhibits chaperone-like activity during muscle creatine kinase [36].

Some proteins identified are associated with the metabolism of nucleic acids. For example, histone H3 (Bmb012751 and Bmb028559) forming the eukaryotic nucleosome octomer core winds approx. 146 DNA base-pairs [37]. Bmb004885 (nuclear receptor co-activator 7) and Bmb007243 (the protein associated with mitotic cell cycle G_2/M transition DNA damage checkpoint) are the proteins related to cell division. Some proteins are related to carbohydrate metabolism. For example, Bmb027946 (lactate or malate dehydrogenase), Bmb020504 (lactate or malate dehydrogenase), Bmb006175 [GAPDH (glyceraldehyde-3-phosphate dehydrogenase)], Bmb005299 [PDH (pyruvate dehydrogenase)], Bmb019850 (glucose/ribitol dehydrogenase), Bmb007149 (citrate synthase) and Bmb025249 (glutamic-oxaloacetic transaminase) are associated with carbohydrate metabolism [38]. Some proteins are responsible for fatty acid catabolism, such as Bmb033769 (enoyl-CoA hydratase/isomerase) and Bmb011880 (acetyl-CoA C-acyltransferase) [38]. Bmb039014 [FAMeT (famesoic acid *o*-methyltransferase)] can catalyse the conversion of farnesoic acid into MF (methylfarnesoate) by the mandibular organ of crustaceans [39]. Previous studies indicated that eyestalk neuropeptides may negatively regulate MF biosynthesis through FAMeT in the mandibular organ [40]. In insects, the regulation of JH (juvenile hormone) III production in corpora allata and allatostatin-mediated inhibitory action on JH III biosynthesis is facilitated by FAMeT [41].

Some proteins are associated with energy metabolism in the midgut. For example, in the function of cytochrome oxidase (Bmb039215), cytochrome *c* (Bmb004448) binding Fe^{2+} can send the electrons from QH_2 to O_2. Bmb016550 (cytochrome P450 4d2) plays a major role in the synthesis and degradation of insect hormones (ecdysteroids and juvenoids) and the activity [42] or detoxification of such chemicals as plant toxins and insecticide [42–45]. The F_1F_o-ATP synthase complex exists only in the form of dimeric and oligomeric complexes in the inner mitochondrial membrane [46]. It consists of F_o subunits responsible for proton transport and F_1 subunits performing the task of ATP synthesis/hydrolysis. From Table 4, Bmb011628, Bmb008836, Bmb016644 and Bmb005542 all are the components forming ATP synthase and have the capability of proton transport and binding ATP; however, only Bmb027800 can bind ATP but does not have the capability of proton transport. In *Saccharomyces cerevisiae*, Bmb033352 (ATP synthase E)'s N-terminal hydrophobic region plays an important role in the subunit e-dependent processes of mitochondrial DNA maintenance, modulation of mitochondrial morphology and stabilization of the dimer-specific F_o subunits, subunits g and k. Its C-terminal coiled-coil region of subunit e functions to stabilize the dimeric form of detergent-solubilized ATP synthase complexes [47].

There are other functional proteins such as signal transduction protein, transport protein, skeleton proteins etc. Bmb028514 is a kind of adenine nucleotide translocator, which is a target protein of nitric oxide, peroxynitrite or 4-hydroxynonenal, which is involved in the pathological demise of cells via apoptosis [48]. Transferrin (Bmb000857), glycoprotein, is a blood plasma protein for iron ion delivery, which reversibly binds iron very tightly [49]. Learned from Interpro annotation (Supplementary Table S3), Bmb013324 (porin) not only transports anion but also has voltage-dependent ion-selective channel activity. Protein kinase catalyses the phosphorylation of serine, threonine or tyrosine residue of target protein, which change the conformation and function of the target protein. Bmb020823 (serine/threonine protein kinase) and Bmb004994 (tyrosine protein kinase) play an important role in many cellular processes such as division, proliferation, apoptosis and differentiation [50]. Microtubule, one of the basal ingredients of eukaryotic cytoskeleton, is made of tubulin, adimeric proteins composed of α-tubulin and β-tubulin. We identified a kind of α-tubulin (Bmb004930) and three kinds of β-tubulin (Bmb008789, Bmb020953 and Bmb003475). Tau proteins are microtubule-associated proteins that are involved in microtubule assembly and stabilization and cellular shape, motility and signal transduction. Bmb000655 is a kind of tau protein. Learned from Interpro annotation (Supplementary Table S3), it contains a proline-rich region, so the protein is phosphorylated by many kinases, changes the protein's conformation and affects its ability to bind microtubules, the tubulin-binding domain and other cytoskeletal elements. In addition, tau mRNA is expressed predominantly in neurons, and particularly in their axons. According to Interpro annotation (Supplementary Table S3), Bmb019690 is an axon guidance receptor, but tau does not appear to be an essential protein, since transgenic mice lacking tau appear to develop a normal nervous system with only mild alterations in the structure of certain small-calibre axons [51].

REFERENCES

1 Kurin, R. (2002) The silk road: connecting cultures, creating trust. In Talk Story, Vol. 21 (Fall), pp. 1–11, Smithsonian Center for Folklife and Cultural Heritage

2 Marian, R. G., Toru, S. and Hiroaki, A. (2005) The genetics and genomics of the silkworm, *Bombyx mori*. Annu. Rev. Entomol. 50, 71–100

3 Wu, X. F., Yin, Z. Z., Cao, C. P., Lu, X. M., Huang, L., Lu, X. M., Liu, J. X. and Cui, W. Z. (2004) Expression of human VEGF165 in silkworm (*Bombyx mori* L.) by using a recombinant baculovirus and its bioactivity assay. J. Biotechnol. 111, 253–261

4 Wang, Y. Z., Wu, X. F., Liu, G. F., Cao, C. P., Huang, H. Q., Xu, Z. R. and Liu, J. X. (2005) Expression of porcine lactoferrin by using recombinant baculovirus in silkworm, *Bombyx mori* L., and its purification and characterization. Appl. Microbiol. Biotechnol. 69, 385–389

5 Wu, X. F. and Cao, C. P. (2004) Targeting of human aFGF gene into silkworm, *Bombyx mori* L. through homologous recombination. J. Zhejiang Univ. Sci. 5, 644–650

6   Xia, Q. Y., Zhou, Z. Y., Lu, C., Cheng, D. J., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C. et al. (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mo*ri). Science 306, 1937–1940

7   Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y. et al. (2004) The genome sequence of silkworm, *Bombyx mori*. DNA Res. 11, 27–35

8   Xia, Q. Y., Cheng, D. J., Huan, J., Wang, G. H., Cheng, T., Zha, X., Liu, C., Zhao, P., Dai, F., Zhang, Z. et al. (2007) Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*. Genome Biol. 8, R162

9   Souza, G. A., Godoy, L. M. F. and 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

10 Robertson, C. and Ronald, C. B. (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466–1467

11 Marcin, A., Thomas, B., Rajasree, M., Lennart, M., Henning, H., Chris, T., Gilbert, S. O., David, J. and States, D. (2005) Data management and preliminary data analysis in the pilot phase of the HUPO Plasma Proteome Project. Proteomics 5, 3246–3261

12 David, J. S., Gilbert, S. O., Thomas, W. B., Damian, F., Jimmy, E., David, W. S. and Samir, M. H. (2006) Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. Nat. Biotechnol. 24, 333–338

12a Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. Anal. Biochem. 72, 248–254

13 Damian, F., Baxter, B. A., Thomas, W. B., Rajasree, M., Adamski, M., Xu, Y., Ulintz, P., Omenn, G. S. and States, D. J. (2006) Novel gene and gene model detection using a whole genome open reading frame analysis in proteomics. Genome Biol. 7, R35

14 Zhang, P. B., Yamamoto, K., Wang, Y. Q., Banno, Y., Fujii, H., Miake, F., Kashige, N. and Aso, Y. (2004) Utility of dry gel from two-dimensional electrophoresis for peptide mass fingerprinting analysis of silkworm proteins. Biosci. Biotechnol. Biochem. 68, 2148–2154

15 Kajiwara, H., Ito, Y., Imamaki, A., Nakamura, M., Mita, K. and Ishizaka, M. (2005) Protein profile of silkworm midgut of fifth-instar day-3 larvae. J. Electrophor. 49, 61–69

16 Hou, Y., Guan, J., Zhao, P. and Liu, H. L. (2007) Proteomics analysis of midgut from silkworm, *Bombyx mori*. Acta Sericol. Sin. 33, 216–222

17 Yi, E. C., Marelli, M., Lee, H., Purvine, S. O., Ruedi, A., John, A. D. and David, G. R. (2002) Approaching complete peroxisome characterization by gasphase fractionation. Electrophoresis 23, 3205–3216

18 Ruedi, A. and Matthias, M. (2003) Mass spectrometry-based proteomics. Nature 422, 198–207

19 Utleg, A. G., Yi, E. C., Xie, T., Shannon, P., White, J. T., Goodlett, D. R., Hood, L. and Lin, B. Y. (2003) Proteomic analysis of human prostasomes. The Prostate 56, 150–161

20 Xiong, Y., Chalmers, M. J., Gao, F. P., Cross, T. A. and Marshall, A. G. (2005) Identification of *Mycobacterium tuberculosis* H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. J. Proteome Res. 4, 855–861

21 Scott, L. H. and Jeffrey, B. T. (2005) Invertebrate muscles: muscle specific genes and proteins. Physiology 85, 1001–1060

22 Nakazawa, H., Tsuneishi, E., Ponnuvel, K. M., Furukawa, W., Seiichi, F., Ai, A., Hiromitsu, T., Jun, I. and Minoru, Y. (2003) Antiviral activity of a serine protease from the digestive juice of *Bombyx mori* larvae against nucleopolyhedrovirus. Virology 321, 154–162

23 Reginald, H. G. and Charles, M. G. (2005) Biochemistry (3rd edn), Brooks/Cole Publishing Company, Pacific Grove, CA

24 Rawlings, N. D. and Barrett, A. J. (1995) Evolutionary families of metallopeptidases. Methods Enzymol. 248, 183–228

25 Steele, M. I., Lorenz, D., Hatter, K., Park, A. and Sokatch, J. R. (1992) Characterization of the mmsAB operon of *Pseudomonas aeruginosa* PAO encoding methylmalonate-semialdehyde dehydrogenase and 3-hydroxyisobutyrate dehydrogenase. J. Biol. Chem. 267, 13585–13592

26 Wilkinson, B. M., Purswani, J. and Stirling, C. J. (2006) Yeast GTB1 encodes a subunit of glucosidase II required for glycoprotein processing in the endoplasmic reticulum. J. Biol. Chem. 281, 6325–6333

27 Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. and Davies, G. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc. Natl. Acad. Sci. U.S.A. 92, 7090–7094

28 May, P., Woldt, E., Matz, R. L. and Boucher, P. (2007) The LDL receptor-related protein (LRP) family: an old family of proteins with new physiological functions. Ann. Med. 39, 219–228

29 Bolognesi, R., Ribeiro, A. F., Walter, R. T. and Clélia, F. (2001) The peritrophic membrane of *Spodoptera frugiperda*: secretion of peritrophins and role in immobilization and recycling digestive enzymes. Archiv. Insect Biochem. Physiol. 47, 62–75

30 Tellam, R. L. (1996) The peritrophic matrix. In Biology of the Insect Midgut (Lehane, M. J. and Billingsley, P. F., eds), pp. 86–113, Chapman and Hall, London

31 Tellam, R. L. and Eisemann, C. (2000) Chitin is only a minor component of the peritrophic matrix from larvae of *Lucilia cuprina*. Insect Biochem. Mol. Biol. 30, 1189–1201

32 Maguire, B. A. and Zimmermann, R. A. (2001) The ribosome in focus. Cell 104, 813–816

33 Chandra, S. S. and Liljas, A. (2000) The end of the beginning: structural studies of ribosomal proteins. Curr. Opin. Struct. Biol. 10, 633–636

34 Young, J. C., Moarefi, I. and Hartl, F. U. (2001) Hsp90: a specialized but essential protein-folding tool. J. Cell Biol. 154, 267–273

35 Zhao, T. J., Ou, W. B., Xie, Q., Liu, Y., Yan, Y. B. and Zhou, H. M. (2005) Catalysis of creatine kinase refolding by protein disulfide isomerase involves disulfide cross-link and dimer to tetramer switch. J. Biol. Chem. 280, 13470–13476

36 Ou, W. B., Luo, W., Park, Y. D. and Zhou, H. M. (2001) Chaperone-like activity of peptidyl-prolyl *cis*–*trans* isomerase during creatine kinase refolding. Protein Sci. 10, 2346–2353

37 Thatcher, T. H. and Gorovsky, M. A. (1994) Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res. 22, 174–179

38 Wang, X. C. (2000) Biochemistry, Tsinghua University Press, Beijing

39 Wainwright, G., Webster, S. G. and Rees, H. H. (1999) Involvement of adenosine cyclic-3′,5′-monophosphate in the signal transduction pathway of mandibular organ-inhibiting hormone of the edible crab, *Cancer pagurus*. Mol. Cell. Endocrinol. 154, 55–62

40 Tobe, S. S., Young, D. A. and Khoo, H. W. (1989) Production of methyl farnesoate by the mandibular organs of the mud crab, *Scylla serrata*: validation of a radiochemical assay. Gen. Comp. Endocrinol. 73, 342–353

41 Yagi, K. J., Konz, K. G., Stay, B. and Tobe, S. S. (1991) Production and utilization of farnesoic acid in the juvenile hormone biosynthetic pathway by corpora allata of larval *Diploptera punctata*. Gen. Comp. Endocrinol. 81, 284–294

42 Li, B., Xia, Q. Y., Lu, C., Zhou, Z. Y. and Xiang, Z. H. (2005) Analysis of cytochrome P450 genes in silkworm genome (*Bombyx mori*). Sci. China Ser. C Life Sci. 48, 414–418

43 Scott, J. G. (1999) Molecular basis of insecticide resistance: cytochromes P450. Insect Biochem. Mol. Biol. 29, 757–777

44 Schuler, M. (1996) The role of cytochrome P450 monooxygenasas in plant–insect interaction. Plant Physiol. 112, 1411–1419

45 Berge, J. B., Feyereisen, R. and Amichot, M. (1998) Cytochrome P450 monooxygenase and insecticide resistance in insect. Biol. Sci. 353, 1701–1705

46 Gavin, P. D., Prescott, M., Luff, S. E. and Devenish, R. J. (2004) CRM1 and IFN-α1 mRNA: nuclear export partners. J. Cell Sci. 117, 2333–2343

47 Junge, W., Lill, H. and Engelbrecht, S. (1997) ATP synthase: an electrochemical transducer with rotary mechanics. Trends Biochem. Sci. 22, 420–423

48 Vieira, H. L. A., Belzacq, A. S., Haouzi, D., Bernassola, F., Cohen, I., Jacotot, E., Ferri, K. F., Hamel, C. E., Bartle, L. M., Melino, G. et al. (2001) The adenine nucleotide translocator: a target of nitric oxide, peroxynitrite, and 4-hydroxynonenal. Nature 20, 4305–4316

49 Parkkinen, J., Bonsdorff, L., Ebeling, F. and Sahlstedt, L. (2003) Function and therapeutic development of apotransferrin. Vox. Sang. 83 (Suppl. 1), 321–326

50 Manning, G., Plowman, G. D., Hunter, T. and Sudarsanam, S. (2002) Evolution of protein kinase signaling from yeast to man. Trends Biochem. Sci. 27, 514–520

51 Dehmelt, L. and Halpain, S. (2005) The MAP2/tau family of microtubule-associated proteins. Genome Biol. 6, 204