



Physiology of digestion and the molecular characterization of the major digestive enzymes from *Periplaneta americana*



Fábio K. Tamaki^{a,1}, André C. Pimentel^{a,1}, Alcides B. Dias^a, Christiane Cardoso^a, Alberto F. Ribeiro^b, Clélia Ferreira^a, Walter R. Terra^{a,*}

^a Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, C.P. 26077, 05513-970 São Paulo, Brazil

^b Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, C.P. 11461, 05513-970 São Paulo, Brazil

ARTICLE INFO

Article history:

Received 26 June 2014

Received in revised form 22 August 2014

Accepted 23 August 2014

Available online 1 September 2014

Keywords:

Proteases

Carbohydrases

Digestive enzymes sequences

Secretory mechanism

Enzyme immunocytolocalization

ABSTRACT

Cockroaches are among the first insects to appear in the fossil record. This work is part of ongoing research on insects at critical points in the evolutionary tree to disclose evolutionary trends in the digestive characteristics of insects. A transcriptome (454 Roche platform) of the midgut of *Periplaneta americana* was searched for sequences of digestive enzymes. The selected sequences were manually curated. The complete or nearly complete sequences showing all characteristic motifs and highly expressed (reads counting) had their predicted sequences checked by cloning and Sanger sequencing. There are two chitinases (lacking mucin and chitin-binding domains), one amylase, two α - and three β -glucosidases, one β -galactosidase, two aminopeptidases (none of the N-group), one chymotrypsin, 5 trypsins, and none β -glucanase. Electrophoretic and enzymological data agreed with transcriptome data in showing that there is a single β -galactosidase, two α -glucosidases, one preferring as substrate maltase and the other aryl α -glucoside, and two β -glucosidases. Chromatographic and enzymological data identified 4 trypsins, one chymotrypsin (also found in the transcriptome), and one non-identified proteinase. The major digestive trypsin is identifiable to a major *P. americana* allergen (Per a 10). The lack of β -glucanase expression in midguts was confirmed, thus lending support to claims that those enzymes are salivary. A salivary amylase was molecularly cloned and shown to be different from the one from the midgut. Enzyme distribution showed that most digestion occurs under the action of salivary and midgut enzymes in the foregut and anterior midgut, except the posterior terminal digestion of proteins. A counter-flux of fluid may be functional in the midgut of the cockroach to explain the low excretory rate of digestive enzymes. Ultrastructural and immunocytochemical localization data showed that amylase and trypsin are released by both merocrine and apocrine secretion mainly from gastric caeca. Finally, a discussion on Polyneoptera digestive physiology is provided.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Cockroaches, which are among the first neopteran insects to appear in the fossil record, are extremely generalized in most morphological features. They are usually omnivorous and are included in the sub-order Blattodea that together with Mantodea (mantids) form the Order Dictyoptera. After extensive molecular phylogenetic analyses, Inward et al. (2007) showed that termites are social cockroaches, no longer deserving classification as a separate order (Isoptera) from cockroaches. Actually, termites pertain to a sister family (Termitidae) of that of the woodroach *Cryptocercus*

(Cryptocercidae) (Lo et al., 2000). The branch Cryptocercidae-Termitidae is a sister of Blattidae, forming Blattoidea that is a sister of Blaberoidea (Blattellidae plus Blaberidae), which in addition to Polyphagoidea form the Blattodea.

The organization of the digestive process in the different insect orders that corresponds to the basic plans of the ancestral forms was reviewed several times (Terra, 1988, 1990; Terra and Ferreira, 1994, 2012). Dictyoptera is supposed to be derived from the Polyneoptera ancestors. Hence, its basic digestive organization should be alike that of the Polyneoptera ancestor. One of the aims of this paper is to provide support in this direction.

Data on compartmentalization of midgut pH and digestion in Dictyoptera are fragmentary and contradictory (see Elpidina et al., 2001 and references therein). In part this is a consequence

* Corresponding author. Fax: +55 11 3838 2186.

E-mail address: warterra@iq.usp.br (W.R. Terra).

¹ These authors contribute equally to this paper.

of differences among the cockroaches. For example, starch digestion occurs mainly in foregut and anterior midgut and protein digestion takes place in posterior midgut in *Nauphoeta cinerea* (Blaberidae) (Elpidina et al., 2001), whereas in several other cockroach species there is not a marked difference in starch digestion in foregut and midgut (Vinokurov et al., 2007). There is a lack of data on cockroach terminal digestion of both proteins and carbohydrates and on the presence (or absence) of a midgut countercurrent flux of fluid to propel digestive enzyme recycling (reviews: Terra, 1988, 1990; Terra and Ferreira, 1994, 2012). Thus, a re-examination of *Periplaneta americana* (Blattidae) digestive enzyme compartmentalization attempting to those details and using only one sex (males) would reduce variability and result in a broader picture of cockroach digestive physiology.

There are numerous papers dealing with cockroach digestive enzymes, including from *P. americana*, such as chitinase (Powning and Irzykiewicz, 1963), glycosidases (Scrivenner et al., 1989; Koffi et al., 2012), peptidases (Baumann, 1990; Lopes and Terra, 2003; Hivrale et al., 2005) and also some works regarding the distribution of polymer hydrolases in cockroach gut, exemplified by Elpidina et al. (2001) and Vinokurov et al. (2007). Nevertheless, there is a lack of data on digestive enzyme sequences.

In this paper, we combined enzymological and transcriptomic data to identify the sequences of the major digestive enzymes. Furthermore, enzyme distribution studies showed how the digestive process is spatially organized and that a countercurrent flux of fluid may be functional in the *P. americana* midgut to explain the low excretory rate of digestive enzymes. Finally, secretory mechanisms of digestive enzymes were identified by immunocytolocalization of trypsin and amylase.

2. Materials and methods

2.1. Animals and enzyme samples

P. americana (Dictyoptera) were laboratory reared feeding on chayote (*Sechium* sp.) and oats (*Avena sativa*) and maintained under a natural photoregime at room temperature. Only adult males were used in the experiments, because females do not routinely feed, lay eggs (that may affect results), and have variable sizes. Adult males were immobilized in a carbon dioxide chamber for 10 min. The antennae, legs and wings were removed, the insects placed on ice and then dissected in cold 220 mM NaCl. The rinsed guts were transferred to a glass slide. The gut was divided into foregut (crop), midgut caeca, midgut and ileum + colon. Midgut was separated into sections of identical lengths: anterior and posterior midgut. Midgut sections were separated into tissue and contents.

Midgut sections of tissue and contents, as well as foreguts and ileum, were homogenized in double distilled water with the aid of a Potter-Elvehjem homogenizer and centrifuged at 20,000×g for 30 min at 4 °C. The supernatants were filtered in glass wool and the pellets (except those of midgut contents) were resuspended in double distilled water. The pellets were regarded as cell membrane fraction, because the hypotonic homogenization medium ruptures all cells. The samples were stored at –20 °C until use. No enzyme inactivation was detected on storage.

2.2. Protein determination, enzymatic assays, excretion rates, effect of pH on enzymatic activity and luminal pH determination

Protein was determined according to Bradford (1976) using ovalbumin as a standard.

Enzymatic assays were performed at 30 °C in specific conditions (Table 1). Proteinase profiles were determined using synthetic

substrates (Table 1) or the natural substrate casein-FITC (Twining, 1984), combined or not with inhibitors: 17 μM SBTI, 100 μM benzamidine, 100 μM chymostatin or 20 μM TPCK.

The effect of pH on enzymatic activities were studied in 50 mM sodium citrate–phosphate (pH 4.0–7.0) or 50 mM sodium phosphate (pH 7.0–8.0) for α-glucosidase, β-glucosidase, β-galactosidase, cellobiase, and trehalase. For amylase, trypsin, maltase and aminopeptidase assays were performed in 100 mM sodium acetate (pH 3.0–4.0), 100 mM piperazine (pH 4.0–6.0), 100 mM HEPES (pH 6.0–8.0), 100 mM TAPS (pH 8.0–9.0) and 100 mM CAPS (pH 9.0–10.0). All buffers included 100 mM NaCl.

The pH of midgut contents sections were separately determined using a universal pH indicator (E. Merck, Darmstadt, pH 4.0–10.0). For this, midgut section contents were dispersed in 5 μL of dissecting saline and then added to 5 μL of a 5-fold dilution of the pH indicator. The resulting colored solutions were compared with suitable standards.

Excretion rates were calculated using the equation $\{[Activity\ Ileum + Colon \times (Volume\ Midgut / Volume\ Ileum + Colon)] / Activity\ Midgut\}$.

2.3. Resolution of midgut enzymes by chromatography and polyacrylamide-gel-electrophoresis in native conditions

Midgut proteins were resolved by ion-exchange chromatography in a High-Q column connected to a Econo-pac System (Bio-Rad, USA) in 20 mM Tris–HCl pH 7.5. Elution was carried by a linear gradient of 0–0.6 M NaCl, followed by a washing step with 10 mL 1 M NaCl. Fractions of 1 mL were collected and assayed with different substrates for peptidases in the presence and absence of specific enzyme inhibitors.

Midgut homogenates were applied onto a 7.5% polyacrylamide gel (8 cm), and proteins were electrophoretically resolved at 4 °C using 2.5 mA per cylinder (Hedrick and Smith, 1968). Afterwards, gels were washed for 1 h in 20 mM citrate–phosphate pH 6.0 to remove the Tris–HCl buffer (which inhibits carbohydrases) at 4 °C and subsequently sliced in a 50 mM citrate–phosphate pH 6.0 buffer. Samples were collected and proteins were eluted for 3 h at 4 °C, before being assayed with different substrates for carbohydrases (Table 1).

2.4. SDS–PAGE and Western blotting

SDS–PAGE was carried in 12% (w/v) polyacrylamide gel electrophoresis containing 0.1% SDS in a discontinuous pH system (Laemmli, 1970), in a Mini-Protein II equipment (Bio-Rad (U.S.A.)). Samples were mixed with sample buffer containing: 60 mM Tris–HCl buffer pH 8.8, 2.5% (w/v) SDS, 10% glycerol, 0.05% (w/v) bromophenol blue and 0.36 mM 2-mercaptoethanol. The samples were heated at 95 °C for 3 min, before being loaded onto gel. Electrophoresis was carried out at 200 V and silver-stained (Blum et al., 1987).

Western blottings were performed after SDS–PAGE separation of midgut homogenate proteins (see above) and electrophoretic transfer to nitrocellulose membrane filter (pore size 0.45 μm, Bio-Rad, U.S.A.). The transfer efficiency was verified using a pre-stained molecular mass marker (Bio-Rad, U.S.A.). After blocking the membrane with milk in TBS-T (Tris-buffered saline containing 0.05% Tween 20), it was reacted for 4 h at room temperature with polyclonal *Tenebrio molitor* anti-amylase (Cristofolletti et al., 2001) or polyclonal *Musca domestica* anti-trypsin (Jordão et al., 1996) sera, both diluted 500-fold in TBS-T. After washing 3 times with TBS-T, the membrane was reacted with anti-rabbit IgG coupled with peroxidase diluted 1000-fold in TBS-T for 2 h at room temperature. After extensive washing, the membranes were maintained in 0.08% 4-chloro-1-naphthol in TBS containing 0.1% hydrogen

Table 1Assays conditions and methods used in the determination of hydrolases in *P. americana* males.

Enzyme	[Substrate]	Tested pH	References
Amylase	0.5% starch with 10 mM NaCl	4.5–6.0	Noelting and Bernfeld (1948)
Aminopeptidase	1 mM Leu-pNa	7.0	Erlanger et al. (1961)
α -Glucosidase	5 mM pNp- α -glucoside	6.0	Terra et al. (1979)
β -Glucosidase	5 mM pNp- β -glucoside	6.0	Terra et al. (1979)
β -Galactosidase	5 mM pNp- β -galactoside	6.0	Terra et al. (1979)
Cellobiase	7 mM cellobiose	6.0	Dahlqvist (1968)
Chymotrypsin	10 μ M Suc-AAF-MCA	7.0–8.0	Alves et al. (1996)
Cysteine proteinase	1 μ M ϵ -NH ₂ -csb-C-MCA	6.0	Alves et al. (1996)
Maltase	7 mM maltose	6.0	Dahlqvist (1968)
Trehalase	7 mM trehalose	6.0	Dahlqvist (1968)
Trypsin	10 μ M Z-FR-MCA	7.0–8.0	Alves et al. (1996)

Assays were accomplished at 30 °C at the specified pH values. The buffers used were: 50 mM citrate–phosphate (pH 4.0–7.0) or 50 mM sodium phosphate (pH 7.0–8.0) for α -glucosidase, β -glucosidase, β -galactosidase, cellobiase, and trehalase. For amylase, trypsin, maltase and aminopeptidase assays were performed in 100 mM sodium acetate (pH 3.0–4.0), 100 mM piperazine (pH 4.0–6.0), 100 mM HEPES (pH 6.0–8.0), 100 mM TAPS (pH 8.0–9.0) and 100 mM CAPS (pH 9.0–10.0). Assays were carried out at different periods of time and initial velocities were calculated. All assays were performed at conditions that they were proportional to time and to protein concentration. One unit of enzyme is the amount of enzyme that cleaves one μ mol of substrate or (bond) per min. In amylase assays, NaCl was added to become 10 mM.

peroxide until gray bands appeared where antigens were recognized. The reagent was prepared immediately before use by the addition of one volume of 0.5% chloro-1-naphthol in methanol to five volumes of TBS with hydrogen peroxide.

2.5. Pyrosequencing and annotation of a cDNA midgut library from *P. americana*

mRNA was isolated from total RNA with Dynabeads mRNA purification kit (Life technologies) according to manufacturer's instructions. Total RNA and mRNA quality was checked with Agilent 2100 Bioanalyzer system with usual procedures. cDNA library construction and pyrosequencing were performed following the procedures of Roche using a platform 454 Genome Sequencer FLX (454 Life Sciences/Roche). The resulting files (.sff) containing reads were processed by GS De Novo Assembler (Newbler) (Margulis et al., 2005), forming contigs. Further manipulation of contigs and their annotation was performed using dCAS software (<http://exon.niaid.nih.gov>). Pyrosequencing of the cDNA library from midgut tissues generated 270613 reads with average size of 510.6 bp. The relative number of reads indicated the expression level of each gene in the midgut.

The selected sequences were manually curated by multiple sequence alignment (Bioedit version 7.1.3.0, Hall, 1999) with reference sequences. The following reference sequences were used: aminopeptidases pertaining to family M1, M17 and M24 were aligned respectively with human aminopeptidase N (UniProt P15144), leucyl aminopeptidase from *Bos taurus* (UniProt 00727) and aminopeptidase P from *Escherichia coli* (UniProt P15034); α -amylases were aligned with that of *T. molitor* (TM_Amy, accession number P56634.1); α -glucosidases from family 31 were aligned with α -glucosidase II from *Bombyx mori* (BAM78681.1) and those from family 13 with the enzyme from *Culex pipiens* (AAL05443.1); β -glucosidases with *Spodoptera frugiperda* family 1 β -glucosidase (AAC06038.1); β -galactosidases with *Homo sapiens* family 35 β -galactosidase (AAA518119.1); chitinases with *T. molitor* digestive chitinase (AAP92418.1) and *Manduca sexta* molting fluid chitinase (AAB53952.1); chymotrypsin with bovine chymotrypsin (UniProt P00766) and, finally, trypsin with *Helicoverpa armigera* trypsin (ADA83702.1).

2.6. Expression analysis by RT-PCR

Adults immobilized on ice were dissected using gloves, sterile forceps and glassware treated with diethylpyrocarbonate. Salivary

glands, Malpighian tubules, fat bodies, and midgut sections were separated and total RNA was extracted using Trizol (Invitrogen), following the manufacturer's instructions based on Chomczynski and Sacchi (1987). Extracted RNA was DNase treated (DNase I, Invitrogen), before used as template to synthesize a cDNA library correspondent to mRNAs, by using oligo-dT as primer and the reverse transcriptase from the Superscript kit (Invitrogen). Polymerase chain reaction (PCR) was performed using TAQ DNA Polymerase (Invitrogen) in a reaction buffer containing 1.5 mM MgCl₂. PCR conditions were 26 cycles (log phase) of 30 s at 94 °C (denaturation), 30 s at 64 °C (annealing) and 1 min at 72 °C (synthesis) using the resulting cDNA as templates. The primers used are listed in Supplementary Table 1.

2.7. Computational analysis

The prediction of sequence features was done as follows: signalP (Emanuelsson et al., 2007), NetOGlyc (Julenius et al., 2005), big-PI Predictor (Eisenhaber et al., 2000), NetNGlyc (Gupta et al., 2004). The presence of transmembrane regions was analyzed by (TMHMM Server v. 2.0, <http://www.cbs.dtu.dk/services/TMHMM-2.0>; HMMTOP (Tusnady and Simon, 2001) and SOSUI (<http://bp.nuap.nagoya-u.ac.jp/sosui>). All the analysis can be done from the EXPASY Proteomics Server (<http://expasy.org/>).

2.8. Cloning and sequencing the cDNAs coding for the major digestive enzymes

RNA was extracted from the midgut (or salivary gland in the case of salivary amylase) and treated as described in Section 2.6. RNA quality was checked with Agilent 2100 Bioanalyzer system with appropriate procedures, before the cDNA synthesis by reverse transcriptase from the Superscript kit (Invitrogen). The design of the primers was based on our transcriptome, except for the salivary amylase that was amplified using degenerated primers (Supplementary Table 1).

RT-PCR for each selected sequence was performed as described above and the resulting amplified bands were purified and cloned in pGEM-T Easy (Promega), following the manufacturer's instructions. Plasmids were added to a *E. coli* XL1-Blue aliquots, and plated in LB-Agar containing 50 μ g/mL Ampicilin, 0.5 mM IPTG and 80 μ g/mL X-Gal after heat shock transformation (90 s at 42 °C, 3 min in ice and 1 h recuperation at 37 °C with LB). White colonies were grown in LB containing 50 μ g/mL Ampicilin, and plasmids were purified using Wizard Plus MiniPrep DNA Purification System

(Promega), as described in manual, and sequenced with the DNA kit Big Dye Terminator Cycle Sequencing (PE Applied Biosystems).

The Per a 10 allergen gene, shown here to be the major digestive trypsin, was amplified using a specific primer ([Supplementary Table 1](#)).

2.9. Purification of the major midgut trypsin and microsequencing

The major digestive trypsin activity was purified by ion-exchange chromatography as described by [Lopes and Terra \(2003\)](#), and trypsin purity was ascertained by SDS–PAGE. N-terminal microsequencing of an internal peptide of trypsin (originated by autolysis on storage) was performed by Edman degradation using a PROCISE 491 peptide sequencer (Applied Biosystems) by Dr. Izaura Y. Hirata at the Department of Biophysics from the Federal University of São Paulo (UNIFESP).

Table 2

Volume of diverse gut regions of *P. americana*.

Region	Volume (μL)	Relative volume (%)
Foregut (Crop)	104 ± 26	58.2
Midgut caeca	15 ± 4	8.5
Midgut	16 ± 2	9.2
Ileum	34 ± 9	18.7
Rectum	10 ± 2	5.5

The figures are means and SEM corresponding to determinations carried out in 3 feeding males. The determinations were calculated from digital photos of the dissected guts. The volume of the epithelial layer was disregarded.

2.10. Electron microscopy

For ultrastructural analysis at the transmission electron microscope, the midgut was fixed in 3% glutaraldehyde solution in 0.1 M

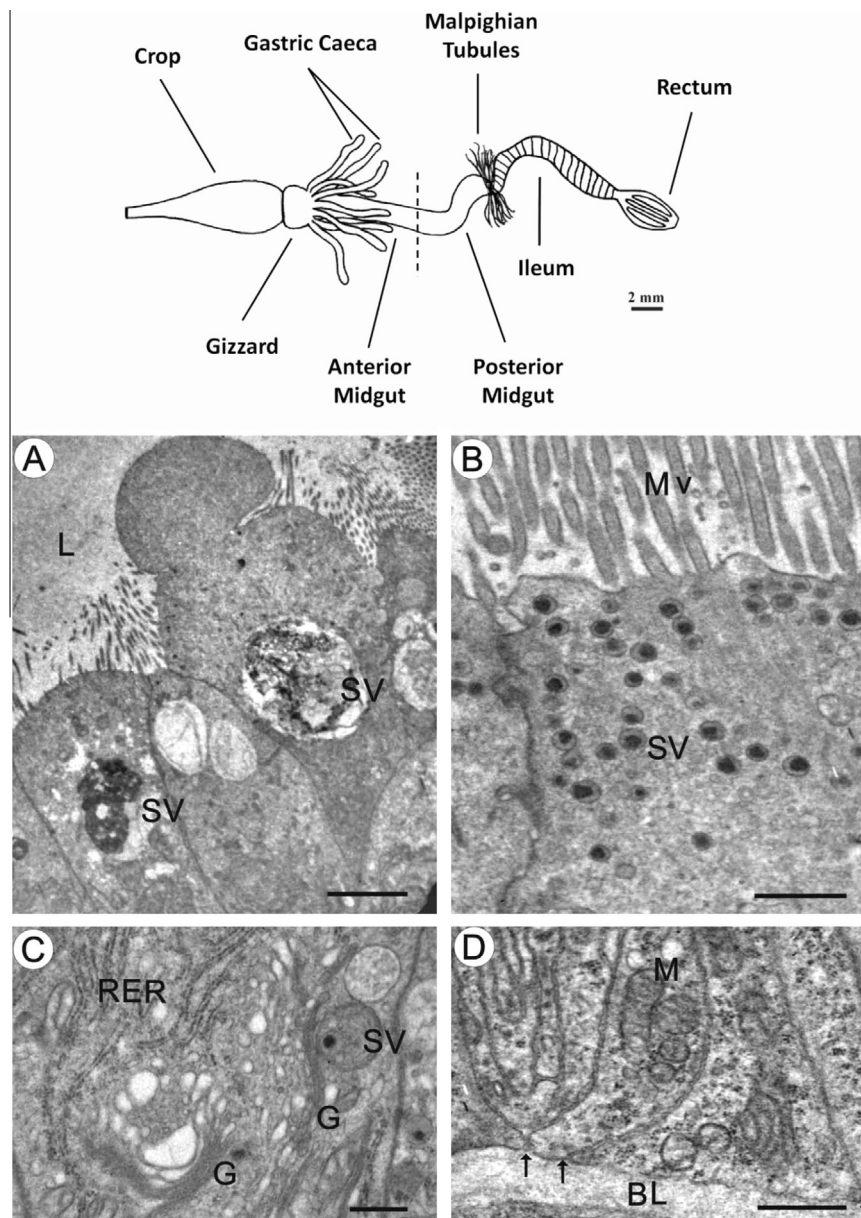


Fig. 1. Diagrammatic representation of *P. americana* gut and ultrastructural details of midgut caeca cells. Foregut includes the crop and gizzard. Midgut limits are the gizzard and the Malpighian tubules. Hindgut includes the ileum and rectum. Note in (D) apertures in basal membrane infoldings (arrows). BL, Basal lamina; G, Golgi complex; L, lumen; M, mitochondria; Mv, microvilli; RER, rough endoplasmic reticulum; SV, secretory vesicle. Bar = 5 μm (A); μm (B, D); 0.5 μm (C).

sodium cacodylate buffer pH 7.4, postfixed in 1% osmium tetroxide and embedded in Spurr resin. Samples were examined in a Zeiss EM 900 electron microscope as detailed elsewhere (Silva et al., 1995).

For immunolabeling of trypsin and amylase at the ultrastructural level, midgut pieces were fixed (paraformaldehyde–glutaraldehyde), embedded in L.R. White acrylic resin, incubated respectively with the primary *M. domestica* anti-trypsin (Jordão et al., 1996) and *T. molitor* anti-amylase (Cristofolletti et al., 2001) and secondary (goat anti-rabbit IgG coupled to 15 nm gold) antibodies and examined in a Zeiss EM 900 electron microscope as detailed elsewhere (Silva et al., 1995). As controls, sections were incubated with pre-immune serum under the same conditions.

3. Results

3.1. Morphology and pH of *P. americana* gut

The general morphology of the gut of *P. americana* and the major ultrastructural features of the midgut cells have been described before (Bignell, 1981; Ma et al., 2009), and therefore, only a brief description will be given, emphasizing only the features important for this work and that have not been reported before.

The gut of *P. americana* presents in the foregut a large crop, comprising 58% of total gut volume, and a proventriculus (gizzard). The midgut is divided in midgut caeca and anterior and posterior

midgut, both presenting 9% of total midgut volume. The hindgut is divided in ileum and rectum (Fig. 1). The volume of each individual region is shown in Table 2. The results are similar to those of Ma et al. (2009).

Gut luminal contents are slightly acidic along the whole gut. The pH values ($n = 10$) of section contents are: foregut, 5.6 ± 0.4 ; midgut caeca, 5.8 ± 0.3 ; anterior midgut, 5.8 ± 0.3 ; posterior midgut, 6.0 ± 0.3 ; ileum, 6.1 ± 0.3 . These figures agree with those obtained by Bignell (1981).

Midgut ultrastructural analysis showed columnar cells presenting several infoldings of the basal plasma membrane that form extracellular compartments (Fig. 1D). The nuclei are central in the cells and Golgi areas and stacks of rough endoplasmic reticulum can be seen around them (Fig. 1C). The apical portion of these cells shows a dense array of microvilli. The regenerative cells are basal-located and are undifferentiated and lack infoldings of the plasma membrane. They have reduced cytoplasm and the nuclei occupy about half of the cell content. Gastric caeca (Fig. 1A) and anterior midgut cells are characterized by the presence of large vesicles and apical extrusions, which are almost absent from posterior midgut cells. Cells in caeca (Fig. 1B) and along the midgut also have electron dense secretory granules.

3.2. Fractionation and properties of proteinases and carbohydrases active in *P. americana* midgut

Midgut proteinases of *P. americana* were resolved by ion-exchange chromatography. There are six major peaks of proteolytic

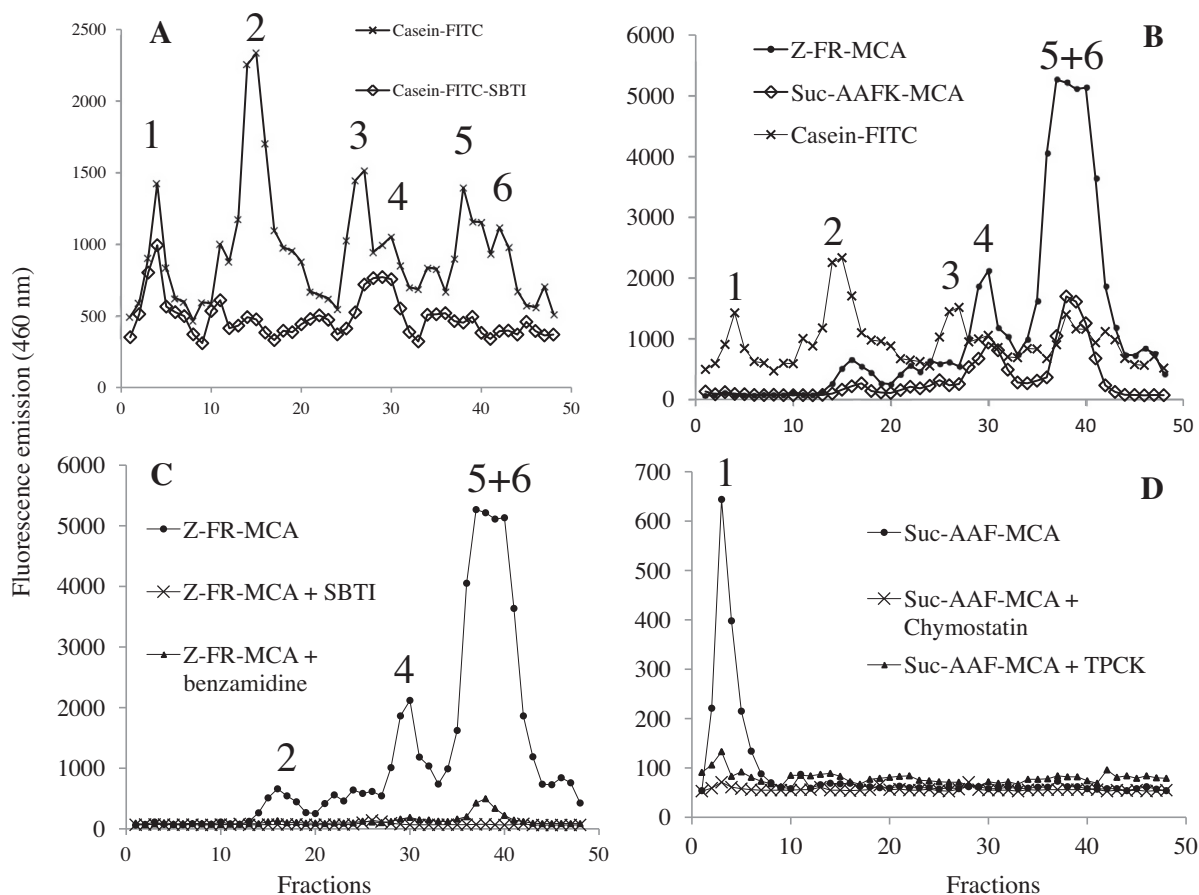


Fig. 2. Ion-exchange chromatography of proteins of midgut homogenates of *P. americana*. The same fractions were assayed in the presence of different fluorescent substrates and inhibitors. Casein-FITC, a fluorescent general substrate for proteinases; SBTI, soybean trypsin inhibitor; Z-FR-MCA, carbobenzoxy-Phe-Arg-methylcoumarin and Suc-AAFK-MCA, succinyl-Ala-Ala-Phe-Lys-MCA, specific substrates for trypsin; Suc-AAF-MCA, succinyl-Ala-Ala-Phe-MCA, specific substrate for chymotrypsin; benzamidine, inhibitor of trypsin; chymostatin, inhibitor of chymotrypsin; TPCK, tosyl phenylalanine chloromethyl ketone, inhibitor of chymotrypsin. All assays were performed in 50 mM sodium phosphate buffer at pH 8.0. The results displayed are typical from a total of 3 independent experiments.

activity against casein and peaks 2, 5 and 6 were completely inhibited by SBTI, indicating the presence of serine proteinases (Fig. 2A). Indeed, peaks 2, 4, 5 and 6 have trypsin activity on synthetic substrates (Z-FR-MCA and Suc-AAFK-MCA, Fig. 2B), and the presence of benzamidine or SBTI inhibited the activity of these peaks when assayed with Z-FR-MCA, confirming that these 4 peaks correspond to trypsin activity (Fig. 2C). These assays also demonstrated that the relative trypsin activity on natural substrates is different from that observed for synthetic substrates. For example, most trypsin activity on casein-FITC is observed in peak 2, whereas on Z-FR-MCA, most activity is found in peak 5 + 6. Peak 1 corresponds to chymotrypsin activity, because it is active on the specific synthetic chymotrypsin substrate, Suc-AAPF-MCA, and it was inhibited by the specific inhibitor chymostatin and by the enzyme chymotrypsin modifying agent TPCK (Fig. 2D). Finally, peak 3 corresponds to an yet unidentified proteinase.

Carbohydrase activities of *P. americana* midgut homogenates were resolved by gel electrophoresis into one peak of amylase and trehalase, and two peaks of α -glycosidase and 3 peaks of β -glucosidase activities (Fig. 3).

The effect of pH on activity was studied for the enzymes resolved by chromatography and electrophoresis, except in the case of aminopeptidase, for which midgut homogenates were used. Trypsin and aminopeptidase present an alkaline pH optimum (pH optima 8.6 and 7.1, respectively), whereas the carbohydrases, excepting trehalase (4.8) have pH optima around 6.0: amylase (6.6), α -glycosidase (6.1), maltase (5.3), β -glycosidase (5.3), cellobiase (6.3), β -galactosidase (5.4) (Fig. 4).

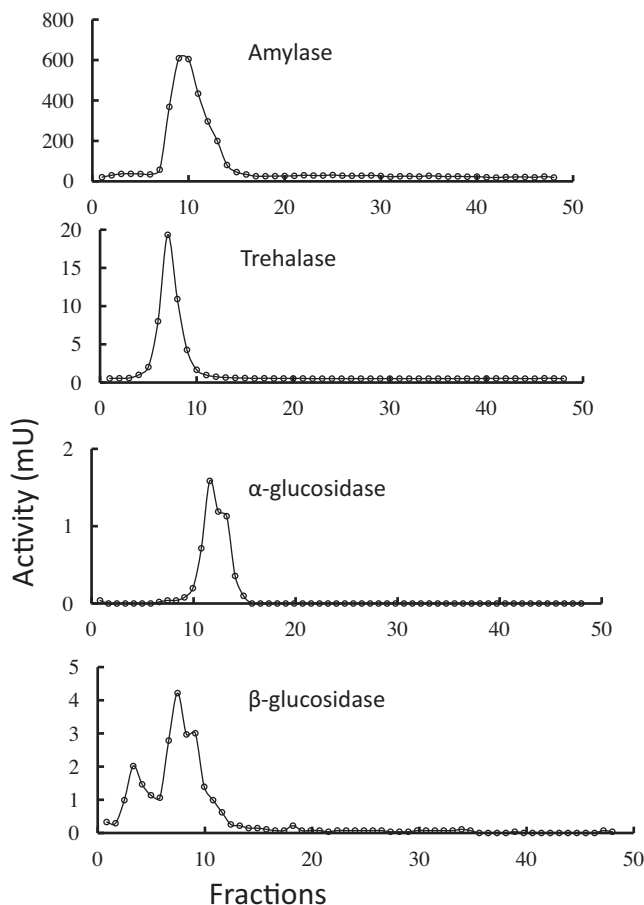


Fig. 3. Resolution of midgut carbohydrases of *P. americana* by polyacrylamide gel (7.5%) electrophoresis. After the run the gel was sliced, maintained in 50 mM citrate-phosphate buffer pH 6 for 3 h at 4 °C and then the samples were assayed with the different substrates, as detailed in Section 2.

3.3. Amino acid sequences of the major *P. americana* digestive enzymes

The *P. americana* midgut transcriptome was at first automatically annotated. Then, the annotation of selected sequences was manually curated using multiple sequence alignments (Bioedit Version 7.1.3.0), as detailed in Section 2. The following sequences were chosen for manual curation: aminopeptidase, α -amylase, α -glucosidase, β -glucosidase, β -galactosidase, chitinase, chymotrypsin, and trypsin. Putative non-functional enzymes were intentionally discarded, because they lacked one or more of their characteristic features. These features are identified below on referring to each enzyme. No β -glucanase was found in the midgut transcriptome. The features of the contigs are summarized in Supplementary Table 2 and the corresponding FASTA sequences are shown in Supplementary Table 3.

At first, 12 coding sequences for aminopeptidases were recovered from the transcriptome (Supplementary Tables 2 and 3) and the predicted sequences were manually curated. Sequences preliminary annotated as pertaining to family M1, M24 and some from family 17 of proteolytic enzymes (Rawlings et al., 2014) were discarded, because they lacked catalytic and/or zinc-binding residues identified by alignment. Only two sequences of leucyl aminopeptidase from family 17 (contig 280 and 2898) have all the important amino acid residues. Their features are described in Table 3. Contig 280 was partially cloned and sequenced (KJ632666.1). Both aminopeptidase sequences have no predicted signal peptide and hence they are unlikely to be exposed to the N-glycosylation machinery and thus may not be glycosylated even though they contain potential motifs.

There is a single sequence of amylase in the midgut and a different one in the salivary glands, as judged by differences in their sequences (Fig. 5) and transcription sites (Fig. 6). This means that there are orthology relationships between gut vs salivary amylases. Both amylases present all essential residues (Fig. 5) and the cDNA coding sequence of the midgut amylase was cloned and sequenced (other details in Table 3). The single cDNA coding for β -galactosidase (details in Table 3) was also cloned and sequenced. It pertains to family 35 of glycosyl hydrolases (Henrissat, 1991).

Eleven sequences of α -glucosidase were found in the transcriptome (Supplementary Tables 2 and 3). Most sequences (8) were very short and lowly expressed. Three sequences (contigs 1753, 2792, and 2829) are complete and highly expressed, but contig 2792 lacks one of the catalytic residues, indicating it is not a α -glucosidase. Contigs 1753 (family 31 of glycosyl hydrolases) and 2829 (family 13) were molecularly cloned and sequenced and their sequences deposited and featured in Table 3.

The transcriptome presents 5 β -glucosidase sequences (family 1 of glycosyl hydrolases) (Supplementary Tables 2 and 3) from which two (contigs 1701 and 1698) were complete or nearly complete and highly expressed. The cDNA coding for the contig 1701 was cloned and sequenced and its features are described in Table 3.

There are 3 chitinase and 2 chymotrypsin sequences in the transcriptome (Supplementary Tables 2 and 3). The chitinase sequences lack mucin and chitin-binding domains (CBD) and two of them (contigs 3012 and 1762) are highly expressed. Their coding cDNA sequences were cloned, sequenced and featured in Table 3. One of the chymotrypsin sequences (contig 4045) was complete, molecularly cloned, sequenced and featured in Table 3.

A large number of trypsin sequences (43 contigs) were recovered from the transcriptome, from which 8 sequences were complete (Supplementary Tables 2 and 3). Nevertheless, only 4 of the complete sequences were highly expressed (contigs 725, 2967, 3624, 3823). The most expressed trypsin (contig 725) had its coding cDNA cloned and sequenced, whereas the second most

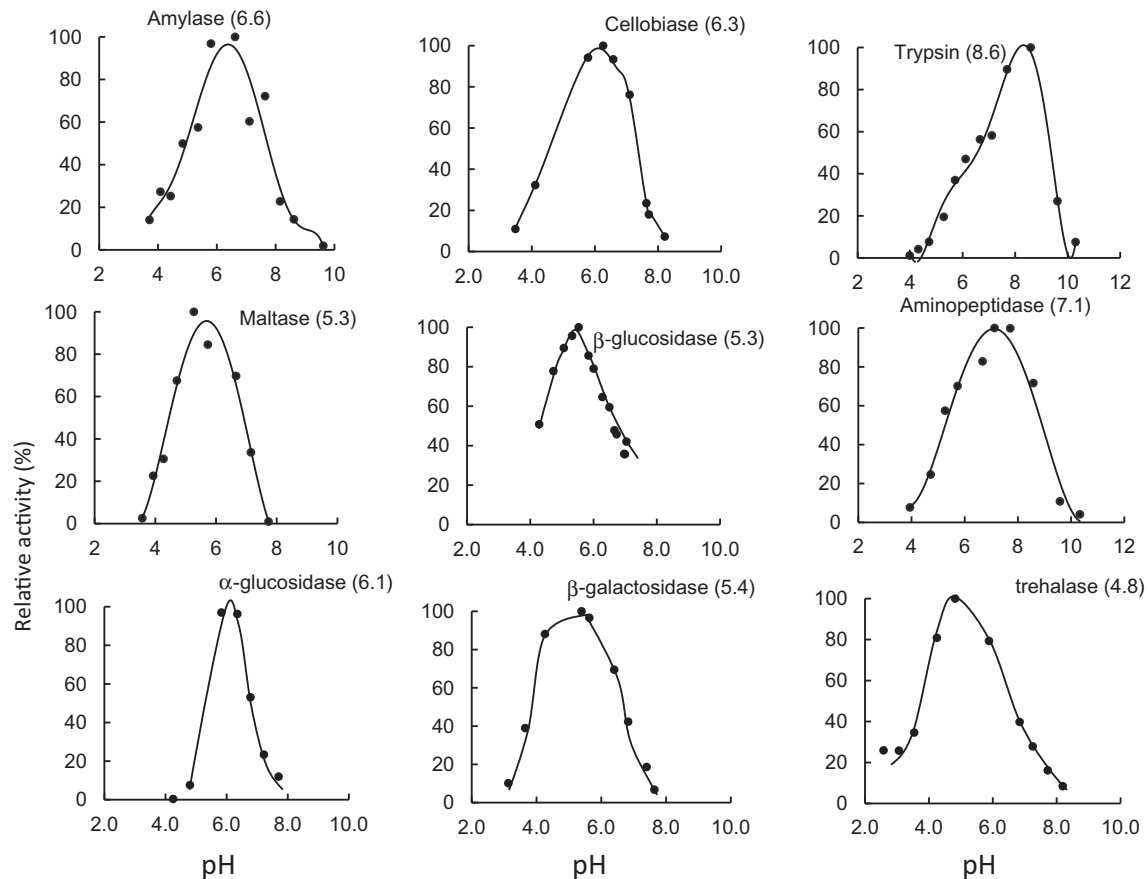


Fig. 4. Effect of pH on the activities of *P. americana* midgut hydrolases. Figures in parentheses are the pH optima. For details see Section 2.

Table 3
Major *P. americana* midgut digestive enzymes.

Contig	GenBank number	Predicted protein	Reads	SP	TMH	GPI	N-Gly	O-Gly	Observations
280	KJ632666.1	Aminopeptidase	43	N	N	N	Y	Y	Partial
2898		Aminopeptidase	30	N	N	N	Y	Y	Complete
2849	KJ576831	Amylase	35	Y	N	N	Y	N	Partial ^a
2829	KJ576830	α-Glucosidase	100	Y	N	N	Y	Y	Complete
1753	KJ576837	α-Glucosidase	283	Y	N	N	Y	Y	Partial ^a
2736	KJ576834	β-Galactosidase	82	Y	N	N	Y	N	Complete
1701	KJ576835	β-Glucosidase	229	Y	N	N	Y	Y	Complete
1698		β-Glucosidase	107	Y	–	–	Y	–	Partial
3012	KJ576833	Chitinase	248	Y	N	N	Y	Y	Complete
1762	KJ576832	Chitinase	663	Y	N	N	Y	Y	Complete
4045	KJ576836	Chymotrypsin	29	Y	N	N	N	Y	Partial ^a
725	KJ569093	Trypsin	2502	Y	N	N	N	N	Complete
2967		Trypsin	178	Y	N	N	N	Y	Complete
3624		Trypsin	367	Y	N	N	Y	Y	Complete
3823	AY792954	Trypsin	586	Y	N	N	N	Y	Complete

SP, signal peptide; TMH, transmembrane helix; GPI, glycosyl phosphatidylinositol anchor; N-gly, N-glycosylation predicted; o-gly, o-glycosylation predicted; Y, found; N, not found; –, the length of the sequence does not permit the search. Except for AY792954, all sequences deposited in GenBank originate from this research. The sequence KJ632666.1 is a fragment, whereas the others that were deposited are complete or at least have complete mature sequences (see note below). See Section 2 for prediction algorithms.

^a These sequences lack a few amino acid residues at the N-terminus, although a signal peptide is predicted. This means that those enzymes have complete mature sequences.

expressed (contig 3823) corresponds to the Per a 10 allergen, with a previously published sequence (Sudha et al., 2008) (Table 3 for details). Both contigs 725 and 3823 are expressed mostly in the gastric caeca and apparently in fewer amounts also along the midgut (Fig. 6).

3.4. Purification of the major trypsin activity

The major trypsin activity with the synthetic substrate Z-FR-MCA was purified to homogeneity, as described by Lopes and Terra (2003), and N-terminal microsequencing of an internal

TM_Amy	1	OKDANEASGRNSIVHLFEWKNDIADECDFLQPGCGGVCI	SPPNEYLVAD	--	CRPWWD	58
PA_Amy_MG	1	OKDPNVWPDRSIVHLFEWREDDIADECDFLAPHGYAGVCT	SPVBEYLVIT	SVDRPWWD		60
PA_Amy_SG	1	-----	-----	-----	RPWWD	5
TM_Amy	59	RYQPVSYIINIRSGDESAFTDMTRRCNDAGVRIYVDVINHMTG	--	MNCVGTSGSSADHD		116
PA_Amy_MG	61	RYQPVSYFIISRSRSGDEIAFRDMVRRCNAGVRIYVDVLNHNMTGNWDNAVGT	CGSTADTY			120
PA_Amy_SG	6	RYQIVSYNIYSRSGDEIAFRDVTVRRCNVGTRIYVDVVLNHNMTGNWDNARGAGTADTY				65
TM_Amy	117	GMNYPAPVPCSGDFHSPOEVNNYODADNVRNCELVGLRDLNQGSDYVRCLIDYANHEMID				176
PA_Amy_MG	121	NYSYPCVPEDEHSEFHPYCIINDYNSPETVRNCELSGLHDLQGSODYVRCKLIDEINRLVD				180
PA_Amy_SG	66	NLOYPAVPYPCGDFHSCSITNYSIDANVRNCELVGLHDLNQGSDYVRCKIIEYLNRLVD				125
TM_Amy	177	LGVAGERVDAAKEMSPCDLSVIESGLKNIINTDYGFAICARPF	IYQEVII	LCGEALISKNEY		236
PA_Amy_MG	181	AGVAGERVDAAKEVWPADLEYIYGVVNNINTDYGFAICSRPF	IYQEVIII	QNDVSKFEY		240
PA_Amy_SG	126	CGIAGERVDAAKHWPADLEYIYNQVNDLSTDHGFASGIRPF	IYQEVII	-----		174
TM_Amy	237	TGFGCVLEFQSGVSLGNATCGGNOIKNLANWCPDWGLLEG	LDAVVEVDNHDNCR	TGGSOI		296
PA_Amy_MG	241	NSFCRVTEEKEDCMCYAFGGQYPLAALHTEDPINWLLDS	GDALVEVDNHDNCR	CONNAI		300
PA_Amy_SG	174	-----	-----	-----		174
TM_Amy	297	LTYKNPKPYKATAFEMLAHPYCTTRIMSSIEDTLINQGPPO	DGSGNLI	SPGINDNTCSN		356
PA_Amy_MG	301	LTYKTPKNYKAGISFELAHYCYPRVMSSEDEAHQGPPO	SDKNIL	HVTTNELGSGCN		360
PA_Amy_SG	174	-----	-----	-----		174
TM_Amy	357	QYVCEHRWRQVYQVVGERNAVEGTQVDNWSNDENQIAFS	RSGSGEVAET	NGGDINONL		415
PA_Amy_MG	361	QWVCEHRWHSAAVVRERNVVRGTETTNWWDNGNHQIG	DARGDRGEVAEV	VNDNLKCTI		420
PA_Amy_SG	174	-----	-----	-----		174
TM_Amy	416	NTGLPAGTYCDVISGELSGCSCTGKSVTVGONGSADIS	LGSAEDDGVAIEVNAKI			471
PA_Amy_MG	421	CTGLPAGTYCDVISGCKNGSSCTGKTIIVNELGTAYIEIL	ISDDDGVAIEHLEKI			476
PA_Amy_SG	174	-----	-----	-----		174

Fig. 5. Multiple sequence alignments of amylases from *P. americana* (PA) and *Tenebrio molitor* (TM_Amy, accession number P56634.1). Amy, amylase; MG, midgut; SG, salivary glands. Catalytic (*), calcium-binding (+), and chloride-binding (#) residues are marked above sequences.

sequence of it revealed the sequence VSPAFSYGTG, which matched only the sequence of contig 3823 (Fig. 7). This contig was annotated as trypsin and corresponds to the Per a 10 allergen of *P. americana* (accession AY792954). As this allergen was shown to have trypsin activity (Sudha et al., 2008), the sequence of the purified trypsin should correspond to that of the allergen, despite the short sequence used in its identification.

3.5. Protease and carbohydrase activities along the *P. americana* gut and their excretory rates

Amylase, trypsin, maltase and aminopeptidase activities were determined along the gut (Fig. 8). Amylase and maltase activities are mainly (56% and 59%, respectively) present in the crop, indicating that this is the place of most carbohydrate digestion. Amylase activity was also found in the soluble fraction of caeca (18%) and in midgut contents (26%).

Trypsin activity is mainly present in the crop (62%), whereas the remaining activity is distributed along midgut tissue soluble fractions and contents (32%). Aminopeptidase activity is equally distributed along midgut regions and a major activity (50%) was found in gut contents, besides the presence of activities in soluble (29%) and in membrane fractions (21%) of midgut tissues.

The activities listed in Table 4 were determined before and after midgut contents were maintained at 25 °C in pH 6 for 8 h, the time the food bolus usually takes to go through the *P. americana* gut. The results showed that the enzymes are completely stable in those conditions, which approximate those found in *P. americana* midgut. Actually, the midgut samples were diluted 5.6 times in relation to *in vivo* concentrations. Digestive enzymes in midgut samples at higher dilutions were unstable. RT-PCR data (E.A. Machado, personal communication) showed that the enzymes listed in Table 4 are not expressed in foregut and hindgut. Thus, hindgut activities, as well as gut volume measurements of *P. americana*, were used to calculate enzymatic excretory rates. It was observed that the

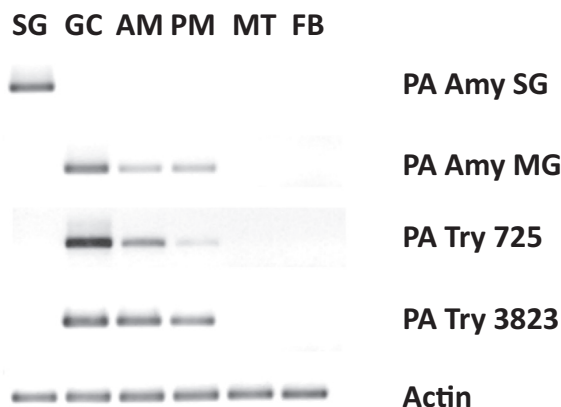


Fig. 6. Expression patterns of salivary (Amy SG) and midgut (Amy MG) amylase and two trypsins (contigs 725 and 3823) in different tissues, as shown by reverse transcription-PCR (RT-PCR). SG, salivary glands; GC, gastric caeca; AM, anterior midgut; PM, posterior midgut; MT, Malpighian tubules; FB, fat body. The number and amplification cycles were chosen after several trials so that the amplification was in log phase and resulting in a clearly visible amplified band.

excretory rates for all assayed enzymes (except amylase) was equal to or less than 15% in both feeding and starving males (Table 4).

3.6. Immunocytolocalization of trypsin and amylase along the *P. americana* gut

Western blots using heterologous antibodies recognized a single band of trypsin and amylase present in the gut (combined protein extract of crop, caeca and midgut) of *P. americana* (Fig. 9).

Trypsin was immunolocalized in caeca cells inside large vesicles, originating from the Golgi areas (Fig. 10A and B), and amylase inside small vesicles from the same cells (Fig. 10C and D). Trypsin and amylase were also found in small and large vesicles, respectively (not shown). Large vesicles are presumed to be released along the apical cytoplasm now lacking visible structures (apocrine secretion), whereas the contents of the small vesicles are supposed to be discharged by exocytosis. Data for anterior midgut is similar to those described for caeca (not shown).

4. Discussion

4.1. Digestive enzyme activities in *P. americana* guts and their putative amino acid sequences

The effect of pH on the activity on pNPβGal is the same as that on pNPβGlu (pH optimum 5.4 and 5.3, respectively) and is different from on cellobiose (pH optimum 6.3). Thus, the major activity on

pNPβGlu resolved by electrophoresis should be a β-glucosidase more active on β-galactosides than on β-glucosides. The pH-activity profile with cellobiose as substrate is different from that of the pNPβGlu one, indicating the existence of a β-glucosidase which prefers cellobioses (cellobiase) to aryl β-glucosides. Thus, enzymatic assays suggest that there are at least two β-glucosidase and one β-galactosidase activities. This agrees with the finding in the transcriptome of a sequence coding for a β-galactosidase (contig 2763) from the family 35, differing thus from the two β-glucosidases (contigs 1701 and 1698) from family 1.

Koffi et al. (2012) purified from whole *P. americana* adults a β-glucosidase more active on pNPβGlu than on cellobiose, thus corresponding to one of the activities described here. Nevertheless, the pH optimum they found for the enzyme is 3.6, which differs from the one described here (pH optimum 5.3) and from the *P. americana* midgut luminal pH (5.8–6.1). It is not clear why their results are different from ours.

Enzymological evidence suggests that there are at least 2 α-glucosidases (resolved by gel electrophoresis, Fig. 3): one of which (maltase) prefers maltose as substrate with a pH optimum of 5.3 (Fig. 4), whereas the other prefers aryl α-glucosides (pH optimum 6.1, Fig. 4). These findings parallel those of transcriptome data which reveals two highly transcribed α-glucosidases (contigs 1753 and 2829).

Powning and Irzykiewicz (1963) identified two chitinases by chromatography of *P. americana* midgut contents and purified one of them and named it chitinase I. This enzyme is incapable of splitting chitobiose, splits chitotriose very slowly and higher oligosaccharides and chitin at a considerable faster rate. The enzyme is thus a non-processive endoenzyme. Chitinase I should correspond to contig 1762 (the most expressed chitinase), whereas the other chitinase may correspond to contig 3012 (details in Table 3). The lack of mucin and CBD in both chitinases is in accordance to their presumed role in microbiota digestion without causing harm to the peritrophic membrane. The absence of CBD decreases the ability of the chitinase to binding to chitin, except if in high concentrations, as in dietary fungi, similarly to what was described for *T. molitor* (Genta et al., 2006).

Liquefying and saccharifying amylases perform, respectively, a single or multiple rounds of catalysis each time they attach to the substrate. In other words, liquefying amylases are non-processive, whereas saccharifying amylases are processive enzymes. Preliminary work done in our laboratory suggested that in *P. americana* midgut there are two amylases, one liquefying and the other saccharifying. As only one amylase was found in the midgut transcriptome, the expression of amylase in salivary glands was looked for. A partial amylase sequence, different from the midgut amylase, was obtained from salivary gland samples, in agreement with the preliminary work. RT-PCR data confirmed that the salivary amylase is expressed only in this tissue, whereas

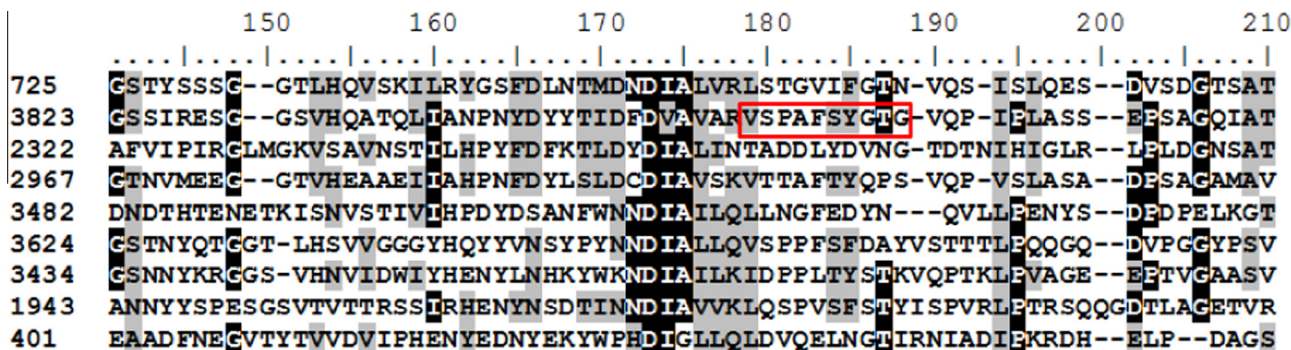


Fig. 7. Multiple sequence alignments of part of the nine most expressed trypsins from *P. americana* midguts. Only contig 3823 has a sequence (boxed) that is matched by a microsequenced N-terminus of an internal peptide from the major trypsin purified from *P. americana* midgut contents.

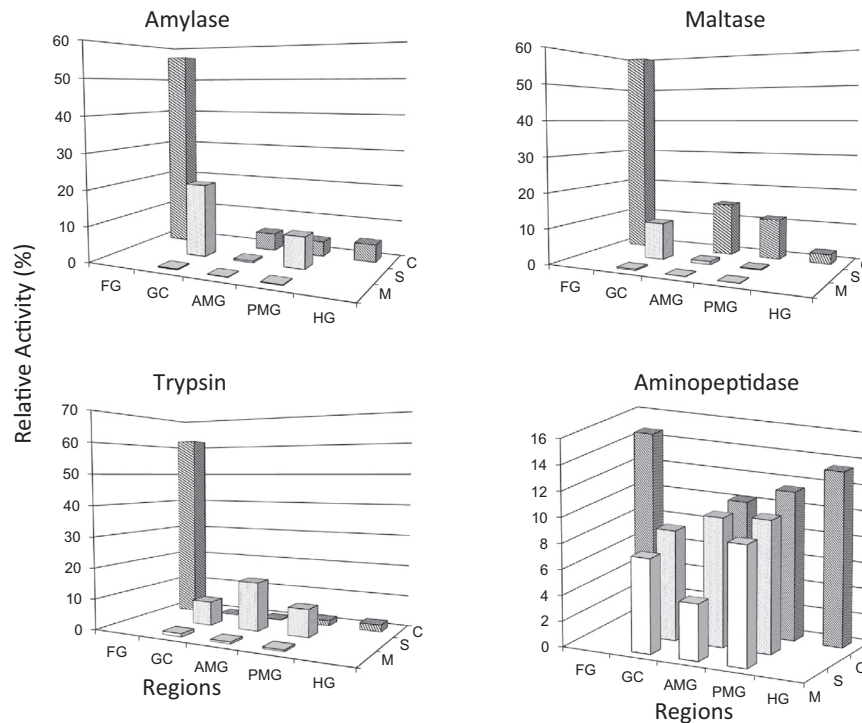


Fig. 8. Distribution of the major hydrolases along the gut of *P. americana*. C, contents of the midgut; M, membrane fraction of tissues; S, soluble fraction of tissues; FG, foregut (crop); GC, gastric (midgut) caeca; AMG, anterior midgut; PMG, posterior midgut; HG, Hindgut. Determinations were performed in 3 different preparations obtained from 5 insects each. SEM were found to be 10–20% of the means. The activities (mU/animal) in whole gut homogenates were: amylase, 13,800; maltase, 20,241; trypsin, 308; aminopeptidase, 34. Amount of proteins in gut sections ($\mu\text{g}/\text{animal}$) were: FG, 3300; AMG, 450; PMG, 270; hindgut, 1410.

Table 4
Digestive enzyme activities and their excretory rates.

	Activity				Excretory rate (% midgut contents)	
	Feeding		Starving		Feeding	Starving
	Midgut	Ileum + colon	Midgut	Ileum + colon		
Amylase	1440	802	2940	800	26.3	12.8
Maltase	6300	620	9093	2630	4.6	13.6
Trypsin	108	7.3	35	11	3.2	14.8

The figures correspond to the means of activities determined along the gut in 3 feeding and 5 starving males. SEM was 5–15% of the means. Starving males are those having empty foreguts. The excretory rate is the ratio of soluble activities in ileum + colon and midgut with correction. The correction was calculated assuming that when the midgut empties, it fills a volume 2.12 smaller than that of ileum + colon. The last assumption is based on a comparison of the volumes of midgut and ileum + colon (Table 2). Other details in Section 2.

the other amylase is expressed only in midguts. Presently, we are trying to identify which of the preliminarily characterized amylases is salivary and which one is intestinal. A guess based on vertebrate findings is that the salivary amylase is liquefying and the intestinal one is saccharifying.

A single chymotrypsin was identified in this paper after chromatographic separation, in agreement with Baumann (1990), who purified to homogeneity also a single chymotrypsin. The most remarkable characteristic of *P. americana* chymotrypsin is a resistance to tosyl-phenylalanyl chloromethyl ketone (TPCK) inactivation (Baumann, 1990). Such TPCK inactivation resistance was shown to be characteristic of lepidopteran chymotrypsins, whereas enzymes from other insects are usually sensitive (Lopes et al., 2009), in spite of the finding of insensitive chymotrypsins among some Coleoptera (Vinokurov et al., 2006). This demands further investigation. Baumann (1990) detected a second (minor) chymotrypsin in electrophoresis under native conditions. It is possible that the chymotrypsin purified by Baumann corresponds to our trypsin detected by chromatography and is coded by contig

4045. The minor chymotrypsin resolved by Baumann may correspond to contig 5314 (Supplementary Table 2).

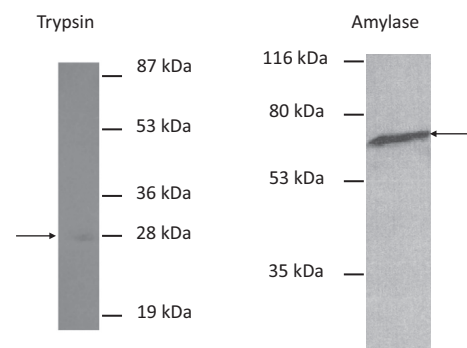


Fig. 9. Western blot using anti-trypsin and anti-amylase serum against *P. americana* gut homogenate after SDS-PAGE.

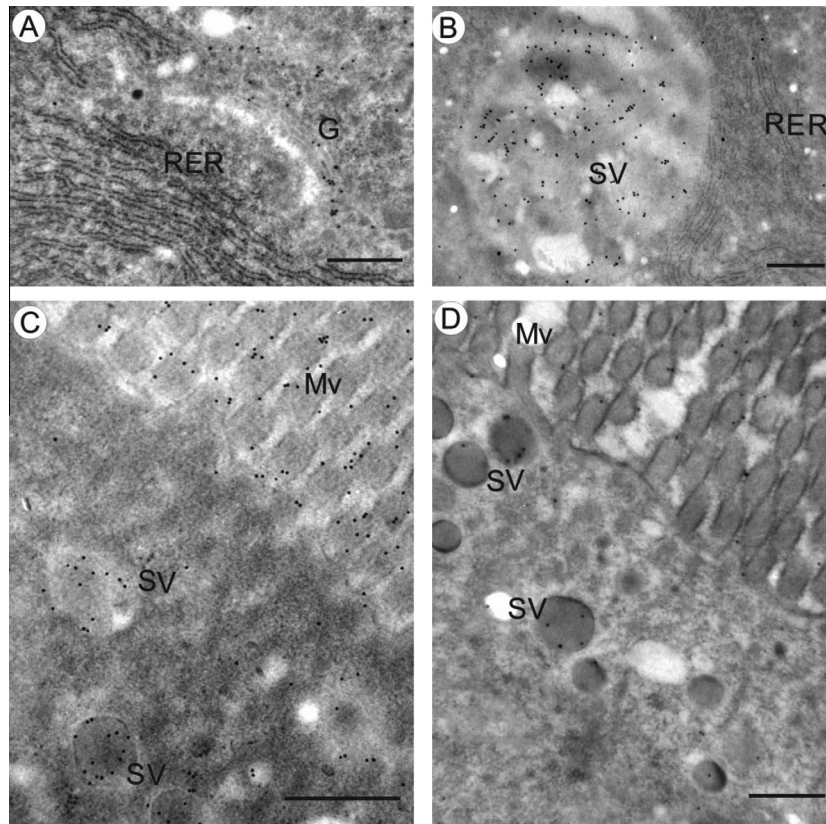


Fig. 10. Immunolocalization of *P. americana* trypsin (panel A and B) and amylase (panel C and D) in caeca cells. Abbreviations as in Fig. 1. Antibody dilutions: 1:500 (A, B, and D); 1:1000 (C). Bars: 0.5 μ m.

The single major trypsin purified to homogeneity from *P. americana* adults midguts by Lopes and Terra (2003) was microsequenced and an internal sequence matches only contig 3823 (Fig. 7), which corresponds to Per a 10 allergen that was shown to be a trypsin (Sudha et al., 2008). This trypsin, as well as the most expressed one (contig 2502) are expressed mostly in midgut caeca and in apparently less amounts along the midgut (RT-PCR data, Fig. 6). The large number of other trypsin sequences identified in *P. americana* transcriptome, from which contigs 725, 2967, 3624, and 3823 are the most expressed, originate small activity peaks in the chromatogram reported in this paper. Thus, apparently they are less translated than expected. Nevertheless, Hivrale et al. (2005) found in *P. americana* collected in natural environments a large number of different trypsins, which may correspond to the other sequences identified in the transcriptome. The difference between our results and those of Hivrale et al. (2005) may result from the different diets used to raise the cockroaches.

The aminopeptidases (contig 280 and 2898) are considered to be digestive enzymes acting in *P. americana* gut lumen. Nevertheless, they have no predicted signal peptide (Table 3), which may imply that they are cellular enzymes. In spite of that, they may be released into the midgut lumen together with the apical cytoplasm, when the cells undergo apocrine secretion (Fig. 10). None of the aminopeptidase sequences showed predicted transmembrane helix or GPI-anchor to correspond to membrane-bound aminopeptidase detected in the *P. americana* midgut (Fig. 8). This demands further research.

4.2. Organization of digestion in *P. americana*

Transit times for the passage of solid food through *P. americana* are typically 8 up to 20 h, although solid food is retained in the gut of starving insects for as long as 100 h (Bignell, 1981).

The crop is the largest gut compartment and enzyme distribution data demonstrated that this is the place where most amylase and α -glycosidase activities occur (around 60% of total activity for each enzyme), indicating the crop importance in carbohydrate digestion. The same is true for protein initial digestion, since it is where most trypsin activity (60%) is found, as seen previously (Lopes and Terra, 2003). The major digestive trypsin was purified and shown to correspond to the Per a 10 allergen, important in human allergies and displaying trypsin activity.

Despite digestive enzymes are found in the crop, this is not the site of their expression and secretion, since it is covered by chitin. Thus, crop digestive enzymes must be produced in other tissues and somehow be transported to the crop. Molecular cloning and RT-PCR data showed that most amylase and trypsin are expressed in gastric caeca cells and in lesser amounts also in anterior midgut. It is supposed that the same is true for the other digestive enzymes, with the exception of salivary enzymes (see below).

It is noteworthy that the pH in *P. americana* gut regions is around 6.0, which is suitable for carbohydrases and aminopeptidases activities, but is low for serine proteinases (trypsins and chymotrypsins). It is not clear the meaning of this.

Two cellulases and three laminarinases that open plant cell walls and lyse fungi cells was previously reported in *P. americana* saliva (Genta et al., 2003). Furthermore, in this paper the cloning of the cDNA coding for a salivary amylase was reported, thus supporting the preliminary studies of Agrawal (1981).

Combining the salivary and midgut enzymes, the cockroach is prepared to start the initial digestion of its detritus-rich diet in the crop. Terminal digestion of carbohydrates (represented by maltase action) starts in the crop and continues at the midgut lumen.

Protein terminal digestion (indicated by aminopeptidase activity) occurs in the caeca and midgut by luminal, glycocalix-associated and membrane-bound enzymes. It is not clear why

the cockroaches use, in the terminal digestion of protein, a cytosolic-like aminopeptidase, instead of the usual aminopeptidase N.

Ingested dyes are concentrated in the caeca, suggesting that this tissue absorbs ions and water. This, combined with indirect evidence that a Malpighian tubule fluid may enter the midgut, strongly suggests that a counter-current flux of fluid occurs in *P. americana* midgut (Dow, 1986). This hypothesis is supported by the low excretory rates (less than 15% of midgut contents at each midgut emptying) observed for 3 enzymes in *P. americana*, with a single exception. This exception is the amylase excretory rate (26.3%) in fed animals, which may be accounted for enzyme adsorption to excessive dietary starch, as shown to be the case for trypsin in excessive dietary protein in *Rhynchosciara americana* and *M. domestica* (Terra and Ferreira, 1994).

In summary, the main features of *P. americana* digestion is their ability to deal with carbohydrates (including those in plant cell walls and fungi cells) in the crop, under the action of salivary and midgut enzymes. Protein initial digestion depends on midgut enzymes and takes place in the crop, whereas the terminal digestion occurs in the midgut under the action of enzymes largely associated with the apical surface of caeca and midgut cells. Due to an ectoperitrophic (between the peritrophic membrane and tissue) counter-current flux of fluid, most digestive enzymes are recovered from the endoperitrophic (inside the peritrophic membrane) space before excretion.

4.3. Comparative aspects of Polyneopteran digestive physiology

The basic plan of the digestive physiology of the Polyneoptera (Dictyoptera, Phasmida and Orthoptera) ancestors is believed to conserve the characteristics of the plan of the Neoptera ancestors. This basic plan was inferred from studies carried out in Orthoptera, both from grasshoppers (Caelifera, see Dow, 1981a,b; Ferreira et al., 1990; Marana et al., 1997) and crickets (Ensifera, see Woodring et al., 2007; Biagio et al., 2009) and from stick bugs (Phasmida, see Monteiro et al., 2014). In Polyneopteran ancestors, the major part of digestion is carried out in the crop by digestive enzymes propelled by antiperistalsis forward from the midgut. Saliva plays a variable role in carbohydrate digestion. After a while, following ingestion, the crop contracts, transferring digestive enzymes and partly digested food into the midgut. The anterior midgut is acid and has high carbohydrase activity, whereas the posterior midgut is alkaline and has high proteinase activity. The food bolus moves backward by peristalsis in the insect midgut. As soon as the polymeric food molecules are digested to become small enough to pass through the peritrophic membrane, they diffuse with the digestive enzymes into the ectoperitrophic space (outside the peritrophic membrane). The enzymes and nutrients are then displaced toward the caeca with a countercurrent flux by secretion of fluid at the Malpighian tubules and its absorption back by cells at the caeca, where final digestion is completed and nutrient absorption occurs. When the insect starts a new meal, the caeca contents are moved into the crop. As a consequence of countercurrent flux, digestive enzymes occur as a decreasing gradient in the midgut, and lower amounts are excreted.

Digestion in grasshoppers is very close to that proposed for the Polyneoptera ancestor, except that part of the terminal digestion of proteins occurs at the surface of midgut caeca cells and a midgut countercurrent flux is lacking in feeding insects (Dow, 1981a,b; Ferreira et al., 1990; Marana et al., 1997). The loss of the midgut countercurrent, and hence of the enzyme recycling, is a consequence of the remarkable salivation during grasshopper feeding, which saturates the water absorption site in the caeca (Dow, 1981a,b). Digestion in crickets differs from that of grasshoppers only in presenting more membrane-bound enzymes, a higher involvement of midgut in protein digestion and in the occurrence

of enzyme recycling in both feeding and starving insects (Biagio et al., 2009).

The plan of the digestive physiology of stick bugs (Phasmida) is alike that of the Polyneoptera ancestor, except that terminal digestion occurs only in the midgut by soluble carbohydrases and membrane-bound peptidases (Monteiro et al., 2014). It is worth mentioning the occurrence of peculiar midgut tubules, apparently involved in midgut alkalization (Monteiro et al., 2014), although they may excrete several different substances (Shelomi and Kimsey, 2013).

Digestion in Dictyoptera (one of the Polyneopteran orders), is expected to derive from the Polyneoptera ancestor. In agreement with that, digestion in cockroaches occurs as described for that ancestor, except that part of the terminal digestion of proteins occurs at the surface of caeca and midgut cells, as glycocalyx-associated or membrane bound enzymes. The saliva of *P. americana* has two cellulases and three laminarinases that open plant cells and lyse fungi cells (Genta et al., 2003). This is in accordance with the detritivorous habit of this insect. The wood roach, *Panesthia cri-bata* (Blaberidae), also has its own cellulase (Scrivener et al., 1989; Tokuda et al., 1999). However, the differentiation of pH along the midgut described for the Polyneoptera ancestor, is not conserved among cockroaches like *P. americana* (Blattidae), but is maintained in others exemplified by the blaberid *N. cinerea* (Elpidina et al., 2001). Another difference observed is the enlargement of hindgut structures (Terra, 1988), noted mainly in wood-feeding cockroaches. These hindgut structures harbor bacteria producing acetate and butyrate from ingested wood or other cellulose-containing materials. Acetate and butyrate are absorbed by the hindgut of all cockroaches, mainly of wood roaches (Terra and Ferreira, 1994). Cellulose digestion may be accomplished in part by bacteria in the hindgut of *P. americana* or by protozoa in *Cryptocercus punctulatus* (Cryptocercidae) (Bignell, 1981).

Secretory mechanisms of digestive enzymes deserve comments. These mechanisms include merocrine secretion through exocytosis, where secretory vesicles fuse with the apical membrane, releasing their contents and apocrine secretion, in which case there is a loss of part of the apical cytoplasm during the release of secretory vesicles. Microapocrine secretion is a special case of apocrine secretion involving the pinching-off of double membrane small vesicles occurring along the microvilli. Apocrine secretion and microapocrine secretion of digestive enzymes in anterior (water-absorbing) midgut cells may be an adaptation to enhance the dispersion of secretory vesicle contents into the midgut. As posterior midgut cells secrete fluid, no problem arises in the dispersion of material released by exocytosis by these cells. Based on known examples, where digestive enzyme secretion was followed by immunocytolocalization techniques, apocrine mechanisms appear to be widespread among less evolved Holometabola insects, whereas more-derived Holometabola use microapocrine mechanisms in water-absorbing tissues (Cristofolletti et al., 2001 and references therein).

In Polyneoptera insects, secretory mechanisms, also observed by immunocytolocalization techniques, present some differences regarding those of Holometabola. Thus, in Phasmida (stick bugs) exocytic processes are found in anterior and middle midgut cells. Microapocrine processes seen in posterior midgut in these insects are not involved in amylase and trypsin secretions, as judged by the fact those enzymes are not immunolocalized there. It is not known what the microapocrine processes are secreting (Monteiro et al., 2014). Orthoptera Caelifera (grasshoppers) may show apocrine mechanisms in anterior midgut (Heinrich and Zebe, 1973), or exocytosis, where the secretory vesicles discharge a compact mass that are dispersed only far from the cells (Marana et al., 1997), thus fulfilling the same function as vesicles released by apocrine/microapocrine processes. Finally, in Orthoptera Ensifera

(crickets) (Biagio et al., 2009) and Dictyoptera (cockroaches) (this paper) both apocrine secretion and exocytosis processes are observed in caeca and anterior midgut.

Acknowledgments

This work was supported by the Brazilian research agencies FAPESP (Temático) and CNPq. We are indebted to W. Caldeira, and M.V. Cruz for technical assistance. F.K. Tamaki and A.B. Dias were, and A.C. Pimentel still is, graduate fellows of FAPESP. C. Ferreira and W.R. Terra are staff members of their respective departments, research fellows of the CNPq, and members of the Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jinsphys.2014.08.007>.

References

- Agrawal, O.P., 1981. Amylase activity in the alimentary tract and salivary glands of *Periplaneta americana* L. *Acta Physiol. Pol.* 32 (1), 29–36.
- Alves, L.C., Almeida, P.C., Franzoni, L., Juliano, L., Juliano, M.A., 1996. Synthesis of N-protected aminoacyl 7-amino-4-methyl-coumarin amide by phosphorus oxychloride and preparation of specific fluorogenic substrates for papain. *Peptide Res.* 9, 92–96.
- Baumann, E., 1990. Isolation and partial characterization of a chymotrypsin-like endoprotease from the cockroach intestinal system. *Insect Biochem.* 20, 761–768.
- Biagio, F.P., Tamaki, F.K., Terra, W.R., Ribeiro, A.F., 2009. Digestive morphophysiology of *Gryllos sigillatus* (Orthoptera: Gryllidae). *J. Insect Physiol.* 55, 1125–1133.
- Bignell, D.E., 1981. Nutrition and digestion. In: Bell, W.J., Adiyodi, K.G. (Eds.), *The American Cockroach*. Chapman and Hall, pp. 57–86.
- Blum, H., Beier, H., Gross, H.J., 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8, 93–99.
- Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of protein – dye binding. *Anal. Biochem.* 72, 248–254.
- Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium triocyanate–phenol–chloroform extraction. *Anal. Biochem.* 162, 156–159.
- Cristofaletti, P.T., Ribeiro, A.F., Terra, W.R., 2001. Apocrine secretion of amylase and endocytosis of trypsin along the midgut of *Tenebrio molitor* larvae. *J. Insect Physiol.* 47, 143–155.
- Dahlqvist, A., 1968. Assay of intestinal disaccharides. *Anal. Biochem.* 22, 99–107.
- Dow, J.A.T., 1981a. Localization and characterization of water uptake from the midgut of the locust, *Schistocerca gregaria*. *J. Exp. Biol.* 93, 269–281.
- Dow, J.A.T., 1981b. Countercurrent flow, water movements and nutrient absorption in the locust midgut. *J. Insect Physiol.* 27, 579–585.
- Dow, J.A.T., 1986. Insect midgut function. *Adv. Insect Physiol.* 19, 187–328.
- Eisenhaber, B., Bork, P., Yuan, Y., Loeffler, G., Eisenhaber, F., 2000. Automated annotation of GPI anchor sites: case study *C. elegans*. *TIBS* 25, 340–341.
- Elpidina, E.N., Vinokurov, K.S., Gromenko, V.A., Rudenskaya, Y.A., Dunaevsky, Y.E., Zhuzhikov, D.P., 2001. Compartmentalization of proteinases and amylases in *Nauphoeta cinerea* midgut. *Arch. Insect Biochem. Physiol.* 48, 206–216.
- Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. *Nat. Protoc.* 2, 953–971.
- Erlanger, B.F., Kokowsky, N., Cohen, W., 1961. The preparation and properties of two new chromogenic substrates of trypsin. *Arch. Biochem. Biophys.* 95, 271–278.
- Ferreira, C., Oliveira, M.C., Terra, W.R., 1990. Compartmentalization of the digestive process in *Abracris flavolineata* (Orthoptera: Acrididae) adults. *Insect Biochem.* 20, 267–274.
- Genta, F.A., Blanes, L., Cristofaletti, P.T., do Lago, C.L., Terra, W.R., Ferreira, C., 2006. Purification, characterization and molecular cloning of the major chitinase from *Tenebrio molitor* larval midgut. *Insect Biochem. Mol. Biol.* 36, 789–800.
- Genta, F.A., Terra, W.R., Ferreira, C., 2003. Action pattern, specificity, lytic activities, and physiological role of five digestive β -glucanases isolated from *Periplaneta americana*. *Insect Biochem. Mol. Biol.* 33, 1085–1097.
- Gupta, R., Jung, E., Brunak, S., 2004. Prediction of N-glycosylation sites in human proteins. <<http://www.cbs.dtu.dk/services/NetNGlyc/>>.
- Hall, T.A., 1999. Bioedit: userfriendly alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41, 95–98.
- Hedrick, J.L., Smith, A.J., 1968. Size and charge isomer separation and estimation of molecular weights of proteins by disc gel electrophoresis. *Arch. Biochem. Biophys.* 1968 (1), 155–164.
- Heinrich, D., Zebe, E., 1973. Zur Feinstruktur der Mitteldarmzellen von *Locusta migratoria* in verschiedenen Phasen der Verdauung. *Cytobiologie* 7, 315–326.
- Henrissat, B., 1991. A classification of glycosyl hydrolases based on amino-acid sequence similarities. *Biochem. J.* 280, 309–316.
- Hivrale, V.K., Chougule, N.P., Chhabda, P.J., Giri, A.P., Kachole, M.S., 2005. Unraveling biochemical properties of cockroach (*Periplaneta americana*) proteinases with a gel X-ray film contact print method. *Comp. Biochem. Physiol. B* 141, 261–266.
- Inward, D., Beccaloni, G., Eggleton, P., 2007. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* 3, 331–335.
- Jordão, B.P., Terra, W.R., Ribeiro, A.F., Lehane, M.J., Ferreira, C., 1996. Trypsin secretion in *Musca domestica* midguts: a biochemical and immunocytochemical study. *Insect Biochem. Mol. Biol.* 26, 337–346.
- Julienius, K., Mølgaard, A., Gupta, R., Brunak, S., 2005. Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiol.* 15, 153–164.
- Koffi, Y.G., Konan, H.K., Kouadio, J.P.E.N., Dabonné, S., Dué, E.A., Kouamé, L.P., 2012. Purification and biochemical characterization of beta-glucosidase from cockroach *Periplaneta americana*. *J. Animal Plant Sci.* 13, 1747–1757.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.
- Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., Maekawa, K., Bandi, C., Noda, H., 2000. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Curr. Biol.* 10, 801–804.
- Lopes, A.R., Sato, P.M., Terra, W.R., 2009. Insect chymotrypsins: chloromethyl ketone inactivation and substrate specificity relative to possible coevolutionary adaptation of insects to plants. *Arch. Insect Biochem. Physiol.* 70, 188–203.
- Lopes, A.R., Terra, W.R., 2003. Purification, properties and substrate specificity of a digestive trypsin from *Periplaneta americana* (Dictyoptera) adults. *Insect Biochem. Mol. Biol.* 33, 407–415.
- Ma, H., Liu, Z.-G., Bao, Y., Ran, P.-X., Zhong, N.-S., 2009. Morphology and three-dimensional reconstruction of the digestive system of *Periplaneta americana*. *J. Med. Entomol.* 46, 165–168.
- Marana, S.R., Ribeiro, A.F., Terra, W.R., Ferreira, C., 1997. Ultrastructure and secretory activity of *Abracris flavolineata* (Orthoptera: Acrididae) midguts. *J. Insect Physiol.* 43, 465–473.
- Margulis, M., Egholm, M., Altman, W.E., Attiya, S., et al., 2005. Genome sequence in microfabricate high-density picoliter reactors. *Nature* 437, 376–380.
- Monteiro, E.C., Tamaki, F.K., Terra, W.R., Ribeiro, A.F., 2014. The digestive system of the “stick bug” *Cladomorphus phyllinus* (Phasmida, Phasmatidae): a morphological, physiological and biochemical analysis. *Arthropod Struct. Dev.* 43, 123–134.
- Noelting, G., Bernfeld, P., 1948. Sur les enzymes amylolytiques. III. La β -amylase: dosage d’activité et contrôle de l’absence d’ α -amylase. On amylases. The beta-amylase: enzyme assay and check for the absence of alpha-amylase. *Helv. Chim. Acta* 31, 286–290.
- Powning, R.F., Irzykiewicz, H., 1963. A chitinase from the gut of the cockroach *Periplaneta americana*. *Nature* 200, 1128.
- Rawlings, N.D., Waller, M., Barrett, A.J., Bateman, A., 2014. Merops: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42, D503–D509.
- Scrivener, A.M., Slaytor, M., Rose, H.A., 1989. Symbiont-independent digestion of cellulose and starch in *Panesthia cribata* Saussure, an Australian wood-feeding cockroach. *J. Insect Physiol.* 35 (935–441).
- Shelomi, M., Kimsey, L.S., 2013. Vital staining of the stick insect digestive system identifies appendices of the midgut as novel system of excretion. *J. Morphol. Early View*. <http://dx.doi.org/10.1002/jmor.20243>.
- Silva, C.P., Ribeiro, A.F., Gulbenkians, S., Terra, W.R., 1995. Organization, origin and function of the outer microvillar (perimicrovillar) membranes of *Dysdercus peruvianus* (Hemiptera) midgut cells. *J. Insect Physiol.* 41, 1093–1103.
- Sudha, V.T., Arora, N., Gaur, S.N., Pasha, S., Singh, B.P., 2008. Identification of a serine protease as a major allergen (Per a 10) of *Periplaneta americana*. *Allergy* 63, 768–776.
- Terra, W.R., 1988. Physiology and biochemistry of insect digestion: an evolutionary perspective. *Braz. J. Med. Biol. Res.* 21, 675–734.
- Terra, W.R., 1990. Evolution of digestive systems of insects. *Annu. Rev. Entomol.* 35 (1881–200).
- Terra, W.R., Ferreira, C., De Bianchi, A.G., 1979. Distribution of digestive enzymes among the endo- and ectoperitrophic spaces and midgut cells of *Rhynchosciara americana* and its physiological significance. *J. Insect Physiol.* 25, 487–494.
- Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: properties, compartmentalization and function. *Comp. Biochem. Physiol.* 109B, 1–62.
- Terra, W.R., Ferreira, C., 2012. Biochemistry and molecular biology of digestion. In: Gilbert, L.I. (Ed.), *Insect Molecular Biology and Biochemistry*. Academic Press, London, pp. 365–418.
- Tokuda, G., Lo, N., Watanabe, H., Slaytor, M., Matsumoto, T., Noda, H., 1999. Metazoan cellulase genes form termites: intron/exon structures and sites of expression. *Biochim. Biophys. Acta* 1447, 146–159.
- Tusnady, G., Simon, I., 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17, 849–850.

- Twining, S.S., 1984. Fluorescein isothiocyanate-labeled casein assay for proteolytic enzymes. *Anal. Biochem.* 143, 30–34.
- Vinokurov, K.S., Elpidina, E.N., Oppert, B., Prabhakar, S., Zhuzhikov, D.P., Dunaevsky, Y.E., Belozersky, M.A., 2006. Diversity of digestive proteinases in *Tenebrio molitor* (Coleoptera: Tenebrionidae) larvae. *Comp. Biochem. Physiol. B* 145, 126–137.
- Vinokurov, K., Taranushenko, Y., Krishnan, N., Sehnal, F., 2007. Proteinase, amylase, and proteinase-inhibitor activities in the gut of six cockroach species. *J. Insect Physiol.* 53, 794–802.
- Woodring, J., Hofman, K., Lorenz, M.W., 2007. Activity, release and flow of digestive enzymes in the cricket, *Grillus bimaculatus*. *Physiol. Entomol.* 32, 56–63.

Supplementary Table 1. Cloning and sequencing primers

Name	Nucleotide sequence
Amylase_degen_FW	AGACCWTGGTGGGAGAGGTA
Amylase_degen_RV	GATCAATRACTTCYTGAWAGA
PA_AG2829_FW	GATGTTTCGCGTGGCTGTAGC
PA_AG2829_RV	GAGGTATGCTACATTCTTTC
PA_AG2829_FW1	CCTAGCTTTGGCACACTGGA
PA_AG2829_FW2	TGTGGACATGGAACGACCAG
PA_AG2829_FW3	ATCAAGGAATGGCGGTCGTT
PA_AG2829_RV1	GTGTGCAGACTGCCTTGTTG
Pa_AG1753_FW	AACTGCAGATAGCTCTATGCTT
Pa_AG1753_RV	ATAAAGCAAATCGCTCATGCTG
Pa_AG1753_FW1	AGTACTGGCCGCTAGAGGAA
Pa_AG1753_FW2	TTCAAAGCCAAGAACGCTGC
Pa_AG1753_FW3	TTCGGTCCGATGGTAGAGGT
Pa_AG1753_RV1	ACCTCTACCATCGGACCGAA
PA_Chi1_1762_FW	GAACAATGGCGTCTATCTGCAG
PA_Chi1_1762_RV	TTCGAAGATATTAAAGTCGCCC
PA_Chi1_1762_FW1	GGCTAAGAACGGTGGTGTCA
PA_Chi1_1762_FW2	GCGCTAGCGATAAGGCAAAC
PA_Chi1_1762_FW3	AATTCACGCAGGAAGCTGGA
PA_Chi1_1762_RV1	GCAGAAACGGCAACAGACAG
PA_Chi2_3012_FW	CAAAGTTGTACGATGAAGCT
PA_Chi2_3012_RV	ATCTAGATGCTTTATACTCGC
PA_Chi2_3012_FW1	AGAGAAGTTCGCCGACAGTG
PA_Chi2_3012_FW2	AACACATGGCTGAAGGGAGG
PA_Chi2_3012_RV1	CACTGTCGGCGAACTTCTCT
PA_APM17_280_FW	CAGCAGCAAGAGCAGTGGAG
PA_APM17_280_RV	CTAGTGTGTCATTGGGTAAAC
PA_APM17_280_FW1	AAGCGGTGGACTCTGAAGTG
PA_APM17_280_FW2	AGATGTGCCTGAGAAAGCCC
PA_APM17_280_FW3	TCGAGAGACAAGTGTGGTGC
PA_BetaGal2736_FW	GGTATGTTCAACATGACACT
PA_BetaGal2736_FW1	ACTTCGAGCTGCTGGTTTCA
PA_BetaGal2736_FW2	GTACGGTCTTCGCAGTCACA
PA_BetaGal2736_FW3	ACACCCTGACAAACCAGCAA
PA_BetaGal2736_Fw4	TGGCATATAACGCTGCACCA
PA_BetaGal2736_FW5	TGGGCATTGTGATGCTGGAT
PA_BetaGal2736_FW6	GCTTCACGCCACTTTGGAAG
PA_BetaGal2736_RV	TGACTTCGCTATCTCTGCTG
PA_BtGlu1698_FW	GTAAGGCTGTCAATTTCAGC
PA_BtGlu1698_FW1	CGCACAACAAGTCCAACCTG
PA_BtGlu1698_FW2	AACGGAATTCAGCCGATGGT
PA_BtGlu1698_FW3	TTGCTATGAACAGCGAGGCA
PA_BtGlu1698_FW4	CACGTAGTGCCTTGGGGATT

PA_BtGlu1698__RV	CTTGGCATCTTATTTGTTTACTC
PA_Chymo4045_FW	ATTCTGGTGACACTGCTCTCT
PA_Chymo4045_RV	TTACGTTTCGATCGGTGTTTTG
PA_Try725_FW	AGTGAGCGGTATACAAAATAC
PA_Try725_FW1	GGTCGTTTCGACCTCAACACT
PA_Try725_RV	GTTCACTGCATGTTGTAGT
PA_AlpaAmy2849_FWI	AGCTGTTGTTTTGGTAGCA
PA_AlpaAmy2849_FWII	CAGAAAGACCCCAATGTGTG
PA_AlpaAmy2849_RV	TGTGTTGTTGGTTCTTGTA

RT-PCR primers

Name	Nucleotide sequence
PA_AMY_MG_FW	GCCTATGTCTTATAGAATTATCAGC
PA_AMY_MG_RV	AAAGGACGAGATCCTTCTGCGAAC
PA_AMY_GS_FW	ATTAGTCTCGTACAACATTTACAGC
PA_AMY_GS_RV	AACGGTCTGGTCCCTGAAGCAAAT
PA_Try725_FW	AGTGAGCGGTATACAAAATAC
PA_Try725_RV	GTTCACTGCATGTTGTAGT
PA_Try3823_FW	AGCTCTGCTACATTCCGAGC
PA_Try3823_RV	CCTTGGCAAGCGTCTTTTCC
PA_actin1156_FW	TCCAGGGCAACATAGCACAG
PA_actin1156_RV	GGTGATGAGGCCCAAGTCAAA

Supplementary Table 2. Features of the digestive enzyme sequences found in the *P. americana* transcriptome.

Contig	Predicted Protein	Reads	Length	SP	TMH	GPI	N-Gly	O-Gly	
2849	Alpha-Amylase	35	483	-	-	-	Y	-	Partial
1753	Alpha-Glucosidase	283	595	Y	N	N	Y	Y	Complete
2792	Alpha-Glucosidase	122	563	N	N	N	Y	N	Complete
2829	Alpha-Glucosidase	100	543	Y	N	N	Y	Y	Complete
3232	Alpha-Glucosidase	16	375	-	-	-	-	-	Partial
4940	Alpha-Glucosidase	8	139	-	-	-	-	-	Partial
2908	Alpha-Glucosidase	21	484	-	-	-	-	-	Partial
2910	Alpha-Glucosidase	18	577	-	-	-	-	-	Partial
4268	Alpha-Glucosidase	13	244	-	-	-	-	-	Partial
4659	Alpha-Glucosidase	7	213	-	-	-	-	-	Partial
4869	Alpha-Glucosidase	10	199	-	-	-	-	-	Partial
5285	Alpha-Glucosidase	8	170	-	-	-	-	-	Partial
280	Aminopeptidase	43	526	N	N	N	Y	Y	Complete
3256	Aminopeptidase	18	398	-	-	-	Y	Y	Partial
2324	Aminopeptidase	7	82	-	-	-	-	-	Partial
2325	Aminopeptidase	8	119	-	-	-	-	-	Partial
4015	Aminopeptidase	9	271	-	-	-	-	-	Partial
4127	Aminopeptidase	8	260	-	-	-	-	-	Partial
4370	Aminopeptidase	7	235	-	-	-	-	-	Partial
5056	Aminopeptidase	16	69	-	Y	-	-	-	Partial
5255	Aminopeptidase	11	159	-	-	-	-	-	Partial
5636	Aminopeptidase	6	70	-	-	-	-	-	Partial
2898	Aminopeptidase	30	506	N	N	N	Y	Y	Complete
4544	Aminopeptidase	8	209	-	-	-	Y	-	Partial
2736	Beta-Galactosidase	82	667	Y	N	Y	Y	N	Complete
1699	Beta-Glucosidase	39	387	-	-	-	-	-	Partial
1700	Beta-Glucosidase	9	101	-	-	-	-	-	Partial
1702	Beta-Glucosidase	31	379	-	-	-	-	-	Partial
1701	Beta-Glucosidase	229	505	Y	N	N	Y	Y	Complete
1698	Beta-Glucosidase	107	495	Y	-	-	Y	-	Partial
3012	Chitinase	248	407	Y	-	N	Y	N	Complete
1762	Chitinase	231	123	-	-	-	-	-	Partial
3726	Chitinase	10	262	-	-	-	-	-	Partial
5314	Chymotrypsin	13	160	-	-	-	-	Y	Partial
4045	Chymotrypsin	29	251	Y	N	N	N	Y	Complete
617	Trypsin	34	196	-	-	-	-	-	Partial
3433	Trypsin	96	249	-	-	-	-	Y	Partial
620	Trypsin	327	254	-	-	-	-	-	Partial
1654	Trypsin	33	354	-	-	-	-	-	Partial
1710	Trypsin	365	299	-	-	-	-	-	Partial
1942	Trypsin	6	164	-	-	-	-	-	Partial
1944	Trypsin	13	84	-	-	-	-	-	Partial
2277	Trypsin	4	100	-	-	-	-	-	Partial
3010	Trypsin	47	353	-	-	-	-	-	Partial
3127	Trypsin	24	131	-	-	-	-	-	Partial
438	Trypsin	10	135	-	-	-	-	-	Partial
4947	Trypsin	8	193	-	-	-	-	-	Partial
619	Trypsin	25	200	-	-	-	-	-	Partial
724	Trypsin	1223	47	-	-	-	-	-	Partial

1941	Trypsin	253	278	-	Y	-	-	-	Partial
1945	Trypsin	198	109	-	-	-	-	Y	Partial
2593	Trypsin	178	253	-	Y	-	-	Y	Partial
3851	Trypsin	45	268	-	-	-	-	-	Partial
4043	Trypsin	20	200	-	-	-	-	-	Partial
616	Trypsin	10	201	-	-	-	-	Y	Partial
618	Trypsin	4	57	-	-	-	-	-	Partial
2758	Trypsin	50	243	-	-	-	Y	Y	Partial
2966	Trypsin	32	215	-	-	-	Y	Y	Partial
3779	Trypsin	28	254	-	-	-	Y	Y	Partial
4033	Trypsin	7	222	-	-	-	Y	Y	Partial
415	Trypsin	102	255	-	-	-	Y	-	Partial
4260	Trypsin	21	227	-	-	-	Y	Y	Partial
4335	Trypsin	15	215	-	-	-	Y	Y	Partial
5034	Trypsin	7	138	-	-	-	Y	Y	Partial
640	Trypsin	60	225	-	-	-	Y	Y	Partial
2967	Trypsin	178	256	Y	N	N	N	Y	Complete
3823	Trypsin	586	256	Y	N	N	N	Y	Complete
725	Trypsin	2502	252	Y	N	N	N	N	Complete
3434	Trypsin	72	259	Y	N	N	Y	Y	Complete
3482	Trypsin	64	260	Y	N	N	Y	N	Complete
3624	Trypsin	367	258	Y	N	N	Y	Y	Complete
401	Trypsin	42	272	Y	N	N	Y	Y	Complete
2322	Trypsin	51	254	Y	N	Y	Y	N	Complete
1940	Trypsin	106	159	Y	-	-	-	-	Partial
3824	Trypsin	20	225	Y	-	-	-	-	Partial
1943	Trypsin	955	315	Y	-	-	-	Y	Partial
4190	Trypsin	19	250	Y	-	-	-	Y	Partial
3554	Trypsin	18	254	Y	-	-	Y	-	Partial

SP, signal peptide; TMH, transmembrane helix; GPI, glycosyl phosphatidylinositol anchor; N-gly, N-glycosylation predicted; o-gly, o-glycosylation predicted; Y, found; N, not found; -, the length of the sequence does not permit the search. Contig 2792 is homologous of α -glucosidase, but lacks one of the catalytic residues.

Supplementary Table 3. FASTA sequences of all contigs listed at Supplementary Table 2.

>280

MTAPARNLFQAVIVAAAKAHQHHPHHHQKIRDGIVTVVPTGGEGPTHPELKKSFSFAFSKL
DKAVDSEVTLPFCDLPAGRVISSPTGPLNDDYADVRSFTDAKKGIARALKSGVTAPLLV
LPSPAFKHSELVTLLGALEALYVNLQMREDVPEKAQKVKTGLGVYSENTENLPNIIKLAS
ALESGRYVARDIGGSDPERMCPPKVEAYIRELFKNSNISIEVITDPAVLKREYPLFDAVD
RAASHIERHRGRIIYLTYPSSGPVDKTLFLVGKGVTYDTGGADIKAGGVMAGMSRDKCG
AAAVAGFMQILTMLKPRGVKVIGALPVVRNSVGADAYVADEIILSRNKLVRIGNTDAEG
RMIMGDVLCKMRELAEKAVNPHLMTIATLTGHAVLTVGAGYSIVMDNGPARKASHAQLLQ
AAGEELGDMLEISTIRREDIAFHKGKGMFDDIYQANNLPSSRTPRGHQGPAGFLILVSGL
DKHGSESKIPLNYSHIDIAAASSGDVPHDATGAPILGLVSLHLKDRY

>401

MKFQIVLFATIVTSCLGADVDESDDIDSRIIGDIYAEDDLPSVVYITKGKSGETATTI
QCVGTLINPVFVLTSASCIVGGLPEDFHVVAEAAADFNEGVYTVVDVIPHENYEDNYEKY
WPHDIGLLQLDVQELNGTIRNIADIPKRDHELDPAGSVYGWGATCGFCGPTSGLKVVNLQ
VYNRQDCTSTYYDGKAPADSSFCAGNIENMLGVCDGDSGAPLLYNGTIYGVASWMSYPC
GSPPAVFTDVQSHVEWIVRKMSTTPLTPLVSA

>415

CLYLLAGFLVLSAQGIPRRIGKVPDAGRIVNGTDAAPNQFPYQVSLQHEGSHSCGGTILS
DKWILTAAHCVEEIEANLFSVLAGTNVLSQGGERRNVIRYHMHEEYDADNNYNNDIAVVE
LETPLSMTGDMKVVTLPQTGGQDTAGDLTATVIGWGLLYTDGPSSDTLKMVDIKTYSDEDC
LKVMPPGYVRPNNICAGVPEGGKGQCRGDSGGPLIVNGYQAGIVSWSVKPCGLAPYPGVYT
EVSHYIDWIKSKTNL

>438

VNIINTGISALPKRGHVRPNTGSVYGWGANCGFCGPSSYLKVLNDAFISNSDDCSSLFYD
GKGPADSSICAGSLTGFGVCDGDSGAPIVSKGTIYGVASWLNYPGGSVPVFTNVQSHV
EWIVRKMSATPLVPL

>616

ILSENWALTAHCVDGSSTSSITLYAGSIHLSIGGTRHYVSQIIVHENYDEYDSWINDIA
VIKVSNNPFKIDGVNVNPNVSLPAQGNPADGSTATVIGWGRLYTS GPIPNNLQKVDIAIVN
QKECNTVYGNMNYPIYDQGICADVPQGNQGCNGDSGGPLFVDGTVVGLVSWAKGCATPG
YPTVYTRVSAYRDWIKQKTGV

>617

LAILVAAGTQGKTIPRIPVDRGSRIINGDNAVRSQFPQQLSLQISSHSCGASILNENWA
LTAHCQVQWNNASSMTLYAGSINLSSGGTIHNVSEIVIHEYSYDAYDSYINDIAVLKVSNP
FQIDGVDVKPVSLPAQGQTPADNTVATVVGWGEHGEHGFSDTLQRVNIIIVNHDQCNSI
YEFYSYHVDYSQICAG

>618

PQGIGGSCRGGDDGGPLFIDGNVGLVSWSRYPARPEYPTVYTRVSAYIDWIKETGA

>619

LNENWALTAHCVERNNASSMTLYAGSIYLSGGGTTHNVSEIVIHEYSYDAYDSYINDIAV
LKVSNNPFQIDGVDVKPVSLPAEGQNPDEMLAEVIGWGEHGEHGFSDTLQKVNLMIVNQ
DQCRSIYEFYSEHVYDSQICAGLPEGLGGSCRGGDDGGPLFVDGNVGLVSWSRYPARPEY
PTVYTRVSDYVDWIKQKTGA

>620

VAILVATGTQGKIIKRIPIVRDNRIINGVDASPGFEFPSQLSLQRPTSHSCGASILNENWA
LTAHCVDGSSTSSITLYAGSIYLSGGGTTHNVSEIVIHEYSYDAYDSYINDIAVLKVSNP
FQIDGVNVKPISLPAQGQNPADGALATVIGWGRIEEDGPIPDTLQKVDIAIVNQDECNSV
YEIIGYPIYDQICAAVPDEVKGSCNGDSGGPLFVDGTLVGLVSWANGCARIGYPTVYTR
VSAYRDWIRQNTGV

>640

KFILTAAHCIYSGIIDNYTIRAGTNNYKRGSVHKVWHVYHEL YEENIKVMHGHRSK
EYPFQVSLQIEGEHYCGATILNEXXXDIAVLEVDPLTFRPRVQPTRLPSEGEEDPDVGAK
SVAIGWGSKNFYGPCRPQLHEVSLNIYSHQECLDVYWNLPPTDMVCSGAPYEQQQVCSGD
SGGPLLVNGIQVGITSWGAFPCVSYPTVWTRVSHYRSWIREKSGM

>724

GGDQSNILKEADVPVVG YHSCNTSYATYGGITDNMLCAGYNEGGVDA

>725

MAAIRFLVLAFLFACGASEPINEKEGRIIGGTTTTITSYPYIASLHKGTEHICGATIINSEWLTAASCISIP
NAREYLVRVGSTYSSSGGNVYQVSKIYRHEFDLNTMDNDIALVRLSTGVIFGTNVQSI TLQESDVS DGT SATV
LGWGTISEGGAQSNILKQVVVPVVG FHSCNTSYATYGGITDNMLCAGYNEGGVDACQGD TGGPLVASGHLIGI
VSWGAGCGHSTFGVYTKVSAYRDWIKDKTGV

>1654

MLAAIVVILLHAYTDAHVVNTTVAISNNMVLETTGTTAAPRTAPTVTIEATKGAETELS
TATISLKTAPTVDYKNLSIRNETNNSVLSSGRPEPKIVGGRDAQQGEFPYQVSLRRNGWT
LSHYCGGSVIDEIHVL TAAHCVVGMAPS DIKVVVG DIMDRSLCTSVSRVS AIFIHEKY
NKSTFENDIALLRVYEEFPKDSEFIMAIRLQESPVPEGTWCNASGWGV LSSGASEMPNNL
QYVELPVMSYADCYKANGIYDGMICAGSMEGGKDSCQGD SGGPLQCNGYL VGIVSWGVGC
AYPNYPGVYTEVSIFYDWNATKY LKGSASMC RQNAFYFPSILALLVYSMSKYV

>1698

GLCLVALVAAASSAAVVHDAQSKAYTFPEGFL LGVATASYQIEGAWNEDGKGVNIWDTQV
HNRSYLIVDHSTGDVACDAYHKYKDDVQLLKEMGVDIYRFSFSWSRILPNGHINEINQPG
IDYYNNLINELIANGIQPLASIFHWDLPQPLQDLGGWPNPVIADYYEDYARVLF TNFGDR
IKYWLTFNEPATFTSGYEDAEGHAPSIGANGFGRYLASHTVIKAHARAYHLYDDEF RATQ
QGKVSITLDVGWCEPLENSTEYVEACERQQEFELGLYANPIYSAEGDYP AVVKERVAAIS
KAQGFTTSRLPTFTQEEIEYIRGTHDFFGMNHYTTSLGSPSDDSENPPSFYKDIGVIKTQ
DPSWPTSASSWL YVVPWGFRKQLVAIAQKYN NPPLIVTENGFS DHGELNDSGRVDYYTSY
LTEMLKAINEDGCNVFGYTAWSFMDNFEWNRGYTERFGMYVDFDDPERPRTIKESGNVY
TEIIKTRQIPERFRK

>1699

VQRCSTHERNGVDFYRFSFSWSRILPNGHVNVINQAGIDFYNNLINELLANGIQPLATI
YHWDLPQPLQDLGGWPNPIIADYYEDYARVLF TNFGDRIKYWITFGRYLASHTLIKSHAR
AYHLYDDEF RATQQGKVGITL NIGWCEPLENSTEYEEACERQQQFEMGLYANPIFTDEGD
YPAVVKERVAAISKAQGFTSSRLPTFTQEEIEYIRGTHD FGLNHYSTYLGSPFDDSGNPP
AFYKDMGTSTTQDPSWPGSASSWLAVVPWGFRKQIASIAQRYNNPPLIVTENGFS DHGEL
NDNGRIDYLSYLTEMLKAINEDGCNVFGYTAWSFMDNFEWNSGYTERFGMYVDFDDPE
RPRTIKDSANVYAEIIKTRQIPERFRK

>1700

ENGFS DHGELNDTDRINYVTSYLTEMLKAINEDGCNVFGYTLWTLMDNFEWNEGYTERFG
LYYVDFEDPERPRTIKESAKVYAEIISTAQIPERFRKNNEM

>1701

MESHMLLFLFSLATGLFGAVHGAPEDKVPDSLKDYAFPDGFLFGTATSSYQVEGAWLED
GKSLNIWDTLTHNKSNI LSDRSNGDVACDSYHKYKEDVQLLKE LGVNFYRFSVSWSRILP
TGHINVVNQAGIDYYNNLINELLANGIQPMVTMYHWDLPQTLQDLGGWPNQVMAQYFEDY
ARVLF TNYGDRVKYWITFNEPSVFTAGYESVAFHAPNVGATGFGQYLATHTVLKAHARAY
HLYDNEFRAAQQGKIGMAFNINWCEPRDNITEDIAACNRMQEFNLGMYAHPVFSPEGDFP
TVVKERVAMNSEAEGFTQSRLPSLTQEEIEYIKGTADFFGLNHYTTFYGSPLTYTGEP TF
TKDVGILMMPDFSWPGSASIWLHVVPWGFRKQLNRIAEIYNNPPVIITENGFS DHGELND
TGRINYLSYLTEMLNAIHEDGCNVVGYTAWSLIDNFEWNNGYTQRFGLYHVD FEDPNRA
RTIKESAKVYAEIIATKQIPERFREE

>1702

GVATASYQIEGAWNEDGKGVNIWDTHTHTRSDLIADQSTGDIACDSYHKYKEDVQLMKEL
GVDFYRFSFSWSRILPNGHVNEVNQAGIDYYNNLIDELTANGIQPYATMYHWDLPQPLQD
IGGWPNPVIAEYFEDYARILFSNFGDRVKYWITFNEPSTFTGGYESEKGHAPSIGALVG
KYLATHTVLKAHARVYHLYDENFRASQQGKIGMSINVPWCEPLDNAAIYIDACERQQQFE
MGLYAHPIYSEEGDYPDIVKERVAENSLNEGFTRSRLPTFTQEEVAYIKGTHDFFGLNH
STYLGFP LPSDGEPSVIKDMGMTMILPDPSWQGSASVWL YVVPWGLRKQLNRVKEEYNNP
AVIITENGFS DHGELNDTD

>1710

PSENCKDGKLLVSLTGDTSLSDGHTYCGSGNFTLESESNQLTMQLITSYFNRGAF LCSIT
ATQQQT PPPPTCDCGWKESRIVNGEETGVNEYPMAGLVDFVNEDLFCGGTIIAKRYVL
TAAHCLLNRAVNITGVLVGDHNITTGDDTSSARVLLAESFQAHPDYN SATQMNDIAVIRV
NKAIIFS YEVGPA CLPFRFPNINVGAVTVL GWGTTEQAGPKSDVLLQTKLYIIP LKNCT
DTFGSKVTSNQICTFNAGTDACQADSGGPGLWLGP GDRLQLAIVISYKGKCAFKYPGVN

>1753

LQIALCLLILMTVTDVLP RSIPHVSV DITPGFLGHR SITIKRDEIPVLTGDLGVTSS THL
SIARDGSC THMSQTAETASDVLEHCFDLGDAHWFGGGEQFYQYWPLEEATLVDYSYVTKQ
QDNQGIAEPYWLNSDGVYFHVDDGVPLFIDLNNGNKSLCFKAKNAAPYPPTTIPFLNYSI
CTYADAREAHEHAIQTVLGKPTGYPDERMIRHPIWSTWARYKANVNESVVLTFAKEIVEN
GFNNSQLEIDDDWEVCYGALTFRESKFPNMKNLTDTLKAMGFRVTLWVHPFINLDCEPWH
SEALEKGYFVKNL DGNVNSTWWNGAAS AIDFTNPEAVEWYVNR LKQM QEENGIDSLKFDA
GETSWLPQLPVLNSTDIQNPSIYTTRYVDTV SQFGPMVEVRVGQRSQDLPIFVRMLDKDT
KWNWDNGLKTLVTTLLEMMVGYPLVLPDMIGGNGYGDIVTKELFIRWLQANVFMP SLQY
SYVPWDFDEETIDICKKFTDLHAQYADKIIELAAKATVDGSPVNPPIWVVDPTDETALAT
NSGFLLGEDILVAPILDQGATS RDIYLPKGQWRDEL RQETVTGPITLSNYSVALDELAYF
TRVSA

>1762

MASICSFSLICFIAIFSTSFQGS EAASDKSVFCYFGSWATYRWGIALFEVSNIDTTLC T
HMVYMYTGLKNGEIASLDEYNDYEENFGKGNIKKFVTLAKNGGVKALVGIGGWNEGSVKY
SLMAASAAGRETFADSAVAFIQKFGFDGLELDWEYPASNGGGASDKANYVLLLQTLSEKF
AENNLILSVAVSASAEVAEAAAYEFVDVAKFADHII LFSYDYHSFYDGVTA VSSPLTGDDK
LNVNYSVHQWLDAGVPSNKIVIGVPLYGR TYHLADPSKHGLGAAVTGPGTAGQFTQEAGL
LTYLEYCTSATWTVVQSDDPLYTYAYHDSEWISYDDPQTIEGKAQYVISNELGGMVFSL
ESDDFGGNCGPAYPL LNAVNTGLGRL

>1940

MKTFVALALLVTFGLMAEARLYPGL LKGIQAPKAPKAHIKDISSRITHGETASRGQFPYQ
VALSVDNSWFCGGLISNDWVLTAGHCIGNSYQVLLGANSVNNPESGAVVVTSSTSIQHE
DYANDHDIGLVKVSVSYSTYISPIALPSRSDQGNTFVGQ

>1941

VSNMKT FVALALLVTFGLMAEARLYPGL LKGIKPVGKVHVRDASSRITNGQTASRGQFPW
QAALHVDNSWFCGGLISNEWVLTAAHCIGSTYRITLGANNVNSPESGVS VSSSSSIRH
SGYNSDTLANDIGLVKIPSVSYSNYISPISLVPRSDQGN SFAGESVRVSGWGKNSDSASG
ITAQLQYVDLTVITNQECNSVYGIITSGMLCVSTPGGRSTCSGDSGGPLIHSQNGGYVQI
GVVSFVASSGCEGLPAGFTRVTSYLDWIQSN TGIATK

>1942

RDASSRITHGQTASRGQFPYQAGLIVDNSWFCGGLISNNWILTAAHCIGSTYQVILGAN
NVNNPESGTVVISSSSSIIHSGYSSNTLANDIGLVRI PSVSYSTYISPISLPSRSDQGNS
YVGETVIVTGWGLNSGSGTGITGQLQYWNLT VITILKLTSDTCS

>1943

ACSMKT FVALALLATFGLMAEALQSRNIVMPINPRYTAE EWRTIMAAQLESPEQYPGFKK
GSEPQPAFAGKVHVHDVSSRITNGQIASRGQFPWQVALLVDNSWFCGGSVISNQWVLTAA
HCAGSSYQVILGANNYSPESGSVTVTRSSIRHENYNSDTINNDIAVVKLQSPVSFSTY
ISPVRLPTRSQQGD TLAGETVRVSGWGKISDSASSVTDQLRFVDLTVITNSVCAGTYGSV
ITSSKLCVSTPGGRSTCSGDSGGPLVHLESSGVYTEVGIVSFGAAAGCELGYPAAFTRVT
SYLDWIETNTGISIG

>1944

FQLQYWNLT VITNQECNLVYGIISSSILCALTPDGESTCIGDSGGPLLHSENGGYVQIGI
VSFTAAAGCEAGLPALFTRVTSYL

>1945

ARASGWLTSDGSSSIAELLQYADLTVITNDECNAVYGGIHS GILCVSTPGGKGTCSGDS
GGPLVLGGIQIGIVSFGAAAGCEAGLPDGFARVTNYLDWIQSHTGIIIN

>2277

VNQAAALRAAIDGADSVTMKYLENARDKVL MGPERSRIPDEEANVITAYHEGGHAI VAYY
TKDSHPLHKVTIIPRGHSLGHTAYIPEKERYHVTKS QLLA

>2322

MLRILAVAPILFACCLGLPDQPSNRILDGVDTDAESYTYMASVEYNKNHRCAGSII SYNW
ILTVAHCVHGLDTELLRVRSAFVPIRGLMGKVS AVNSTILHPYFDFKTLDYDIALINTA
DDLYDVNGTDTNIHIGLRPLDGN SATVFGWGARIIGTWFTKFLTNQTLTIVYFNECKYS
YESVGNPVTERMFCAGLVGGGPCQGDSDPLIYGNRLIGMMSWSYGCGERRFP AVFTDVT
QFRFWIESTAGILL

>2324

YLQVPKIEEWLAMQLPSGDVISADPKIIPYTDWILWDEYFTKRNLRL EPLYDNLVDMIWG

AAEGRPPYPSDPIFVLDTKYTG

>2325

LSWQDKISAVRQTMAEKGFNVLVLTALDEIAWLFNLRGSDVPFTPVFQSYAVVEMSGATL
YLSPGKSTPQVEEHLNAHSSNMDDSVRIKNYTDIWNDLAELKDDAVRVWLPYPFSYARG

>2593

MKVLVLLAALVACVAGVPYILQKPLPDGRIVGGTTTTITNFPYQLSLRYSGSHICGASII
SANWAVTAGHCIIGSASSLTLAGSTYSNTGGTVHQVSQAIRHGSYNANTLDYDIAVLKV
SSAFSFGSSVQAIPLETGSVSAGTSAVVSGWGTTSEGGASATTLRQVAVPIVSDSSCNSA
YSSYGGITARMICAGYTSGGKDACQGDSSGGLPVANGKLIGIVSWGAGCARPNYPGVYTKV
SALRSWISSNTGV

>2736

MFNMTLTSRIVVAVFTIVLLTAVSGSLVKGKYAQLPTLYEYYTEGGILAGLETLPDFTL
LNNKNITLFSGTIHYFRVHPQYWRDLRLKRAAGFNAVETYVPWNLHEPQQGVDFGNGG
NDMSEFLDIVTYLKTAQEEDLFVILRPSPICAEWEFGGMPSWLLRYPDLVVRSSQSQFV
DLVKEYNANLIPLLVDLQFTKGGPIIAVQLENEYGSLAGDGSEEDVAYLSFLKEDYEKYG
IVELLFTSDSSNADKGGPLPGVLQTANFQSNPEGMLSQKLNHPDKPAMAMEYWTGWFDHW
FESHHTTSVAGFRDVLERILVFPASVNLVFMHGGTSFGFLNGANGEGNSYGADVTSYDID
APLSEAGDYTDKYDACVETVAKYNKVLTKVPSPPEISKKMAYNAAPITGQLDFNQIIDKI
PAKDRANSNDNVHMEALPVNNGNGQSYGYVLYRHADTSVTTSSVLQVKNDIYDMGIVML
DGERKTEKLTSVSQVGDFGFWILKDATLVMNSTTSGTKTLDILVENWGRVNFGRSTFDQ
RKGLRENTIELIDNNEITGWEVISLQFKSNWVNSLDGWQPVSSSEMIAPKLLHATLEVDT
YDTFLDMRGWGRGAVFINGFNIGRYFSAGPSHTLYVPAPLLKTGTNDIAVFELFTAGNEI
VFADQPYL

>2758

GHPHPAPSITSGERIVNGTNASPGQYPFMASLRYWKNHVCGASIINECYVLTASHCVIWK
LPEELSVRVGSVDLSSGGSLHDVSEIIMHSEYDPDNMHVHDIALKLSSPITLNKYAQP
VLPHQLQDTKGGSAATLVGWGSSETDGEVMEHLQQVRLRVYTDEACRSLHYRIHMSNIC
AGIPEGGKGQCGGDSGGPLVAHGHQVGIVSWSVKPCTVAPYPGVYTEVSYYADWIKHHTS
GSC

>2792

MAFLSTAAAAPAAEALDWWQTTVFYQVYPRSFMDADGDGVGDLAGIAQKLDHLQDAGVGA
VWLSPIYKSPMADFGYDIANFTDVPDIFGTMKDFEDLKARAQELGIKLVLDFVPNHASNE
SEWFAKSLQKVDPTYDYFVWVDPKTDDEGNRIPPNNWLSAFGGSAWEWREERQQYYLHQF
AIQQADLNYYNNPNLVEEMKDVLRFWMNKGVGDGRVDAVPYLFDEPELKDEPLSGDPNAQP
GEAAYLSHIYTQNLPGTYDMVEQWREVADAKKAEDGQTRLLMLETYASLEQVMAYYGTDE
RPGGHFFPNFLLITDLNNGSTALDFDETVHKWMDNMPAGRWANWVIGNHDNHRVATRYGR
ELADGLNMIAQLLPGSGVTYNGEEIGMEDTFISWEDTVPAGVNAGPDYQLFSRDPERT
PFQWDDSVSAGFSTNATTWLPVNENYKQLNLEAQAADVSHYKVYQRLLEARTLPVIQTG
GLQTYAASEEVFAFSRTLEGSSPFVVVNLGNSQEEVNLATLADVPEDLVVYVASIGSNR
QAGAAVSKSNFVLDPKEGLVLT

>2829

MFRVAVAVALMGAVLAEQDPLDWWKTTVVYQVYPRSFKDSNGDGNGDLNGITEKLDHLKD
LGVGAVWLSPYIKSPMVDNGYDISDFYDVDPSTFGTLDLDDFAKLKARAKELDIMVLDVFPN
HSSDEHEWFEKSVQRIEPTYDYVWHDGKIDPDTGIRVEPSNWISIFSGSMWTWNDQRQQ
YYLHQFGVKQPDNLFRNPVMEEMKNVLRFWVDRGVGDGRVDAVPHIDEDAAFTDEERTN
YPGVPPTDYNLYHVHTANRPETYDVIKEWRSFLDGEFKDKDNHTRVIMTEVYASIEETM
AYYGNDTNPGAYFTFNFNLVNSAGSWSTANDFNNVIHSWVDNIPEGKQSNWVVGNDNSR
VANRYGNTMVDVAVNMLVTLPGTAVTYGEEELGMEDLAMDSPDNDRPERTPFQWDNTTS
AGFSTSNVTWLPVNPYKDLNLVSEKAAEKSHYHVYKQLLQARQEPAIQQGSLHTNVIGD
VFSFSRELEGAAIFVVVMNLSDSEQTVLDLEAPYPSLPQSPLVHTASINSNLTQGQVRN
LTLGPRAVVLTS

>2849

LLFLVAVLAFQAQKQDPNVWPDORSVIVHLFEWRFDDIADECENFLAPHGYAGVQTSPVHE
YLIVTSVDRPWWERYQPMYSYRIISRGDETAFRDMVRRCNAGGVRIYVDVLLNHMTGNWD
NAVGTGGSTADTYNYSYPGVFDHSHFHPYCILNDYNSPEIVRNCELSGLHDLQSQDYV
REKLIDFLNRLVDAGVAGFRVDAAKHVWPADLEYIYGKVNLLNTEHGFAEGSRPFYQEV
IDFGNDVSKFEYNSFGRVTEFKHGEQMCYAFGGQYPLAALHTFDPNWNLLDSGDALVFI
DNHDNQRGGNNAILTYKTPKNYRMGISFILAHYTYGYPRVMSSFDFAHDQGGPPQDSKNI

LHVTTNEDGSCGNGWVCEHRWHSAAAMVFRFRNVVRGTEITNWWDNNGNHQIGFARGDRGFV
AFVVNDNDLKQTIQTGLPAGTYCDVISGGKNGSSCTGKTITVNEDGTAYIEILISDDDG
IAIHLEEK

>2898

MFSSEQLKEKAGLVLGYYTADTTGQMKLTSTAQTFNEKANGQLIQHIKLAGP
SIKKGKGQVFYNQDSEFPVTAVVGLGTEDAGYNEQEEIDEKLENIRIAASEGCKKLQEVG
VTKVVLDGMGNPEAAAEGATLGLWLYQGFKDKQKQKVAPQIQLCSENEAERTAWERGVK
AQAQNVARNLMDTPANYMTPRIFAQNAATELLTAHDVKVTVRDKSWAESKKMGSSFLSVARG
SIEPPVFLEISYTGASSQYKPIALVGKGVTFDSGGISLKPPAKMDEMRAADMGGGACVVAA
IYAAARLKLKVNLIIGLIPLTENLPSSGATKPGDVVHAMNGKSIIVDNTDAEGRILILADAL
CYAQEFNPQLTVDIATLTGAIRIALGSAAAGVYSNSTPLFQKLQQAGALTGDRVWRMPLW
QHYTKAMTDTQSADLLNVSKVAGAGSCTAAAFLEKFIIPSGNWIHMADIAGVMSTSNKLPYL
SKGMTGRPARTLIEFISNLEEGSKSC

>2908

EGVVEPYWLTSSGIAVFPDPVPLFVSQNKTHLCLAAKIESPYLPDDTGEVSLQYHLCSSG
SDPVTTHMHATQSFLPLPRDIPAERVFRHPIWSTWARYKKNINTSIVSDFADQILHYGFN
NSQLEIDDNWETCYGSAEFDVSKFEDPKGLVQSLKNRGFHVTLWTHPFFNLDCPAREVAL
QKHYLCTGANNSETLTKWNGVASMINITNPEAATWWADRLVNIIDRYNFDDFKYDAGEA
GYLPEVGRNLNAPIRLQPNVFTTQYVQTVASISDQKAAEVVVYRNQGEPIFARMSDKKST
WDTGSGLKSLVTLLAMNMVGYSFVMPDMVGGNGYGNFPDKELFIRWLQANTFMPTIQFS
FAPWDYDDETVEIAKKFTSLHYQYSDKILELARKKVADGTPINLPLWWLDPTDEETLTID
SEYLLGEDILVAPILDQGALERDIYLPAGSWRDELHLNVTYPGRQWLRNYRVELHQLAYF
TRVP

>2910

LGNPSVCASEELDLCKWVRRQNWTSSCTTREGALAHPSLATKNVITVPADKRIVACFG
LQGIQLYGGAVTSTQYWPSQRNNKSESAYVTGIDGNEGVVEPYWLTSSGSAIFVPPDVPL
FVSQNDSHLCLTSMIQQPYLPDDTEKVSLEKLYLCSGSDPVATHMHATQSFLPRPRDIPDE
RMFCHPIWSTWARYKRLINTFIVTNFAQQIVNYGFNNSQLEIDDNWETCYGSAEFFNATF
SDPKSLVQSLKQKGRVTLWTHPFINIFCPAHAVAIQNRYLAMTPDMKRTLTMWWDLGG
VLNFTNADAVTWWTGRLKNLSMRYGIDAFKYDAGEAAYLPGLVGRQLQPNVFTTRYVQT
VANVSTTRLAEVRVVYRNQGEPIFVRMSDKSSTWSIENGLKALVTLLAMNMVGYSFVMP
DMVGGNGYSNQPTKELFIRWLQANTFMPTIQFSYVPWDYDTETVEIAKKFTSLHYQYSDT
ILELAKKKVADSTPINLPLWWLDPTDEVALKIDSEFLGEDILVAPVLDEGARHRDIYLP
AGRWRDELHPDILHSGKRWLLNYTVQLDELAYFTRVS

>2966

VSFEFFGSHHCGAAIISELWVVTAGHCVDGIPVSTVTFRVGSNTIERGGTVHEATDIINH
PNYDYHTEDCDISVAKVSTAFVYQSAVQPISLASVEPPAGAMAVVSGWGTTYSGGPMSP
LLQVSVPIVDFEACNASYSDHGGITRNMLCAGEDGGKDSCKGDSGGPLVVEKELVGIVS
FGAGCGVRGYPGIYASIPVLRDFVFSNTDVQVQLK

>2967

MFRIAI FVSVLFASCWGASLIPRRPRHDGRIVGGEPDTITNYPYQVSFEFFGSHHCGAAI
ISELWVVTAGHCVDGIPVSTVTFRAGTNVMEEGTVHEAAEIIAHPNFDYLSLDCDIAVS
KVTTAFTYQPSVQPVSLASADPSAGAMAVVTGWGSTSSGGALSPLLLKVSVPVVSFDACN
ASYSKYGGITSNMMCAGADEGGKDSCKGDSGGPLVVEKELVGIVSWGAGCGVRGYPGVYA
SVAVLREFVTSNTDVQ

>3010

HRGARLSVTMLLFSVLACAALIVGTVDSSATTSTAKPETTTTRKVLPEKPRVVGGKPANRGQ
FPYQVIVYVEVEGGKSLCGGTIIAEQYILTAACHVFDSTPRAVEVWAGDIDRGKKQISGI
VSEIHYHENYINGRYNDVAVLKLQTPYVLNGYSITKLQLRDNFVSPNXXXXXTLCTVSG
WGARKERGGSSKILQFVPVPIVDLKNCSANHGKELYEGEICAGYSTGGKDSCKGDSGGPL
VCDGYLTGVVSWGDGCARVNKPGVYANVSYYKSWIDNIISNQNLIPVPTKDTRDQTNHI
AGATESTLRRTTPFNRSDKSPNSASAPNYHLITIFFVIGLMLYTGYYLDYLHE

>3012

MKLTSNIIYVGLVTIALILQTCCKGFIKDFKHRRRLQEKGDRAVFCYYGSWASYRW
GVGTFNVDNIDTRLCSHIVYAFTGLRDGVIVSLDQYNDLEENWGKGLMKKFTTLARNNGI
KSLLAIGGWNEGSTKYSEMAATQEGREKFADSVVSFVEKQGFNGLDLDEYPPARRGGVPE
DKDHFTLLLKVLSEKLHARGHLITIAVSADPKTIANGYDIPNVGKYADYITVMTFDYHTP
SSDVTGLNSPLDSLQDGTGYNKELNVKNSINTWLKGGAPPNKLGLTPLYGRTYRLWYE

DQHDLGSPILGPGDAGLYTQEAGFLAYYEICSNPDWNIVWNATSSVLAYKTYQWLSYDD
PNTITIKTKWALEKNLGGVMVWSLESDDFHGNCCKGTYPPLTTVNKALQRV

>3127

FFDGLQSANMTLLEKELCEDANKKLSGQFCVGYSEDQRRSCLGDAGGPLVCDTNLTGIT
WGPVGCDDAGLTDPAFFTDVSLFAGSMNVSSPAVFTDVSFYKEWITKQISSSATIVPCLT
TLLIMLFVALK

>3232

MRGSQSNRTVVFGCRKMYLHLFTFILIELHLSRSSEILRPRGVSISKASLYVPDKDFTCF
DGSRSIPFAHVNDYDCDCQDGSDEPGTPACPNGTFFHCTNAGHRPLNIPSSRVNDGICDCC
DTSDEYVSQATCVNNCNELGKAARVEAQKLAKLTKLGSEMRIQLSKKGKQLKQEKKERLE
QLEKDKEEAEGVKKEKELLKTAEELENKALEKYRKIEEEVKQRQVEEERKASEQEAFEH
FHRLDSNQDGKIDISELQTQQTFDQNKDGVVSEDEAKFFFDSDHDEL TWEDFLATGWPRMK
PFLMLDKGLFKPPVRETSEEPTGPQGGHDAISQDDAAEFSSSEGDEADKHYYEEDQDGLQ
EEEEEEEEEEEEETG

>3256

MSSTKLQFMGSLTKTDPQSHPVVIIGQVKNLAKVKFSDIKVKLEPRVTEEVYLTALASLH
PSPTDTCSLYLNLATLAALPVKCSRHNAPSRAHAITKIVRSTTVGADESIVIVCERSDVF
ASACAVARAYPLFSRKTSSNPAGQTPNNTTVSVEFLLVPSAECQVTVESNGSLDDKLALSA
DDLKCLESACTGVRLAARIVDAPCNDMNVDHFLQEVRAVADSLGVTSTIIRGEELRVRGF
GGIYGVGKAASVPPALAVLSHNPEGATQTIWVGKGIVYDTGGLSIKAKTAMPGMKRDCG
GAAGVLGAFYA AVRQGFQNLHAFVCLAENAVGPDATRPDDIHLLYSGRTVEINNTDAEG
RLVLSDGVSYAQKDLKANIILDMATLTGAQGISTGKYHG

>3433

LLLGCLAAPKIEGVLIDLARNVRIVNGQPAEDGEYPFQVSLQNSGRHYCGATVLNNEYI
LTAHHCYGGNIDSYTVRAGTNNYNSGGSVHQQVVEWYHQEYQNQAYWYNDVAVVRVDP
LTYSSKVQPTKLPAAAGEEPAGGAASVAIGWGS GCYGC SG IADLQEVNLRISYSECLDIY
GDGPTADMVCSGIP EEGKGVCSGDSGGALLVDGVQVGITSWTRMPCASYPVWTKVSHYR
DWIKEQSGV

>3434

MKLHLSVFSTLLLSCLGSKADFQRPHADVARNGWVKHARRASEGKYPFQVSLQVFSEHEC
GATILDQEFLLTAAHHCYNGSMDDYTIRAGSNKYKRGGSVHNVIDWIYHENYLNHKYWK
DIAILKIDPPLTYSTKVQPTKLPVAGEEPTVGAASVAIGWGS KCYGCESIRHLHEVSLEI
YSYEQCLDSMQQPTTDMICSGIPGEAEVCSGDSGGALLVNGTQVGITSWLVVPCAFNP
AVWTRVSHYRDWIRNKSGV

>3482

MKKFVAFVGTLLVLLYLGTLTDGVTGDI F PETRIAEGEDSAFDEFYVVS IQYQTEEKGYI
HRCAGAMINSSYVTVAHCFQSVDT SKYRVVNDNDTH TENETKISNVSTIVIHPDYDSAN
FWNNDIAILQLLNGFEDYNQVLLPENYS DPDP ELK GTTLGWGSGEMNGQP VQMLQRLDDL
NIFDSTACEDLYIFGP NENSICAGNEDGSTNIGVCSGDSGNPLIINGTLYGLLSWLRI PC
FQLPGVYTKVSNYVEWIGHN

>3554

VSTILLACFLGVAVNANS DSKSHRTVNGSPVNGGEYPSQLSLQFFNIHICGASILNEIWA
LTTATCVHEYPTSQLSLLAGTINLQNIGSATRHRVQETVKYENYNAEKFWVNDIALVKVS
EPFILIGGNLKAIPLAQDQGVPEGS KATIIGWGAANGEILSTLQKIEIEIAAMDRCNMT
YSLITRPLDNTNICAGSVMDSKSICMGDYGGLIVNGGVVGIASVYVNCGKPIYLSICTS
VPTFRNWITLKTNL

>3624

MQLYLALLSTLVGSGYAPTNTKTIHVDFARNVRIVNGYYASPGEFFP FIVSLQYNGGHYCG
ATILNSEWILTAAHHCYGNIGAYTVRAGSTNYQTGGTLHSVVG GYHQYYVNSYPYNN
IALLQVSPPFSF DAYVSTTTLPQQGQDVP GGYP SVAVGWSSCDDCGSNVNLMEVDLYTY
SYAECVAIYGS GPTTDMVCSGNPSEGKGICFGDSGGPLL VSGVQVGVT SWTGT PCASVPA
VWTKVSSYRNWISANTGV

>3726

YAKGQKLEIVQQGLWYVREVGHEDSGGESYRSCELI SGSMFWQLTYPCGENVDYFGRGS
KQLSWNYNYGPFSSQVFGNPAVLLDEPAAVADSWLNFASGFWFSVTPQSPKPPMLWIDG
TWVPNEADKAAGRTPGFGATTNIINGGLECDGKGSSHPGPKNRAEAYKHFMREFGATVDP
DEDLDCQKMKPFDQTSSAATEAYYTMNWSGDAKCDLTSSQTAWQISRGDYARCVKYFFR
SHWFKDGT LVSDYGLTPGSSYS

>3779

MFYIILVTVLGVSFAGPLLRDVSDDWRIINGTNAEPGQYPYHVSLQSNGEHF CGATILTP
 TWCLTAAHCVDGKDV EKL SVLAGTVLLSKGGVRQAVQKYKVHEGYNSSNNWENDIALIQL
 QTPLEFNDLVQPLPQLQQGEETSPRTSVTVVGWGRSTSTGGERPDNLQTV DILTF SDEECR
 SILNDIIHTSQICAGLPEGHKGECNGDSSGALVADGFQVGIVSWSYKPCTKPPRPSGVYT
 QVSYYIDWIKDNIE

>3823

MLRYLVLASLIACSLSAVPKAKRPRLDGRIVGGRPADIADYPYQLSFEYYGSHMCGASII
 SPNWVVTAAHCVDGVSASSATFRAGSSIRESGGSVHQATQLIANPNYDYYTIDFDVAVAR
 VSPAFSYGTGVQPIPLASSEPSAGQIATVSGWGTTSEGGSTLPSQLQVVSVPPIVSRSECN
 QAYSDYGGITDNMICAAEQGGKDACQGDSSGGLVVNGQLAGIVSWGVGCAEQGYPGVYS
 NVASLKG FITEQTGVN

>3824

MLRFLIIASLVACSLGAAPRIRRPRLDGRIVGGSDADIANYPYQLSFEYNGSHRCGASII
 SENWVLTAAHCVDGISANNVQFRAGSSIRSGGSLHRASQLIENPAFDYYTLDSDVAVAR
 VSTPFSYGNVQPISLATSEPSAGEIATVTGWGTTSSGGSLPSQLQVVS LPIVSRAECNQ
 DYSSYGGITDTMICAAEQGGKDCQGDSSGGLVVNGKQAGIVSW

>3851

DWSNLLSLEPQLPNIPLPKVLPNAGSRITGGEEAALHQFPYQAAVLVHTPDGTGFCGGSL
 LSPEWVLTAAHCADPADSFTVILGAHRVMETEDTQVRFDVTEKHVHPAWDPLRISNDVAL
 LRLPSSVQLSQHIQTLRLPTLSQKDET FADVM AVASGWGRSDNSEISPVLRWVRTPIT
 TNYMCGVL YFGAINDGHICVSGTGGKSTCNGDSSGGLVVEDAEGRPTQVGVSFSGISFGC
 EIGWPAAFARVTTYLDWILSTTGIAPRA

>4015

SVHNLTFPRKNTRL MGLYASNQLLV TQCEPEGFRITYYIDQPSNMAKFTVKLVGSKVDF
 PYLLSNGNEISKEDLGDKHAATFVDPFPKPGHLFAICAGELTFEKANFTTMGGRKVDIYV
 MAEPPYAGQLKWALDMAVAFAFWDEVTFGREYDLDTFRAIGVSDFNMGAMENKSLNLFNI
 SAIVAKEDRTRDWFYKRAHVVAHEYFHNWSGNRVSVQDWFQVQLKEGFTSFRESRFIED
 TFGKGLTRISQVNDLRSRQFTEDAGPTSHSI

>4033

QISLQQVIFIFRQHICGGSII SPTWVVTAGHCVVIGDLEV VAGEHSLNNNDGTEQTVAVA
 QAIVHPNYAGEVGPYDIAVLKLSSALVFNDYVQSIALPPAGSIPSGSAILTGWGSTSQTN
 NPSMPDILQTATLPIVEYDECVRTYGEDSPLDPTNVCTGPLSGGMSACSGDSGGPLAVRR
 SDGTFELIGIVSWGEIPCGQSNAPSVYTRVSAYIDWINNNAK

>4043

VLNDQWALTAAHCEGLVPEDLILLAGTVDLKTGGTTHQVVEIIMHEGYTPEDSWMNDIA
 VVRVSN SFPIDGVNVEPISLPAQGQNPADGSTATFIGWGYNENSVLPTILSKADIAIVNQ
 GTCSNAYASQGQNIYDGNICAEAPQGNKGACTGDDGGPLFVEGTSVGLMSWMNACNDPGY
 PTVFTRVSYRDWIKKHTGV

>4045

ILVTL LSLSGALGAPAADTDLGERIVGGEDAPLGQYPFQISLQYNNQHMCGGSILTENYC
 LTAHCVDGVYVGLSVLAGTVSLSSGGQRRVVASHTMHEGYDPYDQWINDIAVVKLSSPL
 ELGSYVQPVPLLGQGVETPGGTPATVIGWGRLATGGSAPDKLQMV DILVYSDEECNALHN
 GATHPTNICAGVPEGKGQC SGDSGGPLIADGYQVGIVSWSVKPCTVAPYPGVYTQVTSY
 IDWIKQNTDRT

>4127

IPSDDAHQSEYLASCDRRRAYISGFTGSAGTAIVTENEACMWT DGRYFLQASEEMDKNWT
 LMKEGLPSTPTQAVWLTKNLPPGSRVGTDPNLVSYSIWKPLVTQLEAAGLSLVPITTNLI
 DVIWEDRPTPPCNTIQPLSVKYTGKTKAKVEDVRTQMAEKGARLLVVSALDEAAWLLNL
 RGSDIQYNPVFFSYVIVTKTEVHLFVDEAKVSLAVYNHFNEEDLQVTVHPYEKFQTF LMD
 QISEKDGKIWISPESSYAVS

>4190

MYLLALASLLACTGAVSLAGRPNTDGRIVGGTPAEITNYPYQVSMEYFFSHTCGASII SP
 EWWVTAHCVDKKEGQLSVVRFRAGSTYRQQGGTLHNAAEIIMHPQYSDYDYDIAVVRVK
 EPFEYSDSVQPI SLTSTAPATGTSVVVTGWGRLAEDGEVPYQLQQVQVNIVDFEECNSDY
 WLYGGITTRMICAGVPDGGKDACEGDSGGPLVAEGELVGVSWSGSGCAQPGYPGVYADVA
 VFKSWTSVT

>4260

SNGRIVGGTDADIANYPYQVSMQVNNGHFCGGSIIISTEWVTAHCVKDIPINVVSFRAG
SSNQGTGGTVHPAEQITYNPNFSEQTADYDIAVAKVSVLFVFSNTVQPIKFTSVAPPTGT
PVVVTGWGALYEDGPAPEVLQQVQINIVDFEEDSVYGGITQRMICAGVPQGGKGVCGGD
SGGPLVANGSLVGVVSWGAFCAAKGYPVFSDVAVLSSWVTFITGVS

>4268

GIGDLKGITENFDHLKDIGVDGVWLSPIFPSPMADFGYDISDFRNIDPMFGTLEDFDALL
AKANELGIKLILDFVPNHSSDEHEWFVKSLQKIDPYTDYYTWNDGKLDESGNRVPPNNWQ
AEFGGPAWAWRDERQQYYLHRFDPKQPDNLNYSNPVVEEMKDVLRFWMDRGVDGFRVDAV
PWLFEGPIDVPDSETVAPQNLPEYDMVVKWRKVLDEKSAELGKTEMFMLEVYGNTEQVM
GYYG

>4335

DYLYQLSLQSTWQHCKGAVMISNEWAITAARCTYGRNVVYMSVRTRQSSLSSRMGRRVGL
ERILYYNNNETYENDIAAIQLYLPSRASFEWIPMRTSPVEEGTTGFVLGWGKLKEDGT
SDVLQKVSVTIQNTDKCNKILQDNGLGDNMHGQMCSMDGKNVCQGNYGTPFIVDGKLA
GMVSWGKGCGETGNPIVYTEISYYYNNWLKNVIAEG

>4370

KYIGDEDFRKGMLYLTRHQYRNTFTEDLWSALEEASNKPVGAVMSTWTKQMGFPVVKVT
NEQMGNDRLKLVITQEKLLCADGSSNSNGCLWMIPVMISTSEKPGEVVHHTVLQTHSVEVI
LKNVASHEWVKVNPVGTGYYRTQYPPDMLAQFVDAIRDKTLPPLDRLGLLDDL FAMVQAG
HTPTVEVLKLMKAFSNEDNYTVWSSISNCLSKLSILLSHTEYEDLFKAYGRQLMQ

>4544

ASVLASALAVADLDLPDEVHFIGMCCANLVDQHAYLVSDVVTASNGKTIEVTNTDAEGR
LALCDGLVFAENVMGGPGKGTIVDVATLTGGALVALGPDMAAVMSNKQELADAIQQVSDA
TGELTWQIPMREAYMEHLKSKTADIVNVNMGWPNASTMVGGFLKQFVSKETAWAHVDA
GPMWNGSENAGTGWGVYLITQLLLSQTKN

>4659

VNNGTGRQEITVFSLDKQLVVLVRGGFIIPVQEPKVIVRETLSPTYLLVALKCTVSNRC
DASGSLYVDSLDLFSFTATEQGVIEQSSCNDSTSSPILDAVTYGLDYGIGCDSVNIVNGS
AMPCFLKDQVLKITDMNLQLCNREGIIMEWNLTNIPESTTLKTTAPTIPSTSDDTIDSE
NTDDSTDTAPSTTSPETDDSATDTGPSTISP

>4869

RYGPELVDGMNFLGQLLPGIGVTYDGEEIGMENTYITWDQCRDPQGINAGPEHYLERTRD
LERTPFQWDDTTSAGFSTNASTWLPVNPYKELNLAAEKSAEKSHYKVYKQLIELRKPT
IQGGGTQVDALSEKVLTFARFLEGESYIVLVNLGDETETVNLKNTFDWLGDVTIYAVN
IQSQHMGVHVHSTTDFEIA

>4940

WDATTSAGFSTNATTWLPVNSNYKELNLDAQKATEKSHYWVYKKLVQLRQSQSIQYGD
LQIVAFSEQIFAFSRSYPDSESYVVIVNLGSSVESVTIDSSIQTLVYRVASVQSMRNEGDEV
EAIYFSMSPKEALVLSTSP

>4947

KNHSQEIRPIVGGKKANISDFPYQVSLAYYLAEDSEKNQETSHYCGGSIIISVEWVLTAAH
CVHDINPEDIVVRAGSCVRNRGGQLHKVALIIHDQYDYDTNDFDIALLKVSKPFKFRK
VQPIALNRNPSLKAGGLAVVSGWGDTEGGELAEELQVVKIPLPHWECKWLYKPKKITT
NMFCAGRSEKDAC

>5034

QLRKEEAPLGLNCTTSGWGLLDSEEYPDILQIVSVPIPEDECRQIYELNNYTENEIYPG
MTCAGYEQGGKDACAGDSGGPLQCGLLTGIVSWGAGCAEPNYPGVYADVYYKEWIATQ
LAKSRIKPLSIPVQHITS

>5056

LGDFPGLTWHVDMAYPAYIGERATGYGVALLVALFGVYSQDRMLRSLSPVSRNCGPEMG
NACKKLKRN

>5255

IKGNCLDTLARKFLWDVGLDYMHTGHGIGMYLNVHEGPMGVSWRVYPNDPGLQEGMFIS
NEPGYYEDGKFGFRIEDIIRIVPAKTPHNFKDRGYLTFETVSLVPIQTKMLVPEMLTEKE
IRHLNEYHNRCREIVGPHLKKMGHNEAREWLYRETEPIG

>5285

NYQGSTDLYIWNDMNEPSVFNGPEVTMHKDCVHHGGWEHRDIHNIYGLLHTMATYEGLL
QRSGGKLRFILTRSHYAGSQRFSTSVWTDGNTAEWGHRLRITIPMCLSLAVAGISFCGADV

GGFFRNPDSELFTRWYQAGAFQPFMRTHSHIDTKRREPWLFDEQTLTLIR

>5314

SSHEVSSITVHSGYSAANSWRNDIALKLLVPIAFTSLQSAVTLPAQGENAGAGTACIAS
GWGYTSYPGAGLPNNLQKVSLSVISNADCQASHTGTIYSTICASAGAGKGVCNGDSGGP
LTSNGKQIGIVSWGRPCATGVDPVYTRVSEYITWINQNAV

>5636

GGFLLARLVKYTTENFASEEHATEVEAFFRDHYSPGTERTVQQSCETIRLNAAWLQRDRE
NIHKYLSSGK