CHARACTERIZATION OF THE TRANSCRIPTOME AND THE PEPTIDASE AND THE NON-PEPTIDASE HOMOLOGS OF CALLOSOBRUCHUS MACULATUS

BY

ALICE M. VOSSBRINCK

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Entomology in the Graduate College of the University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Professor Barry Pittendrigh
ABSTRACT

The cowpea bruchid, *Callosobruchus maculatus* is a major pest of the cowpea, which is an important crop in many areas of Africa and is also consumed in other parts of the world. The cowpea utilizes serine peptidase inhibitors as a defense against herbivorous insects. Peptidases have been considered a target for pest control. The cowpea bruchid in turn employs cysteine peptidases as its major digestive peptidase to avoid the plant’s defense. Important not only in digestion, but in other functions such as immune response and molting, the peptidase and non-peptidase homologs were annotated. Here a de novo transcriptome assembly was built from 840,766 Roche 454 reads using CLC Genomics workbench. The assembly resulted in 35,111 contigs that were analyzed using Blast2GO.

Important not only in digestion, but in other functions such as immune response and molting, the peptidase and non-peptidase homologs were annotated. A local BLAST search against sequences downloaded from the MEROPS database found 264 putative peptidase and non-peptidase homologs for *C. maculatus*. Illumina sequencing from the same sample produced 295,690,882 reads which were mapped back to the assembly using BWA software. While *C. maculatus* does not appear to have an increased number of cysteine peptidases compared with other insects it does appear have more highly expressed cysteine peptidases compared to serine peptidases. Examination of the peptidase and non-peptidase homolog composition of *C. maculatus* might ultimately lead to innovations in pest management or other biological insights.
ACKNOWLEDGEMENTS

I would like to thank my advisor Barry Pittendrigh and my committee members Hugh Robertson, and Stewart Berlocher for their help and guidance. I would also like to thank Kim Walden and Carmen Valero for their advice on methods and helping me to understand genomic techniques. I am grateful to Weilin Sun, Tolulope Agunbiade and Brett Olds for teaching me to use the software and their day to day advice on the minutia of this project. I would like to say thanks to my parents for their edits and advice. Finally I would like to acknowledge my fellow graduate students who were extremely generous in their time, advice, expertise, troubleshooting, edits and most of all their constant encouragement and support.
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARACTERIZATION OF THE TRANSCRIPTOME AND THE PEPTIDASE AND THE NON-PEPTIDASE HOMOLOGS OF CALLOSOBURCHUS MACULATUS</td>
</tr>
<tr>
<td>Introduction.................................................................................................................. 1</td>
</tr>
<tr>
<td>Materials and Methods................................................................................................. 4</td>
</tr>
<tr>
<td>Results.......................................................................................................................... 8</td>
</tr>
<tr>
<td>Discussion..................................................................................................................... 12</td>
</tr>
<tr>
<td>Tables.......................................................................................................................... 21</td>
</tr>
<tr>
<td>Figures......................................................................................................................... 24</td>
</tr>
<tr>
<td>References................................................................................................................... 30</td>
</tr>
</tbody>
</table>

| APPENDIX.......................................................................................................................... 34 |
CHARACTERIZATION OF THE TRANSCRIPTOME AND THE PEPTIDASE AND THE NON-PEPTIDASE HOMOLOGS OF CALLOSOBRUCHUS MACULATUS

Introduction

Originating in the savannah of West and Central Africa, cowpea (*Vigna unguiculata*) is a vital crop in many tropical and subtropical areas of the world. Africa produces about 94% of the world's cowpeas (Jackai, 1986; Langyintuo *et al.*, 2003, Abate *et al.*, 2011) with Nigeria and Niger being the top producers in West Africa (Jackai, 1986). Cowpeas are a particularly important crop because they provide a low cost source of protein in areas where income ranges from $140-$1060 US dollars per year with 6.2 million metric tons produced annually (Jones, 2002; Abate *et al.*, 2011). Cowpeas are able to produce yields in harsh conditions, growing in poor soils with as little as 300 mm of rain and often remain productive when other crops fail due to drought (Jackai, 1986; Langyintuo *et al.*, 2004; Timko *et al.*, 2008). Additionally, cowpeas are important to food security in West Africa since they mature before the grain harvest and provide a food source when grain stores are running out (Langyintuo *et al.*, 2004). The cowpea bruchid, *Callosobruchus maculatus* (Fabricius) (Coleoptera: Chrysomelidae) is the primary pest of cowpeas during storage and is widespread with a global distribution (Janzen, 1977; Jackai, 1986). *C. maculatus* has a rapid rate of population growth and can achieve 100% infestation of a store in only 3-5 months (Boeke *et al.*, 2004). It causes at least $30 million dollars in cowpea losses each year in Nigeria alone (Jackai, 1986). The importance of this pest has caused several international agricultural organizations such as the International Institute of Tropical Agriculture (IITA) and Consultative Group on International Agricultural Research (CGAIR) to dedicate resources to controlling the cowpea bruchid. A variety of methods including chemical and physical control are employed to combat the cowpea bruchid. Production of transgenic cowpeas has also been studied as a way to control cowpea bruchids (Poswal and Akpa, 1991; Popelka *et al.*, 2004). New
and potentially more effective strategies for controlling *C. maculatus* will require a more in-depth knowledge of the ecology and genomics of this important pest.

One proposed strategy for managing *C. maculatus* is to inhibit its ability to acquire nutrients from its host. Many plants including cowpea produce peptidase inhibitors as a means of defense against insect herbivory. Serine peptidases are the most common and most studied insect peptidase, but *C. maculatus* and other Coleoptera rely primarily on cysteine peptidases, which are active in the acidic conditions of the *C. maculatus* gut, as their principal enzymes for protein digestion (Murdock *et al.*, 1987). In response to ingestion of cysteine peptidase inhibitors, *C. maculatus* compensates in three ways, first by producing more cysteine peptidases that are insensitive to the inhibitor, second by utilizing an increased number of peptidases of other classes and third by production of enzymes to degrade the inhibitors (Zhu-Salzman *et al.*, 2003). Additionally, *C. maculatus* increases production of enzymes associated with stress when exposed to peptidase inhibitors (Nogueira, 2012). When challenged with a dietary soybean cysteine protease inhibitor soyacystatin N (scN), 94 transcripts were shown to be up or down regulated. This included genes that encoded protein and carbohydrate digestive enzymes, some of the most important of which were the cathepsin B-like cysteine peptidases (Moon *et al.*, 2004). To gain a better understanding of the diversity and composition of peptidases in *C. maculatus* an investigation of the genome and its transcriptome products is necessary. Genomic and transcriptomic information may also lead to other insights about the insect’s biology that are important to understanding the pest and its life-history.

The genome for *C. maculatus* has not been sequenced but is estimated to be about 1 giga base pairs (personal communication with S.J. Johnston). Before taking on such a large endeavor an understanding of the nature of the *C. maculatus* genome is essential. Sequencing and
annotating a full-body transcriptome for *C. maculatus* will facilitate annotation of the genome. A transcriptome assembled from 454 data from all life stages of *C. maculatus* will provide insights both specifically into the composition of the peptidases and generally into the expressed portion of the genome. Here I present a brief description of the *C. maculatus* transcriptome with a detailed analysis of the peptidase composition.
Materials and Methods

Rearing of Insects

*C. maculatus* lab colonies were collected from locations in Indonesia, Nigeria, Cameroon, Niger, Ghana, USA and Puerto Rico and have been maintained in culture for over two decades. Insects were kept at 24° C at 35% +/- relative humidity and reared on cowpea seeds (*V. unguiculata*). A mixed lab culture was created by combining individuals from all populations and allowing them to breed for at least three generations. The mixed lab strain was used since the more inbred populations are prone collapse due to inbreeding and also because these datasets were used to investigate potential molecular markers beyond the scope of this project. All life stages of both sexes from a subset of the population were used and all insects were pooled for RNA extraction. Eggs were also included in this mixture. Larvae from all life stages were identified using a biomonitoring system that uses sensitive microphones to identify feeding behavior and molting periods (Pttrandrig, 1997).

RNA isolation

Total RNA was extracted from the pooled sample using the Qiagen RNeasy extraction kit for whole body extraction (Qiagen, Valencia, CA, USA) according to the manufacturer’s instructions and genomic DNA was removed by on-column digestion with DNase. RNA quality was assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific Inc. Wilmington, DE).

The RNA sample was submitted to the Keck Center at the University of Illinois at Urbana-Champaign where library construction was performed according to the GS FLX Titanium General Library Preparation Method Manual (Roche). The cDNA library was normalized using the Trimmer Direct Kit (Evrogen, Russia) as described by Dassanayake et al., (2009). Samples were assessed for purity using an Agilent Bioanalyzer and sequenced on a half
plate on the Roche 454 sequencer (Roche, Indianapolis, IN, USA). Raw sequence files were downloaded in the format .sff. A second portion of the same RNA sample was submitted for Illumina sequencing. The library was prepared with Illumina’s TrueSeq RNAseq Sample prep kit and quantified by qPCR. The sample was sequenced with the Illumina HiSeq2000 (San Diego, CA, USA) using 100 sequencing cycles for each end in one lane using a TruSeq SBS sequencing kit version 3. The paired end reads were analyzed with Cassava 1.8 (pipeline 1.9) and downloaded in the format .fastq.

Sequencing and de novo transcriptome assembly

The raw sequence reads from the whole-body RNA extractions were trimmed and analyzed with CLC Genomics Workbench version 6.0.1 (CLC Cambridge, MA, USA). The adapters (454- 5’ CCATCTCATCCCTGCGTGTCTCCGACTCAG-MID-, 5’ CCTATCCCCTGTGCTTGCAGTCTCAG-, Illumina AGATCGGAAGAGCACACGTCTCAGTGATGTGATCTCGTATGCCGTCTTCTGCTTG, AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCT) were removed and the reads were trimmed with a minimum Phred quality score of 20, and minimum read lengths of 100 bases for 454 and 50 bases for Illumina.

The 454 trimmed sequences were de novo assembled using CLC with a k-mer (word) size of 21, minimum contig length of 200, 80% similarity, mismatch cost of 2 and an insertion or deletion cost of 3. The threshold of 80% similarity was used for assembly since a mixed population was used and assemblies with higher similarity did not yield better results.

Human and bacterial sequence contamination was examined using the web-based version of DeconSeq with a query coverage and sequence identity thresholds of 90% and 94% DNA identity respectively (Schmieder and Edwards, 2011).
The assembled contigs were searched for open reading frames using CLC-Genomics Workbench 6.5. The parameters were any start codon, no open-ended sequences, searching both strands and a minimum length of 100 codons.

**Functional annotation of peptidases**

The publicly available Blast2GO software was used to acquire BLAST and GO terms (Gotz et al., 2008). The assembled transcripts were subject to a similarity search using the BLASTx algorithm against NCBI’s non-redundant database with a cut off e value of e^{-6} to avoid false positives (Atschul et al., 1997).

The assembled CLC-assembled contigs were subjected to a similarity search against the MEROPS database, which contains all known proteolytic enzymes and their inhibitors ([http://merops.sanger.ac.uk/index.shtml](http://merops.sanger.ac.uk/index.shtml)) (Rawlings, 2012). A local database of peptidases was created for insects with sequenced genomes by downloading their peptidase sequences from the MEROPS database. To determine which insect species had the greatest peptidase similarity to *C. maculatus*, the *C. maculatus* de novo assembly was queried against the insect proteins using the BLASTx algorithm in CLC. The insect species with the most sequence similarity in the Orders Coleoptera, Hymenoptera, Diptera and Lepidoptera were chosen for further investigation. Sequence similarity searches were carried out in CLC by querying the *C. maculatus* de novo assembly sequences against the local peptidase database using the BLASTx and tBLASTn algorithms. Since *Tribolium castaneum* is in the same order as *C. maculatus* an e value of $10^{-40}$ was the cut off for accepting a match. For the other insects (*Apis mellifera, Nasonia vitripennis, Danaus plexippus, Anopheles gambiae, and Drosophila melanogaster*) an e value of $10^{-15}$ was used as the cut off value.
The local BLAST results from *T. castaneum* and *D. melanogaster* along with the seven sequences previously annotated for *C. maculatus* were used to annotate the putative peptidases and non-peptidase homologs. When there were multiple acceptable hits that spanned different regions of a gene the *C. maculatus* contigs were aligned and combined to create more complete transcripts.

Names assigned to the putative peptidases and non-peptidase homologs were based on the name assigned to the homolog in the MEROPS database. For the ease of reading in tables the peptidases for *C. maculatus* were given ‘nicknames’. These nicknames consist of first the species indicator, CM for *C. maculatus* followed by a period. Next is the MEROPS assigned ‘species name’ or identifier for a unique peptidase (Rawlings *et al.*, 2012). Finally, some nicknames were followed by an underscore and a number that was then assigned when sequences were subunits of a peptidase and not given a unique species name.

**Relative transcript levels of peptidases**

The number of reads of a transcript varied greatly and this variation was used as a proxy for the level of expression of transcripts. The Burrows-Wheeler Alignment (BWA) software (Li and Durbin 2009) was used to calculate the relative expression levels of the contigs created for the *C. maculatus de novo* assembly using the set of Illumina trimmed reads mapped to the CLC-454 *de novo* assembly. The MEM algorithm was chosen because it performed better with Illumina reads of a length 70-100 bases than the original algorithm. Only the peptidase and non-peptidase homologs were examined for difference in expression levels. Expression level is reported as number of Reads Per Kilobase Million Mapped (RPKM) to correct for the difference in transcript length (Mortazavi *et al.*, 2008).
Results

Sequencing and transcriptome assembly

Roche 454 sequencing of the *C. maculatus* larval and adult transcriptome produced 904,805 reads with an average length of 356.6 base pairs. After trimming there were 840,766 reads with an average length of 336.5 base pairs. Using the CLC Genomics Workbench *de novo* assembler resulted in an assembly with 36,113 contigs covering a total of 27,122,286 base pairs. Contig size ranged from 25 to 14,282 base pairs with an average contig size of 751 base pairs and an N50 of 955 base pairs. Before analysis contigs below 200 base pairs were discarded, this left a subset of 35,118 contigs greater than 200 base pairs. A search for open reading frames (ORFs) resulted in 21,353 ORFs. The sequencing and assembly results are summarized in Table 1. Contamination of the sample was negligible; only 7 contigs (0.02%) were found to have bacterial or human contamination using the program DeconSeq.

The Illumina sequencing produced 295,690,882 reads with an average length of 100.0 base pairs. After trimming there were 272,588,974 reads with an average length of 94.0 base pairs.

Functional Annotation

Contigs were analyzed using the Blast2GO V.2.6.6 program. Blast2GO did not process sequences greater than 8000 base pairs. Seven of the *C. maculatus* contigs were greater than 8 kb. Of the 35,111 contigs analyzed 18,224 (52%) had BLAST hits, 15,036 (43%) had GO mapping results and 10,298 (29%) had annotation results. The greatest number of BLAST hits for *C. maculatus* were from *T. castaneum* followed by *Dendroctonus ponderosae* (Figure 1). Both species have sequenced genomes and belong to the order Coleoptera and should have greater similarity to *C. maculatus* than other organisms. The majority the BLAST hits were to other
insect species (Figure 1). The seven sequences too large for Blast2GO were submitted to a BLASTx search at NCBI; five of them had significant hits with \( e^{-40} \) or less and similarity greater than 75%. They aligned fully to a sequence in the NCBI database and despite their size were not chimeric sequences.

The GO results are subdivided into functional categories at several levels. In the level 2 GO hits for *C. maculatus* the most abundant categories were binding and catalytic activity in the cellular components (CC), intracellular functions in the molecular function (MF) category and metabolic and cellular processes in the biological processes (BP) category (Figure 2).

**Peptidases and non-peptidase homologs**

Plants, through the use of peptidase inhibitors, target insect digestive peptidases, in order to decrease insect herbivory. The most studied peptidases are those with a digestive function and these are considered a potential target for insect pest control. Peptidases have other functions besides digestion including immune defense, molting and degradation of cell products. Non-peptidase homologs lack proteolytic activity but are homologous with peptidases and serve functions other than breaking peptide bonds such as maintaining structural integrity of proteins. Figure 3 shows the number of peptidase and non-peptidase homologs that *C. maculatus* has in common with insects from several orders. As one would expect the greatest number of homologs was shared with the beetle *T. castaneum*. The two Diptera representatives had the second greatest similarity with *C. maculatus* followed by the Hymenoptera and finally the Lepidoptera.

Annotation of the *C. maculatus* transcriptome revealed 264 peptidases and non-peptidase homologs. This number of peptidases is comparable to that of other insects being about half the number of peptidase and non-peptidase homologs annotated in *T. castaneum*, but two and a half times the number found in *Bombyx mori* (Table 2). Peptidases are typically categorized by
catalytic site into cysteine, serine, aspartic, threonine and metallopeptidases. MEROPS categorizes peptidases and non-peptidase homologs based on phylogenetic relationships of the peptidase and non-peptidase homolog genes with a mixed category containing peptidases with serine, cysteine and threonine catalytic sites (Rawlings, 2012). The number of peptidase and non-peptidase homologs belonging to the different catalytic types varied among insects. Figure 4 shows the percent composition of catalytic types of peptidase and non-peptidase homologs in seven different insects from five different orders while Figure 5 shows the total number of annotated peptidases and non-peptidase homologs by catalytic type. The greatest number of peptidases and non-peptidase homologs for C. maculatus with 84 homologs fell into the metallopeptidase category followed by 66 homologs for mixed, 55 homologs in the serine category, 49 homologs in the cysteine category and 10 homologs in the aspartic category. Within the mixed peptidases and non-peptidase homologs found in the C. maculatus transcriptome seven were of the cysteine catalytic type, 41 were of the serine type and 18 were of the threonine type.

**Relative transcription levels of peptidase and non-peptidase homologs**

The BWA analysis program was used to align the trimmed Illumina reads to the 454-CLC de novo assembly to examine the relative expression levels as number of reads per kilobase per million mapped reads (RPKM) in the peptidase and non-peptidase homologs of C. maculatus. These ranged from 101 to 1719089 RPKM. Of the twenty most expressed peptidases and non-peptidase homologs (Table 3) ten were of the metallo catalytic type, five were of the cysteine type, three of the mixed group all with a threonine catalytic type, one of the serine type and one of the aspartic type.
Of the different catalytic types of peptidase and non-peptidase homologs the mixed category had the greatest number of RPKM with an average of 33893 (+/-592 SE) RPKM. Of the mixed category the three catalytic types varied greatly in their expression level. The threonine catalytic type at 109577 RPKM had the highest RPKM followed by cysteine at 107357 RPKM and then serine 4621 RPKM. Second highest were the aspartic with 14045 (+/-9092 SE) RPKM followed by cysteine with 10977 (+/- 2953 SE) RPKM and the metallopeptidases and non-peptidase homologs with 10328 (+/- 1662) RPKM. The lowest average RPKM was of the serine peptidases at 4605 (+/- 704). Figure 6 shows the distribution of number of RPKM for the five types of peptidase and non-peptidase homologs.
Discussion

The CLC de novo assembly of C. maculatus with 840,766 reads covering 36,113 contigs provides adequate coverage to ensure reliable transcripts. The T. castaneum genome contains 16,404 genes and the recently completed draft genome of the mountain pine beetle (D. ponderosae) has 13,088 genes (Tribolium Genome Sequencing Consortium, 2008; Keeling et al., 2013). The 36,113 contigs in the C. maculatus transcriptome do not each represent a unique gene, but a transcript. Since the assembly was not created using an inbred line, gene polymorphisms may appear as different transcripts. Additionally one gene may not always produce the same transcript, when different arrangements of exons occur. This is known as alternative splicing and results in transcripts that do not align with other transcripts from the same gene (Lopez, 1998).

The number of contigs assigned BLAST hits was 18,224 (52%). Other recent transcriptome studies reported a similar level of BLAST matches to assembled contigs. In the midgut of the poplar leaf beetle Chrysomela tremulae 56.1% of transcripts had BLAST hits (Pauchet et al., 2009), in the western bean cutworm, Striacosta albicosta 39.4% had BLAST hits (Miller et al., 2012) and 33.0% had BLAST hits in the salt marsh beetle Pogonus chalceus (Van Belleghem et al., 2012). A large portion of eukaryotic genomes consists of non-coding regions and likely accounts for a significant portion of the transcripts without BLAST hits (Andolfatto, 2005).

In C. maculatus the Blast2GO level 2 terms share a similar distribution to that of other beetle transcriptomes from D. ponderosae, P. chalceus, T. castaneum and C. tremulae. Across all five transcriptomes in the biological processes category the metabolic and cellular processes GO terms have the highest distribution, under the cellular components category the binding and catalytic activity has the most GO terms and in the molecular functions category the intracellular
functions are the most numerous for all five insects (Van Belleghem et al., 2012; Pauchet et al., 2009; Park, 2008; Keeling et al., 2012).

The *C. maculatus* peptidase and non-peptidase homologs were annotated with a local BLAST against *T. castaneum* and *D. melanogaster*. *C. maculatus* had 227 transcripts in common with *T. castaneum* and 169 transcripts in common with *D. melanogaster*. With a total of 264 annotated peptidases for *C. maculatus* (see appendix) there is a great deal of overlap within the common transcripts (Figure 3). When comparing the percent composition of the different peptidases with that of *C. maculatus* (Figure 4) there is no obvious expansion in the cysteine peptidases or reduction in serine peptidases. Figure 5 showing the total number of peptidases in each insect also does not display any great difference in the number of cysteine or serine peptidases and non-peptidase homologs when looking at the overall picture.

The number of *C. maculatus* genes of each catalytic type from all life stages can only give us a partial understanding of the relative importance of the different catalytic types. *C. maculatus* relies primarily on cysteine peptidases for digestion (Silva and Xavier-Filho, 1991). Other functions such as molting and immune defense also require peptidases. The expression levels of the different transcripts can give further insight into how much *C. maculatus* relies on cysteine peptidases compared to serine peptidases. Using the BWA analysis to calculate RPKM as a proxy for expression level shows a general trend of cysteine peptidases and non-peptidase homologs expressed at higher levels than serine peptidases and non-peptidase homologs. Figure 6 shows the expression profile of all 264 annotated peptidases in groups of twenty. The 20 *C. maculatus* peptidases and non-peptidase homologs with the highest RPKM are given in Table 3. Five of the top ten are of the cysteine catalytic type, while one of the top twenty is of the serine catalytic type.
By examining the function of the ten most expressed peptidases and non-peptidase homologs in the *C. maculatus* whole body, all life stage, pooled sample we see what proteolytic functions are most utilized. Peptidases are involved in maintaining cell function, immune response, molting and protein digestion. Some peptidases may have multiple functions.

Cellular maintenance involves the constant production and degradation of products. Five of the ten transcripts with the most RPKM are involved in cellular maintenance. With the highest number of RPKM, the transcript, CM.T01.012 (mixed-threonine) is one of the catalytic subunits of the 20S proteasome, which is involved in the degradation of intracellular proteins (Seemuller *et al.*, 2004). CM.M76.001 (metallo) had the third highest RPKM and is ATP23 peptidase, which is involved in the assembly of mitochondrial ATPase (Wilmes and Langer, 2012). CM.A01.009 (aspartic), with the fourth highest RPKM, is cathepsin D, which is found in humans and mice as well as insects. Knock down of the gene in *D. melanogaster* leads to neuronal ceroid lipofuscinoses, which are caused by a buildup of lipofuscin (lipopigments) indicating that this peptidase is needed to break down lipofuscin (Myllykangas *et al.*, 2005). It is also utilized in digestion in the Colorado potato beetle, (*Leptinotarsa decemlineata*) (Brunell *et al.*, 2004). CM.M17.UPW (metallo) has the seventh highest RPKM and belongs to a group of peptidases called leucine aminopeptidases, which have been found in mammals and microbial organisms. These proteins are involved in maintaining the cell, but evidence of use in defense, membrane transport and meiosis has also been found (Matsui *et al.*, 2006). CM.M16.003 (metallo), ranked ninth in RPKM, is a mitochondrial processing beta-subunit peptidase (Seemuller *et al.*, 2004).

Two of the ten transcripts were involved in immune function. With the fifth highest RPKM CM.C01.UPA_1 (cysteine) was shown in amphioxus, a fish-like marine chordate, to
likely be involved in immune defense (Wang et al., 2005). CM.C01.067 (cysteine), which had the eighth highest RPKM, functions optimally at pH 6 but is active in a range of pH 4-8 (Saito et al., 1992). This protein has been isolated from larval hemocytes, which are involved in immune defense and in Manduca sexta have been shown to be down regulated when parasitized, strongly indicating that this peptidase has a role in immune function (Saito et al., 1992, Serbielle et al., 2009).

The remaining four transcripts all had different functions. With the second highest RPKM, CM.C01.UPA_3 (cysteine) is a cathepsin-L like proteinase demonstrated to be involved in molting and metamorphosis in a lepidopteran (Wang et al., 2010; Beton et al., 2012). CM.M24.UNW_2 (metallo) had the sixth highest RPKM and is a non-peptidase homolog with an unknown function. CM.M01.013_1 (metallo), ranked tenth in RPKM, is an aminopeptidase N that acts as the receptor of the Cry1Ac toxin from Bacillus thuringiensis (Sivakumar et al., 2007). As mentioned above CM.A01.009 (aspartic), cathepsin D has also been known to act in insect digestion in addition to cellular maintenance. Two know digestive peptidases had low RPKM in the C. maculatus transcripts. Cathepsin O had 375 RPKM and chymotrypsin m-type 2 had 242 RPKM which is at the very low end of RPKM. A third digestive enzyme, cathepsin B has an RPKM of 45,484, which is towards the middle range. This may indicate that the expression of digestive enzymes is generally lower than that of other peptidases. Alternatively C. maculatus may not rely on these two digestive enzymes for its primary digestive functions.

Of the transcripts with the highest number of RPKM the function of one is unknown, four are involved in cellular maintenance and breaking down proteins the organism no longer needs, and one is involved in both cellular maintenance and digestion. Two of the transcripts are involved in immune function, one is involved in molting and metamorphosis and one is a target
for pesticides. Peptidases are essential in digestion and immune function but the bulk of insect peptidase activity may be devoted to cell maintenance.

Within the insect species represented in Table 2, which contains the total number of annotated peptidase and non-peptidase homologs, there is considerable variation. Diet and life history might account for a large amount of this diversity. Peptidases are important in both digestive and immune function and a large expansion in the number of peptidases might arise from challenges to the insect digestive or immune system or both.

The species with lowest number of peptidase and non-peptidase homologs at only 100 is the silk moth, *B. mori*. This species has been domesticated for thousands of years, does not exist in the wild and is oligophagous feeding primarily on mulberry leaves (Zhou, 1958; Zhou et al., 2008). The low number of peptidase and non-peptidase homologs could be a result of a limited diet, conditions with a lowered exposure to pathogens compared to wild insects or it could be an artifact of breeding. Alternatively this low number could be due to a less thorough annotation of the *B. mori* genome. With almost three times as many peptidases *A. mellifera* (honey bee) and *D. plexippus* (monarch butterfly) have limited diets. Adult *A. mellifera* eat pollen, honey and nectar while larvae eat special food produced by adult bees derived from the adult diet (Haydak, 1970). Larval *D. plexippus* feed on plants in the genus *Asclepias* and as adults they eat nectar from a wide variety of flowers (Oberhauser, 1997; Lavoie and Oberhauser, 2004). Both insects had comparable numbers of peptidases and non-peptidase homologs (*A. mellifera* 287, *D. plexippus* 278); one hypothesis is that their similar breadth of diets might account for this, however, this hypothesis remains to be tested. *Solenopsis invicta* (red imported fire ant) has 276 annotated peptidase and non-peptidase homologs, but feeds on arthropods, seeds, nectar, pollen and other plant material and has a much more diverse diet than *A. mellifera* and *D. plexippus* (Vogt et al.,
A largely predatory diet may not require a large diversity of peptidases and non-peptidase homologs, or immune requirements, not diet might be the greater driver peptidase expansion.

With 314 and 317 peptidase and non-peptidase homologs respectively *Megachile rotundata* (alfalfa leafcutter bee) and *Pediculus humanus* (human head/body louse) do not have a large increase over the previous three species. *M. rotundata* like *A. mellifera* eats pollen and nectar and it is not surprising they have a similar number of peptidase and non-peptidase homologs (O’Neil et al., 2004). *P. humanus* feeds entirely on human blood and so has a very narrow diet but a high chance at contacting a wide variety of pathogens (Burgess, 1995). Like *B. mori*, *P. humanus* has a very specific food source, but the greater exposure of *P. humanus* to pathogens might account for its having three times as many peptidases and non-peptidase homologs as *B. mori*. Additionally extracting nutrients from a blood meal, which contains a lot of protein, may be more challenging for an insect than obtaining nutrients from mulberry and may also account for an increase in peptidase and non-peptidase homologs. *Harpegnathos saltator* (Jerdon's jumping ant) is an ant like *S. invicta* but with 362 peptidases and non-peptidase homologs it has almost 100 more than *S. invicta*. Both of these ants are largely predatory, but unlike many ants *H. saltator* has small colonies with a looser caste differentiation (Bonasio et al., 2010). It has been proposed that *A. mellifera* uses colony level behavioral controls to detoxify its food source which could account for a decrease in detoxification enzymes, however it would be a stretch to suggest that a greater level of social behavior in *S. invicta* could lead to a lower need for peptidase and non-peptidase homologs (Claudianos et al., 2006).

*N. vitripennis* is a hymenopteran like *A. mellifera*, *S. invicta*, *H. saltator* and *M. rotundata*, but with 463 peptidase and non-peptidase homologs *N. vitripennis* has a significantly greater number than the other hymenopterans. Unlike the other hymenopterans *N. vitripennis* is
an ectoparasitoid that develops under the puparium of dipteran hosts from a variety of families including Calliphoridae, Sarcophagidae and Muscidae (Whiting, 1967; Darling & Werren, 1990). Unlike predators parasitoids need to be able to overcome their host’s immune defenses as well as any pathogens the host may have and this need may have fueled *N. vitripennis* peptidase expansion compared with the other hymenopteras.

The coleopteran *T. castaneum* has exactly 500 annotated peptidase and non-peptidase homologs. *T. castaneum* is a stored product pest and can survive on a wide variety of stored products such as wheat, rice, soy and corn and will also engage in cannibalism (Sokoloff, 1966; Via, 1999). Feeding on such a variety of plants *T. castaneum* likely encounters a variety of defensive compounds, which peptidases may help to degrade. Living in stored products *T. castaneum* may also encounter fungus or bacteria growing on the grains.

The three dipteran species *A. aegypti*, with 710, *D. melanogaster* with 752 and *A. gambiae* 836 had the highest number of peptidases and non-peptidase homologs. *D. melanogaster* is one of the best annotated genomes which might lead to more complete annotation of the dipteran species, but this would not likely account for the entire increase in peptidase and non-peptidase homologs. *D. melanogaster* lives on yeast and sugars of fermenting foods with some tolerance for ingesting alcohol (Baumberger, 1917). Fermenting foods foster the growth of a large number of microorganisms, many of which may challenge the immune system of *D. melanogaster*. Living in this challenging environment might be aided by having a large number of peptidases. *A. gambiae* and *A. aegypti* are both mosquitoes and have to meet several dietary and immune challenges. As larvae they are aquatic eating detritus or small organisms, while as adults they acquire sugars from nectar, honeydew, sap and damaged fruit and the females require a blood meal to produce eggs (Impoinvil, 2004). Taking blood meals exposes the
mosquito to a wide variety of pathogens and both these species are important disease vectors (Ahmed et al., 2002).

A more thorough investigation would be needed to make any conclusions, but there does seem to be a trend of an increase in the number of peptidase and non-peptidase homologs with an increase in complexity of diet or exposure to pathogens. It is unclear whether diet or pathogens are more important in driving the expansion of peptidases. Our conservative annotation of *C. maculatus* resulted in 264 peptidase and non-peptidase homologs. Feeding only on Fabaceae this number might be close to the total amount of peptidase and non-peptidase homologs for *C. maculatus*; however with only a transcriptome it is impossible to know exactly how many peptidase and non-peptidase homologs are contained in the *C. maculatus* genome. Like *T. castaneum*, *C. maculatus* is a stored product pest and it would be expected that *C. maculatus* would face similar challenges of fungus and bacteria. Underestimation can occur due to rejection of peptidases with less similarity to known peptidases than our cuts offs allowed, or the failure to resolve into separate genes *C. maculatus* gene groups with very high similarity.

A transcriptome and similarity based analysis of the peptidases and non-peptidase homologs of *C. maculatus* has been presented here. A sequenced genome will be necessary before a complete list of peptidases can be compiled. Even though *C. maculatus* does not appear to have an obvious expansion in its number of cysteine peptidases the cysteine peptidases are expressed at higher levels than the serine peptidases, which many other insect species rely upon heavily for digestion. The mixed and metallopeptidases and non-peptidase homologs made up a large portion of the total peptidase and non-peptidase homolog composition of *C. maculatus*. Peptidases are necessary for digestion but other functions such as immune defense and maintenance of the cell may require greater proteolytic activity. Further investigation into the
function of various peptidases will help to get a better understanding of the variety of functions that peptidase and non-peptidase homologs have which may lead to more effective pest control.

Coleopteran pests are a huge problem in agriculture and are targets for genetically modified plants, yet there is currently only one complete genome and one draft genome available for Coleoptera. Genomic and transcriptomic studies can offer insight on how to manage these pests. Transcriptome studies are more time and cost effective than genome projects and can provide valuable information for genome projects.
### Tables

Table 1. *Callosobruchus maculatus* transcriptome sequencing, assembly and annotation summary.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number of 454 reads</th>
<th>904805</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average length of reads</td>
<td>356.6 bp</td>
</tr>
<tr>
<td></td>
<td>Bases</td>
<td>322653463</td>
</tr>
<tr>
<td>Sequencing</td>
<td>Number of 454 reads</td>
<td>601179</td>
</tr>
<tr>
<td></td>
<td>Average length of reads</td>
<td>258.6 bp</td>
</tr>
<tr>
<td></td>
<td>Bases</td>
<td>155464889</td>
</tr>
<tr>
<td>Trimming</td>
<td>Total number of contigs</td>
<td>36113</td>
</tr>
<tr>
<td></td>
<td>Number of contigs &gt; 200bp</td>
<td>35118</td>
</tr>
<tr>
<td></td>
<td>Average length of contigs</td>
<td>763 bp</td>
</tr>
<tr>
<td></td>
<td>N50 of contigs</td>
<td>977 bp</td>
</tr>
<tr>
<td></td>
<td>Max length of contigs</td>
<td>14281 bp</td>
</tr>
<tr>
<td></td>
<td>Number of contigs with ORF</td>
<td>34575 bp</td>
</tr>
<tr>
<td></td>
<td>Bases</td>
<td>29703894</td>
</tr>
<tr>
<td>CLC assembly</td>
<td>Contigs with BLAST hits</td>
<td>18224</td>
</tr>
<tr>
<td></td>
<td>Contigs with GO terms</td>
<td>15036</td>
</tr>
<tr>
<td></td>
<td>Contigs with Annotation</td>
<td>10298</td>
</tr>
</tbody>
</table>
Table 2. The total number of peptidase and non-peptidase homologs in the MEROPS database for selected insect species.

<table>
<thead>
<tr>
<th>Order</th>
<th>Family</th>
<th>Species</th>
<th>Total Number of peptidase and non-peptidase homologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepidoptera</td>
<td>Bombycidae</td>
<td>Bombyx mori</td>
<td>100</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>Nymphalidae</td>
<td>Danaus plexippus</td>
<td>278</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Apidae</td>
<td>Apis mellifera</td>
<td>287</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Megachilidae</td>
<td>Megachile rotundata</td>
<td>314</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Pteromalidae</td>
<td>Nasonia vitripennis</td>
<td>463</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Formicidae</td>
<td>Solenopsis invicta</td>
<td>276</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Formicidae</td>
<td>Harpegnathos saltator</td>
<td>362</td>
</tr>
<tr>
<td>Phthiraptera</td>
<td>Pediculidae</td>
<td>Pediculus humanus</td>
<td>317</td>
</tr>
<tr>
<td>Diptera</td>
<td>Drosophilidae</td>
<td>Drosophila melanogaster</td>
<td>752</td>
</tr>
<tr>
<td>Diptera</td>
<td>Culicidae</td>
<td>Anopheles gambiae</td>
<td>836</td>
</tr>
<tr>
<td>Diptera</td>
<td>Aedes</td>
<td>Aedes aegypti</td>
<td>710</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Tenebrionidae</td>
<td>Tribolium castaneum</td>
<td>500</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Chrysomelidae</td>
<td>Callosobruchus maculatus</td>
<td>265</td>
</tr>
</tbody>
</table>
Table 3. The twenty most expressed peptidases and non-peptidase homologs in the *Callosobruchus maculatus* transcriptome.

<table>
<thead>
<tr>
<th>Name</th>
<th>Contigs</th>
<th>Group</th>
<th>Average reads/kilobase</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM.T01.012</td>
<td>contig11207 &amp; 3190</td>
<td>Mixed-Threonine</td>
<td>1719089.00</td>
</tr>
<tr>
<td>CM.C01.UPA_3</td>
<td>contig6861, 11080</td>
<td>Cysteine</td>
<td>108547.90</td>
</tr>
<tr>
<td>CM.M76.001</td>
<td>contig9139</td>
<td>Metallo</td>
<td>94504.52</td>
</tr>
<tr>
<td>CM.A01.009</td>
<td>contig92</td>
<td>Aspartic</td>
<td>93451.95</td>
</tr>
<tr>
<td>CM.C01.UPA_1</td>
<td>contig3897</td>
<td>Cysteine</td>
<td>82730.12</td>
</tr>
<tr>
<td>CM.M24.UNW_2</td>
<td>contig2947</td>
<td>Metallo</td>
<td>63152.87</td>
</tr>
<tr>
<td>CM.M17.UPW</td>
<td>contig702</td>
<td>Metallo</td>
<td>52182.30</td>
</tr>
<tr>
<td>CM.C01.067</td>
<td>contig116, 10338 &amp; 1099</td>
<td>Cysteine</td>
<td>51366.22</td>
</tr>
<tr>
<td>CM.M16.003</td>
<td>contig2272</td>
<td>Metallo</td>
<td>48478.84</td>
</tr>
<tr>
<td>CM.M01.013_1</td>
<td>contig2090</td>
<td>Metallo</td>
<td>46512.27</td>
</tr>
<tr>
<td>CM.C01.060</td>
<td>contig2334</td>
<td>Cysteine</td>
<td>45484.12</td>
</tr>
<tr>
<td>CM.M16.UNB_2</td>
<td>contig2970, 11479</td>
<td>Metallo</td>
<td>33880.38</td>
</tr>
<tr>
<td>CM.M16.UNB_2</td>
<td>contig847 &amp; 900</td>
<td>Metallo</td>
<td>32822.65</td>
</tr>
<tr>
<td>CM.C01.084</td>
<td>contig6674</td>
<td>Cysteine</td>
<td>32591.72</td>
</tr>
<tr>
<td>CM.S33.UPW_2</td>
<td>contig16677 &amp; 5452</td>
<td>Serine</td>
<td>32369.86</td>
</tr>
<tr>
<td>CM.T01.987</td>
<td>contig9462</td>
<td>Mixed-Threonine</td>
<td>31256.71</td>
</tr>
<tr>
<td>CM.T01.UNA_3</td>
<td>contig8362</td>
<td>Mixed-Threonine</td>
<td>29890.80</td>
</tr>
<tr>
<td>CM.M02.UPW_1</td>
<td>contig2325, 2326 &amp; 2271</td>
<td>Metallo</td>
<td>28480.50</td>
</tr>
<tr>
<td>CM.M14.A18</td>
<td>contig3115</td>
<td>Metallo</td>
<td>25583.70</td>
</tr>
<tr>
<td>CM.M01.UPW_2</td>
<td>contig592</td>
<td>Metallo</td>
<td>24504.52</td>
</tr>
</tbody>
</table>
Figures

Figure 1. The species with the highest number of top BLAST hits to the *Callosobruchus maculatus* CLC-assembled contigs using Blast2GO.

**Top-Hit species distribution**

<table>
<thead>
<tr>
<th>Species</th>
<th>BLAST Top-Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tribolium castaneum</td>
<td>10,000</td>
</tr>
<tr>
<td>Dendroctonus ponderosae</td>
<td>9,000</td>
</tr>
<tr>
<td>Bombyx mori</td>
<td>8,000</td>
</tr>
<tr>
<td>Acyrthosiphon pisum</td>
<td>7,000</td>
</tr>
<tr>
<td>Camponotus floridanus</td>
<td>6,000</td>
</tr>
<tr>
<td>Danaus plexippus</td>
<td>5,000</td>
</tr>
<tr>
<td>Aedes aegypti</td>
<td>4,000</td>
</tr>
<tr>
<td>Nasonia vitripennis</td>
<td>3,000</td>
</tr>
<tr>
<td>Harpegnathos saltator</td>
<td>2,000</td>
</tr>
<tr>
<td>Megachile rotundata</td>
<td>1,000</td>
</tr>
<tr>
<td>Strongylocentrus purpuratus</td>
<td>0</td>
</tr>
<tr>
<td>Hydra magnipapillata</td>
<td></td>
</tr>
<tr>
<td>Branchiostoma floridanus</td>
<td></td>
</tr>
<tr>
<td>Solenopsis invicta</td>
<td></td>
</tr>
<tr>
<td>Anopheles gambiæ</td>
<td></td>
</tr>
<tr>
<td>Pediculus humanus</td>
<td></td>
</tr>
<tr>
<td>Callosobruchus maculatus</td>
<td></td>
</tr>
<tr>
<td>Caprela teleta</td>
<td></td>
</tr>
<tr>
<td>Culex quinquefasciatus</td>
<td></td>
</tr>
<tr>
<td>Daphnia pulex</td>
<td></td>
</tr>
<tr>
<td>Ceratitis capitata</td>
<td></td>
</tr>
<tr>
<td>Acromyrmex echinatoides</td>
<td></td>
</tr>
<tr>
<td>Lepidoptera decemlineata</td>
<td></td>
</tr>
<tr>
<td>Chrysomela tremula</td>
<td></td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td></td>
</tr>
<tr>
<td>Metaseulcusoccidentalis</td>
<td></td>
</tr>
<tr>
<td>Glycine max</td>
<td></td>
</tr>
<tr>
<td>Nematostoma vectens</td>
<td></td>
</tr>
<tr>
<td>Bombus impatiens</td>
<td></td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
</tr>
</tbody>
</table>

Others include species with less than 1,000 BLAST top hits.
Figure 2. Distribution of Blast2GO annotated level 2 GO terms for biological process (BP), cellular component (CC), and molecular function (MF) categories for the transcriptome of *Callosobruchus maculatus*.
Figure 3. The number of peptidase and non-peptidase homologs in common with *Callosobruchus maculatus* for the species shown.
Figure 4. Distribution of catalytic type by percent of total peptidase and non peptidase homologs annotated in the MEROPS database for the species shown compared with the transcriptome results from *Callosobruchus maculatus*. 
Figure 5. Distribution of catalytic type by total number of peptidase and non peptidase homologs annotated in the MEROPS database for the species shown compared with the transcriptome results from *Callosobruchus maculatus*. 
Figure 6. Distribution of the *Callosobruchus maculatus* peptidase and non-peptidase homologs ranked by number of reads per kilobase. The peptidase and non-peptidase homologs are separated by catalytic type and total number of peptidase and non-peptidase homologs are shown graphed in increments of 20.
References


Appendix

CM.A01.009
MRLLLVFLCALVAVNCEFHRIPLYKFKSVRRTFQEVGDVSQVVLNGNKYRNLGGPVP
EPLSNYLDAQYYPISIGTPQFTKVIFDTSNVLWPSKRCHFTNIACLLHNKDYSSSS
TYKKNGTAFAYRGSGLDFSTHDVSFGGLKVENQTFAEAMNEPGMAFVAAKFDI
LGMSGYSRIAADVGVPPVYNMVSQKLVSQPVPSFYMNBADPAPQGGEILDILGSGDKAHYK
GEFTYLPVDLQAAYWQFKMDKVQVPGPETTLCACKGEIAADTGTSLIAQPSEEVKAI
GATPIMGGEYLVSCSIPKPLTINFLGKGPFALGKDYILRVSQAGQTLCLSGFMGIDIP
PPNGPLWILGDVFIGRYTEFDLGNNRVGFAEAV

CM.A22.003
MADLANTIVAQAENVRTARNATGKTPSTPEGMAYAVGSLVMALLPIFFGYSYRSVV
CHKEKKPEKMTKRDAAIFPIMASCALFALYIVFKLFSKEYINLLLTGYYFFFGLYLALTHLL
SPIVSKLVPAAVNPIYHMFKQEGDASAHVLVDYRFSTYDDVALVACLLVGAWYLQ
KHVVANLFLGFAVNAVELLHLNNVTGCILLSGLFFYDFWWWGTDMVTVAKSFE
APIKLVFPQDLLQQGLQAANFAMLGSLGDIVPGIFIALLRLDFHLSKRDGTKYFHASVIAY
FLGLTATIFVMHVFKHAQPALLYLVPACIGTPLGVLAVKGDLAAMFKY

CM.C01.060
FLSIIALTSSEISENENYLSDEFDIFNSKAQSSWATGRNFHENIDISYIJKLMGVLPNY
KNYMPPVKEHDLEALEIPDSDFDAREQWPHCRIPTEREIRDQGSCGSCWAFAVEAMSDRVCC
IHSKGVKHFRIFSADDLVCSCWCTGMGNCNGYPGAAWHYVWVRKGLVSGGAYGSHIGC
RPYEIPCEHHVNGSRPSCESGESGKTPRCDKQCESGYDVPHYDKDKHYGASAHSISDDPK
QIQAEIMGKNGPVEGAFSVYSDFYNYKTGYYVHAKGQLLGHAIRILGAVGQNDTPYW
VANSWNSDWGDNGTFKLIRGBKDHVIESIGVAGIPKD

CM.C01.067
ANTGPVWSQVYSKVGLHIPYAEIDEFPFAYAWFDAPSQSRIDYYGDMVKTYQLGGNGY
GTSIKIAPITTHEHELNVQTCVQVTNGTAEDKILPQTVLPAFLNFQYIGGETVNGIETQKWR
VETVGEKENKYMWMYKDNPDGIDKIAVPVRYEMKGFNSLLGSHYDHYYLDYDTD
STDDIPGEIFQINSMTCSSFGPGDKHDHYTNFPMAEFVKPETTGHLDDEFDKFRKHKY
QYEQQTEHLMLREIFQRNIFHSVNHQNGSKYSLSVNHLDADKTPQELALKRGKQYSGU
YNGSQFPPYGKTINTENLPPQFDWRIYGAFTPVDQSVCGSCWSTGIAEGAYFLCNG
GKLVRLSQALVDCSWGYGNNGCGDGEDFRAVSWIHKHEGVPTEEDYGPYLQGDGY
CHADVKVPVAPIKGVNWSTNDELRAISIFKHGPISSVAIDASHRTGFYSNGVYYEPK
CGNKVEELEHDHALVAVGYSYSLNGKDKYWLKNSWNSYWNNDGVYILMSAKNNCGVMTT
PTYVTM

CM.C01.084
LNKELVDKFKSVFYADEKNLLAQNVCSRTIDIFDVCLSRKTEETQHVYNHKIDAEAKPV
TNQKSSGRCSWLFAALNVIRLPIKHFNLEDFFSQGQLYFYWDLERSNYFLNAIVETAKR
NEPVDGRTVAWLLHDPICTDGGQWDMILNLVRKHGLMPKKHFPESFSCESARMNAILK
SKLREYAKAIRDLVSAGSDADISSLIEQPMNNIYRVVGICLGVPDETFWTYDDKNKA
FHSVGPITKFKEVKHKPFLNVDDKCLVTDPRTNEYGKSYTVDCGLGNMVGRRCIY
NNQPVELLELTAKSIKEGEAVWFCDVSKRSTYKQIQDNLKIHNFPVFGVDIQNPNLTK
ADRILYGESAMTHAMVFTAVIDNINDGTEIKFRVENSWGDRGEGKGYLLMTTDWFEKFT
FEVVIDKKFVPEELVRFLEDPIVLPAPWDMGTTLA

CM.C01.A30
DLDKLWEEFKATFKNYSDYHDNSRQAWEEENLAKIKKHNEEADKGIHSYYIRANPLS
DMTTDYLQKMAKLRSTRTHKVDPVGDVDHMHIEMSSWAVDKFVKIPSYNQKD
CGSCYAFSVATAIQGQIFKQTEKLPSEQQIVDCSVSYNGCAGGLSLRNLRYLEKTG
GLMTYNQYPSKQCKFDHRTINITTWWAVLPAITLELAVKIGPVAAASINAS
PHTFQLYHRGIYDDLSCSSTHNVAMLLIVGTYKEAWILKNWWGKHWENGYMRLRNN
RNRCGIANYAA

CM.C01.UNA_1
VEELRYSIFQDNVRKIKEHNGEYDKGVVSYAMKVTQFADMTEREFMDLKLKTQSKGLK
SLKTGEVSTGIAEADSTRHKEVGATVENQMLCSCWAFSAVGAMEQIFKLKNGLS
DSLSPQNLVDCATSVYNNDCNGGLMDNAFQYVIKNGIETEQKYKYGVEDEKCRQKE
HTHKKFKEYLDIDPDEKEIAKALSGPISAMAMEAVALRPFYSHIGVDSKCSNHPSDLNH
GVLLVGYQTEGDDDYWIVKNSWGTWDEGQFGYFKLRNVRNACGIKWYVSFPQL

CM.C01.UNA_2
EQLSLEQFQNGFKEFNLKVYSSAEELHRKEIFENVKISRMNTEYASLDRGFVAKINPF
ADSTFHEFFNQLNGFNRSSHGARNRQAKGSSFIPPAWVFETDVDRGFAGVPVEYQG
K*RRLVLGFESQLQGHWRFTKRQKQLPDVSVQNLIDCTLPYGNSCNGGLMTPAYEYV
RDNDGVDSEQYRFEETGCRFRREDVVVTCSGNIGYVEIPVGDKEALIEAVGITYGS
VGIDAGKQTQFQIFSGYIYDEPTCNCNPDDTAVLNVGTVGYYGREDGREGYWIKNSYGPQ
WGIGGMYKIAKANGNLGCAKASYPLV

CM.C01.UPA_1
GEESFAMKVTQFADMTHEFLLDLLLQKVGVPALSPAVFDNFDTIEEKAVIDWRKEGAVTPVKNQGKCGSCWAFSAVGIESIAQFFKKNRTLWLSAQELVDCADEHYRNKGCQ
GGTMNADFDFVRDEIQTEESYPYREAMSTCRKNNGGYVTKVKKYVYSSNEQIEARIVSS
EGPVSAADASGLQLYSHGIADTEKTCQDSLFLSNHIGLTVGYYGSEDGYDYYWIVKNSWGA
DWGEDGYFKLKRAGKVCGIGTYNTPL

CM.C01.UPA_2
EEWAKFKSREHGKTYSRTLEDLKRFNIFKQNLQKIQDHNAKYEKGEVSYAMKVTTFSDMTLEEFKDFLSKSGNISKITDGSHLTSAVLDYADDGEDDTSFKPAPWAKVHKAGAVTK
VKNQGLCGSCWAFSAVGALEAQLYQKNGTLEDLSAQDLVDCAADSAGNMGCAGGLMSNAF DYVKDKGLGETEYPYKGISDICRPQKQEFGFQKIQSYVNLAINERALA KAVAVQGVPVSCAMDATYMQSYSGYIINSSGCESERFLNHGVLV VGYQEGKIEYWLWKNSWGTGWGEDYFRLLKNDNRNTCGIATLSSYPI

CM.C01.UPA_3
MKFLVVATVLVAGASSVVEEQQKFKLDHGKTYRSLVEEKRFSFVKNLIIDIQEHNKKYERGEVSFAKKTQFADMTHEFLDLLLKLQGVPALPSNAHFDNFEDTDMEKEKDADVWREEEAVTPKQANCSCWAFSAVGAIEGQFFKKGNSLDLLSVDQLVDCAGGSYGNMGCNGGLMNAYDYVKDRGVLTEKEYPYEGITGTCKQMNANGFKITSYVNITIQEDEKALAHAVATVGPISCAMDASLYSSYSHGIDISSLCGCGERFLLNHGVLVVGYGEEKGEVYWLWKNSWGTGWGEKGYFKLLKNDKNTCGIATLSSYP

CM.C12.005
MAATAGNWCLIESDGVFTELIREFGCKGVQVEEELWSDLQEGFESLKPIHGLIFLFKWTN
DNEPTASLVDQSRLEKIFFARQVIENACATQAILSVLNCRHQDLKLGSTSDLKDFCQG
FDANMKGLTISNSPVIRTVHNFSVARQQLFEDPSMANKSEDVFHVFSYVPIDGRLYELDG
LKGQPIDLGAIPADTETWDIVRIPIEIEKPSIYRESEGIEIFNLMAISDRKLYELEKIQ
MEASGMETDQQAIEAKLRLMISDEEEKRQQYQVENIRKRKHNYPLIVEILKILAEGKLMPLYE

CM.C12.008
MDLFPESNPEVTLKFIHVLGVPDNYKLVDVYGLDNESLWLRPRVIAFILLFPCEKEFY
EHCKQEDDEIAKPKPISSHNVWMKQYVSNACGTIALHISIAERLHLKDGVFKKWLDE
VKEKTPDERGHLLAEGSDSEAFKLISAHQQMLAMEGQTEVNPKVNHFFIAILEKDG
ELFELDGRQHPVYRGTTDDKDVFVHDAANNICKQYMERDSEDINFTVLALTGDS

CM.C13.005
KFANSNHTNWWAVLVDTSRWFNFYHRANVLSIYRSVKRLGIPDSQILMIADDMACNP
RNRPATVFNANQHINVYGDDEVYDYGYETVRENFVRLLTGRLPPGPTSPKQLLSDDGNS
ILYLTGHHGDGBKQDSEEITNQEMSADLEDQWMKQRYHEVFIFIDTCQASM
YERFSPANLAVGSSLVGDSLSHVDPAINVYIIDRYTYYALEFLESIKPNSDRTVAEFL
KVCPSVCIISTVGYRDLYRRNPEDVPVTFDDGSVRPVMSSGSICYKCNPQPHFARECS
QPGRDSGSRSEGRGKRECKKCNKGVGFARECEKEQARCRCYEGHFAKDCLQSP
MPSYCNCRKGVGLARFCPEGNDRMANETCNQRIQGHISRCNPENTKTCYLCQHPGH
KRECCQENDYRK

CM.C14.015
NNGDGLNSEOEGVNGDQVNVDVDGALGRFSACRGGEVHARMPVAKYATHYKMNH
RKRGMOVIFNEFVQCGGLKSRSGTNEDECKNLKECLTSLGFVHGFKDLNYQDIETRIRE
VSKLHSDHDCDCLIAVLSHGEMGIVYAKDMPYKDPYLWTPTADKCPITALGPKIFFLQ
ACQGDKLDGGINLSTTEDGIEJYNTYRIPVQADFLIVYSTVKGYYSWRNTTGSWFIQSL
CEELKKRAHEYDLMTILTFSQEVRADFENSEVPDNHSMHRQKQIPCIMSMLTRLIHF

CM.C14.040
MGNFLSSMSKYVSEGEILSDSDEFDNVDNSADTKTPSTSLNTADASLSKSCVLPDPQAI
ITLDBKVSSIQKHLDEQDMISLYLLFQPLYALQEIAFMENRSENSGAL*VG*FIKWKQRI
YLEKSAARGTMYSELPHFKKLYGDKDHNLINFPHSPYATVYVNKLRLMKLYIICEKLDS
SKTEVLLEYVRKDMHKSMEPNENFNDPIALELLYWESCQIYKQTVNENLMKAFKV
MEEEQLYNILDSYHKLQPLDKQAYITSPEVSGPKKSARSQGLSNEIPAVTSEETHA
SNLFHASLWDDPREEKGRRISTSVNDQQLNVRDQALRVRGLGDPELETNLFEPTVEE
NNRFYINPEPGKILINIQEYFYETMELYRH LLPENSEQLERRGTDNKMRLASTFKKF
GFEQKFENDKLKMDMDVDCIKKTVASITNESSLFVILSHNKGIVYANSRCPVPSEIQD
MKCQHFAKPGVLIQLQSCQFGCFNLNDRKDQQQEEIVESQPLATDGATSLPHHANULL
TFWAVTPGYAAIRHKEKGSWFIQSLCEKEVAYQDKIHFDICTRIRKDAEKKWIKGAA
EYKMCSEIQTTFFLDQDFLPKRM

CM.C14.UPA_1
MLPEYKIERIKNFVPLMKECDVRILTKLLVEKKVFTDNEINFKSTNDARHNKRVLFFEI
QKKEQRADFALVESLERTNQGHLADLQHKRTQAVKFNQAVCEDDDGDISLKLSPHIIDP
TTSPLNIKEVSRNFFDDTEHNGFEYYYRSKRTRGRALIINNYDFGTKMEYRNGAMVD
NNNLKSLFKQMGWQDMHQKTMATRLMEQMKNFAQDQKKNNYHNCVIFMSHGCEN
NNKTIYIGIDSYSCBEKIQEEFFNENHCLKLFITPKVLIHFVCRGNNLDYSTKQVETTD
AVRDMEMLNMRMTQEDMLIGYSTULLYGTKHTRDAHRGSWYIEILCEKFMKYAHDPVDD
LLSLIDQGLRRRMEYTMQTAEVSN

CM.C14.UPA_2
KFHERTHFYTEKLEYPRHGDEPGFIFINQEQKSDHRPEKRLGSTRDVNEIVLCMQRLGYN
IERNSNLTDGGTDDVMKLDKLWESMDNFSRYNSLFIFFLTHGHEHDGLETKDFGFIHVVD
MGPFRAANGTLDKPKIFVFQACQGVASTTFEGQAPESPLVCTKFEDSYMPDTLLLVF
STVEGTASYKRVLGLGSFWQIQEMCRNFVSYGRKEDVISLFIRTTKCVSGNYYFTEDKITFK
KQLPMLISTLKKFYLNKDKFRHFFLH

CM.C19.015
MPQYKVVKVKWGETFDIEVNTDEPPFQAKLFAALTGVQPERQKVMIKGVTLDTLD
DNLKKLDGVTTILMGSKEEDVPKPEKTVFVEDMNENSLASALELPAQLNLGNTCYMM
NATVQCLKVPPEREALQSFQGGVATGNgVPAQSTAAALRDLYSSMDKGNIIPIVLLL
QVLMHMAFPRFAEKNASHGFGFSQQUANEWCWTELVRMLQQKLPAKKEGDNSQEQRSRSLID
QYFGGTFEVEWKCVEAEDEPVTSKEQFLQLSCFISQDYHMSSGLKNMQEQTITKKSP
SLDRDAVYIKTSISRLPAYLTVQFVRFYFYFKEKESINAKILDKVFPLDFDAFDFLCTPELQ
EKLSPMRAKFKFELEDTQLEQQAAKAIKENKGDGKRKDVKKYYAPYWFENDLGSNNSG
YYTLQAVLTHKGRSSSGYGHVGVWVRQSDDNWKCDINTERNCPVTTEEILKLSGGGDWH
CAYVLLYGPRLELPSEEK

CM.C19.A23
MGANISQLERDIGSEQFPCHEHYFGLVNGNTCSNVSVLQALYFCKPFRDRVLEYKAKN
KRTKETLTTCLADLFHSLPTTKKVGSIAKPKFIALRFKEKEEFDNYMQDAHEFNLFLI
NHISEIIAERQQNNSTTNNKSKGAGENGTAHNSPEPTWHEIFQGILTSETRCLNCENVVSS
KDEDFDLQVQDIEQNTSITHCLRCFSNTELCSDNKEFCDNCSMYQEAKQMKRMVKKLP
MILALHLKRFKYMEOYNRIKHVSHRNPPLELRLFNTSDALNPDLRYDLVAVVIHCGSGP
NPNGHYYISVKSHEFWLWFLFDDDDQVFDIADTIEDFYGTLATDTQKSSETGYLFIQY

CM.C26.A22
MLRFLALASSFIFSSHADVPIIGLISQETYILKNYIENGHHSFIVASYVKFLESAGARV
IPIWIGKDDDYTHVLYNTGVLFPGGTYPFNETGGYGEAAARKLYRLAKATNDEKGTYY
PILGVCVGMQVLGFAEKGDDIRTNCELKNVAVPLRFDAYEKESRLFSKAPDVTIILKAL
KNTYNFHRYCLTEETLNNWLPEWKLSTDKDINGIEFISSMENKNKYFPGYVQFGHPEK
NIEFEMKMGVIGPHEVEAVKTSQYPANFFVEECKNGNRFDHDMESRTLNYNTVYTGIG
KGSVYEQTYVFEDTD

CM.C44.975
CGIFAYLNYLTTPKRAEIEELLVNLKRLERYRGYDSAGAVADAAANSKDIAIK
KTKGVALLLEETKSKSNELETDVLHCGIAHTRWATHGVPEVNCHPHRSDTDNGF
VTVHNGIITKEKFETLLEKFCGCEFSEDTVEIAKLSHIFQHYQPHEFPFRELVESVVRQ
LEGAFALCFKSKYPFECVATRGSPLLVIKATYTRALTDDPVILYGMDFRPDGRADYQT
LPRTGSTSEEPLEDKQVEYFFASDASEAIIEHTNVIYFEDDDVAVKNGLSHRLRRG
DDPHKREIHTKMLQQIMKGSYDFYMFMKEIFEQPDVSVNTMRGRVNFFQTAKVTLGGI
KDYIPEIKCRRLMLICGTSHASIAATRQLLEELTELPMVPAEDFDRTPRVDV
CFFISQSGETADTMALRYCKQRQGITNTVGSSICRESHCGVHAINAGPEIGVASTKAYTSQF
ISLVMFALVMSERLSMKQRREEIEGKSLQSIQIREVLKLDQYEVKRLAEDLYKKSSLI
MGRGYNFATCLEALKVKELTYMHSEGIMAGELKHGPLALIDESMPVMMIMKDPVYYA
KCMNQLQQVTARDGKPIIIICQUEGEKDTEMATMAFHGKSLLQVPKTVDCLQGILTVPMPQ
SSLFSFHIAVLKGCNVCPRNLAKSSTVE

CM.C45.A01
MPVNDSTGRRQCIPILYTKGTHYEVGYDMGRTFSDMIHNFKLSTLNCYLPCYDMPE
GRKAYEDTLNCVTKTNFPQYVRELEGIADAGKVFPKFLLLHMNIVAYPCRKECEAKP
QGCStICVNEKQGEILGHTEDALSETLNHYYYVSAHIsDKPEGKWKVTTEEKFTSLCYAG
HLPGYTMGYNHHGLVYSINTISAKTLRPGKNPRHFLTRALLSADNFAKAQEILRDRGCG
VDGACSVNMTFLNQEGDRLFHNAELGPADKANESQLNIFTASPGEYIAHCNQFLRMKL
EADSEMKSSVRLAEFKHAPQKDVLVEMLDTSNCKYPVFRDDEPREPVMTIAVG
IFNCVERTWSLYSDNPKKNELAVLP

CM.C54.A04
MDCMFMEACMEATLEPPDIPKTTEPVWILGKYYSAIQDLNKIRQDIVSKIWLTYRKNNFVP
GGGEGLTSDKGWGCMLRCGQMQVMALANVTLHLGRDWWLPQTRDPTYLKILKKQFE
RRPAPYSIHQIALMGASLGKSVQGQFPNTAVQLKLVYDEWSRLAIHVALDNTIV1
SEIRDLCITYQLQTKLDGTVNLSRTREWKLPLLIVPLRLGLTDINPIYISGLQKFQFQLVGG
KPNVALYFIGCVGREIVYLDPHNTQKTGFVENKETESEQIMDLTYHCKYASRINMLMD
PSVAVCSFVVLKLGDFNDCLRKLIDLIDPEKEPMFEVTYSKPEEWSPTLENIVEAAATTSA
adyqgrhydqdqefel

CM.C56.UNW
MAKKALVFLAPGAEEMEFTTADILVRAGIQQYTVAGLPGSVECRRGNIKPTVPIGNA
KSGSSPVDVLVPLPGGLSGSKALASSKKEVMGELKEQEFAQRLIAIAACTALKAHAIGAAIG
KVTYSPPSMQMEGGQYTLEDKVVVDGVNVTSRPGTAFDFGLALVEQLVGEKE
KVEVAKGLLLS

CM.C65.001
MGDKGESTGGCTMNNGNQNQDELILAQRQIEIKEISESVAMLGAEIAITTLDKEYSTD
KIYLEEKVDASKYKYYRRTDGPNCFFRAFSAASYANLERLLDTPNELNEFNYIAETSAIL
VLGDFPQFTVEFDYTDYMEVLKLRSDIKEDAKKELHSIFNEQGSYDVMLVYLLRTLS
GQLQKIDQDFYDSCFIEGDRTITDFCHQVEPEMVYESHDHIIMEACAAINAGVRRYMNR
GTGKNCTEHDIEGRIPIVTWLLYRPGHYDITY

CM.C85.007
SIALRVKTGQQYVNTLSTQSTIKELKQVLSSIASIPLERLHVLSSGFPKTLDISQDNL
SLGDGIASTGDLILEEKAQPDVTPPLSTAPSPKTPVAIEEDTNSLENYCSGILMRQV
VPADNSCLFTSFNLNGKVDSDATVAPYMRKLVAETINAERFLYDEAIVLGPVPDDYC
AWIQQQRSWAGIAELALSNYYYGIEAIADTMNAINRFGEKDRKYPHORVFLMDGIGYDP
LYLEEPADGSKIRTIFEPSDTEGTILKQAEQLAQEANTSRQFTDMEKFTLMCVCQQQLKGQ
AEARQHAASTGHNIFGEV

CM.C86.UPW_1
METIFHEKQEGSLCAQHCLNSLLQGYPYFTAVDLSLAQKLDDEERQRAEMCGEESEDYQ
KFLQQPQSNMODDDSYFQVQVSSALQVWGLFVPYGESERIKAADDPTKMNIFCNYK
DHWLTVRKIANQWFNLNSLPKPEISNTLYLVLALQLRNDGYSIFVFVELPECTADEL
LKNNPITTVPTPAYRSTISNESDPDlQAALQLSLKDDIPSTSSFKAPEDDDLQKALRLS MENFNSPEEDQDDEESLRLRKAINMSL

CM.C86.UPW_2
MGSNEIYHEKQIKELCALHALNNLFFQARDAFTKEELDYICHSLSPDNWINPHKSVLGLG NYDINVIMKVLQSRYGAEIWFDKRKRDPSCSLNLGNICGFVLINPSDYKISFITLPLRKKHW AIREINGFFYNLDSKLPAPQILIGQESELLILYREELDDKEKQLFLVTEEVGKDHSWILD QNIYDALNNNINDADIVELKDINRD

CM.C97.002
MFSNGLACNLPSFCMSINRDSGDELLPSKMSREPVLLNVYDMYKINEFTTNIGLVF HSGVEIYGETAYGHHQYPFTGIFEINPRDETDLGEQFFKQRTQVQITDFTEDDV KRIIYELGKEFRGDRYHLMNINCNHFGAFTKILCGQDIAPWNRALAYFSSWVPFLERC LPKEWLMPMALQHSLSRHDSTCEAST

CM.M01.013_1
DDTTEILNIVYPDTVQVDVLEVITYQGKLSTEDMDGFYRSSYTENGKVKNALATQFEP VYARRAFPCFDEPTFKAPFKVSITHPKQYTAIISNLASQKVSTSTVTDMMKQETFETTPKMS SYLVAFVVFSEVKAKSDSSASQDYKWNVEARPSEKSNMTALKYTPQLVDMKGKWT QGKYTALGNEQFQVAPIFDAAGAMENWGLITYREVDLLDEKQTSANSKQUITIQTIAH ELSHQWFGDSVTLDWWSDLWNEGAFATYFQAHLANLVENIMELDKQFVTWVHKA FYQDALVTTTPISNKANEINTPTEINKFDISYKAGSVIRMIHEFGLKDFQFTALKSYL AQNAKKTSPXDFKHFDEAQQKYYKNEEIMHNWFYEAQIFPLVNVWNETACVISQER FNSNASTKWAYPITYATSEDDDFSSTTTKTLLEPGKTVIELSANETAWIIALTQVEF QVGEECHKYVTEQFEAYKKNTSIDDYDKRVTFCYALKK

CM.M01.013_2
GSANFKRGIQQLSTNKNKNTKPNALWTASLVAVDNTVSMPLATLPVMQNWIEKAG FPLTIVTNKNTLALKQERFLFGSDTDDTKWYVPVTYTSTVDATAKEKTSFHLWIEPDK EATIQVPDGASWIILNNQQTFGYRNYDVYDIALWAEIKAESDFGGAELESQIVDDLF NLAKPPKVSYKALNINFISNDISYYTWAYAARRGFSSFLDIEKGFESDLGKAIADVKL MKDYKVSVPYTDLKPTDDLSTLQTEVIGLACRLQHPECSEMKS

CM.M01.A02
MVEKVITYSSIPASEALEENNSNOKKTYVMNRRGREPSKSVISRPPLICMLGIGALLLAILVGLV VFFLVPRLSACEETPSALTIAVTDVLGLKRSEDVDERLPRSIRPHYIRIRVYPLNT SSTSGWVSMVLKAEDETDSSIFHARNISIDKHSVTRKNNSSQIKINAPQYAEQDKYKIRL DETLSKGDCEVSLDYTGILDRHLKGFYKGYISARNNESYFASQFSDARKAFPCFD
EPYFKAKFTISLPNMTTLSNMLPGKVEAYEHDKSWYWDYETPDMPTYLVAFMVSDLQGSTQSTDHRIKMWANDLIHTAYAGDLAPKILRYFEFEGTNNFPLPKIDIVAVPEFGFSAMENWGLITFRENVLLIDPATSTNSDKREEAALVGHEIAHQPWFQGNLVPKWNWDLKKEFATYLEYFGVDSANPDWKIAEELFTLTDDTHRAMGVDALESMARASIFEMVNSKQIRQAFDDISAYAKGCIRMMNNFLGEAYFKTGLNLFHLQYQVAEDREDLFTSLLTEAKHNGLPPNVTVGIGEWTERTPGFVVHSHIADYENKLQSRQNRFLLSGQKQEKSWWIPISFTAKKTLDENDFSETKPYWMRGEHIEDADLRNITWTLLNVNHGTYFIVNYDEKNWRALSNDNIMLPLPLTRAQLISDSMDALSALISYDIPLRMIARMATQDKMIMIIPLATLNLKDFLNNMLYNTPAFGLFEFHSKIFKQTYTNYTQFEGLVDVYTNWRIRVLMVESCRSSISCAHEARHRFRERMIHTVHEPVSTVYCTAREGEGDVKEYRRLDTPSISEKNIIDALGCTKEKWLRLSRYLDSLTAGSSIRYQDAKVFSSVAANPDGMPIAFDFLRKRNWNA LLNRYGDGSIVAKMRSLATEHMENTFEQLSEFRKDSVKQMNSTTASFDAAIEQVRANVWNMRNYYEVEDWLVQQHRDYF

CM.M01.A24
ITKPTDEIVLNALDLTFGGEVEKLKDVLTPSKEVSQMSSECETATIKFASSVAPGPYEILND
FGEQEINDMKGLYRSKYTNEKGEAFAAVTQFEOANDRCCFCPDGEPAKATFDIRLTV
VPNHLVALSNMVPKSHIPFADLVTYDFEPEKTSMSTLYVAAVIGEYVYDEKSSDGVLR
VYTPKGKNEQGQLFALEVATKLVPYYKEYFIDAYLPKIDLIAADSFAGAMENWGLVTY
RESCLVDPQNTSASSKQWAIVGHEIAHQPWFQGNLVTMEWWTLLNVENYASFVEFLCVNFLFPEYDIWTQFVNDSYIKALELDSLKNSHPIEVPGNPSEIDEIFDDISYNGASVI
RMLHYIGDADFRFGHMLYLTHQYSQNTFEDLWAAEASESNPKPVKDMSTWTKQM
GFPVVKTSQPDKGDKGVSLTLTQSDFADGSKPDEQLWMVPIGSTSKDPSKEVVNTV
LKTREASVVPDIDPDEWKVPNGTIFGYRTQYTMELKFTAIRNKTLPFIRDLGLVV
LFAMVKAHTUNTVEVLKLLKAYEDTDYNVWSSAGNLGRLTLQMDYTDCKDFKQY
QKQLKKVYIRLGNQPGHETLDTLRGLVLGRLAWLDDENTVEAREKRFKHKVSE
ELPQADLRSAICYKTVLRAGGKEEFETLLKLYRATDLHEEKDRIGSGLAACKDPLLGGVK
LEFAMS

CM.M01.A27
GLSPLDPSYRPELAIVTNIFNLNVNFHTKILAGYVTLDVNTDLDDASKELDDTLKLN
ISSYVAENNEKEFLKLDDEEVEPVEGSKLTQVLPHKESNYKIVGKYETDPTASGLQWLSPRATAGKKHPFMTHFEARTHARSVPLCQDTPAVENTKYSATHISAPPEYTVLMSAICNGTKTLHGGKSHFVQNVPPVSYLAVGLESRTLGPRTVWAEEKIEECAYEFARTEKQLEARAEEICGYPWEIGYDLVLPPSFPFGMENPCLTFVTPTLLAGDRSALANVVAHEIAHWTTGNYTNRFHEFWLNEGFTVFEVRKIGKLESQPSQDFDAYLHVSELRNERIEGRSDNP
LTQLVNVNKLGKHPPDASVREVPIEGKQFTRLYLENVGGPDVFEPFLKRKYETKYKSID
TNDFKKNFESHSFSSNISISSDWEWLSLYKPGMPVIPDYDKSLAVFCNLDLLEKFLKWNLEGEPITSADRDQMSSAQKIPvFLDALSEAEQPSLEKALNLF
CM.M01.UNW
IIVGVVQAATAAAKHKRSIDLIAAHLIQLPSGQERPLHDTAPSGLHLFRPDSESGFGFENVI
INVTVQKHYTEIVLHAKDLIQELKNTQVAPDDDELKKNSTEADVFTPQVLITIET
AEKVPKKPYQIAFGSPIPKPGAQLQVYIQFGPLNFLDNEGFRSYPDPVTKKAKWLA
THMRPNMARTVFPCYDEPAKYVMVQVARPKHMAISNMPEVKATADDPKLDWLRD
FQKTPPIATYSIGVVISELSSLBYNLTGTAEVPDVRVSVIPRAVDINAVIDSTKIDSKLI
LSELWGTPYLPKIDIFALPNYQATKPSDAWGLILFKESLSGRGSHIIQELVYYQWIA
LATPFWDSADHINNHVYRTAVYTTLQMHEINEISNWPMTMLYSIYYEFSKRYPHGKN
TAIKQDSATSAKTELIFRLMLNYTLGEETFGRKLQRFMADRQYKTFFGDIWLCITEQARFD
EKLVPNPIVTNEIAASWITKDRPLPVVTTRNYQSNATLTQQVYLRLERPMDVDQEKYLW
WIPIVLVRQDKLDFHNSTHPISEREGIQDGLGDSFIIVNPEEGIPFPVNYDANKN
WMLSNFLNSPDRTQIPVTTRAKLHLHDANWFALGNLSFIALDMLTLFKDEREYLAWDP
FTLDHIHGKIDSSPKVHTKFQNYVRLLSPYEELEGESEPKDENGDSHLRLRSSKFLCQA
GYQPCVDEAQTTAYRKWMEAKNPDEGNPVPNQYIPCVFQOKMEWQGFLQRVLPF
FRQKQERTYLLKTLAGCNPANEKIIKLLNVAVLEKNVFSNDIYLLFSMSINTANGYN
TTFEFKNWNNIISRFLKPYLWNMSVTAATNSFKEEGLDRVRELVEKQNEFGTA
DFVIEKALVSLNEEVRSKENLPVIPKELWNYIA

CM.M03.006_1
VSIWSPLEAEAFGRNPMQLQRKLKRNTGFISELSYDSFYMLQDKCIFKTEKLI
SEALSNSRKRKMVEIFDDMSDCLCQVADLAEFIRLAPESNYRNAESAICITVSGIVEKLT
NIKLYDALKHAEEQGDLPTEDTKLVGEFLFLDEQCGIHLPEDERNRVTULTNMLNLGQ
QFVAGTSTPRLVKKSQLPSTLRNVSSDGDITLITGLYTDLDLAVAREMAYRIFLHPDAT
QDKLNLQLQRDELAKVGFTNYSERALRGSIFDRPGKLVSLEKLGSIIWAKKDF
ETMSTMKLEENAINGPLPWDVYFTQAKKAKDDLKVSSKEYSPFFSLGAEMGLNMIF
QSLYRIQLVNTETKGCWSPDPYKLAVEHEKGLLYICYDFYERSGKPSQDCCHTIR
GGRALQDGSYQIPIVVLMLNLSPPWSSPLLTTPSMVDDLFHEHMGAMHSMLARTQYQ
HSVGRCTSTDFAEVPISFACDHRVLKKFAHRFTQEPMDPMVQRLCASKHFT
ASETQLQVFYAAIDQAYHGTHPPLGTTETEVTALAEQSFYGIPYVANTAWQHRFSHLVG
YGAKYYSYLVSRALAYSINWKYFETDPFNAEGSNKYERTQCLAHGKSN

CM.M03.006_2
MFMRDRDRTICLLLINVIIVGSGQNFLKYSTVQDRPKRPELTVWTCTISDDEQRKCN
SFSMANERDQIKVGYETVNTCKQARKNEECMIQLDEEAAATMTLDAVASVGFGRHSL
LPIAQEVLEGHHNYYFSVAVIKKGTDLDDVVSLQRLGRKAFCFGPETAGFWVPINTLM
KEGGMIEIVDCNHKSTHYFGPSCSVNCLSDDKYPJGDSNKLCQLCIGKIPGRCRTDT
DPYAGYEGAKCCLAGEJAEFLKHSNSPEIILDKLFGTGSIDSQFLCLKKGDSRSSRPI
SDYLCMNGKVPSDAIWSSATSEFILRLKQLEKFKSKKYPKASHNNYGRNGGSTEDPD
ARPHVLPNQVLYVNDIPYDKDVQYNPDNRDPYNRPYNGAQPDSNEYGPNNRTNGTF
YESFSLFESSPRYGDRSNNLFFQDATRHIAPLPLENLQFTSTFLGTDGRTMDG

TTLVKLQSSARPIYRSQVAKQLSTESSDIQLFKEGQEYHGFLVKSVEDISEFRITAVL
MQHEKTKAQYLHLYRNSNNTFSINFRTTPMNSSGQPHELHTVLGCSELYPVKDPFFK
MLRSRLATFMNALTGDLPMTYPFSTLYNDYRLRIYLDVFRPLLREPDRQEGWRL
ENTDPQDIKSPLIKGVVYNEKMGVFSENENIIQGIQKSCIPLPDHTYGVISGDPPKEIPNL
YGDVLVKFHEKHYPSNARFYSYGNFPLLPSLEYINNEYLSKYAYSPMTHTQVPNQKRW
TEPRKEHIACRYDAMGDAIAKQNTITISLLLSNNTDGYETFMLMQFLTEIIKGPNSFYKS
MIEPNFSGGFTSPSTGDFTPQKDPQDCIFTGLQQLKKEQFQVEKIFDETIASVVNGFEPQHIE
SVLHRYELSVKHEIHOSTNFLHLFGLTPWPNHTENEIVEQVRNPDTLAKLQKEKMRWSK
TTLVKLQSSARPIYRSQVAKQLSTESSDIQLFKEGQEYHGFLVKSVEDISEFRITAVL
MQHEKTKAQYLHLYRNSNNTFSINFRTTPMNSSGQPHELHTVLGCSELYPVKDPFFK
MLRSRLATFMNALTGDLPMTYPFSTLYNDYRLRIYLDVFRPLLREPDRQEGWRL
ENTDPQDIKSPLIKGVVYNEKMGVFSENENIIQGIQKSCIPLPDHTYGVISGDPPKEIPNL
YGDVLVKFHEKHYPSNARFYSYGNFPLLPSLEYINNEYLSKYAYSPMTHTQVPNQKRW
TEPRKEHIACRYDAMGDAIAKQNTITISLLLSNNTDGYETFMLMQFLTEIIKGPNSFYKS
MIEPNFSGGFTSPSTGDFTPQKDPQDCIFTGLQQLKKEQFQVEKIFDETIASVVNGFEPQHIE
SVLHRYELSVKHEIHOSTNFLHLFGLTPWPNHTENIV

CM.M16.UNB_1
WNNLTPRPKNKTDMMPMSEPIKGRLPRPIYANFKEEDQVTEVTTLNSGLRVASENRFGE
FCTGVIVIIGSRYEVAVPYGISHFLKELAFNSTKPYPNDDLAILENHGGICDQSARD
TFIYAAASYSGLGDVIRLLEAEATIRPQILKEEVDGAKMAVKFELETLGMRPEQETLMM
DMIHAAYRDNTLGLKPLCPPENIHKIREILFTYLSHHTPKRMVIAVGVEHSLVDA
VQNYFVDDLPTWEADKGLTISNKQLDVSIDHIAQYTGGMVQEECDIPQFASAGLPVLT
VMIGLECSHQDEFMACVLNMMSGGGSGASAGPKGMYTRFLYTNVLNYHWMY
SATAYNHAYSDGFLFCIHASAPPETHLDRDMEVIVKEMVMNMTGNISSQELRAKRATQLQS
MLMNLESRPVMFEDIGRQVLATGHRKRPKYFIDEIKEISKTDIHNVAKRLRSLAPAVAARA
GDLRKMPSELEYIQAGLVDESEGKMPSGRKLSLFR

CM.M16.UNB_2
AARSYSQALPVCGNPSPDLLSTVLQKNKLVATLNDISSARIISIVFRAGSRNETADSLGATHVLRTAAGLSTKNASQFAQIIRNIQEVGASLTATSDRETIAAYLTEGTSAIEQALPFTEV
ATQQVFKPWELADLSRKLDLITRTLPRTVIDLHNAARSGGLNNSLFIPKNIKGKSETSLHQHYAVSTFLSGLRCAVVGSGVNHSHTQYANALKLNNGSVCATPSYKGGELRNSNGKDFAFVAIAGEGAPNASNIKELACAVLQRALGTGPHIKWSTHDNLGIAKSIQGTPPEGFAS
TAINVNYSDTGLVGVLVAAPAKCACNIQVAAYNSVMKCGSITADVNRGKNQLKTLQLVDAECCGATAIHDGIQAVLTGNPQTAAQLAAGVDSITAADVQPSALKKAGQKLSVAAMGNLVDVPMDELK

CM.M17.011
YDGIVLVTAPDLQLVSQKILHDAITEAQYQYDASLTLTTELAVLPLPSKLKLVHSPPTGIDPD
YGDVRAFKKAASGIKRALKSGMKTPLLVLEEHPRFQRAELVTLLGALQAIYVPIQLRE
SGSPAQVKVHLGVTNDIEKTKLHIAIDISGGRVACDIDADPERMAPQKVEYYVRRVHDGIIKLDVSDLQQKKEPLFEPVNRASAVVERHRGRVFIFLEYDPKDVTETLYLVKGVGVTYDTGAGDGNGMSRDKitMGSAAAAVAGFMEMVRLLQPKHIRVVAAMSVMRNSIGSNSYVADEVITARGSGKRVIRNTAEGRMVMADVLCFKFEADAVNAVPNHLFTIATLTGHACLTGEGYSIVMDNGPARKSNYSQAQLKKSSELIGDPFEVSTVRQEDFEAYKHMMEGDDVTSQPSEKHGQAFLILASGLDEHGSRSPLKYSHLDIAASSAGSLPLPATGSPVNALAEQYLL

CM.M17.UPW
KKGLILGLYEGENDEPVFTKAADFKNDHTSGKLYDVKRAPDIKRGGARIFNNIDKEFWSIAVVGGLKEKMGGYNELEANIDEAGMEAVREAAGMGAQVLKREGVTRILVEGLGHPEQAEEGSALALWRYQENRSQHWRIIPTLELFDODCAEFGQRGLFKAESQNLARLSDTPAQMTPHLQETVNELCPCGIKVMVHDKDWEIQKKLQAFLMVARGSCEPPVIEISYCGAPQDQKPVLMVGTKTADFAGGLKLKREGMDRADMGSAAIADPAASALSLPIEAIJLCENMSMGDKCDVGMNGKSIMDTDNEGRLIALDALVYQQAQYKPRLVIDVATLPGVRAALGSAMGTVSNTSQTMSQTRKAGVITGRVWRLPLWWKFTKKVTYFRSHDVMNGKQGGLSSCIAAAALNFKIEIFCVDWHDFTTGVMKSEHKVYPYLTGKRMGRPTRTLIQLLYQLSCPAEGARKL

CM.M20.005
ALPEPLPKVFKKIDENKAKYISALKDAVAIQSVSAWPHTRPEITKMAQWIADKCNQLGIKTELRPLGVQQKLPGSELPLPPAVLGLSLITDPQKKTILYIYGLHDLVQPAKLEDGDTEPFVLTENKGLYGRSGDDDKGPVICWLNAEAYQSLGLEIPVNLKVFEGMEEGSEGSLSELEEAENKFLSGVVDYVCSNDNYLGTKEPCITYGLRIGICYFYIEVECAAADHLHSGMGGETVYEAMTDLVHNQLVDRNGKILVPNYKEVAPLTDTEKALYKIEFDVIEDYRKINGSSKLHNTKEDILMHRWRYPSLSIHGIEGAFSDPGQKTVPKVIGKFSRIRVPNQEPEQIEKYVVDYVQQKFQELGSPNTMKVFMAHGGPTSEDPTHYPYQAGARATKHAYGVEDPDSERGASIPVTLDDLQATAGKNVNLPPVGADGDGAHSQEKEIDVRNYIEGVR

CM.M24.009
AKPTTDLKQLRGLMKDPQYVSEPINAYIVPSEDAHFSEYIAQCEYRAFISGFNGSWGTLITENEALLWDGQRIYQQASMQIDNNWTLMKEGSPTDLKLGWALAKHLAPSGKVLVDPRLYTYQTQYKLHHAENVLLEVWIKDRPCRANPAPNLPQETYGTG1RDKFSRVSEQIKEKNNANHLVTALDEIAYFLNLRGSDILNPVFAYLIISRDNFTLFVNKQKVSSDVRHHLGLEGQNYVIKDYGEVETHNLAVSQKEDRWFVSFSENSLALINVIPKKNRVMTDITPICMCMKAICKNPVEALGMRNAHIKDAVALLCYFWLENSVGKIKITLSAAKLQEFKRMQKDYVGDGSIASSVGAHGAIHYHPDSSSDAEIVDQMYLCLICKSAQYRDGTDDVTRTLHFPGTPKFKECYTKVLKUGQKLGRTIFPRKILSNYLDFAFEFLQWQSLD/YAHGTHGIGSYLNVHEGPIRSWLPEDPGLEAGMFISNEPGYQDGEFIRIEDIVEV
CM.S01.316_1
LLLKVFIIFSVIRHGLINVNGQLIDAGALKKYGPKGFNPWNASNKPPHRHTPPSPCRCRC
GEKNDDTRIVGGKPSENEFPWMARLSYFNRFYCQGMLINDRYVTAAHCVKGFMWF
MIKVTFGEHNRCDKKNDQSPDSSFGRS*GPSVSYFDHIDAILLRLNDRVPITQFIPICLPS
DKNHQYVGAGTKASGWGTLKEDGKPSCLQKVGDVPVMSNKDCCRNTNYSSKMSDNL
CAGYPDGTKKDSCQGDSGGPVLVTRKPNSSRYEVIGIVSWGGNCARPYGYPGTYTRTVQY
LDWIKQNAADGFCEN

CM.S01.316_2
VTEKVVEQQISIKENSSNVQVDPQQPQKGNEWFYGIVGAKPVETNQPGEQQSQVIS
GCPPCKCCEVLKVRIVGGETLNEFQMTALMYNNRIFCGASLNNYRLTAAHCV
NGFSKERRISAVFLDHDRSTATETQITRKKIKNHRISHSRYGSYNNDIAVQLQEDDVPL
TGKLQPVCPLPPTGKSFGSTGVAVGWATKQNGQNTASKLQEVEKVPIMSNVDCKKTGY
GSRTDNMMCAGYPEGKKDSCQGDSGGPVLVQNGTAYQIVGVSWEGCAQPNNPGV
YTR

CM.S01.413
HPKEEKDKLDTPSYSHIRRRRLYLGLGKKFKLNKRKYQECVPAGDAKGGHCKHIIYPAGLY
KDVASLEYLCIERRFVGVCPPDITSSNRVSQLIMDLPPAGY DYEDENNFTGCGLSA
NSMFSKSKKSAKMQIWPWLVALYKIKPFQEQGKENQFCGALLTDTHVLSAAHCFTGVT
ASDIRARLGEYSEFQHSESRRARYDVAEIIVHENYVSATYENIAVRLVSPTSFNAYIW
PICLPIEGTYENEMPIVVGWQLYYGGPTSDVPMQAVPVPWPPQKCIDAYIQKVTNDT
LCAAGYEGNKKDSCLGDSGGPMLMYQLDNGRWITHITGIVSVWIGCGKQKNTPGIYTQYI
PWIIRHTMTK

CM.S01.422
RIVDGTDAAEGEFPYAVSLRHNHSHKICGGTILEKSEYILTAHCVCSNNGPRDPQNYTIQY
DMLKITEDEVNVTQVSEINCHDFNVEKLTYDSAVLELESPLPEGKWKPVKIAKDFTPSEG
QKGTIIGWGRIFHGMPISKLQKIEVEVYDDETCGKKFNKEHHLICLGAAPPLGGACNGDSG
TALIVDGQVQGIAFSITNIC

CM.S01.461
FSDYVDVFDPKEQRYVSHTPPRAVTRWPIPIPTTDPADYIFLEYDDGEPILKPESII
YAEYKKKKVCVKGDFYCSSKMCVTAMVCDCGHRDCEKRDEKGCCEEFIGRTASKNA
QLEVLEKQRWINVSYYTTCAMLCLCIQNTNFCSRFSNRYKDYRATCFSLDNLNGLTGALKVYN
PSDYYELSAGTIDSCDSKSYECNSSKCCLEIREQLCDGTDDCDGRQDEKNCLKPDFGYSIK
LAGAKRKNQGRIEVTAFGKTGYVCDDQFGLKEANIVCKELEGFEMAGDFQSYFAKDK
VKENNTLYMDDLCFNENTELLECNFAWGIHDCDQLEIVGVICKTPK同意CGKFWK
CATNCEVPLFMCDDLNCDSDDSDESQYCEAPEIRLVNGSSSHGRLLEVSHYGJWG
TVCDDDFNEADAHKVCRLGYKASTVKKDSYFGAGSGPIWLDQVSCGSEELDLKDCT
CM.S01.492
MSLAGLLNLMPMSIAIDQNCDFDQNLMLMGQYEYIHNVKEYPQGYYPGISRWTARSPQ
GTKIILTCDVDLPSSPNCQDDKLISLSGHPNFYDAKYNCGTGLSVDNSNLKTVGFLS
SRIGGGKFVCTLAIAIWDDGGTTTAPNVCDCGWKQGVRIVGGEETGVNEFSPMAGILD
ATTADVFCGCTISNYVLTAAHICWRNVKVNQGLVLGVDNHNISGLTSVASYRVDAY
EVHPMFVSPQQYGYDIGILRTEITIQFNLYVGPACLPRFNTQDFDGQETVMTONGLGWTLES
GPRPDVLQKVNVKTVNPVECSYHTERFIVEPIQICTYSPGKDACQRDSGGPLLWMDTSA
ARLHLVGIISYNGNCASQPKSINTRVTSYLWIVSRTGDATYCIK

CM.S01.505
FKCANGESINDQSONCDGVVDCSDGSDETDSCKNNVICPGYVFTCAYGACIDKKSVCVDG
NQDCRDNSDEANCSKLIDNRFKCKSGEIVDEDLHNCNGHEDCSDGSDETDACKRDICPG
YVFTCAYGACIDFKRCDGNKCDKDGSDENDCDIHTTRKPVPTSTEANHPSNGKNCR
LPNHPEGAYISLALQKELPPTLTVDTSSLFRVKCDGSYQPTPHFLIDCVGWSWPALPPK
CEKTCPXISDTHLECKNRNGTRIPCSQATSTITALEFAYHPADKQYDKRVCIDG
KWNWPLPQCIPDCGERELPGSPLIVHGNDTMKGSPWSTIIYMKDGDITMFKICGGMIS
VRFILTAACVTTDDGNIIFDNSLFEAVGKYYSNYSSTEPSAQHLKIQIIHKEYRGDK
RGYQADIALMKTVMQVMTLFFVQPVTINLCEVTSHVDGLVTGWAGAAEPRDDQRTDSLK
VATYPVKBDECLNELPNKNTFDYLYTLADKVCAGYQQKIDSVCKGDSGSLVFNPVNDNS
DRYYIQGGIVSLGHRLKFCGCTNQAALFTNVKYYYE

CM.S01.511
VPEGTPCATPDQQIGPCVSIEYSESPLLQLLKKRGADSEARSLYQRSVCGYNYRDTPRVCPC
EATSKNFEDGGGGSNPQRKDKTIPTPTSTSKKETLQELPSTCPGFSNVTTARVVGVPAS
ELGQPNSWMLGYNRSVPKSPLCGLGMITIDHVLTAHCYKTVDLYLARGED
LYEQNDCGAYPEDIPVQAKMHCENYSKVSTNDAILTLARRPRNDMVWPCILPHAELRS
QSFQKYDCFVAGVAVYRNPSSISLRTAEIPVPLTENCKNAFKQPNIVIDERILCAGW
STGKIDACQGDGGPLMYGKTEDNNVRYYQIGVVSFGYRCARERDFPGAYTRVTHFDWI
QVRNV

CM.S01.53A
YCDDLSCPQLLLNLGNLRQSLCFLSEPGVCCPKSPNTSSSSSSATPIVTTTTTH
KPVTFAAIYINTAEIPTSASRPGIEQLPITANLQLTPINVNYVDPQECCQPESAKFRVV
YNDIAVLRIDKPVDFAKNPHISPACLPHPRDDYTGRCTWTTGWGKDAFGDFGKYQNILK
EVDVPVLNNDGVCQRQM QTTRLGYDFKLHPFGICAGGEKGDKACKGDGGGPMVCERG
GTWQVVGIV

CM.S01.UPA_1
FNHLFVSFVFYNVCVKCQGWLVDDDCRTDRGPVGVSCIAIKSCVSIMEFLENEPKLSHR
VQALNRYTCTGYTSDEVKVCCSPSPDIRIPGSGENGNSASNNPNGNIGNHRNIRLLPTNC
GLMNSDNIKNGENALGNEFPMWALYSRTSRPDRFRCGGTIINDRWTAAHICIKGNQ
RQLLGVRVGEHDISTRDTDCETQADGSRKCSPPPQDVIESAYPHPSFDSNSLSDDIGLLK
VNRIINIVNIMPVLPTGTPATAKFNHVDTVGWFGETGRKSVQILQKVSLPVSLLDEC
QRVYNLNQPQNLKITYKQLCAGGRNDFKDCSGDSSGGLPVSTFLNDDTRYIQGIVSF
HRFCGEEGFPGVYTIAYMDWILD

CM.S01.UPA_2
LFICIVLSAVSVCEGASEYVGAACSNSSRAGTCRKRVTDCPPAVETLKRKHSHDFKRCWD
DDKTEVICCPNLTSRPDDEDGSDSIYKKPHPKPNRQKTSSKRKCEQACEEIKKGASC
VHFNIVRGENALQDEPHMAALGFNENGRMNWDCCASVSNKFLSAAHCFVIKGVT
KVRGLVLKITSDDLISKYSKHDEYDRTTKHNDIVALVEVEQIEIFSDAVRPACLYWN
NEDPEDEMWIIGWGKTIEQDDGDRTIVLQQARVSASVKECNSTYRASMHTRTIRAS
QICAIIGNSSDACGGDSGSGGLPVVENGTCSFSIVGIVSYMGCCGVGVGPGVYTRVSYIDWI
EIQNWPP

CM.S01.UPA_3
RIIGGHEAVPHAYPYAGIVIRGDACFGGSSLRNSFVLTAAHCVDGLHNDLNILGAHR
IFEDESTEVRQRADWIVIHENYNRHSYENDIALIKLSVPVELVDIVKTIPLPESEGGVDI
YSEQDVISTGWGLTRDVPQPSINDCPVLQVLQTVILQISSYSSCKAFAKDEEHDGLVYVT
SKNMCTSGIGIKGTCDGDSGSGGLQFNGTQIGIVSIERTCEGASIFTRVDKLWDIQRN
SDI

CM.S01.UPA_4
WKYFDTKHCIGTIKYDEQKTRIVGVEGAKMGEFPWMVRLGHRYKTNDALIFFNCACTLI
NTRYVLTIAAHCIGNTNTMARIEGHDVNVTDCEGSICAPPVQDRKIEKYFGFYGKEIRK
RDMLVLLEEVPKVKNEDVPPACLPRGVLEADYLGKTVQIAGGWILNSTLGSIPDLMS
LKPILKNVCDKJYHVHLDESQFCAGKYKTGRDCAGDSGPGVTKRLKLGGRGQHYVL
GISYGGK*CGDGPAYTVKAYFIELDKIE

CM.S01.UPA_5
LCLCLVPTVSVAAGIPHLNGRIIGGTNSILQFPYQAISRRGGSHICGTTIFHYSHVLS
AAHCTSMGTVMYSVRVGTDTVNYGTVSVSVCSIKRHEKFRLITMEGDAIFTLCLIPLTF
NERLVPAVLQPWDVTTPGTEALVSGWGVQPEGGSTSRLOAIRIPIISVEVCDRLYGD
GISRNMLCAGFVGRGEKDACQGDSSPGPLLADGKLFGIVSWGYGCDADPNFPGVYTVNVA
KYRAWIRQTTN

CM.S08.090
MKDJPVNTEFPISGLPKKETGVTSFLNKYPNYDGRDTVAIILSDGIDPGAPGLQQTIDG
KIYIERFDCSGCDVNTSITPKEGYETLTGK3LKI3PSWNKPNPEGIYRRGKNAFDL
YPRDLKERMKSDYKKHWDEHRAVSDVRELAFDTKHYPNNSLTAEKLGKLEN
AKQDVLANYEKKYHDGPGVYDCELFHIDGKWVCIDTSEDGLSKCPLVGEYSVTHQY
EPLTKQDKMMFSINVHDDGNLELVGVCSSHGHVASIAAGYFPQSHEPNGAQPVG
SLTIDGRLGSMETGTALIRAMIKITELKKKM3NHINMSYGEHAVWVDAGRI
GDLDVNE
VNVKYGVWVVSAGNPALGGTITPSDINNEPIISVGAYVSPEMMVAEYSMRQKPLGS
PYTWSSRGPTIDGGVGPHICAPRGAITSVPNFTLYRQLMNGTSMAPHAAGYSVLISG
LLQQNLPSYPYNIARRALETATFIDDEIAPAQQSGSLIQVEKAFEHHTNYNEPEVIRFQV
QCGSNSSKGIYMRSKVHNPVQFTKISVEPHFLDENVDPDKRIYNYQKFLVCTNASYVS
CPETHDLNSVMAFMVATIDTSREGLYSTFINAYDVKGVCGPKPVFKIPITVQIPKEVTEP
KFVVSYSKINFKPNTKRHYTVPHMATAWVLKTSDEDSGRFVIHTLQAIQRHCEALE
MNKTVPTSKADAVFYFPVKGDLVLELVIAKWAYNLGEAHLTYLSFHVGPKPNPiI
MHAADGHTVEVKTLQGEDISPVNKLTVQILKEEHIKAPSLTRDIIIPNRQIYELVLV
YFTLAKQCEVSPNLALLSGMYSEYESQFWFLYDNSKQLMGCGDAYPKSYLVKLEK
DYSIRLQVRHDKKEYVEKNEAFLNKL NLTTIDTYLYSQAIGGKKTGVTCSNRP
YASIPFYIAELSDKFQVKSNSNAHYLTGTYSYAKDEFGKADVYPIKYLMGDI
SGNHTSEKSKQEEYNDLRDLJKTQWLSKMEASVASSMYEELTTQYPDHSVHAYLQ

CM.S08.A58
FTFVYILSLFLCAINSNSQIVKENEIPTEDILESPLCCNLTAEHVDELYSSKIVE
YIVMFNGYYNQARIINJNTALNASGISNWKIIERENPASDYPDFDVILTDKT
LGLNASHPAVKRVTQAQRMSRTLKFINPEYVRYGRKSLNSK5QFWQATGRHT RSRRR
QIATAQLADSLWNMGITKGKI2VAFDGTLSKTHHPFRKIKERTWNTNEKDYG
LHGFIVVGIASSKECGLFAPDSELHIFRVTNNQSYTVSTWFLDANYAILKKINL
LSIGGPDFKDFPVFKWELTANRVMVSAINGDGPLYGTLNPPADQMVIGVG
INFEDIACKFSSRGMTTWELPQGYGRVKPDIVITYGSAVKSHIKGGCR
LSGTVASPVVAGAVTLLYSGVLHGRDVNAPSMQALMASARLPGVNMFEQGHGKLN
MAYQILNSYQPSASLSPSYVDSLCECPYMWPYCQPLLGYAQPVIVNVTVLNGLV
GSGRIVSKPEWHYPAPOHGLLDVAVSYSELLWPWSGWMAHVLSVAEAGY
EGLAHGHSVTVESPPPGHEQPRRSTVNLPIRAKIIPTPPRHKRILWDQYHLRYPG
FYPRDKLKI5NDPLDWNGDHIHNTKFDMYQNLNRTGYYVEVGLSGPTCFRRVT
VYFYICVPDEEEYFPEEILKLKRVDAGLS
VIIFADWNYVTMVKEYFDETRQENwuMDGTGGANVPAINDLLASWNIQLGNRVLEG
HKLNDHEMYYASGTIRKFDPKDVGIVSVSANLNDQGAGILGDADTSKHSAI
GLMQTKSDKSRIIVYGDNCIADASHLETPCYWMLDAMEYTVSTSLPVSFKDQNMK
WEGAIET
EIPSRMEGNRLYRFSKILEGNGESQAKLPQCPHLVWTQPIPLVNSAPLNYQSQKLSSL
IEEVLNLHPNVDNSIKDASGE

CM.S09.001
ACYTSKMVFNYPKARRDETVVDNYFGTAVDYPRWLEDPSDEETKLFIDAQNAITKPY
LDGCOMYKDKIKSKITKLWNYPKYPVFCHGNRYYQRNRTGQLNQSFIVFQDKDLQSEAE
VFNLNPMSDDGTIALSKTAFEDGSTFYAGISNSGSDWVEIKFKDVKTGIDHKETLKRV
KYSSMAWMDNIFGGYAGYLDQTGGKAGDSETTSSDQKLYYHRGLDQSEDIMTVEFE
DTQLRIGASISHCGEYLVTPTKGCYKLVLVFYAKFSADKKIDGKQLTSSVTEFVADYEE
VHNIGSKFFFRNRTDAAANYKITVIFDDLLTSCDILLPEHPKDVLWDHAVINTMNLVACY
LTVDKNTLVIDLETTGKIHDFKVDVGTIASIAGKYYHKQMFSFFSFLPTNTHQVDFKE
GEIITDFHETKVGDPSLEIFQIYKSKDDDIPMFIIHKKQGLQDQGQNPCLYGYYGG
FNVSFLPAFSVRSLVFIRNFNGIYATANIRGGGEYGDKWHNGGRFEKKQNCFDDFQWA
AKYLVKENCYTTREKLSIMGGSNGGLLVACINQAPLOGASAASQVGDMLRYHKFTI
GYAWTDDYGCADVKEQFEYLYKYSPLHNIKEEKEQYPALELLTADHDDRVPVLHSLKF
AELQHKVGLPQPPQKNNLRVENVTAK صحغگکپتسکل

CM.S09.004
MSTKADKFVKTYKQLSQIAPAALGGAIRSNGVINTLSQRDIIEKGGKTTFIKKNFTTDLKP
VADVTPIIDVSNEILSOFSKSEFKAPRLNFKQYLEIWHQHTIRTVDLNALDVHSVY
SDGEFASFEWSPDEKLLYVAEEKPISPEFYKRKCPSTNNEGNGDEEKPKGEELYLFIQ
DWGEQMVGKYSLAAYDEIENDSVTLKGLPDDICV AQPKYSDNLGIMVGVAYQVEPR
KLGLIYCTNRLSTIFKLDLNGLVSLPLKDSVKSPIFSCDGKTLLWLQRKAGGPHASC
MLVKTENPLTESSSVVKVPIVEIKINGKIFYLYNGSFLKRPWASGNRLVNTNQK
YAVNSYVINCIENGDISEMITYDDGSQIQLDYDDLAVQRTYLKSDKLVLGKLASSGAD
ITWNLDTSEVEGLEHICYQLDLDTEGDCNSFATYLLGPKDGKEKSPLIVWHPGH
PHGSFTNNILQGAMYLKFGYAILFINFRGSGLGCGQKCVDLPKIGDTDVKCICLATDT
ALEKFSWLNPMALVGGSGHGFVLAHLGQYDPDKFHAVARNPVIDVSSMMLSIDIPD
WCHVEAGKEYTQKGVEVDNELLKMKRVSPIHAKHKVKAPTMIHGKDLRVPHHQGTE
YHLRLKANGVTPRLHIYDDHNPLGTVPNEFDHIIINSMLWISEH

CM.S09.A67
IAAACLAPPEPTEYEFVSDENGKCSFALTERAEWQFSEREKENYEVEVEFFTRTNRGNRIACL
FVRCSSTARFTILSFHGNAVAIDLQMSMFFLYLGSRMCNIFSYDYSGYGVSGAKPSEKNL
YADIDAAWQALRTRYGISPFENIIYQGQSIGTVPTVLARAAEYVGAVILHSPLMSGMRVAF
PNNTKRTWFLMPF

CM.S09.UNW_1
QIKGECKSHQSNTYESFTGVPFAEPPVECQKLPQCPHLVWTQPIPLVNSAPLNYQSQKLSSL
IEEVLNLHPNVDNSIKDASGE

IFQGAFIVGSEDCLYLNVYTPTVAKNSSKLLPVLLWGFHGGGFFQGFSGIHMMLFAFGSFDFK
DPAFFMDEEIVVSPNYRLGALGFLTTEDDVIPANLGHDDQLLALKWQDQNIELFGGDPS
KDIVIAGESSGSASVYCHLLLLSVPEINAVTGAIMISGTCIAPWAFVTNATENAFDVGGKL
GSDYSSDHNSASLRLFQSSTEDFLHAAEMYPIRKTDLGLSRGKWAPTYAKDFSAPAEP
MTDVIDQGKFHKVPLLGFNNSEECLSPLYLQQLPQIYRKAIWHDRPSKMIENLTNVRD
RRKAGEDMKAIYTNSSFLKDIAALVKYCQDDQFTPILVGRHAESASKYGVPVYMYQMA
YHFVPHTPGIEGAGHAEDMFLYWDFTLKLKLIPKFDLVSRRMVRLLSNFVKHKKPI
VEK

CM.S09.UNW_2
PPVRTHNGLVEGTQEHSEQGRVYFAFRGIPFAKPPVGRFRRFKPERADAENWTHYLA
AAHIIKQNYLINDPKVEGSEDCLYNVYTPERPSGGRDRSISQNLNLSVMVFHHWWGGFC
GYSDSKYLGPQYFMDKNVVTVFNVRVGVGLFNLSTNDDGSPNYGLKDQVMALQWV
QDNIDNFGGGNKAQVTFIQGSAGASSVHYHMLSLPSRGLFRQGEISESGASALNLWARPFP
TQNOGLIIIAAQATIGCPCDPKTTAMVDCLRKDANIVNSGDSKFTSFVDPNLVYGPSIE
KKTVQNPQFPFITQHPVEYIRNKQFQHPGWGINNEEALLRQSETRETNRNFD
SIMNLMGLESVKKEIKSINISQKDFYLEGRNKNVSNFGLNFETDLYSRFAYS
VYQAALFHNFQGHENIIFWYNSQGYSQVWATENIDGYWGVSHSDELYLF
KTPALFPSLENNDKMITCMCLNLWNTNFATKGDP

CM.S09.UNW_3
GVLQGTYMETKNGRFSFSFMGIPIYACPPIDLRFPQPPVPWKhKALDATKVKHAVCPQ
RDVYRRSIVEEGECEDCLYNVYTPKLEKFQDSPDKLVPMIFFHGGWLCSGNSMWYGP
EVLLDKDVLVLVTNYRLGALGFLMTTGDEVVPNGLSKQLQNLARWIRNNIKYFGGDPD
SUTIFGESAGASQAQHMLSLPSRGLKRAILQSTGFACTWAQSEQNVTNSKCLKIIF
NCPTDCTKGMVDCCLKKVDASKIIEQEDEKFMRLWDYDMPIFKFPIEPKLDGALTFHEPMK
LITSHKYENIPVMVGLTEDGKLVAGMSQLEEFKEKDELAPMSELYDKFLDDVT
VTKNLKRFYFGKEKIDQTKVEKLIDYTDGWFDMAADTAVMHYKHTSPVYVYLFY
YRGSASFKTIFGGGDTDFVCHADELYLFPIGGLDFDKPSEDKRITEIMTTLWTNF
AKTDGPDTPRTTEEISERKPYTAKSENYYIDSNSTINKSGFLFDRAALLWRSFTKNYDKK
YIKDEL

CM.S09.UNW_4
GHIKGSSLTSRLGKTIYAFRIGYAEPPVGELRFKPPVPLTETWNGTYDATKDGPLCPQPT
AGDPVSEDCLMVYSTKLPDKWNPKRPCSVHILPGGFYASAGSVSWAYACPQYFMDQD
IVLVTNFYRLGSGLFLATGDKEPQNGYGLKDQVVLKKWVQKNIAFSGGDPSNTLLGYS
AGSWIIFHMVSPMSKGLFHKGIAAMGSSVGSWPLPTQMDVAKKQARLVCPPDDTSDK
NIKCLKTSAAELGNSFQFREGFSPVIEPVQFERIDPNRIMNHHLQHIPFITGQTTDFSYDAFAVNNETLTKENEDWDRIAPIAFYERGTNQSLKVSRAIRKFY
LNDKPVSKGNREALGKITYDSIETFPINRAVKLMAGQYNKSCHVVYYRFSFQGRYSHFYLP
GTNNTVPYGVHDDILIYLFYIQKIFKPKNATSRPEIMDVEKLTTLHTNFATKG

57
EELNKLGFLESHSSEFNVRTNYLDWLTSLPWGVQSENMLQRASEILDQDHYGMDIK RRILEFIAVSQLKGSTQGKILCFHGPGVGKTSIAISARIALNREYFRFSVGGMTDAEIK GHRRTYVGAMPGKVQICLLKTRTENPLVILEDVDKIGKYSGDPSSALLELLELDEAFNSN FLDHYLDVPVDSKVLFICTIONVVDITPEPLRDRMEMIDMSGYVAEEKLAIACKYLLPQ AMRDSGLKQDDQIVMEALNMLIKNSCREGVRNLQHIEKVVRKVAVKVIDAEASF VQVDSNLQEFVGKPVFTHDRMYPVPTPGVVMGLAHTAMMGGSTYIETTTRRRAPEKE TDPSSLVTLGHLGEVMKESAKALTAVARNFLKNEPENFLYKHNHLHLHVPEGATPKDGP SAGCTIAATALLSLAKGESIRQDVAMTGEISLMGRILPVGGIKEKQLQLSDLG*NA*FYQK KTRRTSMIFLASSLKGLEVHFVNTFDEI

CM.S26.023
SLFDEVKRMKNRQFLYQVLFSGMIVSSAMLMIWKGLMVGTGSESPIVVVLGSMEPAFHR GDLLFLTNFPEEPVREGIEVFKVEGRDIEPEHRVLKLHEKNNGTVKFLLLTKGDNSVDD RGLYAPGQLWLTKQDVGRARGFLPYVGMVTILMNEYPKFYAVLGLGFYVVLHRE

CM.S26.A09
KSVFFGIPIGITIDCVGYIARVDGISMQPALNPDKNQTDYVFLNKWAIVKSEYIQRGDI SLVSPKDPHQIIKRVVGVEGDIATLGYKNIDRVRPEGHCWIEGDHKGHSMDSNKFVPV ALGLVTAKASCIVWPPSRQIIITPFLPSAPVNLNPSSSS

CM.S28.UNW
MRFIIILISLYGATADQYYTTKKYIDVPLHDHSYTSNLTFLKLKYINLDTHYHEKGGPIFF YTGNEGPIETFAQNSSFINDITPKFNALIVFAEHRYYGESLPFNKSYTSPQYLGYLSSE QALADFVYLLNLIQKEHASEGRRKDKLVPVIFAGSYGGMLSAWLRMKYPSVVGAIATS APIWQFDLVPDCNYRITTNVYNVIGGTCVDNIARSWKDRLAKNDTGKMNITKAW SVCNPLTDTDINDLISLEYINLMVKMNYPYPTSLAPLPAHIPEFCSRNKPYKTDTDE MLTSIGSAIVFTNFTGTKNDKNIKNDPDTLGLATGWQFTCTEMIMPMPSKDNMFSFY EWNFKSYSDTFCFKKYQVRPDVDPILQYYGKIDKASNIVFNSGMLMDPWSGGVLSNI SSTVVVIIPDGAHHLDDLGRHDMDPSAVKNAREFHIRHLEKWLHSYYLEK

CM.S33.UNW
KIFIGLAICVIYIGWKVNNFMAKPLLPPVEDKWGWPLKPRQVDSIRKFTIKIRDEVLE DLQYRLDHHRPMPPTPLENTPQQYYGMNTLVKIKVDWRTKYNWREREVFLNKYQPF R VNIDGLDVHFIHLKPDNPVSPKIKVLPNLVHPGVGSLREFYDILPTPPREGYDFVEV VAPHPGYGFSQGAVRQQLGGPIHANMFRLKMERLGHKFYMQAGDFHIILQHLAIWF PENVLGHFSFVCVFPTPSYIKHLSIASPSFFKVEEHILNLYPLSRIYLYLLETGYMHLQ ATKPDITVGVARDSPVGLAAYILEKFTTWTNKEWHDLEDGGLTKKYTLTNDNVMV YWLNRCTRITSSMRLYSEAFGGEYYGQQHWRDLPKEVPVPGCLARHELAFSPDFVLEGEKY ENLIHLEDYDNVGHFAAFEEPELFTRDVY
CM.S33.UPW_1
FVEQLGYKFETVPVTTKDGYLELLHHVYLNTSTNPVMMPGLLSSSADFLITGKNSRLALRLVDQGYDVWNIANNRGTTHYRHKALNADKDREQFWEFSFHELGYDSPACVDYILDRTKAKTLSYIGHSQGTTQYFVMSTRPEYNKSVSIMVALAPVAFMTHMPQLFFQFLKYHIIILEWRLKHHVFVEFHIYFLMTIGKLCREGSKVQWICLFYSFILQGFIDSETDMKLLPTILENFPAGSALKELHLYLQGINSGRFRQDYDYGVRINLELYQNQSSPDPYDVSKitPSVALYGKNDLVTPKDValesKNSPCQG

CM.S33.UPW_2
HKTVVDLIVEDGYPVESHYVTTEDGYILNMHRIPHGGLSGKSNKGVAFLQHGVLVLESSADF LISGPGRALGILADEGYDVWLGNFRGNRSYRNRHTNLNPWGGAKFWKDWHQGIIDLPTMDYVLKQTGVDAYYAGHSQGTTSYFVMTSLRPAYEKNKHVSALAPIFMHMTSPLRRIISWWDGPLDILLGLGMIHIFMPSSKFLDTNSFCSHGITYLCENSLFLACGFSPQ EMNITLLATVMAHTPAGAKSTQILHYQQGISSFGFRFDGWSKNKQIYGKLPDEPEYNLQQVTPYPYLYISRDWLAENDVEKLSNRGNCVGLFMSDYAFNHLVTDYIFGIANAHKVYPKVLGLFSRH

CM.S51.002
VHGYGFLDYSANDICNILTNDNVSSVLFIPYAKRHNDEYLRLAEKFKKGWNISIHTQNPQKAIQEAALVFVGGGNTFRLKLTLYDVLVKTNRRVLEEGVYIGTSAGSNVATRSIHTNDMPIVWPPTLSALNLVVPNHPFVDSPKDSHTHPREDRILEYHEEPSSTVGLREGATLYVDFNQAIKGIYARLYIRNEEPDKIKVGSVDVSYL

CM.S54.012
GHTILEDKRLRTIFNKCDKDRGYITVKELYLDIESREYEEDIPPQVVRHIHESHDQNRDKRLDNFSDMVNPNNLQHFHGYLNYRINWIIIPCPPATTFTTDGYEEETYCMPPAVGMIIFSLIEIIFCVDATAEFTSTNTASGPMATLFIYEPRRREIWRITYMLVHGPFHLVINLQLVQLMYPLVEMVLYAQLVGSLTSVIDPTVKLASSVGVS LVTAGIASIAMNWKEMSYPALQLVVFLTVIGDLAMSREADINGVQRCWRLRHSNGVGYVAHLAG ALAGVPCGLAGAEEFQANKETYLWVVAVFTFSVMLGAMVNLNFVYDTWRK

CM.T01.011
SVLCPDPSPGFGFENCRRNALLESKGLAFPATAQKTGTTIVGITYFIDGVLGADTRATEDTVADKNSKIHYPAINMYCCTGATAADTTEMIDMISSQLALHKLNTNVARVCTANMMCLKQYLFRYQGFIAALILGVDTEGSHLMIYPHGSSDKLPPVTMGSGSLAAIAVFESRWRKDMDEEIEQKQLVDAIAAGIFNLDGGSVNDLICIKKGEVDYRRTYEEANKKGV QKQNYKYPKGATSILSTKVLVEDVMVTIELEMDTA

CM.T01.012
CM.T01.971
MSRGSSAGFDRHITFSPEGRLYQVEYAFAKAINQAQLTSTVAVKGVDTACCVTQRKIPDKLIDAGTITHFLQLTDHSGCVMTMIIADSKVQVRARYEAAQF KYKYGYEMPIDTLCCRRVADISQVYTVQNAEMRPLGCSMVILIGYDSELGCPVYKADPAGYFCGYRAVSGAQLKTQTEANSYLEKMKLKDQDQLQHDDVIQLAI SCLSVLSVDFKPSEIEVGVSSEENKFRKLTEQIEIDRHLTAIAEKD

CM.T01.972
MASERSYFSLTTFSPSGKLQVEIQALAAAVAAGGASVGSIASNGVVIATENKHKISLYEHSVHKVEMITNHGMYGSGMIPDYRRVLKQARMAQQYYLYVKE PIPTVQLQVRVAVMQETYQSGGVRFGVSILICGWDTDRPLFQCDPSGAFAWKATAMGKNYNQFLEKRYSEDLELDIADAVHTAILTLKESFGQM TADMENIVGICDASGFRLDPSTVKDYLANIP

CM.T01.973
MARRYDTRTTSPEGRLYQVEYAAMEAISHAGTCGLANDGILAAERRNTNKLLDEVFTSEKIYKLNDMDVVCAGITSANDNLRLIQQYRLFQYGESIPCQ EQLVSWLDCVKQAYTQYGGKRPFVGVSILYMGWDKHYQYGQLQSDPSGNYSWKTACIGNNASSAISSQKEYEKEMMLDKALAKVKLVKTDMAEKLTSEKV EMATLTKNDKTNILSSKHEEELYEYK

CM.T01.974
MSSRYDRAITVFSPDGHLLQVEYAQEAVRKGSTAVGVRANAVALVGVEKSKVAKLEERTVRKICLDDLHVMAFAGLTADADILRAQIECQSHKLTVEDPVTE LLEYITRIASLQKYQSNRRPFGISCLLGGFDYDKPQITQPSGIYYEWKANATGRSAKTVREFLEKYTYAEVEATEKTVKLIAKALLEVEQSGQKNLEIAVMR HGEPLVMLDGDTI

CM.T01.975
MFLTREYDRGVTFSPEGRLFQVEYAIEAIKLGSTAIGICTSEGVLAVEKIRISPLMEMPTTEKIVEVDKIHIGCAVSLMADSMRTMRARIIECQHNFVYN ENISVESCAQGVSNLIQFGDSDDDAGAMSRPFVGAILFAGIDEKGPQOLYHMDPSGFVFQFDAAIGGSGEGAQQSLQEVYHRSMTLDEATTSALTIL KQVMEEKLNASNVEVMTPKDLFHMFTEQEEV

IKKLA

CM.T01.987
FIKMFDQLGSPPALWKNGPISGPFIQTKPNEPITHSKSPVTATSVIAIAYDKGV
VMAGDLLASYGSLARYKNCPRILQVNELNGSADGADOWQYKSFIEQKIIIDEDCMGDNIKLLPKSLYCWLTRVMYQRKKFDFFWNLLAVIGQDGEPLQGDVKFGTAHYTARICT
GYAAHAVLPIIREDFLDKPNPTLTTEAKALVIKCMELVLYRDRSYSKFQIGIMDKEAGSSI
EGPLACQENWEIA

CM.T01.UNA_1
SILAYNGGAMVAMKGDQCVAIAAADRRFGVQAQTISTNFKEIKFNMGMGLCGLPGLATD
TQTVMELKLKFRKYLKELKNNRMSPKVSAMLSMNLEYERFGPFFVEVPAEGLMPKTYEPYICNMDLICINAPDFFVVGTTASAQLYGMECIALWQPNLGPDDLFETASQALINAF
DRDA

CM.T01.UNA_2
MECLLGIFKFDFVLIAADMTASQSIIVMKTDENKLFLKSDKLVMAVAGESGDTTQFAEY
IARKIQLYKMRNTRYELPSKEAASRTRNLDALTIRLTMSPYHVNLALAGYDEKDGPQOLLYY
MDYLASAVSDYAAHGYGYYFSLSIMDRNYLKNLNEKQAYDLLKCVKEIQLRLAINL
PNFKVQAIDKNGIHMDPNTLDLK

CM.T01.UNA_3
MSLLERPYYPYSSAPIQHPQFSPYADNGGVSVAIAAGDFVVAIAADTRLSTFGSIYTREMQKLFPALSQGILGCACGCWCTDLTTLTRILKSRMQMNYQHEHNTMSTTAMAQMLSTLLYY
RFFPPYYISNVLVGFDTEGKCGVFYDPHCEKTKYKAGSAGALLQPLLDNQIAHTNLDKVPKEPLTVEALKALLDKVFISAADERIYTDGDNLINIITKDIKEEHEFDFLRKD

CM.T01.UNA_4
MSSTGTGYDLSASQFSDFGRVQVEYAMAKAVENSGTVIGLGTKGDIVLAAEKLLLHHEPNTNRRYINVRHIGMAFSGLIADARQIEJARKEAANYRQYGVISLKLYNDRVSMYMHAYTLYASRVYSGCSVILASYDDKEPNMYLIDPSGVYGYHCATGKAKQSKAETEE
KLKLGQMTCRELVKEAAKIYTVHDELKDKNFELELSWVCQESNYYHVRVPDVYCDAEKAQAMEADSESDTEDM

CM.T02.001
TIILTLPTGLKAYKGAUSVSNDLCAEIGKIKLEKGGGTADDAAAIYAVLFCGTVLPECMGLG
GGFLLTYQYSTNVEICLSNRETPAPAQQHARDMNYNGNRFASQGLLLAVAVPGELIYWTLYNRFKKGVPWADLVFQSIDLCKEKFPGVGGFLASKFDLEPDRLIGPETKELVHPHTKK
PYKPEIILKRPRLAKTLRIIAKHGMALHDGVLTKGFVKDVQDRGILTVEDMRNYKIPWNVQQTTSNHCTLYSIPVGSGAVLSSILNLDNLMRGDFYTPINQHRMIESFKFAY
GARTKLGDATTPFMRLISGLSMDKSKYAIKELRSYIDKNTDFSVAYYGANTTFEDHT
AHMSVLAPNGDAAVTSTINGRWGAYFGSPSTGIILNEMDDFSSPGFQNYLPHYGLPPSPSN
FIQPGRRISSMSPIIIVNEKGDVVMIGAAGSKITTSVQAIIKHLWYIGDLKRAMDIA
RVHHQLMPNVLVVEDNFKIEYKLLHQIEKIGHNIEFKYPADGFCAVTSISHVNRGNVV
GVPDIRR

CM.T02.001_1
INTWKFQHAAQKAWSVLSRGEDLALDAITAGCEVCQDEQCDFTVGFGGSPDENGETTL
DAMIFDGDTMDMGAVGGMRNIAKAARYVLERTEHSLLVGSLATQFARGFGLPEES
LSTNYSRNWENWKQHSCQPNFWKNVRPNPKACGPYKTLDIGKQWENTNKFNTGHNH
DTIGMIVISRTGHIVAGTGSTNATFKIPGRVGDSPIPGAGAYADSRIGAAAATGDGDIMM
RFLPSFLAVEQLRQGASPSKAAKIAISRIVEKYPKKFGGIIVTDDKGEIGAACNGMDKF
PTIANTVRFPEGLTKYVTC