Diversity of cysteine cathepsins in two coleopterans, *Tenebrio molitor* and *Tribolium castaneum*

Alexander G. Martynov

Faculty of Bioengineering and Bioinformatics, Moscow State University,

Leninskiye Gory 1-73, Moscow 119991, Russian Federation, agmart@mail.ru

Brenda Oppert

USDA Agricultural Research Service, Center for Grain and Animal Health Research

1515 College Avenue, Manhattan, KS 66502, USA bso@k-state.edu

Elena N. Elpidina

A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University

Leninskiye Gory 1-40, Moscow 119991, Russian Federation, elp@belozersky.msu.ru

Tenebrio molitor and *Tribolium castaneum* are two closely related species of darkling beetles (Tenebrionidae) with similar spatial organization of digestion and diet. The major role in their protein digestion belongs to secreted cysteine peptidases (CPs) from papain family, mostly cathepsins L and B, that have been discovered in mammals as lysosomal enzymes.

This study explored diversity in papain-like CPs in *T. molitor* and *T. castaneum*. Primary structures of *T. molitor* proteins were obtained from high-throughput Illumina sequencing of cDNA, synthesized on midgut mRNA samples. *T. castaneum* sequences were obtained from sequenced *T. castaneum* genome [1]. Composition of substrate binding sites of predicted proteins were derived by homology with well investigated human and animal CPs based on multiple sequence alignment and comparison of modeled and experimentally obtained 3D structures. We also studied possible alternative localization of cathepsins to lysosomes or secretion to the midgut lumen using structural markers of mannose-6-phosphate pathway. Levels of cathepsins genes expression were assessed by the high-throughput Illumina sequencing data for the midgut transcriptome in both insect larvae.

Seventeen sequences similar to cathepsins L and 15 similar to cathepsins B were found in *T. molitor* midgut. Nine of them were found in the previous paper [2], and 23 were new. From *T. castaneum* genome we used 15 cathepsin L-like sequences and 9 cathepsin B-like sequences [1]. *T. molitor* cathepsins L set contained 3 peptidases that showed binding

subsites composition corresponding to human lysosomal cathepsin L, while in *T. castaneum* there was found only one typical cathepsin L peptidase. Composition of subsites in most of obtained peptidases from both organisms did not correspond to any of described cathepsin types. A large proportion of sequences had tryptophan residue in S2 subsite lacking in human cathepsins. In some peptidases S2 subsite contained hydrophilic and even positively and negatively charged amino acid residues, while classical cysteine cathepsins usually have hydrophobic residues. Furthermore, 3 sequences that have major changes in S1 and S2 binding subsites (compared to typical) have changes in S1 subsite residues that are potentially responsible for substrate specificity. Cathepsin B-like sequences can also be separated into classical cathepsins B (3 sequences in each organism) and a group of predicted proteins that have significant differences in the structure of occluding loop, which is very important in classical cathepsins B due to the presence of additional active center. Moreover, cathepsin B-like sequences also showed variety in S2 subsite containing both hydrophobic and hydrophilic residues including charged amino acids.

All obtained sequences can be divided into two groups according to the expression levels of their genes: low-expressed and high-expressed. The closest to human cathepsin L sequences in both organisms belong to low-expressed group and they are predicted to be lysosomal. Moreover, it is possible that there is a correlation between the level of expression and localization of cathepsins: high-expressed genes code secreted proteins and lysosomal enzymes are mostly coded by low-expressed genes. For instance, cathepsin L with the highest expression level in both organisms was described biochemically in *T. molitor* larvae as the major cysteine digestive peptidase.

This work was supported by RFBR (grants # 12-04-01562-a and 12-04-31451-mol-a) and ISTC (grant #3455).

1. *Tribolium* Genome Sequencing Consortium (2008) The genome of the model beetle and pest *Tribolium castaneum, Nature* **452(7190):**949-955

2. B. Oppert et al. (2012) *Bacillus thuringiensis* Cry3Aa protoxin intoxication of *Tenebrio molitor* induces widespread changes in the expression of serine peptidase transcripts. *Comp. Biochem. Physiol.*, **7D(3):**233-242.