

Structure and function analysis of protein HD73_0859 produced by *Bacillus thuringiensis*

Dandan Wang, Nan Zhang and Shuyuan Guo*

School of Life Science, Beijing Institute of Technology, Beijing 100081, China

Abstract. HD73_0859 is a secretory protein from *Bacillus thuringiensis* (Bt). In this study, its structural feature was analyzed and its function was predicted. The results revealed that HD73_0859 is a peptidase consists of a signal peptide, three well-separated SH3_3 domains and one Peptidase_M23 domain. The peptidase belongs to the M23/37-metallopeptidase family, and might have enzymatic activity to cell wall peptidoglycan and some elastins. This paper improves the understanding of both the protein HD73_0859 and the M23/37-metallopeptidases family, and lays a foundation for further experimental testing and theoretical study.

Keywords: *Bacillus thuringiensis*, HD73_0859, structure, function, analysis

1. Introduction

Bacillus thuringiensis (Bt) has been widely exploited as biological insecticides due to the insecticidal properties of the crystal proteins produced during its sporulation. The proteins identified to be insecticidal are δ -endotoxins (named as Cry and Cyt toxins based on their primary sequence identity). These toxins are highly specific to many species of insects including agricultural and forest pests and several vectors of human and animal diseases [1]. Presently these proteins have been classified as Cry1 to Cry55; Cyt1 and Cyt2. The Cry group is also classified into many subgroups (Cry1Aa, Cry1Ab, Cry1Ac...) [2]. With its wider application, new requirements have been put forward. For example, low tolerance to environmental stresses in the field (such as UV radiation, rain and temperature) is hindering the broader applications of these insecticidal proteins [3]. Several protective methods (such as the encapsulation of Bt strain or Cry proteins) have been developed, but their application is still yet to be tested because of the high costs and the risks of environmental contamination. Therefore, finding a reliable, safe and feasible alternative method is still crucial.

Recent studies showed that cell envelope can protect the insecticidal crystal proteins of *B. thuringiensis* against external environmental stresses, while still maintain the insecticidal activity [4]. In order to improve the effectiveness and tolerability of insecticidal proteins, it is possible to construct

*Corresponding author: Shuyuan Guo, School of Life Science, Beijing Institute of Technology, Beijing 100081, China. Tel.: 68914495-803; Fax: (010)-68915956; E-mail: guosy@bit.edu.cn.

B. thuringiensis mutants that the mother cell lysis is blocked but Cry protein production is not affected [5]. Some genes were found to be potential candidates for the mutant construction. The complete genome sequence of *B. thuringiensis* HD73 has been determined [6]. *B. thuringiensis* HD73 could produce Cry1Ac toxin which is highly specific to some insects of lepidoptera (for example, the Asian corn) but is innocuous to humans or plants. Among the numerous open reading frames of HD73, there are many ones whose biological functions are still unknown, and gene *HD73_0859* is one of them. The structure and possible physiological function of protein HD73_0859 encoded by gene *HD73_0859* were analyzed in this study. Protein HD73_0859 was predicted to be a peptidase belonging to M23/37 family.

Peptidase family M23/37 contains zinc metallopeptidases with a variety of digestion specificities. Members of family M37 are Gly-Gly endopeptidases, and members of family M23 are also endopeptidases which could digest peptidoglycan of cell wall and some elastin. For example, LytM, an autolysin produced by *Staphylococcus aureus*, is a zinc metallopeptidase with preference for glycyl-glycine peptide. After being processed with trypsin, it shows clear activity to peptidoglycan and splits pentaglycine into di- and triglycine [7]. Bacterial metallopeptidases of family M23/37 contain Zn^{2+} , and the systematic database searches show that the most common ligand for Zn^{2+} is histidine, followed by glutamic acid, aspartic acid and cysteine [8].

With the development of science, a growing number of disciplines and technologies have been applied with biology, such as computer technology, databases, mathematics, physics and so on. They complement each other and make a research more accurate and in-depth. In this study, computational approaches were used to analyze the protein's structure and function, which could provide a basis for future experimental studies.

2. Methods and materials

NCBI database and the Basic Local Alignment Search Tool (BLAST) were used for obtaining the DNA sequence and initial analysis. Then the gene sequence was translated into a protein sequence by Primer 5, and conserved domain analysis was done by CDD [9]. MEROPS [10] was used for functional identification, and it provided in-depth information about the peptidase family M23/37. Various physical and chemical parameters such as the molecular weight, theoretical pI, amino acid composition of the protein were calculated by ProtParam [11]. SignalP 4.1 was used for signal peptide and cellular location identification [12]. Finally, PSIPRED [13] and Phyre2 [14] were used to predict the secondary structure; SWISS-MODEL [15] was used to predict the tertiary structure; and 3DLigandSite [16] was used for the analysis of 3D construction and ligand sites of the protein.

3. Results

3.1. Conserved domain analysis

The analysis of open reading frame showed that the whole HD73_0859 protein is consisted of 384 amino acids with a molecular weight of 42.283kDa. SignalP 4.1 predicted that there is a signal peptide of 32 amino acids located at the N-terminal of the protein. The conserved domain analysis predicted four domains: three Src homology 3 (SH3) domains and a Peptide_M23 domain, as shown in Figure 1.



Fig. 1. Conserved domain analysis of protein HD73_0859.

SH3_3 domain is involved in binding to the cell wall; it may help the protein locate to the cell wall before it plays its enzymatic role.

Peptidase_M23 domain is highly conserved among the members of this family, and this domain is a zinc-dependent catalytic unit. However this domain’s substrate-binding pocket is usually blocked by a pro-peptide in the N-terminal region, which is quite common among peptidase family M23. The presence of Peptidase_M23 domain suggests that the wild-type full-length protein is expressed as a proenzyme in an inactive conformational state and needs further processing to become active. For example, wild-type LytM is inhibited by its own N-terminal loop and it can be strongly activated by the peptidase trypsin which has no physiological relation with it [17]. The full-length NMB0315 (a protein of *Neisseria meningitidis*) is also auto-inhibited, and the inhibition of its active site is achieved by an N-terminal helix [18]. It seems that the proteins in this family share the conformational feature of using a segment of the N-terminal to block the active site in the C-terminal. Therefore it is possible

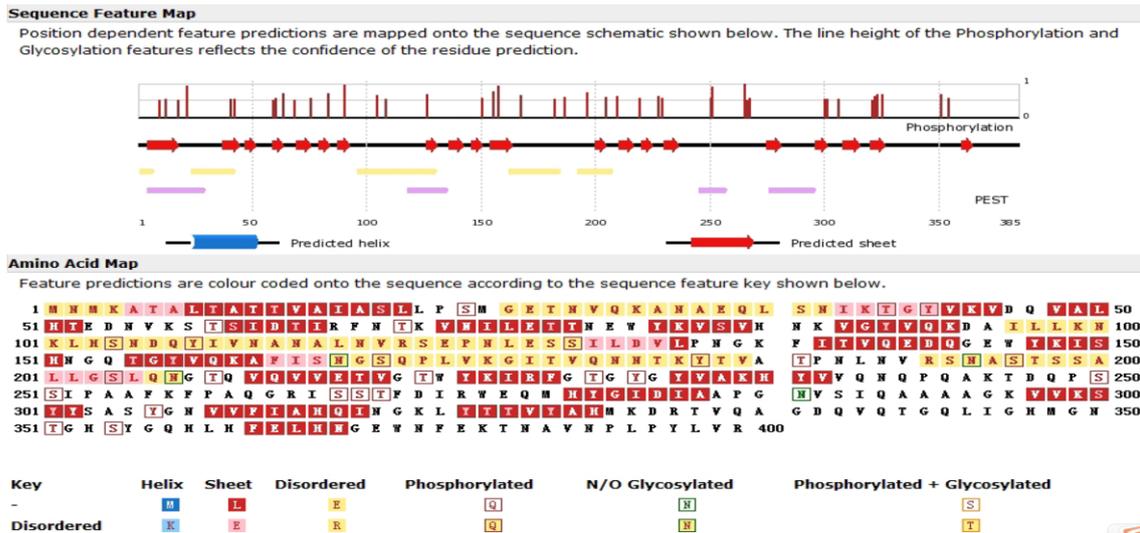


Fig. 2. Secondary structure assignment of protein HD73_0859. Helices are shown in blue, sheets are in red, and disordered regions are in yellow. The upper part of the figure visually shows that the protein consists of a lot of sheets; the lower part of the figure is “amino acid map”, which displays each amino acid’s position in the secondary structure correspondingly.

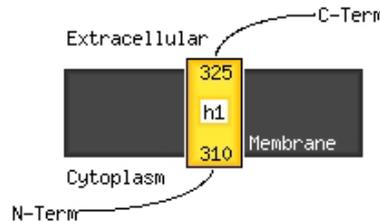


Fig. 3. Predicted transmembrane topology of protein HD73_0859.

that the wild-type full-length protein HD73_0859 from Bt is also a latent form. If this is true, searching for a method to activate the proenzyme will be significant.

3.2. Sequence feature analysis

Protein HD73_0859 is composed of a large number of sheets. Also no predicted helix was found in the whole sequence (Figure 2).

Phyre2 search of deduced amino acid sequences of HD73_0859 suggested this protein has a transmembrane topology region (the region from amino acids 310 to 325), when the N-terminal end of the protein is located in the cytoplasm and the C-terminal end is located in the extracellular (Figure 3). However, the region from amino acids 275 to 378 is homologous to the members of the peptidase family M23. It indicates that the SH3_3 domains exist in the cytoplasm, while the C-terminal region containing the M23 domain sequence might be located in the cytoplasm, periplasm or extracellular matrix, because one predicted membrane-spanning region exists in this domain.

3.3. Tertiary structure and ligand site analysis

Based on protein sequence alignments, two models were built for the HD73_0859 protein molecule. The structure of the N-terminus was built on the foundation of the N-acetylmuramoyl-L-alanine amidase and no ligand site was found in this region. However, the C-terminus of the protein shows large similarity to glycine-glycine endopeptidase lytM which belongs to family M23, and a ligand site was predicted in this domain (Figure 4). This figure also displays that HD73_0859 protein contains only sheets.

The predicted active site is organized around a divalent metal cation (Zn^{2+}). Together with the results of protein sequence alignment, ligand site analysis also revealed that both “HXXXD” and “HXH” short zinc binding consensus sequences [19] are present in the Peptide_M23 domain of HD73_0859 protein. And there is a Zn^{2+} in the internal structure of the protein itself. This Zn^{2+} is surrounded by three spatially adjacent residues (Figure 5); two of them come from the H-X (3)-D motif (His276 and Asp280), while the third is from the H-X-H motif (His360, the second histidine of this motif). A sequence alignment revealed that these three residues, and the organization of the active sites are highly conserved among the members of peptidase family M23.

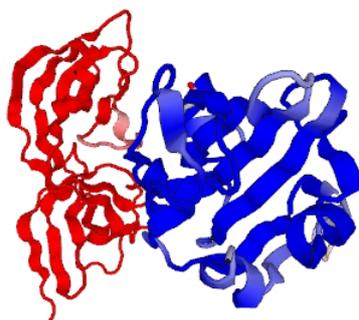


Fig. 4. One HD73_0859 protein molecule. The molecule is colored with the N-terminus in red and the C-terminus in blue.

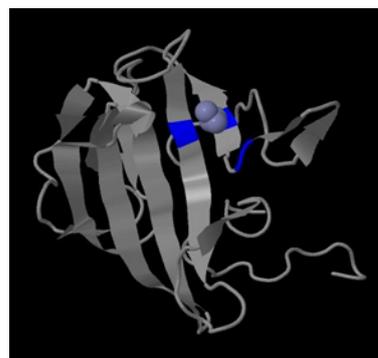


Fig. 5. Ligand site analysis of HD73_0859 protein.

4. Discussion

Our results suggest that HD73_0859 protein has three Src homology 3 (SH3) domains involved in cell wall binding, and a Peptide_M23 domain as a catalytic unit. The protein is suggested to be anchored in the cell wall by the three SH3_3 domains and a transmembrane region in the Peptide_M23 domain. It is structurally and bio-chemically characterized as zinc metalloendopeptidases of family M23/37. Moreover it is speculated that the HD73_0859 protein might have a proteolytic activity toward peptidoglycan cell wall matrix or some elastin, like some others in this family.

It should not be overlooked that full-length HD73_0859 might be in an auto-inhibited conformational state. However, no certain way has been found to activate the proenzyme of this family so far. Only a few members have been activated by different ways with no common rules to follow. For example, LytM is strongly activated by the peptidase trypsin which has no physiological relation with it, while amidase EnvC which has a LytM domain is activated via the induction of a reversible conformational change rather than through a proteolytic processing event [20]. The difficulty to get the mature form with biological activity is the biggest limiting factor for the study of peptidases in this family. This paper improves the understanding of the M23/37-metallopeptidases family, and lays a foundation for further experimental testing and theoretical study.

Although the biological function of protein HD73_0859 is still not verified, the finding in this study can deepen our understanding of the protein and provide a theoretical basis for future study. Presumptively it could digest peptidoglycans on the cell wall; it also could be involved in numerous cellular processes including cell growth, cell-wall turnover, peptidoglycan maturation, cell division, separation and mother cell lysis. Moreover the *HD73_0859* gene could be a potential candidate for mutation to develop a new way to protect the Cry proteins from degradation in the field and enhance their persistence and tolerance.

Acknowledgement

This work is supported by grants from the National Natural Science Foundation (Grant No. 31171911).

References

- [1] I. Arenas, A. Bravo, M. Soberon and I. Gomez, Role of alkaline phosphatase from *Manduca sexta* in the mechanism of action of bacillus thuringiensis Cry1Ab toxin, *The Journal of Biological Chemistry* **285** (2010), 12497–12503.
- [2] N. Crickmore, D.R. Zeigler, J. Feitelson, E. Schnepf, J.V. Rie, D. Lereclus, J. Baum and D.H. Dean, Revision of the nomenclature for the bacillus thuringiensis pesticidal crystal proteins, *Microbiology and Molecular Biology Reviews* **62** (1998), 807–813.
- [3] W. Yang, K. He, J. Zhang and S. Guo, Ph-controlled *Bacillus thuringiensis* Cry1Ac protoxin loading and release from polyelectrolyte microcapsules, *PloS One* **7** (2012), e45233. doi: 10.1371/journal.pone.0045233
- [4] V. Sanchis, M. Gohar, J. Chaufaux, O. Arantes, A. Meier, H. Agaisse, J. Cayley and D. Lereclus, Development and field performance of a broad-spectrum nonviable asporogenic recombinant strain of bacillus thuringiensis with greater potency and UV resistance, *Applied and Environmental Microbiology* **65** (1999), 4032–4039.
- [5] J. Yang, Q. Peng, Z. Chen, C. Deng, C. Shu, J. Zhang, D. Huang and F. Song, Transcriptional regulation and characteristics of a novel N-acetylmuramoyl-L-alanine amidase gene involved in bacillus thuringiensis mother cell lysis, *Journal of Bacteriology* **195** (2013), 2887–2897.

- [6] G. Liu, L. Song, C. Shu, P. Wang, C. Deng, Q. Peng, D. Lereclus, X. Wang, D. Huang, J. Zhang and F. Song, Complete genome sequence of *Bacillus thuringiensis* subsp. *kurstaki* strain HD73, *Genome Announcements* **1** (2013), e0008013. doi: 10.1128/genomeA.00080-13
- [7] S.G. Odintsov, I. Sabala, M. Marcyjaniak and M. Bochtler, Latent LytM at 1.3Å resolution, *Journal of Molecular Biology* **335** (2004), 775–785.
- [8] B.L. Vallee and D.S. Auld, Zinc coordination, function, and structure of zinc enzymes and other proteins, *Biochemistry* **29** (1990), 5647–5659.
- [9] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng and S.H. Bryant, CDD: A conserved domain database for the functional annotation of proteins, *Nucleic Acids Research* **39** (2011), D225–229.
- [10] N.D. Rawlings, M. Waller, A.J. Barrett and A. Bateman, MEROPS: The database of proteolytic enzymes, their substrates and inhibitors, *Nucleic acids Research* **42** (2014), D503–509.
- [11] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel and D.F. Hochstrasser, Protein identification and analysis tools in the ExPASy server, *Methods in Molecular Biology* **112** (1999), 531–552.
- [12] T.N. Petersen, S. Brunak, G. von Heijne and H. Nielsen, SignalP 4.0: Discriminating signal peptides from transmembrane regions, *Nature Methods* **8** (2011), 785–786.
- [13] D.W. Buchan, F. Minneci, T.C. Nugent, K. Bryson and D.T. Jones, Scalable web services for the PSIPRED protein analysis workbench, *Nucleic Acids Research* **41** (2013), W349–357.
- [14] L.A. Kelley and M.J. Sternberg, Protein structure prediction on the Web: A case study using the phyre server, *Nature Protocols* **4** (2009), 363–371.
- [15] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T.G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede, SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Research* **42** (2014), W252–258.
- [16] M.N. Wass, L.A. Kelley and M.J. Sternberg, 3DLigandSite: Predicting ligand-binding sites using similar structures, *Nucleic Acids Research* **38** (2010), W469–473.
- [17] M. Firczuk, A. Mucha and M. Bochtler, Crystal structures of active LytM, *Journal of Molecular Biology* **354** (2005), 578–590.
- [18] X. Wang, X. Yang, C. Yang, Z. Wu, H. Xu and Y. Shen, Crystal structure of outer membrane protein NMB0315 from *neisseria meningitidis*, *PloS One* **6** (2011), e26845. doi: 10.1371/journal.pone.0026845
- [19] N.M. Hooper, Families of zinc metalloproteases, *Federation of European Biochemical Societies* **354** (1994), 1–6.
- [20] N.T. Peters, C. Morlot, D.C. Yang, T. Uehara, T. Vernet and T.G. Bernhardt, Structure-function analysis of the LytM domain of EnvC, an activator of cell wall remodelling at the *Escherichia coli* division site, *Molecular Microbiology* **89** (2013), 690–701.