

Three-dimensional (3D) structure prediction and function analysis of the chitin-binding domain 3 protein HD73_3189 from *Bacillus thuringiensis* HD73

Yiling Zhan and Shuyuan Guo*

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

Abstract. *Bacillus thuringiensis* (Bt) is capable of producing a chitin-binding protein believed to be functionally important to bacteria during the stationary phase of its growth cycle. In this paper, the chitin-binding domain 3 protein HD73_3189 from *B. thuringiensis* has been analyzed by computer technology. Primary and secondary structural analyses demonstrated that HD73_3189 is negatively charged and contains several α -helices, aperiodical coils and β -strands. Domain and motif analyses revealed that HD73_3189 contains a signal peptide, an N-terminal chitin binding 3 domains, two copies of a fibronectin-like domain 3 and a C-terminal carbohydrate binding domain classified as CBM_5_12. Moreover, analysis predicted the protein's associated localization site to be the cell wall. Ligand site prediction determined that amino acid residues GLU-312, TRP-334, ILE-341 and VAL-382 exposed on the surface of the target protein exhibit polar interactions with the substrate.

Keywords: *Bacillus thuringiensis*, chitin binding domain 3 protein, 3D structure, functional analysis

1. Background

Bacillus thuringiensis (Bt) is a Gram-positive protein, whose mother cells can produce insecticidal crystal proteins during the stationary phase of the growth cycle and release it in the decline phase. These proteins can efficiently terminate specific insects, and thus control certain insect species [1]. The insecticidal mechanism of Bt has been studied worldwide [2, 3]. To improve its pest control capacity, it is of great significance to comprehensively understand the insecticidal action mode of Bt.

Chitin is the second most abundant natural biopolymer, which belongs to a group of water-insoluble biopolysaccharides composed of β -1, 4-linked GlcNAc units. Chitin presents in crustaceans, fungal cell walls, protozoa, algae, nematodes and other simple organisms, and can be degraded by chitinases [4, 5]. Chitin is not only an important component of the insect cuticle, but also an important component of insect peritrophic matrices (PM). A PM is a membrane structure inside the insect midgut, which can separate the gut contents from the epithelial cells of the insect midgut and thus protect the insect from damage caused by viruses, bacteria and other organisms [6-8]. Chitin plays an

* Address for correspondence: 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.

important physiological role in the protection of insects which makes it the target for insect termination mechanisms. Chitin binding proteins (CBP) are a group of proteins that interact highly specifically with chitin, which is widely found in animals, plants and micro-organisms. CBP is believed to be an important functional protein for bacteria.

Computer technology used for the sequence analysis and characterization of proteins has been widely used in recent years. The use of computational tools can provide information about a variety of structural and physicochemical properties of proteins. The analytical results of computational tools are extremely reliable, which can help us better understand the structure-function relationship of proteins and substrate-protein interactions.

2. Methods and materials

DNA and the amino acid sequence of the chitin-binding domain 3 protein HD73_3189 were obtained from the NCBI database, CDS: YP_007422288.1. The Conserved Domain Database (CDD) [9] of NCBI's Protein Database was used to perform conserved domain analysis of the target protein. Physico-chemical parameters of the protein sequence (amino-acid and atomic compositions, isoelectric point, etc.) were analyzed by the ProtParam tool of ExPaSY [10]. The presence and location of signal peptide cleavage sites in the target amino acid sequences were predicted by SignalP 4.1 server [11]. PSORTb 3.0 [12] was used to predict the subcellular localization (SCL) of the protein. Secondary structure analysis of the query protein was performed by PSIPRED [13] and Phyre2 Server [14]. The 3D model of the target protein was predicted and automatically generated by SWISS Model, which was compared to available homologous proteins of the target protein in PDB by using the visualization tool of Swiss-PdbViewer [15, 16].

3. Results and discussion

3.1. Primary and secondary structure analysis

Primary sequences analyses of the target protein have shown that the chitin-binding domain 3 protein HD73_3189 contains 455 amino acids, and has a molecular weight of 49783.6 Daltons and an isoelectric point (pI) of 6.22, while the total number of negatively (Asp+Glu) and positively (Arg+Lys) charged residues are 46 and 42, respectively. Results determined that protein HD73_3189 is a negatively charged protein. Secondary structure analysis was performed by various servers, and the protein was shown to contain several α -helices, aperiodical coils and β -strands; the predicted percentage of β -strands was higher than α -helices, as shown in Figure 1.

3.2. Domain and motif analysis

The SignalP 4.1 software predicted that the 40 amino acids located at the N-terminus of the deduced amino acid sequence demonstrated typical attributes of a signal peptide with a maximum cleavage site probability (C-score 0.574) between Ala-40 and His-41 on the target protein. Phyre2 server analysis of the amino acid sequence of the target protein suggested the presence of a transmembrane topology region (Thr-32 through Pro-47) in the protein. Furthermore, the 1-25 Signal peptide at the N-terminal of the protein is located in the extracellular matrix, and the C-terminal of the protein is located in the

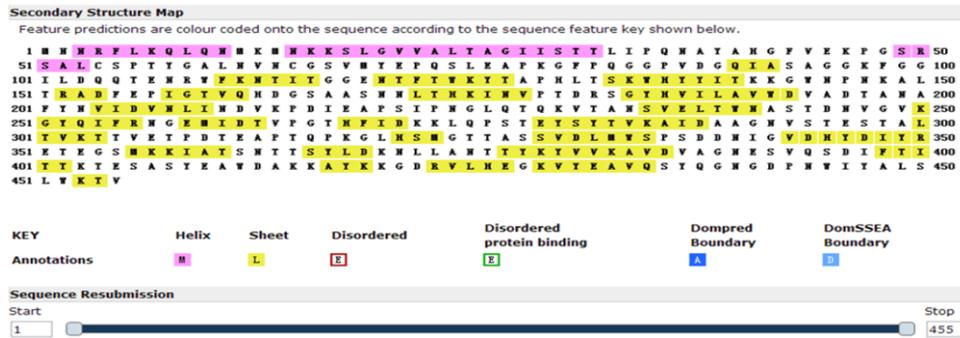


Fig. 1. Secondary structural analysis of protein HD73_3189.

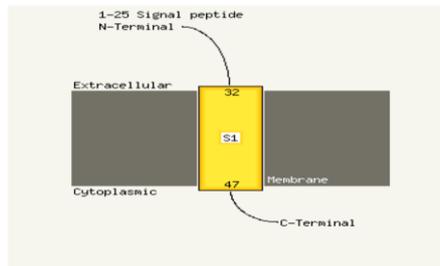


Fig. 2. Transmembrane topology prediction of protein HD73_3189.

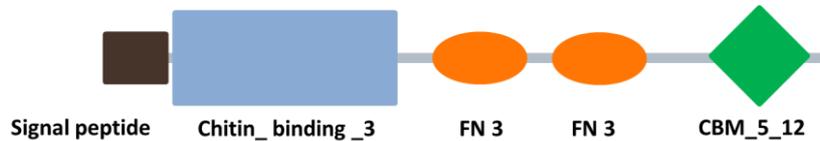


Fig. 3. Conserved domain analysis of protein HD73_3189.

cytoplasm, as shown in Figure 2. PSORTb 3.0 also predicted the presence of the signal peptide in the protein, which moreover predicted the protein’s associated localization site in the cell wall.

Conserved domain and motif analyses predicted the presence of four domains in the protein: an N-terminal chitin_binding_3 domain (CBD_3, His-41 through Asn-208), two copies of a fibronectin-like domain 3 (FN3, Pro-220 through Thr-304, Pro-314 through Thr-402), and a C-terminal carbohydrate binding domain (Tyr-409 through Pro-443) classified as CBM_5_12, as shown in Figure 3. The chitin_binding_3 domain is associated with a wide variety of cellulose binding domains. Chitin_binding_3 domain is a specific chitin binding domain, with a chitin binding function. The CBM_5_12 domain is found in various glycosyl hydrolase enzymes, and is presumed to have a carbohydrate binding function. This domain consists of six aromatic groups that may be important to the binding process. The presence of a CBM_5_12 domain and a Chitin_binding_3 domain may assist the target protein to bind with some functional structure that is composed of chitin or polysaccharides. The FN 3 domain is one of three types of internal repeats found in the plasma protein fibronectin. Its tenth fibronectin type III repeat contains an RGD cell recognition sequence in a flexible loop between two strands. Approximately 2% of all animal proteins contain the FN3 repeat; including extracellular and intracellular proteins, membrane spanning cytokine receptors, and adhesion molecules; FN3-like domains are also found in bacterial glycosyl hydrolases. Its presence in the protein HD73_3189 demonstrates that this protein may have a similar function to bacterial glycosyl hydrolases.

These results suggest that protein HD73_3189 may be inserted in the cell membrane, and spans into the extracellular matrix. Furthermore, it is located in the cell wall near the two chitin binding domains (CBM_5_12 and CBD_3).

3.3. Molecular homology modeling and ligand site prediction

The amino acid sequence of the target protein was submitted to the online server SwissModel to search for templates and build a model. Then, the template with maximum sequence identity and a minimal e-value was chosen. The most reliable 3D model of the target protein built by ProMod Version 3.70 was chosen based on its QMEAN4 value, GRE, RMSD, and the Errat quality factor estimated by Swiss-PdbViewer tools. The template selected for the N-Terminal domain of protein HD73_3189 is 2BEM (Seq Identity 56.02%, the CBP21 from *Serratia marcescens*), and the template selected for the C-Terminal domain is 1ed7.1 (Sequence Identity: 35.7%, the chitin binding domain of *Bacillus circulans* WL-12 chitinase).

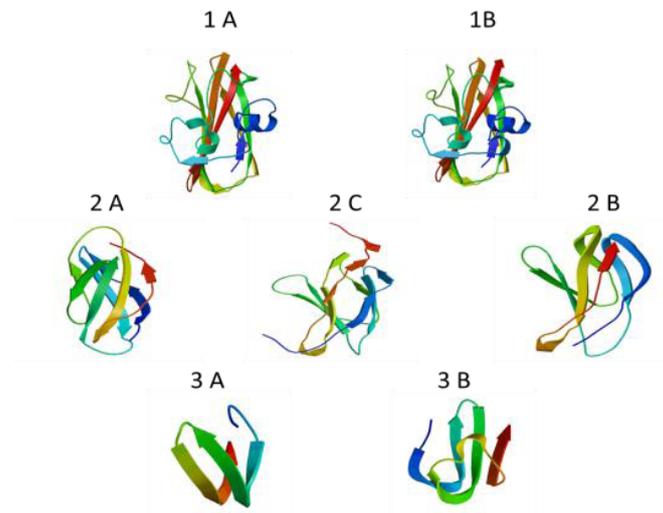


Fig. 4. Templates and molecular homology model: 1A) N-Terminal domain of protein HD73_3189 generated by SwissModel; 1B) CBP21 (2BEM) downloaded from PDB; 2A) the first fibronectin-like domain 3 of protein HD73_3189 generated by SwissModel; 2C) Template (2dju.1) downloaded from PDB; 2B) the second fibronectin-like domain 3 of protein HD73_3189 as generated by SwissModel; 3A) C-Terminal domain of protein HD73_3189 generated by SwissModel; 3B) template (1ed7.1) downloaded from PDB.

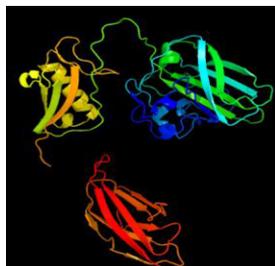


Fig. 5. Molecular homology model of protein HD73_3189 generated by 3Dligandsite.

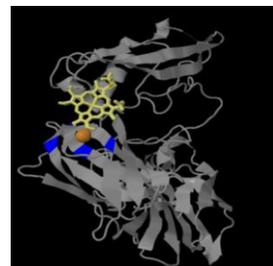


Fig. 6. Structural view of binding site prediction; the residues in blue are predicted components of the binding site.

The template selected for the FN3 domain is 2dju.1 (Sequence Identity: 32.14%, the FN3 domain of human receptor-type tyrosine-protein phosphatase F). The 3D models of the four domains are shown in Figure 4; the 3D model of the target protein is shown in Figure 5. The CBP21 (2BEM) from *Serratia marcescens* has been studied extensively, and has been shown to bind to the insoluble crystalline substrate and strongly promote the hydrolysis of crystalline chitin by chitinases [17]. It can be seen in Figure 4 that the structure of protein HD73_3189 (1A) and the structure of CBP21 (2BEM) from *Serratia marcescens* (1B) demonstrate a high similarity; they can therefore be speculated that protein HD73_3189 may demonstrate functional similarities to CBP21.

The binding site prediction of the target protein was performed by the 3Dligandsite server. The prediction determined that the binding site of the protein is GLU-312, TRP-334, ILE-341 and VAL-382, as shown in Figure 6. Multiple sequence alignments have shown that these residues are highly conserved as the active sites among chitin binding proteins. Some CBPs such as CBP24, CBP50 from *Bacillus thuringiensis* also contains these binding sites [18, 19].

Some studies have shown that CBP acts as an important functional protein in bacteria with a variety of physiological and biochemical functions. Dong Liu [20] and Muhammad Aamer Mehmood [21] found that CBP inhibited the spore germination of certain species of fungi. A chitin binding protein from *Bacillus thuringiensis* HD-1 has shown that this chitin binding protein is expressed in the spore mother cell and interacts with Cry1Ac to potentiate its insecticidal activity and facilitate the propagation of the *bacillus* strain in the environment by inhibiting the growth of certain fungi [22]. However, the 3D structure of this chitin binding protein is uncertain. According to the result of the multiple sequence alignment of protein HD73_3189, the HD-1 chitin binding protein previously described was found to share a Sequence Identity of 100% with protein HD73_3189. The structure of protein HD73_3189 may be similar to the structure of the HD-1 chitin binding protein, and may have some similar characteristics and properties. Although the biological function of protein HD73_3189 is not yet verified, the findings in this study can help us better understand the proteins structure and provide a theoretical basis for future study. Presumptively, it could bind with some functional structures that are composed of chitin or polysaccharides to potentiate insecticidal activity. The HD73_3189 gene could be a potential candidate for mutation to develop a new way to potentiate the insecticidal activity of Bt.

Acknowledgment

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

References

- [1] E. Schnepf, N. Crickmore, J. Vanrie, D. Lereclus, J. Baum, J. Feitelson, D.R. Zeigler and D.H. Dean, *Bacillus thuringiensis* and Its Pesticidal Crystal Proteins, *Microbiology and Molecular Biology Reviews* **62** (1998), 775–806.
- [2] Andre Luiz de Almeida Melo, Vanete Thomaz Soccol and Carlos Ricardo Soccol, *Bacillus thuringiensis*: Mechanism of action, resistance, and new applications: A review, *Crit Rev Biotechnol* **29** (2014), 0738–8551, 1-10.
- [3] A.I. Aronson and Y. Shai, Why *Bacillus thuringiensis* insecticidal toxins are so effective: Unique features of their mode of action, *FEMS Microbiology Letters* **195** (2001), 1–8.
- [4] E. Cohen, Chitin synthesis and degradation as targets for pesticide action, *Archives of Insect Biochemistry and Physiology* **22** (1993), 245–261.
- [5] J.A. Souza-Neto, D.S. Gusmao and F.J.A. Lemos, Chitinolytic activities in the gut of *Aedes aegypti* (Diptera: Culicidae)

- larvae and their role in digestion of chitin-rich structures, *Comparative Biochemistry and Physiology Part A* **136** (2003), 717–724.
- [6] M.J. Edwards and M. Jacobs-Lorena, Permeability and disruption of the peritrophic matrix and caecal membrane from *Aedes aegypti* and *Anopheles gambiae* mosquito larvae, *Journal of Insect Physiology* **9** (2000), 1313–1320.
- [7] W.R. Terra, The origin and functions of the insect peritrophic membrane and peritrophic gel, *Archives of Insect Biochemistry and Physiology* **47** (2001), 47–61.
- [8] S. Li, M. Devenport and M. Jacobs-Lorena, The peritrophic matrix of hematophagous insects, *Archives of Insect Biochemistry and Physiology* **47** (2001), 119–125.
- [9] A. Marchler-Bauer, C.J. Zheng, F. Chitsaz, M.K. Derbyshire, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, S.N. Lu, G. H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, D.C. Zhang and S.H. Bryant, CDD: Conserved domains and protein three-dimensional structure, *Nucleic Acids Research* **41** (2013), D348–D352.
- [10] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel and A. Bairoch, Protein identification and analysis tools, on the ExPASy Server, in: *The Proteomics Protocols Handbook*, J.M. Walker, ed., Humana Press Inc., Totowa, NJ, 2005, pp. 571–602.
- [11] T.N. Petersen, S. Brunak, G. von Heijne and H. Nielsen, SignalP4.0: Discriminating signal peptides from transmembrane regions, *Nature Methods* **8** (2011), 785–786.
- [12] N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, L.J. Foster and F.S.L. Brinkman, PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, *Bioinformatics* **26** (2010), 1608–1615.
- [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] N. Guex, M.C. Peitsch and T. Schwede, Automated comparative protein structure modeling with Swiss-model and Swiss-Pdbviewer: A historical perspective, *Electrophoresis* **30** (2009), S162–S173.
- [16] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, L. Bordoli and T.N. Schwede, Swiss-model: Modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Research* **42** (2014), W252–W258.
- [17] G. Vaaje-Kolstad, S.J. Horn, D.M.F. van Aalten, B. Synstad and V.G.H. Eijsink, The Non-catalytic Chitin-binding Protein CBP21 from *Serratia marcescens* is essential for chitin degradation, *The Journal of Biological Chemistry* **280** (2005), 28492–28497.
- [18] U. Sehar, M.A. Mehmood, S. Nawaz, S. Nadeem, K. Hussain, I. Sohail, M.R. Tabassum, S.S. Gill and A. Saqib, Three dimensional (3D) structure prediction and substrate-protein interaction study of the chitin binding protein CBP24 from *B. thuringiensis*, *Bioinformatics* **9** (2013), 725–729.
- [19] U. Sehar, M.A. Mehmood, K. Hussain, S. Nawaz, S. Nadeem, M.H. Siddique, H. Nadeem, M. Gull, N. Ahmad, I. Sohail, S.S. Gill and S. Majeed, Domain wise docking analyses of the modular chitin binding protein CBP50 from *Bacillus thuringiensis* serovar konkukian S4, *Bioinformatics* **9** (2013), 901–907.
- [20] D. Liu, J. Cai, C.C. Xie, C. Liu and Y.H. Chen, Purification and partial characterization of a 36-kDa chitinase from *Bacillus thuringiensis* subsp. *colmeri*, and its biocontrol potential, *Enzyme and Microbial Technology* **46** (2010), 252–256.
- [21] M.A. Mehmood, X. Xiao, F.Y. Hafeez, Y.B. Gai and F.P. Wang, Molecular characterization of the modular chitin binding protein Cbp50 from *Bacillus thuringiensis* serovar *konkukian*, *Antonie van Leeuwenhoek* **100** (2011), 445–453.
- [22] N. Arora, B. Sachdev, R. Gupta, Y. Vimala and R.K. Bhatnagar, Characterization of a chitin-binding protein from *Bacillus thuringiensis* HD-1, *PloS One* **8** (2013), e66603.