In silico Studies of Parasporin Proteins: Structural and Functional Insights and Proposed Cancer Cell Killing Mechanism for Parasporin 5 and 6

*Nasima Aktar[a,b], Muhammad Manjurul Karim[a], Shakila Nargis Khan[a], Mustafizur Rahman[c], Anowara Begum[a], Md. Mozammel Hoq[a]**

Abstract

Background: Cancer is the leading cause of death in the world and the new types of cancer are diagnosed regularly but the advancement in their treatment is relatively slow and not to mention very costly. Parasporins (PS), parasporal inclusion proteins from *Bacillus thuringiensis*, possess specific cytotoxicity against different cancer cells which has suggested them to be potential for cancer treatment due to their specific binding to cancer cells.

Methods: Computational investigation were performed to exploit their physicochemical characteristics, structural properties including three dimensional (3D) model, model quality analysis, phylogenetic assessment and functional analysis along with the cancer-cell killing mechanism of PS-5 and PS-6 proteins using standard tools of bioinformatics.

Results: PS proteins were found to be slightly acidic based on their isoelectric points i.e., pI ranging from 5.12-6.19, and the instability indices (29.03- 42.31) indicate their highly stable nature in test tubes and higher aliphatic indices (62.54-94.75) indicate their thermostability, a feature suitable for high-level industrial production. In silico analysis of cellular localization predicts that the parasporins are mostly located in the cytoplasm and few in the plasma membrane but devoid of any signal peptide. The generated 3D models of PS proteins upon verification by Ramachandran plot analysis confirmed that our prediction lies in the good quality model range and facilitated the understanding of the very protein folding, assembly into complexes and cell killing mechanisms. It could be hypothesized that the PS-5 protein might induce apoptosis or act as β- pore forming toxin to kill specific cancer cells while PS-6 might act simply as pore forming toxin.

Conclusions: The theoretical overview of this research would facilitate the researchers with valuable insights of the PS protein structures, cancer cell killing mechanism of PS-5 and PS-6 proteins eventually in tumor micro-environment and their receptor molecules with a view to develop anti-cancer drugs.

Keywords: Parasporin, Domain, Motif, Cytotoxicity, Molecular docking.

Abbreviations: Bt, *Bacillus thuringiensis*; PS, Parasporin; β-PFT, β- Pore Forming Toxin; CD, Circular Dichroism; Cry, Crystal; pI, Isoelectric point; AI, Aliphatic Index; GRAVY, Grand average of hydropathy; PI, Phosphatidyl Inositol; GPI, Glycosyl Phosphatidyl Inositol; GOL, Glycerol; 13D, 1, 3-Diaminopropane; Br, Bromide; NGA, N- Acetyle- D- Galactosamine; U1, Uracil; E64, N- [N- [1- Hydroxycarboxyethyl- Carbonyl] Leucylamino- Butyl]- Guanidine; HEA, Hydroxyethylamine; MN, Manganese (II) ion.

Introduction

Bacillus thuringiensis is a gram-positive, spore-forming bacterium that produces large crystalline parasporal inclusion protein (δ-endotoxin) with specific cytotoxicity against agriculturally and medic-

Significance | **Structural and functional insights of anti-cancer proteins from *Bacillus thuringiensis*.**

*Correspondence: **Md. Mozammel Hoq**, PhD, Professor, Department of Microbiology, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: bahareh.nowruzi@srbiau.ac.ir; Contact no.: 00989113710956

Edited by **El-Sayed A. El-Sheikh, PhD, Zagazig University,** Zagazig, Egypt, and accepted by the Editorial Board February 21, 2019 (Received for review April 6, 2018)

Author Affiliation:
a. Department of Microbiology, University of Dhaka, Dhaka- 1000, Bangladesh
b. Basic and Applied Research on Jute, Dhaka- 1207, Bangladesh
c. Infectious Diseases Division and Virology Laboratory, ICDDR'B, Dhaka- 1212, Bangladesh.

Please cite this article:
Aktar N, Karim MM, Khan SN, Rahman M, Begum A, Hoq MM (2019). *In silico* Studies of Parasporin Proteins: Structural and Functional Insights and Proposed Cancer Cell Killing Mechanism for Parasporin 5 and 6. *Microbial Bioactives, 2*(1), 082-090.

ally important insects of several orders including Lepidoptera, Diptera, and Coleoptera etc. (Beegle & Yamamoto, 1992). Earlier, the isolation of insecticidal *B. thuringiensis* strains from Bangladesh harboring crystal (Cry1A) protein was reported which demonstrated high level of toxicity against Melon fruit fly, *Bactocera cucurbitae* from Diptera order (A Shishir et al., 2015; Asaduzzaman Shishir et al., 2012, 2014; M. A. Shishir et al., 2012; M. A. Shishir, Akter, et al., 2015; M. A. Shishir, Pervin, Sultana, Khan, & Hoq, 2015). The presence of non-insecticidal Bt strains is as abundant as their insecticidal counterpart in nature nevertheless very few extensive studies were attempted (Hastowo, Lay, & Ohba, 1992; Martin & Travers, 1989; Meadows, Ellis, Butt, Jarrett, & Burges, 1992; Mizuki et al., 1999; M. Ohba, 1996; Michio Ohba & Aizawa, 1986; Roh et al., 1996). Interestingly, the parasporal inclusions proteins from certain non-insecticidal Bt strains had been reported to exert preferential cytocidal activity against human cancer cells and thus making them potential candidates in cancer treatment (Ferdous, Shishir, Khan, & Hoq, 2018; Mizuki et al., 1998, 1999; Okassov, Nersesyan, Kitada, & Ilin, 2015).

To date, 19 parasporin proteins are grouped into 6 classes under two main categories i.e. three-domain type protein (PS 1, 3 and 6) and non three-domain type protein (PS 2, 4 and 5) according to the sequence homology as described by the Committee of Parasporin Classification and Nomenclature, 2006 (Adang, Crickmore, & Jurat-Fuentes, 2014; Ferdous et al., 2018; S. Okumura et al., 2010).

Parasporin proteins are primarily produced as inactive precursor proteins which are eventually activated upon proteolytic digestion. N-terminal digestion occurs in PS-1 and PS-6 belonging to three-domain classes, C- terminal digestion occurs in PS-4 and PS-5 of non three-domain classes and digestions at both termini occur in PS-2 and PS-3 (Michio Ohba, Mizuki, & Uemori, 2009). PS-1 is not a pore forming toxin. Cancer cell killing mechanism of PS1Aa1 starts by rapid increase of the intracellular Ca^{2+} concentration and inducing apoptosis in HeLa cells through degradation of apoptosis-related proteins, procaspase-3 and poly ADP-ribose polymerase (Michio Ohba et al., 2009). PS2Aa1 exerts cytotoxicity by increasing the plasma membrane permeability of susceptible cancer cells through receptor proteins binding and forming oligomers in lipid rafts of plasma membranes leading to pore formation and cell lysis. The oligomerization occurs in the presence of membrane proteins, lipid bilayer and cholesterols. Substantial homology exists in amino acid sequences between PS2Aa1 and *Clostridium perfringens* epsilon toxin whose cell-killing mechanism involves the toxin oligomerization in lipid rafts and pore formation in plasma membrane (Michio Ohba et al., 2009). Likewise, PS3Aa1 also acts as a pore-forming toxin eventually causing increased membrane permeability (Yamashita et al., 2005). PS-4 shares homologies with Cry15Aa, α-toxin, aerolysin, and ε-toxin and a circular dichroism (CD) spectrum of PS-4 contains 51% of β-structure and is a unique cholesterol-independent β-PFT (Shiro Okumura, Saitoh, Ishikawa, Inouye, & Mizuki, 2011). PS-5 and PS-6 are two newly discovered proteins and their mode of action is yet to be described completely (Ekino et al., 2014; Ferdous et al., 2018; Nagamatsu, Okamura, Saitou, Akao, & Mizuki, 2010; Shiro Okumura et al., 2011).

In order to wield Parasporin proteins as anti-cancer drug, translating the molecular mechanism underlying the specificity of ligand- receptor binding is very important.

Here, we have analyzed the physico-chemical parameters, functional domains, 3D structures and phylogenetic relationship of Parasporin proteins. The analyses also addressed the active sites of PS proteins as well as its receptor molecules on the susceptible cancer cells and the cancer cell killing mechanisms for PS-5 and PS-6 protein were hypothesized.

Materials and methods

Retrieval of sequences and homology tree construction

Parasporin protein sequences of *B. thuringiensis* were retrieved in FASTA format from NCBI. The sequences were subjected for 'Multiple sequence alignment' using clustal omega.

Bioinformatic analyses of parasporin

Physico-chemical properties of parasporin were computed using ExPASy's ProtParam tool (Wilkins et al., 1999). Cellular localization and solubility of parasporin proteins were determined using PSORT and SOSUI respectively (Hirokawa, Boon-Chieng, & Mitaku, 1998; Mitaku & Hirokawa, 1999; Mitaku, Hirokawa, & Tsuji, 2002; Nakai & Horton, 1999). GPI (Glycosylphosphatidyl Inositol) anchoring possibility was examined through big PI (Phosphatidyl Inositol) which also predicted potential sites for PI (Eisenhaber, Bork, & Eisenhaber, 1998, 1999; Eisenhaber, Bork, Yuan, Löffler, & Eisenhaber, 2000; Sunyaev et al., 1999). Signal peptide of parasporin protein was predicted by SignalIP-4.1 Server for Biological Sequence Analysis (CBS) platform (Petersen, Brunak, von Heijne, & Nielsen, 2011). Transmembrane region and orientation of the transmembrane region was detected using Tmpred.

3D structure Elucidation

The 3D structures of parasporin proteins were predicted using protein structure homology-modelling server, SWISS MODEL, which models oligomeric structures of target proteins including evolutionary conserved ligands (essential cofactors or metal ions) along with model-quality estimates based on a QMEAN potential and GMQE values, specifically re-parameterized for models built by SWISS-MODEL. Target-template alignment (sequence identity, sequence similarity, HHbits score, agreement between predicted secondary structure of target and template, agreement between predicted solvent accessibility between target and template; all normalized by alignment length) was used to predict the expected quality of most suitable structure (Arnold, Bordoli, Kopp, & Schwede, 2006; Biasini et al., 2014; Guex, Peitsch, & Schwede, 2009; Kiefer, Arnold, Kunzli, Bordoli, & Schwede, 2009). We have also used Phyre2 for 3D structure determination. Ramachandran plot by PDBsum was used for quality check of 3D structure.

Domain Analysis

InterPro and Pfam were used for domain analysis (Finn et al., 2016; Mitchell et al., 2015) while specific domain analysis was performed using Clustal omega to compare relatedness among different classes of parasporin (Li et al., 2015; McWilliam et al., 2013; Sievers et al., 2014).

Phylogenetic Analysis

Here, phylogenetic analysis was performed using domain sequences of all parasporin proteins. Domain sequences were retrieved from motif search and those sequences were converted to nucleotide sequences using Emboss transeq (Emboss; EMBL-EBI bioinformatics. 2013) then phylogenetic tree was constructed using MEGA 7 (Kumar, Stecher, & Tamura, 2016).

Binding site Detection

Binding sites for parasporin protein was calculated using RaptorX along with different confidence score (Score, P-value, Score and uGDT (GDT) and Pocket Multiplicity for potential target molecule prediction (Källberg et al., 2012). Moreover, binding sites were also identified from PDB database.

Deducing Molecular Docking

Ligand-receptor binding model for parasporin protein and receptor molecules of the specific cancer cells were deduced by molecular docking using ZDOCK server (Eisenhaber et al., 2000). We have also performed ligand-receptor binding for parasporin protein using PatchDock server and viewed the model using FireDock server.

Results

Retrieval and analysis of parasporin protein sequences

Anti-cancer drug design requires full understanding of the parasporin protein, their chemical composition, chemical properties, structure and conformation analysis. Amino acid analysis revealed that Parasporin 1 was rich in serine and leucine, whereas Parasporin 2 and 5 were enriched with threonine and serine, parasporin 3 was dominated by asparagine and leucine. Threonine and glycine were predominant in parasporin 4 while asparagine and isoleucine were dominant in parasporin 6 (Fig. 1). The relative abundance of amino acid residues helped to figure out Aliphatic index (Ikai, 1980), indicating the thermostability of the proteins. A highest AI to class 6 PS proteins (94.79) ranked them as most thermo stable, followed by class 1 (79.34-82.64) and 4 (81.16), while classes 3 (75.16-76.22), 5 (72.85) and 2 (62.54- 75.00) were less stable, due to the presence of Aspartic acid, Serine and Threonine.

Isoelectric point of parasporin proteins lie in the acidic range (5.12-6.19). Instability index calculates the stability of the protein in a test tube and except PS1Ad1 all parasporin proteins are stable with value less than 40 (Guruprasad, Reddy, & Pandit, 1990). The GRAVY value for a protein or a peptide is calculated by adding the hydropathy values of each amino acid residues and dividing by the number of residues in the sequence or length of the sequence where increasing positive score indicates a greater hydrophobicity (Kyte & Doolittle, 1982). Negative values of Grand Average of Hydropathicity Index [GRAVY] corroborates the hydrophilic nature of parasporin proteins where PS3Ab1 (-0.508) and PS4Aa1 (-0.171) were most and least hydrophilic respectively (Table 1).

In silico analysis revealed that parasporin proteins are cytoplasmic except two of PS-3 and four of PS-1 classes are localized in cytoplasmic membrane. Negative average hydrophobicity of parasporin proteins determined by SOSUI indicates that these proteins are soluble in aqueous solution and none of them have any PI (Phosphatidyl Inositol) sites which indicates that none of the parasporin proteins are GPI (Glycosyl Phosphatidyl Inositol) anchored proteins and are not directly accessible from the cytosolic face of the membrane. SignalP-4.1 web server discriminates signal peptides from non-signal peptides through D value and parasporin proteins with D-value less than 0.45 indicates no possession of signal peptide i.e. the non-secretory inclusion nature of parasporin protein. Moreover, three-domain proteins contain some potential transmembrane helices having TMPred value higher than 500 while non-three domain proteins possesses some insignificant transmembrane helices with less than 500 TMPred value (Table 1).

3D structure

Tertiary structure analysis revealed that all parasporin proteins had lower QMEAN4 values except PS2 (Fig. 2 and Sup. Table 2). The experimentally proved X-ray crystallography crystal structures of two members of this class are available in PDB database (PDB: 2ZTB &

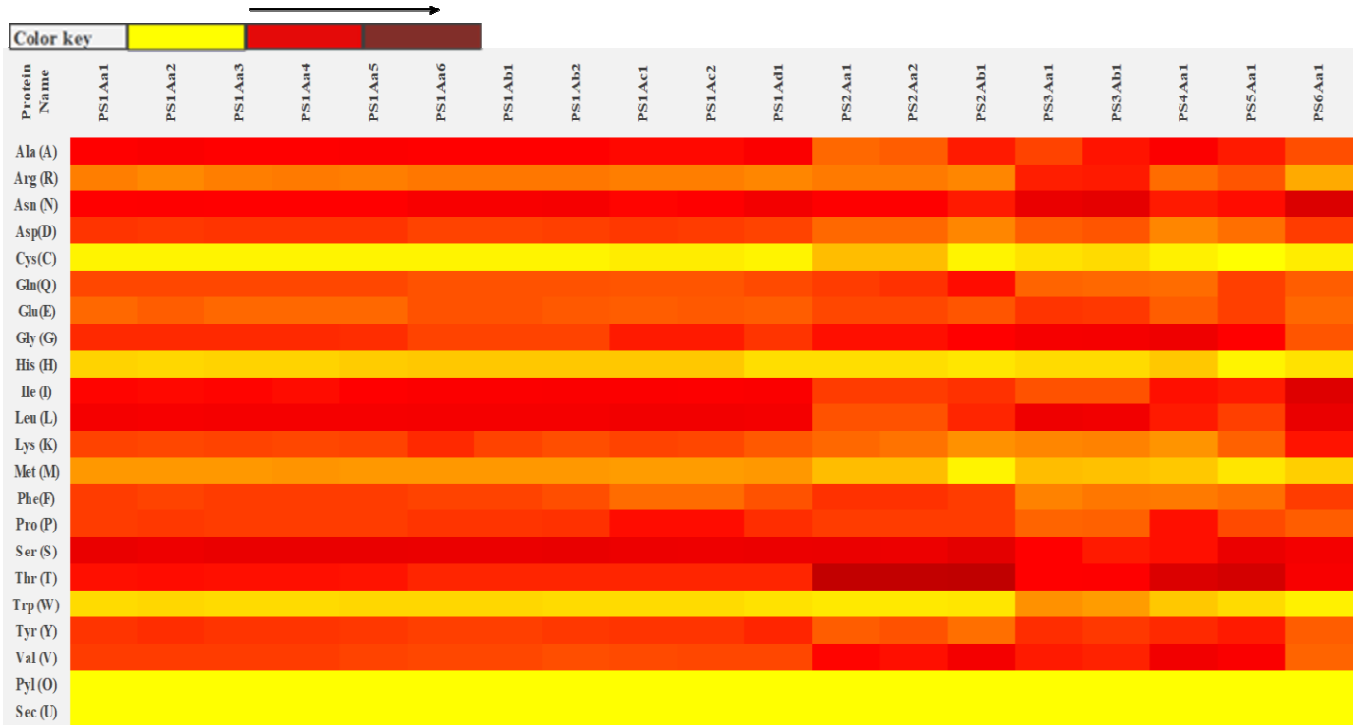

Figure 1| **Variability in amino acid richness in different parasporin protein class. Yellow-low value, Red-medium value and Maroon-Highest value.**

2D42) (Poupon & Janin, 2010), which confirm their higher QMEAN4 value. These structures were used as reference standard to compare our in silico analyses. Given the fact that QMEAN4 value of -4.0 or lower indicates poor quality of the structure, our results indicate that classes 2, 4 and 5 have good quality scores, while classes 1, 3 and 6 have slightly higher negative value than -4.0 indicating poor quality structure.

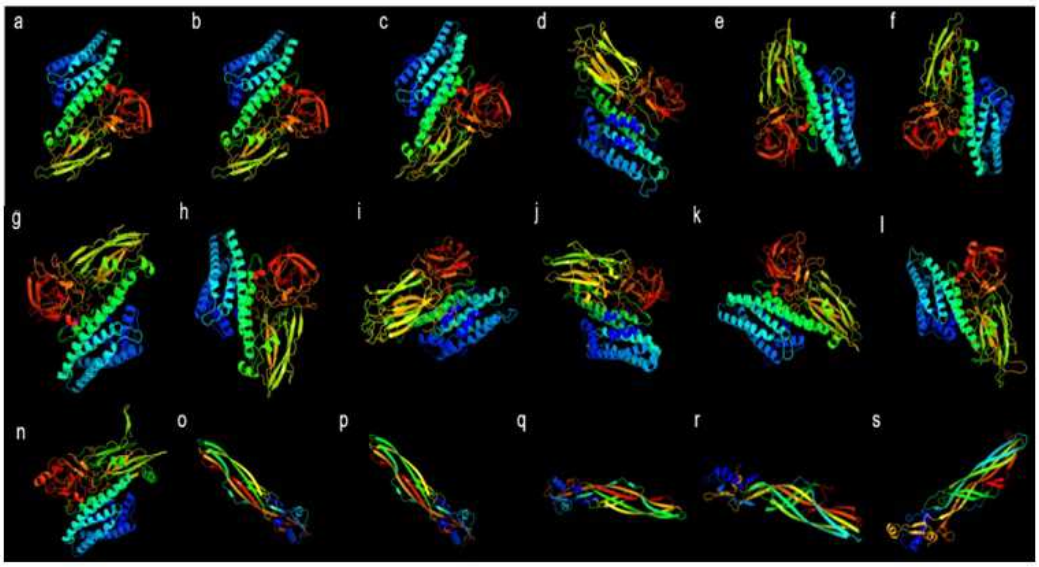

Figure 2| **3-D structure of parasporin protein; Three-domain parasporin (a) PS1Aa1 (b) PS1Aa2 (c) PS1Aa3 (d) PS1Aa4 (e) PS1Aa5 (f) PS1Aa6 (g) PS1Ab1 (h) PS1Ab2 (i) PS1Ac1 (j) PS1Ac2 (k) PS1Ad1 (l) PS3Aa1 (m) PS3Ab1 (n) PS6Aa1; Non-three-domain parasporin (o) PS2Aa1 (p) PS2Aa2 (q) PS2Ab1 (r) PS4Aa1 (s) PS5Aa1.**

However, all the predicted structures possessed higher GMQE score ranging 0.41-0.69 indicating the good quality of modeled structure. The phi-psi torsional angles analysis through Ramachandran plot revealed that the structure of PS2Aa2, PS4Aa1 and PS5Aa1 are more stable having more than 90% residues in most favored region. Rest of the structures of parasporins was also stable, because nearly 90% residue resided in most favored and additional allowed region (Fig. 3) (Sup. Table 3).

Domain Analysis of parasporin protein

Domain analysis with InterPro and pfam showed three-domain parasporin except PS6Aa1 contain Endotoxin_N and Endotoxin_C domains. The Galactose-binding domain in parasporin protein (Kitada et al., 2005) showed that Glycosylphosphatidylinositol (GPI)-anchored proteins are involved in the cytocidal actions and glycosylation alteration in the glycolipids are the basis for cancer cell formation (Daniotti, Vilcaes, Torres Demichelis, Ruggiero, & Rodriguez-Walker, 2013). Galactose-binding domain might play a role in specific binding with cancer cells. Moreover, parasporin 3 (PS3Aa1, PS3Ab1) contains some additional Ricin_B, lectin domain that binds with simple sugars (galactose or lactose) of cancer cells containing multi-antennary structures than normal cells (Raymond W. Ruddon, 2003). InterPro analysis revealed presence of Aerolysin like toxin, beta complex domain in non-three domain protein, while Pfam analysis showed *Clostridium* epsilon toxin ETX/*Bacillus* mosquitocidal toxin MTX2 but internal database links of Pfam showed similarity with Aerolysin (Sup. Table 4). These types of toxin bind to eukaryotic cells and aggregate to form holes in lipid bilayers leading to the destruction of membrane permeability and osmotic lysis (Aerolysin, IPR005830). Domain analysis of PS2Aa1, having two experimentally determined crystal structures (Akiba et al., 2009) showed similar domain finding compared to InterPro and Pfam (Wilkins et al., 1999).

Phylogenetic tree construction

Multiple sequence alignment shows that three domain and non-three domain parasporin proteins differ in amino acid sequences. Parasporin 1 and 6 shares higher sequence similarity. Non- three domain parasporin (PS 2, 4 and 5) shares sequence similarity among themselves. Percent Identity Matrix analysis reveals that parasporin proteins in each class share more than 80% sequence similarity and lesser sequence similarity among the classes but interestingly share similar toxicity towards similar cancer cell lines (Michio Ohba et al., 2009). Parasporin 6 shares slightly higher sequence similarity with class 1 compared to other classes and also shares higher toxicity to HeLa cell lines which is similar to Class 1. It is interesting that although being a three-domain protein, parasporin 3 shares sequence similarity with non-three domain parasporin 2, which is not reflective in Percent Identity Matrix analysis. Therefore, we performed multiple sequence alignment with specific domain sequence rather than the whole protein sequence and found that parasporin 1, 3 and 6 resides closely in the phylogenetic tree and they all showed toxicity towards HeLa cell containing the Endotoxin_N domain. Although being a non-three domain, parasporin 5 ETX/MTX2 domain resides with the Endotoxin_N domain, which explains their toxicity towards HeLa cells. However, ETX/MTX2 domain of class 2 resides distantly from class 5 and shows cytotoxicity towards HL-60 cells. Aerolysin domain of class 2 and 3 resides closely in the phylogenetic tree (Fig. 4) and have cytotoxicity towards HL-60 cells (Aldeewan, Zhang, & Su, 2014).

Prediction of binding sites

We have detected binding sites of parasporin protein but found only three pockets with more than 40 pocket multiplicity in parasporin PS3Aa1 and PS3Ab1. Binding site prediction with PDB revealed that all three-domain classes contain GOL (Glycerol) and 13D (1, 3- Diaminopropane) ligand. RaptorX too predicted Br^- (Bromide Ion), NGA (N-Acetyle-D-Galactosamine), SO_4^{2-} along with GOL and 13D. On the other hand, non-three domain class harbors U1 in addition SO_4^{2-}, E64 (N- [N- [1- Hydroxycarboxyethyl- Carbonyl] Leucylamino- Butyl]- Guanidine), HEA (Hydroxyethylamine), Mn^{2+} was found in PS2. In PDB database, PS2Aa1 (PDB ID: 2ZTB) has

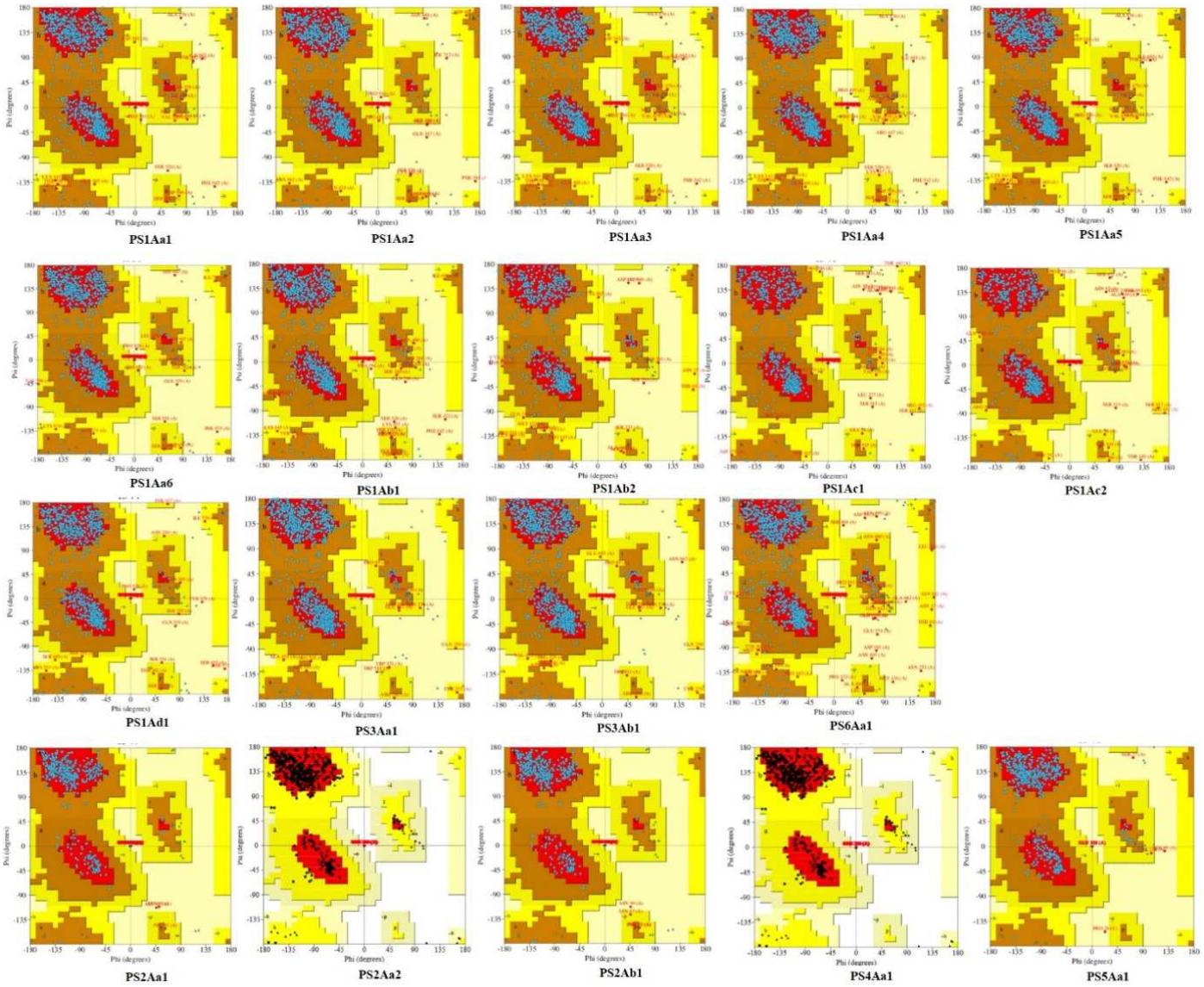

Figure 3| **Ramachandran plot analysis. PS 2, PS 4 and PS 5 class have more than 90% residue in the favored zone indicating their more structural stability. However, predicted structures of other classes are also stable as they have near 90% residue in the favored region.**

similar binding sites identified by RaptorX binding sites corroborating the correct binding site prediction by RaptorX (Sup. Table 5). Moreover, RaptorX predicted some additional binding sites because PDB showed ligands based on homologous searches of model and RaptorX predicted the ligands on the basis of amino acid sequences. However, two protein of Parasporin 1 class (PS1Ac1, Ps1AC2) also have SO_4^{2-} and Ca^{2+}- ligand which might help in dock in different receptors.

Prediction of Receptor on Cancer cells

We have taken the cancer cells against which each parasporin protein showed highest toxicity (Michio Ohba et al., 2009). PS 1, 5 and 6 showed higher toxicity towards HeLa cells. PS 2, PS3 and PS 4 showed higher toxicity towards HL-60 (Michio Ohba et al., 2009). Hence, we have taken their receptor molecule from Endogenous GPCRs in Common Cell Lines and their PDB ID and binding ligands were determined using PDB (Sup. Table 6).

Molecular Docking

We have predicted binding orientation of different parasporin protein to receptors of target cancer cells for understanding cancer cell killing mechanism. Endotoxin_N domain of parasporin 1 and 6 contains GOL, SO_4^{2-} ligand molecule and HeLa cell contains receptor molecule (Andregenic β2, Endothelin Type-B, Lysophospholipid) that tends to bind with this ligands. Enotoxin_N and ETX_MTX2 domain of Parasporin 2 and 3 contains ligand GOL and is cytotoxic towards HL-60 cells containing receptor molecule Chemokine CCR1. Parasporin 5 contains U1, PO_4 ligand molecule on ETX_MTX2 domain through which they bind with Histamine H1 and Muscarinic receptor of HeLa cell. Parasporin 4 have U1, GOL ligand molecule on ETX_MTX2 domain and shows cytotoxicity through binding with Somatostatin receptor of Caco-2 cells (Sup. Fig. 1) (Sup. Table 5 & 6).

Cancer Cell Killing Mechanism

Domain based phylogenetic analysis revealed that parasporin 1, 3 and 6 contains Endotoxin_N domain. However, PS-6 has greater sequence similarity with PS-1 and closely related in the phylogenetic tree, and were reported for their higher cytotoxicity against HeLa cells (Nagamatsu et al., 2010; Michio Ohba et al., 2009). Both of them have similar type of activated protein product after protease digestion (Nagamatsu et al., 2010; Michio Ohba et al., 2009) and posses similar

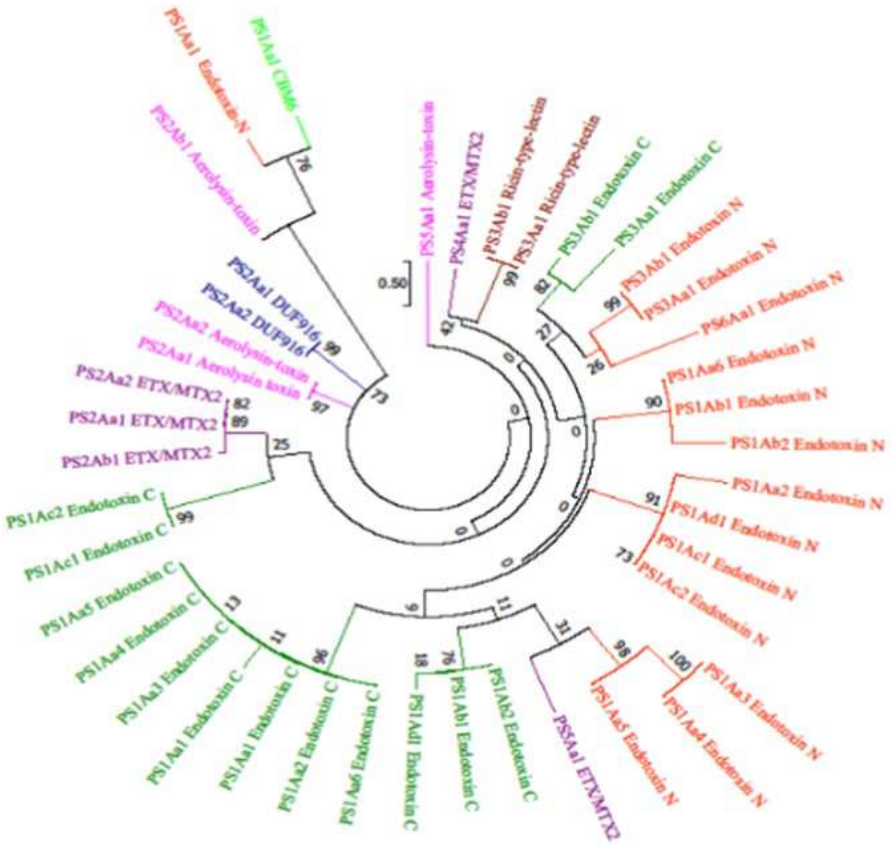

Figure 4| **Phylogenetic relationship among domains of different parasporin proteins by Neighbor-Joining method. Numbers at nodes represent the percentages of occurrence of nodes in 1000 bootstrap trials Three Domains parasporin proteins are clustered together and non-three domain proteins domains are clustered together. However, PS 5 ETX/MTX2 is clustered with three domain proteins, and PS 3 Ricin-type-lectin are grouped with PS 4 and 5.**

ligand molecule (GOL, 13D), and they might follow similar cancer cell killing mechanism. Briefly, upon activation, PS-6 binds to the receptor molecule and increases intracellular Ca^{2+}, protein and DNA synthesis followed by vacuole formation. As a result, cell swelling occurs leading to degradation of apotosis related protein (Procaspase-3 and Poly ADP-ribose polymerase) and apotosis of the cells.

three domain protein which lead to hypothesize that PS-5 might be a β-PFT protein. Upon activation, they bind to specific receptor molecule and oligomerize to form pore and increase cell permeability and lyse the cell (Fig. 5).

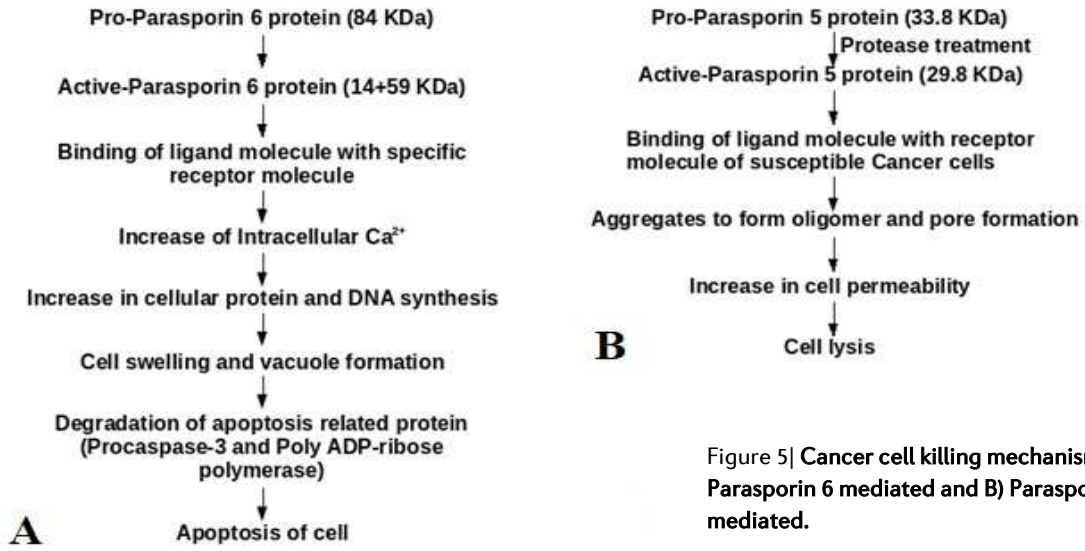

Figure 5| **Cancer cell killing mechanism. A) Parasporin 6 mediated and B) Parasporin 5 mediated.**

Unlike other small parasporin i.e. PS-2 and PS-4, PS-5 shows cytotoxicity towards HeLa cells (Ekino et al., 2014). However, they have highest sequence similarity with PS-4 protein and likely cytotoxic towards MOLT-4, TCS, HL-60 and Swano cells. Again, they are non-

Discussion

Cancer with leading cause of deaths remains the matter of health concern for both developed and developing countries (WHO media center, 2017). Despite the advances in cancer treatments, mortality

rate due to this deadly disease is still very high. Owing to the development of resistance by cancer cells towards current anti-cancer chemotherapeutic drugs, there is an urgent need to add new weapons in the anti-cancer drug arsenal to fight with this deadly disease. Parasporal inclusion proteins from *B. thuringiensis* having specific cytotoxicity against specific cancer cells together with their low industrial production cost could shed lights of hope. The thermostable nature of PS proteins in aqueous solution would make them suitable for industrial production. Variability in amino acid richness was found among different classes of parasporin protein, Parasporin 1 and 3 showing richness in leucine while Parasporin 4 and 6 in glycine and iso-leucine respectively can be categorized as potential anticancer protein as anticancer protein tend to be rich in these amino acids (Tyagi et al., 2013). We have predicted quality 3D structure of the parasporin proteins and validated with Ramachandran plot analysis to buttress structural quality which is important to determine the function of the protein. We have predicted their binding sites and also detected the receptor molecules to which they are likely to bind. On the basis of literature review and our analysis we have proposed cell killing mechanism of parasporin 5 and 6 protein which needs further validation from experimental evidence.

Conclusion

Although parasporin protein has specific cytotoxicity towards different cancer cell lines they can not be used directly for treatments because of the unpredicted immunological reactions to the patients. This necessitates detailed study of molecular structure–function relationships involving the unique action of parasporin. Here, we have performed in silico analyses to elucidate the composition, structural orientation of parasporin protein along with interaction with cancer cells. The functionally modular domain structure of the toxin is obviously advantageous for such a purpose. But The safety assessment of these proteins needs to be performed on a greater variety of non-target species to experimentally demonstrate their specificity and safety.

Author Contributions

NA, MMK and MMH conceived the idea; NA performed the analysis; NA, MMH, MMK, SNK, MR and AB prepared and revised the manuscript. All authors approved the final paper.

Acknowledgment

We thank Md. Sabbir Alam for his thorough and critical review of the manuscript.

Competing financial interests

Authors disclose no potential conflicts of interest.

References

Adang, M. J., Crickmore, N., & Jurat-Fuentes, J. L. (2014). Diversity of Bacillus thuringiensis Crystal Toxins and Mechanism of Action. In T. S. Dhadialla & S. S. Gill (Eds.), *Advances in Insect Physiology* (Insect Mid, Vol. 47, pp. 39–87). Oxford: Academic Press: Elsevier Ltd. https://doi.org/10.1016/B978-0-12-800197-4.00002-6

Akiba, T., Abe, Y., Kitada, S., Kusaka, Y., Ito, A., Ichimatsu, T., … Harata, K. (2009). Crystal structure of the parasporin-2 Bacillus thuringiensis toxin that recognizes cancer cells. *Journal of Molecular Biology*, *386*(1), 121–33. https://doi.org/10.1016/j.jmb.2008.12.002

Aldeewan, A., Zhang, Y., & Su, L. (2014). Bacillus thuringiensis Parasporins Functions on Cancer Cells. *International Journal Of Pure & Applied Bioscience*, *2*(4), 67–74.

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)*, *22*(2), 195–201. https://doi.org/10.1093/bioinformatics/bti770

Beegle, C. C., & Yamamoto, T. (1992). History of Bacillus thuringiensis Berliner Research and Development. *The Canadian Entomologist*, *124*(04), 587–616. https://doi.org/10.4039/Ent124587-4

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., … Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(Web Server issue), W252-8. https://doi.org/10.1093/nar/gku340

Daniotti, J. L., Vilcaes, A. A., Torres Demichelis, V., Ruggiero, F. M., & Rodriguez-Walker, M. (2013). Glycosylation of Glycolipids in Cancer: Basis for Development of Novel Therapeutic Approaches. *Frontiers in Oncology*, *3*, 306. https://doi.org/10.3389/fonc.2013.00306

Eisenhaber, B., Bork, P., & Eisenhaber, F. (1998). Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein Engineering Design and Selection*, *11*(12), 1155–1161. https://doi.org/10.1093/protein/11.12.1155

Eisenhaber, B., Bork, P., & Eisenhaber, F. (1999). Prediction of potential GPI-modification sites in proprotein sequences. *Journal of Molecular Biology*, *292*(3), 741–58. https://doi.org/10.1006/jmbi.1999.3069

Eisenhaber, B., Bork, P., Yuan, Y., Löffler, G., & Eisenhaber, F. (2000). Automated annotation of GPI anchor sites: case study C. elegans. *Trends in Biochemical Sciences*, *25*(7), 340–1. https://doi.org/10.1016/S0968-0004(00)01601-7

Ekino, K., Okumura, S., Ishikawa, T., Kitada, S., Saitoh, H., Akao, T., … Mizuki, E. (2014). Cloning and characterization of a unique cytotoxic protein parasporin-5 produced by Bacillus thuringiensis a1100 strain. *Toxins*, *6*(6). https://doi.org/10.3390/toxins6061882

Ferdous, U. T., Shishir, M. A., Khan, S. N., & Hoq, M. M. (2018). Bacillus spp.: Attractive Sources of Anti-cancer and Anti-proliferative Biomolecules. *Microbial Bioactives*, *1*(1), E033–E045. https://doi.org/10.25163/microbbioacts.11005B0408130818

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279-85. https://doi.org/10.1093/nar/gkv1344

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*, *30 Suppl 1*(S1), S162-73. https://doi.org/10.1002/elps.200900140

Guruprasad, K., Reddy, B. V, & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering*, *4*(2), 155–61. https://doi.org/10.1088/1751-8113/44/8/085201

Hastowo, S., Lay, B. W., & Ohba, M. (1992). Naturally occurring Bacillus thuringiensis in Indonesia. *Journal of Applied Bacteriology*, *73*(2), 108–113. https://doi.org/10.1111/j.1365-2672.1992.tb01695.x

Hirokawa, T., Boon-Chieng, S., & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/14.4.378

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*, *88*(6), 1895–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7462208

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web

server. *Nature Protocols*, *7*(8), 1511–1522. https://doi.org/10.1038/nprot.2012.085

Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, *37*(Database), D387–D392. https://doi.org/10.1093/nar/gkn750

Kitada, S., Abe, Y., Ito, A., Kuge, O., Akao, T., & Mizuki, E. (2005). Molecular Identification and Cytocidal Action of Parasporin, a Protein Group of Novel Crystal Toxins Targeting Human Cancer Cells. *6th Pacific Rim Conference on the Biotechnology of Bacillus Thuringiensis and Its Environmental Impact*, 6–10.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–32. https://doi.org/10.1016/0022-2836(82)90515-0

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., … Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*, *43*(W1), W580–W584. https://doi.org/10.1093/nar/gkv279

Martin, P. A., & Travers, R. S. (1989). Worldwide Abundance and Distribution of Bacillus thuringiensis Isolates. *Applied and Environmental Microbiology*, *55*(10), 2437–42. Retrieved from http://aem.asm.org/content/55/10/2437

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., … Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, *41*(Web Server issue), W597-600. https://doi.org/10.1093/nar/gkt376

Meadows, M. P., Ellis, D. J., Butt, J., Jarrett, P., & Burges, H. D. (1992). Distribution, frequency, and diversity of Bacillus thuringiensis in an animal feed mill. *Applied and Environmental Microbiology*.

Mitaku, S., & Hirokawa, T. (1999). Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. *Protein Engineering*, *12*(11), 953–7. https://doi.org/10.1093/protein/12.11.953

Mitaku, S., Hirokawa, T., & Tsuji, T. (2002). Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. *Bioinformatics*, *18*(4), 608–616. https://doi.org/10.1093/bioinformatics/18.4.608

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., … Finn, R. D. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, *43*(Database issue), D213-21. https://doi.org/10.1093/nar/gku1243

Mizuki, E., Ohba, M., Akao, T., Yamashita, S., Saitoh, H., & Park, Y. S. (1999). Unique activity associated with non-insecticidal Bacillus thuringiensis parasporal inclusions: in vitro cell-killing action on human cancer cells. *Journal of Applied Microbiology*, *86*(3), 477–486. https://doi.org/10.1046/j.1365-2672.1999.00692.x

Mizuki, E., Ohba, M., Ichimatsu, T., Hwang, S. H., Higuchi, K., Saitoh, H., & Akao, T. (1998). Unique appendages associated with spores of Bacillus cereus isolates. *Journal of Basic Microbiology*, *38*(1), 33–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9542106

Nagamatsu, Y., Okamura, S., Saitou, H., Akao, T., & Mizuki, E. (2010). Three Cry toxins in two types from Bacillus thuringiensis strain M019 preferentially kill human hepatocyte cancer and uterus cervix cancer cells. *Bioscience, Biotechnology, and Biochemistry*, *74*(3), 494–8. https://doi.org/10.1271/bbb.90615

Nakai, K., & Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in Biochemical Sciences*, *24*(1), 34–35. https://doi.org/10.1016/S0968-0004(98)01336-X

Ohba, M. (1996). Bacillus thuringiensis populations naturally occurring on mulberry leaves: a possible source of the populations associated with silkworm-rearing insectaries. *Journal of Applied Bacteriology*, *80*(1), 56–64. https://doi.org/10.1111/j.1365-2672.1996.tb03190.x

Ohba, M., & Aizawa, K. (1986). Insect toxicity of Bacillus thuringiensis isolated from soils of Japan. *Journal of Invertebrate Pathology*. https://doi.org/10.1016/0022-2011(86)90158-8

Ohba, M., Mizuki, E., & Uemori, A. (2009). Parasporin, a new anticancer protein group from Bacillus thuringiensis. *Anticancer Research*, *29*(1), 427–433.

Okassov, A., Nersesyan, A., Kitada, S., & Ilin, A. (2015). Parasporins as new natural anticancer agents: a review. *Journal of Balkan Union of Oncology*, *20*(1), 5–16.

Okumura, S., Ohba, M., Mizuki, E., Crickmore, N., Côté, J.-C., Nagamatsu, Y., … Shin, T. (2010). Parasporin nomenclature. Retrieved December 10, 2017, from http://parasporin.fitc.pref.fukuoka.jp/

Okumura, S., Saitoh, H., Ishikawa, T., Inouye, K., & Mizuki, E. (2011). Mode of action of parasporin-4 , a cytocidal protein from Bacillus thuringiensis. *Biochimica et Biophysica Acta*, *1808*(6), 1476–1482. https://doi.org/10.1016/j.bbamem.2010.11.003

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–6. https://doi.org/10.1038/nmeth.1701

Poupon, A., & Janin, J. (2010). Analysis and Prediction of Protein Quaternary Structure. In *Methods in molecular biology* (Vol. 609, pp. 349–364). https://doi.org/10.1007/978-1-60327-241-4_20

Raymond W. Ruddon. (2003). What Makes a Cancer Cell a Cancer Cell? In D. W. Kufe, R. E. Pollock, R. R. Weichselbaum, R. C. Bast, T. S. Gansler, J. F. Holland, & E. Frei (Eds.), *Holland-Frei Cancer Medicine* (6th ed.). Hamilton (ON): BC Decker; 2003. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK12516/

Roh, J. Y., Park, H. W., Jin, B. R., Kim, H. S., Yu, Y. M., & Kang, S. K. (1996). Characterization of novel non-toxic Bacillus thuringiensis isolates from Korea. *Letters in Applied Microbiology*, *23*(4), 249–252.

Shishir, A., Akter, A., Hassan, M. H., Kibria, G., Ilias, M., Khan, S. N., & Hoq, M. M. (2012). Characterization of locally isolated Bacillus thuringiensis for the Development of Eco-friendly Biopesticides in Bangladesh. *Journal of Biopesticides*, *5*(Supplementary), 216–222.

Shishir, A., Bhowmik, A. A., Akanda, N. R., Al Mamun, A., Khan, S. N., & Hoq, M. M. (2015). Efficacy of Indigenous Bacillus thuringiensis Strains for Controlling Major Vegetable Pests in Bangladesh. *Egyptian Journal of Biological Pest Control*, *25*(3), 729–734. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=asx&AN=112957578&site=eds-live

Shishir, A., Roy, A., Islam, N., Rahman, A., Khan, S. N., & Hoq, M. M. (2014). Abundance and diversity of Bacillus thuringiensis in Bangladesh and their cry genes profile. *Frontiers in Environmental Science*, *2*(June), 20. https://doi.org/10.3389/fenvs.2014.00020

Shishir, M. A., Akter, A., Bodiuzzaman, M., Aktar, N., Rahman, M., Khan, S. A., … Hoq, M. M. (2012). Molecular characterization of indigenous Bacillus thuringiensis kurstaki isolates with effective pesticidal activity against Bactrocera cucurbitae from Bangladesh. In R. Kada, J. Sasaki, C. R. Ahsan, & M. L. Bari (Eds.), *Proceddings of 1st AFSA Conferences* (Vol. 1, pp. 143–148). Osaka, Japan: Asian Food Safety & Security Association (AFSSA). Retrieved from http://www.afsahome.org/wp-content/uploads/2017/05/Food_Safety_and_Food_Security_2012-00.pdf

Shishir, M. A., Akter, A., Bodiuzzaman, M., Hossain, M. A., Alam, M. M., Khan, S. A., … Hoq, M. M. (2015). Novel Toxicity of Bacillus thuringiensis Strains against the Melon Fruit Fly, Bactrocera cucurbitae (Diptera: Tephritidae). *Biocontrol Science*, *20*(2), 115–123. https://doi.org/10.4265/bio.20.115

Shishir, M. A., Pervin, S., Sultana, M., Khan, S. N., & Hoq, M. M. (2015). Genetic Diversity of Indigenous Bacillus thuringiensis Strains by RAPD-PCR

to Combat Pest Resistance. *Bt Research*, *6*(8), 1–16. https://doi.org/10.5376/bt.2015.06.0008

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539–539. https://doi.org/10.1038/msb.2011.75

Sunyaev, S. R., Eisenhaber, F., Rodchenkov, I. V., Eisenhaber, B., Tumanyan, V. G., & Kuznetsov, E. N. (1999). PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Engineering*, *12*(5), 387–94. https://doi.org/10.1093/protein/12.5.387

Tyagi, A., Kapoor, P., Kumar, R., Chaudhary, K., Gautam, A., & Raghava, G. P. S. (2013). In silico models for designing and discovering novel anticancer peptides. *Scientific Reports*, *3*(1), 2984. https://doi.org/10.1038/srep02984

WHO media center. (2017). Cancer fact sheet. Retrieved from http://www.who.int/mediacentre/factsheets/fs297/en/

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology (Clifton, N.J.)*, *112*, 531–52. https://doi.org/10.1088/1751-8113/44/8/085201

Yamashita, S., Katayama, H., Saitoh, H., Akao, T., Park, Y. S., Mizuki, E., … Ito, A. (2005). Typical three-domain cry proteins of Bacillus thuringiensis strain A1462 exhibit cytocidal activity on limited human cancer cells. *Journal of Biochemistry*, *138*(6), 663–672. https://doi.org/10.1093/jb/mvi177