In-silico prediction and modeling of ApxA exotoxins of *Actinobacillus pleuropneumoniae:* ApxIA, -IIA, -IIIA and -IVA

Myung Whan Oh [1], **Han Sang Yoo** Corresp. [1]

[1] Department of Infectious Diseases, College of Veterinary Medicine, Seoul National University, Korea

Corresponding Author: Han Sang Yoo
Email address: yoohs@snu.ac.kr

Background. *Actinobacillus pleuropneumonia* is a gram-negative bacterium that serves as the major etiological agent for porcine pleuropneumonia, one of critical diseases causing substantial socio-economic losses in swine rearing industry world-wide. Apx exotoxins are the members of RTX-toxin family secreted by gram-negative bacteria that facilitates Type-I Secretion System (T1SS), reported as the major virulence factor in the pathogenesis. This study was conducted to demonstrate successful pre-experimental approach by predicting the tertiary structures of ApxA exotoxins and further characterizing their structural and functional annotations via application of appropriate *in silico* methods.

Methods. The sequences of ApxA exotoxins were retrieved from the National Centre for Biotechnology Information database (NCBI) for bioinformatics analyses in this study. ApxA exotoxins were subjected to several computational analyses to characterize in different aspects. I-TASSER and Phyre2 servers were used to predict 3-D structures of ApxA toxins which were further validated by ProSA and SAVES servers. Antigenic epitopes of each ApxA exotoxin were predicted using the BepiPred 2.0 program.

Results. Based on predictions on secondary structure compositions, hydropathicity, domain boundaries, tertiary structures and their structural analogs, ApxIA, -IIA and –IIIA shared remarkable resemblance in structural and functional aspects while ApxIVA exhibited distinct and complex characteristics.

Discussion. 3-D modeling of ApxA exotoxins were executed to best interpret the proteins in their structure and functions to gain novel insights of their pathogenesis. The domain-wise interpretation of the proteins indicates the likely roles of each toxin with conjectured antigenic epitopes. The structural and functional annotations suggested in this study via *in silico* predictive approach may provide insights to prevent and control *A. pleuropneumoniae* by development of diagnostic methods and vaccine candidates.

In-silico prediction and modeling of ApxA exotoxins of *Actinobacillus pleuropneumoniae:* ApxIA, -IIA, -IIIA and –IVA

Myung Whan Oh[1] and Han Sang Yoo[1]

[1]Department of Infectious Diseases, College of Veterinary Medicine, Seoul National University, Seoul 08826, Korea

Corresponding Author:

Han Sang Yoo[1]

E-mail address: yoohs@snu.ac.kr

ABSTRACT

Background. *Actinobacillus pleuropneumonia* is a gram-negative bacterium that serves as the major etiological agent for porcine pleuropneumonia, one of critical diseases causing substantial socio-economic losses in swine rearing industry world-wide. Apx exotoxins are the members of RTX-toxin family secreted by gram-negative bacteria that facilitates Type-I Secretion System (T1SS), reported as the major virulence factor in the pathogenesis. This study was conducted to demonstrate successful pre-experimental approach by predicting the tertiary structures of ApxA exotoxins and further characterizing their structural and functional annotations via application of appropriate *in silico* methods.

Methods. The sequences of ApxA exotoxins were retrieved from the National Centre for Biotechnology Information database (NCBI) for bioinformatics analyses in this study. ApxA exotoxins were subjected to several computational analyses to characterize in different aspects. I-TASSER and Phyre2 servers were used to predict 3-D structures of ApxA toxins which were further validated by ProSA and SAVES servers. Antigenic epitopes of each ApxA exotoxin were predicted using the BepiPred 2.0 program.

Results. Based on predictions on secondary structure compositions, hydropathicity, domain boundaries, tertiary structures and their structural analogs, ApxIA, -IIA and –IIIA shared remarkable resemblance in structural and functional aspects while ApxIVA exhibited distinct and complex characteristics.

Discussion. 3-D modeling of ApxA exotoxins were executed to best interpret the proteins in their structure and functions to gain novel insights of their pathogenesis. The domain-wise interpretation of the proteins indicates the likely roles of each toxin with conjectured antigenic epitopes. The structural and functional annotations suggested in this study via *in silico* predictive

approach may provide insights to prevent and control *A. pleuropneumoniae* by development of diagnostic methods and vaccine candidates.

Keywords in silico, bioinformatics, *Actinobacillus pleuropneumoniae*, computational analysis, 3-D structure prediction, I-TASSER, RTX toxin, antigenic epitope

INTRODUCTION

A. pleuropneumoniae is a highly infectious and contagious disease that elicits various clinical degrees of pleuropneumoniae in swine, ranging from peracute to chronic where infected pigs typically display a hemorrhagic, necrotizing pneumonia associated with fibrinous pleuritis (*Shin et al, 2011*). The virulence factors of *A. pleuropneumoniae* involved in the pathogenesis include capsular polysaccharides, lipopolysaccharides, outer membrane proteins, adhesion factors, proteases, and exotoxins, giving *A. pleuropneumoniae* the multifactorial virulence in infection (*Wiles et al, 2013*). Despite the ambiguity in respective contributions by the listed virulence factors in infection, ApxA exotoxins have been reported to be strongly related to the pathogenesis of *A. pleuropneumoniae*, demonstrating loss of virulence by impairing the production or secretion of Apx exotoxins via mutation vice versa in pigs and mice (*Shin et al, 2011*; *Inzana & Fenwick, 2011; Tascon et al, 1994*). Also, the variations in virulence between the serotypes of *A. pleuropneumoniae* dictated by the Apx toxin profiles support the generally accepted claim that the Apx exotoxins play decisive roles in inducing protective immunity in APP infection (*Shin et al, 2011*).

There are four different Apx exotoxins secreted by *A. pleuropneumoniae*, ApxIA, -IIA, -IIIA and IVA, showing different degrees in hemolytic and cytotoxic activities against the phagocytic cells: ApxI is both strongly hemolytic and cytotoxic, ApxII is both weakly hemolytic and

cytotoxic, ApxIII is non-hemolytic but strongly cytotoxic (*Bouke K. H. L Boekema et al, 2004*; *Shin et al, 2011; Stanley, Koronakis & Hughes, 1998*). As a member of RTX-toxin family, Apx toxin is encoded by the *apx* operon comprised of four genes, *apxCABD*, where *apxC* codes for post-translational activation protein, *apxA* for structural toxin protein, and *apxBD* for secretion-related proteins (*Wiles et al, 2013*; *Frey et al, 1995*). The secreted ApxA exoproteins of RTX family exhibit pathogenic activities against the macrophages upon phagocytosis. ApxA exotoxins share the presence of RTX repeat blocks in their amino acid sequences, namely the Glycine-rich nonapeptide tandem array L/V-X-G-G-X-G-N/D-D-X, that play a major role in virulence (*Bumba et al, 2016*). The cytotoxic and hemolytic activities of RTX-toxins including the ApxA have been studied extensively with main focus on RTX determinant identifications.

Current diagnostic methods utilized by several commercially available kits apply a few strategies to diagnose *A. pleuropneumoniae* via different types of Enzyme-Linked Immunosorbent Assay (ELISA) technique. The LPS antigen based ELISA strategy attempts to differentially diagnose the serotypes of *A. pleuropneumoniae*, though facing the struggle to provide the specificity due to high cross-reactivity (*Inzana & Fenwick, 2011*). Other commercial kits use antigen candidates unspecified by the manufacturer as the diagnostic agent, which also struggle from low specificity. Moreover, these diagnostic kits fail to measure the antibody titers against each of ApxA exotoxins and fail to discriminate the naturally infected pigs from the vaccinated pool (*Shin et al, 2011*). Considering such weaknesses in current methods, development of new diagnostic strategies by identification of the novel antigenic epitopes of ApxA exotoxins is necessary to effectively diagnose and control APP infection (*Shin et al, 2011*).

Proteins serve as an integral element in biology, and rational understanding of proteomics rely heavily on the structural aspects compared to the comprising amino acid sequences. The

substantial discrepancy between the load of sequential data and rather scant structural data on database emphasizes the necessity of structural characterizations (*Petrey D & Honig,* 2005). Bioinformatics screening methods have been applied increasingly to report over a thousand of RTX family members. Most of their biological functions remain yet to be characterized aside from the few already discovered, which range from pore-forming RTX leukotoxins to multifunctional enzymatic toxins, bacteriocins, secreted hydrolytic enzymes exhibiting metalloprotease, lipase and more (*Linhartova, 2010)*. In the structural aspects of the Apx exotoxins, the parallel β-roll motif appears universally in all of the four apxA toxins, characterized by the well-established glycine- and aspartate- rich nonapeptide Ca^{2+}-binding turns, L/V-X-G-G-X-G-N/D-D-X (*Wiles et al, 2013; Linhartova, 2010; Shin et al, 2011)*. However, the knowledge of tertiary structures of ApxA proteins is far from sufficient to successfully elucidate the cellular and molecular ontology behind pathogenesis. Also, although the current reports have designated ApxA exotoxins as a pivotal virulence factor and decisive immune-determinant in *A. pleuropneumoniae* infection, the following pathogenic mechanism against the host immune cells is yet fully understood (*Wang et al 2009*). Thus, the structural characterization of Apx exotoxins by three-dimensional modeling may help remove this veil and assist in characterizing the attributes of Apx exotoxins in *A. pleuropneumoniae* pathogenesis.

MATERIALS AND METHODS

Sequence retrieval

The sequences of the four Apx structural toxins, ApxIA (AF363361.1), ApxIIA (AF363362.1), ApxIIIA (AF363363.1) and ApxIVA (HM021153.1), were selected as the representative sequences for the purpose of this study from the NCBI (National Center for Biotechnology

Information) nucleotide database in FASTA format.

Sequence alignments

Retrieved sequences of ApxIA, -IIA, -IIIA and IVA were submitted to PSI-Blast at https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=Proteins&PROGRAM=blastp&RUN_PSIBLAST=on against non-redundant protein sequences database to collect sequences with significant hits. Multiple sequence alignments were then generated by ClustalW via Bioedit sequence alignment editor software (*Antonis et al, 2003*).

Structural and functional analysis

Bioinformatics analyses were executed via online web services and softwares to execute physiochemical characterizations of the RTX toxins of *A. pleurponeumoniae*. Molecular weight, theoretical pI, amino acid and atomic composition, molecular formula, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) were predicted by ProtParam at http://web.expasy.org/protparam/, and Compute pI/Mw at http://web.expasy.org/compute_pi/. Kyte and Doolittle suggests that the transmembrane proteins spanning the membrane has the tendency to show higher GRAVY scores than globular proteins (*Seddigh & Darabi, 2017*). Further, the hydropathy plot of each apx structural toxins was constructed by using the ProtScale server at http://web.expasy.org/protscale/, which validated the hydropathic characteristics of toxins (*Seddigh & Darabi, 2017; Rahbar et al, 2012*).

Motif and domain annotation

Several web-based computational tools were used in attempt to obtain and analyze the conserved

motifs and domains of apx structural exotoxins. Motif Search tool of Genomenet server at http://www.genome.jp/tools/motif, ProDom at http://prodom.prabi.fr/prodom/current/html/home.php, Dompred at http://bioinf4.cs.ucl.ac.uk:3000/dompred (*Marsden et al.,2002*) were used. MOTIF search tool of GenomeNet server is an in-house-developed search system that allows sequence motif search against major motif libraries such as Pfam, NCBI-CDD, and PROSITE (*Kanehisa et al, 2016*). Motif and domain search by GenomeNet specified position, score and E-value of each domain. Putative domain boundaries were predicted by the Dompred server which employs an algorithm to identify the locations of amino acid residues to which the N- and C-termini matching database hits are aligned to the sequences. (*Rahbar et al, 2012*).

Secondary structure prediction

The secondary structures constituting the apxA exotoxins were calculated by SOPMA server at http://npsa-pbil.ilbcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html, and the color coded graph describing the constitution of 4 secondary structures, α-helices, extended strand, β-turn and random coils, for each apxA toxin was retrieved (*Geourjon C, Deleage G, 1995*). Furthermore, Psipred software at http://bioinf.cs.ucl.ac.uk/psipred/ was used to supplement the secondary structures of Apx exotoxins. The computed secondary structure predictions can also be referred to validate the predicted three-dimensional models of each exotoxin.

Topology prediction

TMHMM server at http://www.cbs.dtu.dk/services/TMHMM-2.0/ was also used to confirm any transmembrane helices within the Apx structural toxins (*Geourjon C, Deleage G, 1995*). ApxA sequences were submitted to support vector machine SRTpred program for classification of the

exotoxins as secretory or non-secretory at http://crdd.osdd.net/raghava/srtpred/ (*Garg, Raghava, 2008*). Subcellular localizations of the ApxA toxins were predicted by PSLpred (*Bhasin et al., 2005*) and CELLO (*Yu et al., 2006*) at http://crdd.osdd.net/raghava//pslpred/ and http://cello.life.nctu.edu.tw/.

Three-dimensional modeling

Two major modelling programs were used in attempt to predict the three-dimensional structures of the four ApxA structural exotoxins, I-TASSER and Phyre2. I-TASSER server at https://zhanglab.ccmb.med.umich.edu/I-TASSER/predicts and provides predicted protein structures and functions via hierarchical approach, utilizing both threading and *ab initio* methods (*Roy A, Kucukural A & Zhang Y, 2010*). The 3D-structure modelling by I-TASSER bases on the *ab initio* and multiple-threading alignments. I-TASSER incorporates SPICKER strategy to identify the near-native folds with low free-energy states via clustering the simulation decoys (*Roy A, Kucukural A & Zhang Y, 2010*). The individual domains within each of ApxA exotoxin parsed based on the Dompred server were submitted to I-TASSER program for in-depth analysis.

Validation

The predicted three-dimensional structures were then submitted to PROCHECK (*Laskowski, MacArthur, Moss, & Thornton, 1993*) and ProSA-web (*Wiederstein & Sippl, 2007*) servers to validate the qualities of each structural outputs.

Ligand binding sites

The I-TASSER software simultaneously defined the binding sites of the structural analogs to the

predicted model. Along with this integrated result, 3DligandSite at http:/ http://www.sbg.bio.ic.ac.uk/3dligandsite/ was used to predict the binding sites on resulting hypothetical models of Apx exotoxins (*Roy A, Kucukural A & Zhang Y, 2010*).

B-cell epitope prediction

Linear B-cell epitopes in each of Apx exotoxin were predicted using the Bepipred which combines the hidden Markov model with two best propensity scale methods (*Levitt, 1978; Parker et al., 1986*). Discontinuous epitopes of Apx toxins were predicted using the Ellipro server at http://tools.iedb.org/ellipro/ by incorporating the predicted 3D models submitted in PDB files (*Rahbar et al, 2010*).

RESULTS

Primary analysis

The theoretical pI values were computed to be ranging from 4.75 to 5.88, and the instability index ranged from 18.60 to 25.74 (Table 1). The GRAVY scores of ApxIA, -IIA, -IIIA and –IVA were -0.352, -0.365, -0.370 and -0.615, respectively (Table 1). The instability indices for four of the ApxA exotoxins were 25.74, 25.58, 18.60, 20.98, and aliphatic indices were 88.63, 91.20, 87.46, 76.59 in order of –IA to –IVA (Table 1). Though the GRAVY scores tend to deliver some of information on structural ques of proteins of interest, hydropathicity plot by the Kyte & Doolittle scale was constructed by the ProtScale server for comprehensive analysis (*Seddigh & Darabi, 2017*; *Kyte & Doolittle 1982*). The hydropathicity plots of ApxIA, -IIA and –IIIA uniformly showed hydrophobic tendency on the N-terminal regions (Fig.1). ApxIVA also

showed a slightly greater hydrophobicity on the N-terminal region, though displaying uniformity at neutral scores after a notable drop near 600th amino acid residue. In addition, C-terminal regions generally showed slight depressions with notably smaller peaks compared to N-terminal regions. Though this tendency is not as clear in ApxIVA, strong downward peaks with scores near -3 were shown repeatedly near the C-terminal region.

Topology prediction

Subcellular localizations of ApxA proteins were predicted as extracellular by CELLO program with scores of 4.465, 4.820, 4.828 in the order of ApxIA, -IIA, -IIIA (Table 2). CELLO, incorporating multi-class SVM classifications, generated reliability score of 2.498 for outer membrane and 1.629 for extracellular for ApxIVA. However, PSLpred server determined the subcellular locations of all four ApxA toxins as extracellular with reliability indices of 5, 5, 5 and 4 in the order of ApxIA, -IIA, -IIIA and –IVA, with over 90% accuracy. As expected, TMHMM detected none of transmembrane regions.

Alignments

Multiple sequence alignments of each Apx exotoxins were generated by ClustalW embedded in BioEdit software by using the sequences retrieved from the PSI-Blast. The multiple alignment of Apx toxins and other RTX toxins such as the bifunctional adenylate cyclase toxin-hemolysin (CyaA) of *Bordetella pertussis*, leukotoxin (LtxA) of *Aggregatibacter actinomycetemcomitans* and luekotoxin (LktA) of *Mannhemia haemolytica*, shows that all ApxA exotoxins possess the C-terminal secretion signal downstream of RTX repeat blocks that were modelled as the parallel β-roll by I-TASSER. It was evident that the RTX repeat blocks of ApxA exotoxins well correlate

to the Ca^{2+}-binding RTX domains. Strong homology was observed including the RTX-toxin signature Glycine- and Aspartate- rich tandem L/V-X-G-G-X-G-N/D-D-X, acylation sites of posttranslational modification and C-terminal secretion signal reported by previous studies (Fig. 2) (*Bumba et al, 2016; Stanley, Koronakis & Hughes, 1998*).

Structural and functional analysis

Secondary structures of ApxA proteins were analyzed by the SOPMA and PHD tools as a joint prediction method suggested by NPSA (Network of Protein Sequence Analysis) (*Geourjon, Deleage,* 1995) SOPMA bases on the consensus prediction from multiple alignments, and PHD uses the evolutionary information in the form of multiple sequence alignments. Compositions of secondary structure types are organized and illustrated as graphical plot (Fig.3) (Table 3). SOPMA searches were set to discover 4 conformational states, helix, sheet, turn and coil, and PHD search only provided searches for 3 conformational states, helix, turn and coil. In general, ApxIA, -IIA and –IIIA shared significant resemblance in local distributions of secondary conformations (Table 3). The N-terminal consisted densely of α-helices while β-strands were dispersedly distributed in the rest of the protein sequences. Secondary structure distribution pattern of ApxIVA were distinct from the rest of ApxA proteins. ApxIVA contained segments of α-helix regions in the N-terminus that is evidently segregated by an emphasized region of random coils approximately at 600^{th} residue flanked by multiple short peaks of α-helices. Notably, adjacent to this evident random coil region is the region heavily concentrated with α-helices, shown as a prominent peak in SOPMA diagram.

Motif and domain analysis

A conserved sequence motif in a protein is a repeatedly recurring segment or a pattern of amino acids that is conjectured to suggest biological significance in a group of related proteins. Motif discovery and analysis were carried out by using different methods, MOTIF of GenomeNet, DomPred and ProDom programs. ProDom server predicted domain compositions of ApxA toxins almost identical to GenomeNet MOTIF search results, and ProDom data was not included. The GenomeNet generated a cartoon diagram of each ApxA domain compositions predicted based on three different databases (Fig. 4). The relative distributions of predicted domains within each ApxA toxin are listed in detail with E-values, scores and positions in Table 4. DomPred program predicted domain boundaries within each of ApxA exotoxin (Fig. 5). ApxIA were parsed into total of 5 domains ranging from 1-480, 481-721, 722-844, 845-941 and 942-1022. The individual domains of ApxIIA were defined as 1-476, 477-720, 721-807, 808-956, total of 4 domains. ApxIIIA domains were defined as 1-487, 488-634, 635-735, 736-816, 817-943 and 944-1052, total of 6 domains. ApxIVA domains ranged from 1-412, 413-595, 596-1084, 1085-1226, 1227-1368, 1368-1499, 1500-1644, 1645-1773, 1774-1863 and 1864-1951, making total of 10 domains. Based on the plots generated by the Dompred program, the number of domains were determined to be 5, 4, 5 and 10 for ApxIA, -IIA, -IIIA and –IVA (Table 5).

3D modelling

Three of the I-TASSER ranked structure predictions for each ApxIA, ApxIIA and ApxIIIA as whole proteins are listed with respective C-scores (Fig. 6). The TM-scores were only provided for the first ranked proteins. The I-TASSER generated models of ApxIA, -IIA and –IIIA shared great morphological similarity, displaying bent β-roll shape in majority of ranked models. Variations in 3-D model prediction of ApxIA, -IIA and –IIIA proteins were concentrated in N-

terminal regions mostly occupied by α-helices. The PDB hit provided by the threading programs included, crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*, S-layer protein RsaA from *C. crescentus*, X-ray crystal structure of a hyperactive, Ca^{2+}-dependent, β-helical antifreeze protein from an Antarctic bacterium, crystal structure of LipA from *Serratia marcescens*, and crystal structure of extracellular lipase from *Pseudomonas sp*. MIS38, X-ray crystal structure of a hyperactive, Ca^{2+}-dependent (data not shown). The structural analogs of individual domains on PDB, listed based on the TM-scores in Table 5 (Fig. 7). The 1A1, 2A1 and 3A1 domains of ApxIA, -IIA and –IIIA proteins were analogous to the talin triple domain module. The RTX domain block of Ca^{2+}-binding β-roll was present unanimously in all ApxA toxins as expected. ApxIA, -IIA and –IIIA contained metallo-protease domain, and ApxIVA contained the domain structurally analogous to *Pseudomonas* sp. MIS38 lipase. ApxIVA also had a domain structurally analogous to human nuclear pore complex domain (Table 5).

Antigenic epitope predictions

For antigenic epitope identification, multiple parameters were considered (*Kanehisa et al, 2016*). Presence of energetic frustration in native proteins, antigenic propensity, ligand binding sites, hydrophilicity, flexibility, accessibility, β-turns and surface exposure were correlated with the expected locations of continuous epitopes in each ApxA toxins (*Larsen, Lund, Nielsen, 2006*). Conformational epitopes are generally localized at highly frustrated regions, which often correspond to the binding sites of ligands, macromolecules and antibodies (*Ferreiro et al., 2007*). Potential B-cell epitopes were predicted by Bepipred 2.0, the sequence-based epitope prediction tool, with the threshold of 0.5, giving a specificity and sensitivity of 0.57158 and 0.58564 as

search parameters (*Jespersen et al, 2017*) (Table 6). Characterization of linear epitopes composed of small lengths of peptides in each toxin can provide an efficient step towards discovering novel antigenic determinants that trigger immune response in host cells.

Structural validation

Structural validation via the method of energy frustration serves as a crucial step for pre-experimental design of novel antigenic epitopes and vaccine candidates by characterizing the probable conformational epitopes in antigen-antibody interaction (*Rahbar et al, 2012; Negi & Braun, 2009*). The calculation of local energy frustration of ApxA exotoxins requires the tertiary structures predicted with reasonable TM-scores. The predicted models of ApxIA, -IIA and –IIIA were subjected to the energy frustration computation (*Rahbar et al, 2012*) (Fig. 8). The TM-scores of I-TASSER predicted structures of ApxIA, -IIA and –IIIA were also taken into consideration prior to the structural validation by ProSA program. The calculated TM-score (*Zhang & Skolnick, 2004*) of >0.5 generally indicates a model of correct topology. The validity of the ApxIA, -IIA and –IIIA exotoxin models was confirmed by ProSA program quantified by the z-scores that were computed as 1.78, 1.75 and 1.71 for ApxIA, -IIA and –IIIA (Fig. 9).

Discussion

In recent years, bioinformatics tools have provided promising approaches in biological researches by offering *in silico* studies as pre-experimental procedures especially in discovery of immunogenic epitopes and other essential properties of proteins. There are growing number of studies that incorporate the benefits of *in silico* studies to design the clinical trials (Clermont et

al., 2004) such as developing the vaccines and diagnostic agents (*Aagaard et al., 2003; Bannantine et al., 2002; Korber et al., 2006*). Olfa et al. (2008) conducted *in silico* analysis to study OmcB protein of *Chlamydia tracomatis* to discover the specific and immunogenic antigen to perform serodiagnosis of *C. trachomatis* infections, which they successfully validated the application of *in silico* approach by ELISA tests (*Frikha-Gargouri et al., 2008*). Also, Zhang et al (2002) successfully characterized the intertype epitopes on human adenovirus (HAdV) hexon based on computerized analyses including alignment, antigenicity prediction and 3-D structure characteristics, in which the identified epitopes were validated by Western blot and immunofluorescent assay of synthesized recombinant proteins (*Zhang et al., 2002*). As such, *in silico* studies combined with *in vivo* and *in vitro* validations have been proven to be a compelling methods in research.

The theoretical pI and instability indices of ApxIA, -IIA, -IIIA and -IVA predicted the proteins to be stable. The theoretical isoelectric point, indicating the pH at which the protein carries neutral net charge, can be of a good information in cases of protein purification and accumulation (*Seddigh & Darabi, 2017*). The instability indices of all four ApxA toxins were well below 40, which is often used to determine the stability of proteins in test tubes or *in vivo* environment by providing as a measure of *in vivo* half-life of a protein (*Rogers et al. 1986*). The GRAVY scores ranged from -0.352 to -0.615, where GRAVY values below 0 suggests globular proteins while values above 0 indicate membranous proteins (*Kyte & Doolittle 1982*). Thus, all four ApxA proteins were predicted as expected exotoxins in nature. The relatively high aliphatic index signified the overall thermostability of ApxA exotoxins based on the percent content of aliphatic amino acid side chains, which suggested higher hydrophobicity. The aliphatic indices

were potential measures for determining relatively high protein stability of ApxA exotoxins at high temperatures or chemical detergents. The topology of ApxA exotoxins were validated as exotoxins, in which the ApxA toxins inflict pathogenic activities in host macrophages via extracellular export by T1SS pathway (*Thomas, Holland & Schmitt, 2014*)

To enhance the C-scores of predicted structures, each ApxA toxins were parsed into individual domains based on the suggested domain boundaries. The predicted putative domain boundary patterns within ApxIA, -IIA and –IIIA proteins were remarkably similar. The similar patterns in domain boundaries of ApxIA, -IIA and –IIIA was considered as an implication of similar building blocks comprising the three specific proteins. In domain-wise interpretation the process of domain selection among the ranked models also relied on the implication that each corresponding domains between the exotoxins share structural or functional resemblance. Domains 1A1, 2A1 and 3A1, which span the N-terminal regions of each ApxIA, -IIA and –IIIA exotoxins (1-480, 1-476, 1-488 respectively) were very similar to the triple domain module of talin protein with TM-scores of 0.90, 0.885 and 0.873, which was proven to play a central role in regulating the integrin-mediated signaling and maintaining the quaternary structure of talin protein (*Zhang et al., 2016*). Similarly, the structural analogs of domains 1A2, 2A2 and 3A2 (481-721, 477-720, 488-735 of ApxIA, -IIA, and –IIIA, respectively) were all related to protease activity, also supported by strong TM-scores of 0.825, 0.880 and 0,767 in the same order. These consensus structural analogs of metallo-protease activity are conserved in the serralysin of *Serratia marcescens*, alkaline protease of *Pseudomonas aeruginosa* and protease C (PrtC) of *Erwinia chrysanthemi* (*Oberholzer et al., 2009; Baumann et al., 1995; Hege & Baumann, 1998*). These proteins also commonly possess the Ca^{2+}-binding β-roll motif with the glycine-rich

nonapeptide tandems (*Oberholzer et al., 2009*). ApxIA, -IIA and –IIIA exotoxins, much like the hemolysin (HlyA) of *Escherichia coli*, are also known to undergo posttranslational modification via acylation at conserved KI and KII positions, which according to the domain assignments by this data, both acylation sites reside in these 1A2, 2A2 and 3A2 domains (*Stanley, Koronakis & Hughes, 1998)*. In contrast, ApxIVA exotoxin showed distinct characteristics in the aspects of domain-analysis. The N-terminal domain, 4A1, of ApxIVA exotoxin was predicted with the C-score of -0.49, showing TM-score of 0.943 to the lipase of *Pseudomonas* sp. MIS38. This lipase domain of *Pseudomonas* sp. MIS38 is known to be greatly similar to the I.3 lipase (LipA) of *Serratia marcenscens*, that assumes lid-opening mechanism by the helical hairpin structure at the N-terminal end of the domain (*Angkawidjaja et al., 2007)*. Whether the 4A1 domain of ApxIVA contribute the similar lipase activity based on the lid-opening mechanism requires further investigations. The predicted model of 4A3 domain showed C-score of -1.39 with the TM-score of 0.943 to the structural analog of human nuclear pore complex. The presence of this domain within the ApxIVA exotoxin may explain the pore-forming activity of *A. pleuropneumoniae*. The predicted domains 1A3, 1A4, 2A3, 3A3, 3A4, 4A4, 4A5, 4A6, 4A7, 4A8, 4A9 and 4A10 with given C-scores in Table.10 were given the PDB analogs of RTX domain block V, a well-studied RTX signature calcium ion binding and glycine-rich nonapeptide motif that is necessary for the virulence of *A. pleuropneumoniae* upon Ca^{2+}-binding. As such, the computed 3-D structures of individual domains in ApxA exotoxins generated promising structural and functional annotations that may reveal the pathogenic mechanisms of *A. pleuropneumoniae*.

3-D models predicted by I-TASSER programs demonstrated validity both by local frustration and ProSA-web. In general, the presence of only a small fraction of highly frustrated regions and

prominent distributions of minimally frustrated and neutral regions in the measure of local frustration density throughout the ApxIA, -IIA and –IIIA exotoxins support the *in vivo* stability of the predicted structures, which also support the structural validity of each models (*Rahbar et al, 2012*). The energy landscape theory-inspired algorithm quantifies the location of frustration within molecules where sites of frustration often indicate regions of biological importance such as bindings (*Parra et al., 2016*). Thus, the energy frustration throughout the predicted protein models can also be used to conjecture the regions of antigenic epitopes. However, the ProSA-web generated plots indicate poor validity of I-TASSER predicted models of ApxIA, -IIA and –IIIA exotoxins, often observed for incorrect structures (*Wiederstein & Sippl, 2007*). The z-scores of typical native proteins of similar sizes to ApxA whole proteins generally lie near -12, which is far below the z-scores of ApxIA, -IIA and –IIIA structures.

Glycine-rich repeat motifs were detected as functional epitopes by Bepipred 2.0 along with several predicted epitopes, ranging from 16 to 20 epitopes. As antigenic epitopes are often stretched within the range of 5 to 25 amino acid residues, further pre-experimental analysis of ApxA exotoxins' epitopes is thought to be required (*Kringelum et al., 2013*).

Conclusion

The prediction of 3-D structure of a protein from its amino acid sequence has been the grand challenge in structural biology. Despite the difficulties that rise from the lack of sizable structural data registered in database, constant efforts and applications of bioinformatics methods as pre-experimental research are being conducted as introduced in this study. Though the predicted whole structures of ApxIA, -IIA and –IIIA were validated poorly, interpretations via structural

analogs to individual domains conducted in this study may provide more invaluable insights in pathogenicity of *A. pleuropneumoniae*. Thus, with careful analysis and data interpretation, bioinformatics approach in the drug or vaccine design as well as the antigenic epitope discovery in the means of diagnostic purpose of certain diseases is clearly an alternative to the conventional time consuming and costly methods. In addition, the structural and functional findings of ApxA exotoxins in this study may aid in the prevention and control of the *A. pleuropneumoniae* infection.

Reference

Aagaard C, Govaerts M, Meng OL, Andersen P, Pollock JM. 2003. Genomic approach to identification of *Mycobacterium bovis* diagnostic antigens in cattle. *Journal of Clinical Microbiology* **8**:3719–3728 DOI 10.1128/JCM.41.8.3719-3728.2003

Angkawidjaja C, You DJ, Matsumura H, Kuwahara K, Koga Y, Takano K, Kanaya S. 2007. Crystal structure of a family I.3 lipase from *Pseudomonas* sp. MIS38 in a closed conformation. *FEBS Letter* **581**:5060–5064 DOI 10.1016/j.febslet.2007.09.048

Baltes N, Hennig-Pauka I, Gerlach GF. 2002. Both transferrin binding proteins are virulence factors in *Actinobacillus pleuropneumoniae* serotype 7 infection. *FEMS Microbiology Letters* **209**:283–287 DOI 10.1111/j.1574-6968.2002.tb11145.x

Bandara AB, Lawrence ML, Veit HP, Inzana TJ. 2003. Association of *Actinobacillus pleuropneumoniae* capsular polysaccharide with virulence in pigs. *Infection and Immunity* **71**:3320-3328 DOI 10.1128/IAI.71.6.3320–3328.2003

Bannantine JP, Baechler E, Zhang Q, Li L, Kapur V. 2002. Genome scale comparison of *Mycobacterium avium* subsp. *paratuberculosis* with *Mycobacterium avium* subsp. *avium* reveals potential diagnostic sequences. *Journal of Clinical Microbiology* **40**:1303–1310

Bhasin M, Garg A, Raghava GP. 2005. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics* **21**:2522–2524

Boekema BK, Kamp EM, Smits MA, Smith HE. 2004. Both ApxI and ApxII of *Actinobacillus pleuropneumoniae* serotype 1 are necessary for full virulence. *Veterinary Microbiology* **100**:17-23 DOI 10.1016/j.vetmic.2003.09.024

436 **Baumann U, Bauer M, Letoffe S, Delepelaire P, Wandersman C. 1995.** Crystal structure
437 of a complex between *Serratia marcescens* metallo-protease and an inhibitor from *Erwinia*
438 *chrysanthemi*. *Journal of Molecular Biology* **248**:653–661
439 **Bumba L, Masin J, Macek P, Wald T, Motlova L, Bibova I, Klimova N, Bednarova L,**
440 **Veverka V, Kachala M, Svergun DI, Barinka C, Sebo P. 2016.** Calcium-driven folding of
441 RTX domain β-rolls ratchets translocation of RTX proteins through type I secretion ducts.
442 *Molecular Cell* **62**:47–62 DOI 10.1016/j.molcel.2016.03.018
443 **Clermont G, Bartels J, Kumar R, Constantine G, Vodovotz Y, Chow C. 2004.** In silico
444 design of clinical trials: a method coming of age. *Critical Care Medicine* **32**:2061-70
445 **Frey J. 1995.** Virulence in *Actinobacillus pleuropneumoniae* and RTX toxins. *Trends in*
446 *Microbiology* **3**:257-261 DOI 10.1016/S0966-842X(00)88939-8
447 **Frikha-Gargouri O, Gdoura R, Znazen A, Gargouri B, Gargouri J, Rebai A,**
448 **Hammami A. 2008.** Evaluation of an in silico predicted specific and immunogenic antigen
449 from the OmcB protein for the serodiagnosis of *Chlamydia trachomatis* infections. *BMC*
450 *Microbiology* **8**:217
451 **Gao J, Faraggi E, Zhou Y, Ruan J, Kurgan L. 2012.** BEST: Improved prediction of B-cell
452 epitopes from antigen sequences. *PLOS ONE* **7**:e40104 DOI 10.1371/journal.pone.0040104
453 **Garg A, Raghava GP. 2008.** A machine learning based method for the prediction of
454 secretory proteins using amino acid composition, their order and similarity-search. *In Silico*
455 *Biology* **8**:129–140
456 **Geourjon C, Deleage G. 1995.** SOPMA: significant improvements in protein secondary
457 structure prediction by consensus prediction from multiple alignments. *Computer*
458 *Applications in the Biosciences: CABIOS* **11**:681–684 DOI 10.1093/bioinformatics/11.6.681

Hege T, Baumann U. 2001. Protease C of Erwinia chrysanthemi: the crystal structure and role of amino acids Y228 and E189. *Journal of Molecular Biology* **314**:187–193

Inzana TJ, Fenwick B. 2011. Serologic detection of *Actinobacillus pleuropneumoniae* in swine by capsular polysaccharide-biotin-streptavidin enzyme-linked immunosorbent assay. *Journal of Clinical Microbiology* **39**:1279–1282 DOI 10.1128/JCM.39.4.1279-1282.2001

Inzana TJ, Todd J, Ma J, Veit H. 1991. Characterization of a non-hemolytic mutant of *Actinobacillus pleuropneumoniae* serotype 5: role of the 110 kilodalton hemolysin in virulence and immunoprotection. *Microbial Pathogenesis* **10**:281-296 DOI 10.1016/0882-4010(91)90012-Y

Jespersen MC, Peters B, Nielsen M, Marcatili P. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research* **45**:W24-W29 DOI 10.1093/nar/gkx346

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Ressearch* **44**:D457-D462 DOI 10.1093/nar/gkv1070

Kyte J, Doolittle RF. 1982. A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology* **157**:105-132

Korber B, LaBute M, Yusim K. 2006. Immunoinformatics comes of age. *PloS Computational Biology* **2**:484–492 DOI 10.1371/journal.pcbi.0020071

Kringelum JV, Nielsen M, Padkjær SB, Lund O. 2013. Structural analysis of B-cell epitopes in antibody:protein complexes. *Molecular Immunology* **53**:24–34

Kyte J, Doolittle RF. 1982. A Simple Method for Displaying the Hydropathic Character of

a Protein. *Journal of Molecular Biology* **157**:105-132

Larsen JE, Lund O, Nielsen M. 2006. Improved method for predicting linear B-cell epitopes. *Immunome Research* **2**:2.

Laskowski RA, Macarthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**:283-291 DOI 10.1107/S0021889892009944

Linhartova I, Bumba L, Masin J, Basler M, Osicka R, Kamanova J, Prochazkova K, Adkins I, Hejnova-Holubova J, Sadilkova L, Morova J, Sebo P. 2010. RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiology Reviews* **34**:1076-1112. DOI 10.1111/j.1574-6976.2010.00231.x

Maier E, Reinhard N, Benz R, Frey J. 1996 Channel-Forming Activity and Channel Size of the RTX Toxins ApxI, ApxII, and ApxIII of *Actinobacillus pleuropneumoniae. Infection and Immunity* **64**:4415–442

Marsden RL, McGuffin LJ, Jones DT. 2002. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Science* **11**:2814–2824

Negi SS, Braun W. 2009. Automated detection of conformational epitopes using phage display peptide sequences. *Bioinformatics and Biology Insights* **3**:71-81

Oberholzer AE, Bumann M, Hege T, Russo S, Baumann U. 2009. Metzincin's canonical methionine is responsible for the structural integrity of the zinc-binding site. *Biological Chemistry* **390**:875–881 DOI 10.1515/BC.2009.100

Parra RG, Schafer NP, Radusky LG, Tsai MY, Guzovsky AB, Wolynes PG, Ferreiro DU. 2016. Protein Frustratometer 2: A tool to Localize Energetic Frustration in Protein Molecules, Now with Electrostatics. *Nucleic Acids Research* **44**:W356–360

Pellequer JL, Westhof E, Van Regenmortel MH. 1991. Predicting location of continuous epitopes in proteins from their primary structures. *Methods in Enzymology* **203**:176–201

Persson B, Argos P. 1996. Topology prediction of membrane proteins. *Protein Science* **5**:363–371 DOI 10.1002/pro.5560050221

Petrey D, Honig B. 2005. Protein structure prediction: inroads to biology. *Molecular Cell* **20**:811–819 DOI 10.1016/j.molcel.2005.12.005.

Rahbar MR, Rasooli I, Gargari SL, Sandstrom G, Amani J, Fattahian Y, Jahangiri A, Jalali M. **2012.** A potential in silico antibody-antigen based diagnostic test for precise identification of *Acinetobacter baumannii*. *Journal of Theoretical Biology* **294**:29–39 DOI 10.1016/j.jtbi.2011.10.026

Rahbar MR, Rasooli I, Mousavi Gargari SL, Amani J, Fattahian Y. 2010. In silico analysis of antibody triggering biofilm associated protein in *Acinetobacter baumannii*. *Journal of Theoretical Biology* **266**:275–290 DOI 10.1016/j.jtbi.2010.06.014

Rogers S, Wells R, Rechsteiner M. 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**:364–368

Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**:798-804

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Natural Protocols* **5**:725–738

Seddigh S, Darabi M. 2017. Functional, structural, and phylogenetic analysis of mitochondrial cytochrome b (cytb) in insects. *Mitochondrial DNA A DNA Map Sequencing and Analysis* **24:**1-17 DOI 10.1080/24701394.2016.1275596

Shin MK, Cha SB, Lee WJ, Yoo H. Predicting genetic traits and epitope analysis of

528 apxIVA in Actinobacillus pleuropneumoniae. *The Journal of Microbiology* **49**:462-8 DOI
529 10.1007/s12275-011-0449-y
530 **Shin MK, Kang MR, Cha SB, Lee WJ, Sung JH, Yoo HS. 2011.** An immunosorbent
531 assay based on the recombinant ApxIa, ApxIIa, and ApxIIIa toxins of *Actinobacillus*
532 *pleuropneumoniae* and its application to field sera. *Journal of Veterinary Diagnostic*
533 *Investigation* **23**:736-42 DOI 10.1177/1040638711407889
534 **Stanley P, Koronakis V, Hughes C. 1998.** Acylation of *Escherichia coli* hemolysin: a
535 unique protein lipidation mechanism underlying toxin function. *Microbiology and Molecular*
536 *Biology Reviews* **62**:309–333
537 **Tascon RI, Vazquez-Boland JA, Gutierrez-Martin CB, Rodriguez-Barbosa I,**
538 **Rodriguez-Ferri EF. 1994.** The RTX haemolysins ApxI and ApxII are major virulence
539 factors of the swine pathogen *Actinobacillus pleuropneumoniae*: evidence from mutational
540 analysis. *Molecular Microbiology* **14**:207–216 DOI 10.1111/j.1365-2958.1994.tb01282.x
541 **Thomas S, Holland IB, Schmitt L. 2014.** The Type 1 secretion pathway-the hemolysin
542 system and beyond. *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research*.
543 **1843**:1629-1641 DOI 10.1016/j.bbamcr.2013.09.017
544 **Vandepoele K, Jeroen R, Veylder LD, Rouze P, Rombauts S, Inze D. 2002.** Genome-
545 Wide Analysis of Core Cell Cycle Genes in Arabidopsis. *The Plant Cell* **14**:903-916 DOI
546 10.1105/tpc.010445
547 **Wang C, Wang Y, Shao M, Si W, Liu H, Chang Y, Peng W, Kong X, Liu S. 2009.**
548 Positive role for rApxIVN in the immune protection of pigs against infection by
549 *Actinobacillus pleuropneumoniae Vaccine* **27**:5816–5821 DOI
550 10.1016/j.vaccine.2009.07.065

Welch RA. 1991. Pore-forming cytolysins of gram-negative bacteria. *Molecular Microbiology* **5**:521-528 DOI 10.1111/j.1365-2958.1991.tb00723.x

Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* **35**:W407–W410 DOI 10.1093/nar/gkm290.

Wiles TJ, Mulvey MA. 2013. The RTX pore-forming toxin alpha-hemolysin of uropathogenic Escherichia coli: progress and perspectives. *Future Microbioly* **8**:73-84 DOI 10.2217/fmb.12.131

Yang J, Zhang Y. 2015. Protein structure and function prediction using I-TASSER. *Current Protocols in Bioinformatics* **52**:5–8

Yu CS, Chen YC, Lu CH, Hwang JK. 2006. Prediction of protein subcellular localization. *Proteins* **64**:643–651 DOI 10.1002/prot.21018

Zhang H, Chang YC, Huang Q, Brennan ML, Wu J. 2016. Structural and Functional Analysis of a Talin Triple-Domain Module Suggests an Alternative Talin Autoinhibitory Configuration. *Structure* **24**:721–729 DOI 10.1016/j.str.2016.02.020

Zhang T, Xin R, Zhang Q, Bao Y, Wu J. 2002. Characterization of intertype specific epitopes on adenoviruses. *Chinese journal of experimental and clinical virology* **16**:44-7

Table 1(on next page)

Physiochemical properties of ApxA proteins by ProtParam and Compute pI.

Table 1 Physiochemical properties of ApxA proteins by ProtParam and Compute pI.

GenBank ID	Name	ProtParam								Compute pI
		No. amino acids	Formula	Total no. of atoms	Gravy	Aliphatic index	Instability index	Ex. Coefficient	Mol. weight	Theoretical pI
AAK50051.1	RTX toxin IA	1022	$C_{4903}H_{7744}N_{1340}O_{1553}S_4$	15544	-0.352	88.63	25.74	90080	110439.71	5.50
AAK50052.1	RTX toxin IIA	956	$C_{4474}H_{7203}N_{1279}O_{1468}S_3$	14427	-0.365	91.20	25.58	58790	102495.27	5.88
AAK50053.1	RTX toxin IIIA	1052	$C_{4963}H_{7911}N_{1373}O_{1616}S_6$	15869	-0.370	87.46	18.60	82630	112862.99	5.62
ADG63470.1	RTX toxin IVA	1951	$C_{9667}H_{14830}N_{2608}O_{3135}S_{22}$	30262	-0.615	76.59	20.98	252010	218451.00	4.75

Table 2(on next page)

Subcellular localization prediction.

Table 2 Subcellular localization predictions of ApxA exotoxins by CELLO v2.5 and PSLpred.

SVM	ApxIA		ApxIIA		ApxIIIA		ApxIVA	
	Localization	Reliability	Localization	Reliability	Localization	Reliability	Localization	Reliability
Amino Acid Comp.	Extracellular	0.971	Extracellular	0.984	Extracellular	0.989	Extracellular	0.591
N-peptide Comp.	Extracellular	0.952	Extracellular	0.961	Extracellular	0.959	Extracellular	0.409
Partitioned seq. Comp.	Extracellular	0.985	Extracellular	0.984	Extracellular	0.990	Outer membrane	0.965
Physico-chemical Comp.	Extracellular	0.925	Extracellular	0.925	Extracellular	0.926	Inner membrane	0.374
Neighboring seq. Comp.	Extracellular	0.632	Extracellular	0.966	Extracellular	0.963	Outer membrane	0.545
CELLO prediction	Extracellular	4.465*	Extracellular	4.820*	Extracellular	4.828*	Outer membrane	2.498*
	Periplasmic	0.164	Periplasmic	0.072	Periplasmic	0.074	Extracellular	1.629
	Cytoplasmic	0.160	Cytoplasmic	0.058	Cytoplasmic	0.050	Inner membrane	0.594
	Outer membrane	0.152	Outer membrane	0.028	Outer membrane	0.028	Cytoplasmic	0.150
	Inner membrane	0.060	Inner membrane	0.023	Inner membrane	0.020	Periplasmic	0.130
PSLpred	Extracellular	5	Extracellular	5	Extracellular	5	Extracellular	4

Table 3(on next page)

Percent secondary structure compositionsof ApxIA, ApxIIA, ApxIIIA and ApxIVA retrieved by SOPMA and PHD tools.

Table 3 Percent secondary structure compositions of ApxIA, ApxIIA, ApxIIIA and ApxIVA retrieved by SOPMA and PHD tools.

Secondary Structure	ApxIA		ApxIIA		ApxIIIA		ApxIVA	
	SOPMA	PHD	SOPMA	PHD	SOPMA	PHD	SOPMA	PHD
a-helix	45.69	44.52	47.38	44.35	44.49	43.44	32.29	27.78
Extended strand	18.59	15.56	17.47	15.27	18.44	15.21	23.22	23.01
b-turn	11.84	N/A	8.89	N/A	10.74	N/A	9.48	N/A
Random coil	23.87	39.92	26.26	40.38	26.33	41.35	35.01	49.21

Table 4(on next page)

List ofprotein motifs retrieved by Genomenet Motif search server.

Listed motifs are provided with the responsible database and accession id, name of the motif, position, score and E-value.

Table 4 List of protein motifs retrieved by Genomenet Motif search server. Listed motifs are provided with the responsible database and accession id, name of the motif, position, score and E-value.

Toxin	Entry	Motif	Position	Score	E-value
ApxIA	308152	RTX N-terminal domain	9-652	1017	0.0
	311996	RTX C-terminal domain	869-1018	149	1e-40
	225483	RTX toxins and related Ca^{2+}-binding proteins [Secondary metabolites biosynthesis, transport, and catabolism]	721-845	75.4	6e-14
	285719	Peptidase M10 serralysin C terminal	739-780	50.8	9e-07
	306789	Haemolysin-type calcium-binding repeat	756-773	31.6	0.39
	223083	Ammonia permease [Inorganic ion transport and metabolism]	342-406	34.1	0.40
	235773	aspartate kinase; Provisional	391-434	33.4	0.67
	310829	Protein of unknown function (DUF1093)	559-611	30.7	0.79
ApxIIA	308152	RTX N-terminal domain	6-654	1001	0.0
	311996	RTX C-terminal domain	810-942	151	2e-41
	225483	RTX toxins and related Ca^{2+}-binding proteins [Secondary metabolites biosynthesis, transport, and catabolism]	703-812	59.6	5e-09
	285719	Peptidase M10 serralysin C terminal	737-826	53.6	7e-08
	226018	IrpA, Uncharacterized iron-regulated protein [Inorganic ion transport and metabolism]	139-206	40.7	0.004
	214853	Inhibitor_I29, Cathepsin propeptide inhibitor domain	427-473	31.8	0.36
ApxIIIA	308152	RTX, RTX N-terminal domain	11-668	757	0
	225483	RTX toxins and related Ca^{2+}-binding proteins [Secondary metabolites biosynthesis, transport, and catabolism]	750-862	77.2	2e-14
	285719	Peptidase M10 serralysin C terminal	754-795	51.5	5e-07
	311996	RTX_C, RTX C-terminal domain	886-1049	50.1	8e-07
	310219	ApoL, Apolipoprotein L	134-192	36.0	0.089
	309442	YdjM, LexA-binding, inner membrane-associated putative hydrolase	376-436	35.1	0.11
	224093	AraH, Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components [Carbohydrate transport and metabolism]	368-417	34.8	0.25

	119321	**TM_PBP1_transp_AraH_like, Transmembrane subunit (TM) of Escherichia coli AraH and related proteins**	**368-417**	**34.3**	**0.26**
	310976	**ThrE, Putative threonine/serine exporter**	**366-417**	**34.0**	**0.33**
	206776	**RNase_PH_MTR3, MTR3 subunit of eukaryotic exosome**	**370-415**	**32.9**	**0.59**
	185066	**PRK15111, antimicrobial peptide ABC transporter permease SapC; Provisional**	**247-331**	**33.1**	**0.78**
	311816	**SpoU_sub_bind, RNA 2'-O ribose methyltransferase substrate binding**	**44-97**	**30.7**	**0.88**
ApxIVA	**225483**	**RTX toxins and related Ca^{2+}-binding proteins [Secondary metabolites biosynthesis, transport, and catabolism]**	**1370-1885**	**82.0**	**1e-15**
	285719	**Peptidase_M10_C, Peptidase M10 serralysin C terminal**	**1093-1135**	**44.2**	**3e-04**
	188993	**M3_like, Peptidase M3-like family, a zincin metallopeptidase, includes M3 and M32 families**	**839-902**	**35.4**	**0.31**
	310889	**HCBP_related, Haemolysin-type calcium binding protein related domain**	**1873-1905**	**32.4**	**0.45**
	311996	**RTX_C, RTX C-terminal domain**	**1857-1923**	**33.2**	**0.87**

3

4

Table 5(on next page)

5 PDB registeredstructural analogs to parsed domain structures of each ApxA exotoxins predictedby I-TASSER.

Table 5 PDB registered structural analogs to the parsed domain structures of each ApxA exotoxins predicted by I-TASSER.

Toxin	Domain ID	Start	End	# of res.	PDB hit	TM-score	RMSD[a]	IDEN[a]	Cov	Name
ApxIA	1A1	1	480	480	5ic0A	0.90	1.10	0.107	0.912	Structural analysis of a talin triple domain module
	1A2	481	721	241	3hb2P	0.825	2.63	0.086	0.913	PrtC methionine mutants: M226I
	1A3	722	844	123	5cxlA	0.700	3.32	0.265	0.911	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	1A4	845	941	96	5cvwA	0.775	1.24	0.284	0.835	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	1A5	942	1022	81	4rqoA	0.545	3.48	0077	0.901	Crystal structure of L-Serine Dehydratase from *Legionella pneumophila*
ApxIIA	2A1	1	476	476	5ic0A	0.885	1.61	0.098	0.918	Structural analysis of a talin triple domain module
	2A2	477	720		1smpA	0.880	1.96	0.115	0.963	Crystal structure of a complex between *Serratia marcescens* metallo-protease and an inhibitor from *Erwinia chrysanthemi*
	2A3	721	807	87	5cxlA	0.876	1.88	0.276	1.00	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	2A4	808	956	148	3dyoD	0.492	4.67	0.071	0.778	*E. coli* (lacZ) beta-galactosidase (H418N) in complex with IPTG
ApxIIIA	3A1	1	487	487	5ic0A	0.873	2.39	0.085	0.938	Structural analysis of a talin triple domain module
	3A2	488	735		1k7Ia	0.767	2.92	0.128	0.887	PrtC from *Erwinia chrysanthemi*: Y228F mutant
	3A3	736	816	81	5cxlA	0.951	0.86	0.407	1.000	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	3A4	817	943	127	5cvwA	0.901	0.81	0.356	0.929	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	3A5	944	1052	108	2yevA2	0.700	3.16	0.074	0.982	Structure of caa3-type cytochrome oxidase
ApxIVA	4A1	1	412	412	2z8Za	0.963	1.44	0.123	0.985	Crystal structure of a platinum-bound S445C mutant of *Pseudomonas* sp. MIS38 lipase
	4A2	413	595	182	2hq0A	0.749	3.08	0.092	0.940	Structure of RafE from *Streptococcus pneumoniae*
	4A3	596	1084	488	5a9q4	0.943	2.02	0.094	0.982	Human nuclear pore complex
	4A4	1085	1226	141	5cvwA	0.927	0.81	0.333	0.951	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	4A5	1227	1368	142	5cxlA	0.925	0.69	0.328	0.944	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	4A6	1369	1499	131	5cxlA	0.926	0.51	0.293	0.939	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	4A7	1500	1644	145	5cvwA	0.916	0.87	0.314	0.945	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	4A8	1645	1773	129	5cxlA	0.912	0.74	0.298	0.938	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	4A9	1774	1863	90	5cvwA	0.884	2.00	0.367	1.000	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*
	4A10	1864	1951	89	5cvwA	0.658	0.69	0.433	0.682	Crystal structure of RTX domain block V of adenylate cyclase toxin from *Bordetella pertussis*

Table 6(on next page)

B-cell epitopepredictions by Bepipred 2.0 program for ApxIA, -IIA, -IIIA and -IVA.

Table 6 B-cell epitope predictions by Bepipred 2.0 program for ApxIA, -IIA, -IIIA and –IVA with threshold of 0.5.

ApxIA			ApxIIA			ApxIIIA			ApxIVA		
Start	End	Peptide length	Start	End	Peptide length	Start	End	Peptide length	Start	End	Peptide length
5	22	18	56	64	9	19	28	10	15	42	28
52	59	8	77	88	12	65	73	9	66	99	34
74	88	15	94	103	10	87	102	16	107	115	9
119	138	20	124	141	18	132	147	16	174	182	9
162	173	12	161	173	13	221	247	27	192	204	13
196	205	10	196	204	9	338	357	20	217	242	26
214	223	10	214	221	8	368	382	15	254	272	19
225	236	12	225	233	9	417	493	77	291	326	36
327	346	20	256	264	9	501	507	7	328	342	15
423	575	153	328	357	30	512	647	136	349	366	18
583	636	54	367	374	8	666	758	93	376	412	37
644	725	82	415	465	51	768	826	59	419	448	30
747	781	35	473	498	26	835	842	8	458	524	67
790	810	21	503	609	107	848	855	8	563	636	74
825	838	14	641	741	101	862	903	42	647	675	29
840	865	26	790	807	18	905	1048	144	687	810	124
869	885	17	809	903	95				816	1094	279
887	957	71							1098	1111	14
968	980	13							1114	1787	674
1006	1019	14							1830	1943	114

Figure 1

Kyle & Doolittlehydropathy plots for ApxA exotoxins. Apolar residues are assigned negativevalues while positive values indicate hydropathicity.

ApxIA, -IIA and –IIIA display resembling hydropathy patterns. (A) Hydropathy plot for ApxIA toxin (B) Hydropathy plot for ApxIIA toxin (C) Hydropathy plot for ApxIIIA toxin (D) Hydropathy plot for ApxIVA.

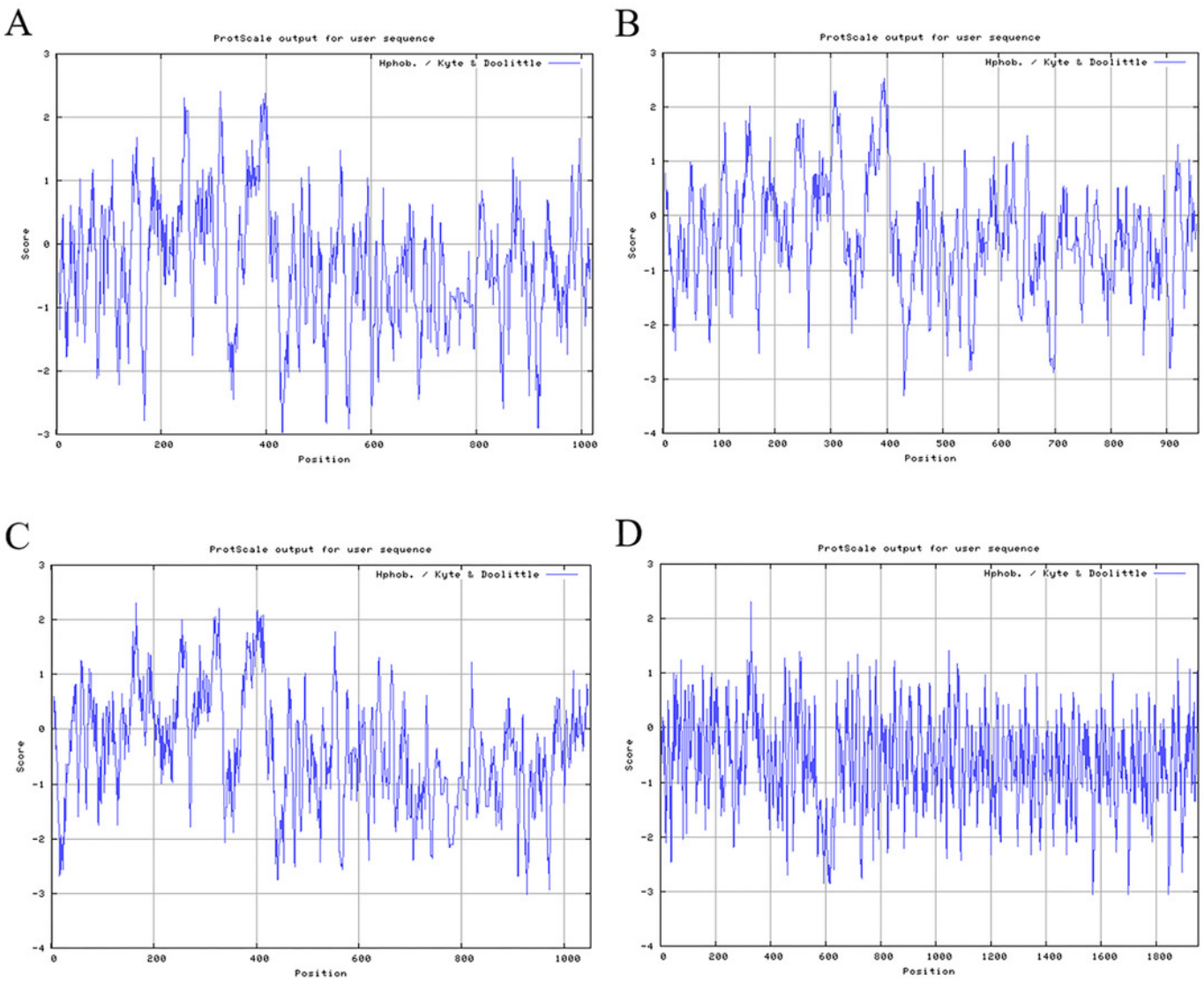


Figure 2

Sequence alignments of ApxA exotoxins and other RTXtoxins by ClustalW alignment program.

(A) Significant homology of Glycine-rich RTX tandem repeats is shown with consensus C-terminal secretion signal motif. The conserved residues involved in initiation of Ca^{2+} -induced folding of the RTX b -roll motif are indicated by red box. (B) Consensus posttranslational acylation modification sites are indicated by red boxes in alignments.

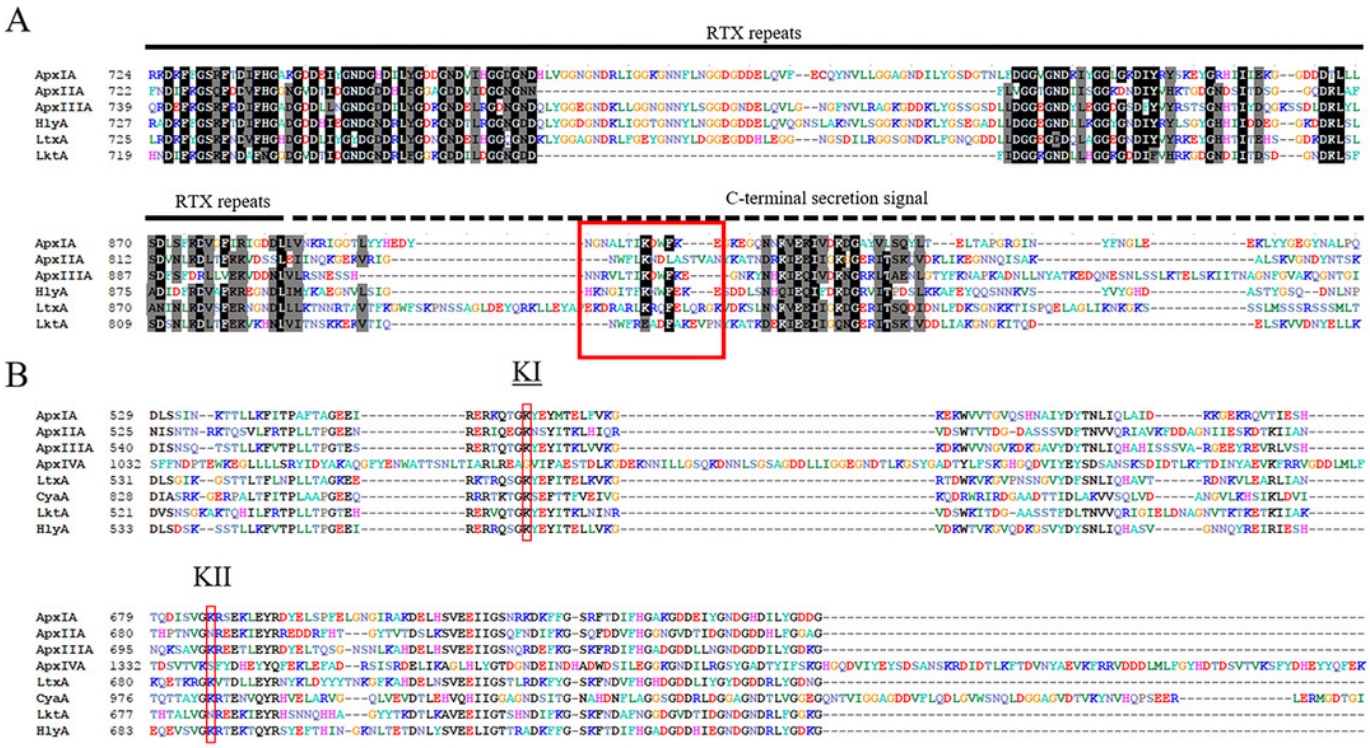


Figure 3

SOPMA (significant improvements in proteinsecondary structure prediction by consensus prediction from multiplealignments) and PHD (Prediction of protein secondary structure at better than70% accuracy) secondary structure predictions

SOPMA secondary structure prediction graph where blue, red, green and purple each represents helix, extended strand, beta-turn and random coil respectively. PHD secondary structure prediction graph where blue, red and purple each denotes for helix, extended strand and random coil respectively. (A) SOPMA and PHD result of ApxIA (B) SOPMA and PHD result of ApxIIA (C) SOPMA and PHD result of ApxIIIA (D) SOPMA and PHD result of ApxIVA.

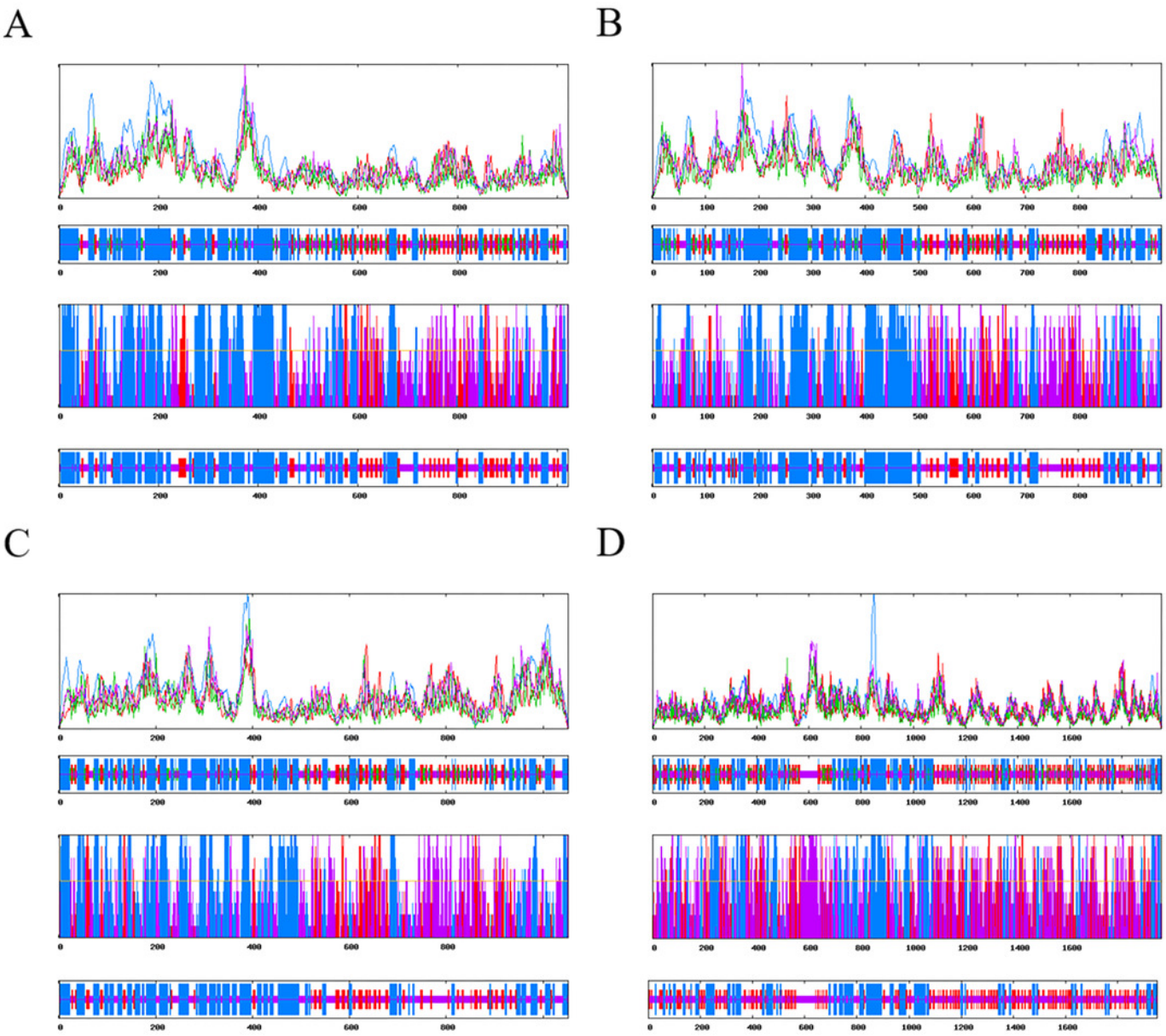

A
B
C
D

Figure 4

MOTIF search results by Genomenet serverdemonstrating general domain distributions within each of ApxA exotoxin.

Green, red and blue bars represent protein motifs/domains retrieved by Prosite, NCBI-CDD and Pfam databases, respectively. Domain annotations by different databases among the four ApxA exotoxins are significantly similar. (A) Protein domain annotation in ApxIA toxin by NCBI-CDD and Prosite database shows general domain distributions (B) General domain distribution within the ApxIIA exotoxin, displaying similar distribution to other ApxA toxins (C) Domain distributions within ApxIIIA toxin (D) Domain distribution within ApxIVA toxin shows distinct patterns compared to the rest of ApxA toxins.

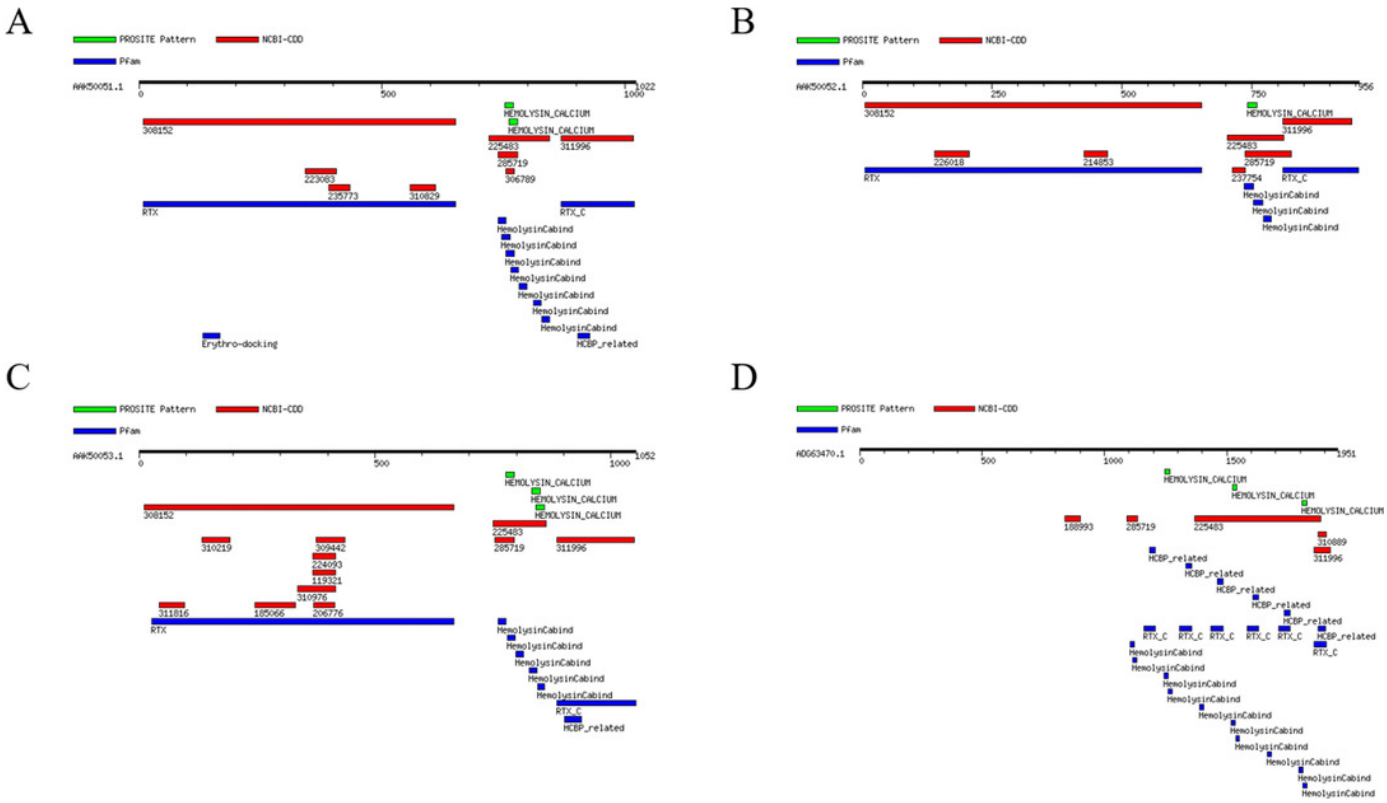
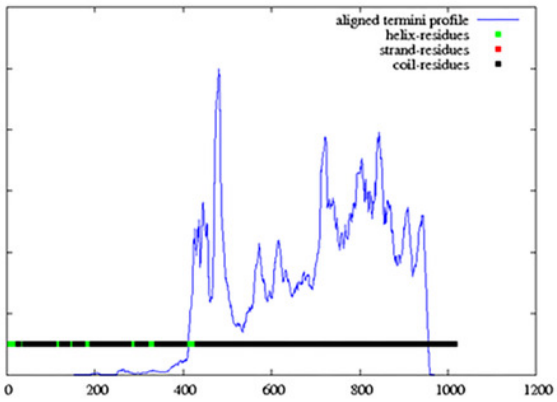


Figure 5

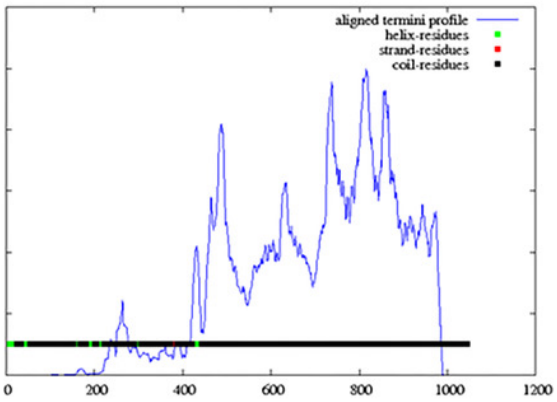
Putativedomain boundaries plot constructed by incorporating an algorithm to identifypositions of amino acids where N and C termini of matching database hits arealigned to the query sequence.

The N and C termini positions from the PSI-BLAST matches are summed along the input sequence, giving higher weightings to the regions where both N and C termini hits are found. (A) Domain boundaries of ApxIA were defined as 480, 721, 844 and 941. (B) Domain boundaries for ApxIIA were 476, 720, 807. (C) Domain boundaries for ApxIIIA were 487, 634, 735, 816, 943 (D) ApxIVA domain boundaries were 412, 595, 1085, 1227, 1368, 1500, 1645, 1774 and 1864.

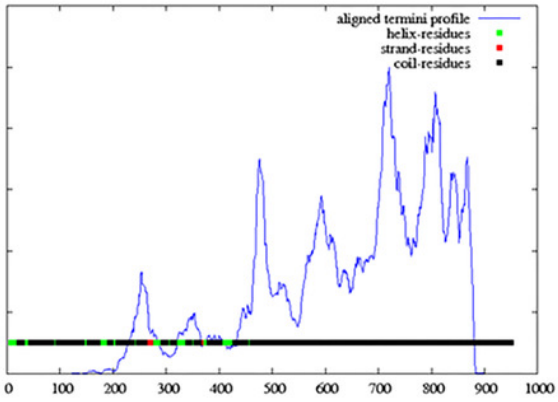
A


aligned termini profile
helix-residues
strand-residues
coil-residues
0
200
400
600
800
1000
1200

B


aligned termini profile
helix-residues
strand-residues
coil-residues
0
200
400
600
800
1000
1200

C


aligned termini profile
helix-residues
strand-residues
coil-residues
0
100
200
300
400
500
600
700
800
900
1000

D

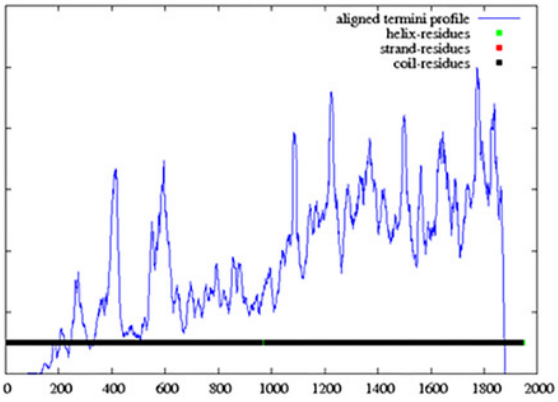
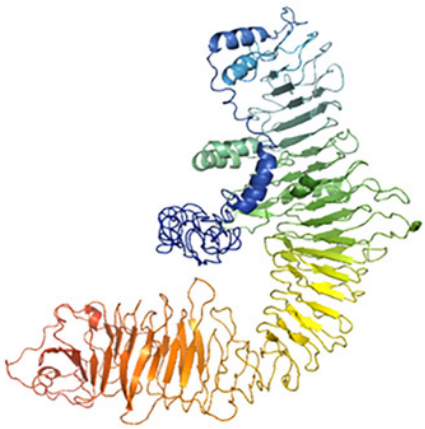

aligned termini profile
helix-residues
strand-residues
coil-residues
0
200
400
600
800
1000
1200
1400
1600
1800
2000

Figure 6

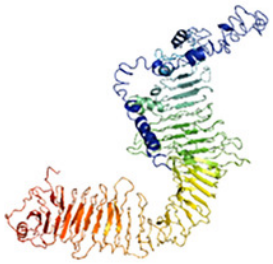
Predicted3-D structures of ApxIA, -IIA and –IIIA exotoxins displayed with three versionsof different C-scores for each toxin.

The C-score of a predicted model signifies the confidence level. A) Predicted models of ApxIA. B) Predicted models of ApxIIA. C) Predicted models of ApxIIIA.

A



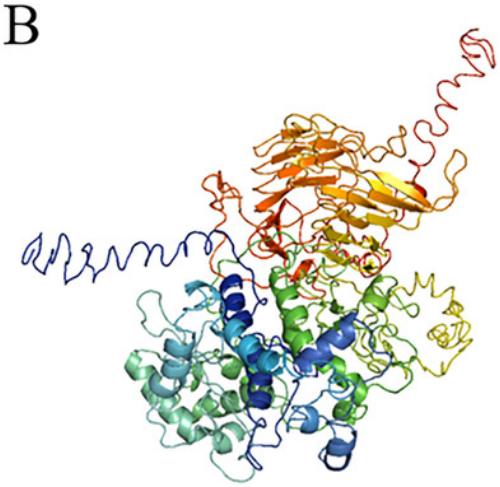
C-score = -2.66
TM-score = 0.41±0.14



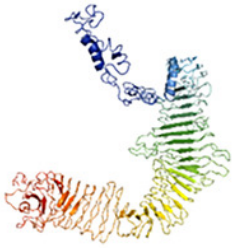
C-score = 2.71

C-score = -2.87

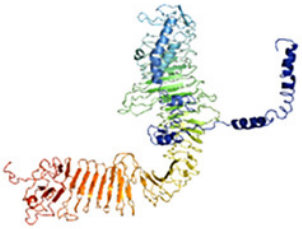
B



C-score = -2.77
TM-score = 0.4±0.13



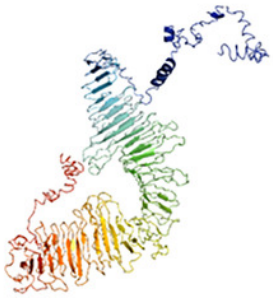
C-score = -3.06



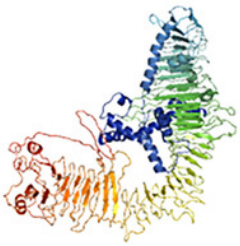
C-score = -3.14

C

C-score = -1.88
TM-score = 0.49±0.15



C-score = -2.95



C-score = -1.51

Figure 7

Predicted3-D models of individual domains that comprise ApxIA, -IIA, -IIIA and -IVA exotoxins.

The ApxA exotoxins were parsed into individual domains based on the domain boundaries computed by Dompred. Each of domains is given a code for convenience purpose, first numbers indicating the type of ApxA toxin and last number indicating the domain within the corresponding toxin. Partial structures are organized from (A) to (X).

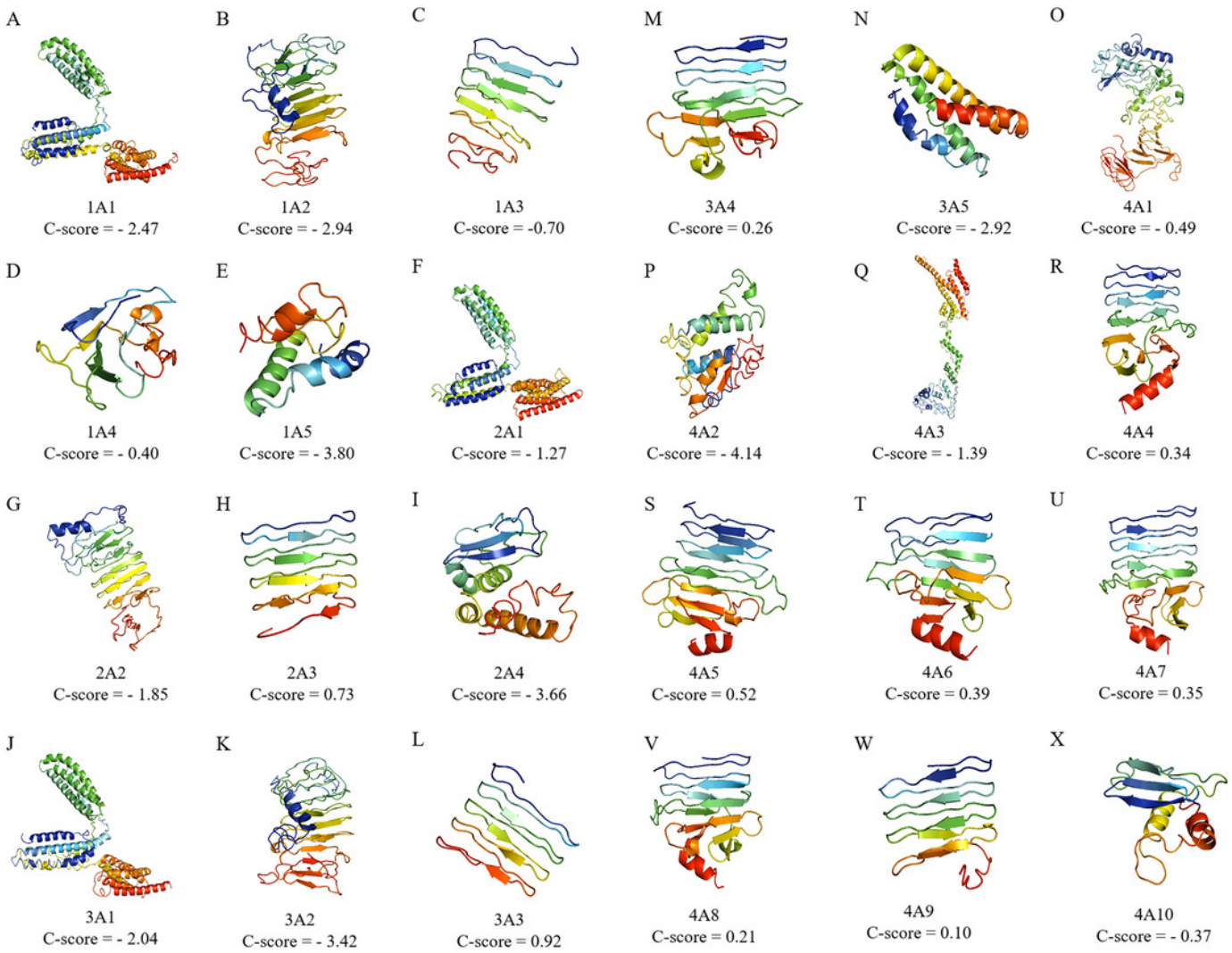


Figure 8

Localfrustration of ApxIA, -IIA and –IIIA exotoxins.

Frustratometer plots are color coded to display minimally frustrated and highly frustrated regions of the proteins. A) Local frustration plot for ApxIA. B) Local frustration plot for ApxIIA. C) Local frustration plot for ApxIIIA.

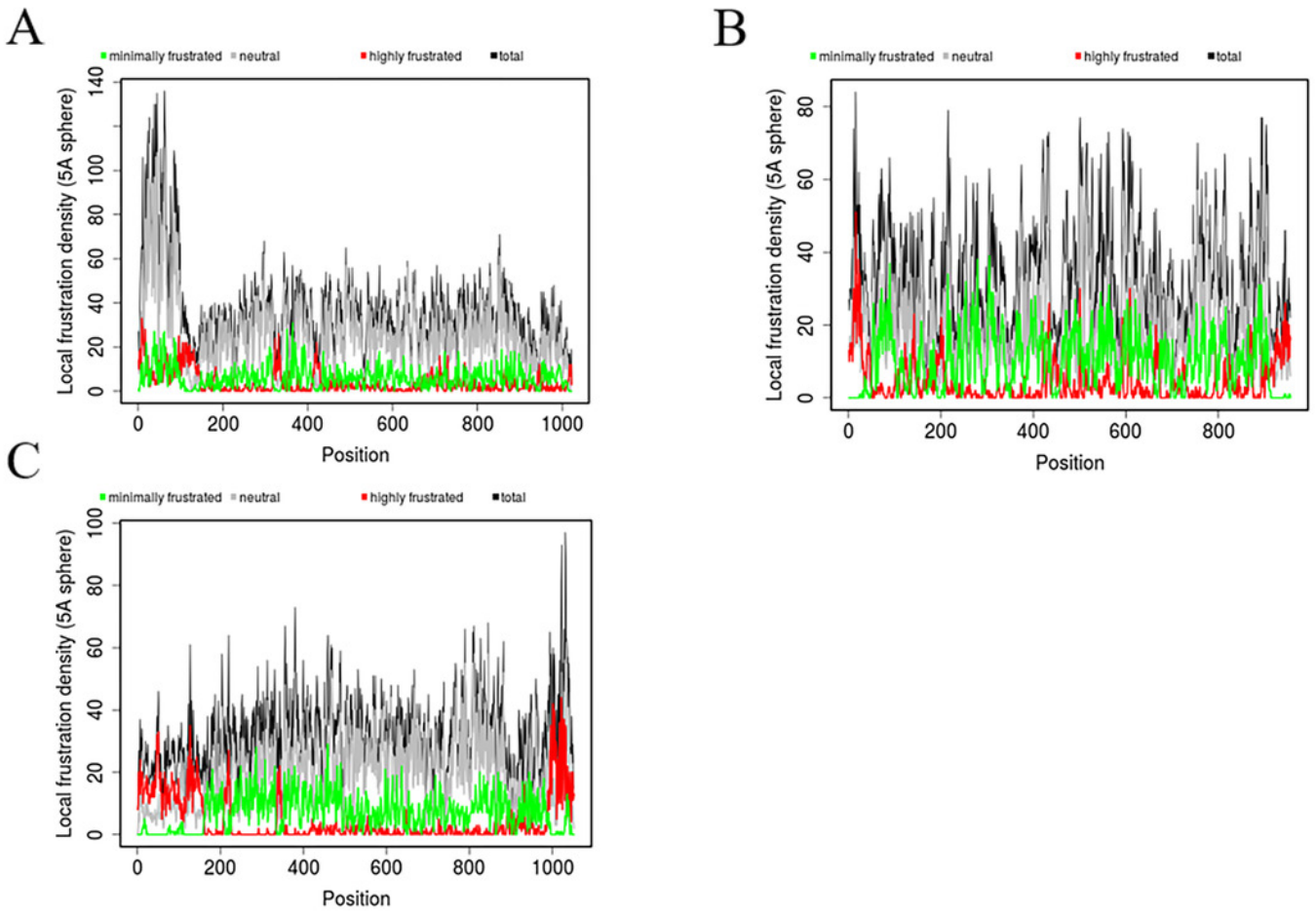


Figure 9

Z-score plots of predicted models ofApxIA, -IIA and –IIIA by ProSA.

Values for each structure are displayed in plots which contain the z-scores of all experimentally registered proteins in PDB. X-ray and NMR sources are distinguished by colors in z-score plots. The z-score plots can be used as references to see where input structures lie in respect to z-scores compared to native proteins of similar sizes typically range.