The Pharma Innovation Journal 2018; 7(9): 67-71

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2018; 7(9): 67-71

www.thepharmajournal.com
Received: 07-07-2018
Accepted: 09-08-2018

Shruti Mishra
Department of Computational Biology & Bioinformatics, Jacob Institute of Biotechnology & Bioengineering, Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, Uttar Pradesh, India

Prashant Ankur Jain
Department of Computational Biology & Bioinformatics, Jacob Institute of Biotechnology & Bioengineering, Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, Uttar Pradesh, India

Raghvendra Raman Mishra
Department of Medical Laboratory Technology, Deen Dayal Upadhyay Kaushal Kendre, Banaras Hindu University (BHU)Varanasi, U.P India

Correspondence
Prashant Ankur Jain
Department of Computational Biology & Bioinformatics, Jacob Institute of Biotechnology & Bioengineering, Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, Uttar Pradesh, India

Antigenic sites characterization and *In Silico* annotation of the S*taphylococcus aureus* Enterotoxins

Shruti Mishra, Prashant Ankur Jain and Raghvendra Raman Mishra

Abstract
Food poisoning is one of the sensitive infection with a rapid spread and bulk victims in one epidemic. Though seems to be a small problem it would result in a major effect of high concern. Food poisoning is caused by the enterotoxins secreted by the bacterium *staphylococcus aureus.* The toxins are of various types and cause a random infection which makes the treatment less effective and needs the diagnosis of the type of enterotoxin causing the specific symptom. Owing to the importance of *staphylococcus aureus* and its role in food poisoning, the current work was undertaken to analyse the Sequence, Structural and Antigenic characteristics of the 5 enterotoxin coding genes SEA, SEB, SEC, SED and SEE. Various *In silico* tools and software have been used to perform the analysis. This data can be employed further for the development of a drug or enterotoxin specific therapy.

Keywords: Snail, bovine, porcine, physicochemical properties, mucin, mucoadhesives

Introduction
Staphylococcus aureus is a human major pathogen known to produce a wide array of enterotoxins causing intestinal infections. Staphylococcal enterotoxins called SE's are of nine major serotypes of heat stable proteins being a major cause of gastrointestinal infections [1]. These enterotoxins are released by the bacteria into its medium thereby causing a poisonous effect. The major enterotoxins include SEA, SEB, SEC, SED and SEE. These are super antigens provoking the nonspecific Tell activation and proliferation in the immune system. These enterotoxins bind to class II MHC molecules on antigen presenting cells and stimulate large populations of T cells which share variable regions on the β chain of the T cell receptor [2]. Each toxin is coded by a different gene and has its unique sequence. These are known of their structural and phylogenetic similarities. Several *In silico* tools have been employed to characterize all the five enterotoxin coding genes.
NCBI, a database for genes, proteins [4] and other research has been used in the current work. Tools like Blast, Clustal Omega (5), Emboss Antigenic [6, 7] etc have been used to analyse the sequences and compare their characteristics. In order to analyse the structural characteristics of the enterotoxins tools like Protparam, SOPMA and Phyre have been used. To study the antigenic property of the peptides tools like Antigenic Emboss and PVS have been employed. 3D Structural visualization is done using Rasmol [8] visualization software. Based on the results of the above tools and software the final summary has been made.

Materials and Methods
Protein Sequence Retrieval from NCBI
Protein Sequence Retrieval of all the Enterotoxins SEA, SEB, SEC, SED and SEE using NCBI database. Along with the gene sequence the length of the sequence has also been noted for further analysis. The search term was limited to the enterotoxin proteins of.
Staphylococcus aureus only. Several sequences of the same toxin were obtained among which the sequence with correct length were saved. The sequences of all the 5 enterotoxins were collected.

Multiple sequence alignment using Crustal Omega
In order to detect the sequence similarity that may exist among the 5 sequences Multiple Sequence Alignment using Clustal omega was performed and the results were studied. This step is required to analyse the sequence similarity of any among the five enterotoxins so as to study their relation. This tool is available at EBI website. The clustal tools are a series of

multiple sequence alignment tools of bioinformatics available at various sites working of different principles. The MSA is used to analyze the evolutionary, sequence level, structural or functional relationship. This tool employs the seeded guide trees and HMM profile-profile techniques to generate alignments.

Identification of Foreignness using BLASTP

In order to identify the similarity between the enterotoxin proteins and the human proteome BLAST analysis has been performed. BLAST is an online tool provided by NCBI database which identifies the degree of similarity between any two user submitted sequences. It is of several types based on the purpose like Blastn, Tblasty, Blastp etc. In the current study Blastp has been used as both the sequences under comparison are proteins. BLAST can be directly used from NCBI webpage. The result are displayed in Graph, Table and Description methods. All those proteins that do not share any similarity are also mentioned in the results. Such result indicates a foreign protein.

Antigenic Site Identification using Emboss Antigenic from EBI

Antigenicity of any peptide is because of its specified regions on the peptide known as antigenic sites. The antigenic sites on the peptide region has to be identified so as to prepare them either for vaccine development or to target them for docking with a drug. Identification of these sites is performed by an *In silico* tool called EMBOSS ANTIGENIC. This would locate all the sites possible on a given protein sequence. It would first classify the antigenic regions available for a given protein within which the sites are marked. Kolaskar and Tongaonkar method is employed by the tool to predict these sites of antigenicity.

Structural Analysis of Proteins

Protparam: is an online tool used to characterize the protein both physically and chemically so as to analyse its properties. The tool would use a protein sequence as an input and perform all the calculations using the same sequence so as to calculate the properties like Sequence length, Molecular weight, isoelectric point, Hydropathi city, Stability etc. The tool is available at Expasy [9] server.

Sopma: is a tool for the characterization of proteins secondary structure so as to summarize the structural conformations within the protein. It would provide the information like the total types of conformations possible for the input protein and the number of each type of amino acid participating in a different structural confirmation. It enables the user to identify the conformation of each and every amino acid in the submitted protein. It would provide a complete structural insight about the protein under study. Tool is provided by the Expasy software.

Phyre: is another tool that would work similar to the BLAST search, however the comparison preformed here are for the identification of PDB ID'S of the protein. The input for the tool being the protein sequence, the output delivered is the list of PDB ID's along with the other structural details of each entry. This tool can be specially sued to identify the proteins 3D structure based on the comparison between the sequences in the database and the user entered sequence. The 3D structure can be downloaded using from PDB databank using the ID provided by phyre.

Rasmol Visualization

The final 3D structure of the protein downloaded from the PDB database can be visualized using a special software called visualization software, one of which is Rasmol. The software can be downloaded from its website and installed. The 3D structure of the required protein can be studied in the Rasmol using its PDB structure script. The software is used to analyze different type of secondary structures present in the protein, the ligands if any, the co-factors of the protein if present, the polar and non-polar regions etc. However the software is limited to visualization and cannot be used for manipulation or editing the structure.

Results and Discussion

Protein Sequences of SEA, B, C, D and E were retrieved from NCBI and were used for all the other analysis. It was found that all the sequences have different number of amino acids as shown in the table below.

Table 1: Sequence length of enterotoxins

Type of Enterotoxin	Length
SEA	146aa
SEB	266aa
SEC	266aa
SED	207aa
SEE	154aa

Multiple Sequence alignment and Phylogenetic Study

Clustal Omega was used to study the sequence similarity among the 5 sequences which was further used for the development of dendrogram to study the phylogenetic relation among the same. The figure 1(a) shows the Multiple Sequence alignment and 1(b) shows the evolutionary relationship by Dendrogram

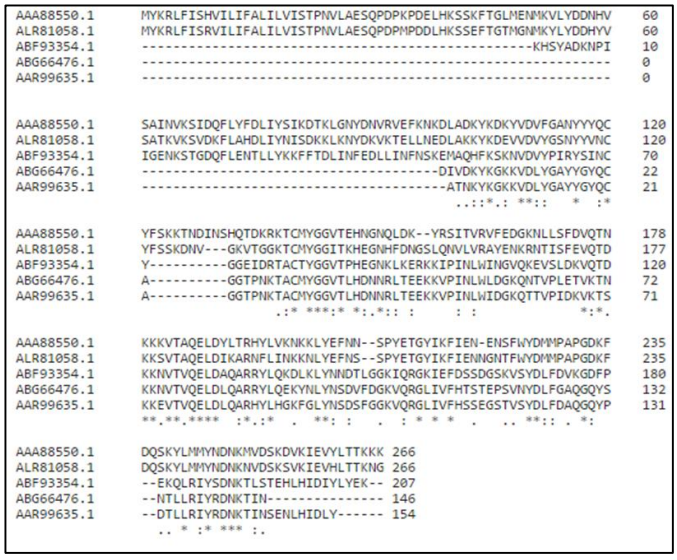

Fig 1(a): Showing sequence similarity and Conservation

An * (asterisk) indicates position which has a single, fully conserved residue

A. (point) indicates conservation between groups of weakly similar properties

A: (colon) indicates conservation between groups of strongly similar properties

The above alignment shows a less degree of sequence similarity among the selected five sequences, however it also shows the regions sharing similar properties

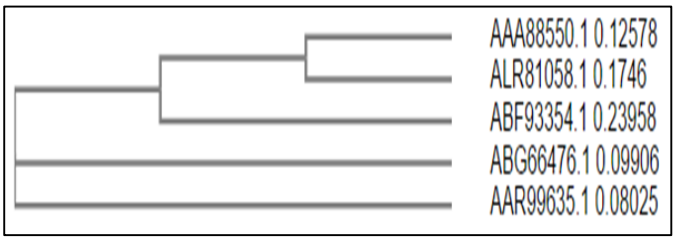

Fig 1(b): Showing Evolutionary relationship among sequences

In the above dendrogram the accession number used are as follows:
AAA88550: Enterotoxin B, ALR81058: Enterotoxin C, ABF93354: Enterotoxin D, ABG66476: Enterotoxin A, AAR99635: Enterotoxin E
From the above Sequence alignment and phylogenetic tree it can be observed that SEA and SEE share maximum similarity among them. However all the other sequence share considerably less degree of similarity.

BLASP for Foreignness Identification
In order to be considered as antigen it is important that the human body considers the peptide as foreign so that it can subsequently produce antibody against it. The results of BLAST P indicated that all the 5 proteins share no similarity to the human proteome. Thus all are considered to be foreign to the human.

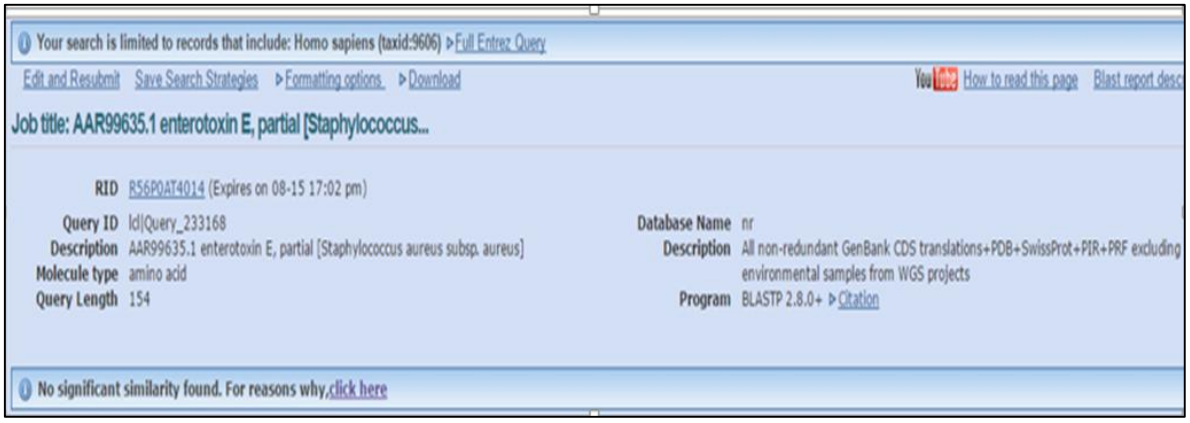

Fig 2: Showing the Sequence not sharing any similarity to human proteome as calculated by BLASTP

Identification of Antigenic Sites in the Protein
Using the EMBOSS ANTIGENIC tool all the antigenic sites present in the protein are identified and tabulated below.

Table 2: Total Antigenic regions and sites as shown by

Protein	Number of Antigenic Regions	Antigenic Regions	Antigenic Sites
SEA	8	106-114, 8-24, 63-69, 50-56, 75-88, 31-40, 117-129, 133-138	108, 20, 67, 54, 79, 34, 122, 137
SEB	7	4-29, 106-123, 252-263, 182-199, 52-80, 170-175, 158-164	13, 119, 258, 193, 75, 173, 164
SEC	6	4-29, 105-123, 155-163, 252-262, 54-80, 180-187	13, 119, 157, 258, 75, 183
SED	10	195-204, 58-73, 109-117, 23- 30, 165- 178, 77- 87, 123- 130, 138- 144, 181- 187, 38- 44,	201, 62, 117, 29, 170, 82, 127, 140, 185, 44
SEE	8	105-113, 7-23, 74-92, 61- 68, 116-129, 49-54, 30-39, 131-137	111, 19, 85, 66, 121, 53, 33, 132

Emboss Antigenic
The above table clearly depicts the total list of antigenic regions possible in each Enterotoxin and the antigenic sites within these regions. SED was found to have maximum number of antigenic sites.

Structural analysis of the enterotoxins
The five selected enterotoxins being protein in nature were analysed for their Primary, Secondary and tertiary structure. The primary structure analysis done by Protparam yielded the information related to their physicochemical behaviour.

Table 3: Physicochemical properties of the enterotoxins

	Molecular Weight	PI Value	No of negatively charged aa	No of positively charged aa	Instability index	Aliphatic Index	Hydropathi city
SEA	16710.84	8.94	16	20	33.95	78.01	-0.729
SEB	31435.87	8.65	36	40	34	74.66	-0.664
SEC	30628.78	8.22	34	36	35.09	75.04	-0.588
SED	24001.09	7.84	31	32	23	80.48	-0.781
SEE	17390.57	8.73	17	20	35.54	75.91	-0.653

From the above table of physicochemical parameters of all the proteins it can be concluded that except for the length of the peptides all the other properties are nearly the same. The negative Hydropathi city renders them hydrophilic in nature.
Secondary Structure Prediction:
SOPMA has been used for secondary structure analysis of enterotoxins wherein the structural motifs like the alpha helixes, extended strands, beta sheets, random coils, turns, loops etc are analyzed and results were summarized as follows:

```
SOPMA :
   Alpha helix      (Hh) :     23 is  15.75%
   3₁₀ helix        (Gg) :      0 is   0.00%
   Pi helix         (Ii) :      0 is   0.00%
   Beta bridge      (Bb) :      0 is   0.00%
   Extended strand  (Ee) :     55 is  37.67%
   Beta turn        (Tt) :     12 is   8.22%
   Bend region      (Ss) :      0 is   0.00%
   Random coil      (Cc) :     56 is  38.36%
   Ambiguous states (?)  :      0 is   0.00%
   Other states          :      0 is   0.00%
```

3a) SEA

```
SOPMA :
   Alpha helix      (Hh) :     80 is  30.08%
   3₁₀ helix        (Gg) :      0 is   0.00%
   Pi helix         (Ii) :      0 is   0.00%
   Beta bridge      (Bb) :      0 is   0.00%
   Extended strand  (Ee) :     80 is  30.08%
   Beta turn        (Tt) :     24 is   9.02%
   Bend region      (Ss) :      0 is   0.00%
   Random coil      (Cc) :     82 is  30.83%
   Ambiguous states (?)  :      0 is   0.00%
   Other states          :      0 is   0.00%
```

3b) SEB

```
SOPMA :
   Alpha helix      (Hh) :     83 is  31.20%
   3₁₀ helix        (Gg) :      0 is   0.00%
   Pi helix         (Ii) :      0 is   0.00%
   Beta bridge      (Bb) :      0 is   0.00%
   Extended strand  (Ee) :     76 is  28.57%
   Beta turn        (Tt) :     27 is  10.15%
   Bend region      (Ss) :      0 is   0.00%
   Random coil      (Cc) :     80 is  30.08%
   Ambiguous states (?)  :      0 is   0.00%
   Other states          :      0 is   0.00%
```

3c) SEC

```
SOPMA :
   Alpha helix      (Hh) :     41 is  19.81%
   3₁₀ helix        (Gg) :      0 is   0.00%
   Pi helix         (Ii) :      0 is   0.00%
   Beta bridge      (Bb) :      0 is   0.00%
   Extended strand  (Ee) :     65 is  31.40%
   Beta turn        (Tt) :     11 is   5.31%
   Bend region      (Ss) :      0 is   0.00%
   Random coil      (Cc) :     90 is  43.48%
   Ambiguous states (?)  :      0 is   0.00%
   Other states          :      0 is   0.00%
```

3d) SED

```
SOPMA :
   Alpha helix      (Hh) :     23 is  14.94%
   3₁₀ helix        (Gg) :      0 is   0.00%
   Pi helix         (Ii) :      0 is   0.00%
   Beta bridge      (Bb) :      0 is   0.00%
   Extended strand  (Ee) :     57 is  37.01%
   Beta turn        (Tt) :     11 is   7.14%
   Bend region      (Ss) :      0 is   0.00%
   Random coil      (Cc) :     63 is  40.91%
   Ambiguous states (?)  :      0 is   0.00%
   Other states          :      0 is   0.00%
```

3e) SEE

Fig 3: Secondary structure analysis of the five enterotoxins

The above results of secondary structure prediction indicate that in SEA, SED and SEE Extended strands were far more than alpha helices while in SEC helices were more than strands. In SEB both of them were equally distributed.

Tertiary structure analysis and visualization

Phyre was used for the tertiary structure analysis of the enterotoxins and it was found that all five of them were sharing a common PDB IDi.e1sxt indicating that all the 5 proteins share a very high structural similarity corresponding to a common PDB ID**.** The structure was downloaded and viewed in Rasmol Visualization tool.

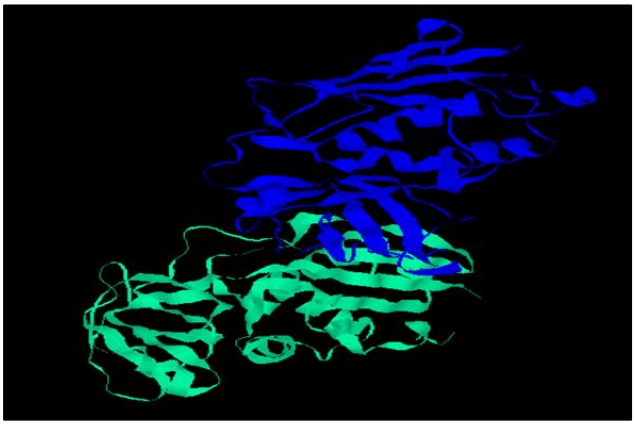

Fig 4: Shows the 3D structure of the id 1sxt in Rasmol

Conclusion

The major focus of the work was to identify the antigenic sites among the 5 enterotoxins of *Staphylococcus aureus* that imparts its toxicity. *In silico* analysis was performed at the sequence and structural level using various tools and software. BLASTP and Clustal Omega were used to identify the evolutionary relation among the 5 proteins which showed a very close relationship for all the 5. The structural analysis of these proteins revealed that though the sequences vary in length their properties like IP, Stability, Hydropathi city etc all remained nearly the same. The Phyre tool used to identify the PDB ID corresponding to the sequences of the proteins showed that all the 5 can be represented by the single ID of 1SXT. Rasmol software was further used to visualize the structure of 1sxt. The final conclusion obtained firm the analysis is that though all the 5 protein sequences of enterotoxins differ they all share similarity in their physicochemical and structural properties.

References

1. https://www.sciencedirect.com/science/article/pii/S0168160500003779
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3153290
3. http://www.bepls.com/bepsljan2014/3.pdf
4. Romiti M, Entrez Nucleotide, Entrez Protein FAQs. 2006 Updated 2010 May 12. In: Entrez Sequences Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US), 2010.
5. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011; 7:539.
6. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. Febs Letters. 1990; 276:172-174.
7. Parker JMR, Guo D, Hodges RS. Biochemistry. 1986;

25:5425-5432.
8. Roger Sayle, E James. Milner-White. Ras Mol: Biomolecular graphics for all", Trends in Biochemical Sciences (TIBS), September, 1995; 20(9):374.
9. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. Ex PASy: the proteomics server for in-depth protein knowledge and analysis Nucleic Acids Res. 2003; 31:3784-3788