Signature Pattern Mining of Type VI Effector Proteins

Rishika Sen, Losiana Nayak, Rajat K. De
{rishikasen_r,losiana_t,rajat}@isical.ac.in
Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, West Bengal

1. Introduction to the Problem

- Effector proteins of bacteria infect their hosts by secretion systems (SS).
- Six ([T1-T6]SS) such secretion systems have been identified in gram negative bacteria [1].
- Effector proteins of the T6SS of many species are yet to be discovered.
- No signature pattern have yet been discovered to differentiate T6 effector proteins from the non-effectors ones.
- Such signature patterns will speed the discovery of T6 effector proteins in many gram-negative bacteria.

2. Secretion Systems

- Pathogens have developed numerous ways of transporting protein between locations via secretion apparatus.
- Bacterial secretion apparatuses (systems) can be divided into different classes, based on their structures, functions, and specificity.
- Protein secretion systems are essential for the growth of bacteria and are used in an array of processes.
- Secretion systems are used by bacterial pathogens to manipulate the host and establish a replicative niche.

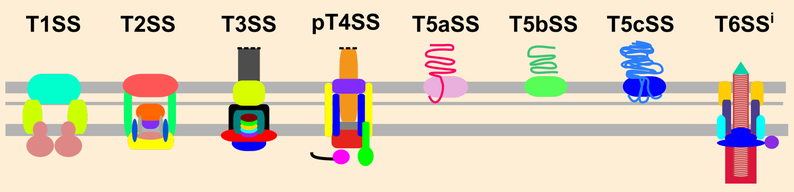


Figure 1: Schematic diagram of the secretion systems [2]. Gram negative secretion can be split roughly into two classes: Sec dependent and Sec independent. Sec is a highly conserved membrane protein found in every living cell (including, of course, humans) where it carries out the majority of protein translocation across membranes. Under sec-dependent class falls the T2SS and the T5SS. Under sec-independent class falls the T1SS,T3SS and T4SS.

3. Type VI secretion System

T6SS is a recently identified secretion system. This system facilitates transport and transfer of various proteins required for bacterial growth and infection in host environment. The arrangement of the component proteins of T6SS is depicted below.

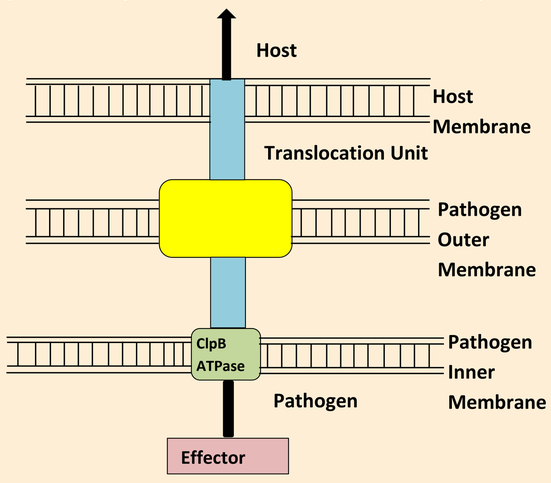
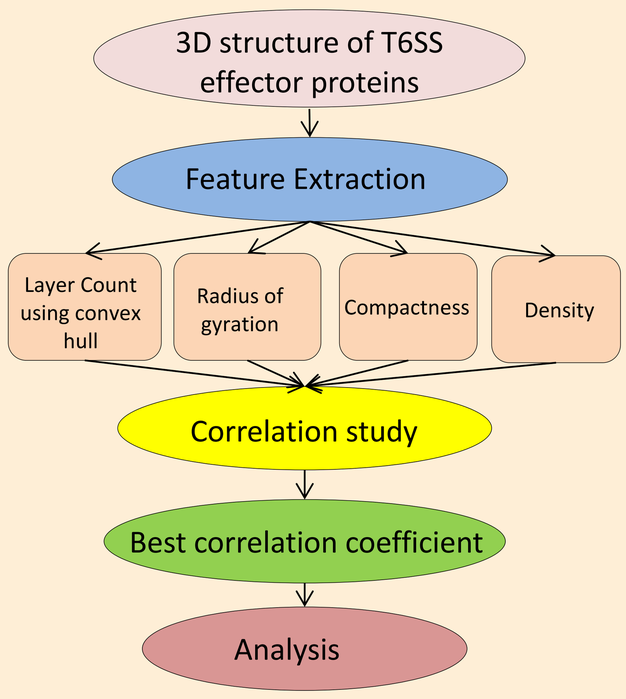


Figure 2: Systematic view of T6SS

4. Methodology



5. Basic Concept of Convex Hull Theory

In our work the 3D structure of the established T6 effector proteins have been analyzed. Some of the analyzes have been carried out by applying convex hull theory [3]. Convex hull is used to explore the interaction interfaces of proteins. The convex hull of a set of points S in the Euclidean plane or in a Euclidean space (or, more generally, in an affine space over the reals) is the smallest convex set that contains S. A convex hull of a finite point set S is defined by the formula:

$$Conv(S) = \{\sum_{|S|}^{i=1} \alpha_i x_i | (\forall i : \alpha_i \geq 0) \wedge \sum_{|S|}^{i=1} \alpha_i = 1\}$$

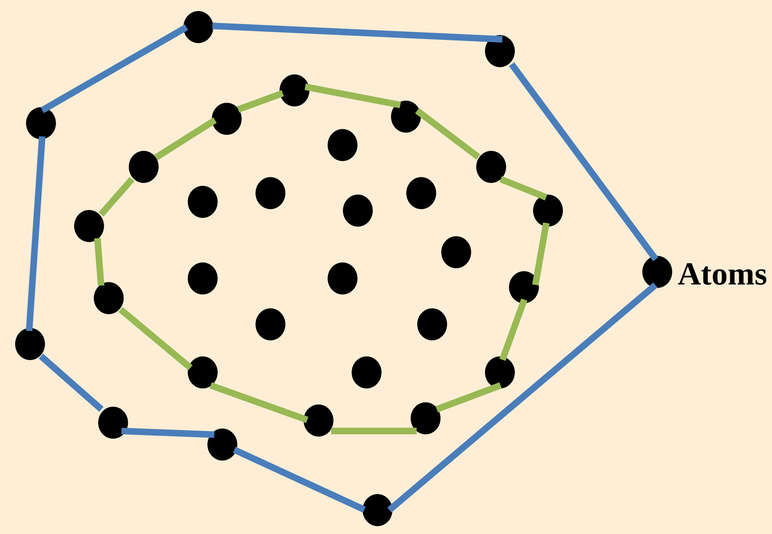
where α_i is the weight assigned to the point x_i, in such a way that the weights assigned are all non-negative and sum to one, and these weights are used to compute a weighted average of the points. For each choice of the coefficients, the resulting convex combination is a point in the convex hull, and the whole convex hull can be formed by choosing coefficients in all possible ways.

6. Feature Extraction from Effector Proteins

The component proteins of secretion systems, the transferred proteins and any other bacterial protein component like the flagellar proteins which help in effecting bacterial infection in host, are known as effector proteins. In this respect we have considered the known effector proteins of a few species from an established T6 effector protein database and have tried to find a signature pattern that would differentiate the effector proteins from the non-effectors ones.

Four features have been extracted, among which the 'Layer count' and the 'Radius of Gyration' of proteins are of significance. A protein can be perceived as a graph with its atoms as vertices and the bonds among the atoms as edges. Such a graph (protein graph) has been used with application of convex hull theory to generate the following features.

- Layer count - It is defined as the number of possible convex hull layers that can be formed in a protein graph. Starting from the initial protein 3D structure, a convex hull is formed, and then the vertices that formed the convex hull, are removed, and again the next layer is formed form existing vertices. The process continues till no more points exist to form a convex hull. If S is the set of all vertices in a protein, then the next set of vertices on which convex hull can be formed is $S^1 = S - Conv(S)$, and so on $S^2 = S^1 - Conv(S^1)$, ..., $S^x = S^{x-1} - Conv(S^{x-1})$, where x is the number of convex hull layers.



The points connected by the blue lines constitute the first set of points forming the first convex hull, '$Conv(S)$'. Similarly, the points connected by the green lines constitute the second set of points forming the second convex hull, '$Conv(S^1)$'.

- Radius of gyration - Radius of gyration is generally calculated for finding a central point of a protein. Radius of gyration (Rg) provides quantitative estimates for the dimensions of the protein. The radius of gyration of a protein is calculated by using the formula

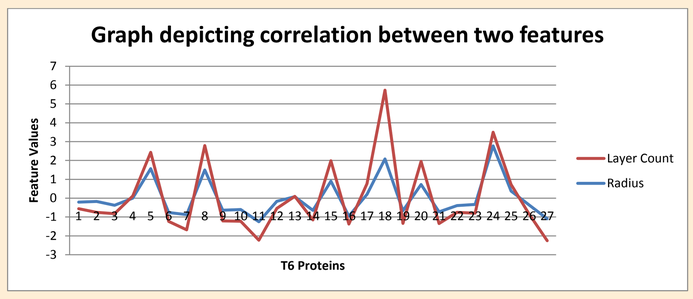
$$(X, Y, Z)_{centroid} = \left(\frac{\sum_i^n (x_i)}{n}, \frac{\sum_i^n (y_i)}{n}, \frac{\sum_i^n (z_i)}{n} \right)$$

where x_i, y_i, z_i are the coordinates of each of the atoms.

7. Results

Proteins	Features			
PDB ID	Radius	Layer Count	Compactness	Density
2IOY	28.17	75	1.386	0.04553
3EAA	28.66	57	1.202	0.02314
3ESW	25.94	68	1.508	0.04328
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
5TGL	16.41	10	1.489	0.01543

8. Analysis of some Results



A high correlation coefficient of 0.84 was noticed among the features Layer count and Radius of gyration.

9. References

[1] T. R. Costa, C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter, and G. Waksman: *Secretion systems in Gram-negative bacteria: structural and mechanistic insights*, Nature Reviews Microbiology (2015).

[2] Abby, Sophie S and Cury, Jean and Guglielmini, Julien and Néron, Bertrand and Touchon, Marie and Rocha, Eduardo PC: *Identification of protein secretion systems in bacterial genomes*, Scientific reports(2016)

[3] Dafas, Panos and Bolser, Dan and Gomoluch, Jacek and Park, Jong and Schroeder, Michael: *Using convex hulls to extract interaction interfaces from known structures*, Bioinformatics (2004).

Acknowledgements

Losiana Nayak, one of the authors, acknowledges University Grants Commission (`http://www.ugc.ac.in/`), India for a UGC Post-Doctoral Fellowship [No. F.15-1/2013-14/PDFWM-2013-14-GE-ORI-19068(SA-II)].

Author's Comments

The results shown here are part of an ongoing project. Their full analysis are yet to be done. Thank you for your interest in our poster.