Insilico Studies on HA 33 Subunit of Haemagglutinin of Clostridium botulinum – Construction of Three Dimensional Model and Validation

Nirmal Rajah.R
Independent researcher
PO.BOX.62226, Dubai, United Arab Emirates

L.Rufus Auxilia
Research scholar,
Department of Botany and Microbiology
Lady Doak College, Madurai

ABSTRACT

Botulinum neurotoxin (BoNT) complexes consist of neurotoxin and neurotoxin-associated proteins. The Clostridia releases these as protein aggregate, termed as the progenitor toxins. Heamagglutin – 33 (HA-33) plays an essential role in the agglutination activity of the botulinum progenitor toxin; it assists in binding and translocating the progenitor toxins through the alimentary epithelial barriers. A three-dimensional homology model of HA-33 is demonstrated in this paper based upon the crystal structure of HA1 subcomponent of botulinum type C progenitor toxin complexed with galactose. The three dimensional model was designed using MODELLER (9v11) a comparative modelling tool that built a model based on an alignment of a sequence of HA-33 based on the known structure of botulinum type C progenitor toxin complexed with N-acetylgalactosamine. The modelled structure was then evaluated using SAVES. The structural dynamics of the modelled protein was performed by elNemo and based on the scores obtained, it was found to be a reliable and accurate three dimensional structure of the Hemeagglutin component. It gives an idea about the three dimensional structure of a protein which will would serve as a guide to a detailed understanding of its participation in the process of agglutination. This could also aid in molecular drug designing against botulism.

Key Words

Clostridum botulinum, homology modelling, HA-33.

1. INTRODUCTION

Clostridium botulinum is a gram negative, spore forming [22] anaerobic [8] rod shaped food poisoning bacteria, causing Botulism a fatal paralytic illness. Clostridium botulinum strains produce immunologically distinct neurotoxins from A to G [10].The Botulinum Neurotoxins(BoNTs) are the most potent toxins known

in nature, they act by inhibiting the neurotransmitters released from cholingeric synapses [11]. The neurotoxins associate with nontoxic components in culture fluids to form larger complexes termed as the progenitor toxins [10]. The toxin enters the body in one of any four ways. 1) Colonizing of the bacteria in the digestive tracts of children (Infant Botulism) 2) Adult (Adult intestinal Toxemia) 3) by ingestion of toxin from food stuffs (Food Borne botulism) or 4) by contamination of a wound by the bacterium [11].

Haemagglutinin(HA) are found with the progenitor toxins of serotypes A,B,C and B, but absent in E and F[8]. Botulinum toxin consists of three HA subcomponents HA-70, HA33 and HA17, labelled as HA with their molecular mass (HA- 33 kda). HA 33 component selected for this paper assists chiefly in binding to the microvilli and translocates the progenitor

toxins through the alimentary epithelial barrier, hence facilitating an efficient absorption of the toxin [22].

HA33 was selected as the protein of interest for Homology modelling to get a better understanding of HA complexes with respect to the Botulism and related mechanisms of toxin action. Homology modelling is the three dimensional modelling of a target protein with the template amino acid sequences.

A structure of a protein is determined by its amino acid sequence [4]. Knowing the sequence would enable to obtain its structure, and during the course of evolution the structure of the protein changes very slowly than its associated sequence, resulting in similar structures adopting similar sequences [2] [20]. With this rule the sequence of HA-33[17] was compared with an identical sequence of HA1 (HA33) mutant F1791 of botulinum type C progenitor toxin complexed with N-acetyl galactosamine, bound at site II (PDB ID 3Aj6) with a known structure obtained through X ray diffraction [15]. Based on the high sequence similarity of 74 % and structure of 3Aj6 as an approximation to the structure of HA 33, the homology modelling was done using MODELLER 9v11. Results were validated using different online servers and the physiochemical properties and secondary structures were determined.

2. MATERIALS AND METHODS
2.1 Template selection

The HA-33 component of botulinum toxin protein sequence was retrieved from NCBI (http://www.ncbi.nlm.nih.gov/protein/BAB71747.1), in a FASTA format [13] the 285 amino acid long protein with a molecular mass of 33kda was entered under the accession number GenBank: BAB71747.1.

2.2 Protein family prediction

Selected protein(HA33) family was identified using an online server pfam (pfam.sanger.ac.uk). The results are listed in the table.1

2.3 Protein secondary structure prediction

A protein's secondary structure includes helixes, turns, coils and sheets. SOPMA, an Expasy online server was used to predict the percentage of these secondary structures present in the HA 33 protein and listed out in the table.2

2.4 Physio chemical properties prediction

Expasy server also provides a tool called "PROTPARAM", using this physiochemical properties like molecular weight, isoelectric point, extinction coefficient, estimated half life, instability index, aliphatic index and Grand Average of Hydropathicity (GRAVY) were predicted and tabulated in table.3

2.5 Target selection

The sequence of BAB71747.1 was assigned as the query sequence and a template sequence was identified through the BLAST program against PDB (Protein Data Bank). The resulting hits of sequences were analyzed for the high amount of sequence similarity and the sequence of the protein, HA1 (HA33) mutant F1791 of botulinum type C progenitor toxin complexed with N-acetylgalactosamine, bound at site II (PDB ID 3Aj6) was identified with a 74% sequence similarity, resolution of 1.48 Å, the sequence of the template was retrieved as an approximation to the structure of HA-33.

2.6 Three dimensional modelling

The sequence thus obtained was used to construct a model for HA-33 using MODELLER 9V11, a comparative protein structure modelling software [19] under Windows operating system. The constructed model was saved as a PDB file and was subjected to assessment.

2.7 Model Validation

Every homology model contain errors [9], depending either on the sequence identity, where the rms error in the modelled coordinates can be large with considerably large local errors or with the errors based on the templates used. Hence verification of the model is proved to be very essential in the homology modelling process. The resultant model was subjected to examination to prove its accuracy.

The overall stereo chemical quality of the model was assessed using the Ramachandran plot analysis, using the computer program PROCHECK, which provides a detailed check on the stereochemistry of a protein structure outputs comprising a number of plots in PostScript format and a comprehensive residue-by-residue listing. In the Ramachandran Plot [16] analysis, the residues were categorized according to its regions in the quadrangle. The Ramachandran plot shows the phi-psi torsion angles for all residues in the structure (except those at the chain termini). All the residues are represented as squares except Glycine residues which are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other side chain types [18]. The graph for Ramachandran plot was obtained and results were discussed.

VERIFY 3D [1][14] is a protein structure verification computer program that analyzes the compatibility of a Three dimensional atomic model with this own amino acid sequence. Each residue of the protein is assigned a structural class based on its location and environment (Alpha, Beta, Polar, Non-polar etc). As a reference a set of good known structures are used and a score for each of the twenty amino acids in this structural class are obtained. The scores of a sliding 21 residue window from (-10 to +10) are added and plotted for each individual residues.

ERRAT is a protein structure verification program written in FORTRAN which uses an algorithm to differentiate between correctly and incorrectly determined regions of protein structure based on characteristic atomic interactions, ERRAT considers distributions of non-bonded atoms according to atom type and distance, and the energy strain method [3]. The Error values are plotted as a function of the position of a sliding 9-residue window. The error function is based on the statistics of non-bonded atom-atom interactions in the reported structure comparing to a database of reliable high-resolution structures.

elNemo – the Elastic Network Model is an online tool that calculates the B-Factor of a structure [21], elNemo is used to calculate the mobility of the atoms present in the structure and to predict the dynamics of the protein.

3. RESULTS AND DISCUSSION

Table: 1 Protein Family Prediction

Sl.no	Family	HMM	
		From	To
1.	Ricin type B trefoil lectin domain	**7**	124
2	Ricin type B trefoil lectin domain 2	13	104

Table 2 Secondary structure Prediction

Sl.no	Secondary Structure	Percentage
1.	Alpha helix	5.61%
2.	Beta Sheet	43.51%
3.	Turns	7.37%
4.	Coils	43.51%

Table:3 Physiochemical properties

Sl.no	Properties	Value
1	Molecular Weight	33039.8
2	Theoretical P^I	8.77
3	Extinction Coefficient	82975 82850
4	Half Life	30 hrs
5	Instability Index	22.16
6	Aliphatic Index	75.54
7	GRAVY	-0.593

Using the Hidden Markov Model (HMM) pfam searches the similar protein families from the databases. The amino acids from 7 to 124 and 13 to 104 have the similarity of Ricin type B trefoil lectin domain, Ricin type B trefoil Lectin 2 respectively**(table 1)** .The clan has 15 protein families including Agglutinin

SOPMA Server [6] predicted the secondary structures of the protein. In the **Table.2** Alpha helix (5.61%), Beta sheet (43.51%) and turns (7.37%), Coils (43.51%).It gives a clear idea of the percentage of secondary structures present in the HA33.

Protoparam is an Expasy server tool which computes the physiochemical properties of a protein sequence. HA 33 sequence was analysed using the tool and the results were tabulated in the **Table 3.** The calculated Molecular weight is 33039.8. P^I is Iso Electric Point. At particular P^H, the net charge of the protein will be Zero. In that P^H the protein will

be very stable and compact [18]. The $P^{I\ of}$ HA 33 is 8.77 which is slight alkaline. The computed extinction coefficient helps in the quantitative study of protein- protein and protein ligand interaction in solution [18]. The extinction coefficient is 82975 $M^{-1}cm^{-1}$ which consider all the cys as cysteine residue and 8285 $M^{-1}cm^{-1}$ which consider all the cys as reduced residues. The half life of the protein is 30 hours in mammalian reticulocytes, in vitro.

PROVE PLOT was used to calculate the atoms in the modeled structure of HA-33 and the Z- score mean was found to be 0.124, with a Z score standard deviation of 1.432 and a Z score RMS of 1.436 **(Fig.4)**
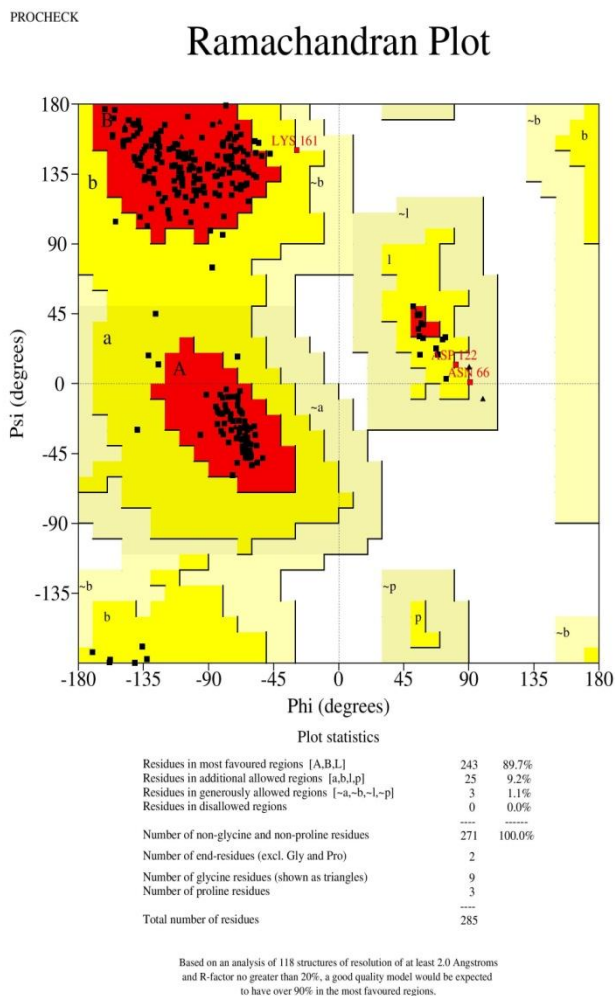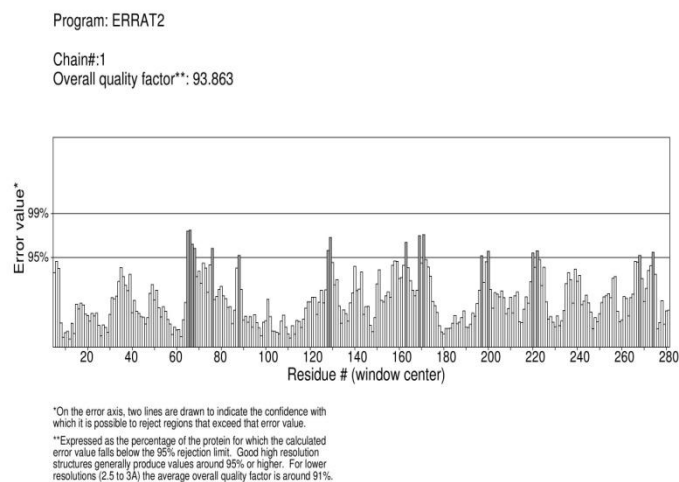

Fig: 1 Ramachandran Plot for the modelled protein HA33

Program: ERRAT2

Chain#:1
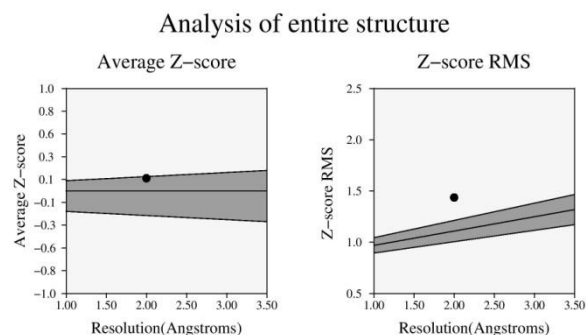Overall quality factor**: 93.863

*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3A) the average overall quality factor is around 91%.

Fig : 2 Errat Results page of HA-33.

Fig:3 PROVE Plot for the HA-33.
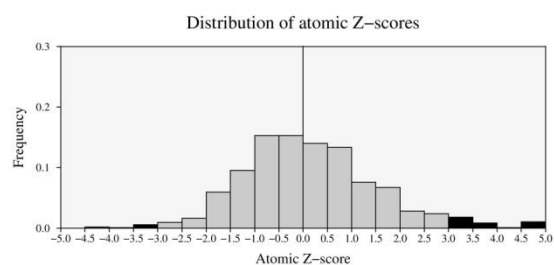
Z–score mean	0.124
Z–score stddev	1.432
Z–score RMS	1.436
# scored atoms	1217
# outliers	51
% outliers	4.200

Fig:4 Z scores of HA 33 obtained from the PROVE Plot.

Instability index explains whether the protein is stable or instable using a constant value 40. The instability index below 40 indicates that the protein is stable; whereas the index above 40 shows that the protein is instable [7]. The instability index of the HA33 is 22.16 which indicate that the protein is relatively stable. Aliphatic index which explains the thermal stability of a protein is calculated as 75.54 which reveal that the protein can withstand high temperatures to a certain

extent. GRAVY (Grand Average of Hydropathicity) value for a protein is calculated by the sum of the hydropahty values of all amino acids which is divided by the number of residues in the sequence [18]. The GRAVY value is -0.593. The value indicated the better interaction of protein with water

The protein quality was predicted through the LG Score which for HA-33 was 3.914 and with a maxsub score of 0.368 confirms that it's a reasonably good model.
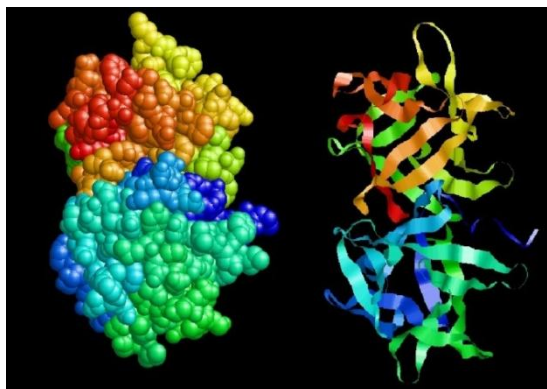

Fig 5.Predicted 3-D structure of HA-33 (Spacefill and ribbon representation)

The results (Fig:1) revealed that the modeled structure of HA-33, had 89.7% residues in the allowed regions and no residues in the disallowed regions, making it a good and reliable model. The results of VERIFY 3D for the modeled protein revealed that 89.86% of the residues had an averaged 3D-1D score greater than 0.2%. A single output plot is produced that gives the value of the error function *vs.* position of a 9-residue sliding window. By comparison with statistics from highly refined structures, the error values have been calibrated to give confidence limits. The ERRAT analysis **(Fig.2 &3)** for the modeled protein revealed that the overall quality factor of the protein as 93.863 comparing with its resolution 1.48 Å of making it a good quality model.

B factors calculated from the elNemo analysis shows that the B factor of the modelled protein as -0.105 for 285 c alpha atoms.
The modeled protein structure was visualized using RASMOL **(Fig.5)**

4. CONCLUSION

With the help of comparative modelling the three dimensional structure of the Hemagglutin component 33 (HA-33) was determined in the above study, the predicted protein structure was subjected to relevant tests to prove its accuracy and its validity as a structure with an accuracy that is comparable to the best results achieved experimentally. The model was proven to be precise. Considering the role of HA 33 its effect on the botulinum neurotoxin and the issue of the fatal food poisoning by the clostridia a better understanding of its structure would open the doors for future drug designing and targeting.

5. REFERENCES

[1] Bowie, J.U., Luthy, R., and Eisenberg,D. (1991). A method to identify protein sequences that fold into a known three dimensional structure. Science. 253(5016) 164-170

[2] Chothia,C and Lesk, A.M(1986). The relation between the divergence of sequence and structure in proteins . EMBO journal 5, 823-826.

[3] Colovos, C and Yeates, T.O.(1993). Verification of protein structure: patterns of non – bonded atomic interactions. Protein Science.2(9):1511- 1519.

[4] Epstein , C.J., Goldberger, R.F and Anfinsen , C.B (1963). The genetic control of tertiary protein structure: studies with model systems. Cold Spring Harbor Symposium on Quantitative Biology. 28: 439-444.

[5] Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,s., Willkins, M.r., Appel, R.D., and Bairoch,A. Protein identification and analysis tools on the Expasy server. John Walker(ed)

[6] Geourjon, C and Deleage, G (1995). SOPMA: Significant improvement in protein secondary structure prediction.CABIOS.11,681-684.

[7] Guruprasad K, Reddy BVP, Pandit MW (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. Protein engineering 4: 155-164.

[8] Harry. B. Hines, Frank Lebeda, Martha Hale andErnstE.Bruggermann.(2005).Characterization of Botulinum progenitor toxins by Mass Spectrometry. Applied and Environmental Microbiology.71(8): 4473-4486.

[9] Hooft,R.W.W., Sander,C., Vriend,G. and Abola,E. (1996a) Errors in protein structures. Nature, 381, 272.

[10] Inoue, K., Y. Fujinaga, K. Honke, H. Arimitsu, N. Mahmut, Y. Sakaguchi, T. Ohyama, T. Watanabe, K. Inoue, and K. Oguma. 2001. *Clostridium botulinum* type A haemagglutinin-positive progenitor toxin (HA$^+$-PTX) binds to oligosaccharides containing Galβ1-4GlcNAc through one subcomponent of haemagglutinin (HA1). Microbiology 147**:**811-819.

[11] Kimiko Hasegawa, Toshihiro Watanabe, Tomonori Suzuki, Akihito Yamano, Tetsuo Oikawa, Yasuhiko Sato, Hirokazu Kouguchi, Tohru Yoneyama, Koichi Niwa, Toshihiko Ikeda and Tohru Ohyama (2007). A novel subunit structure of Clostridium botulinum serotype D toxin complex with three extended arms. Journal of Biological Chemistry.282 :24777- 24783.

[12] Koteswara Reddy, G., Mohan Kalyan Reddy, K., Nagamalleswara Rao, K and Gyana R.Satpathy . Comparative modelling of CDP- diacyl glycerol- serine o- Phosphatidyl transferase in *Clostridium botulinum*: A potent target of botulism. International journal of Drug Discovery.2 (2): 11-16.

[13] Lipman, D.J and Pearson, W.R (1985). Rapid and sensitive protein similarity searches. Science. 227, 1435-1441.

[14] Luthy,R., Bowie,J.U., and Einsenberg,D.(1992). Assessment of protein models with three dimensional profiles. Nature. 356(6364): 83-85.

[15] Nakamura, T., Tonozuka, T., Ito, S., Takeda, Y., Sato, R., Matsuo, I., Ito, Y., Oguma, K., Nishikawa, A.(2011) Molecular diversity of the two sugar-binding sites of the β-trefoil lectin HA33/C (HA1) from Clostridium

botulinum type C neurotoxin Arch.Biochem.Biophys. 512**:** 69-77

[16] Ramachandran,G.N.,Ramakrishana,C., and Sasisekharan,V.(1963). Stereochemistry of polypeptide chain configurations. Journal of Molecular biology.7, 95-99.

[17] Sagane, Y., Koguchi,H., Watanabe,T., Sunagawa,H., Inoue,K., Fujinaga, Y., Oguma,K., and Ohyama,T.(2001). Role of c termainal region, of HA-33 component of botulinum toxin in hemagglutination. Biochemical and Biophysical Research Communication.288, 650-657.

[18] Sahay A, Shakya M (2010) *In silico* Analysis and Homology Modelling of Antioxidant Proteins of Spinach. J Proteomics Bioinform 3: 148-154. doi:10.4172/jpb.1000134

[19] Sali A & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.

[20] Sander,C and Schneider,R (1991). Database of homology derived protein structures and the structural meaning of sequence alignment. Proteins : Structure, Function and Bioinformatics 9(1):56-68.

[21] Suhre.K & Y.H. Sanejouand, ElNemo: a normal mode web-server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research*, **32**, W610-W614, 2004.

[22] Yukako Fujinaga, Kaoru Inoue, Sadahiro Watanabe, Kenji Yokota, Yoshikazu Hirai, Eiko Nagamachi and Keiji Oguma (1997). The haemagglutinin of *Clostridium botulinum* type *C* progenitor toxin plays an essential role in binding of toxin to the epithelial cells of guinea pig small intestine, leading to the efficient absorption of the toxin. Microbiology .143, 3841-3847