Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.2414

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **4 (6):** 47-58 (2016)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Insilico Strategy of Epitope Design in *Clostridium botulinum*

Khalida Naveed*, Naila Jawaid and Syeda Tatheer Fatima

Baqai Institute of Information Technology, Baqai Medical University, Karachi, Pakistan

*Corresponding Author E-mail: khalidanaveed@baqai.edu.pk

Received: 30.11.2016 | Revised: 15.12.2016 | Accepted: 17.12.2016

ABSTRACT

*Clostridium botulinum is the name of a collection of bacteria that frequently habitat in soil. These are rod-shaped organisms that rose under anaerobic condition. They have the ability to produce spores for their survival under unfavorable conditions. The C. botulinum genome comprises of 3,886,996 BP, 28.24, GC%, 3650 coding sequence, 9rRNA and 80tRNA. Identifying a legend in a drug discovery project assumes that the drug target is known and has been characterized. The majority of available drugs have protein molecules as their targets for botulinum toxins are available for botulinum toxins. In our current project, potential epitope were identified in C. botulinum genome (An epitope is the part of the antigen that binds to a specific antigen receptor on the surface of a B cell) the predicted epitope by using available bioinformatics tools i.e. Vaxijen, TMHMM, Bcpred, MHCpred, Propred I, Propred, T-epitope designer, phyre and Pepitope. The two screening steps were adopted in this work. Previously identified 22 essential membrane proteins sequence were taken and in the first step screening the Vaxijen, TMHMM and Bcpred were used for selecting C. botulinum proteins as antigenic exo-membrane B-epitope. The two epitope, one from each protein have been designed in such a way that each epitope is highly likely to bind determined the number of (HLA molecules comprising of both the MHC-I and II) and interacts with most frequent HLA alleles (A*0201, A*0204, B*2705, DRB1*0101, and DRB1*0401) in human population. Therefore, our selected epitope are highly potential to induce both the B-cell and T-cell mediated immune responses. Of course, these selected epitope require further experimental validation.*

Key words: *insilico strategy,epitope prediction, Clostridium botulinum*.

INTRODUCTION

Clostridium botulinum is the name of a collection of bacteria that normally habitats in soil.These are rod-shaped organisms that rose under anaerobic condition. They have the ability to produce spores for their survival under unfavorable conditions. *C. botulinum* produces a toxin known as botulinum neurotoxin toxin (BoNTs) that is associated with life threatening neuro-paralysis. BoNTs are grouped into seven different sub-types designated with alphabetic letters from A to G. Among them A, B, E and F are pathogenic to humans (CDC, 2015). This toxin is among the list of bioterrorist agent with highest potential risk to the masses and has been used as biowarfare weapon by US military since Second World War.

Cite this article: Naveed, K., Jawaid, N. and Fatima, S.T., Insilco Strategy of Epitope Design in *Clostridium botulinum*, *Int. J. Pure App. Biosci.* **4(6):** 47-58 (2016). doi: http://dx.doi.org/10.18782/2320-7051.2414

The most effective treatment against BoNTs is vaccinations. Since 1940s efforts has been directed for the development of effective vaccines against botulism[10]. To date, under the supervision of Investigational New Drugs (INDs) amonovalent and a pentavalent vaccine is available against serotype F and A-E respectively[1]. While various DNA-recombinant vaccines are under clinical trial[3]. But, at present, there are no approved vaccines are available against[4]. Many different approaches are under consideration for the improved and effective diagnosis and development of treatment including vaccination, synthetic small peptides and small-molecule inhibitors[11]. The most effective approach for the prevention of botulism could be the vaccines consisting of suitable antigen that can elicit appropriate immune response. The process of development of effective vaccine involves recognition and identification of epitope that can stimulate strong protective humeral response[17].

The pathogenicity of a pathogen is often due to its pathogenic secretory and surface proteins that have also antigenic properties too[14]. Therefore, can be considered as one of the potential candidates against vaccine development. B-cell generates antibodies against antigenic and pathogenic proteins after recognizing the B-cell epitope over these proteins. B-cell epitope mapping can not only beneficial for effective diagnostic procedures but can also serve as a initial step towards vaccine designing[2]. Likewise, the specificity and diversity between antigen and human leukocyte antigen (HLA) alleles binding also contributes to specific and selective immune response.

The epitope and peptide based vaccine developmental procedures are easy, specific and harmless as compared to orthodox procedures. The peptide based vaccine developmental processes has accelerated as a result bacterial and human genome projects. The effective prediction and mapping of B-cells and T-cells from pathogenic bacteria with the aid of computational and bioinformatics tools is a significant achievement in the course of effective vaccine design[17].

Various membrane proteins are reported in literature through bioinformatics based approach against *C. botulinum* for potential vaccine targets[14]. In this study, we have explored 22 different essential membrane associated proteins for peptide based vaccine design with the aid of insilco approach together with simulation and verification of fold levels. This will lead to the identification of best possible and potential epitope for both B-cell and T-cell based immunity[14].

MATERIAL AND METHODS

Antigenic B-epitope selection: The twenty two essential membrane proteins of *C. botulinum* were earlier identified by using of subtractive genomics approach and used as best vaccines candidates were selected for the current study and a novel approach of epitope designing was adopted where an epitope should produce both B-cell and T-cell mediated immune response[5].

Prediction of B-cell epitope: The complete amino acids sequence of each protein was retrieved from NCBI database and all proteins sequence were analyzed at Vaxijen server for the prediction of antigenic proteins[9]. And default parameters with threshold value (>0.4, ACC output) were used against bacterial species to check the antigenicity of each full length protein. Proteins having antigenic score >0.5 were selected[9]. Then each full length proteins were subjected at TMHMM v 0.2 prediction server in order to identify the

surface exposed amino acids for each full length protein[12].

In the processes of B-cell epitope prediction each full length protein was subjected at Bcpred server to Use for the identification and characterization of B-cell epitope to make a novel vaccines against *C. botulinum*. And all predicted B-cell epitope (20 mers) having a Bcpred cutoff score >0.8 were selected[15]. The selected were then subsequently checked for membrane topology by comparing with TMHMM results for exo-membrane amino acids sequences. Then selected B-cell was further analyzed at Vaxijen to check the antigenicity. Then finally 44 were selected for use in T-epitope selection.

Screening one step: The T- cell was anticipated from B-cell. Then two screening steps were approved for sequence based and QSAR simulation approaches were used in the interpretation of T- selection. In the first screening step the considerations were. The first criteria was sequence should bind with MHC class I, and class II and sequence must be >15 was selected and second sequence must bind with HLA-DRB1*0101 and must be antigenic based on Vaxijen score. Then after this the sequence was bind with Propred I (47 MHC Class-I alleles)[15], and propred 51 MHC Class II alleles[16]. And both servers were used for the determination of alleles bind with above criteria. The QSAR simulation approach was used for check the inhibitor concentration of antigenic epitope at the platform of Propred I, Propred and MHCpred server (DRB1*0101) and Vaxijen, respectively[6]. With highest antigenicity and those bind more than 15 MHC molecules including with class I and II and less than 100nm (IC50) scores were selected.

Screening step two: The second screening step was subjected by using T-epitope designer tool in which the principles was >1000 HLA fragments, peptides should bind with 75% of total HLA molecules and must bind with high score to A*0201, A0204 and B*2705 and must bind with DRB1*0101, and DRB1*0401 and MHCpred .The final list of was made with non-overlapping peptide sequences that pass these above mentioned criteria and Vaxijen and IC50 scores was used for selection of DRB1*0101 and DRB1*0401 binding peptides were selected for design epitope for fold level topology examination[7].

Homology modeling for epitope analysis: For the homology modeling of each protein as a whole was subjected to Phyre version 2.0 Web-server[8]. Followed by best model prediction on the basis of super families and e-value of template. The exo-membrane topologies of the predicted were examined through their 3D structures and clustering of with the aid of Pepitope server[13]**.** This sever utilizes 3D structure of Phyre server for the prediction of linear alignment of with their respective proteins and also for prediction their epitope pockets and exo-membrane position in 3D structure of proteins.

RESULTS AND DISCUSSION

Antigenicity and topology of selected protein:

By using Vaxijen the anticipated possible proteins having antigenic score >0.4 were selected as an antigenic determinant and lastly 17 proteins sequence were selected as antigenic determinants shown in **(table 1).** The basic principle of a decent epitope is that it must be showing to cell outside and TMHMM result showed that all 17 probable antigen having surface exposed peptides shown in **(table 2).**

Table1: VexiJen predicted probable antigen and non-probable antigen

S. No.	Protein id	VaxiJen score	Comments
1	YP_001386116.1	**0.6282**	Probable ANTIGEN
2	YP_001386895.1	0.3078	**Probable NON-ANTIGEN**
3	YP_001387255.1	**0.4396**	Probable ANTIGEN
4	YP_001386503.1	**0.4003**	Probable ANTIGEN
5	YP_001386587.1	**0.4173**	Probable ANTIGEN
6	YP_001386886.1	**0.4931**	Probable ANTIGEN
7	YP_001387027.1	**0.7008**	Probable ANTIGEN
8	YP_001387384.1	**0.5959**	Probable ANTIGEN
9	YP_001387457.1	**0.4304**	Probable ANTIGEN
10	YP_001387707.1	**0.5763**	Probable ANTIGEN
11	YP_001387745.1	0.3927	**Probable NON-ANTIGEN**
12	YP_001388265.1	**0.5384**	Probable ANTIGEN
13	YP_001389313.1	0.2271	**Probable NON-ANTIGEN**
14	YP_001387020.1	**0.5164**	Probable ANTIGEN
15	YP_001387133.1	**0.4053**	Probable ANTIGEN
16	YP_001387258.1	0.3751	**Probable NON-ANTIGEN**
17	YP_001387314.1	**0.5220**	Probable ANTIGEN
18	YP_001386372.1	**0.4096**	Probable ANTIGEN
19	YP_001387500.1	0.3101	**Probable NON-ANTIGEN**
20	YP_001388739.1	**0.6415**	Probable ANTIGEN
21	YP_001388796.1	**0.7614**	Probable ANTIGEN
22	YP_001389355.1	**0.6086**	Probable ANTIGEN

Bcpred result based on Vaxijen and TMHMM Scores: The Bcpred results recommended that predicted peptides among 22 proteins only 19 are having Bcpred score >0.8 shown **(Table 2).** These 19 proteins were exposed to TMHMM again to identify exo-membrane constituencies and from them 15 proteins were concluded with Bcpred score >0.8 and exo-membrane topology. The Superficial exposed B-cell epitope peptides of 15 finalized proteins were analyzed using Vaxijen (threshold=0.4) to check the antigenicity and all peptides except one are having Vaxijen score >0.4 results are précised in Table 3. So 15 membrane proteins, among 22 were finalized as antigenic exo-membrane B-epitope targets shown in **(Table 3).**

Table 2: TMHMM result of 17 finalized proteins from VexiJen score			
S.No.	**PROTEIN**	**TMHMM RESULTS**	**LENGTH OF PROETIN**
1	YP_001386116.1	1– exo peptide	1-133
2	YP_001387255.1	2- exo peptide	1-90 209-572
3	YP_001386503.1	5- exo peptide	46- 59 117 - 167 231 - 239 298- 349 400 - 433
4	YP_001386587.1	5- exo peptide	1 - 111 164 - 196 262- 275 342 - 350 409 - 422
5	YP_001386886.1	3- exo peptide	30 - 55 208 - 232 110 - 123
6	YP_001387027.1	4-exo peptide	30 - 67 125 - 128 200 - 203 253 - 268
7	YP_001387384.1	2-exo peptide	30 - 43 102 - 190
8	YP_001387457.1	1-exo peptide	1-212
9	YP_001387707.1	2- exo peptide	186 - 211 393- 396
10	YP_001388265.1	1-exo peptide	29 - 556
11	YP_001387020.1	5-exo peptide	1 - 14 106 - 147 203 - 211 323 - 352 405 - 408
12	YP_001387133.1	3-exo peptide	36 - 97 156 - 174 259 - 277
13	YP_001387314.1	3-exo peptide	1 - 19 77 - 90 157 - 188
14	YP_001386372.1	7-exo peptide	28 - 41 87 - 109 176 - 189 256 - 302 381 - 394 461 - 474 544 - 546
15	YP_001388739.1	1-exo peptide	1 - 3
16	YP_001388796.1	1-exo peptide	1 - 3
17	YP_001389355.1	4-exo peptide	46 - 64 175- 216 267 - 270 325 327

Table 3: Summarized results of BCPred, TMHMM and VexiJen of finalized 15 proteins

S.No.	Protein name	Total no of epitope	Amino acids position	Bcpred epitope sequence	Bcpred score	Vaxijen score
1	YP_001386116.1	1	41	NMITYLKPTITKYIDLNGNK	0.845	0.6981
2	YP_001387255	4	306	SMEETGASAEEMSATSEEIE	0.973	1.3210
			119	EKNLNEFRQRREEVFKVAKE	0.925	0.1372
			22	IIVGITGYYFNAKSNKAIKK	0.901	0.6987
			370	IDSATKDAIEKSKTIEEINV	0.82	0.6195
				VTTNIISKSIINPLKESVEY	0.872	
3	YP_001386503	2	111	IGTLIINPTKGVDPSTMQKI	0.906	**0.4247**
			400	DAPATLLNSTGNTVCAMMIT	0.851	**0.5935**
4	YP_001386587	3	32	FNIKSEEKANDEEVKSIKGV	0.985	**1.5689**
			445	GFDDPVEQNKEEQKKVMEEN	0.98	**0.7432**
			84	SSKDNNEKKKEGIVDRVLSV	0.957	**1.2129**
5	YP_001386886	1	211	TRMKTEIIPTSIFLNMSTGD	0.879	**0.4487**
6	YP_001387027	1	217	TYPIVMFPFIQSGDRTIASS	0.873	**1.2066**
7	YP_001387384	1	162	TLMVAGNIPGKTQTIPTAIY	0.761	**0.5783**
8	YP_001387457	3	18	IEIFKTININTRELEDKGRK	0.963	**0.8082**
			103	LPSFKLDQGYNRKKIATKYP	0.873	**1.0699**
			55	KAKDVETYVEYGAADIGIVG	0.83	**0.4053**
9	YP_001387707	3	41	NITNKNIFKIRSEKITLKDC	0.825	**0.9307**
			296	FKENIKNLNKNIKDGKSISI	0.823	**0.7517**
			193	FFYKRYKIKYAVDNLKIKTP	0.814	**0.7133**
10	YP_001388265	2	117	KSGKDIKNTCTTLGANRFQT	0.998	**0.9145**
			51	KKIMIQINHTLMSIPPVLMG	0.893	**0.7043**
11	YP_001387020	10	79	VPSAFKDNKEEKDEEKLLTI	0.996	**0.7198**
			524	VVFIPKIKTKGESAGTTSVP	0.996	**1.1954**
			192	NRVKDNKTSKIIYKKDLQGE	0.992	**1.2277**
			106	NEKYDITQNQTIDSSGKNYY	0.978	**0.9687**
			213	ISKEAEEPKSNINPKLTLDS	0.962	**1.3843**
			128	DKITYEKLKDIKGVKGFYTY	0.953	**0.6478**

			313	SFKCTGEFESNKKGTHGSFS	0.928	**0.8250**
			376	EQQGAIQMEKGEKPNLSDGT	0.909	**0.7079**
			487	KTGTNERQEVNDKGKMEHLS	0.905	**1.3061**
			149	KQEVDKNINDKKEAWKLENM	0.881	**0.8601**
12	YP_001387133	3	223	DKIIKEPAGGAHKNLNKMAE	0.853	**0.5951**
			25	ARLKERPTALDYINIIFDDF	0.823	**0.5965**
			119	DTQGAFCGIDAEERGQGEA	0.782	**1.0672**
13	YP_001387314	5	84	IIGVPVAIAFGGPGAVFWMW	1	**0.9976**
			242	IFKGAFTPQAAVGGFGGATL	0.994	**0.4582**
			266	RWGTARGTYSNEAGMGSAPI	0.95	**0.6914**
			48	KETFGKMFSKPAEGEGTITP	0.933	**0.6455**
			152	MCSFCFMIEIIPSISTQSLS	0.768	**1.0821**
14	YP_001388739	1	43	KPNKIHIMGTDSMGRDVFSR	0.802	**0.6468**
15	YP_001388796	3	92	DTDKKEDMDKKEEKSAKEK	0.999	**1.7980**
			24	YVIIQTGPDRVKVQISKSGI	0.994	**0.5201**
			64	PEKKRKKKYNEMLNSISVND	0.921	**0.9828**

B –cell derived T-cell: Each selected B-cell epitope was examined for identification of T-cell within the B-cell epitope sequence. For the first level screening, Propred-I (47 MHC Class-I alleles), Propred (51 MHC Class-II alleles), and Mhcpred (DRB1*0101 allele) were used to isolate common T-cell epitope that share B-cell epitope sequence, can interact with both the MHC classes with highest number, and specifically interact with DRB1*0101(as the DRB1*0101 is commonest bound allele, therefore the interaction epitope produce better antigenic response).

At the second level of screening, identified peptides in the first screen were used to predict their binding abilities to >1000 MHC alleles using T-Epitope Designer and that bind to >100% alleles were selected. Similarly, as **A*0201, A*0204, and B*2705** alleles are mostly used in various prediction methods, we set the cut off that selected peptides must bind to these three HLA molecules and T-epitope Designer was also used for this purpose. Then selected epitope % work done at T-epitope designer tool **(Table 5).** Since the frequency of **DRB1*0101 and DRB1*0401** alleles of MHC Class-II is 20-50% in all**.** We selected T-epitope that interacted with these two HLA molecules using Mhcpred as described in methods. The final list of epitope was made with non-overlapping peptide sequences that confirm the above mentioned criteria and Vaxijen and IC50 scores. with highest antigenicity and those bind more than 15 MHC molecules comprising of both the MHC class I and II alleles and less than 100 nM IC50 scores for DRB1*0101 were selected. From the result 4 proteins with 5 peptides among 15 were fulfilled the criteria and predicted as T-epitope candidates are shown in **(Table 4 as *).**

Table 4: Common epitopes from each protein that can produce both the B- and T-cell mediated immunity are represents along with their various parameters. Epitopes selected as vaccine candidates are labeled as *

S.No.	Protein id	Bcpred predicted epitope sequence	Number of MHC class 1 binding alleles Propred1	Number of MHC class II binding alleles Propred	Mhcpred predicted epitope sequence	Vexijen score based on Mhcpred sequence	IC50 value of epitopes for DRB1*0101 (MHCPRED)	Total number of MHC binding alleles
1	YP_001386111	NMITYLKPTITKYIDLNGNK	0	23	YLKPTITKY TITKYIDLN NMITYLKPT KYIDLNGNK MITYLKPTI ITKYIDLNG	**0.7397** **0.6176** **0.1075** **1.5123** **0.2926** **1.0295**	0.24 30.90 34.12 34.83 40.36 53.58	23
2	YP_001387255	SMEETGASAEEMSATSEEIE	0	0	MEETGASAE EMSATSEEI EETGASAEE SMEETGASA EEMSATSEE ETGASAEEM SAEEMSATS AEEMSATSE GASAEEMSA	**2.0598** **0.9682** **1.7880** **1.3794** **1.5282** **1.3154** **1.2224** **1.2644** **1.0501**	13.52 29.85 30.20 37.58 38.99 55.85 57.15 61.80 68.71	0
		EKNLNEFRQRREEVFKVAKE	0	1	EEVFKVAKE NLNEFRQRR FRQRREEVF QRREEVFKV KNLNEFRQR NEFRQRREE LNEFRQRRE REEVFKVAK EFRQRREEV	**0.8125** **0.4323** **0.1051** **0.0243** **0.7049** **1.5227** **1.1472** **0.8793** **0.3925**	13.80 15.74 25.82 28.71 29.38 39.17 43.25 89.54 96.61	1
		IIVGITGYYFNAKSNKAIKK	0	25	YFNAKSNKA YYFNAKSNK IIVGITGYY GYYFNAKSN NAKSNKAIK GITGYYFNA	**1.7583** **1.4327** **0.3207** **1.0890** **1.2196** **0.9462**	0.82 4.37 12.39 33.19 54.95 75.34	25
		VTTNIISKSIINPLKESVEY	0	7	ISKSIINPL SIINPLKES IINPLKESV IISKSIINP INPLKESVE VTTNIISKS NIISKSIIN KSIINPLKE	**-0.8236** **-1.0420** **0.8115** **-0.6858** **-0.3249** **-0.2138** **-0.3920** **-1.5567**	1.83 9.73 13.90 17.02 19.32 28.71 52.48 69.02	7
		IDSATKDAIEKSKTIEEINV	0	1	IDSATKDAI IEKSKTIEE SATKDAIEK ATKDAIEKS	**0.9744** **0.5307** **0.3668** **0.3223**	9.12 22.86 26.67 37.33	1
3	YP_001386503	IGTLIINPTKGVDPSTMQKI	0	1	IGTLIINPT IINPTKGVD INPTKGVDP KGVDPSTMQ GVDPSTMQK NPTKGVDPS	**-0.8689** **1.3924** **1.3938** **1.6148** **0.7043** **1.4531**	5.00 7.60 30.90 48.53 55.21 81.47	1
		DAPATLLNSTGNTVCAMMIT	0	8	DAPATLLNS GNTVCAMMI NSTGNTVCA STGNTVCAM	**0.3571** **1.0599** **1.0920** **0.4179**	1.05 8.63 30.20 62.37	8
4	YP_001386587	FNIKSEEKANDEEVKSIKGV	5 0	1 0	KSEEKANDE KANDEEVKS NIKSEEKAN EEKANDEEV FNIKSEEKA IKSEEKAND DDPVEQNKE QNKEEQKKV FDDPVEQNK KEEQKKVME EEQKKVMEE GFDDPVEQN	**1.8137** **2.2451** **0.9825** **1.8589** **1.4696** **1.8361** **0.2730** **1.8256** **0.6671** **0.4609** **0.2955** **0.4641**	3.26 19.36 35.97 56.62 81.66 86.30 11.32 16.48 16.94 50.47 95.72 96.38	6 0
		GFDDPVEQNKEEQKKVMEN						
		SSKDNNEKKKEGIVDRVLSV	0	0	GIVDRVLSV EGIVDRVLS	**-0.1952** **0.4301**	9.44 23.17	0
5	YP_00138686	TRMKTEIIPTSIFLNMSTGD	0	4	IPTSIFLNM EIIPTSIFL PTSIFLNMS IIPTSIFLN RMKTEIIPT IFLNMSTGD	**0.1837** **0.7217** **-0.2156** **0.2835** **0.0357** **0.3659**	6.19 7.55 7.71 12.33 16.60 73.96	4
6	YP_0013870.1	TYPIVMFPFIQSGDRTIASS	0	45	SGDRTIASS IVMFPFIQS PIVMFPFIQ FIQSGDRTI MFPFIQSGD YPIVMFPFI	**0.7581** **1.2194** **2.2772** **0.3658** **0.7988** **3.3778**	4.55 12.33 16.87 22.70 38.02 40.93	

					QSGDRTIAS IQSGDRTIA TYPIVMFPF	**0.9732** **1.2340** **3.1066**	58.88 60.81 65.16	
7	YP_001387384	TLMVAGNIPGKTQTIPTAIY	0	14	KTQTIPTAI TLMVAGNIP PGKTQTIPT TQTIPTAIY NIPGKTQTI	**0.4205** **0.0788** **0.5960** **0.5508** **0.2571**	26.98 29.92 37.76 47.86 61.94	14
8	YP_001387457*	IEIFKTININTRELEDKGRK*	6	38	NINTRELED IEIFKTINI EIFKTININ INTRELEDK IFKTININT ININTRELE FKTININTR TININTREL NTRELEDKG	**1.1550** **-0.3309** **0.4764** **1.4586** **1.0627** **2.3701** **1.3949** **1.4760** **0.8280**	10.19 27.73 30.13 31.48 34.43 35.48 37.67 72.28 93.76	44
		LPSFKLDQGYNRKKIATKYP	0	17	NRKKIATKY SFKLDQGYN FKLDQGYNR GYNRKKIAT YNRKKIATK PSFKLDQGY	**0.6706** **0.5881** **0.7713** **-0.4442** **0.6158** **0.7628**	4.32 26.73 47.53 64.71 68.55 85.31	17
		KAKDVETYVEYGAADIGIVG*	8	8	YVEYGAADI TYVEYGAAD KDVETYVEY KAKDVETYV EYGAADIGI YGAADIGIV	**0.1827** **0.2998** **0.2379** **0.7194** **0.6886** **0.8185**	5.86 31.70 42.17 43.95 59.16 97.72	16
9	YP_001387707*	NITNKNIFKIRSEKITLKDC	0	24	SEKITLKDC IRSEKITLK KNIFKIRSE IFKIRSEKI NITNKNIFK NIFKIRSEK TNKNIFKIR RSEKITLKD KIRSEKITL ITNKNIFKI	**1.1059** **1.5512** **0.7365** **1.8222** **0.2600** **0.7830** **0.9905** **1.1855** **0.9130** **0.5455**	22.18 28.05 29.04 38.02 41.98 42.17 53.09 55.98 80.35 89.74	24
		FKENIKNLNKNIKDGKSISI*	3	27	NIKDGKSIS NIKNLNKNI FKENIKNLN ENIKNLNKN IKDGKSISI	**0.8192** **-0.5160** **0.4539** **0.1772** **2.1246**	6.61 9.29 25.41 50.35 98.40	30
		FFYKRYKIKYAVDNLKIKTP	0	23	FYKRYKIKY KIKYAVDNL YAVDNLKIK RYKIKYAVD KYAVDNLKI YKRYKIKYA AVDNLKIKT	**0.8634** **-0.2952** **2.1610** **1.0271** **0.9084** **1.1921** **1.8388**	1.02 11.40 16.37 30.62 35.73 35.73 54.33	0
10	YP_001388265	KSGKDIKNTCTTLGANRFQT	0	8	CTTLGANRF TCTTLGANR KNTCTTLGA DIKNTCTTL NTCTTLGAN SGKDIKNTC TLGANRFQT	**0.3731** **0.6461** **0.5892** **0.6486** **1.0142** **1.0292** **0.4561**	3.49 6.49 8.22 16.87 21.13 32.43 79.07	0
		KKIMIQINHTLMSIPPVLMG	0	50	QINHTLMSI INHTLMSIP KIMIQINHT MIQINHTLM TLMSIPPVL HTLMSIPPV NHTLMSIPP IQINHTLMS MSIPPVLMG	0.1541 -0.1214 0.9198 1.0504 1.4452 0.6236 0.3938 0.7890 0.7154	3.79 10.35 15.35 25.35 26.18 29.51 30.27 37.41 68.55	50
11	YP_001387020*	VPSAFKDNKEEKDEEKLLTI	0	0	KDEEKLLTI EEKDEEKLL PSAFKDNKE DNKEEKDEE FKDNKEEKD EKDEEKLLT	**-0.5857** **0.4140** **-0.2012** **2.2560** **2.4049** **-0.1784**	18.62 31.70 33.96 35.56 45.19 89.95	0
		VVFIPKIKTKGESAGTTSVP*	2	42	KIKTKGESA FIPKIKTKG ESAGTTSVP	**1.4006** **0.0516** **0.9316**	15.63 35.16 76.03	44
		NRVKDNKTSKIIYKKDLQGE	0	3	IYKKDLQGE IIYKKDLQG KIIYKKDLQ DNKTSKIIY RVKDNKTSK NKTSKIIYK	**1.5027** **1.2054** **0.1733** **-0.2053** **2.0850** **0.0903**	14.39 25.82 26.67 30.69 37.33 69.82	3
		NEKYDITQNQTIDSSGKNYY	0	13	KYDITQNQT IDSSGKNYY TIDSSGKNY TQNQTIDSS DITQNQTID	**0.7658** **1.8068** **0.9657** **0.8574** **0.5253**	2.88 16.60 18.28 19.05 24.10	13

					QNQTIDSSG ITQNQTIDS NEKYDITQN EKYDITQNQ NQTIDSSGK YDITQNQTI	**1.3929** **01.220** **0.8988** **1.1749** **1.1139** **0.5493**	39.26 39.36 45.92 76.91 79.25 90.99	
		ISKEAEEPKSNINPKLTLDS	0	32	NINPKLTLD EAEEPKSNI INPKLTLDS EPKSNINPK ISKEAEEPK AEEPKSNIN	**2.3057** **0.4077** **1.7361** **1.3404** **1.4756** **0.7647**	2.96 14.59 17.66 21.38 86.10 87.90	32
		DKITYEKLKDIKGVKGFYTY	0	4	KITYEKLKD YEKLKDIKG ITYEKLKDI KLKDIKGVK DIKGVKGFY TYEKLKDIK	**0.5463** **0.8069** **-0.0591** **1.0099** **0.2432** **0.2710**	7.21 12.30 16.11 21.68 40.18 88.31	4
		SFKCTGEFESNKKGTHGSFS	0	6	FKCTGEFES TGEFESNKK FESNKKGTH SFKCTGEFE	**0.1155** **0.1017** **2.6654** **-0.1207**	54.45 55.72 59.57 76.38	6
		EQQGAIQMEKGEKPNLSDGT	0	0	GEKPNLSDG MEKGEKPNL EQQGAIQME EKGEKPNLS IQMEKGEKP QQGAIQMEK	**0.5964** **1.4655** **0.6455** **1.5979** **0.4918** **0.4918**	39.99 44.67 55.21 61.38 67.14 83.56	0
		KTGTNERQEVNDKGKMEHLS	0	7	QEVNDKGKM TGTNERQEV TNERQEVND NERQEVNDK NDKGKMEHL EVNDKGKME	**0.6192** **1.5187** **0.9566** **1.5214** **0.6646** **0.6720**	19.91 37.33 44.87 46.24 50.93 98.40	7
		KQEVDKNINDKKEAWKLENM	0	0	NDKKEAWKL NINDKKEAW EVDKNINDK QEVDKNIND INDKKEAWK KQEVDKNIN KNINDKKEA KEAWKLENM	**0.6709** **-0.6496** **1.5209** **0.8674** **0.3138** **1.0295** **1.2620** **0.3980**	8.20 20.37 30.97 33.57 55.59 56.36 60.67 70.79	0
12		DKIIKEPAGGAHKNLNKMAE	0	0	IIKEPAGGA GAHKNLNKM	**0.9651** **0.4107**	3.65 23.01	1
	YP_001387133	ARLKERPTALDYINIIFDDF	0	3	KERPTALDY YINIIFDDF DYINIIFDD TALDYINII ERPTALDYI RLKERPTAL RPTALDYIN	**0.1009** **1.3204** **0.7424** **-0.1944** **0.2703** **0.3706** **-0.0005**	0.75 10.23 11.38 12.56 13.61 20.28 74.13	3
		DTQGAFCGIDAEERGQGEA	0	2	FCGIDAEER GIDAEERGQ	**1.0394** **1.1701**	26.36 41.88	2
13		IIGVPVAIAFGGPGAVFWMW	0	18	IGVPVAIAF FGGPGAVFW IIGVPVAIA	**1.0727** **1.0641** **0.7140**	18.54 20.04 27.73	18
		IFKGAFTPQAAVGGFGGATL*	4	15	IFKGAFTPQ AFTPQAAVG GAFTPQAAV FTPQAAVGG FKGAFTPQA	**0.0284** **1.0686** **0.8912** **1.4438** **0.0834**	12.91 13.68 35.32 42.56 69.02	19
	YP_001387314*	RWGTARGTYSNEAGMGSAPI	0	2	TYSNEAGMG WGTARGTYS YSNEAGMGS EAGMGSAPI TARGTYSNE ARGTYSNEA	**1.3142** **1.3114** **1.8533 0.5664** **1.1848** **1.1406**	22.49 31.92 61.52 62.09 62.95 93.33	2
		KETFGKMFSKPAEGEGTITP	0	10	FGKMFSKPA FSKPAEGEG KETFGKMFS TFGKMFSKP	**0.2997** **2.2657** **-0.2576** **-0.5614**	4.98 30.20 43.25 73.11	10
		MCSFCFMIEIIPSISTQSLS	0	34	PSISTQSLS IIPSISTQS FMIEIIPSI CFMIEIIPS EIIPSISTQ MIEIIPSIS FCFMIEIIP SFCFMIEII	**1.1298** **0.7081** **0.7752** **1.4870** **0.5050** **0.5922** **1.7171** **1.8738**	3.66 5.65 10.59 16.83 20.99 24.55 30.55 30.55	34
14	YP_001388739	KPNKIHIMGTDSMGRDVFSR	0	31	HIMGTDSMG MGTDSMGRD DSMGRDVFS KIHIMGTDS TDSMGRDVF SMGRDVFSR IMGTDSMGR	**1.1178** **2.0171** **0.2259** **1.1923** **0.8134** **0.1891** **1.3634**	23.33 50.82 56.89 57.68 64.57 84.53 95.50	31

15	YP_001388796	DTDKKEDMDKKEEKSAKEK	0	0	KEDMDKKEE DMDKKEEKS KKEEKSAKE EDMDKKEEK	**1.7079** **1.7928** **2.2276** **2.0821**	38.11 65.77 68.08 89.95	0
		YVIIQTGPDRVKVQISKSGI	0	28	RVKVQISKS IIQTGPDRV PDRVKVQIS YVIIQTGPD QTGPDRVKV IQTGPDRVK KVQISKSGI	**0.7288** **1.1728** **0.8000** **0.3088** **0.9551** **1.6724** **0.1868**	2.07 12.85 20.61 26.24 29.58 50.93 71.94	28
		PEKKRKKKYNEMLNSISVND	0	8	KYNEMLNSI YNEMLNSIS MLNSISVND KKYNEMLNS PEKKRKKKY EMLNSISVN KRKKKYNEM NEMLNSISV	**0.1323** **0.2446** **1.0659** **0.4477** **1.6571** **1.1054** **0.7381** **0.1941**	3.68 9.25 28.12 34.28 39.54 46.34 55.21 58.21	8

Table 5: Final selection of T-epitopes from 22 proteins *Clostridium botulinum*

S.No.	Protein	Bcpred predicted epitope sequence	Epitopes	T-Epitope Designer A*0201	T-Epitope Designer A*0204	T-Epitope Designer B*2705	Mhcpred (IC50 value) DRBI*0101	MHCPred (IC50 values) DRBI*0401
1	YP_001387457	IEIFKTININTRELEDKGRK	NINTRELED	414.90	-238.08	50.09	10.19	774.46
			EIFKTININ	-666.14	-581.60	268.50	30.13	505.82
			INTRELEDK	387.71	-3.97	2050.80	31.48	1778.28
			IFKTININT	692.05	698.62	-109.28	34.43	2233.57
			ININTRELE*	564.12	160.73	1472.59	35.48	1169.50
			FKTININTR	116.12	-404.02	1737.06	37.67	246.60
			TININTREL	288.37	-282.39	575.60	72.28	458.14
2		KAKDVETYVEYGAADIGIVG		476.60	-0.30	-194.58	43.95	1545.25
			KAKDVETYV	-1176.99	-832.31	-1901.75	59.16	195.88
			EYGAADIGI	-1421.67	-1347.90	-555.62	97.72	405.51
			YGAADIGIV					
3	YP_001387707	FKENIKNLNKNIKDGKSISI	NIKDGKSIS	-422.70	-1335.83	-405.58	6.61	1256.03
			FKENIKNLN*	680.97	152.62	1577.69	25.41	228.56
			IKDGKSISI	44.32	-36.26	-200.68	98.40	1479.11
4	YP_001387020	VVFIPKIKTKGESAGTTSVP	KIKTKGESA	302.16	-379.47	994.38	15.63	993.12
			ESAGTTSVP	-1076.85	-1386.53	-117.11	76.03	258.82
5	YP_001387314	IFKGAFTPQAAVGGFGGATL	IFKGAFTPQ	-9.95	-129.75	364.70	12.91	788.86
			AFTPQAAVG	-826.56	-926.19	-267.74	13.68	---
			GAFTPQAAV	-203.67	-108.78	-865.36	35.32	797.99
			FTPQAAVGG	-417.68	-450.27	681.70	42.56	193.20
			FKGAFTPQA	-484.89	-403.49	768.31	69.02	96.38

Clusters and folding: Homology modeling for each full length protein was carried out using Phyre version 2.0 Web-server and best models were selected based on super families and E-values of templates. The 3D folding and clusters of in folded protein were analyzed to confirm the exo-membrane topology of this using Pepitope server. Pepitope was fed with Phyre derived 3D structure of each protein and all identified from the same protein to analyze the linear alignment of on the corresponding protein and to determine the epitope clusters and exo-membrane position of in the folded proteins shown in (Figure 1). And best clusters calculated by Pepitope were having scores 45.551 residues no 12 for peptide **ININTRELE** and 10.122 residues no 9 for peptide **FKENIKNLN (Table 6).**

Table 6: Screening of T-epitopes using T-epitope Designer

Protein names	Epitope	T-Epitope Designer (no of binders with HLA molecules)	comments	lowest score	highest score
YP_001387457	ININTRELE	100%	almost all values are positive .mostly bind with A,B,C alleles.	24.09(A) 82.35(B)	2025.15(B*1561) 1635.00(A*3401)
YP_001387707	FKENIKNLN	90%	mostly positive for all B and C alleles	1.82(A-alleles) 87.07(B alleles)	1100.77(A alleles) 1848.06(B alleles

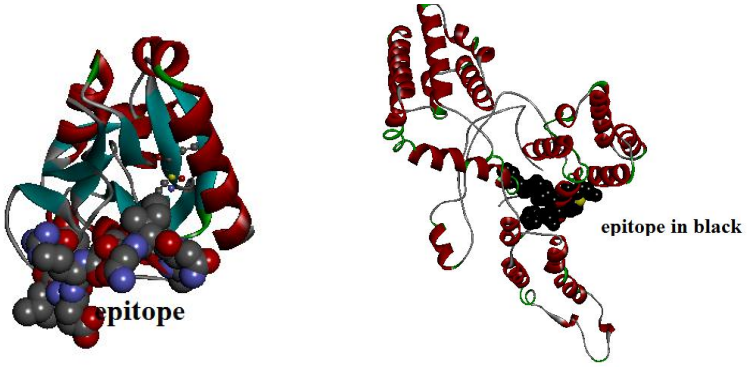

Fig. 1: Fold level characterization of cluster and topology of best using Pepitope (in red and black)

CONCLUSION

In this study by using computational approaches based on sequence structure, QSAR, simulation, and fold level analysis, we identified two potential T- derives from antigenic B-cells epitope of twenty two exo-membrane essential proteins of ***C. botulinum.*** Selected T- **["INTRELEDK**" from **YP_001387457**, **"FKENITKNL"** from **YP_001387707]** are antigenic and have much potential to interact most common human HLA alleles **(**A*0201, A*0204, B*2705, DRB1*0101, and DRB1*0401). These are also found to interact with >75% of HLA molecules in a binding screening using T-Epitope Designer (that contains >1000 HLA molecules). Therefore these selected are likely to induce both the B-cell and T-cell mediated immune responses. Homology and simulation results also support the suitability of these as vaccine candidates. However, there are several pitfalls in developing a good vaccine and moreover there is lack of proper experimental disease model for botulinum suitable animal model should be used for experimental validation of these to confirm these **"Insilco"** results.

REFERENCES

1. Arnon, S, *et al*., Jama **285(8):** 1059-1070 (2001).
2. Barh, D. *et al*., *Bioinformation* **5(2):** PMC: 3039994 (2010).
3. Byrne and Smith *et al.,* Biochimie**82(9):** 955-966 (2000).
4. Cerdino-tarraga, A.M. *et al., Nucleic Acids Res* **31:** 6516 [PMID: 14602910] (2003).
5. Doytchinova, I.A. and Flower, D.R., *BMC Bioinformatics*, **8:** PMID:17207271 (2007).
6. Guan, P. *et al. Nucleic Acids Res* **31:** PMID: 12824380 (2003).
7. Kangueane, P. *et al., Bioinformation* **1:** PMID: 17597847 (2005).
8. Kelley, L.A. and Sternberg, M.J., *Nat Protoc* **4:** PMID: 19247286 (2009)
9. Krogh, A. *et al*., *J Mol Biol* **305:** PMID:11152613 (2001).
10. Kukreja, R. and Singh, B.R., *Springer***,** 469-504 (2015)
11. Larsen, J.E. *et al*., *Immunome Res* **2:** 2 [PMID: 16635264] (2006).
12. Manzalawy, Y.E.L. *et al*., *J Mol Recogni*., **21:** PMID: 18496882 (2008).
13. Mayrose, I., *et al., Bioinformatics*, **23:** 3244 PMID: 17977889 (2007).
14. Reddy, G.K. *et al*., *International Journal of Pharmaceutical Studies and Research*, **2**: 48-54 (2011).
15. Singh, H. and Raghava, G.P., *Bioinformatics*, **12:** PMID: 11751237 (2001).
16. Sturniolo, T. *et al*., *Nat Biotechnology* **17:** PMID: 10385319 (1999).
17. Zhao, B. *et al., Int. J. Integrative Bi*, **2**: 127 [Google Scholar] (2007).