European Journal of Biological Sciences 5 (3): 68-76, 2013
ISSN 2079-2085

DOI: 10.5829/idosi.ejbs.2013.5.3.7282

Screening of *Bacillus anthracis* Plasmid Px01 Proteins to Identify Novel Antigenic Peptides-an Immunoinformatics Approach

[1]M.A. Rasheed, [2]S. Anuraj Nayarisseri, [2,3]Mukesh Yadav, [4]Aditya Jain, [3]Priya Sharma, [4]Suparna Roy and [3]Shailesh Saket

[1]Deparment of Biotechnology, Sreenidhi Institute of Science and Technology, Hyderabad, India, 500029
[2]In silico Research Laboratory, Eminent Biosciences, Vijaynagar, Indore, India, 452010
[3]Department of Biotechnology, Softvision College, Vijaynagar, Indore, India, 452010
[4]Manipal Life Sciences Center, Planetarium Complex, Manipal University, Manipal, India - 576104

Abstract: *Bacillus anthracis* is a gram positive bacterium and the etiologic agent of anthrax, a common disease of livestock and, occasionally, of humans and the only obligate pathogen within the genus Bacillus. Hence the aim of the current study was to identify the novel antigenic peptides from the whole plasmid PX01 proteins of *Bacillus anthracis*. The plasmid proteome were analyzed using various online bioinformatics algorithms. The current study performed was based on two facts i) the toxicity of *B. anthracis* is by secretion of exotoxins encoded from plasmid pX01 into the host cell and ii) the plasmid pX01 has a distinct 70kbp region. Hence the work was initiated by considering the proteins that are secreted out of the *B. anthracis* cell via the classical pathway. The proteins were analyzed for the transmembranes, antigenecity and their regions followed by two protein blasts across the human and *bacillus* database. The blast against the human database showing their absence in humans confirmed their antigenecity for humans and blast across *bacillus* had shown the *B. anthracis* specificity. The complete plasmid genome had 1695 protein entries at the genbank. At the end 10 potential proteins were found using the Bioinformatics algorithms. During the process two other observations were found, one showing an antigenic protein sequence specific to nearly 100 *Bacillus* species and the second was a protein sequence that had low similarity for the well-identified breast cancer gene BRCA1.

Key words: *Bacillus antracis* · Anthrax · Antigenecity Prediction · Signal Peptide Prediction

INTRODUCTION

Bacillus anthracis is the etiological agent of anthrax [1]. Its life cycle includes vegetative and endospore morphologies [2]. The endospores of *Bacillus* which are the infectious particles of anthrax are dormant bacterial morphotypes [3], resistant to a variety of physical extremes including heat, desiccation, UV and irradiation, and oxidation. The ability to switch between this cell type and the rapidly dividing vegetative form provides the *bacilli* with a highly effective strategy for persistence in the environment. The regulated progression of this cycle is critical for successful establishment of disease and subsequent survival of the organism after death of the host [2]. Anthrax occurs in three forms depending on the route of infection: Cutaneous anthrax (through a cut or abrasion on the skin), gastrointestinal anthrax (by consumption of undercooked meat from contaminant animals) [4] and inhalational anthrax (due to inhalation of spores) [5].

The plasmids pX01 and pX02 are responsible for the virulence of *B. anthracis* [4, 5]. Full virulence requires the presence of an antiphagocytic capsule of plasmid pX02 and three toxin components of plasmid pX01. The capsule

Corresponding Author: S. Anuraj Nayarisseri, In silico Research Laboratory, Eminent Biosciences, Vijaynagar, Indore, India, 452010, Tel: +97-52295342.

inhibits host phagocytosis while the three toxins of plasmid pX01 are individually nontoxic and they act in binary combinations causing edema and cell death [4]. A 70 kbp region on pX01 is distinct from the remainder of the genome and includes these toxin genes [6].

Anthrax vaccine adsorbed (AVA; BioThrax™) is the FDA-licensed human anthrax vaccine used in the United States [7]. AVA is a sterile suspension from an avirulent and non-capsulated strain of *B. anthracis*. It contains no dead or live bacteria. Protective antigen protein is the principal antigen responsible for the induction of immunity [4] that induces anti-PA antibody production which is capable of neutralizing the toxicity. Immunization with Biothrax elicits Th2 and Th1 immune response [8]. AVA is safe and efficacious but has limitations of standardization and schedule of administration. This justifies the need for development of an improved vaccine [9]. The aim and focus of this research is to search the pX01 genome for proteins/peptides that are novel and could be helpful in the vaccination programme against Anthrax. The process was developed and designed on this exotoxin secretion mechanism of the *bacilli* into the host cell and the immunization studies.

Hemalatha *et al*. [10] discussed a similar strategy in their research. The novel amino acid sequence repeats and domains identified were discussed. In our project the Computational tools utilized and the analysis strategies adapted were different. As the aim and the protocol differed entirely, the results were certainly different. Even by this approach, we also found the BRCA1 gene related sequence similar to that reported by the Hemalatha *et al*.'s repeat and domain recognition methodology. The complete genome of *Bacillus anthracis* including the plasmids is sequenced [11] and is available at the NCBI database. The proteome sequence of plasmid pX01 was downloaded and further analyzed using the following Software tools and Servers: The prediction of signal peptides was carried out at the PREDISI server which is well suited for the evaluation of whole proteomes as it predicts the signal peptide sequences and their cleavage positions [12]. Expert Protein Analysis System is a *proteomics* server, providing access to a variety of databases and analytical tools [13]. The DAS and TMHMM were accessed through ExPASy. Dense Alignment Surface (DAS) is a good method for predicting transmembrane segments in prokaryotes [14] and TransMembrane Hidden Markov Model (TMHMM) is a protein topology prediction method [15]. The accuracy of DAS is calculated to be 90-95% [14] whereas TMHMM prediction is 97-98% accurate [15]. Immunogenecity of the protein was the next characteristic to be determined. This was done by using the Antigenecity prediction server and EMBOSS software. Antigenecity prediction server uses the method of Kolaskar and Tongaonkar for antigenecity prediction. EMBOSS is "The European Molecular Biology Open Software Suite". It integrates a range of tools for sequence analysis [16] among which is the antigenic suit available. Basic Local Alignment Search Tool (BLAST) is a program for sequence similarity searching. BLAST is a heuristic method that attempts to optimize a specific similarity measure for comparing gene and protein sequences against others in public databases [17]. The *blastp* program compares protein queries to protein databases [17]. The Clustal programs are widely used for carrying out automatic multiple alignments of nucleotide or amino acid sequences. ClustalX is the most familiar version with graphical user interface and some powerful graphical utilities for aiding the interpretation of alignments and is the preferred version for interactive usage [18]. The predictions obtained from various softwares were compared and they were found to be the same for transmembrane prediction at DAS and TMHMM as well as for antigenecity detection by Antigenecity prediction server and EMBOSS.

MATERIALS AND METHODS

Mass Screening of the Plasmid Proteome Sequences: The proteome of plasmid pXO1 from *B. anthracis* were retrieved from Genbank. The dataset contains 1695 protein entries which were obtained and were given for the prediction of signal peptides at the predisi online server. Those proteins shown to have signal peptides were separated and looked for the presence of Transmembrane region. Transmembrane prediction has been predicted using DAS and TMHMM. Each individual protein result was inspected manually and all the proteins showing Transmembrane regions at both the servers were considered for the further analysis.

Identification of Immunogenic Proteins: Identification of the novel antigenic peptides and proteins were carried out using antigenicity predicting tools, EMBOSS and BlastP. The antigenic peptides predictions were kept dependent on immunax antigenic prediction server build at Harvard University. Simultaneously the presence of the antigenic regions in the extracellular region was checked using TMHMM server, upon that the absence of those regions in the human genome would contribute to the confirmation of its antigenic nature. Hence the Protien

BLAST was carried out only for the human genome (Taxid: 9606) against Non-redundant protein database. The proteins that either does not have antigenecity in outside region or have low similarity hits in outside antigenic regions against Humans were ignored. Another set of BLASTP was carried out for proteins obtained from the above BLAST to exclude the proteins shared in *Bacillus* species. This Protein-protein BLAST was carried out by limiting the Entrez query to '*Bacillus'*. The proteins obtaining no hits or low similarity hits were subjected to further analysis.

Alignment of Identical Proteins: It is elementary here to characterize the identified proteins. The manual sequence observation of all the proteins showed the reoccurrence of proteins. Hence Multiple Sequence Alignment of such proteins was carried out by the ClustalX and the repetitive proteins were discarded. The EMBOSS algorithm has been used to obtain primary scoop of the proteins. The non hit proteins in BLAST and good antigenic score regions were considered to be novel peptides of the pX01 proteins. An integral flowchart of systematic analysis is given in Fig.1.

RESULTS AND DISCUSSIONS

Mass Screening of the Plasmid Proteome Sequences: The 181.5 kb plasmid pX01 of *B. anthracis* had 1695 protein sequence entries. The 1695 sequences were obtained and subjected to signal peptide prediction and 234 sequences gave a positive signal. Each of the 234 protein sequences was submitted to DAS server. The proteins having regions of strict cutoff i.e. value greater than 2.2 were taken into consideration few had regions with loose cutoff i.e. 1.7. A graph was obtained showing the regions present inside, outside and in the transmembrane regions. Most of the proteins covered three regions (Fig. 2).

There were around 209 proteins that were dispersed across the three regions and had score value greater than 2.2. For efficient prediction, these proteins were loaded on TMHMM server, which provided the peptide sequences found in each of the three regions (Fig. 3). Here, it was found that the proteins were divided into two groups, half of the proteins were spread across three regions and the rest half were totally extracellular. The first group 101 proteins were taken; additionally the 14 Protective Antigen proteins were picked though they fell in the 2nd category. This ended the first stage giving 115 proteins out of 1695.

Identification of Immunogenic Proteins: At beginning of the 2nd stage all the proteins were analyzed on 3 parameters simultaneously. The outside regions predicted by TMHMM (115 proteins). The antigenic regions of proteins were found by immunax antigenic prediction server build at Harvard University. (Fig. 4 and Tab. 1) and the EMBOSS suite. The proteins with no antigenecity in the outside region were excluded at the earliest (9 proteins).

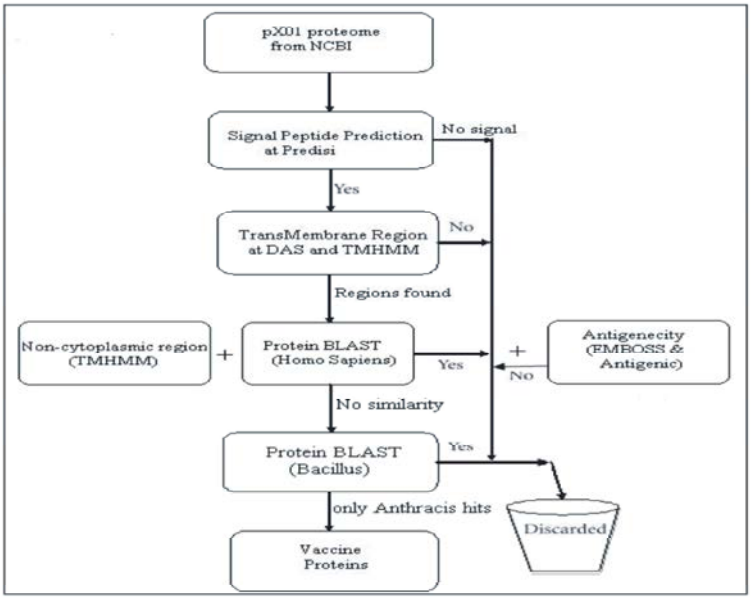

Fig. 1: Flowchart of systematic analysis of pX01 proteome.

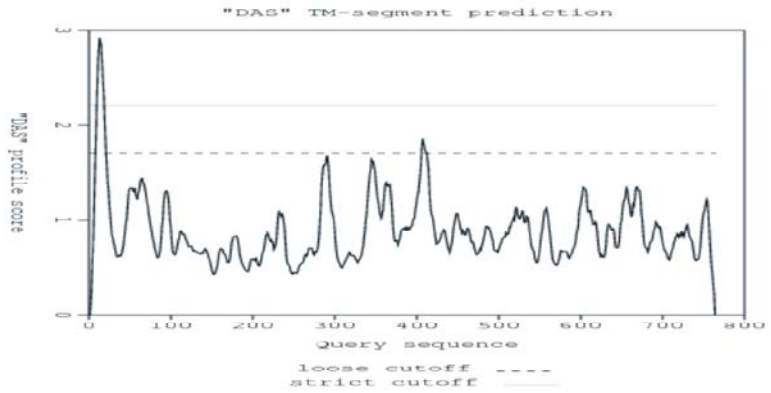

Fig. 2: The Transmembrane predictions by DAS server (The DAS curve for Protective antigen precursor protein)

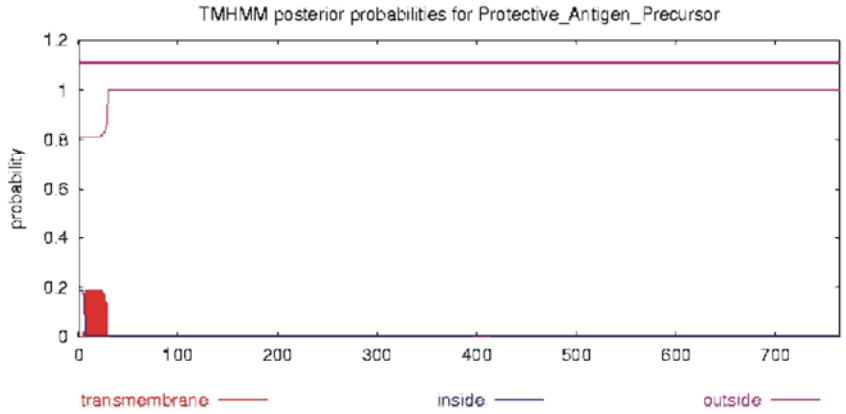

Fig. 3: The Transmembrane prediction by TMHMM Server.

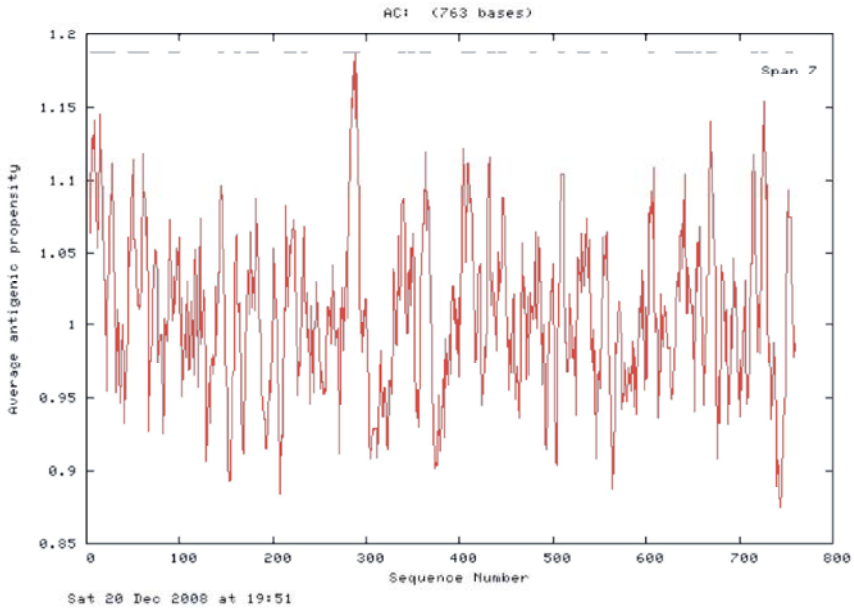

Fig. 4: Antigenic peptides are determined using Antigeneicty predicton server by the method of Kolaskar and Tongaonkar et al.[1990.] There are 27 antigenic determinants in the sequence by antigenic server. The sequence length is 763 amino acid residues. Average antigenic propensity for Protective antigen precursor protein is 1.0065.

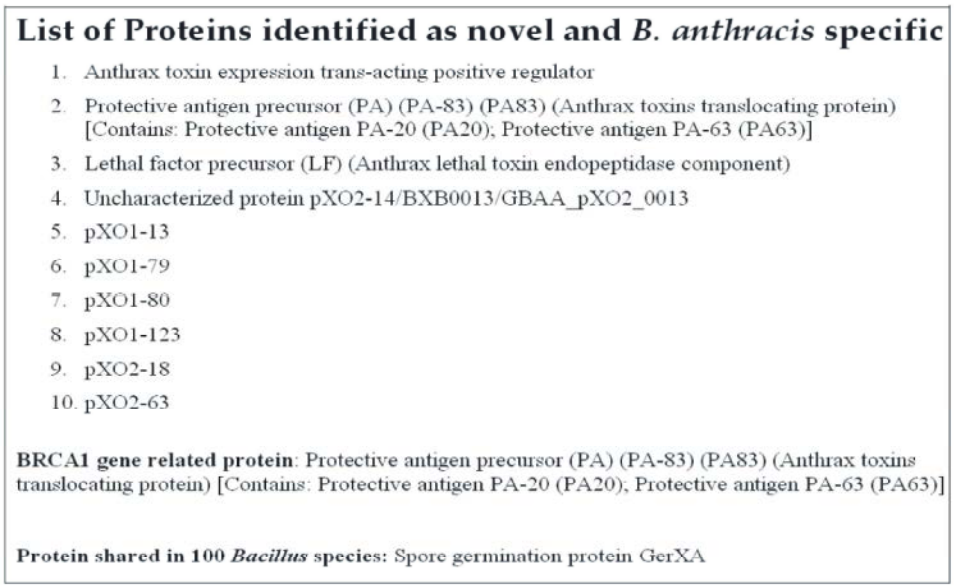

Fig. 5: List of proteins identified as novel and *B. anthracis* specific.

Pairwise alignment against nr database has been performed using Blastp, the result proteins were obtained in four categories: good identity, low similarity, no significant similarity and no hits. Good identity proteins are those having more identity in the outside region (7 proteins) It is note worthy that none of the 114 proteins had high identity. *Low similarity* refers to similar regions found in outside antigenic region (27 proteins). No significant hit refers to those proteins which has similarity with human proteins but not in the extracellular regions of 60 proteins. No hit clearly shows the absence of hits in BLAST results which contain 11 proteins. All the no significant hit and no hit proteins were considered even the proteins with very low similarity were taken into account counting to a total of around 80 proteins. The significance of all these proteins is that they are all good antigen proteins for humans and they can be further exploited and studied. The second set of Blast was carried out with 80 proteins by limiting the Entrez query against Bacillus species, to eliminate those proteins that are not specific to *B. anthracis* i.e. found in other *Bacillus* species. These BLAST results were of three types of proteins: *Bacillus* identical, *Bacillus* similar and *B. anthracis* specific. *Bacillus identical* proteins were found with 95-100% identity in different species, predominantly in *B.Thurengenesis, B. cereus* and also in *B. Subtilis, B. leichiemins, B. Pubmilans* and many others. From this category 9 similar proteins having only 1 identical hit for *B. Thurengenesis* were considered (Anthrax toxins

Table 1: Antigenic determinants in Anthrax toxin expression trans-acting positive regulator. Average propensity of Anthrax toxin expression trans-acting positive regulator protein is 1.0359

S.No.	Start	Sequence	End	Score
1	4	PISIEK	9	1.056
2	10	EHIRLINLLHF	20	1.125
3	32	ELSDYLQVAD	41	1.134
4	43	TVRKYLKLLED	53	1.111
5	57	PSWNLLVQK	65	1.090
6	67	KGIYLKK	73	1.105
7	75	LNESLSFVES	84	1.096
8	86	ILRKSLNLQICEELVF	101	1.141
9	107	QSLAQKLHLQVGALYPIIN	125	1.142
10	131	IQSSHLNIKKKPLE	144	1.059
11	150	QDVRVFMLRLYCNI	163	1.125
12	187	KILNVQMYTYSKHKLCVLFAITISRL	212	1.229
13	220	NVSGLILV	227	1.186
14	233	HYKTVASIT	241	1.087
15	246	NSFGVTLHE	254	1.089
16	256	EISFLALALLLSL	268	1.197
17	286	TIMPLAK	292	1.031
18	298	IEHKLQLG	305	1.082
19	309	DESFLTYVVLIIKK	322	1.213
20	328	FIQYYNYNIKFIRHIK	343	1.074
21	352	TIQECISNLNYTVYSHFDCYEISLLT	377	1.155
22	392	KKIYVYTSQGCIHREYISALLE	413	1.113
23	416	YNGLIKIVRN	425	1.088
24	440	IDIIISNVNLPIKNIPIVQIS	460	1.133

Table 2: Antigenic determinants in the Protective antigen precursor protein. Average antigenic propensity for Protective antigen precursor protein is 1.0065.

S.No.	Start	Sequence	End	Score
1	4	RKVLIPLMALSTILVSS	20	1.146
2	22	GNLEVIQAE	30	1.111
3	44	SSQGLLGYYFSDLNFQAPMVVTS	66	1.118
4	72	LSIPSS	77	1.053
5	88	YFQSAIWSGFIKV	100	1.073
6	141	RLYQIKI	147	1.096
7	159	LDFKLYW	165	1.063
8	171	KKEVISSDNLQLPEL	185	1.087
9	212	IPDSLEVEGYTVDV	225	1.083
10	230	TFLSPWI	236	1.068
11	276	VSPEARHPLVAAYPIVHV	293	1.187
12	334	GNAEVHASF	342	1.087
13	345	IGGSVSA	351	1.064
14	358	SSTVAIDHSLSL	369	1.119
15	402	PIYNVLPTTSLVLGK	416	1.121
16	429	QLSQILAPNN	438	1.115
17	441	PSKNLAPIAL	450	1.089
18	475	KQLRLDTD	482	1.064
19	506	WSEVLPQ	512	1.104
20	526	DLNLVERRIAAVNP	539	1.074
21	552	LKEALKI	558	1.065
22	602	NIYTVLDK	609	1.109
23	632	IAVGADESVVKEAHREVIN	650	1.104
24	653	TEGLLLN	659	1.068
25	664	IRKILSGYIVE	674	1.141
26	712	DKLPLYIS	719	1.117
27	723	YKVNVYAV	730	1.154
28	750	IKKILIFS	757	1.094

Table 3: Antigenic determinants in the Lethal factor precursor protein. Average antigenic propensity for Lethal factor precursor protein is 1.0153

S.No.	Start	Sequence	End	Score
1	5	KEFIKVISMSCLVTAITLSGPVFIPLVQG	33	1.198
2	73	MKHIVKI	79	1.137
3	92	AEKLLEKVPSDVLEM	106	1.127
4	111	GGKIYIVD	118	1.075
5	122	TKHISLEA	129	1.098
6	142	GKDALLHEHYVYA	154	1.158
7	157	GYEPVLVIQSSE	168	1.180
8	176	KALNVYYEIGK	186	1.105
9	188	LSRDILSKINQPYQKFLDVLN	208	1.112
10	218	GQDLLFT	224	1.077
11	232	TDFSVEF	238	1.078
12	245	EVQEVFAKAFAYYIE	259	1.102
13	261	QHRDVLQLYAPE	272	1.170
14	285	INLSLEE	292	1.056
15	307	IKQHYQHWS	315	1.069
16	325	LLKKLQIPI	333	1.111
17	338	DDIIHSL	344	1.100
18	352	LKRIQIDSSDF	362	1.089
19	370	FLKKLQI	376	1.088
20	389	LLNRIQVDS	397	1.100
21	408	FLKKLKLDIQPY	419	1.090
22	431	LIDSPSINLDVRK	443	1.073
23	451	NIDALLHQSIGS	462	1.121
24	464	LYNKIYL	470	1.097
25	483	LGADLVD	489	1.085
26	514	SNYMIVD	520	1.046
27	548	ENGKLILQR	556	1.067
28	560	LEIKDVQIIK	569	1.094
29	576	IRIDAKVVPK	585	1.122
30	605	KALGLPKYTKLITF	618	1.095
31	625	ASNIVESAYLIL	636	1.125
32	644	QSDLIKKVTNYLV	656	1.113
33	682	IYEQVHSKGLYVPESRSILLHGP	704	1.102
34	717	FIHEFGHAVDDYAGYLLD	734	1.089
35	737	QSDLVTN	743	1.033
36	769	FFAEAFRL	776	1.041
37	784	ERLKVQK	90	1.065

Table 4: Antigenic determinants in Uncharacterized protein pXO2-14/pXO2_0013 sequence. Average antigenic propensity for Uncharacterized protein pXO2-14/pXO2_0013 protein is 1.0092.

S.No.	Start	Sequence	End	Score
1	12	QSKKALVIGILAVITLFVSFTFTPIAYAFDW	42	1.180
2	44	EFLGLKD	50	1.092
3	52	DTEKVVAF	59	1.159
4	65	IQYADFLGSIWQ	76	1.086
5	82	IIKTLLYIVSALEG	95	1.182
6	102	SFFDILKD	109	1.079
7	116	ASSIIKGLFYALFVLVIVWLGIRTIIQHKP	145	1.270
8	147	RFKSVGVNILVMIGLLGG	164	1.172
9	193	LAWDLVK	199	1.091
10	202	TADLIYLS	209	1.108
11	239	KDIFLKAQLGDVVTPKVIEM	258	1.172
12	268	ETEYLVYK	275	1.135
13	301	PGGYVRY	307	1.056
14	314	IFWGLLALGVAYLFTVFVFVMTIFE	338	1.190
15	341	MKKIIAPIVFVT	352	1.184
16	361	KMVIQDIFKGFLVFAFTGLNLRVYTIIVNY	390	1.145
17	396	INAFLYIVAMVCLTVAL	412	1.236
18	417	ESILKYFGVDVGL	429	1.120
19	535	VAETVGG	541	1.056
20	549	KVSGVAKGVSD	559	1.082
21	561	VKGTVSE	567	1.072
22	722	GSQTVQREVNVLDT	735	1.117
23	749	AKAGVATI	756	1.066
24	824	EAVESAVRAGS	834	1.090
25	851	EAVESSVRAGS	861	1.083
26	884	VRAGATTVQRDV	895	1.097
27	905	SEVTADVRVGT	915	1.123
28	918	VQRDVNVQDNVKE	930	1.097

Table 5: Antigenic determinants in pXO1-13 sequence. Average antigenic propensity for pXO1-13 protein is 1.0223

S.No.	Start	Sequence	End	Score
1	12	KYEDIIK	18	1.039
2	23	GNLIIFVGAGISMLLGL	39	1.138
3	45	YAQEILE	51	1.052
4	54	IEKGVIDS	61	1.096
5	72	PKKILTICKM	81	1.105
6	105	YKAIYEK	111	1.081
7	113	YSMNAIYI	120	1.034
8	124	YDECLDVLTT	133	1.140
9	158	SETKVFIDHTELLE	171	1.064
10	178	NVIHIHGS	185	1.112
11	190	ESMLVTVQDYLER	202	1.138
12	210	SHPEVSVFLDQ	220	1.154
13	222	FNSNYVVLFM	231	1.151
14	252	SHTISHYM	259	1.085
15	273	LYKKYYSNFGVELIPF	288	1.095
16	297	NLIPII	302	1.070
17	304	EWSKVLGE	311	1.098
18	332	DDTAFEVRSRSVIK	345	1.098
19	351	ESLALYFF	358	1.131
20	377	DPEDVPQP	384	1.110
21	401	YLRKVVNQ	408	1.119
22	418	DILGIVKSISF	428	1.108
23	442	VWNQFVEVLEQ	452	1.126
24	472	SMFSVDYIS	480	1.107
25	484	GEKLVPKF	491	1.128
26	505	KIIECLVIL	513	1.207
27	518	NQVKLGK	524	1.043
28	536	VKKIAEKCSMQLLEKIMQQLKEY	558	1.099
29	573	KPYTLEIAKIKK	584	1.064
30	586	YITRIVEHGI	595	1.096
31	631	IEDSVKSIYID	641	1.115
32	647	MLSLQY	652	1.064
33	662	FELILYLFKDLCNY	675	1.155
34	681	MQPFLIN	687	1.085
35	690	NQEYLLFTKILLDVIG	705	1.140
36	723	EGTIVFQEYYLESELKLILEQ	743	1.130
37	751	QESLLLYLIE	760	1.189
38	778	RLKALRHIPLFEKRYLKL	795	1.112
39	797	EITKVDP	803	1.063
40	805	LSSKAKVIGSKFVRIVSPVSQPEL	828	1.179
41	836	IAKYLREYVES	846	1.087
42	850	SITTVYSLGQAIR	862	1.111
43	872	MVEGLKQ	878	1.056
44	881	NTDYYYIHQIILGLCDV	897	1.170
45	908	KLLNFIISY	916	1.098
46	925	GKLLIN	930	1.052
47	935	DYDCKVVISIICKLFV	950	1.235
48	967	KAKQILLLVFKKIIDLE	983	1.199
49	991	SISYFLNSVKGNVLEALLKMS	1011	1.140
50	1027	FKALYNE	1033	1.073
51	1035	LELNVGEAYSLLG	1047	1.085
52	1088	INEDIYILM	1096	1.049

Table 5: Continued

S.No.	Start	Sequence	End	Score
53	1115	TRLSKVLVEIYLR	1127	1.204
54	1149	FIKEFIDVFCFI	1160	1.164
55	1196	DKKLLSKIILLTS	1208	1.142
56	1217	NCDLIKRAIPFVDNYFDVVIELLYVKI	1243	1.204
57	1254	IIGELVDVWVSDLTPHKYCPE	1274	1.167
58	1277	MLYIIKYLYA	1286	1.151
59	1291	ELFELANRVCNK	1302	1.121
60	1305	ENDFIFLRDLF	1315	1.096

Table 6: Antigenic determinants in pXO1-79 sequence. Average antigenic propensity for pXO1-79 protein is 1.0162

S.No.	Start	Sequence	End	Score
1	4	KIIDWVGGLIIHMVKTLLAPIY	25	1.121
2	29	DTIPVLVF	36	1.176
3	62	VMFKFVAFCVLVSVVFTAAKYSAT	85	1.315
4	100	KDLMIISICLWHLDFFYNFVF	120	1.153
5	146	QQIPLLKSGLSD	157	1.097
6	173	YMFIIFGLFIEA	184	1.109
7	190	ANFYYLMRAVTLYVLMLLGP VMVGMWLFPQ	219	1.200
8	221	KQQTLYW	227	1.056
9	232	MGAVLIQAIHAITLWLVITLIG	253	1.160
10	256	QNPIIKLVLLAMFIPVGEIVKSFIGL	281	1.195
11	300	ALASVAGLVKG	310	1.153
12	361	KAGQVGSAL	369	1.067
13	387	PGAMVAGSQVGSALGAAPG	405	1.087
14	408	AGRSVAAVGEGAVALGK	424	1.108
15	471	NSAVASRLKQAF	482	1.094
16	604	FDKSYVNK	611	1.099
17	613	AFASQLALGKVGKAVEG	629	1.109
18	644	QVGALVGAST	653	1.127
19	699	YAQQAA	704	1.033
20	705	ESAGIVQ	711	1.074
21	723	AFASHLSQQL	732	1.095
22	742	GISSAAGIQKAVNG	755	1.069
23	769	QAKQVSNVGNLVQA	782	1.106
24	827	FKAVAQQAAVQAGAMPVQA	845	1.141
25	863	AFASQLTKGLQSHASLGTVSQQVGS	887	1.101
26	889	ITSAIQGVRSHGVAA	903	1.099
27	907	INKAVYAAT	915	1.094
28	923	RAGDLSVAE	931	1.083
29	936	YADASTSVANI	946	1.067
30	955	IPTSAVGR	962	1.085
31	989	TVSNAYFSAASAVQ	1002	1.099
32	1008	AVQRVIS	1014	1.126
33	1031	RHYNVTKGLSYAAGVIGG	1048	1.115
34	1077	QEIQQMV	1083	1.052
35	1091	GAQSVVKGAVRL	1102	1.147
36	1118	GMTQVVSR	1125	1.105
37	1134	RSGQVVYQD	1142	1.153
38	1159	ASPAYIQ	1165	1.076
39	1181	NPNQLVG	1187	1.092
40	1191	QRPLQQKLVFNEPPLTVNH	1209	1.133

Table 7: Antigenic determinants in pXO1-80 sequence. Average antigenic propensity for pXO1-80 protein is 1.0595

S.No.	Start	Sequence	End	Score
1	4	VVAAVGVPAALIMGVIIAVCVILVSLVPF	32	1.290
2	48	EVSMALN	54	1.099
3	56	DWKEVVAF	63	1.098
4	77	PHDAISYF	84	1.099

Table 8: Antigenic determinants in pXO1-123 sequence. Average antigenic propensity for pXO1-123 protein is 1.0369

S.No.	Start	Sequence	End	Score
1	4	KHIVTTTALSFGLIALGG	21	1.121
2	28	AADILRN	34	1.079

Table 9: Antigenic determinants in pXO2-18 sequence. Average antigenic propensity for pXO2-18 protein is 1.0545

S.No.	Start	Sequence	End	Score
1	4	FTIFAMKLLTY	14	1.080
2	17	FFDSVSSAL	25	1.092
3	31	KLQGLGIAVIIFCVCIIAF	49	1.284
4	62	KKWLLYIVVGGVLLW	76	1.210
5	81	FASTVQG	87	1.092

Table 10: Antigenic determinants in sequence pXO2-63. Average antigenic propensity for pXO2-63 protein is 1.0631

S.No.	Start	Sequence	End	Score
1	5	IWITVLLVGTIGVIS	19	1.174
2	22	PNWAVIIKNSLYSRSLGVSTVAG	44	1.125
3	46	ATGSLIHIVYCLIGIGVIISKSILLFNT	73	1.231
4	75	KWIGVAYLLYIGIKLLRS	92	1.203
5	95	QSPAAI	100	1.075
6	123	PKATLFYLAIFTQVI	137	1.151
7	141	TNIFVQSVYGLTVWSVEIL WHMVLVFFLTH	170	1.194
8	172	SVRNYFLSISHWI	184	1.112
9	188	TGTALILLGIRLA	200	1.142

translocating proteins) and the rest were discarded. *Bacillus similarity* proteins are those showing not significant hits. These 10 proteins were taken for further analysis. *B. anthracis specific* proteins are those that brought hits only for anthracis strains. Being the subject of interest, these 11 proteins were taken for further analysis.

Alignment of Identical Proteins: The basic analysis of proteins had shown the repetitions of sequences. Five proteins were found to be unique and the rest were found as repeats of the other five proteins. Therafter, Multiple Sequence Alignment was carried out using ClustalX, which further had shown these sequences 99-100% identical and we successfully obtained 10 protein sequences. The list of novel proteins identified and B. anthracis specific is given in Fig. 5. Tables 1-10 show the peptide sequences of these proteins with antigenic property.

CONCLUSIONS

The computational methodology employed here yielded ten proteins of the *Bacillus anthracis* plasmid pX01 that are novel and unique. These ten proteins are specific to the *B. anthracis* species and are not shared by any *Bacillus cereus* group organism unlike most other proteins that share an ancestral similarity. Additionally these proteins are extracellular in nature and hence these proteins may be helpful in i) Administering immunity to the suspected group of people by inducing both the humoral and cell mediated immune responses to eliminate the host bacterium. ii) Diagnose the presence of *B. anthracis* in the sample. iii) Develop a biosensor kit for detection of the bacterium in the environment. The BRCA1 gene related sequence is a low similarity protein, pointing to the BRCA1 subunit 3. This can be employed in breast cancer studies as BRCA1 is the principle gene involved. The spore germinating protein is found in vast species of *Bacillus* genus and various other bacteria. Thus it can be used as a biological indicator for any of the *bacillus* species.

REFERENCES

1. Michèle Mock and Agnès Fouet, 2001. ANTHRAX Annual Review of Microbiology, 55: 647-671.
2. Nicholas H. Bergman, C. Erica Anderson, E. Ellen Swenson, M. Matthew Niemeyer, D. Amy Miyoshi, and C. Philip Hanna, 2006.Transcriptional Profiling of the Bacillus anthracis Life cycle In Vitro and an Implied Model for Regulation of Spore Formation J. Bacteriol. 6092–6100 188: 17.
3. Hongbin Liu and H. Nicholas Bergman, 2004. Brendan Thomason, Shamira Shallom, Alyson Hazen *et al*. Formation and composition of the Bacillus anthracis endospore J. Bacteriol., 186(1): 164-78. 14679236.
4. Stefan Riedel, M.D. and PhD Anthrax, 2005. a continuing concern in the era of bioterrorism Proc (Bayl Univ Med Cent), 18(3): 234-243.
5. Dixon, T.C., M. Meselson, J. Guillemin and P.C. Hanna, 1999. Anthrax. New England Journal of Medicine., 341(11): 815-826. [PubMed]

6. Okinaka, R., K. Cloud, O. Hampton *et al*., 1999. Sequence, assembly and analysis of pX01 and pX02 Journal of Applied Microbiology, 87: 261-262
7. Sarah C. Taft and A. Alison, 2008. Weiss Neutralizing Activity of Vaccine-Induced Antibodies to Two Bacillus anthracis Toxin Components, Lethal Factor and Edema Factor Clin Vaccine Immunol. January; 15(1): 71-75.
8. Mingtao Zeng, Qingfu Xu and E. Michael, 2007. Pichichero Protection against Anthrax by Needle-Free Mucosal Immunization with Human Anthrax Vaccine. Vaccine. 4; 25(18): 3588-3594.
9. Leppla, S.H., J.B. Robbins, R. Schneerson and J. Shiloach, 2002. Development of an improved vaccine for anthrax. J. Clin. Investig., 110: 141-144.
10. Hemalatha, G.R., D. Satyanarayana Rao and L. Guruprasad, 2007. Identification and Analysis of Novel Amino-Acid Sequence Repeats in Bacillus anthracis str. Ames Proteome Using Computational Tools Comp Funct Genomics., pp: 47161.
11. Read, T.D., S.N. Peterson, N., Tourasse, *et al*., 2003. The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. Nature.; 423(6935): 81-86. [PubMed]
12. Karsten Hiller, Andreas Grote, Maurice Scheer, Richard Münch and Dieter Jahn, 2004. PrediSi: prediction of signal peptides and their cleavage positions Nucleic Acids Research, pp: 32.
13. Ron D. Appel, Amos Bairoch and F. Denis Hochstrasser, 1994. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server Trends in Biochemical Sciences, 19(6): 258-260.
14. Miklos Cserzo, 1997. Erik Wallin, Istvan Simon, Gunnar von Heijne, Arne Elofsson Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: Application of the Dense Alignment Surface method. Protein Engineering, 10(6): 673-676.
15. Krogh, A., B. Larsson, G. von Heijne and E.L.L. Sonnhammer, 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology, 305(3): 567-580.
16. Rice, P., I. Longden and Bleasby A. EMBOSS, 2000. The European Molecular Biology Open Software Suite Trends in Genetics, 16(6): 276-277.
17. Stephen F. Altschul, L. Thomas Madden, A. Alejandro Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller and J. David Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs Nucleic Acids Research, 25(17): 3389-3402.
18. Thompson, J.D., T.J. Gibson and D.G. Higgins, 2002. Multiple sequence alignment using ClustalW and ClustalX. (Aug-) Current protocols in bioinformatics, Chapter 2: Unit 2.3.