Interdiscip Sci Comput Life Sci (2013) 5: 127–135
DOI: 10.1007/s12539-013-0160-2

Characterization of *Bacillus thuringiensis* Cry1 Class Proteins in Relation to Their Insecticidal Action

Brindha SENTHIL KUMAR, Zothansanga RALTE, Ajit Kumar PASSARI, Vineet Kumar MISHRA, Biku Moni CHUTIA, Bhim Pratap SINGH, Gurusubramanian GURUSWAMI, Senthil Kumar NACHIMUTHU*
(Departments of Biotechnology and Zoology, Mizoram University, Aizawl 796004, Mizoram, India)

Received 14 June 2012 / Revised 27 September 2012 / Accepted 1 November 2012

Abstract: Thirty nine *Bt* Cry1 subgroup protein sequences were retrieved from NCBI and analyzed for physicochemical properties, active site and relationship in relation to their variations in toxicity. Cry1 proteins were found to be hydrophilic and stable. SOSUI server predicted presence of two transmembrane regions in Ag and a single transmembrane region from Aa to Ae. EMBOSS PepWheel tool analysis of the transmembrane regions showed that there were 23 highly conserved residues towards the N terminal which are hydrophobic and more than half of the residues were neutrally charged. No signal peptide was detected which classifies the Cry1 group proteins as non-secretory proteins. Cry1 proteins have very high composition of neutral amino acids and might transform into negative charge after solubilization in alkaline environment (insect midgut). The negatively charged protein might misfold causing difficultly to digest and thereby be toxic to lepidopteran. Active sites of Cry1 proteins with more than 50% neutral amino acids showed wide insecticidal spectrum and further positive correlation ($r = 0.7731$) was observed between neutral amino acids and insect species affected ($Y = -138.21 + 2.907X$). Similarity of sequences was found between Cry1 proteins based on their high or low spectrum of insecticidal activity.
Key words: *Bacillus thuringiensis*, neutral amino acids, Cry1 protein, receptor site, evolutionary relationship, lepidoptera, protein misfolding, insecticidal spectrum.

1 Introduction

Bacillus thuringiensis (*Bt*) produces insecticidal crystal proteins during its sporulation phase which becomes active in the insect midgut (Rowe and Margaritis, 1987). Different classes of Cry proteins having structural and functional similarities have been characterized and are known to have high specificity against particular group of insects (Schnepf *et al.*, 1998; Crickmore *et al.*, 1998). Cry1 toxins are the largest family among more than 200 crystal proteins characterized so far and cryptic genes are known to exist among Cry toxins (Masson *et al.*, 1998; Letowski *et al.*, 2005). Cry1 protein has specific insecticidal activity against lepidopteran insects and this activity is also found in Cry2 and Cry9 proteins (Schnepf *et al.*, 1998; Nariman, 2007; Bobrowski *et al.*, 2002). Currently, there are about 12 different subclasses of Cry1 protein group, each subclass has a specific range of activity against different lepidopteran insects.

Bt has proved itself to exist in diverse environments and has been shown to have interrelationships between habitat and their biological characteristics such as serovar, Cry proteins and toxicity (Martinez and Caballero, 2002). Moreover, PCR detection of *cry*1 genes indicated that Ab, Ac, C, D, I were present both on chromosomal and plasmid DNA, whereas Aa was present only on plasmid DNA (Thaphan *et al.*, 2008). In the present study, *in silico* analysis of the Cry1 protein subclasses and their relationship with insecticidal spectrum has been carried out.

2 Materials and methods

Cry1 subclass proteins sequences were retrieved from the National Centre for Biotechnology Information (NCBI) database and Crickmore's webpage (www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/intro.html) in FASTA format and used for analysis.

2.1 Computational tools and servers

ClustalW and BLASTp programs were used on individual subclass sequences for searching consensus sequences within each subclass. In total, 133 Cry1 sequences were retrieved; out of which 38 sequences were chosen, representing one consensus sequence per

*Corresponding author.
E-mail: nskmzu@gmail.com

Table 1 Physicochemical properties of Cry1 proteins

S. No.	Protein (Cry1)	Acc. No.	Half Life	Ext. coefficient T 280 nm	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Aliphatic Index	Grand average of hydropathicity (GRAVY)	Molecular weight Daltons	Theoretical pI	Hydrophobic %	Hydrophilic %
1.	Aa	AAP40639.1	> 10 hrs	180220	158	108	82.22	−0.405	133683.11	−48.77	44.55	48.61
2.	Ab	CAA38701.1	> 10 hrs	177075	151	106	81	−0.427	130622.57	−43.47	43.72	49.26
3.	Ac	AAA73077.1	> 10 hrs	191430	152	107	82.23	−0.402	133295.705	−43.75	44.74	48.3
4.	Ad	AAA22340.1	> 10 hrs	177575	153	106	82.42	−0.395	133376.09	−45.71	44.53	48.35
5.	Ae	AAA22410.1	> 10 hrs	183660	156	107	80.19	−0.428	133737.12	−47.77	44.03	48.77
6.	Af	AAB82749.1	> 10 hrs	126140	103	87	84.73	−0.35	103180.07	−14.59	44.13	49.29
7.	Ag	AAD46137.1	> 10 hrs	186055	151	108	84.52	−0.364	133392.405	−42.01	45.58	47.45
8.	Ah	AAQ14326.1	> 10 hrs	199660	151	113	81.88	−0.387	134030.775	−36.63	44.5	48.65
9.	Ba	CAA29898.1	> 10 hrs	194785	171	117	79.63	−0.491	139646.445	−52.03	43.49	50
10.	Bb	AAA22344.1	> 10 hrs	200160	162	112	79.09	−0.48	139769.48	−48.27	43.78	49.96
11.	Bc	CAA86568.1	> 10 hrs	203265	158	117	78.84	−0.501	140451.3	−39.24	43.23	50.45
12.	Bd	AAD10292.1	> 10 hrs	200160	160	111	80.48	−0.461	139654.22	−47.17	43.95	49.8
13.	Be	AAC32850.1	> 10 hrs	194200	163	114	80.42	−0.45	139083.84	−47.27	43.52	50.37
14.	Bf	CAC50778.1	> 10 hrs	197180	164	111	9.47	−0.469	139306.925	−50.87	43.16	50.73
15.	Ca	AYO78160.1	> 10 hrs	175875	166	115	83.6	−0.434	134714.46	−49.92	44.41	48.36
16.	Cb	AAG35409.1	> 10 hrs	175875	158	110	81.7	−0.424	132867.205	−46.63	44.05	48.72
17.	Da	X54160.1	> 10 hrs	179775	156	113	84.27	−0.384	132480.31	−41.64	45.49	48.24
18.	Db	CAA80234.1	> 10 hrs	173940	156	108	83.62	−0.372	130969.5	−46.5	45.52	48.1
19.	Ea	M73252.1	> 10 hrs	174720	165	115	85.9	−0.421	133265.95	−48.66	44.06	49.7
20.	Eb	M73253.1	> 10 hrs	170710	159	113	85.22	−0.408	133601.465	−44.47	43.95	49.49
21.	Fa	M63897.1	> 10 hrs	182740	162	115	81.27	−0.452	133621.105	−45.47	43.95	49.49
22.	Fb	Z22512.1	> 10 hrs	175540	155	111	1.35	−0.455	133349.925	−42.1	43.61	49.49
23.	Ga	YO9326.1	> 10 hrs	187670	152	124	82.12	−0.437	133723.61	−26.68	44.84	48.85
24.	Gb	U70725.1	> 10 hrs	161450	165	119	82	−0.44	132904.49	−44.69	44.65	48.85
25.	Ha	Z22513.1	> 10 hrs	175180	161	113	86.42	−0.398	132979.585	−46.27	44.97	48.81
26.	Hb	U35780.1	> 10 hrs	172450	156	113	84.25	−0.449	131061.37	−41.4	44.59	48.83
27.	Ia	AJ315121.1	> 10 hrs	113930	78	65	81.49	−0.347	81202.52	−11.25	44.23	49.65
28.	Ib	EU233027.1	> 10 hrs	113805	77	70	83.41	−0.352	81294.8	−5.38	44.37	49.65
29.	Ic	AAC62933.1	> 10 hrs	113805	73	72	82.17	−0.376	81209.8	0.71	43.81	49.93
30.	Id	AFO47579.1	> 10 hrs	81402.59	75	62	79.19	−0.368	81402.59	−11.61	44.09	50.21
31.	Ie	AF211190.1	> 10 hrs	112315	77	66	84.62	−0.304	81024.38	−9.28	44.51	49.65
32.	If	AAQ52382.1	> 10 hrs	112315	71	63	78.94	−0.375	80074.255	−6.19	43.38	50.14
33.	Ja	AAA22341.1	> 10 hrs	157885	164	115	82.93	−0.422	132759.52	−47.23	43.96	49.44
34.	Jb	AAA98959.1	> 10 hrs	166045	168	122	79.56	−0.468	133552.595	−44.75	44.62	48.97
35.	Jc	AAQ52372.1	> 10 hrs	175305	163	122	82.96	−0.477	133277.215	−39.43	44.35	48.89
36.	Jd	CAC50779.1	> 10 hrs	175555	158	122	80.63	−0.456	133028.115	−34.46	44.35	48.97
37.	Ka	AAB00376.1	> 10 hrs	188005	159	115	79.98	−0.445	137378.12	−42.12	43.62	49.47
38.	La	AAS60191.1	> 10 hrs	171085	163	124	84.7	−0.419	133290.245	−37.66	44.02	50
Average	–	–	> 10 hrs	168393	144	105	78	−0.418	125217	−37.38	44.21	49.26

subclass, for further analysis (Table 1). The physicochemical parameters, theoretical pI, molecular weight, total number of positive and negative residues, extinction co-efficient (Gill and van Hippel, 1989), aliphatic index and instability index (Ikai, 1980), GRAVY (Kyte and Doolittle, 1982), composition of charged amino acids, hydropathicity, hydrophilicity and net hydrophobic content and half life (Bachmair *et al.*, 1986; Gonda *et al.*, 1989; Tobias *et al.*, 1991; Ciechanover and Schwartz, 1989) were computed using PROT-PROP software (Brindha *et al.*, 2012).

Identification of transmembrane region was done using SOSUI server (Takatsugu *et al.*, 1998). The predicted transmembrane region was visualized and analyzed using helical wheel generated by EMBOSS PepWheel program. The 3D structures of Cry 1 subclasses were generated by homology modeling using Swiss Model. The model 3D structures were validated using RAMPAGE (Lovell *et al.*, 2003). The CastP sever was used to visualize the active sites (Joe *et al.*, 2006). Active sites with maximum area and volume were only taken into consideration for analysis of neutral amino acids. FATCAT server (Ye and Godzik, 2003) was used to align the structures based on pairwise alignment. The secondary structure of the active site amino acid sequences was obtained from YASPIN (Lin *et al.*, 2005).

The number of insects screened against Cry1 subfamily of proteins was obtained (Diego and Graciela, 2008) (Table 2). Cry1 subgroup proteins were divided into three categories based on their insecticidal spectrum: high spectrum (Aa, Ab, Ac), moderate spectrum (Da, Ea, Fa) and no spectrum (Ga, Ha, La). The selected sequences were clustered using MEGA 5.0 software for understanding their evolutionary relationship (Tamura *et al.*, 2007). All the aligned sequences were manually checked and edited to minimize the gaps. Cry2 and Cry9 sequences were used as an out group as they are also toxic to lepidopteran insects (Nariman, 2007). The SPSS software was used to determine regression correlation between percentage of neutral amino acids and groups of insect affected (SPSS for Windows).

3 Results and discussion

3.1 Physicochemical analysis

The molecular weight of lepidopteran-active Cry1 protein is 125 kDa (Table 1). Extinction co-efficient at 280 nm is 168,393 due to the poor concentration of tryptophan and cysteine residues and cannot be analyzed using UV spectral methods (Gill and van Hippel, 1989). The instability index (II) and Aliphatic Index of Cry1 proteins are less than 35 and 78, respectively and are stable for wide range of temperatures. The average theoretical pI is very low at 7.0 pH (−37.38) indicating that Cry1 proteins are acidic, resulting in higher number of negatively charged residues than positively charged residues (Vaseeharan and Valli, 2011) (Table 1; Fig. 1). Cry1 proteins are not strongly hydrophilic (GRAVY score: −0.418) resulting in moderate interaction in water and hence, they are highly soluble in the alkaline environment.

The Cry1 subclass proteins are rich in neutral amino acids with a net composition of 541 amino acid residues. The neutral amino acid residues of Cry1 protein comprise leucine (95), valine (73), isoleucine (69), threonine (77), serine (80), asparagine (74) and glycine (73) (Fig. 1). They are hydrophobic (237 amino acid residues), hydrophilic (231 amino acid residues) and aliphatic (73 amino acid residues) with a ratio of 3:3:1 (Fig. 2 and Fig. 3). According to Le Chatelier Principle, the neutral amino acids might donate a hydrogen atom and become negatively charged in alkaline environment (insect midgut) (Fig. 4). This may lead to significant increase in net hydrophilic-negatively charged residues in Cry1 group proteins. Solubilization of Cry proteins in alkaline environment is one of the significant steps for exhibiting toxicity against lepidopteran insects (Devendra *et al.*, 2006). After solubilization, the high negatively charged protein tends to misfold which

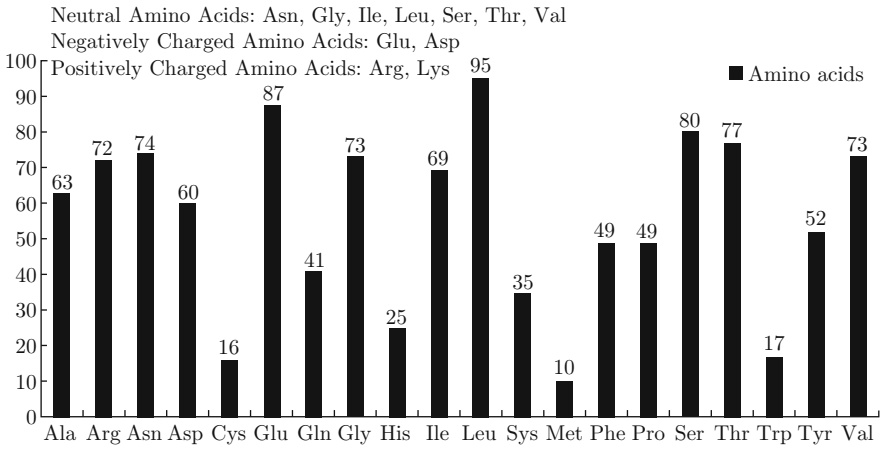

Fig. 1 Average amino acid composition of Cry1 subclass proteins.

Table 2 Insecticidal spectrum of Cry1 class proteins

Lepidopteran	Cry1 subclass Proteins
Plutella xylostella	Aa, Ab, Ac, Ad, Ba, Bb, Bd, Be, Bf, Ca, Da, Ea, Eb, Fa, Gb, Gc, Ia, Ib, Id, Ie, If, Ja, Jc
Trichoplusia ni	Aa, Ab, Ac, Ad, Ae, Ba, Bb, Bd, Be, Bf, Ca, Cb, Ea, Eb, Fa, Gc, If, Jb
Ostrinia nubilalis	Aa, Ab, Ac, Ba, Bb, Be, Bf, Fa, Gc, Ia, If, Ja, Jb, Jc
Conopomorpha cramerella	Aa, Ab, Ac, Ba, Ca, Cb, Dd, Ea, Fa, Ia
Heliothis virescens	Aa, Ab, Ac, Ae, Be, Ca, Fa, If, Ja, Jc
Choristoneura fumiferana	Aa, Ab, Ac, Ad, Ba, Ca, Da, Ea, Fa
Cacyreus marshalli	Aa, Ab, Ac, Ba, Ca, Da, Ea, Fa, Ja
Lymantria dispar	Aa, Ab, Ac, Ba, Bb, Ca, Da, Ja
Bombyx mori	Aa, Ab, Ac, Da, Db, Ea, Fa, Ia
Choristoneura rosaceana, Pandemis pyrusana, Platynota stultan	Aa, Ab, Ac, Ad, Ca, Ea, Eb, Fa
Spodoptera exigua	Ab, Ad, Be, Ca, Da, Fa, If, Ja
Malacosoma disstria, Manduca sexta	Aa, Ab, Ac, Ba, Ca, Da, Ea
Sesamia nonagrioides	Aa, Ab, Ac, Ba, Ca, Da, Fa
Pectinophora gossypiella	Aa, Ab, Ac, Bb, Ca, Da, Ja
Choristoneura occidentalis	Aa, Ab, Ac, Ad, Ca, Ea, Fa
Cydia pomonella	Aa, Ab, Ac, Ba, Da, Fa, Ia
Argyrotaenia citrana	Aa, Ab, Ac, Ca, Ea, Eb, Fa
Helicoverpa zea	Aa, Ab, Ac, Be, If, Ja, Jc
Pseudoplusia includes	Aa, Ab, Ac, Ba, Bb, Jb
Helicoverpa armigera	Aa, Ab, Ac, Fa, Ia
Orgyia Leu costigma	Aa, Ab, Ac, Ca, Da
Pieris brassicae, Chilo supperssalis, Epinotia aporema	Aa, Ab, Ac, Ba, Ca
Thaumetopoea pityocampa	Aa, Ab, Ac, Ba
Earias vitella, Marasmia patnalis	Aa, Ab, Ac
Spodoptera frugiperda	Bb, Be, Ca, Da, Fa, If, Ja
Spodoptera littoralis	Ca, Ea, Fa
Helicoverpa punctigera	Ab, Ac, Fa
Mamestra brassicae	Aa, Ab, Ac
Artogeia rapae	Ba, Ia, Ka
Cnaphalocrocis medinalis, Choristoneura pinus pinus	Aa, Ab, Ac
Sciropophaga incertulas	Aa, Ac, Ca
Maruca vitrata	Aa, Ab, Ca
Diatraea saccharalis	Aa, Ab
Agrotis ipsilon	Gc, Jc
Eldana saccarina, Busseola fusca, Danaus plexippus, Sesamia calamistis	Ab, Ac
Spodoptera exempta	Ca, Ea
Spodoptera litura	Db, Ia
Plodia interpunctella	Ab, Ca
Ephestia kuehniella, Rachiplusia nu, Agrotis segetum, Anticarsia gemmatalis, Planotortrix octo, Ctenopusestis obliquana, Ephiphyas postvitta, Phthorimaea operculella, Tecia solanivora, Spilosoma virginica	Ac
Hyphantria cunea	Aa
Mamestra configurata	Ab
Lambia fiscellaria	Da
Wiseana cervinata, Wiseana copularis, Wiseana jocose	Ba
Ostrinia furnacalis	Ie

causes difficulty for the insect larvae to assimilate the misfolded protein resulting in death of the larvae (Coen *et al.*, 2007). One main factor for causing protein misfolding is strong interaction of negatively charged residues with membrane phospholipids (Tran *et al.*, 2001). The widest spectrum insecticidal Cry 1 proteins are Ac, Ab, and Aa which are rich in tryptophan and valine residues; threonine and serine; and glycine and asparagines, respectively. These amino acid residues might contribute to the toxicity against wide range of insects because valine, threonine, serine, glycine are neutral amino acids and asparagines is negatively charged.

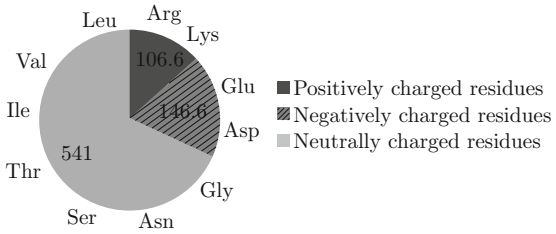

Fig. 2 Net composition of charged and neutral amino acids.

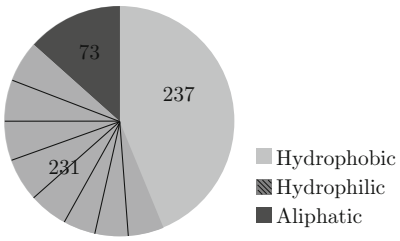

Fig. 3 Properties of neutral amino acids in Cry1 proteins.

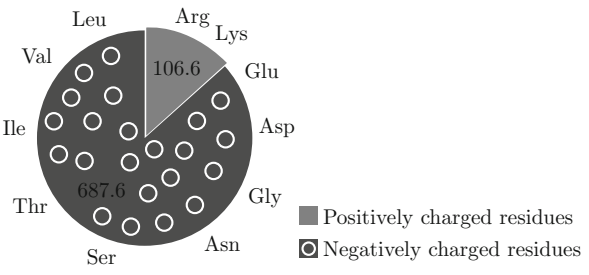

Fig. 4 Neutral amino acids after solubilization in insect midgut (alkaline environment).

One of the most increased and effectively absorbed amino acids in lepidopteran larval midgut is leucine (Giordana *et al.*, 2002) and Cry1 subgroup proteins are very rich in leucine (neutral amino acid) which probably forms a good binding for Cry1 proteins in the larval midgut (Fig. 1). The toxic nature of leucine was enhanced in magainin 2, frog antimicrobial peptide, by using leucine zipper like motifs (Pandey, 2011).

3.2 Transmembrane and signal peptide analysis

Even though all Cry1 group proteins are membrane proteins, 29% of them have one transmembrane region (Aa, Ab, Ac, Ad, Ae and Ga). Cry1Ag1 has two transmembrane regions (Table 3). The presence of transmembrane domain in Cry1A subgroups may play a major role in the stability of secondary and tertiary structures of proteins (Hormaeche *et al.*, 2006). This along with the presence of abundant neutral amino acids in the transmembrane regions might also contribute to the wide range of insecticidal spectrum of Cry1A group.

Table 3 Transmembrane regions of Cry proteins

Name	Transmembrane
Cry1Aa	QFLLSEFVPGAGFVLGLVDIIWG
Cry1Ab	QFLLSEFVPGAGFVLGLVDIIWG
Cry1Ac	QFLLSEFVPGAGFVLGLVDIIWG
Cry1Ad	QFLLSEFVPGAGFVLGLVDIIWG
Cry1Ae	QFLLSEFVPGAGFVLGLVDIIWG
Cry1Ag	QFLLSEFVPGAGFVLGLVDIIWG AAFAVYTLRASLFLLVLLIHAR
Cry1Ga	LLEAAVPEAGFALGLFDIIWGALG

Cry1 subclass proteins do not show presence of signal peptide which classifies them as non-secretory proteins. One of the properties of non-secretory proteins in bacteria is low composition of cysteine and leucine residues (Garg and Raghava, 2008). Though Cry1 subgroup proteins are non-secretory proteins and the composition of cysteine is low, leucine is very high which is effectively absorbed in the insect midgut.

3.3 Active site analysis

The 3D structures having scores greater than 98% accuracy in RAMPAGE were only used for active site analysis. The result of regression analysis of Cry1 subgroup proteins ($Y = -138.21 + 2.907X$) revealed that the number of insect species affected (Y) was positively correlated with the neutral amino acids present in receptor site of Cry1 proteins (X), having a regression coefficient $r = 0.7731$, $r^2 = 0.5977$, SD of residuals -11.653, $F_{1,37} = 54.97$, $p < 0.0001$. The results of regression analysis (95% confidence level) revealed that the efficacy of the Cry1 subclass protein increased with the number of neutral amino acids present in the receptor site. Higher slope value (2.907) and lower fiducial limits at 95% were observed for insect groups affected (Fig. 5). Interestingly, active sites which have 50% or more of neutral amino acids have high toxic spectrum activity against lepidopteran insects as shown with Cry1A (54-73% of lepidopterous insects affected) and 1C (43% of lepidopterous insects affected) subclass proteins. The Cry1B, 1F, 1E and 1D subclasses showed medium (22-30%) toxic spectrum activity followed by 1J (14%) and 1I (12%). Low toxic spectrum activity

was observed in Cry1G (2-6%). The Cry1 subclass proteins having < 50% of neutral amino acids did not exhibit toxicity against lepidopterous insects and can be designated as non-toxicants (Af, Ag, Ah, Bc, Fb, Ga, Ha, Hb, Ic, Jd, La) (Fig. 6). This shows that the neutral amino acids may play vital role in Cry1 protein toxicity. Elimination of the trypsin cleavage site of *Bt* wild-type protoxin at position 164 by substitution with a neutral amino acid resulted in increased toxicity against lepidopteran insects (Lambert *et al.*, 1996). Cry1Aa and Ac proteins comprised of 58 % and 67% neutral amino acids in the active site showing a toxic spectrum activity of 54 % and 73%, respectively (Fig. 6 and Table 5). There are interesting signature sequences in different Cry protein sequences according to their insecticidal spectral activity (Fig. 7).

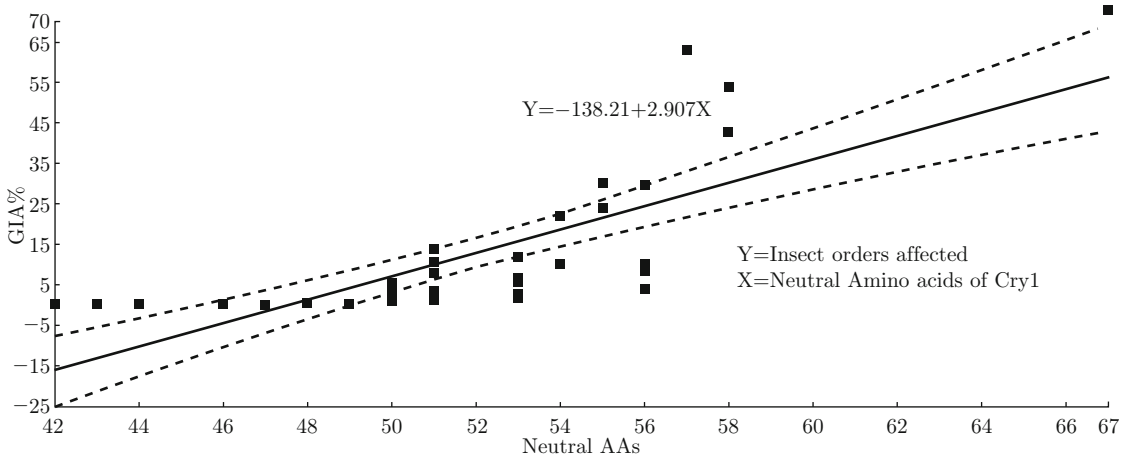

Fig. 5 Correlation between neutral amino acids of Cry1 (neural AAs) and insect orders affected (% GIA). Solid line: linear regression; Dashed line: 95% confidence limit.

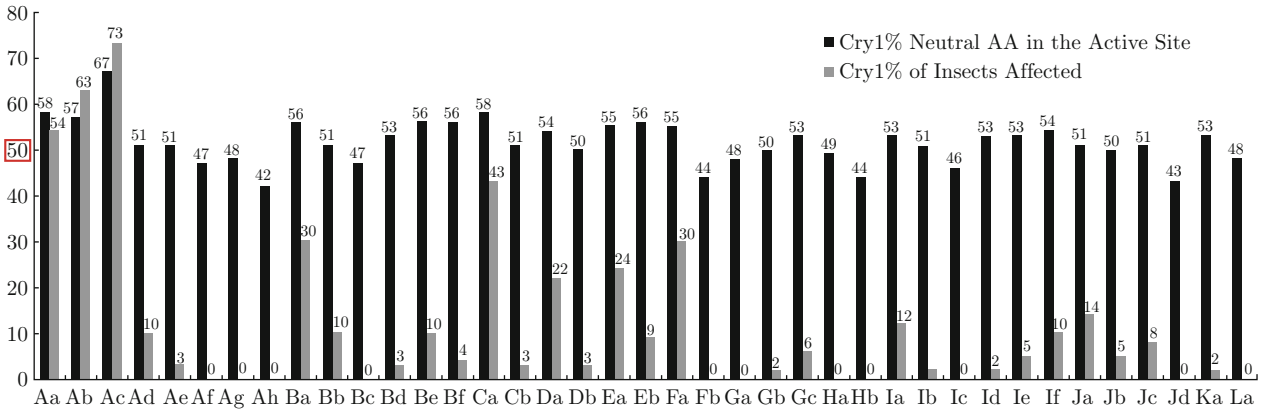

Fig. 6 Percentage of amino acids in active site and insects affected.

Insect spectrum range	Cry1	Active site Amino acids Alignment
High spectrum	Aa	VYNLYNRRTLDALTREIYEIRQFHMHTTSGT----KGFGTGGD 39
	Ac	VYNLY--RTLDDLTREIYEIRSFHMDSIVS------G-GTGGD 34
	Ab	VYNLEY-RTLDS-TREIYEIRSFHMDILQRLSIFTKGFGTGGD 41
		**** **** ********.***. * *****
Moderate spectrum	Ea	---VYNL--RTLDSLREVY-IRSPHMD----------SVKG--TGGD---- 29
	Fa	---LYNLNRRTLDATREIYEVRPPHMDNHST----------AVRGPGTGGD---- 38
	Da	TNVDYNLRYRTLDATREVYSIRSPHVDFLFSSVPIRLCHTVFSTAVKKGPGFTGDPLTS 59
		*** ****: **:* :*.**:* :* **
Nil spectrum	Ha	-AFNL--------RYL-QRTLDA-TREIYQ----------LREPHMDCHRVST----- 32
	Ga	RAYNLNEIGGISRRYLDQRTLDA-TREIYTSPVANINFLSIALRAPHMDDQVTDTDECLA 59
	La	----------YLIQRTLDALTREIFS-----VVVAYEVRRAPHFCGVVVSG----- 36
		** ****** ****: * **: ..

Fig. 7 Active site amino acids alignment of high, moderate and nil spectral insect ranges.

Table 4 FATCAT pairwise alignment for 3D structures

Structures	RMSD	P-VALue (P-**VAL**ue < 0.05 are significantly similar)	Alignment Model
Cry1Ac with Cry1La	0.92	0.00	Flexible & Rigid
Cry1Ac with Cry1Fa	0.61	0.00	

Table 5 Amino acids involved in the active sites of Cry1 proteins

Insect Spectrum Range	Cry1	Amino Acids Involved in Active Site
High Spectrum	Aa	**VAL**208 TYR211 **ASN**212 **LEU**215 TYR229 **ASN**230 ARG233 ARG234 **THR**237 **LEU**241 ASP242 ALA245 **LEU**263 **THR**264 ARG265 GLU266 **ILE**267 TYR268 GLU288 **ILE**291 ARG292 GLN293 PRO294 HIS295 MET297 HIS433 **THR**435 **THR**451 **SER**453 **GLY**455 **THR**476 LYS489 **GLY**490 PRO491 **GLY**492 **THR**494 **GLY**495 **GLY**496 ASP497
	Ab	VAL208 TYR211 ASN212 LEU215 GLU216 TYR229 ARG233 THR237 LEU241 ASP242 SER245 THR264 ARG265 GLU266 ILE267 TYR268 GLU288 ILE291 ARG292 SER293 PRO294 HIS295 MET297 ASP298 ILE299 LEU300 GLN353 ARG429 LEU430 SER431 ILE456 PHE462 THR477 LYS490 GLY491 PRO492 GLY493 THR495 GLY496 GLY497 ASP498
	Ac	**VAL**208 TYR211 **ASN**212 **LEU**215 TYR229 ARG233 **THR**237 **LEU**241 ASP242 ASP245 **LEU**263 **THR**264 ARG265 GLU266 **ILE**267 TYR268 GLU288 **ILE**291 ARG292 **SER**293 PRO294 HIS295 MET297 ASP298 **SER**431 **ILE**455 **VAL**476 **SER**488 **GLY**489 **GLY**491 **THR**493 **GLY**494 **GLY**495 ASP496
Moderate Spectrum	Ea	**VAL**206 TYR209 **ASN**210 **LEU**213 ARG232 **THR**236 **LEU**240 ASP241 **SER**244 **LEU**263 ARG264 GLU265 **VAL**266 TYR267 **ILE**291 ARG292 **SER**293 PRO294 HIS295 MET297 ASP298 **SER**427 **VAL**447 LYS483 **GLY**484 **THR**488 **GLY**489 **GLY**490 ASP491
	Da	**THR**203 **ASN**204 **VAL**207 ASP208 TYR210 **ASN**211 **LEU**214 ARG215 TYR228 ARG232 **THR**236 **LEU**240 ASP241 ALA244 **THR**263 ARG264 GLU265 **VAL**266 TYR267 **SER**291 **ILE**294 ARG295 **SER**296 PRO297 HIS298 **VAL**300 ASP301 PHE302 **LEU**303 PHE327 **SER**356 **SER**358 **VAL**359 PRO360 **ILE**361 ARG422 **LEU**423 CYS424 HIS425 **THR**427 **VAL**441 PHE442 **SER**443 **THR**445 ALA449 **VAL**466 LYS467 LYS479 **GLY**480 PRO481 **GLY**482 PHE483 **THR**484 **GLY**485 ASP487 PRO549 **LEU**550 **THR**551 **SER**552
	Fa	**LEU**206 TYR209 **ASN**210 **LEU**213 **ASN**228 ARG231 ARG232 **THR**235 **LEU**239 ASP240 ALA243 **THR**262 ARG263 GLU264 **ILE**265 TYR266 GLU288 **VAL**291 ARG292 PRO293 PRO294 HIS295 MET297 ASP298 **ASN**424 HIS425 **SER**447 **THR**449 ALA453 **VAL**470 ARG483 **GLY**484 PRO485 **GLY**486 **THR**488 **GLY**489 **GLY**490 ASP491
Nil Spectrum	Ha	ALA210 PHE213 **ASN**214 **LEU**217 ARG226 TYR227 **LEU**228 GLN231 ARG232 **THR**235 **LEU**239 ASP240 ALA243 **THR**262 ARG263 GLU264 **ILE**265 TYR266 GLN282 **LEU**285 ARG286 GLU287 PRO288 HIS289 MET291 ASP292 CYS425 HIS426 ARG428 **VAL**446 **SER**448 **THR**450 **VAL**471 LYS484 **GLY**485 PRO486 **GLY**487 **THR**489 **GLY**490 **GLY**491 ASP492
	La	TYR213 **LEU**223 **ILE**227 GLN230 ARG231 **THR**234 **LEU**238 ASP239 ALA242 **LEU**260 **THR**261 ARG262 GLU263 **ILE**264 PHE265 **SER**267 **VAL**270 **VAL**271 **VAL**277 ALA278 TYR279 GLU280 **VAL**282 ARG283 ARG284 ALA285 PRO286 HIS287 PHE289 CYS422 **GLY**427 **VAL**441 **VAL**443 **VAL**468 **SER**480 **GLY**481 **GLY**483 **THR**485 **GLY**486 **GLY**487 ASP488
	Ga	ARG83 ALA206 TYR209 **ASN**210 **LEU**213 **ASN**214 GLU215 **ILE**216 **GLY**217 **GLY**218 **ILE**219 **SER**220 ARG221 ARG222 TYR223 **LEU**224 ASP225 GLN227 ARG228 **THR**231 **LEU**235 ASP236 ALA239 **THR**259 ARG260 GLU261 **ILE**262 TYR263 **THR**264 **SER**265 PRO266 **VAL**268 ALA269 **ASN**271 **ILE**272 **ASN**273 PHE274 **LEU**276 **SER**277 **ILE**278 ALA279 **LEU**282 ARG283 ALA284 PRO285 HIS286 MET288 ASP289 ASP349 GLN351 **VAL**377 **THR**397 ASP398 **THR**399 ASP401 GLU402 CYS419 **LEU**424 ALA436 **ILE**438 PHE439 TRP441 **THR**442 **VAL**463 **VAL**475 LYS476 **GLY**477 PRO478 **GLY**479 **THR**481 **GLY**482 **GLY**483 ASP484 **ILE**485 ASP546 **LEU**547 TYR585

3.4 Structure and sequence alignment

The structures of three Cry1 sequences which have high (Ac), moderate (Fa) and no (La) insecticidal spectrum were taken for comparing structure and sequence via their alignments. The N-terminal and catalytic domain regions of high insecticidal spectrum Cry1 proteins (Aa, Ab, Ac) were highly conserved, whereas there was very few conserved regions between Ac, Fa and La sequences (data not shown). The Cry1 sequence alignments show that Ac and Fa have higher similarities than Ac and La. But, Ac with Fa and Ac with La structures were significantly similar having p-values 0.00 (less than 0.5; without twists). The Root Mean Square Deviation (RMSD) values are 0.92 and 0.61 for the above case for both rigid and flexible alignment models (Table 4). Hence, Cry1Ac exhibits toxicity based on the active site and its neutral amino acid composition but not based on structures. The secondary structures of active site sequences were similar in high, moderate and no insecticidal spectrum crystal proteins.

3.5 Evolutionary tree analysis

The proteins clustered according to their toxicity spectrum. Cry1 subgroup radiated into six different clades based on their toxic spectrum activity against lepidopterous insects. Clade 1 consists of high spectrum activity Cry proteins (Aa, Ab, Ac) and are grouped together. The Clades 2, 3 and 4 consisted of Cry1 proteins Ea, Da and Fa having moderate spectrum activity. The Cry1 proteins having nil spectrum activity branched into Clades 5 (La, Ha) and 6 (Ga) (Fig. 8). The outgroups (Cry2Aa and Cry9Aa) rooted separately. The sequence analysis of Cry1 proteins showed that the subclasses were more dissimiliar to each other. The branch lengths reflect the number of changes taken to evolve the proteins and nodes representing the ancestral divergence points. The relationship between Cry1 toxic spectrum activity and composition of neutral amino acid as mentioned above was in corroboration with the phylogenetic tree. The high, medium and nil range Cry1 proteins have unique sequences in their active site and they also cluster in the phylogenetic tree within themselves according to their insecticidal spectral range.

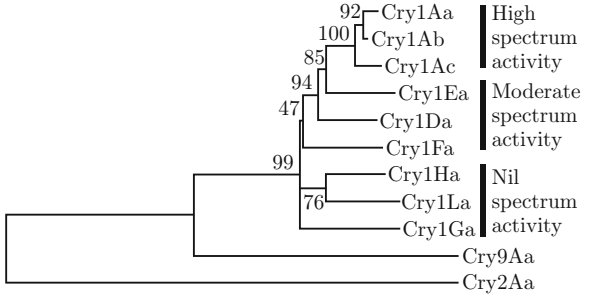

Fig. 8 Evolutionary relationship selected Cry1 proteins by Maximum Likelihood method based on the Poisson correction model (500 replicates).

4 Conclusion

This study reveals that the neutral amino acids may play a vital role in the Cry1 subgroup proteins to exhibit toxicity in the alkaline environment of insect midgut. Based on neutral amino acid residues, the Cry1 subclass proteins are divided into toxicants ($> 50\%$ neutral amino acids) and non-toxicants ($< 50\%$ neutral amino acids). The amino acid residues are in neutral charge before solubilization of Cry1 proteins and become negatively charged residues after solubilization process in the midgut of insects. The N terminal and catalytic domains are highly conserved. Variations in the functional motif were observed in toxicant and non-toxicant Cry1 subclass proteins. The toxic spectrum activity is based on active site and not on structure of the proteins. Cry1Ac (exhibiting high insecticidal spectrum) shows difference in sequence level but not in tertiary structure level, the same as Cry1Fa (moderate insecticidal spectrum) and Cry1La (no insecticidal spectrum). The amino acids present in the receptor sites of Cry1Ac, 1Ab, 1Aa may also be important for its high insecticidal spectrum.

Acknowledgements The authors are thankful to Department of Biotechnology (DBT), Government of India, New Delhi for the financial support in the form of Bioinformatics Infrastructure Facility (BTISNeT).

References

[1] Bachmair, A., Finley, D., Varshavsky, A. 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science 234, 179-186.

[2] Baig, D.N., Mehnaz, S. 2010. Determination and distribution of cry-type genes in halophilc *Bacillus thuringiensis* isolates of Arabian Sea sedimentary rocks. Microb Res 165, 376-383.

[3] Bobrowski, V.L., Pasquali, G., Bodansese-Zanettini, M.H., Pinto, L.M.N., Fiuza, L.M. 2002. Characterization of two *Bacillus thuringiensis* isolates from south Brazil and their toxicity against *Anticarsia gemmatalis* (Lepidoptera: Noctuidae). Biol Control 25, 129-135.

[4] Ciechanover, A., Schwartz, A.L. 1989. How are substrates recognized by the ubiquitin mediated proteolytic system? Trends Biochem Sci 14, 483-488.

[5] Coen, M., Suzanne, H., Barend, B., Wim, J., Martijn, F.B.G. 2007. A role for protein misfolding in immunogenicity of biopharmaceuticals. J Biol chem 282, 2229-2236.

[6] Crickmore, N., Zeigler, D.R., Feitelson, J., Schnepf, E., Van Rie J., Lereclus D., Baum, J., Dean, D.H. 1998. Revision of the nomenclature for the *Bacillus thurin*

giensis pesticidal crystal proteins. Microbiol Mol Biol Rev 62, 807-813.

[7] Devendra, J., Udayasuriyan, V., Indra Arulselvi, P., Dev, S.S., Sangeetha, P. 2006. Cloning, characterization, and expression of a new cry2Ab gene from *Bacillus thuringiensis* strain 14-1. Appl Biochem Biotech 128, 185-194.

[8] Diego, H.S., Graciela, B.B. 2008. *Bacillus thuringiensis*: General aspects. An approach to its use in the biological control of lepidopteran insects behaving as agricultural pests. Rev Argent Microbiol 40, 124-140.

[9] Garg, A., Raghava, G.P.S. 2008. A machine learning based method for the predicttion of secretory proteins using amino acid composition, their order and similarity-search. In Silico Biol 8, 129-140.

[10] Gill, S.C., van Hippel, P.H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182, 319-326.

[11] Giordana, B., Forcella, M., Leonardi, M.G., Casartelli, M., Fiandra, L., Hanozet, G.M., Parenti, P. 2002. A novel regulatory mechanism for amino acid absorption in lepidopteran larval midgut. J Insect Physiol 48, 585-592.

[12] Gonda, D.K., Bachmair, A., Wunning, I., Tobias, W.J., Lane, W.S., Varshavsky, A. 1989. Universality and structure of the N-end rule. J Biol Chem 264, 16700-16712.

[13] Hormaeche, I., Segura, R.L., Vecino, A.J., Goni, F.M., de la Cruz F., Alkorta, I. 2006. The transmembrane domain provides nucleotide binding specificity to the bacterial conjugation protein TrwB. FEBS Lett 580, 3075-3082.

[14] Ikai, A. 1980. Thermostability and aliphatic index of globular proteins. J Biochem 88, 1895-1898.

[15] Joe, D., Zheng, O., Jeffery, T., Andrew, B., Yaron, T., Jie, L. 2006. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucl Acid Res 34, 116-118.

[16] Kyte, J., Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol 157, 105-132.

[17] Lambert, B., Buysse, L., Decock, C., Jansens, S., Piens, C., Saey, B., Seurinck, J., Audenhove, K.V., Rie, J.V., Vliet, A.V., Peferoen, M. 1996. A *Bacillus thuringiensis* insecticidal crystal protein with a high activity against members of the family Noctuidae. Appl Environ Microbiol 62, 80.

[18] Letowski, J., Bravo, A., Brousseau, R., Masson, L. 2005. Assessment of cryI gene contents of *Bacillus thuringiensis* strains by use of DNA microarrays. Appl Environ Microbiol 71, 5391-5398.

[19] Lin, K., Simossis, V.A., Taylor, W.R., Heringa, J. 2005. A simple and fast secondary structure prediction method using hidden neural networks. Bioinformatics 21, 152-159.

[20] Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.W., Michael, W.J., Prisant, M.G., Richardson, J.S., Richardson, D.C. 2003. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins Str Func Genetics 50, 437-450.

[21] Martinez, C., Caballero, P. 2002. Contents of cry genes and insecticidal toxicity of *Bacillus thuringiensis* strains from terrestrial and aquatic habitats. J Appl Microbiol 92, 745-752.

[22] Masson, L., Erlandson, M., Puzstai-Carey, M., Brousseau, R., Juarez-Perez, V., Frutos, R. 1998. A holistic approach for determining the entomopathogenic potential of *Bacillus thuringiensis* strains. Appl Environ Microbiol 64, 4782-4788

[23] Nariman, A.H.A. 2007. PCR detection of cry genes in local *Bacillus thuringiensis* isolates. Aus J Basic and App Sci 1, 461-466

[24] Pandey, B.K., Srivastava, S., Singh, M., Ghosh, J.K. 2011. Inducing toxicity by introducing a leucine zipper like motif in frog antimicrobial peptide magainin 2. Biochem J 436, 609-620.

[25] Rowe, G.E., Margaritis, A. 1987. Bioprocess developments in the production of bioinsecticides by *Bacillus thuringiensis*. Critical Rev in Biot 6, 87-127.

[26] Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R., Dean, D.H. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol Mol Biol Rev 62, 775-806.

[27] Senthilkumar, B., Sailo, S., Guruswami, G., Nachimuthu, S. 2012. Prot-Prop: J-tool to predict the subcellular location of proteins based on physiochemical characterization. Interdiscip Sci Comput Life Sci 4, 296-301.

[28] Takatsugu, H., Boon-Chieng, S., Mitaku, S. 1998. SOSUI: Classification and secondary structure prediction system for membrane proteins. Bioinformatics 14, 378-379.

[29] Tamura, K., Dudley, J., Neil, M., Kumar, S. 2007. MEGA4: Molecular evolutionary genetics analysis. Mol Biol Evo 24, 1596-1599.

[30] Thaphan, P., Keawsompong, S., Chanpaisaeng, J. 2008. Isolation, toxicity and detection of cry gene in *Bacillus thuringiensis* isolates in Krabi province, Thailand. Songklanakarin J Sci Technol 30, 597-601.

[31] Tobias, J.W., Shrader, T.E., Rocap, G., Varshavsky, A. 1991. The N-end rule in bacteria. Science 254, 1374-1377.

[32] Tran, L.B., Vachon, V., Schwartz, J.L., Laprade, R. 2001. Differential effects of pH on the pore-forming properties of *Bacillus thuringiensis* insecticidal crystal toxins. App Environ Microbiol 67, 4488-4494.

[33] Vaseeharan, B., Valli, S.J. 2011. *In silico* homology modeling of prophenoloxidase activating factor Serine Proteinase gene from the haemocytes of *Fenneropenaeus indicus*. J Proteom Bioinform 4, 53-57.

[34] Ye, Y., Godzik, A. 2003. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 19, 246-255.