Short Communication

Journal of
Molecular Microbiology
and Biotechnology

J Mol Microbiol Biotechnol 2011;20:204–210
DOI: 10.1159/000329824

Published online: July 19, 2011

Tandem Repeats in a New Toxin Gene from *Bacillus thuringiensis* and in Other *cry11*-Like Genes

Olga Melnikov[a] Nadine Baranes[a] Monica Einav[a] Eitan Ben-Dov[b, c]
Robert Manasherob[a] Mark Itsko[a] Arieh Zaritsky[a]

[a]Department of Life Sciences, Ben-Gurion University of the Negev, Be'er-Sheva, [b]Department of Life Sciences, Achva Academic College, Achva, and [c]National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Be'er-Sheva, Israel

Key Words
Tandem repeats · *cry11Bb2* · *Bacillus thuringiensis*, a field isolate · Mosquito larvae

Abstract
A new gene, *cry11Bb2* from a field isolate of *Bacillus thuringiensis,* was cloned for expression in *Escherichia coli*. The encoded protein, with a deduced molecular mass of 89.5 kDa, exhibits 97 and 79% identities with the overlap regions of Cry11Bb1 from *B. thuringiensis* ssp. *medellin* and Cry11Ba1 from ssp. *jegathesan*, respectively. It is however longer than Cry11Bb1 by 42 amino acids in its carboxy-terminus, of which 32 comprise 2 tandem repeats additional to the 5 existing in the latter polypeptide. Possible roles for this recurrent motif among Cry toxins and their accessory proteins, and for their encoding genes are proposed.

Copyright © 2011 S. Karger AG, Basel

Various subspecies of the aerobic, Gram-positive, endospore-forming bacterium *Bacillus thuringiensis* are better than all other microbial control agents against larvae of insect pests found to date. During sporulation, each ssp. produces large quantities of a certain combination of insect larvicidal proteins known as δ-endotoxins that are aggregated in parasporal crystalline bodies [de Maagd et al., 2003]. *B. thuringiensis*-based formulations have long been used intensively worldwide to control various insect pests [Van Frankenhuyzen, 1993].

Isolates of *B. thuringiensis* toxic to mosquito larvae have been classified into three groups by Delecluse et al. [2000]. Group 2 includes ssp. *medellin* and *jegathesan* that display crystal protein patterns different than that of ssp. *israelensis* and are nearly as toxic [Margalith and Ben-Dov, 2000], each contains a different Cry11 toxin: -Aa1 in *israelensis*, -Ba1 in *jegathesan*, and -Bb1 in *medellin*.

Isolation and Characterization of a New *cry11Bb*-Like Open Reading Frame

Preliminary PCR screening of a series of *B. thuringiensis* field isolates [Ben-Dov et al., 1997] with a new pair of universal primers for *cry11*, Ucry11(d) (CCAGCATTAATAGCAGCTCC) and Ucry11(r) (GTACACATCTGAGTAAAAAG), discovered several *cry11-like* genes. The pair Ucry11(d) and cry11Bb(r) (CAATATGCCACCAATATCT), designed for *cry11Bb1* from ssp. *medellin* [Orduz et al., 1998], revealed a *cry11Bb*-like gene in isolate K34. When the pair of primers based on the sequence of

This article is dedicated to the memory of our colleague and friend, the late Professor Yoel Margalith.

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2011 S. Karger AG, Basel
1464–1801/11/0204–0204$38.00/0

Accessible online at:
www.karger.com/mmb

Dr. Eitan Ben-Dov
Department of Life Sciences
Achva Academic College
MP Shikmim 79800 (Israel)
Tel. +972 8 646 1920, E-Mail etn@bgu.ac.il

```
 N   T   S   S   G   Y   E   Q   G   Y   N   D   N   Y   N   Q

AAT ACA AGC AGT GGG TAT GAG CAA GGA TAT AAC GAT AAT TAT AAC CAA
AAT ACA AGT AGT GGG TAT GAG CAA GGA TAT AAC GAT AAT TAT AAC CAA
AAT ACA AGT AGT GGG TAT GAG CAA GGA TAT AAC GAT AAT TAT AAC CAA
AAT ACA AGT AGT GGA TAT GAG CAA GGA TAT AAC GAT AAT TAT AAC CAA
AAT ACA AGC AGT GAG TAT GAG CAA GGA TAT AAC GAC AAT TAT AAC CAA
AAT ACA AGT AGT GGA TAT GAG CAA GGA TAT AAC GAT AAT TAT AAC CAA
AAT ACA AGT AGT GGA TAC GAG CAA GGA TAT ATT GAT AAT TAT AGG CCA
```

Fig. 1. Tandem repeats of cry11Bb2. Nucleotides alignment of the 7 repeats, each encoding the 16 amino acids motif (top line; symmetrical tyrosines are bold-faced, and asparagines underlined). Mismatched nucleotides are marked in bold and italics. Codons of mismatched amino acids are underlined.

cry11Bb1, 11B-Nco(d) (CAATAAATTTAAGCAGGAACCATGGTAAATTC) and 11B-XhoBam(r) (GTATTTTGGATCCTAAGTTCTCGAGTACCTTGCTCATACTCCACTGC), was used to amplify the whole gene, a truncated reading frame of 2,135 bp, with no stop codon, was discovered (accession # HM068615). This sequence (named Δ*cry11Bb2*) was cloned for expression in *Escherichia coli* at the *Nco*I-*Bam*HI sites of pUHE-24S [Ben-Dov et al., 1995] to yield plasmid pHNt-B [Baranes-Sela, 2005]. The missing 3′-end of the gene was amplified with 11Bb1387(d) (CGCAGGGTCTTTACTATGGATGGAATA; starting from nucleotide # 1,387) and 11Bb-Bam(r) (AAAACGGATCCATAATGTGCTTGG) downstream *cry11Bb1* [Orduz et al., 1998]. The 1,530-bp amplicon containing the native *Kpn*I site (nt 1,448–1,453) was ligated into pHNt-B by replacing its shorter (687 bp) *Kpn*I-*Bam*HI fragment to yield pHN-B with 2,874 bp encompassing the whole new *cry11Bb1-like* ORF (between nucleotides 210–2,588).

The new ORF (2,379 bp; accession # HM068615) was analyzed (on both strands) with DNA sequencer ABI model 373A System (Applied Biosystems). Two putative ribosome-binding sites with adjacent in-frame ATG codons were found in *cry11Bb1* [Orduz et al., 1998] but in vivo translation in *B. thuringiensis* ssp. *medellin* starts at the second (accession # AF017416). The first ORF is indeed excluded in the new *cry11Bb2* because it contains two tandem T's (# 141–142; accession # HM068615) rather than 3 in *cry11Bb1* that would form a short ORF (9 amino acids long) from the first ATG, concordant with Cry11Ba1 and Cry11Aa1. This may stem from an artifact 'T' in *cry11Bb1*. The identity of the N-terminal amino acids of both Cry11Bb's (-1 and -2) confirms this explanation.

The N-terminus of the deduced encoded protein of 792 amino acids with a molecular mass of 89.486 kDa and pI of 5.5 (http://workbench.sdsc.edu) is 60% identical to Cry11Aa1 of *B. thuringiensis* ssp. *israelensis* (overlapping 651 amino acids), 79% to Cry11Ba1 of ssp. *jegathesan* (overlapping 728 amino acids) and 97% to Cry11Bb1 of ssp. *medellin* (overlap of 750 amino acids). The protein was named Cry11Bb2 (accession # HM068615), but the conventional nomenclature [Crickmore et al., 1998] does not take into account additional amino acids at protein termini.

Western blot analysis (using antibodies against Cry11Aa1, kindly provided by Sarjeet Gill) of the recombinant *E. coli* strains expressing full-length *cry11Bb2* and its truncated version from pHN-B and pHNt-B respectively, displayed about 110 and 70 kDa proteins (data not shown). Cross-reaction of Cry11Ba1 has previously been observed with anti-Cry11Bb1 [Delecluse et al., 1995; Orduz et al., 1998] and even with antibodies raised against the more distant Cry4Aa1 [Delecluse et al., 1995]. Toxicities to early 3rd instar *Aedes aegypti* larvae of the same recombinant *E. coli* strain [as in Ben-Dov et al., 1995] expressing the truncated or the intact *cry11Bb2* were comparable to that of *cry11Aa1* [Baranes-Sela, 2005; Khasdan et al., 2001; data not shown].

Repeated Motifs

The 126 bp at the 3′-end of *cry11Bb2*, additional to *cry11Bb1*, completing a sequence of 7 tandem repeats of 16 amino acids (NTSSGYEQGYNDNYNQ) near the C-terminus (fig. 1), are located between amino acids 665–776 of the deduced sequence (accession # HM068615). The 5th repeat contains a single mismatch (at amino acid # 733) and the last has 3 (# 771, 775 and 776). Cry11Bb1 (accession # AF017416) includes 5 such regions, whereas Cry11Aa1 (accession # AL731825) contains none, and Cry11Ba1 (accession # X86902) has 3 with varying numbers of mismatches.

Repeated blocks have been found in other mosquito larvicidal δ-endotoxins [de Maagd et al., 2003]: Cry20Aa and Cry27Aa. The roles of such tandem repeat motifs are however unknown.

Table 1. Correlation between number of repeats and observed molecular mass in currently known versions of Cry11

Cry11 version	Derived MW, kDa	Observed MW, kDa	ΔMW (= obs – der) (% difference)	Repeats, n	Reference
–Bb2	89.5	110	20.5 (22.9%)	7	This study
–Bb1	88.2	94	5.8 (6.6%)	5	Orduz et al., 1998
–Ba1	81.3	80	–1.3 (0.0%)	3	Delecluse et al., 1995
–Aa1	72.3	65	–7.3 (–10.1%)	0	Dai et al., 1993
		72	0 (0%)	0	Donovan et al., 1988
–ΔBb2	77.2	70	–7.2 (–9.3%)	0	This study

MW = Molecular weight.

The highly conserved 16-mer motif of 7 tandem repeats in Cry11Bb2 (accession # HM068615) is reflected in a similar conservation of its encoding nucleotide sequence of 48-mer motif (fig. 1). In particular, only 6 transitions exist among the 288 nucleotides composing the first 6 repeats, one of which alone encodes a different amino acid (G→E). The other 6 modifications, all in the last (7th) repeat, contain 3 transversions. Intriguingly, the degree of accuracy in repeat sequence is lower in the last copy in both amino acids and nucleotides. The latter may result of an in vivo mechanism that mimics the in vitro microgene polymerization reaction [Itsko et al., 2009, 2011].

The SDS-PAGE-observed molecular mass (110 kDa; not shown) of the full-length Cry11Bb2 is higher by ca. 20% than the deduced (89.5 kDa; accession # HM068615). This difference is significantly larger than the analogous differences observed with other Cry11 versions (table 1). The ORF for the truncated polypeptide ΔCry11Bb2 predicts a protein of 77,196 kDa – about 9% higher than the sequence-derived mass. The mobility of currently known Cry11 proteins is inversely related to the copy number of the repetitive motif (table 1). Orf2 from *B. thuringiensis* ssp. *kurstaki* displays an even larger anomaly: its electrophoretic mobility corresponds to a molecular mass of 50 kDa despite a predicted size of 29 kDa [Widner and Whiteley, 1989]. This phenomenon was also recorded in small spore-coat proteins of *Bacillus subtilis* [Zhang et al., 1993]. A higher apparent molecular weight may be due to residual secondary structures of these proteins retarding migration on the gel [Manasherob et al., 2001]. The question whether it means that the repeated motif (NTSSG**Y**EQG**Y**NDN**Y**NQNTSSG**Y**EQG**Y**NDN**Y**NQ) imposes some constraint on the denatured protein remains moot, but its symmetries around 3 tyrosine residues (**Y*****Y*****Y**) and 4 asparagines (N*N*N*N) are noteworthy. The possible roles of repeated motif in protein structure, stability and mobility are still to be resolved.

Repeating motifs in natural proteins are widely documented and described (http://www.ncl-india.org/trips/), and their possible activity mechanisms discussed [Katti et al., 2000]. The fact that they derive of encoding genes is sometimes not sufficiently considered. During studies with the entomopathogenic *B. thuringiensis* subspecies, we encountered several such cases where DNA repeats exist, reflecting their encoded protein motifs. Specifically, helper Orf2 proteins of some *cry* operons, *cry2Aa1* [Widner and Whiteley, 1989], *cry2Ac1* [Wu et al., 1991] and *cry39Aa1* [Ito et al., 2002] include different numbers of repetitive motifs. For example, Orf2$_{cry2Aa}$, encoded by *orf2* of the *cry2Aa* operon, contains 11 tandem repeats of a 15-amino-acid motif and accounts for two-thirds of all its residues [Widner and Whiteley, 1989]. The suggestion that Orf2 is involved in Cry2 crystallization [Crickmore and Ellar, 1992] was later confirmed for Cry2Aa1 [Ge et al., 1998; Staples et al., 2001]. This motif is encoded by a 45-nt repeat in its gene *orf2*$_{cry2Aa}$ (fig. 2a) that includes only 6 transitions and a single transversion among the 450 nt composing the first 10 repeats. All 5 additional changes are located at the 3′ end of the last (11th) repeat. *orf2*$_{cry2Aa3}$ [Sasaki et al., 1997] (fig. 2b) includes 13 repeats with the same basic unit as *orf2*$_{cry2Aa}$. In this gene, the first 9.5 repeats are separated from the rest by a 12-nt long 'linker' (highlighted dark gray with black letters in figure 2b).

A second type of repeat unit among the known *orf2* genes is found in *orf2*$_{cry2Ca}$ [Wu et al., 1991] and in *orf2*$_{cry2Ac}$ (accession # AY007687). It differs from the repeat of *orf2*$_{cry2Aa}$ by two triplets that surround a conserved T(A/G)C found in both kinds of repeats and by an

orf2$_{cry2Aa}$ (Access. # M23723)

```
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TAC ACA CAA GAT TTA ATT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
a ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTG AAT GTA
```

orf2$_{cry2Aa3}$ (Access. # D86064)

```
ACG TAT AAC CAA AGT CAG AAT GTA TGC GCG CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC ACA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC ACA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC ACA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC ACA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC GCG CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC ACA CAA GAT TTA GAA GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC ACA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC GCG CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AATGGTGTCTACAAG CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC GCG CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC ACA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT ACT TAT ATA CAA GAT TTA ATT GAT
b AAG TAT AAT CAA AGT CAG AAT
```

orf2$_{cry2Ca}$ (Access. # X57252)

```
ACG TAT AAC CAA AGC CAA AATTGGCTG   TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAC CAG AGC CAA AAT GGT GCC TGC GCG CAA GAT TTA ATG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GCG CAA GAT TTA GTG AAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TGC GCG CAA GAT TTA ATG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GCG CAA GAT TTA GTG AAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TGC GCG CAA GAT TTA ATG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GCG CAA GAT TTA GTG AAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG AAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG AAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG AAT
ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA GTG GAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TAC GCG CAA AAT TTA GTG GAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TGC GTG CAA GAT TTA GTG AAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TGC GTG CAA GAT TTA GTG GAT
c ACG TAT AAC CAA AGT CAG AAT GGT GCC TAC G
```

orf2$_{cry2Ac}$(Access. # AY007687)

```
                             GT GCCTTAC GCG CAAAGAT TTA GTG GAAT
ACGGAAT AACCCAAAAGT CAG AAT GGT GCCNTAC GCG CAANGAT TTA AGT GAAT
ACG TATTAANCCAA AGT CAG AAT GGT GCC TAC GCG CAA GAT TTA AGT GNAT
ACGGTATAATC CAA AG  CAG AAT GGG CCC TAC GCG CAA AAT TTA GTG GAT
ACG TAT AAT CAA AGT CAG AAT GGT GCC TGC GTG CAA GAT TTA GTG AAT
ACG TAT AAT CAA AG  CAG AAT GGT GCC TGC GTG CAA GAT TTA GTG GAT
d ACG TAT AAC CAA AGT CAG AAT GGT
```

cry11Bb2 (Access. # HM068615)

```
            CAATAAT ACA AGC AGT GGG TAT GAG CAA GGA TAT AAC GAT
AAT TAT AAC CAA AAT ACA AGT AGT GGG TAT GAG CAA GGA TAT AAC GAT
AAT TAT AAC CAA AAT ACA AGT AGT GGG TAT GAG CAA GGA TAT AAC GAT
AAT TAT AAC CAA AAT ACA AGT AGT GGA TAT GAG CAA GGA TAT AAC GAT
AAT TAT AAC CAA AAT ACA AGC AGT GAG TAT GAG CAA GGA TAT AAC GAC
AAT TAT AAC CAA AAT ACA AGT AGT GGA TAT GAG CAA GGA TAT AAC GAT
AAT TAT AAC CAA AAT ACA AGT AGT GGA TAC GAG CAA GGA TAT ATT GAT
e AAT TAT AGG CCA
```

Fig. 2. Individually aligned repeated motifs found in different *B. thuringiensis* (Bt) subspecies. **a** *orf2$_{cry2Aa}$* from Bt ssp. *kurstaki* HD-1. **b** *orf2$_{cry2Aa3}$* from Bt ssp. *sotto*. **c** *orf2$_{cry2Ca}$* from Bt S_1. **d** *orf2$_{cry2Ac}$* from Bt ssp. *thuringiensis*. **e** *cry11Bb2* from Bt strain K34. Fonts: black, bold, italics – pyrimidine transitions; black, bold – purine transitions; black, bold, underlined – transversions. Highlights: light gray with white letters, the conserved triplet in the middle of the motif; light gray with black letters, region of varied length (modulus 3); black with white letters, sequence homology between *cry11Bb2* and *orf2*; dark gray with black letters, non-homologous insertions; N = non-identified nucleotides.

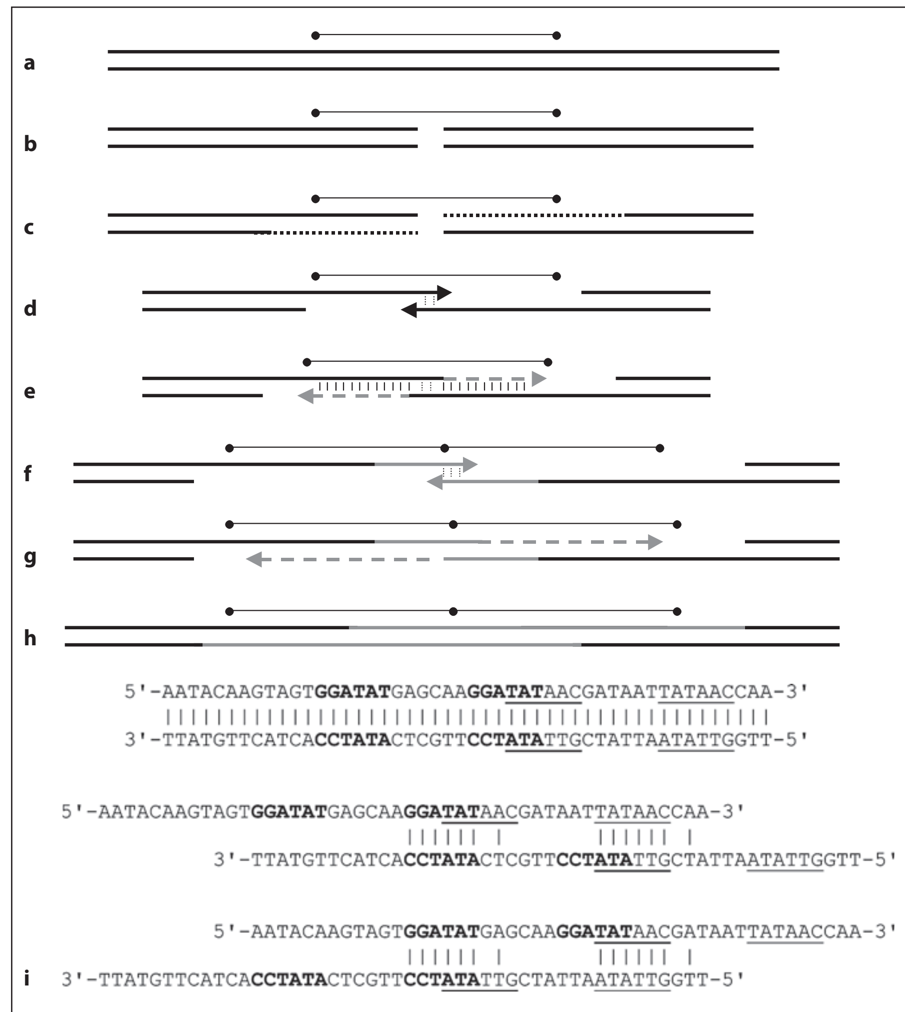

Fig. 3. Generation of sequence duplication due to DSB repair via the NHEJ-like pathway. **a**, **b** Chromosome region (**a**) suffers a DSB (**b**). **c** Trans-resection of DNA strands flanking the break by a dedicated exonuclease. **d** Bridging the broken ends in staggered mode by occasional Watson-Crick bonds due to partial complementarity. **e** Filling-in single-stranded regions by DNA polymerase. **f** Back slippage of the broken ends under repair that is mediated by a weak 2nd rank complementarity. **g** Filling-in single-stranded regions by DNA polymerase. **h** Restored integrity of the chromosome with the generated duplicated region. Dumbbells designate schematic location of the sequence undergoing duplication. Newly synthesized DNA is designated in gray. **i** Fully aligned and two slipped structures existing between constituent strands of *cry2Bb2* repeat motif that may assist to bridge the DSB ends. Bold and underlined regions designate two sets of complementarity.

extra adjacent triplet. A string of 991 nt in $orf2_{cry2Ca}$ encoding almost 21 repeats of a 16-mer motif in Cry2Ca contains only 35 transitions (fig. 2c), whereas the shorter (7 repeats) string in $orf2_{cry2Ac}$ is much less conserved and contains many deletions and insertions (fig. 2d).

Addition of four endogenous nucleotides to the 5′ end *cry11Bb2* encoding nucleotide sequence as displayed in figure 1, arranges 7 tandem repeats of 48-mer motif somewhat differently (fig. 2e). Seven (highlighted black with white letters) out of 16 triplets constituting the repetitive motif of Cry11Bb2 are identical/homologous in composition and location to those found in previously described *orf2* genes (fig. 2a–d).

The spectrum of replacements in the repeats of $orf2_{cry2Aa}$ and $orf2_{cry2Ca}$ is reciprocal on two sides of the conserved T(A/G)C triplet: in its 5′ part there is a bias to pyrimidine transitions whereas purine transitions are exclusively observed at its 3′ part. Intriguingly, such pattern of changes may point to occurrence of double strand break (DSB) in the middle of the repeat. Repair of DSB is mediated by trans-resection of opposite strands surrounding it in a way that transiently exposes generated ssDNA stretches to increased mutagenesis with characteristic pattern of transitions encompassing the DSB [Yang, 2008]. Moreover, considerable evidence has accumulated in eukaryotic cells that point to blunt dsDNA ends as key intermediates in the process leading to gene amplification [Mondello et al., 2010; Pace et al., 2009].

One of the error-prone processes exploited by cells to survive DSB is non-homologous end-joining (NHEJ) [Mahaney et al., 2009] that may contribute to sequence amplifications. It was regarded as exclusive prerogative of eukaryotic cells. However, it has recently been found in prokaryotes by identifying bacterial homologues of Ku

protein [Brissett and Doherty, 2009; Doherti et al., 2001]. Moreover, a functional NHEJ repair pathway is essential for spore viability in *B. subtilis* under conditions that yield DSBs [Weller et al., 2002; Moeller et al., 2007]. It is noteworthy that many of the bacteria that contain the Ku ligase system are capable of sporulation (*B. subtilis*, *Streptomyces coelicolor*) or spend long periods of their life cycle in the stationary phase (*Mycobacterium tuberculosis*, *Mesorhizobium loti*, *Sinorhizobium loti*) [Weller et al., 2002]. The sporulating *B. thuringiensis* may contain a NHEJ repair pathway as well.

Exogenous oligonucleotides complementary to the broken ends can efficiently target DSB for repair in yeasts [Storici et al., 2006]. We have recently proposed [Itsko et al., 2011] that partial complementarity between sequences flanking the broken ends may assist their sealing and contribute to repeat generation via slipped structures (fig. 3a–h). An example involving putative 2nd rank complementarity that can participate in the process to seal the broken ends was revealed in Cry2Bb2 (fig. 3i).

Once the initial doublet has been generated inside the chromosome, the repeats may propagate during DNA replication since the constituent strands may slide over each other between the multiple complementary regions. The alternative mechanism for the propagation is possible recurrence of DSBs in the middle of the generated repeats. Bridging the broken ends by occasional Watson-Crick bonds due to partial complementarity (fig. 3d) can bring about nucleotides insertions/deletions during fill-in by DNA polymerase. Selection for functionality of a given coding region can restore the lost frame by generating an additional triplet or dropping an existing one. This can result in the two/three triplets varied region (highlighted light gray with black letters in figure 2) surrounding the conserved T(A/G)C triplet in the 4 *orf2* versions and in *cry11Bb2* listed here (highlighted light gray with white letters in figure 2). The inserted linker in that region as in $orf2_{cry2Aa3}$ (highlighted dark gray in figure 2b) is also consistent with this explanation.

It is intriguing that the AGT triplet in $orf2_{cry2Ca}$ (fig. 2c) is always adjacent to CAG whereas its 5′-replaced AGC is adjacent to CAA. Both combinations encode the same amino acids: serine by AGT/C and glutamine by CAG/A. This coincidence cannot be explained by selection for suppression of one transition by the other. Such replacements do affect stability of slipped structures presumably emerging during replication of this region. Without the replacements, almost triplet repetition AAT CAA AGT CAA AAT (fig. 2c) would be explicitly observed that could putatively propagate in an uncontrolled way through slipped structures, bringing about genetic instability. Strong selection for silent transitions that save triplet coding but prevent deleterious expansion by disrupting repetition pattern would thus be anticipated to stabilize the genome.

A sequence of 13 bp long (bp 2,601–2,613; accession # HM068615) starting 12 bases downstream the stop codon shares 100% identity with those of the 16-bp inverted repeat of the insertion sequence IS*240*A that has initially been found to flank *cry4Aa1* [Delecluse et al., 1989]. Most *B. thuringiensis* ssp. tested, including all known mosquitocidal strains, possess at least one IS*240*-related element [Rosso and Delecluse, 1997]. IS*240* represents a family of insertion sequences with several variants [Leonard et al., 1997] that might be involved in the dispersion of toxin genes among various *B. thuringiensis* strains [Leonard et al., 1997; Rosso and Delecluse, 1997]. They may also be instrumental to duplicate such motifs as found here.

Concluding Remarks

A new representative of Cry proteins that contains tandem repeats, Cry11Bb2, was discovered, described and discussed. The high degree of conservation of nucleotide repeats in *cry11Bb2* that code for the corresponding amino acids tandem repeats, points to their possible generation from a common template apparently by slippage between partly complementary regions during DNA replication. The motif repetitions generate periodic structures. Proteins that contain such a structure can serve as scaffolding centers, which assist toxin crystallization during the sporulation stage.

Molecular participants in the repeat expansion process and reaction conditions accompanying it in vivo may shed further light on the basic genetic mechanism as well as the crystallization process, stability and toxicity of the δ-endotoxins.

Acknowledgments

This investigation was partially supported by grants of the United States-Israel Binational Science Foundation (BSF, # 2001-042 and 2007-037), Jerusalem, Israel (to A.Z.), and two Levi Eshkol scholarships (one to each, O.B. and M.I.) from the Israeli Ministry of Science, Culture and Sports. Yoel Margalith and Sarjeet Gill are gratefully acknowledged for the free supply of *Aedes aegypti* eggs and for antibodies raised against Cry11Aa1, respectively.

References

Baranes-Sela N: An attempt to improve mosquito larvicidal activity of *Bacillus thuringiensis* subsp. *israelensis;* MSc thesis, Ben-Gurion University of the Negev, Be'er-Sheva 2005.

Ben-Dov E, Boussiba S, Zaritsky A: Mosquito larvicidal activity of *Escherichia coli* with combinations of genes from *Bacillus thuringiensis* subsp. *israelensis*. J Bacteriol 1995; 177:2851–2857.

Ben-Dov E, Zaritsky A, Dahan E, Barak Z, Sinai R, Manasherob R, Khamraev A, Troyetskaya E, Dubitsky A, Berezina N, Margalith Y: Extended screening by PCR for seven *cry*-group genes from field-collected strains of *Bacillus thuringiensis*. Appl Environ Microbiol 1997; 63:4883–4890.

Brissett NC, Doherty AJ: Repairing DNA double-strand breaks by the prokaryotic nonhomologous end-joining pathway. Biochem Soc Trans 2009;37:539–545.

Crickmore N, Ellar DJ: Involvement of a possible chaperonin in the efficient expression of a cloned CryIIA δ-endotoxin gene in *Bacillus thuringiensis*. Mol Microbiol 1992;6:1533–1537.

Crickmore N, Zeigler DR, Feitelson J, Schnepf E, Vanrie J, Lerecluse D, Baum J, Dean DH: Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiol Mol Biol Rev 1998;62:807–813.

Dai S-M, Gill SS: In vitro and in vivo proteolysis of the *Bacillus thuringiensis* subsp. *israelensis* CryIVD protein by *Culex quinquefasciatus* larval midgut proteases. Insect Biochem Mol Biol 1993;23:273–283.

Delecluse A, Bourgouin C, Klier A, Rapoport G: Nucleotide sequence and characterization of a new insertion element, IS*240*, from *Bacillus thuringiensis israelensis*. Plasmid 1989;21: 71–78.

Delecluse A, Rosso ML, Ragni A: Cloning and expression of a novel toxin gene from *Bacillus thuringiensis* subsp. *jegathesan* encoding a highly mosquitocidal protein. Appl Environ Microbiol 1995;61:4230–4235.

Delecluse A, Juarez-Perez V, Berry C: Vector-active toxins: structure and diversity; in Charles JF, Delécluse A, Nielsen-le Roux C (eds): Entomopathogenic Bacteria: From Laboratory to Field Application. Dordrecht, Kluwer Academic, 2000, pp 101–125.

De Maagd RA, Bravo A, Berry C, Crickmore N, Schnepf HE: Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu Rev Genet 2003;37:409–433.

Doherty AJ, Jackson SP, Weller GR: Identification of bacterial homologues of the Ku DNA repair proteins. FEBS Lett 2001;500:186–188.

Donovan WP, Dankocsik C, Gilbert MP: Molecular characterization of a gene encoding a 72-kilodalton mosquito-toxic crystal protein from *Bacillus thuringiensis* subsp. *israelensis*. J Bacteriol 1988;170:4732–4738.

Ge B, Bideshi D, Moar WJ, Federici BA: Differential effects of helper proteins encoded by the cry2A and cry11A operons on the formation of Cry2A inclusions in *Bacillus thuringiensis*. FEMS Microbiol Lett 1998;165:35–41.

Ito T, Sahara K, Bando H, Asano S: Cloning and expression of novel crystal protein genes *cry39A* and *39orf2* from *Bacillus thuringiensis* subsp. *aizawai* Bun1–14 encoding mosquitocidal proteins Insect Biotechnol Sericol 2003;71:123–128.

Itsko M, Rabinovitch A, Zaritsky A: Kinetics of repeat propagation in the microgene polymerization reaction. Biophys J 2009;96:1866–1874.

Itsko M, Ben-Dov E, Rabinovitch A, Zaritsky A: Tandem DNA repeats: generation and propagation in the microgene polymerization reaction and in vivo; in Tadashi M (ed): Application of Thermodynamics to Biological and Materials Science. Rijeka, InTech Open Access Publisher, 2011, pp 175–202.

Katti MV, Sami-Subbu R, Ranjekar PK, Gupta VS: Amino acid repeat patterns in protein sequences: their diversity and structural-functional implications. Protein Sci 2000;9: 1203–1209.

Khasdan V, Ben-Dov E, Manasherob R, Boussiba S, Zaritsky A: Toxicity and synergism in transgenic *Escherichia coli* expressing four genes from *Bacillus thuringiensis* subsp. *israelensis*. Environ Microbiol 2001;3:798–806.

Leonard C, Chen Y, Mahillon J: Diversity and differential distribution of IS231, IS232 and *IS240* among *Bacillus cereus, Bacillus thuringiensis* and *Bacillus mycoide*. Microbiology 1997;143:2537–2547.

Mahaney BL, Meek K, Lees-Miller SP: Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J 2009;417:639–650.

Manasherob R, Zaritsky A, Ben-Dov E, Saxena D, Barak Z, Einav M: Effect of accessory proteins P19 and P20 on cytolytic activity of Cyt1Aa from *Bacillus thuringiensis* subsp. *israelensis* in *Escherichia coli*. Curr Microbiol 2001;43:355–364.

Margalith Y, Ben-Dov E: Biological control by *Bacillus thuringiensis* subsp. *israelensis*; in Rechcigl JE, Rechcigl NA (eds): Insect Pest Management: Techniques for Environmental Protection. Boca Raton, CRC Press, 2000, pp 243–301.

Moeller R, Stackebrandt E, Reitz G, Berger T, Rettberg P, Doherty AJ, Horneck G, Nicholson WL: Role of DNA repair by non-homologous end joining in *Bacillus subtilis* spore resistance to extreme dryness, mono- and polychromatic UV and ionizing radiation. J Bacteriol 2007;189:3306–3311.

Mondello C, Smirnova A, Giulotto E: Gene amplification, radiation sensitivity and DNA double-strand breaks. Mutat Res 2010;704: 29–37.

Orduz S, Realpe M, Arango R, Murillo LA, Delecluse A: Sequence of the *cry11Bb1* gene from *Bacillus thuringiensis* subsp. *medellin* and toxicity analysis of its encoded protein. Biochim Biophys Acta 1998;1388:267–272.

Pace JK II, Sen SK, Batzer MA, Feschotte C: Repair-mediated duplication by capture of proximal chromosomal DNA has shaped vertebrate genome evolution. PLoS Genet 2009;5:e1000469.

Rosso ML, Delecluse A: Distribution of the insertion element IS*240* among *Bacillus thuringiensis* strains. Curr Microbiol 1997; 34:348–353.

Sasaki J, Asano S, Hashimoto N, Lay BW, Hastowo S, Bando H, Iizuka T: Characterization of a cry2A gene cloned from an isolate of *Bacillus thuringiensis* serovar *sotto*. Curr Microbiol 1997;35:1–8.

Staples N, Ellar D, Crickmore N: Cellular localization and characterization of the *Bacillus thuringiensis* Orf2 crystallization factor. Curr Microbiol 2001;42:388–392.

Storici F, Snipe JR, Chan GK, Gordenin DA, Resnick MA: Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. Mol Cell Biol 2006;26:7645–7657.

Van Frankenhuyzen K: The challenge of *Bacillus thuringiensis*; in Entwistle PF, Cory JS, Bailry MJ, Higgs SR (eds): *Bacillus thuringiensis,* an Environmental Biopesticide: Theory and Practice. Chichester, Wiley, 1993, pp 1–35.

Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, Devine SK, Day JP, Wilkinson A, di Fagagna FD, Devine KM, Bowater RP, Jeggo PA, Jackson SP, Doherty AJ: Identification of a DNA non-homologous end-joining complex in bacteria. Science 2002;297:1686–1689.

Widner WR, Whiteley HR: Two highly related insecticidal crystal proteins of *Bacillus thuringiensis* subsp. *kurstaki* possess different host range specificities. J Bacteriol 1989; 171:965–974.

Wu X, Cao XL, Bai YY, Aronson AI: Sequence of an operon containing a novel N-endotoxin gene from *Bacillus thuringiensis*. FEMS Microbiol Lett 1991;81:31–36.

Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA: Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. PLoS Genet 2008; 4:e1000264.

Zhang J, Fitz-James PC, Aronson A: Cloning and characterization of a cluster of genes encoding polypeptides present in the insoluble fraction of the spore coat of *Bacillus subtilis*. J Bacteriol 1993;175:3757–3766.