CROATICA CHEMICA ACTA CCACAA **68** (3) 533–542 (1995)

ISSN 0011-1643
CCA–2252

Original Scientific Paper

The Coding Region of the Equinatoxin II Gene Lacks Introns

Gregor Anderluh,[a]* *Jože Pungerčar,*[b] *Borut Štrukelj,*[b] *Peter Maček*[a] *and Franc Gubenšek*[b]

[a]*Department of Biology, Biotechnical Faculty, University of Ljubljana, Večna pot 111, 61000 Ljubljana, Slovenia*

[b]*Department of Biochemistry and Molecular Biology, Jožef Stefan Institute, Jamova 39, 61000 Ljubljana, Slovenia*

Received March 27, 1995; revised April 24, 1995; accepted May 19, 1995

Sea anemones produce several toxic peptides and proteins. Equinatoxins (Eqt) isolated from the sea anemone *Actinia equina* are basic cytolytic proteins with molecular masses of approximately 20 kDa. Of the three Eqt purified so far, EqtII is the most abundant and well characterized. Its gene organization has not yet been studied. In order to obtain the first information about the EqtII gene structure and sequence, genomic DNA was isolated from *A. equina* and the target DNA fragment amplified by the polymerase chain reaction (PCR) using three different pairs of oligonucleotide primers deduced from the preserved regions of Eqt cDNA clones. The sequence of the PCR product obtained after amplification of genomic DNA, using an oligonucleotide specific for EqtII, was almost indistinguishable from that of EqtII cDNA. As the DNA fragments derived from PCR of genomic DNA were of the same length as those from control PCR reactions performed on an *A. equina* cDNA library and EqtII cDNA, the EqtII gene was proved to be intronless, at least within the amplified preproprotein region. The presence of such an intronless gene coding for this cytotoxic protein might be explained by the relative low position of cnidarians in the evolutionary tree or by the advantage provided by a potentially higher rate of gene expression.

* Author to whom correspondence should be addressed.

INTRODUCTION

From the tentacles and bodies of the sea anemone *Actinia equina* L., Maček and Lebez[1] isolated three lethal hemolytic toxins, designated equinatoxins (Eqt) I, II and III. They all belong to a larger heterogeneous group of protein cytolysins produced by sea anemones (ordo *Actiniaria*). To date, about 30 such toxins have been characterized[2], with molecular masses ranging from 10–20 kDa. Eqt are basic, single-chain proteins, devoid of cysteine residues, with a molecular mass of approximately 20 kDa. In *Actinia equina*, EqtII is the most abundant of the three equinatoxins. It exhibits cytotoxic and cytolytic pore-forming activity[3–6] and several other pharmacological effects[7–9].

In a single species of sea anemones, several closely related basic cytolysins can be found. There are reports on purification of these toxins including isolation of equinatoxins I, II and III from *Actinia equina*[1], tenebrosins-A, B, C from *Actinia tenebrosa*[10], and cytolysins I, II, III and IV from *Stichodactyla helianthus*[11]. Comparison of their N-termini shows a high percentage of identical residues indicating that several isoforms differing only in a few residues might be present. Additionally, a high level of similarity between basic cytolysins from different species was also observed. Moreover, the amino acid sequence of EqtII[12] is nearly identical to that of tenebrosin-C, a cytolysin from *Actinia tenebrosa*[13]. Both proteins are also very similar in their biochemical and biological properties.[14] Although the protein sequences of the two toxins match almost perfectly within all of the 179 amino acid residues, the sequence of the genes encoding EqtII and tenebrosin-C might be different. In the present work, we performed PCR analysis of Eqt genes using different pairs of oligonucleotide primers deduced from Eqt cDNA sequences in order to investigate the structure and sequence of the EqtII gene within the protein coding region.

EXPERIMENTAL

Isolation of genomic DNA

Genomic DNA was independently isolated from two specimens of the sea anemone *Actinia equina* L. The whole body was immersed in liquid nitrogen and ground in a mortar to a fine powder. Ten volumes of extraction buffer (10 mM Tris-HCl, pH 8.0, 0.1 M EDTA, 0.5% SDS, 100 μg/ml of proteinase K) were added to the powder and incubated overnight at 55 °C. The viscous solution was extracted seven times with an equal volume of phenol equilibrated with 0.5 M Tris-HCl (pH 8.0). The water phase was dialyzed three times against TE buffer (50 mM Tris-HCl, 10 mM EDTA, pH 8.0). The DNA was stored at 4 °C. The concentration of the DNA was estimated from its absorbance at 260 nm.

PCR amplification

Oligonucleotides deduced from Eqt cDNA sequences (Pungerčar, J. *et al.*, unpublished) were synthesized on an Applied Biosystems 381A DNA synthesizer and partially purified by precipitation with a one tenth volume of 3 M NaOAc, pH 5.2, and five volumes of 100% EtOH. The following oligonucleotides were used:

E1(+), 5'-CGAATTCAATCGCACGCTTTAACTCAG-3';
E1(–), 5'-GTGGGATCCTAGTTTTACATCAGTATTCTCGAG-3';
E2(+), 5'-GGAATTCATATGTCCGCAGACGTGGCT-3';
E2(–), 5'-CAGAATTCGGATCCTATCAAGCTTTGGTCACGT-3';
E3(+), 5'-CGGAATTCGGGCAAGACGTGGACCGC-3'; and
E3(–), 5'-GTGGGATCCATTGTCCCCTCGAAATGGAGA-3'.

Either an *Eco*RI or *Bam*HI restriction site was added to the 5' end of the primers in order to subclone the PCR amplified DNA fragments into the pUC19 vector.

PCR reaction was performed in a total volume of 100 μl using a Perkin-Elmer DNA Thermal Cycler. The reaction mixture typically consisted of 1x PCR buffer II (Perkin-Elmer Cetus), 8 μl of 20 mM $MgCl_2$, 16 μl of dNTP mix (1.25 mM each), 50 pmoles of each of the sense and antisense oligonucleotides and 1 μg of the isolated genomic DNA. The mixture was overlaid with 35 μl of mineral oil, incubated in boiling water for 10 min, after which 2.5 U of AmpliTaq DNA Polymerase (Perkin-Elmer Cetus) was added. Amplification was carried out for 25 cycles, each cycle consisted of incubations for 1 min at 96 °C for denaturation, 1 min at 40 °C for annealing and 6 min at 70 °C for primer extension. The PCR products were analyzed by electrophoresis on a 1.5% agarose gel. In the case of amplification of non-genomic DNA samples, a 20-μl aliquot of an *A. equina* cDNA library or 0.3 μg of an EqtII cDNA clone in pUC19, respectively, was used instead of 1 μg of genomic DNA and amplified under the same conditions.

Cloning and nucleotide sequencing

The PCR products of genomic DNA samples obtained by amplification using E2(+) and E2(–) oligonucleotides were digested with *Eco*RI (New England BioLabs) and ligated into pUC19. The DNA fragment was sequenced[15] using T7 DNA polymerase (T7 sequencing kit, Pharmacia) and [^{35}S]dATPαS (Amersham). Nucleotide and amino acid sequences obtained were analyzed on a computer by DNASIS (Pharmacia).

RESULTS

DNA amplification

Six different oligonucleotides complementary to three different Eqt cDNAs, including the one specific for EqtII, E2(+), were used to amplify different segments of the coding, 5'- and 3'-untranslated regions of the EqtII gene (Figure 1). For PCR amplification of genomic DNA, the oligonucleotide

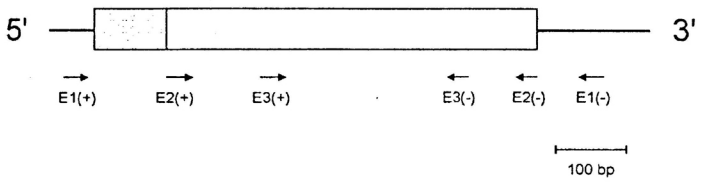

Figure 1. Oligonucleotide localization according to the EqtII cDNA sequence. The shaded box represents the prepropeptide, the open box the mature protein, and the lines the 5'-untranslated and 3'-untranslated flanking regions.

primers were combined in sets of two in different ways. PCR reaction was performed using genomic DNA isolated from a single specimen of the sea anemone *A. equina*. Two clones were sequenced in order to exclude the possibility of error made by *Taq* DNA polymerase. Additionally, applying the same conditions, the PCR reaction was carried out using genomic DNA isolated from another specimen of *A. equina*. As control reactions, an *A. equina* cDNA library and EqtII cDNA were also amplified. Figure 2 shows the analysis of some of the PCR products by agarose gel electrophoresis. The DNA fragments resulting from DNA amplification of both specimens with E2(+) and E2(–) primers were of the same length, approximately 560 base pairs (data shown for one specimen only, lane 1). This length also perfectly matches the length of the PCR products obtained after amplification of the cDNA library (lane 2) and EqtII cDNA clone in pUC19 (lane 3) using the same primers.

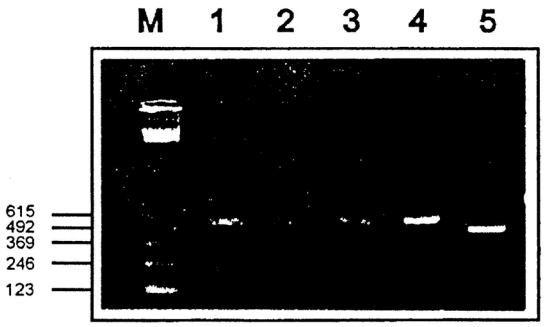

Figure 2. Analysis of PCR products by agarose gel electrophoresis. DNA was amplified with different pairs of oligonucleotides. Lane **1**, E2(+)/E2(–), genomic DNA (570 bp); lane **2**, E2(+)/E2(–), cDNA library of *A. equina* (570 bp); lane **3**, E2(+)/E2(–), EqtII cDNA clone (570 bp); lane **4**, E1(+)/E3(–), genomic DNA (580 bp); lane **5**, E3(+)/E1(–), genomic DNA (490 bp); **M**, 123-bp DNA marker.

In lanes 4 and 5, the products of PCR amplifications of genomic DNA using two different pairs of oligonucleotides, E1(+)/E3(–) and E3(+)/E1(–), respectively, are shown. In each case, one of the oligonucleotides of the particular pair was located within the protein coding region (E3) and the other oligonucleotide (E1) either within the 5'-untranslated or within the 3'-end of Eqt cDNAs. Both PCR products were of the expected size, according to cDNA sequences, approximately 580 bp and 490 bp, respectively.

Sequence of the PCR product

The DNA product resulting from PCR amplification of genomic DNA, using E2 sense and antisense primers, was sequenced after cloning into pUC19. In Figure 3, the nucleotide and deduced amino acid sequences of the EqtII gene in the mature protein region are shown. This coding region comprises 540 bp ended by a TGA stop codon. It is neither interrupted by termination codons nor by any intervening sequences and encodes a protein of 179 amino acids. The nucleotide sequence nearly perfectly matches the EqtII cDNA sequence, the only difference being an ACC codon, as a consequence of the E2(–) primer, coding for Thr177 instead of AGC coding for Ser177, like in the case of EqtII cDNA. Thus, the deduced amino acid sequence is thus actually the same as the recently published protein sequence of EqtII[12] where, at position 81, by protein sequencing equimolar amounts of Pro and Asp were found (Figure 4). In our case, we deduced a Pro residue at this position. Interestingly, EqtII from *A. equina* (family *Actiniidae*) shares only about 60% of amino acid indentity with cytolysin III from *Stichodactyla helianthus,*[16] which belongs to the other family of sea anemones, *Stichodactylidae*.

DISCUSSION

In evolution, cnidarians (coelenterates), including sea anemones, separated early from other, more complex metazoans to represent a distinct sidebranch in the phylogenetic tree. The presence of several closely related cytolytic toxins in a single or in different species of sea anemones, differing in their biological activities, indicates a strong positive selection pressure on the toxic proteins taking place during the evolution of these rather primitive organisms. Isolation and characterization of the toxins, together with amino acid sequencing, showed that the number of the members within a certain group of cytolysins and the differences between them exceed the possible variability due to allelic polymorphism. Therefore, it seems reasonable to assume that a single gene coding for such a toxic protein had to be duplicated and re-arranged to fullfil all the different demands of a predator. One can also speculate that expression of such toxins has to be fast in order to be effective. At present, there is little information about structure of cnidarian genes.

```
TCCGCAGACGTGGCTGGCGCCGTTATTGATGGTGCAAGTCTGTCCTTCGACATCCTGAAGACT    63
 S  A  D  V  A  G  A  V  I  D  G  A  S  L  S  F  D  I  L  K  T  21

GTCCTCGAAGCACTCGGTAATGTCAAACGAAAGATTGCCGTCGGTGTCGACAACGAGTCGGGC   126
 V  L  E  A  L  G  N  V  K  R  K  I  A  V  G  V  D  N  E  S  G  42

AAGACGTGGACCGCATTGAACACATACTTCCGTTCTGGCACCTCTGATATCGTCCTTCCCCAT   189
 K  T  W  T  A  L  N  T  Y  F  R  S  G  T  S  D  I  V  L  P  H  63

AAAGTTCCACATGGTAAGGCACTACTCTACAACGGTCAGAAAGATCGTGGTCCAGTTGCGACT   252
 K  V  P  H  G  K  A  L  L  Y  N  G  Q  K  D  R  G  P  V  A  T  84

GGCGCGGTTGGAGTACTTGCTTATCTCATGAGCGATGGAAACACCCTGGCCGTTTTATTCAGC   315
 G  A  V  G  V  L  A  Y  L  M  S  D  G  N  T  L  A  V  L  F  S  105

GTTCCCTATGACTATAACTGGTACAGCAACTGGTGGAATGTCAGGATCTATAAGGGAAAGAGG   378
 V  P  Y  D  Y  N  W  Y  S  N  W  W  N  V  R  I  Y  K  G  K  R  126

CGAGCAGACCAGAGGATGTACGAGGAACTCTACTACAATTTGTCTCCATTTCGAGGGGACAAT   441
 R  A  D  Q  R  M  Y  E  E  L  Y  Y  N  L  S  P  F  R  G  D  N  147

GGCTGGCACACCAGGAATCTTGGATATGGGTTGAAGAGCCGTGGATTCATGAACAGCTGTGGA   504
 G  W  H  T  R  N  L  G  Y  G  L  K  S  R  G  F  M  N  S  S  G  168

CATGCCATACTGGAAATTCACGTGACCAAAGCTTGA                               540
 H  A  I  L  E  I  H  V  T  K  A  *                                179
```

Figure 3. The nucleotide and deduced amino acid sequences of the mature protein region of a genomic EqtII PCR clone. The regions complementary to the oligonucleotides E2(+), E3(+), E3(–) and E2(–) are underlined.

```
                10        20        30        40        50
gEqtII  SADVAGAVIDGASLSFDILKTVLEALGNVKRKIAVGVDNESGKTWTALNT
EqtII   ..................................................
TenC    ..................................................
SthIII    AL..TI.A....T.QV.DK...E..K.S.------------------

                60        70        80        90       100
gEqtII  YFRSGTSDIVLPHKVPHGKALLYNGQKDRGPVATGAVGVLAYLMSDGNTL
EqtII   ............................*.....................
TenC    ..................................................
SthIII  ---..T.VI..EF..NT.....S.R.DT........AAFQ.Y..S.....

               110       120       130       140       150
gEqtII  AVLFSVPYDYNWYSNWWNVRIYKGKRRADQRMYEELYYNLSPFRGDNGWH
EqtII   ..................................................
TenC    ..................................................
SthIII  G.M....F.........D.K..S.......G...D...G-N.Y....GHW

                160       170       179
gEqtII  TRNLGYGLKSRGFMNSSGHAILEIHVTKA
EqtII   ........................S..
TenC    ........................#..
SthIII  EK......RMK.I.T.A.EAKMQ.KISR
```

Figure 4. Amino acid sequence alignment of an EqtII genomic clone (**gEqtII**) and protein sequences of EqtII (**EqtII**)[12], tenebrosin-C[13] (**TenC**) and *Stichodactyla helianthus* cytolysin III[16] (**SthIII**). A dot indicates the same amino acid residue, the asterisk (*) indicates Pro or Asp residue, the sign # denotes Thr or Ser residue. A gap introduced in order to optimize the alignment is shown by dashes (–).

In the course of our study, we determined the structure and sequence of a cnidarian gene encoding EqtII by use of PCR. The deduced amino acid sequence of a genomic EqtII clone within the mature protein region was practically the same as that of an EqtII cDNA or protein sequence of EqtII (Figure 4). Additionally, this sequence also nearly perfectly matches that of tenebrosin-C isolated from the Australian sea anemone *A. tenebrosa*. All these sequences differ at two positions only, 81 (Pro or Asp) and 177 (Ser or Thr). Our genomic clone, together with protein and cDNA sequences, shows that this heterogeneity could be explained by different allelic forms

of EqtII present in the sea anemone *A. equina*. However, it is still difficult to explain the presence of the same toxic protein in two different species of sea anemones, *A. equina* and *A. tenebrosa*[15]. Recent results of N-terminal sequencing of EqtI and EqtIII (Križaj, I. *et al.*, unpublished) and different cDNA sequences of Eqt indicate that the number of different Eqt, basic cytolysins of about 20 kDa, is higher than three. These Eqt have to be products of different genes.

The length and sequence of a genomic clone of EqtII, in comparison with its cDNA sequence, clearly indicate that this gene is intronless, at least within the preproprotein region. Additionally, since not even a single mutation was found, it is not likely to be a pseudogene. Furthermore, in our PCR experiments, all but one oligonucleotide primer were synthesized to amplify genomic sequences of at least three Eqt. Since these oligonucleotides could anneal to different genomic Eqt sequences, not just to that of EqtII, we can further conclude that other Eqt genes do not possess intervening sequences.

Most eukaryotic genes are interrupted by introns. It is believed that introns were already present in the primordial genes and were later differentially lost during evolution[17–19]. Intron loss was the most drastic in prokaryotes. For example, one can hardly find any introns in bacteria, and the situation is quite similar in fungi[20]. In these evolutionary lower organisms, loss of introns is correlated with a high rate of cellular divison. In eukaryotes, however, the genome is bigger and the replication rate is lower. Additionally, introns could also contribute to the development of more complex proteins. As a consequence, the selection pressure against eukaryotic introns was smaller. It may be therefore expected that the higher an organism is placed in the evolutionary tree the fewer intronless genes it will possess. Cnidarians are evolutionary primitive organisms. On the basis of comparison of the 18S ribosomal RNA from different species, Hendriks *et al.*[21] placed them in their own lineage just after the branching of fungi and other metazoans. We may, thus, assume that more intronless genes are present in these animals than in higher organisms, especially vertebrates. Indeed, there are some reports on the isolation of such genes from different classes of the phylum Cnidaria[22–24]. In this respect, Eqt genes are just another example of cnidarian genes lacking intervening sequences, reflecting their common origin. In genome evolution, multiplication of complete genes is likely to be easier in the absence of introns. Further studies will show how Eqt genes are organized in the genome of *A. equina*. On the other hand, the presence of intronless genes encoding important cytotoxic proteins could be advantageous for these predatory animals in order to respond quickly to different environmental stimuli. Splicing out introns could decrease the gene expression rate and slow down the rapid accumulation of these toxins, re-

quired for efficient action. Structural and functional analyses of the promotor regions of Eqt genes, together with thorough physiological studies, will help us to understand better the expression of these interesting toxic proteins.

Acknowledgements. – This work was supported by a grant from the Ministry for Science and Technology of Slovenia. We are grateful to Dr. Roger H. Pain for critical reading of the manuscript.

REFERENCES

1. P. Maček, and D. Lebez, *Toxicon* **26** (1988) 441.
2. P. Maček, *FEMS Microbiol. Immunol.* **105** (1992) 121.
3. P. Maček, and D. Lebez, *Toxicon* **19** (1981) 233.
4. R. Zorec, M. Tester, P. Maček, and W. T. Mason, *J. Membrane Biol.* **118** (1990) 243.
5. U. Batista, P. Maček, and B. Sedmak, *Cell Biol. Int. Rep.* **14** (1990) 1013.
6. G. Belmonte, C. Pederzolli, P. Maček, G. Menestrina, *J. Membrane Biol.* **131** (1993) 11.
7. D. Šuput, N. Rubly, and H. Meves, in: *Progress in Venom and Toxic Research* P. Gopalakrishnakone, and C. K. Tan, (Eds.), The Venom and Toxin Research Group, National University of Singapore, Republic of Singapore, 1988, p. 467.
8. C. I. Ho, J. L. Ko, H. M. Lue, C. Y. Lee, and I. Ferlan, *Toxicon* **25** (1987) 659.
9. W. M. Lafranconi, I. Ferlan, F. E. Russel, and R. J. Huxtable, *Toxicon* **22** (1984) 347.
10. R. S. Norton, G. Bobek, J. O. Ivanov, M. Thomson, E. Fiala-Beer, R. L. Moritz, and R. J. Simpson, *Toxicon* **28** (1990) 29.
11. W. R. Kem, and B. M. Dunn, *Toxicon* **26** (1988) 997.
12. G. Bellmonte, G. Menestrina, C. Pederzolli, I. Križaj, F. Gubenšek, T. Turk, and P. Maček, *Biochim. Biophys. Acta* **1192** (1994) 197.
13. R. J. Simpson, G. E. Reid, R. L. Moritz, C. Morton, and R. S. Norton, *Eur. J. Biochem.* **190** (1990) 319.
14. R. S. Norton, P. Maček, G. E. Reid, and R. J. Simpson, *Toxicon* **30** (1992) 13.
15. F. Sanger, S. Nicklen, and A. R. Coulson, *Proc. Natl. Acad. Sci. U. S. A.* **74** (1977) 5467.
16. K. M. Blumenthal, and W. R. Kem, *J. Biol. Chem.* **258** (1983) 5574.
17. W. Gilbert, M. Marchionni, and G. McKnight, *Cell* **46** (1986) 151.
18. M. Marchionni, and W. Gilbert, *Cell* **46** (1986) 133.
19. M.-C. Shih, P. Heinrich, and H. M. Goodman, *Science* **242** (1988) 1164.
20. J. D. Hawkins, *Nucl. Acids Res.* **16** (1988) 9893.
21. L. Hendriks, Y. Van de Peer, M. Van Herck, J.-M. Neefs, and R. De Wachter, *FEBS Lett.* **269** (1990) 445.
22. D. J. Miller, J. McMillan, A. Miles, M. ten Lohuis, and T. Mahony, *Gene* **93** (1990) 319.
23. J. Piatigorsky, J. Horwitz, and B. L. Norman, *J. Biol. Chem.* **268** (1993) 11894.
24. M. Naito, H. Ishiguro, T. Fujisawa, and Y. Kurosawa, *FEBS Lett.* **333** (1993) 271.

SAŽETAK

Kodirajuća regija gena za ekvinatoksin II nema introna

Gregor Anderluh, Jože Pungerčar, Borut Štrukelj, Peter Maček, i Franc Gubenšek

Morske vjetrenjače proizvode nekoliko toksičkih peptida i proteina. Ekvinatoksini (Eqt) izolirani iz morske vjetrenjače *Actinia equina* bazični su citolitički proteini čija je molekulska masa približno 20 kDa. Od triju Eqt koji su do sada izolirani, EqtII je najraširenij i dobro je opisan. Njegova genska organizacija još nije proučena. Da bi dobili informaciju o genskoj strukturi i sekvenci EqtII, izolirana je genomska DNA iz *A. equina*. Ciljna DNA umnožena je lančanom reakcijom polimerizacije (PCR) uz upotrebu tri različita para oligonukleotida, određena na osnovi sačuvanih regija cDNA Eqt. Sekvenca PCR proizvoda koju smo dobili upotrebljavajući oligonukleotid specifičan za EqtII se gotovo nije razlikovala od cDNA EqtII. DNA fragmenti genomske DNA dobiveni upotrebom PCR iste su dužine kao i oni iz kontrolnih PCR reakcija koje smo izveli na cDNA biblioteci *A. equina* i cDNA EqtII. To pokazuje, da je gen za EqtII bez introna u umnoženoj preproproteinskoj regiji. Prisustvo ovakog gena bez introna, koji kodira za ovaj citotoksički protein, može se objasniti pojavom knidarija relativno nisko na evolucijskom drvetu ili prednošću, koju omogućava brza ekspresija gena.