

Nucleotide sequence of cloned cDNA coding for honeybee prepromelittin

Reinhard VLASAK, Claudia UNGER-ULLMANN, Günther KREIL, and Anna-Maria FRISCHAUF

Institut für Molekularbiologie, Österreichische Akademie der Wissenschaften, Salzburg; and
Europäisches Institut für Molekularbiologie, Heidelberg

(Received March 31, 1983) – EJB 83 0324

Total mRNA from venom glands of young queen bees was transcribed into cDNA and cloned into the *Pst*I site of plasmid pBR322. The nucleotide sequence of two clones with inserts containing genetic information for prepromelittin is presented. The longer insert encompasses 374 base pairs including 52 nucleotides before the initiation codon and a 3' non-coding region of 112 base pairs. The 70 amino acids of prepromelittin represent the total coding capacity of the mRNA from which this insert is derived. Southern blot analysis with this cloned cDNA has shown that it hybridizes with a single *Eco*RI fragment of honeybee DNA which contains about 3000 base pairs.

Highly specialized cells, producing one or a few polypeptides in large excess, have played a major role in biochemical research. These were the first sources for the purification of specific mRNAs, which were used extensively for studies on protein synthesis in cell free systems. More recently mRNA preparations obtained from such cells served as starting material for the construction of cDNA clones, which then opened the way towards the isolation of the corresponding genes.

In our laboratory we have in the past years studied the biosynthesis of melittin, a peptide comprised of 26 amino acids, which represents the main constituent of honeybee venom [1]. This lytic peptide is liberated from a larger precursor, termed prepromelittin, through several processing steps [2–5]. For these studies we have used mRNA isolated from venom glands of young queen bees [6], which represent an extreme case of specialisation in that they synthesize almost exclusively the melittin precursor [7]. We have estimated that melittin mRNA, which contains about 400 nucleotides and a short poly(A) tail, represents more than 80% of the total mRNA of these glands [8].

In this communication we describe the construction of cDNA clones from venom gland mRNA using the plasmid pBR322. The nucleotide sequence of two of these clones, of which the larger contains the entire coding region of prepromelittin mRNA, is presented.

MATERIALS AND METHODS

Materials

Isotopes were purchased from Amersham International (England), restriction endonucleases, DNA polymerase and T₄ DNA ligase were from New England Biolabs (Frankfurt, FRG) or Bethesda Research Laboratories (Neu-Isenburg, FRG). Reverse transcriptase from avian myeloblastosis virus was kindly provided by Dr J. W. Beard (St Petersburg, USA), polynucleotide kinase and calf intestinal phosphatase were from Boehringer (Mannheim, FRG). S₁ nuclease, prepared as described by Vogt et al. [9], was a gift from H. Lehrach (EMBL, Heidelberg). Total mRNA from venom glands of newly emerged queen bees was prepared as described previously [6].

Preparation of double-stranded cDNA

About 3 µg poly(A)-rich RNA from queen bee venom glands and an equal amount of (dT)₁₀ were incubated with reverse transcriptase, the second strand was synthesized with DNA polymerase using published procedures [10].

Molecular cloning

One-half of the double-stranded cDNA was cloned into the *Pst*I site of plasmid pBR322 by the dG,dC-tailing procedure [11]. The other half of the dC-tailed double-stranded cDNA was annealed into the dG-tailed *Cl*aI site of pBR322. After transformation into *Escherichia coli* X1776, bacteria were grown on tetracycline plates in the former and on ampicillin plates in the latter case. To test for insertion into the *Pst*I site transformants were screened for sensitivity towards ampicillin, while tetracycline sensitivity was used to check for insertion into the *Cl*aI site.

Colony hybridization and characterization of hybrid plasmids

The cDNA clones were hybridized with ³²P-labeled cDNA prepared from total venom gland mRNA [12]. Of 100 colonies, 41 gave a strong hybridization signal and from these, minipreps were made as described by Birnboim and Doly [13]. Eight recombinant plasmids with inserts larger than 200 base pairs were isolated and purified by gel filtration over Biogel A50. Mapping with restriction endonucleases was done as described by Smith and Birnstiel [14]. The reaction conditions for the various enzymes were as recommended by the suppliers.

Cloning into pUC8

The insert of clone pBM13 was cut out with *Pst*I and ligated into the *Pst*I site of plasmid pUC8 [15] at a DNA concentration of 10 µg/ml. Recombinant plasmids were transformed into *E. coli* JM103 and grown on X-gal plates. The white colony pUM13/4 was selected for sequence analysis.

DNA sequencing

DNA fragments were end-labeled and sequenced according to Maxam and Gilbert [16]. The 5' ends were dephosphorylated

150

GGGGG-AGCGAATTAACAGAAATTAACAGGAAGGAAGGAAGGATCGGAGAAATC

100

ATG AAA TTC TTA GTC AAC GTT GCC CTT GTT TTT ATG GTC GTG TAC ATT TCT TAC ATC TAT GCG
Met Lys Phe Leu Val Asn Val Ala Leu Val Phe Met Val Val Tyr Ile Ser Tyr Ile Tyr Ala/

GCC CCT GAA CCG GAA CCG GCA CCA GAG CCA GAG GCG GAG GCA GAC GCG GAG GCA GAT CCG GAA
Ala Pro Glu Pro Glu Pro Ala Pro Glu Pro Glu Ala Glu Ala Asp Ala Glu Ala Asp Pro Glu

200

GCG GGA ATT GGA GCA GTT CTG AAG GTA TTA ACC ACA GGA TTG CCC GCC CTC ATA AGT TGG ATT
Ala/Gly Ile Gly Ala Val Leu Lys Val Leu Thr Thr Gly Leu Pro Ala Leu Ile Ser Trp Ile

300

AAA CGT AAG AGG CAA CAG GGT TAG TCGGATCCATCGATGCCGATTATCGATCTATCGAATCGTGAAAAATCTT
Lys Arg Lys Arg Gln Gln Gly END

374

ATTGCAACTTGAAGTAACATGTATACATGCTGATAATATAATTTCTCATTATTCTpolyA

Fig. 1. Nucleotide sequence of the cDNA inserts of clones pBM22 and pUM13/4 and the amino acid sequence of prepromelittin. Only the coding strands with the poly(dA) tail are shown. The d representing deoxy and the hyphens representing phosphodiester links have been omitted. The sequence is from clone pUM13/4, clone pBM22 starts at (*→). The bars (/) in the amino acid sequence mark the boundaries between the signal peptide, the pro part, and melittin (taken from [2])

with calf intestinal phosphatase and labeled with polynucleotide kinase and [γ - ^{32} P]ATP. For this reaction the recombinant plasmids were usually purified by an additional equilibrium centrifugation in CsCl. The 3' ends were labeled using the Klenow fragment of DNA polymerase I and [α - ^{32} P]deoxynucleoside triphosphates. Labeled fragments were isolated by electrophoresis in agarose gels and subsequently electroeluted into dialysis bags. After the base-specific cleavages, fragments were analyzed on 6%, 8% and 20% gels with a thickness of 0.2 mm.

Nick translation

Recombinant plasmid DNA was digested with *Pst*I and inserts were isolated by electrophoresis in agarose or polyacrylamide gels. For the nick translation a commercially available kit was used (Amersham International).

Isolation of honeybee DNA

Bee larvae obtained from local beekeepers were powdered under liquid nitrogen. The powder was suspended in 0.5 M EDTA, pH 8.0, containing 0.5% sarcosyl and incubated for 2 h in the presence of 0.1 mg/ml proteinase K. After three extractions with phenol/chloroform/isoamyl alcohol, the DNA was dialyzed against 50 mM Tris/HCl/10 mM EDTA/10 mM NaCl at pH 8.0 [17]. Finally the DNA was purified by equilibrium centrifugation in CsCl.

Genomic Southern blots

Bee DNA was completely digested with different restriction enzymes and fractionated on agarose gels (10–15 μ g digested DNA/slot). DNA fragments were transferred to nitrocellulose filters by the method of Southern [18] and hybridized with ^{32} P-labeled cDNA or with nick-translated inserts of different clones.

RESULTS

In view of the high abundance of melittin mRNA, it appeared likely that the corresponding cDNA should be present in a sufficient number of clones in a cDNA library made

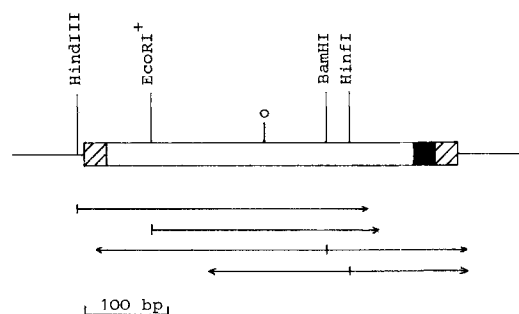


Fig. 2. Restriction endonuclease cleavage map and sequence analysis strategy for the insert of clone pUM13/4. Arrows indicate direction of sequencing and length of sequenced region. Cross-hatched boxes indicate dG·dC tails, the filled-in box the poly(dA) tail. bp = base pairs

from total venom gland mRNA so that no specific screening procedure would be required. Therefore, we simply selected those clones which strongly hybridized with total cDNA and chose from this group the ones with the largest inserts. The second of the clones selected for sequencing, pBM22, already had an insert derived from melittin mRNA. As shown in Fig. 1, the insert has a poly(dA) tract on one end and, 15 nucleotides away from it, the modified polyadenylation signal d(A-A-T-A-T-A-A-A). The clone has a single open-reading frame, which starts at the 5' end with residue 49 of prepromelittin, i.e. the fifth amino acid of melittin [2]. After the COOH-terminal glycine, found in prepromelittin synthesized *in vitro* [3], an amber stop codon is present.

A second clone with a larger insert, pBM13, was also found to contain the coding sequence for prepromelittin. In order to facilitate the sequencing the insert of the clone was cut out of pBR322 with *Pst*I and re-cloned into the *Pst*I site of plasmid pUC8, which contains a synthetic DNA segment with several restriction sites [15]. The sequencing strategy for the new clone, pUM13/4, is shown in Fig. 2. Its 3'-terminal half is identical with the insert found in pBM22. At the 5' end pUM13/4 starts with a stretch of 52 nucleotides, after which it contains a triplet for an AUG codon and an open-reading frame of 69 triplets followed by that for the amber stop codon already found in pBM22 (see Fig. 1). The amino acid sequence corresponding to this open-reading frame is identical to the sequence of prepro-

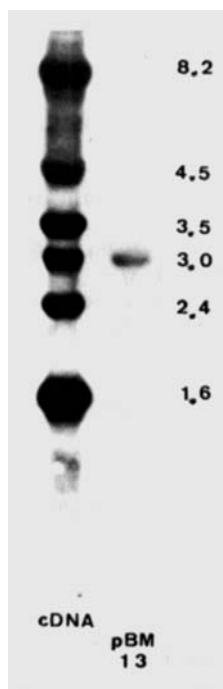


Fig. 3. Southern blot hybridization of honeybee DNA with venom gland cDNA and the cDNA insert of clone pUM13/4. *Eco*RI-digested bee DNA was separated on 0.8% agarose gels, transferred to nitrocellulose and hybridized with labeled cDNA and the nick-translated insert of clone pUM13/4. Numbers refer to 1000 base pairs

melittin determined by peptide analysis. Residue 17, which could not be determined previously, is found to be serine; residue 7 has recently been identified to be asparagine and this has been confirmed by the cDNA sequence. The total length of the insert in pUM13/4, excluding the poly(dA) tail, is 374 nucleotides and we assume that this is quite close to the total size of melittin mRNA.

Search for the melittin gene

Honeybee DNA was isolated as described under Materials and Methods, exhaustively digested with *Eco*RI, fractionated by electrophoresis in agarose gels and then used for Southern blot analysis. Hybridization with cDNA prepared from venom gland mRNA gave a complex pattern of six major bands corresponding to *Eco*RI fragments of 1600–8200 base pairs (see Fig. 3). Despite the high content of melittin mRNA in these glands, other mRNAs are apparently present in sufficient quantity to give strong hybridization signals under these conditions, where a large excess of radioactive cDNA is used. Using the nick-translated insert of clone pBM13 a much clearer result was obtained. As shown in Fig. 3, a single *Eco*RI fragment of the honeybee genome with length of approximately 3000 base pairs hybridized with the labeled insert.

DISCUSSION

The present communication describes the analysis of two cDNA clones derived from melittin mRNA, which are of different length but have an identical nucleotide sequence in the region present in both clones. The larger clone contains an insert of 374 nucleotides, excluding the poly(dA) tail, which is quite close to the 400 bases estimated for the mRNA on the

basis of electrophoretic mobility. This clone, pUM13/4 has 52 base pairs before the initiating methionine and, by comparison with the 5'-untranslated regions of other mRNAs, it seems unlikely that melittin mRNA is much longer than the insert found in this clone. The amino acid sequence of the single open-reading frame corresponds exactly to the sequence of prepro-melittin determined by analysis of the labeled polypeptide and its fragments. In addition, residue 17 could be identified as serine. The nucleotide sequence of clone pUM13/4 shows that melittin mRNA can only code for a polypeptide of 70 amino acids and not for any still larger precursors, a notion that could not rigorously be excluded by analysis of the translation product of this mRNA *in vitro* [2, 3]. In particular, after the COOH-terminal glycine found in prepromelittin, a stop codon is present in the mRNA. It has been suggested that this terminal glycine is involved in the formation of the amidated α -carboxyl group present at the end of the final product melittin [3]. Similar situations have since been found in a number of other precursors in that a glycine is present adjacent to the amino acid, which becomes the amidated carboxyl end in the mature product [19–23]. In all instances, however, this glycine residue is linked to one [23] or a pair [19–22] of basic amino acids and it has in fact been suggested that a Gly-B sequence, where B can be either lysine or arginine, might be required for the formation of terminal amides [20]. This possibility is certainly excluded for the case of melittin as no codons for basic amino acids are found next to the one for the terminal glycine. Moreover, it has recently been demonstrated that a partially purified enzyme preparation from pig pituitary slowly converts peptides with a terminal glycine to the amidated peptides lacking this amino acid [24]. This makes it very likely that not only in honeybee venom glands but also in mammalian cells a terminal glycine, exposed during the processing of peptide precursors, is the recognition signal for the amidation reaction.

Southern blots of genomic bee DNA digested with different restriction enzymes were hybridized with radioactive cDNA prepared from total venom gland mRNA. This yielded a complex pattern suggesting the presence of more than one melittin gene in the honeybee genome. However, the cloned melittin cDNA hybridized with only a single *Eco*RI fragment of about 3000 base pairs (see Fig. 3). The other strong bands obtained with total cDNA must, therefore, correspond to the genes for other mRNAs present in venom gland cells. Clearly the intensity of these hybridizing cDNAs does not reflect their abundance in the mRNA population, which must be very much lower than the content of melittin mRNA. Upon translation of total venom gland mRNA in cell-free systems prepromelittin is by far the most prominent product and only one other component with a molecular weight of around 20000 is detectable upon gel electrophoresis (K. Richter, unpublished experiments). Translation in presence of radioactive histidine, an amino acid not present in prepromelittin, reveals the formation of a small amount of another polypeptide which comigrates with the melittin precursor and which might be preprosecapin [25, 26]. Still, more than 80% of the radioactivity incorporated into polypeptides is present in prepromelittin and the corresponding mRNA is the only species detectable after gel electrophoresis of total venom gland mRNA and subsequent hybridization with radioactive poly(U) [8].

We are currently attempting to clone the *Eco*RI fragment containing the melittin gene in order to analyze its structure. In addition other cDNA clones, which do not contain inserts coding for melittin, are being analyzed and the results obtained with two of these, which code for preprosecapin, will be presented in a forthcoming publication.

The mRNA used for the original cloning experiments was kindly provided by Dr I. Kindas-Mügge (Institut für Krebsforschung, Vienna). We thank Dr H. Lehrach (EMBL Heidelberg) for help and advice and L. Reifberger for technical assistance. This work was supported by grant 4174 from the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung*.

REFERENCES

1. Habermann, E. (1972) *Science (Wash. DC)* 177, 314–318.
2. Suchanek, G., Kreil, G. & Hermodson, M. A. (1978) *Proc. Natl Acad. Sci. USA*, 75, 701–704.
3. Suchanek, G. & Kreil, G. (1977) *Proc. Natl Acad. Sci. USA*, 74, 975–978.
4. Kreil, G., Haiml, L. & Suchanek, G. (1980) *Eur. J. Biochem.* 111, 49–58.
5. Mollay, C., Vilas, U. & Kreil, G. (1982) *Proc. Natl Acad. Sci. USA*, 79, 2260–2263.
6. Kindas-Mügge, I., Lane, C. D. & Kreil, G. (1974) *J. Mol. Biol.* 87, 451–462.
7. Bachmayer, H., Kreil, G. & Suchanek, G. (1972) *J. Insect Physiol.* 18, 1515–1521.
8. Kindas-Mügge, I. (1979) *Trends Biochem. Sci.* 4, 135–137.
9. Vogt, P. M. (1973) *Eur. J. Biochem.* 33, 192–200.
10. Lehrach, H., Frischauf, A. M., Hanahan, D., Wozney, J., Fuller, F. & Boedtker, H. (1979) *Biochemistry*, 18, 3146–3152.
11. Rowekamp, W. & Firtel, R. A. (1980) *Dev. Biol.* 79, 409–418.
12. Grunstein, M. & Wallis, J. (1979) *Methods Enzymol.* 68, 379–389.
13. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acid Res.* 7, 1513–1523.
14. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acid Res.* 3, 2387–2393.
15. Vieira, J. & Messing, J. (1982) *Gene*, 19, 259–268.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499–560.
17. Blin, N. & Stafford, D. W. (1976) *Nucleic Acid Res.* 3, 2303–2308.
18. Southern, E. M. (1975) *J. Mol. Biol.* 98, 503–517.
19. Amara, S. G., David, D. N., Rosenfeld, B. A., Ross, B. A. & Evans, R. M. (1980) *Proc. Natl Acad. Sci. USA*, 77, 4444–4448.
20. Yoo, O. J., Powell, C. T. & Agarwal, K. L. (1982) *Proc. Natl Acad. Sci. USA*, 79, 1049–1053.
21. Land, H., Schütz, G., Schmale, H. & Richter, D. (1982) *Nature (Lond.)* 295, 299–303.
22. Hoffmann, W., Bach, T. C., Seliger, H. & Kreil, G. (1983) *EMBO J.* 2, 111–114.
23. Furutani, Y., Morimoto, Y., Shibahara, S., Noda, M., Takahashi, H., Hirose, T., Asai, M., Inayama, S., Hayashida, H., Miyata, T. & Numa, S. (1983) *Nature (Lond.)* 301, 537–540.
24. Bradbury, A. F., Finnie, M. D. A. & Smyth, D. G. (1982) *Nature (Lond.)* 298, 686–688.
25. Gauldie, J., Hanson, J. M., Shipolini, R. A. & Vernon, C. A. (1978) *Eur. J. Biochem.* 83, 405–410.
26. Kudelin, A. B., Martynov, V. I., Kudelina, I. A. & Miroshnikov, A. I. (1978) *Abstr 15th Eur. Peptide Symp. (Gdansk)* p. 84.

R. Vlasak, C. Unger-Ullmann, and G. Kreil,
Institut für Molekularbiologie der Österreichischen Akademie der Wissenschaften,
Billrothstrasse 11, A-5020 Salzburg, Austria

A.-M. Frischauf, Europäisches Laboratorium für Molekularbiologie,
Postfach 102209, Meyerhofstraße 1, D-6900 Heidelberg, Federal Republic of Germany