The EMBO Journal vol.3 no.9 pp.2071–2075, 1984

Insect immunity. Isolation and sequence of two cDNA clones corresponding to acidic and basic attacins from *Hyalophora cecropia*

K.Kockum, I.Faye, P.v.Hofsten, J.-Y.Lee, K.G.Xanthopoulos and H.G.Boman

Department of Microbiology, University of Stockholm, S-106 91 Stockholm, Sweden

Communicated by H.G.Boman

The Cecropia moth has three known classes of antibacterial immune proteins, attacins, lysozyme and cecropins (earlier referred to as P5, P7 and P9, respectively). Six attacins with different isoelectric points have been purified. The N-terminal sequences for five of these forms imply that only two different genes exist. We have now isolated and sequenced two cDNA clones, one for the basic attacin and one for the acidic form. The two mature proteins show 76% homology at the nucleotide level, while the regions beyond the stop codons are 36% homologous. The differences in the content of aspartic acid accounts for the difference in net charge between the acidic and basic attacin. Further differences in charge can be obtained by post-translational removal of a lysine-containing tetrapeptide at the C-terminal end of the two proteins. Evidence for a prepro form of the basic attacin is presented.

Key words: attacin/immune protein P5/cDNA clones/DNA sequencing/insect immunity

Introduction

The induction of immunity in diapausing pupae of the Cecropia moth is a suitable model system for the study of selective gene expression. The genes for immunity are activated by an injection of live bacteria and the insect responds by synthesizing three classes of antibacterial proteins; attacins, lysozyme and cecropins (earlier called P5, P7 and P9, respectively; see review by Boman and Steiner, 1981). Three main forms of cecropins exist, the A, B and D forms. The complete amino acid sequences imply the existence of three closely related cecropin genes originating from gene duplications (Steiner *et al.*, 1981; Hultmark *et al.*, 1982). Six forms of attacin with different isoelectric points were found in immune hemolymph, but the N-terminal sequences of five of the purified proteins suggested that only two different genes exist, one for a basic form and one for an acidic form (Hultmark *et al.*, 1983). The preparation of a cDNA bank and the isolation of a small attacin clone has recently been described (Lee *et al.*, 1983). We now report the isolation and sequence of two cDNA clones containing the full coding information for each of the two main forms of attacin. The basic attacin clone contains additional coding information for a precursor form of the mature protein. The parallel work on the amino acid sequence of attacin F is described in the accompanying paper by Engström *et al.*

Results

Our cDNA bank was screened for attacin clones in two ways. Initially we used our earlier clone pCP510 (Lee *et al.*, 1983) and later a synthetic probe (5'-AA$^{A}_{G}$TTNGG$^{C}_{T}$TCCCA-3') corresponding to amino acid residues 176–180 of the acidic attacin F (see Engström *et al.*, accompanying paper). The latter procedure gave one single clone denoted pCP521. The screening with the insert of pCP510 revealed 10 clones which were further characterized by dot-blot hybridization to the insert of pCP521 and by restriction enzyme analysis. The plasmid with the largest insert, pCP517, was selected for further studies.

Characterization and sequence of cDNA clones

Figure 1 shows the restriction maps obtained for pCP521 and pCP517. Suspected attacin clones were characterized with *Aval* and *Acc*I, both specific for the acidic attacin clone pCP521, and with *Kpn*I which cleaves both the acidic and the basic cDNA inserts. Restriction sites, other than those used in the screening, are also indicated in the figure. These restriction enzymes were used in the construction of fragments for sequencing and primers for extending the cDNA.

The results from the DNA sequencing are given in Figure 2. The solid line between the two amino acid sequences is interrupted whenever a base substitution occurs; when they give rise to amino acid substitutions, these are indicated by boxes. A comparison of the two protein sequences coded by the two different clones shows a homology of 79% at the amino acid level. The corresponding DNA homology is 76%, which is in contrast with the 36% found beyond the stop

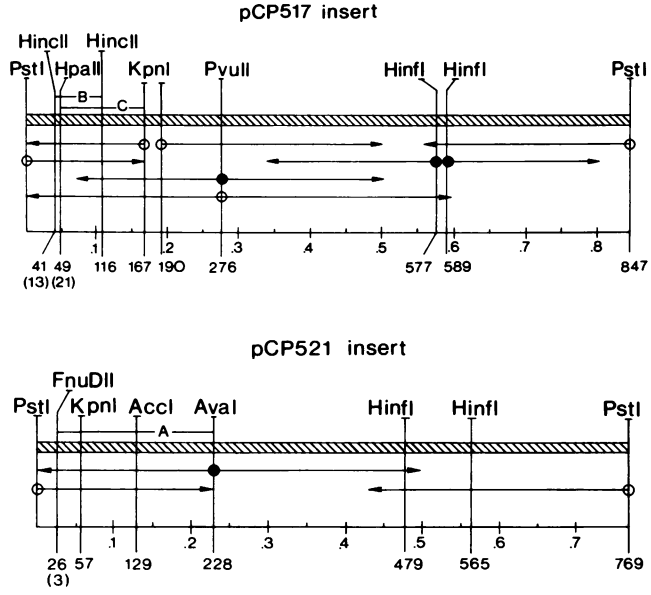

Fig. 1. Restriction analysis of the inserts in clones pCP517 and pCP521. Fragments A, B and C in the upper part of the maps were used in the primer extension experiments. The arrows in the lower part indicate the length and direction of the sequencing experiments. A filled circle in one end of an arrow denotes labelling at the 5' end, an open circle at the 3' end of the fragment. Numbers in parenthesis show the distance to the G/C tail.

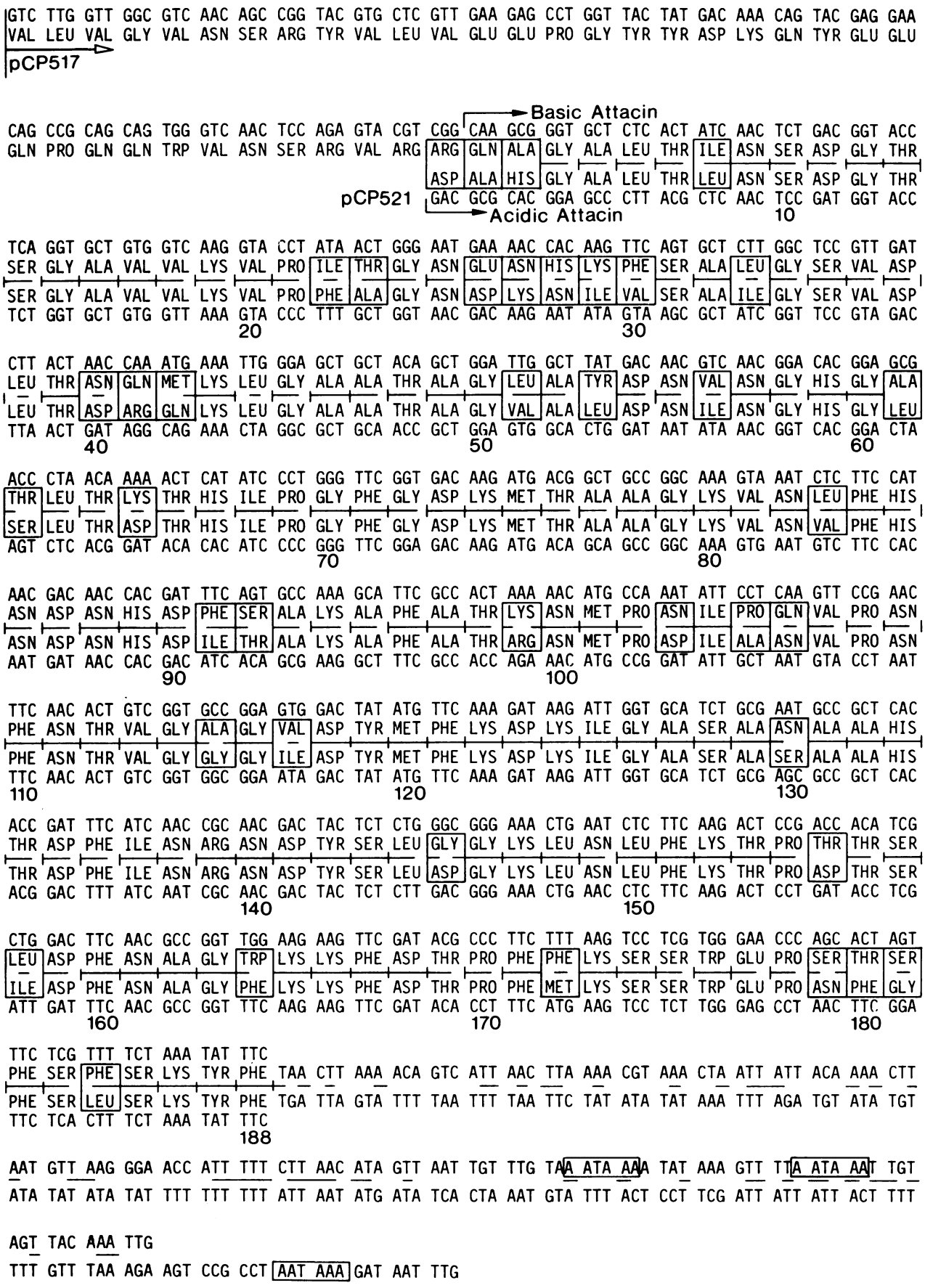

Fig. 2. Nucleotide sequences of the inserts in two attacin clones pCP517 (uppermost line) and pCP521 (bottom line). The amino acid sequence belonging to pCP517 is given below the nucleotide sequence, for pCP521 above. The line between the amino acid sequences is broken when the nucleotide sequences differ. Amino acid substitutions and the signals for polyadenylation are boxed in. The numbers indicate amino acid positions for both attacins.

Table I. Amino acid composition and codon usage in acidic and basic attacin

Amino acid	Residues per mature molecule		Base in wobble position (%)			
	Basic attacin	Acidic attacin	T	C	A	G
Asp	12	19	48	52	–	–
Asn	18	17	34	66	–	–
Thr	17	14	35	26	23	16
Ser	14	13	48	30	7	15
Glu	2	1	–	–	67	33
Gln	3	1	–	–	75	25
Pro	8	7	47	27	6	20
Gly	18	19	40	16	30	14
Ala	20	20	43	27	15	15
Cys	0	0	0	0	–	–
Val	9	10	16	26	32	26
Met	4	4	–	–	–	100
Ile	6	10	31	44	24	–
Leu	11	11	23	27	18	32
Tyr	4	3	71	29	–	–
Phe	16	15	13	87	–	–
His	6	6	17	83	–	–
Lys	16	14	–	–	47	53
Arg	1	3	0	50	25	25
Trp	2	1	–	–	–	100
Sum:	187	188				
Average codon usage (%):			30	36	16	18

All values are obtained from the sequences in Figure 1. The data for basic attacin assume that Gln is N-terminal (see Discussion).

signals. A common feature in the C terminus of both amino acid sequences is a tetrapeptide, Ser-Lys-Tyr-Phe. This structure is not present in attacin F (Engström *et al.*, accompanying paper).

The basic attacin clone (pCP517) codes for 36 amino acid residues at the N terminus. Since this sequence is not present in the mature protein it probably belongs to a prepro part of the molecule. The mol. wts. obtained from the sequences are 20 009 and 19 996 for the mature acidic and basic attacin, respectively (based on Asp as N-terminal amino acid in the acidic attacin and Gln in the basic one, Figure 2). These data agree well with the 20–23 K estimated by SDS-polyacrylamide gel electrophoresis (Hultmark *et al.*, 1983).

The regions beyond the stop codons differ in several respects. The basic attacin clone carries two signals for polyadenylation (Proudfoot and Brownlee, 1976) while only one is found in the acidic attacin clone. The two clones also differ in their potential for forming secondary structures which is much more prevalent in pCP521.

Table I shows the amino acid compositions of the basic and the acidic attacins. Both of the proteins lack cysteine, a fact which was evident already in the experimental analysis by Hultmark *et al.* (1983). These results could not distinguish glutamine and asparagine from the respective dicarboxylic acids and it was therefore not possible to explain any differences in isoelectric points. However, Table I shows that varying amounts of aspartic acid account for this difference between the acidic and the basic attacin. The differences in lysine and arginine content do not give rise to any change in net charge. The difference in aspartic acid content should not contribute significantly to the net charge at pH 4.0. In corroboration, electrophoresis at this pH does not separate the basic and acidic forms. Table I also shows C and T to be the most frequent bases in the wobble base position. Comparative data for many other eucaryotic systems show C and G to be the most frequent bases in this position (Wain-Hobson, 1981).

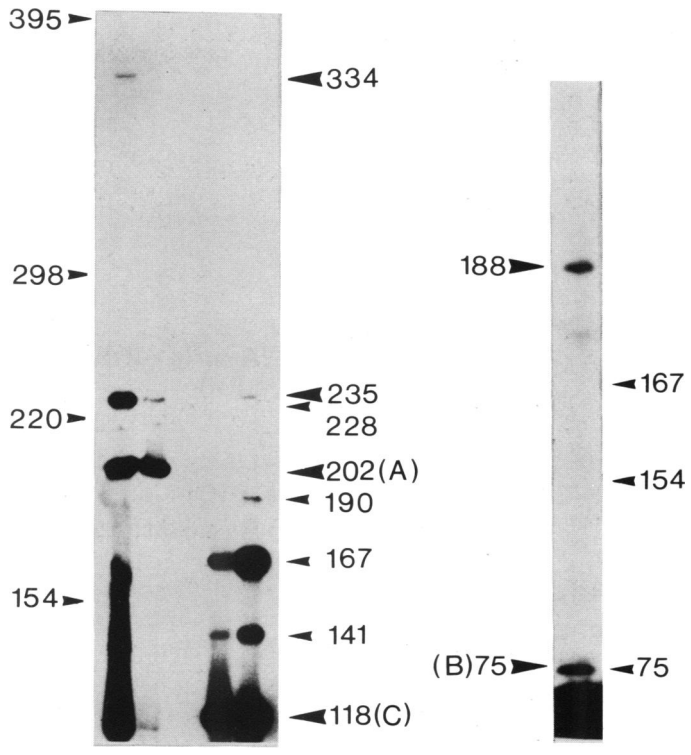

Fig. 3. Determination of missing mRNA information by primer extension. The gel (6% polyacrylamide, 7 M urea) to the right was for pCP517 alone, the one to the left was for pCP517 and pCP521. Size markers (arrows on the outer sides of the gels) were pBR322, cleaved with *Hinf*I. Details for the labelling are given in the text. Numbers refer to lengths in bases. Larger arrowheads point at fragments A, B and C and their extended product. Small arrowheads point at labelled fragments which were not extended. The two middle lanes of the left gel are control samples not extended.

Size of mRNAs for the two attacins

In neither of the two clones studied did we obtain the N-terminal start signal or the mRNA leader sequence (Figure 2). An attempt was therefore made by primer extension to estimate the amount of missing mRNA information. As primers we used fragments A, B and C (Figure 1). The 202-bp fragment A was produced by *Fnu*DII and *Ava*I cleavage of pCP521, while fragments B (75 bp) and C (118 bp) were both obtained from pCP517, B by digestion with *Hinc*II and C by use of *Hpa*II and *Kpn*I. T4 polynucleotide kinase and [γ-^{32}P]-ATP was used in the labelling of fragments A and B, while fragment C was labelled with terminal transferase and [α-^{32}P]ddATP. By the latter technique the 5′ ends of the DNA were also weakly labelled. In the case of fragments A and C we first isolated and labelled a larger fragment obtained by use of *Pst*I together with either *Ava*I or *Kpn*I, depending on the target plasmid. The labelled fragments were then split by *Fnu*DII and *Hpa*II, respectively. Primer extension was then carried out essentially as by Ghosh *et al.* (1978) with the modifications specified in Materials and methods. Figure 3 shows that fragment A gave an extended product of ~335 bases. This result means that ~130 bases are missing between the start of the mRNA and its corresponding first base of the cDNA. For fragments B and C similar exper-

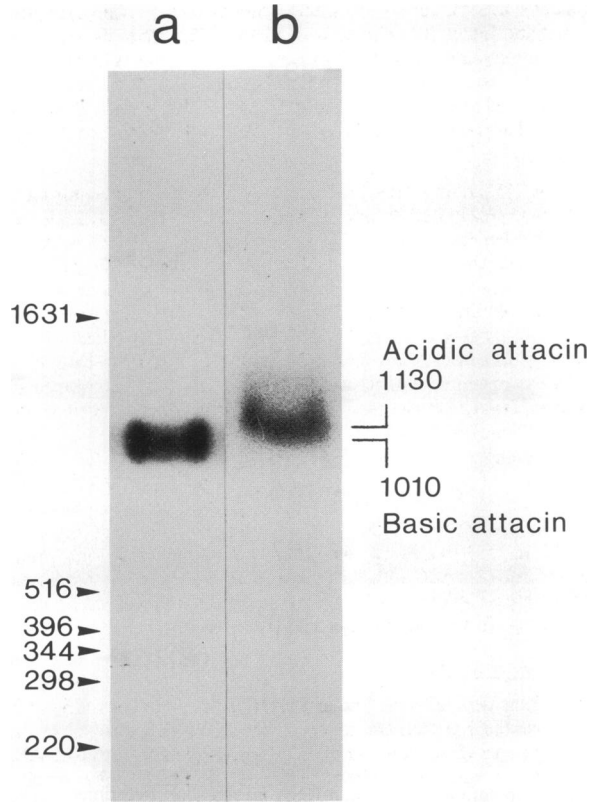

Fig. 4. Northern blot estimation of the size of mRNA for basic and acidic attacin. **Lane a**, basic attacin; **lane b**, acidic attacin. Size markers (arrows) were pBR322 cleaved with *HinfI*.

iments gave extended products of ~185 and 230 bases, respectively. This means that pCP517 lacks the first 95–100 bases of the mRNA information.

A Northern blotting experiment was performed to compare the mRNAs of the acidic and basic attacin (Figure 4), which we determined to be ~1130 and 1010 bases in length, respectively. This is in agreement with the 950 bases found with pCP510 (Lee *et al.*, 1983). These size determinations of the attacin mRNAs make it possible to estimate the number of mRNA bases which are missing from the 3′ region of the cDNA. The mRNA for basic attacin was found to extend ~100 bases downstream; whereas the acidic attacin mRNA is extended ~250 bases.

Discussion

Our previous work has shown immune hemolymph to contain six different forms of attacin, four basic forms (A–D) and two acidic ones (E and F). However, three of the four basic attacins showed identical N-terminal sequences (the fourth one could not be analyzed) and the same was true for the two acidic forms (Hultmark *et al.*, 1983). These results implied the existence of only two different genes. We have now isolated and analyzed two cDNA clones, one for each of the main forms of attacins. The coding parts for the mature proteins show a high degree of homology (76% on the nucleotide level and 79% on the amino acid level), while for the region beyond the stop codons the homology was only 36%. The two forms may therefore have evolved from a gene duplication stretching from the beginning of the mature attacin and extending to the stop signal. Alternatively, the data could possibly be interpreted to mean that the duplication is longer and that the difference in homology is due to a large difference in the selection pressure between the coding region and the non-coding parts of the genes.

There is a complete agreement between the amino acid sequence of attacin F (Engström *et al.*, accompanying paper) and residues 1–184 of the acidic attacin structure predicted from the DNA analysis. However, in both of the clones there is coding information for a C-terminal tetrapeptide with the sequence Ser-Lys-Tyr-Phe. This structure is missing from attacin F (Engström *et al.*, accompanying paper) and probably also from attacins C and D. It is therefore possible that this terminal peptide is removed proteolytically. At present it cannot be judged if this step is an artifact or a reaction of physiological significance. The assumed removal of the peptide would also eliminate one positive charge from the protein, a change which may explain the occurrence of the two acidic forms and of two of the basic forms. It is interesting that in the accompanying paper, Engström *et al.* have found that the N terminus of the basic attacins can be blocked by a pyroglutamate group. This finding is supported by the fact that in the mRNA a glutamine precedes the alanine residue proposed earlier as N terminus (Hultmark *et al.*, 1983). If indeed the mature protein starts with this glutamine residue, then the N terminus becomes adjacent to two arginine residues. This is a typical site for post-translational processing by trypsin-like enzymes (Steiner *et al.*, 1975; Richter, 1983). It should also be noted that the formation of pyroglutamate will remove one positive charge and the reaction can therefore, in combination with the removal of the tetrapeptide Ser-Lys-Tyr-Phe, explain all six forms of attacin.

Electrophoresis of the products formed after *in vitro* translation of crude mRNA and immunoprecipitation shows three not fully resolved bands with sizes of 27–28 K (Lee *et al.*, 1983, and results not shown), as opposed to the 20 K calculated from the sequence. This discrepancy agrees with the fact that the coding information in pCP517 extends 36 amino acids beyond the N-terminal glycine residue of attacin A/B. In this sequence there is a potential cleavage site for signal peptidase between amino acid seven and eight in agreement with the model of von Heijne (1983). Moreover, since 36 amino acids cannot account for a 7–8 K difference, we conclude that a prepro-attacin must exist as a precursor of the basic attacin. Further support for this comes from the primer extension experiments, which showed that pCP517 lacks ~100 bases that are present in the complete mRNA molecule. In the case of pCP521, the 5′ end coincides with the N-terminal region of the mature acidic attacin. Also, we expect that the 130 bases revealed by the primer extension experiments contain information for a prepro sequence. This is as far as our present data lend themselves to interpretations and the final assignment of the precursors to different attacins must await the isolation and sequencing of genomic clones, a project which is under way.

Materials and methods

Isolation and purification of DNA fragments

Restriction enzyme cleavage was as recommended by the suppliers. Agarose (Sigma) gel separations were done in a buffer containing 40 mM Tris-HCl, 30 mM sodium acetate and 2 mM EDTA adjusted with acetic acid to pH 7.8 (Loening, 1967). The gel contained 0.5 μg/ml of ethidium bromide. The samples were analyzed at 85–100 V for 15–45 min using a minigel apparatus (100 mm x 80 mm gel size). DNA was eluted from agarose or polyacrylamide gels by smashing the gel slice through a syringe without a needle into 20 ml of 0.3 M NaCl in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 (1 x TE buf-

fer). The sample was incubated for 18–48 h at 37°C. DNA was then adsorbed to a DE-52 cellulose (Whatman) column equilibrated with 1 x TE. The column was first washed with 1 x TE buffer, then with 0.3 M NaCl in 1 x TE and the DNA was finally eluted with 800 μl of 1 M NaCl in 1 x TE. After two subsequent ethanol precipitations, the DNA was considered pure enough for either sequencing, secondary restriction enzyme digestions, or radioactive labelling.

Nick-translation of probes

Labelling of probes by nick-translation was carried out according to Rigby *et al.* (1977). The DNA polymerase (New England Biolabs) was diluted 100 times before adding 4–5 units. 50 μCi of [α-^{32}P]dATP (New England Nuclear or Amersham) were used to label 0.1–1 μg of DNA. The dNTPs were from Pharmacia. Maximum labelling was obtained when the DNase I (0.05 pg, Boehringer-Mannheim) was added to the mixture of isotope, DNA, buffer, and cold nucleotides before the DNA polymerase. The probe was separated from free radioactive dATP on a G-50 column and precipitated with ethanol.

Screening of cDNA clones by dot-blot and synthetic primers

A synthetic oligomer, 5′-AA$\frac{A}{G}$TTNGG$\frac{C}{T}$TCCCA-3′ (kindly provided by KabiGen AB, Stockholm) was used to screen the cDNA bank for attacin clones. The oligomer was complementary to residues 176–180 in the C terminus of acidic attacin. The probe was labelled with T4 polynucleotide kinase and [γ-^{32}P]ATP, and subjected to G-50 purification. Colony hybridization was according to Thomas (1980). Dot-blot hybridization was performed by binding ~5 μg of plasmid DNA to a nitrocellulose filter as follows. The DNA was dissolved in 0.1 M Tris-HCl (pH 7.4). To this solution was added NaOH followed by the addition of 3 M NaCl, 0.3 M tri-sodium citrate (20 x SSC) to final concentrations of 0.2 M NaOH and 0.9 M NaCl, 0.09 M tri-sodium citrate (6 x SSC). The samples were then heated at 80°C for 20 min. Afterwards they were quickly chilled in ice water and 1 M Tris-HCl (pH 7.4) was added to a final concentration of 0.2 M. Using a manifold, small portions of the DNA solution (total volume 760 μl) were then bound to the filter under vacuum. After baking for 120 min at 80°C, the filters were hybridized to a probe as for colony hybridization (Thomas, 1980).

Sequencing of DNA

DNA sequencing was carried out according to Maxam and Gilbert (1980), with the addition of a T-specific reaction (Rubin and Schmid, 1980). End-labelling of generated 5′ ends (Maxam and Gilbert, 1980) of restriction fragments was done with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (New England Nuclear), while 3′ ends were labelled according to Amersham with terminal deoxynucleotidyl transferase (New England Biolabs) and [α-^{32}P]ddATP (Amersham). The sequence reactions were separated on polyacrylamide gels (8 or 20%, 7 M urea, Maxam and Gilbert, 1980) on an LKB Macrophor Electrophoresis Unit, essentially according to Garoff and Ansorge (1981). The glass plate was treated with Bind-Silane (LKB), and the thermostatic plate with Repel-Silane (LKB). Kodak X-OMAT RP1 films and Cronex Lightning Plus intensifying screens were used for autoradiography of the gels. Routine computer analysis (restriction sites, overlaps etc.) of the sequencing data was performed on an Apple IIe microcomputer using the Sequence Analysis Program, version 2.1 by Roger Larson, Department of Biochemistry, University of Minnesota.

Preparation of mRNA

The interior of four immunized pupae were thoroughly washed with 0.9% KCl and ground to a fine powder in a mortar with liquid nitrogen. After allowing the liquid nitrogen to evaporate, the powder was suspended in 50 mM Tris-HCl (pH 8.0), 10 mM EDTA and 1% SDS to a final volume of 60 ml. To this was added an equal volume of phenol/chloroform/isoamylalcohol (24:24:1). The mixture was shaken for 20 min at room temperature. After the organic phase was spun down (25 000 *g*, 20 min), it was again extracted for 5 min at room temperature with 10 ml of 50 mM Tris-HCl, 10 mM EDTA and 1% SDS. The two combined water phases were then extracted with 10 ml of phenol/chloroform/isoamylalcohol (24:24:1) for 20 min at room temperature. The nucleic acids (the water phase) were dialyzed at 4°C against 1 x TE (pH 8.0). The poly(A)-containing RNA fraction was then isolated by oligo(dT) chromatography (Aviv and Leder, 1972).

Primer extension

Extended cDNA clones were prepared according to Ghosh *et al.* (1978) with the following modifications. The appropriate ^{32}P-end-labelled DNA primer (10–50 x 10^3 c.p.m.) was combined with 5 μg of Cecropia mRNA, precipitated with ethanol, and dissolved in 10 μl containing 0.1 M NaCl, 20 mM Tris-HCl (pH 7.9), and 0.1 mM EDTA. The samples were heated at 100°C for 2 min in glass capillaries and then incubated at 60°C for 5–10 h. An equal volume was added, containing 80 mM Tris-HCl (pH 6.8) with 10 mM $MgCl_2$, 4 mM DTT, 0.4 mM of each dNTP (Pharmacia) and 5 units of AMV reverse transcriptase (New England Biolabs). After 5 min on ice, the sample was incubated at 37°C for 30–45 min, phenol extracted once, ether extracted three times and finally precipitated with ethanol. The products were analyzed on a 6% polyacrylamide gel with 7 M urea. *Hinf*I fragments of pBR322 to be used as mol. wt. markers, were labelled by a 'fill-in' reaction with [α-^{32}P]dATP and the Klenow fragment of DNA polymerase I. Bromophenol blue was used as tracker dye and allowed to run to the bottom of a 50 cm gel.

Northern blot

The mRNA (15 μg) was separated by electrophoresis on denaturing formamide agarose gels in MOPS buffer (Maniatis *et al.*, 1982). The gel was immediately blotted (Southern, 1975) to a nitrocellulose filter, essentially as described by Maniatis *et al.* (1982) with no intermediate washings. The nitrocellulose filter was soaked in 20 x SSC in order to have the same salt concentration as the blotting solution. After the blot was completed (usually within 16 h), the filter was washed in 2 x SSC and baked as described earlier. To detect any traces of remaining RNA, the gel was stained twice with 50 μg/ml of acridine orange in 10 mM phosphate buffer, pH 7.0 for 30 min each time at room temperature. If necessary, the gel was destained in the same buffer. Filters were then hybridized to the denatured probe at 42°C, in 5 x SSC, 50% deionized formamide, 50 mM sodium phosphate buffer (pH 6.5), 10% dextran sulphate, 100 μg/ml denatured herring sperm DNA, 0.2% SDS, and 5 x Denhardt's solution (Maniatis *et al.*, 1982). The filters were washed twice at room temperature in 2 x SSC, 0.1% SDS for 30 min each time, and then once for 30 min at 50°C. Finally the filters were washed twice for 20 min each time at 50°C in 0.1 x SSC with 0.1% SDS and subjected to autoradiography. The filters were not allowed to dry if they were to be rehybridized. In this case, washing was performed according to Thomas (1980).

Acknowledgements

We thank Hans Bennich and Åke Engström for a continuous flow of information about the work on the structure of attacin F, described in the accompanying paper (Engström *et al.*). Staffan Josephsson, KabiGen AB, has kindly helped us with the synthesis of the oligomers. We also thank Gun Hellström for skilful technical assistance. Financial support was provided by the Swedish Natural Science Research Council, contract BU 2453. K.G. Xanthopoulos was the recipient of a fellowship from 'Alexander S.Onassis' Public Benefit Foundation.

References

Aviv,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408-1412.
Boman,H.G. and Steiner,H. (1981) *Cur. Top. Microbiol. Immunol.*, **94/95**, 75-91.
Garoff,H. and Ansorge,W. (1981) *Anal. Biochem.*, **115**, 450-457.
Ghosh,P.K., Reddy,V.B., Swinscoe,J., Lebowitz,P. and Weissman,S.M. (1978) *J. Mol. Biol.*, **126**, 813-846.
Hultmark,D., Engström,Å., Bennich,H., Kapur,R. and Boman,H.G. (1982) *Eur. J. Biochem.*, **127**, 207-217.
Hultmark,D., Engström,Å., Andersson,K., Steiner,H., Bennich,H. and Boman,H.G. (1983) *EMBO J.*, **2**, 571-576.
Lee,J.-Y., Edlund,T., Ny,T., Faye,I. and Boman,H.G. (1983) *EMBO J.*, **2**, 577-581.
Loening,U.E. (1967) *Biochem. J.*, **102**, 251-257.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Richter,D. (1983) *Trends Biochem. Sci.*, **8**, 278-281.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.
Rubin,C.M. and Schmid,C.W. (1980) *Nucleic Acids Res.*, **8**, 4613-4619.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Steiner,D.F., Kemmler,W., Tager,H.S., Rubenstein,A.H., Lernmark,A. and Zucker,H. (1975) *Cold Spring Harbor Conf. Cell Prolif.*, **3**, 250-255.
Steiner,H., Hultmark,D., Engström,Å., Bennich,H. and Boman,H.G. (1981) *Nature*, **292**, 246-248.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.
von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17-21.
Wain-Hobson,S., Nussinov,R., Brown,R.J. and Sussman,J.L. (1981) *Gene*, **13**, 355-364.

Received on 18 June 1984