Cloning and analysis of the esterase genes conferring insecticide resistance in the peach-potato aphid, *Myzus persicae* (Sulzer)

Linda M. FIELD,* Martin S. WILLIAMSON, Graham D. MOORES and Alan L. DEVONSHIRE
AFRC Institute of Arable Crops Research, Rothamsted Experimental Station, Harpenden, Herts. AL5 2JQ, U.K.

Full-length cDNA clones encoding the esterases (E4 and FE4) that confer insecticide resistance in the peach-potato aphid [*Myzus persicae* (Sulzer)] were isolated and characterized. The E4 cDNA contained an open reading frame of 1656 nucleotides, coding for a protein of 552 amino acids. The FE4 cDNA shared 99% identity with E4 over this region, the most important difference being a single nucleotide substitution resulting in the FE4 mRNA having an extra 36 nucleotides at the 3′ end. The derived amino acid sequences for the N-terminus of E4 and FE4 were identical, with the first 23 residues being characteristic of a signal peptide and the next 40 residues being an exact match to the N-terminal sequence determined by Edman degradation of both purified proteins. The predicted molecular masses of 58.8 and 60.2 kDa for the E4 and FE4 polypeptides were consistent with those previously observed by *in vitro* translation of mRNA. Five potential N-linked glycosylation sites were present in both polypeptides, in accordance with earlier evidence that the native esterases are glycoproteins. Comparison of the aphid esterase protein sequences with other serine hydrolases provided evidence that their activity involves a charge-relay system with a catalytic triad the same as that found in acetylcholinesterase. Restriction mapping and sequencing of cloned genomic DNA showed that the E4 gene is spread over 4.3 kb with six introns and that the previously reported differences between the 3′ ends of the E4 and FE4 genes result from single nucleotide substitutions and not gross differences in the DNA sequences.

INTRODUCTION

The peach-potato aphid (*Myzus persicae*) is an economically important pest of agricultural and horticultural crops throughout the world. As such it has been subjected to extensive treatment with many types of insecticide, and resistance has developed widely in glasshouse and field populations [1].

In all cases studied so far, resistant *M. persicae* overcome the toxic effect of insecticides by the increased synthesis of a carboxylesterase capable of hydrolysing and sequestering insectidal molecules before they reach their target sites [1]. There are two very closely related forms of elevated esterase, E4 and FE4, which have not been found together in one insect; they differ slightly in both their molecular mass (FE4 and E4 are approx. 66 and 65 kDa respectively) and catalytic activity (FE4 is approx. 1.5-fold more active towards some insecticide groups) [2]. The overproduction of esterase protein correlates with high levels of E4 or FE4 mRNA [3], which in turn result from the presence of either amplified E4 or FE4 genes. The unamplified form of esterase enzyme and gene in susceptible aphids is not yet clear [4]. Preliminary restriction mapping of cloned aphid genomic DNA has shown that the amplified E4 and FE4 genes are very similar, probably differing only at the 3′ end [5].

Although there are many examples of insecticide resistance arising from increasing esterase activity, the only other case where this has been shown to result from gene amplification is for the B1 esterase of mosquitoes (*Culex quinquefasciatus*) [6]. Duplicate esterase genes have been reported in the fruitfly, *Drosophila melanogaster*, although here there is no link with insecticide resistance [7].

In the present paper we report the cloning of full-length cDNAs and genomic sequences of the amplified E4 and FE4 genes, and their analysis by restriction mapping and sequencing. The derived amino acid sequences corresponding to E4 and FE4 enzymes are compared with those of other serine hydrolases, including the insect esterases B1 (from *C. quinquefasciatus*), Est-6 (from *D. melanogaster*) and juvenile-hormone esterase (JHE) from the tobacco budworm, *Heliothis virescens*. The amino acid residues of E4/FE4, likely to be involved in the hydrolysis of insecticides, are discussed in the light of proposed models of esterase active sites, especially the catalytic triad of acetylcholinesterase (AChE) which has recently been established by atomic-structure determination [8] and mutagenesis studies [9].

MATERIALS AND METHODS

Aphid material

Two parthenogenic clones of insecticide-resistant aphids were used, 800F and 794J (R_3), containing approx. 60-fold levels of FE4 and E4 proteins respectively. The origins of these clones have been reported elsewhere [10].

Protein purification and N-terminal amino acid sequencing

E4 and FE4 proteins were purified as described previously from R_3 and 800F aphid clones respectively [11]. The first 40 N-terminal amino acids were determined by the AFRC Sequencing Laboratory, Department of Biochemistry, University of Leeds, Leeds, U.K. using automated solid-phase Edman degradation [12] after immobilizing freeze-dried protein on *p*-phenylene diisothiocyanate-derivatized glass beads [13].

Isolation of cDNA clones

Total RNA was extracted from 800F and R_3 aphid clones using guanidine isothiocyanate and purified by centrifugation through

Abbreviations used: AChE, acetylcholinesterase; JHE, juvenile-hormone esterase; SSPE, 0.9 M NaCl/50 mM NaH_2PO_4/5 mM EDTA, pH 7.7.

* To whom correspondence should be sent.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases.

5.7 M CsCl [14]. Polyadenylated [Poly(A)$^+$] RNA was selected by oligo(dT)–cellulose chromatography, and double-stranded cDNA was synthesized as described by Gübler and Hoffman [15]. After addition of *Eco*RI adaptors, the cDNAs were ligated into λZapII arms (Stratagene), packaged *in vitro* and libraries of approx 4×10^5 plaque-forming units (p.f.u.) plated on *Escherichia coli*, strain XL-1Blue. The libraries were screened with an E4 cDNA, pMp24 [4], in a solution containing 10 × Denhardt's, 2 × SSPE (0.9 M NaCl/50 mM NaH_2PO_4/5 mM EDTA, pH 7.7), 0.5% SDS, 200 μg/ml herring sperm DNA and 10 ng/ml ^{32}P-labelled probe. Hybridization was at 60 °C for 16 h, and the filters were washed in 2 × SSPE/0.1% SDS at 65 °C for 2 h. The 800F (FE4) and R_3 (E4) libraries gave 4 and 18 positive signals respectively. The four FE4 clones and four E4 clones, chosen at random, were plaque-purified and their cDNA inserts excised as pBluescript II phagemids using the helper phage VCS-M13 (Stratagene). Full-length cDNAs for FE4 (pFE4.1) and E4 (pE4.3) were identified by restriction-enzyme mapping and partial nucleotide sequencing.

DNA sequencing

Nested deletions of pE4.3 were generated using the ExoIII/Mung Bean deletion protocol (Stratagene) starting at the 5′ end. Dideoxy sequence reactions on double-stranded plasmid DNAs were done using the Sequenase Version 2 Sequencing Kit (United States Biochemical Corp.), with M13 forward primer. Overlapping clones were chosen to give the full 5′→3′ sequence. The reverse strand of pE4.3 was then sequenced using oligonucleotide primers (16-mers) complementary to regions at approx 300 bp intervals. Oligonucleotide primers were also used to sequence the FE4 cDNA, pFE4.1, in both directions and for limited sequencing of genomic DNA, in one direction only, to locate exon/intron borders (see Figure 3 below). The cDNA and derived amino acid sequences were analysed using GCG software [16].

Isolation and mapping of aphid genomic DNA sequences

Two fragments of genomic DNA, containing E4 sequences, were cloned from R_3 aphids; an 8 kb fragment from the *Eco*RI site in the gene (see Figure 3 below) to another *Eco*RI site downstream [17] and a 0.9 kb *Kpn*I/*Eco*RI fragment from the 5′ end of the gene (see Figure 3 below). Likewise, the 4 kb *Eco*RI fragment of FE4 genomic DNA, shown in Figure 3 (below), was cloned from 800F aphids. In each case the target DNA was enriched for the desired fragments [18], before cloning, by digesting with enzymes known not to have sites within these sequences [17], i.e. *Ava*I, *Bam*H1, *Pst*I, *Sal*I and *Xho*I for the *Eco*RI fragments and *Apa*I, *Hin*dIII, *Pst*I and *Xba*I for the *Kpn*I/*Eco*RI fragment. The DNA was then electrophoresed in 0.8%-agarose gels and the appropriate regions of the gel were excised and the DNA electroeluted.

The fragments were ligated into plasmid vectors, cut with the appropriate restriction enzymes; pUC8 was used for the 8 kb and 4 kb fragments and pBluescript for the 0.9 kb fragment. All fragments were cloned in an McrA$^-$ strain of *E. coli* (K802), which is essential in this case to prevent degradation of methylated sequences [19]. Positive clones were selected by probing colony blots and plasmid DNAs with either a partial E4 cDNA (pMp24; approx 1 kb; see [4]) or a 0.7 kb *Eco*RI fragment excised from the 5′ end of the pE4.3 cDNA and gel-purified. Hybridization and washing conditions were as described for the screening of cDNA libraries. Four positive clones of the 0.9 kb and four of the 8 kb E4 fragments were obtained from screening approx 1000 and 200 transformants respectively, and two of approx 500 colonies screened contained the 4 kb FE4 fragment. Restriction maps of the E4 and FE4 genomic fragments were constructed by single and double digests of the cloned sequences and assessment of fragment sizes on agarose gels stained with ethidium bromide.

RESULTS AND DISCUSSION

cDNA sequences and derived amino acid sequences of E4 and FE4

The nucleotide sequence of the E4 cDNA and its derived amino acid sequence are given in Figure 1; the corresponding FE4 sequences are only given where they differ from E4. Up to residue 527, the two genes show 99% identity at the DNA level and 98% identity at the amino acid level. There are 20 single nucleotide differences, nine of which change the corresponding amino acid; all are conservative for charge and six are also conservative for polarity. The close similarity of the predicted proteins is also evident from their identical computer-generated hydropathy and surface probability plots and predictions of secondary structure (results not shown). Most of the differences between E4 and FE4 are towards the 3′ end of the genes, with seven of the nine amino acid replacements being in the last 100 residues.

The most interesting difference is at position 1959, where there is an A in the E4 sequence and a T in the FE4. This results in a stop codon in E4 which is not present in FE4, and adds an extra 36 bases to the 3′ end of the FE4 mRNA before an in-frame stop codon is reached. The FE4 protein therefore contains an additional 12 amino acids with a predicted increase in molecular mass of 1.5 kDa. This is in good agreement with the relative sizes of the polypeptide synthesized *in vitro* from E4 and FE4 mRNAs [3], and with the observed molecular masses of the mature glycosylated FE4 and E4 proteins, which are approx 66 and 65 kDa as judged by SDS/PAGE [2]. In keeping with the glycoprotein nature of the esterases [3] the cDNA sequences show five potential sites of N-linked glycosylation (Figure 1) as predicted by the sequence Asn-Xaa-Ser/Thr, where Xaa is not Pro [20].

The sequences of E4 and FE4 cDNAs predict open reading frames of 1656 and 1692 bases respectively, with an in-frame start (ATG) codon showing good agreement to the consensus eukaryotic initiation codon context [21]. As discussed above, the open reading frames terminate at different stop codons and both are followed by moderately conserved untranslated regions. For E4 there is a putative polyadenylation signal, AATAAA (see Figure 1), at position 1814–1819.

The first 23 amino acids (−23 to −1), which are identical for E4 and FE4 (Figure 1), show all the features of a signal peptide [22], including a short positively charged N-terminal region (a positively charged lysine as part of five polar residues), a central hydrophobic core (residues −17 to −9) and a more polar C-terminus (residues −8 to −1). The Cys and Ala at positions −3 and −1 in E4 and FE4 meet the requirement that these should be small uncharged residues [22] if the cleavage site between the signal sequence and the mature protein is between the Ala and the Ser. This has been confirmed by N-terminal amino acid sequencing of both E4 and FE4 mature proteins, which gave Ser as the first residue. The amino acid sequences of the next 39 residues were identical for both proteins and an exact match for the residues predicted by the nucleotide sequence. This provides good evidence that the cloned sequences correspond to the E4/FE4 proteins. Allowing for the removal of the signal peptide the mature E4 polypeptide will be 529 amino acids and have a molecular mass of 58.8 kDa, and FE4 will be 541 amino acids with a molecular mass of 60.2 kDa.

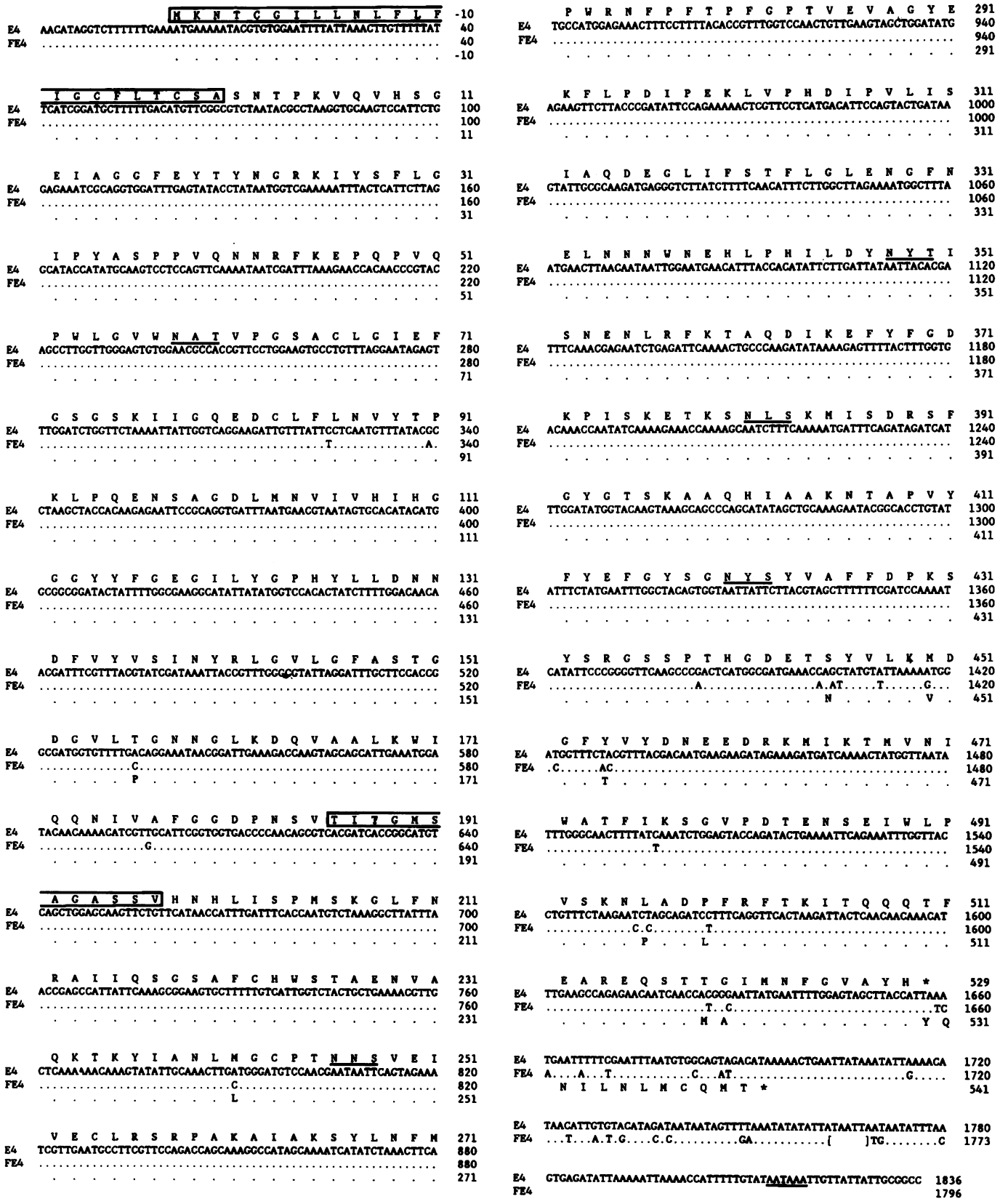

Figure 1 Nucleotide sequences and derived amino acid sequences of aphid esterases E4 and FE4

The nucleotide and amino acid sequences are given in full for E4, and where they differ for FE4 {dots represent the same bases and amino acids and the region in brackets ([ ]) is a deletion in the FE4 sequence}. Amino acid residues of the signal peptide (−23 to −1) and those surrounding the active-site serine residue (191) are boxed. Potential N-linked glycosylation sites and the putative polyadenylation signal (AATAAA) are underlined.

Table 1 Amino acid sequences surrounding the active-site serine residue (*) in some serine hydrolyses

The amino acid sequences from published data are aligned with esterases E4/FE4 (the present study).

Hydrolase	Sequence	Reference
Serine proteinases, subtilisin family		
Subtilisin BPN	AYNGTS*MASPHV	25
Serine proteinases, trypsin family		
Bovine trypsin	SCQGDSGGPVVC	25
Drosophila trypsin-like	ACQGDSGGPLVS	25
Hornet trypsin-like	ACHGDSGGPLVA	25
Serine esterases		
Torpedo AChE	TIFGESAGGASV	8
Human AChE	TLFGESAGAASV	26
Drosophila AChE	TLFGESAGSSSV	27
Mosquito AChE	TLFGESAGGSSV	28
Rat carboxylesterase	TIFGESAGGVSV	29
Mouse carboxylesterase	TIFGESSGGISV	30
Rabbit carboxylesterase	TIFGESAGGQSV	31
Drosophila Est-6	LLVGHSAGGASV	24
Heliothis JHE	TIAGQSAGASAA	23
Mosquito B1 Est	TLAGHSAGAASV	6
E4/FE4	TITGMSAGASSV	The present study

Comparison of E4/FE4 proteins with other serine hydrolases

Of the best ten alignments to a translation of the E4 cDNA [16], nine were to serine hydrolases, including mammalian cholinesterases (EC 3.1.1.8), AChEs (EC 3.1.1.7), other mammalian carboxylesterases (EC 3.1.1.1) and the three insect esterases: JHE (EC 3.1.1.59) from *H. virescens* [23], the amplified B1 esterase (EC 3.1.1.1) from mosquito [6] and *Drosophila* Est-6 (EC 3.1.1.1) [24]. The similarities between these sequences were confined to the N-terminal halves of the proteins. In addition, there was similarity to thyroglobulin, as reported previously for other carboxylesterases [23,24].

The similarity between E4/FE4 and other serine hydrolases suggests that a comparison of conserved residues might implicate those involved in the catalytic activity of E4/FE4. Table 1 shows the amino acid residues surrounding the active-site serine residue in the published sequences of some serine hydrolases aligned with the corresponding serine in E4 and FE4 (where all 12 residues are identical). Considerable similarity is seen between the aphid esterases and AChEs (eight out of 12 identical residues), mammalian carboxylesterases (seven or eight out of 12) and the insect esterases (eight out of 12 for JHE and B1, but only six for Est-6).

Although the active-site serine residue is well established in serine hydrolases, the role of other amino acid residues in catalysis in less certain. Both the 'trypsin' and 'subtilisin' families of serine proteinases have a triad of precisely orientated active-site amino acid residues (His^{57}, Asp^{102} and Ser^{195} for the trypsin family, and Asp^{32}, His^{64} and Ser^{221} in subtilisin) which stabilize the catalytic transition state and have been designated as the 'charge-relay' system. The role of these three residues has been confirmed by site-directed mutagenesis [32].

The elucidation of the three-dimensional structure of AChE from *Torpedo californica* has established unequivocally that in this case the catalytic triad consists of Ser^{200}, Glu^{327} and His^{440}, with Glu replacing Asp, and having the opposite 'handedness' to the serine proteinases [8]. The Glu and His residues are highly conserved in other AChEs, and recently Schafferman et al. [9] provided biochemical evidence (site-directed mutagenesis) for the involvement of the corresponding residues, Ser^{203}, Glu^{327} and His^{447}, in catalysis by human AChE. It has been proposed that the catalytic function of mammalian carboxylesterases also involves a His residue [29,30].

As far as insect esterases are concerned, the two AChEs from *Drosophila* [27] and mosquito [28] contain Ser, Glu and His residues corresponding to those of *Torpedo* and human AChE, and it seems likely that they function in the same way. However, it has been proposed that the charge-relay system in *Drosophila* Est-6 involves an Asp^{160} and Ser^{188} [33] with either an Arg on the N-terminal side of the Ser or an unspecified His on the carboxy side [34]. For JHE, site-directed mutagenesis has recently been used to predict Ser^{201}, Glu^{332} and His^{447} as the catalytic triad analogous to that of AChE [35].

So where do E4 and FE4 fit into the various models of catalytic function? Figure 2 shows an alignment of E4 amino acid residues with those of *Torpedo* AChE, *Drosophila* Est-6 and *Heliothis* JHE. E4 shows conservation of residues Ser^{191}, Glu^{316} and His^{440} corresponding to those purported to be involved in catalysis by AChE and JHE, and it is also apparent that E4 as well as *Torpedo* AChE and JHE, have conserved residues corresponding to the proposed catalytic Asp^{160} of Est-6. The three-dimensional structure of *Torpedo* AChE ([8]; analysed by R. Taylor, personal communication) indicates that for AChE the equivalent Asp^{172} along with Arg^{44} and Glu^{92} are not located in the catalytic 'gorge' but are associated with a cavity, approx. 2.0–2.5 nm (20–25Å) away from the catalytic serine, and that within this cavity Glu^{92} is salt-bridged with Arg^{44}, and Asp^{172} with another Arg, Arg^{149}, which is also conserved in the esterases shown in Figure 2. This cavity is the only characteristic feature of the three-dimensional structure that is conserved across several cholinesterases and is probably important for structural reasons.

Thus it seems likely that E4 and FE4, like *Torpedo* AChE, have residues Arg^{43}, Glu^{81}, Arg^{141} and Asp^{163} conserved for structural reasons and that it is the Ser^{191}, Glu^{316} and His^{440} of E4 and FE4 that function as the catalytic triad analogous to that of AChE and JHE.

Another aspect of three-dimensional structure is the formation of disulphide bridges, and here we can again compare E4/FE4 with other esterases. *Torpedo* AChE has bonds Cys^{67}–Cys^{94}, Cys^{254}–Cys^{265} and Cys^{402}–Cys^{521} with an unpaired Cys^{231} and a Cys^{572} involved in intersubunit bonding [36]. By analogy E4/FE4 could have pairings between Cys^{66} and Cys^{83} and between Cys^{243} and Cys^{254}, with Cys^{222} being unpaired (see Figure 1). There are no other residues to constitute a third disulphide bond, and this is comparable with other mammalian carboxylesterases where only four cysteines are present [26–28]. *Drosophila* Est-6 [7] and JHE [23] have six cysteine residues, and the mosquito B1 esterase has ten [6], but there is no evidence for specific pairings to generate disulphide bonds.

Genomic organization of E4 and FE4 genes

Maps of the E4 and FE4 genes showing the relative positions and sizes of exons and introns are shown in Figure 3. The E4 coding sequences are interrupted by six introns and spread over 4.3 kb, and the FE4 gene is identical from the *Eco*RI site at the 5′ end of the cloned fragment. The *Hin*dIII and *Bgl*II restriction sites present in FE4, but not E4, are the result of single nucleotide substitutions and not gross differences between the two genes. This is contrary to the previous suggestion that there might be a rearrangement within the gene caused by the chromosome translocation present in aphid clones with amplified E4 genes [6].

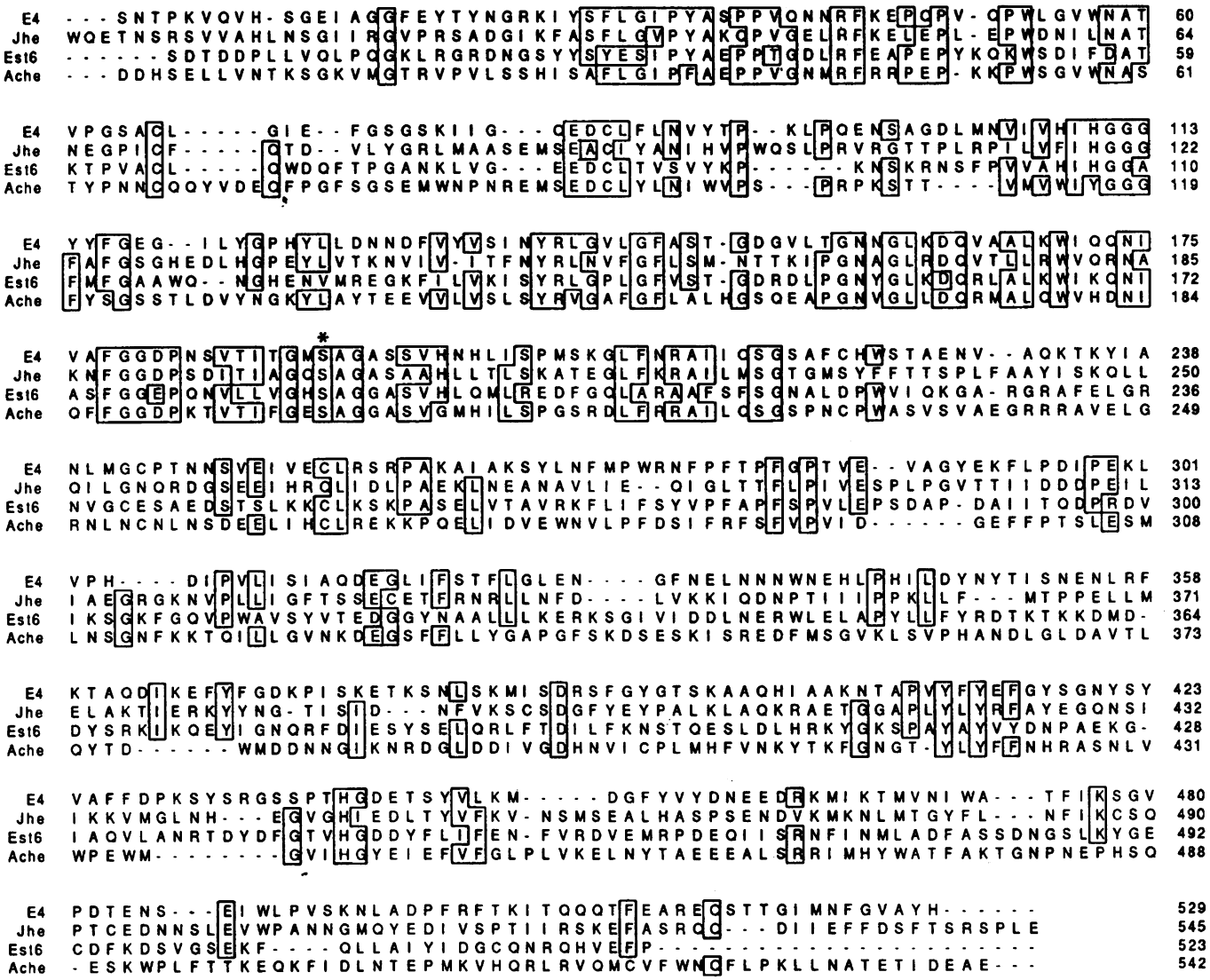

Figure 2 Alignment of the E4-derived amino acid sequence with other serine esterases

The E4 sequence is aligned with the published sequences for juvenile-hormone esterase (Jhe) from *Heliothis virescens* [23], esterase 6 (Est-6) from *Drosophila melanogaster* [24] and acetylcholinesterase (Ache) from *Torpedo californica* [8]. Residues conserved in at least three of the sequences are boxed, and the active-site serine residue is marked with an asterisk.

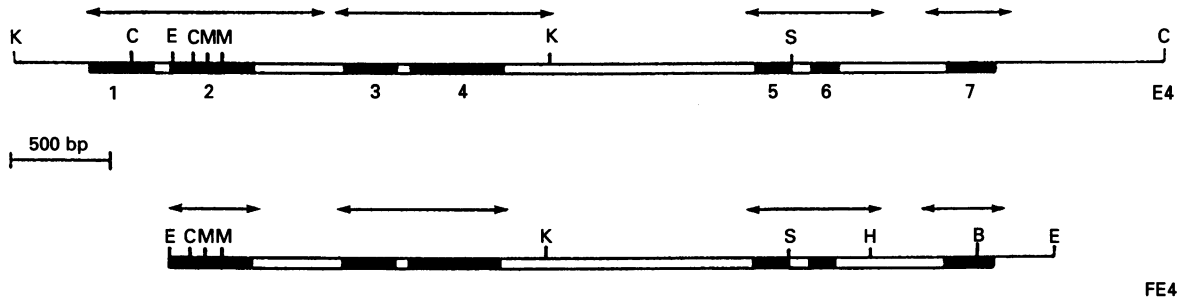

Figure 3 Restriction maps and location of introns in E4 and FE4 genomic sequences

The cloned genomic DNAs were mapped by single and double restriction digests. Limited sequencing in one direction, for the regions indicated ↔, was used to locate exon/intron borders; exons 1–7 are indicated by solid bars. Abbreviations for restriction enzymes: B, *Bgl*II; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; M, *Msp*I; and S, *Sma*I.

The strong similarities between the E4 and FE4 genes suggest that before amplification they were allelic. It is known by restriction mapping that the flanking DNA of many aphid clones is the same for all FE4 types and for all E4 types, but that the two types differ (L. M. Field, A. L. Devonshire and N. Javed, unpublished work). It is not yet clear whether the differences result from point mutations or from different stretches of DNA flanking the two genes. Current work should provide information on the evolution of the amplified aphid esterase genes and the corresponding esterase gene present in susceptible aphids.

It has previously been reported that an *Msp*I site at the 5′ end of E4 and FE4 genes and the *Sma*I (*Msp*I) site present in both, contain 5-methylcytosine (as a CpG) in aphids with expressed amplified esterase genes, but that this methylation is absent in aphids with unexpressed amplified genes [10]. It is now clear from the sequence data that there are in fact two *Msp*I sites in the second exon of E4 and FE4 (as shown in Figure 3) that are not resolved by restriction mapping, and that these must both be methylated in the expressed genes. The methylated *Sma*I site lies in exon 5 very close to the intron border. If or how these sites might affect expression of the aphid esterase genes remains unclear and we are currently cloning and sequencing upstream regions of the E4/FE4 genes to identify any other methylation sites that might affect the binding of regulatory proteins.

We thank Mary Stribley for rearing the aphid clones, D. Pappin (AFRC Sequencing Laboratory) for sequencing the esterase proteins and R. Taylor (ICI) for providing an insight of the three-dimensional structure of AChE. The Resistance Group at Rothamsted is a member of the European Network for Insect Genetics in Medicine and Agriculture (ENIGMA).

REFERENCES

1 Devonshire, A. L. (1989) in Aphids, Their Biology, Natural Enemies and Control (Minks, A. K. and Harrewijn, P. eds.), pp. 123–139, Elsevier, Amsterdam
2 Devonshire, A. L. (1989) Pestic. Sci. **26**, 375–382
3 Devonshire, A. L., Searle, L. M. and Moores, G. D. (1986) Insect Biochem. **16**, 659–665
4 Field, L. M., Devonshire, A. L. and Forde, B. G. (1988) Biochem. J. **251**, 309–312
5 Field, L. M. and Devonshire, A. L. (1991) in Resistance '91: Achievements and Developments in Combating Pesticide Resistance (Denholm, I., Devonshire, A. L. and Hollomon, D. W., eds.), pp. 240–250, Elsevier Applied Science, London and New York
6 Mouches, C., Pauplin, Y., Agarwal, M., Lemieux, L., Herzog, M., Abadon, M., Beyssat-Arnaouty, V., Hyrien, O., Robert De Saint Vincent, B., Georghiou, G. P. and Pasteur, N. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 2574–2578
7 Collet, C., Nielsen, K. M., Russell, R. J., Karl, M., Oakeshott, J. G. and Richmond, R. C. (1990) Mol. Biol. Evol. **7**, 9–28
8 Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L. and Silman, I. (1991) Science **253**, 872–879
9 Schafferman, A., Kronman, C., Flashner, Y., Leitner, M., Grosfeld, H., Ordentlich, A., Gozes, Y., Cohen, S., Ariel, N., Barak, D., Harel, M., Silman, I., Sussman, J. L. and Velan, B. (1992) J. Biol. Chem. **267**, 17640–17648
10 Field, L. M., Devonshire, A. L., ffrench-Constant, R. H. and Forde, B. G. (1989) FEBS Lett. **243**, 323–327
11 Devonshire, A. L. and Moores, G. D. (1982) Pestic. Biochem. Physiol. **18**, 235–246
12 Wachter, E., Machleidt, W., Hofner, H. and Otto, J. (1973) FEBS Lett. **35**, 97–102
13 Laursen, R. A. (1971) Eur. J. Biochem. **20**, 89–102
14 Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
15 Gübler, U. and Hoffman, B. J. (1983) Gene **25**, 263–269
16 Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395
17 Field, L. M. (1989) PhD Thesis, Council for National Academic Awards
18 Nicholls, R. D., Hill, A. V. S., Clegg, J. B. and Higgs, D. R. (1985) Nucleic Acids Res. **13**, 7569–7577
19 Raleigh, E. A., Murray, N. E., Revel, H., Blumenthal, R. M., Westaway, D., Reith, A. D., Rigby, P. W. J., Elhai, J. and Hanahan, D. (1988) Nucleic Acids Res. **16**, 1563–1575
20 Bause, E. (1983) Biochem. J. **209**, 331–336
21 Kozak, M. (1984) Nucleic Acids Res. **12**, 857–872
22 von Heijne, G. and Abrahmsén, L. (1989) FEBS Lett. **244**, 439–446
23 Hanzlik, T. N., Abdel-Aal, Y. A. I., Harshman, L. G. and Hammock, B. D. (1989) J. Biol. Chem. **264**, 12419–12425
24 Oakeshott, J. G., Collet, C., Phillis, R. W., Nielsen, K. M., Russell, R. J., Chambers, G. K., Ross, V. and Richmond, R. C. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 3359–3363
25 Dayhoff, M. O. (1978) in Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed.), vol. 5, suppl. 3, pp. 73–93, National Biomedical Research Foundation, Washington
26 Soreq, H. and Prody, C. A. (1989) in Computer-Assisted Modelling of Receptor–Ligand Interactions: Theoretical Aspects and Applications to Drug Design, (Rein, R. and Golombeck, A., eds.), pp. 347–359, Alan R. Liss, New York
27 Hall, L. M. C. and Spierer, P. (1986) EMBO J. **5**, 2949–2954
28 Hall, L. M. C. and Malcolm, C. A. (1991) Cellular and Molecular Neurobiology **11**, 131–141
29 Long, R. M., Satoh, H., Martin, B. M., Kimura, S., Gonzalez, F. J. and Pohl, L. R. (1988) Biochem. Biophys. Res. Commun. **156**, 866–873
30 Ovnic, M., Tepperman, K., Medda, S., Elliott, R. W., Stephenson, D. A., Grant, S. G. and Ganschow, R. E. (1991) Genomics **9**, 344–354
31 Korza, G. and Ozols, J. (1988) J. Biol. Chem. **263**, 3486–3495
32 Carter, P. and Wells, J. A. (1988) Nature (London) **332**, 564–568
33 Cooke, P. H. and Oakeshott, J. G. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 1426–1430
34 Myers, M., Richmond, R. C. and Oakeshott, J. G. (1988) Mol. Biol. Evol. **5**, 113–119
35 Ward, V. K., Bonning, B. C., Huang, T., Shiotsuki, T., Griffith, V. N. and Hammock, B. D. (1992) Int. J. Biochem. **24**, 1933–1941
36 MacPhee-Quigley, K., Vedvick, T. S., Taylor, P. and Taylor, S. S. (1986) J. Biol. Chem. **261**, 13565–13570

Received 25 January 1993/7 April 1993; accepted 13 April 1993