Functional Conservation of the Active Sites of Human and *Drosophila* Angiotensin I-Converting Enzyme[†]

David Coates,[‡] R. Elwyn Isaac,*[,‡] Joel Cotton,[§] Richard Siviter,[‡] Tracy A. Williams,[||] Alan Shirras,[⊥] Pierre Corvol,[||] and Vincent Dive[§]

School of Biology, University of Leeds, Leeds LS2 9JT, U.K., Department d'Etudes et d'Ingéniérie des Protéins, Commissariat à l'Energie Atomique, CE-Saclay, 91191 Gif-Sur-Yvette, France, INSERM Unit 36, College de France, 3 Rue d'Ulm, 75005 Paris, France, and Department of Biological Sciences, University of Lancaster, Lancaster LA1 4YQ, U.K.

Received March 15, 2000; Revised Manuscript Received May 16, 2000

ABSTRACT: Human somatic angiotensin I-converting enzyme (sACE) has two active sites present in two homologous protein domains, resulting from a tandem gene duplication. It has been proposed that the N- and C-terminal active sites can have specific in vivo roles. In *Drosophila melanogaster*, *Ance* and *Acer* code for two ACE-like single-domain proteins, also predicted to have distinct physiological roles. We have investigated the relationship of Ance and Acer to the N- and C-domains of human sACE by genomic sequence analysis and by using domain-selective inhibitors, including RXP 407, a selective inhibitor of the human N-domain. These phosphinic peptides were potent inhibitors of Acer, but not of Ance. We conclude that the active sites of the N-domain and of Acer share structural features that permit the binding of the unusual RXP407 inhibitor and the hydrolysis of a broader range of peptide structures. In comparison, Ance, like the human C-domain of ACE, displays greater inhibitor selectivity. From the analysis of the published sequence of the *Adh* region of *Drosophila* chromosome 2, which carries *Ance*, *Acer*, and four additional ACE-like genes, we also suggest that this functional conservation is reflected in an ancestral gene structure identifiable in both protostome and deuterostome lineages and that the duplication seen in vertebrate genomes predates the divergence of these lineages. The conservation of ACE enzymes with distinct active sites in the evolution of both vertebrate and invertebrate species provides further evidence that these two kinds of active sites have different physiological functions.

Mammalian angiotensin I-converting enzyme (ACE,[1] EC 3.4.15.1) is a dipeptidyl carboxypeptidase responsible for the conversion of angiotensin I to the powerful vasoconstrictor, angiotensin II, and the inactivation of the vasodilatory peptide, bradykinin by the removal of C-terminal dipeptides (*1*). Human somatic ACE is a two-domain protein resulting from a tandem gene duplication (*2*), with each domain possessing dipeptidyl carboxypeptidase activity and distinctive substrate specificity (*3*). Because of its central role in the metabolism of vasoactive peptides, ACE has attracted intense interest from the point of view of the development of orally active ACE inhibitors to treat hypertension and because the *ACE* gene is a candidate gene for several cardiovascular diseases (*4*).

There are two forms of mammalian ACE, somatic ACE (sACE, M_r 150000−180000) and testicular or germinal ACE (gACE, M_r, 100000−110000), which is unique to developing spermatids (*5*, *6*). Both isoforms are primarily plasma membrane enzymes, anchored by a common C-terminal hydrophobic region. gACE is coded for by exons 13−26 of the *ACE* gene, with a short amino-terminal sequence unique to gACE being encoded by exon 13, which is spliced from the somatic ACE mRNA (*2*). The physiological substrate for gACE has not been identified, but the enzyme is known to play a key role in reproduction, since male mice lacking gACE display reduced fertility (*7*).

Ever since the discovery that sACE has two active sites, there has been much speculation about the functional significance of this duplication. Although the two domains of sACE have a relatively broad substrate specificity and share some enzymatic properties, such as the ability to hydrolyze bradykinin and angiotensin I, there are several biochemical features that differentiate between the two active sites (*8*). For example, the haemoregulatory peptide, *N*-acetyl-Ser-Asp-Lys-Pro (AcSDKP), a peptide that inhibits the recruitment of pluripotent haematopoietic stem cells and normal early progenitors into the S-phase of the cell cycle is an in vivo substrate for the N-domain, but not the C-domain, of human ACE (*9*). In vitro experiments, have shown that angiotensin 1−7 and Met-enkephalin-Arg-Phe are also preferentially cleaved by the N-domain (*10*, *11*). The N- and C-domains can also be differentiated by their responses to chloride ions and by the relative potencies of some inhibitors (*3*, *12*). Recently, a screen of a combinatorial

[†] R.E.I., D.C., and A.D.S. thank the BBSRC (24/S09564, 89/S09563, and ISIS) for financial support.

* To whom correspondence should be addressed. Phone: 0113 2332903. Fax: 01132332835. E-mail: r.e.isaac@leeds.ac.uk.

[‡] University of Leeds.

[§] Department d'Etudes et d'Ingéniérie des Protéins.

[||] INSERM Unit 36.

[⊥] University of Lancaster.

[1] Abbreviations: ACE, angiotensin I-converting enzyme; Ance, *D. melanogaster* angiotensin I-converting enzyme; Acer, *D. melanogaster* ACE-related; EST, expressed sequence tag; Dpa, N3 (2,4,-dintrophenyl) L-2,3-diamino propionyl.

10.1021/bi000593q CCC: $19.00
Published on Web 07/06/2000

library of phosphinic peptide inhibitors led to the discovery of a potent and highly selective inhibitor (RXP 407) of the N-domain of human sACE (*13*). These differences in the substrate and inhibitor specificities of the two domains of the tandem ACE protein leads to the suggestion that the duplication of the ancestral ACE gene has allowed the somatic enzyme to perform efficiently in more than one peptide signaling pathway without compromise.

Reports of the existence of ACE-like proteins in several insects (*14*−*19*), a crab (*20*), a tick (*21*−*23*), an annelid (*24*), and a mollusc (*25*) has established that ACE is an evolutionarily ancient enzyme. Of the invertebrate ACEs, the two *Drosophila melanogaster* enzymes, Ance and Acer, have been studied in most detail (*14*, *17*, *18*, *26*). Unlike mammalian sACE, they are single-domain proteins without a membrane anchor. There is around 60% similarity in the amino acid sequences of both Ance and Acer to the N- and C-domains of mammalian ACEs and therefore it is not surprising that Ance and Acer share some properties with the two domains of mammalian sACE. For example, both Ance and Acer hydrolyze bradykinin and the ACE synthetic substrate Hip-His-Leu, which mimics the C-terminal dipeptide sequence of AI, and these activities are inhibited by mammalian ACE inhibitors (*14*, *26*, *27*) (X. Houard, unpublished data). However, a recent study of the enzymic properties of Acer has revealed some important differences between the substrate specificities of these two *D. melanogaster* enzymes, with Ance showing a stronger similarity to the C-domain of human ACE (*27*). The question arises as to whether the distinct characteristics of human N- and C-domains are peculiar to higher mammals or whether these two differing activities are the result of an ancient functional divergence occurring before the separation of the ecdysozoa and deuterostome lineages,

To answer these questions, a phosphinic peptide library, previously used to probe the differences between the human N- and C-domains, was screened to compare Acer and Ance. In addition, the sequencing of the *D. melanogaster* genome, and especially of the recently sequenced *Adh* region of chromosome 2, in which *Ance* maps, has allowed us to perform a detailed analysis of the region around the *Ance* gene (*28*). This has identified a further 4 *D. melanogaster* ACE-like genes, two of which flank *Ance*. Of these six genes, only two (*Ance* and *Acer*) are predicted to be active zinc metallopeptidases, but two others (*Ance-2* and *Ance-3*) have a genomic structure reminiscent of the two domain vertebrate sACE coding region. This analysis, coupled with biochemical comparisons of the active *D. melanogaster* ACE-like gene products and the human N- and C-domains, suggests that the duplication event which gave rise to the vertebrate tandem protein predates the divergence of the ecdysozoa and deuterostomes, and that the differing activities of the N- and C-domains have been maintained over long periods of evolutionary time. Understanding these physiological roles of Ance and Acer in insects might provide valuable insights into additional functions of sACE and gACE in mammals.

MATERIALS AND METHODS

Inhibitor Studies. Phosphinic peptide inhibitors and the substrates, Mca-Ala-Ser-Asp-Lys-DpaOH (Dpa, N3 (2,4,-dintrophenyl) L-2,3-diamino propionyl) and Mca-Ser-Asp-Lys-DpaOH, were synthesized and purified as described previously (*13*). Recombinant Ance and Acer were purified from cultures of *Pichia pastoris* expressing Ance and Acer cDNA (*26*, *27*). Purity of the enzymes was confirmed by SDS−PAGE and the molar concentration of Ance and Acer was determined by titration with the ACE inhibitor fosinoprilat. The enzyme activity of Acer and Ance was assayed using Mca-Ala-Ser-Asp-Lys-DpaOH and Mca-Ser-Asp-Lys-DpaOH as substrates, respectively, and were carried out at 20 °C in 200 μL of 50 mM Hepes, pH 6.8, and 200 mM NaCl. Assays, performed in 96 well microassay plates, recorded the increase in fluorescence (390 nm) induced by the hydrolysis of the Mca-substrate, using a Fluorolite 1000 microassay fluorometer (Dynatech). Inhibitors were preincubated with the enzyme (Ance, 3 pmol, and Acer 2.6 pmol) for 40 min at 20 °C prior to the addition of substrate.

Sequence Analysis of the Adh Region of Chromosome 2. Human germinal ACE protein sequence (NCBI accession P29966, ACET_HUMAN) was used to search the Berkeley Drosophila Genome Project (BDGP) sequence database using the website Blast server (http://www.fruitfly.org). Sequences containing regions of significant homology were downloaded from the BDGP website, and iteratively used to search the *D. melanogaster* genome database to ensure all representatives of the group had been identified (sequence data set as of March 26th, 2000). Further analysis of the database on May 12th, 2000 revealed no further ACE-like or M2 peptidase sequences in the *D. melanogaster* genome.

Gene and exon predictions were made using FGENESH, at the Sanger Centre (http://genomic.sanger.ac.uk/gf/gf.shtml), using web-based submission of sequence, and accepting the default parameters. The robustness of these predictions was checked using several other prediction methods, including HMMgene v.1.1 (*29*), Genscan (*30*), and webgene (*31*), all of which agreed in terms of internal exons, though differed sometimes in their predictions of first and last exon (data not shown). Gene structure predictions were refined using the Artemis annotation program (http://www.sanger.ac.uk/software/artemis). Sequence alignments were created using ClustalX v.1.64b (*32*), bootstrap analysis used the SEQBOOT, NEIGHBOUR, and CONSENSE programs fom the PHYLIP suite (*33*) and the trees generated were viewed using TreeView (*34*). Sequence alignments were carried out using the core regions of the human N- and C-domains, and of Ance and Acer, as previously described (*2*, *14*, *18*), and trimming alignments of the newly predicted Ance-2−Ance-5 protein products to define their core comains. Potential signal peptides in the first 60 amino acid residues of the predicted protein sequences were searched for using the SignalP server (*35*). Transmembrane predictions were carried out using the TMpred membrane-span prediction programs (*36*), and identification of the predicted proteins as members of the M2 zinc metalloprotease family was carried out using the protein families and alignment database Pfam (*37*).

Intron Analysis. Intron analysis involved identifying intron/exon fusion sites in the predicted protein sequences and determining the relative positions in each predicted protein. Presence or absence of each intron in the core duplicated region was analyzed using the discrete state analysis program DOLLOP from the PHYLIP suite of programs (*33*). Bootstrap values were generated using SEQBOOT, DOLLOP, and CONSENSE. For this analysis, the first and last introns

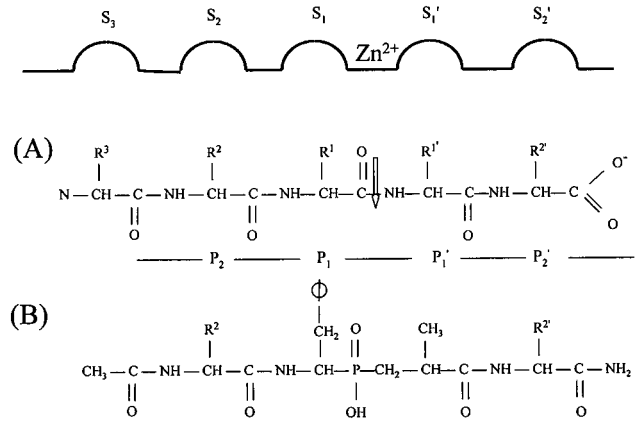

FIGURE 1: Schematic model of the ACE active site showing the relative positions of the enzyme subsites (S_3, S_2, S_1, S_1', S_2'), the residues of the phosphinic peptide inhibitors (P_3, P_2, P_1, P_1', P_2') and the catalytic zinc.

in each prediction were ignored, to avoid any complications arising from the possibility of alternative splicing giving two domain protein products.

RESULTS

Inhibition of Ance and Acer by Phosphinic Peptide Libraries. Quenched fluorogenic substrates (Mca-Ala-Ser-Asp-Lys-DpaOH and Mca-Ser-Asp-Lys-DpaOH) were used to determine the activity of Ance and Acer and the potency of the phosphinic acid peptide inhibitors. Mca-Ala-Ser-Asp-Lys-DpaOH (K_m of 15 μM) was cleaved twice as fast as Mca-Ser-Asp-Lys-DpaOH by Acer and was chosen as the substrate for routine measurement of Acer activity in this study. Mca-Ser-Asp-Lys-DpaOH (K_m of 19 μM) was the preferred substrate for Ance and was therefore used for measuring the enzyme activity in determining the potency of inhibitors.

The inhibitors were mixtures of phosphinic peptides with the general formula AcYaa-Pheψ(PO_2-CH_2)Ala-Yaa′-NH_2 , where Yaa is the P_2 amino acid interacting with the S_2 subsite and Yaa′ signifies a mixture of 20 different amino acids at P_2' (Figure 1).

Acer was generally more susceptible than Ance to inhibition by the 20 mixtures of phosphinic peptides (50 nM, final inhibitor concentration) with 13 mixtures inhibiting Acer by over 50% compared to only five inhibiting Ance by 50% (Figure 2A). AcAsp-Pheψ(PO_2-CH_2)Ala-Yaa′-NH_2 was exceptional in showing no inhibition of Ance, but significant Acer inhibition. Increasing the final concentration of Ac-Asp-Pheψ(PO_2-CH_2)Ala-Yaa′-NH_2 to 200 nM resulted in 66% inhibition of Acer, but still no significant inhibition of Ance activity, thus confirming the selectivity of this mixture of phosphinic peptides.

The selectivity of the Ac-Asp-Pheψ(PO_2-CH_2)Ala-Yaa′-NH_2 for Acer was investigated further by testing 20 phosphinic peptides, where Yaa′ is a single amino acid, as inhibitors of Acer and Ance. The inhibitors were used at two concentrations, 500 nM for Acer and 5 μM for Ance. Phosphinic peptides with Phe, Trp, Ala, and Pro at P_2' were the more potent inhibitors of Acer (Figure 2B).

The importance of N-terminal acetylation and C-terminal amidation for the selectivity Ac-Asp-Pheψ(PO_2-CH_2)Ala-Ala-NH_2 (RXP 407). Among the most potent inhibitors of Acer, reported in Figure 2B, Ac-Asp-Pheψ(PO_2-CH_2)Ala-Ala-NH_2 was selected for further structure/function analysis. A comparison of the inhibitory properties of Ac-Asp-Pheψ-(PO_2-CH_2)Ala-Ala-NH_2, containing a mixture of four diastereoisomers, and analogues revealed some of the structural features that contribute toward the observed selectivity of RXP 407 (Table 1). The substitution of the N-terminal Asp by Ala, had the greatest effect, reducing the selectivity of the inhibitor by 20-fold. The presence of an acetyl group at the N-terminus and a C-terminal amide group also made a significant contribution to the selectivity of Ac-Asp-Pheψ-(PO_2-CH_2)Ala-Ala-NH_2 toward Acer. Asp-Pheψ(PO_2-CH_2)-Ala-Ala-OH with a free terminal carboxyl group showed increased potency toward both Ance and Acer, compared with RXP 407. The pure diastereoisomer Ac-Asp-$_{(L)}$Pheψ-(PO_2-CH_2)$_{(L)}$Ala-Ala-NH_2 exhibits a K_i value of 12 nM and 2200 nM, respectively, toward Acer and Ance.

Table 1: Inhibition of Ance and Acer by RXP 407 and Related Phosphinic Peptides

inhibitor	K_i (nM) Acer[a]	Ance[a]
Ac-Asp-$_{(L)}$Pheψ(PO_2-CH_2)$_{(L)}$Ala-Ala-NH_2 (RXP 407)	12	2200
Ac-Asp-Pheψ(PO_2-CH_2)Ala-Ala-NH_2	98	9000
Ac-Asp-Pheψ(PO_2-CH_2)Ala-Ala-OH	0.3	33
Asp-Pheψ(PO_2-CH_2)Ala-Ala-NH_2	36	915
Ac-Ala-Pheι(PO_2-CH_2)Ala-Ala-NH_2	9	50

[a] Acer and Ance were assayed using Mca-Ala-Ser-Asp-Lys-DpaOH (25 μM) and Mca-Ser-Asp-Lys-DpaOH (25 μM) as substrates, respectively, as described in Materials and Methods.

Analysis of the Drosophila melanogaster Genome Sequence. Blastx analysis of the BDGP sequence database identified four cosmid sequences mapping to chromosome 2 with predicted homologies to human gACE and the C-domain of sACE (Table 2). Analysis of three of these cosmids identifies gene predictions (Acer, Ance-4 and Ance-5; Table 2), which would appear to code for simple, single domain duplications, with little difficulty in predicting ACE-like proteins. The predicted proteins of these genes include potential signal peptides, but none have C-terminal transmembrane regions, analogous to the membrane anchoring domains positioned close to the C-terminus of mammalian sACE and gACE. Unlike Acer, neither Ance-4 nor Ance-5 appear to have functional active sites − indeed all three of the critical residues required for coordinating the zinc atom have been substituted. Both Ance-4 and Ance-5 are expressed, as expressed sequence tags (ESTs) have been identified which map to these predictions (Ance-4 with BDGP EST clot no. 4148, Ance-5 with BDGP EST clot no. 11487). Both products retain the same pattern of cysteine residues found in the core regions of the human N- and C-domains, as well as in Ance and Acer.

Analysis of the Adh Region. The fourth cosmid sequence identified from the genome sequence (BG:DS00180, NCBI accession no. AC001660) includes part of the coding region for Ance and shows a much more complex set of predicted homologies. This region forms part of an extensive genomic segment which includes the *adh* locus and whose preliminary annotation was recently published (*28*). In that description, two ACE-like genes were identified: *Ance* (which starts in the next cosmid in the series, AC001664), and *BG*:

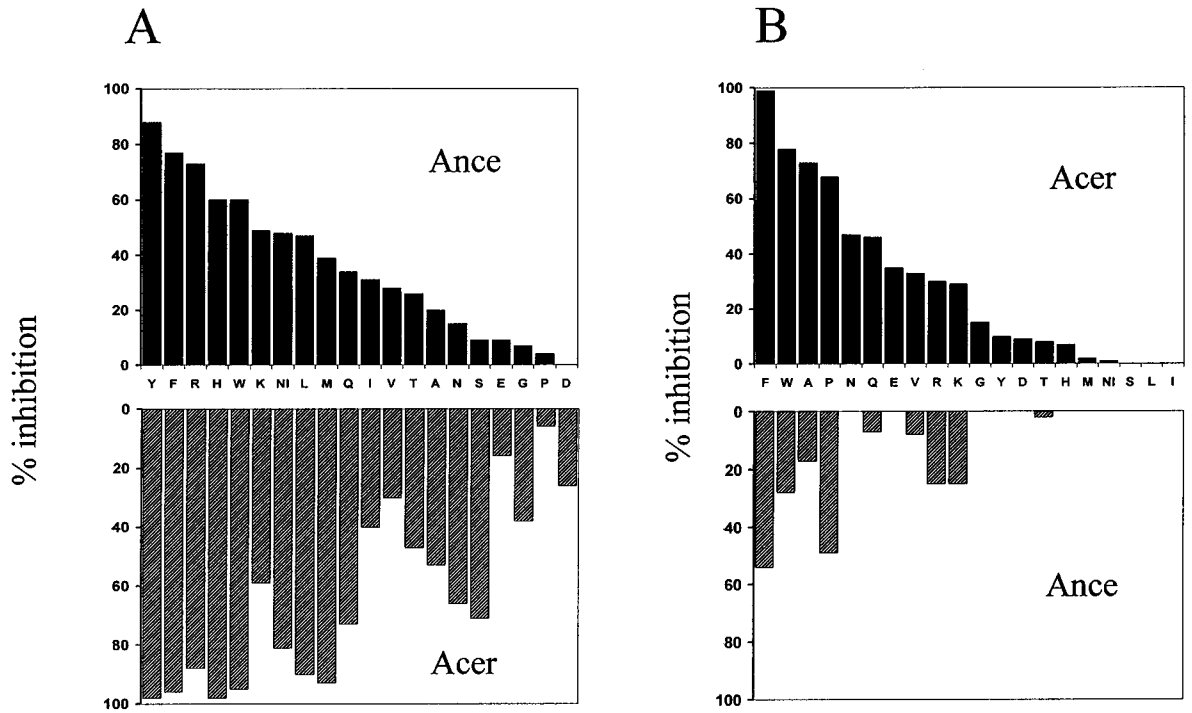

FIGURE 2: Influence of the P_2 position on the inhibition of Ance and Acer by mixtures of phosphinic peptides (50 nM) of the general formula AcYaa-Pheψ(PO_2-CH_2)Ala-Yaa′-NH_2 where Yaa′ represents a mixture of 20 amino acids and where Yaa is the individual amino acid indicated between the two histograms (A). The influence of the P_2' position on the inhibition of Ance and Acer by mixtures of phosphinic peptides of the general formula Ac-Asp-Pheψ(PO_2-CH_2)Ala-Yaa′-NH_2 where Yaa′ is the individual amino acid indicated between the two histograms (B). The final concentration of phosphinic peptides in (B) was 500 nM for Acer and 5 μM for Ance.

DS00180.5, which shows homology to the C-terminal end of vertebrate gACE, including a potential C-terminal transmembrane anchor. The original description of *BG:DS00180.5* suggested that it coded for an ACE-like protein lacking an active site (*28*).

Two further predictions can be made for ACE-like genes in the *Adh* region (Figure 3). Immediately downstream of *Ance*, there is another predicted gene (termed *Ance-2*) which would code for an ACE-like protein, with a putative signal peptide but with no C-terminal membrane anchor, very like Ance and Acer, but lacking crucial active site residues (Table 2). Next to this prediction are a set of open reading frames which, when combined with the C-terminal exons 4 and 5 of *BG:DS00180.5*, potentially codes for a further ACE-like protein, albeit with a large (36.5 kbp) intron. We have called this potential splice variant of *BG:DS00180.5*, *Ance-3*. Large introns, containing other gene coding sequences, are not unknown in *D. melanogaster*, but may explain why the *Ance-3* prediction was not identified in the original annotation, as the prediction programs have been designed to predict nonoverlapping, sequential gene sets. The predicted protein Ance-3 does have an intact active site motif (HExxH), as well as a predicted signal peptide and includes a C-terminal transmembrane anchor, the only such anchor among the six predicted proteins. However, we would predict that *Ance-3*, if expressed, would probably not be an active zinc-peptidase, as it specifically lacks the equivalent of E_{395} of Ance (equivalent to the zinc-coordinating E_{418} and E_{1016} of the human N- and C-domains, respectively), despite high conservation of the other residues in the active site region (Table 2).

Neither *Ance-2* nor *Ance-3* have ESTs associated with them. Currently, about 10 000 EST clots have been identified, corresponding to about 50% of the predicted number of genes, so it is possible that both these predicted genes are expressed, but at low levels, or in a limited number of cells. These two gene predictions are particularly interesting, in that alternative splicing of the predicted exons gives a two domain, membrane-bound ACE-like product, highly reminiscent of the vertebrate somatic ACE (Figure 3), by fusion from an internal 5′ splice site in exon 3 of *Ance-2* to exon 2 of *Ance-3*. The resultant predicted protein is 1315 residues in length (compared to 1340 residues for human somatic ACE), has a predicted signal peptide, a C-terminal transmembrane domain, and a duplication of the core domain of ACE.

This leads us to ask whether this *Ance-2::Ance-3* duplication might be the remnants of the ancestral gene structure, i.e., equivalent to the vertebrate duplication. We tested this hypothesis by comparing sequence relatedness of the six *D. melanogaster* ACE-like core domain protein sequences with the core domains of the human N- and C-domains and by assessing the pattern of intron loss and gain, using discrete state analysis to identify the most parsimonious tree giving rise to the current patterns of introns (Table 3 and Figure 4). Both sequence and intron analysis imply that, of the six genes in *D. melanogaster*, *Ance-3* is closest in similarity to the vertebrate domains, though the analysis does not distinguish between the N- and C-domains. Pairwise sequence comparisons of *Ance-3* with human sACE always give an alignment of *Ance-3* with the C-domain (data not shown).

DISCUSSION

Previous studies have shown that insect Ance is more similar enzymatically to the C-domain than the N-domain of human ACE (*26*). Ance has a high optimal chloride concentration for the hydrolysis of Hip-His-Leu and displays

Table 2: Characteristics of Known and Predicted ACE-Like Genes in *D. melanogaster*

D.melanogaster gene product	chromosome	locus	cosmid[a]	no. of residues	SP[b]	TM[c]	active site region[d]	activity[e]	EST[f]	Cys[g]
Acer	2L	29D4−29D5	DS05577	629	yes	no	HHELGHIQYY...HEAVG	yes	yes	yes
Ance	2L	34E4−34E5	DS08220/| DS00180	611	yes	no	HHELGHIQYF...HEAVG	yes	yes	yes
Ance-2	2L	34E4−34E5	DS00180	607	yes	no	FEAQSDLQYY...SDAIG	no	no	yes
Ance-3	2L	34E4−34E5	DS00180	844	yes	yes	HHEMAHIQYF...HQAVG	no	no	yes
Ance-4	2R	44F12−45A4	DS05998	622	yes	no	HGTMAELQYH...GAAIA	no	yes	yes
Ance-5	2R	60E1−60E2	DS04807	609	yes	no	HSHMARVYYA...EFAVG	no	yes	yes
		N-domain (*H. sapiens*)			yes	no	HHEMGHIQYY...HEAIG	yes		
		C-domain (*H. sapiens*)		732	yes	yes	HHEMGHIQYF...HEAIG	yes		

[a] Cosmid refers to the BDGP cosmid sequence in which the prediction lies. [b] SP, signal peptides, where not known, were predicted using the SignalP server. [c] TM refers to prediction of C-terminal membrane spanning helices by TMPred (*36*). [d] The active site region includes two zinc-coordinating histidines (underlined). [e] Yes in activity is the prediction of zinc-metallopeptidase activity based on presence or absence of the zinc-coordinating residues. [f] EST refers to the presence or absence of expressed sequence tags in the BDGP database. [g] Cys refers to the presence of cysteine residues conserved between human N-domain ACE, human C-domain ACE (germinal ACE) and Ance. The equivalent codes for these predicted genes in the *D. melanogaster* genome database (http://www.fruitfly.org) are *Ance*, CG8827(U25344); *Acer*, CG10593 (X96913); *Ance-2*, CG16869; *Ance-3*, CG17988; *Ance-4*, CG8196; *Ance-5*, CG10142.

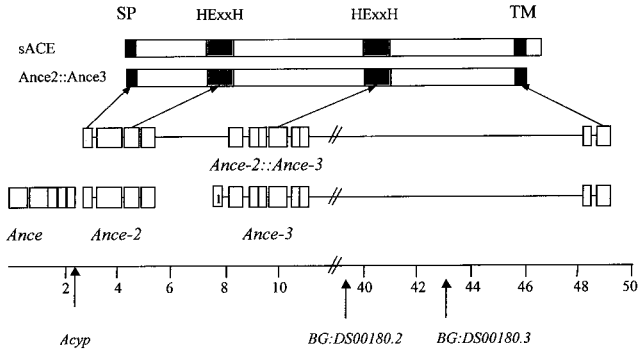

FIGURE 3: Gene and protein predictions for sequence BG:DS00180 (not to scale). The bottom scale bar is in kilobasepairs, and the arrows indicate approximate locations of three other known (*Acyp*) and predicted (*BG:DS00180.2* and *.3*) genes. The joined boxes identify the predicted exons for *Ance*, *Ance-2*, and *Ance*-3(*BG: DS00180.5*). The first exon of *Ance-3* (shaded) is equivalent to exon 13 of the human *ACE* gene, which codes for the N-terminal region (including signal peptide) of gACE, and is not used for the two domain cDNA. The upper part of the figure shows the exon structure for the putative two domain *D. melanogaster* ACE-like cDNA, with the protein product drawn diagrammatically above (Ance-2::Ance-3), showing the exons which code for the signal peptide (SP, in black) the active site regions (HExxH, dotted) and the C-terminal transmembrane anchor (TM, shaded), compared to human sACE.

Table 3: Intron Congruences for the Human N-Domain (hN) and C-Domain (hC) ACE

Gene product	Intron
hC	.XXXXXX.XX.XXX[a]
hN	.XXXXXX.XX.XXX
Acer	....X.....
Ance	....X..X....XX
Ance2	.......X....X.
Ance3	.X..X..X..XXXX
Ance4	............X.
Ance5	..........X.XX

[a] (X) The presence of an intron. (.) The absence of an intron.

a very low rate of hydrolysis of the N-domain-specific substrate, AcSDKP (*26*, *27*). These properties are very similar to those of the human C-domain which also requires a high chloride concentration for optimal hydrolysis of the synthetic peptide Hip-His-Leu, whereas the human N-domain is only weakly activated by chloride with an optimal chloride concentration of 10 mM (*3*).

Acer on the other hand has much less in common with the C-domain of human ACE. The hydrolysis of Hip-His-Leu by Acer is not influenced by chloride concentration and, although Acer can cleave Hip-His-Leu, it cannot hydrolyze AI (*27*). Acer also differs from both Ance and the human C-domain in the efficiency of hydrolysis of Hip-His-Leu. Acer cleaves Hip-His-Leu with a k_{cat}/K_m ratio of 0.015 s^{-1}, μM^{-1}, which is much lower than the k_{cat}/K_m obtained for human ACE C-domain (0.23 s^{-1}, μM^{-1}) and Ance (0.28 s^{-1}, μM^{-1}), but similar to k_{cat} /K_m for the hydrolysis of the same substrate (0.02 s^{-1}, μM^{-1}) by the human N-domain (*27*).

In the present study, both the analysis of phosphinic peptide mixtures and analogues of RXP 407 confirms the strong functional similarity between Ance and the C-domain of human ACE and establishes that Acer shares unique binding properties with the human N-domain. The human N-domain selectivity of RXP 407 is due to the presence of a carboxamide group at the C-terminus, an Asp at the P_2 position and an N-terminal acetyl group, all of which act repulsively at the C-domain active site (*13*). Similarly, the selectivity of RXP 407 for Acer can be attributed to the same three structural features, which are poorly tolerated by *D. melanogaster* Ance. These results indicate that the Acer active site, like the human N-domain, has a more relaxed specificity compared to that of Ance and the human C-domain. This proposal is also supported by the results

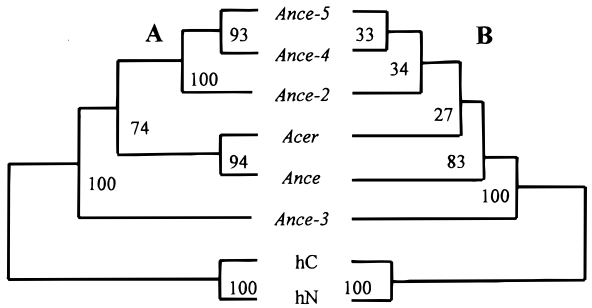

FIGURE 4: Dendrogram showing the relationships between the core protein sequence domains (A) and the intron congruence (B) between the six *D. melanogaster* ACE-like genes and the human N- (hN) and C- (hC) coding regions. The figures at the branch points are percent bootstrap values, based on 1000 iterations of the PHYLIP NEIGHBOUR (A), or DOLLOP (B) algorithms.

reported in Figure 2A, showing that the S_2' subsite of Acer A tolerates many more P_2' substitutions than the S_2' subsite of Ance.

Further valuable information on the evolutionary relationship of Ance and Acer has been gleaned from the analysis of the published *Adh* region of chromosome 2. The similarities between *Ance-3* and the vertebrate genes, both in sequence and gene structure, suggest that this is the nearest of the *D. melanogaster* genes to the ancestral form. The presence of a C-terminal transmembrane spanning region further implies that this shares a common ancestor with vertebrate gACE. The presence of a duplicated region upstream, giving a second gene which itself lacks a C-terminal anchor, but can be spliced to give a two-domain protein product, suggests that this whole region shares a common ancestor with vertebrate sACE, and that the duplication therefore predates the deuterostome/ecdysozoa divergence. Subsequently, at least in the line leading to the Diptera, a number of further duplications have occurred to give *Ance*, *Acer*, *Ance-4*, and *Ance-5*. The biochemical evidence and comparative sequence analysis would suggest that *Ance* is a duplication of the ancestral C-domain, now *Ance-3*, but without the terminal exon that codes for the C-terminal transmembrane anchor. The provenance of the other copies is more complex, and depends on whether the biochemical characteristics of the N-domain-like *Acer* gene product are derived or ancestral. If ancestral, then the prediction would be that *Ance-2* gave rise by duplication to *Acer*, and that *Ance-2* subsequently mutated to lose the critical active site residues required for zinc binding. Following this a further set of duplications, either sequential or independent, gave rise to *Ance-4* and *Ance-5*. The clustering of *Acer* with *Ance* would then be explained by selection to maintain the biochemical properties of the active enzyme. The direct sequence comparisons of the several *D. melanogaster* ACE-like gene products imply that *Ance* and *Acer* shared a common ancestor; likewise *Ance-2*, *Ance-4*, and *Ance-5* shared a common ancestor, but the data cannot tell us whether any were directly derived from the same common ancestor as either the N- or the C-domain. This may reflect functional differences, in that *Ance* and *Acer* are functional peptidases, whereas the other group are not, and so selection criteria might be different between the groups, but gene conversion and purifying selection could be occurring within the groups, hiding the true ancestral pathway. Intron analysis likewise gives us closest similarity between *Ance-3* and the vertebrate genes, and clear separation between *Ance* and the other genes, but after that bifurcation there is not enough data to build a robust tree.

The role(s) of *Ance-4* and *Ance-5*, which should lack peptidase activity, but are expressed, would necessarily involve different selection criteria and hence give rise to their significant sequence divergence from *Acer*. If, however, the N-domain characteristics are derived, then it is possible that *Acer* is a duplication of *Ance-3* or *Ance*, i.e., has the C-domain ancestor to explain the sequence similarity between *Ance* and *Acer*, but that convergent evolution has operated to select for N-domain-like characteristics. It is intriguing to note that, within the active site residues of the vertebrate ACE domains, there is a consistent difference between the N- and C-domains, a substitution of tyrosine (Y) in the C-domain by phenylalanine (F) in the N-domain (Table 2), which exists in all vertebrate N- and C-domains for which there is sequence data (*Mus musculus*, P09470; *Rattus norvegicus*, P47820; *Homo sapiens*, A31759; *Gallus gallus*, Q10751; *Bos taurus*, P12820; *Oryctolagus cuniculus*, P12822), and that this difference can also be seen between *Ance* and *Ance-3* (F, same as C-domain), and *Acer* and *Ance-2* (Y, same as N-domain). By analogy with thermolysin, whose crystal structure has been solved (*41*), the HExxH region forms a helix with the two coordinating histidines pointing into the active site. Propagation of the helix for one more turn along the strand also aligns the Y/F residue with the active site. This region would then constitute the P_2 subsite, whose composition is critical for the differential inhibition of N- and C-domains by RXP 407. Conservation of these critical alternate residues not only supports the *Ance-2*:: *Ance-3* region as sharing a common ancestor with the human sACE gene, but also that *Ance* is derived from *Ance-3* and is therefore an orthologue of gACE. One surprise is the high level of sequence conservation between these different genes, especially for the two for which there is no evidence of expression from EST data, *Ance-2* and *Ance-3*. This may reflect a low level, or highly tissue specific expression pattern, or it may be that this whole series of duplications is relatively recent. Of the two published arthropod ACE protein sequences, *Haematobia irritans exigua* [buffalo fly, NCBI accession Q10715 (*19*)], which is also a dipteran, lacks the C-terminal anchor, whereas *Boophilus microplus* [cattle tick, NCBI accession AAB32556 (*22*)] has a C-terminal hydrophobic region, which argues that the shift from a membrane bound form as the major form, to the soluble form epitomised by *Ance*, has occurred since the divergence of the Chelicerata and the Tracheata.

Taken together, the data suggest that the duplication seen in the vertebrates occurred before the divergence of the protostomes and the deuterostomes, and that the functional difference between the N- and C-domains may also be ancient. Conservation of these functional differences in both vertebrate and invertebrate species during the course of evolution suggests that these distinct active sites have different physiological roles. The availability of RXP 407, the first selective inhibitor of the mammalian N-domain and of Acer, may help to establish whether this proposal is true. Ance not only has C-domain characteristics, but is derived from the ancestral C-domain-like gene, differing mainly in the loss of the C-terminal anchor as a result of a subsequent duplication in the Diptera. This shift from membrane-bound

ACE activity in most phyla, to soluble ACE activity in the Diptera may reflect a fundamental developmental or functional difference which is as yet unclear.

REFERENCES

1. Erdos, E. G. (1990) *Hypertension 16*, 363−370.
2. Hubert, C., Houot, A. M., Corvol, P., and Soubrier, F. (1991) *J. Biol. Chem. 266*, 15377−15383.
3. Wei, L., Alhencgelas, F., Corvol, P., and Clauser, E. (1991) *J. Biol. Chem. 266*, 9002−9008.
4. Soubrier, F., Wei, L., Hubert, C., Clauser, E., Alhencgelas, F., and Corvol, P. (1993) *J. Hypertens. 11*, 599−604.
5. Ehlers, M. R. W., and Riordan, J. F. (1989) *Biochemistry* 28, 5311−5318.
6. Soubrier, F., Hubert, C., Testut, P., Nadaud, S., Alhencgelas, F., and Corvol, P. (1993) *J. Hypertens. 11*, 471−476.
7. Hagaman, J. R., Moyer, J. S., Bachman, E. S., Sibony, M., Magyar, P. L., Welch, J. E., Smithies, O., Krege, J. H., and Obrien, D. A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 2552−2557.
8. Corvol, P., Williams, T. A., and Soubrier, F. (1995) *Methods Enzymol. 248*, 283−305.
9. Azizi, M., Rousseau, A., Ezan, E., Guyene, T.-T., Michelet, S., Grognet, J.-M., Lenfant, M., Corvol, P., and Menard, J. (1996) *J. Clin. Invest. 97*, 839−844.
10. Deddish, P. A., Marcic, B., Jackman, H. L., Wang, H. Z., Skidgel, R. A., and Erdos, E. G. (1998) *Hypertension 31*, 912−917.
11. Deddish, P. A., Jackman, H. L., Wang, H. Z., Skidgel, R. A., and Erdos, E. G. (1997) *Biochem. Pharmacol. 53*, 1459−1463,
12. Wei, L., Clauser, E., Alhencgelas, F., and Corvol, P. (1992) *J. Biol. Chem. 267*, 13398−13405.
13. Dive, V., Cotton, J., Yiotakis, A., Michaud, A., Vassiliou, S., Jiracek, J., Vazeux, G., Chauvet, M. T., Cuniasse, P., and Corvol, P. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 4330−4335.
14. Cornell, M. J., Williams, T. A., Lamango, N. S., Coates, D., Corvol, P., Soubrier, F., Hoheisel, J., Lehrach, H., and Isaac, R. E. (1995) *J. Biol. Chem. 270*, 13613−13619.
15. Lamango, N. S., Sajid, M., and Isaac, R. E. (1996) *Biochem. J. 314*, 639−646.
16. Schoofs, L., Veelaert, D., DeLoof, A., Huybrechts, R., and Isaac, E. (1998) *Brain Res. 785*, 215−227.
17. Tatei, K., Cai, H., Ip, Y. T., and Levine, M. (1995) *Mech. Dev. 51*, 157−168.
18. Taylor, C. A. M., Coates, D., and Shirras, A. D. (1996) *Gene 181*, 191−197.
19. Wijffels, G., Fitzgerald, C., Gough, J., Riding, G., Elvin, C., Kemp, D., and Willadsen, P. (1996) *Eur. J. Biochem. 237*, 414−423.
20. Smiley, J. W., and Doig, M. T. (1994) *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol. 108*, 491−496.
21. Ribeiro, J. M. C., and Mather, T. N. (1998) *Exp. Parasitol. 89*, 213−221.
22. Riding, G. A., Jarmey, J., McKenna, R. V., Pearson, R., Cobon, G. S., and Willadsen, P. (1994) *J. Immunol. 153*, 5158−5166.
23. Jarmey, J. M., Riding, G. A., Pearson, R. D., McKenna, R. V., and Willadsen, P. (1995) *Insect Biochem. Mol. Biol. 25*, 969−974.
24. Laurent, V., and Salzet, M. (1995) *Neuroscience Lett. 198*, 60−64.
25. Laurent, V., Stefano, G., and Salzet, M. (1997) *Regul. Pept. 69*, 53−61.
26. Williams, T. A., Michaud, A., Houard, X., Chauvet, M. T., Soubrier, F., and Corvol, P. (1996) *Biochem. J. 318*, 125−131.
27. Houard, X., Williams, T. A., Michaud, A., Dani, P., Isaac, R. E., Shirras, A. D., Coates, D., and Corvol, P. (1998) *Eur. J. Biochem. 257*, 599−606.
28. Ashburner, M., Misra, S., Roote, J., Lewis, S. E., Blazej, R., Davis, T., Doyle, C., Galle, R., George, R., Harris, N., Hartzell, G., Harvey, D., Hong, L., Houston, K., Hoskins, R., Johnson, G., Martin, C., Moshrefi, A., Palazzolo, M., Reese, M. G., Spradling, A., Tsang, G., Wan, K., Whitelaw, K., Kimmel, B., Celniker, S., and Rubin, G. M. (1999) *Genetics 153*, 179−219.
29. Krogh, A. (1997) in *Fifth International Conference on Intelligent Systems for Molecular Biology* (al, G. T. e., Ed.) pp 179−186, AAAI Press, Menlo Park, CA.
30. Burge, C., and Karlin, S. (1997) *J. Mol. Biol. 268*, 78−94.
31. Milanesi, L., and Rogozin, L. B. (1998) in *Guide to Human Genome Computing* (Bishop, M. J., Ed.) pp 215−259, Academic Press, Cambridge.
32. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res. 25*, 4876−4882.
33. Felsenstein, J. (1989) *Cladistics 5*, 164−166.
34. Page, R. D. M. (1996) *Comput. Appl. Biosci. 12*, 357−358.
35. Nielsen, H., Engelbrecht, J., Brunak, S., and vonHeijne, G. (1997) *Protein Eng. 10*, 1−6.
36. Hofmann, K., and Stoffel, W. (1993) *Biol. Chem. Hoppe-Seyler 347*, 166.
37. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D., and Sonnhammer, E. L. L. (1999) *Nucleic Acids Res. 27*, 260−262.
38. Holmes, M. A., and Matthews, B. (1982) *J. Mol. Biol. 160*, 623−639.

BI000593Q