## Two putative active centers in human angiotensin I-converting enzyme revealed by molecular cloning

(endothelial cells/metallopeptidase/transmembrane domain/genomic DNA hybridization)

Florent Soubrier\*, François Alhenc-Gelas\*, Christine Hubert\*, Jacqueline Allegrini\*, Marie John<sup>†</sup>, Geoffrey Tregear<sup>†</sup>, and Pierre Corvol\*

\*Institut National de la Santé et de la Recherche Médicale, Unité 36, 17 rue du Fer à Moulin, 75005 Paris, France; and <sup>†</sup>Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville, Victoria 3052, Australia

Communicated by Alexander G. Bearn, August 24, 1988 (received for review June 10, 1988)

The amino-terminal amino acid sequence and ABSTRACT several internal peptide sequences of angiotensin I-converting enzyme (ACE; peptidyl-dipeptidase A, kininase II; EC 3.4.15.1) purified from human kidney were used to design oligonucleotide probes. The nucleotide sequence of ACE mRNA was determined by molecular cloning of the DNA complementary to the human vascular endothelial cell ACE mRNA. The complete amino acid sequence deduced from the cDNA contains 1306 residues, beginning with a signal peptide of 29 amino acids. A highly hydrophobic sequence located near the carboxylterminal extremity of the molecule most likely constitutes the anchor to the plasma membrane. The sequence of ACE reveals a high degree of internal homology between two large domains, suggesting that the molecule resulted from a gene duplication. Each of these two domains contains short amino acid sequences identical to those located around critical residues of the active site of other metallopeptidases (thermolysin, neutral endopeptidase, and collagenase) and therefore bears a putative active site. Since earlier experiments suggested that a single Zn atom was bound per molecule of ACE, only one of the two domains should be catalytically active. The results of genomic DNA analysis with the cDNA probe are consistent with the presence of a single gene for ACE in the haploid human genome. Whereas the ACE gene is transcribed as a 4.3-kilobase mRNA in vascular endothelial cells, a 3.0-kilobase transcript was detected in the testis, where a shorter form of ACE is synthesized.

Peptidyl-dipeptidase A (EC 3.4.15.1) plays an important role in blood pressure homeostasis by hydrolyzing angiotensin I, the inactive peptide released after cleavage of angiotensin by renin, into angiotensin II (1). Accordingly, this Zn metallopeptidase is designated angiotensin I-converting enzyme (ACE), although being the same enzyme as kininase II, it is also able to hydrolyze bradykinin and various other peptides (2, 3). This enzyme is a widely distributed peptidase, occurring, for example, as a membrane-bound ectoenzyme on the surface of vascular endothelial cells and renal epithelial cells and as a circulating enzyme in plasma (3–5). We report here the amino acid sequence of ACE as deduced from the nucleotide sequence of DNA complementary to the ACE mRNA.<sup>‡</sup>

## **MATERIALS AND METHODS**

**Purification and Sequencing of ACE and Preparation of Oligodeoxyribonucleotide Probe.** The cortex of fresh postmortem human kidneys (600 g) was homogenized (54:100, wt/vol) in 20 mM potassium phosphate buffer (pH 8) containing 250 mM sucrose and a mixture of protease inhibitors, cells debris was discarded, and the particulate fraction was sedimented by centrifugation at  $105,000 \times g$  for 1 hr. The pellet was resuspended in 200 ml of 150 mM potassium phosphate buffer (pH 8; buffer I) and treated for 18 hr with the detergent 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS, 8 mM; Serva). The supernatant obtained after centrifugation at 105,000  $\times$  g for 1 hr was dialyzed extensively against buffer I and then applied at a flow rate of 24 ml/hr to a  $3 \times 80$ -cm column of phenyl-Sepharose 4B (Pharmacia). Roughly 60% of the enzyme activity was retained on the column and was eluted as a single peak by a linear gradient of 0-10 mM CHAPS in buffer I (800 ml). ZnSO<sub>4</sub> and KCl were added to final concentrations of 0.1 mM and 330 mM, respectively, and the purification was completed by affinity chromatography on Lisinopril  $\{N^{\alpha}-[(S)-1-\text{carboxy}-$ 3-phenylpropyl]-L-lysyl-L-proline} coupled to Sepharose 4B as described (6). The protein isolated after this last step (1.2)mg) was analyzed, after reduction by 5% (vol/vol) 2-mercaptoethanol and denaturation by boiling, by NaDodSO<sub>4</sub>,/ 6% polyacrylamide gel electrophoresis (7).

Treatment with CNBr (2.0 mg together with 0.5 mg of ACE in 0.2 ml of 70% formic acid for 16 hr at room temperature) was used to cleave the enzyme at methionine residues and generate fragments for sequencing. Peptide fragments were isolated by reverse-phase HPLC on Vydac C<sub>4</sub> (average pore size, 30 nm; particle size, 5  $\mu$ m; column dimensions, 4.6 × 250 mm; The Separations Group, Hesperia, CA), using a trifluoroacetic acid/water/acetonitrile gradient system. Several peptides were separated, lyophilized, and sequenced with an Applied Biosystems (Foster City, CA) model 470A protein sequencer. The native protein was also analyzed to obtain an amino-terminal sequence.

A 64-fold-redundant mixed oligonucleotide, HACE6, whose sequence  $(3'-TACACCCG_ATT_{TAC}^TGGACCCT_TTT-A_GTA-5')$  was based (8) on the peptide sequence Met-Trp-Ala-Lys-Ser-Trp-Glu-Asn-Ile (sequence CN8b, see *Results*) was synthesized on an automated DNA synthesizer (Gene Assembler, Pharmacia) using phosphoramidite chemistry.

**Preparation and Screening of cDNA Libraries.** A human umbilical vein endothelial cell cDNA library primed with oligo(dT) and constructed in bacteriophage  $\lambda$ gt11 (Clontech, Palo Alto, CA) was screened with the <sup>32</sup>P-labeled HACE6 probe (specific activity  $\approx 5 \times 10^6$  cpm/pmol). Hybridizations with HACE6 probe were performed in 6× SSC (1× SSC is 150 mM NaCl/15 mM sodium citrate, pH 7) containing 0.1% NaDodSO<sub>4</sub>, 50 mM sodium phosphate buffer (pH 6.8), 5× Denhardt's solution (1× is 0.02% Ficoll/0.02% polyvinylpyr-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

Abbreviation: ACE, angiotensin I-converting enzyme.

<sup>&</sup>lt;sup>‡</sup>The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04144).

rolidone/0.02% bovine serum albumin), and 0.1 mg of denatured salmon sperm DNA per ml at 50°C. Filters were washed twice in  $2 \times SSC/0.1\%$  NaDodSO<sub>4</sub> at 55°C (15 min per wash). To obtain clones corresponding to the 5' end of ACE mRNA, another cDNA library was constructed using 5  $\mu$ g of poly(A)<sup>+</sup> RNA from cultured endothelial cells of human umbilical vein. Cells were grown in the presence of 20% fetal bovine serum, heparin at 100  $\mu$ g/ml, and fibroblast growth factor (gift of D. Gosporadowicz, Cancer Research Institute, University of California, San Francisco) at 2 ng/ml (9). Total RNA was extracted from the cells after three subcultures (10) and purified on (dT)<sub>7</sub>-cellulose (Pharmacia). The cDNAs were specifically primed with a 17-base oligomer (CP5-21) complementary to a sequence located near the 5' extremity of an ACE cDNA clone isolated from the first endothelial cell library (see *Results*), together with  $(dT)_{12-18}$  (Pharmacia). Subsequent steps for library construction in  $\lambda gt10$  were performed according to Koenig et al. (11). Recombinant phages  $(1.5 \times 10^6 \text{ clones})$  were screened under high-stringency conditions with a 300-base-pair (bp) human genomic DNA restriction fragment (see Results) labeled at high specific activity (12).

Nucleotide Sequencing. Phage inserts were subcloned in both orientations into the *Eco*RI site of plasmid vector pBluescript (Stratagene, San Diego, CA). Single-stranded DNA was prepared by infection of cultures with M13 helper phage KO7 and sequenced by the chain-termination method (13). Sequencing reactions were primed either with the universal primer or with 20-mer primers synthesized according to a confirmed cDNA sequence. Both strands of all regions were sequenced by using the modified T7 polymerase (Sequenase, United States Biochemicals, Cleveland).

**RNA and Genomic DNA Hybridization with ACE cDNA.** Testicular, renal, and endothelial cell  $poly(A)^+$  RNAs were isolated as described above. Northern blots were prepared with glyoxalated RNA according to the procedure of Thomas (14). Human high molecular weight DNA from placental nuclei (14 µg) was digested with *Hind*III, electrophoresed in a 0.7% agarose gel, and analyzed by Southern blot hybridization under high-stringency conditions (final washing conditions: 0.1× SSC/0.1% NaDodSO<sub>4</sub>, 45 min at 65°C) (15).

## RESULTS

The purified ACE migrated as a single band at 170 kDa on polyacrylamide gel electrophoresis (data not shown). Determination of the enzymatic activity with furylacryloyl-L-phenylalanylglycylglycine as substrate (16) showed apparent kinetic constants equal to 136  $\mu$ M for  $K_m$  and 22,100 min<sup>-1</sup> for  $k_{cat}$ , similar to those measured for the rabbit pulmonary enzyme (16). Amino-terminal sequence analysis revealed a high degree of similarity with amino-terminal sequences of other mammalian ACEs (Fig. 1). Thirteen internal peptide sequences were obtained (Table 1). Sequence CN8b, completed by the putative preceding methionine, was selected to design oligonucleotide probe HACE6.

Two clones,  $\lambda$ HEC1922 [3.3 kilobases (kb)] and  $\lambda$ HEC2111 (2.8 kb), were obtained by screening 1.5 × 10<sup>6</sup> phages of the

HUMAN		L	D	Р	G	L	Q	Р	G	(D)	F	s	А	D	Е	A	G
RABBIT	T,	L	D	Ρ	G	L	L	Ρ	G	D	F	А	Α	D	N	A	G
CALF	Е	L	D	Ρ	А	L	Q	Ρ	G	N	F	Ρ	А	D	Е	A	G
PIG		L	D	s	Α	L	Q	Ρ	G	С	F	т	А	D	Е	A	G
MOUSE		L	D	Ρ	G	L	Q	P	G	N	F	S	Ρ	D	Е	А	G

FIG. 1. Amino-terminal amino acid sequence of human kidney ACE. Alignment with amino-terminal sequences of rabbit lung (17), calf lung (18), pig kidney (19), and mouse kidney (20) ACEs. One-letter amino acid notation is used. The aspartic acid shown in parentheses probably resulted from deamidation of the asparagine predicted by the nucleotide sequence (see Fig. 2). Boxes indicate amino acids conserved in all species.

Table 1.	Amino acid	sequence	data	for CNB	r fragments	of
human kio	iney ACE					

•		
HPLC elution time, min	Sequence*	Location in predicted protein <sup>†</sup>
5.6	LKQGWTPRR	882-890
9.7	GHIQYF	962967
10.9	PPEFWEGS	310-317
12.1	ALEKIAFLPFGYLVDQ-R-GVF	429-450
12.3	VVPFPDKPNLDVTST-LQ	268-285
12.8	ETTYSVATV	719–727
13.0	KLGFSRPWPEA	1143–1153
14.0	LE(Q)PTDGREVV-HA	917–930
14.0	WA(K)SWENIYD	257-266
19.3	ATSRKYEDLL	746–755
20.6	VGLDALDAQ	560-568
22.2	LLFAWEGWHNAAGIPLKPL	156–174
22.2	YETPSL-QDLERL	800-812
	HPLC elution time, min 5.6 9.7 10.9 12.1 12.3 12.8 13.0 14.0 14.0 19.3 20.6 22.2 22.2	HPLC elution time, minSequence*5.6LKQGWTPRR 9.79.7GHIQYF10.9PPEFWEGS12.1ALEKIAFLPFGYLVDQ-R-GVF12.3VVPFPDKPNLDVTST-LQ12.8ETTYSVATV13.0KLGFSRPWPEA14.0LE(Q)PTDGREVV-HA14.0WA(K)SWENIYD19.3ATSRKYEDLL20.6VGLDALDAQ22.2LLFAWEGWHNAAGIPLKPL22.2YETPSL-QDLERL

\*The two residues assigned from the protein sequence data that differ from those predicted from the nucleotide sequence are shown in parentheses (see text). Dashes indicate unidentified residues. \*Location in the sequence predicted from the cDNA (Fig. 2).

oligo(dT)-primed library. These clones contained a nucleotide sequence identical to the probe HACE6 except for one base, which changed a lysine to a glutamine residue. All the other available peptide sequences of ACE, except the aminoterminal sequence, were present in the protein sequence deduced from the clones (Table 1). A single inversion in assignment of amino acids during sequencing of two peptides that were coeluted in the same HPLC peak (sequences CN8a and CN8b, Table 1) explains the apparent discrepancies in position 259 corresponding to HACE6 and in position 919. The two clones overlapped by 2323 bp and spanned 3840 nucleotides of the ACE mRNA. The cDNA corresponding to the 5' end of the mRNA was obtained from another endothelial cell cDNA library (see Materials and Methods) primed with the oligomer CP5-21, which is complementary to nucleotides 234-250 of ACE cDNA (see Fig. 2) located near the 5' end of clone  $\lambda$ HEC2111, and with oligo(dT). This library was screened with a 300-bp DNA restriction fragment, isolated in this laboratory from a human genomic library and selected because it hybridized both with CP5-21 oligonucleotide and with a synthetic 44-base oligomer designed after the aminoterminal sequence of human ACE. From several independent positive clones obtained, clone  $\lambda$ CHDT32, 250 bp long, was found to overlap by 60 bp with clone  $\lambda$ HEC2111 and to contain the 5' end of the coding sequence of ACE mRNA together with a short nontranslated region.

The nucleotide sequence of ACE cDNA obtained by sequencing the three overlapping clones (Fig. 2) comprises 4024 nucleotides. The 3' end of the sequence does not extend to the polyadenylylation signal. The open reading frame from the first ATG codon to the stop codon TGA encodes 1306 amino acids. The amino-terminal amino acid, leucine (see Fig. 1), is located after a signal peptide of 29 residues. The calculated molecular mass for the mature enzyme is 146.6 kDa.

Clone  $\lambda$ HEC1922 was used as a probe to study ACE gene expression by Northern blot analysis. Cultured endothelial cells from umbilical veins contained a mRNA with an estimated size of 4.3 kb that hybridized with the probe (Fig. 3, lane c). In the testis, a mRNA of 3.0 kb was detected (lane b).

Southern blot analysis of human DNA under stringent conditions was performed with a 300-bp EcoRI(end)-Bgl II restriction fragment from the 5' end of clone  $\lambda$ HEC2111. This probe hybridized with a single, 9.5-kb restriction fragment

GCCGAGCACCGCGCACCGCGTC 22

						CURGERCEGEGERCEGEGIE	~~~
ATGGGGGCCGCCTCGGGCCGCGGGGCCGGGGCCGC MetGlyAlaAlaSerGlyArgArgGlyProGlyLeuL	TGCTGCCGCTGCCGCTGC euLeuProLeuProLeuL	TGTTGCTGCTGCCGCC euLeuLeuLeuProPi	CGCAGCCCGCCCTG roGlnProAlaLeu	GCGTTGGACCCCGGG AlaLeuAspProGly	CTGCAGCCCGGCAAC LeuGlnProGlyAsn	TTTTCTGCTGACGAGGCCGGG PheSerAlaAspGluAlaGly	157
-29 -20 GCGCAGCTCTTCGCGCAGAGCTACAACTCCAGCGCCG	AACAGGTGCTGTTCCAGA	-10 SCGTGGCCGCCAGCT(	GGCGCACGACACC	AACATCACCGCGGAG	AATGCAAGGCGCCAG	GAGGAAGCAGCCCTGCTCAGC	292
AlaGlnLeuPheAlaGlnSerTyr <u>AsnSerSer</u> AlaG 20	luGlnValLeuPheGlnSe 30	erValAlaAlaSerTı 4(	rpAlaHisAspThr )	AsnIleThrAlaGlu	AsnAlaArgArgGin 50	GluGluAlaAlaLeuLeuSer 60	
CAGGAGTTTGCGGAGGCCTGGGGCCAGAAGGCCAAGG GlnGluPheAlaGluAlaTrpGlyGlnLysAlaLysG 70	AGCTGTATGAACCGATCT luLeuTyrGluProIleT 80	GGCAGAACTTCACGG/ rpGln <u>AsnPheThr</u> A: 0	ACCCGCAGCTGCGC spProGlnLeuArg	AGGATCATCGGAGCT ArgIleIleGlyAla 90	GTGCGAACCCTGGGC ValArgThrLeuGly	TCTGCCAACCTGCCCCTGGCT SerAlaAsnLeuProLeuAla 100	427
AAGCGGCAGCAGTACAACGCCCTGCTAAGCAACATGA LysArgGlnGlnTyrAsnAlaLeuLeuSer <u>AsnMetS</u> 110	GCAGGATCTACTCCACCG erArglleTyrSerThrA 120	CCAAGGTCTGCCTCCC laLysValCysLeuPi	CCAACAAGACTGCC ro <u>AsnLysThr</u> Ala	ACCTGCTGGTCCCTG ThrCysTrpSerLeu	GACCCAGATCTCACC AspProAspLeuThr 140	AACATCCTGGCTTCCTCGCGA AsnIleLeuAlaSerSerArg 150	562
AGCTACGCCATGCTCCTGTTTGCCTGGGAGGGCTGGC SerTyrAlaMetLeuLeuPheAlaTrpGluGlyTrpH	ACAACGCTGCGGGGCATCCG isAsnAlaAlaGlyIlePi	CGCTGAAACCGCTGT roLeuLysProLeuTy	ACGAGGATTTCACT yrGluAspPheThr	GCCCTCAGCAATGAA AlaLeuSerAsnGlu	GCCTACAAGCAGGAC AlaTyrLysGlnAsp	GGCTTCACAGACACGGGGGGCC GlyPheThrAspThrGlyAla	697
TACTGGCGCTCCTGGTACAACTCCCCCACCTTCGAGG TyrTrpArgSerTrpTyrAsnSerProThrPheGluA	ACGATCTGGAACACCTCT spAspLeuGluHisLeuT	ACCAACAGCTAGAGCO yrGlnGlnLeuGluPi	CCCTCTACCTGAAC roLeuTyrLeuAsn	CTCCATGCCTTCGTC LeuHisAlaPheVal	CGCCGCGCGCACTGCAT ArgArgAlaLeuHis	CGCCGATACGGAGACAGATAC ArgArgTyrGlyAspArgTyr	832
ATCAACCTCAGGGGACCCATCCCTGCTCATCTGCTGG	GAGACATGTGGGCCCAGA	GCTGGGAAAACATCT	20 ACGACATGGTGGTG	CCTTTCCCAGACAAG	CCCAACCTCGATGTC	ACCAGTACTATGCTGCAGCAG	967
11eAsnLeuArgG1yPro11eProAlaHisLeuLeuG 250	lyAspMetTrpAlaGInSo 20	erTrpGluAsnlleTy 60	yrAspMetValVal	270	ProAshLeuAspval	280	
GGCTGGAACGCCACGCACATGTTCCGGGTGGCAGAGG GlyTrp <u>AsnAlaThr</u> HisMetPheArgValAlaGluG 290	AGTTCTTCACCTCCCTGG luPhePheThrSerLeuG 300	AGCTCTCCCCCATGCC luLeuSerProMetPr 31	CTCCCGAGTTCTGG roProGluPheTrp 10	GAAGGGTCGATGCTG GluGlySerMetLeu	GAGAAGCCGGCCGAC GluLysProAlaAsp 320	GGGCGGGAAGTGGTGTGCCAC GlyArgGluValValCysHis 330	1102
GCCTCGGCTTGGGACTTCTACAACAGGAAAGACTTCA AlaSerAlaTrpAspPheTyrAsnArgLysAspPheA 340	GGATCAAGCAGTGCACAC rgIleLysGlnCysThrA 3	GGGTCACGATGGACC <i>i</i> rgValThrMetAspG 50	AGCTCTCCACAGTG lnLeuSerThrVal	CACCATGAGATGGGC HisHisGluMetGly 360	CATATACAGTACTAC HislleGlnTyrTyr	CTGCAGTACAAGGATCTGCCC LeuGlnTyrLysAspLeuPro 370	: 1237
GTCTCCCTGCGTCGGGGGGGCCAACCCCGGCTTCCATG ValSerLeuArgArgGlyAlaAsnProGlyPheHisG 380	AGGCCATTGGGGACGTGC luAlalleGlyAspValL	TGGCGCTCTCGGTCTC euAlaLeuSerValS	CCACTCCTGAACAT erThrProGluHis	CTGCACAAAATCGGC LeuHisLysIleGly	CTGCTGGACCGTGTC LeuLeuAspArgVal	ACCAATGACACGGAAAGTGAC Thr <u>AsnAspThr</u> GluSerAsp 420	:1372
ATCAATTACTTGCTAAAAATGGCACTGGAAAAAATTG IleAsnTyrLeuLeuLysMetAlaLeuGluLysIleA 430	CCTTCCTGCCCTTTGGCT LaPheLeuProPheGlyT	ACTTGGTGGACCAGT yrLeuValAspGlnT: 40	GGCGCTGGGGGGGTC rpArgTrpGlyVal	TTTAGTGGGCGTACC PheSerGlyArgThr 450	CCCCCCTTCCCGCTAC ProProSerArgTyr	AACTTCGACTGGTGGTATCTT AsnPheAspTrpTrpTyrLeu 460	1507
CGAACCAAGTATCAGGGGATCTGTCCTCCTGTTACCC ArgThrLysTyrGlnGlyIleCysProProValThrA 470	GAAACGAAACCCACTTTG rg <u>AsnGluThr</u> HisPheA 480	ATGCTGGAGCTAAGT spAlaGlyAlaLysPl	TTCATGTTCCAAAT heHisValPro <u>Asn</u> 90	GTGACACCATACATC ValThrProTyrIle	AGGTACTTTGTGAGT ArgTyrPheValSer 500	TTTGTCCTGCAGTTCCAGTTC PheValLeuGlnPheGlnPhe 510	1642
CATGAAGCCCTGTGCAAGGAGGCAGGCTATGAGGGCC HisGluAlaLeuCysLysGluAlaGlyTyrGluGlyP 520	CACTGCACCAGTGTGACA roLeuHisGlnCysAspI	TCTACCGGTCCACCA LeTyrArgSerThrL	AGGCAGGGGGCCAAG ysAlaGlyAlaLys	CTCCGGAAGGTGCTG LeuArgLysValLeu 540	CAGGCTGGCTCCTCC GlnAlaGlySerSer	AGGCCCTGGCAGGAGGTGCTG ArgProTrpGlnGluValLeu 550	1777 1
AAGGACATGGTCGGCTTAGATGCCCTGGATGCCCAGC LysAspMetValGlyLeuAspAlaLeuAspAlaGlnP	CGCTGCTCAAGTACTTCC roLeuLeuLysTyrPheG	AGCCAGTCACCCAGT InProValThrGlnT	GGCTGCAGGAGCAG rpLeuGlnGluGln	AACCAGCAGAACGGC	GAGGTCCTGGGCTGG GluValLeuGlyTrp	CCCGAGTACCAGTGGCACCCG ProGluTyrGlnTrpHisPro	;1912 ,
CCGTTGCCTGACAACTACCCGGAGGGCATAGACCTGG ProLeuProAspAsnTyrProGluGlyIleAspLeuV	TGACTGATGAGGCTGAGG alThrAspGluAlaGluA	CCAGCAAGTTTGTGG laSerLysPheValG	aggaatatgaccgg luGluTyrAspArg	ACATCCCAGGTGGTG ThrSerGlnValVal	TGGAACGAGTATGCC TrpAsnGluTyrAla	GAGGCCAACTGGAACTACAAC GluAlaAsnTrpAsnTyrAsn	: 2047
b10 ACCAACATCACCACAGAGACCAGCAAGATTCTGCTGC Thr <u>AsnIleThr</u> ThrGluThrSerLysIleLeuLeuG	6 AGAAGAACATGCAAATAG lnLysAsnMetGlnIleA	20 CCAACCACACCCTGA/ la <u>AsnHisThr</u> Leuly	AGTACGGCACCCAG ysTyrThrGlnAla	630 GCCAGGAAGTTTGAI ArgLysPheAspVal	GTGAACCAGTTGCAG AsnGlnLeuGln <u>Asn</u>	AACACCACTATCAAGCGGATC ThrThrIleLysArgGlyIle	2182
ATAAAGAAGGTTCAGGACCTAGAACGGGCAGCGCTGC IleLysLysValGInAspLeuGluArgAlaAlaLeuP	CTGCCCAGGAGCTGGAGG roAlaGlnGluLeuGluG	AGTACAACAAGATCC LuTyrAsnLysIleL	rGTTGGATATGGAA euLeuAspMetGlu	ACCACCTACAGCGTG ThrThrTyrSerVal	GCCACTGTGTGCCAC AlaThrValCysHis	CCGAATGGCAGCTGCCTGCAG Pro <u>AsnGlySer</u> CysLeuGln	2317
CTCGAGCCAGATCTGACGAATGTGATGGCCACATCCC LeuGluProAspLeuThrAsnValMetAlaThrSerA	GGAAATATGAAGACCTGT rgLysTyrGluAspLeuL	TATGGGCATGGGAGG euTrpAlaTrpGluG	GCTGGCGAGACAAG lyTrpArgAspLys	GCGGGGGAGAGCCATC AlaGlyArgAlaIle	CTCCAGTTTTACCCG	AAATACGTGGAACTCATCAAC LysTyrValGluLeuIleAsn	2452
CAGGCTGCCCGGCTCAATGGCTATGTAGATGCAGGGG GlnAlaAlaArgLeuAsnGlyTyrValAspAlaGlyA	750 ACTCGTGGAGGTCTATGT spSerTrpArgSerMetT	ACGAGACACCATCCC	50 IGGAGCAAGACCTG euGluGlnAspLeu	GAGCGGCTCTTCCAG	GAGCTGCAGCCACTC GluLeuGlnProLeu	TACCTCAACCTGCATGCCTAC TyrLeuAsnLeuHisAlaTyr	2587
790 GTGCGCCGGGCCCTGCACCGTCACTACGGGGCCCAGC ValArgArgAlaLeuHisArgHisTyrGlyAlaGlnH	86 ACATCAACCTGGAGGGGCG isIleAsnLeuGluGlyPi	00 CCATTCCTGCTCACC rolleProAlaHisLe	IGCTGGGGAACATG euLeuGlyAsnMet	810 TGGGCGCAGACCTGG TrpAlaGlnThrTrp	STCCAACATCTATGAC SerAsnIleTyrAsp	820 TTGGTGGTGCCCTTCCCTTCA LeuValValProPheProSer	2722
830 GCCCCCTCGATGGACACCACAGAGGCTATGCTAAAGC AlaProSerMetAspThrThrGluAlaMetLeuLysG	840 AGGGCTGGACGCCCAGGA lnGlyTrpThrProArgA	85 GGATGTTTAAGGAGG rgMetPheLysGluA	50 CTGATGATTTCTTC laAspAspPhePhe	ACCTCCCTGGGGCTG ThrSerLeuGlyLeu	860 CTGCCCGTGCCTCCT LeuProValProPro	870 GAGTTCTGGAACAAGTCGATG GluPheTrp <u>AsnLysSer</u> Met	2857
880 CTGGAGAAGCCAACCGACGGGGGGGGGGGGGGGGGGCCGGC	8 ACGCCTCGGCCTGGGACT	90 TCTACAACGGCAAGG	ACTTCCGGATCAAG	900 CAGTGCACCACCGTG	AACTTGGAGGACCTG	910 GTGGTGGCCCACCACGAAATG	2992
LeuGluLysProThrAspGlyArgGluValValCysH 920	isAlaSerAlaTrpAspPl 930	heTyrAsnGlyLysA: 94	spPheArgIleLys 40	GlnCysThrThrVal	AsnLeuGluAspLeu 950	ValValAlaHisHisGluMet 960	
GGCCACATCCAGTATTTCATGCAGTACAAAGACTTAC GlyHisIleGlnTyrPheMetGlnTyrLysAspLeuP 970	CTGTGGCCTTGAGGGAGG roValAlaLeuArgGluG 9	GTGCCAACCCCGGCT lyAlaAsnProGlyPl 80	TCCATGAGGCCATT heHisGluAlaIle	GGGGACGTGCTAGCC GlyAspValLeuAla 990	CTCTCAGTGTCTACG LeuSerValSerThr	CCCAAGCACCTGCACAGTCTC ProLysHisLeuHisSerLeu 1000	3127
AACCTGCTGAGCAGCGAGGAGGAGGAGGAGGACGAGGATG AsnLeuLeuSerSerGluGlyGlySerAspGluHisA 1010	ACATCAACTTTCTGATGA splleAsnPheLeuMetL 1020	AGATGGCCCTTGACA ysMetAlaLeuAspLy 1	AGATCGCCTTTATC ysIleAlaPheIle 030	CCCTTCAGCTACCTC ProPheSerTyrLeu	CGTCGATCAGTGGCGC WalAspGlnTrpArg 1040	TGGAGGGTATTTGATGGAAGC TrpArgValPheAspGlySer 1050	3262
ATCACCAAGGAGAACTATAACCAGGAGTGGTGGAGCC IleThrLysGluAsnTyrAsnGlnGluTrpTrpSerL 1060	TCAGGCTGAAGTACCAGG euArgLeuLysTyrGlnG 1	GCCTCTGCCCCCAG lyLeuCysProProV 070	TGCCCAGGACTCAA alProArgThrGln	GGTGACTTTGACCCA GlyAspPheAspPro 1080	AGGGGGCCAAGTTCCAC DGlyAlaLysPheHis	ATTCCTTCTAGCGTGCCTTAC IleProSerSerValProTyr 1090	3397
ATCAGGTACTTTGTCAGCTTCATCATCCAGTTCCAGT IleArgTyrPheValSerPheIleIleGlnPheGlnP 1100	TCCACGAGGCACTGTGCC heHisGluAlaLeuCysG 1110	AGGCAGCTGGCCACA lnAlaAlaGlyHisT 1	CGGGCCCCCTGCAC hrGlyProleuHis 120	AAGTGTGACATCTAC LysCysAsplleTyr	CCAGTCCAAGGAGGCC GlnSerLysGluAla 1130	GGGCAGCGCCTGGCGACCGCC GlyGlnArgLeuAlaThrAla 1140	3532
ATGAAGCTGGGCTTCAGTAGGCCGTGGCCGGAAGCCA MetLysLeuGlyPheSerArgProTrpProGluAlaM	TGCAGCTGATCACGGGCC etGlnLeuIleThrGlyG	AGCCCAACATGAGCG 1nPro <u>AsnMetSer</u> A 160	CCTCGGCCATGTTG laSerAlaMetLeu	AGCTACTTCAAGCCO SerTyrPheLysPro	GCTGCTGGACTGGCTC DLeuLeuAspTrpLeu	CGGCACGGAGAACGAGCTGCAT ArgThrGluAsnGluLeuHis 1180	3667 5
GGGGAGAAGCTGGGCTGGCCGCAGTACAACTGGACGC GlyGluLysLeuGlyTrpproGlnTyr <u>AsnTrpThr</u> P 1190	CGAACTCCGCTCGCTCAG roAsnSerAlaArgSerG 1200	AAGGGCCCCTCCCAG luGlyProLeuProA 1	ACAGCGGCCGCGTC spSerGlyArgVal 210	AGCTTCCTGGGCCTC SerPheLeuGlyLeu	GGACCTGGATGCGCAC AspLeuAspAlaGlr 1220	CAGGCCCGCGTGGGCCAGTGG GlnAlaArgValGlyGlnTrp 1230	3802
CTGCTGCTCTTCCTGGGCATCGCCCTGCTGGTAGCCA LeuLeuPheLeuGlyIleAlaLeuLeuValAlaT 1240	<u>ccctgggcctcagccagc</u> hrLeuGlyLeuSerGlnA 1	GGCTCTTCAGCATCC rgLeuPheSerIleA 250	GCCACCGCAGCCTC rgHisArgSerLev	CACCGGCACTCCCAC hisArghisSerHis 1260	CGGGCCCCAGTTCGGC SGlyProGlnPheGly	TCCGAGGTGGAGCTGAGACAC SerGluValGluLeuArgHis 1270	3937 8
TCCTGAGGTGACCCGGCTGGGTCGGCCCTGCCCAAGG	GCCTCCCACCAGAGACTG	GGATGGGAACACTGG	TGGGCAGCTGAGGG	CGGT 4024			

Ser \* 1277

FIG. 2. Nucleotide sequence and deduced amino acid sequence of human endothelial cell ACE cDNA. The nucleotide sequence is numbered on the right, starting from the first nucleotide of clone  $\lambda$ CHDT32. The amino acid sequence, presented below the nucleotide sequence, is numbered from the leucine found at the amino terminus of the mature protein, which is preceded by a 29-residue signal peptide. The termination codon TGA is denoted by an asterisk. The peptide sequence used for construction of probe HACE6 is overlined (peptide sequence data had indicated lysine rather than the glutamine predicted here at position 259; see text). Potential sites of asparagine-linked glycosylation are underlined (21). The putative transmembrane segment near the carboxyl terminus is boxed.

(Fig. 3, lane a). Similar results were obtained when a cDNA fragment located close to the 3' end was used (data not shown). These results are consistent with the presence of one gene for ACE per haploid human genome.

## DISCUSSION

A striking feature of the cDNA sequence is the presence of a high degree of internal homology (Fig. 4 Top). Two



FIG. 3. (Left) Genomic Southern blot analysis (lane a). Human placental DNA (14  $\mu$ g), digested with *Hind*III, was hybridized with a 300-bp *Eco*RI(end)-*Bgl* II fragment from the 5' end of clone  $\lambda$ HEC2111. (*Right*) Northern blot hybridization. Human testis poly(A)<sup>+</sup> RNA (20  $\mu$ g; lane b) and cultured human endothelial cell poly(A)<sup>+</sup> RNA (15  $\mu$ g; lane c) were hybridized with clone  $\lambda$ HEC1922.

homologous domains surrounded by nonhomologous regions can be delimited: a continuous nucleotide sequence identity >60% is observed between positions 700–1770 and positions 2495–3565 (overall identity, 70.1%). At the amino acid level the overall similarity calculated for the corresponding 357 residues is 67.7%. The identity is maximal in the central part of these two domains; for example, an identity of 89% is observed for amino acids 361-404 and 959-1002, segments comprising essential residues of the putative active sites (see below). The number and positions of the cysteine residues are the same in the two domains (Fig. 4 *Middle*). There are 17 potential asparagine-linked glycosylation sites (21) in the molecule, mostly grouped in the amino-terminal region of the protein and in the region located at the junction between the two homologous domains (Figs. 2 and 4).

Two particular regions of the protein display a high degree of hydrophobicity, the signal peptide and the region located just before the carboxyl terminus (Fig. 4 *Bottom*). A search for membrane-associated helices (24) indicates that these two regions have the highest probability of being involved in membrane insertion. Since amino-terminal sequence analysis of the protein indicates that the signal peptide is cleaved off during processing, the mature enzyme is most probably anchored to the cell membrane by its carboxyl-terminal hydrophobic segment, as suggested by Hooper *et al.* (19).

A computer-assisted search revealed no clear significant homology of ACE with any other protein found in the



FIG. 4. (Top) Dot matrix comparison (22) of the amino acid sequence of ACE with itself (horizontal and vertical axes). Diagonal bars are generated by segments of 60 amino acids that show  $\geq 60\%$  sequence identity between the horizontal and vertical axes. The long diagonal line represents the line of identity resulting from the comparison of the enzyme sequence with itself. The two offset diagonals denote the internal homology present within the enzyme sequence. (Middle) Schematic diagram showing the cysteine positions, the potential asparagine-linked glycosylation sites, and the positions of the putative residues of the active site (one-letter code). (Bottom) Hydropathy analysis done according to the method of Kyte and Doolittle (23) with a window size of 10 residues. Positive values indicate increasing hydrophobicity.

National Biomedical Research Foundation data bank (release 15.0). Alignment with the sequence of another Zn protease, thermolysin (25), revealed short segmental identities only in regions comprising residues involved in catalytic activity. The x-ray structure analysis of thermolysin (26) showed that the His-142, His-146, and Glu-166 coordinate the Zn atom. Corresponding residues (His-361, His-365, and Glu-389; His-959, His-963, and Glu-987) are found within short conserved regions in the two domains of ACE (Fig. 5). Another glutamic acid (Glu-143) of thermolysin, which acts as a general base in catalysis, is again found in the two domains of ACE (Glu-362 and Glu-960). Another possible homology is observed around His-231 of thermolysin, which stabilizes the transition state by formation of a hydrogen bond to the water molecule with the oxygen of the scissile bond to be cleaved. This histidine residue is conserved in the two domains of ACE (Fig. 5). Similar short sequences are also observed in the corresponding regions of rat or rabbit neutral endopeptidase (27, 28) (Fig. 5), a Zn metalloendopeptidase able to hydrolyze enkephalins, kinins, and several neuropeptides, like ACE, but with different efficiency and sensitivity to inhibitors (3, 30). Segmental homology was also found with the sequence reported (29) for the Zn protease collagenase from human skin fibroblasts (Fig. 5). The structural similarity between the active site of true endopeptidases (thermolysin, neutral endopeptidase, and collagenase) and ACE most likely explains why this enzyme can act both as a peptidyl-dipeptidase and as an endopeptidase (3).

In ACE the regions homologous to active-site sequences of thermolysin and neutral endopeptidase are present twice in the molecule, included in regions highly similar in the two domains. Whether both of these putative active sites of ACE are involved in catalysis cannot be determined from sequence analysis. However, it has been reported (31, 32) that each ACE molecule contains one Zn atom, suggesting that only one domain is able to coordinate the metal. Likewise, binding studies with competitive ACE inhibitors suggest a single class of sites (33, 34). These observations may indicate that despite its repetitive structure, ACE has only one functional active site per molecule.

The sequence of ACE cDNA indicates that the ACE gene most likely resulted from a gene duplication. Since other known mammalian ACEs are of a similar size, it is likely that they also resulted from a gene duplication that occurred before mammalian radiation. Our Southern hybridization results are consistent with the presence of one ACE gene per human haploid genome, encoding the different forms of ACE. The plasma enzyme probably originates from the vascular endothelium and is enzymatically and immunologically identical to the solubilized membrane-bound enzyme (35, 36). The sequence described here is that of a protein possessing a hydrophobic anchor in its carboxyl-terminal region. The

hACE hACE THERM rNEP hCOLL	T V V V V	V A V I I	H H A G G	H H H H H	ल ल ल ल ल	M L L L	G G T T G	H H H H	365 963 146 581/588 222	E A I E A I E A I	391 989 168
hACE hACE THERM rNEP hCOLL	H H H H H	L L I L	H H N R	K S - Y	- GGR	I L I I I	– I E	G N N N N	409 1007 238 635/643 120		

FIG. 5. Homology among amino acid sequences (one-letter notation) of human ACE (hACE), thermolysin from Bacillus thermoproteolyticus (THERM) (25), rat or rabbit neutral endopeptidase (rNEP) (27, 28), and human skin fibroblast collagenase (hCOLL) (29). The first and second lines refer to the first and second domains of ACE. Numbers refer to the last amino acid position in each protein segment. Identical residues are boxed; conservative amino acid changes are indicated by broken lines. Gaps are indicated by dashes.

plasma enzyme may derive from this form by a posttranslational modification or result from transcription of a mRNA different at its 3' end, although we have not been able. by Northern blot analysis, to detect a second mRNA species in human endothelial cells. It may also be that the enzyme is released by leakage from the plasma membrane. The observation of a 3-kb mRNA in the testis is in agreement with the observation that testicular ACE is synthesized as a shorter polypeptide chain than the vascular endothelial enzyme (37). The results of the present study suggest that the testicular ACE mRNA and the endothelial ACE mRNA (4.3 kb) result from differential splicing of the gene transcript.

We thank T. Sarda for help in DNA sequencing, F. Pinet and A. Michaud for cell cultures, I. Sauvaget (Unité d'Informatique Scientifique, Institut Pasteur) for assistance in the computer search, A. M. Houot for technical assistance, and E. Clauser for useful suggestions. We thank B. Roques and F. Rougeon for helpful discussions during preparation of the manuscript.

- Skeggs, L. T., Jr., Khan, J. R. & Shumway, N. P. (1956) J. Exp. Med. 1. 103, 295-299
- Yang, H. Y. T., Erdös, E. G. & Levin, Y. (1970) Biochim. Biophys. Acta 2. 214, 374-376.
- Erdös, E. G. & Skidgel, R. A. (1987) Lab. Invest. 56, 345-348.
- Caldwell, P. R. B., Seegal, B. C., Hsu, K. C. & Soffer, R. L. (1976) 4. Science 191, 1050-1051
- Ryan, U. S., Ryan, J. W., Whitaker, C. & Chiu, A. (1976) Tissue Cell 8, 5. 125-145.
- Bull, H. G., Thornberry, N. A. & Cordes, E. H. (1985) J. Biol. Chem. 6. 260. 2963-2972
- Laemmli, U. K. (1970) Nature (London) 227, 680-685. 7
- Lathe, R, (1985) J. Mol. Biol. 183, 1-12. 8.
- Gosporadowicz, D., Cheng, J. & Lirette, M. (1983) J. Cell. Biol. 97, 1677-9. 1685.
- 10. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochemistry 18, 5294-5299.
- 11. Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C. & Kunkel, L. M. (1987) Cell 50, 509-517. Feinberg, A. P. & Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
- 12
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. 13. USA 74, 5463-5467.
- Thomas, P. S. (1983) Methods Enzymol. 100, 255-266. 14.
- 15.
- Southern, E. M. (1975) J. Mol. Biol. 98, 503-517. Holmquist, B., Bünning, P. & Riordan, J. F. (1979) Anal. Biochem. 95, 16. 540-548
- Iwata, K., Blacher, R., Soffer, R. L. & Lai, C. Y. (1983) Arch. Biochem. 17. Biophys. 227, 188-201. St. Clair, D. K., Presper, K. A., Smith, P. L., Stump, D. C. & Heath,
- 18. E. C. (1986) Biochem. Biophys. Res. Commun. 141, 968-972.
- Hooper, N. M., Keen, J., Pappin, D. J. C. & Turner, A. J. (1987) 19. Biochem. J. 247, 85-93.
- 20. Bernstein, K. E., Martin, B. M., Striker, L. & Striker, G. (1988) Kidney Int. 33, 652-655
- Bause, E. (1983) Biochem. J. 209, 331-336. 21.
- 22.
- Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961. Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132. 23.
- Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1982) Nature (London) 24. 299. 371-374.
- Titani, K., Hormodson, M. A., Ericsson, L. H., Walsh, K. A. & Neurath, H. (1972) Nature (London) New Biol. 238, 35–37. 25.
- Kester, W. R. & Matthews, B. W. (1977) Biochemistry 16, 2506–2516. Malfroy, B., Schofield, P. R., Kuang, W. J., Seeburg, P. H., Mason, 26
- 27. A. J. & Henzel, W. J. (1987) Biochem. Biophys. Res. Commun. 144, 59-66
- 28. Devault, A., Lazure, C., Nault, C., Le Moual, H., Seidah, N. G., Chretien, M., Kahn, P., Powell, J., Mallet, J., Beaumont, A., Roques, B. P., Crine, P. & Boileau, G. (1987) EMBO J. 6, 1317-1322
- Goldberg, G. I., Wilhelm, S. M., Kronberger, A., Bauer, E. A., Grant, 29. G. A. & Eisen, A. Z. (1986) J. Biol. Chem. 261, 6600-6605.
- Matsas, R., Kenny, A. J. & Turner, A. J. (1984) Biochem. J. 223, 433-30. 440.
- 31. Das, M, & Soffer, R. L. (1975) J. Biol. Chem. 250, 6762-6768
- Bünning, P. & Riordan, J. F. (1985) J. Inorg. Biochem. 24, 183-198. 32.
- Strittmatter, S. M. & Snyder, S. H. (1986) *Mol. Pharmacol.* 29, 142–148. Cumin, F., Alhenc-Gelas, F. & Corvol, P. (1987) *Clin. Exp. Hypertens.* 33. 34.
- 11-A9, 1892 (abstr.). Das, M., Hartley, J. L. & Soffer, R. L. (1977) J. Biol. Chem. 252, 1316-35.
- 1319. Alhenc-Gelas, F., Weare, J. A., Johnson, R. L., Jr., & Erdös, E. G. 36. (1983) J. Lab. Clin. Med. 101, 83-96.
- 37. El-Dorry, H. A., Pickett, C. B., MacGregor, J. S. & Soffer, R. L. (1982) Proc. Natl. Acad. Sci. USA 79, 4295-4297.