Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 5635–5639, June 1992
Biochemistry

Sequence and molecular characterization of human monocyte/neutrophil elastase inhibitor

(cDNA cloning/inflammation/serpins/ovalbumin family)

EILEEN REMOLD-O'DONNELL*, JEAN CHIN†, AND MONICA ALBERTS

The Center for Blood Research, Harvard Medical School, Boston, MA 02115

Communicated by John T. Edsall, March 18, 1992

ABSTRACT **cDNA encoding human monocyte/neutrophil elastase inhibitor (EI), a $M_r \approx 42,000$ protein with serpin-like functional properties, has been sequenced. The 1316-base-pair sequence was obtained from overlapping clones and amplified DNA from libraries of monocyte-like and neutrophil-like cells. Hybridization with EI cDNA identified three EI mRNA species of 1.5, 1.9, and 2.6 kilobases in U937 monocyte-like cells and no hybridizing mRNA in lymphoblastoid cells lacking detectable EI. The cDNA open reading frame encodes a 379-amino acid protein, of which 167 residues were confirmed by tryptic peptides. Although EI may function extracellularly as well as intracellularly, its deduced sequence lacks a typical cleavable N-terminal signal sequence. Sequence analysis established that EI is a member of the serpin superfamily. EI has greatest homology (50.1% identity of amino acids) with plasminogen activator inhibitor 2, also a monocyte protein, and ovalbumin and gene Y, which were previously grouped as an ancient branch of the serpin superfamily. The extent of EI identity with the functionally related serpin α1 antitrypsin is only 30.1%. Sequence alignment indicates that the reactive center P_1 residue is Cys-344, consistent with abrogation of elastase inhibitory activity by iodoacetamide and making EI a naturally occurring Cys-serpin. The cleavable bond, Cys-Met, suggests an oxidation-sensitive molecule capable of inhibiting more than one serine protease. Oxidation sensitivity would limit the place of action of EI to the immediate vicinity of carrier cells. The molecular structure will help clarify the likely role of EI in regulating protease action and preventing tissue damage by phagocytic cells.**

The ability of neutrophils to protect against pathogens is based on cytolytic, oxidative, and proteolytic components. The best studied of the proteolytic components is the serine active site enzyme called elastase or neutrophil elastase, which is synthesized in myelomonocytic precursor cells and found at high levels in neutrophils and at lower levels in monocytes and some macrophages (1). Elastase is stored as active enzyme in azurophil granules, together with the related proteases cathepsin G (2) and proteinase 3 (3, 4). Elastase is a potent broad-specificity enzyme, which degrades phagocytized and extracellular substrates (5). In addition to defense functions, elastase has a well-publicized role as pathological agent in arthritic joint diseases, skin diseases including psoriasis, and inflammatory lung diseases including cystic fibrosis and emphysema (5).

Several human proteins function as potent fast-acting inhibitors of elastase. Each of these has a characteristic physiological location, likely equivalent to its principal site of inhibitory action. These molecules are secretory cell protein that inhibits leukocyte proteases (SLPI), a M_r 12,000 acid-stable polypeptide produced by secretory cells and found in bronchial mucus, cervical mucus, seminal plasma, and salivary gland secretions [also known as antileukoprotease and human seminal proteinase inhibitor (HUSI)] (6, 7); elafin, a M_r 7000 acid-stable polypeptide found in skin (8); α1 antitrypsin (α1-AT; α1 proteinase inhibitor), the M_r 50,000 prototype of serpin superfamily protease inhibitors, found in plasma at high concentrations from which it enters tissue sites (9); and elastase inhibitor (EI), the M_r 42,000 protein found within the same cells as elastase—i.e., neutrophils, monocytes, and macrophages. Human EI, the subject of this study, was detected initially as an activity in the cytoplasmic fraction of blood leukocytes and lung macrophages (10, 11); it was purified from monocyte-like cells (12, 13). When incubated with elastase (13), pure EI forms a covalent E–EI complex, a reaction described only for serpin inhibitors, suggesting that EI is a member of that superfamily (13).

We report here the cDNA cloning and sequence of human EI‡ and characterize it as a Cys-serpin, a naturally occurring serpin molecule with an active site cysteine residue (14).

MATERIALS AND METHODS

Cells and Protein Analysis. U937 monocyte-like cells and CEM lymphoblastoid cells were grown without stimulating agents (13). EI was purified from U937 cells (13). Forty to 80 μg was heat-inactivated (85–90°C for 9 min) and treated with 0.3–1% (wt/wt) trypsin (Sigma) at 37°C for 18 hr. Tryptic peptides were separated by C_{18} reverse-phase HPLC (VYDAC, Hesperia, CA). N-terminal sequences were determined by automated Edman degradation at the Dana Farber Cancer Institute Microchemistry Facility using an Applied Biosystems 470A gas-phase sequencer.

RNA and DNA Isolation and Analysis. RNA was isolated from guanidinium thiocyanate- and mercaptoethanol-lysed cells by CsCl centrifugation (15). Poly(A)$^+$ RNA was enriched by oligo(dT) chromatography (16). RNA was size-fractionated by formaldehyde/agarose gel electrophoresis (17), transferred to nitrocellulose, and hybridized with the insert of an EI cDNA clone labeled with [^{32}P]dCTP by random priming (18). cDNA was generated from poly(A)-enriched U937 cell RNA and purified as described (19).

Generation of Probes by Mixed Oligonucleotide Primed cDNA Amplification. Oligonucleotides corresponding to EI peptide and containing deoxyinosine substitution, most fre-

Abbreviations: EI, elastase inhibitor of human monocytes, neutrophils, and macrophages; α1-AT, α1 antitrypsin (α1 proteinase inhibitor); SLPI, secretory cell protein that inhibits leukocyte proteases, also known as antileukoprotease and human seminal proteinase inhibitor (HUSI); PAI-2, plasminogen activator inhibitor 2; AT_{III}, antithrombin III.

*To whom reprint requests should be addressed.
†Present address: Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892.
‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M93056).

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.

quent codons, and mixtures of bases, according to guidelines of Moremen (19), were synthesized on an Applied Biosystems model 380B synthesizer. U937 cell cDNA was amplified (19) in 100 μl containing 0.5 μM primers and "amplification reagents" [10 mM Tris·HCl, pH 8.3/50 mM KCl/1.5 mM $MgCl_2$/0.01% gelatin/200 μM of each dNTP/2.5 units of Taq polymerase (Perkin–Elmer/Cetus) for 30 cycles of 92°C for 1 min, 50°C for 2 min, and 72°C for 3 min (thermal cycler; Perkin–Elmer)]. The products were electrophoresed on 1% low-melting agarose and eluted from gel slices (17).

Isolation of Cloned cDNA. A cDNA library prepared from poly(A)-enriched U937 cells RNA by unidirectional insertion of cDNA into λ-ZAP II (Uni-ZAP; Stratagene) was screened by hybridization with a 550-base-pair (bp) amplified cDNA product labeled with [^{32}P]dCTP by random priming. Repeatedly positive plaques were isolated (17), and pBluescript plasmids containing EI cDNA inserts were obtained by *in vivo* excision as described by Stratagene. DNA inserts were isolated by restriction digestion and electrophoresis in low-melting agarose.

Amplification of the 5′ cDNA Region. A λgt11 cDNA library derived from U937 cells (Clontech) and a λgt10 cDNA library from HL60 promyelocytic cells induced toward granulocytic differentiation by dimethylformamide (20) (provided by Stuart Orkin, Children's Hospital, Boston) were used as sources of cDNA for amplification (21). Pelleted bacteriophage (10^5–10^7 plaque-forming units) were denatured and incubated in 100 μl with 1 μM antisense primer (clone U-10, residues 590–607) and flanking vector sense primer (New England Biolabs no. 1218 or no. 1231) and amplification reagents for 40 cycles of 92°C for 1 min, 55°C for 2 min, and 72°C for 3 min. Products of ≈800 bp were excised from agarose gels and used as template in second amplifications with a "nested" antisense primer (clone U-10, residues 182–199). Single dominant products of ≈460 bp were excised from gels, reamplified in six to eight parallel 100-μl reaction mixtures, and purified by agarose electrophoresis, phenol/chloroform extraction, and ethanol precipitation.

cDNA Sequencing. Double-stranded DNA (amplified DNA or pBluescript plasmids) was sequenced by the dideoxy chain-termination/extension reaction (22) with Sequenase reagents (United States Biochemical), vector-specific primers, and 18-mer primers based on EI sequence.

RESULTS AND DISCUSSION

Peptides of EI. Human EI purified from U937 monocyte-like cells (≈40 μg of two preparations) was subjected to N-terminal amino acid sequencing. No sequence was obtained, suggesting a blocked N terminus. The protein was inactivated and trypsinized, and sequence was established for 11 internal peptides (Fig. 1). Several of the peptides showed partial identity with regions of serpin molecules (FASTA and LFASTA analysis, ref. 23) (data not shown).

Isolation of cDNA. Large EI-specific probes were prepared by amplifying U937 cell cDNA with mixed oligonucleotide primer pairs corresponding to EI tryptic peptides, which could be tentatively positioned within the molecule based on their partial identity with regions of known serpins. Four primer pairs formed from sense primers to peptides D and E and antisense primers to peptide I, and contiguous peptides O plus F gave single dominant amplification products, D-I, D-OF, E-I, and E-OF, with the sizes predicted by peptide localization. Product E-OF, 550 bp in length, was used as probe without further characterization.

Hybridization of 700,000 clones of a unidirectional λ-ZAP cDNA library from U937 cells with probe E-OF identified 14 repeatedly positive clones. The clones with the largest inserts [>0.8 kilobase (kb)], U-1, U-5, U-10, and U-11, were partially

```
Peptide A:  E-A-T-T-N-A-P-F-R
Peptide B:  K-T-I-N-Q-(-)-V-K
Peptide D:  (T/A)-F-H-F-N-T-V-E-E-V-H-S-(R)
Peptide E:  (T)-Y-G-A-D-L-A-S-V-D-F-Q-H-A-S-E-D-A
Peptide F:  L-G-V-Q-D-L-F-N-S
Peptide G (residues 7-20):  P-E-N-L-D-F-I-E-V-N-V-S-L-P
Peptide H:  (T)-Y-N-F-L-P-E-F-L-V-S-T-Q-K
Peptide I:  V-L-E-L-P-Y-Q-G-E-E-L-S-M-V-I-L-L-P-D-D-I-E
Peptide M:  K-I-E-E-Q-L-T-L-E-K
Peptide O:  F-K-L-E-E-S-Y-T-L-N-S-D-L-A-(R)
Peptide Z:  F-A-Y-G-Y-I-E-D-L-K
```

FIG. 1. Sequence of tryptic peptides of human EI.

or totally sequenced. These encode EI peptides but do not contain the N-terminal region.

To complete the sequence, total DNA from other cDNA libraries was amplified by use of flanking vector sense primer and clone U-10 antisense primer. Size-selected product (*Materials and Methods*) was subjected to a second amplification with a nested U-10 antisense primer. The amplified cDNA product from a U937 cell library (U′-Ampl) and that from a neutrophil-like HL-60 cell library (HL-Ampl) each contain the missing 5′ region and overlap with the inserts of clones U-10 and U-1.

Nucleotide Sequence of EI. The composite sequence of clones and amplified products is 1316 nucleotides long and includes one open reading frame beginning at position 49 with the first methionine codon, part of the optimal eukaryotic initiation sequence ACCATGG (24) (Fig. 2). The stop codon at position 1186 is followed by a short 3′ untranslated region including one poly(A) signal, AATAAA, at residue 1277 and a poly(A) tail. All overlapping residues are identical except 3′ untranslated residue 1298, which is an A in clone U-10 and a C in U-1 and U-5.

Derived Amino Acid Sequence. The open reading frame encodes a 379-amino acid protein of M_r 42,741 with the

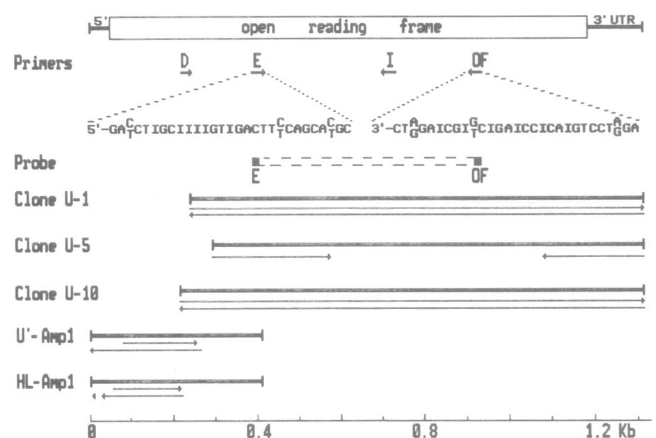

FIG. 2. Sequencing strategy. The upper line represents the 1316-bp composite sequence with the open box indicating the protein coding region. Positions are indicated for primers D, E, I, and OF and sequences for E and OF. A dashed line indicates the likely position of the screening reagent, amplified product E-OF. Inserts of clones U-1 (residues 241–1316), U-5 (292–1316), and U-10 (216–1316) and amplification products U′-Ampl and HL-Ampl (each 1–414) are indicated. Lines with directional arrows indicate regions sequenced by overlapping reactions. UTR, untranslated region.

```
                                                         CCCGCAGCTCGGAGCCCGGAGCGTCCTCGGCGGCTGTCGGTTTTCACC   48
                                  10                                      20                                      30
Met Glu Gln Leu Ser Ser Ala Asn Thr Arg Phe Ala Leu Asp Leu Phe Leu Ala Leu Ser Glu Asn Asn Pro Ala Gly Asn Ile Phe Ile
ATG GAG CAG CTG AGC TCA GCA AAC ACC CGC TTC GCC TTG GAC CTG TTC CTG GCG TTG AGT GAG AAC AAT CCG GCT GGA AAC ATC TTC ATC  138
                                  40                                      50                                      60
Ser Pro Phe Ser Ile Ser Ser Ala Met Ala Met Val Phe Leu Gly Thr Arg Gly Asn Thr Ala Ala Gln Leu Ser Lys Thr Phe His Phe
TCT CCC TTC AGC ATT TCA TCT GCT ATG GCC ATG GTT TTT CTG GGG ACC AGA GGT AAC ACG GCA GCA CAG CTG TCC AAG ACT TTC CAT TTC  228
                                  70                                      80                                      90
Asn Thr Val Glu Glu Val His Ser Arg Phe Gln Ser Leu Asn Ala Asp Ile Asn Lys Arg Gly Ala Ser Tyr Ile Leu Lys Leu Ala Asn
AAC ACG GTT GAA GAG GTT CAT TCA AGA TTC CAG AGT CTG AAT GCT GAT ATC AAC AAA CGT GGA GCG TCT TAT ATT CTG AAA CTT GCT AAT  318
                                 100                                     110                                     120
Arg Leu Tyr Gly Glu Lys Thr Tyr Asn Phe Leu Pro Glu Phe Leu Val Ser Thr Gln Lys Thr Tyr Gly Ala Asp Leu Ala Ser Val Asp
AGA TTA TAT GGA GAG AAA ACT TAC AAT TTC CTT CCT GAG TTC TTG GTT TCG ACT CAG AAA ACA TAT GGT GCT GAC CTG GCC AGT GTG GAT  408
                                 130                                     140                                     150
Phe Gln His Ala Ser Glu Asp Ala Arg Lys Thr Ile Asn Gln Trp Val Lys Gly Gln Thr Glu Gly Lys Ile Pro Glu Leu Leu Ala Ser
TTT CAG CAT GCC TCT GAA GAT GCA AGG AAG ACC ATA AAC CAG TGG GTC AAA GGA CAG ACA GAA GGA AAA ATT CCG GAA CTG TTG GCT TCG  498
                 ◇               160                                     170                                     180
Gly Met Val Asp Asn Met Thr Lys Leu Val Leu Val Asn Ala Ile Tyr Phe Lys Gly Asn Trp Lys Asp Lys Phe Met Lys Glu Ala Thr
GGC ATG GTT GAT AAC ATG ACC AAA CTT GTG CTA GTA AAT GCC ATC TAT TTC AAG GGA AAC TGG AAG GAT AAA TTC ATG AAA GAA GCC ACG  588
                                 190                                     200                                     210
Thr Asn Ala Pro Phe Arg Leu Asn Lys Lys Asp Arg Lys Thr Val Lys Met Met Tyr Gln Lys Lys Lys Phe Ala Tyr Gly Tyr Ile Glu
ACG AAT GCA CCA TTC AGA TTG AAT AAG AAA GAC AGA AAA ACT GTG AAA ATG ATG TAT CAG AAG AAA AAA TTT GCA TAT GGC TAC ATC GAG  678
                                 220                                     230                                     240
Asp Leu Lys CYS Arg Val Leu Glu Leu Pro Tyr Gln Gly Glu Glu Leu Ser Met Val Ile Leu Leu Pro Asp Asp Ile Glu Asp Glu Ser
GAC CTT AAG TGC CGT GTG CTG GAA CTG CCT TAC CAA GGC GAG GAG CTC AGC ATG GTC ATC CTG CTG CCG GAT GAC ATT GAG GAC GAG TCC  768
                                 250                                     260                                    ◇270
Thr Gly Leu Lys Lys Ile Glu Glu Gln Leu Thr Leu Glu Lys Leu His Glu Trp Thr Lys Pro Glu Asn Leu Asp Phe Ile Glu Val Asn
ACG GGC CTG AAG AAG ATT GAG GAA CAG TTG ACT TTG GAA AAG TTG CAT GAG TGG ACT AAA CCT GAG AAT CTC GAT TTC ATT GAA GTT AAT  858
                                 280                                     290                          ◇          300
Val Ser Leu Pro Arg Phe Lys Leu Glu Glu Ser Tyr Thr Leu Asn Ser Asp Leu Ala Arg Leu Gly Val Gln Asp Leu Phe Asn Ser Ser
GTC AGC TTG CCC AGG TTC AAA CTG GAA GAG AGT TAC ACT CTC AAC TCC GAC CTC GCC CGC CTA GGT GTG CAG GAT CTC TTT AAC AGT AGC  948
                                 310                                     320                                     330
Lys Ala Asp Leu Ser Gly Met Ser Gly Ala Arg Asp Ile Phe Ile Ser Lys Ile Val His Lys Ser Phe Val Glu Val Asn Glu Glu Gly
AAG GCT GAT CTG TCT GGC ATG TCA GGA GCC AGA GAT ATT TTT ATA TCA AAA ATT GTC CAC AAG TCA TTT GTG GAA GTG AAT GAA GAG GGA 1038
                                 340             ⇩                       350 ◇                                   360
Thr Glu Ala Ala Ala Ala Thr Ala Gly Ile Ala Thr Phe CYS Met Leu Met Pro Glu Glu Asn Phe Thr Ala Asp His Pro Phe Leu Phe
ACA GAG GCG GCA GCT GCC ACA GCA GGC ATC GCA ACT TTC TGC ATG TTG ATG CCC GAA GAA AAT TTC ACT GCC GAC CAT CCA TTC CTT TTC 1128
                 ◇               370
Phe Ile Arg His Asn Ser Ser Gly Ser Ile Leu Phe Leu Gly Arg Phe Ser Ser Pro ***
TTT ATT CGG CAT AAT TCC TCA GGT AGC ATC CTA TTC TTG GGG AGA TTT TCT TCC CCT TAG AAG AAA GAG ACT GTA GCA ATA CAA AAA TCA 1218

AGCTTAGTGCTTTATTACCTGAGTTTTTAATAGAGCCAATATGTCTTATATCTTTACCAATAAAACCACTGTCCAGAAACAAAAAAAAAAAAAAAAA                   1316
                                                                                     A
```

FIG. 3. cDNA sequence and deduced amino acid sequence for human EI. The five potential N-glycosylation sites (155, 270, 298, 351, 365) are marked by open diamonds, and the two cysteine residues (214 and 344) are indicated by bold uppercase letters. The proposed reactive center Cys-Met, at positions 344–345, is marked by a broad arrow. The poly(A) signal is double-underlined. The Fig. 1 peptides are D (57–69), H (97–110), E (111–128), B (130–137), A (178–186), Z (204–213), I (216–237), M (245–254), G (261–274), O (276–290), and F (291–299). Additional mixed and short tryptic peptides correspond to positions 48–56, 92–96, 163–168, and 312–317. [One 17-residue tryptic peptide (not shown) is not present in the deduced sequence and is not in a data base.]

composition $Asn_{23}Asp_{18}Gln_{11}Glu_{30}His_{7}Lys_{29}Arg_{14}Ala_{30}$-$Cys_{2}Gly_{20}Tyr_{10}Val_{18}Ile_{19}Leu_{41}Phe_{27}Met_{12}Trp_{3}Pro_{12}Ser_{31}$-$Thr_{22}$, values close to those determined for purified EI (13). The deduced sequence is unique since the National Biomedical Research Foundation data base (Release 29, June 1991) contains no identical or nearly identical sequence. The composite cDNA sequence is confirmed as encoding EI because the deduced protein includes exact sequence of 15 tryptic peptides prepared from purified EI totalling 167 amino acids, and each peptide sequence follows lysine or arginine (Fig. 3).

Structure of EI. Comparison with a data bank identified 14 proteins that have >30% amino acid identity with EI. All of these are members of the serpin superfamily (25, 26). Serpin molecules are thought to share a common tertiary structure defined by x-ray diffraction for the prototype molecule α1-AT (27). This highly ordered globular structure consists of nine α-helical regions and three β-sheets arranged in a "stressed configuration" with the reactive center, the most variable region, located in an exposed loop, where it acts as a bait for target protease (26–29). By aligning sequences of 20 serpins, Huber and Carrell (29) defined 51 positions with conserved residues. These residues are distributed throughout the primary sequence, but located in internal regions and surface niches in the α1-AT crystal structure. Forty-nine of these 51 residues are conserved in the EI sequence (Fig. 4), strongly indicating that EI has the highly ordered serpin tertiary structure.

"Ovalbumin Branch" of the Serpin Superfamily. The proteins with greatest homology with EI are PAI-2 (33–36) and the chicken proteins gene Y (30) and ovalbumin (31) (Fig. 4). The PAI-2 amino acid sequence is 50.1% identical with EI, gene Y is 42.7% identical, and ovalbumin is 39.6% identical (Fig. 4).[§] These three proteins were previously grouped as an ancient conserved branch of the serpin superfamily based on identities of amino acid sequence and exon–intron organization (33, 37). The identities within this subgroup extend to the N terminus, a region not strongly conserved in the larger superfamily (Fig. 4). The existence of an "ovalbumin family" within the larger serpin superfamily was predicted in 1980 by Hunt and Dayhoff (25).

Homology of EI and PAI-2. Particularly outstanding is the high homology of EI and PAI-2, both of which are monocyte proteins. Their 50.1% identical amino acid sequence alignment requires only two gaps (Fig. 4). Noteworthy also is their common motif, GYIED, at positions 230–234 in strand B1 (Fig. 4), a region that has minimal identities in the larger superfamily (29). In contrast, α1-AT has only 30.1% identity with EI (seven gaps), even though α1-AT and EI share the functional property of elastase inhibition (Fig. 4).

Glycosylation Sites. Whereas EI purified from U937 cell lysates is nonglycosylated (13), the deduced sequence contains five potential N-glycosylation sites (Fig. 3). The possible existence of glycosylated EI molecules has not been specifically investigated, nor have extracellular fluids been examined for possible content of EI. The closely related monocyte molecule PAI-2 has three potential N-glycosyla-

[§]Other molecules identified by the FASTA (23) search of the National Biomedical Research Foundation data base (Release 29) include AT_{III}, 39.6% identity with EI (39.0% by Fig. 4 methods); mouse contrapsin, 36.4%; rat growth hormone-regulated proteinase inhibitor, 34.3%; α1-antichymotrypsin, 34.0%; heparin cofactor II, 33.8%; protein C inhibitor, 33.5%; protease nexin I, 32.2%; α2-antiplasmin, 31.3%; corticosteroid-binding globulin, 30.9%; α1-AT, 30.3% (30.1% by Fig. 4 methods); plasminogen activator inhibitor 1, 30.1%; thyroxine-binding globulin, 29.3%; and C1 inhibitor, 27.4%.

FIG. 4. Comparison of EI with other serpins. Manual alignment was based on the α1-AT structure (27) with gaps and inserts placed at positions tolerated in the larger serpin superfamily (29). Position numbers refer to α1-AT. Shading indicates identity with EI. Structural regions of the α1-AT crystal structure (27) are indicated. The reactive site is at positions 358–359. The α1-AT and antithrombin III (AT$_{III}$) signal sequences and the AT$_{III}$ N-terminal 28 residues are not shown. EI lacks the variable-length loop between helices C and D (residues 86–89 in α1-AT) and differs in this respect from plasminogen activator inhibitor 2 (PAI-2), gene Y, and ovalbumin (Oval.), which have loops of 37, 14, and 14 residues, respectively, at this position (omitted from the figure). The 51 dark bars indicate residues categorized as conserved in the larger serpin superfamily (29). EI has 49 of these conserved residues but does not have valine at position 388 and asparagine at position 390, a deviance found also in PAI-2, gene Y, and ovalbumin. The motif GYIED at positions 230–234 in EI and PAI-2 (enclosed box) is not found in other serpins. The α1-AT, gene Y, ovalbumin, and AT$_{III}$ sequences are from refs. 9 and 30–32; that of PAI-2 is from refs. 33, 34, 36, and 37.

tion sites (33–36) and is an apparently novel case in which a single mRNA sequence supports synthesis of a nonglycosylated cytoplasmic protein and a secreted glycoprotein ("facultative polypeptide translocation"; ref. 46).

N Terminus. The EI sequence lacks a typical cleavable hydrophobic signal peptide. Its 20 N-terminal residues include five leucines and three alanines but also include hydrophilic residues glutamine, asparagine, glutamic acid, aspartic acid, and arginine. Moreover, a peptide recovered from one EI trypsinization gave the sequence --LDLSLA, corresponding with one deviation to deduced residues 13–18 (LDLFLA). In addition, the deduced EI sequence aligns precisely with ovalbumin and PAI-2, which lack cleavable peptides (38, 39), and begins, relative to α1-AT, in helix A with the first residue buried in the globular structure (Fig. 4).

Cysteine Residues. The presence of at least one free sulfhydryl cysteine was indicated by adherence of EI to sulfhydryl exchange resins (13) and, independently, by abrogation of EI activity by the sulfhydryl reagent iodoacetamide (13). Since the deduced sequence contains only two cysteines (Cys-214 and Cys-344), it follows that both have free sulfhydryls and EI lacks disulfide bonds.

Reactive Center. Within the serpin superfamily, the exposed reactive center is particularly susceptible to variation, and, in many cases, the P_1 residue—i.e., the N-terminal residue of the cleavage site—mirrors the specificity of the target protease and thereby defines inhibitory specificity (26). Alignment with serpin sequences identified the EI P_1 residue as Cys-344, consistent with abrogation of activity by iodoacetamide, and the P_1' residue as Met-345. The EI reactive center thus does not have the "ideal" P_1 residue, valine, which would mirror the specificity of the target protease elastase (40). Similar observations have been made for the plasma elastase inhibitor α1-AT, which has Met-Ser as P_1–P_1' residues.

These findings suggest that EI and α1-AT constitute independent cases in which evolution and natural selection have yielded elastase inhibitors with reactive sites that are less than ideal. An alternative interpretation is that the reactive sites are ideal, but the function of the molecule(s) has been inadequately perceived. According to the latter hypothesis, the *in vivo* function of EI (and α1-AT) is to inhibit several serine proteases, one of which is elastase. The supporting evidence includes (*i*) the presence of three serine proteases (cathepsin G and protease 3 in addition to elastase) in neutrophil granules and the failure thus far to identify other neutrophil/monocyte serpin inhibitor molecules, (*ii*) the inhibition of trypsin as well as elastase by the apparent guinea pig counterpart of EI (41), and (*iii*) the inhibition of elastase and cathepsin G by a highly purified human neutrophil fraction containing "cytosolic inhibitor," which is likely identical to EI (42).

Counterpart Molecules in Other Species. Molecules that are possible counterparts of human EI have been studied in horse, pig, cattle, guinea pigs (references in ref. 13), sheep (43), and mice (44). The horse molecule shares 16 of 20 sequenced residues (45) with human EI residues 343–362 (not shown). Its elastase inhibitory activity is significantly decreased by the oxidant *N*-chlorosuccinimide (45). On the other hand, its P_1–P_1' residues, Ala-Met (45), differ from the Cys-Met of human EI; its M_r of 35,000 also differs from the M_r of 42,000 of human EI. The guinea pig molecule resembles human EI in molecular weight (41), adherence to thiol resins, and abrogation of activity by iodoacetamide (E.R.-O., unpublished findings), the latter suggesting a P_1 cysteine residue. The guinea pig molecule is present in extracellular fluid as well as intracellularly (41).

mRNA. The abundance of EI in monocytes, neutrophils, lung macrophages, and monocyte-like U937 cells and its

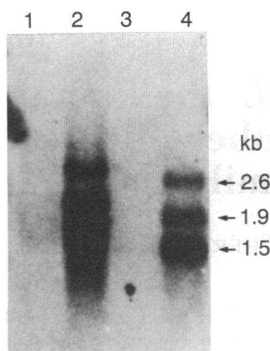

FIG. 5. Northern blot. Total RNA (30 μg) from CEM lymphoblastoid cells (lane 1) or unstimulated U937 cells (lane 2) or poly(A)-selected RNA (2 μg) from CEM cells (lane 3) or U937 cells (lane 4) was hybridized with the insert of clone U-10.

absence in the lymphocyte cell line CEM (13) suggest that expression is tightly regulated. Northern assay of total RNA with U-10 cDNA revealed multiple hybridizing mRNA species in U937 cells and no hybridizing mRNA in CEM cells (Fig. 5). On analysis of poly(A)-enriched U937 cell RNA, the hybridizing mRNA species were characterized as molecules of 2.6, 1.9, and 1.5 kb (Fig. 5). The composite cDNA defined in this study is similar in size to the 1.5-kb mRNA, suggesting that the cDNA sequence is nearly complete. It will be of interest to determine whether EI or a related serpin inhibitor is expressed in cells containing other granule serine proteases—i.e., mast cells, natural killer cells, and cytolytic T lymphocytes.

CONCLUSIONS

The composite cDNA sequence encodes EI, an abundant protein in monocytes, neutrophils, and macrophages, which functions *in vitro* as a fast-acting inhibitor of elastase. Amino acid sequence identities established that EI is a member of the serpin superfamily and is highly likely to have the characteristic stressed serpin secondary structure. Sequence analysis indicated that EI is a member of the serpin subgroup that includes chicken ovalbumin and gene Y and the human molecule PAI-2. This subgroup (ovalbumin family) was previously categorized as an ancient branch of the superfamily based on shared exon–intron organization, a feature otherwise atypical among serpin molecules (33, 37). The homology of EI with PAI-2, also a monocyte molecule, amounts to 50.1%. The deduced EI sequence encodes a 379-amino acid protein with five potential N-glycosylation sites and lacking disulfide bonds. The reactive center P_1–P_1' residues, Cys-Met, suggest that the new serpin molecule (*i*) inhibits more than one protease and (*ii*) is likely oxidation-sensitive. The latter feature would limit the place of action of EI to the immediate vicinity of carrier cells. All of these features are consistent with EI being a major agent terminating and/or regulating the proteolytic action of phagocytic cells, particularly neutrophils.

We thank Philip Auron, R. Alan Ezekowitz, Ignacio Molina, Kelley Moremen, Jon Nixon, Stuart Orkin, Helene Rosenberg, Takashi Sugimori, Daniel Tenen, and Andrew Webb for reagents and/or advice; Ruth Steinbrich for amino acid sequencing; Bonnie Faulkner, Michael Maxim, and Tina Pate for technical assistance; and Richard Lundberg for computer analysis of tryptic peptides. This study was supported by National Institutes of Health Grant HL41579.

1. Takahashi, H., Nukiwa, T., Basset, P. & Crystal, R. G. (1988) *J. Biol. Chem.* **263**, 2543–2547.
2. Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C. & Travis, J. (1987) *Biochemistry* **26**, 2289–2293.
3. Kao, R. C., Wehner, N. G., Skubitz, K. M., Gray, B. H. & Hoidal, J. R. (1988) *J. Clin. Invest.* **82**, 1963–1973.
4. Campanelli, D., Melchior, M., Fu, Y., Nakata, M., Shuman, H., Nathan, C. & Gabay, J. E. (1990) *J. Exp. Med.* **172**, 1709–1715.
5. Janoff, A. (1985) *Annu. Rev. Med.* **36**, 207–216.
6. Seemüller, U., Arnhold, M., Fritz, H., Wiedenmann, K., Machleidt, W., Heinzel, R., Appelhans, H., Gassen, H. G. & Lottspeich, F. (1986) *FEBS Lett.* **199**, 43–48.
7. Thompson, R. C. & Ohlsson, K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6692–6696.
8. Wiedow, O., Schröder, J. M., Gregory, H., Young, J. A. & Christophers, E. (1990) *J. Biol. Chem.* **265**, 14791–14795.
9. Carrell, R. W., Jeppsson, J. O., Laurell, C. B., Brennan, S. O., Owen, M. C., Vaughan, L. & Boswell, D. R. (1982) *Nature (London)* **298**, 329–334.
10. Janoff, A. & Blondin, J. (1971) *Proc. Soc. Exp. Biol. Med.* **136**, 1050–1053.
11. Blondin, J., Rosenberg, R. & Janoff, A. (1972) *Am. Rev. Respir. Dis.* **106**, 477–479.
12. Remold-O'Donnell, E. (1985) *J. Exp. Med.* **162**, 2142–2155.
13. Remold-O'Donnell, E., Nixon, J. C. & Rose, R. M. (1989) *J. Exp. Med.* **169**, 1071–1086.
14. Remold-O'Donnell, E., Chin, J. & Alberts, M. (1992) *FASEB J.* **6**, 1974 (abstr.).
15. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
16. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
18. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
19. Moremen, K. W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5276–5280.
20. Parkos, C. A., Dinauer, M. C., Walker, L. E., Allen, R. A., Jesaitis, A. J. & Orkin, S. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3319–3323.
21. Rosenberg, H. F., Corrette, S. E., Tenen, D. G. & Ackerman, S. J. (1991) *BioTechniques* **10**, 53–54.
22. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
23. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
24. Kozak, M. (1986) *Cell* **44**, 283–292.
25. Hunt, L. T. & Dayhoff, M. O. (1980) *Biochem. Biophys. Res. Commun.* **95**, 864–871.
26. Carrell, R. & Travis, J. (1985) *Trends Biochem. Sci.* **10**, 20–24.
27. Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984) *J. Mol. Biol.* **177**, 531–556.
28. Carrell, R. W., Pemberton, P. A. & Boswell, D. R. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 527–535.
29. Huber, R. & Carrell, R. W. (1989) *Biochemistry* **28**, 8951–8966.
30. Heilig, R., Muraskowsky, R., Kloepfer, C. & Mandel, J. L. (1982) *Nucleic Acids Res.* **10**, 4363–4382.
31. McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M. & Brownlee, G. G. (1978) *Nature (London)* **273**, 723–728.
32. Bock, S. C., Wion, K. L., Vehar, G. A. & Lawn, R. M. (1982) *Nucleic Acids Res.* **10**, 8113–8125.
33. Ye, R. D., Ahern, S. M., Le Beau, M. M., Lebo, R. V. & Sadler, J. E. (1989) *J. Biol. Chem.* **264**, 5495–5502.
34. Webb, A. C., Collins, K. L., Snyder, S. E., Alexander, S. J., Rosenwasser, L. J., Eddy, R. L., Shows, T. B. & Auron, P. E. (1987) *J. Exp. Med.* **166**, 77–94.
35. Schleuning, W. D., Medcalf, R. L., Hession, C., Rothenbühler, R., Shaw, A. & Kruithof, E. K. O. (1987) *Mol. Cell. Biol.* **7**, 4564–4567.
36. Antalis, T. M., Clark, M. A., Barnes, T., Lehrbach, P. R., Devine, P. L., Schevzov, G., Goss, N. H., Stephens, R. W. & Tolstoshev, P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 985–989.
37. Samia, J. A., Alexander, S. J., Horton, K. W., Auron, P. E., Byers, M. G., Shows, T. B. & Webb, A. C. (1990) *Genomics* **6**, 159–167.
38. Palmiter, R. D., Gagnon, J. & Walsh, K. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 94–98.
39. Ye, R. D., Wun, T. C. & Sadler, J. E. (1988) *J. Biol. Chem.* **263**, 4869–4875.
40. McRae, B., Nakajima, K., Travis, J. & Powers, J. C. (1980) *Biochemistry* **19**, 3973–3978.
41. Remold-O'Donnell, E. & Lewandrowski, K. (1983) *J. Biol. Chem.* **258**, 3251–3257.
42. Thomas, R. M., Nauseef, W. M., Iyer, S. S., Peterson, M. W., Stone, P. J. & Clark, R. A. (1991) *J. Leukocyte Biol.* **50**, 568–579.
43. Junger, W., Hallström, S., Redl, H. & Schlag, G. (1988) *Biol. Chem. Hoppe-Seyler* **369**, Suppl., 63–68.
44. Takeuchi, K. H. & Swank, R. T. (1989) *J. Biol. Chem.* **264**, 7431–7436.
45. Potempa, J., Dubin, A., Watorek, W. & Travis, J. (1988) *J. Biol. Chem.* **263**, 7364–7369.
46. Belin, D., Wohlwend, A., Schleuning, W. D., Kruithof, E. K. O. & Vassalli, J. D. (1989) *EMBO J.* **8**, 3287–3294.