Molecular Evolution of Serpins: Homologous Structure of the Human α_1-Antichymotrypsin and α_1-Antitrypsin Genes†

Jia-ju Bao,‡ Richard N. Sifers, Vincent J. Kidd,§ Fred D. Ledley, and Savio L. C. Woo*

Howard Hughes Medical Institute, Department of Cell Biology, and Institute of Molecular Genetics, Baylor College of Medicine, Houston, Texas 77030

Received January 22, 1987; Revised Manuscript Received June 12, 1987

ABSTRACT: α_1-Antichymotrypsin belongs to a supergene family that includes α_1-antitrypsin, antithrombin III, ovalbumin, and angiotensinogen. The human chromosomal α_1-antichymotrypsin gene has been cloned and its molecular structure established. The gene is approximately 12 kb in length and contains five exons and four introns. The locations of the introns within the α_1-antichymotrypsin gene are identical with those of the human α_1-antitrypsin and angiotensinogen genes. Other members of this supergene family contain introns located at nonhomologous positions of the genes. The homologous organization of the α_1-antichymotrypsin and α_1-antitrypsin genes corresponds with the high degree of homology between their protein sequences and suggests that these loci arose by recent gene duplication. A model is presented for the evolution of both the genomic structure and the protein sequences of the serine protease inhibitor superfamily.

α_1-Antichymotrypsin (ACT) is a plasma serine protease inhibitor with affinity toward chymotrypsin-like enzymes (Travis et al., 1978a,b). Its target substrates include neutrophile cathepsin G, mast cell chymase, and proteases that convert angiotensin I to the biologically active vasoconstrictor angiotensin II in vitro (Reilly et al., 1982; Wintroub et al., 1981; Tonnensen et al., 1982). It is a glycoprotein of 68 000 daltons and is structurally related to α_1-antitrypsin (AAT) (Morii & Travis, 1983a,b), which is a specific inhibitor of neutrophile elastase and protects the lung elastin fibers from degradation by this protease (Olsen et al., 1975; Tuttle & Jones, 1975). The physiological function of ACT has not been clearly defined. However, it has been shown that neutrophile cathepsin G enhances the rate of elastin digestion by neutrophile elastase (Reilly & Travis, 1978) and that ACT is selectively concentrated in the bronchial lumen of patients with chronic infections (Ryley & Brogan, 1972). Thus, ACT could be involved in the maintenance of the overall protease–antiprotease balance in the lung. ACT has also been shown to markedly reduce the natural cytotoxic activity of T-cell killer lymphocytes (Graragha et al., 1982), presumably by binding to a chymotrypsin-like enzyme on the surface of these cells. In addition, it is a sensitive marker of hepatocellular carcinoma (Orodonez & Manning, 1984), and high-level production has been demonstrated in both benign and malignant breast epithelial cells (Tokes et al., 1981).

In man, the normal serum level of ACT is only about one-tenth that of α_1-antitrypsin (AAT), and both inhibitors are acute-phase reactants. In response to inflammation or infection, the plasma concentration of ACT increases to 4 times its basal level within 8 h, while the concentration of AAT increases slowly and doubles in 3 days (Laurell, 1972).

In order to determine the structure of the ACT gene and the evolutionary relationship between ACT and other serine protease inhibitors, we have previously cloned and sequenced a full-length human ACT cDNA clone (Chandra et al., 1983). Comparison of the amino acid sequence of ACT with that of AAT revealed an overall homology of 42%. Moreover, their reactive sites are located in the same position of the protein sequence though the active site is Leu-Ser in ACT (Morii & Travis, 1983) instead of Met-Ser in AAT (Johnson & Travis, 1978). In the present paper we describe the molecular structure of the human chromosomal ACT gene and present a model for evolution of the serine protease inhibitor superfamily on the basis of both the organization of these gene loci and their protein sequences.

Materials and Methods

Materials. Restriction endonucleases, T4 DNA ligase, and polynucleotide kinase were purchased from New England Biolabs. *Escherichia coli* DNA polymerase I (Klenow fragment) and calf intestine alkaline phosphatase were products of Boehringer Mannheim. [α-^{32}P]dCTP (sp act. >3000 Ci/mmol) for nick translation and [α-^{35}S]dATP (sp act. >650 Ci/mmol) for sequencing were obtained from Amersham Corp. [γ-^{32}P]dATP (sp act. >7000 Ci/mmol) was purchased from ICN.

Isolation of the Human Chromosomal ACT Gene. A human genomic DNA library was constructed with the cosmid vector pHC79 and screened with the full-length human ACT cDNA clone phACT235 (Chandra et al., 1981) as probe according to the procedure of Lau and Kan (1983). Positive colonies were purified by secondary and tertiary screening. A 19-kb *Hin*dIII fragment containing the entire ACT gene was isolated from low-melting agarose gel and subcloned into pUC18 at the *Hin*dIII site to generate subclone phACT19.

Identification of Exonic Regions. *Eco*RI and *Bam*HI sites were mapped according to the method of Smith and Birnsteil (1976). Exons were localized by digestion of phACT19 DNA with *Eco*RI and/or *Bam*HI followed by Southern blotting with a ^{32}P-labeled nick-translated ACT cDNA probe. Exon I was identified by oligonucleotide hybridization of dried agarose

† This work was supported by NIH Grant HL27509. J.B. is the recipient of a WHO fellowship on leave from the Department of Biophysics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences. R.N.S. is a recipient of Postdoctoral Fellowship HL07343 from the National Institutes of Health. S.L.C.W. is an Investigator and F.D.L. is an Assistant Investigator of the Howard Hughes Medical Institute.

‡ Present address: Department of Biophysics, Institute of Basic Medical Sciences, 5 Dong Dan San Tiao, Beijing, People's Republic of China.

§ Present address: Department of Cell Biology and Anatomy, University of Alabama—Birmingham, Birmingham, AL 35294.

0006-2960/87/0426-7755$01.50/0

gels containing cloned ACT DNA fragments as described (DiLella et al., 1986).

Nucleotide Sequence Analysis. Exon-containing fragments of the human ACT gene were inserted into M13mp18 (Messing, 1983) and sequenced by the dideoxynucleotide chain termination method of Sanger et al. (1977).

Primer Extension Analysis. Primer extension was performed with human liver poly(A) RNA as template as described previously (Chandra et al., 1981). The primer consisted of a 57-bp *Fnu*4HI/*Hin*fI fragment from the 5′ end of the ACT cDNA clone phACT235 which was end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase at the complementary strand. The fragment was denatured, allowed to hydridize with total human liver poly(A) RNA, and extended toward the 5′ end of the mRNA with reverse transcriptase and unlabeled nucleoside triphosphates. The largest resulting fragment was isolated by polyacrylamide gel electrophoresis and sequenced according to the procedure of Maxam and Gilbert (1980). A 26-mer oligonucleotide probe corresponding to this sequence was synthesized and used as a probe for exon I.

S1 Nuclease Analysis. A 60-mer oligonucleotide complementary to exon I and suspected upstream noncoding DNA sequences was synthesized. The oligonucleotide was end-labeled as described above to a specific activity of 7×10^8 cpm/μg of DNA. The probe was annealed with total liver RNA in 40 μL of 70% formamide, 0.3 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 0.03 M 1,4-piperazinediethanesulfonic acid (Pipes), (pH 6.5) by denaturation at 70 °C for 15 min followed by incubation at 37 °C for 3 h. Unprotected probe was digested by the addition of 0.32 mL of S1 buffer (0.3 M NaCl, 2 mM $ZnCl_2$, 50 mM NaOAc, pH 6.5), containing 50 μg/mL sheared and denatured herring sperm DNA (Sigma) and 2000 units of S1 nuclease (Miles). Following incubation at 37 °C for 2 h, nucleic acids were precipitated with 2.5 volumes of cold ethanol, washed with 70% ethanol, and dried under vacuum. Precipitates were dissolved in a solution containing 80% formamide, 50 mM NaOH, 1 mM EDTA, 0.01% xylene cyanol, and 0.01% bromphenol blue. Following incubation at 65 °C for 3 min, samples were electrophoresed in a 10% polyacrylamide/7 M urea sequencing gel. Following electrophoresis, radiolabeled bands were detected by autoradiography. The end-labeled 60-mer oligonucleotide was subjected to specific chemical degradation (Maxam & Gilbert, 1980) and electrophoresed in the gel as a size marker.

Computer Analysis of Sequence Homology and Evolution. Computer analysis was performed by use of the on-line facilities of the Protein Identification Resource (PIR) and methods previously described (Dayhoff, 1972, 1976). Sequences were obtained from the PIR, EMBL, and GENBANK repositories.

Results

Organization of the Human Chromosomal ACT Gene. Approximately 5×10^5 recombinants of the human genomic cosmid library were screened, and three positive clones were obtained. One positive clone, ACTcosI, contained an insert of approximately 38 kb and was selected for further study. Within ACTcosI, there was a single *Hin*dIII fragment of 19 kb that hybridized to ACT cDNA. This 19-kb *Hin*dIII fragment was subcloned to pUC18 to generate phACT 19 and proved to contain the entire human ACT gene. Exon I was identified by hybridization with the oligonucleotide 5′-CCTCAGGGAGCTGGAGAGCTGGTGGA-3′, which corresponds to the 5′-untranslated region in ACT mRNA as-

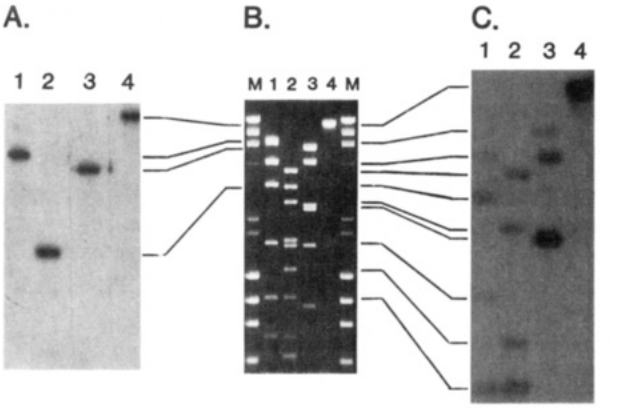


FIGURE 1: Southern hybridization analysis of phACT19. Plasmid DNA (lane 4) was digested with *Eco*RI (lane 1), *Bam*HI (lane 3), and both (lane 2). (Panel A) DNA in dried gels hybridized to a 26-mer oligonucleotide from the first exon of ACT; (panel B) digested DNA from phACT19 stained with ethidium bromide; (panel C) DNA transferred to nitrocellulose paper and probed with the full-length cDNA showing fragments containing exons II–V.

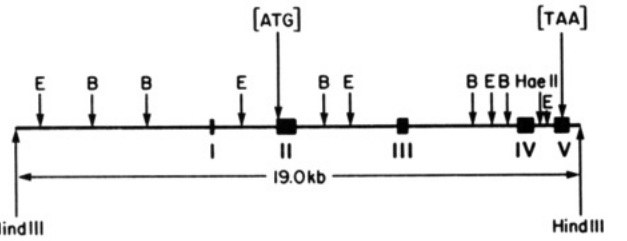


FIGURE 2: Restriction map of *Hin*dIII, *Eco*RI, *Bam*HI, and *Hae*II sites in the human chromosomal ACT gene. Exons are denoted by solid boxes and marked I–V.

certained by primer extension, and a single hybridizing band was detected (Figure 1A). Additional exons were identified by Southern blot hybridization analysis of phACT19 DNA digested with *Eco*RI and/or *Bam*HI (Figure 1B) and probed with ACT cDNA (Figure 1C). Four hybridizing fragments were detected, representing four different exons. The overall structure of the human ACT gene was established by additional restriction mapping and Southern blot hybridization as shown in Figure 2. The gene contains five exons spanning approximately 12 kb of genomic DNA.

All five exon-containing regions of the gene were sequenced. Figure 3 shows the sequences at each intron–exon boundary. The results of this analysis were consistent with established consensus splice sites (Mount, 1982). Three types of intron–exon junctions were observed in the human ACT gene. Type O junctions (those between codons) occur once, whereas type I and II junctions (interrupting codons after the first and second nucleotide, respectively) occur a total of 3 times. These observations are in accordance with intron–exon boundaries of other eukaryotic genes. Of particular interest are positions of the introns in the ACT gene that are identical with those in the highly homologous human AAT gene (Figure 3).

Exon Structure of the Human ACT Gene. The transcription initiation region was identified by an S1 nuclease protection analysis using a ^{32}P end-labeled 60-mer oligonucleotide as probe. Following hybridization with human liver RNA and S1 nuclease treatment, the predominant protected DNA fragment appeared to contain a guanidine residue at its 5′ terminus (Figure 4A). This corresponds to a cytidine as the transcription initiation site in the sense strand of the ACT gene. Compilation of several eukaryotic transcription start sites has revealed that most begin with an adenine, preceded by a cytidine (Bucher & Trifonov, 1986). Thus, we have assigned

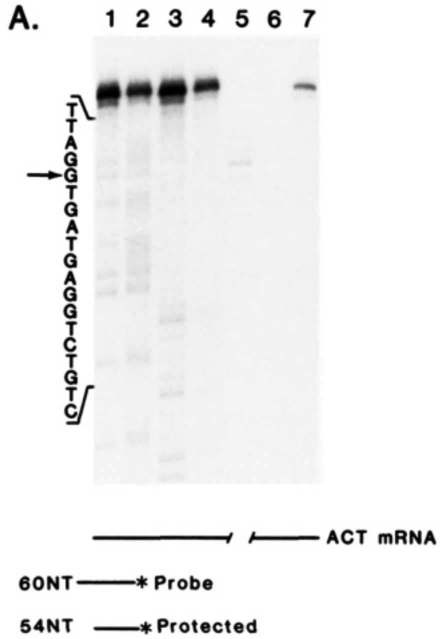
```
               Exon I                                                Exon II
                                                                     -24 -23
                                       Intron A                      Met Pro
AAT . . . TGG GAC AGT GAA TC  gtaagt . . . . . . . ttgcag      G ACA ATG CCG . . .
ACT . . . TCC CTG AGG CAG     gtaatc . . . . . . . tcacag    TTG AGA ATG GAG . . .
                                                                     Met Glu
                                                                     -25 -24

               Exon II                                               Exon III
          189 190 191                                          192 193 194 195
          Phe Phe Lys                  Intron B                Gly Lys Trp Glu
AAT . . . TTC TTT AAA G       gtaagg . . . . . . . ctccag       GC AAA TGG GAG . . .
ACT . . . TTC TTT AAA G       gtgagt . . . . . . . cttcag       CC AAA TGG GAG . . .
          Phe Phe Lys                                          Ala Lys Trp Glu
          187 188 189                                          190 191 192 193

               Exon III                                              Exon IV
          179 180 181 182                                          283 284 285
          Glu Asp Arg Arg              Intron C                    Ser Ala Ser
AAT . . . GAA GAC AGA AG      gtgatt . . . . . . . cgacag        A TCT GCC AGC . . .
ACT . . . CTG GAG TTC AG      gtgatt . . . . . . . ttctag        A GAG ATA GGT . . .
          Leu Glu Phe Arg                                          Glu Ile Gly
          278 279 280 281                                          282 283 284

               Exon IV                                               Exon V
          328 329 330 331                                      332 333 334 335
          Lys Leu Ser Lys              Intron D                Ala Val His Lys
AAT . . . AAG CTC TCC AAG     gtgaga . . . . . . . ctccag      GCC GTG CAT AAG . . .
ACT . . . GCA GTC TCC CAG     gtgagt . . . . . . . gacgag      GTG GTC CAT AAG . . .
          Ala Val Ser Gln                                      Val Val His Lys
          328 329 330 331                                      332 333 334 335
```

FIGURE 3: Exon–intron junction sequences of the human ACT and AAT genes. Exon sequences are in capital letters; intron sequences are in lower-case letters. Exon–intron junctions were positioned by the gt–ag splice rule (Mount, 1982). The numbers shown denote the positions of the corresponding amino acids in the appropriate genes, as deduced from the previously published ACT and AAT cDNA sequences (Chandra et al., 1983; Long et al., 1984).



B.

```
-400  CTGTCTCAAA ATAAAAATAA AAAATAAAAA GAAATAAAAA AGAAATATAC
-350  CAAAATGTTA GCTGGGGTCT TCTCTGGGTA GTAAAGTGCT GGGGGATATT
-300  TTCCAAAGTC CTTCTTTACA TTCTCTGAGT TTTTCCATGT TCTTCAATGA
-250  GTATTTAATA AGCAGATAAA AACTAATACA ACAAAGGATT TTTTCTGTGT
-200  GCTTTTTTGA CCTTGGAGGA AGAGATTAGA GCTAGTCCCA TAACCAGGTT
-150  ATTTGAGTAG GTCTAACAAG CCCTATTACC AGAAATTATC ATCTGGTCAT
-100  TTCCAGTCCG AGAACAGAAC ACTTGGTTGT CCTGGCATTT CCCAAGCAGT
 -50  GGGAGGAGTT CTCTGCAGGA ATAAATAAGC CTCAGCATTC ATGAAAATCC
   1  ACTACTCCAG ACAGACGGCT TTGGAATCCA CCAGCTACAT CCAGCTCCCT
  51  GAGGCAGGTA ATCCATGATG TTTTACATCC TGGGAGCGGA GGAATCTGTT
 101  TTTCCAGGAG AGTTTTAGGC AGCAGCCTGG AGTGTGTGGA GTGTGAGGGG
 151  TAAGCAGAG
```

FIGURE 4: Determination of the transcription initiation site of the human ACT gene. (A) A ^{32}P end-labeled 60-mer oligonucleotide was incubated with 100 μg of human liver RNA. Following treatment with S1 nuclease, protected probe fragments were separated in a 10% polyacrylamide/7 M urea sequencing gel and detected by autoradiography. (Lane 5) Probe after hybridization and S1 nuclease treatment; (lane 6) probe plus tRNA after S1 nuclease treatment; (lane 7) probe without S1 nuclease. The end-labeled probe was also subjected to Maxam–Gilbert sequencing as shown: (lane 1), G; (lane 2) G+A; (lane 3) T+C; (lane 4) C. An arrow depicts the apparent transcription start site. (B) Boundaries of exon I and identification of upstream sequences. Transcription begins at position +1 and proceeds to the right. Exon I of the ACT gene is underlined, followed by intronic sequences. A candidate TATA box is located upstream of the transcription start site and is depicted within a box.

the adjacent adenine as the proposed transcription start site in the ACT gene (Figure 4B). The existence of a more 5′ exon is excluded since this adenine is not proceded by an AG dinucleotide, which is part of the cononical intron–exon splice acceptor site (Mount, 1982). A candidate TATA box is located upstream of the transcription initiation site (Figure 4B); however, no apparent CAAT box was identified.

Exon II contains the initiation methionine codon and codes for 214 amino acids. Exon III contains 268 nucleotides and codes for 89 amino acids. Exon IV contains 151 nucleotides and codes for 50 amino acids. The 3′-most exon, exon V, is 442 bp in length. It contains 231 bp of protein-coding sequence, the termination codon, and 208 bp of 3′-untranslated sequence, which includes the AATTAA polyadenylation sequence as previously reported (Chandra et al., 1986).

Computer Analysis of Sequence Homology. As reported previously, there is extensive homology among the proteins in the serine protease inhibitor superfamily. The evolutionary distance in PAMs (accepted point mutations) among these proteins was made on the basis of the assumption that sequence evolution occurs by nucleic acid substitution at a constant rate (Table I). The rate of evolution of the serine protease inhibitors may be calculated for angiotensinogen sequences (51 PAMs, 75 million years) and for AAT on the basis of the distance between human and baboon sequences (9 PAMs, 13 MYR). Both calculations give an approximate rate of evolution of 60–70 PAMs/100 million years.

Discussion

α_1-Antichymotrypsin is one member of a superfamily of serine protease inhibitors (Hunt & Dayhoff, 1980). Five other members of this superfamily have been described and the organization of the genetic loci determined: (1) α_1-antitrypsin (AAT), which is an inhibitor of neutrophil elastase; (2) antithrombin III, which is a major physiological regulator of thrombin and several other activated coagulation factors; (3) chicken ovalbumin, which is the major egg white protein; (4) Y protein, which is closely related to ovalbumin but has unknown function; (5) angiotensinogen, the precursor for angiotensin II, which is a critical ingredient in the regulation of arteriolar vasconstriction and aldosterone secretion. There is extensive sequence homology between various members of this superfamily. The amino acid homology of ACT and AAT is 42%, and there is a high degree of homology between the nucleic acid sequences of these genes as well (Kurachi et al., 1981; Chandra et al., 1983; Doolittle, 1983). Other members of this superfamily exhibit somewhat lower levels of homology, and angiotensinogen is the most distantly related member of this superfamily with only 10% homology to the other members.

Despite the homology among these sequences, there are significant differences in the organization of the genetic loci. The number of introns varies between four (AAT, ACT, and

Table I: Evolutionary Distance among Proteins in the Serine Proteinase Inhibitor Superfamily[a]

proteins	AT-III (human)	AAT (human)	AAT (baboon)	ACT (human)	OVAL (chicken)	Y (chicken)[b]	ANGIO (rat)[b]	ANGIO (human)[b]
AT-III (human)		76	73	75	76	73	90	89
AAT (human)	202		9	65	75	75	90	90
AAT (baboon)	180	9		65	75	75	89	89
ACT (human)	197	131	132		77	77	91	92
OVAL (chicken)	201	192	194	215		43	91	92
Y (chicken)	179	197	191	215	62		91	92
ANGIO (rat)	512	482	441	595	548	599		38
ANGIO (human)	471	505	461	708	661	718	51	

[a] Optimal alignments between eight proteins in the serine protease inhibitor superfamily were constructed. The difference between sequences was determined, and the evolutionary distance PAMs (accepted point mutations) were calculated with algorithms as described (Dayhoff et al., 1972; Dayhoff, 1976). Sequences were from the Protein Identification Resource (PIR). [b] PAMs computed from percent differences.

angiotensinogen) and seven (ovalbumin and Y protein), and the introns interrupt the mRNA coding sequence at nonhomologous positions (Leicht et al., 1982) as well as some overlapping positions (Prochownik et al., 1985). Paradoxically, the number and position of the introns of the AAT and angiotensinogen genes are identical despite the fact that these proteins exhibit the least homology (Tanaka et al., 1984).

In the present work we determined the position of the introns in the human ACT gene and compared their placement with other members of this superfamily (Figure 5). The position of the introns in ACT was found to be identical with those in AAT. The first intron, intron A, is located at the 5′-untranslated region of both genes and precedes the initiation codon by 4–6 bp. Introns B, C, and D interrupt the protein-coding sequences of the two genes at homologous regions. Furthermore, the types of the intron–exon junctions between these two genes are exactly the same (Figure 3). The similar organization of the ACT and AAT genes is consistent with the high degree of homology between these sequences and suggests that the divergence of these two loci is a relatively recent event. This interpretation is also consistent with the observation that both genes are located at the q31–q32 region of human chromosome 14 (Schroeder et al., 1985; Rabin et al., 1986).

The positions of the introns within all five members of the superfamily are shown in Figure 5A. Positions a, c, and d correspond to intron A of ACT, AAT, angiotensinogen, ovalbumin, and Y protein. These introns are in homologous positions within the 5′-untranslated region of each gene. The first intron of antithrombin III, intron A, is in a different position (b) and interrupts the signal peptide. Intron B of antithrombin III, ovalbumin, and Y protein occurs at an identical position (e) within the nucleotide sequence (Prochownik et al., 1985). Positions f, g, and h correspond to introns C, D, and E of ovalbumin and Y protein, and there are no introns in this position in other members of the superfamily. Introns B, C, and D in ACT, AAT, and angiotensinogen are in identical positions (i, l, and n, respectively). Positions j and k correspond to introns F and G of ovalbumin and Y protein. Positions m and o correspond to introns D and E of antithrombin III.

Cornish-Bowden (1982) suggested that the duplication of an ancestral gene containing many introns, and sequential deletion of introns during evolution, could produce related genes whose introns differ and are an overlapping subset of the original intron complement. Alternatively, duplication of an ancestral intron-free gene, and sequential insertion of introns, could give rise to the same present-day pattern (Leicht et al., 1982). Both models account for the presence of introns at nonhomologous positions in genes with homologous sequences. Moreover, according to these models the evolution of a gene superfamily might be reconstructed by grouping

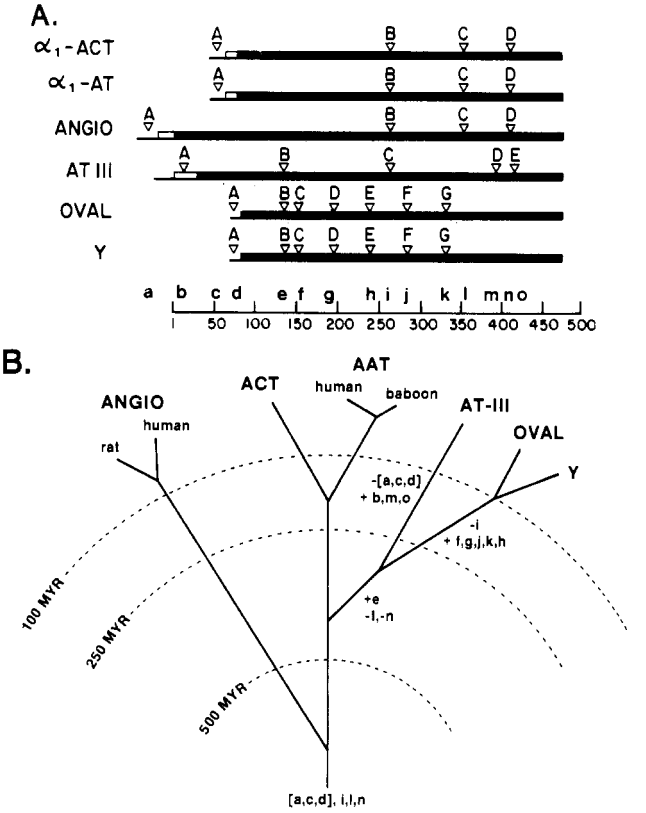


FIGURE 5: (A) Intron positions in genes of the serine protease inhibitor superfamily. Introns for each individual gene are indicated in capital letters. A composite map of intron positions utilized in all members of this superfamily is indicated by small letters a–o below. This schematic demonstrates that many introns are in identical positions in several members of this superfamily, while other introns are present in unique positions (see text). (B) Schematic model of an evolutionary tree for the serine protease inhibitor superfamily based on insertion and deletion of introns as well as degree of protein sequence homology. Introns designated [a,c,d], i, l, and n are assumed to be present in the primordial ancestor. Insertion (+) and deletion (−) of introns at each segment of the tree are indicated. This evolutionary tree is proposed on the basis of the data in Table I with approximate 500, 250 and 100 million year intervals indicated.

genes that have introns in identical positions as the product of recent gene duplications. Models for evolution of the serine protease inhibitor superfamily based on either addition or deletion have been described (Prochownik et al., 1985).

The difficulty with the previous proposed models of intron insertion or deletion in the serine protease inhibitor superfamily is that angiotensinogen, which exhibits the least sequence homology among any of the members of this superfamily, has a genomic organization identical with that of AAT and ACT. Models based on intron insertion from an intron-free progenitor or intron deletion from an intron-rich progenitor place the divergence of angiotensinogen, ACT, and AAT as a recent

event. This conclusion contradicts evolutionary models based on protein sequence data, which place the divergence of angiotensinogen before the divergence of the other protease inhibitor genes. The rate of mutation acceptance for angiotensinogen appears to be the same as for AAT. Thus, there is no evidence for a more rapid divergence of this sequence that could account for a rapid loss of homology following a recent gene duplication.

These two approaches toward reconstructing the evolution of the protease inhibitor superfamily may be reconciled by assuming that introns may be both added and deleted during the course of evolution and that the immediate progenitor of the serine protease inhibitor superfamily had a genetic organization similar to that of ACT, AAT, and angiotensinogen (Figure 5B). According to this model, angiotensinogen arose from a gene duplication greater that 500 million years ago as suggested by protein sequence data and still exhibits the primordial intron organization. A second gene duplication 250–500 million years ago may have given rise to the progenitor for the ACT and AAT genes and the progenitor for the ovalbumin and antithrombin III genes. Several intron addition and deletion events must have occurred in the progenitor of the ovalbumin and antithrombin III genes before a subsequent gene duplication approximately 250 million years ago gave rise to separate lineages for ovalbumin and antithrombin III. More recent gene duplications within this superfamily gave rise to separate loci for the ovalbumin and Y protein loci and separate loci for AAT and ACT. This model of serine protease inhibitor superfamily evolution is consistent with evolutionary distances calculated from protein sequence data and is parsimonious in that it entails a minimum number of insertion/deletion events and does not require repetition of events at the same locus. Obviously, more extensive calculations of evolutionary rate for the different arms of this evolutionary tree will be required to further test and refine this model.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Drs. Andrzej Elvanowski and David George of the Protein Identification Resource in sequence analysis. We also thank Dr. John Collins for the human genomic cosmid library and Lori Reed, Anne Hodgson, and Yi Li for technical assistance. We also express our appreciation to Kelly Porter and Debbie Martin for typing the manuscript.

Registry No. α_1-Antichymotrypsin, 9004-07-3; α_1-antitrypsin, 9041-92-3.

REFERENCES

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Bucher, P., & Trifonov, E. N. (1986) *Nucleic Acids Res. 14*, 10009–10026.

Chandra, T., Kurachi, K., Davie, E. W., & Woo, S. L. C. (1981) *Biochem. Biophys. Res. Commun. 103*, 751–758.

Chandra, T., Stackhouse, R., Kidd, V. J., Robson, K. J. H., & Woo, S. L. C. (1983) *Biochemistry 22*, 5055–5061.

Cornish-Bowden, A. (1982) *Nature (London) 297*, 625–262.

Dayhoff, M. O. (1976) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 35*, 2132–2138.

Dayhoff, M. O., Eck, R. V., & Park, C. M. (1972) *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, pp 89–99, National Biomedical Research Foundation, Washington, DC.

DiLella, A. G., Marvit, J., Lidsky, A. S., Guttler, F., & Woo, S. L. C. (1986) *Nature (London) 322*, 799–803.

Doolittle, R. F. (1983) *Science (Washington, D.C.) 222*, 417–419.

Graragha, P., Gianazza, E., Arnand, D., Neels, M., & Ades, E. W. (1982) *J. Reticuloendothel. Soc. 32*, 125–130.

Hohn, B., & Collins, J. (1980) *Gene 11*, 291–298.

Johnson, D., & Travis, J. (1978) *J. Biol. Chem. 253*, 7142–7144.

Kurachi, K., Chandra, T., Degen, S. J. F., White, T. T., Marchioro, T. L., Woo, S. L. C., & Davie, E. W. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6826–6830.

Lau, Y. F., & Kan, Y. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5225–5229.

Laurell, C. B. (1972) in *Pulmonary Emphysema and Proteolysis* (Mittman, C., Ed.) pp 161–166, Academic, New York.

Leicht, M., Long, G. L., Chandra, T., Kurachi, K., Kidd, V. J., Mace, M., Davie, E. W., & Woo, S. L. C. (1982) *Nature (London) 297*, 655–659.

Long, G. L., Chandra, T., Kurachi, K., Woo, S. L. C., & Davie, E. W. (1984) *Biochemistry 23*, 4828–4237.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65 (Part 1)*, 497–559.

Messing, J. (1983) *Methods Enzymol. 101*, 20–78.

Morii, M., & Travis, J. (1983a) *Biochem. Biophys. Res. Commun. 111*, 438–443.

Morii, M., & Travis, J. (1983b) *J. Biol. Chem. 258*, 12749–12752.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Olsen, G. N., Harris, J. O., Castle, J. R., Waldeman, R. H., & Karmgard, H. J. J. (1975) *J. Clin. Invest. 55*, 427–430.

Ordonez, N. J., & Manning, T., Jr. (1984) *Am. J. Gastroenterol. 79*, 959–963.

Prochownik, E. V., Bock, S. C., & Orkin, S. H. (1985) *J. Biol. Chem. 260*, 9608–9612.

Rabin, M., Watson, M., Kidd, V., Woo, S. L. C., Breg, W. R., & Ruddle, F. H. (1986) *Somatic Cell Mol. Genet. 12*, 209–214.

Reilly, C. F., & Travis, J. (1978) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 925* (Abstr.).

Reilly, C. F., Tewksury, D. A., Schechter, N. M., & Travis, J. (1982) *J. Biol. Chem. 257*, 8619–8622.

Ryley, H. C., & Brogan, T. D. (1972) *J. Clin. Pathol. 26*, 852–856.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Schroeder, W. T., Miller, M. F., Woo, S. L. C., & Saunders, G. F. (1985) *Am. J. Hum. Genet. 37*, 868–872.

Smith, H. O., & Bernstiel, M. L. (1976) *Nucleic Acids Res. 3*, 2387–2398.

Tanaka, T., Oh Kubo, H., & Nakanishi, S. (1984) *J. Biol. Chem. 259*, 8063–8065.

Tokes, Z. A., Gendler, S. J., & Dermer, G. B. (1981) *J. Supramol. Struct. 17*, 69–77.

Tonnesen, M., Klempner, M., Austen, K., & Wintroub, B. (1982) *J. Clin. Invest. 69*, 25–30.

Travis, J., Garner, D., & Bowen, J. (1978a) *Biochemistry 17*, 5647–5651.

Travis, J., Bowen, J., & Baugh, R. (1978b) *Biochemistry 17*, 5651–5656.

Tuttle, W. C., & Jones, R. K. (1975) *Am. J. Clin. Pathol. 64*, 477–482.

Wintroub, B., Klickstein, L., Kaempter, C., & Austen, K. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 1204–1208.