Location of two of the introns in the antithrombin-III gene

Susanne Jagd, Karen Vibe-Pedersen and Staffan Magnusson

Department of Molecular Biology and Plant Physiology, University of Aarhus, DK-8000 Århus C, Denmark

Received 5 September 1985

At least two of the introns in the antithrombin-III (AT-III) gene are located in positions different from those of the other three proteins in this superfamily for which the gene structures are known, namely, ovalbumin, α_1-antitrypsin and angiotensinogen. In another part of the 3'-portion of the AT-III gene there is no intron where each of the other three gene structures has one.

Antithrombin-III superfamily *Intron* *Homology*

1. INTRODUCTION

Antithrombin-III (AT-III, heparin cofactor) inhibits most of the serine proteinases involved in blood clotting (review [1]). Its primary structure was determined in 1977 [2]. The amino acid sequence has also been deduced from cDNA sequences [3–5]. Other serine proteinase inhibitors from blood plasma, namely α_1-antitrypsin [6] and α_1-antichymotrypsin [7], ovalbumin from hen egg white [8,9], angiotensinogen from blood plasma [10,11] and Z-protein from barley endosperm [12] have turned out to have amino acid sequences which are sufficiently homologous with that of AT-III to prove that all these proteins belong to the same superfamily. The gene structures of some of the members of this superfamily have been elucidated, namely those of ovalbumin [13], α_1-antitrypsin [14] and angiotensinogen [15]. Because of the distant evolutionary origin and apparently different functions of at least some of these proteins it is interesting to compare the gene structures in this superfamily of proteins. In essence that of ovalbumin with 7 exons has turned out to be quite significantly different from those of α_1-antitrypsin and angiotensinogen, which on the other hand are mutually quite similar [15].

In a study of a gene clone of human antithrombin-III we have now found that at least part of the AT-III gene near the 3'-end is quite different not only from that of ovalbumin but also from those of α_1-antitrypsin and angiotensinogen.

2. MATERIALS AND METHODS

2.1. *Cloning of the AT-III gene*

The cosmid pHC79-H-AT-III with a 32.4 kbp insert containing the AT-III gene was a kind gift from Dr S.L. Woo. Its isolation is to be published elsewhere (S.L. Woo et al., in preparation).

2.2. *Isolation of subclones*

The pHC79-H-AT-III cosmid was transformed into competent *E. coli* 8767 and cosmid DNA was prepared using standard procedures [16]. The DNA was digested with *Eco*RI (Boehringer, Mannheim) and the *Eco*RI fragments were eluted from preparative agarose gels using the procedure of Dretzen et al. [17] or with DEAE paper (Schleicher and Schuell, Dassel, FRG) using the directions of the manufacturer. The isolated *Eco*RI fragments were ligated into pBR328 [18] which had been linearized with *Eco*RI and treated with alkaline phosphatase (Boehringer). The plasmids were then transformed into competent *E. coli* K803 [16]. Those recombinant clones were selected which grow on agar plates with tetracyclin, but not chloramphenicol, since the *Eco*RI site in pBR328 is located in the chloramphenicol resistance gene. Recombinant clones containing an

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

insert of 2.4 kbp were grown on a large scale (pA2.4) as were the clones with inserts of 4.7 kbp (pA4.7) and 3.8 kbp (pA3.8) [16].

2.3. *Restriction mapping*

The mapping of pHC79-H-AT-III with *Eco*RI, *Hin*dIII and *Bam*HI is to be published elsewhere (S.L. Woo et al., in preparation). The pA2.4, pA4.7 and pA3.8 subclones were mapped by a combination of single and double enzyme digestions. The digests were examined by agarose gel electrophoresis and compared with digests of pBR328 obtained with the same enzyme combinations.

2.4. Pst*I*/Eco*RI M13 clones*

The pA4.7 and pA3.8 clones were digested with *Pst*I and *Eco*RI and fragments were eluted from an agarose gel as described already (0.4 kbp from pA4.7 and 0.5 kbp from pA3.8). The fragments were ligated into M13mp8 and M13mp9 (which had been digested with *Pst*I and *Eco*RI) and transformed into competent *E. coli* JM103 [16].

2.5. Fnu*DI*/Eco*RI M13 clones*

The pA4.7 clone was digested with *Eco*RI and the insert of 4.7 kbp was eluted from an agarose gel. The fragment was digested with *Fnu*DI (Biolab), ligated into M13mp8 (digested with *Sma*I (Biolab) and *Eco*RI) and transformed into competent *E. coli* JM103.

Single-stranded DNA was prepared from recombinant M13 plaques using standard procedures [16] and sequenced by the dideoxy method [19] using the standard 15-mer primer, which was a kind gift from R. Garrett, University of Aarhus.

3. RESULTS AND DISCUSSION

The insert of the cosmid pHC79-H-AT-III contains the coding part of the human AT-III gene (mapped in fig.1). The *Eco*RI fragments were subcloned and partly sequenced. Restriction mapping and DNA sequencing of the clones containing the 4.7 kb *Eco*RI fragment (pA4.7) and the 3.8 kb *Eco*RI fragment (pA3.8) showed the position of 2 introns in the 3′-end of the gene. The sequences around the *Pst*I site in pA4.7 and from the 3′-end of the insert of pA3.8 are shown in fig.2A and B and compared with the corresponding parts of the

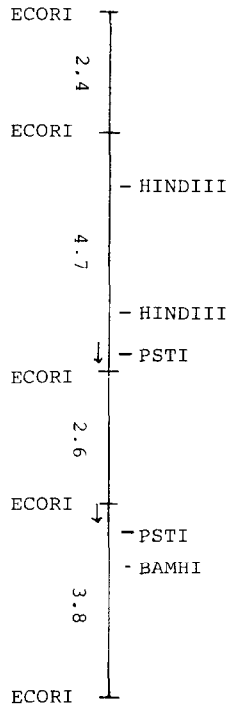

Fig.1. Restriction map of part of the insert of the pHC79-H-AT-III cosmid clone containing the coding part of the antithrombin-III gene. The numbers refer to the sizes of the *Eco*RI fragments in kbp. The arrows indicate the areas sequenced. In the 4.7 kbp fragment there is an *Fnu*DI site (not shown) 125 bp upstream from the *Pst*I site. This *Fnu*DI site was used for subcloning the DNA into M13 for sequencing (see section 2).

cDNA sequence published by Chandra et al. [5]. The sequenced areas are shown with arrowheads in fig.1. The sequences show the presence of 2 introns, at positions 1153 (fig.2A) and 1218 (fig.2B), in the cDNA sequence of AT-III. They also show that there are no introns present in that part of the AT-III gene which corresponds to the base stretches 882–1153 and 1218–1442 in the cDNA of AT-III.

There are 2 silent A-to-G transitions corresponding to positions 981 and 1011 in the cDNA sequence [5]. The latter creates the *Pst*I site used in the sequence determination and was observed as a polymorphism in the cDNA sequence of Prochownik et al. [4].

The positions of the 2 introns in the AT-III gene identified here correspond to the positions after

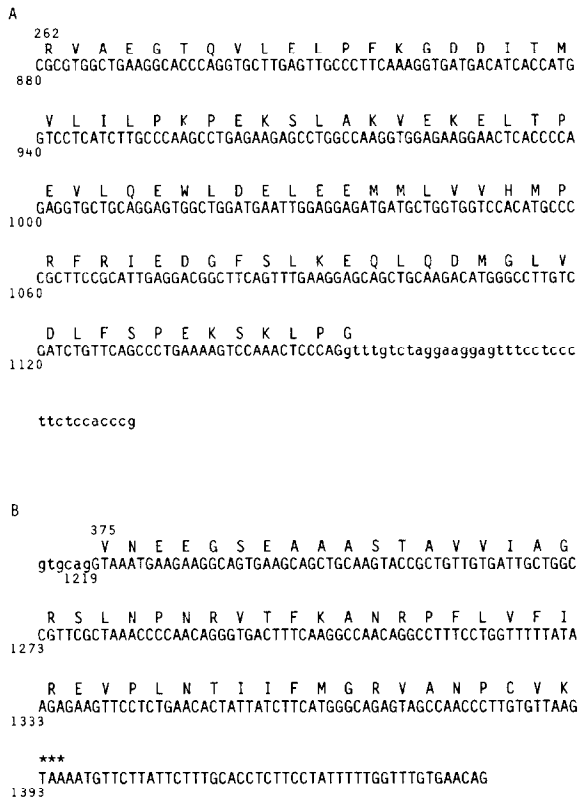

Fig.2. (A) Partial nucleotide sequence of the pA4.7 subclone containing part of the human antithrombin-III gene. The sequence corresponds to bases 880–1153 in the cDNA sequence [5], coding for amino acid residues 262–353 [2], followed by the sequence of the first 37 bases of the intron at position 1153. (B) Partial nucleotide sequence of the pA3.8 subclone, containing the last 6 bases of the intron at position 1219 of the cDNA [5]. This area codes for the last part of AT-III (residues 375–432) and contains 47 bases of the 3′-non-coding region. The stop codon is marked with 3 asterisks.

Fig.3. Alignment (as in [11]) of the amino acid sequence of residues 261–432 of human antithrombin-III (AT-III) with the homologous stretches of rat angiotensinogen (AGTR) (residues 313–477), human α_1-antitrypsin (A1AT) (residues 230–394) and chicken ovalbumin (OVCH) (residues 220–385). Intron positions are marked with arrowheads.

the first base of Gly-352 and between Glu-374 and Val-375 in the amino acid sequence. In fig.3 an alignment is shown of the C-terminal part of the protein sequences of angiotensinogen [10], α_1-antitrypsin [6], ovalbumin [8] and AT-III [2] with the positions of their introns [13–15]. The area between amino acid residues 262 and 352 in AT-III is all encoded on one exon, whereas the corresponding area is interrupted by one intron in each of the other 3 members of the superfamily. This area contains the primary proteinase inhibitor sites of AT-III and α_1-antitrypsin and the subtilisin-susceptible site in ovalbumin cleaved during conversion to plakalbumin [20,21]. Of the 2 introns found at positions in the gene corresponding to residues 352 and 374 in the C-terminal part of AT-III, neither has an equivalent in any of the genes for the other 3 proteins.

Previous results showed that genes coding for members of this superfamily of proteins have 2 different intron patterns, one for ovalbumin and another for α_1-antitrypsin and angiotensinogen. Our results provide evidence showing that the intron pattern for antithrombin-III is different from the others, thus adding new information to the current discussion of evolutionary origin of introns [22,23] favouring the opinion that introns are of relatively late origin [22].

ACKNOWLEDGEMENTS

This work was supported by NIH grant HL 16238 (S.M.), the Danish Cancer Society (K.V.-P.), and a Senior Fellowship from the University of Aarhus, Science Faculty (K.V.-P.). We thank Dr Savio L. Woo for giving us the pHC79-H-AT-III clone, Rita R. Hansen for excellent technical assistance and Dorte Abildsten and Erik Østergaard Jensen for valuable discussions.

REFERENCES

[1] Damus, P.S., Hicks, M. and Rosenberg, R.D. (1973) Nature 246, 355–357.
[2] Petersen, T.E., Dudek-Wojciechowska, G., Sottrup-Jensen, L. and Magnusson, S. (1979) in: The Physiological Inhibitors of Coagulation and Fibrinolysis (Collen, D. et al. eds) pp.43–54, Elsevier/North-Holland, Amsterdam.
[3] Bock, S.C., Wion, K.L., Vehar, G.A. and Lawn, R.M. (1982) Nucleic Acids Res. 10, 8113–8125.
[4] Prochownik, E.V., Markham, A.F. and Orkin, S.H. (1983) J. Biol. Chem. 258, 8389–8394.
[5] Chandra, T., Stackhouse, R., Kidd, V.J. and Woo, S.J. (1983) Proc. Natl. Acad. Sci. USA 80, 1845–1848.
[6] Carrell, R.W., Jeppsson, J.-O., Laurell, C.-B., Brennan, S.O., Owen, M.C., Vaughan, L. and Boswell, D.R. (1982) Nature 298, 329–334.
[7] Chandra, T., Stackhouse, R., Kidd, V.J., Robson, K.J.H. and Woo, S.L.C. (1983) Biochemistry 22, 5055–5061.
[8] McReynolds, L., O'Malley, B.W.O., Nisbet, A.D., Fothergill, J.E., Givol, D., Fuields, S., Robertson, M. and Brownlee, G.G. (1978) Nature 273, 723–728.
[9] Hunt, L.T. and Dayhoff, M.O. (1980) Biochem. Biophys. Res. Commun. 95, 864–871.
[10] Ohkubo, H., Kageyama, R., Ujihara, M., Hirose, T., Inayama, S. and Nakanishi, S. (1983) Proc. Natl. Acad. Sci. USA 80, 2196–2200.
[11] Doolittle, R.F. (1983) Science 222, 417–419.
[12] Hejgaard, J., Rasmussen, S.K., Brandt, A. and Svendsen, I. (1985) FEBS Lett. 180, 89–94.
[13] Dugaiczyk, A., Wood, S.L.C., Lai, E.C., Mace, M.L. jr, McReynolds, L. and O'Malley, B.W. (1978) Nature 274, 328–333.
[14] Leicht, M., Long, G.L., Chandra, T., Kurachi, K., Kidd, V.J., Mace, M. jr, Davie, E.W. and Woo, S.L.C. (1982) Nature 297, 655–659.
[15] Tanaka, T., Ohkubo, H. and Nakanishi, S. (1984) J. Biol. Chem. 259, 8063–8065.
[16] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor, NY.
[17] Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112, 295–298.
[18] Soberon, X., Covarrubias, L. and Bolivar, F. (1980) Gene 9, 287–305.
[19] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[20] Linderstrøm-Lang, K. and Ottesen, M. (1947) Nature 159, 807–808.
[21] Ottesen, M. and Wollenberger, A. (1952) Nature 170, 801–802.
[22] Rogers, J. (1985) Nature 315, 458–459.
[23] Gilbert, W. (1985) Science 218, 823–824.