Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221-226.
Ntambi, J. M., Marini, J. C., Bangs, J. D., Hajduk, S. L., Jimenez, H. E., Kitchin, P. A., Klein, V, K., & Englund, P. T. (1984) *Mol. Biochem. Parasitol. 12*, 273-286.
Pope, L. E., & Sigman, D. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3-7.
Portugal, J., & Waring, M. J. (1987) *FEBS Lett. 225*, 195-200.
Radic, M. Z., Lundgren, K., & Hamkalo, B. A. (1987) *Cell 50*, 1101-1108.
Ryder, K., Silver, S., Delucia, A. L., Fanning, E., & Tegtmeyer, P. (1986) *Cell 44*, 719-725.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning*, p 5.75, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sigman, D. S., Spassky, A., Rimsky, S., & Buc, H. (1985) *Biopolymers 24*, 183-197.
Snyder, M., Buchman, A. R., & Davis, R. W. (1986) *Nature 324*, 87-89.
Spassky, A., & Sigman, D. S. (1985) *Biochemistry 24*, 8050-8056.
Sturm, N. R., & Simpson, L. (1990) *Cell 61*, 879-884.
Travers, A. A. (1987) *Trends Biochem. Sci. (Pers. Ed.) 12*, 108-112.
Trifonov, E. N., & Sussman, J. L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3816-3820.
Trifonov, E. N., & Ulanovsky, L. (1988) in *Unusual DNA Structures* (Wells, R. D., & Harvey, S. C., Eds.) pp 173-187, Springer-Verlag, New York.
Tullius, T. D., Dombroski, B. A., Cherchill, M. E. A., & Kam, L. (1987) *Methods Enzymol. 155*, 537-558.
Ulanovsky, L., & Trifonov, E. N. (1987) *Nature 326*, 720-722.
Ulanovsky, L., Bonder, M., Trifonov, E. N., & Choder, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 862-866.
Wu, H. M., & Crothers, D. M. (1984) *Nature 308*, 509-513.
Yoon, C., Prive, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332-6336.
Zahn, K., & Blattner, F. R. (1985) *Nature 317*, 451-453.
Zahn, K., & Blattner, F. R. (1987) *Science 236*, 416-422.

cDNA Cloning and Sequencing of Rat α_1-Macroglobulin[†,‡]

Birgitta Wärmegård, Nancy Martin, and Staffan Johansson*
Department of Medical and Physiological Chemistry, The Biomedical Center, Box 575, S-751 23 Uppsala, Sweden
Received September 6, 1991; Revised Manuscript Received November 20, 1991

ABSTRACT: cDNA clones coding for the plasma protease inhibitor α_1-macroglobulin were isolated from a rat liver library. The obtained cDNA sequence contained 4701 nucleotides and had an open reading frame coding for a 1500 amino acid long protein, including a 24 amino acid signal peptide. The identity of the deduced protein sequence as α_1-macroglobulin was established by comparison with published peptide sequences of the protein. The mature protein shares 53% and 57% overall amino acid identity with the two other identified members of the rat α-macroglobulin family, α_1-inhibitor 3 and α_2-macroglobulin. A sequence typical for an internal thiol ester was identified. Of the 24 cysteines, 23 are conserved with α_2-macroglobulin. However, instead of the two most C-terminal cysteines in α_2-macroglobulin, which forms a disulfide bridge in the receptor binding domain, α_1-macroglobulin contains phenylalanine. One mRNA species hybridizing with the α_1-macroglobulin probe was observed in rat and mouse liver RNA (~6.2 kb), whereas no corresponding transcript was detected in RNA from human liver.

α-Macroglobulins (αMs)[1] are large plasma proteins which act as inhibitors of proteases of all subclasses [for reviews, see Sottrup-Jensen (1987, 1989)]. Upon cleaving of a particular peptide stretch in the αM, the "bait" region, the protease becomes trapped due to a conformational change of the inhibitor. In addition, an internal thiol ester in the αM is activated and may form a covalent linkage with the protease. A recently identified cell-surface receptor (Moestrup & Gliemann, 1989; Strickland et al., 1990) recognizes activated αMs and mediates internalization of the protease–αM complex by endocytosis (Van Leuven et al., 1979). Interestingly, also the complement factors C3 and C4, which are structurally related to the αMs (Sottrup-Jensen, 1987), are activated by limited proteolysis, bind their target proteins by use of a labile thiol ester, and are recognized by specific receptors after complex formation.

In rat, three members of the αM protein family have been described (Lonberg-Holm et al., 1987). α_2M and α_1M are tetramers of 180-kDa subunits whereas α_1I_3 is monomeric. α_1M differs from the other two proteins in that each monomer is composed of a heavy chain (~140 kDa) and a light chain (~40 kDa) which are held together by disulfide bonds (Lonberg-Holm et al., 1987). While the primary structures of α_2M and α_1I_3 are known (Gehring et al., 1987; Braciak et al., 1988), only limited peptide and cDNA data of α_1M have

[†] This work was supported by grants from the Swedish Medical Research Council (7147), the King Gustaf V's 80th Anniversary Fund, and the Medical Faculty at Uppsala University.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number JO5359.

[1] Abbreviations: αMs, α-macroglobulins; α_1M, α_1-macroglobulin; α_2M, α_2-macroglobulin; α_1I_3,α_1-inhibitor 3; PZP, pregnancy zone protein; bp, base pair(s); IPTG, isopropyl β-D-thiogalactopyranoside; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl and 15 mM sodium citrate.

0006-2960/92/0431-2346$03.00/0

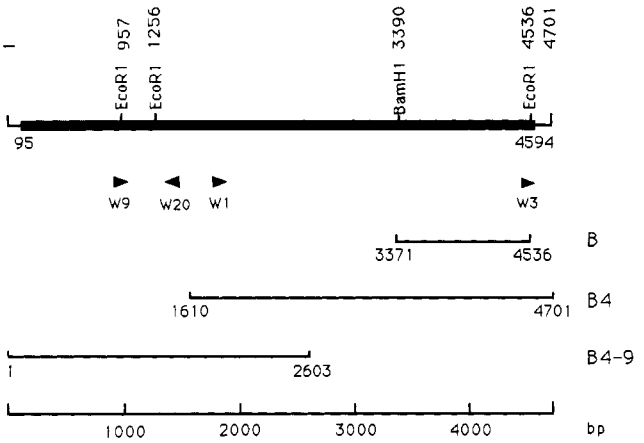


FIGURE 1: Structure of the α_1M cDNA. The coding and noncoding regions of the sequenced cDNA are represented as bold and thin lines, respectively, in the upper part of the figure. The numbers below refer to the nucleotides at the start and the end of the coding region. The positions of the three λ clones (B, B4, B4–9) used for the determination of the nucleotide sequence are indicated. The position and orientation of oligonucleotides mentioned in the text are marked by arrowheads pointing in the 5′ to 3′ direction. Relevant restriction sites are shown.

been published (Lonberg-Holm et al., 1987; Sottrup-Jensen, 1987; Sottrup-Jensen et al., 1989; Enghild et al., 1990). In this paper, the complete amino acid sequence of rat α_1M is reported, allowing for a direct structural comparison of the three proteins.

Materials and Methods

Screening of the cDNA Library. A rat liver cDNA library constructed in λgt11 (6.8×10^5 independent clones) was immunoscreened according to the recommendations of the manufacturer (Clontech), with hen IgG raised against rat hepatocyte fibronectin receptor (integrin $\alpha_5\beta_1$) (Johansson et al., 1988). Briefly, phages adsorbed to *Escherichia coli* strain Y1090 were plated on agar plates containing ampicillin. After incubation at 42 °C for 3.5 h, the plates were overlayed with IPTG-saturated nitrocellulose membranes at 37 °C for 3.5 h. The membranes were then sequentially incubated with the primary antibody and horseradish peroxidase conjugated rabbit (anti-hen IgG) IgG (Imms, Uppsala, Sweden) at 10 μg/mL, before addition of the color substrate 4-chloro-1-naphthol. The anti-fibronectin receptor IgG had been affinity-purified on Sepharose conjugated with isolated fibronectin receptor and eluted from the column with 50 mM diethylamine, pH 11.5. Both the primary and the secondary antibodies were adsorbed against *E. coli* antigens as described by Young and Davis (1983), prior to use in immunoscreening of the library. Positive clones were purified by three additional screening cycles.

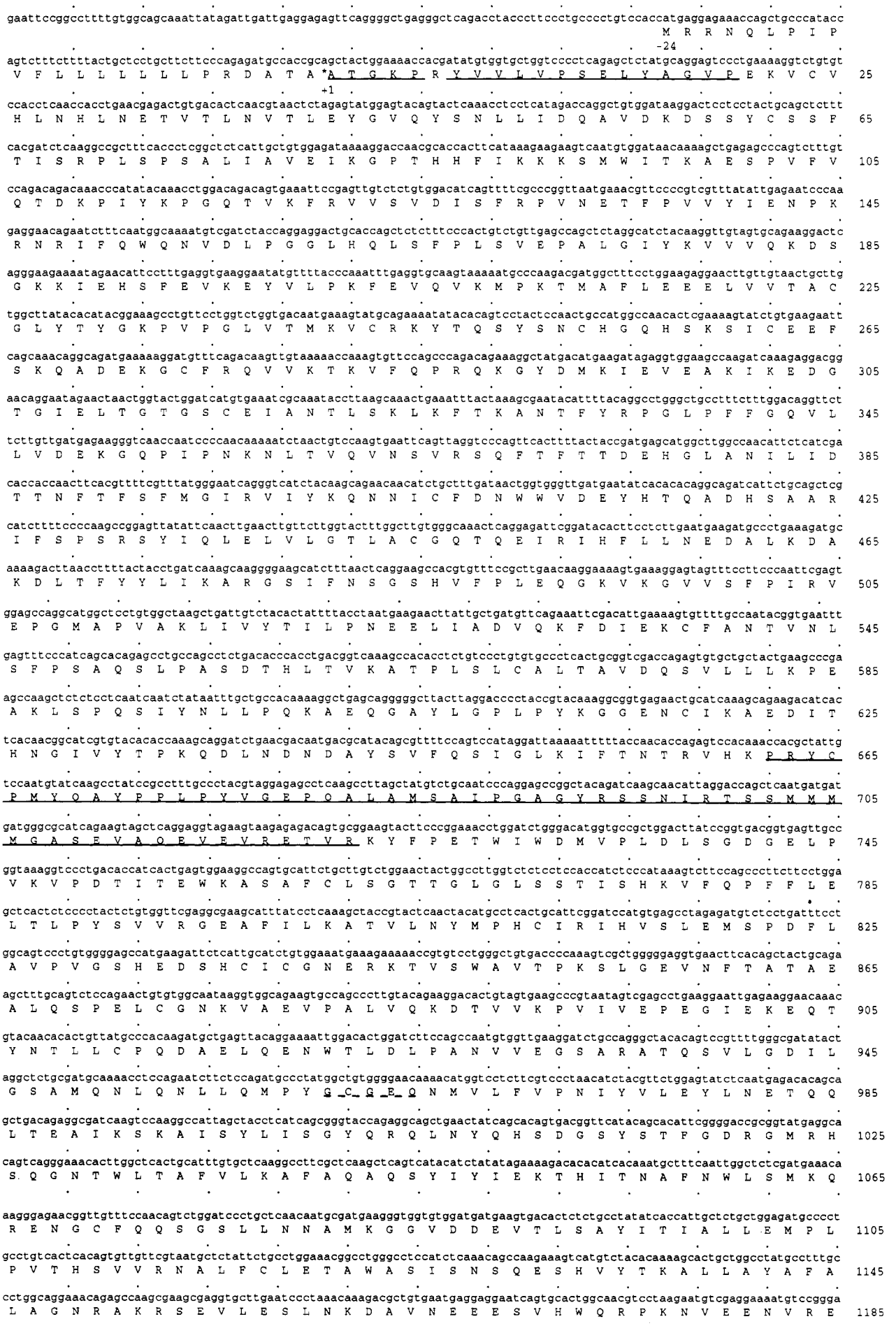
Rescreening of the library was performed with DNA fragment B (Figure 1) (Sambrook et al., 1989), which was labeled with [α-^{32}P]dCTP using the RPN.1601Y multiprime labeling system (Amersham International). The library was also screened with oligonucleotide W1 (Figure 1), labeled at the 5′ end with [γ-^{32}P]ATP using T4 polynucleotide kinase. Hybridization of the membranes with the labeled oligonucleotide was performed at 65 °C in 6 × SSC, 3 × Denhardt's solution, 0.5% SDS, and 0.1 mg/mL single-stranded salmon sperm DNA, followed by washing in 2 × SSC/0.5% SDS at 42 °C.

Subcloning and DNA Sequencing. cDNA inserts of λ clones were isolated by agarose gel electrophoresis and subcloned into pUC19 or M13 mp18/19. The complete nucleotide sequence was determined independently on both strands using the dideoxy chain termination reaction with [α-^{35}S]dATP and the modified T7 DNA polymerase (Sequenase, U.S. Biochemical Corp.) (Tabor & Richardson, 1987). The sequencing primers used were universal oligonucleotide primers which bind close to the polylinker of pUC19 and M13 (U.S. Biochemical Corp.), λ primers (forward and reverse) which bind close to the cloning site in λgt11 (Clontech), and specific α_1M primers, synthesized on Gene Assembler Plus (KABI-Pharmacia AB). To sequence across the internal *Eco*RI sites of the cDNA, two fragments were generated with polymerase chain reactions using λ DNA as template. The oligonucleotides W9 (nucleotides 895–911) together with W20 (antisense sequence of nucleotides 1425–1440) (Figures 1 and 2) and W3 (nucleotides 4481–4498) in combination with the λ reverse primer were used to amplify fragments from B4–9 and B4, respectively. The generated fragments were used as templates for asymmetric amplification with the same pairs of primers. The second set of reactions was carried out in a thermal cycler (Gene ATAQ controller; KABI-Pharmacia AB) in 10 mM Tris buffer, pH 8.3, containing 3 mM $MgCl_2$, 1 mM each dNTP, 2.5 units of Amplitaq DNA polymerase (Perkin-Elmer Cetus), 0.15 μM primer in excess, and 0.006 μM limiting primer in a total volume of 100 μL. The samples were incubated for 40 cycles, each cycle including denaturation at 95 °C for 1 min, annealing at 37 °C for 1 min, and extension at 42 °C for 3 min, followed by one cycle at 42 °C for 10 min. The four different products, each containing an excess of one strand of the two DNA fragments, were sequenced using the complementary oligonucleotide as primer.

Northern Blot Analysis. Total RNA was isolated from rat hepatocytes, mouse liver, and human liver biopsies by the LiCl/urea/SDS method (Sambrook et al., 1989). The hepatocytes had been obtained by collagenase perfusion of rat liver in situ and purified as described (Forsberg et al., 1990). Samples of RNA (20 μg) were denaturated and fractionated by electrophoresis in 1.2% agarose gels containing 2.2 M formaldehyde (Shambrook et al., 1989) and transferred to a nitrocellulose membrane by vacuum blotting. Hybridization was carried out at 65 °C in 10 mM Hepes buffer, pH 7.4, 1 × Denhardt's solution, 6 × SSC, 0.1% SDS, 1 mg/mL yeast RNA, 0.1 mg/mL single-stranded salmon sperm DNA, and 2 mM EDTA. The probe (fragment B, Figure 1) was ^{32}P-labeled by random priming as described above. The final wash of the membrane was performed in 0.2 × SSC/0.1% SDS.

Production of Fusion Protein and Immunization. The *Bam*HI–*Eco*RI fragment of the B clone (nucleotides 3390–4536, Figure 1) was cloned into the bacterial expression vector pGEX-2T (KABI-Pharmacia AB) and transformed into JM109. Production of the fusion protein, containing glutathione *S*-transferase (26 kDa; Smith & Johnson, 1988) and 382 amino acids of α_1M, was induced by the addition of 1 mM IPTG. Three hours after induction, the bacteria were lysed and the proteins separated by SDS–PAGE. The fusion protein band was excised and homogenized in Freund's complete adjuvant before intramuscular immunization of a rabbit.

SDS–PAGE and Immunoblotting. Electrophoresis was performed on polyacrylamide gradient gels (7–15%) in SDS according to the method of Blobel and Dobberstein (1975). In order to minimize autolytic degradation of thiol ester containing proteins, the samples of rat plasma were incubated with 10 mM methylamine at 37 °C for 30 min prior to the addition of SDS–PAGE sample buffer. Immunoblot analysis of proteins electrophoretically transferred to nitrocellulose membranes from the acrylamide gels was performed as described (Burnett, 1981). After incubation of the membranes with antiserum at 1:50 dilution, the specifically bound anti-



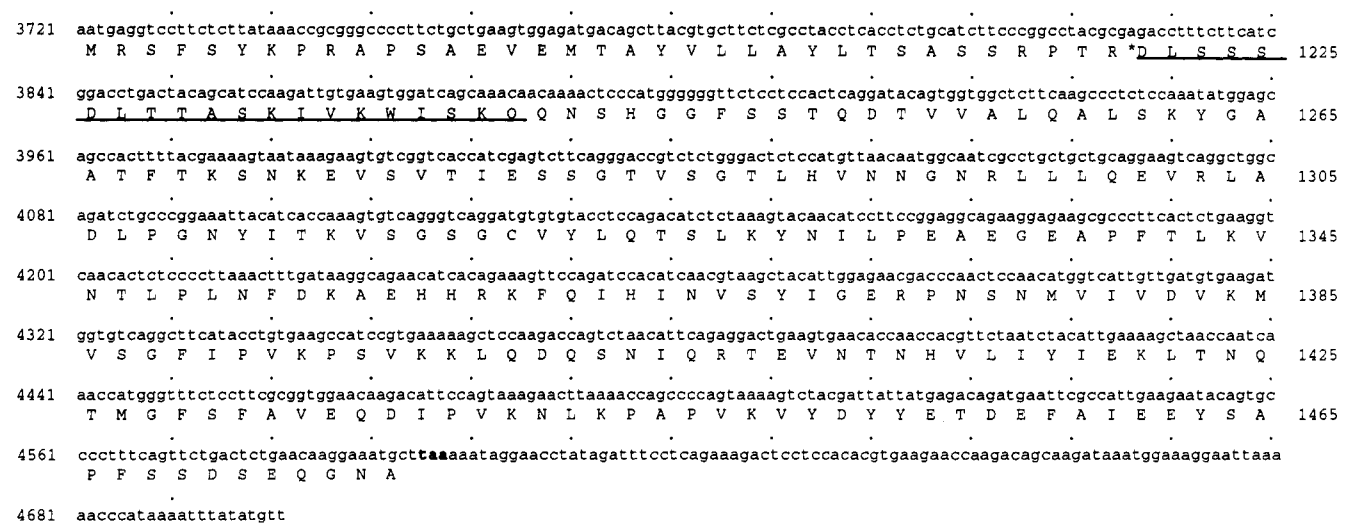


FIGURE 2: Nucleotide sequence and predicted amino acid sequence of rat α_1M. The cDNA sequence of α_1M is shown in the 5′ to 3′ direction. The taa-stop codon (nt 4595–4597) is in boldface letters. Amino acids of the signal peptide are numbered from −24 to −1 and residues of the mature protein from +1 to 1476. The N-terminals of the heavy and light chains of the mature protein (Lonberg-Holm et al., 1987; Sottrup-Jensen, 1987) are indicated by an asterisk. The dashed line marks the conserved thiol ester site. Underlined amino acid sequences have previously been identified at the protein level (Lonberg-Holm et al., 1987; Sottrup-Jensen et al., 1989; Sottrup-Jensen, 1987).

Table I: Comparison of Rat and Human αM Proteins

	rat			human	
	α_1M	α_2M[a]	α_1I_3[b]	α_2M[c]	PZP[d]
precursor (aa[e])	1500	1472	1477	1474	1482
signal peptide (aa)	24	27	24	23	25
mature protein (aa)	1476 (1220 + 256)	1445	1453	1451	1457
molecular mass (Da) (mature peptide backbone)	164402	160670	161053	160798	161028
cysteines (mature protein)	24	25	23	25	25
cysteins conserved with human α_2M	23	25	22		25
overall aa identity with α_1M		57	53	60	56
overall as identity with α_1I_3				61	59
overall aa identity with rat α_2M				73	65

[a] Data from Braciak et al. (1988). [b] Data from Gehring et al. (1987). [c] Data from Kan et al. (1985). [d] Data from Devriendt et al. (1991). [e] Amino acids.

bodies were allowed to react with ^{125}I-labeled protein A. The antigens recognized were detected by autoradiography.

Results

For the purpose of isolating cDNA clones corresponding to fibronectin receptor mRNA, a rat liver cDNA library constructed in the λgt11 expression vector was screened with affinity-purified antibodies against the receptor. One clone (B), which was purified by three additional screening cycles with the antibodies, turned out to contain a cDNA insert which was unrelated to the fibronectin receptor. Determination of the nucleotide sequence of the 1.2 kbp insert instead revealed significant homology with proteins of the α_2M gene family.

Isolation and Characterization of cDNA Clones Coding for α_1M. In order to obtain the complete primary structure of this protein, the rat liver library was rescreened with the B insert as hybridization probe. This resulted in the isolation of 10 new clones, of which B4 had the largest insert (2.9 kbp). After being subcloned into pUC 19, the nucleotide sequence of the 2.9 kbp fragment was determined. It was found to extend from the B clone in the 5′ direction, while the 3′ end was identical to that of B (Figure 1). The B4 sequence had one open reading frame all the way through, but it contained neither a stop codon nor a start methionine. When the inserts of the other isolated λ clones were sequenced, several of them were found to terminate at the same 3′ nucleotide as B, indicating the presence of an internal *Eco*RI site in the cDNA sequence at this position. Since the λ clones thus could contain additional sequence information, intact λ DNA of B4 was sequenced, using a primer (W3) which would hybridize close to the *Eco*RI site. This resulted in a 170-nucleotide sequence extending beyond the *Eco*RI site, which coded for 20 additional amino acids and contained the stop codon.

λ clones containing the 5′ end of the coding sequence were isolated by one further screening of the same liver library, this time using a synthetic oligonucleotide (W1) derived from B4 (Figure 1). Of the new series of clones, B4–9 was found to contain the missing part. In Figure 2, the entire cDNA sequence obtained is shown along with the deduced amino acid sequence. In addition to noncoding sequences of 94 and 107 nucleotides in the 5′ and the 3′ ends, respectively, the 4701-nucleotide sequence encoded a protein of 1500 amino acids. On the basis of complete homology with published peptide sequences (underlined in Figure 2), the protein was identified as α_1M. The N-terminal sequence of the light chain (40 kDa) of α_1M was encompassed in the deduced sequence, demonstrating that the 180-kDa subunit is synthesized as one protein which subsequently is cleaved into two parts (Figure 2). Other features of the sequence include a 24 amino acid signal peptide and a thiol ester activation site (amino acids 961–965).

Comparison of αM Sequences. A comparison of the predicted primary structure of α_1M with that of rat α_2M and α_1I_3 is shown in Figure 3. In addition, Table I lists some characteristics of these proteins and of the two human αMs, α_2M and PZP. Of the three rat αMs, α_1M exhibited the lowest degree of homology with both of the two known human αMs (Table I). The overall homology of α_1M with rat α_2M and α_1I_3 is 57% and 53%, respectively, for the mature proteins. The

```
α1I3   1    LN NSK M    Q  TET   I L  Y      VTAS ISQRGTRK FDELV    LF 60
α1M    1    ATGKPRYVVLVPSELYAGVPEKVCVHLNHLNETVTLNVTLEYGVQYSNLLIDQAVDKDSS 60
α2M    1   S PQ  I M M   L H  T   A FLFS      AVR S  SVHGNQS FT LV    LF 61

       61   H V    P LP SEEEESLDIN E AK K SERRVVLVKNK  V        M     S  120
       61   YCSSFTISRPLSPSAL--IAVEIKGPTHHFIKKKSMWITKAESPVFVQTDKPIYKPGQTV 118
       62   H T   VPQSS DELMF-FT QV  A  E RRQSTVLVK K  L  A              120

       121       M KNLH L  L  LA   D  M   M   D KTEN  K    S  A  IQ P   180
       119  KFRVVSVDISFRPVNETFPVVYIENPKRNRIFQWQNVDLPGGLHQLSFPLSVEPALGIYK 178
       121  R     L E  H L  LI LL  QD  N   A    FN E   K        S  TQ S   180

       181  I IL Q  V E    T M F   R G D  V NAISVYD IIN    AT         H  240
       179  VVVQKDSGKKIEHSFEVKEYVLPKFEVQVKMPKTMAFLEEELVVTACGLYTYGKPVPGLV 238
       181    IRTE  RTV  P S E F       R TV E ITI    MN SV  I          R  240

       241  KISL HG-NPTF S----ET  G K EDSRL NN  ST E NITE  LKENYLK HQAFH 295
       239  TMKVCRKYTQSYSNCHGQHSKSICEEFSKQADEKGCFRQVVKTKVFQPRQKGYDMK--IE 296
       241   VNI    SNP-   F EE VAF  KL Q L GR   S L    S  LKRQE E Q--LD 297

       276   N TVT E   S FS S RI VER RN FL L  DSHF H I   VK R   I  D     355
       297  VEAKIKEDGTGIELTGTGSCEIANTLSKLKFTKANTFYRPGLPFFGQVLLVDEKGQPIPN 356
       298   H   Q E   V E  K LTK TR IT  S VNVDSHF Q I   V       GR T   Y 357

       356  EQVLIKARDAGYTNAT   Q    KFS   NGISDYSLN K YH EESS IHSSCTA R  415
       357  KNLTVQVNSVRSQFTFTTDEHGLANILIDTTNFTFSFMGIRVIYKQNNICFDNWWVDEYH 416
       358  ETIFIGADEANLYINT   K    RFS N DDIMGTSLTV AK  DS A YGFR LT EN 417

       416  AE H T YAVY L K   Y DTEA V P N IHTVQA  I KGQV GVLQQIV H  VM  475
       417  TQADHSAARIFSPSRSYIQLELVLGTLACGQTQEIRIHFLLNEDALKDAKDLTFYYLIKA 476
       418  VE W T YAV      FLH  SLPDK R D  L VQA YI  GE MQEL E V    MMA 477

       476  Q   LQT N THQV P ESQ Q NFALE P  FS V    MLI     DG V   SV  Q 535
       477  RGSIFNSGSHVFPLEQG--KVKGVVSFPIRVEPGMAPVAKLIVYTILPNEELIADVQKFD 534
       478  K G VRA T  L  K   --QMR HF IL SM TDL    R VL A    G VVG TA YE 535

       536  V   LR K H   SPS      Q  MR T S Q   G R        Q    E   SL   595
       535  IEKCFANTVNLSFPSAQSLPASDTHLTVKATPLSLCALTAVDQSVLLLKPEAKLSPQSIY 594
       536    N L  K D V RPNSG   TRAL S M S Q   G R        M   TE  ASL   595

       596  D PGMQDSNFIASSND FEDEDY- LMYQP AR------------EK V RYVRET  MA 642
       595  NLLPQKAEQGAYLGPLPYKGGEN-CIKAEDITHNGIVYTPKQDLNDNDAYSVFQSIGLKI 653
       596  D   V DLT FPQ ADQREEDT G V QN TYI   L S V NT EE M GFLKDM   V 655

       643     LKIKL T  NT--D DMV LAVPAV  DS TDR MYESLPVVAVK PLPQEPPRKDP 700
       654  FTNTRVHKPRYCPMYQAYPPLPYVGEPQALAMSAIPGAGYRSSNIRTSSMMMMGAS--EV 711
       656     SNIR  KV ERLRDN--------------KG  A YHLV QSHMDAFLESSE PT-- 699

       701  PPKDP I  I N         L TVNS  VT  EMT         G L   ND       760
       712  AQEVEVRETVRKYFPETWIWDMVPLDLSGDGELPVKVPDTITEWKASAPCLSGTTGLGLS 771
       700  -------  R S         L VV SA VA VE T         G     ND       752

       761   VA FQA     V   M    I     T        L TSLPMA L  A    T   EN 820
       772  STISHKVFQPFFLELTLPYSVVRGEAFILKATVLNYMPHCIRIHVSLEMSPDFLAVPVGS 831
       753  PVVQFQA     V   M    I     T        L T   VA Q  A      A EEK 812

       821  NQ  Y LGA G H S  L       N   SVS   R   GP  SE  T  ETGR      880
       832  HEDSHCICGNERKTVSWAVTPKSLGEVNFTATAEALQSPELCGNKVAEVPALVQKDTVVK 891
       813  EQR     M Q H A    I     N    VS    N K     E PV  EQGK   II 872

       881  VL       K  H FSS   AS    S TLS L  PT  KD    HF  M    S  IK  940
       892  PVIVEPEGIEKEQTYNTLLCPQDAELQENWTLDLPANVVEGSARATQSVLGDILGSAMQN 951
       873  SLL     L N V F S    MG  VS LIA K  SD   E    SVT             932

       941  T   I             A      K          K     LG  RA    E   K K 1000
       952  LQNLLQMPYGCGEQNMVLFVPNIYVLEYLNETQQLTEAIKSKAISYLISGYQRQLNYQHS 1011
       933  T D  K             A      D         QE  T   A  NT      K R  992

       1001      A   HN-GQG            S    RAF F DES   D  T   KQ KDS     1059
       1012 DGSYSTFGDRGMRHSQGNTWLTAFVLKAFAQAQSYIYIEKTHITNAFNWLSMKQRENGCF 1071
       993       A   KPG N-HA           S    RK  F DEV   Q LL   QQ KD      1051

       1060 RS                 I       M    SS  D DP  SK  S    SS EN EQGGN 1119
       1072 QQSGSLLNNAMKGGVDDEVTLSAYITIALLEMPLPVTHSVVRNALFCLETAWASIS-NSQ 1130
       1052 RS                 E                S     P          D   K ARGGAG 1111

       1120 G F      M         QE  N I K  D E IK DN I  E  QKPTKS----EGYL 1175
       1131 ESHVYTKALLAYAFALAGNRAKRSEVLESLNKDAVNEEESVHWQRPKNVEENVREMRSFS 1190
       1112 G                  QDTKK I K  DEE  K  D    T  QKPSVSV----GLW 1167

       1176  T Q S      S   V  R  AQPAPSPE  AL MG  -----I  LT     Y       1230
       1191 YKPRAPSAEVEMTAYVLLAYLTSASSRPTRDLSSSDLTTASKIVKWISKQQNSHGGFSST 1250
       1168  Q Q T                TEPA-  QE----   A ML    LT              1222

       1231       D          S  Q TP   VQ    SF QKFQ DKS      Q S PYI    1290
       1251 QDTVVALQALSKYGAATFTKSNKEVSVTIESSGTVSGTLHVNNGNRLLLQEVRLADLPGN 1310
       1223       H      S    RAK AAQ   R     F TKFQ   N Q     R T PTV  D 1282

       1291  TVS   E    A  T R  V- LEKQQPA  A   Q V  TCNNPKGQNS   SLEI  M 1349
       1311 YITKVSGSGCVYLQTSLKYNILPEAEGEAPFTLKVNTLPLNFDKAEHHRKFQIHINVSYI 1370
       1283  TV  T E           SV  R- E F   AVV Q   GTCEDPKA TS   SL I  T 1341

       1350  S  A     A    L     L  T    ERLGHVS    T  N  L LDQV    LS   1409
       1371 GERPNSNMVIVDVKMVSGFIPVKPSVKKLQDQSNIQRTEVNTNHVLIYIEKLTNQTMGFS 1430
       1342  S SE   A A          L  T  M ERSVHVS    SN       LD VS   VNL  1401

       1410  IIQ        Q  I            V FA   S C   DQNV--               1453
       1431 FAVEQDIPVKNLKPAPVKVYDYYETDEFAIEEYSAPFSSDSEQGNA                 1476
       1402  T Q    IRD    V         K    VAK    C T Y--                  1445
```

FIGURE 3: Comparison of the deduced amino acid sequences of rat αMs. Only the amino acid residues in rat α_1I_3 (Braciak et al., 1988) and rat α_2M (Gehring et al., 1987) which differ from α_1M are shown above and below the α_1M sequence, respectively. The numbers indicate the amino acid residues of the mature proteins. Gaps inserted to optimize the alignment of the sequences were identified by the command "BESTFIT" of the GCG program (Devereux et al., 1984).

bait regions (amino acids 666–718 in α_1M) showed minimal homologies as reported previously (Sottrup-Jensen et al., 1989). The C-terminal half of α_1M (amino acids 775–1475), containing the domain which is recognized by the cellular receptor for activated αMs (Van Leuven et al., 1986; Enghild et al., 1989), exhibited 66% and 63% homology with the corresponding region in rat α_2M and α_1I_3, respectively. Of the 24 cysteines in α_1M, 23 are conserved in rat and human α_2M. For the human protein, the disulfide formations have been

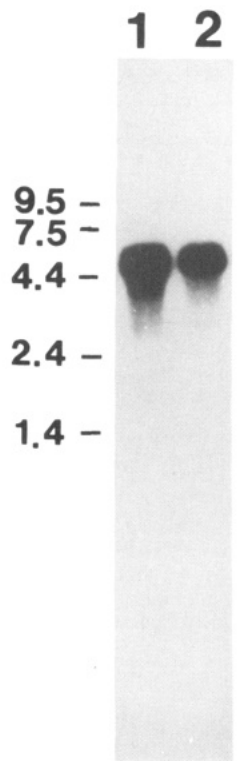


FIGURE 4: Northern blot analysis. Total RNA from rat hepatocytes (lane 1) and mouse liver (lane 2) was separated by agarose electrophoresis and hybridized with ^{32}P-labeled B fragment as described under Materials and Methods. Sizes of RNA markers (Bethesda Research Laboratory) are indicated in kilobases.

elucidated (Jensen & Sottrup-Jensen, 1986). The α_1M sequence contains one cysteine (amino acid 316) which has no counterpart in the other αMs. Its location in a narrow region together with the two cysteines postulated to be used for interchain bridging of two monomers in α_2M may suggest that the dimer units of α_1M are held together by three cysteines. The C-terminal region contains two phenylalanines (amino acids 1376 and 1490) which replace two cysteines in α_2M and α_1I_3, suggested to form an internal disulfide bridge (Jensen & Sottrup-Jensen, 1986; Braciak et al., 1988).

Northern Analysis. Northern blot analysis of rat hepatocyte RNA using the B fragment as probe resulted in the detection of one hybridizing band (Figure 4, lane 1). The α_1M mRNA had an approximate size of 6.2 kb, obviously containing more noncoding sequences than those shown in Figure 2. The probe also hybridized with high stringency to an RNA species of similar size from mouse liver RNA (Figure 4, lane 2), while no cross-hybridization was obtained with human liver RNA even under low-stringency conditions (6 × SSC/0.1% SDS, 42 °C) (not shown).

Relation of α_1M to Fibronectin Receptor. The major part of the original B clone (the *Bam*HI–*Eco*RI fragment, Figure 1), which was selected by the fibronectin receptor antibodies, was expressed as a fusion protein with glutathione *S*-transferase, after subcloning into the pGEX-2T vector. A rabbit antiserum was raised against the fusion protein and tested in immunoblotting against rat plasma. In unreduced samples, the α_1M dimer (360 kDa) was faintly detected by the antiserum (Figure 5, lane 1). Unfolding of disulfide bonds was apparently important for recognition, since the antiserum reacted strongly with the 40-kDa peptide of reduced α_1M. In addition, the antiserum showed a weak reactivity against the 140-kDa fragment of reduced α_1M (consistent with the sequence expressed in the fusion protein). The faint 190-kDa

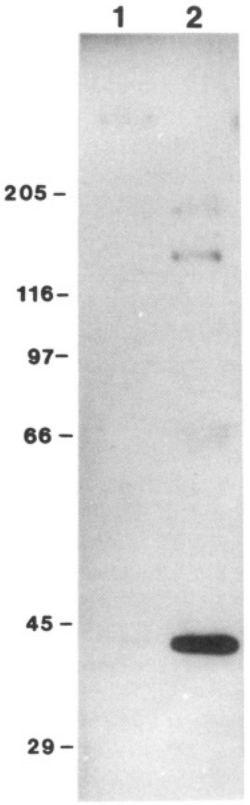


FIGURE 5: Western blot analysis. Rat plasma was subjected to SDS–PAGE in nonreduced (lane 1) and reduced form (lane 2). Immunoblotting was performed as described under Materials and Methods using the antiserum raised against a fusion protein which contained the C-terminal part of α_1M. Size markers (Sigma) are indicated in kilodaltons.

band (Figure 5, lane 2), also recognized by the antiserum, was identified as α_1I_3 on the basis of comigration with the isolated protein. Indeed, immunoblotting showed that the fusion protein antiserum had a weak cross-reactivity with α_1I_3 (not shown). Several preparations of pure fibronectin receptor and a crude fraction used for receptor isolation (wheat germ agglutinin binding proteins of solubilized hepatocytes) (Johansson et al., 1987) were also analyzed in immunoblotting. However, no immunoreacting material was detected in any of the samples either before or after reduction (not shown).

Discussion

While three structurally related plasma proteins of the αM family are expressed in rats (α_2M, α_1M, and α_1I_3), only two have been identified in humans (α_2M and PZP). The relations between the rat and human proteins have not been fully clear in lack of complete sequence data for all proteins. PZP has been suggested to be the human counterpart of rat α_1M (Sand et al., 1985). On the other hand, the expression of both human α_2M and rat α_1M is only moderately affected during acute phase reactions, whereas there is a dramatic up-regulation of rat α_2M biosynthesis by the same stimuli (Gehring et al., 1987; Lonberg-Holm et al., 1987). The recently reported cDNA sequence for PZP (Devriendt et al., 1991) and our α_1M sequence now allow direct comparisons of the primary structure of all five proteins.

Such an analysis revealed that the proteins are highly homologous but suggests that neither of the human proteins has a direct equivalent among the three rat proteins. Of the rat αMs, both PZP and human α_2M show the highest similarity to α_2M (65% and 74% identical amino acids, respectively). Both human proteins were also more homologous to α_1I_3 (59% and 60%) than to rat α_1M (55% and 59%). Since PZP, human α_2M, and rat α_2M are all expressed during strikingly different physiological conditions, they may have evolved to meet distinct functional requirements.

The specific functions of α_1M are poorly understood. The protein has been shown to act as an inhibitor of a broad spectrum of proteases and would seem to have considerable overlapping capability with α_2M and α_1I_3, although the markedly dissimilar bait regions have been suggested to provide the three proteins with somewhat different specificities (Kan et al., 1985; Gehring et al., 1987; Braciak et al., 1988; Sottrup-Jensen et al., 1989). However, in addition to regulation of extracellular proteolysis, αMs may have other functions. For example, α_2M has been reported to bind several growth factors and hormones (Huang et al., 1985; O'Connor-McCourt & Wakefied, 1987; Matsuda et al., 1989; Borth & Luger, 1989; Danielpour & Sporn, 1990), possibly mediating endocytosis of these compounds via the αM receptor. Similarly, α_1I_3 has been discussed as a "general clearance protein" which would opsonize damage proteins to facilitate their elimination by endocytosis (Braciak et al., 1988). The basis for the latter suggestion was three cases where antibody screening of λgt11 libraries resulted in the isolation of clones containing α_1I_3 cDNA in addition to the anticipated cDNA clones. Thus, our isolation of an α_1M cDNA clone by use of antibodies directed against the fibronectin receptor (integrin $\alpha_5\beta_1$) seemed to follow an analogous pattern.

To test the hypothesis (Braciak et al., 1988) that immunogenic fragments derived from activated αMs may become associated with various proteins to mediate recognition by clearance receptors, we raised an antiserum against a fusion protein corresponding to the α_1M cDNA of the λ clone isolated by screening with the fibronectin receptor antiserum. The fusion protein antiserum, which was found to react strongly with the C-terminal part of α_1M (the αM–receptor binding region), did not detect any immunoreactive material in our preparations of fibronectin receptor. These results do not exclude the possibility that trace amounts of α_1M-derived fragments were associated with a small fraction of receptor molecules and that it could have elicited antibody production during the immunization. However, since the antibodies used in screening of the liver library had been affinity-purified on fibronectin receptor-Sepharose, an alternative explanation may be that these two large proteins actually share some cross-reacting epitope even though no obvious homology in their primary structure was found.

Acknowledgments

A sample of α_1I_3 was kindly provided by Dr. Bo Åkerström, University of Lund, Sweden.

References

Blobel, G., & Dobberstein, B. (1975) *J. Cell. Biol. 67*, 835–851.

Borth, W., & Luger, T. A. (1989) *J. Biol. Chem. 264*, 5818–5825.

Braciak, T. A., Northemann, W., Hudson, G. O., Shiels, B. R., Gehring, M. R., & Fey, G. H. (1988) *J. Biol. Chem. 263*, 3999–4012.

Burnette, W. N. (1981) *Anal. Biochem. 112*, 195–203.

Danielpour, D., & Sporn, M. B. (1990) *J. Biol. Chem. 265*, 6973–6977.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acid Res. 12*, 387–395.

Devriendt, K., Van den Berghe, H., Cassiman, J.-J., & Marynen, P. (1991) *Biochim. Biophys. Acta 1088*, 95–103.

Enghild, J. J., Thørgersen, I. B., Roche, P. A., & Pizzo, S. V. (1989) *Biochemistry 28*, 1406-1412.
Enghild, J. J., Thørgersen, I. B., Salvesen, G., Fey, G. H., Figler, N. L., Gonias, S. L., & Pizzo, S. V. (1990) *Biochemistry 29*, 10070-10080.
Forsberg, E., Paulsson, M., Timpl, R., & Johansson, S. (1990) *J. Biol. Chem. 265*, 6376-6381.
Gehring, M. R., Shiels, B. R., Northemann, W., de Bruijn, M. H. L., Kan, C.-C., Chain, A. C., Noonan, D. J., & Fey, G. H. (1987) *J. Biol. Chem. 262*, 446-454.
Huang, J. S., Huang, S. S., & Deuel, T. F. (1985) *Proc. Natl. Acad. Sci. U.S.A. 81*, 342-346.
Jensen, P. E. H., & Sottrup-Jensen, L. (1986) *J. Biol. Chem. 261*, 15863-15869.
Johansson, S., Forsberg, E., & Lundgren, B. (1987) *J. Biol. Chem. 262*, 7819-7824.
Kan, C.-C., Solomon, E., Belt, K. T., Chain, A. C., Hiorns, L. R., & Fey, G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2282-2286.
Lonberg-Holm, K., Reed, D. L., Roberts, R. C., Hebert, R. R., Hillman, M. C., & Kutney, R. M. (1987) *J. Biol. Chem. 262*, 438-445.
Matsuda, T., Hirano, T., Nagasawa, S., & Kishimoto, T. (1989) *J. Immunol. 142*, 148-152.
Moestrup, S. K., & Gliemann, J. (1989) *J. Biol. Chem. 264*, 15574-15577.
O'Connor-McCourt, M. D., & Wakefield, L. M. (1987) *J. Biol. Chem. 262*, 14090-14099.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning: a laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sand, O., Folkersen, J., Westergaard, J. G., & Sottrup-Jensen, L. (1985) *J. Biol. Chem. 260*, 15723-15735.
Smith, D. B., & Johnson, K. S. (1988) *Gene 67*, 31-40.
Sottrup-Jensen, L. (1987) in *The Plasma Proteins. Structure, Function and Genetic Control* (Putman, F. W., Ed.) Vol. V, pp 191-291, Academic Press, Orlando, FL.
Sottrup-Jensen, L. (1989) *J. Biol. Chem. 264*, 11539-11542.
Sottrup-Jensen, L., Sand, O., Kristensen, L., & Fey, G. H. (1989) *J. Biol. Chem. 264*, 15781-15789.
Strickland, D. K., Ashcom, J. D., Williams, S. S., Burgess, W. H., Migliorini, M., & Argraves, W. S. (1990) *J. Biol. Chem. 265*, 17401-17404.
Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767-4771.
Van Leuven, F., Cassiman, J.-J., & Van den Berghe, H. (1979) *J. Biol. Chem. 254*, 5155-5160.
Van Leuven, F., Marynen, P., Sottrup-Jensen, L., Cassiman, J.-J., & Van den Berghe, H. (1986) *J. Biol. Chem. 261*, 11369-11373.
Yuong, R. A., & Davis, R. W. (1983) *Science 222*, 778-782.

Alternative Splicing of the mRNA Encoding the Human Cholesteryl Ester Transfer Protein†

Akihiro Inazu,‡ Elaine M. Quinet, Suke Wang, Maryanne L. Brown, Susan Stevenson, Mark L. Barr, Philippe Moulin, and Alan R. Tall*

Division of Molecular Medicine, Department of Medicine, Columbia University College of Physicians and Surgeons, New York, New York 10032

Received October 17, 1991; Revised Manuscript Received December 3, 1991

ABSTRACT: The plasma cholesteryl ester transfer protein (CETP) is known to facilitate the transfer of lipids between plasma lipoproteins. The human CETP gene is a complex locus encompassing 16 exons. The CETP mRNA is found in liver and small intestine as well as in a variety of peripheral tissues. While the CETP cDNA from human adipose tissue was being cloned, a variant CETP cDNA was discovered which excluded the complete sequence encoded by exon 9, but which was otherwise identical to the full-length CETP cDNA, suggesting modification of the CETP gene transcript by an alternative RNA splicing mechanism. RNase protection analysis of tissue RNA confirmed the presence of exon 9 deleted transcripts and showed that they represented a variable proportion of the total CETP mRNA in various human tissues including adipose tissue (25%), liver (33%), and spleen (46%). Transient expression of the exon 9 deleted cDNA in COS cells or stable expression in CHO cells showed that the protein encoded by the alternatively spliced transcript was inactive in neutral lipid transfer, smaller, and poorly secreted compared to the protein derived from the full-length cDNA. Endo H digestion suggested that the inactive, cell-associated protein was present within the endoplasmic reticulum. The experiments show that the expression of the human CETP gene is modified by alternative splicing of the ninth exon, in a tissue-specific fashion. The function of alternative splicing is unknown but could serve to produce a protein with a function other than plasma neutral lipid transfer, or as an on-off switch to regulate the local concentration of biologically active protein.

The plasma cholesteryl ester transfer protein (CETP)[1] is a hydrophobic glycoprotein, which facilitates neutral lipid and phospholipid transfer between the plasma lipoproteins (Tall, 1986; Hesler et al., 1987). The CETP appears to be a member

† This work was supported by NIH Grant HL43165.
* Address correspondence to this author.
‡ Present address: Division of Molecular Medicine and Lipoprotein Research, The Second Department of Internal Medicine, School of Medicine, Kanazawa University, Takara Machi 13-1, Kanazawa, 920, Japan.

[1] Abbreviations: CETP, cholesteryl ester transfer protein; bp, base pair(s); PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; HDL, high-density lipoprotein(s); LDL, low-density lipoprotein(s); VLDL, very-low-density lipoprotein(s); PMSF, phenylmethanesulfonyl fluoride.

0006-2960/92/0431-2352$03.00/0