Biochemistry

 Volume 28, Number 23 November 14, 1989

Perspectives in Biochemistry

Implications of the Three-Dimensional Structure of α_1-Antitrypsin for Structure and Function of Serpins

R. Huber*,‡ and R. W. Carrell§

Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG, and Department of Haematology, University of Cambridge MRC Centre, Cambridge CB2 2QH, U.K.

Received June 6, 1989

There is now much interest in a newly recognized superfamily of proteins, the serpins (Carrell & Travis, 1985; Carrell et al., 1987a,b). More than 40 members of the family have been identified in viruses and plants as well as higher organisms. The serpins have developed by divergent evolution over a period of some 500 million years (Hunt & Dayhoff, 1980), most of the members retaining the presumed function of the original ancestral protein as serine proteinase inhibitors. Some, however, have lost this function and developed specialized roles as carriers of lipophilic molecules (thyroxine- and cortisol-binding globulins) or as peptide hormone precursors (angiotensinogen) or have no recognized function (ovalbumin).

The best studied members are those in human plasma where there is a diversity of inhibitory specialization that illustrates the way in which the serpins have evolved in parallel with their cognate proteases: antithrombin with thrombin, C_1-inhibitor with C_1-esterase, antiplasmin with plasmin, and so on. A key plasma serpin is α_1-antitrypsin; this is an efficient inhibitor of trypsin, but its prime physiological role is as an inhibitor of the elastase released by leukocytes. Interest focused on α_1-antitrypsin because its common genetic deficiency is associated with the development of premature lung degeneration (Laurell & Eriksson, 1963).

The establishment of α_1-antitrypsin as the archetype of the serpins was strengthened by the determination of its crystallographic structure, in a modified form, by Löbermann et al. (1984). It seems timely now to show how it can act as a general template for the other serpins. We look here at the common structural features of the family: the location of insertions and deletions and their compatibility with the three-dimensional template, the conserved amino acid residues and their relevance for the integrity of the spatial structure, the location of cysteine residues and disulfide bridges, and the sites of glycosylation. We well also discuss binding sites of functional modulators of some serpins (i.e., heparin in antithrombin) and ligand binding sites in serpins with carrier function. Finally, we demonstrate the overall validity of the α_1-antitrypsin model by showing how it provides a general explanation of the molecular pathology associated with diverse variants of the human serpins.

Molecular Structure of α_1-Antitrypsin

Fortunately, α_1-antitrypsin has turned out to be a typical member of the serpin family (Carrell et al., 1982). It is a glycoprotein of 394 residues with MW 51 000 and functions by forming a tight complex with its target protease. The serpins are believed to function as ideal substrates with association rates of the order of 10^4 M^{-1} s^{-1} or more and negligible dissociation rates (Travis & Salvesen, 1983). In particular, the sequence at the reactive center helps define specificity by providing a putative cleavage site for the target proteinase. Thus, the methionine 358 reactive center residue of α_1-antitrypsin provides a cleavage site of choice for leukocyte elastase, whereas in antithrombin the homologously aligned reactive center arginine 393 provides a specific cleavage site for thrombin. The critical role of the reactive center residues was highlighted by the finding of a pathological variant of α_1-antitrypsin in a child with a bleeding disorder in which methionine 358 had been substituted by an arginine, thus converting the protein from an inhibitor of elastase to a highly effective inhibitor of thrombin (Owen et al., 1983).

The molecular structure of α_1-antitrypsin as reported by Löbermann et al. (1984) is based on that of the cleaved molecule subsequent to release from the complex with chymotrypsin. It crystallizes in three different crystal forms which have been analyzed and found to be based on very similar molecular structures (Löbermann et al., 1984; Engh et al., 1989). The surprising feature of the structure was the separation of methionine and serine at the cleaved 358–359 reactive

‡ Max-Planck-Institut für Biochemie.
§ University of Cambridge MRC Centre.

0006-2960/89/0428-8951$01.50/0

Table I: Secondary Structural Elements in α_1-Antitrypsin[a]

helixes	sheets	turns	bulges
hA: 20–44 (kink at 28 Pro)	s6B: 49–53	thAs6B: 45–48	169–172
	s5B: 380–389	thBhC: 68–70	171–174
hB: 53–68	s4B: 369–378	thChD: 81–88 (lh: 81)	173–176 (series of overlapping bulges)
hC: 69–81	s3B: 247–255		
hC1: 83–87	s2B: 236–245	thDs2A: 105–110	
hD: 88–105	s1B: 228–233	ts2AhE: 122–127	
hE: 127–139	s6A: 290–299	thEs1A: 139–140 (lh: 139)	bs5B: 382–385
hF: 149–166	s5A: 326–342		
hF1: 200–203 (one open turn)	s4A: 343–356	ts1AhF: 146–149	bs5A: 329–332
	s3A: 181–194	thFs3A: 166–181 (lh: 166) (series of bulges)	
hF2: 232–236 (one open turn)	s2A: 109–121		
	s1A: 140–146		
hG: 259–264	s4C: 203–212	ts3AhF1: 194–199	
hH: 268–278	s3C: 213–226	ts4Cs3C: 211–214	
hI: 299–306	s2C: 283–289	ts3Cs1B: 226–228	
hI1: 309–312 (one open turn)	s1C: 362–367	ts1Bs2B: 233–236 (lh: 236)	
hI2: 376–380 (one open turn)		ts2Bs3B: 244–248	
		ts3BhG: 256–259	
hI3: 390–393 (one open turn)		thHs2C: 278–283	
		thIIs5A: 318–325	
		ts5As4A: 341–344	
		ts4Bs5B: 377–380 (lh: 380)	
		ts5Bc-ter: 389–394	

[a] Residues at the termini of helixes are included if at least one of their main-chain conformational angles is canonical; strands of sheets are defined similarly; appropriate hydrogen bonds are not always made by these residues. hX, helix X; sXY, strand X in sheet Y; thXhY, turn between helix X and helix Y; bsXY, bulge in strand X of sheet Y; lh, left-handed helical conformation.

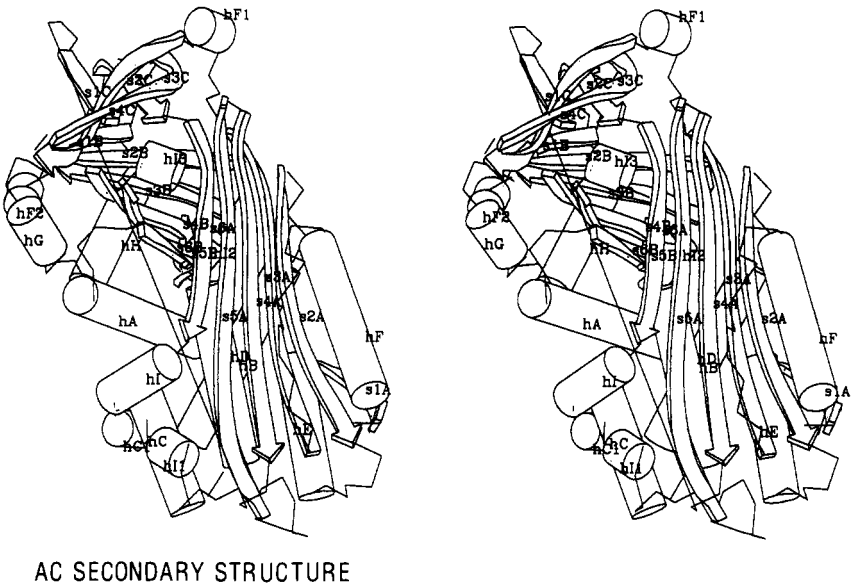

FIGURE 1: Polypeptide chain folding of α_1-antitrypsin with the secondary structure elements represented by arrows (sheet strands) and cylinders (helixes) and marked according to Table II. Residue numbers refer to α_1-antitrypsin and the alignment in Table II. Residues C-terminal to 358 have 100 added to their numbers in the plot.

center to opposite poles of the molecule with a distance of 69 Å. The significance of this change and the reconstruction necessary to give the native, uncleaved, protein are discussed in a subsequent section.

Table I gives a concise account of the secondary structural elements. The structure as a whole is remarkably ordered, with 80% of the amino acids in eight well-defined helixes (A–H) and three large β-sheets (A–C), as shown in Figure 1.

The α-helixes are regular, but hA, hB, and hC have 3_{10} geometry at their N- and C-termini. These helixes are fully or partly buried. The short helixes hF2 and hI2 are helical turns between β-strands. The β-sheets are antiparallel, except for short parallel strands added to sheet A (s1A) and C (s1C). They are regular, and their strands are only twice interrupted by bulges bs5B and bs5A. Because of these bulges, a proline residue, P382,[1] is accommodated in s5B and a lysine residue, K331, is turned to the surface of s5A, respectively.

The segment 169–176, which is antiparallel to α-helix hF, forms a sequence of bulges, by which the peculiar series of apolar and polar residues D171, L172, V173, K174, E175, and L176 are appropriately positioned internally and externally, respectively. It resembles the sequence of reverse turns in lysozyme (residues 17–22) but lacks the $O_j \rightarrow N_{j+3}$ hydrogen bonds characteristic for type I and II turns. The conserved N158 in hF makes hydrogen bonds to the main chain of V173 which may be crucial for the stabilization of this segment. The

[1] The one-letter code for amino acids is used. The amino acid numbers are based on the α_1-antitrypsin sequence and the alignment of Table II. When two numbers are given, the first refers to the protein discussed and the second to the homologous residue in α_1-antitrypsin.

turns are β-hairpins (ts4Cs3C, ts1Bs2B, ts2Bs3B, ts5As4A, ts4Bs5B), other sharp turns (ts1AhF, ts3BhG, thEs1A, thFs3A, thDs2B, thI1s5A), or wide excursions (ts2AhE, thFs3A, ts3AhF1, thHs2C). Of the β-hairpins only ts1Bs2B and ts5As4A have a regular $O_j \rightarrow N_{j+3}$ hydrogen bond.

Consistently in Ramachandran plots of all α_1-antitrypsin crystal structures, six non-glycine residues are outside the favorable region of conformational space, of which five have the slightly unfavorable left-handed α-helical conformation. This conformation is infrequent but clearly established in some proteins where it occurs in turns (Huber et al., 1987). Also, in α_1-antitrypsin the left-handed helical residues mediate turns: N81, at the end of hC, abruptly changes the chain direction and buries L80 and F82 in the interior of the protein. H139 is at the C-terminus of hE and allows an abrupt change of chain direction into s1A. Similarly, Q166 ends hF and leads into the antiparallel segment and series of bulges 169–176. It is adjacent to invariant T165. S236 ends the helical turn hF2 to lead into s2B. K380 ends the helical turn hI2 and leads into s5B so that the side chain of K380 projects to the surface.

A70 is in a high-energy conformation in all crystal forms. It is located in the turn thBhC and well defined. The carbonyl oxygen of the preceding K69 is hydrogen bonded to O^{γ} of T72 and to N of H73. These interactions enforce a strained conformation of A70, as observed similarly in other proteins [see, e.g., Bode et al. (1989)]. Eight out of 18 sequences in Table II have A or G at this position.

In the completely buried and apolar segments (hB, s4B, s5B, s6B), which are strongly conserved within the serpin superfamily, all hydrogen bonds are satisfied, mainly by interactions of main-chain atoms. Some irregularities in secondary structure occur where neutral polar amino acid side chains participate in hydrogen bonds to main-chain atoms: N49 by its O^{δ_1} and N^{δ_2} side-chain atoms forms hydrogen bonds to main-chain O and N atoms of N390, to N of V389, and to N^{ϵ_2} of Q393. E376 is at the N-terminus of hI2 and forms hydrogen bonds to the main-chain N atoms of N378 and T379 and also balances the charge of the helical dipole. N158 is hydrogen bonded by N^{δ_2} and O^{δ_1} to the main-chain N and O atoms of 173 and anchors hF to the series of bulges. O^{γ} of S53 initiates hB and is hydrogen bonded to N of S56. It also stabilizes bs5B by bonding to O of L383. T165 O^{γ} is hydrogen bonded to the main-chain O atom of V161 and N atom of I169 and acts as a clamp for the helical turn thFs3A. N186 is hydrogen bonded by its side chain to O and O^{γ} of S56 and to O^{γ} of T59. W194 has its N^{ϵ} hydrogen bonded to the main-chain O of E341. Y297 O^{ϵ} is hydrogen bonded to N of 51Phe. H334 has its N^{δ} hydrogen bonded to O of A350 and its N^{ϵ} to N of S53. These internal and polar residues are rather conserved or conservatively replaced in the serpin superfamily.

Family Relationships

Table II shows the alignment of 20 members of the superfamily modified from that of Bock et al. (1986) and Carrell et al. (1987a,b) and reevaluated on the basis of three-dimensional modeling using FRODO (Jones, 1978). Some 20 other serpins are not included because of either incomplete sequences or their recent elucidation as with the vaccinia virus serpins [e.g., Kotwal and Moss (1989)].

There is considerable variation in length due to long additions at the N- (C1 inhibitor) and C-termini (antiplasmin) and a 33-residue insertion between hC and hD (placental PAI). Other species have shorter insertions or deletions in the same area. Both N- and C-termini of modified α_1-antitrypsin protrude into solution so that insertions can be accommodated. The shortest species begin with residue 23 and end with residue 391, which are indeed the first and the last residues buried in the globular structure. hC and hD are connected by a wide loop including short helical segment hC1 so that long insertions can be incorporated without conflict with the globular fold of the molecule. Also, deletions between hC and hD are possible by eliminating hC1 as in endothelial PAI and antiplasmin. Short insertions or deletions occur in various members of the superfamily usually in turns between secondary structural elements: One- or two-residue insertions or deletions occur in thAs6B, a surface turn and also a carbohydrate attachment site. Up to three-residue deletions or two-residue insertions occur in thDs2A where they can be accommodated without affecting the helix and sheet elements. Similarly, two-residue insertions are found in ts1AhF where they can be accommodated between the series of bulges (ending with L176) and the start of s3A (T181). One- or two-residue insertions occur in ts3Cs1B. Up to five-residue insertions are found in ts3BhG (256–259). P255 and L260 are strongly conserved and in contact. It is likely that the insertions are arranged as a protruding loop. A two-residue insertion occurs in thHs2C, which is a wide exposed loop. It is unlikely to affect the adjacent secondary structure A one-residue insertion occurs between s6A and hI and can be accommodated by extending hI by one residue. A two-residue insertion is found near A325 at the beginning of s5A, which is exposed. The variations in lengths near the site of specific cleavage (358–359) are likely to be tolerable because it must be exposed, as discussed later.

Heparin cofactor II, antithrombin III, and C_1-inhibitor have extensions of 81, 25, and 93 residues at their N-termini, respectively. In the latter two cases they are linked by disulfide bonds to the rest of the molecule so that their general locations are defined and will be discussed later.

In Table III we note conserved residues and indicate their locations and environments to provide clues as to their importance for the structural integrity. All conserved residues are internal or in surface niches. They are clustered. Four segments, hB, s3A, s4B, and s5B, have more than four conserved residues. hB, s4B, and s5B constitute the core of the molecule, while s3A represents the spine (Figure 2).

Disulfide Linkages

Disulfide bonds absent in α_1-antitrypsin but present in family members provide direct evidence for structural similarity, but very few of them have been chemically defined. Two occur in conserved parts: C216–C392 in antithrombin III replacing V and T in α_1-antitrypsin (Petersen et al., 1979) and C27–C381 in placental PAI which replace T and S (U. Kiso, A. Henschen, I. Leander, and B. Astedt, unpublished results). Figure 3 shows that these disulfide bonds can be formed by substitution in the α_1-antitrypsin structure without significant structural perturbations. This observation is relevant also with regard to the structural change upon limited proteolysis.

Other disulfide linkages are in areas with deletions or insertions: C87h–C133 in ovalbumin. The C^{α} atoms of residues 90 and 133 in α_1-antitrypsin are 11 Å apart so that the disulfide bond can be easily made if the nine-residue insertion between hC and hD is in an appropriate conformation, perhaps extending hC1. Defined cystine residues are also in antithrombin III and C_1-inhibitor in the long N-terminal extensions, which they link to the globular domain. They serve to define the general course of the N-termini in these molecules as shown in Figure 3. The N-terminus of C_1-inhibitor is longer by 93 residues compared to that of α_1-antitrypsin. It is constrained from residue 9 on by the disulfide bonds 9–319 and 14b–88. In the α_1-antitrypsin structure, H20 is the first spatially defined residue. Its distance to residue 88 is 16.5 Å,

Table II: Alignment of Amino Acid Sequences of 20 Members of the Serpin Family[a]

```
                                                                    -90
C1-INHIBITOR          5.                                            NPNATSS
                            -80       -70       -60       -50       -40
secondary structure (Table 1)
ANTITRYPSIN HUMAN     1.
ANTITRYPSIN BABOON    2.
HEPARIN COFAC II      3.     GSKGPLDQLEKGGETAQSADPQWEQLNNKNLSMPLLPADFHKENTV
ANTITHROMBIN          4.                              mysnvigtvtsgkrkvyllsll
C1-INHIBITOR          5. SSQDPESLQDRGEGKVATTVISKMLFVEPILEVSSLPTTNSTTNSATKIT
ENDOTHELIAL PAI       6.
THYROXINE B G         7.
ANTICHYMO MOUSE       8.
ANTICHYMO HUMAN       9.
ANTIPLASMIN          10.
OVALBUMIN            11.
GENE Y PROTEIN       12.
PLACENTAL PAI        13.
ANGIOTENS RAT        14.              mtptgaglkatifciltwvsltagDRVYIHPFHLLY
ANGIOTENS HUMAN      15.      mrkrapqsemapagvslratilcllawaglaagDRVYIHPFHLVI
BARLEY Z PROTEIN     16.
CORTICOSTEROID BG    17.
PROTEIN C INHIB.     18.
PROTEASE NEXIN       19.
RAB ORF1             20.
                            -30       -20       -10       -11        10
secondary structure
ANTITRYPSIN HUMAN     1.            mpssvswgilllaglcclvpvslaEDPQGDAAQKTDTS-
ANTITRYPSIN BABOON    2.                        lllaglccllpgslaEDPQGDAAQKTDTP-
HEPARIN COFAC II      3. TNDWIPEGEEDDDYLDLEKIFSEDDDYIDIVDSLSVSPTDSDVSAGNIL-
ANTITHROMBIN          4. ligfwdcvtcHGSPVDICTAKPRDIPMNPMCIYRSPEKKATEDEGSEQK-
C1-INHIBITOR          5. ANTTDEPTTQPTTEPTTQPTIQPTQPTTQLPTDSPTQPTTGSFCPGPVTL
ENDOTHELIAL PAI       6.                                  mqmspaltclvlglalvf
THYROXINE B G         7.                           mspflylvllvlglhatihcASPEGKVTA
ANTICHYMO MOUSE       8.
ANTICHYMO HUMAN       9.              mermlpllalgllaagfcpavlchpNSPLDEENLTQE-
ANTIPLASMIN          10.                       NQEQVSPLTLLKLGNQEPGGQTALKSPPGV
OVALBUMIN            11.
GENE Y PROTEIN       12.
PLACENTAL PAI        13.
ANGIOTENS RAT        14. YSKSTCAQLENPSVETLPEPTFEPVPIQAKTSPVDEKTLRDKLVLATEK-
ANGIOTENS HUMAN      15. HNESTCEQLAKANAGKPKDPTFIPAPIQAKTSPVDEKALQDQLVLVAAK-
BARLEY Z PROTEIN     16.
CORTICOSTEROID BG    17.                         mplllytcllwlptsglwtvqaMDPNA-
PROTEIN C INHIB.     18.                       mqlfllclvllspqgaslHRHHPREMKK-
PROTEASE NEXIN       19.                                    MNWHLPLFLLASVTL-
RAB ORF1             20.
                           20        30        40        50        60
secondary structure         ^^^^^^^^^^^^hA^^^^^^^^^^^      -s6B-^^^^^^hB^
ANTITRYPSIN HUMAN     1. -HHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNS-TNI-FFSPVSIATAF
ANTITRYPSIN BABOON    2. -PHDQNHPTLNKITPSLAEFAFSLYRQLAHQSNS-TNI-FFSPVSIATAF
HEPARIN COFAC II      3. -QLFHGKSRIQRLNILNAKFAFNLYRVLKDQVNTFDNI-FIAPVGISTAM
ANTITHROMBIN          4. -IPEATNRRVWELSKANSRFATTFYQHLADSKNDNDNI-FLSPLSISTAF
C1-INHIBITOR          5. CSDLESHSTEAVLGDALVDFSLKLYHAFSAMKKVETNM-AFSPFSIASLL
ENDOTHELIAL PAI       6. gegsaVHHPPSYVAHLASDFGVRVFQQVAQASKD-RNV-VFSPYGVASVL
THYROXINE B G         7. CHSSQPNATLYKMSSINADFAFNLYRRFTVETPD-KNI-FFSPVSISAAL
ANTICHYMO MOUSE       8.
ANTICHYMO HUMAN       9. -NQDRGTHVDLGLASANVDFAFSLYKQLVLKALD-KNV-IFSPLSISTAL
ANTIPLASMIN          10. CSRDPTPEQTHRLARAMMAFTADLFSLVAQTS-TCPNL-ILSPLSVALAL
OVALBUMIN            11.          MGSIGAASMEFCFDVFKELKVHHAN-ENI-FYCPIAIMSAL
GENE Y PROTEIN       12.          MDSISVTNAKFCFDVFNEMKVHHVN-ENI-LYCPLSILTAL
PLACENTAL PAI        13.          MEDLCVANTLFALNLFKHLAKASPT-QNL-FLSPWSISSTM
ANGIOTENS RAT        14. -LEAEDRQRAAQVAMIANFMGFRMYKMLSEARGVASGA-VLSPPALFGTL
ANGIOTENS HUMAN      15. -LDTEDKLRAAMVGMLANFLGFRIYGMHSELWGVVHGATVLSPTAVFGTL
BARLEY Z PROTEIN     16.
CORTICOSTEROID BG    17. -AYVNMSNHHRGLASANVDFAFSLYKHLVALSPK-KNI-FISPVSISMAL
PROTEIN C INHIB.     18. -RVEDLHVGATVAPSSRRDFTFDLYRALASAAPS-QNI-FFSPVSISMSL
PROTEASE NEXIN       19. -PSICSHFNPLSLEELGSNTGIQVFNQIVKSRP-HDNI-VISPHGIASVL
RAB ORF1             20.   MKYLVLVLCLTSCACRDIGLWTFRYVYNES---DNV-VFSPYGLTSAL
                                 70        80
secondary structure         ^^^^^^^^^^^^hC^^^^^. ^ hC1^
ANTITRYPSIN HUMAN     1. AMLSLGTKADTHDEILEGLNFN-LTEI----------------------
ANTITRYPSIN BABOON    2. AMLSLGTKADTHSEILEGLNFN-LTEI----------------------
HEPARIN COFAC II      3. GMISLGLKGETHEQVHSILHFKDFVNA----------------------
ANTITHROMBIN          4. AMTKLGACNDTLQQLMEVFKFDTISEK----------------------
C1-INHIBITOR          5. TQVLLGAGQNT-KTNLESILSYPKDFT----------------------
ENDOTHELIAL PAI       6. AMLQLTTGGETQQQIQAAMGFK----------------------------
THYROXINE B G         7. VMLSFGACCSTQTEIVETLGFN-LTDT----------------------
ANTICHYMO MOUSE       8.
ANTICHYMO HUMAN       9. AFLSLGAHNTTLTEILKASSSP-HGDL----------------------
ANTIPLASMIN          10. SHLALGAQNHTLQRLQQVLH-----------------------------
OVALBUMIN            11. AMVYLGAKDSTRTQINKVVRFDKLPGF----------------------
GENE Y PROTEIN       12. AMVYLGARGNTESQMKKVLHFDSITGA----------------------
PLACENTAL PAI        13. AMVYMGSRGSTEDQMAKVLQF-NEVGANAVTPMTPENFTSCGFMQQIQKG
ANGIOTENS RAT        14. VSFYLGSLDPTASQLQVLLGVPVKEGD----------------------
ANGIOTENS HUMAN      15. ASLYLGALDHTADRLQAILGVPWKDKN----------------------
BARLEY Z PROTEIN     16.
CORTICOSTEROID BG    17. AMLSLGTCGHTRAQLLQGLGFN-LTER----------------------
PROTEIN C INHIB.     18. AMLSLGAGSSTKMQILEGLGLN-LQKS----------------------
PROTEASE NEXIN       19. GMLQLGADGRTKKQLAMVMRY----------------------------
RAB ORF1             20. SVLRIAAGGNTKREIDVPESVVED-------------------------
```

Table II (Continued)

```
                                        90        100         110        120
secondary structure                      ^^^^^^^hD^^^^^^^^^    -  ----s2A-----
ANTITRYPSIN HUMAN  1. ----------PEAQIHEGFQELLRTLNQPDSQ--LQLTTDGGLFLSEGLK
ANTITRYPSIN BABOON 2. ----------PEAQVHEGFQELLRTLNKPDSQ--LQLTTGNGLFLNKSLK
HEPARIN COFAC II   3. ------SSKYEITTIHNLFRKLTHRLFRRNFG--YTLRSVNDLYIQKQFP
ANTITHROMBIN       4. ----------TSDQIHFFFAKLNCRLYRKANK-SSKLVSANRLFGDKSLT
C1-INHIBITOR       5. ----------CVHQALKGFTTKG-------------VTSVSQIFHSPDLA
ENDOTHELIAL PAI    6. ---------IDDKGMAPALRHLYKELMGPWNK--DEISTTDAIFVQRDLK
THYROXINE B G      7. ----------PMVEIQHGFQHLICSLNFPKKE--LELQIGNALFIGKHLK
ANTICHYMO MOUSE    8.
ANTICHYMO HUMAN    9. ----------LRQKFTQSFQHLRAPSISSSDE--LQLSMGNAMFVKEQLS
ANTIPLASMIN       10. --------AGSGPCLPHLLSRLCQDLG-PG-----AFRLAARMYLQKGFP
OVALBUMIN         11. -GDSIEAQCGTSVNVHSSLRDILNQITKPND--VYSFSLASRLYAEERYP
GENE Y PROTEIN    12. -GSTTDSQCGSSEYVHNLFKELLSEITRPNA--TYSLEIADKLYVDKTFS
PLACENTAL PAI     13. SYPDAILQAQAADKIHSSFRSLSSAINASTGD--YLLESVNKLFGEKSAS
ANGIOTENS RAT     14. -CTSRLDGH-KVLTALQAVQGLLVTQGGSSSQTPLLQSTVVGLFTAPGLR
ANGIOTENS HUMAN   15. -CTSRLDAH-KVLSALQAVQGLLVAQGRADSQAQLLLSTVVGVFTAPGLH
BARLEY Z PROTEIN  16.
CORTICOSTEROID BG 17. ----------SETEIHQGFQHLHQLFAKSDTS--LEMTMGNALFLDGSLE
PROTEIN C INHIB.  18. ----------SEKELHRGFQQLLQELNQPRDG--FQLSLGNALFTDLVVD
PROTEASE NEXIN    19. ----------GVNGVGKILKKINKAIVSKKNK--DIVTVANAVFVKNASE
RAB ORF1          20. ------------------------------S--DAFLALRELFVDASVP

                        130       140         150       160       170
secondary structure      ^^^^^^hE^^^^^-- s1A--  ^ ^^^^^^hF^^^^^^^^^
ANTITRYPSIN HUMAN  1. LVDKFLEDVKKLYHSE-AFTVNFGD-TEEAKKQINDYVEKGTQGKIVDLV
ANTITRYPSIN BABOON 2. VVDKFLEDVKNLYHSE-AFSVNFED-TEEAKKQINNYVEKGTQGKVVDLV
HEPARIN COFAC II   3. ILLDFKTKVREYYFAE-AQIADFSD--PAFISKTNNHIMKLTKGLIKDAL
ANTITHROMBIN       4. FNETYQDISELVYGAK-LQPLDFKENAEQSRAAINKWVSNKTEGRITDVI
C1-INHIBITOR       5. IRDTFVNASRTLYSSS-PRVLSNN--SDANLELINTWVAKNTNNKISRLL
ENDOTHELIAL PAI    6. LVQGFMPHFFRLFRST-VKQVDFSE-VERARFIINDWVKTHTKGMISNLL
THYROXINE B G      7. PLAKFLNDVKTLYETE-VFSTDFSN-ISAAKQEINSHVEMQTKGKVVGLI
ANTICHYMO MOUSE    8.
ANTICHYMO HUMAN    9. LLDRFTEDAKRLYGSE-AFATDFQD-SAAAKKLINDYVKNGTRGKITDLI
ANTIPLASMIN       10. IKEDFLEQSEQLFGAK-PVSLT--GKQEDDLANINQWVKEATEGKIQEFL
OVALBUMIN         11. ILPEYLQCVKELYRGG-LEPINFQTAADQARELINSWVESQTNGIIRNVL
GENE Y PROTEIN    12. VLPEYLSCARKFYTGG-VEEVNFKTAAEEARQLINSWVEKETNGQIKDLL
PLACENTAL PAI     13. FREEYIRLCQKYYSSE-PQAVDFLECAEEARKKINSWVKTQTKGKIPNLL
ANGIOTENS RAT     14. LKQPFVESLGPFTPAIFPRSLDLSTDPVLAAQKINRFVQAVTGWKMNLPL
ANGIOTENS HUMAN   15. LKQPFVQGLALYTPVVLPRSLDF-TELDVAAEKIDRFMQAVTGWKTGCSL
BARLEY Z PROTEIN  16.
CORTICOSTEROID BG 17. LLESFSADIKHYYESE-VLAMNFQDW-ATASRQINSYVKNKTQGKIVDLF
PROTEIN C INHIB.  18. LQDTFVSAMKTLYLAD-TFPTNFRD-SAGAMKQINDYVAKQTKGKIVDLL
PROTEASE NEXIN    19. IEVPFVTRNKDVFQCE-VRNVNFED-PASACDSINAWVKNETRDMIDNLL
RAB ORF1          20. LRPEFTAEFSSRFNTS-VQRVTFN--SENVKDVINSYVKDKTGGDVPRVL

                            180       190       200       210       220
secondary structure            ------s3A-----     ^hF1-- -s4C-------s3C-
ANTITRYPSIN HUMAN  1. KELDRD--TVFALVNYIFFKGKWERPFEVKDTEE-EDFHVDQVTTVKVPM
ANTITRYPSIN BABOON 2. KELDRD--TVFALVNYIFFKGKWERPFEVEATEE-EDFHVDQATTVKVPM
HEPARIN COFAC II   3. ENIDPA--TQMMILNCIYFKGSWVNKFPVEMTHN-HNFRLNEREVVKVSM
ANTITHROMBIN       4. PSEAINELTVLVLVNTIYFKGLWKSKFSPENTRK-ELFYKADGESCSASM
C1-INHIBITOR       5. DSLPSD--TRLVLLNAIYLSAKWKTTFDPKKTRM-EPFHFKNSV-IKVPM
ENDOTHELIAL PAI    6. GKGAVDQLTRLVLVNALYFNGQWKTPFPDSSTHR-RLFHKSDGSTVSVPM
THYROXINE B G      7. QDLKPN--TTMVLVNYIHFKAQWANPFDPSKTEDSSSFLIDKTTTVQVPM
ANTICHYMO MOUSE    8.           VVLVNYIYFKGKWKISFDPQDTFE-SEFYLDEKRSVKVPM
ANTICHYMO HUMAN    9. KDP--DSQTMMVLVNYIFFKAKWEMPFDPQDTHQ-SRFYLSKKKWVMVPM
ANTIPLASMIN       10. SGLPED--TVLLLLNAIHFQGFWRNKFDPSLTQR-DSFHLDEQFTVPVEM
OVALBUMIN         11. QPSSVDSQTAMVLVNAIVFKGLWEKAFKDEDTQA-MPFRVTEQESKPVQM
GENE Y PROTEIN    12. VSSSIDFGTTMVFINTIYFKGIWKIAFNTEDTRE-MPFSMTKEESKPVQM
PLACENTAL PAI     13. PEGSVDGDTRMVLVNAVYFKGKWKTPFEKKLNGL-YPFRVNSAQRTPVQM
ANGIOTENS RAT     14. EGVSTD--STLFFNTYVHFQGKM-RGFSQ-LTGL-HEFWVDNSTSVSVPM
ANGIOTENS HUMAN   15. MGASVD--STLAFNTYVHFQGKM-KGFSL-LAEP-QEFWVDNSTSVSVPM
BARLEY Z PROTEIN  16.
CORTICOSTEROID BG 17. SGLDS--PAILVLVNYIFFKGTWTQPFDLASTRE-ENFYVDETTVVKVPM
PROTEIN C INHIB.  18. KNLDS--NAVVIMVNYIFFKAKWETSFNHKGTQE-QDFYVTSETVVRVPM
PROTEASE NEXIN    19. SPDLIDGVTRLVLVNAVYFKGLWKSRFQPENTKK-RTFVAADGKSYQVPM
                                 L
RAB ORF1          20. DASLDRD-TKMLLLSSVRMKTSWRHVFDPSFTTD-QPFYSGNV-TYKVRM

                            230       240       250             260
secondary structure      ------   ----^^hF2^----s2B--  ----s3B--        ^^^
ANTITRYPSIN HUMAN  1. MKRLGMF--NIQHCKK-LSSWVLLMKYL-GNANAIFFLPD-----EGKLQ
ANTITRYPSIN BABOON 2. MRRLGMF--NIYHCEK-LSSWVLLMKYL-GNATAIFFLPD-----EGKLQ
HEPARIN COFAC II   3. MQTKGNF--LAANDQE-LDCDILQLEYV-GGISMLIVVPHK----MSGMK
ANTITHROMBIN       4. MYQEGKF--RYRRVAE--GTQVLELPFKGDDITMVLILPKP----EKSLA
C1-INHIBITOR       5. MNSKKYP-VAHFIDQT-LKAKVGQLQLS-HNLSLVILVPQNL--KHRLED
ENDOTHELIAL PAI    6. MAQTNKFNYTEFTTPDGHYYDILELPYHGDTLSMFIAAPYE---KEVPLS
THYROXINE B G      7. MHQMEQY--YHLVDME-LNCTVLQMDYS-KNALALFVLPK-----EGQME
ANTICHYMO MOUSE    8. MKMKLL-TTRHFRDEE-LSCSVLELKYT-GNASALLILPD-----QGRMQ
ANTICHYMO HUMAN    9. MSLHHL-TIPYFRDEE-LSCTVVELKYT-GNASALFILPD-----QDKME
ANTIPLASMIN       10. MQARTYP-LRWFLLEQ-PEIQVAHFPFK-NNMSFVVLVPTH---FEWNVS
OVALBUMIN         11. MYQIGLF--RVASMAS-EKMKILELPFASGTMSMLVLLPDE----VSGLE
GENE Y PROTEIN    12. MCMNNSF--NVATLPA-EKMKILELPYASGDLSMLVLLPDE----VSGLE
PLACENTAL PAI     13. MYLREKL--NIGYIED-LKAQILELPYA-GDVSMFLLLPDEIADVSTGLE
ANGIOTENS RAT     14. LSGTGNF--QHWSDAQ-NNFSVTRVPL-GESVTLLLIQPQ----CASDLD
ANGIOTENS HUMAN   15. LSGMGTF--QHWSDIQ-DNFSVTQVPF-TESACLLLIQPH----YASDLD
BARLEY Z PROTEIN  16.       YISSSDNLK-VLKLPYAKGHDKRQFSMYILLPG----AQDGLW
CORTICOSTEROID BG 17. MLQSSTI--SYLHDSE-LPCQLVQMNYV-GNGTVFFILPD-----KGKMN
PROTEIN C INHIB.  18. MSREDQY--HYLLDRN-LSCRVVGVPYQ-GNATALFILPS-----EGKMQ
PROTEASE NEXIN    19. LAQLSVFRCGSTSAPNDLWYNFIELPYHGESISMLIALPT---ESSTPLS
RAB ORF1          20. MNKIDTL-KTETFTLRNVGYSVTELPYKRRQTAMLLVVP------DDLGE
```

Table II (Continued)

```
                                  270       280         290       300
secondary structure          hG^   ^^^^^hH^^^^      --s2C-----s6A---^^ ^hI^^^
ANTITRYPSIN HUMAN  1. HLENELTHDIITKFLENEDR--RSASLHLPKLSITGTYDLK-SVLGQLGI
ANTITRYPSIN BABOON 2. HLENELTHDIITKFLENENR--RSANLHLPKLAITGTYDLK-TVLGHLGI
HEPARIN COFAC II   3. TLEAQLTPRVVERWQKSMTN--RTREVLLPKFKLEKNYNLV-ESLKLMGI
ANTITHROMBIN       4. KVEKELTPEVLQEWLDELEE--MMLVVHMPRFRIEDGFSLK-EQLQDMGL
C1-INHIBITOR       5. MEQALSPSVFKAIMEKLEMSKFQPTLLTLPRIKVTTSQDML-SIMEKLEF
ENDOTHELIAL PAI    6. ALTNILSAQLISHWKGNMTR--LPRLLVLPKFSLETEVDLR-KPLENLGM
THYROXINE B G      7. SVEAAMSSKTLKKWNRLLQK--GWVDLFVPKFSISATYDLG-ATLLKMGI
ANTICHYMO MOUSE    8. QVEASLQPETLRKWRKTLFPS-QIEELNLPKFSIASNYRLEEDVLPEMGI
ANTICHYMO HUMAN    9. EVEAMLLPETLKRWRDSLEFR-EIGELYLPKFSISRDYNLN-DILLQLGI
ANTIPLASMIN       10. QVLANLSWDTLHPPLVWE----RPTKVRLPKLYLKHQMDLV-ATLSQLGL
OVALBUMIN         11. QLESIINFEKLTEWTSSNVMEERKIKVYLPRMKMEEKYNLT-SVLMAMGI
GENE Y PROTEIN    12. RIEKTINFDKLREWTSTNAMAKKSMKVYLPRMKIEEKYNLT-SILMALGM
PLACENTAL PAI     13. LLESEITYDKLNKWTSKDKMAEDEVEVYIPQFKLEEHYELR-SILRSMGM
ANGIOTENS RAT     14. RVEVLVFQHDFLTWIKNPPP--RAIRLTLPQLEIRGSYNLQ-DLLAQAKL
ANGIOTENS HUMAN   15. KVEGLTFQQNSLNWMKKLSP--RTIHLTMPQLVLQGSYDLQ-DLLAQAEL
BARLEY Z PROTEIN  16. SLAKRLSTEPEFIENHIPKQTVEVGRFQLPKFKISYQFEAS-SLLRALGL
CORTICOSTEROID BG 17. TVIAALSRDTINRWSAGLTS--SQVDLYIPKVTISGVYDLG-DVLEEMGI
PROTEIN C INHIB.  18. QVENGLSEKTLRKWLKMFKK--RQLELYLPKFSIEGSYQLE-KVLPSLGI
PROTEASE NEXIN    19. AIIPHISTKTIDSWMSIMVP--KRVQVILPKFTAVAQTDLK-EPLKVLGI
RAB ORF1          20. IVRALDDLSLVRFWIRNMRK---DVCQVVMPKFSVESVLDLR-DALQRLGV

                     310       320         330       340       350
secondary structure   ^hI1                -------s5A----------s4A-------
ANTITRYPSIN HUMAN  1. TKVFSNGAD-LSGVTEEA--PLKLSKAVHKAVLTIDEKGTEAAGAMFLEA
ANTITRYPSIN BABOON 2. TKVFSNGAD-LSGVTEDA--PLKLSKAVHKAVLTIDEKGTEAAGAMFLEA
HEPARIN COFAC II   3. RMLFDKNGN-MAGISDQR---IAIDLFKHQGTITVNEEGTQATTVTTVGF
ANTITHROMBIN       4. VDLFSPEKSKLPGIVAEGRDDLYVSDAFHKAFLEVNEEGSEAAASTAVVI
C1-INHIBITOR       5. FD-FSYDLN-LCGLTEDP--DLQVSAMQHQTVLELTETGVEAAAASAISV
ENDOTHELIAL PAI    6. TDMFRQFQADFTSLSDQE--PLHVAQALQKVKIEVNESGTVASSSTAVIV
THYROXINE B G      7. QHAYSENAD-FSGLTEDN--GLKLSNAAHKAVLHIGEKGTEAAAVPEVEL
ANTICHYMO MOUSE    8. KEVFTEQAD-LSGIIETK--KLSVSQVVHKAVLDVAETGTEAAAATGVIG
ANTICHYMO HUMAN    9. EEAFTSKAD-LSGITGAR--NLAVSQVVHKAVLDVFEEGTEASAATAVKI
ANTIPLASMIN       10. QELF-QAPD-LRGISEQ---SLVVSGVQHQSTLELSEVGVEAAAATSIAM
OVALBUMIN         11. TDVFSSSAN-LSGISSAE--SLKISQAVHAAHAEINEAGREVVGSAEAGV
GENE Y PROTEIN    12. TDLFSRSAN-LTGISSVD--NLMISDAVHGVFMEVNEEGTEATGSTGAIG
PLACENTAL PAI     13. EDAFNKGRANFSGMSERN--DLFLSEVFHQAMVDVNEEGTEAAAGTGGVM
ANGIOTENS RAT     14. STLLGAEAN-LGKMGDTN--PRVGEVLNSILLELQAGEEEQPTESAQQPG
ANGIOTENS HUMAN   15. PAILHTELN-LQKLSNDR--IRVGEVLNSIFFELEA-DEREPTESTQQLN
BARLEY Z PROTEIN  16. QLPFSEEAD-LSEMVDSS-QGLEISHVFHKSFVEVNEEGTEAGAATVAMG
CORTICOSTEROID BG 17. ADLFTNQAN-FSRITQDA--QLKSSKVVHKAVLQLNEEGVDTAGSTGVTL
PROTEIN C INHIB.  18. SNVFTSHAD-LSRISNHS--NIQVSEMVHKAVVEVDESGTRAAAATGTIF
PROTEASE NEXIN    19. TDMFDSSKANFAKITTGSE-NLHVSHILQKAKIEVSEDGTKASAATTAIL
RAB ORF1          20. RDAFDPSRADFGQASPSN--DLYVTKVLQTSKIEADERGTTASSDTAITL

                        360             370       380       390 394
secondary structure     -       --      s1C- ---s4B-^hI2^---s5B---^hI3
ANTITRYPSIN HUMAN  1. IP-MSIPPE-----VKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK
ANTITRYPSIN BABOON 2. IP-MSIPPE-----VKFNKPFVFLMIEQNTKSPLFIGKVVNPTQK
HEPARIN COFAC II   3. MP-LSTQVR-----FTVDRPFLFLIYEHRTSCLLFMGRVANPSRS
ANTITHROMBIN       4. AG-RSLNPN--RVTFKANRPFLVFIREVPLNTIIFMGRVANPCVK
C1-INHIBITOR       5. A--RTLLV------FEVQQPFLFVLWDQQHKFPVFMGRVYDPRA
ENDOTHELIAL PAI    6. SA-RMAPEE-----IIMDRPFLFVVRHNPTGTVLFMGQVMEP
THYROXINE B G      7. SD-QPENTFLHPI-IQIDRSFMLLILERSTRSILFLGKVVNPTEA
ANTICHYMO MOUSE    8. GIRKAILPA-----VHFNRPFLFVIYHTSAQSILFMAKVNNPK
ANTICHYMO HUMAN    9. TL-LSALVETRTI-VRFNRPFLMIIVPTDTQNIFFMSKVTNPKQA
ANTIPLASMIN       10. S--RMSLSS-----FSVNRPFLFFIFEDTTGLPLFVGSVRNPNPSAPREL
OVALBUMIN         11. DA-ASVS-EE----FRADHPFLFCIKHIATNAVLFFGRCVSP
GENE Y PROTEIN    12. NIKHSLELEE----FRADHPFLFFIRYNPTNAILFFGRYWSP
PLACENTAL PAI     13. TG-RTGHGG---PQFVADHPFLFLIMHKITKCILFFGRFCSP
ANGIOTENS RAT     14. SP--------EVLDVTLSSPFLFAIYERDSGALHFLGRVDNPQNVV
ANGIOTENS HUMAN   15. KP--------EVLEVTLNRPFLFAVYDQSATALHFLGRVANPLSTA
BARLEY Z PROTEIN  16. VA-MSMPLKVDLVDFVANHPFLFLIREDIAGVVVFVGHVTNPLISA
CORTICOSTEROID BG 17. NL-TSKPII-----LRFNQPFIIMIFDHFTWSSLFLARVMNPV
PROTEIN C INHIB.  18. TF-RSARLN--SQRLVFNRPFLMFIV---DNNILFLGKVNRP
PROTEASE NEXIN    19. IA-RSSPPW-----FIVDRPFLFFIRHNPTGAVLFMGQINKP
RAB ORF1          20. IP-RNALTA-----IVANKPFMFLIYHKPTTTVLFMGTITKGEKVIYDTE

                     400       410       420       430       440   446
ANTITRYPSIN HUMAN  1.
ANTITRYPSIN BABOON 2.
HEPARIN COFAC II   3.
ANTITHROMBIN       4.
C1-INHIBITOR       5.
ENDOTHELIAL PAI    6.
THYROXINE B G      7.
ANTICHYMO MOUSE    8.
ANTICHYMO HUMAN    9.
ANTIPLASMIN       10. KEQQDSPGNKDFLQSLKGFPRGDKLFGPDLKLVPPMEEDYPQFGSPK
OVALBUMIN         11.
GENE Y PROTEIN    12.
PLACENTAL PAI     13.
ANGIOTENS RAT     14.
ANGIOTENS HUMAN   15.
BARLEY Z PROTEIN  16.
CORTICOSTEROID BG 17.
PROTEIN C INHIB.  18.
PROTEASE NEXIN    19.
RAB ORF1          20. GRDDVVSSV
```

Table II (Continued)

[a] The secondary structural elements are indicated in the headline. Most amino acid sequences were taken from MIPSX Database. *α_1-Antitrypsin precursor—Human.* Bollen, A., Herzog, A., Cravador, A., Herion, P., Chuchana, P., Vander Straten, A., Loriau, R., Jacobs, P., & Van Elsen, A. (1983) *DNA 2*, 255–264. Rosenberg, S., Barr, P. J., Najarian, R. C., & Hallewell, R. A. (1984) *Nature 312*, 77–80 (this sequence differs from that shown in having His 125, Gly 139, Asn 140, Thr 273, and Ile 326). Carrell, R. W., Jeppsson, J.-O., Laurell, C.-B., Brennan, S. O., Owen, M. C., Vaughan, L., & Boswell, D. R. (1982) *Nature 298*, 329–334 (sequence of residues 25–418; this sequence differs from that shown in having Gly 139, Asn 140, and Thr 273). *α_1-Antitrypsin—Baboon.* 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. *Heparin cofactor II precursor—Human.* Blinder, M. A., Marasa, J. C., Reynolds, C. H., Deaven, L. L., & Tollefsen, D. M. (1988) *Biochemistry 27*, 752–759. *Antithrombin III precursor—Human.* Bock, S. C., Wion, K. L., Vehar, G. A., & Lawn, R. M. (1982) *Nucleic Acids Res. 10*, 8113–8125. Chandra, T., Stackhouse, R., Kidd, V. J., & Woo, S. L. C. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1845–1848. Prochownik, E. V., Markham, A. F., & Orkin, S. H. (1983) *J. Biol. Chem. 258*, 8389–8394 (this sequence differs from that shown in having Arg 97). Petersen, T. E., Dudek-Wojciechowska, G., Sottrup-Jensen, L., & Magnusson, S. (1979) *The Physiological Inhibitors of Coagulation and Fibrinolysis* (Collen, D., Wiman, B., & Verstraete, M., Eds.) pp 43–54, Elsevier/North-Holland Biomedical Press, Amsterdam. *Complement C_1-inhibitor precursor—Human.* Carter, P. E., Dunbar, B., & Fothergill, J. E. (1988) *Eur. J. Biochem. 173*, 163–169. *Plasminogen activator inhibitor 1 precursor—Human.* Pannekoek, H., Veerman, H., Lambers, H., Diergaarde, P., Verweij, C. L., van Zonneveld, A. J., & van Mourik, J. A. (1986) *EMBO J. 5*, 2539–2544. Ginsburg, D., Zeheb, R., Yang, A. Y., Rafferty, U. M., Andreasen, P. A., Nielsen, L., Dano, K., Lebo, R. V., & Gelehrter, T. D. (1986) *J. Clin. Invest. 78*, 1673–1680. *Thyroxine-binding globulin precursor—Human.* Flink, I. L., Bailey, T. J., Gustafson, T. A., Markham, B. E., & Morkin, E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7708–7712. *Contrapsin—Mouse (fragment).* Hill, R. E., Shaw, P. H., Boyd, P. A., Baumann, H., & Hastie, N. D. (1984) *Nature 311*, 175–177. *α_1-Antichymotrypsin precursor—Human.* Chandra, T., Stackhouse, R., Kidd, V. J., Robson, K. J. H., & Woo, S. L. C. *Biochemistry 22*, 5055–5061 (partial sequence derived from the 3′ half of the m-RNA). Hill, R. E., & Hastie, N. D. (1987) *Nature 326*, 96–99 (this sequence is reported wherever there are differences). Morii, M., & Travis, J. (1983) *J. Biol. Chem. 258*, 12749–12752 (inhibitory site). *α_2-Antiplasmin precursor—Human.* Holmes, W. E., Nelles, L., Lijnen, H. R., & Collen, D. (1987) *J. Biol. Chem. 262*, 1659–1664. *Ovalbumin—Gallus gallus.* Woo, S. L. C., Beattie, W. G., Catterall, J. F., Dugaiczyk, A., Staden, R., Brownlee, G. G., & O'Malley, B. W. (1981) *Biochemistry 20*, 6437–6446. Benoist, C., O'Hare, K., Breathnach, R., & Chambon, P. (1980) *Nucleic Acids Res. 8*, 127–142. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4853–4857. Gannon, F., O'Hare, K., Perrin, F., LePennec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., & Chambon, P. (1979) *Nature 278*, 428–434. McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M., & Brownlee, G. G. (1978) *Nature 273*, 723–728. Robertson, M. A., Staden, R., Tanaka, Y., Catterall, J. F., O'Malley, B. W., & Brownlee, G. G. (1979) *Nature 278*, 370–372. *Gene Y protein (ovalbumin-related)—Chicken.* Heilig, R., Muraskowsky, R., Kloepfer, C., & Mandel, J. L. (1982) *Nucleic Acids Res. 10*, 4363–4382. *Placental plasminogen activator inhibitor, type II—Human.* Ye, R. D., Wun, T. C., & Sadler, J. E. (1987) *J. Biol. Chem. 262*, 3718–3725. *Angiotensinogen precursor—Rat.* Ohkubo, H., Kageyama, R., Ujihara, M., Hirose, T., Inayama, S., & Nakanishi, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2196–2200. Bouhnik, J., Clauser, E., Strosberg, D., Frenoy, J. P., Menard, J., & Corvol, P. (1981) *Biochemistry 20*, 7010–7015 (sequence of residues 25–41). *Angiotensinogen precursor—Human.* Kageyama, R., Ohkubo, H., & Nakanishi, S. (1984) *Biochemistry 23*, 3603–3609 (it is uncertain whether Met-1 or Met-10 is the initiator). Tewksbury, D. A., Dart, R. A., & Travis, J. (1981) *Biochem. Biophys. Res. Commun. 99*, 1311–1315 (sequence of residues 34–58; this sequence differs from that shown in having Ser 51 and Asp 58, residue 47 was not determined). *Protein Z—Barley (fragment).* Hejgaard, J., Rasmussen, S. K., Brandt, A., & Svendsen, I. (1985) *FEBS Lett. 180*, 89–94. Nielsen, G., Johansen, H., Jensen, J., & Hejgaard, J. (1983) *Barley Genet. Newslett. 13*, 55–57 (map position). *Human corticosteroid binding globulin mRNA, complete cds.* Hammond, G. L., Smith, C. L., Goping, I. S., Underhill, D. A., Harley, M. J., Reventos, J., Musto, N. A., Gunsalus, G. L., Bardin, C. W. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5153–5157. *Protein C inhibitor precursor—Human.* Suzuki, K., Deyashiki, Y., Nishioka, J., Kurachi, K., Akira, M., Yamamoto, S., & Hashimoto, S. (1987) *J. Biol. Chem. 262*, 611–616. *Glia-derived neurite promoting factor precursor—Human.* Gloor, S., Odink, K., Guenther, J., Nick, H., & Monard, D. (1986) *Cell 47*, 687–693 (sequence). *Hypothetical protease inhibitor—Rabbit plasmid.* Upton, C., Carrell, R. W., & McFadden, G. (1986) *FEBS Lett. 207*, 115–120. MIPSX: F. Pfeiffer, Martinsried Institute for Protein Sequences (MIPS), unpublished.

which is 21 Å away from residue 319. Both disulfide bonds can therefore be made if the segments 14b–20 and 9–14b have appropriate conformations. The N-terminus of C_1-inhibitor is rich in prolines and may prefer the conformation of a polyproline II helix with a rise of 3.1 Å per residue to span considerable distances. In antithrombin III two disulfide bonds link the N-terminus to the globular molecule. The segment from residue −5 to −18 is linked to 69 and 101, respectively, which are 29.7 Å apart. It is rich in prolines (3 out of 15) and can easily span the long distance. It is obvious that in both molecules the N-termini are located close to helix hD.

Carbohydrate Attachment Sites

α_1-Antitrypsin has three branched carbohydrate chains linked to asparagines 46, 83, and 247. These are located in turns thAs6B, thChD, ts2Bs3B, respectively. Most of the members of the serpin family are glycoproteins but only in a few of them are the sites of attachment chemically defined. In others the presence of consensus amino acid sequences, Asn-X-Ser/Thr, in relation to the content of carbohydrate determined chemically suggests sites of attachment.

These are shown in Figure 4, projected onto the α_1-antitrypsin structure except for the glycosylation sites at the N-terminal extension in C_1-inhibitor. Carbohydrate is apparently distributed over most of the surface of the molecule with no obvious preference. All glycosylation sites are external with the asparagine side chains as sites of attachment projecting into solution. Most of them are located in turns; a few are on the hydrophilic side of peripheral α-helixes.

Structure of the Active Inhibitor

The crystallographic structure obtained by Löbermann et al. (1984) is that of antitrypsin cleaved at the reactive center 358 methionine. Although the separation of the cleaved ends to opposite poles of the molecule suggests that the modified molecule has undergone a major conformational change, other evidence indicates that the changes are relatively limited. In the first place, the known structure of the cleaved form puts constraints on the structural transition, and as described here, the cleaved structure is compatible with the requirement of the native protein for the placement of oligosaccharide attachments and for the heparin binding site. The cleaved structure also provides a satisfying explanation for the functional changes observed in mutants of native serpins. In particular, the disulfide linkages observed in well-defined conserved areas of antithrombin and placental PAI, as in Figure 3, indicate that the central part of the molecule is not involved in significant conformational shifts.

What then is the change that occurs at the reactive center on cleavage? There is good evidence that a reconstruction of the native reactive center involves an extraction of much of the central 4A strand of the A sheet in which methionine 358 forms the C-terminus. Exposure of the native serpins to a range of proteases shows that they are consistently vulnerable to cleavage at sites within s4A (Carrell et al., 1987a,b). This is compatible with the model of Löbermann et al. (1984) in which the reconstructed reactive center is situated near the position of serine 359 in the cleaved molecule with removal of s4A followed by annealing of the A sheet. The evidence

Table III: Conserved Residues in the Serpin Family

F33	internal; close to conserved I57
hA	
Y/F38	internal; O^{ϵ} hydrogen bonded to O^{ϵ}E264(conserved); close to conserved F52,M385
N49	internal; side chain hydrogen bonded to main chain of V380 and N390; close to conserved I293,V388
s6B	
S53	internal; initiates hB; O^{γ} hydrogen bonded to main chain of S56 and L383
P54	internal; initiates hB; close to conserved P382,F52
S56	internal; O^{γ} hydrogen bonded to side chain of conserved, internal N186
I57	internal; close to conserved F33,H334; in pocket formed by these residues and by F61,M351,L37
hB	
T59	internal; O^{γ} close to side chain of conserved N186
L66	in surface niche; close to conserved F130,Y138
G67	terminates hB; a side chain would interfere with conserved F130,G320
T72	internal; O^{γ} hydrogen bonded to O K69, N H73, O N317
F/Y119	close to conserved I157; in pocket formed by I157,Y160, F143
s2A	
F/Y130	close to G67; in pocket formed by L118,V321
hE	
Y138	internal; ends hE; O^{ϵ} hydrogen bonded to N^{δ}H93
I157	internal tight pocket; close to conserved F(Y)119
N158	internal; side chain hydrogen bonded to main chain of V173; holds hF to antiparallel segment of bulges 169-176
hF	
T165	O^{γ} hydrogen bonded to main chain of K168,I169,V161 which are conserved; the adjacent residue Q166 is in left handed helical conformation
I169	internal; in tight pocket
L184	internal;
N186	internal; side chain hydrogen bonded to side chain of conserved S56; close to T59 and N116
I188	internal pocket; close to conserved F384,S56
F/Y190	internal; close to conserved F384,M374
s3A	
G192	a side chain would interfere with conserved F190
W194	internal; tightly packed, close to conserved F198,Y244; N^{ϵ} hydrogen bonded to main chain D341 close to conserved E342

Table III (Continued)

-	F198	internal; close to conserved W194
T	T203	ends hF1; O^{γ} is hydrogen bonded to main chain O of V200 and to the side chain of the invariant E342
s4C		
¦	F208	internal pocket formed by conserved P391,F370
-		
T	V218	internal; close to conserved P391, F208, M220
¦	M220	internal; close to conserved F208
s3C		
¦	M221	internal; close to conserved T203,E342,K290,F198
-		
-	P255	ends s3B and initiates ts3BhG; close to conserved L260
T	L260	in niche; close to conserved P369
hG		
¦	E264	salt link to K387 in hydrophobic surface pocket;
-		hydrogen bonded to O^{ϵ} of conserved Y38
-	P289	internal; close to conserved F370, M220, F208
T	Y297	O^{ϵ}hydrogen bonded to main chain of F51
hI		
¦	L303	ties hI to molecule; close to conserved F312, F33
-		
-	F312	ties hI1 to molecule
-	L318	
-	G320	see G67
T	H334	internal; side chain hydrogen bonded to main chain S53 and A355
s5A		
¦	E342	salt link to K290 and hydrogen bonded to side chain of conserved T203
s4A		
-	G344	a side chain would interfere with conserved W194 and Y244
T	P369	turns s1C into s4B; close to conserved F208,P255,K387
¦	F370	
s4B		close to opposite strand s5B with conserved F384,G386,V388;
¦		also close to Y38
¦	V/L371	
¦	F372	
-		
T	F384	see F370,V371,F372
¦	G386	
s5B		
¦	V388	internal; close to conserved F370,F372,F208
¦	N390	$O^{\delta 1}$ hydrogen bonded to N, O^{γ} T392; initiates hI3
¦	P391	internal; close to conserved F208; initiates one
-		turn C-terminal helix (hI3)

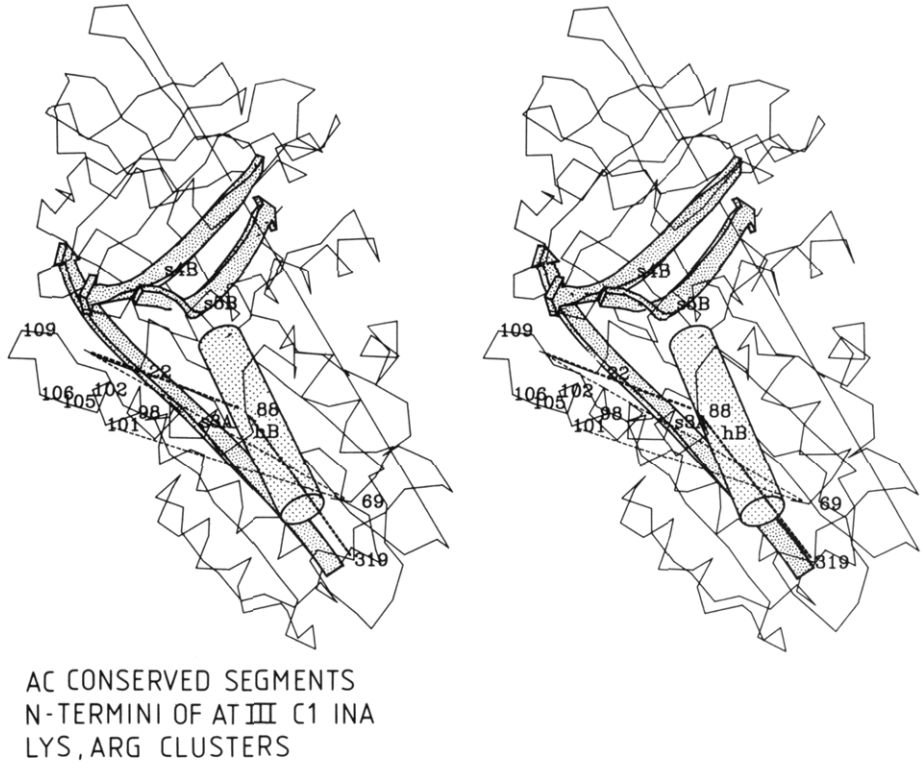

FIGURE 2: Polypeptide chain folding of α_1-antitrypsin with the locations of the most conserved secondary structure elements drawn as dotted arrows and cylinders (hB, s3A, s4B, s5B). The broken lines indicate the general course of the N-termini in antithrombin III (thin line) and C_1-inhibitor (thick line) as derived from the disulfide linkages made to 69 and 101 and to 88 and 319, respectively. Residue numbers 22, 98, 102, 105, 106, and 109 mark the cluster of lysine and arginine residues in antithrombin III. Residue numbers refer to α_1-antitrypsin and the alignment in Table II. Residues C-terminal to 358 have 100 added to their numbers in the plot.

from proteolysis is clearly incompatible with the alternative reconstruction of Toma et al. (1987) involving a movement of serine 359 along with strand s1C to methionine 358 fixed in the A sheet.

Confirmation of the Löbermann et al. model comes from their prediction that the cleaved structures will have increased conformational stability over that of the native structures. Experimental evidence confirms this and shows that the serpins in general undergo a striking change in thermal stability on cleavage, with an increase in denaturation temperature from 58 °C in the native inhibitors to 80 °C or above in the cleaved inhibitors (Carrell & Owen, 1985). This so-called stressed (S) to relaxed (R) conformational change on cleavage at various sites within the exposed s4A segment has been demonstrated in α_1-antitrypsin, C_1-inhibitor, α_1-antichymotrypsin, antithrombin, and cortisol- and thyroxine-binding globulins (Pemberton et al., 1988) and quantitatively characterized in some of these by proton NMR and circular dichroism studies (Gettins & Horton, 1988; Bruch et al., 1988). It does not occur with the noninhibitors ovalbumin and angiotensinogen (Gettins, 1989; Stein et al., 1989); this is compatible with our proposal that the stressed conformation has been conserved in an evolutionary sense as a requirement for the inhibitory function of the reactive center.

A problem intimately related to the structure of the active, intact inhibitor concerns the docking to its cognate enzyme, leukocyte elastase. This is related to specific interactions of the substrate binding area of leukocyte elastase (Bode et al., 1986) with the primary binding segment but probably also to electrostatic attractions acting between more extended surface areas of the molecules. Calculations of the electrostatic potential of α_1-antitrypsin and human leukocyte elastase (HLE) using the Atanasov and Karshikov (1985) approach based on the Tanford and Kirkwood (1957) theory confirm the very dipolar character of the inhibitor with the positive pole at the S359 and the negative pole at the M358 end. Conversely, HLE has a negative electrostatic potential in the substrate binding area. This suggests that docking takes place at the S359 end of the inhibitor and is in accord with the view that s4A is removed from sheet A to approach S359 in the intact inhibitor.

The proposed docking area encompasses the site of access to the barrel formed by residues 190–300 and 359–394, i.e., β-sheets B and C. The barrel is the proposed ligand binding site in serpins with carrier function. Bound ligands may therefore modulate protease binding. Reciprocally, limited proteolysis is expected to affect ligand binding.

Carrier Function

Plasma proteins which function as carriers of thyroxine or corticosteroids have been identified as members of the serpin family (Flink et al., 1986; Hammond et al., 1987). There is direct information on the binding site of corticosteroid ligands in the carrier protein. Corticosteroid binding protein has two cysteine residues (69 and 237, α_1-antitrypsin numbering), one of which is affinity labeled by a ligand analogue but has not been identified (Defye et al., 1980; Kahn & Rosner, 1977). C69 is located in thBhC with no obvious ligand binding cavity, while C237 is in s2B at the mouth of a deep pocket formed by a barrel of β-strands, characteristic of other ligand binding sites (Katsunuma et al., 1980; Ragg, 1986). The binding site of thyroxine in thyroxine-binding globulin has been labeled by the chemical cross-linking of ligand and receptor. A labeled residue is K256 (D in α_1-antitrypsin) (Tabachnik & Perret, 1987) at the mouth of the proposed pocket of the β-barrel very close to residue 237 (Figure 5). Protein families which exhibit general ligand binding properties analyzed so far are immunoglobulins (Huber, 1984), retinol and bilin binding proteins

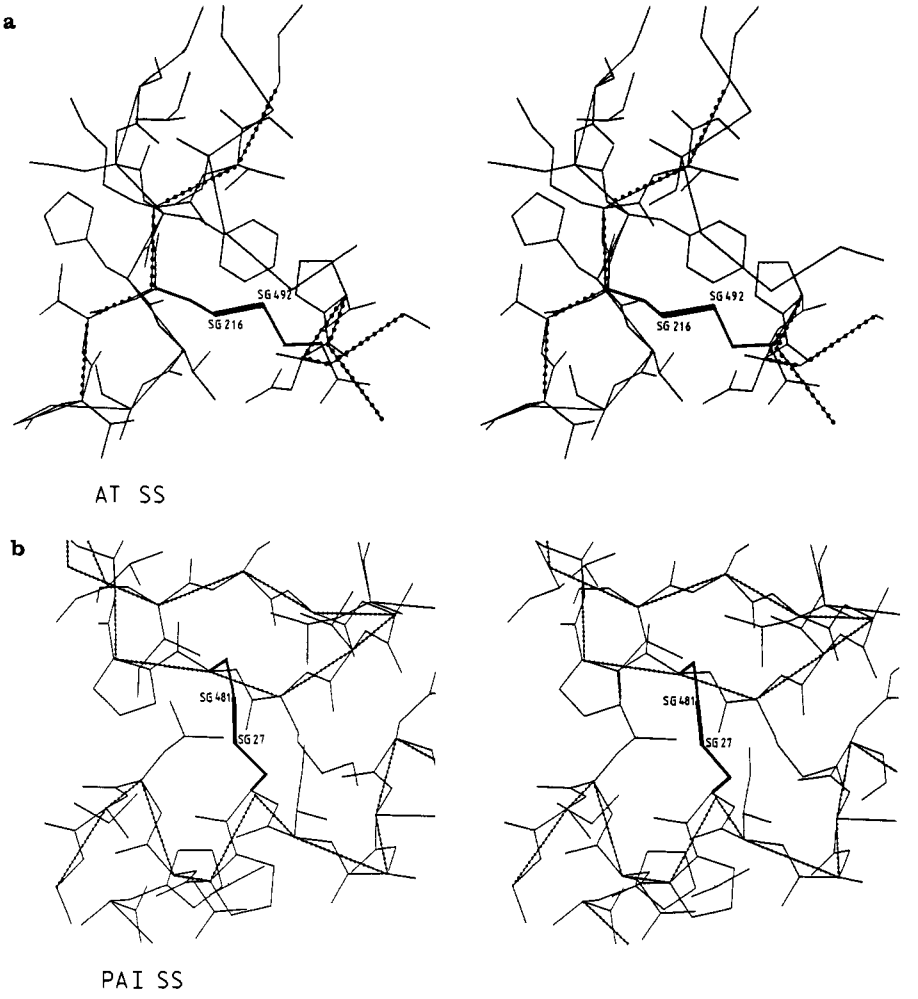

FIGURE 3: Model of antithrombin III and placental PAI at the (a) C216–C392 and the (b) C27–C381 disulfide groups, respectively. The dotted lines mark the positions of the C^{α} atoms in α_1-antitrypsin. Residue numbers refer to α_1-antitrypsin and the alignment in Table II. Residues C-terminal to 358 have 100 added to their numbers in the plot.

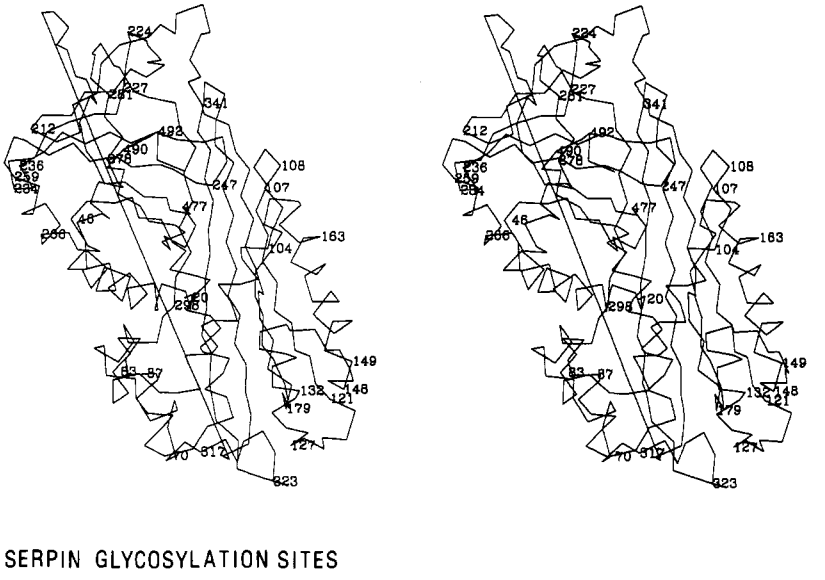

FIGURE 4: Potential glycosylation sites in members of the serpin superfamily projected onto the α_1-antitrypsin structure. Residue numbers refer to α_1-antitrypsin and the alignment in Table II. Residues C-terminal to 358 have 100 added to their numbers in the plot.

(Newcomer et al., 1984; Huber et al., 1987), and prealbumin (Blake & Oatley, 1977). Prealbumin deserves particular attention as it is also a thyroxine binding protein. Though seemingly unrelated in detailed folding topology and probably also in evolution, all three protein classes have β-barrel structures in common, which in the latter two cases are the ligand binding sites. There is a similar structural motif in α_1-antitrypsin in the S359 end of the molecule consisting of residues 190–300 and 359–394, as shown in Figure 5 in a comparison with bilin binding protein. In α_1-antitrypsin the entrance to the barrel as a putative ligand binding site is blocked by bulky W238, which in thyroxine-binding globulin

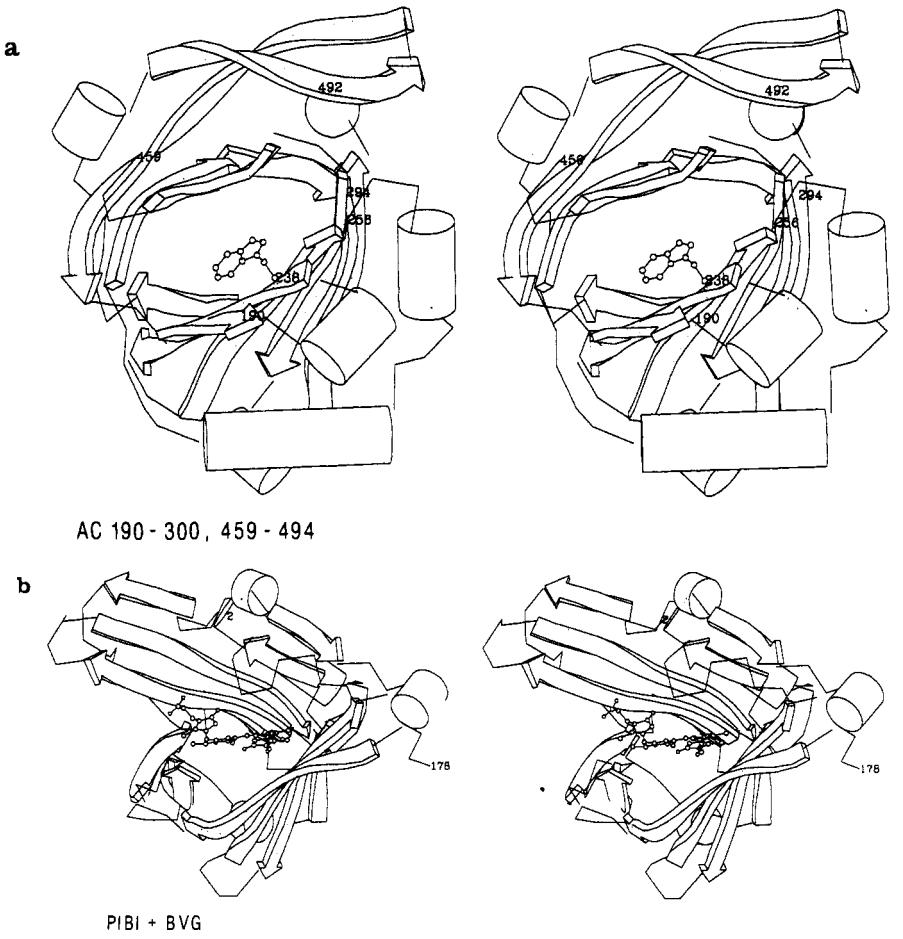

FIGURE 5: Bilin binding protein oriented with a view onto the bound biliverdin IXγ and residues 190–300 and 359–394 of α_1-antitrypsin with a similar view into the putative ligand binding pocket. Residue numbers refer to α_1-antitrypsin and the alignment in Table II. Residues C-terminal to 358 have 100 added to their numbers in the plot.

is T and in corticosteroid binding protein is Q. These small residues may permit access to the barrel.

Heparin Binding Site

Four of the human serpins, antithrombin III, heparin cofactor II, protease nexin I, and protein C inhibitor, are relatively inefficient inhibitors until they are activated by specific sulfated polysaccharides. The best studied of these inhibitors is antithrombin III, which undergoes a 1000-fold increase in its association constant with thrombin in the presence of heparin. The binding function of heparin is dependent on a pentasaccharide sequence with four sulfate groups (Choay et al., 1983; Lindahl et al., 1983; Beetz & Van Boeckel, 1986). Evidence as to the binding site of the pentasaccharide on antithrombin has come primarily from human variant antithrombins with decreased heparin affinity. Three of these variants were shown to be due to mutations of a single arginine at position 47/22 to cysteine, to histidine, and to serine (Koide et al., 1984; Owen et al., 1987; Borg et al., 1988). This finding together with other chemical evidence (Peterson et al., 1987) implicating K125/98 supported the inference that the binding of heparin to antithrombin was primarily due to salt bridging between the sulfates of the heparin and basic residues on the protein. Carrell et al. (1987a,b) compared the aligned serpins to determine which basic residues were uniquely conserved in antithrombin, in heparin cofactor II, and, subsequently, in protease nexin I. When these mutually conserved arginines and lysines are projected on the three-dimensional α_1-antitrypsin template, they are seen to form a band of positive charge stretching from the base of the A helix and across the underside of the D helix (Figure 6). Further support comes from the identification of other human heparin affinity variants. In the variant P41/16 → L (Chang & Tran, 1986) heparin binding is affected. This mutation is spatially closed to the suggested binding site. Also, Brennan et al. (1987, 1988) showed that the heparin affinity of antithrombin was decreased by the presence of oligosaccharide side chains at either the upper or lower ends of the proposed site. The change at the upper end of the site occurs due to aberrant glycosylation of N135/108, to give the high-affinity β-antithrombin normally present as a minor component in plasma. The change at the lower end of the site occurs in a mutant where I7/−19 is replaced by an asparagine which consequently is subject to glycosylation. The position of I7 can be approximately fixed on the antitrypsin model since it is adjacent to C8/−18, which is linked to C128/101. This places asparagine 7, and hence its bulky attached oligosaccharide, near the base of the proposed heparin site where it would predictably interfere with the binding of the heparin pentasaccharide.

On the basis of these results, we suggest that the primary heparin binding site on these three serpins is as shown in Figure 6, involving in the case of antithrombin arginines 47, 129, and 132 and lysines 125 and 133. Support for this conclusion is given by Smith and Knauer (1987) and Griffith et al. (1985), who show that the isolated antithrombin peptide 114–159 is preferentially bound by heparin. A further confirmation has been given by the recent identification of a low heparin affinity variant of heparin cofactor II in which the residue homologous to antithrombin's R129 (R189 in heparin cofactor) has been replaced by a histidine (Blinder et al., 1989).

The story is of course more complex than this because heparin activates these serpins as well as binding to them. The

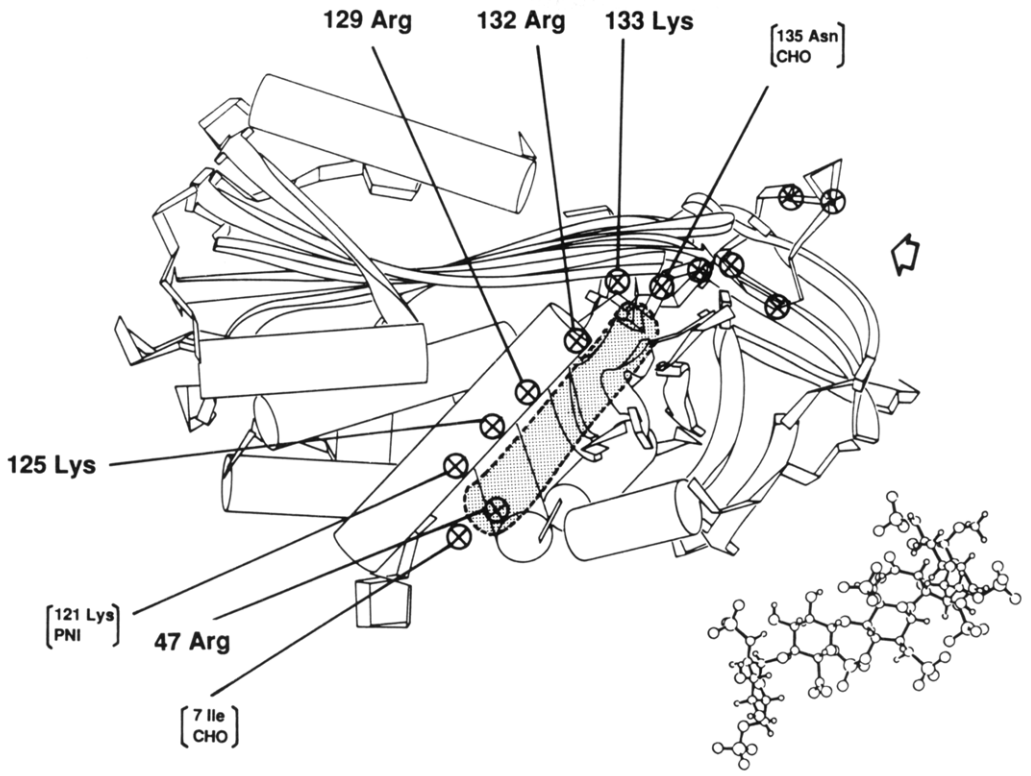

FIGURE 6: Heparin site. Schematic view of antithrombin projected onto the structure of cleaved α_1-antitrypsin, showing the prime (pentasaccharide) binding site (shaded) formed by the side chains of lysines and arginines 47, 125, 129, 132, and 133. The site extends around the molecule to the reactive center with K136, K228, R235, and K236 (shown but not labeled). The estimated region of the reactive center is arrowed. The prime site is flanked by oligosaccharide attachment points 135 and 7 (Asn in variant Rouen-III). In protease nexin I the equivalent of R47 is absent, but at position 121 is a lysine whose side chain overlaps that of 47. The size of the pentasaccharide is indicated in the lower right. Residue numbers refer to antithrombin III. Residues C-terminal to 358 have 100 added to their numbers in the plot.

mechanism of activation of antithrombin is uncertain, but it is likely to involve both a conformational change on the binding of heparin, as occurs with the activation by the pentasaccharide of factor Xa inhibition, and also a direct influence on the reactive site as probably occurs with the large heparins (12–15 units) necessary for the activation of thrombin inhibition. The latter type of activation is compatible with the observation from the model that the primary basic binding extends, in a less well-defined way, to the reactive center pole of the molecule—hence, the longer heparins may influence the electrostatic environment of the apposition site of thrombin and antithrombin. More appropriate to this paper is the conformational change that occurs on the binding of the pentasaccharide (Villanueva & Danishefsky, 1977). Antithrombin has an initial low affinity for the pentasaccharide with a subsequent high-affinity state (Olson et al., 1981). The binding constant for the low-affinity state is the same as the binding constant of cleaved antithrombin (Björk & Fish, 1982). We have also noted that some reactive center mutants of antithrombin are locked in the high-affinity state (Owen et al., 1988). Our deduction from this (with our colleagues M. C. Owen, P. B. Christey, and J.-Y. Borg) is that antithrombin has two conformational states, on the basis of changes at the primary heparin site, which influence the reactive center. The native state is the low-affinity conformation, as present in the locked form in cleaved antithrombin and represented sterically by the Löbermann et al. antitrypsin structure. The high-affinity state results from a conformational change at the binding site induced by the approach of heparin and transmitted reciprocally to the reactive center, hence the high-affinity state of the reactive center variants.

A detailed discussion of the likely mechanisms is not appropriate here, but we believe that the model of antithrombin, derived from cleaved antitrypsin, provides prospects of determining the likely conformational contribution to heparin-induced activation.

Natural Serpin Variants

A confirmation of the structure–function deductions made from the α_1-antitrypsin model is provided by natural variants of the serpins. Some 40 of these have now been characterized (Table IV, listing and references), and the mutations in each can be assessed in terms of the model. Overall there is a convincing correlation between the structural consequences predicted and the actual functional changes as observed in the affected individuals. Where no such functional changes are apparent, as in the physiologic polymorphisms, the underlying mutations are indeed seen to involve minimal structural perturbations on the model as in the substitutions of amino acids with similar properties in sterically noncritical areas. Examples in antitrypsin are the major common polymorphism V213 → A in s3C, the polymorphism E376 → D in s4B, the M2 polymorphism R101 → H in hD, and the nonpathological variants E204 → K in s4C and Christchurch E363 → K in s1C.

The pathological variants, i.e., those associated with significant abnormalities, are due to mutations that fall into three groups: those that directly affect the expression of the protein (which are not relevant to this discussion); those that affect a critical functional site in the molecule; those that affect the integrity of the overall structure. Examples of variants affecting critical functional sites include the mutations at the reactive center. These have been identified and show predicted functional consequences, in antitrypsin, antiplasmin, antithrombin, and C_1-inhibitor; they include alterations in the s4A reactive site loop as well as at the reactive center. The other major functional site is that of heparin binding; as previously

Table IV: Natural Serpin Variants[a]

structure	serpin and name	mutation	consequence	disease[b]	ref
prehA	antithrombin Rouen-III	7 Ile → Asn	new carbohydrate; heparin affinity decreased	thrombosis	Brennan et al., 1988
prehA	antithrombin Basel	41 Pro → Leu	perturbation of heparin site	thrombosis	Chang & Tran, 1986
hA	antithrombin Rouen-I	47 Arg → His	loss of heparin binding	thrombosis	Owen et al., 1987
hA	antithrombin Toyama	47 Arg → Cys	loss of heparin binding	thrombosis	Koide et al., 1984
hA	antithrombin Rouen-II	47 Arg → Ser	loss of heparin binding	thrombosis	Borg et al., 1988
hA	antitrypsin I	39 Arg → Cys	loss of salt bridge; deficiency	emphysema	Kalsheker et al., 1989
hA	antitrypsin M Procida	41 Leu → Pro	helix distortion; unstable	emphysema	Takahashi et al., 1987
s6B	antitrypsin I Malton	52 Phe deleted	misfolding; nonexpression	emphysema	Kalsheker et al., 1989
hD	thyroxine BG Gary	96 Ile → Asn	impaired T4 binding	?	Mori et al., 1986
hD	antitrypsin M2	101 Arg → His	apparently normal	polymorphism	
hD	heparin cofactor II	189 Arg → His	decreased dermatan sulfate binding		Blinder et al., 1989
s2A	antitrypsin Null Newport	115 Gly → Ser	steric distortion?; deficiency	emphysema	Kalsheker et al., 1989
hDs2A	antithrombin β	135 Asn	loss of carbohydrate; decreased heparin affinity	physiological	Brennan et al., 1987
hF	antitrypsin Null Granite Falls	160 Tyr → Stop	nonexpression	emphysema	Nukiwa et al., 1987
s4C	antitrypsin X	204 Glu → Lys	apparently normal	nil	Jeppsson & Laurell, 1988
s3C	antitrypsin polymorphism	213 Val → Ala	apparently normal	nil	Carrell et al., 1982
s3C	antitrypsin Null Bellingham	217 Lys → stop	nonexpression	emphysema	Satoh et al., 1988
ts3BhG	antitrypsin Null Cardiff	256 Asp → Val	impaired folding	emphysema	Kalsheker et al., 1989
hG	antitrypsin S	264 Glu → Val	loss of salt bridge; deficiency	emphysema	Owen & Carrell, 1976
thI1s5A	antitrypsin Null Hong Kong	318 Leu, TC deleted	nonexpression	emphysema	Sifers et al., 1988
s5A	antitrypsin Z	342 Glu → Lys	loss of salt bridge; incomplete processing	emphysema cirrhosis	Jeppsson, 1976
s4A	antithrombin Hamilton	382 Ala → Thr	loop variant; nonfunctional	thrombosis	Devraj-Kisuk et al., 1988
s4A	antitrypsin Null Mattawa	insert 353	termination position 376	emphysema	Curiel et al., 1988
s4A	antiplasmin Enschede	Ala insert 353–357	? loss of loop stress; nonfunctional	bleeding	Holmes et al., 1987
reactive center	antitrypsin Pittsburgh	358 Met → Arg	changed inhibitory specificity	bleeding	Owen et al., 1983
reactive center	C_1-inhibitor	444 Arg → His	nonfunctional	angioedema	Aulak et al., 1988b
reactive center	C_1-inhibitor	444 Arg → Cys	nonfunctional	angioedema	Aulak et al., 1988a
reactive center	antithrombin Glasgow	393 Arg → His	nonfunctional	thrombosis	Lane et al., 1987
reactive center	antithrombin Northwick Park	393 Arg → His	nonfunctional	thrombosis	Erdjument et al., 1988
reactive site	antithrombin Denver	394 Ser → Leu	reduced activity	thrombosis	Stephens et al., 1987
s1C	antitrypsin Christchurch	363 Glu → Lys	normal function	nil	Brennan & Carrell, 1986
s1Cs4B	antitrypsin M Heerlen	369 Pro → Leu	turning-point instability	emphysema	Hofker et al., 1987
s1Cs4B	antithrombin Utah	407 Pro → Leu	as above; disrupts turn	thrombosis	Bock et al., 1985
s4B	C_1-inhibitor	458 Met → Val	apparently normal	polymorphism	Bock et al., 1986
s4B	antitrypsin M3	376 Glu → Asp	normal function	polymorphism	Jeppsson & Laurell, 1988

[a] Adapted from Carrell et al. (1989). The sequence numbers refer to the residue numbers of the given protein. [b] Disease, i.e., predisposition to.

discussed, the variants of antithrombin and heparin cofactor II have helped to define this site but, conversely, also demonstrate the validity of the antitrypsin structure for the other serpins.

Further support for the model is given by the third group of variants, in which the overall integrity of the molecule is affected. Two good examples are the much-studied S and Z deficiency variants of antitrypsin which occur commonly in Northern Europeans. Both are due to substitutions of residues involved in key salt bridges—as seen on the model and as indicated by the consistent conservation of the residues involved. The Z variant has replacement of a glutamate in the bridge E342–K290 by a lysine; the S variant has replacement of the glutamate in the bridge E264–K387 by a valine. This variant has been studied crystallographically (Engh et al., 1989). In both variants there is evidence of decreased structural stability and in the Z variant the additional problem of a failure in transport/solubility at the final stage of the vesicular synthesis pathway. A second mutation of K290 → E in the Z variant appears to correct the defect (Brantley et al., 1988).

Another deficiency mutant, the I variant of antitrypsin, also has a similar consequence to that seen with the S mutation. This is in keeping with the structural change of the replacement, by cysteine, of the arginine at position 39 that contributes to the stabilization of the turn between helixes G and H and donates hydrogen bonds to the carbonyl oxygens projecting from the C-terminus of hG. The correlation between observed changes and the conformational predictions is also seen with other variants that affect the overall structure. The mutation of P369 → L, at the beginning of s4B, occurs in variants of both antitrypsin (M Heerlen) and antithrombin III (Utah, 407 in antithrombin numbering) with consequent gross dysfunction leading to a predisposition to emphysema with one and thrombosis with the other. A final example is that of two antitrypsin null variants—so-called because there is no circulating gene product. DNA analysis shows that one variant (Null Malton) has F52 deleted from an internal strand, s6B; the other (Null Cardiff) has the mutation D256 → V. In the first example β-sheet B would be obstructed; in the second a salt bridge between D256 and H231 was eliminated and turn ts3BhG destabilized. In both, misfolding of the molecule may result to explain the apparent nonexpression of the protein. In conclusion, experience in correlating and predicting the changes consequent on single amino acid substitutions supports our confidence that the Löbermann et al. model of α_1-antitrypsin provides a generally valid model of the native molecule in vivo.

References

Atanasov, B. P., & Karshikov, A. D. (1985) *Stud. Biophys. 105*, 11–22.

Aulak, M. S., Pemberton, P. A., Rosen, F. S., Carrell, R. W., Lachmann, P. J., & Harrison, R. A. (1988a) *Biochem. J. 253*, 615.

Aulak, M. S., Lachmann, P. J., Rosen, F. S., & Harrison, R.

A. (1988b) *Complement 5*, 181.

Beetz, T., & Van Boeckel, C. A. A. (1986) *Tetrahedron Lett. 27*, 5889–5892.

Björk, I., & Fish, W. W. (1982) *J. Biol. Chem. 257*, 9487–9493.

Blake, C. C. F., & Oatley, S. J. (1977) *Nature 268*, 115–120.

Blinder, M. A., Andersson, T. R., Abildgaard, U., & Tollefsen, D. M. (1989) *J. Biol. Chem.* (in press).

Bock, S. C., Harris, J. F., Schwartz, C. E., Ward, J. H., Hershgold, E. J., & Skolnick, M. H. (1985) *Am. J. Hum. Genet. 37*, 32.

Bock, S. C., Skriver, K., Nielsen, E., Thogersen, H. C., Wiman, B., Donaldson, V. H., Eddy, R. L., Marrinan, J., Radziejewska, E., Huber, R., Shows, T. B., & Magnusson, S. (1986) *Biochemistry 25*, 4292–4301.

Bode, W., Wei, An-Zhi, Huber, R., Meyer, E., Travis, J., & Neumann, S. (1986) *EMBO J. 5*, 2453–2458.

Bode, W., Greyling, H. J., Huber, R., Otlewsky, J., & Wilusz, T. (1989) *FEBS Lett. 242*, 285–292.

Borg, J. Y., Owen, M. C., Soria, C., Caen, J., & Carrell, W. R. (1988) *J. Clin. Invest. 81*, 1292–1296.

Brantley, M., Courtney, M., & Crystal, R. G. (1988) *Science 242*, 1700–1702.

Brennan, S. O., & Carrell, R. W. (1986) *Biochim. Biophys. Acta 873*, 13.

Brennan, S. O., George, P. M., & Jordan, R. E. (1987) *FEBS Lett. 219*, 431–436.

Brennan, S. O., Borg, J.-Y., George, P. M., Soria, C., Soria, J., Caen, J., & Carrell, R. W. (1988) *FEBS Lett. 237*, 118.

Bruch, M., Weiss, V., & Engel, J. (1988) *J. Biol. Chem. 32*, 16626–16630.

Carrell, R. W., & Owen, M. C. (1985) *Nature 317*, 730–732.

Carrell, R. W., & Travis, J. (1985) *Trends Biochem. Sci. 10*, 20.

Carrell, R. W., Jeppson, J.-O., Laurell, C.-B., Brennan, S. O., Owen, M. C., Vaughan, L., & Boswell, D. R. (1982) *Nature 298*, 329–334.

Carrell, R. W., Pemberton, P. A., & Boswell, D. R. (1987a) *Cold Spring Harbor Symp. Quant. Biol. 52*, 527–535.

Carrell, R. W., Christey, P. B., & Boswell, D. R. (1987b) in *Thrombosis and Haemostasis* (Verstraete, M., Vermylen, J., Lijnen, H. R., & Arnout, J., Eds.) pp 1–15, Leuven University Press, Leuven, Belgium.

Carrell, R. W., Aulak, K. S., & Owen, M. C. (1989) *Mol. Biol. Med.* (in press).

Chang, J.-Y., & Tran, T. H. (1986) *J. Biol. Chem. 261*, 1174–1176.

Choay, J., Petitou, M., Lormeau, J. C., Sinay, P., Cosu, B., & Gatti, G. (1983) *Biochem. Biophys. Res. Commun. 116*, 492–499.

Curiel, D., Brantly, M., Curiel, E., Stier, L., & Crystal, R. G. (1988) *Am. Rev. Respir. Dis. 137*, 210.

Defye, G., Basset, M., Monnier, N., & Chambaz, E. M. (1980) *Biochim. Biophys. Acta 623*, 280–294.

Devraj-Kizuk, R., Chui, D. H. K., Prochownik, E. V., Carter, C. J., Ofosu, F. A., & Blajchman, M. A. (1988) *Blood 72*, 1518.

Engh, R., Löbermann, H., Schneider, M., Wiegand, G., Huber, R., & Laurell, C.-B. (1989) *Protein Eng. 2*, 407–415.

Erdjument, H., Lane, D. A., Panico, M., Di Marzo, V., & Morris, H. R. (1988) *J. Biol. Chem. 263*, 5589.

Eriksson, U., Rask, I., & Peterson, D. A. (1984) *EMBO J. 3*, 1451–1454.

Flink, I. L., Bailey, T. J., Gustafson, T. A., Markham, B. E., & Markin, E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7708–7712.

Gettins, P. (1989) *J. Biol. Chem. 264*, 3781.

Gettins, P., & Horton, B. (1988) *Biochemistry 27*, 3634–3639.

Griffith, M. J., Noyes, C. M., & Church, F. C. (1985) *J. Biol. Chem. 260*, 2218–2225.

Hammond, G. L., Smith, C. L., Goping, I. S., Underhill, D. A., Harley, M. J., Revetos, J., Musto, N. A., Gunsalus, G. L., & Bardin, C. W. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5153–5157.

Hofker, M. H., Nukiwa, T., Van Paassen, H. M. B., Nelen, M., Frants, R. R., Klasen, E. C., & Crystal, R. G. (1987) *Am. J. Hum. Genet. 41*, A2200.

Holmes, W. E., Lijnen, H. R., Nelles, L., Kluft, C., Nieuwenhuis, H. K., Rijken, D. C., & Collen, D. (1987) *Science 238*, 209.

Huber, R. (1984) *Behring Inst. Mitt. 76*, 1–14.

Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H., & Kayser, H. (1987) *J. Mol. Biol. 198*, 499–513.

Hunt, L. T., & Dayhoff, M. O. (1980) *Biochem. Biophys. Res. Commun. 95*, 864–871.

Jeppsson, J.-O. (1976) *FEBS Lett. 65*, 195.

Jeppsson, J.-O., & Laurell, C.-B. (1988) *FEBS Lett. 231*, 327.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Kahn, M. S., & Rosner, W. (1977) *J. Biol. Chem. 252*, 1895–1900.

Kalsheker, N. A., Newton, C., Graham, A., Bamforth, F. J., Powell, S., & Markham, A. (1989) *J. Med. Genet.*, Abstract (in press).

Katsunuma, T., Tsuda, M., Kusumi, T., Ohkubo, T., Mitoni, T., Nakasaki, H., Tajima, T., Yokoyama, S., Kamiguchi, H., Kobayashi, K., & Shinoda, H. (1980) *Biochem. Biophys. Res. Commun. 93*, 552–557.

Koide, T., Odani, S., Tokahashi, K., Ono, T., & Sakuragawa, N. (1984) *Proc. Natl. Acad. U.S.A. 81*, 289–293.

Kotwal, G. J., & Moss, B. (1989) *J. Virol. 63*, 600.

Lane, D. A., Lowe, G. D. O., Flynn, A., Thompson, E., Ireland, H., & Erdjument, H. (1987) *Br. J. Haematol. 66*, 523.

Laurell, C.-B., & Eriksson, S. (1963) *Scand. J. Clin. Lab. Invest. 15*, 132–140.

Lindahl, U., Bäckström, G., & Thunberg, L. (1983) *J. Biol. Chem. 258*, 9826–9830.

Löbermann, H., Tokuoka, R., Deisenhofer, J., & Huber, R. (1984) *J. Mol. Biol. 177*, 531–556.

Mori, Y., Refetoff, S., Seino, S., Flink, I. I., & Murata, Y. (1986) *N. Engl. J. Med. 314*, 694 (Abstract).

Newcomer, M. E., Jones, T. A., Åqvist, J., Sundelin, J., Eriksson, U., Rask, I., & Peterson, P. A. (1984) *EMBO J. 3*, 1451–1454.

Nukiwa, T., Takahashi, H., Brantly, M., Courtney, M., & Crystal, R. G. (1987) *J. Biol. Chem. 262*, 11999.

Olson, S. T., Srinivasan, K. R., Björk, I., & Shore, J. D. (1981) *J. Biol. Chem. 256*, 11073–11079.

Owen, M. C., & Carrell, R. W. (1976) *Br. Med. J. 1*, 130.

Owen, M. C., Brennan, S. O., Lewis, J. H., & Carrell, R. W. (1983) *N. Engl. J. Med. 309*, 694.

Owen, M. C., Borg, J. Y., Soria, C., Soria, J., Caen, J., & Carrell, R. W. (1987) *Blood 69*, 1275–1279.

Owen, M. C., Beresford, C. H., & Carrell, R. W. (1988) *FEBS Lett. 231*, 317–320.

Pemberton, P. A., Stein, P. E., Pepys, M. B., Potter, J. M., & Carrell, R. W. (1988) *Nature 336*, 257-258.
Petersen, T. E., Dudek-Wojciechowska, G., Sottrup-Jensen, L., & Magnusson, S. (1979) in *Physiological Inhibitors of Coagulation and Fibrinolysis* (Collen, D., Wiman, B., & Verstraete, M., Eds.) pp 43-54, Elsevier/North-Holland Biomedical Press, Amsterdam.
Peterson, C. B., Noyes, C. M., Pecon, J. M., Church, F. C., & Blackburn, M. N. (1987) *J. Biol. Chem. 262*, 8061-8065.
Ragg, H. (1986) *Nucleic Acids Res. 14*, 1073-1088.
Satoh, K., Nukiwa, T., Brantly, M., Garver, R. I., Jr., Hofker, M., Courtney, M., & Crystal, R. G. (1988) *Am. J. Hum. Genet. 42*, 77.
Sifers, R. N., Brashears-Macatee, S., Kidd, V. J., Muensch, H., & Woo, S. L. C. (1988) *J. Biol. Chem. 263*, 7330.
Smith, J. W., & Knauer, J. (1987) *J. Biol. Chem. 262*, 11964-11972.
Stein, P. E., Tewkesbury, D. A., & Carrell, R. W. (1989) *Biochem. J.* (in press).
Stephens, A. W., Thalley, B. S., & Hirs, C. H. W. (1987) *J. Biol. Chem. 262*, 1044.
Tabachnik, M., & Perret, V. (1987) *Biochem. Int. 15*, 409-417.
Takahashi, H., Nukiwaa, T., Ogushi, F., Brantly, M., Courtney, M., & Crystal, R. G. (1987) *Am. Rev. Respir. Dis. 135*, A292.
Tanford, C., & Kirkwood, J. G. (1957) *J. Am. Chem. Soc. 79*, 5333-5339.
Toma, K., Yamaomoto, S., Deyashiki, Y., & Suzuki, K. (1987) *Protein Eng. 1*, 471-475.
Travis, J., & Salvesen, G. S. (1983) *Annu. Rev. Biochem. 52*, 655-709.
Villanueva, G. B., & Danishefsky, I. (1977) *Biochem. Biophys. Res. Commun. 74*, 803-809.

Accelerated Publications

Structure of the Human Hepatic Triglyceride Lipase Gene[†,‡]

Sheng-Jian Cai, Daphne M. Wong, San-Hwan Chen, and Lawrence Chan*
Departments of Cell Biology and Medicine, Baylor College of Medicine, Houston, Texas 77030
Received August 7, 1989; Revised Manuscript Received August 30, 1989

ABSTRACT: The structure of the human hepatic triglyceride lipase gene was determined from multiple cosmid clones. All the exons, exon–intron junctions, and 845 bp of the 5′ and 254 bp of the 3′ flanking DNA were sequenced. Comparison of the exon sequences to three previously published cDNA sequences revealed differences in the sequence of the codons for residues 133, 193, 202, and 234 that may represent sequence polymorphisms. By primer extension, hepatic lipase mRNA initiates at an adenine 77 bases upstream of the translation initiation site. The hepatic lipase gene spans over 60 kb containing 9 exons and 8 introns, the latter being all located within the region encoding the mature protein. The exons are all of average size (118–234 bp). Exon 1 encodes the signal peptide, exon 4, a region that binds to the lipoprotein substrate, and exon 5, an evolutionarily highly conserved region of potential catalytic function, and exons 6 and 9 encode sequences rich in basic amino acids thought to be important in anchoring the enzyme to the endothelial surface by interacting with acidic domains of the surface glycosaminoglycans. The human lipoprotein lipase gene has been recently reported to have an identical exon–intron organization containing the analogous structural domains [Deeb & Peng (1989) *Biochemistry 28*, 4131–4135]. Our observations strongly support the common evolutionary origin of these two lipolytic enzymes.

Hepatic triglyceride lipase is a lipolytic enzyme synthesized by the liver parenchymal cells and is localized primarily on the sinusoidal surface of the liver. It catalyzes the hydrolysis of tri-, di-, and monoglycerides, acyl-CoA thioesters, and phospholipids (Kuusi et al., 1982; Jensen et al., 1982; Jackson, 1983; Laboda et al., 1986). The enzyme hydrolyzes the triglycerides of intermediate-density lipoproteins to produce low-density lipoproteins and triglycerides and phospholipids of high-density lipoproteins ($HDL)_2$ to produce HDL_3 (Rao et al., 1982; Kinnunen et al., 1984). There is an inverse relationship between hepatic lipase activity and plasma HDL levels (Kuusi et al., 1983, 1987). Patients with familial hepatic triglyceride lipase deficiency accumulate high levels of HDL_2 in plasma, leading to hyper-α-triglyceridemia (Breckenridge et al., 1982; Little & Connelly, 1986; Carlson et al., 1986). Furthermore, studies in vivo and in vitro suggest that hepatic lipase is involved in the delivery of HDL phospholipid and cholesterol to the liver (Kussi et al., 1979; Jansen et al., 1980; Bamberger et al., 1983, 1985). The observation that high HDL levels protect against the development of atherosclerosis (Barr et al., 1951; Miller & Miller, 1975) has stimulated considerable interest in the potential role of hepatic lipase in HDL regulation and atherogenesis.

Cloned cDNAs of rat and human hepatic lipase have been isolated in a number of laboratories (Komaromy & Schotz, 1987; Stahnke et al., 1987; Datta et al., 1988; Martin et al., 1988; Semenkovich et al., 1989). The human gene for hepatic

[†]This research was supported by National Institutes of Health Grant HL 16512 and by grants from the March of Dimes Birth Defects Foundation and the Juvenile Diabetes Foundation.
[‡]The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02882.
*To whom correspondence should be addressed at the Department of Cell Biology.

0006-2960/89/0428-8966$01.50/0