Primary Structure of Tetranectin, a Plasminogen Kringle 4 Binding Plasma Protein: Homology with Asialoglycoprotein Receptors and Cartilage Proteoglycan Core Protein†

Jannie Fuhlendorff,*,‡ Inge Clemmensen,§ and Staffan Magnusson‡

Department of Molecular Biology and Plant Physiology, University of Aarhus, DK-8000 Århus C, Denmark, and Department of Clinical Microbiology, Hvidovre Hospital, DK-2650 Hvidovre, Denmark

Received February 23, 1987; Revised Manuscript Received May 28, 1987

ABSTRACT: Tetranectin binds to plasminogen and to isolated kringle 4 [Clemmensen, I., Petersen, L. C., & Kluft, C. (1986) *Eur. J. Biochem. 156*, 327–333], apparently to its lysine-binding site. Each of the four identical chains consists of 181 amino acid residues. The three intrachain disulfide bonds connect Cys residues 50–60, 77–176, and 152–168. The tetranectin sequence is homologous (17–24% identical positions) with those parts of the asialoglycoprotein receptor family that are considered to be extracellular. Tetranectin has no structures corresponding to those parts of the receptors considered to be intracellular and membrane anchoring. The sequence of tetranectin is also homologous (22–23% identical positions) with the C-terminal globular domain of the core protein of the cartilage proteoglycan. All six Cys residues in tetranectin are located at positions that are also Cys residues in this proteoglycan. Therefore, a plausible disulfide bond pattern can now be proposed for both the asialoglycoprotein receptors and the C-terminal domain of the proteoglycan core protein. No covalently bound carbohydrate has been found.

Tetranectin was first identified as a contaminant in α_2-antiplasmin preparations and shown to bind to plasminogen and more specifically to kringle 4 isolated from elastase digests of plasminogen (Clemmensen et al., 1986). The 11 kringle structures described until now are mutually homologous domains that have been found in five different blood plasma glycoproteins. All five are serine proteinase zymogens with functions in blood coagulation or fibrinolysis, namely, prothrombin (Magnusson et al., 1975), plasminogen (Sottrup-Jensen et al., 1978), urokinase (Günzler et al., 1982), tissue plasminogen activator (Pennica et al., 1983), and Hageman factor (factor XII) (Cool et al., 1985). Certain binding affinities of these proteins have been associated with some of the kringles, such as Ca^{2+} and phospholipid binding with the first (Esmon & Jackson, 1974; Esmon et al., 1974) and factor V binding with the second (Bajaj et al., 1975) kringle of prothrombin. Lysine binding has been found associated with one of the kringles 1, 2, and 3 and with kringle 4 of plasminogen (Sottrup-Jensen et al., 1978), the primary binding of α_2-antiplasmin with K1+2+3[1] (Wiman & Collen, 1978; Wiman et al., 1979), and tetranectin binding with kringle 4 of plasminogen (Clemmensen et al., 1986), as well as the fibrin binding with kringle 2 of tissue plasminogen activator (van Zonneveld et al., 1986). All the interactions of kringle structures with these different proteins has not yet made it possible to derive a general understanding of the function of these structures. Therefore, it was thought that knowledge of the primary structure of tetranectin would contribute to our understanding of kringle function.

Materials and Methods

Chemicals and Reagents. TPCK-treated trypsin was from Cooper Biomedical (A. Johnson, Worthington, Freehold, NJ), chymotrypsin and soybean trypsin inhibitor were from Millipore (Worthington, Freehold, NJ), and *Staphylococcus aureus* V8 protease was from Boehringer (Mannheim, FRG). Guanidine hydrochloride, iodoacetic acid, and DTE were purchased from MERCK (Darmstadt, FRG). Citraconic acid anhydride and polybrene were from Pierce (Rockford, IL). Radiolabeled iodoacetic acid was from Amersham International (Amersham, U.K.), DTNB from Fluka (Buchs, Switzerland), and Tris from Ferak (Berlin, FRG). SDS was obtained from BDR Chemicals Poole, U.K.). DEAE-Sephacel and CNBr-activated Sepharose 4B were from Pharmacia (Uppsala, Sweden), and Sephadex G-25 was a manufactured column (PD-10) from Pharmacia. Ultrogel ACA-34 was from LKB (Bromma, Sweden). TFA was from Rathburn (Peebleshire, U.K.). Spherisorb 5S ODS2 for reversed-phase HPLC was from Phase Separations (Deeside, U.K.), Vydac C_{18}, 5 μm, was from the Separation Group (Hesperia, CA), and Nucleosil C_{18}, 5 μm, and C_4, 5 μm, were from Macherey-Nagel (Düren, FRG). Gas-phase-grade reagents and solvents for the gas-phase sequencer were from Applied Biosystems (Forster City, CA). Ethyl acetate and *n*-butyl chloride were of HPLC grade from Rathburn. Membrane filters were from Amicon (Lexington, KY) and rabbit anti-

† This work was supported by grants from the National Institutes of Health (HL-16238 to S.M.), the Danish Cancer Society, The Danish Medical and Science Research Councils (S.M.), the NOVO Foundation (scholarship to J.F.), Dakopatts (I.C.), and the Danish Medical Research Council (12-4924 to I.C.).

* Correspondence should be addressed to this author at the Vaccine Department, State Serum Institute, DK-2300 København S, Denmark.
‡ University of Aarhus.
§ Hvidovre Hospital.

[1] Abbreviations: ASGR, asialoglycoprotein receptors; CCP, chicken cartilage proteoglycan core protein; CHL, chicken hepatic lectin; CMCys, *S*-(carboxymethyl)cysteine; DTE, dithioerythritol; DTNB, 5,5′-dithiobis(2-nitrobenzoate); EtOH, ethanol; H1 and H2, human asialoglycoprotein receptors 1 and 2; HPLC, high-performance liquid chromatography; K1+2+3, elastase fragment of plasminogen containing the first three kringles; K4, fourth kringle of plasminogen; PITC, phenyl isothiocyanate; PTH, phenylthiohydantoin; RCP, rat cartilage proteoglycan core protein; RHL, rat hepatic lectin; SDS, sodium dodecyl sulfate; TPCK, N^{α}-tosylphenylalanine chloromethyl ketone; TFA, trifluoroacetic acid; TN, tetranectin; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/87/0426-6757$01.50/0

human plasma proteins from Dakopatts (code A134, Copenhagen, Denmark).

Tetranectin. Tretranectin was isolated by a procedure different from that already published (Clemmensen et al., 1986). Serum (500 mL) from rabbits immunized with TN was precipitated with ammonium sulfate and ion exchanged as described (Harboe & Ingild, 1973). The IgG (15 g) was coupled to 50 g of CNBr-activated Sepharose 4B. The anti-TN-coupled Sepharose was incubated with 1.8 L of cryoprecipitate-depleted plasma containing 15–20 mg of TN as measured by electroimmunoassay (Clemmensen et al., 1986) at 20 °C for 1 h. The gel was washed with 0.5 M NaCl–0.05 Tris-HCl buffer (pH 7.4) until $A_{280nm} < 0.010$. TN was eluted from a column (2.6 × 28 cm) with 3 M $MgCl_2$ and dialyzed against the Tris-HCl buffer. The TN solution (~300 mL; 45 mg/L) was concentrated 80 times by ultrafiltration through Amicon PM-10 membranes and then run through a rabbit anti-human plasma protein column (100 mL of the Tris buffer). At this stage the yield of TN was approximately 10–11 mg. The TN was then gel filtered on a column of Ultrogel ACA-34 (2.6 × 90 cm) in 0.05 M Tris-HCl–0.1 M NaCl (pH 7.4). The elution was monitored at every step by electroimmunoassay with anti-TN. The final yield of TN was approximately 5–6.5 mg. The material was finally dialyzed against distilled water and lyophilized.

Sequence Strategy. Six enzymatic digests were performed on a total amount of 5.3 mg of tetranectin.

Reduction and S-^{14}C-Carboxymethylation (Hirs, 1967a,b). Prior to enzymatic digestion, 2.2 mg of TN was reduced and S-carboxymethylated. The mixture was tested for free thiol groups as in Maeda et al. (1970). An amount of 1.4 mg of TN was desalted by gel filtration on a column of Sephadex G-25 (9.1 mL) in 0.1 M NH_4HCO_3 (pH 8.3) and lyophilized before incubation with *S. aureus* V8 protease.

An amount of 1.4 mg of TN was reduced and S-carboxymethylated as above but dissolved in 1 mL of 6 M guanidinium chloride–0.1 M sodium pyrophosphate, pH 8.3.

Selective Cleavage. Arginyl bonds were cleaved by tryptic digestion of 1.4 mg of N-citraconylated protein (Dixon & Perham, 1986). The protein was desalted by gel filtration on a column of Sephadex G-25 (9.1 mL) in 0.1 M NH_4HCO_3 (pH 8.3), lyophilized, redissolved in 300 μL of 0.1 M NH_4HCO_3 (pH 8.3), and treated with 14 μg (1% w/w) of trypsin for 3 h at 37 °C. The reaction was terminated by addition of 9 μg of soybean trypsin inhibitor. The lysine residues were deblocked in sodium citrate buffer (pH 2.1) overnight at 20 °C. The mixture of peptides was fractionated by ion-exchange chromatography on a column of DEAE-Sephacel (8.3 mL, 0.9 × 13.0 cm) with a linear gradient of NH_4HCO_3 (0.1–1.0 M), pH 8.3. The peptides were further purified by reversed-phase HPLC on a column of Nucleosil C_{18}.

Lysyl and arginyl bonds were cleaved by digestion of 0.6 mg of TN with trypsin (6 μg of trypsin, 1% w/w) for 3.5 h at 37 °C in 0.6 mL of 0.1 M NH_4HCO_3 (pH 8.3). The reaction was terminated by addition of 4 μg of soybean trypsin inhibitor. The peptide mixture was fractionated by reversed-phase HPLC on a column of Vydac C_{18}.

Glutamyl bonds were cleaved by digesting 2.2 mg of reduced and S-carboxymethylated TN with *S. aureus* V8 protease (Houmard & Drapeau, 1972). An amount of 1.4 mg of reduced and S-carboxymethylated TN was incubated with *S. aureus* V8 protease (2% w/w) for 18 h at 37 °C in 0.5 mL of 0.1 M NH_4HCO_3 (pH 8.0). For initial peptide separation, the mixture was diluted 10 times with water and applied to a column of DEAE-Sephacel (8.0 mL, 0.9 × 12.5 cm). The column was eluted by a linear gradient of NH_4HCO_3 (10 mM–1.0 M, pH 8.3). The resulting fractions were lyophilized and further separated by reversed-phase HPLC on a column of Vydac C_4 or Vydac C_{18}. After reduction and S-carboxymethylation, 0.8 mg of TN was diluted with water to less than 4 M urea and then incubated with *S. aureus* V8 protease (2% w/w) for 24 h at 37 °C. The resulting peptides were separated directly by reversed-phase HPLC on a column of Nucleosil C_{18}.

Chymotryptic Digestion. An amount of 0.6 mg of TN was cleaved with chymotrypsin (6 μg of chymotrypsin, 1% w/w) in 0.6 mL of 0.1 M NH_4HCO_3 (pH 8.3) for 3 h at 37 °C. The reaction was stopped with 4.4 μg of phenylmethanesulfonyl fluoride in 99% EtOH (0.8 μg/μL), and the resulting peptides were separated by reversed-phase HPLC on a column of Nucleosil C_{18}.

An amount of 0.5 mg of TN was digested with both chymotrypsin and trypsin (5 μg of trypsin and 5 μg of chymotrypsin, 1% w/w) for 4 h at 37 °C in 0.2 mL of 0.1 M NH_4HCO_3 (pH 8.0). The peptides were immediately separated by reversed-phase HPLC on a column of Vydac C_{18}.

Peptide Separation by Reversed-Phase HPLC. The solvents used are (A) 0.1% TFA in H_2O and (B) 96% EtOH. The column dimensions were 4.6 × 250 mm.

Amino Acid Analysis. Amino acid compositions of the intact protein and the peptides were determined on a Beckman 121 MB amino acid analyzer. Hydrolyses were performed at 110 °C for 18 h in 6 M HCl–0.1% phenol in glass tubes sealed at low pressure. Corrections for handling loss in the hydrolysates of the intact protein were made by including norleucine as an internal standard.

Performic Acid Oxidation. This was carried out with freshly prepared performic acid (Hirs, 1967a,b).

Sequence Analysis. Automated sequence analysis was performed on a gas-phase instrument (Applied Biosystems 470A protein "sequencer") (Hewick et al., 1981). Peptide samples of 0.5–4.0 nmol were used for sequence analysis. For N-terminal sequencing of the intact protein, 2 nmol was used, dissolved in distilled water containing 0.1% SDS. The PTH derivatives of amino acids were identified by reversed-phase HPLC on a column of Spherisorb ODS2 (Sottrup-Jensen et al., 1980).

Cysteine residues were identified as PTH–CMCys, as the methyl ester of PTH-CMCys, and as radioactivity incorporated in a certain step as counted in a liquid scintillation counter (Intertechnique SL 30 liquid scintillation spectrometer). The radioactivity of PTH derivatives of ^{14}C-labeled CMCys residues was measured with the remaining 36% of the PTH derivatives from the sequencer dissolved in 50 μL of 96% EtOH and 4.5 mL of scintillation liquid. The scintillation liquid was xylene containing 65 g of 2,5-diphenyloxazole/25 L of xylene and Triton X-100, 5:1.

Esterification of PTH derivatives was performed for 15 min at 37 °C in 50 μL of 1 M methanolic HCl prepared as in Hunkapiller and Hood (1983).

Disulfide-Linked Peptides. All pools from HPLC separations containing $^1/_2$-Cys after hydrolysis were sequenced.

Search for Homologous Sequences. The PIR database (updated May 1985) (NBRL Ver 5.0 Riverside Scientific Enterprise, Seattle, WA) and a file of "interesting" newer sequences were searched for sequences homologous with TN. The program GENEPRO was used.

Results

The strategy used to deduce the primary structure of tetranectin was to determine the amino acid sequences of as many

FIGURE 1: Summary of the sequence evidence for tetranectin. The proven sequences of specific peptides are given in one-letter code (IUPAC–IUB Commission on Biochemical Nomenclature, 1968) below summary sequences (bold type). Prefixes A, C, R, and T denote peptides generated by digestion of tetranectin with *S. aureus* V8 protease, chymotrypsin, trypsin after citraconic anhydride treatment, and trypsin, respectively. Peptide sequences written in upper-case letters were determined by Edman degradation; residues in lower-case letters have been tentatively identified. Underlined lower-case letters denote residues deduced only from compositional analysis. Those not identified in a peptide are shown by dashes. An arrow indicates that the peptide is longer but that no further residues were identified by Edman degradation.

as possible of a complete set of peptides generated by specific cleavage at arginyl bonds and then to align these peptides by use of overlaps identified in other digests (Figure 1). Overlap peptides were obtained by cleavage of TN with trypsin, chymotrypsin, and *S. aureus* V8 protease. To locate the disulfide bonds, a tryptic and a chymotryptic digest were made on intact TN. Finally, a combined tryptic and chymotryptic digest was used to obtain evidence for the last missing disulfide bond.

Attempted Cleavage at Arginyl Bonds. Tryptic digestion of N-citraconylated TN yielded 11 peptides representing 80% of the sequence (Figure 1). Of these 11 peptides (R1–R11) the sequences of R8 and part of R3 are contained in that of R7; that of R10 is contained in that of R9. Tryspin cleaved all the arginyl bonds, but obviously, not all lysine side chains were fully blocked by citraconic acid anhydride, and as a result 6 of the 16 lysyl bonds were completely cleaved. The N-terminal part of the sequence (residues 1-29) was not found as tryptic peptides, and no further search for this part was performed.

The 11 peptides were obtained in yields of 2–30 nmol after separation by HPLC (Table I). In some cases the low yield was caused by poor recovery from the chromatographic system. This is especially true for the peptides in pool R3, which initially yielded ~30 nmol, but after a further separation by reversed-phase HPLC (Vydac C_{18} column, data not shown) the peptide already sequenced was in a yield lower than 1 nmol. Another reason for low yields was the incompletely blocked lysine side chains, resulting in overlapping variant peptides. Both peptides R7 and R8 showed heterogeneity in residues 34 and 37 giving PTH derivatives of both Ala and Ser as residue 34 and of both Val and Met as residue 37. Peptide R3 was probably contaminated with more than one peptide, which, however, did not show up in the sequencing.

Cleavage at Glutamyl Bonds. Cleavage of TN with *S. aureus* V8 protease gave four useful peptides, A1–A4, A1–A3, and A4 from two digests, respectively. The peptides from the first digest were separated on a column of DEAE-Sephacel before purification by reversed-phase HPLC, giving A1–A3.

Table I: Amino Acid Composition[a] of Peptides from Tetranectin and Intact Tetranectin[b]

	R1	R2	R3[c]	R4	R5	R6	R7[d]	R8	R9	R10	R11
CMCys (CMC)		2.0 (1)[e]	1.4 (1)[f]	0.6 (1)	1.0 (1)		1.2 (1)		4.4 (2)[e]	g (1)	
Asx (N/D)		1.3 (1)	0.8 (0)	1.1 (0)	3.6 (4)	1.7 (2)	1.3 (1)	1.0 (1)	1.3 (1)	1.2 (1)	3.6 (4)
Thr (T)		1.1 (1)	1.2 (1)	1.7 (2)	2.8 (3)	2.7 (3)	1.8 (2)	0.9 (1)			1.7 (2)
Ser (S)		1.7 (2)	1.1 (0)	0.8 (0)	1.3 (1)	1.5 (1)	0.4 (0.5)	0.2 (0.5)			1.1 (1)
Glx (Q/E)		2.0 (2)	3.3 (4)	1.6 (1)	4.8 (4)	3.2 (3)	4.4 (6)	2.2 (2)	1.7 (2)	2.0 (2)	4.1 (4)
Pro (P)			0.5 (0)	0.4 (0)	1.3 (1)	1.1 (1)			0.7 (1)	1.2 (1)	
Gly (G)			0.9 (0)	0.6 (0)	3.3 (4)	3.2 (4)			1.0 (1)	0.9 (1)	3.3 (4)
Ala (A)	1.1 (1)	1.0 (1)	1.2 (1)	1.3 (1)	4.0 (4)	1.2 (1)	3.0 (2.5)	1.7 (1.5)			3.6 (4)
1/2-Cys (C)											
Val (V)		0.7 (1)	1.0 (1)	1.2 (1)			2.3 (1.5)	1.1 (0.5)	0.8 (1)[h]	0.3 (1)[h]	2.3 (2)
Met (M)			0.2 (1)				(0.5)	(0.5)			0.7 (2)
Ile (I)	0.9 (1)	1.0 (1)	0.4 (0)	0.5 (0)	1.3 (1)				1.2 (2)[h]	0.3 (2)[h]	1.1 (1)
Leu (L)			2.0 (2)	1.2 (1)	1.3 (1)	3.0 (3)	5.2 (6)	2.9 (4)	1.1 (1)	0.8 (1)	2.2 (2)
Tyr (Y)	1.1 (1)			0.4 (0)	0.5 (0)	2.0 (2)			1.0 (1)	0.6 (1)	
Phe (F)		1.0 (1)	0.4 (0)	1.4 (2)	0.4 (0)				1.2 (1)	1.0 (1)	
Lys (K)	1.0 (1)		1.1 (1)	2.0 (2)	1.7 (2)		1.5 (2)	1.0 (1)			
His (H)		1.1 (1)		0.6 (1)							
Trp (W)					(1)						(2)
Arg (R)		1.1 (1)	0.5 (0)	0.4 (0)		1.0 (1)	0.3 (0)		1.0 (1)		1.0 (1)
total residues	4	12	11	13	28	21	23	12	14	12	29
yield (nmol)	7	3	30	3	26	14	4	7	8	4	6
position	131–134	69–80	42–52	56–68	135–162	81–101	30–(52)	30–41	168–181	170–181	102–130
contaminating peptide				131–162	161–(167)		28–41				

	T1	T2	T3	C1	C2	C3	C4a[i]	C5	C6	C7	C8
CMCys (CMC)											
Asx (N/D)	1.0 (1)		0.6 (0)	1.2 (1)	1.1 (1)	0.3 (0)	1.1 (1)	6.3 (6)	3.1 (0)	0.9 (1)	1.7 (2)
Thr (T)			2.8 (2)	0.9 (1)	0.9 (0)	0.9 (1)		3.6 (3)	2.8 (2)		0.9 (1)
Ser (S)			0.6 (0)	0.4 (0)	0.8 (1)		0.9 (1)	1.4 (1)		0.6 (1)	
Glx (Q/E)		1.7 (2)	5.3 (5)	1.6 (2)	0.3 (0)			8.6 (5)	3.2 (1)	3.4 (4)	1.2 (1)
Pro (P)				2.2 (3)				3.0 (2)	1.3 (0)		
Gly (G)			0.5 (0)		1.4 (2)	1.0 (1)	1.9 (2)	4.6 (4)	1.6 (1)	0.8 (1)	2.0 (2)
Ala (A)	0.4 (0)		2.3 (2)		2.0 (2)	2.0 (2)	2.0 (2)	5.6 (4)	1.9 (0)	1.3 (1)	2.4 (2)
1/2-Cys (C)			(2)					(2)	3.3 (2)		
Val (V)			1.2 (1)	1.0 (1)	0.3 (0)			2.0 (1)	2.4 (2)	1.0 (1)	
Met (M)		1.3 (1)	0.3 (0)						0.1 (1)		0.3 (1)
Ile (I)				0.6 (1)		0.8 (1)		1.4 (1)	0.8 (0)	0.7 (1)	
Leu (L)		1.3 (1)	3.1 (2)					5.3 (2)	2.8 (1)	0.9 (1)	1.7 (2)
Tyr (Y)						0.8 (1)		1.4 (1)	0.7 (0)	0.6 (1)	
Phe (F)	1.0 (1)	1.0 (1)	2.0 (2)		0.4 (0)		1.0 (1)	3.0 (1)	1.7 (1)		
Lys (K)	1.0 (1)	1.0 (1)	2.1 (2)	2.8 (3)	1.5 (1)		1.1 (1)	4.6 (4)	4.0 (3)	0.5 (0)	
His (H)									0.6 (1)		
Trp (W)	(1)				(1)		(1)	(2)		(1)	(1)
Arg (R)	1.0 (1)					0.9 (1)		2.0 (2)	1.1 (0)	0.6 (1)	
total residues	5	6	20	12	8	7	9	41	15	14	12
yield (nmol)	9	8	6	3	2	3	20	3	3	9	3
position	163–167	22–27	42–52, 60–68	1–12	156–163	127–133	156–164	134–(164), 165–174	47–51, 52–61	98–111	112–123
contaminating peptide	131–134		42–59		65–70			1–23[j]	134–(155), 164–174		

	A1	A2[d]	A3	A4		intact protein[b]
CMCys (CMC)					Cya	6.6
Asx (N/D)	1.6 (1)	3.9	2.0 (2)	1.8 (1)	Asx (N/D) (8/9)	14.8–18.1
Thr (T)	1.7 (1)	2.7	1.7 (2)	0.6 (0)	Thr (T) (16)	12.9–17.2
Ser (S)	1.5 (1.5)	1.8		0.9 (1)	Ser (S) (6.5)	10.1–14.2
Glx (Q/E)	3.3 (2)	7.5	0.8 (1)	2.4 (3)	Glx (Q/E) (11/15)	24.0–28.8
Pro (P)	0.5 (0)	1.7		0.5 (0)	Pro (P) (6)	5.0–9.4
Gly (G)	0.7 (0)	4.8	1.8 (2)	1.0 (1)	Gly (G) (14)	13.6–18.1
Ala (A)	1.5 (0.5)	3.3	1.8 (2)	1.0 (1)	Ala (A) (15.5)	13.9–17.3
1/2-Cys (C)		1.3			Cys (C) (6)	1.5–4.0
Val (V)	0.4 (0)	1.7	1.1 (1)	1.2 (1)	Val (V) (9.5)	7.5–11.0
Met (M)	0.2 (0)		0.3 (1)		Met (M) (4.5)	2.2–6.2
Ile (I)		1.4	0.9 (1)		Ile (I) (7)	5.5–7.6
Leu (L)	3.0 (0)	3.4		1.3 (1)	Leu (L) (15)	12.9–15.9
Tyr (Y)		1.0	1.0 (1)	0.7 (1)	Tyr (Y) (4)	3.9–5.6
Phe (F)		2.0		0.5 (0)	Phe (F) (6)	5.8–7.2
Lys (K)	1.3 (1)	3.2	1.1 (1)	1.0 (1)	Lys (K) (16)	13.4–16.9
His (H)		0.8			His (H) (2)	1.8–3.4
Trp (W)			(2)		Trp (W) (4)	
Arg (R)	1.0 (1)	1.6	0.8 (1)	1.0 (1)	Arg (R) (6)	5.3–7.7
total residues	11	(76)[k]	17	11		
yield (nmol)	11	4	19	4		
position	26–36	43–(72), 73–(92), 151–(170), 157–(181)	121–137	99–109		
contaminating peptide	121–(150)		1–16, 17–24(25)			

[a] Residues per peptide by amino acid analysis (6 M HCl, 0.1% phenol, 110 °C, 18 h). Values less than 0.3 are not reported except for Met and in the cases of heterogeneity. Values in parentheses indicate that the peptide was not sequenced to the end. Ratios are calculated to the integral value underlined. One-letter amino acid abbreviations (IUPAC–IUB Commission on Biochemical Nomenclature, 1968) are in parentheses. Values are not corrected for identified

Footnotes to Table I continued

contaminants. [b]Residues per polypeptide by amino acid analysis (conditions as above). The values are given as intervals on the basis of values from seven analyses of seven batches. Cya represents cysteinesulfonic acid determined by performic acid oxidation of tetranectin. [c]The peptide is not pure; further purification has been tried, but there was too little material. [d]Peptide not sequences to the end. [e]Decomposing CMCys in the standard mixture could account for the varying values in peptides. [f]Value for CMCys calculated by adding the value of CMCys and $^1/_2$-Cys in the composition. [g]Counts in peptide but no sign of CMCys in the composition. [h]Low value due to incomplete hydrolysis of Ile-Val (residues 180–181). [i]Hydrolysis of peptide C4 after further purification by HPLC. [j]Edman sequence analysis was performed on the purified disulfide-linked peptide C5a. [k]A mixture of four long peptides, none of which was sequenced to the end.

Peptide A1 was not pure, and A2 was a mixture of four peptides, which could all be analyzed by Edman degradation. The recoveries of A1 and A2 were too poor to allow further separation by HPLC. The heterogeneity in residue 34 was resolved by the evidence from A1. Only peptide A4 from the second V8 digest was obtained in sufficient amount.

The conditions for *S. aureus* V8 protease were chosen with the intent to restrict cleavage to glutamyl bonds. No aspartyl bonds were in fact cleaved. Anomalous cleavage by *S. aureus* V8 protease had generated one of the four peptides in A2 resulting from spurious cleavage at Ser-156.

Peptides with Intact Disulfide Bonds. (A) Tryptic Digestion. Digesting unreduced TN with intact disulfide bonds with trypsin and separating the peptides by reversed-phase HPLC gave three peptides T1–T3, which were further analyzed. Although T1 and T3 were not pure, it was possible to deduce their amino acid sequences. T1 was contaminated with the tetrapeptide R1 (residues 131–134). T3 contained the disulfide bond Cys(1)–Cys(2) (residues 42–52 and 60–68), a variant of the same peptide (residues 42–59 and 60–68), and some contaminant that did not interfere with the Edman degradation. T1 and T2 filled in two stretches of the sequence not obtained as pure peptides in the tryptic digest of citraconic acid anhydride treated TN.

(B) Chymotryptic Digestion. The chymotryptic peptides C1–C4 and C6–C8 were not further purified, but fraction C5 was. The resulting peptide C5a was sequenced without preceding compositional analysis, giving proof that residues 134–155 and 165–174 are disulfide linked [Cys(4)–Cys(5)]. Fraction C6 was a mixture of four peptides linked two by two.

The yields of the eight chymotryptic peptides obtained ranged from 2 to 20 nmol (Table I). This wide variation of yields could be explained by overlapping cleavage variants, C5a and C6, C1 and C5, and C2 and C4. Other cleavage variants were sequenced in peptide mixtures and are not shown here. The results can be explained in terms of the sequence deduced. Residue 4 has been deduced only from compositional analysis of C1, and though sequenced many times, the fourth step from the N-terminal gave no PTH amino acid derivative, indicating that Thr-4 is probably posttranslationally modified. The composition of C1 showed no sign of *N*-acetylgalactosamine.

(C) Concurrent Trypsin and Chymotrypsin Digestion. Trypsin and chymotrypsin were used simultaneously to provide the disulfide-linked peptide containing Cys(3)–Cys(6). The major peaks were analyzed for composition, and those containing even trace amounts of $^1/_2$-Cys were selected for Edman sequencing. One of the peaks, CT1 (~2.2 nmol), gave a double sequence covering residues 71–80 and 75–78 and therefore must contain Cys(3) and Cys(6). The peptides that contributed evidence for the disulfide bonds are CT1, C5a, C6, and T3.

Compositional Analysis. The amino acid compositions of intact TN from seven batches are summarized in Table I. This shows a serine content that is too high compared with that found by sequencing. This is true for all six batches of TN used for the primary structure determination. The serine content was somewhat variable, indicating a content of 10–14 serines for 6.5 found in the sequence. The "best" compositional analysis of the intact protein was obtained on a batch further purified on an antifibrinogen column (data not shown), which reduced a high molecular mass contaminating band in SDS–PAGE to a level not detectable. The amino acid composition of this batch showed a Ser content of 10.1 per monomer of TN, so the high molecular mass contaminant did not account for all the extra serine residues in the amino acid composition of the TN batches.

The N-terminal sequencing of the intact protein was disturbed by the three Pro residues present in steps 2, 3, and 7. Extending the cleavage time to 30 min instead of 15 did not reduce the "carry-over" due to incomplete cleavage of prolyl–anilinothiazolinone.

Search for Homology. When stretches of 25 residues of tetranectin were compared with the sequences in the database, no relevant homologies were found to other proteins with a window of 9 identities in 25 residues. The homology with the asialoglycoprotein receptors was found by coincidence with GENEPRO software system.

Homology with Asialoglycoprotein Receptors. An alignment of the amino acid sequences for TN and a family of homologous proteins, the asialoglycoprotein receptors, known from chicken, chicken hepatic lectin (CHL) (Drickamer, 1981), from rat, rat hepatic lectins 1 and 2/3 (RHL-1 and RHL-2/3) (Drikamer et al., 1984), and from man, human asialoglycoprotein receptors 1 (Spiess et al., 1985) and 2 (Spiess & Lodish, 1985) (H1 and H2), is shown in Figure 2.

In TN and the asialoglycoprotein receptors there are 21 of 181 residues (12%) that are identical in all the aligned sequences, and TN is identical with at least one of the asialoglycoprotein receptors in each of 55 positions (30% of the total compared). If the region containing residues 11–36 in TN is compared to the corresponding region of H2, 13 residues of 26 are identical, or 50%. An observation that strengthened the homology is the conservation of the cysteine residues and partly the tryptophan residues in the sequences. The duplicated domain in RHL-1, RHL-1*, is lying in front of the region with the common cysteine residues. On the basis of the disulfide bonds identified in TN and the extensive homology, it is possible to predict that the Cys residues corresponding to Cys(1)–Cys(2), Cys(3)–Cys(6), and Cys(4)–Cys(5) in TN are also linked in the five homologous receptors (or lectin) structures. One pair of cysteines in the asialoglycoprotein receptors (ASGR) is not found in TN; most likely, that pair is linked together in the ASGR.

Homology with the Core Protein of Rat Cartilage Proteoglycan. The homology between tetranectin and the C-terminal globular domain of the two cartilage proteolgycan core proteins from chicken and rat concerns essentially the same amino acid residues that are conserved between tetranectin and the asialoglycoprotein receptors. Amino acid residues 9–184 of the rat cartilage protein (RCP) (Doege et al., 1986) and 108–292 of the chicken cartilage protein (CCP) (Sai et al., 1986) have been aligned with the 181 residues of tetranectin (Figure 2). With seven deletions in RCP and six in CCP, there are 42 and 40 identities, respectively, corresponding to 23% and 22% identity. The two cartilage proteins have 114 identical positions in 176 common residues (65%

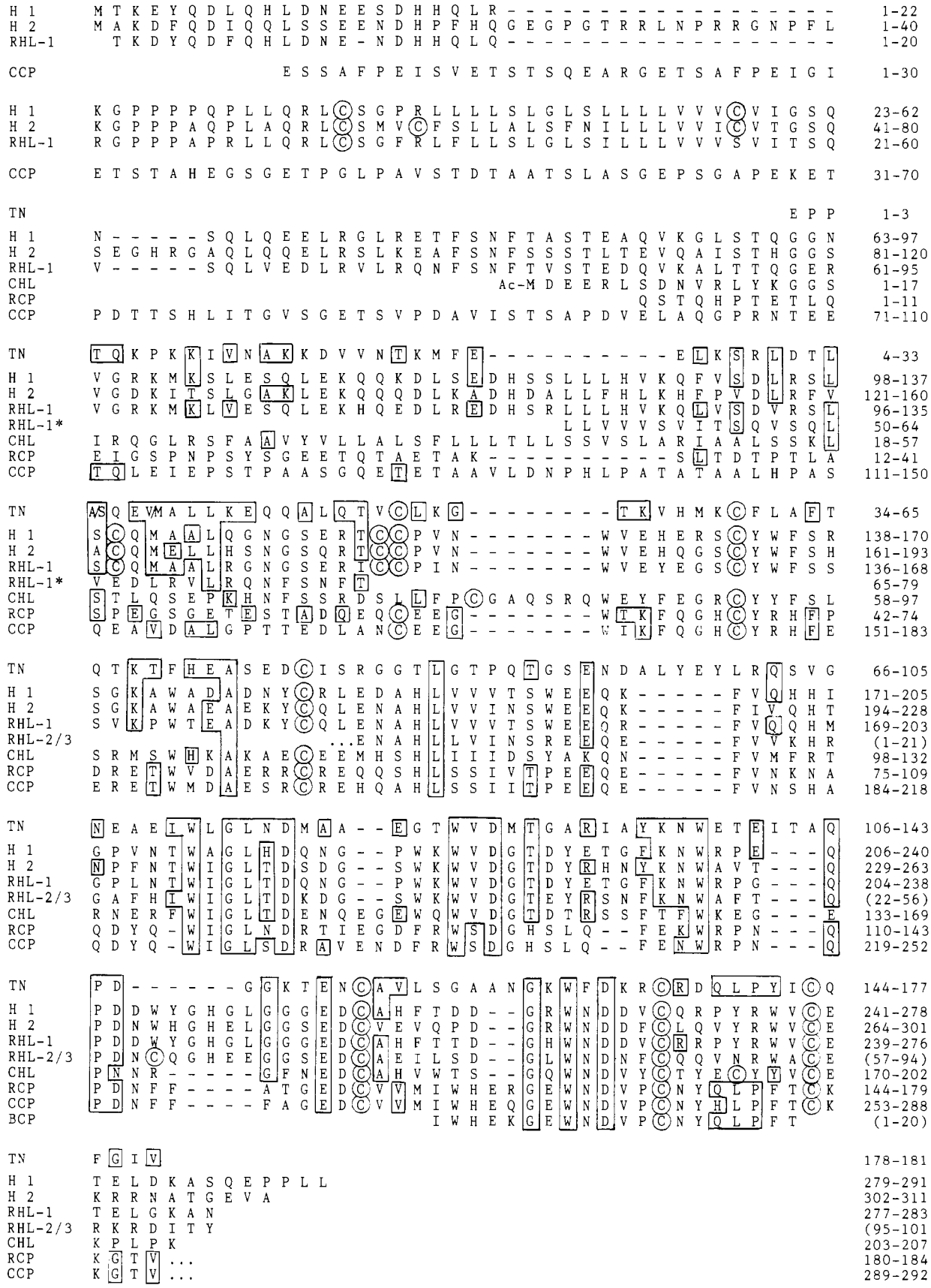

FIGURE 2: Alignment of the amino acid sequences of tetranectin, human asialoglycoprotein receptors 1 and 2, rat hepatic lectins 1 and 2/3, chicken hepatic lectin, rat cartilage proteoglycan core protein, chicken cartilage proteoglycan core protein, and a fragment of bovine cartilage proteoglycan core protein. TN: tetranectin (181 residues). H1 and H2: human asialoglycoprotein receptors 1 and 2 (291 and 311 residues, respectively) (Spiess et al., 1985; Spiess & Lodish, 1985). RHL-1: rat hepatic lectin (283 residues). RHL-1* is an internally duplicated domain (residues 50–70) of RHL-1 homologous to residues 121–150 (Drickamer et al., 1984). RHL2/3: rat hepatic lectin 2 and 3 partial sequence (101 residues) (Drickamer et al., 1984). CHL: chicken hepatic lectin (207 residues) (Drickamer, 1981). RCP: rat cartilage proteoglycan core protein, C-terminal globular domain (the first 184 residues) (Doege et al., 1986). CCP: chicken cartilage proteoglycan core protein, C-terminal globular domain (the first 292 residues) (Sai et al., 1986). BCP: bovine cartilage proteoglycan core protein, C-terminal globular domain, 20 residues (Périn et al., 1981), used to design a synthetic probe for cloning RCP and CCP.

identity). In the parts of their sequences that have been aligned with tetranectin, both RCP and CCP have six Cys residues in exactly the same positions as the six Cys residues in tetranectin. Thus, we can predict that the C-terminal domain of the proteoglycan core protein has three disulfide bonds corresponding to the three in tetranectin.

Discussion

The primary structure of tetranectin has been determined. The monomer consists of 181 amino acid residues with three disulfide bonds connecting Cys-50 to Cys-60, Cys-77 to Cys-176, and Cys-152 to Cys-168. No hexasamines were found in the amino acid composition analysis. Thr-4 was not identified during sequencing, indicating that it may be post-translationally modified. Phosphate analysis (data not shown) was carried out on two batches of TN. One of the batches was completely free of phosphate. The amino acid composition of TN was found to be in agreement with the sequencing results except that more serine was present than was found by sequencing. This discrepancy has not been explained. The heterogeneities found in positions 34 and 37 may be due to genetic polymorphism, each of the amino acids found in each position representing one allele, indicating the presence of two alleles. The TN used had been purified from pooled plasma, each batch being prepared from approximately eight blood donors. Our data indicate that the two alleles occur with similar frequencies. Each polymorphism can be explained as a single base mutation (A → C for Ser → Ala and T → C for Met → Val).

The binding of tetranectin to kringle 4 of plasminogen (plg) is probably mediated by the lysine binding site of kringle 4, because TN can be eluted from K4–Sepharose with the same kind of Lys analogoues as elute plg from Lys–Sepharose. If plg is removed from plasma by running the plasma through a column of Lys–Sepharose, plg binds to the Lys–Sepharose without TN being bound to plg. This is used to separate plg and TN in the purification procedure developed by I. Clemmensen (Clemmensen et al., 1986). It is not possible to draw any conclusions about the function of TN in plasma. Asialoglycoprotein receptors are hepatic parenchymal membrane proteins involved in the clearance of disialylated glycoproteins from circulation (Ashwell & Morell, 1974). These "lectins" appeared to mediate the specific binding phase of the clearance phenomenon (Pricer & Ashwell, 1977). They recognize the terminal galactose residues exposed upon removal of sialic acid from glycoproteins. Although this carbohydrate recognition property is quite widely distributed in mammalian species (Hudgin et al., 1974; Tanabe et al., 1979; Baenzigen & Maynard, 1980), birds differ in some respects (Kawasaki & Ashwell, 1977). The plasma of the birds studied contain significant levels of glycoproteins whose carbohydrate moieties lack sialic acid. These glycoproteins (with exposed terminal galactose) are not cleared from avian circulation unless *N*-acetylglucosamine residues have been uncovered by removal of galactose. Correspondingly, a chicken hepatic lectin that recognizes terminal *N*-acetylglucosamine (instead of galactose) has been sequenced by Drickamer (1981).

This class of membrane receptors is responsible for the selective uptake of macromolecules by cells, and an example of the general process of receptor-mediated endocytosis (Goldstein et al., 1979).

Drickamer and Mamon (1982) have shown that chicken hepatic lectin (CHL) is phosphorylated at Ser-7 and have suggested that CHL is a transmembrane protein with the N-terminal exposed to the cytoplasm on the basis of this phosphorylation and the existence of a hydrophobic stretch of 25 uncharged, hydrophobic residues in positions 24–48. CHL has 207 amino acids and three potential carbohydrate attachment sites.

The primary structure of the major rat asialoglycoprotein receptor (RHL-1) has also been determined by Drickamer et al. (1984). RHL-1 has 283 residues and an apparent molecular mass of 41 500. Two less abundant species are found in rat having higher molecular masses [49 000 (RHL-2) and 54 000 (RHL-3)]. RHL-2 and RHL-3 have been partly sequenced (Drickamer et al., 1984), and 53% identity in the region available for comparison has been found. CHL and RHL-1 show 28% identity in the common region.

A cDNA encoding the human asialoglycoprotein receptors has been sequenced by Spiess et al. (1985). The corresponding polypeptide of 291 residues would have an M_r of 33 122 and 79% of the residues identical with corresponding residues in RHL-1. Two of the three potential N-glycosylation sites were actually found to carry carbohydrate, namely, Asn-79 and Asn-147. Spiess and Lodish (1985) have also found and sequenced a clone encoding a second human asialoglycoprotein receptors with a protein sequence identity to H1 of 58%. The corresponding polypeptide has 311 residues.

The sequence homology found between tetranectin and the asialoglycoprotein receptors concerns that part of the latter which is considered to be extracellular. Those parts of the receptors that are considered to be membrane-anchoring and intracellular have no counterparts in the TN sequence. On the basis of the sequence homology, it is to be expected that homologous residues in these proteins will turn out to have similar functions.

The sequence of tetranectin is also homologous with those of the C-terminal globular domain of the proteoglycan core protein from cartilage that have recently been deduced from rat (Doege et al., 1986) and chicken (Sai et al., 1986) cDNA sequences cloned with the aid of synthetic probes designed on the basis of a short bovine peptide (20 residues) sequence (Pêrin et al., 1984). In general, those residues that have been conserved between the receptors and tetranectin have also been conserved in this core protein domain.

The rat and chicken proteoglycan proteins have 67% mutual sequence identity whereas the identities between tetranectin and these two proteins amount to 23 and 22%, respectively. It can therefore be excluded that tetranectin is the C-terminal globular domain of the corresponding human cartilage proteoglycan core protein. However, a fair number of different proteoglycans have been described (Poole, 1986; Gallagher et al., 1986; Hassell et al., 1986), and until their structures are at least partly elucidated, one cannot exclude the possibility that tetranectin is a C-terimal domain of one of these other proteoglycan core proteins, particularly since for some of them it is known that only about 30% of the core protein units of a proteoglycan molecule have the C-terminal globular domain still attached while 70% lack this domain (Jürgen Engel, personal communication). If tetranectin turns out to be such a C-terminal domain of a proteoglycan core protein, its "physiological" function might be to anchor plasminogen to that structure in connection with its activation to plasmin so that unlimited spreading of fibrinolysis is avoided.

Acknowledgments

We thank Torben E. Petersen for handling the computer search for homology, H. Frede Hansen for performing an initial gas-phase sequence determination, Peter Højrup and Margit S. Jensen for valuable discussions, and Birthe Hother Vesterager and Bodil Hansen for typing the manuscript.

Registry No. Plasminogen, 9001-91-6.

REFERENCES

Ashwell, G., & Morell, A. G. (1974) *Adv. Enzymol. Relat. Areas Mol. Biol. 41*, 99–128.

Baenzigen, J. U., & Maynard, Y. (1980) *J. Biol. Chem. 255*, 4607–4613.

Bajaj, S. P., Botowski, R. J., & Mann, K. G. (1975) *J. Biol. Chem. 250*, 2150–2156.

Clemmensen, I., Petersen, L. C., & Kluft, C. (1986) *Eur. J. Biochem. 156*, 327–333.

Cool, D. E., Edgells, C.-J. S., Louie, G. V., Zoller, M. J., Brayer, G. D., & MacGillivray, R. T. A. (1985) *J. Biol. Chem. 260*, 13666–13676.

Dixon, H. B. F., & Perham, R. N. (1968) *Biochem. J. 109*, 312–314.

Doege, K., Fernandez, P., Hassell, J. R., Sasaki, M., & Yamada, Y. (1986) *J. Biol. Chem. 261*, 8108–8111.

Drickamer, K. (1981) *J. Biol. Chem. 256*, 5827–5839.

Drickamer, K., & Mamon, J. F. (1982) *J. Biol. Chem. 257*, 15156–15161.

Drickamer, K., Mamon, J. F., Bins, G., & Leung, J. O. (1984) *J. Biol. Chem. 259*, 770–778.

Esmon, C. T., & Jackson, C. M. (1974) *J. Biol. Chem. 249*, 7791–7797.

Esmon, C. T., Owen, W. G., & Jackson, C. M. (1974) *J. Biol. Chem. 249*, 8045–8047.

Gallagher, J. T., Lyon, M., & Steward, W. P. (1986) *Biochem. J. 236*, 313–325.

Goldstein, J. L., Anderson, R. G. W., & Brown, M. S. (1979) *Nature (London) 279*, 679–685.

Günzler, W. A., Steffens, G. J., Ötting, F., Kim, S.-M. A., Frankus, E., & Flohé, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem. 363*, 1155–1165.

Harboe, N., & Ingild, A. (1973) *Scand. J. Immunol., Suppl. 2*, 161–164.

Hassell, J. R., Kimura, J. H., & Hascall, V. C. (1986) *Annu. Rev. Biochem. 55*, 539–567.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E., & Dreyer, W. J. (1981) *J. Biol. Chem. 256*, 7990–7997.

Hirs, C. H. W. (1967a) *Methods Enzymol. 11*, 59–62.

Hirs, C. H. W. (1967b) *Methods Enzymol. 11*, 197–199.

Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3506–3509.

Hudgin, R. L., Pricer, W. E., Jr., Ashwell, G., Stockert, R. J., & Morell, A. G. (1974) *J. Biol. Chem. 249*, 5536–5543.

Hunkapiller, M. W., & Hood, L. E. (1983) *Methods Enzymol. 91*, 386–493.

IUPAC–IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem. 5*, 151–153.

Kawasaki, T., & Ashwell, G. (1977) *J. Biol. Chem. 252*, 6536–6543.

Maeda, H., Glase, C. B., & Meinhofer, J. (1970) *Biochem. Biophys. Res. Commun. 39*, 1211–1218.

Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., & Claeys, H. (1975) in *Proteases in Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E., Eds.) pp 123–149, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., & Collen, D. (1983) *Nature (London) 301*, 214–221.

Pêrin, J. P., Bonnet, F., Jollès, J., & Jollês, P. (1984) *FEBS Lett. 176*, 37–42.

Pricer, W. E., Jr., & Ashwell, G. (1976) *J. Biol. Chem. 251*, 7539–7544.

Sai, S., Tanaka, T., Kosher, R. A., & Tanzer, M. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5081–5085.

Sottrup-Jensen, L., Claeys, H., Zajdel, M., Petersen, T. E., & Magnusson, S. (1978) *Prog. Chem. Fibrinolysis Thrombolysis 3*, 191–209.

Sottrup-Jensen, L., Petersen, T. E., & Magnusson, S. (1980) *Anal. Biochem. 107*, 456–460.

Spiess, M., & Lodish, H. F. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6465–6469.

Spiess, M., Schwartz, A. L., & Lodish, H. F. (1985) *J. Biol. Chem. 260*, 1979–1982.

Tanabe, T., Pricer, W. E., Jr., & Ashwell, G. (1979) *J. Biol. Chem. 254*, 1038–1043.

Van Zonneveld, A. J., Veerman, H., & Pannekoek, H. (1986) *Proc. Natl. Sci. Acad. U.S.A. 83*, 4670–4674.

Wiman, B., & Collen, D. (1978) *Nature (London) 272*, 549–550.

Wiman, B., Lijnen, H. R., & Collen, D. (1979) *Biochim. Biophys. Acta 579*, 142–154.