Biol. Chem. Hoppe-Seyler
Vol. 368, pp. 1533–1544, December 1987

Structural Relationship of an Apolipoprotein (a) Phenotype (570 kDa) to Plasminogen: Homologous Kringle Domains are Linked by Carbohydrate-Rich Regions

Hartmut KRATZIN[a], Victor W. ARMSTRONG[b], Marion NIEHAUS[b], Norbert HILSCHMANN[a] and Dietrich SEIDEL[b]

[a] Abteilung Immunchemie, Max-Planck-Institut für experimentelle Medizin, Göttingen
[b] Abteilung Klinische Chemie, Universitäts-Klinikum Göttingen, Göttingen

(Received 23 October 1987)

Summary: At least six allelic forms of apolipoprotein(a), differing in molecular mass, could be detected by immunoblot analysis. One of these phenotypes with a molecular mass of 570 kDa has been investigated. After reduction and carboxymethylation it was digested with trypsin and the resulting peptides were separated by gel filtration and reverse phase HPLC. The tryptic fragments sequenced comprised a total of 356 amino acids. The N-terminus of apo(a) was highly homologous to the start of the kringle 4 domain from human plasminogen and the majority of the tryptic peptides isolated was also homologous to sequences from this kringle. At least five homologous "kringle 4" domains are present in apolipoprotein(a) whereby one domain occurs more frequently than the others.

A carbohydrate-rich peptide was also obtained in high yield. This glycopeptide connects two "kringle 4" domains and contains one N-glycoside within the kringle and six potential O-glycosides in the linking region. From the recovery it can be estimated that this peptide occurs several times within the whole apolipoprotein(a) sequence. The high carbohydrate content is in sharp contrast to that of human plasminogen. Other peptides sequenced indicate that apo(a) also contains domains homologous to the kringle 5 and protease regions of plasminogen. No unique peptides were found.

These studies suggest that apolipoprotein(a) could have arisen through duplication of specific regions from the human plasminogen gene. The size heterogeneity of apo(a) might then be explained by differences in the numbers of gene duplications.

Strukturelle Verwandtschaft eines Apolipoprotein (a)-Phänotyps (570 kDa) mit Plasminogen: Homologe „Kringle"-Domänen sind durch kohlenhydratreiche Regionen miteinander verbunden

Zusammenfassung: Durch „Immunblotting" wurden mindestens 6 allelische Formen von Apolipoprotein (a) entdeckt, die sich in der Molekularmasse unterscheiden. Einer dieser Phänotypen mit einer Molekularmasse von 570 kDa wurde gereinigt, reduziert, carboxymethyliert, mit Trypsin verdaut und die Peptide mittels Gelfiltration und Umkehrphasen-HPLC aufgetrennt. Die sequenzierten tryptischen Bruchstücke umfassen insgesamt 356 Aminosäuren. Der N-Terminus dieses Apo(a) war homolog zum Beginn der „Kringle"-4-Domäne des menschlichen Plasminogens. Auch der überwiegende Teil der isolierten tryptischen Peptide zeigte eine deutliche Homologie zu Sequenzen dieser „Domäne". Daher existieren in Apo(a)

Abbreviations:
Lp(a), lipoprotein(a); apo(a), apolipoprotein(a); EDTA, disodium salt of ethylenediamine tetraacetic acid; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; TFMSA, trifluoromethanesulfonic acid; PTH, phenylthiohydantoin; HPLC, high-performance liquid chromatography; FPLC, fast-protein liquid chromatography; LDL, low-density lipoprotein; HDL, high-density lipoprotein; IgG, immuno γ-globulin.

mindestens fünf zu „Kringle" 4 homologe Domänen, von denen eine häufiger auftritt als die übrigen.

Unter anderem wurde ein kohlenhydratreiches Peptid in großer Ausbeute isoliert. Dieses Glycopeptid verbindet einige dieser Domänen miteinander und enthält einen N-glycosidischen Zuckerrest am Ende der vorausgehenden „Kringle"-Region und sechs O-glycosidische Kohlenhydratseitenketten in der Zwischenregion, auf die dann der N-Terminus einer neuen „Kringle"-4-Region folgt. Von der Ausbeute her kann vermutet werden, daß dieses Peptid wiederholt in der Apo(a)-Sequenz vorkommt. Der hohe Kohlenhydrat-Gehalt stellt das Charakteristikum dieses Apo(a)-Phänotyps dar und steht in krassem Gegensatz zu dem geringen Zuckergehalt des menschlichen Plasminogens. Die Sequenzen der übrigen Peptide zeigen, daß Apo(a) außer den zu „Kringle" 4 homologen Regionen mindestens eine zu „Kringle" 5 homologe Domäne enthält, auf die – wie im Plasminogen – eine Protease-Region folgt.

Diese Ergebnisse lassen den Schluß zu, daß Apo(a) durch Duplikation bestimmter Regionen des menschlichen Plasminogen Gens entstanden ist. Die Heterogenität in der Molekularmasse der Apo(a)-Phänotypen könnte durch eine unterschiedliche Anzahl von Gen-Duplikationen erklärt werden.

Key words: Apo(a), lipoprotein(a), plasminogen, kringle, amino acid sequence.

Since its discovery by Berg in 1963[1], numerous clinical studies have revealed a strong correlation between elevated serum concentrations of the lipoprotein Lp(a) and cardiovascular disease[2–6]. The structure of Lp(a) is that of a lipoprotein particle similar in properties to LDL, to which the glycoprotein apo(a) is attached through disulfide bonds to apoB-100[7–12]. After reduction of Lp(a) with dithiothreitol, this lipoprotein particle can be separated from apo(a) by various methods including ultracentrifugation[9,10], affinity chromatography on heparin sepharose[9] or immunoaffinity chromatography on anti-apo(a)-sepharose[11]. It contains apoB-100 as its sole apolipoprotein and is taken up and degraded as efficiently by the LDL-receptor in cultured human fibroblasts as LDL itself[9]. In contrast, normal unreduced Lp(a) is a much weaker ligand for the LDL-receptor than its reduced form suggesting that apo(a) prevents the effective binding of the native lipoprotein to the receptor[9].

Apo(a) shows remarkable size heterogeneity and several different molecular-mass forms have been described[8–13] ranging from 280 000 to 710 000 Da. In the most detailed study yet reported on over 200 persons, Utermann et al.[13] detected six apo(a) phenotypes with apparent molecular masses between 400 000 and 700 000 Da. From these observations they suggested that the various apo(a) phenotypes are controlled by a series of autosomal alleles at a single locus. Seman and Breckenridge[12] also reported 6 different apo(a) phenotypes in 20 subjects with molecular masses of 500 000 to 710 000 Da.

To obtain more information on the chemical structure of apo(a) we have therefore performed protein sequencing studies on an apo(a) phenotype with an apparent molecular mass of 570 000 Da. While these studies were in progress, Eaton et al.[14] reported the partial sequence of an apo(a) with an apparent molecular mass of 280 000 Da. Intriguingly, all peptides sequenced by these investigators and ourselves show a high homology to specific domains of human plasminogen.

Materials and Methods

Isolation of Lp(a) and apo(a)

Plasma from a healthy male subject was stabilised with 0.02% sodium azide and 0.001 M EDTA and then its density was adjusted to 1.15 g/m*l* with solid KBr. After centrifugation at 150000 × *g* for 24 h at 10 °C, the lipoproteins of $d < 1.15$ g/m*l* were collected. This material was used as the dense solution in forming a linear gradient with a d 1.03 g/m*l* KBr solution. After centrifugation in a Beckman 75 Ti rotor at 49000 rpm for 20 h (10 °C) the gradient was fractionated and the Lp(a)-containing fractions were pooled and then dialysed against 0.01 M Tris/$Cl^{\ominus}$, pH 8.0, 0.05 M NaCl, 0.001 M EDTA, 0.02% sodium azide. Traces of HDL in this Lp(a) preparation were removed by affinity chromatography over heparin sepharose 4B equilibrated with the same buffer. HDL were not retained by this column and after all the A_{280} absorbing material had eluted, the NaCl concentration in the buffer was raised to 0.5 M in order to release bound Lp(a). This was then dialysed against 0.02 M Tris/$Cl^{\ominus}$, pH 7.4, 0.15 M NaCl, 0.001 M EDTA, 0.02% sodium azide and stored at 4 °C.

To obtain pure apo(a), Lp(a) was reduced as previously described[9]. Typically, dithiothreitol was added to a final concentration of 0.01 M to a solution of Lp(a) containing 1–2 mg/m*l* total protein. After 60 min at 37 °C this solution was then adjusted to d 1.063 g/m*l* with KBr and centrifuged at 150000 × *g* for 20 h (10 °C). Apo(a)

sedimented as a pellet under these conditions. After rinsing with distilled water the pellet was solubilised in 0.01M Tris/$Cl^{\ominus}$, pH 8.0. 0.15M NaCl, 0.001M dithiothreitol, 1% sodium dodecyl sulfate. It was then extensively dialysed against 0.001M EDTA, pH 7.0 and stored at − 20 °C.

Electrophoretic procedures

Polyacrylamide gel electrophoresis in the presence of SDS was performed in 4–30% gradient gels (Pharmacia, Sweden) as described previously[9].

For detection of apo(a) phenotypes in serum or plasma, SDS-PAGE was performed according to Neville[15] using 1 mm slab gels with a 6.1% separating gel and a 3.2% stacking gel. Samples were prepared by treating 20 µl serum with 250 µl of 0.0054M Tris/$SO_4^{2\ominus}$, pH 6.14, 0.17M SDS, 0.3M 2-mercaptoethanol and heating for 10 min at 95 °C. After addition of 1% bromophenol blue solution (5 µl) and 87% glycerol (20 µl), 20 µl of the sample solution were applied to the sample wells. Electrophoresis was carried out in a Mini PROTEAN II cell (Bio Rad) at a constant voltage of 150 V for 150 min at room temperature. The electrophoresis buffers had been precooled to 4 °C prior to starting the electrophoresis. After SDS-PAGE immunoblotting was performed according to Towbin et al.[16] using a Mini-Trans-Blot cell (Bio Rad). Apo(a) bands were detected by a double antibody technique using either rabbit anti-Lp(a) or anti-apo(a) IgG's as the first antibody and a goat anti-rabbit IgG horseradish-peroxidase conjugate (Bio-Rad) as second antibody.

Amino-acid analysis

The amino-acid composition of apo(a) was derived after gas-phase hydrolysis[17] using a Durrum analyser (Dionex, Palo Alto). Serine, threonine and tyrosine values were corrected with the factors 1.11, 1.05, and 1.05 respectively.

Amino-acid sequencing

Purified apo(a) (9 mg) in 6M guanidinium hydrochloride, 0.1M Tris/$Cl^{\ominus}$, pH 8.1 (12 ml) was treated with 15 mg dithiothreitol at 37 °C for 4 h. Alkylation was then performed with 75 mg iodoacetic acid at pH 8.6 for 30 min and the mixture was subsequently dialysed against 8M urea, 0.1M Tris/$Cl^{\ominus}$, pH 8.1, followed by 0.05M ammonium hydrogencarbonate. The reduced carboxymethylated apo(a) was digested with 90 µg TosPheCH_2Cl-treated trypsin (Merck, Darmstadt) for 2 h. After addition of another 90 µg trypsin and further incubation for 1 h the solution was concentrated under a stream of nitrogen. This tryptic hydrolysate was size-fractionated by FPLC on a Superose 12 column (Pharmacia, Freiburg) equilibrated with 0.75% ammonium hydroxide. The effluent was monitored simultaneously at 214 nm and 280 nm with a Waters M 440 dual wavelength detector (Millipore, Eschborn). Pooled fractions were further purified by reverse phase HPLC on self-packed columns (250 mm × 4.6 mm) filled with Hypersil ODS (120 Å pore size, 5 µ particle size) material (Shandon, Runcorn, England). The column was developed over 40 min using 5mM potassium phosphate, pH 6.0 as starting buffer and a gradient of 0–36% acetonitrile. The separation was performed at 60 °C. A final purification of the peptides was performed by rechromatography of the HPLC fractions in 0.1% aqueous trifluoroacetic acid with 0.1% trifluoroacetic acid/acetonitrile as the organic modifier at 40 °C.

Material from FPLC pool 2 was digested further with *Staphylococcus aureus* V8 protease. After drying in a stream of nitrogen the residue was taken up in 0.05M ammonium hydrogencarbonate, pH 7.8 and V8 protease was added to a final enzyme/substrate ratio of 1:50. The solution was left at 37 °C for 24 h and peptides were then purified by reverse phase HPLC on Hypersil ODS (120 Å pore size, 5 µ particle size) eluted with the potassium phosphate/acetonitrile buffer system followed by a second purification step on Hypersil WP-300 butyl using a gradient developed with 0.1% trifluoroacetic acid and 0.1% trifluoroacetic acid in 70% n-propanol at ambient temperature.

Glycopeptides were deglycosylated with trifluoromethanesulfonic acid (TFMSA) by the procedure of Edge et al.[18]. Automated gas-phase sequencing was performed on an Applied Biosystems Model 470A sequencer with on-line PTH-analysis using an Applied Biosystems 120A PTH analyser.

Results

Polymorphism of apo(a)

Whole plasma was reduced with 2-mercaptoethanol in SDS and then subject to SDS-PAGE in homogeneous 6.1% acrylamide gels according to Neville[15]. After electroelution onto nitrocellulose[16], apo(a) phenotypes were detected with polyclonal antibodies raised against either Lp(a) or purified apo(a). Using this technique we identified at least 6 different apo(a) phenotypes in a collective of 73 individuals. Typical examples are shown in Fig. 1. In addition to the apo(a) bands, all individuals showed an additional band with a mobility identical to that of apoB-100 in agreement with Utermann et al.[13]. This band was used to assign the different apo(a) phenotypes. Of the 73 individuals, 52% showed only one apo(a) band, while the remaining 48% possessed two. The immunoblotting technique was reproducible, individuals always showing the same phenotype(s) on different runs.

For the sequencing studies on apo(a), the glycoprotein was purified as described in Materials and Methods from a single healthy male donor with a serum Lp(a) concentration of around 100 mg/dl. This apo(a) was of the phenotype 3 (lane 7, Fig. 1). After reduction and carboxymethylation, gradient SDS-PAGE revealed a single protein-staining band of apparent molecular mass 570 000 Da; estimated using apoB-100 (514 000 Da), myosin (205 000 Da), β-galactosidase (116 000 Da) and phosphorylase *b* (96 400 Da) as standards.

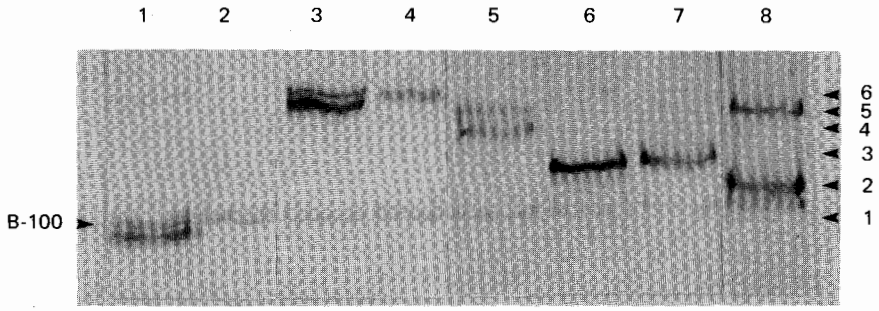

Fig. 1. Immunoblotanalysis with anti-Lp(a) of individual sera after SDS-PAGE in 6.1% polyacrylamide gels.

Six apo(a) bands can be detected and a band common to all samples with a mobility similar to that of apoB-100: lane 1) apo(a) bands 1/6; lane 2) serum with Lp(a) concentration (< 5 mg/dl) below detection limit of the immunoblot analysis; lane 3) apo(a) bands 5/6; lane 4) apo(a) band 6; lane 5) apo(a) bands 4/5; lane 6) apo(a) bands 3/6; lane 7) apo(a) band 3; lane 8) apo(a) bands 2/5.

Amino-acid analysis of apo(a)

The amino-acid composition of our apo(a) is presented in Table 1 together with previously published results on other different molecular mass forms of apo(a). These results were obtained on both purified single apo(a) phenotypes[19,20] and mixtures of different molecular-mass forms[11,12]. Despite the apparent size heterogeneity there is a remarkable similarity in the amino-acid compositions reported. Apo(a) appears to have a rather high content of Trp, Pro, Thr and a low content of Lys, Phe in comparison to an "average" protein[20].

N-Terminal sequence of apo(a)

Edman degradation of our purified apo(a) gave the following N-terminal sequence for the first 14 amino acids: EQSHVVQDCYHGDG. An identical sequence was obtained using purified apo(a) of the same phenotype from a second healthy female donor. This sequence shows a high degree of positional identity (79%) to resi-

Table 1. Amino-acid composition of our apo(a) with $M_{r(app.)}$ = 570000 compared to previously published data for various apo(a) isoforms.

ND = not determined.

Amino-acid	Bersot et al.[19]			Fless et al.[20]	Seman and Breckenridge[12]	Gaubatz et al.[11]		
	$(a)^1_v$	$(a)^1_1$	$(a)^1_2$	$(a)^2$	$(a)^3$	$(a)^4_A$	$(a)^4_B$	$(a)^5$
Lys	1.6	1.9	1.5	1.5	0.9	3.3	1.9	1.3
His	2.5	3.0	3.1	2.8	3.6	2.4	2.9	2.9
Arg	5.1	5.8	6.2	6.2	6.6	5.2	5.3	6.4
Asp	8.7	8.7	8.5	8.2	7.6	8.6	9.7	8.1
Thr	10.1	10.4	11.0	10.0	12.3	9.5	10.7	10.8
Ser	9.1	8.1	8.2	6.3	8.2	8.1	8.7	10.1
Glu	12.1	12.4	12.3	12.0	13.4	13.7	13.7	12.1
Pro	10.0	10.4	11.0	11.4	10.4	8.1	10.2	12.2
Gly	11.1	8.0	7.6	7.4	10.5	8.6	8.2	7.4
Ala	8.4	7.7	7.6	7.0	8.7	7.1	7.3	7.6
Val	5.9	6.2	6.4	6.1	7.7	6.7	6.8	5.4
Met	1.6	1.7	1.8	2.0	ND	1.9	1.9	2.1
Ile	2.3	2.5	2.4	2.0	1.8	2.9	1.9	1.4
Leu	4.9	5.5	4.9	4.2	4.4	6.7	4.5	4.2
Tyr	5.2	6.1	6.6	5.8	4.7	4.8	5.3	7.0
Phe	1.3	1.4	1.0	1.0	0.8	2.4	1.0	1.0
Trp	ND	ND	ND	3.3	ND	ND	ND	ND
Cys	ND	ND	ND	3.0	ND	ND	ND	ND

[1] The proteins were purified by preparative SDS-PAGE. $Apo(a)_v$ was derived from postprandial $d < 1.006$ lipoproteins. $Apo(a)_1$ and $apo(a)_2$ were from Lp(a). Their apparent molecular masses decreased in order $(a)_v > (a)_1 > (a)_2 >$ B-100.

[2] The apparent molecular mass of this apo(a) was 280000 Da.

[3] The data presented by the authors are the means from three different apo(a) preparations.

[4] Apo(a) was purified from two separate donors A and B. $Apo(a)_A$ showed three bands on SDS PAGE and $apo(a)_B$ two bands, all with apparent molecular masses equal to or greater than B-100.

[5] Data from our purified apo(a) with an apparent molecular mass of 570000 Da.

dues 349–362 (Fig. 5) from human plasminogen[21]: ELTPVVQDCYHGDG. Eaton et al.[14] have reported the N-terminal sequence of a lower molecular mass (280 000 Da) form of apo(a). This differed from our sequence by a single amino acid, Pro being substituted for Ser at the third position in the sequence.

Partial amino-acid sequence of apo(a)

The reduced carboxymethylated apo(a) was digested with trypsin and then size-fractionated by FPLC on a Superose 12 column resulting in the separation shown in Fig. 2. Fractions were pooled as indicated giving 7 pools in all. Pool 1 eluted in the void volume and contained only a very small amount of protein. It was not investigated further.

Pool 2 from the size fractionation: On reverse phase HPLC, the material from pool 2 eluted as a single broad shouldered peak with a low k' value. This behaviour is typical for glycopeptides[22]. On gas-phase hydrolysis[17] a dark brown precipitate formed which had to be separated by centrifugation before amino-acid analysis. Under these conditions only small

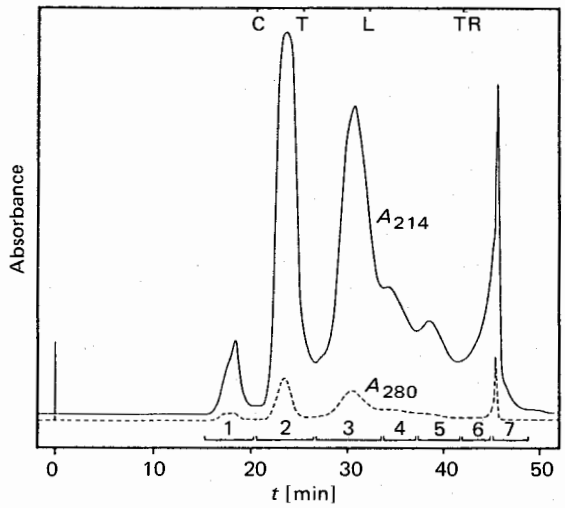

Fig. 2. Gelpermeation chromatography of the tryptic hydrolysate of reduced, carboxymethylated apo(a) using Superose 12 in 0.75% ammonium hydroxide.

Absorbance was monitored at both 214 nm (——) and 280 nm (– – –). Fractions were pooled (pools 1 to 7) as indicated below the absorbance curves. The column was calibrated with the following molecular mass markers: C = carbonate dehydratase (29000 Da); T = soybean trypsin inhibitor (21500 Da); L = lysozyme (14400 Da); TR = trasylol (6500 Da).

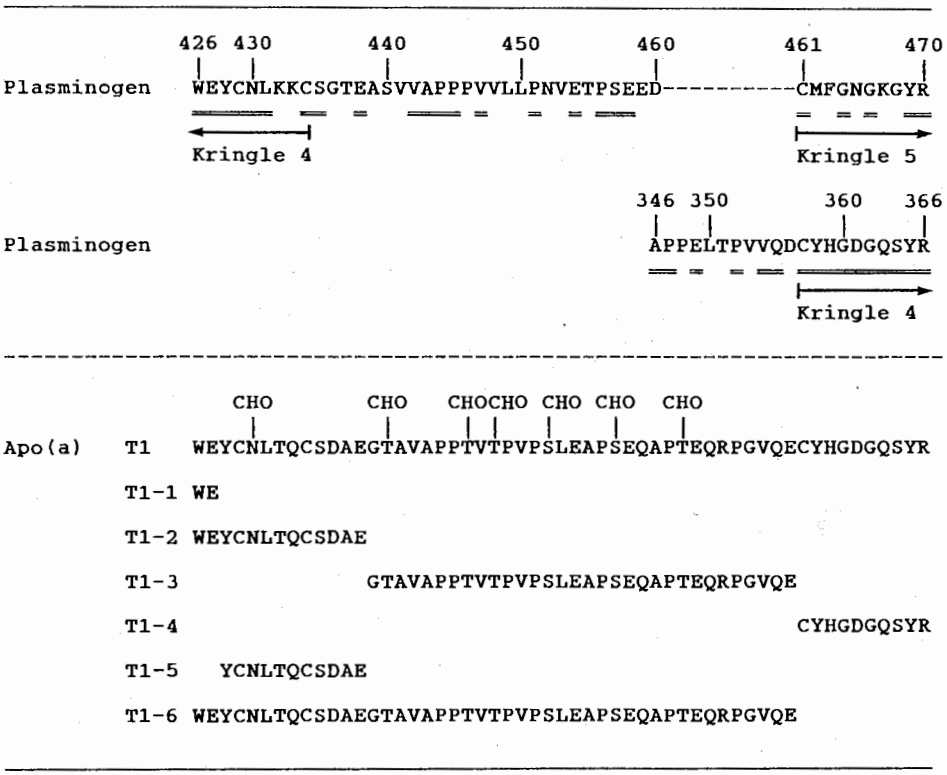

Fig. 3. Homology of the carbohydrate-rich tryptic peptide T1 from FPLC pool 2 to the plasminogen sequences 426–470 and 346–366.

Peptides T1-1 to T1-6 were obtained after *Staphylococcus aureus* V8 digestion of T1. Amino acids common to both apo(a) and plasminogen have been underlined in the plasminogen sequences. The proposed glycosylation sites are indicated by the abbreviation CHO.

amounts of ninhydrin-positive sugars were detected. These increased dramatically, however, when hydrolysis was carried out with 4M HCl for 6 h, and both glucosamine and galactosamine could be identified.

We were able to unambiguously identify 43 residues from the intact tryptic glycopeptide by gas-phase sequencing. The complete sequence was derived after a further V8 digestion followed by reverse phase HPLC purification and sequence analysis of the smaller V8 peptides. The results are summarised in Fig. 3 and Table 2. The complete sequence consists of 55 amino acids. Seven gaps were observed in the automatic sequence analysis. Although high yields were obtained in steps 4 and 6 during the Edman degradation, no PTH derivative was detected in position 5 of this glycopeptide. This result together with the fact that a threonine residue was observed in position 7 thus fulfilling the Asn/Xxx/Thr or Ser rule[23,24] indicates an asparagine residue at position 6 to which a carbohydrate side chain is N-glycosidically linked. When peptide T1-2 was pretreated with TFMSA[18] and then analysed, a broad peak appeared during PTH identification of step 5 which gradually disappeared in steps 6 and 7. A sharp peak corresponding to a PTH derivative of an Asn containing a single covalently bound monosaccharide as seen after TFMSA treatment of carcinoembryonic antigen[25] was not detected. The remaining six amino acids could be identified as four Thr and two Ser by comparing the actual amino-acid compositions of the peptides in Fig. 3 to their compositions as determined from the sequence analysis (Table 2). The assignments of these residues were tentatively made after TFMSA hydrolysis of the glycopeptide T1-3.

The first six amino acids of the glycopeptide T1 are identical to the sequence 426–431 from the end of the kringle 4 region in plasminogen. A comparison of the plasminogen sequence 426–460 to amino acids 1–35 in T1 reveals a high degree of similarity. The former sequence connects kringle 4 to kringle 5 which begins with the Cys at position 461. There is no Cys, however, at a corresponding position in peptide T1, whereas a Cys is to be found after a further eleven amino acids and the subsequent sequence of CYHGDGQSYR is identical to the N-terminal sequence of kringle 4 from plasminogen (Fig. 3). It is also noteworthy that there is a high degree of positional identity between the eleven amino acids preceding this sequence in T1 and plasminogen residues 346–356.

On the basis of the amount of T1 recovered from the reverse phase HPLC it could be estimated that 160 nmol (assuming 100% recovery) must have been present in pool 2 from the tryptic digest of 9 mg purified apo(a). Assuming a molecular mass for our apo(a) of 570 000 Da, then a maximum of 15.8 nmol of T1 would have been expected if it occurred only once in the complete sequence of apo(a). This peptide must therefore be replicated several times in the apo(a) sequence.

It was also possible to detect T1 immunochemically using anti-apo(a). After SDS-PAGE in 4–30% gradient gels and immunoblotting, a single band was observed (Fig. 4) with an esti-

Table 2. Comparison of the numbers of individual amino acids in T1 and its V8 peptides calculated from the amino-acid composition (A) with those obtained from the sequence analysis (S).

	T1-1		T1-2		T1-3		T1-4		T1-5		T1-6		T1	
	A	S	A	S	A	S	A	S	A	S	A	S	A	S
C			1.46	2			0.82	1	1.32	2	1.47	2	1.83	3
D(N)			2.26	2			1.15	1	1.94	2	2.38	2	3.14	3
T[a]			0.89	1	3.36	–			0.73	1	4.66	1	4.91	1
S[a]			0.88	1	1.87	–	0.81	1	0.79	1	3.17	1	3.61	2
E(Q)	1.00	1	3.02	3	7.38	7	1.04	1	1.75	2	10.41	10	10.57	11
P					6.37	7					6.91	7	7.14	7
G					2.24	2	2.11	2			2.33	2	4.08	4
A			0.88	1	3.78	4			0.88	1	4.90	5	5.00	5
V					3.62	4					3.73	4	3.55	4
L			1.12	1	1.12	1			1.00	1	2.26	2	2.21	2
Y			0.88	1			1.76	2	0.71	1	0.98	1	3.01	3
H							1.04	1					1.11	1
R					0.88	1	1.00	1			0.97	1	1.89	2
W[b]		1		1								1		1

[a] From the differences between the sequence analysis (S) and the amino-acid composition (A) it can be deduced that there are an additional 4 Thr and 2 Ser in the T1 sequence.

[b] Tryptophan is degraded by the acid hydrolysis necessary for the determination of the amino-acid compositions of the peptides.

mated molecular mass of 28 000 Da. A similar estimate of 26 000 Da was also derived from calibration of the Superose 12 column (Fig. 2). Since the peptide sequence only accounts for around 6 000 Da, the remaining 20 000–22 000 Da must be due to the carbohydrate content, and/or other posttranslational modifications. The one N-linked and six O-linked sugars would normally be expected to contribute around 6 000 Da to the total molecular mass. The discrepancy between the measured values of 26 000–28 000 Da (by gel filtration and SDS PAGE) and the calculated value of around 12 000 Da cannot be explained at present, although it should be noted that a high carbohydrate content can lead to an erroneous estimate of protein molecular masses by SDS PAGE[26].

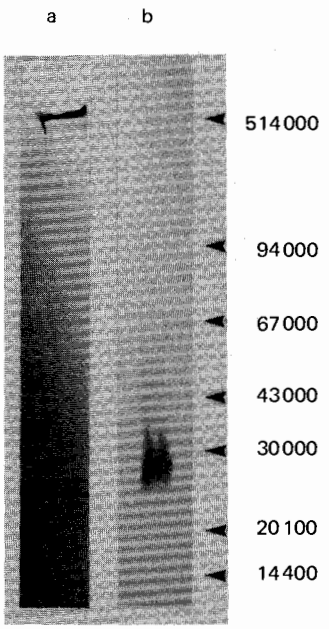

Fig. 4. Immunoblot analysis with anti-apo(a) of (a) intact apo(a) and (b) the material from FPLC pool 2. SDS-PAGE was performed on a 4–30% polyacrylamide gradient gel and apoB-100 as well as Pharmacia (Sweden) LMW standards were used as molecular mass markers.

Since apo(a) has such a low lysine content, we investigated the possibility that the peptide T1 may have been an oligomer with desmosine cross-links as is seen in elastin[27]. However, no desmosine could be detected by either amino-acid analysis[28] or UV spectroscopy.

Pools 3–7 from the size fractionation: Material from pools 3–7 was further purified by reverse phase HPLC. Many of the manually collected fractions represented analytically pure peptides as indicated by amino-acid analysis and/or Edman degradation. The sequences of sixteen different peptides (T2–T17) are shown in Fig. 5, all of which showed great similarity (43–100% positional identity) to various regions between residues 349 and 425 of human plasminogen. A further nine tryptic peptides analysed (results not shown) proved to be shorter chymotryptic fragments identical in sequence to one of the above peptides. Furthermore, the first 14 amino acids of peptide T2 are identical to the N-terminal sequence of the apo(a), confirming the sequence obtained by Edman degradation of intact apo(a). The residues 349–425 from human plasminogen include the greater part of the kringle 4 domain (residues 357–429). Noteworthy is the fact that five different tryptic peptide sequences (T6–T10) corresponded to plasminogen residues 392–407 and a further five (T12–T16) were similar to the sequence 408–425. In Table 3 we have compared the similarity of these peptides to the corresponding sequences from all five kringle domains of human plasminogen. In the case of peptides T6, T7, T12, T13 and T14 we have only considered the partial sequences for the sake of comparison to the shorter peptides. The pep-

Table 3. Comparison of apo(a) tryptic peptide sequences to the corresponding plasminogen kringle sequences.
The numbers of amino-acid differences between the apo(a) sequences and the appropriate kringles are indicated.

Peptide	Apo(a) amino-acid sequence	Kringle number				
		1	2	3	4	5
T6	TPENYPNDGLTM	8	8	6	1	6
T7	TPEYYPNAGLIM	8	8	7	2	5
T8	TPENYPDAGLTR	8	9	6	2	5
T9	TTEYYPNGGLTR	9	9	8	4	7
T10	TPAYYPNAGLIK	7	7	8	4	5
T12	NPDADTGPWCFTT	4	4	5	1	4
T13	NPDAEIRPWCYTM	6	4	6	5	6
T14	NPDPVAAPWCYTT	5	5	4	5	5
T15	NPDPVAAPYCYTR	7	7	6	7	7
T16	NPDAVAAPYCYTR	7	7	6	6	7

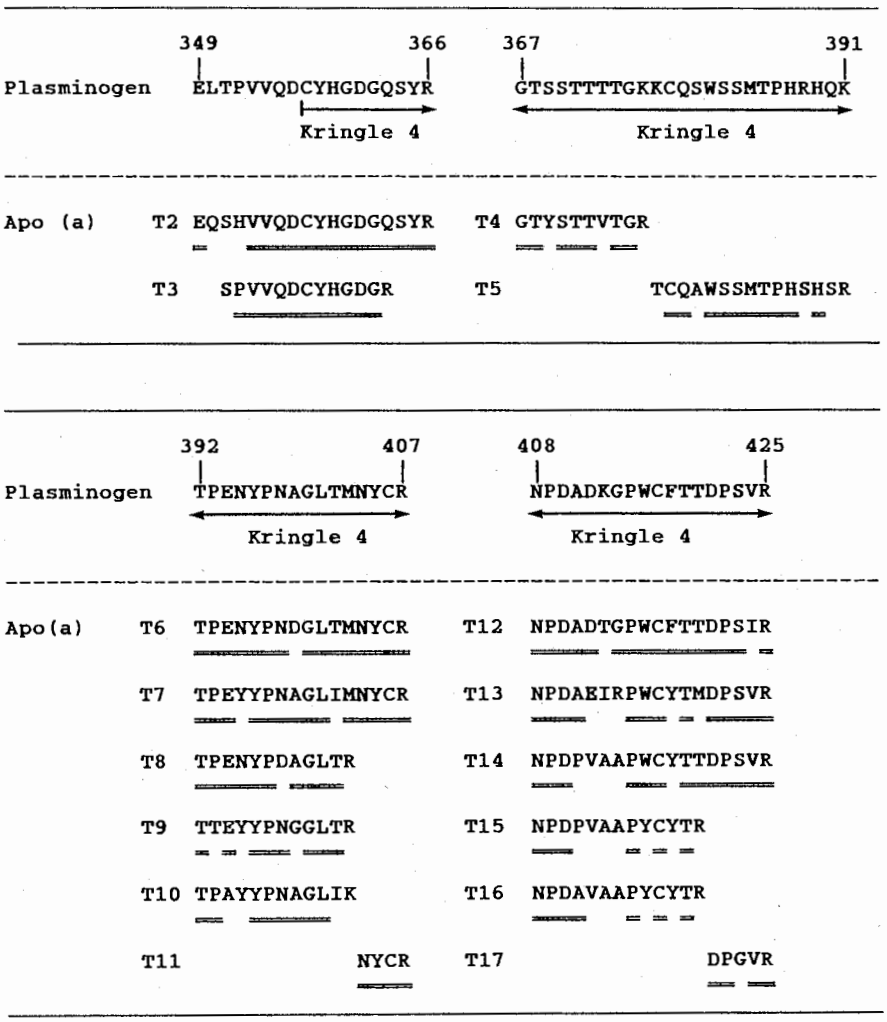

Fig. 5. Homology of apo(a) tryptic peptides to plasminogen sequences between residues 349–425.

Amino acids common to both apo(a) and plasminogen have been underlined in the apo(a) peptide sequences.

tides T6 to T10 clearly showed greatest similarity to kringle 4. Of these 5 peptides, T7 was recovered in much higher yield than the other peptides suggesting that this sequence occurs more often in apo(a).

In the case of those peptides corresponding to residues 408–420 from plasminogen, only T12 showed a clear positional identity to kringle 4, with only one amino-acid substitution in this sequence. The remaining four peptides did not reveal a distinct similarity to any one particular kringle domain. The highest levels of similarity were as follows: T13 to kringle 2; T14 and T15 to kringle 3; T16 to both kringles 3 and 4. In the case of peptides T13 and T14, the remaining sequence of DPSVR in both peptides, not considered in Table 3, was identical to residues 421–425 of kringle 4 and showed either 2 or 3 substitutions to the appropriate sequences DPNKR and NSQVR from kringles 2 and 3, respectively. As with the previous five peptides (T6–T10), one peptide T16 from this series was present in much higher yield than the other four peptides, indicating that it occurs more frequently in apo(a).

The sequence 408–425 from plasminogen contains an Asp at position 412 which together with Arg 425 is believed to be essential for the binding of ω-aminocarboxylic acids[29]. This residue is, however, only retained in one (T12) of the five apo(a) tryptic peptides which correspond to this region of plasminogen.

In contrast to tryptic peptides T2–T17, the peptides T18 and T19, isolated from FPLC pool 3 by reverse phase HPLC revealed a striking homology not to kringle 4 but to kringle 5 (Fig. 6). The peptide T18 had only one amino-

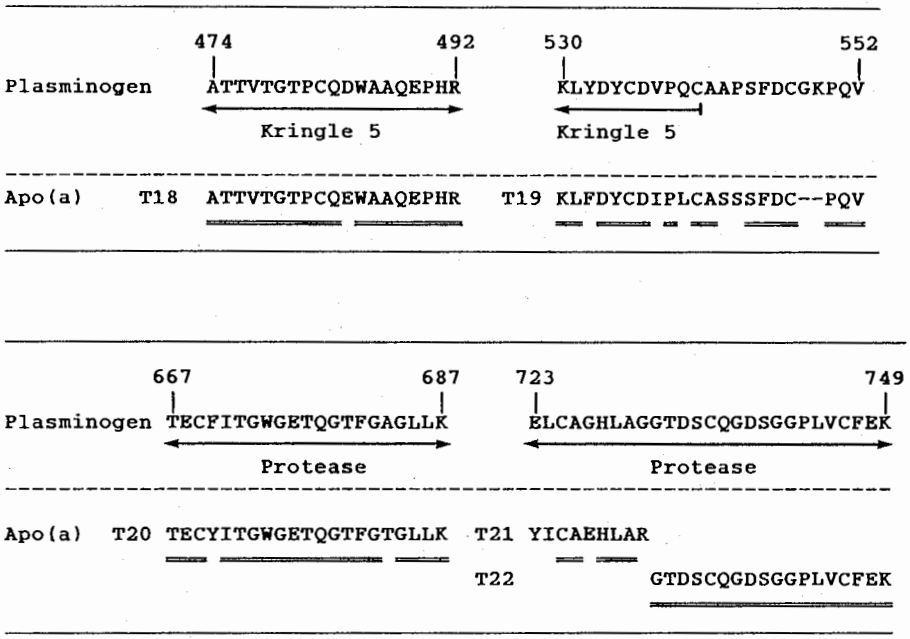

Fig. 6. Homology of apo(a) tryptic peptides to partial sequences from the kringle 5 and protease domains of plasminogen.

Amino acids common to both apo(a) and plasminogen have been underlined in the apo(a) peptide sequences.

acid substitution in comparison to the appropriate sequence (474–492) from kringle 5 whereas there were a total of 11, 11, 9 and 9 substitutions compared to kringles 1, 2, 3 and 4, respectively. Peptide T19 is similar to the final eleven amino-acid residues of kringle 5 and the start of the protease region in human plasminogen.

A further three tryptic peptides T20, T21 and T22 were isolated that showed greatest similarity to specific regions of the protease domain in plasminogen (Fig. 6). Peptide T22 was identical to the sequence 732–749 from plasminogen and therefore has retained a serine analogous to Ser^{740}, the active-site serine essential for the proteolytic activity of plasminogen. We do not know, however, whether or not the other essential amino acids, His^{602} and Asp^{645}, at the active site of plasminogen are also retained in this apo(a). The occurrence of these five peptides T18–T22 in the tryptic digest of apo(a) would suggest that there is a sequence in apo(a) analogous to the kringle 5 and protease region ranging from residues 461–790 in human plasminogen.

Discussion

The protein apo(a) from the plasma lipoprotein Lp(a) is a large molecular mass glycoprotein containing between 24% and 50%[12,20] carbohydrate by weight and displaying remarkable inter- and intraindividual size heterogeneity[8–13]. In the largest population study so far, Utermann et al. using SDS-PAGE and immunoblotting[13] identified six different molecular-mass phenotypes of apo(a) with apparent molecular masses ranging from 400 000–700 000 Da. We are able to confirm these findings in a collective of over 70 randomly selected persons with plasma Lp(a) concentrations greater than 10 mg/d*l*. Seman and Breckenridge[12] also reported six apo(a) species with molecular masses between 500 000–710 000 Da in a smaller collective of around 20 persons. In these three studies, no individual exhibited more than two apo(a) bands suggesting that the different phenotypes are under genetic control. On the basis of their family studies Utermann et al.[13] have therefore postulated that the apo(a) phenotypes are controlled by a series of autosomal alleles at a single locus.

In order to obtain further information on the structural basis of the apo(a) heterogeneity, we have performed protein sequencing studies on an apo(a) species purified from a single individual. This phenotype had an apparent molecular mass of 570 000 Da. During our investigation Eaton et al.[14] published the sequences of 8 peptides derived from an apo(a) with an apparent molecular mass of 280 000 Da. This form has not been observed in other studies and may therefore represent an unusual variant

of apo(a). Nevertheless, despite the large difference in their molecular masses, all of the peptides so far sequenced from both our apo(a) phenotype and that of Eaton et al.[14] have demonstrated high homology to specific sequences of human plasminogen[21]. The latter is a single-chain glycoprotein with a molecular mass of about 92 000 Da containing only 2% carbohydrate. It is characterized by five homologous kringle domains linked by short connecting sequences and a C-terminal trypsin-like protease domain. Each kringle consists of approximately 80 amino acids including six Cys residues. Disulfide bridges occur between cysteines 1 and 6, 2 and 4, and 3 and 5, thus building a characteristic triple loop structure. Such kringle domains also occur in prothrombin[30], tissue plasminogen activator[31], urokinase[32] and factor XII[33].

Although the present studies do not provide us with the complete primary sequence of apo(a), the high homology of the tryptic peptides to specific regions of the plasminogen molecule, notably the kringle 4 domain, implies a close evolutionary relationship between these two proteins. The fact that five different primary sequences were obtained from apo(a) corresponding to the sequences 392–407 and 408–425 from human plasminogen suggests that there are at least five kringle domains in apo(a) with different primary structures. Two of the peptides, T7 and T16 were obtained in far greater yield than the others. We can therefore deduce that these two peptides are to be found in kringle units which are repeated several times. Interestingly the peptide tryptic 2 derived from the low-molecular mass apo(a)[14] represents an overlap between T7 and T16. Plasminogen contains several structural domains known as lysine binding sites characterised by their ability to bind ω-aminocarboxylic acids[29]. These sites are thought to be responsible for the interaction of plasminogen (and plasmin) with both fibrin and α_2-antiplasmin and therefore probably play a crucial role in the regulation of fibrinolysis. One of these binding sites is located in kringle 4 and the amino acids Asp^{412} and Arg^{425} have been shown to be essential for binding activity[29]. Of the five peptides (T12–T16) sequenced from apo(a) that were homologous to the plasminogen sequence containing Asp^{412}, only one (T12) had retained this particular amino acid. The lower molecular mass apo(a)[14] also contains the corresponding sequences with this Asp residue and can bind to lysine Sepharose. However, the possibility that apo(a) phenotypes may occur in which this residue has been completely lost and which cannot bind ω-aminocarboxylic acids deserves further investigation especially in view of the fact that apo(a) could direct the cholesterol-rich lipoprotein Lp(a) to fibrin deposits at sites of vascular injury.

In contrast to plasminogen, apo(a) is rich in carbohydrate. In the case of our apo(a) phenotype, carbohydrate was associated with a proline-rich 55-amino acid peptide, the first nine amino acids WEYCNLTQC of which showed high homology to the last nine amino acids of kringle 4 from plasminogen, and the final 10 C-terminal amino acids of which were identical to the start of a new kringle 4 domain. This peptide probably contains one N-linked oligosaccharide within the kringle sequence and six O-linked oligosaccharides in the connecting peptide. The latter is similar to the proline-rich hinge region of IgA1[34] which contains several O-glycosides of the type Gal-GalNAc-[35] and links two immunoglobulin domains (C_H1 and C_H2). In contrast to apo(a), plasminogen contains only one N-linked oligosaccharide (Asp^{288}) in kringle 3 and one O-linked oligosaccharide (Thr^{345}) in the connecting region between kringles 3 and 4[36]. However, there is a remarkable similarity between the sequence ELAPT preceding Thr^{345} of plasminogen and the sequence EQAPT preceding the sixth glycosylated threonine of the carbohydrate-rich peptide T1 from apo(a). On accout of the high recovery of this carbohydrate-rich peptide we can deduce that this sequence must be repeated several times in the apo(a) molecule. The kringle domains are linked by a 36 amino acid sequence, in contrast to the 26 amino acid sequence-linking kringles 4 and 5 in plasminogen. Whereas the first 25 amino acids of the connecting sequence in apo(a) show a high degree of positional identity to the plasminogen sequence 426–459, the remaining sequence of eleven amino acids is similar to plasminogen residues 346–356. In this respect it should be noted that an exon with flanking splice junctions has been detected in the human plasminogen gene that encodes amino acid residues 346–399 in human plasminogen[37].

In addition to the "kringle 4" sequences, there is evidence that at least one kringle is present in apo(a) that is highly homologous to kringle 5 of plasminogen. Similarly there also appears to be a protease domain analogous to that in plasminogen. It cannot be ascertained as to whether these sequences are repeated more than once in apo(a), but considering the yields of the peptide recoveries from these two regions they do not represent a large proportion of the apo(a) molecule.

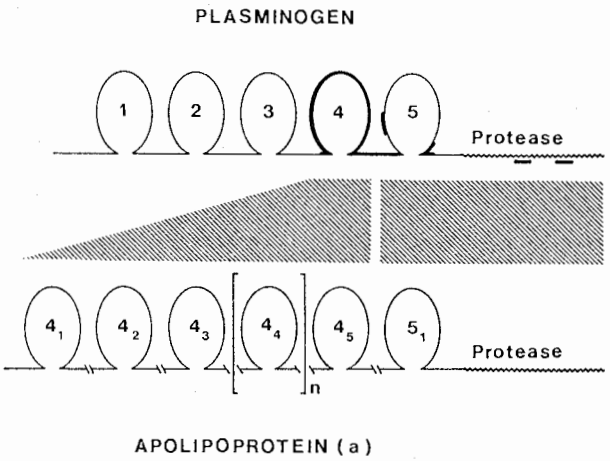

Fig. 7. Apolipoprotein (a) derives from human plasminogen kringles 4 and 5 as well as the protease region. The regions in plasminogen homologous to the tryptic peptides isolated from apo(a) have been marked by thick lines in the plasminogen model. The domains homologous to kringle 5 and the protease region probably only occur once. The remaining part of the molecule consists of domains homologous to kringle 4. These are connected by a proline-rich region which contains 6-*O*-glycosidically linked carbohydrate side chains. The yield of the connecting peptide indicates that most of the "kringle 4" domains are linked by this glycopeptide. An N-linked glycoside was also detected towards the end of a "kringle 4" domain. According to the yields of the peptides kringle 4_4 must be present at least 10 times.

On the basis of our data, we consider the most likely structure for apo(a) to be that of several duplicated "kringle 4" domains with differences in their primary structures (Fig. 7). One of these domains is probably duplicated more frequently than the others. In contrast to plasminogen, apo(a) contains a large amount of carbohydrate. This carbohydrate is clustered in short sequences of amino acids connecting the kringle domains and will therefore be distributed over a large portion of the apo(a) molecule rendering it highly hydrophilic. Finally, by analogy to plasminogen, the C-terminal end probably includes a "kringle 5" domain followed by a "protease" domain.

The kringle regions in plasminogen are thought to have arisen from a common ancestor of the first prothrombin kringle region by gene duplication[36]. It can therefore be envisaged that apo(a) has arisen from the plasminogen gene through gene duplication and that differences in the number of duplications are responsible for the structural heterogeneity of apo(a) leading to phenotypes with similar amino-acid compositions and immunological properties but differing in their molecular masses. This does not of course exclude that post-translational modification may also add to the heterogeneity. The preliminary data of Eaton et al.[14] on a different species of apo(a) suggest the same basic structure. However, the N-terminal sequence of this apo(a) differed from our N-terminal sequence by a Ser/Pro replacement at position 3. Whether this indicates differences between apo(a) phenotypes at the primary sequence level remains to be shown.

The elucidation of the structure of apo(a) and its similarity to human plasminogen opens up new aspects with regard to the potential role of Lp(a) in atherogenesis. In particular, the lysine binding sites of plasminogen are thought to play an important role in the regulation of fibrinolysis. It remains to be determined whether or not high levels of plasma Lp(a) can interfere with this regulation. Furthermore since plasminogen can bind fibrin through these sites it is possible that apo(a) could target the cholesterol-rich lipoprotein particle Lp(a) to fibrin deposits at the site of vascular injury.

We gratefully acknowledge the expert technical assistance of Mrs. *D. Hesse*, Mrs. *E. Nyakatura*, Mr. *N. Otte* and Mrs. *A. Möhle*. The amino-acid analyses were kindly performed by Mrs. *M. Praetor* and Mr. *W. Sinn*. We are indebted to *Dr. H. Weber Mandel* (University Clinic, Göttingen) who performed the plasmaphereses on the donor. This work was supported by a grant from the *Deutsche Forschungsgemeinschaft* to VWA.

Literature

1 Berg, K. (1963) *Acta Pathol. Microbiol. Scand.* **59**, 369–382.
2 Berg, K., Dahlen, G.M. & Frick, M.H. (1974) *Clin. Genet.* **6**, 230–235.
3 Kostner, G.M., Avagaro, P., Cazzalato, G., Marth, E. & Bittolo-Bon, G. (1981) *Atherosclerosis* **38**, 51–61.
4 Armstrong, V.W., Cremer, P., Eberle, E., Manke, A., Schulze, F., Wieland, H., Kreuzer, H. & Seidel, D. (1986) *Atherosclerosis* **62**, 249–257.
5 Dahlen, G.H., Guyton, J.R., Attar, M., Farmer, J.A., Kautz, J.A. & Gotto, A.M. (1986) *Circulation* **74**, 758–765.
6 Rhoads, G.G., Dahlen, G.H., Berg, K., Morton, N.E. & Dannenberg, A.L. (1986) *J. Am. Med. Assoc.* **256**, 2540–2544.
7 Utermann, G. & Weber, W. (1983) *FEBS Lett.* **154**, 357–361.
8 Gaubatz, J.W., Heidemann, C., Gotto, A.M., Morrisett, J.D. & Dahlen, G.H. (1983) *J. Biol. Chem.* **258**, 4582–4589.
9 Armstrong, V.W., Walli, A.K. & Seidel, D. (1985) *J. Lipid Res.* **26**, 1314–1323.
10 Fless, G.M., ZumMallen, M.E. & Scanu, A.M. (1985) *J. Lipid Res.* **26**, 1224–1229.
11 Gaubatz, J.W., Chari, M.V., Nava, M.L., Guyton, J.R. & Morrisett, J.D. (1987) *J. Lipid Res.* **28**, 69–79.

12 Seman, L.J. & Breckenridge, W.C. (1986) *Biochem. Cell Biol.* **64**, 999–1009.
13 Utermann, G., Menzel, H.J., Kraft, H.G., Duba, H.C., Kemmler, H.G. & Seitz, C. (1987) *J. Clin. Invest.* **80**, 458–465.
14 Eaton, D.L., Fless, G.M., Kohr, W.J., Mclean, J.W., Xu, Q.-T., Miller, C.G., Lawn, R.M. & Scanu, A.M. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3224–3228.
15 Neville, D.M. (1971) *J. Biol. Chem.* **246**, 6328–6334.
16 Towbin, H., Staehlin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354.
17 Bidlingmeyer, B.A., Cohen, S.A. & Tarvin, T.L. (1984) *J. Chromatogr.* **336**, 93–104.
18 Edge, A.S.B., Faltynek, C.R., Hof, L., Reichert, L.E. & Weber, P. (1981) *Anal. Biochem.* **118**, 131–137.
19 Bersot, T.P., Innerarity, T.L., Pitas, R.E., Rall, S.C., Weisgraber, K.H. & Mahley, R.W. (1986) *J. Clin. Invest.* **77**, 622–630.
20 Fless, G.M., ZumMallen, M.E. & Scanu, A.M. (1986) *J. Biol. Chem.* **261**, 8712–8718.
21 Sottrup, L., Petersen, T.E. & Magnussen, S. (1978) in *Atlas of Protein Sequence Structure*, Vol. 5, Suppl. 3, p. 91.
22 Takahashi, N., Takahashi, Y., Ortel, T.L., Lozier, J.N., Ishioka, N. & Putnam, F.W. (1984) *J. Chromatogr.* **317**, 11–26.
23 Marshall, R.D. (1972) *Annu. Rev. Biochem.* **41**, 673–702.
24 Kornfeld, R. & Kornfeld, S. (1985) *Annu. Rev. Biochem.* **54**, 631–664.
25 Paxton, R.J., Mooser, G., Thompson, J. & Shively, J.E. (1986) in *Methods in Protein Sequence Analysis* (Walsh, K.A., ed.) pp. 443–449, Humana Press, Clifton, N.Y.
26 Produslo, J.F. (1981) *Anal. Biochem.* **114**, 131–139.
27 Gallop, P.M., Blumenfeld, O.O. & Seifter, S. (1972) *Annu. Rev. Biochem.* **41**, 618–672.
28 Glazer, A.N., De Lange, R.J. & Sigman, D.S. (1975) in *Chemical Modification of Proteins, Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T.S. & Work, E., eds.) pp. 48–49, North Holland/Elsevier, Amsterdam.
29 Trexler, M., Vali, Z. & Patthy, L. (1982) *J. Biol. Chem.* **257**, 7401–7406.
30 Magnussen, S., Petersen, T.E., Sottrup-Jensen, L. & Claeys, H. (1975) in *Proteases and Biological Control.* (Reich, E. & Rifkin, D.B. eds.) pp. 123–149, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31 Pennica, D., Holmes, W.E., Kohr, W.J., Harkins, R.N., Vehar, G.A., Ward, C.A., Bennet, W.F., Yelverton, E., Seeburg, P.H., Heyneker, H.L., Goeddel, D.V. & Collen, P. (1983) *Nature (London)* **301**, 214–221.
32 Gunzler, W.A., Steffens, G.J., Otting, F., Kim, S.M., Frankes, E. & Flohe, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem.* **363**, 1155–1165.
33 Cool, D.E., Edgell, C.-J., Louie, G.V., Zoller, M.J., Brayer, G.D. & MacGillivray, R.T.A. (1985) *J. Biol. Chem.* **260**, 13666–13676.
34 Kratzin, H., Altevogt, P., Ruban, E., Kortt, A., Staroscik, K. & Hilschmann, N. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1337–1342.
35 Baenziger, J. & Kornfeld, S. (1974) *J. Biol. Chem.* **249**, 7270–7281.
36 Young, C.L., Barber, W.C., Tomaselli, C.M. & Dayhoff, M.O. (1978) *Atlas of Protein Sequence and Structure*, Vol. 5, pp. 73–91.
37 Malinowski, D.P., Sadler, J.E. & Davie, E.W. (1984) *Biochemistry* **23**, 4243–4250.

Prof. Dr. N. Hilschmann, Dr. Hartmut Kratzin*, Abteilung Immunchemie, Max-Planck-Institut für experimentelle Medizin,
Hermann-Rein-Straße 3, D-3400 Göttingen;

Prof. Dr. Dietrich Seidel, Dr. Victor W. Armstrong*, Frau Marion Niehaus, Abteilung Klinische Chemie, Universitäts-Klinikum Göttingen,
Robert-Koch-Straße 40, D-3400 Göttingen.

* Correspondence should be addressed either to Dr. V.W. Armstrong or Dr. H. Kratzin.