Molecular Basis for the Three Major Forms of Human Serum Vitamin D Binding Protein (Group-Specific Component)†

Jisnuson Svasti,* Alexander Kurosky, Ann Bennett,‡ and Barbara H. Bowman

ABSTRACT: Human vitamin D binding protein, previously known as the group-specific component (Gc), was shown to consist of a single polypeptide chain of molecular weight 52000 as evidenced by gel electrophoresis and gel filtration under denaturing conditions. Quantitative analyses by carboxypeptidase A hydrolysis and by automated sequential Edman degradation identified leucine as the amino and carboxyl terminus in molar yields consistent with a molecular weight of 52000. Peptide maps of tryptic hydrolysates obtained from S-[^{14}C]carboxymethylated Gc revealed the expected number of peptides predicted from amino acid composition. The molecular basis of the major Gc proteins (Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow}) was investigated. In addition to identical carboxyl-terminal sequences, all three major components exhibited an identical amino-terminal sequence of 20 residues. Tryptic peptide maps of these Gc components were similar; however, Gc 1 and Gc 2 differed by two peptides, whereas Gc 1_{fast} and Gc 1_{slow} differed by only one peptide. Sequence analysis of the characteristic peptides indicated that the difference between Gc 1_{fast} and Gc 1_{slow} is posttranslational in nature, involving carbohydrate dissimilarities. However, the difference between Gc 1 and Gc 2 was related to primary structure involving only a small number of amino acids. These results support the hypothesis that the major components of Gc occur as a result of two autosomal alleles at a single locus.

Human group-specific component (Gc)[1] is a serum α-globulin (Hirschfeld, 1959; Putnam, 1977) reported to be involved in the transport of metabolic products of vitamin D (Daiger et al., 1975). Polymorphisms of Gc were established by gel electrophoretic analysis which revealed three major phenotypes: Gc 1-1 (Gc 1), Gc 2-1, and Gc 2-2 (Gc 2) (Hirschfeld et al., 1960). These phenotypes occur in all populations studied (Cleve & Bearn, 1962). Although patterns of inheritance suggest that these phenotypes are the products of two codominant autosomal alleles, Gc^1 and Gc^2, the gene products of both homozygotes appear to exhibit heterogeneity. Gel electrophoretic patterns of Gc from sera of individuals homozygous for the Gc^1 allele revealed two characteristic bands of equal intensity, Gc 1_{fast} and Gc 1_{slow} (Reinskou, 1963; Kitchin & Bearn, 1966). Some investigators have observed that the electrophoretic pattern of Gc from sera of Gc^2 homozygotes is comprised of a major band, Gc 2_{slow}, and of a minor fast band, Gc 2_{fast} (Kitchin & Bearn, 1966); however, others have reported the occurrence of only a single band (Reinskou, 1963). Recent studies have indicated that the electrophoretic heterogeneity of Gc results in part from posttranslational events (Svasti & Bowman, 1978). Binding of vitamin D to the major Gc components (Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow}) increased their anionic mobilities during polyacrylamide gel electrophoresis at pH 8.3. In addition, neuraminidase digestion reduced the electrophoretic mobility of Gc 1_{fast} to that of Gc 1_{slow}, suggesting that Gc 1 heterogeneity may be due to sialic acid differences.

Our earlier studies (Bowman & Bearn, 1965; Bowman, 1969) suggested that the observed heterogeneity of Gc in the electrophoretic patterns of sera from homozygotes resulted from the presence of two nonidentical subunits, whose synthesis was controlled by three gene loci. In this study, we present results from polyacrylamide gel electrophoretic analysis, gel filtration, peptide mapping studies, and quantitative amino- and carboxyl-terminal determinations which strongly support the conclusion that Gc occurs as a single polypeptide chain. We also provide evidence for the posttranslational nature of Gc 1 heterogeneity and for the molecular basis of the Gc^1/Gc^2 polymorphism by analysis of peptides that distinguish Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow}.

Experimental Procedures

Materials

[1α,2α-^{3}H]Cholecalciferol (sp act. 12.3 Ci/mmol), iodo-[2-^{14}C]acetic acid (sp act. 57 mCi/mmol), iodo[1-^{14}C]acetamide (sp act. 57 mCi/mmol), and [4,5-^{3}H]leucine (sp act. 59 Ci/mmol) were from Amersham/Searle Corp. Pth amino acids, Dns amino acids, Cheng Chin polyamide sheets, dithiothreitol, iodoacetic acid, iodoacetamide, and Sequanal grade reagents used in the sequencer were purchased from Pierce Chemical Co. Acrylamide, *N,N'*-methylenebis(acrylamide), and ammonium persulfate were obtained as Electrophoresis Purity grade from Bio-Rad Laboratories. Blue Dextran 2000 and Sephadex G-150 were from Pharmacia Fine Chemicals. Guanidine hydrochloride and urea were purchased as Ultrapure grade from Schwarz/Mann Biochemicals. Trypsin, treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone, and carboxypeptidase A, treated with diisopropylfluoro phosphate, were obtained from Worthington Biochemical Corp. Aquasol used as scintillation fluid was a product of New England Nuclear, while X-ray film was purchased from Eastman Kodak Co.

Methods

Purification of Gc. Human Gc was purified from [^{3}H]-cholecalciferol-labeled plasma fractions of individual homozygotes (kindly provided by Dr. D. R. Barnett) according to the procedure of Svasti & Bowman (1978). Gc 2 preparations obtained by column chromatography on DEAE-

† From the Division of Human Genetics, Department of Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston, Texas 77550. *Received October 23, 1978; revised manuscript received January 11, 1979.* This investigation was supported by National Institutes of Health Grants HD 03321 and CA 17701.

* Present address: Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand.

‡ Present address: Department of Pediatrics, Stanford, Palo Alto, CA.

[1] Abbreviations used: Gc, group-specific component; BAWP, 1-butanol–acetic acid–water–pyridine; HI, hydriodic acid; Pth, phenylthiohydantoin; Dns, dansyl (5-dimethylaminonaphthalene-1-sulfonyl).

0006-2960/79/0418-1611$01.00/0

cellulose contained traces of Gc 2_{fast} in addition to the major component, Gc 2_{slow}. Column isoelectric focusing was required for the preparation of Gc 1_{fast} and Gc 1_{slow}. The purity of these preparations was established by electrophoresis in 7.5% polyacrylamide gel slabs according to the method of Kitchin (1965) by use of the Tris–borate buffer system of Peacock et al. (1965).

Polyacrylamide Gel Electrophoresis of Gc. Gc samples (2 mg) were reduced in 0.1 M Tris-HCl, pH 8.1, containing 7.5 M guanidine hydrochloride and made 25 mM in dithiothreitol for 1 h at 37 °C under N_2 in the dark. The mixture was made 75 mM in iodoacetamide, and alkylation was allowed to proceed for 1 h at 37 °C under N_2 in the dark. The reaction was terminated by addition of 2-mercaptoethanol to a final concentration of 0.3 M. Some carbamoylmethylated Gc preparations were analyzed by electrophoresis in 10% polyacrylamide gels containing 4.5 M urea and 0.9 M acetic acid, pH 2.7 (Panyim & Chalkley, 1969), after dialysis against 9 M urea containing 0.3 M 2-mercaptoethanol for 16 h. Other samples were analyzed by electrophoresis in 10% polyacrylamide gels containing 0.1% sodium dodecyl sulfate in 0.1 M sodium phosphate buffer, pH 7.0 (Shapiro et al., 1967; Weber et al., 1972), after a further dialysis against 1% sodium dodecyl sulfate and 0.3 M 2-mercaptoethanol in 0.01 M sodium phosphate, pH 7.0. In the latter gels, molecular weights were estimated from a standard graph prepared from the following proteins of known molecular weight: bovine serum albumin, IgG heavy and light chains, ovalbumin, pepsin, trypsin, and cytochrome *c*. Controls of unreduced protein were obtained in the same manner except that dithiothreitol and iodoacetamide were absent from the guanidine–Tris-HCl buffer and 2-mercaptoethanol was absent from the dialysis solutions.

Analytical Gel Filtration. The molecular weight of carbamoylmethylated Gc was determined by gel filtration on a Sephadex G-150 column equilibrated in 0.05 M sodium acetate, pH 4.5, containing 6 M guanidine hydrochloride. Samples of 2 mg of Gc were analyzed in pairs, with one sample being maintained in the column buffer at 37 °C for 2 h, while the other sample was reduced and alkylated as described above. The column was calibrated with proteins of known molecular weights that were reduced and alkylated in the same manner. The void volume and the total bed volume in each experiment were determined by the use of standard markers—blue dextran and [4,5-^{3}H]leucine, respectively. Semilogarithmic plots were made of molecular weight against K_d, where the partition coefficient (K_d) (Mann & Fish, 1972) was defined as the ratio of elution volume (V_e) minus void volume (V_0) to total bed volume (V_t) minus void volume (V_0).

Peptide Mapping. Purified Gc 1_{fast}, Gc 1_{slow}, or Gc 2_{slow} was dissolved (5 mg/mL) in 0.1 M Tris-HCl containing 7.5 M guanidine hydrochloride, pH 8.1. Each sample was reduced with 5 mM dithiothreitol as described above and alkylated in 15 mM iodo[2-^{14}C]acetic acid (sp act. 0.5 mCi/mmol). After addition of a drop of 2-mercaptoethanol, each sample was dialyzed against 0.5% NH_4HCO_3 and hydrolyzed with trypsin (enzyme to substrate ratio of 1:50) for 4 h at 37 °C. The strategy for peptide mapping was essentially the same as that used to study mouse immunoglobulin κ chains (Svasti & Milstein, 1972) and involved high-voltage paper electrophoresis at pH 6.5 as a first step. Electrophoretic mobilities ($m_{6.5}$) are expressed as fractions of the distance between aspartic acid and 6-dinitrophenyllysine (Offord, 1966). The pH 6.5 electropherogram was cut into three sections corresponding to the acidic peptides, the neutral and slow-moving peptides, and the basic peptides. Good fractionation of the basic peptides was achieved by descending chromatography in 1-butanol–acetic acid–water–pyridine (15:3:12:10) (Waley & Watson, 1953) at right angles to the original direction. Resolution of the "neutral" peptides required electrophoresis at pH 2.0 prior to the same chromatography step. The acidic peptides did not chromatograph well on paper since they were probably quite large; however, electrophoresis at pH 2.0 at right angles to the original direction was found to give reasonable resolution. Peptides were visualized when sprayed with fluorescamine (Mendez & Lai, 1975) and *S*-[^{14}C]carboxymethylcysteine peptides were detected by radioautography. Peptides characteristic of Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} were eluted with 0.5% acetic acid. The amino acid composition and sequences of these peptides were further studied as described below.

Amino Acid Analysis. Peptides or proteins were hydrolyzed for 20–24 h in 5.7 N HCl at 107 °C. Amino acid compositions were determined by use of either a Beckman 119 or a Beckman 121M analyzer with single-column methodology on Durrum DC-6A resin (Kurosky et al., 1977).

Carboxyl-Terminal Analysis. The carboxyl-terminal sequences of Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} were investigated by carboxypeptidase A hydrolysis (Ambler, 1972) by use of the procedure of Fraenkel-Conrat et al. (1955) for enzyme activation. Samples of Gc protein (0.3–0.6 mg) were digested with 0.25–50 μg of enzyme in 0.25 mL of 0.2 M *N*-ethylmorpholine acetate, pH 8.2, for 10–240 min at 37 °C. The reaction was stopped by lowering the pH to 3 with formic acid, and the mixture was lyophilized. Norleucine (10 nmol) was added as an internal standard, and protein was precipitated by addition of 1% sulfosalicyclic acid in pH 2.2 buffer. Amino acids released were quantitated on a Beckman 121M analyzer programmed for physiological analysis and capable of giving unequivocal identification of asparagine.

Automated Sequence Analysis. The amino-terminal sequences of S-[^{14}C]carbamoylmethylated Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} were investigated by automated sequential degradation with an updated 890B Beckman sequencer by use of an *N*,*N*′-dimethylbenzylamine program (Kurosky et al., 1976) as well as a quadrol program (Beckman program 060275). Samples of protein (3–7 mg) introduced into the cup were quantitated by amino acid analysis. Sequencer products were converted to Pth amino acid residues and identified by gas–liquid chromatography and by amino acid analysis after HI back-hydrolysis. The ethyl acetate and aqueous phases of the thiazolinone conversion mixtures were analyzed separately. Identification of half-cystinyl residues was facilitated by alkylation with iodo[1-^{14}C]acetamide. An aliquot (1/5) of each residue collected was measured for radioactivity. Details of sequencer operation and of the methods of identification and quantitation have been previously described by Kurosky et al. (1976). Confirmation of carboxyl groups was achieved in some cases when glycine ethyl ester was coupled to the free carboxyl groups in the presence of 1-ethyl-3-dimethylaminopropylcarbodiimide according to Gibson & Anderson (1972).

Manual Dns–Edman Sequence Analysis. Manual sequence analysis of peptides was carried out by use of the Dns–Edman procedures (Hartley, 1970; Gray, 1972). Dansylated amino acids were identified by two-dimensional thin-layer chromatography on 5 × 5 cm polyamide sheets (Kurosky & Hofmann, 1976).

Results

Polyacrylamide Gel Electrophoresis of Gc. Preparations of Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} exhibited identical mobilities during urea–acetic acid and sodium dodecyl sulfate–poly-

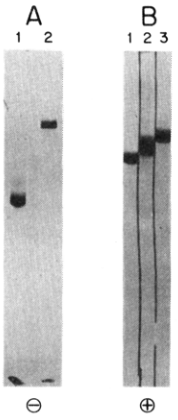

FIGURE 1: Polyacrylamide gel electrophoresis of denatured Gc. (A) Acrylamide gels (10%) in 6 M urea–0.9 M acetic acid, pH 2.7, were subjected to electrophoresis for 4 h at 25 V/cm at 10 °C. Samples were dialyzed against 9 M urea containing 0.3 M 2-mercaptoethanol for 6 h: (1) unreduced control and (2) S-carbamoylmethylated. (B) Acrylamide gels (10%) in 0.1 M sodium phosphate buffer containing 0.1% sodium dodecyl sulfate, pH 7.0, were subjected to electrophoresis for 15 h at 2 mA/tube at 20 °C: (1) unreduced control, (2) reduced directly with 2-mercaptoethanol in 1% sodium dodecyl sulfate, and (3) S-carbamoylmethylated sample from A (2) above was further dialyzed for 14 h against sodium dodecyl sulfate–phosphate buffer containing 0.3 M 2-mercaptoethanol.

acrylamide gel electrophoresis. They also eluted at the same elution volume during gel filtration. Therefore, these Gc types will not be differentiated in the description of the results. Gel electrophoresis in 4.5 M urea–0.9 M acetic acid (Figure 1A) indicated that the unreduced Gc (gel 1) had a considerably faster cathodal electrophoretic mobility when compared to that of reduced Gc (gel 2). Sodium dodecyl sulfate gel electrophoresis (Figure 1B) showed that unreduced Gc (gel 1) migrated more rapidly than reduced Gc (slots 2 and 3). Gc samples reduced directly in sodium dodecyl sulfate (gel 2) had electrophoretic mobilities intermediate between those of unreduced Gc (gel 1) and Gc reduced in guanidine hydrochloride (gel 3). From a standard graph prepared by use of proteins of known molecular weights, the molecular weights of Gc reduced in guanidine, Gc reduced in sodium dodecyl sulfate, and unreduced Gc were estimated to be 55 500 ± 1000, 49 700 ± 800, and 46 100 ± 800, respectively.

Gel Filtration. On Sephadex G-150 gel filtration in 0.05 M sodium acetate buffer, pH 4.75, containing 6 M guanidine hydrochloride, unreduced Gc eluted from the column more rapidly than reduced Gc. The molecular weights of reduced and unreduced Gc were calculated to be 51 500 ± 500 and 36 900 ± 600, respectively. Since amino acid analysis (Bowman, 1969) indicated the presence of approximately 20 mol of half-cystine/mol of Gc protein, the molecular weights of reduced and alkylated Gc include approximately 1500 due to carbamoylmethyl groups. For protein quantitation in subsequent studies on amino- and carboxyl-terminal analyses, an average molecular weight of 52 000 was calculated from results obtained by gel electrophoresis and gel filtration, after subtraction of 1500 for carbamoylmethyl groups.

Amino-Terminal Sequence Analysis. Automated sequential Edman degradation of the amino termini of the three major Gc types indicated a single sequence for each protein. No differences could be detected in the first 20 amino-terminal residues of purified Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} (Table I). The average repetitive yield calculated from gas chromatographic analysis of leucine at positions 1 and 19 was 92.8 ± 1.8%. The average yield of amino-terminal leucine was 0.57 ± 0.15 mol/mol of Gc protein. The initial residue recovery from sequencer analyses of proteins usually ranged from 55 to 90%. A control analysis of sperm whale myoglobin performed close to the same time yielded 0.69 mol of valine/mol of protein. Typically, the incorporation of radioactivity in the *S*-[^{14}C]carbamoylmethyl derivative of Gc was 4000–5000 cpm above background (200 cpm) per cysteinyl residue per total protein analyzed (90–130 nmol).

Table I: Automated Sequence Analysis[a] of the Amino-Terminal Regions of Gc 1_{slow}, Gc 1_{fast}, and Gc 2_{slow}

Edman cycle	residue (nmol)			
	Gc 1_{slow}	Gc 1_{fast}	Gc $1_{slow/fast}$[b]	Gc 2_{slow}
0[c]	– (129)	– (110)	– (100)	– (91)
1	Leu (102)	Leu (45)	Leu (46)	Leu (68)
2	Gln (61)	Gln (35)	Gln (34)	Gln (50)
3	Arg (5)	Arg *d*	Arg (36)	Arg (8)
4	Gly (106)	Gly (47)	Gly (49)	Gly (71)
5	Arg (9)	Arg *d*	Arg (30)	Arg (10)
6	Asn (35)	Asn (29)	Asn (24)	Asn (38)
7	Tyr (59)	Tyr (24)	Tyr (23)	Tyr (46)
8	Glu (52)	Glu (22)	Glu (33)	Glu (46)
9	Lys (19)	Lys (14)	Lys (22)	Lys (14)
10	Asn (43)	Asn *d*	Asn (28)	Asn (46)
11	Lys (23)	Lys (4)	Lys (19)	Lys (15)
12	Val (40)	Val (24)	Val (15)	Val (47)
13	Cys[e] (21)	Cys[e] *d*	Cys[e] (15)	Cys[e] (36)
14	Lys (21)	Lys *d*	Lys (18)	Lys (10)
15	Glu (26)	Glu (13)	Glu (24)	Glu (29)
16	Phe (36)	Phe (19)	Phe (14)	Phe (38)
17	Ser (9)	Ser (7)	Ser (5)	Ser (18)
18	His (3)	His *d*	His (19)	His (3)
19	Leu (35)	Leu (14)	Leu (22)	Leu (36)
20	Gly (30)	Gly (22)	Gly (12)	Gly (35)
21			Lys (10)	
22			Glx (17)	
23			Asx (13)	
24			Phe (8)	
25			Thr (8)	

[a] Pth residues were quantitated by amino acid analysis after HI hydrolysis and by gas chromatography as previously described (Kurosky et al., 1976). Only the results obtained from amino acid analysis are shown. Serine and cysteine were quantitated as alanine and threonine as α-aminobutyric acid. Results given are for a single analysis; however, Gc 1_{slow}, Gc 1_{fast}, Gc 2_{slow}, and Gc $1_{slow/fast}$ were analyzed two, one, two, and three times, respectively. Most analyses were performed on S-[^{14}C]carbamoylmethylated protein. [b] Analysis was performed on an equimolar mixture of Gc 1_{slow} and Gc 1_{fast} by use of the protein–quadrol program. [c] Initial nanomoles of protein applied to sequencer. [d] Sample was lost or poorly hydrolyzed by HI. Residue assignments were deduced from analysis of Gc $1_{slow/fast}$ and Gc 1_{slow}. [e] Confirmed by measurement of radioactivity due to the *S*-[^{14}C]-carbamoylmethyl label.

Carboxyl-Terminal Analysis. The carboxyl-terminal sequences of Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} were investigated after hydrolysis with carboxypeptidase A. The results of four experiments are summarized in Table II. Carboxypeptidase A hydrolysis of native protein at low enzyme concentrations [(0.25 and 2 μg/0.3 mg of Gc)/0.25 mL] released leucine at a slightly faster rate than isoleucine. At higher concentrations of enzyme [(20 and 50 μg/0.2–0.6 mg of Gc)/0.25 mL] a significant amount of asparagine was also released. No other amino acids were released even after prolonged hydrolysis with carboxypeptidase A. Hydrolysis of S-carbamoylmethylated Gc yielded distinctly more leucine than isoleucine or asparagine. The results suggested that Gc 1_{slow} and Gc 2_{slow} had identical carboxyl-terminal sequences: –Asn-Ile-Leu-COOH. A limiting amount of Gc 1_{fast} permitted only one experiment to be performed. Although no definite sequence could be established for Gc 1_{fast}, the results are consistent with the sequence proposed for Gc 1_{slow} and Gc 2_{slow}. The yield of carboxyl-terminal leucine released even after prolonged hy-

Table II: Carboxypeptidase A Hydrolysis of Gc 1_{slow}, Gc 1_{fast}, and Gc 2_{slow}

sample	preparation	enzyme (μg)	hydrolysis duration (min)	amino acid (mol released/mol peptide) Leu	Ile	Asn
(1) Gc 1_{slow}	native	0.25	10	0.25	0.12	
	native	2	10	0.97	0.76	
	native	50	20	0.78	0.78	0.76
(2) Gc $1_{slow/fast}$[b]	native	20	10	0.98	0.93	0.65
	alkylated[c]	20	10	0.79	0.43	0.45
(3) Gc 1_{fast}	native	50	20	0.58	0.56	0.51
(4) Gc 2_{slow}	native	0.25	10	0.24	0.11	
	native	2	10	0.89	0.78	
	native	50	240	0.93	0.93	0.84
	alkylated[c]	20	10	0.83	0.52	0.47

[a] Calculated by assumption of a molecular weight of 52 000 for Gc. No other amino acids were released in yields greater than 0.1 mol/mol of Gc protein. Data are taken from three separate experiments on Gc 1_{slow} and on Gc 2_{slow} and from one experiment each on Gc $1_{slow/fast}$ and on Gc 1_{fast}. Protein concentration was 0.2–0.6 mg/0.25 mL. [b] Equimolar mixture of Gc 1_{fast} and Gc 1_{slow}. [c] Reduced and S-[^{14}C]carboxymethylated protein.

drolysis with 2–50 μg of enzyme was always close to 1 and never exceeded 1 mol/mol of protein.

Peptide Mapping. On the basis of a molecular weight of 52 000, the amino acid composition of Gc (Bowman, 1969) indicated 35–38 lysyl and 14–16 arginyl residues per molecule. Therefore, cleavage by trypsin would be expected to yield 50–55 peptides. Highly reproducible peptide maps for Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} are shown in Figure 2. The maps of each protein included approximately 19–20 major spots among the acidic peptides, 11–13 major spots among the neutral peptides, and 20–21 major spots among the basic peptides. The total number of major peptides (50–54) was in good agreement with that predicted by the amino acid composition. The number of *S*-[^{14}C]carboxymethylcysteinyl peptides detected by radioautography was 12–14 among the acidic peptides, 2 among the neutral peptides, and 3 among the basic peptides, compared to the 19–20 half-cystinyl residues determined by amino acid analysis (Bowman, 1969). These results are consistent with the occurrence of Gc as a single polypeptide with a molecular weight of approximately 52 000.

Peptide maps of the neutral and slow-moving peptides were highly distinctive for each Gc protein (Figure 2C). The Gc 2_{slow} preparations contained two characteristic peptides, 2N1 and 2N2. Peptide 2N2 was completely absent from the maps of Gc 1_{fast} and Gc 1_{slow}, while only minor traces of a peptide of similar mobility to peptide 2N1 could be detected in the maps of Gc 1 proteins. Similarly, Gc 1_{fast} contained one peptide (1fN6) that was absent from both Gc 1_{slow} and Gc 2_{slow}, while Gc 1_{slow} had one peptide (1sN5) that was absent from Gc 2_{slow} and present in diminished quantities in Gc 1_{fast}. In addition, although no differences could be observed in the basic peptides of the different Gc proteins (Figure 2B), one acidic peptide, 1sfA1, was present in low yields in both Gc 1_{fast} and Gc 1_{slow} but was absent from Gc 2_{slow} (Figure 2A). This characteristic pattern of peptides was consistently observed in preparations of Gc purified from two Gc 1 and two Gc 2 homozygotes. The amino acid compositions, amino termini, and electrophoretic mobilities of these peptides are shown in Table III. These compositions were calculated on the basis of one lysine. The low levels of the usual contaminants such as glycine and serine in the compositions and the absence of other amino-terminals after dansylation indicated that the

Table III: Amino Acid Compositions of Characteristic Peptides from Different Gc Types[a]

	2N1	2N2	1sfA1	1sN5	1fN6
Asp		1.05 (1)	0.43		
Thr		0.99 (1)	2.67 (3)	0.95 (1)	0.95 (1)
Ser		0.23	0.38		
Glu	0.92 (1)	0.21	3.02 (3)	1.09 (1)	1.06 (1)
Pro		1.97 (2)	1.87 (2)	0.83 (1)	0.90 (1)
Gly		0.27	0.45	0.17	
Ala	1.03 (1)	1.01 (1)	2.95 (3)	1.55 (2)[b]	1.51 (2)[b]
Leu	1.01 (1)	0.87 (1)	1.93 (2)	1.08 (1)	1.13 (1)
Lys	0.98 (1)	1.08 (1)	1.19 (1)	1.09 (1)	1.10 (1)
total[c] residues	4	7	14	7[b]	7[b]
$m_{6.5}$[d]	0	0	+0.29	0	+0.18
yield[e] (%)	32	9	6	35	23
amino terminus[f]	Glu	Leu	Leu	Ala	Ala

[a] Results are expressed as moles of amino acid per mole of peptide, by assumption of the presence of one lysine. Amino acids present in yields <0.15 mol/mol of peptide have not been included. Closest integers are shown in parentheses. Data are derived from two different preparations of each peptide. When one preparation was obviously more pure than the other, the composition given is for the purer preparation; otherwise, an average of the two compositions is shown. [b] See Discussion. [c] Sum of closest integers. [d] Electrophoretic mobility at pH 6.5, relative to aspartic acid. [e] Yield of peptide per mole of Gc protein hydrolyzed. [f] Determined by reaction with dansyl chloride.

Table IV: Amino Acid Sequence of Three Characteristic Peptides of Gc 1 and Gc 2[a]

peptide[b]	sequence[c]
2N1	Glu-Leu-Ala-Lys
2N2	Leu-Pro-Asp-Ala-Thr-Pro-Lys
1sN5	Ala-Thr-Leu-Pro-Glu-Ala-Lys

[a] Sequence is obtained by Dns-Edman methodology. [b] Compositions and $m_{6.5}$ are given in Table III. [c] Peptides are arranged to emphasize possible alignment.

peptides were pure. The yields of alanine in peptides 1sN5 and 1fN6 were consistently low—about 1.5 mol/2 mol of peptide. This may be due in part to the fact that alanine was amino terminal and was partially modified by reaction with fluorescamine.

The amino acid sequences obtained for the peptides characteristic for Gc 2, i.e., 2N1 and 2N2, are compared to the sequence of peptide 1sN5 characteristic of Gc 1_{slow} in Table IV. Sufficient material was not available for sequence determination of 1fN6 and 1sfA1.

Identical amino acid compositions and amino termini of peptides 1sN5 and 1fN6 (Table III) led us to reinvestigate the suggestion (Svasti & Bowman, 1978) concerning the involvement of sialic acid in Gc 1_{fast}/Gc 1_{slow} differences at the peptide level. Thus, the characteristic neutral peptides of Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow} were hydrolyzed with neuraminidase and compared by electrophoresis at pH 2.0 (Figure 3) to controls incubated with the hydrolysis buffer only. Although the Gc 1 peptides smeared slightly on paper, it was clear that the cathodal electrophoretic mobility of peptide 1fN6 (lanes 1 and 2) was greatly increased after neuraminidase digestion, suggesting that peptide 1fN6 did indeed contain sialic acid. No other peptides demonstrated increased cathodal electrophoretic mobilities after neuraminidase digestion. The electrophoretic mobility of peptide 1fN6 after neuraminidase hydrolysis was only slightly slower than that of peptide 1sN5 (lane 3), indicating that most if not all of the electrophoretic difference between the two peptides can be accounted for by sialic acid. The presence of a 1sN5-like peptide in Gc 1_{fast}

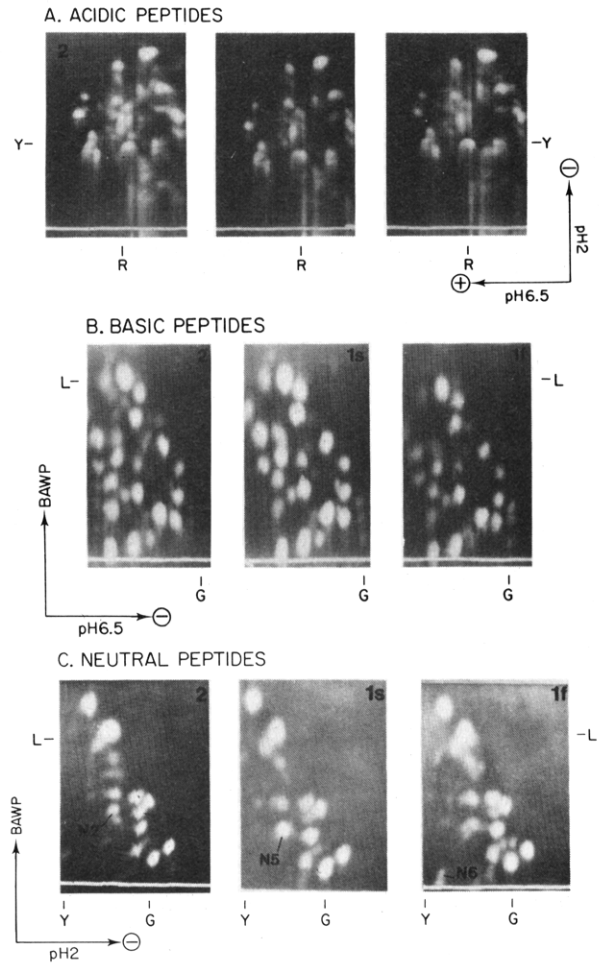

FIGURE 2: Tryptic peptide maps of S-[^{14}C]carboxymethylated Gc. Digests of Gc 1_{fast} (1f), Gc 1_{slow} (1s), and Gc 2_{slow} (2) were initially fractionated by electrophoresis at pH 6.5 for 40 min at 3 kV. The paper was cut into three sections corresponding to the acidic peptides ($m_{6.5}$ from +1.16 to +0.185), neutral and slow-moving peptides ($m_{6.5}$ from +0.185 to −0.09), and basic peptides ($m_{6.5}$ from −0.09 to −1.00). (A) Acidic peptides, fractionated by electrophoresis at pH 2.0 for 75 min at 2 kV at right angles to the pH 6.5 direction; (B) basic peptides, fractionated by chromatography in solvent BAWP at right angles to pH 6.5 electrophoresis; and (C) "neutral" peptides, fractionated by electrophoresis at pH 2.0 in the pH 6.5 direction, followed by chromatography in solvent BAWP at right angles to electrophoresis. Peptides were visualized when stained with 0.01% fluorescamine in acetone. The positions of external markers leucine, 6-dinitrophenyllysine, methyl green, and phenol red are shown by L, Y, G, and R, respectively. Arrows indicate peptides characteristic of different Gc proteins (peptides 2N1, 2N2, 1sN5, 1fN6, and 1sfA1).

preparations may thus be explained by the loss of sialic acid.

Discussion

In this study, we have provided three lines of evidence that strongly indicate the presence of a single polypeptide chain in human Gc. This corrects previous studies (Bowman & Bearn, 1965; Bowman, 1969) in which the conclusion was drawn that Gc consisted of two similar but nonidentical subunits. The molecular weights of carbamoylmethylated Gc (51 500 by gel filtration and 55 500 by gel electrophoresis) are in good agreement with the values reported by others (51 000–56 000) for denatured human Gc (Kuusela & Pihko, 1972; Peterson, 1971; Haddad & Walgate, 1976; Bouillon et al., 1976; Imawari et al., 1976). Artifactual cleavage by the reagents employed during the extended periods of dialysis in 5 M guanidine and 8 M urea could possibly have been responsible for the results reported previously (Bowman & Bearn, 1965; Bowman, 1969). In the present studies, considerable care was taken to use the purest chemicals available to avoid any such artifactual cleavage. In addition, we have used a 50-fold molar excess of dithiothreitol over half-cystine content, where a two- to threefold molar excess is generally sufficient for a wide variety of proteins, for example, in mouse immunoglobins (Svasti & Milstein, 1972) and in human protamines (Pongsawasdi & Svasti, 1976). Samples reduced and alkylated in guanidine were further dialyzed against 9 M urea containing 0.3 M 2-mercaptoethanol for 6 h and against 1% sodium dodecyl sulfate containing 0.3 M 2-mercaptoethanol for 14 h. Amino acid analysis of these proteins confirmed complete alkylation and absence of any disulfide bonds. It has also been possible to substantiate these results by extensive peptide mapping, by sodium dodecyl sulfate electrophoresis, and by amino- and carboxyl-terminal analysis.

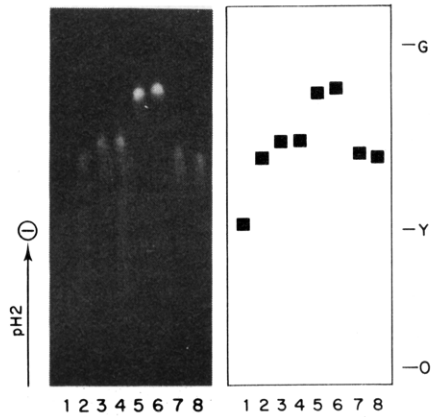

FIGURE 3: Effect of neuraminidase hydrolysis on the electrophoretic mobilities of the neutral characteristic peptides of Gc. The interpretation of the electropherogram is shown schematically on the right. Peptides (1.5–5 nmol) were digested with 2 μg of neuraminidase in 0.05 M pyridine acetate buffer, pH 5.0, for 14 h at 37 °C and compared to controls incubated in buffer alone. Electrophoresis was carried out at pH 2.0 for 75 min at 2 kV, and peptides were stained with 0.01% fluorescamine in acetone. (1) 1fN6 control; (2) 1fN6 digested; (3) 1sN5 control; (4) 1sN5 digested; (5) 2N1 control; (6) 2N1 digested; (7) 2N2 control; and (8) 2N2 digested. The positions of the origin, 6-dinitrophenyllysine, and methyl green are idicated by O, Y, and G, respectively.

Thus, despite the fact Gc was reduced with a high concentration of dithiothreitol (50-fold excess over half-cystine content) in a strongly denaturing solvent, no dissociation of Gc into subunits could be observed by gel electrophoresis (Figure 1) or by gel filtration. Reduced Gc appeared to have a significantly higher molecular weight than unreduced Gc, suggesting that in the absence of reducing agents Gc maintains a relatively compact shape even in 7.5 M guanidine.

Further support that Gc is a single polypeptide was derived from the amino-terminal analysis of Gc 1_{fast}, Gc 1_{slow}, and Gc 2_{slow}, which indicated that a single sequence (Table I) was obtained in each case with a molar yield approaching that obtained for myoglobin. We did observe that carbamoylmethylated Gc was more recalcitrant to automated sequence analysis than were most proteins. Alkylated Gc readily formed a hydrophobic, sticky precipitate after dialysis. Highly nonpolar proteins are known to respond poorly to automated sequence analysis (Bailey et al., 1977). Poor HI hydrolysis and insufficient amounts of Gc 1_{fast} did not permit complete characterization of the first 20 residues; however, the blank positions were established a priori from analysis of Gc 1_{slow} and Gc $1_{slow/fast}$. The quantitation of these residues (3, 5, 10, 13, 14, and 18) in Gc $1_{slow/fast}$ is compatible with the sequence shown for Gc 1_{fast}. Moreover, repeated analysis of Gc $1_{slow/fast}$ did not indicate any other residues at these positions. Measurement of radioactivity of the Pth derivatives of Gc 1_{fast}

confirmed *S*-[^{14}C]carbamoylmethylcysteine only at position 13 in the 20 residues analyzed.

Carboxyl-terminal analysis with carboxypeptidase A released leucine at the carboxyl terminus in yields close to 1 mol/mol of Gc protein in agreement with the presence of a single subunit in Gc with a molecular weight of approximately 52 000.

The third line of evidence supporting the existence of a single polypeptide with a molecular weight of approximately 52 000 stems from good agreement between the number of tryptic peptides detected by peptide mapping and that predicted from the amino acid composition (Bowman, 1969). In our present results, the 50–55 tryptic peptides predicted from amino acid analysis were detected in fingerprints of S-carboxymethylated Gc. These results differed from previous experiments (Bowman, 1969) and from those of others (Ruoslahti, 1967; Ruoslahti et al., 1971) where 25–30 tryptic peptides of the 70 expected from S-aminoethylated Gc were detected. The differences observed in earlier work probably resulted from the use of inadequate fractionation procedures involving electrophoresis at pH 3.5, followed by chromatography in pyridine–isoamyl alcohol (Bowman, 1969; Ruoslahti, 1967; Ruoslahti et al., 1971). Peptide fractionation procedures described here were far more extensive. Although fingerprints were cut into sections, we took great care to avoid obtaining the same peptides on two sections of paper. We compared the capacity of our chromatographic solvent (BAWP) with that of the pyridine–isoamyl alcohol solvent used previously to resolve neutral and basic peptides. The neutral and basic peptides contained only two and three *S*-carboxymethylcysteine peptides, respectively, which were not expected to be present in S-aminoethylated Gc. In contrast to the chromatographic solvent used in this study which resolved 11–13 neutral peptides and 20–21 basic peptides, only six neutral peptides and about five basic peptides moved in the pyridine–isoamyl alcohol solvent.

While the close similarities in the peptide maps of the three major Gc components suggest that they are very similar in amino acid sequence, consistent differences could be observed that are characteristic of each. The results indicated that Gc 1 and 2 differed by at least two peptides, while Gc 1_{fast} and Gc 1_{slow} appeared to differ by one peptide. One of the two characteristic peptides of Gc 2_{slow}, peptide 2N1, has the same amino acid composition as that isolated by Ruoslahti et al. (1971). However, peptides 1sN5 and 1fN6, which distinguish Gc 1_{slow} and Gc 1_{fast}, respectively, have slightly different compositions to those isolated by these authors. The smaller amounts of contaminants in the compositions of our peptides (Table III) suggest that our peptides were obtained in a pure state and should therefore show reliable compositions. Our improved mapping procedures also showed two hitherto undetected peptides, peptide 2N2, present in Gc 2 but not Gc 1, and peptide 1sfA1, common to Gc 1_{slow} and Gc 1_{fast} but absent from Gc 2_{slow}.

The close similarities in the amino acid compositions of the peptides that distinguish Gc 1_{fast} and Gc 1_{slow} led Rouslahti et al. (1971) to propose that these peptides differ in amide content. However, the pK of the γ-carboxyl group of glutamic acid (about 4.5) is too high to explain the large differences in the electrophoretic mobilities of peptides 1sN5 and 1fN6 at pH 2.0 (Figure 2C). Importantly, peptide 1fN6 could not possess a negative charge at pH 6.5, as it does, unless one or more of the residues were somehow modified. Our experiments indicate that the differences in electrophoretic mobility are largely eliminated after neuraminidase treatment (Figure 3). This suggests that, consistent with our experiments on the intact protein (Svasti & Bowman, 1978), the Gc 1_{slow} and Gc 1_{fast} peptides differ in the number of sialic acid residues and possibly in other sugar residues as well.

It is of interest that peptide 1fN6 contained no asparagine (Table III) in view of the fact that carbohydrate is strongly suggested by the change of its electrophoretic mobility after neuraminidase digestion (Figure 3). Although the *N*-acetylglucosaminylasparaginyl linkage predominates in plasma glycoproteins, the carbohydrate linkage in peptide 1fN6 is predicted to be *N*-acetylgalactosaminylthreonyl since asparagine is missing in its composition. This bond has been found in mucins (Anderson et al., 1964), in fetuin (Spiro & Bhoyroo, 1974), and in immunoglobulins (Clamp & Johnson, 1972).

In contrast to the posttranslational nature of the Gc 1_{fast}/Gc 1_{slow} heterogeneity, Gc^1/Gc^2 polymorphism appears to be due to differences in primary structure. The amino acid compositions of the Gc 1 and of the Gc 2 peptides show some similarities, suggesting that they may be derived from similar regions of the polypeptide chains of Gc 1 and of Gc 2. At the present time, however, we are unable to make a confident and complete alignment of the Gc 1 and Gc 2 peptides to pinpoint all differences in these proteins. This will require determination of the complete sequences surrounding the characteristic peptides. Furthermore, other characteristic peptides may still remain undetected (especially in the acidic region of our maps where the peptides were not completely resolved).

A computer search of the first 20 amino-terminal residues of Gc [program SEARCH, Dayhoff (1976)] failed to show chemical similarity with any previously reported sequence. Although only a small segment was compared, these results suggest that Gc is a unique gene product.

In conclusion, our results indicate that, in contrast to the posttranslational nature of the Gc 1_{fast}/Gc 1_{slow} heterogeneity, the basis of the Gc^1/Gc^2 polymorphism is related to dissimilarities of primary structure. Therefore, the major polymorphism of Gc, i.e., Gc 1 and Gc 2, can be explained by the existence of two autosomal alleles at a single locus.

Acknowledgments

We thank Horace D. Kelso, Billy Touchstone, and Linda Merryman for expert technical assistance. We are especially indebted to Dr. Don R. Barnett for kindly providing the plasma fractions used in the isolation of Gc and for many helpful discussions.

References

Ambler, R. P. (1972) *Methods Enzymol. 25*, 143–154.

Anderson, B., Seno, N., Sampson, P., Riley, J. G., Hoffman, P., & Meyer, K. (1964) *J. Biol. Chem. 239*, PC2716–PC2717.

Bailey, G. S., Gillett, D., Hill, D. F., & Petersen, G. B. (1977) *J. Biol. Chem. 252*, 2218–2225.

Bouillon, R., van Baelen, H., Rombauts, W., & de Moor, P. (1976) *Eur. J. Biochem. 66*, 285–291.

Bowman, B. H. (1969) *Biochemistry 8*, 4327–4335.

Bowman, B. H., & Bearn, A. G. (1965) *Proc. Natl. Acad. Sci. U.S.A. 53*, 722–729.

Clamp, J. R., & Johnson, J. (1972) *Glycoproteins 5A*, 612–652.

Cleve, H., & Bearn, A. (1962) *Prog. Med. Genet. 2*, 64–82.

Daiger, S. P., Schanfield, M. S., & Cavalli-Sforza, L. L. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2076–2080.

Dayhoff, M. O. (1976) in *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 2, p 3, National Biomedical

Research Foundation, Washington, DC.
Fraenkel-Conrat, H., Harris, J. I., & Levy, A. L. (1955) *Methods Biochem. Anal. 2*, 359–425.
Gibson, D., & Anderson, P. J. (1972) *Biochem. Biophys. Res. Commun. 49*, 453–459.
Gray, W. R. (1972) *Methods Enzymol. 25*, 333–344.
Haddad, J. G., Jr., & Walgate, J. (1976) *J. Biol. Chem. 251*, 4803–4809.
Hartley, B. S. (1970) *Biochem. J. 119*, 805–822.
Hirschfeld, J. (1959) *Acta Pathol. Microbiol. Scand. 47*, 160–168.
Hirschfeld, J., Jonsson, B., & Rasmuson, M. (1960) *Nature (London) 185*, 931–932.
Imawari, M., Kida, K., & Goodman, DeW. S. (1976) *J. Clin. Invest. 58*, 514–523.
Kitchin, F. D. (1965) *Proc. Soc. Exp. Biol. Med. 119*, 1153–1155.
Kitchin, F. D., & Bearn, A. (1966) *Am. J. Hum. Genet. 18*, 201–214.
Kurosky, A., & Hofmann, T. (1976) *Can. J. Biochem. 54*, 872–884.
Kurosky, A., Kim, H.-H., & Touchstone, B. (1976) *Comp. Biochem. Physiol. B 55*, 453–459.
Kurosky, A., Markel, D. E., & Peterson, J. W. (1977) *J. Biol. Chem. 252*, 7257–7264.
Kuusela, P., & Pihko, H. (1972) *Biochim. Biophys. Acta 263*, 445–447.
Mann, K. G., & Fish, W. W. (1972) *Methods Enzymol. 26*, 28–42.
Mendez, E., & Lai, C. Y. (1975) *Anal. Biochem. 65*, 281–292.
Offord, R. E. (1966) *Nature (London) 211*, 591–593.
Panyim, S., & Chalkley, R. (1969) *Arch. Biochem. Biophys. 130*, 337–346.
Peacock, A. G., Bunting, S. L., & Queen, K. G. (1965) *Science 147*, 1451–1453.
Peterson, P. A. (1971) *J. Biol. Chem. 246*, 7748–7754.
Pongsawasdi, P., & Svasti, J. (1976) *Biochim. Biophys. Acta 434*, 462–473.
Putnam, F. W. (1977) *Plasma Proteins 3*, 333–357.
Reinskou, T. (1963) *Acta Pathol. Microbiol. Scand. 59*, 526–532.
Ruoslahti, E. (1967) *Ann. Med. Exp. Biol. Fenn. 45*, 1–13.
Ruoslahti, E., Simons, K., & Ehnholm, C. (1971) *Eur. J. Immunol. 1*, 45–48.
Shapiro, A. L., Vinuela, E., & Maizel, J. V. (1967) *Biochem. Biophys. Res. Commun. 28*, 815–820.
Spiro, R. G., & Bhoyroo, V. D. (1974) *J. Biol. Chem. 249*, 5704–5717.
Svasti, J., & Milstein, C. (1972) *Biochem. J. 128*, 427–444.
Svasti, J., & Bowman, B. H. (1978) *J. Biol. Chem. 253*, 4188–4194.
Waley, S. G., & Watson, J. (1953) *Biochem. J. 55*, 328–337.
Weber, K., Pringle, J. R., & Osborn, M. (1972) *Methods Enzymol. 26*, 3–27.

Apolipoprotein A-II: Chemical Synthesis and Biophysical Properties of Three Peptides Corresponding to Fragments in the Amino-Terminal Half†

T. C. Chen, J. T. Sparrow,‡ A. M. Gotto, Jr., and J. D. Morrisett*,‡

ABSTRACT: Three peptide fragments of apolipoprotein A-II corresponding to residues 17–31, 12–31, and 7–31 have been synthesized by solid-phase techniques and purified to apparent homogeneity. Each of these fragments contains residues 18–30, a region previously proposed to possess potential amphipathic helical properties. Secondary structural changes of these synthetic fragments accompanying their interaction with phospholipid have been studied by circular dichroism. The magnitude of this interaction has been evaluated from the yields and stoichiometry of lipid–protein complexes isolated by density gradient ultracentrifugation. Fragment 17–31, the smallest peptide containing the proposed amphipathic helix, did not interact with dimyristoylphosphatidylcholine (DMPC) single bilayer vesicles at 24 °C; upon addition of DMPC, no ellipticity change could be detected nor could a stable lipid–peptide complex be isolated. However, fragments 12–31 and 7–31 did interact with phospholipid; in the absence of lipid, both fragments had primarily disordered structures, but when isolated as DMPC–peptide complexes, both fragments possessed increased helical structure. The phospholipid:peptide molar ratio was 14:1 for fragment 12–31 and 27:1 for fragment 7–31. Studies of space-filling models of these fragments suggest that hydrophobicity and/or length are important properties of phospholipid binding apoproteins.

Apolipoprotein A-II (apoA-II)[1] is one of the major protein components of human serum high density lipoproteins (HDL). It possesses two identical polypeptide chains each containing 77 amino acids. These chains are cross-linked by a disulfide bond at cystine-6 (Brewer et al., 1972). The approximate distribution of its secondary structure is 35% α helix, 13% β structure, and 52% disordered structure. Other physical properties of this apoprotein have been recently reviewed (Morrisett et al., 1977).

In this laboratory, considerable effort has been directed toward identifying the structural features of apoA-II which

† From the Departments of Medicine and Biochemistry, Baylor College of Medicine, The Methodist Hospital, Houston, Texas 77030. *Received November 8, 1978.* The research described in this paper was performed under the auspices of the Atherosclerosis, Lipids, and Lipoproteins Section of the National Heart and Blood Vessel Research and Demonstration Center, Baylor College of Medicine; it was also supported by a grant-in-aid (No. 78-1042) from the American Heart Association to J.D.M.

‡ Established Investigator of The American Heart Association.

[1] Abbreviations used: HDL, high density lipoproteins; apoA-II, apolipoprotein A-II (mol wt 17 149), a major apoprotein of HDL; DMPC, dimyristoylphosphatidylcholine; CD, circular dichroism; EDTA, ethylenediaminetetraacetate; Boc, *tert*-butyloxycarbonyl; TLC, thin-layer chromatography; DCC, *N,N'*-dicyclohexylcarbodiimide.

0006-2960/79/0418-1617$01.00/0