# Composition and Subunit Structure of the Inherited Group-Specific Protein of Human Serum*

Barbara H. Bowman†

ABSTRACT: Amino acid composition, disulfide reduction, peptide maps, and end-group analysis of the group-specific protein (Gc) suggested that this serum protein is comprised of three very similar and indivisible subunits, one of which is common to both fast and slow bands of Gc 1-1 and Gc 2-2. Each of two other subunits seems to be unique to the fast and slow bands, respectively. The presence of three similar but not identical subunits implies three separate genetic loci. Carboxypeptidase digestion of the Gc molecule also demonstrates the presence of isoleucine and leucine in equal quantities suggesting that one-half the subunits have C-terminal leucine and one-half have isoleucine in the C terminus. Faint homology is observed in the C-terminal sequence of serum albumin and Gc. This and previous linkage information indicate the possibility of a common genetic ancestor to the albumin and Gc structural genes.

The group-specific protein (Gc) has been observed in all human serums examined, although its function remains unknown. There are three common Gc types; some serums contain two rapidly migrating components, Gc 1-1, others, two slowly migrating components, Gc 2-2, while a third group is characterized by the presence of three migrating components with the middle electrophoretic band in approximately twice the concentration of the other two, Gc 2-1 (Cleve and Bearn, 1962; Bearn *et al.*, 1964). Family studies indicated that these three patterns correspond to the two homozygous and one heterozygous phenotypes of allelic genes. Recently, Azen *et al.* (1969) have been able to separate the heterozygote phenotype Gc 2-1 into four bands, two characteristic of Gc 1-1 and two characteristic of the Gc 2-2 phenotype.

The group-specific protein has been characterized as having a molecular weight of 51,000 with 3.3% carbohydrate and has been found to belong to the small group of serum proteins which contains no sialic acid. Amino acid analyses have been performed on proteins from the three common Gc phenotypes (Cleve *et al.*, 1963).

The presence of some 45 residues of aspartic acid in addition to some 50 glutamic acid residues in all 3 Gc types explained the rapid anodal migration of this protein during electrophoresis.

Electrophoresis of serums from individuals homozygous for Gc 1-1 or Gc 2-2 demonstrates heterogeneity in the protein bands. The heterogeneity of the Gc protein raised the possibility of the existence of polypeptide subunits in the Gc molecule. The existence of subunits was later substantiated by Bowman and Bearn (1965) after finding one-half the expected number of tryptic peptides in fingerprints and a decrease in molecular weight of the Gc protein following reduction with mercaptoethanol in the presence of urea. Thus far, the subunits have been characterized by analytical ultracentrifugation, where the average molecular weight was 25,000 ± 5,000 and in gel filtration, where the molecular weight was decreased after reduction. The number of tryptic peptides present on fingerprints of Gc indicated that this molecule was comprised of similar subunits.

The duplication of an ancestral structural gene of Gc followed by genetic alterations in some of the codons would result in a system of similar but nonidentical subunits. Whether or not the structural genes controlling the synthesis of the Gc subunits are linked on the same region of the chromosome is unknown. It has been shown, however, that the structural gene governing the synthesis of human serum albumin is located adjacent to the Gc structural gene system (Weitkamp *et al.*, 1966). This has raised the question whether or not the albumin and Gc structural genes may have arisen through gene duplication from a common ancestor gene. Similarities in the structure of albumin and Gc have not heretofore been investigated.

Investigation of the structure of this protein has been greatly hindered by the difficulties encountered in its purification. These have been due to the crowded electrophoretic region where the Gc bands migrate and to the low concentration in human serum. From each 250-ml unit of plasma of serum, approximately 1–5 mg of Gc was obtained after purification. The concentration of Gc in human serum ranges from 0.4 mg/ml (Cleve and Dencher, 1966) to 0.7 mg/ml (Kitchin and Bearn, 1965). Therefore, only about 1–5% of the Gc present in serum was recovered in isolation procedures. The Gc material used in this study was purified from approximately 12,000 ml of human plasma or serum.

This paper presents a detailed account of the peptide maps and amino acid composition of the purified bands of the group-specific protein and the results of end-group analyses. These data suggest that each electrophoretic Gc band is a dimer of closely similar, but not identical subunits. Structural investigations of Gc and serum albumin revealed some similarity in the carboxyl termini of these two proteins.

* From the University of Texas Medical Branch, Galveston, Texas. *Received March 20, 1969.* This work was supported in part by a grant from The National Foundation and in part by U. S. Public Service Grants HD-03321 and AM-01542.

† Inquiries are to be made to B. H. B., Department of Human Genetics, University of Texas Medical Branch.

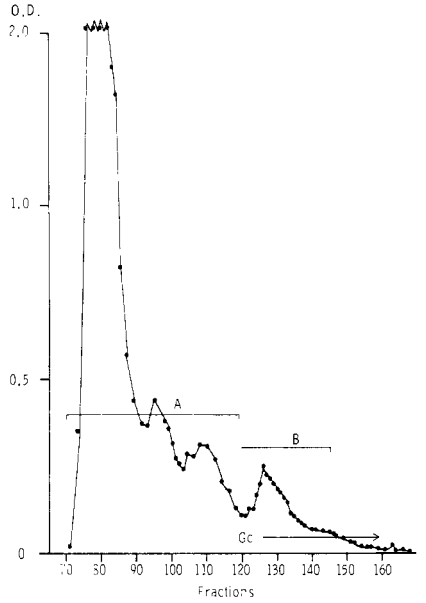


FIGURE 1: Gel filtration of the $\alpha_2$ protein from preparative starch block electrophoresis. Column (4 × 85 cm) of Sephadex G-100. "A" peaks comprised contaminating $\alpha_2$ proteins. Peak "B" contained Gc and a contaminating $\alpha_1$ protein. "B" was recycled into gel once more and collected.

## Materials and Methods

*Preparation of Gc.* Isolation of Gc 1-1, 2-1, and 2-2 from plasma or serum was carried out according to the following procedure. Plasma (or serum) from an individual of known phenotype was concentrated at 5° by ultrafiltration to one-half its original volume. The concentrated serum or plasma was dialyzed for 24 hr against the barbital buffer used in starch block electrophoresis (0.1 M, pH 8.6) and was centrifuged at 25,000 rpm for 30 min. An aliquot of 30 ml of the concentrated serum or plasma was pipetted into the origin slot of the preparative starch block. Large starch blocks (51 × 55 cm) were used for the first electrophoretic separation. After electrophoresis, the starch block was cut into segments 1 in. wide and the protein was eluted using a sintered glass filter (coarse) and suction. The Gc protein was identified in the aliquots by acrylamide gel electrophoresis (Kitchin, 1965); Gc 1-1 was found to extend from the cathodal edge of the albumin band to the anodal edge of the $\alpha_2$-globulin peak. Gc 2-2 was present in the $\alpha_2$-globulin peak. The filtrates containing the Gc protein from the starch block were pooled and their contents were concentrated by ultrafiltration to a volume of 3 ml.

After approximately 200–400 mg of Gc with other $\alpha_2$-contaminating proteins had been collected from two or three starch blocks, the aliquots were pooled, concentrated by ultrafiltration, and dialyzed against 0.1 M Tris-HCl–NaCl buffer (pH 8.0). This buffer contained 48.46 g of Tris, 240 ml of 1 M HCl, 9 g of NaCl, and was made up to a volume of 4 l. After dialysis for 16 hr, the protein solution was centrifuged for 30 min at 25,000 rpm. Gel filtration in G-100 (Sephadex) was performed with the Tris-HCl–NaCl buffer. The protein solution was pumped from the bottom to the top of a Sephadex column (4 × 85 cm) at a flow rate of approximately 30 ml/hr. After 110 ml of column effluent was collected, the subsequent effluent was recycled back into the bottom of the column. When 133 ml had recycled through the column, the effluent was collected in a fraction collector with 4.7 ml/fraction. The Gc was eluted between fractions 128 and 166 after the protein had recycled through the column (Figure 1). The exact location of the Gc fractions was determined by double gel diffusion on Ouchterlony plates with wells for 32 samples/plate. The antiserum located in the center well was produced in a rabbit that had been innoculated with a partially pure preparation of Gc. After the fractions containing Gc were pooled and concentrated by ultrafiltration so that the concentration was approximately 5 mg/ml, acrylamide gel electrophoresis of the Gc preparation was carried out (Figure 2). At this point, the Gc preparation usually contained an $\alpha_1$-protein contaminant. This was eliminated by preparative starch block or preparative acrylamide electrophoresis. When the purified fast or slow bands of Gc 1-1 and Gc 2-2 were desired, the preparation was dialyzed against the electrophoresis buffer (Peacock *et al.*, 1965) and submitted to preparative polyacrylamide electrophoresis (Simons and Bearn, 1967). Long starch block electrophoresis was often utilized for eliminating a slow $\alpha_2$ contaminant. This method was not as effective for separating the fast and slow Gc 1-1 bands, although it separated the Gc 2-2 bands when a 70-cm starch block was utilized. The purity of the final preparation was judged by the appearance of only the fast and slow bands characteristic of the Gc phenotypes on acrylamide gel electrophoresis (Figure 3). Gc preparations were then reduced and alkylated, aminoethylated, or used directly.

*Reduction of Gc.* Gc was reduced and alkylated following a modification of the method of Crestfield *et al.* (1963). To 6 mg of Gc 2-2 in 1.6 ml of 5 M guanidinium chloride the following were added: 0.1 ml of $Na_2EDTA$ solution (50 mg/ml), 0.75 ml of 1 M Tris buffer (p H 8.6), and 0.056 ml of mercaptoethanol. A small vial (4 ml) was filled to the brim with 5 M guanidinium chloride, 0.2% EDTA, and covered for 4 hr. After this, 288 mg of iodoacetamide in 1–2 ml of 5 M guanidine–0.25 M Tris (pH 8.5) was added. After standing at 5° in the dark for 30 min, the reduced and alkylated Gc was applied to a G-25 Sephadex column which had been equilibrated in 50% acetic acid.

Some preparations were aminoethylated as described by Rafferty and Cole (1966). The procedure above for reduction and alkylation was followed except 0.4 M ethylenimine was used in place of iodoacetamide.

In order to obtain subunits from the Gc dimer, as indicated by molecular weight determinations, it was necessary to expose Gc for a period of at least 4 hr to 5 M guanidinium chloride before reduction with mercaptoethanol.

*Acrylamide Electrophoresis.* Electrophoresis on polyacrylamide slabs was carried out according to the method of Kitchin (1965), utilizing the buffer of Peacock *et al.* (1965). The preparative acrylamide electrophoretic system of Simons and Bearn (1967) succeeded in separating the fast and slow bands in Gc 1-1 and Gc 2-2 preparations.

*Immunoelectrophoresis and Diffusion.* Immunoelectrophoresis was carried out in agar gel according to the microtechniques of Scheidegger as modified by Hirschfeld (1959). Specific anti-Gc serum was prepared in rabbits by immunization with purified Gc protein. The two-dimensional diffusion system on agar gel as developed by Ouchterlony (1949), was employed.

*Peptide Patterns.* Tryptic digestions of 5–20 mg of Gc preparations which had been either heat denatured, reduced

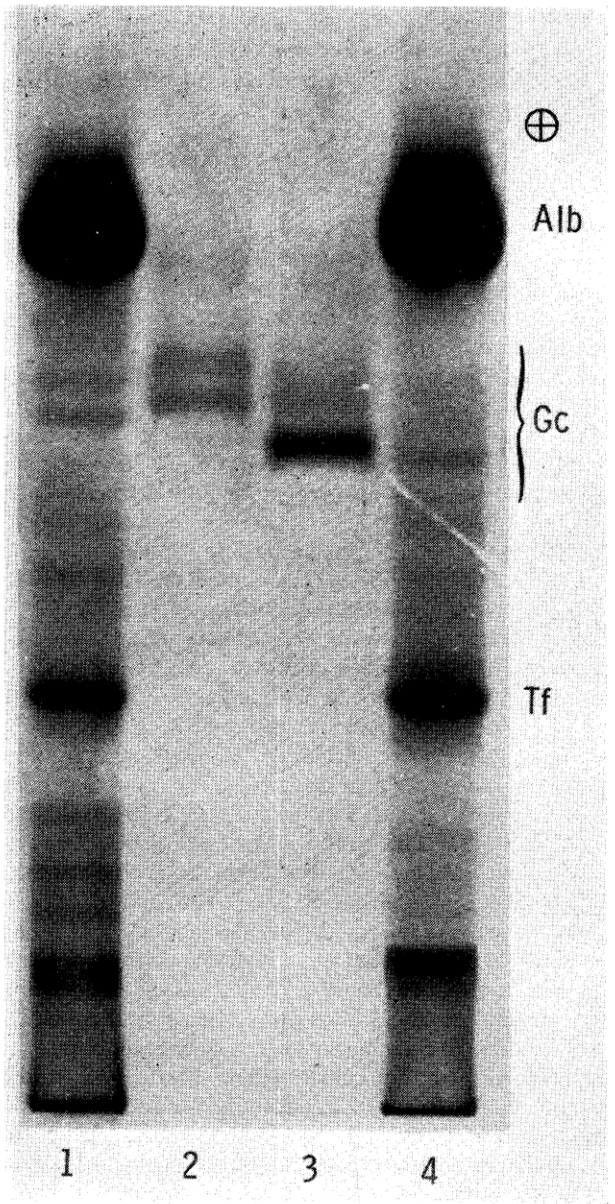


FIGURE 2: Gc preparations after gel filtration. Purified Gc 1-1 (position 2) and Gc 2-2 (position 3) are compared with serums from individuals having Gc 1-1 (position 1) and Gc 2-2 (position 4) phenotypes. Preparations of Gc at this stage contain faint traces of albumin and $\alpha_1$ contaminants. Gc bands migrate slightly faster in partially purified preparations than that in serum. The fast band of Gc 1-1 is characteristically more concentrated than the fast band of Gc 2-2.

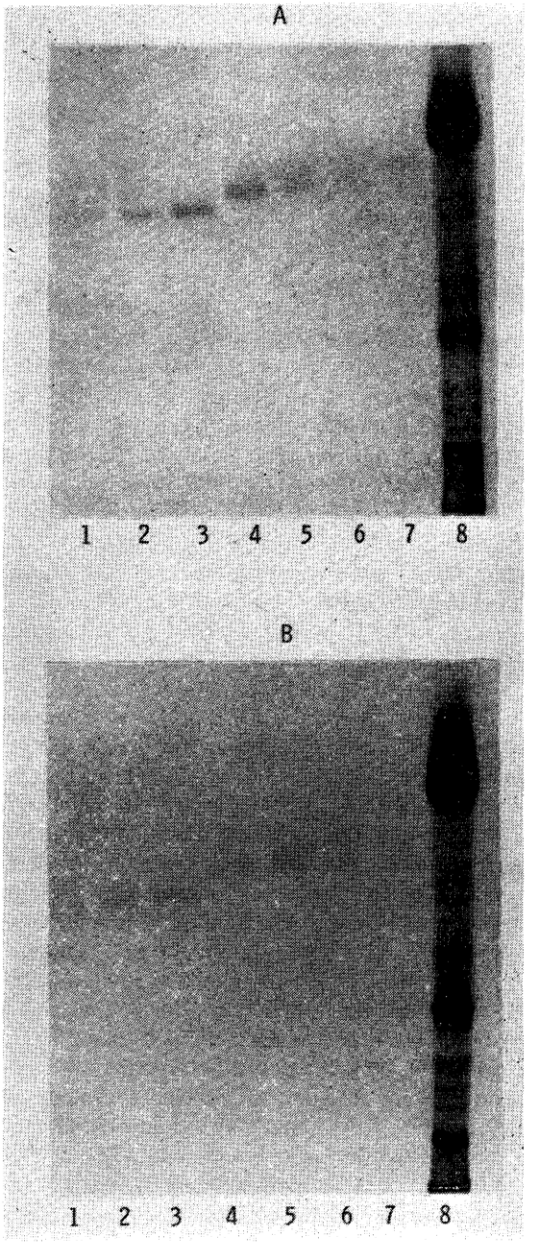


FIGURE 3: Acrylamide gel showing separation of the fast and slow bands of Gc 1-1 (A) and Gc 2-2 (B) by electrophoresis on a long starch block. Fractions A-2, 3 (Gc 1-1 slow) and A-4,5 (Gc 1-1 fast) were pooled separately. Fractions B-1-3 (Gc 2-2 slow) and B-4 (Gc 2-2 fast) were pooled separately.

alkylated, or aminoethylated were performed under a nitrogen barrier in a pH-Stat for 1.5–4 hr. The reaction was carried out at 38°, pH 8.0, and with a concentration of trypsin (Worthington, crystallized) which corresponded to 1 mg of trypsin/100 mg of protein. After tryptic hydrolysis was complete (4 hr for aminoethylated preparations and 2 hr for heat-denaturated or reduced-alkylated preparations), the tryptic peptide solution was centrifuged at 25,000 rpm for 1 hr. Aliquots were concentrated *in vacuo*. Concentrated digest (3–4 mg) was dissolved in 10–20 μl of electrophoresis buffer and applied to Whatman No. 3MM paper in a small spot. The dimensions of the paper have been described (Bowman and Bearn, 1965). High-voltage electrophoresis was carried out at pH 3.5 using a pyridine–acetic acid–water buffer (1:10:189, v/v, respectively) with a voltage of 47 V/cm for 1 hr. This step was performed in a Lucite tank, on a cooled plate, or between glass plates (with lower voltage and longer duration). Ascending chromatography using pyridine–isoamyl alcohol–water solvent (35:34:30, v/v, respectively) was performed after the paper had been dried in an oven at 90° for 10 min.

The peptide patterns were developed by dipping the paper in a cadmium–ninhydrin solution (Heilmann *et al.*, 1957) consisting of cadmium acetate (100 mg), water (10 ml), acetic acid

(5 ml), acetone (100 ml), and ninhydrin (1 g). After remaining in a fume hood for 5 min, the paper was placed in a 90° oven for 5 min. Peptide spots were pink, yellow, and orange, depending upon their amino acid composition.

After one-dimensional electrophoresis, the paper was dried and cut into four to six horizontal strips, each of which was subjected to a different specific stain. On a single one-dimensional separation, it was possible to identify peptides having tyrosine, arginine, tryptophan, histidine, and divalent sulfur.

*Disulfide analyses* were carried out by the procedure developed by Brown and Hartley (1963). A trypsin digest was submitted to high-voltage ionophoresis as a band on Whatman No. 3MM paper. After paper electrophoresis at pH 3.5, a guide strip was stained with ninhydrin and the remaining strip was placed in an atmosphere of performic acid overnight. After the paper had been dried *in vacuo* to remove the acid, it was sewed vertically along a new 25 × 110 cm paper and submitted to electrophoresis at right angles to the original direction. Unaltered peptides would be arranged in a 45° diagonal line, while altered peptides arising from cleavage of a disulfide bond would appear as two new spots off the diagonal.

*Amino Acid Analysis.* Analysis of the protein hydrolysates was carried out by the technique of Spackman *et al.* (1958) using high-sensitivity modifications (Bowman *et al.*, 1967). Some Gc preparations were reduced-alkylated before analysis.

*N-Terminal Analysis.* Qualitative end-group analysis was performed by the dansyl[1] method of Gray (1967), except that the dansylamino acids were separated on polyamide thin layers (Woods and Wang, 1967). The fluorodinitrobenzene method for amino end-group analysis (Frankel-Conrat *et al.*, 1955) and the leucine aminopeptidase reaction described by Hill and Smith (1957) were also used for N-terminal detection.

*C-Terminal Analysis.* To 10–12 mg of Gc was added 0.15 ml of 0.01 M NaOH and 2.25 ml of 0.05 N sodium borate (pH 8.5). This solution was heated at 65° for 30 min and cooled. Aliquots were removed for controls, and then 10 μl of carboxypeptidase A (Worthington Biochemical Corp.) in a concentration of 0.5 mg/ml was added to the protein solution. Aliquots of the enzyme were also removed for controls. Aliquots of the protein–enzyme were removed at 5, 10, 30, 60, and 120 min and 4-hr intervals at room temperature and the reaction was stopped by the addition of glacial acetic acid or by freezing. The aliquots were analyzed on the automatic amino acid analyzer and the amino acid values from intact protein and enzyme controls were subtracted from the values derived from the action of carboxypeptidase A on Gc during the six intervals. Because the amino acid analysis procedure used here does not separate glutamine, asparagine, and serine, samples of this peak eluted from the analytical columns were hydrolyzed with HCl in the usual manner. This hydrolysate was then rechromatographed and analyzed.

Purified serum albumin was analyzed in the same manner.

*Molecular Weight Studies.* A 130 × 2.4 cm chromatography column was calibrated as previously described (Bowman and Bearn, 1965) using Sephadex G-100, or G-75, in 0.1 M Tris or 50% acetic acid. Fractions of 5.8 ml were collected and monitored at 280 mμ.

*Electrophoresis on Urea, Acidic, or Alkaline Gels.* Reduced and alkylated Gc 1-1 was subjected to starch gel electrophoresis at pH 3.5 and 2.1, with and without the presence of 0.2 M mercaptoethanol and 8 M urea (Smithies, 1959). Aminoethylated Gc 2-2 and Gc 1-1 were also subjected to starch gel electrophoresis at pH 2.1, 3.5, and 9.0.

[1] Abbreviation used is: dansyl, 1-dimethylaminonaphthalene-5-sulfonyl chloride.

## Results

*Gc Purification.* Preparative starch block electrophoresis of human serum or plasma succeeded in separating Gc from serum albumin, the β- and γ-globulins, although it contained other $\alpha_2$ proteins. After preparative electrophoresis the Gc fractions were subjected to gel filtration and most of the $\alpha_2$ contaminants were eliminated. The elution curve on Sephadex G-100 is shown in Figure 1. Peak A contained the $\alpha_2$ proteins present while Gc was eluted in peak B which was recycled through the column. Figure 2 demonstrated the electrophoretic separation of the Gc preparation after gel filtration. The fast and slow Gc bands are characteristic of the Gc region in whole serum although partially purified Gc migrates slightly faster than that in serum. The fast band of Gc 1-1 is characteristically more concentrated than the fast band of Gc 2-2. The final step, prolonged electrophoresis on long starch blocks or separation or preparative acrylamide gel eliminated the remaining $\alpha_1$ contaminants in addition to separating the fast and slow bands characteristic of the Gc phenotype. The sequential elutions from prolonged electrophoresis on starch blocks are shown in Figure 3. Fractions A4 and -5 were pooled for Gc 1-1 fast, A2 and -3 for slow. Fraction B-1-3 (slow) and B-4 (fast) contained the bands characteristic for Gc 2-2.

*Amino Acid Composition.* The amino acid composition of the four electrophoretic bands characteristic of the Gc phenotypes is given in Table I. The calculated number of residues per dimer of Gc is based on a molecular weight of 51,000. Half-cystine in Gc was determined in some reduced and alkylated preparations as carboxymethylcysteine. In these analyses there were 19–20 *S*-carboxymethylcysteine residues/molecule. Tryptophan was determined by the reaction of Ehrlich's reagent on fingerprints; 2 tryptophan-containing peptides were found in fingerprints of Gc 1-1, 2-1, and 2-2. In summary, this serum protein is characterized by an usually high aspartic and glutamic acid content, about 110 residues of acidic amino acids, and a relatively low histidine content of some 7–9 residues/molecule. The nonintegral values of some of the amino acids may reflect the existence of similar but nonidentical subunits in the Gc band characterized.

The amino acid composition of the Gc types in Table I are closely similar to the analyses reported by Heimburger *et al.* (1964), Cleve *et al.* (1963), Bowman and Bearn (1965), and Simons and Bearn (1967), except for the number of tryptophan residues and composition of the fast band of Gc 2-2 which have not been reported previously.

*Peptide Mapping.* If native Gc, of molecular weight approximately 51,000, consisted of a single chain or extensively unlike chains, peptide mapping of reduced-aminoethylated protein should reveal approximately 72 ninhydrin-positive spots, if all lysine, arginine, and aminoethylcysteine bonds were hydrolyzed by trypsin. However, fingerprints of the aminoethylated slow band of Gc 1-1 (Figures 4 and 5) contained about 30 ninhydrin-positive spots, roughly one-half the number expected. Approximately 8 of these peptides stain positively for arginine, close to one-half the predicted value from amino acid analyses. Seven tryptic peptides contained histidine

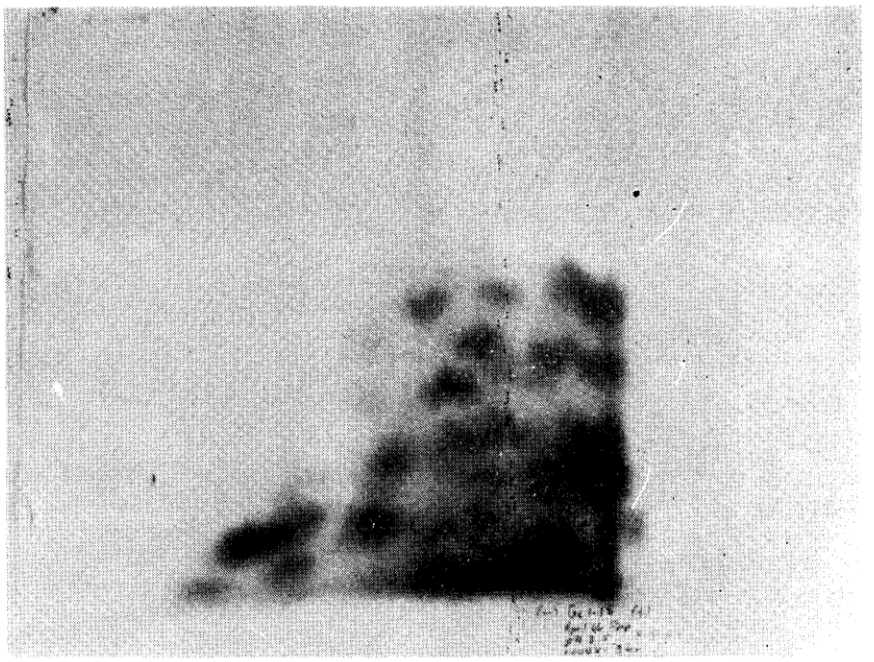

FIGURE 4: Fingerprint of purified slow band of Gc 1-1 which had been reduced aminoethylated; 47 V/cm for 1 hr at pH 3.5 in pyridine–acetic acid–water (1:10:189, v/v). Chromatography in pyridine–isoamyl alcohol–water solvent (35:34:30, v/v).

TABLE I: Amino Acid Analysis (Based on Mol Wt 51,000).

| | Gc 1-1 | | Gc 2-2 | |
|---|---|---|---|---|
| Amino Acid | Fast Band | Slow Band | Fast Band | Slow Band |
| Lysine | 36.8 ± 0.3 | 38.8 ± 0.5 | 35.2 ± 0.5 | 35.7 ± 0.9 |
| Histidine | 6.8 ± 0.2 | 6.9 ± 0.2 | 8.0 ± 0.2 | 9.0 ± 0.3 |
| Arginine | 15.6 ± 1.0 | 14.2 ± 0.7 | 15.7 ± 0.1 | 15.6 ± 0.7 |
| Aspartic acid | 48.8 ± 0.4 | 48.5 ± 0.5 | 46.5 ± 0.3 | 44.9 |
| Threonine | 26.7 ± 0.4 | 26.1 ± 0.1 | 27.2 ± 0.2 | 24.9 ± 0.6 |
| Serine | 35.5 ± 0.8 | 36.1 ± 0.7 | 36.1 ± 2.1 | 38.0 ± 0.1 |
| Glutamic acid | 59.0 ± 0.7 | 60.5 ± 0.1 | 63.3 ± 1.0 | 59.3 ± 0.4 |
| Proline | 28.2 ± 0.8 | 28.0 ± 1.0 | 27.4 ± 3.1 | 29.3 ± 0.9 |
| Glycine | 20.1 ± 0.2 | 21.5 ± 0.7 | 22.9 ± 0.2 | 22.8 ± 0.1 |
| Alanine | 31.9 ± 0.7 | 33.0 ± 0.7 | 34.4 ± 0.1 | 33.0 ± 0.2 |
| Half-cystine | 18.9 ± 1.1 | 19.0 ± 1.0 | 19.0 ± 1.5 | 19.0 ± 1.0 |
| Valine | 23.7 ± 0.2 | 22.1 ± 0.1 | 23.3 ± 0.1 | 25.3 ± 1.6 |
| Methionine | 5.9 ± 1.0 | 4.5 ± 0.4 | 4.1 ± 0.9 | 4.5 ± 0.5 |
| Isoleucine | 9.2 ± 0.1 | 9.8 ± 0.2 | 9.3 ± 0.2 | 10.2 ± 0.4 |
| Leucine | 44.7 ± 0.1 | 45.9 ± 0.4 | 47.3 ± 0.9 | 43.9 ± 0.8 |
| Tyrosine | 17.7 ± 1.0 | 17.4 ± 2.6 | 19.2 ± 0.2 | 17.2 ± 0.2 |
| Phenylalanine | 18.4 ± 0.3 | 18.1 ± 0.3 | 19.4 ± 0.4 | 18.7 ± 1.1 |

as revealed by Pauly's reagent. Tyrosine was present in nine peptides and tryptophan in two.

These results are in agreement with those found in previous works; one-half the expected number of tryptic peptides were observed in fingerprints of Gc 2-1 preparations which had been denatured with heat before tryptic hydrolysis (Bowman and Bearn, 1965) and in Gc 1-1 fingerprints on thin-layer plates (Ruoslahti, 1967). Fingerprints of single Gc bands have not been previously reported.

The divalent sulfur stain detected peptides containing aminoethylcysteine and methionine. The former appeared white against a pink background and appeared first; methionine-containing peptides were faint yellow with this stain and developed later. Ten peptides contained aminoethylcysteine and three appeared to contain methionine. It is unknown whether the dark staining peptide which contains divalent sulfur and migrates half-way up the chromatogram in the neutral band represents unreduced material or whether it is a long peptide which stains intensely with cadmium ninhydrin.

After preparations of approximately 5 mg of each the fast and slow electrophoretic bands of Gc 1-1 had been purified and reduced aminoethylated, they were digested with trypsin and the resulting ionograms were contrasted (Figure 6). Striking similarity was observed in the ninhydrin-stained pep-

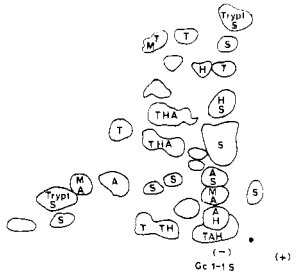

FIGURE 5: Schematic diagram of fingerprint in Figure 4 showing peptides containing specific amino acids. A = arginine, S = aminoethylcysteine, M = methionine, T = tyrosine, Trypt = tryptophan, and H = histidine.

tides. There was, however, in the fast Gc band digest a histidine-containing peptide, no. 2, which migrated slowly toward the cathode at pH 3.5 and which was absent in the ionogram of the Gc slow band. Furthermore, peptide 11, a histidine-containing peptide in the Gc slow band, migrated more rapidly toward the cathode and was absent in the ionogram of the Gc fast band. This indicated that the fast and slow bands of Gc 1-1 have extremely similar tryptic peptides but differ in at least one of them by a net charge difference.

The amino acid analysis of the fast and slow bands of Gc 1-1 point out the extensive similarity shared by the two electrophoretic fractions; peptide patterns of tryptic digests demonstrate that there is at least one peptide difference between the two Gc bands.

*Disulfide Reduction.* A preparation of Gc 1-1 which had been denatured by heat was digested with trypsin and subsequently subjected to electrophoresis. The ionogram was exposed to performic acid vapors and the strip was sewed onto a new piece of filter paper and subjected to electrophoresis at right angles. Figure 7 is a photograph of the pattern after development with ninhydrin. Ten or eleven peptides appear off the diagonal line formed by the unaltered peptides, indicating that five disulfides have been oxidized by performic acid. The same results were obtained from preparations of Gc 2-2 and Gc 2-1. Disulfide reduction was carried out on albumin in the same manner. A total of 33 peptides migrated off the diagonal line formed by the peptides unaltered by performic

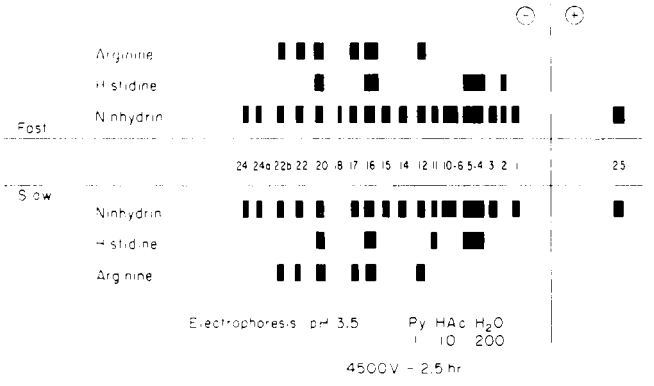


FIGURE 6: Comparison of the tryptic peptides of the fast and slow components of Gc 1-1. One histidine-containing peptide characteristic of fast (no. 2) was missing in slow; however, slow contained a histidine-containing peptide (no. 11) which was missing in fast.

TABLE II: Amino Acids Releasedfr om Group-Specific Protein by Treatment with Carboxypeptidase A.

| Experiment | Incubn Time | Amt of Amino Acid Released (equiv)[b] Ile | Leu | Ser | Asn | Val | Ala |
|---|---|---|---|---|---|---|---|
| 1. Gc 1-1 | 60 min | 1.4 | 1.2 | 1.0 | 0.3 | 0.3 | 0.3 |
| | 240 min | 1.5 | 1.6 | 1.2 | 0.3 | 0.3 | 0.3 |
| 3. Gc 1-1 | 180 min | 0.9 | 1.0 | 0.8 | 0.7 | 0.6 | 0.4 |
| 3. AE[a] Gc 2-1 | 24 hr | 1.0 | 1.2 | 0.8 | 0.3 | 0.3 | 0.3 |
| 4. Gc 2-2 | 60 min | 1.4 | 1.2 | 0.5 | 0.3 | 0.2 | 0.2 |
| 5. Slow band Gc 1-1 | 60 min | 0.7 | 1.0 | 0.3 | 0.2 | 0.3 | 0.2 |

[a] AE, aminoethylated after reduction with mercaptoethanol. [b] Equivalents based on Gc having mol wt 51,000.

acid oxidation indicating the presence of some 16 or 17 disulfide bonds in intact serum albumin.

*C-Terminal Analysis.* The action of carboxypeptidase A on the Gc protein liberates leucine and isoleucine at the same approximate rate (Figure 8); in three experiments leucine was liberated slightly faster than isoleucine and in two experiments isoleucine was liberated faster than leucine (Table II). The concentrations of isoleucine and leucine, however, were in all cases within $\pm 0.3$ equiv of each other. The same C-terminal sequence was found in preparations of Gc 1-1, Gc 2-2, Gc 2-1, and in reduced-aminoethylated Gc. In experiments using mixtures of carboxypeptidases A and B the results were the same as using carboxypeptidase A alone. Insufficient amounts of the purified fast and slow bands of Gc 1-1 and 2-2 were available to perform C-terminal analyses of all four separate bands; however, carboxypeptidase A digestion of 3.2 mg of purified slow band from Gc 1-1 liberated the same amino acids, leucine and isoleucine in the same relative proportions (Table II).

In all experiments approximately 1 mole of each leucine and isoleucine per 51,000 molecular weight of protein dimer was released. Previous analytical ultracentrifugation of Gc in urea and mercaptoethanol indicated the subunit molecular weight to be 20–30 $\times$ $10^3$ (Bowman and Bearn, 1965). These data indicate that 1 mole of C-terminal isoleucine or leucine is present in 1 mole of Gc monomer. The two C-terminal amino acids are followed by roughly equal amounts of serine and asparagine which may be present in the penultimate positions in the two polypeptide chains. Valine and alanine may each represent the third amino acid in the C-terminal sequence on each chain. However, extensive sequence analysis will be necessary to confirm these data.

The action of carboxypeptidase A upon human serum albumin indicated the following C-terminal sequence: Val-Ala-Leu-COOH. These results agreed with those reported by White *et al.* (1955). All three amino acids in the albumin C terminus are present in the carboxyl sequence of the Gc subunits. Like albumin, one subunit has C-terminal leucine. The possible sequence of the terminal three amino acids in the Gc subunits as suggested by their release by carboxypeptidase A are listed with the albumin C-terminal sequence in Figure 9.

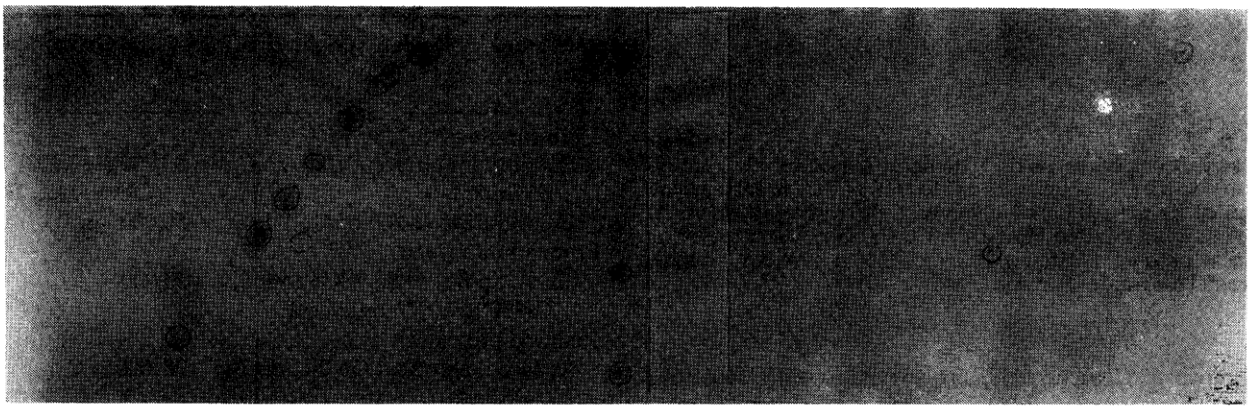

FIGURE 7: Disulfide reduction. Ten peptides appear off the diagonal line formed by the unaltered peptides, indicating that five disulfides have been oxidized by performic acid.

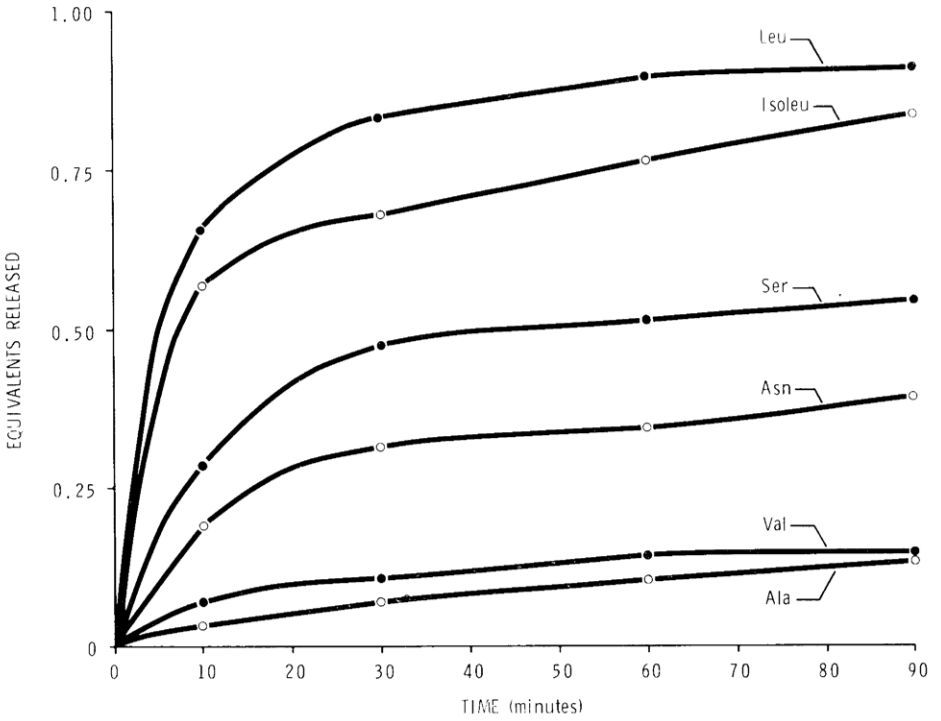


FIGURE 8: Release of amino acids by action of carboxypeptidase A. Details of procedure are in text. The values of amino acids equivalents per gram molecular weight of Gc have been corrected for free amino acids present in the carboxypeptidase solution and present in the intact Gc preparation, without enzyme. A molecular weight of 51,000 for Gc has been used in the calculation.

The first and third amino acids are identical or could be explained as the result of a single base alteration in the nucleotide codon specifying the amino acid in both Gc subunits and albumin. Asparagine appears to be in the penultimate position in one of the Gc subunits; however, it is unrelated to serine or alanine which is present in the other Gc subunit and albumin, respectively.

*N-Terminal Analysis.* The amino-terminal sequence of the group-specific protein was investigated in an attempt to determine the number of N-terminal residues in the different phenotypes. Quantitative dinitrophenylation and dansyl experiments carried out under a wide variety of conditions yielded only submolar quantities of amino-terminal residues. The conditions used included those which lead to the dissociation of Gc into subunits and those which cause extensive unfolding of the polypeptide chains. These results, combined with the lack of reactivity of Gc to leucine aminopeptidase, led to the conclusion that the amino-terminal residues of Gc are not readily available for reaction with amino-terminal labeling reagents and are possibly acetylated.

*Molecular Weight Studies.* Native Gc is eluted on a calibrated G-100 column in elution volume corresponding to mol wt 51,000, which agrees well with previous equilibrium ultracentrifugation (Table III). Furthermore, when Gc is dissolved in 8 M urea and subjected to gel filtration on a calibrated column in 50% acetic acid, it is eluted in a volume corresponding to the same mol wt 51,000. However, when reduced and alkylated Gc was eluted on a calibrated G-75

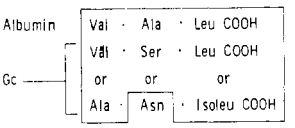


FIGURE 9: Comparison of the C-terminal sequences of serum albumin and the Gc subunits. The C termini of the Gc subunits are suggested by the release of amino acids by carboxypeptidase A. Those residues in albumin and the Gc subunits which are either identical or have nucleotide codons alike in two out of three sequential bases are enclosed in a box.

column in 50% acetic acid, the subunits were eluted in a volume corresponding to mol wt 28,000, close to the value obtained from analytical centrifugation of reduced Gc in previous studies.

*Electrophoretic Separation of Gc Subunits.* It was not possible to separate the 25,000 molecular weight Gc subunits by the electrophoretic methods employed. With 8 M urea and 0.2 M mercaptoethanol incorporated into acidic gels run at pH 3.5 or 2.1 the Gc types separated in all buffers into the fast and slow bands characteristic of Gc 1-1, 2-2, and 2-1 observed in conventional gel electrophoresis. No further division into subunits could be detected. The band pattern of reduced carboxymethylated Gc was also unaffected by preliminary treatment with 8 M urea or 8 M urea and 0.1 M $\beta$-mercaptoethanol when such preparations were subjected to electrophoresis in pH 9.0 or 2.1. All attempts to separate the monomers in Gc 1-1, 2-1, and 2-2 types and subsequently obtain distinct subunit bands failed. The reason for this is unknown; possibly reassociation of the subunits or aggregation of polypeptide chains may have prevented the monomers from migrating separately.

## Discussion

Evidence of heterogeneity of the group-specific protein in the serum of homozygous individuals led to the conclusion that Gc was comprised of similar subunits (Bowman and Bearn, 1965). The data from this study indicate that the subunits of the Gc dimer are composed of polypeptide chains terminating in leucine and isoleucine. All three genotypes, Gc 1-1, 2-1, and 2-2 terminate in their carboxyl ends with equal concentrations of leucine and isoleucine. This would not be the case if, for example, the fast electrophoretic band of Gc 1-1 was comprised of identical subunits $\alpha^1{}_2$ and the slow

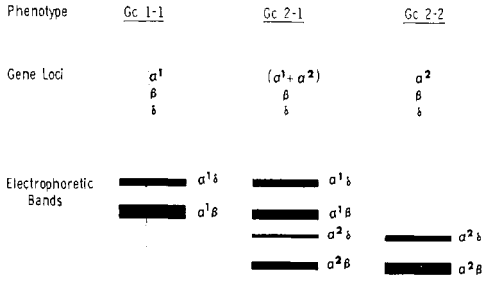


FIGURE 10: Proposed genetic system controlling the heterogeneity of the Gc electrophoretic bands. Three distinct genetic loci are proposed to control the three common Gc phenotypes, Gc 1-1, Gc 2-2, and Gc 2-1.

TABLE III: Molecular Weight Studies of the Group-Specific Protein.

| | |
|---|---|
| Native protein | |
| Equilibrium ultracentrifugation of Gc 2-2 | 50,800 (±2,900) (Cleve *et al.*, 1963) |
| Gel filtration G-100 of Gc 1-1 | 50,000 |
| Protein in 8 M urea | |
| Gc 1-1 in 8 M urea, gel filtration, G-100 in 50% acetic acid | 50,000 |
| Reduced protein | |
| Reduced, alkylated Gc 2-2, on G-75 in 50% acetic acid | 28,000 |
| Gc 1-1 denatured with 6 M guanidinium chloride, centrifuged in 8 M urea–0.1 M mercaptoethanol | 25,000 ± 5,000 (Bowman and Bearn, 1965) |

electrophoretic band was made up of unlike subunits, $\alpha^1\beta$. However, if the slow bands of Gc 1-1 and Gc 2-2 were comprised of $\alpha^1\beta$ and $\alpha^2\beta$ polypeptides, respectively, and if the fast bands of 1-1 and 2-2 were $\alpha^1\delta$ and $\alpha^2\delta$, respectively, equal concentrations of isoleucine and leucine would be detected in C-terminal analysis provided that $\alpha^1$ and $\alpha^2$ contained C-terminal leucine and $\beta$ and $\delta$ each contained C-terminal isoleucine, or *vice versa*. Furthermore, Gc 2-1 which is comprised of four bands corresponding to the fast and slow bands characteristic of Gc 1-1 and Gc 2-2 also has equal amounts of leucine and isoleucine in its C-terminal sequence, as would be expected if the fast bands were $\alpha\delta$ and the slow bands $\alpha\beta$.

It would be expected, therefore, that C-terminal analysis of any of the purified fast or slow Gc bands would yield leucine and isoleucine in approximately equal concentrations. This, in fact, was found when the purified slow band of Gc 1-1 was digested with carboxypeptidases A and B; isoleucine and leucine were released in concentrations of 0.7 and 1.0 equiv, respectively (Table II). Additional direct evidence in support of each of the remaining three bands being comprised of two similar but not identical subunits must await carboxypeptidase digestion of the pure fast and slow bands of Gc 2-2 and the pure fast band of Gc 1-1, which were present in insufficient amounts to analyze in this study.

Several other lines of evidence suggest that Gc is a dimer of very similar, but distinct subunits. The number of amino acids in the composition of Gc types (Table I) compared with the number of tryptic peptides containing the same specific amino acids (Figure 5) is approximately one-half, indicating that the Gc subunits are extensively alike in their sequence. The ionograms of tryptic digests of purified fast and slow bands from the Gc 1-1 phenotype are strikingly similar (Figure 6). However, one histidine-containing peptide differs, migrating faster at pH 3.5 toward the anode in the fast band preparations than in the ionogram of the slow band. This indicates that although these proteins are alike, there is at least one peptide difference between them. The Gc system has, therefore, a structural similarity to the subunit structure of human hemoglobin in which hemoglobins $A_1$, $A_2$, and F are formulated by $\alpha_2\beta_2$, $\alpha_2\delta_2$, and $\alpha_2\gamma_2$, respectively. In both the

hemoglobin and Gc system, one of the polypeptide chains is synthesized in lower concentrations than the other subunits. In the hemoglobin system the concentration of $HbA_2$ is limited by the diminished synthesis of $\delta$ chains, whereas the fast electrophoretic bands of Gc 1-1 and Gc 2-2 are characteristically present in diminished concentrations, less than 30% of the total Gc protein. If the polypeptide chain unique to the fast bands of Gc 1-1 and 2-2 is designated as $\delta$ (following the hemoglobin nomenclature) and the polypeptide chain shared by the fast and slow bands as $\alpha$, the Gc system could be attributed to three genetic loci, with alternate alleles $\alpha^1$ and $\alpha^2$ in one locus and two other loci, $\beta$ and $\delta$ (Figure 10).

The $\delta$ locus would be responsible for the synthesis of $\delta$ chains which dimerize with $\alpha^1$ or $\alpha^2$ chains to comprise the fast, minor bands of Gc 1-1, 2-2, and 2-1. Whereas the $\beta$ locus would produce polypeptide chains capable of dimerizing with the $\alpha^1$ or $\alpha^2$ chains to form the slower, major electrophoretic bands of Gc 1-1, 2-2, and 2-1.

Of the five disulfide bonds in Gc, one or more interchain bonds is involved in the dimerization of the subunits, since both mercaptoethanol and urea are necessary for the reduction from mol wt 51,000–25,000 (Table III). Urea alone, on the other hand, does not reduce the molecular weight, indicating that disulfide bonds are important in holding the Gc subunits together.

Because peptide maps and amino acid analyses demonstrate the structural similarity between $\alpha^1\delta$ and $\alpha^1\beta$ (the fast and slow bands of Gc 1-1) and between $\alpha^2\delta + \alpha^2\beta$ (Gc 2-2), it might be postulated that genetic duplication was the evolutionary mechanism responsible for the Gc system as observed in human serum. Furthermore, the faint similarity seen in the C-terminal sequence of the Gc subunits and serum albumin, in addition to the evidence for close genetic linkage of the structural genes for Gc and albumin (Weitkamp *et al.*, 1966) may reflect the presence in evolution of a common ancestor gene.

## Acknowledgment

Gratitude for excellent technical assistance is expressed to Katherine Wilkerson, Rodney Hite, and Don Barnett, and for a generous gift of anti-Gc antiserum to Doctor F. David Kitchin.

## References

Azen, E. A., Smithies, O., and Hiller, O. (1969), *Biochem. Genet.* (in press).

Bearn, A. G., Kitchin, F. D. and Bowman, Barbara H. (1964), *J. Exptl. Med. 120*, 83.

Bowman, B. H., Barnett, D. R., and Hite, R. (1967), *Biochem. Biophys. Res. Commun. 26*, 466.

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

Brown, J. R., and Hartley, B. S. (1963), *Biochem. J. 89*, 59P.

Cleve, H., and Bearn, A. G. (1962), *Progr. Med. Genet. 2*, 64.

Cleve, H., and Dencher, H. (1966), *Protides Biol. Fluids Proc. 14*, 379.

Cleve, H., Prunier, J. H., and Bearn, A. G. (1963), *J. Exptl. Med. 118*, 711.

Crestfield, A. M., Moore, S., and Stein, W. H. (1963), *J. Biol. Chem. 238*, 622.

Fraenkel-Conrat, H., Harris, J. J., and Levy, A. L. (1955), *Methods Biochem. Anal. 2*, 360.

Gray, W. R. (1967), *Methods Enzymol. 11*, 139, 469.

Heilmann, J., Barollier, J., and Watzke, E. (1957), *Z. Physiol. Chem. 309*, 219.

Heimburger, N., Heide, K., Haupt, H., and Schultze, H. E. (1964), *Clin. Chim. Acta 10*, 293.

Hill, R. L., and Smith, E. L. (1957), *J. Biol. Chem. 228*, 577.

Hirschfeld, J. (1959), *Acta Pathol. Microbiol. Scand. 47*, 160.

Kitchin, F. D. (1965), *Proc. Soc. Exptl. Biol. Med. 119*, 1153.

Kitchin, F. D., and Bearn, A. G. (1965), *Proc. Soc. Exptl. Biol. Med. 118*, 304.

Ouchterlony, O. (1949), *Acta Pathol. Microbiol. Scand. 26*, 507.

Peacock, A. C., Bunting, S. L., and Queen K. G. (1965), *Science 147*, 1451.

Rafferty, M. A., and Cole, R. D. (1966), *J. Biol. Chem. 241*, 3457.

Ruoslahti, E. (1967), *Ann. Med. Exptl. Fenn. 45*, 1.

Simons, K., and Bearn, A. G. (1967), *Biochim. Biophys. Acta 133*, 499.

Smithies, O. (1959), *Biochem. J. 71*, 585.

Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30*, 411.

Weitkamp, L. R., Rucknagel, D. F., and Gershowitz, H. (1966), *Am. J. Human Genet. 18*, 559.

White, W. F., Shields, J., and Robbins, K. C. (1955), *J. Am. Chem. Soc. 77*, 1267.

Woods, K. R., and Wang, K.-T. (1967), *Biochim. Biophys. Acta 133*, 369.