Primary Structure of Human Erythrocyte Glycophorin A. Isolation and Characterization of Peptides and Complete Amino Acid Sequence†

Motowo Tomita,‡ Heinz Furthmayr,* and Vincent T. Marchesi

ABSTRACT: Peptides of glycophorin A^{MN} were prepared by cyanogen bromide cleavage and by chymotryptic and tryptic digestion. Cyanogen bromide cleavage produces three fragments which account for the entire polypeptide chain. Trypsin and chymotrypsin cleave completely at several sites, but incompletely at sites within the glycosylated segment of the polypeptide chain. Some of the latter sites become accessible to proteolysis after desialation in addition to exposure of new sites for cleavage. The amino acid sequence of glycophorin A^{MN} has been determined by manual Edman degradation, using both the direct Edman and the dansyl-Edman procedures simultaneously for determination of glycosylated amino acid residues. The automated procedure was used for sequence determination of a hydrophobic peptide. Glycophorin A is a polypeptide chain of 131 amino acid residues and contains 16 oligosaccharide units attached to the amino-terminal third of the molecule. Fifteen oligosaccharides are linked O-glycosidically to either threonine or serine residues and one complex oligosaccharide unit is attached N-glycosidically to an asparagine residue. Amino-terminal sequences are different for glycophorin A^M and A^N, the two forms of the glycophorin A molecule coded for by genes at the MN locus. The differences in sensitivity to proteases of various sites on glycophorin A seem to be due to heterogeneity in the carbohydrate components and not to differences in the primary structure of the polypeptide chains. This work contains a number of revisions and corrections of earlier preliminary reports [Segrest, J. P., Jackson, R. L., Marchesi, V. T., Guyer, R. B., & Terry, W. (1972) *Biochem. Biophys. Res. Commun. 49,* 964–969; Tomita, M., & Marchesi, V. T. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72,* 2964–2968].

The major sialoglycoprotein fraction of the human red cell membrane (Winzler, 1969; Marchesi et al., 1972) has been extracted from membranes by a number of different procedures (Springer et al., 1969; Marchesi & Andrews, 1971; Blumenfeld & Zvilichovsky, 1972; Hamaguchi & Cleve, 1972; Fukuda & Osawa, 1973). The isolated product contains a number of sialoglycoproteins, some of which carry MNSs[1] blood group and influenza virus binding activities. The properties of these sialoglycoproteins are of great interest to investigators concerned with the molecular organization of cell membranes since these molecules include glycophorin A, one of the principal intrinsic transmembrane proteins of human red cells (Bretscher, 1971; Furthmayr et al., 1975).

Our interest in determining the primary structure of glycophorin A was based in part on the fact that it is a membrane glycoprotein which is readily available and easily purified, and which seemed to have all the characteristics of a transmembrane integral protein. An understanding of its structure has already given us some insight into the ways in which such molecules are arranged in membranes, how they may interact with membrane lipids, and how they may contribute to general membrane functions (Bretscher & Raff, 1975; Marchesi et al., 1976).

Our initial experiments were carried out on sialoglycoprotein preparations extracted from human erythrocyte membranes by the LIS–phenol procedure (Marchesi & Andrews, 1971). During the course of this work we realized that the material designated glycophorin (Marchesi et al., 1972) was composed of at least three different sialoglycoproteins which copurified together and which shared many common properties. A new isolation procedure has been developed for the isolation of the predominant sialoglycoprotein (designated glycophorin A) representing approximately 75–80% of the total glycoprotein fraction and two other sialoglycopeptides, labeled glycophorin B and glycophorin C (Furthmayr et al., 1975; Furthmayr, 1978a). Both glycophorins B and C can be distinguished from the A chain on the basis of their elution pattern on gel filtration columns, their migration on $NaDodSO_4$ gels and their lack of antigenic activity toward antibodies directed to a determinant on the carboxy-terminal region of glycophorin A (Cotmore et al., 1977). Both proteins are different from glycophorin A by amino acid composition and tryptic peptide maps and at least the major amino-terminal glycopeptides have been isolated (Furthmayr, 1978a,b). It was further demonstrated that glycophorin A exists in the membranes of heterozygous individuals in two forms, A^M and A^N. Genes at the MN locus determine a structural difference between these two forms in two positions of the amino acid sequence and this is correlated with M or N blood group activity (Furthmayr, 1978b).

† From the Department of Pathology, Yale University School of Medicine, New Haven Connecticut 06510. *Received September 20, 1977; revised manuscript received May 24, 1978.* This work was supported by U.S. Public Health Service and Membrane Center Grant P17-GM2174 to V.T.M. and grants from the Anna Fuller Fund, American Cancer Society BC-102C, and U.S. Public Health Service GMAM 21714-01C09 to H.F.

‡ Present address: Faculty of Pharmaceutical Sciences, Showa University, Shinagawa-ku, Tokyo, Japan.

[1] Abbreviations used: $NaDodSO_4$, sodium dodecyl sulfate; AMLO, ammonyx-LO (*N,N*-dimethyllaurylamine *N*-oxide); LIS, lithium diiodosalicylate; CB, T, CH, and TH, followed by a number designate cyanogen bromide, tryptic, chymotryptic, and thermolysin derived peptide; d, designates desialation of a peptide before proteolytic treatment; CNBr, cyanogen bromide; Tos-PheCH_2Cl, L-1-tosylamido-2-phenylethyl chloromethyl ketone; AB, alkaline borohydride; glycophorin A^{MN}, A^M, and A^N, superscripts MN, M, and N indicate the MN blood group type of the red cells, from which glycophorin A has been isolated; Pth, phenylthiohydantoin.

0006-2960/78/0417-4756$01.00/0

In this paper we describe the characterization of peptides produced by tryptic, chymotryptic, and cyanogen bromide cleavage and their isolation from either sialoglycoprotein mixtures or purified glycophorin A. The complete amino acid sequence of glycophorin A was determined and the attachment sites of individual oligosaccharides have been obtained. A discussion of the two different forms of glycophorin A, A^M and A^N, and a summary of a recent detailed study on the structure of the hydrophobic region of glycophorin A are included here.

The most difficult part of this study has been to determine the amino acid sequence of multiply glycosylated peptides. Since the phenylthiohydantoins of the glycosylated amino acid residues cannot be identified by conventional methods, and, in some instances, Edman degradation appeared to proceed only with low yields, our strategy has been to use the direct Edman and the dansyl-Edman procedures simultaneously. A further problem was our inability to isolate multiply glycosylated peptides free of small amounts of contaminating peptides. We have tried to overcome this problem by sequencing peptides from the same regions of the polypeptide chains which were generated by different proteolytic enzymes.

Methods

Preparation of the Sialoglycoprotein Fraction and Glycophorin A. Fresh human blood was obtained and membranes were prepared as described (Furthmayr & Marchesi, 1976). The lyophilized membrane preparations from several individuals were pooled, irrespective of blood type. The sialoglycoprotein fraction was isolated as described previously (Marchesi & Andrews, 1971). These preparations contain four bands when analyzed by $NaDodSO_4$–acrylamide gel electrophoresis, which have been labeled PAS-1, -2, -3, and -4 (Steck, 1974). The different glycophorins can also be resolved on $NaDodSO_4$–polyacrylamide gels prepared according to Laemmli (1970), as described by Furthmayr (1978a), Dahr et al. (1976), and Mueller et al. (1976). For some studies this material was fractionated by gel filtration on Bio-Rad A 0.5 equilibrated with 0.1% Ammonyx-LO (Onyx Chem. Co., N.J.) as described previously (Furthmayr et al., 1975) and the material eluting in the first and largest peak from the column was designated glycophorin A. For some studies glycophorin A was isolated by these methods from red cells of individuals homozygous for blood group M and N. These preparations are designated as glycophorin A^M and A^N, respectively. Glycophorin A^{MN} was obtained either from heterozygotes or from a pool of several untyped individuals.

Carbohydrate Analysis. Carbohydrates were determined quantitatively as the trimethylsilyl derivatives of methyl glycosides by gas liquid chromatography (Hewlett Packard 5710B gas chromatograph with an integrator 3373B) as described elsewhere (Reinhold, 1972). Dry samples containing 10–500 μg of carbohydrate were methanolyzed with 1.0 mL of 0.5 N HCl in anhydrous methanol at 65 °C for 16 h. Trisil reagent (Pierce Chemicals) was used for the trimethylsilylation of the methyl glycosides. The following conditions of gas liquid chromatography were used to identify the derivatives: 3% OV17 on Chromosorb WHP 80–100 mesh (6 ft × 2 mm glass columns); zero nitrogen (Precision Gas Products) as carrier gas, at a flow rate of 30 mL/min; the temperature program used was 190 °C for 8 min, followed by a temperature increase at a rate of 4 °C/min to 240 °C and maintaining the temperature at 240 °C for 8 min. A flame ionization detector was used and myoinositol served as an internal standard for the analysis. The carbohydrate content of the glycopeptides is expressed in mol/mol of protein after determination of the protein content of an aliquot by amino acid analysis.

Amino Acid Analysis. Dry samples containing 10–200 μg of amino acid were added to small glass tubes (7 × 60 mm) and hydrolyzed under vacuum in 6 N HCl containing 2% phenol and 1% 2-mercaptoethanol in a sealed glass chamber at 110 °C for 24 h. After hydrolysis the samples were dried in a heated desiccator at 65 °C for 1 h under vacuum and were dissolved in 50–100 μL of the buffer used for loading the amino acid analyzer. The amino acid analyses were performed on a Durrum D500 analyzer. Tryptophan was determined in a 0.1% $NaDodSO_4$ solution by fluorescence (wavelength of excitation 290 nm, of emission 360 nm) using a spectrofluorometer (Farrand Optical Co., Model 801). The data in all the tables are expressed in mol/mol of peptide. Since it is virtually impossible to determine the accurate molecular weight of the glycopeptides by chromatographic and electrophoretic methods and to obtain the size of the peptide portion, this was calculated by quantitative amino acid analysis using norleucine as an internal standard. Peptide samples were dissolved in an accurate volume of water, and 100-μL aliquots of the solution were taken into small test tubes which contained standard amounts of norleucine. From the amino acid analysis the absolute content of each amino acid was obtained, which was used to calculate the yields of the peptides and the molecular weights. This was considered important during the initial phase of this work since the original sialoglycoprotein preparations were contaminated with small amounts of other glycoproteins. By careful quantitation we were able to decide in most cases whether a peptide was derived from glycophorin A or a minor component. In later experiments only purified glycophorin A was used for the isolation of the peptides and the identification of all the major peptides was confirmed.

Enzymatic Hydrolysis. Digestions with trypsin [L-1-tosylamido-2-phenyl ethyl chloromethyl ketone treated, Worthington] were performed using the conditions described by Jackson et al. (1973). In a typical experiment, 300 mg of glycoprotein was dissolved in 30 mL of 40 mM Tris-HCl buffer, pH 8.2, and 9 mg of trypsin in 1.0 mL of the same buffer was added. After a 24-h incubation at 37 °C, trypsin was inactivated by the addition of a tenfold molar excess of *N-p-α*-tosyllysyl chloromethyl ketone (Sigma) and the solution was acidified to pH 3.5 with 1 N HCl. During the digestion period a precipitate usually formed, which was removed by centrifugation at 14 000 rpm for 10 min in a Beckman JA-20 centrifuge. This precipitate contained the peptide T6 as the major component. The supernatant containing soluble peptides and salt was concentrated by lyophilization.

Digestion with chymotrypsin (α-chymotrypsin, Worthington) was done at an enzyme:substrate ratio of 1:100 and a concentration of 10 mg/mL in 0.1 M sodium phosphate buffer, pH 7.0, at 37 °C for 24 h. The incubation mixture was acidified to pH 3 with 1 N HCl and was frozen to facilitate the precipitation of peptide material which was recovered by centrifugation after thawing of the digest at room temperature. The insoluble pellet contained peptide CH6.

Thermolysin (Sigma) digestion was done in 0.1 M ammonium bicarbonate, pH 8.0, at an enzyme:substrate ratio of 1:35 and a glycopeptide concentration of 0.2 mg/mL, for 24 h at 37 °C. The digestion was stopped by lyophilization.

Additional enzymatic digestions were performed on some of the isolated primary peptides and are indicated under Results. This report will deal with the water soluble peptides and will only summarize the results of a more detailed study on the hydrophobic peptide published elsewhere (Furthmayr et al., 1978).

Cyanogen Bromide Cleavage. Seventy-five milligrams of

purified glycophorin A was dissolved in 30 mL of 70% formic acid and after bubbling of nitrogen gas into the solution for 1 min, 50 mg of cyanogen bromide (Pierce) was added. After 48 h, at room temperature (Segrest et al., 1973), the solution was concentrated to 5 mL at 30 °C with a rotary evaporator equipped with a high vacuum pump connected to a dry ice-acetone trap. The digest was directly applied onto a gel filtration column of Sephadex G100 (2.5 × 100 cm) equilibrated with 25% formic acid and the eluted peptides were recovered by lyophilization.

Cyanogen bromide cleavage was also performed on a tryptic glycopeptide (T1, see under Results) using a similar method. Twenty-three milligrams of pooled T1 was dissolved in 2.3 mL of 70% formic acid and after addition of 50 mg of CNBr, the mixture was allowed to stand at room temperature for 24 h. The reaction was stopped by dilution with an excess volume of water and the mixture was lyophilized. The lyophilized peptides were dissolved in 3 mL of 0.05 N acetic acid and subjected to gel filtration on a Sephadex G-50 column (2.5 × 90 cm), equilibrated with the same buffer. The peptides were recovered by lyophilization.

Gel Filtration of Enzymatic Digests of Glycophorin. The soluble peptides from 300 mg of a tryptic or chymotryptic digest of glycophorin were separated by gel filtration on a column packed with Sephadex G-150 superfine (Pharmacia, 2.5 × 150 cm) which was previously equilibrated with 0.1 M ammonium acetate, pH 6.8, at room temperature. A tryptic digest of the chymotryptic peptide CH1 (see under Results) was separated on the same column. More recent fractionations were done by using Ultrogel AcA 54 (LKB) with considerable savings in time and without sacrifice in resolution. All other gel filtration experiments were done on Sephadex G-50 superfine columns (Pharmacia, 2.5 × 90 cm or 2.5 × 150 cm) equilibrated with 0.05 N acetic acid at room temperature.

Ion-Exchange Chromatography. Tryptic peptides T1 and T2 were further purified by chromatography on DEAE-cellulose (DE52, Whatman), packed into a small column (1 × 26 cm), and equilibrated with 0.05 M sodium formate buffer, pH 6.2, at room temperature. The peptides were applied onto the column in the same buffer, and, after an additional 30 mL of starting buffer, the peptides were eluted at a flow rate of approximately 50 mL/h with a linear gradient of NaCl from 0 to 0.5 M over a total volume of 1000 mL.

Tryptic peptides T3 and T4 were separated by DEAE-cellulose chromatography on a column (2 × 23 cm) under the same conditions and on phosphocellulose equilibrated in 1 mM sodium acetate, pH 3.8, using a NaCl gradient from 0 to 0.3 M over a total volume of 500 mL.

Purity of Peptides. Amino acid analysis of the peptides was a useful way to assess purity, but the most reliable method was by quantitative analysis of Pth-amino acids released from the peptide by Edman degradation. $NaDodSO_4$–polyacrylamide gel electrophoresis was not a suitable way to determine the purity of the peptides. Some peptide mixtures have been separated only by gel filtration since attempts to further purify them by ion-exchange chromatography were unsuccessful, presumably due to heterogeneity of the carbohydrates. To control for reliability of sequencing data, particularly of heavily glycosylated peptides, we confirmed the results on most peptide regions by analyzing overlapping peptides obtained by different proteolytic digestions.

Analysis of Glycopeptides in Solution. The peptides from column eluates were determined by absorbance at 226 nm and/or 280 nm using a Gilford spectrophotometer Model 240. Fluorescamine (Hoffmann-La Roche) was used to determine the elution of small peptides which were not detectable by their absorbance at 226 nm, as described elsewhere (Udenfriend et al., 1972). A ratio fluorometer (Turner, Model 111) was used for this assay. Fluorescamine could not be applied to determine the elution of peptides in ammonium acetate buffer which was used in some gel filtration experiments. The thiobarbiturate method was used for the quantitative determination of sialic acid in column eluates (Warren, 1959).

Sialidase Treatment of Glycopeptides. Sialidase was purified from a commercial sialidase (*Clostridium perfringens*, Sigma) according to the method of Cuatrecasas & Illiano (1971). The specific activity of the purified enzyme was assayed using glycophorin as a substrate. In a typical experiment sialoglycopeptide containing about 10 mg of sialic acid was incubated with one unit of the purified enzyme in 1.0 mL of 0.1 N sodium acetate buffer, pH 5.5, supplemented with 0.2 mM calcium chloride and 20 μM EDTA. After incubation at 37 °C for 5 h, the mixture was directly applied to a Sephadex G-10 column equilibrated with 50 mM acetic acid. The peptide peak was lyophilized. Usually more than 95% sialic acid was released as free sialic acid from the glycopeptide.

Edman Degradation of Peptides. Manual Edman degradation was carried out with minor modifications by the method of Peterson et al. (1972). In a typical experiment, 100 nmol of peptide was added to 0.2 mL of 0.4 M dimethylallylamine–trifluoroacetic acid buffer, pH 9.5, in 1-propanol–water (3:2 v/v). Ten microliters of phenyl isothiocyanate (Pierce, sequenator grade) was added under nitrogen and the solution was mixed vigorously (Vortex) and then incubated at 55 °C for 20 min. In order to stop the reaction, 1.0 mL of benzene (Pierce, sequenator grade) was added to the incubation mixture. After mixing, the solution was centrifuged at 3000 rpm for 1 min and the benzene layer was discarded. The aqueous layer was dried with nitrogen and then extracted with 0.5 mL of ethyl acetate (Pierce, sequenator grade) under nitrogen. After centrifugation at 3000 rpm for 1 min, the ethyl acetate layer was discarded and the precipitated peptide was dried completely in a heated desiccator at 65 °C using high vacuum for 10 min. Cleavage was done by addition of 100 μL of trifluoroacetic acid to the phenylthiocarbamoylated peptide and incubation at 55 °C for 8 min. Trifluoroacetic acid was evaporated by a stream of nitrogen while rotating the tube to leave a thin film of peptide on the glass wall. Anhydrous ether containing 10^{-4} M dithiothreitol (1.0 mL) was added, and the mixture was vortexed vigorously. The flakes of peptide formed in this step were recovered by centrifugation at 3000 rpm for 1 min. The ether layer was taken carefully into a small test tube and, after addition of Pth-norleucine (10–40 nmol) as an internal standard, the ether was evaporated by a stream of nitrogen. Conversion to the Pth derivative was done by addition of 0.2 mL of 1 N HCl and incubation at 80 °C for 10 min. Ethyl acetate (0.8 mL) was then added and after vigorous vortexing the reaction mixture was centrifuged at 3000 rpm for 1 min. The ethyl acetate layer was transferred to another test tube and dried with nitrogen. The hydrochloric acid layer was lyophilized directly. A single cycle of the degradation took about 50 min, and the amino acid sequence of up to 20 residues could be determined routinely using 100 nmol of peptide.

Dansyl-Edman Degradation. The method of Gray (1972) was slightly modified for sequencing of multiply glycosylated peptides. The Edman degradation procedure was the same as that described above. Prior to each degradation cycle about 5 nmol of peptide in coupling buffer was taken into a small test tube and dried using a stream of nitrogen. The dry sample was dissolved in 50 μL of 0.1 M sodium bicarbonate and reacted with 15 μL of dansyl chloride solution (Pierce, 5 mg/mL in acetone) for 2 h at room temperature. After drying with ni-

trogen the dansylated peptide was hydrolyzed in 6 N HCl at 110 °C for 16 h using an evacuated hydrolysis chamber. The hydrolyzed sample was dried in a heated desiccator containing sodium hydroxide pellets at 65 °C for 10 min under high vacuum and then extracted with 100 μL of aqueous ethyl acetate. The ethyl acetate extract was evaporated by a stream of nitrogen and, after addition of 10 μL of anhydrous ethyl acetate, aliquots were analyzed by thin-layer chromatography on polyamide sheets (5 × 5 cm, Cheng chin polyamide sheet) according to Gray (1972).

Identification of Amino Acid Phenylthiohydantoins. Pth-amino acids were identified by the following methods: gas liquid chromatography (Pisano & Bronzert, 1969), thin-layer chromatography (Summers et al., 1973), and amino acid analysis after back-hydrolysis (Smithies et al., 1971). Gas liquid chromatography was performed according to the method of Sauer et al. (1974) using a Hewlett-Packard Model 5710B gas chromatograph with integrator 3373B. Eight amino acid phenylthiohydantoins (alanine, threonine, serine, glycine, valine, proline, leucine, and isoleucine) were identified at 190 °C on a column of 10% DC560 on Chromosorb WHP (2 mm × 120 cm). The other eight amino acid phenylthiohydantoins (methionine, phenylalanine, asparatic acid, glutamic acid, asparagine, glutamine, lysine, and tyrosine) were identified as trimethylsilyl derivatives at 230 °C on the same column by a simultaneous injection of *N,O*-bis(trimethylsilyl)acetamide. Using this technique, leucine and isoleucine had an identical retention time, and the reponses of glutamine and asparagine were very low. Identification of tryptophan and histidine derivatives was not attempted by gas liquid chromatography. Thin-layer chromatography of amino acid phenylthiohydantoins was performed on a 5 × 5 cm polyamide sheet by the method of Summers et al. (1973). If a peptide was thought to contain Pth-histidine or Pth-arginine, the lyophilized HCl layer was analyzed for the two derivatives according to the method of Kulbe (1974). Although leucine, isoleucine, and norleucine derivatives were not separated, all the other amino acid phenylthiohydantoins could be identified semiquantitatively by thin-layer chromatography. The hydrolysis of the amino acid phenylthiohydantoins to the original amino acids was performed according to the method of Smithies et al. (1971). Dry samples of the amino acid derivatives were hydrolyzed with 100 μL of 55% HI at 130 °C for 18 h. The hydrolysate was evaporated in a heated desiccator at 65 °C containing NaOH pellets and P_2O_5 under a high vacuum and then injected onto the column of a Durrum D500 amino acid analyzer in loading buffer. By this technique Pth-serine and Pth-threonine were changed to alanine and α-aminobutyric acid, respectively. Methionine and tryptophan phenylthiohydantoins could not be identified because of their decomposition during hydrolysis. The absence of identifiable Pth-amino acids in any position during sequencing of a glycopeptide was indicative of glycosylation. Glycosylated residues were then identified by the dansyl-Edman technique.

Carboxypeptidase Treatment of Peptides. Fifty nanomoles of peptide was hydrolyzed with carboxypeptidase A and/or B (Worthington) as described elsewhere (Ambler, 1972). The total reaction mixture including the released amino acids and the remaining peptide was analyzed on the amino acid analyzer.

Hydrazinolysis. One hundred nanomoles of peptide was subjected to hydrazinolysis according to the method of Schroeder (1972). Free amino acids were separated from the amino acid hydrazines by chromatography on Amberlite CG-50 (1.0 × 4.0 cm column). The fractions which contained the free amino acids were collected, dried on a rotary evaporator, and analyzed for free amino acids on the amino acid analyzer.

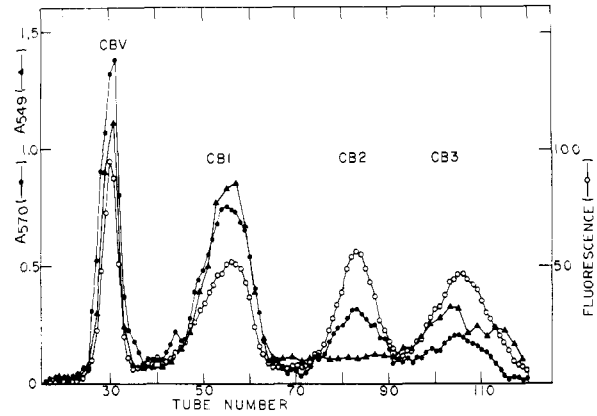

FIGURE 1: Gel filtration of cyanogen bromide peptides of glycophorin A. The peptides were separated at a constant flow rate at room temperature on a Sepahdex G-100 column (2.5 × 100 cm) in 25% formic acid. The fraction volume was 4 mL. Fractions were analyzed by the ninhydrin reaction at 570 nm after alkaline hydrolysis (●), for sialic acid at 549 nm (▲), and by the fluorescamine reaction (○). CBV void volume; CB1, CB2, CB3 cyanogen bromide derived peptides in pooled fractions.

Alkaline Borohydride Treatment of Peptide. Alkaline borohydride degradation was carried out under several different conditions by the method described by Spiro (1974). In a typical experiment, a glycopeptide such as T2 (11.0 mg) was dissolved in 2.0 mL of 0.4 M $NaBH_4$ in 0.2 M NaOH and allowed to stand at room temperature for 36 h. At the end of the incubation period the mixture was acidified to pH 5.0 by dropwise addition of 4 M acetic acid and applied onto a column (2.5 × 90 cm) of Sephadex G-50 equilibrated with 0.05 M acetic acid. The major peptide fractions were recovered by lyophilization.

Methanolysis of Glycopeptides. In a typical experiment a glycopeptide such as T2 (200 nmol) was methanolyzed with 1.0 mL of 0.5 M HCl in absolute methanol at 65 °C for 4 h. At the end of the incubation period methanol was evaporated by a stream of nitrogen. One milliliter of methanol was added to the residue and then evaporated again to remove traces of HCl. The residue was subjected directly to Edman degradation.

Results

Cyanogen Bromide Peptides of Glycophorin A. The gel filtration pattern of the CNBr peptides of glycophorin A is shown in Figure 1. Each of the three pooled fractions, which eluted in the included volume of the column, contains a single peptide in equivalent amounts as judged by their amino acid composition (Table I) and subsequent analysis. The peak eluting in the excluded volume of the column (CBV) apparently contains aggregated material and may be due to incomplete cleavage (Segrest et al., 1973). In earlier studies additional minor components had been described (Segrest et al., 1973) which presumably arose from contaminants of the sialoglycoprotein fraction with glycoproteins other than glycophorin A.

Only CB1 and CB3 contain sialic acid; CB2 lacks this and other carbohydrates as demonstrated earlier (Segrest et al., 1973). Quantitative carbohydrate analysis was not done since carbohydrate and sialic acid in particular are partially hydrolyzed under the conditions of CNBr cleavage.

Isolation of Peptides after Digestion of Glycophorin A with Trypsin. During tryptic digestion of glycophorin A an insoluble precipitate is formed which is removed from the soluble digest

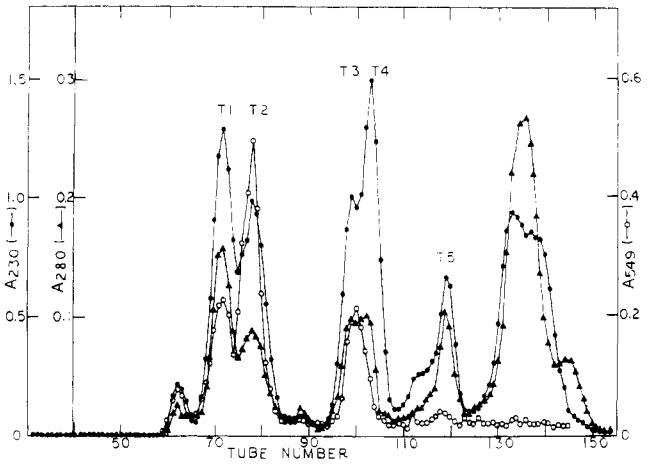

FIGURE 2: Gel filtration of tryptic peptides from the sialoglycoprotein fraction. The water-soluble peptides obtained from 300 mg were applied to a Sephadex G-150 (superfine) column (2.5 × 150 cm), which was previously equilibrated with 0.1 M ammonium acetate at room temperature. The column was eluted with the same buffer at a flow rate of 10 mL/h, and 5.0-mL fractions were collected. The peptides were analyzed by their absorbance at 230 nm (●) and 280 nm (▲) and by the thiobarbiturate method at 549 nm (○). T1–T5 designate the five major pools. Tryptic peptides from isolated glycophorin A gave a similar profile.

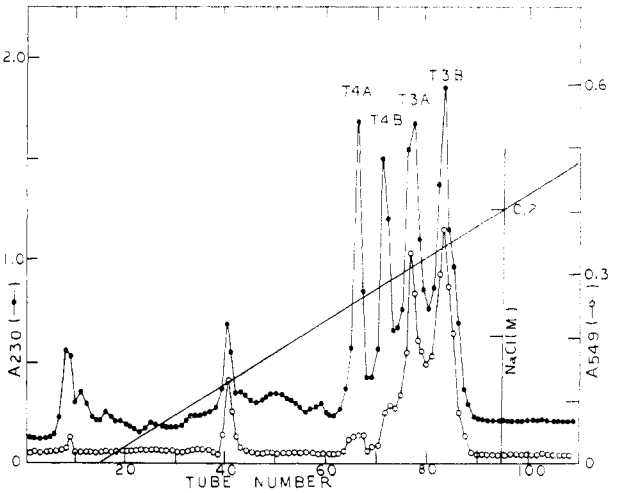

FIGURE 5: Fractionation of tryptic peptides T3 and T4 by DEAE-cellulose chromatography. The peptide pool T3 and T4 of Figure 2 (70 mg) was applied to a DEAE-cellulose DE-52 column (2.0 × 23 cm) equilibrated with 0.05 M sodium formate buffer, pH 6.2, at room temperature. The peptides were eluted with the same buffer at a flow rate of 50 mL/h, and 5.5-mL fractions were collected. At tube 15, a linear gradient was started (0–0.5 M NaCl). Absorbance at 230 nm (●), and sialic acid at 549 nm (○).

TABLE I: Chemical Composition of Cyanogen Bromide Derived Peptides of Glycophorin A.[a]

	CB1 (9–81)[b]	CB2 (82–131)	CB3 (1–8)
Asp	3.6 (3)	5.1 (5)	0.2
Thr[c]	10.2 (10)	3.1 (3)	1.9 (2)
Ser[c]	10 (10)	6.9 (7)	1.6 (1.5)
Glu	10.4 (10)	4.2 (4)	0.5 (0.5)
Pro	4.3 (4)	6.0 (6)	0.2
Gly	2.6 (2)	3.2 (3)	0.6 (0.5)
Ala	4.8 (4)	1.3 (1)	0.9 (1)
Val	5.9 (6)	3.7 (4)	1.0 (1)
Hse	0.8 (1)	*d*	0.6 (1)
Ile	4.7 (5)	5.6 (6)	0.4
Leu	2.2 (2)	5.2 (5)	0.6 (0.5)
Tyr	2.5 (3)	0.6 (1)	
Phe	1.9 (2)		
His	5 (5)		
Lys	2.5 (2)	3.3 (3)	
Arg	3.6 (4)	1.6 (2)	

[a] Expressed in moles/mole of peptide; expected values in parentheses. [b] Positions of peptide within amino acid sequence. [c] Corrected for losses during hydrolysis, factors for Thr 1.12, Ser 1.15. [d] No entry denotes 0.1 residue or less.

by centrifugation (see Methods). The elution profile of the soluble peptides obtained by gel filtration chromatography is shown in Figure 2. All the fragments elute within the included volume of the gel and virtually no undigested material is found. Fractions were combined corresponding to the five major pools, labeled as shown in the figure. The unlabeled peak contains salt, protease inhibitor, phenol red, and small amounts of amino acids. No attempt was made to recover peptide products from this peak. Sialic acid is present predominantly in pools T1, T2, and T3. Pools T1 and T2 were further purified by chromatography on DEAE-cellulose and gave single, symmetrical peaks (not shown). When the entire sialoglycoprotein fraction, which contains other sialoglyproteins in addition to glycophorin A, was digested with trypsin, an identical chromatogram is obtained. However, upon rechromatography of both pools 1 and 2 on DEAE-cellulose, small amounts of a different peptide were separated from the major peptides T1 and T2, corresponding to about 10–15% of the total material (Figures 3 and 4 in supplementary material; see paragraph concerning supplementary material at the end of this paper). These minor peptides are derived from glycophorin B and glycophorin C (Furthmayr, 1978a). Peptides contained in pools T3 and T4 of Figure 2 were separated by ion-exchange chromatography on DEAE-cellulose as shown in Figure 5. Only two of the peaks, designated T3A and T3B, contain sialic acid. Since the peptides are incompletely separated and are contaminated with each other as judged from their amino acid composition, further purification was done by phosphocellulose chromatography (not shown, Cotmore et al., 1977). This system resolves the sialoglycopeptide T3A from T4B (Figure 5) and small amounts of other contaminating peptides. Pool T5 of Figure 2 was not further purified since Edman degradation indicated that the peptide was pure (see below).

The peptide(s) contained in the insoluble precipitate of the digest were not amenable to conventional purification methods because of their hydrophobic nature. Isolation and purification of a homogeneous peptide (T6) from this fraction was achieved by countercurrent distribution in 1-butanol–acetic acid–water (4:1:5, v/v) as described elsewhere (Furthmayr et al., 1978).

The amino acid and carbohydrate compositions of the various peptides are summarized in Table II. All of the peptides are glycosylated with the exception of T4A, T4B, and T6. All of the peptides except T4B and T5 contain more than one Lys/Arg residue, suggesting either the presence of arginyl or lysyl bonds which are resistant to trypsin or the presence of multiple peptides. A second treatment of the isolated fragments with trypsin did not result in further fragmentation, but additional cleavage was observed after sialidase treatment (see below). This indicates that some of the peptide bonds in the native polypeptide are not accessible to cleavage.

Since the peptides T3A and T3B have an essentially identical amino acid composition, a difference in sialic acid content is probably responsible for their separation on DEAE-cellulose. Also the peptides T4A and T4B appear to differ only by the

TABLE II: Chemical Composition of Tryptic Peptides.[a]

	T1 (1-39)[b]	T2 (1-31)	T3A (40-61)	T3B (40-61)	T4A (101-131)	T4B (102-131)	T5 (32-39)	T6[c] (62-97)
Asp	3 (3)	2.2 (2)	0.2	0.2	4.8 (5)	4.6 (5)	1.0 (1)	0.2
Thr[d]	9 (9)	7.3 (7)	2.1 (2)	2.0 (2)	2.0 (2)	2.2 (2)	1.5 (2)	2 (2)
Ser[d]	8.7 (8.5)	8.5 (8.5)	2.2 (2)	2.0 (2)	5.8 (6)	5.6 (6)	0.3	2.1 (2)
Glu	1.7 (1.5)	1.7 (1.5)	6.1 (6)	5.7 (6)	4.3 (4)	4.2 (4)	0.3	3.1 (3)
Pro	0.9 (1)		2.3 (2)	2.1 (2)	6.0 (6)	6.3 (6)	1.0 (1)	1.1 (1)
Gly	0.6 (0.5)	0.6 (0.5)	1.0 (1)	1.0 (1)			0.2	4 (4)
Ala	2.9 (3)	1.2 (1)	1.0 (1)	1.0 (1)			1.7 (2)	2 (2)
Val	2.2 (2)	2 (2)	2.6 (3)	2.8 (3)	3.1 (3)	3.2 (3)		2.7 (3)
Met	0.8 (1)	0.4 (1)						0.8 (1)
Ile	1 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.1 (1)	1.1 (1)		5.9 (7)
Leu	0.5 (0.5)	0.6 (0.5)		0.2	2.1 (2)	2.1 (2)	0.2	4.2 (4)
Tyr	1.7 (2)	0.8 (1)	0.9 (1)	1.0 (1)			0.6 (1)	1 (1)
Phe	e							2.1 (2)
His	2.1 (2)	2 (2)	0.9 (1)	0.9 (1)				2 (2)
Lys	2.1 (2)	2 (2)	0.3		2.1 (2)	1.0 (1)		
Arg	1.7 (2)	0.5 (-/1)	1.9 (2)	1.7 (2)			0.8 (1)	1.7 (1-2)
Fuc	0.4[f]	0.4[f]						
Man	2.7	3.2	0.2	0.1				
Gal	13.5	11.1	2.7	4.2			0.8	
Glc			0.1	0.2			0.2	
GalNAc	10	7.9	3.0	3.0			0.7	
GlcNAc	6.9	6.3						
NANA	12.7	10.7	4.6	7.0			1.5	
yield[g] (nmol)	4500	4100	4400	4400	4100	4600	3500	ND

[a] Expressed in mol/mol of peptide; expected values in parentheses. [b] Numbers in parentheses indicate position of peptide within sequence. [c] Data taken from Furthmayr et al. (1978). [d] Corrected for losses during hydrolysis, factors for Thr 1.12, Ser 1.15. [e] Dash or no entry denotes 0.1 or less. [f] Data taken from Furthmayr (1978a,b). [g] From 300 mg of glycophorin A.

presence of an additional lysine in T4A. In addition the sum of the amino acids of T2 and T5 equals the composition of peptide T1. All of the peptides were obtained in equivalent yields (Table II). The yields of T1 and T2 plus T5 are consistent with the idea that about half of the glycophorin A polypeptides are cleaved at an additional site within the T1 sequence. The higher carbohydrate content of T1 as compared with T2 plus T5 indicates that the additional carbohydrate present in T1 may be responsible for the lack of cleavage.

When glycophorin A is isolated from red blood cells of individuals which are homozygous for the blood group antigens of the MN system, specific differences in amino acid composition of peptides T1 and T2 are observed (Furthmayr, 1978a,b). The peptides A^MT1 and A^MT2 (isolated from glycophorin A^M of homozygotes for M) contain an additional 0.5 mol of serine and glycine and lack 0.5 mol of leucine and glutamic acid. Peptides A^NT1 and A^NT2, on the other hand, contain less serine and glycine and more leucine and glutamic acid than would be expected from a mixture of both forms of glycophorin A. This is reflected in the expected values for these amino acids in some of the tables.

Isolation and Composition of Peptides Obtained by Digestion of Glycophorin A with Chymotrypsin. Digestion of glycophorin A with chymotrypsin also gives an insoluble precipitate which can be removed by centrifugation. The remaining soluble peptide material is resolved by gel filtration chromatography and is illustrated in Figure 6. Five distinct peaks are obtained, labeled CH1-CH5, in addition to an unlabeled peak in the included volume. The first five peaks contain homogeneous peptides as judged from the amino acid composition (Table III) and subsequent analysis. The last peak was found to contain only small amounts of peptide material in addition to salt and was not further analyzed. Only peptides CH1, CH2, and CH3 contain carbohydrate. Thus, these peptides are derived from the same segment of the glycophorin

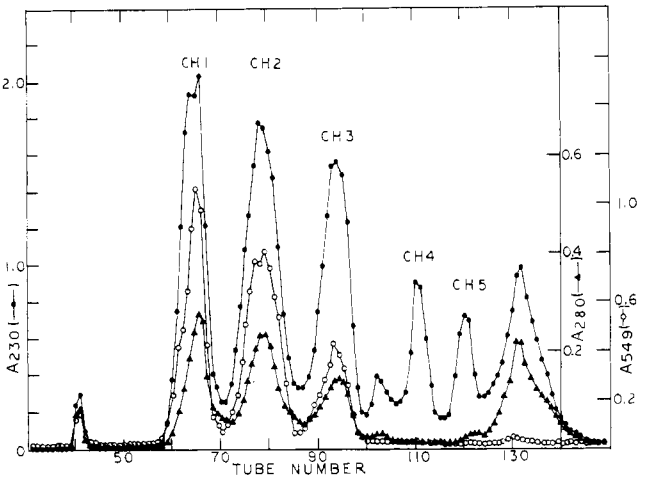

FIGURE 6: Gel filtration of chymotryptic peptides from the sialoglycoprotein fraction. The water-soluble peptides were applied to a Sephadex G-150 (superfine) column (2.5 × 150 cm), which was previously equilibrated with 0.1 M ammonium acetate, pH 6.8, at room temperature. The column was eluted with the same buffer at a flow rate of 10 mL/h, and 5.0-mL fractions were collected. The fractions were analyzed by their absorbance at 230 nm (●) and 280 nm (▲), and by the thiobarbiturate method for sialic acid at 549 nm (○). The peptides obtained by digestion of purified glycophorin A gave a similar profile.

A polypeptide chain as the tryptic glycopeptides T1, T2, T3, and T5.

The peptides CH1 and CH2 contain more than 1 equiv of tyrosine, and these are not cleavable even after a second treatment with the enzyme. The combined amino acids of CH2 and CH3 correspond to peptide CH1 consistent with the idea that these two peptides are produced by partial cleavage of the original polypeptide chain. The two peptides labeled CH4 and

TABLE III: Chemical Composition of Chymotrytic Peptides.[a]

	CH1 (1-64)[b]	CH2 (1-34)	CH3 (35-64)	CH4 (99-118)	CH5 (119-131)	CH6[c] (65-98)
Asp	3.1 (3)	2.8 (3)	0.2	3.1 (3)	2.2 (2)	1.4
Thr	10.3 (11)	7.7 (8)	3.3 (3)	1.1 (1)	1.0 (1)	2.7
Ser	8.6 (10.5)	6.9 (8.5)	2.3 (2)	2.5 (3)	2.5 (3)	2.9
Glu	7.9 (8.5)	2.2 (1.5)	6.9 (7)	0.2	4.1 (4)	2.0
Pro	3.1 (3)		2.4 (3)	5.3 (5)	1.1 (1)	2.7
Gly	1.8 (1.5)	1.2 (0.5)	1.2 (1)		0.2	2.9
Ala	4.5 (4)	1.3 (1)	3.5 (3)		0.3	2.5
Val	6.1 (6)	2.2 (2)	4.6 (4)	2.3 (2)	1.1 (1)	2.3
Met	0.6 (1)	0.6 (1)				0.9
Ile	2.0 (2)	1.1 (1)	1.0 (1)	1.0 (1)	1.1 (1)	4.6
Leu	1.3 (1.5)	0.8 (0.5)	1.0 (1)	2.0 (2)	0.2	3.5
Tyr	2.6 (3)	1.4 (2)	0.7 (1)			1.0
Phe	*d*					1.4
His	3.0 (3)	1.8 (2)	0.9 (1)			1.2
Lys	2.1 (2)	1.8 (2)	0.2	2.9 (3)	0.2	1.0
Arg	3.0 (4)	0.6 (1)	2.6 (3)			1.4
Fuc	1.3	1.5				ND[e]
Man	5.2	4.0	0.1			ND
Gal	24.9	18.7	5.2			ND
Glc			0.6			ND
GalNAc	14.5	12.4	4.0			ND
GlcNAc	8.1	7.3	0.5			ND
NANA	28.7	19.2	7.3			ND
yield[f] (nmol)	2900	5800	7400	4800	6600	ND

[a] Not corrected for losses during hydrolysis. Expressed in mol/mol of peptide; expected values in parentheses. [b] Numbers in parentheses indicate position of the peptide within the sequence. [c] Insoluble pellet without further purification; no expected values given. [d] No entry denotes 0.1 residue or less. [e] Not done. [f] From 300 mg of glycophorin.

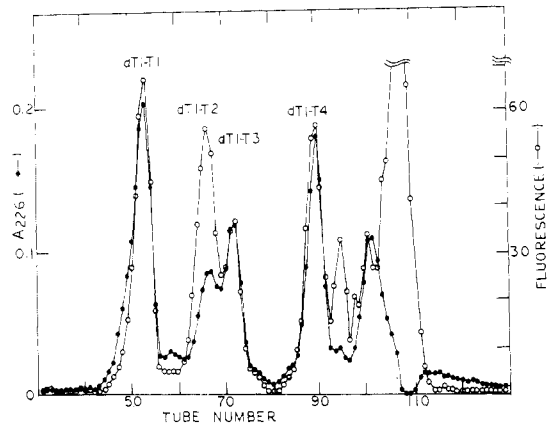

FIGURE 7: Gel filtration of tryptic peptides of desialated T1 (dT1). The peptides were applied to a Sephadex G-50 (superfine) column (2.5 × 90 cm), equilibrated with 50 mM acetic acid at room temperature. The column was eluted with the same buffer at a flow rate of 20 mL/h, and 4.1-mL fractions were collected. Individual fractions were analyzed by their absorbance at 226 nm (●) and by the fluorescamine reaction (○).

CH5 do not contain carbohydrate and their combined amino acid composition is similar to the tryptic peptide T4.

Tryptic Hydrolysis of Desialated Glycopeptides. A second tryptic digestion of peptide T1 does not result in further fragmentation. To explain this result, the possibility was considered that carbohydrates and sialic acid in particular block trypsin sensitive peptide bonds, either by steric hindrance or by interactions between the sugars and amino acid side chains. Alternatively, resistance to trypsin can also be explained by linkage of lysyl or arginyl residues with imino acids or involving modified lysyl residues (Bornstein, 1970). In order to distinguish between these possibilities, peptide T1 was desialated with neuraminidase. Desialated T1 (4 mg) was then dissolved in 1.0 mL of 0.1 M ammonium bicarbonate and 100 μg of Tos-PheCH$_2$Cl-trypsin in 100 μL of the same buffer was added. After incubation for 24 h at 37 °C, the mixture was subjected to gel filtration (Figure 7). Several fragments are separated, some of which are identical with the original tryptic fragments isolated from glycophorin A. The amino acid composition shows that dT1-T1 and dT1-T4 correspond to T2 and T5, respectively, and that the sum of the amino acids of dT1-T2 and dT1-T3 equals T2 (Table IV). Peptide dT1-T1 could not be further fragmented to dT1-T2 and dT1-T3 with trypsin. On the other hand, the desialated peptide T2 can be cleaved completely to produce two peptides, the amino acid compositions of which are identical with those of dT1-T2 and dT1-T3 (data not shown). These results indicate that an internal lysine in T1 is blocked but can be cleaved after desialation, and that an additional lysyl group is only partially cleaved in desialated T1 but can be completely cleaved in T2 after desialation. It is suggested that sialic acid interferes with cleavage, although other carbohydrate moieties may play some role as well. It is also possible that there are additional oligosaccharide chains attached in T1, which are not present in T2.

Cyanogen Bromide Cleavage of Glycopeptides from T1. CNBr digestion of T1 produced two major peptides labeled T1-CB1 and T1-CB2 (Figure 8). T1-CB1 does not contain homoserine (Table V), indicating that this fragment is derived from the C-terminal end of glycopeptide T1 (Table II). The other peptide T1-CB2 contains homoserine. As expected, the amino acid composition of peptides T1-CB1 plus T1-CB2 is nearly identical with that of peptide T1. T1-CB2 is identical with peptide CB3 (Figure 1 and Table I). The same octapeptide was isolated from T2 (not shown) and from the desialated peptide dT1-T2 (Figure 7 and Table IV). The elution pattern of a cyanogen bromide digest of dT1-T2 is illustrated in Figure 9, and fragment dT1-T2-CB3 is identical with the N-terminal peptide CB3 (Table V).

TABLE IV: Amino Acid Composition of Tryptic Fragments of Desialated Glycopeptide T1.[a]

	dT1-T1 (1-31)[b]—T2	dT1-T2 (1-18)	dT1-T3 (19-31)	dT1-T4 (32-39)—T5
Asp	2.2 (2)	0.7	2.0 (2)	1.0 (1)
Thr	7.5 (7)	5.1 (5)	2.3 (2)	2.0 (2)
Ser	7.5 (8.5)	5.0 (5.5)	3.0 (3)	0.2
Glu	1.7 (1.5)	0.9 (0.5)	1.1 (1)	
Pro	0.2			1.0 (1)
Gly	0.8 (0.5)	0.7 (0.5)		0.2
Ala	1.3 (1)	1.2 (1)		2.1 (2)
Val	2.3 (2)	2.1 (2)		
Met	0.7 (1)	0.7 (1)		
Ile	1.0 (1)	0.3	1.0 (1)	
Leu	0.5 (0.5)	0.5 (0.5)		
Tyr	0.7 (1)	0.2	0.8 (1)	0.9 (1)
Phe	[c]			
His	2.5 (2)	1.0 (1)	0.9 (1)	
Lys	1.8 (2)	1.2 (1)	1.0 (1)	
Arg	0.3 (-/1)		0.3 (-/1)	1.2 (1)
yield (nmol)	83	83	58	158
fractions[d] combined	48-55	63-69	70-74	86-91

[a] Expressed in mol/mol of peptide; expected values in parentheses. Threonine and serine were corrected for losses during hydrolysis (factors 1.12, 1.25). [b] Numbers in parentheses indicate the position of the peptide in the sequence. [c] Dash or no entry denotes 0.1 residue or less. [d] Cf. Figure 7.

Digestion of Chymotryptic Glycopeptides with Trypsin. The isolation and characterization of the glycopeptides derived from chymotryptic fragments CH1 and CH2 after digestion with trypsin are included as supplementary material (Figures 10 and 11 and Tables VI and VII). A summary of the findings is included in Figure 12, which gives the order of the various fragments within the amino acid sequence of glycophorin A. The alignment of these peptides will become clear from the elucidation of their amino acid sequence and will be discussed in more detail in a later section.

Preparation of Additional Fragment for Amino Acid Sequence Analysis. Various fragments were isolated from desialated peptide T1 after digestion with chymotrypsin (Figure 13 and Table VIII; see supplementary material), from peptide T3 after digestion with thermolysin (Figure 14 and Table IX; see supplementary material), and by alkaline borohydride treatment of peptide T2 (Figure 15 and Table X; see supplementary material).

Determination of the Amino Acid Sequence of Glycophorin A. Two problems were of major concern to us during our attempts to determine the entire amino acid sequence of this highly glycoslyated protein. The first was our inability to purify smaller glycopeptides to homogeneity. This was due to two major factors: the negative charge of sialic acid residues and heterogeneity in size and charge of the peptides presumably because of variable carbohydrate compositions. Our attempts to purify some of the peptides by ion-exchange chromatography failed either because of lack of separation or because of splitting of the peaks into a number of components which made it difficult to recover sufficient material for sequencing. To overcome this problem we decided to prepare several peptides from the same regions of the molecule by using different proteolytic enzymes so that we could compare the results obtained on one peptide with those from others derived from the same peptide segment.

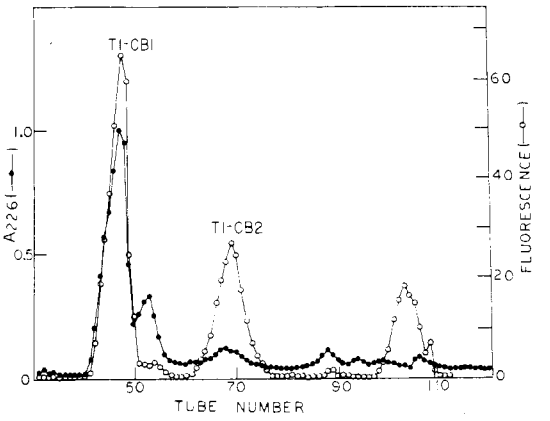

FIGURE 8: Gel filtration of cyanogen bromide derived peptides of T1. The peptides were applied to a Sephadex G-50 superfine column (2.5 × 90 cm) equilibrated with 50 mM acetic acid at room temperature. The column was eluted with the same buffer at a flow rate of 15 mL/h, and 4-mL fractions were collected. Absorbance at 226 nm (●) and by the fluorescamine reaction (○).

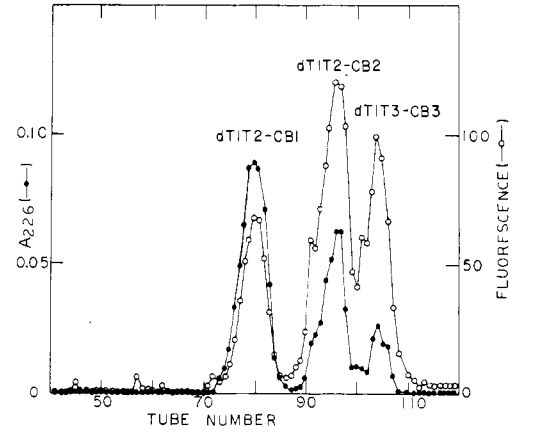

FIGURE 9: Gel filtration of cyanogen bromide cleavage peptides of dT1T2. The reaction mixture was applied to a Sephadex G-50 superfine column (2.5 × 90 cm) equilibrated with 50 mM acetic acid. The column was eluted with the same buffer at a flow rate of 15 mL/h, and 4.1-mL fractions were collected. Absorbance at 226 nm (●) and fluorescamine reaction (○).

Our second concern was to find or develop a procedure for sequencing multiply glycosylated peptides. We have explored several methods (which are outlined in some detail and are included as supplementary material) and have found that our most reliable results were obtained by using simultaneously the direct Edman and the dansyl-Edman degradation.

Amino-Terminal Sequence (Residues 1-20). When the complete glycophorin A molecule (6.0 mg) was subjected to Edman degradation, the amino-terminal sequence was found to be Leu-(X)-(X)-(X)-Gly-Val-Ala-Met-His-(X). As described previously (Segrest et al., 1973), serine was also identified as a minor component at the amino terminus (approximately 20% of leucine) and glutamic acid was found at the fifth position (approximately 30% of glycine). No Pth-amino acids were detected in the second, third, fourth, and tenth steps in significant amounts. Five glycopeptides (T1, T2, CH1, CH2, and CB3) had an amino-terminal sequence identical with that of glycophorin A, but the ratios of leucine to serine at the amino terminus and glycine to glutamic acid at the fifth position varied depending on individual peptide preparations. When two of the glycopeptides (T1 and CH2) were subjected to the dansyl-Edman procedure, the sequence Ser-Thr-Thr was determined for the second to the fourth

TABLE V: Amino Acid Composition of Cyanogen Bromide Cleavage Fragments of Glycopeptides T1 and dT1T2.[a]

	T1-CB1 (9-39)[b]	T1-CB2 (1-8) —CB3	dT1T2-CB1 (1-18) —dT1T2	dT1T2-CB2 (9-18)	dT1T2-CB3 (1-8) —CB3
Asp	2.9 (3)		1.3		
Thr[a]	6.5 (7)	2.0 (2)	4.7 (5)	3.1 (3)	1.7 (2)
Ser[a]	6.3 (6)	1.3 (1.5)	4.9 (5.5)	3.0 (3)	1.3 (1.5)
Glu	1.1 (1)	0.4 (0.5)	1.0 (0.5)	0.2	0.4 (0.5)
Pro	1.0 (1)	0.2			
Gly	0.2	0.8 (0.5)	0.6 (0.5)	0.2	0.6 (0.5)
Ala	2.0 (2)	1.2 (1)	1.0 (1)	0.2	1.0 (1)
Val	1.1 (1)	0.9 (1)	1.7 (2)	1.4 (1)	1.0 (1)
Met	c		0.5 (1)		
Ile	0.9 (1)		0.6		
Leu	0.2	0.3 (0.5)	0.4 (0.5)		0.4 (0.5)
Tyr	1.4 (2)		0.5		
Phe					
His	1.9 (2)		1.1 (1)	1.0 (1)	
Lys	1.9 (2)		1.2 (1)	1.1 (1)	
Arg	1.7 (2)		0.2		
Hse		0.7 (1)			0.6 (1)
yield (nmol)	450[d]	670	100[f]	217	217
fractions[e] combined	45-49	65-75	85-94[g]	100-109	111-117

[a] Thr and Ser are corrected for losses during hydrolysis. Expressed as mol/mol of peptide; expected values in parentheses. [b] Numbers in parentheses indicate the position of the peptide in the sequence. [c] No entry denotes 0.1 residue or less. [d] From 23 mg of T1. [e] Cf. Figure 8. [f] From 4 mg of dT1T2. [g] Cf. Figure 9.

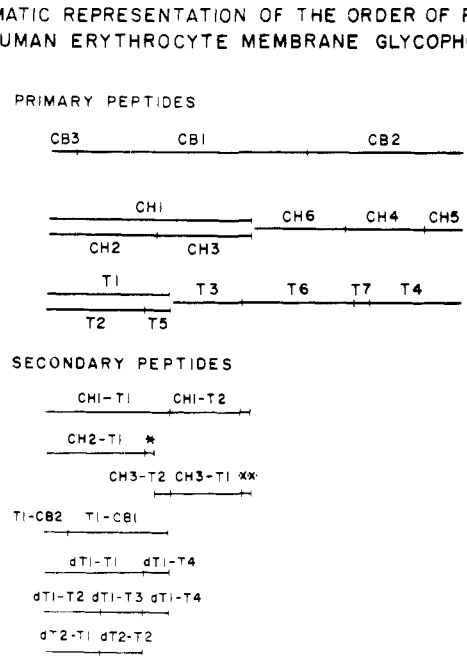

FIGURE 12: Schematic representation of the order of the peptides of glycophorin A. The heavily drawn line indicates the order of the cyanogen bromide derived (CB), chymotryptic (CH), and tryptic peptides (T) of glycophorin A. In addition, the order and nomenclature of secondary peptides is given in fine line drawings.

residues. The sequence Ser-Thr in the second and third positions was not conclusive, however, because the intensity of dansylthreonine on polyamide plates was nearly identical with that of dansylserine for the two steps.

When this study was initiated, we did not recognize the significance of this heterogeneity in positions 1 and 5 but, as described above, recent studies have resolved these ambiguities by demonstrating that glycophorin A exists in two forms, which are genetically determined by the MN locus and which are structurally different (Furthmayr, 1978b). The amino-terminal sequence of A^M was found to be

$$\text{Ser-}\overset{*}{\text{Ser}}\text{-}\overset{*}{\text{Thr}}\text{-}\overset{*}{\text{Thr}}\text{-Gly}$$

and for A^N, serine is replaced by leucine in position 1 and glycine by glutamic acid in position 5.[2] Heterozygous individuals express both molecules on their membranes and since most of the work described here was done with glycophorin A isolated from a pool of donors, this heterogeneity is reflected in the amino acid composition and sequence data on all the peptides derived from this region (cf. Tables I–V, etc.). For the sake of simplicity only the amino acids for A^N are given in Table XI, but the alternative Pth or dansyl derivatives were found for the peptides listed.

The amino acid sequence of three glycopeptides (T1, T2, CH2) was determined up to 18 steps as shown in Table XI. No Pth-amino acids were identified for the 6 residues from the 10th to the 15th cycle, whereas Pth-valine was detected in the 16th cycle with reasonable yield. Since it proved difficult to get reliable results for steps 10–15, dansyl-Edman and direct Edman degradation were applied to the two glycopeptides T1-CB1 (9–39) and dT1-T2 (9–18) and the sequence was found to be

$$\text{His-}\overset{*}{\text{Thr}}\text{-}\overset{*}{\text{Thr}}\text{-}\overset{*}{\text{Thr}}\text{-}\overset{*}{\text{Ser}}\text{-}\overset{*}{\text{Ser}}\text{-}\overset{*}{\text{Ser}}\text{-Val-Ser-Lys-Ser-}$$
$$\text{Tyr}^{20}\text{-Ile-}\overset{*}{\text{Ser}}\text{-}$$

Peptide dT1-T3 (19–31) confirmed the C-terminal part of this sequence and the glycosylated serine in position 22 (Table XI).

Sequence of the Glycopeptides Containing N-Glycosylated Asparagine (Residues 21–32). The amino-terminal peptides T1, T2, CH1, and CH2 contain mannose and *N*-acetylglu-

[2] The asterisk indicates glycosylation.

TABLE XI: Amino-Terminal Sequence of Glycophorin A and the Peptides Used for Sequence Studies.[a]

PEPTIDE		1	②	③	④	5	6	7	8	9	⑩	⑪	⑫	⑬	⑭	⑮	16	17	18	19	20
PROPOSED SEQUENCE		A^M Ser / A^N Leu				-Ser-	Thr-	Thr-	Gly / Glu	-Val-	Ala-	Met-	His-	Thr-	Thr-	Thr-	Ser-	Ser-	Ser-	Val-	Ser-	Lys-	Ser-	Tyr-
Glycophorin		Leu-	( )-	( )-	( )-	Glu-	Val-	Ala-	Met-	His-	( )-										
T1	(1-39)	Leu-	Ser-	Thr-	Thr-	Glu-	Val-	Ala-	Met-	His-	( )-	( )-	( )-	( )-	( )-	( )-	Val-	( )-	Lys-		
T2	(1-34)*	Leu-	( )-	( )-	( )-	Glu-	Val-	Ala-	Met-	His-	( )-	( )-	( )-	( )-	( )-	( )-	Val-	Ser-	Lys-		
CH1	(1-64)	Leu-	( )-	( )-	( )-	Glu-	Val-	Ala-	Met-	His-	( )-										
CH2	(1-34)	Leu-	Ser-	Thr-	Thr-	Glu-	Val-	Ala-	Met-	His-	( )-	( )-	( )-	( )-	( )-	( )-	Val-	( )-	Lys-		
T1CB2	(1-8)	Leu-	( )-	( )-	( )-	Glu-	Val-														
T1CB1	(9-39)*										His-	Thr-	Thr-	Thr-	Ser-	Ser-	Ser-	Val-	Ser-	Lys-	Ser-	Tyr-
dT1T2CB2	(9-18)										His-	Thr-	Thr-	Thr-	Ser-	Ser-	Ser-	Val-				
dT1T3	(19-31)*																				Ser-	Tyr-

		21	㉒	23	24	㉕	㉖	27	28	29	30	31	32	33	34	35	36	㊲	38	39	40
PROPOSED SEQUENCE		-Ile-	Ser-	Ser-	Gln-	Thr-	Asn-	Asp-	Thr-	His-	Lys-	Arg-	Asp-	Thr-	Tyr-	Ala-	Ala-	Thr-	Pro-	Arg-	Ala-
T1CB1	(9-39)	Ile-	Ser-	Ser-																	
dT1T3	(19-31)*	Ile-	Ser-	Ser-	Gln-	Thr-	Asx-	Asp-	Thr-												
dT1CH2	(21-34)	Ile-	Ser-	Ser-	Gln-	Thr-	Asx-	Asp-													
dT1CH5	(30-34)										Lys-	Arg-	Asp-								
T5	(33-39)												Asp-	Thr-	Tyr-	Ala-	Ala-	Thr-	Pro-	Arg-	
dT1CH4	(35-39)															Ala-	Ala-	( )-			
CH3	(35-64)*															Ala-	Ala-	Thr-	Pro-	Arg-	Ala-
CH3T2	(35-39)															Ala-	Ala-	Thr-	Pro-	Arg-	

		40	41	42	43	㊹	45	46	㊼	48	49	㊿	51	52	53	54	55	56	57	58	59	60	61
PROPOSED SEQUENCE		-His-	Glu-	Val-	Ser-	Glu-	Ile-	Ser-	Val-	Arg-	Thr-	Val-	Tyr-	Pro-	Pro-	Glu-	Glu-	Glu-	Thr-	Gly-	Glu-	Arg-	
CH3	(35-64)*		His-	Glu-	Val-	Ser-	Glu-	Ile-	( )-	Val-	Arg-	( )-	Val-	Tyr									
T3A	(40-61)	Ala-	His-	Glu-	Val-	( )-	Glu-	Ile-	( )-	Val-	Arg-	( )-	Val-	Tyr									
T3B	(40-61)	Ala-	His-	Glu-	Val-	Ser-	Glu-	Ile-	Ser-	Val-	Arg-	Thr-	Val-	Tyr-	Pro-	Pro-	Glu-	Glu-	Glu-	Thr-	Gly		
T3TH3	(57-61)																				Glu-	Thr-	Gly-	Glu-	Arg

[a] Circled numbers indicate glycosylation at these positions. Empty parentheses () indicate that no phenylthiohydantoin-amino acid was identified in the direct Edman degradation procedure alone. Underlined residues were identified by the dansyl-Edman technique only. * indicates that quantitative data are given for these peptides in the supplementary materials.

cosamine indicating that they also contain N-glycosylated residue(s) of asparagine. Since it was not possible to identify Pth-amino acids for more than 18 cycles of Edman degradation without ambiguity for these highly glycosylated peptides, smaller glycopeptides containing N-glycosylated asparagine were required. Two glycopeptides designated dT1-T3 (residues 19–31) and dT1-CH2 (21–34) which were prepared from desialated glycopeptide T1 with trypsin (cleavage at Lys^{18}) and chymotrypsin (cleavage at Tyr^{20}), respectively, were suitable for this purpose, although the two glycopeptides were slightly contaminated with other glycopeptides (Figure 4, Table IV and Figure 13, Table VIII in supplementary material). The amino-terminal sequence of dT1-T3 was tentatively determined as Ser-Tyr^{20}-Ile-(X)-Ser-Gln-(X)-(X)-Asp-Thr- by the identification of amino acid phenylthiohydantoins. The dansyl-Edman procedure was used to identify serine for the fourth residue, threonine for the seventh residue, and aspartic acid for the eighth residue (Table XI). Therefore, the eighth residue should be N-glycosylated asparagine. However, the yields of Pth-aspartic acid and Pth-threonine at the ninth and tenth cycles were only about 30% of the expected yield, raising the possibility that the Pth-amino acids in steps 9 and 10 were derived from a contaminating peptide. Since the sequence

$$\text{Ile-}\overset{*}{\text{Ser}}\text{-Ser-Gln-}\overset{*}{\text{Thr}}\text{-Asx-Asp-Thr}$$

was also obtained for the glycopeptide dT1-CH2, which has been prepared in a different way, it is unlikely that the sequence Asp-Thr for positions 27 and 28 is derived from a contaminating peptide. We consider it more likely that Edman degradation does not proceed normally through N-glycosylated asparagine for reasons which are presently unknown. To seek additional confirmation for this sequence, the residual peptide of dT1-T3 after eight cycles of degradation was isolated by gel filtration on Sephadex G-25 (Figure 16). The first peak, dT1-T3-8A has an amino acid composition consistent with incompletely degraded peptides, while the second peak DT1-T3-8B has a composition corresponding to a tetrapeptide composed of histidine, threonine, aspartic acid, and lysine (Table XII). From the elution volume of dT1-T3-8B on Sephadex G-25, it was concluded that the peptide is not glycosylated. Aspartic acid was found at its amino terminus by the dansylation technique. These results further support the interpretation that the sequence Asp^{27}-Thr originates from dT1-T3 peptides and not from contaminants and that the correct sequence is

$$\text{-}\overset{*}{\text{Asn}}\text{-Asp-Thr-}His^{29}\text{-Lys}$$

since the peptide was obtained by tryptic digestion.

Table XI shows the overlaps determined by these peptides. Although histidine in the 29th position could not be identified by Edman degradation, the sequence around $histidine^{29}$ is further supported as Thr-His^{29}-Lys by the following reasoning. First, the peptide dT1-CH5 which was isolated by chymotryptic digestion of desialated glycopeptide dT1 (Figure 13 and

TABLE XIII: Carboxy-Terminal Sequence of Glycophorin A and the Peptides Used for Sequence Studies.[a]

PEPTIDE	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
PROPOSED SEQUENCE		-Val-	Gln-	Leu-	Ala-	His-	His-	Phe-	Ser-	Glu-	Pro-	Glu-	Ile-	Thr-	Leu-	Ile-	Ile-	Phe-	Gly-	Val-
CH3T3 (62-64)		→	→	→																
T6 (62-96)		→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→
CH6 (65-98)					→	→	→	→	→	→										

	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
PROPOSED SEQUENCE	-Met-	Ala-	Gly-	Val-	Ile-	Gly-	Thr-	Ile-	Leu-	Leu-	Ile-	Ser-	Tyr-	Gly-	Ile-	Arg-	Arg-	Leu-	Ile-	Lys-
CB2 (82-131)		→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→
T6 (62-96)	→	→	→	→	→															
CH4 (99-118)																			→	→

	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
PROPOSED SEQUENCE	-Lys-	Ser-	Pro-	Ser-	Asp-	Val-	Lys-	Pro-	Leu-	Pro-	Ser-	Pro-	Asp-	Thr-	Asp-	Val-	Pro-	Leu-	Ser-	Ser-
T4A (101-131)	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→
T4B (102-131)		→	→	→	→	→	→	→	→	→	→	→								
CH4 (99-118)	→	→	→	→	→	→	→	→	→	→	→	→								
CH5 (119-131)*																			→	→

	121	122	123	124	125	126	127	128	129	130	131
PROPOSED SEQUENCE	-Val-	Glu-	Ile-	Glu-	Asn-	Pro-	Glu-	Thr-	Ser-	Asp-	Gln
CH5 (119-131)*	→	→	→	→	→	→	→	→	→	→	←
CH5TH1 (123-131)			→	→	→	→	→	→	→	→	

[a] → denotes residues identified as amino acid phenylthiohydantoins by GLC, TLC and as amino acids after back-hydrolysis. ← denotes identification by amino acid analysis after carboxypeptidase treatment. * indicates that quantitative data are given for these peptides in the supplementary material.

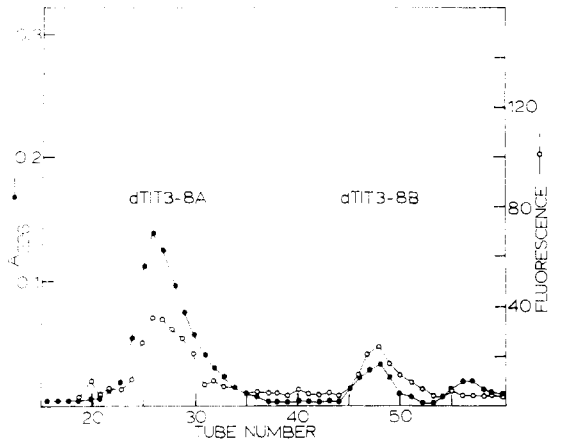

FIGURE 16: Gel filtration of peptide dT1T3 after eight cycles of Edman degradation. The degraded peptide was obtained from 120 nmol of dT1T3. Sephadex G-25 column (1.5 × 90 cm) equilibrated previously with 50 mM acetic acid. Column effluent was monitored by measurement of the absorbance at 226 nm (●) and by the fluorescamine reaction (○).

Table VIII in supplementary material) has the sequence Lys[30]-Arg-Asp(Thr,Tyr), suggesting that this penta peptide was penultimate to the carboxy terminal peptide dT1-CH4 with the sequence Ala-Ala-Thr-Pro-Arg[39] (Table XI). Secondly, it is known that chymotrypsin occasionally cleaves His-X peptide bonds, and the glycopeptide dT1 apparently is cleaved at this histidyllysine linkage to produce the pentapeptide dT1-CH5 and peptide dT1-CH2 with a carboxy-terminal histidine. dT1-CH2 is presumably a mixture of peptides containing residues 21–29 and peptides containing residues 21–34, but no amino acid (including histidine) was released after digestion with carboxypeptidase A. We assume that this is due to steric hindrance by vicinal oligosaccharide units.

Sequence of Residues 32–64. Tables XI and XIII show the amino acid sequence of the peptides derived from this segment of the glycophorin A molecule. The sequence of the peptide T3A was identical with that of peptide T3B, consistent with their identical amino acid compositions. The chymotryptic peptide CH3 was cleaved with trypsin to give three fragments (CH3-T1, CH3-T2, and CH3-T3) as described (Figure 11 and Table VI in supplementary material). The peptide CH3-T1 was identical with the peptides T3A and T3B in amino acid composition and sequence. The amino acid sequence of peptide CH3-T2 was determined as Ala-Ala-Thr-Pro-Arg[39], representing the amino terminal segment of CH3. This sequence was also identical with the carboxy-terminal part of T5 (Table XI). The smallest peptide of the three is CH3-T3, a tripeptide with the sequence Val[62]-Gln-Leu (Table VI in supplementary material) which represented the carboxy-terminal segment of CH3. When glycopeptide CH2 was hydrolyzed with trypsin, only two peptides are obtained from the digest by gel filtration

TABLE XII: Amino Acid Composition of the Degraded Peptides after Eight Cycles of Edman Degradation of Peptide dT1T3.

	peptides			
	dT1T3	dT1T3-7[a]	dT1T3-8A	dT1T3-8B
Asp	2.0[b]	2.0	2.1	1.0
Thr	2.3	2.0	2.2	1.0
Ser	3.0	1.1	1.5	0.3
Glu	1.1	0.7	1.0	0.2
Pro	[c]			
Gly			0.3	0.3
Ala			0.2	
Val			0.2	
Met				
Ile	1.0		0.2	
Leu				
Tyr	0.8			
Phe				
His	0.9	1.3	1.2	1.0
Lys	1.0	0.5	0.3	0.7
Arg	0.3	0.3	0.3	0.3

[a] The degraded peptide after seven cycles of Edman degradation. [b] No correction has been made for losses of amino acids during hydrolysis. Data expressed as residues/peptide. [c] No entry indicates 0.1 or less.

on Sephadex G-50. As shown in Table VII (supplementary material), the smaller of the two peptides, CH2-T2, contains one residue each of aspartic acid, threonine, and tyrosine. Thus the peptide CH2-T2 represents the carboxy terminus of CH2 as well as the amino terminus of T5 (cf. Table XI). It should be noted that the peptide CH2-T2 was obtained in low yield (about 40%), probably due to the heterogeneity of the original CH2 peptide, which appears to be a mixture of peptide 1-29 and peptide 1-34. Since the sequence of neither peptide T3A nor T3B could be determined completely to the carboxy-terminal end, peptide T3B was hydrolyzed with thermolysin to produce small peptides (Figure 14 and Table IX in supplementary material). Peptide T3-TH3 corresponded to the carboxy-terminal region of T3B, and its sequence was determined as shown in Table XI. Three residues of T3B, namely the 5th, 8th, and 11th, could not be determined as Pth-amino acids, but were identified as dansyl-amino acids indicating that residues 44, 47, and 50 in the sequence are glycosylated.

Sequence of the Hydrophobic Peptide Region (Residues 62–98). Prior to this study attempts had been made by Segrest et al. (1972) to deduce the amino acid sequence of this segment of glycophorin A from studies of insoluble peptides obtained by trypsin digestion of the entire sialoglycoprotein fraction. A provisional sequence was obtained, but multiple Pth-amino acids were found at several positions, and it was not clear whether this was due to microheterogeneity of the polypeptide chain or due to the presence of contaminating peptides. Since that study, we have succeeded in purifying a single hydrophobic peptide from glycophorin A by a countercurrent distribution approach which was used to sequence residues 62–85 (Furthmayr et al., 1978). The results obtained from this study are given in Table XIII. Positions 62–74 were confirmed by manual Edman degradation of CB1-T2 (residues 62–81) obtained by digestion of CB1 (9–81) with trypsin. In addition, positions 82–98 in the sequence were confirmed by automated sequence analysis of CB2 (82–131) and the results were identical with earlier reported data (Segrest et al., 1972). In Table XIII the peptides are indicated which were used to establish overlaps. Peptide CH6, although quite contaminated as judged from the amino acid composition (cf. Table III), has the amino terminal sequence Ala-His-His-Phe-Ser-Glu, indicating that chymotrypsin cleaves the Leu^{64}-Ala bond.

Carboxy-Terminal Sequence (Residues 99–131). Two peptides designated T4A and T4B are identical in sequence except for an additional lysyl residue at the amino terminus in T4A. These two peptides also overlapped with CB2 and CH4 and thus represent the C-terminal region of glycophorin A. The order CH4-CH5 is indicated by the overlap of CH4 with T4A as indicated in Table XIII. Partial sequences including these overlaps for T4A, T4B, CH4, and the complete sequence of CH5 (for yields, see Table XVIII in supplementary material) are shown in Table XIII. In addition, a thermolysin fragment derived from CH5 (CH5-TH1; Table VII in supplementary material) was used to confirm the C-terminal sequence. It should be mentioned that none of the peptides derived from this segment of glycophorin A was glycosylated.

Carboxy Terminus of Glycophorin A. Small amounts of glutamine were released from intact glycophorin A by carboxypeptidase A, while neither carboxypeptidase B nor hydrazinolysis produced significant amounts of free amino acids from the same preparations. Small amounts of glutamine were also obtained from peptide CH5 with carboxypeptidase A which is consistent with the proposed carboxy-terminal sequence.

Discussion

Alignment of Peptides. The isolation of different proteolytic fragments which apparently were derived from the same regions of the polypeptide chain produced complex chromatographic elution patterns, but the amino acid composition of the peptides and sequence data indicate that a rather simple model can account for these results (cf. Figure 12). The alignment of the tryptic, chymotryptic, and CNBr-derived peptides was solved by the following considerations.

Both proteolytic digests generate a hydrophobic peptide (although only the trypsin-derived peptide was isolated and purified) which contains one of the two residues of methionine present in the original glycophorin A molecule. CNBr cleaves the methionine residue in this fragment. This was shown by tryptic digestion of CB1, which has a C-terminal homoserine, to give a hydrophobic peptide CB1-T2 containing the homoserine residue (Furthmayr et al., 1978). Since CNBr digestion of the intact molecule gives the octapeptide CB3, which has N-terminal residues identical with the intact molecule the second methionine is located in position 8 of the sequence. CB3 is therefore the N-terminal peptide. Since CB2 does not contain homoserine, the order of the CNBr derived peptides is CB3-CB1-CB2.

CB3 is contained in the large glycopeptides T1, T2, CH1, and CH2 as shown by CNBr digestion of these fragments. These peptides are therefore placed at the N terminus of glycophorin A. Peptides T2 and T5 are fragments of T1 as suggested by the sum of their amino acids. The nonglycosylated peptide T4 is contained in the C-terminal CNBr fragment CB2, as indicated by its high proline content, lack of carbohydrate, and is therefore placed at the C-terminal end of the polypeptide following T6. The overlap in amino acid sequence between T6 and CB2 has been shown previously in preliminary studies (Segrest et al., 1972). The only other peptide isolated from the tryptic digest is the glycosylated fragment T3. Its high glutamic acid content suggests that it is derived from the same segment as CH3. Since no carbohydrate has been found at the cytoplasmic side of the membrane, this fragment T3 should be located at the N-terminal side of the hydrophobic fragment T6. Thus the order T1-T3-T6-T4 and T2-T5-T3-T6-T4 is indicated. Similar reasoning leads to the order CH1-CH6-

CH4-CH5 and CH2-CH3-CH6-CH4-CH5. The order of the peptides CH4 and CH5 was only possible by sequence studies, since both fragments lack tyrosine and phenylalanine. Chymotrypsin cleaves the Leu^{118}-Ser bond within peptide T4. Sequence studies also suggested that a small tri- or tetrapeptide T7 has to be inserted between T6 and T4. However, this peptide has not been isolated.

Both tryptic and chymotryptic digestion of glycophorin A generated double sets of peptides from the glycosylated portion of the molecule, which seem to differ only in their carbohydrate content. These findings could be explained by the presence of two glycophorin molecules differing only in their degree of glycosylation which could influence their susceptibility to proteolytic cleavage. It has not been possible to separate these two forms and thus it remains to be seen whether T3A and T3B have to be placed within the sequence as T1-T3A-, T2-T5-T3A, T1-T3B- or T2-T5-T3B-. Since it has been shown that glycophorin A exists in two structurally different forms with distinct amino acid sequences at the amino-terminal end (Furthmayr, 1978b), it is possible that the heterogeneity discussed above is also correlated with glycophorin A of either M or N type. Preliminary studies have indicated that T1 and T2 fragments are obtained at about the same ratios for glycophorin A^M and A^N (Furthmayr, 1978b, and unpublished results).

Earlier preliminary reports (Marchesi et al., 1972; Jackson et al., 1973) presented models for the arrangement of the glycopeptide components which differ substantially from the more recent data presented here. As was pointed out above these studies were carried out on total sialoglycoprotein fractions which have since been shown to contain at least three different sialoglycopeptides (Furthmayr, 1978a).

The order of the peptides presented in Figure 12 is further supported by more indirect results. Glycopeptides are released by proteases from intact human red blood cells (Jackson et al., 1973) and these can be shown to correspond to T1 and T2 derived from glycophorin A and a glycopeptide T1 from glycophorin C (Furthmayr, 1978a). Recent immunochemical and electron microscopic studies (Cotmore et al., 1977) further support the transmembrane nature of glycophorin A which had been postulated earlier on the basis of labeling studies (Bretscher, 1971). Antibodies specific to T4 (or CH4) conjugated to ferritin localize along the inner surfaces of red cell membranes and thus provide evidence for the localization of T4 at the carboxy terminal end of the polypeptide chains.

Amino Acid Sequence. Our attempts to determine the complete amino acid sequence of human erythrocyte glycophorin A have been complicated by a number of technical pitfalls. As described above double sets of peptides were generated by both tryptic and chymotryptic digestion of the glycosylated portion of this molecule. Once this was realized it was possible to use these overlapping sets to our advantage in confirming the sequences after we found a suitable method to sequence through the glycosylated residues. The Edman degradation technique described by Peterson et al. (1972) was modified slightly for application to the glycopeptides because the yield of the amino acid phenylthiohydantoins using standard techniques was significantly lower for glycopeptides than for the nonglycosylated forms. The important point in the modification was the very careful drying of the samples prior to the trifluoroacetic acid cleavage step because highly glycosylated peptides tend to retain more water molecules. Since such glycosylated peptides are quite soluble in aqueous buffers but insoluble in organic solvents, extraction with a large volume of organic solvent could be used to remove excess phenyl isothiocyanate without significant loss of peptide. Therefore, extraction with 1.0 mL of benzene and 0.5 mL of ethyl acetate was used prior to cleavage. The phenylthiocarbamoylated glycopeptides could be dried effectively in a heated desiccator after the ethyl acetate extraction, and the sample at this step must be white and amorphous but never syrupy. Despite these precautions, the yield of Pth-amino acids from glycopeptides was slightly lower than for nonglycosylated peptides.

Some of the problems encountered during the isolation and purification of the glycopeptides and the rationale for our final approach have already been given in the introductory section and in Results. In sequencing the glycopeptide regions, Edman degradation proceeded normally through O-glycosylated amino acid residues. There are a number of reports which describe amino acid sequence determinations of O-glycosylated peptides (Carlsen et al., 1973; Baenzinger & Kornfeld, 1974; Brewer et al., 1974) but the O-glycosylated amino acid phenylthiohydantoins have not been directly identified. When only one amino acid residue in a peptide is O-glycosylated, that residue can be determined from the amino acid composition (Brewer et al., 1974). However, if more than two residues are O-glycosylated it is not possible to determine O-glycosylated amino acids by identification of Pth-amino acids alone. Since both serine and threonine are glycosylated in the glycophorin A molecule, a different technique was needed to identify the attachment sites of the individual oligosaccharides. Although alkaline borohydride treatment of glycopeptides appeared promising, since this treatment had been used successfully to release O-glycosidically linked oligosaccharides, we could not overcome a major shortcoming inherent in this technique namely random peptide bond cleavage. This occurred too frequently to allow the isolation of pure peptides suitable for Edman degradation. The release of oligosaccharides from glycopeptides was insufficient at conditions below 0.2 M NaOH, at which random cleavage was not apparent. Another disadvantage of this method was that $NaBH_4$ was not satisfactory for the complete reduction of dehydrothreonine to α-aminobutyric acid. We also attempted to cleave O-glycosidically linked oligosaccharides by using methanolysis under mild acid conditions, since it is well known that O-glycosidic linkages are labile under these conditions. We found that the terminal group was not blocked during the acidic methanolysis and peptide bond cleavage was negligible under the conditions used, but the yield of O-glycosylated Pth-threonine was rather low for unknown reasons.

Subtractive Edman degradation also has been used to determine O-glycosylated amino acids (Frangione & Wolfenstein-Todel, 1972; Carlsen et al., 1973). The attachment sites of oligosaccharides were determined by the loss of galactosamine in the degraded peptide (Carlsen et al., 1973). Although this procedure appeared attractive, the results on a highly glycosylated peptide such as T1-CB1 were not conclusive enough even after the degraded peptide was separated from the phenylthiohydantoin derivative by gel filtration. This method seems to be applicable to glycopeptides in which only a few residues are glycosylated. Among all the methods we have tried the most effective was the simultaneous use of the dansyl–Edman and direct Edman degradation.

Although glycosylated Pth-amino acids cannot be analyzed directly, the amino acids in glycosylated positions can be identified as serine or threonine after dansylation. This method not only gave the most conclusive results, but it was also the simplest procedure for the identification of O-glycosylated residues. Since dansylserine and dansylthreonine are relatively more soluble in ethyl acetate than other dansyl derivatives, ethyl acetate extraction is an effective way to increase the sensitivity of detection for these residues. It is obvious that the

interpretation of the data and the identification of serine and threonine as glycosylated residues rely heavily on sensitive methods for the detection of Pth-serine and -threonine since the absence of such residues in the sequence during direct Edman degradation is only suggestive evidence that they are glycosylated. The GLC and TLC methods used were found satisfactory, provided sufficient peptide material was available. It is important to point out that such data are semiquantitative only and do not necessarily allow us to determine the degree of glycosylation at every position. To compensate for this additional confirmation of glycosylated positions was sought, whenever possible, by sequencing peptides from the same region which were prepared by different means.

Quantitation of the yields of Pth-amino acids by GLC for nonglycosylated positions was used throughout this study. However, there is one region within the sequence around the N-glycosylated asparagine[29] which defied all attempts for quantitation. Peptides containing this glycosylated residue were difficult to purify and Edman degradation apparently does not proceed normally through N-glycosylated asparagine. The amide group of asparagine, which is in N-glycosidic linkage with *N*-acetylglucosamine, seems to interfere with the normal cyçlization during Edman degradation, although no systematic study has been done on this problem. In the case of glycophorin A Edman degradation continued with a yield of about 20–30% after the N-glycosylated asparaginyl residue. This was high enough to continue the degradation since sufficient amounts of glycopeptide were available. Because of the low yields and because histidine[29] was not directly analyzed, it is possible that there is some ambiguity in the proposed sequence from asparagine-26 to lysine-30. Our findings are, however, consistent with the suggestion that N-glycosylation occurs only in the sequence

$$\overset{*}{\text{Asn}}\text{-X-Ser(or Thr)}$$

(Spiro, 1974) and no exception has been found thus far.

The data presented here suggest that human red blood cells contain a unique sialoglycoprotein which is under strict genetic control allowing for the expression of two nearly identical polypeptide chains, A^M and A^N. Only rarely have variants been found (Race & Sanger, 1975; Furthmayr, 1978b). There is a second glycoprotein (glycophorin B) associated with human red cell membranes, which shares the amino-terminal structure of glycophorin A^N (Furthmayr, 1978a,b). Correlated with the expression of blood groups M and N are differences in other serological activities, such as reactivity of red cells with lectins like *Vicia graminea*. It is likely that posttranslational modification of the glycophorin A and/or B molecules by glycosyltransferases and their topographical arrangement in the membrane will ultimately be found responsible for these activities.

The transmembrane nature of glycophorin A has been discussed previously in other contexts (Tomita & Marchesi, 1975; Cotmore et al., 1977). The significance of this arrangement as well as the function of these highly glycosylated molecules are not known at present and will provide a challenge for future work on erythrocytes as well as other cells.

Acknowledgment

We are grateful to Mrs. Manyee Tang and Mr. Bruce Proctor for expert technical assistance.

Supplementary Material Available

Additional information on amino acid sequence determinations (including Figures 3, 4, 10, 11, 13, 14, and 15 and Tables VI–X and XIV–XVIII) (20 pages). Ordering information is given on any current masthead page.

References

Ambler, R. P. (1972) *Methods Enzymol. 25,* 143.
Baenziger, J., & Kornfeld, S. (1974) *J. Biol. Chem. 249,* 7270.
Blumenfeld, O. O., & Zvilichovsky, B. (1972) *Methods Enzymol. 28,* 245
Bornstein, P. (1970) *Biochemistry 9,* 2408.
Bretscher, M. S. (1971) *Nature (London), New Biol. 231,* 229.
Bretscher, M. S., & Raff, M. C. (1975) *Nature (London) 258,* 43.
Brewer, H. P., Shulman, R., Herbert, P., Ronan, R., & Wehrly, K. (1974) *J. Biol. Chem. 249,* 4975.
Carlsen, R. B., Bahl, O. P., & Swaminathan, N. (1973) *J. Biol. Chem. 248,* 6810.
Cotmore, S. F., Furthmayr, H., & Marchesi, V. T. (1977) *J. Mol. Biol. 113,* 539.
Cuatrecasas, P., & Illiano, G. (1971) *Biochem. Biophys. Res. Commun, 44,* 1786.
Dahr, W., Uhlenbruck, G., Janssen, E., & Schmalisch, R. (1976) *Blut 32,* 171.
Frangione, B., & Wolfenstein-Todel, C. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69,* 3673.
Fukuda, M., & Osawa, T. (1973) *J. Biol. Chem. 248,* 5100.
Furthmayr, H. (1977) *J. Supramol. Struct. 7,* 121.
Furthmayr, H. (1978a) *J. Supramol. Struct.* (in press).
Furthmayr, H. (1978b) *Nature (London) 271,* 519.
Furthmayr, H., & Marchesi, V. T. (1976) *Biochemistry 15,* 1137.
Furthmayr, H., Tomita, M., & Marchesi, V. T. (1975) *Biochem. Biophys. Res. Commun. 65,* 113.
Furthmayr, H., Galardy, R. E., Tomita, M., & Marchesi, V. T. (1978) *Arch. Biochem. Biophys. 185,* 21.
Gray, W. R. (1972) *Methods Enzymol. 25,* 121.
Hamaguchi, H., & Cleve, H. (1972) *Biochim. Biophys. Acta 278,* 271.
Jackson, R. L., Segrest, J. P., Kahane, I., & Marchesi, V. T. (1973) *Biochemistry 12,* 3131.
Kahane, I., Furthmayr, H., & Marchesi V. T. (1975) *Biochim. Biophys. Acta 426,* 464.
Konigsberg, W. (1972) *Methods Enzymol. 25,* 326.
Kulbe, K. D. (1974) *Anal. Biochem. 59,* 564.
Laemmli, U. K. (1970) *Nature (London) 227,* 680.
Marchesi, V. T., & Andrews, E. P. (1971) *Science 174,* 1247.
Marchesi, V. T., Tillack, T. W., Jackson, R. L., Segrest, J. P., & Scott, R. E. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69,* 1445.
Marchesi, V. T., Furthmayr, H., & Tomita, M. (1976) *Annu. Rev. Biochem. 45,* 667.
Mueller, T. J., Dow, A. W., & Morrison, M. (1976) *Biochem. Biophys. Res. Commun. 72,* 94.
Peterson, J. D., Nehrlich, S., Oyer, P. E., & Steiner, D. F. (1972) *J. Biol. Chem. 247,* 4866.
Pisano, J. J., & Bronzert, T. J. (1969) *J. Biol. Chem. 244,* 5597.
Race, R. R., & Sanger, R. (1975) *Blood Groups in Man,* pp 92 and 463, Blackwell, London.
Reinhold, V. N. (1972) *Methods Enzymol. 25,* 244.
Sauer, R. T., Niall, H. D., Hogan, M. L., Keutmann, H. T., O'Riodan, J. L. H., & Pott, J. T., Jr. (1974) *Biochemistry 13,* 1994.

Schroeder, W. A. (1972) *Methods Enzymol. 25,* 138.
Segrest, J. P., Jackson, R. L., Andrews, E. P., & Marchesi, V. T. (1971) *Biochem. Biophys. Res. Commun. 44,* 390.
Segrest, J. P., Jackson, R. L., Marchesi, V. T., Guyer, R. B., & Terry, W. (1972) *Biochem. Biophys. Res. Commun. 49,* 964.
Segrest, J. P., Jackson, R. L., Kahane, I., & Marchesi, V. T. (1973) *Arch. Biochem. Biophys. 155,* 167.
Smithies, O., Gigson, D., Fannig, E. M., Googfliesch, R. M., Gilman, J. G., & Ballantyne, D. L. (1971) *Biochemistry 10,* 4912.
Spiro, R. (1974) *Adv. Protein Chem.,* 349.
Steck, T. L. (1974) *J. Cell Biol. 62,* 1.
Springer, G. F., Nagai, Y., & Tegtmeyer, H. (1969) *Biochemistry 5,* 3254.
Summers, M. R., Smythers, G. W., & Oroszlan, S. (1973) *Anal. Biochem. 53,* 624.
Tomita, M., & Marchesi, V. T. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72,* 2964.
Udenfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W., & Weigele, M. (1972) *Science 178,* 872.
Warren, L. (1959) *J. Biol. Chem. 234,* 1971.
Winzler, R. J. (1969) in *Red Cell Membrane, Structure and Function* (Jamieson, G. A., & Greenwalt, T. J., Eds.,) p. 157, Lippincott, Philadelphia, Pa.

Circular Dichroism and Magnetic Circular Dichroism of Iron–Sulfur Proteins†

P. J. Stephens,* A. J. Thomson,‡ J. B. R. Dunn, T. A. Keiderling,§ J. Rawlings,# K. K. Rao, and D. O. Hall

ABSTRACT: Circular dichroism (CD) and magnetic circular dichroism (MCD) spectra are reported for the 2-Fe ferredoxins from *Pseudomonas putida* and *Spirulina maxima, Chromatium* HIPIP, the 4-Fe ferredoxin from *Bacillus stearothermophilus,* and the 8-Fe ferredoxin from *Clostridium pasteurianum.* The spectral range spans the near-infrared, visible, and near ultraviolet. In all cases except oxidized 2-Fe ferredoxins, electronic absorption is observed continuously from less than 5000 cm^{-1} to above 30 000 cm^{-1}. The CD spectra of the two 2-Fe ferredoxins are similar. In contrast, the CD of the 4-Fe and 8-Fe proteins, for a given 4-Fe cluster oxidation level, varies considerably with protein. MCD is less sensitive to protein environment than is CD. In the 2-Fe proteins, MCD at 5 T is appreciably smaller than the CD; in the 4-Fe and 8-Fe proteins, MCD and CD are comparable in magnitude. Both CD and MCD are more highly structured than the corresponding absorption spectra. The CD and MCD spectra reported provide a broader base than heretofore available for the characterization of iron-sulfur proteins containing 2-Fe and 4-Fe clusters and for the evaluation of electronic structural models for these clusters.

Iron-sulfur proteins are a class of nonheme iron proteins of widespread occurrence and diverse function (Lovenberg, 1973, 1977). Iron-sulfur proteins can be classified according to the number of iron atoms present. At this time, proteins containing 1,[1] 2, 4, and 8 Fe atoms have been extensively characterized and shown to contain three fundamental types of iron-sulfur cluster: $Fe(SR)_4$, $Fe_2S_2(SR)_4$, and $Fe_4S_4(SR)_4$. S and SR denote inorganic (labile) sulfide (S^{2-}) ions and protein-bound cysteine moieties (SR^-), respectively. Consistent with their function as redox proteins, the iron-sulfur clusters can attain multiple oxidation states. The 1-Fe, $[Fe(SR)_4]^{n-}$ cluster occurs in oxidized and reduced rubredoxins with $n = 1$ and 2, respectively, corresponding to Fe(III) and Fe(II) oxidation states. The 2-Fe, $[Fe_2S_2(SR)_4]^{n-}$ cluster occurs in oxidized and reduced (2-Fe) ferredoxins with $n = 2$ and 3, respectively, corresponding to $[Fe(III)_2]$ and [Fe(III), Fe(II)] formal oxidation states. The 4-Fe, $[Fe_4S_4(SR)_4]^{n-}$ cluster occurs in super-reduced, reduced, and oxidized HIPIP, and in reduced, oxidized, and superoxidized 4-Fe and 8-Fe ferredoxins with $n = 3$, 2, and 1, respectively, corresponding to [Fe(III), $Fe(II)_3$], $[Fe(III)_2, Fe(II)_2]$, and $[Fe(III)_3, Fe(II)]$ formal oxidation states. The definition of these structural and electronic characteristics has been especially aided by X-ray crystallographic studies on *Clostridium pasteurianum* rubredoxin (Watenpaugh et al., 1973), *Chromatium* HIPIP (Carter et al., 1974a,b; Freer et al., 1975), and *Micrococcus aerogenes* ferredoxin (Adman et al., 1973) and by the synthesis of convincing active-site analogue compounds (Holm & Ibers, 1977).

In this paper we discuss the electronic spectroscopy of iron-sulfur proteins. Electronic absorption spectroscopy in the visible-near-UV spectral region has been routinely used in the study of iron-sulfur proteins (Lovenberg, 1973, 1977; Tsibris

† From the Department of Chemistry, University of Southern California, Los Angeles, California 90007 (P.J.S., A.J.T., J.B.R.D., T.A.K., and J.R.), and the School of Biological Sciences, University of London King's College, London, England (K.K.R. and D.O.H.). *Received April 27, 1978.* This work has been supported by grants from the National Institutes of Health, National Science Foundation, and Science Research Council (U.K.). A travel grant (to A.J.T.) from the Royal Society and a Science Research Council Senior Visiting Fellowship (P.J.S.) are gratefully acknowledged.

‡ Permanent address: School of Chemical Sciences, University of East Anglia, Norwich, England.
§ Present address: Department of Chemistry, University of Illinois at Chicago Circle, Chicago, Illinois.
# Present address: Department of Biochemistry, University of Wisconsin, Madison, Wisconsin.

[1] Abbreviations used: CD, circular dichroism; MCD, magnetic circular dichroism; 2-Fe, 4-Fe, and 8-Fe, two-, four-, and eight-iron, respectively.

0006-2960/78/0417-4770$01.00/0