Journal of Chromatography, 443 (1988) 267–284
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 1349

STRUCTURAL CHARACTERIZATION OF GLYCOPROTEINS

FRANK W. PUTNAM* and NOBUHIRO TAKAHASHI*
Department of Biology, Indiana University, Bloomington, IN 47405 (U.S.A.)

SUMMARY

The structural characterization of glycoproteins by high-performance liquid chromatography (HPLC) is often difficult because of microheterogeneity of their oligosaccharide groups. To investigate this phenomenon, a series of human plasma glycoproteins of known amino acid sequence and carbohydrate structure was subjected to comparative study by HPLC. Methods for the isolation of singly glycosylated glycopeptides were developed. The chromatographic behavior in reversed-phase HPLC of mono-, di-, and multiglycosylated glycopeptides was compared with that of unglycosylated peptides of known amino acid sequence. Glycopeptides tended to be eluted earlier than non-glycopeptides, but the major factor governing retention was hydrophobicity. As shown by comparative study of five plasma glycoproteins by four chromatographic methods, microheterogeneity of the oligosaccharides affects chromatographic characterization of glycoproteins. In anion-exchange HPLC, carbohydrate charge heterogeneity is reflected by the broadening or asymmetry of peaks. Hydroxyapatite chromatography is useful for the purification of several forms of ceruloplasmin and other glycoproteins. The effect of carbohydrate is small in reversed-phase chromatography, but some proteins are denatured by the stringent conditions. Carbohydrate does not have much effect in hydrophobic interaction chromatography, which is more gentle. Because the chromatographic behavior of a glycoprotein may vary significantly with the procedure applied, several types of HPLC methods should be used for the characterization of glycoproteins.

INTRODUCTION

Glycoproteins of many kinds, both soluble and insoluble, are ubiquitous in nature; carbohydrate content is the only common denominator. Many insoluble glycoproteins have major structural functions in microbial cell walls, in the intercellular matrix of animal tissues, in the membranes of cells, and as the collagenous fabric of skeleton, tendons, and skin. The chemical structures of the carbohydrate components of many of these structural glycoproteins are known[1,2]. Yet, because of their insolubility and heterogeneity, these glycoproteins have not been well studied by methods

* Present address: Corporate Research and Development Laboratory, Toa Nenryo Kogyo K.K., Iruma-gun, Saitama-ken 354, Japan.

0021-9673/88/$03.50

of protein chemistry, such as sequence analysis. However, the extracellular fluids of higher organisms contain a great variety of soluble glycoproteins that have known structures and specific biological properties of great interest, such as hormones, enzymes, and antibodies. Human plasma proteins are by far the best characterized group of soluble secreted glycoproteins[3] and thus are good examples for study by high-performance liquid chromatography (HPLC). For this reason and also because they are the major focus of research in our laboratory, we will discuss only the structural characterization of human plasma glycoproteins. The main theme is that, because of the combination of several causes of carbohydrate heterogeneity, any particular glycoprotein may have a large number of forms. This leads to a paradox: the more the resolving power of a chromatographic column is increased, the more the separation may become ambiguous[4].

Human plasma proteins comprise the most extensively characterized group of glycoproteins with respect both to their polypeptide structure[3] and also the number, location, structure, and functions of their oligosaccharides[5]. About 100 plasma proteins have been isolated. As of 1986, the amino acid sequence had been determined directly by methods of protein chemistry for almost 40 and had been deduced from the nucleotide sequence for many more[3]. Of course, the presence and kind of carbohydrate cannot be determined from the gene sequence, since addition of carbohydrate is a co- or post-translational event, followed by cytoplasmic processing[5–7]. This emphasizes the importance of new methods of protein chemistry, such as HPLC, for the study of glycoproteins. Another major advance has been the advent of techniques for rapid structural analysis of carbohydrates, such as the use of 360 and 500 MHz ^{1}H NMR[8–10]. This has facilitated determination of the complete structure of hundreds of oligosaccharides from many plant and animal sources[1,6].

Three major problems remain. First, most glycoproteins are multiglycosylated, that is, they have oligosaccharides at two or more positions in the amino acid sequence. Thus, the results are ambiguous, unless a glycopeptide representing just a single position is analyzed. Second, the oligosaccharide at even a single position may be heterogeneous or may be missing from some molecules. Using examples from our own work, this paper will show that HPLC has contributed greatly to the solution of the first two problems. A third problem that has been more difficult to resolve by HPLC is the separation of carbohydrate variants, that is, of protein molecules having identical amino acid sequence but differing in oligosaccharide structure or number. This is a common phenomenon that impedes the purification and characterization of glycoproteins. Examples of this will be given for several plasma glycoproteins.

STRUCTURE, BIOSYNTHESIS, AND HETEROGENEITY OF OLIGOSACCHARIDES OF PLASMA GLYCOPROTEINS

In 1984, Baenziger[5] listed more than 20 plasma glycoproteins for which the oligosaccharide structure had been completely characterized. The number now is at least 30, including several new ones from our laboratory, described later, and just published[11]. Virtually all plasma glycoproteins except serum albumin are glycosylated. Most have several oligosaccharides per polypeptide chain. The carbohydrate content of the well-studied plasma glycoproteins ranges from as low as 2% (w/w) (IgG, apolipoprotein E) to as high as 45% (α_1-acid glycoprotein)[3,5]. With rare ex-

ceptions, the carbohydrate is of two types, either or both of which may be present: N-linked (asparagine-linked) glucosamine (GlcN) oligosaccharides and O-linked (serine or threonine-linked) galactosamine (GalN) oligosaccharides. The glucosamine type is much more common and has a greater potential for heterogeneity. Because much more is known about the biosynthesis, processing, and signal sequence of glucosamine oligosaccharides than for the galactosamine type, emphasis is given here to the glucosamine type.

Structure and heterogeneity of GlcN oligosaccharides

All asparagine-linked oligosaccharides fall into three classes: high mannose, hybrid, and complex[5–9]. All three classes share an identical pentasaccharide core structure, illustrated in Fig. 1. The three classes differ in their branching structure and in the number and kind of constituents on the branches. Many modifications of the three structural types exist that differ in the number of branches (two to four or more), the addition of peripheral sugars, such as fucose, and the number of sialic acid residues[5–9]. These modifications affect the hydrophilicity. Also, the presence of sialic acid affects the charge.

The presence of the common core reflects the fact that all asparagine-linked oligosaccharides originate from a common biosynthetic intermediate, a glucosylated high-mannose oligosaccharide that is attached to a dolichol carrier lipid[5,6]. Co-translational transfer of this large precursor molecule en bloc to a nascent peptide is

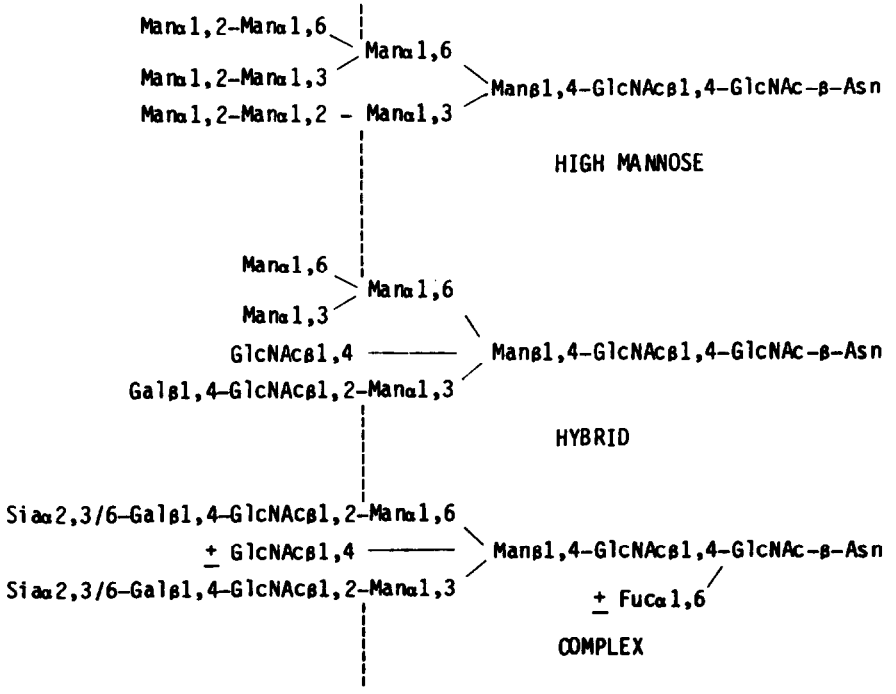

Fig. 1. The three classes of asparagine-linked oligosaccharides. All three classes have the identical core structure to the right of the dashed line. In each class, variations may occur because of the addition of other peripheral sugars or of more branches (from Baenziger[5]).

followed by a series of modifications by glycosidases in the endoplasmic reticulum and in the Golgi apparatus. This post-translational processing produces a large number of intermediates, that are classified into the three types. Which intermediate ends up as the final oligosaccharide at a given site in the sequence of a protein is determined by complex unknown factors, such as the stage of folding of the protein, the length of the polypeptide chain, accessibility to and availability of specific glycosidases, etc. As a result, a single protein, such as IgG, may have a large number of carbohydrate structures. In fact, Parekh *et al.*[12] found at least 30 different complex-type biantennary structures on human IgG, although it has only a single oligosaccharide per heavy chain.

All secreted glycoproteins that have GlcN oligosaccharides undergo the same processing; thus, carbohydrate heterogeneity is a characteristic of most plasma proteins. The heterogeneity is amplified in multiglycosylated proteins such as IgM, which has five GlcN oligosaccharides on each heavy chain and consists of a pentamer of ten heavy chains. To complicate matters further, post-translational modifications of the oligosaccharides that affect the charge may occur, *e.g.*, phosphorylation or sulfation of mannose residues, O-acetylation of sialic residues, and desialiation[6]. Thus, a single glycoprotein having a unique amino acid sequence may consist of a set of molecules having a great variety of carbohydrate structures that differ in size, charge, and hydrophilicity. This greatly complicates the determination of the carbohydrate structure and adversely affects many methods for structural characterization of proteins, including electrophoresis, X-ray diffraction analysis, and HPLC.

Structural requirements

Glucosamine oligosaccharides are always linked to a tripeptide signal sequence which is Asn-X-Ser/Thr; X may be any amino acid, but proline and aspartic acid are unfavorable[13,14]. Several chemical studies and statistical analyses of the conformational and structural requirements of the peptide acceptor sequence have been made[13,14]. Although the requirement for the acceptor sequence is absolute, glycosylation of the asparagine does not occur at 100% efficiency in systems known to have the capacity for N-glycosylation. In one computer analysis of the sequence database, Mononen and Karjalainen[13] found that only about 70% of some 200 potential sites in 105 proteins were glycosylated. The method of analysis could not detect partial glycosylation. As might be expected, study of the predicted secondary structures, both in glycosylated and non-glycosylated sites, showed that most occurred on β-turns (70%) or β-sheets (20%) and only 10% in helical conformations.

Very little has been learned about the three-dimensional structure of oligosaccharides from studies of the crystal structure of glycoproteins because so few have been done, and these have been mainly plasma glycoproteins. In the X-ray structure of α_1-antitrypsin, the three GlcN oligosaccharides are disordered, probably because of their heterogeneity; they are all in bends in the polypeptide chain protruding from the surface of the molecule[15]. In human and rabbit IgG the single GlcN carbohydrate at Asn-297 is at a sharp bend and acts as a spacer to stabilize the two adjacent C_H2 domains[12,16].

Galactosamine oligosaccharides

Much less is known about the mechanism of biosynthesis and the structural

requirements for linkage of galactosamine oligosaccharides. Galactosamine is O-linked to serine or threonine in variations of the general structure Siaα2,3-Galβ1,3-(Siaα2,6)GalNAcα. No acceptor sequence for GalN oligosaccharides has been clearly identified, but the sequence around the site is rich in proline and must have a conformation accessible to glycosylation. Galactosamine oligosaccharides tend to occur at or near the N-terminus of polypeptides or to cluster in serine- or threonine-rich regions that act as hinges. One example is the cluster of 3–5 GalN units in the hinge of human IgD[17].

The GalN oligosaccharides, unlike their GlcN counterparts, are believed to be predominantly synthesized in the Golgi apparatus by the sequential addition of sugars from their nucleotide derivatives by glycosyl transferases[5]. For this reason and because of their small size, the structural variety and, thus, the possible heterogeneity is much less than for GlcN oligosaccharides. The structures of GalN oligosaccharides differ mainly in the degree of sialylation; for example, four species of GalN oligosaccharides, differing in the sites or number of sialic acid residues, are present in close proximity in the IgD hinge region[18].

POLYPEPTIDE AND CARBOHYDRATE STRUCTURE OF REPRESENTATIVE PLASMA GLYCOPROTEINS

Determination of the structure of glycoproteins was often difficult for the protein chemist prior to the development of HPLC methods. The major difficulties lay in: (i) the heterogeneity of the carbohydrate chains, (ii) the variety of overlapping glycopeptides resulting from incomplete enzymatic digestion due to steric hindrance of the sugar chains, (iii) problems in purification of a series of similar large glycopeptides, (iv) technical problems of amino acid sequence determination of glycopeptides[19]. Although the complete amino acid sequence might be attainable, the carbohydrate structure at best was an average over all the sites of a multiglycosylated protein.

From the above it is obvious that for both the protein chemist and the carbohydrate chemist it is desirable to isolate a glycopeptide of known sequence with a single oligosaccharide attached. To achieve this objective, we developed reversed-phase HPLC methods for preparative isolation of monoglycosylated glycopeptides[19–22]. We applied these to a number of plasma glycoproteins for which we determined the complete amino acid sequence, using HPLC to purify both glycopeptides and non-glycopeptides[23–28], and we gave the glycopeptides to collaborators who determined the carbohydrate structure[18,29,30]. We will only highlight the results, because all of the amino acid sequences have been published[17,22–28]; likewise, the carbohydrate structures have either been published[11,18,29] or are in manuscript form[30].

Immunoglobulins

Immunoglobulins offer an excellent example of a family of homologous glycoproteins that differ greatly in their carbohydrate structure and linkage sites[31]. Fig. 2 shows a schematic diagram, presenting the oligosaccharides in all the human immunoglobulin classes. It includes work of our laboratory and others. Normally, all the sugar is on the constant region of the heavy chains, but the carbohydrate content

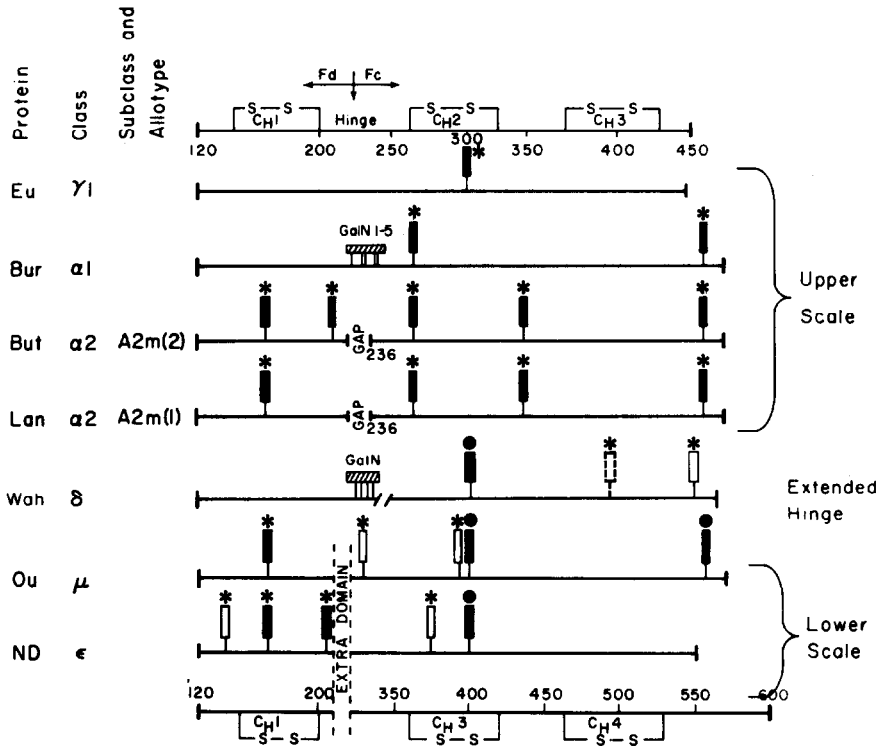

Fig. 2. Oligosaccharides of human heavy chains. Eu, Bur, etc., denote individual myeloma proteins for which the amino acid sequence has been reported[31,32]. Vertical rectangles denote glucosamine oligosaccharides. Shading indicates that these have homologous positions in two or more chains. Solid circles denote the mannose-rich type and asterisks the complex type of GlcN glycan. In the δ chain, the dashed rectangle shows where the oligosaccharide is present on only about half the IgD molecules. The numbers in the upper and lower scales give the residue positions in the chains, but the extra domain ($C\mu 2$ and $C\delta 2$, respectively) has been omitted in the μ and δ chains (from Putnam[32]).

ranges from 3% in IgG to 13% for IgE[32]. Correspondingly, the number of oligosaccharides of the glucosamine type on each heavy chain ranges from one in IgG to six in IgE. There are multiple GalN oligosaccharides (4–5) in the hinge regions of the δ chain of IgD and the α_1 chain of IgA1, but none in IgA2. Most of the glycans are of the GlcN type, and most of these have a complex structure. Some are in homologous positions in different chains; others have no counterpart in other chains. GlcN glycans may differ in structure, even when at homologous positions in several chains. Most of these results were obtained before the great heterogeneity at any given site was recognized by use of ^{1}H NMR and other methods of carbohydrate analysis[6–12]. As an example, we were able to show, by use of HPLC, that at one site (Asn-445) the GlcN oligosaccharide is present on only half the molecules of IgD[20].

GlcN glycopeptides of immunoglobulin D

Structural characterization of the carbohydrate in immunoglobulin D (IgD) offered both an excellent opportunity to test our methods and also some unique problems. The opportunity lay in the fact that this was the first example of a gly-

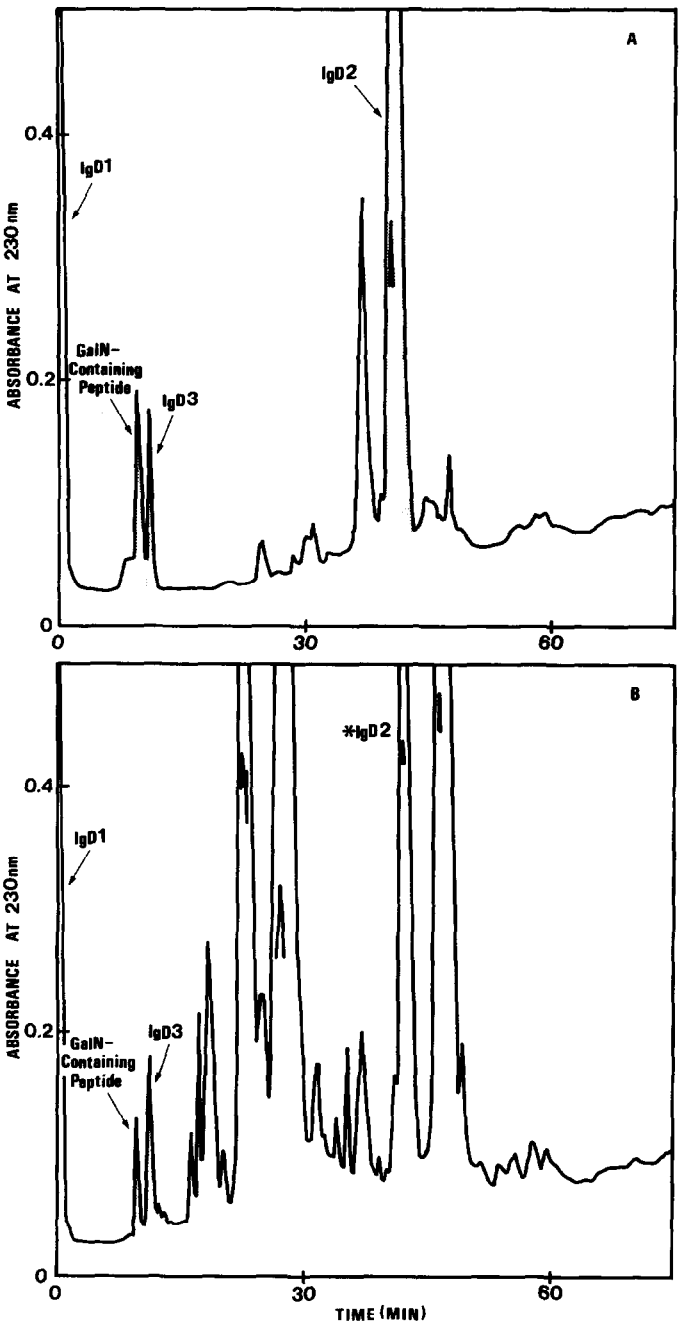

Fig. 3. Purification by reversed-phase HPLC of glycopeptides from a tryptic digest of the Fc fragment of IgD after preliminary fractionation by gel chromatography. Two carbohydrate-containing fractions (A and B) each obtained by gel chromatography were chromatographed by HPLC on a reversed-phase Synchropak RP-P column (25 × 1.0 cm I.D.). Peptides were eluted at a flow-rate of 4.16 ml/min with a linear gradient from 0 to 30% of 1-propanol, containing 0.1% trifluoroacetic acid during 75 min. The GalN-containing peptide is glycosylated at multiple sites. D1, D2 and D3 are different glycopeptides that have a single GlcN oligosaccharide of either the high-mannose or complex type of carbohydrate. The non-glycosylated peptide of D2 is designated *D2 (t_R of D2 = 40.5 min; t_R of *D2 = 43.0 min). Peaks containing glycopeptide are shaded (from Takahashi *et al.*[20]).

coprotein for which the entire amino acid sequence and carbohydrate structure had been determined by using a single source (a myeloma patient)[17,18,31]. The problems were that each δ heavy chain of IgD had three sites for GlcN oligosaccharides, and reports indicated that the hinge region had a variable number (from 3 to 7) GalN oligosaccharides. Because all the carbohydrate is in the Fc half of the δ chain, a tryptic digest of the papain-prepared Fc was made and was separated into five major fractions by initial gel chromatography on a Sephadex G-50 column[20]. Each fraction was chromatographed by reversed-phase HPLC. Three glycopeptides, D1, D2, and D3, that contain GlcN oligosaccharide and correspond, respectively, to Asn-354, Asn-445, and Asn-496, were separated by HPLC and also a GalN-peptide, derived from the IgD hinge region (Fig. 3). The D2 glycopeptide and a non-glycosylated peptide *D2 having the same amino acid sequence were obtained in about equal yield from adjacent gel chromatography fractions. Although the glycopeptide D2 was eluted 2.5 min earlier than the non-glycopeptide *D2, the difference in elution positions was far less than expected from the size of the carbohydrate structure. Nonetheless, this experiment shows that it is possible by use of a combination of gel chromatography and reversed-phase chromatography to separate two peptides that have the same polypeptide structure but differ by the presence or absence of carbohydrate.

GalN glycopeptides of IgD

In previous work[19], the GalN glycopeptides of the hinge region of the delta heavy chain of human IgD were fractionated by use of an RP-P column (25 × 1.0 cm I.D.) with a programmed gradient of 1-propanol, containing 0.1% heptafluorobutyric acid (HFBA). For several reasons, purification of the GalN glycopeptides was much more difficult than the purification of the GlcN glycopeptides: (i) the different nature of the GlcN and GalN oligosaccharides; (ii) the close proximity of multiple GalN glycans in the hinge region, compared to the wide separation of the GlcN glycans in the polypeptide sequence; (iii) the resistance to proteolysis of the GalN-rich hinge peptide; (iv) heterogeneity in the degree of glycosylation of the hinge peptide, which had multiple sites for O-glycosylation of serine or threonine. Thus, in the case of the IgD hinge region, a series of overlapping GalN peptides had to be sequenced. Nonetheless, we determined the complete amino acid sequence of the δ chain[17], and this was later completely confirmed by gene cloning and nucleotide sequence analysis[33]. Recently, in an HPLC study of the proteolytic cleavage of IgD, we isolated a 32-residue GalN-rich peptide[22]. This peptide presented no problem in sequence analysis, which showed that it spanned the hinge region from Ala-235 through Arg-266.

These results show that, although HPLC can be applied to the purification of GalN glycopeptides, difficulties arise in the case of multiply glycosylated proteins where the oligosaccharides are in close proximity. This situation is unlikely for GlcN oligosaccharides; it most often occurs in carbohydrate-rich proteins that have repeating sequences containing serine, threonine, and proline, for example, certain segments of membrane receptor proteins.

Ceruloplasmin glycopeptides

Another example of the power of HPLC to identify partial glycosylation at a single GlcN acceptor site is afforded by our studies of ceruloplasmin[20]. This sky-blue

copper-containing protein, containing 1046 amino acid residues, exists as a major form, Type I, and a minor form, Type II. These can be separated by hydroxylapatite chromatography and were thought to differ only in carbohydrate structure[23]. No apparent difference in the primary structure of Types I and II was identified by our complete amino acid sequence analysis of ceruloplasmin[23]. In previous work[19], the

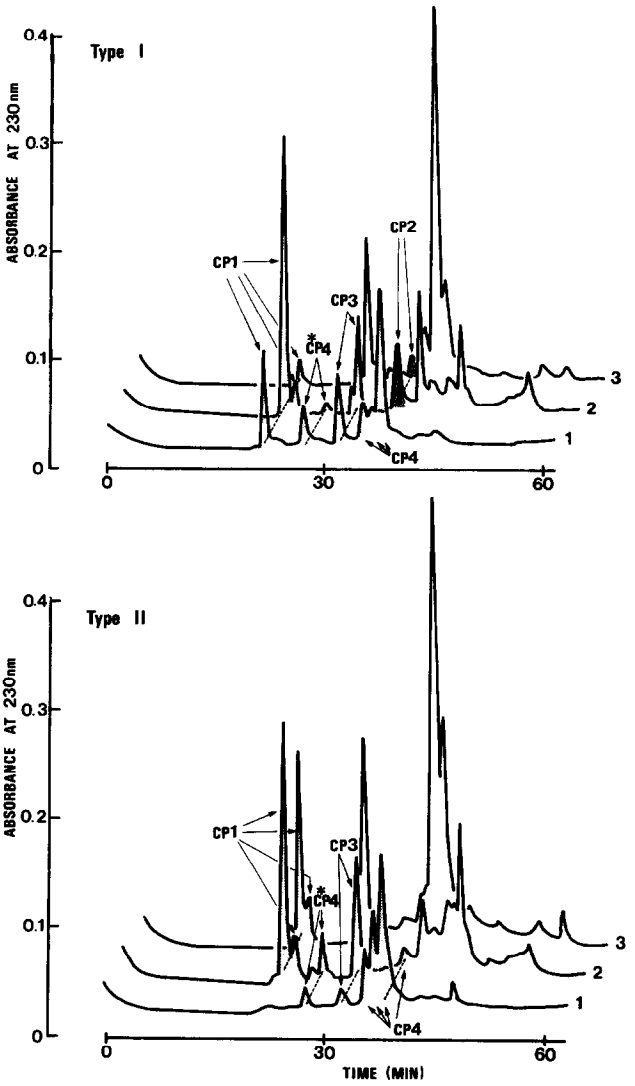

Fig. 4. Three-dimensional visualization of the chromatograms obtained by reversed-phase chromatography of ceruloplasmin glycopeptides on a 25 × 1.0 cm I.D. Synchropak RP-P column. The conditions of chromatography were the same as in Fig. 3. The HPLC profiles designated 1, 2 and 3 are for fractions of tryptic digests of human ceruloplasmin, Type I and Type II, that had been obtained by gel chromatography on a 100 × 1.5 cm I.D. Sephadex G-50 column. Corresponding peptides are connected by a dotted line. Peptide CP2 is cross-hatched in the chromatogram for Type I, but is missing from Type II, There are two forms of glycopeptide CP4 because of incomplete tryptic cleavage after Lys-751 (from Takahashi *et al.*[20]).

glycopeptides of a chymotryptic digest had been purified by HPLC for use in the sequence analysis, but no clear evidence for carbohydrate variants was obtained.

To ascertain whether Types I and II were indeed carbohydrate variants, the combination of gel chromatography and HPLC described above for IgD was modified and was applied to both types of ceruloplasmin[20]. Reversed-phase chromatography on a 25 × 1.0 cm I.D. Synchropak RP-P column was applied to GlcN-containing fractions, obtained by gel chromatography on a 100 × 1.5 cm I.D. Sephadex G-50 column of tryptic digests of Types I and II. Comparison of the HPLC profile of the GlcN-containing fractions is shown as a three-dimensional visualization in Fig. 4. From this comparison it is evident that glycopeptide CP2 is missing from Type II; however, for both Type I and Type II the other three glycopeptides are eluted by HPLC at exactly the same positions and in the same forms. The amino acid sequence of CP2 contains Asn-339, which appears to be glycosylated in Type I but not in Type II. In Fig. 4 the presence of several peaks for each of the glycopeptides probably reflects carbohydrate heterogeneity.

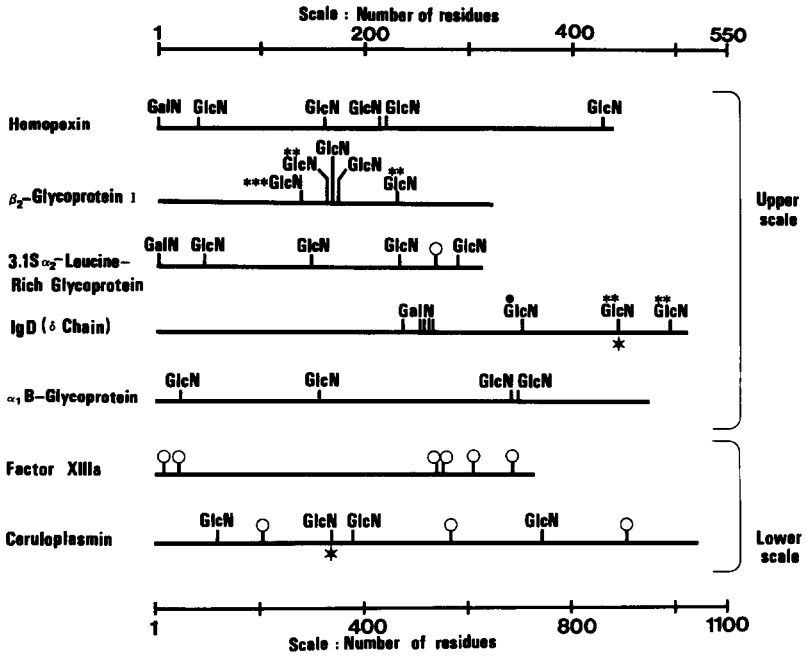

Fig. 5. Linear structural models of human plasma glycoproteins from which glycopeptides were purified. The attachment sites for GalN and GlcN oligosaccharides are indicated by GalN and GlcN. The oligosaccharides for which the structures are known are also distinguished by ** for biantennary, *** for triantennary, and ● for a high-mannose type of GlcN oligosaccharide. If the carbohydrate is missing in some molecules, the site is marked by ★. Open circles identify potential acceptor sites having the Asn-X-Ser/Thr signal sequence which are not glycosylated. In the text, the GlcN glycopeptides of each protein are numbered in their order from the amino terminus, *e.g.*, CP1, CP2, CP3, CP4, in ceruloplasmin. The upper scale for the number of amino acid residues is used for the upper five proteins, and the lower scale is used for Factor XIIIa and ceruloplasmin.

Linear structural models of plasma glycoproteins

Similar structural studies were undertaken on a number of other human plasma proteins, including hemopexin[25], β_2-glycoprotein I[24], 3.1S α_2-leucine-rich glycoprotein[26], blood coagulation Factor XIIIa[28], α_1B-glycoprotein[27], and Zn-α_2-glycoprotein[11]. All of these are multiglycosylated, except for Factor XIIIa, which has six potential sites for N-glycosylation but contains no carbohydrate[28]. Through collaborative studies of monoglycosylated peptides supplied by us, the carbohydrate structure was determined for the GlcN oligosaccharides in ceruloplasmin, hemopexin, and β_2-glycoprotein I. The work on the latter two proteins is still being prepared for publication. However, the results are summarized in Fig. 5, which gives linear structural models for seven plasma glycoproteins. Not shown in Fig. 5 is the structural model for Zn-α_2-glycoprotein, which is a plasma protein related to the major histocompatibility complex antigens on lymphocyte surfaces. The Zn-α_2-glycoprotein has three GlcN glycans of the complex biantennary type[11].

Several generalizations drawn from our work are illustrated by the schematic models in Fig. 5. With respect to GalN oligosaccharides, these are infrequent in most plasma proteins and tend to be attached at or close to the amino terminus or else to cluster in central hinge regions, as in IgD (Fig. 5) or IgA1 (Fig. 2). In fact, the presence of GalN in hemopexin and in the leucine-rich glycoprotein was unsuspected and could only be established by purification of the amino-terminal peptide by HPLC, followed by analysis. In contrast, the GlcN glycans are readily detected, but contrary to prediction[34], they do not tend to predominate towards the amino terminus of the nascent polypeptide chain. In fact, they are distributed non-uniformly and unpredictably along its length— unpredictably, that is, except for the linkage to the Asn-X-Ser/Thr acceptor sequence. However, instances where the signal sequence is not at all glycosylated are also known, *e.g.* in Factor XIIIa[28]. The available data for plasma glycoproteins do not support the conclusion of Pollack and Atkinson[34], who made a survey of glycoproteins and found that the location of glycosylation sites in the polypeptide chain correlates with processing. They concluded that complex type of GlcN glycans tend to appear in the first 100 amino acid residues and high-mannose glycans tend to predominate after residue 200. This is certainly not the case for IgA1, IgA2, IgD (Fig. 2), or β_2I (Fig. 5).

Retention time, hydrophobicity, amino acid sequence, and carbohydrate structure of glycopeptides

The availability of a number of glycopeptides that had different amino acid sequences and were monoglycosylated with glycans of different types offered an opportunity to analyze the effect of carbohydrate on the behavior of glycopeptides in reversed-phase chromatography[20]. Therefore, retention times of glycopeptides (t_R) on a 25 × 1.0 cm I.D. Synchropak RP-P column were plotted against $\ln(1 + H)$ in Fig. 6B, in which H represents the hydrophobicity of the peptide, calculated by the method of Sasagawa *et al.*[35] and using their list of non-weighted retention constants for amino acids. For reference purposes, the t_R values of non-glycosylated peptides from plasma proteins of known amino acid sequence were plotted against $\ln(1 + H)$ (Fig. 6A). Although there is considerable scatter of the points in Fig. 6A and B, in both cases the retention times were linearly related to $\ln(1 + H)$.

In general, the behavior of glycopeptides in reversed-phase chromatography

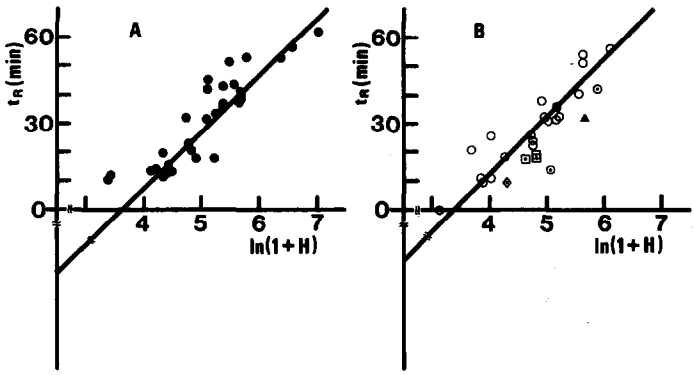

Fig. 6. Relationship between retention time (t_R) and ln(1 + H) in reversed-phase HPLC. The observed retention times are plotted against ln(1 + H), where H indicates the hydrophobicity of the peptide, calculated by the method of Sasagawa *et al.*[35], using their list of non-weighted retention constants. (A) Data for non-glycopeptides; (B) data for glycopeptides. ○, Glycopeptides with a single GlcN oligosaccharide; □, glycopeptides with a single GalN oligosaccharide; ⊙, glycopeptides with two GlcN oligosaccharides; △, glycopeptide with a single GalN and a single GlcN oligosaccharide; ◇, glycopeptide glycosylated at multiple sites for GalN oligosaccharide (from Takahashi *et al.*[20]).

is very similar to that of non-glycopeptides[20]. However, there is some tendency for multiglycosylated glycopeptides to have lower retention times than expected solely from their amino acid composition. This is somewhat surprising, because in most cases the carbohydrate contributes half or more of the total molecular weight of the glycopeptide. For example, the average molecular weight of a GlcN glycan is about 2500 and that of a GalN glycan is about 600. Thus, the effect of carbohydrate on retention time is less than might be expected from its relative size. Nonetheless, as shown by the earlier example of the D2 glycopeptide and the *D2 non-glycopeptide of IgD, a combination of gel chromatography and reversed-phase chromatography can be used to separate two peptides that have the same amino acid sequence and differ only in the presence and absence of carbohydrate.

Reversed-phase chromatography differs in principle from affinity chromatography with immobilized lectins, a procedure that has been widely used for the fractionation and structural characterization of oligosaccharides, glycopeptides, and glycoconjugates[36]. Lectin columns fractionate on the basis of preferential binding of single sugar residues or various types of oligosaccharide chains. However, as shown by Fig. 6, polypeptide structure and hydrophobicity are the major determinants in reversed-phase HPLC. In fact, the chromatographic contact region probably involves only a small number of amino acids, according to studies of Regnier[37] on epitopes of proteins and variants of lysozyme.

Chromatographic purification and characterization of glycoproteins

Microheterogeneity of the oligosaccharides presents a dilemma for HPLC: the greater the separation power of the column, the more ambiguous the separation becomes. There is no single method applicable to purification of all glycoproteins, not only because of the structural microheterogeneity of the carbohydrate, but also because each glycoprotein is unique in amino acid sequence and thus in the kinds of

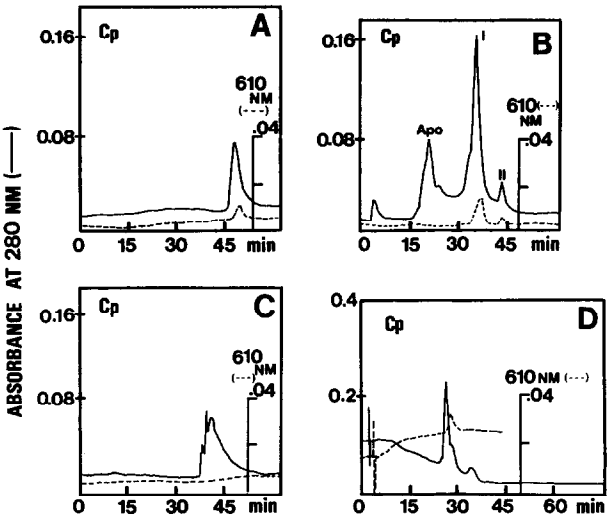

Fig. 7. Chromatographic characterization of human ceruloplasmin (Cp): (A) anion-exchange HPLC on a 25 × 0.4 cm I.D. SynChropak AX-300 column, (B) 10 × 0.75 cm I.D. hydroxyapatite, (C) reversed-phase HPLC on a 25 × 0.4 cm I.D. SynChropak RP-P column, (D) hydrophobic-interaction chromatography on a 7.5 × 0.75 cm I.D. TSK Phenyl 5PW column. In each case, the eluent was monitored by absorbance at 280 nm for protein and at 610 nm for the blue color of ceruloplasmin. In A, the column was eluted at a flow-rate of 1.0 ml/min with a linear gradient for 60 min from 0.02 *M* Tris–acetic acid buffer (pH 8.0), to 0.02 *M* Tris–acetic acid buffer (pH 8.0), containing 1.0 *M* sodium acetate. In B, elution was at a flow-rate of 0.8 ml/min for 60 min with a linear gradient from 0.01 *M* sodium phosphate buffer (pH 6.8), containing 0.3 m*M* calcium chloride to 0.5 *M* sodium phosphate buffer (pH 6.8), containing 0.01 *M* calcium chloride. In C, the column was eluted at a flow-rate of 1.0 ml/min with a linear gradient from 0.1% trifluoroacetic acid to 60% acetonitrile, containing 0.1% trifluoroacetic acid during 60 min. In D, elution was at a flow-rate of 1.0 ml/min with a gradient of decreasing salt concentration from 1.7 *M* ammonium sulfate in 0.1 *M* sodium phosphate buffer (pH 7.0), to 0.1 *M* sodium phosphate (pH 7.0). Because of the low grade of ammonium sulfate used, the baseline decreased gradually with decrease in salt concentration.

oligosaccharides and in the number and location of their potential attachment sites (*e.g.*, see Fig. 5). Therefore, HPLC methods based on different separation modes are needed for the isolation and characterization of glycoproteins.

In order to evaluate the applicability to glycoproteins of columns with different modes of separation, a study was made of five human plasma glycoproteins for which the amino acid sequence and the carbohydrate type and location were known: ceruloplasmin[23], leucine-rich glycoprotein[26], β_2I-glycoprotein[24], IgD[17], and hemopexin[25]. The types of chromatography used were: anion-exchange HPLC (Synchropak AX-300), spherical hydroxyapatite (Toa Nenryo), reversed-phase HPLC (Synchropak RP-P), and hydrophobic-interaction HPLC (TSK Phenyl 5PW). The column sizes are given in the legend to Fig. 7. The results of the application of three or more of these procedures to the same sample of each protein are shown successively as Figs. 7–11. When the four methods are applied to the same protein, as in the panel of Fig. 7 for ceruloplasmin, the varying behavior of the protein in different chromatographic systems is brought out. Equally interesting is the comparison of a single chromatographic system among the whole set of glycoproteins[4].

Ceruloplasmin

Ceruloplasmin is a large, single-chain, multicopper oxidase (1046 residues), that contains six Cu^{2+} ions of three different types[23]. Also, it exists in two forms, Type I and Type II, that differ in the number of GlcN oligosaccharides [four and three, respectively (see Fig. 4)]. Thus, ceruloplasmin presents several challenges for HPLC. In anion-exchange chromatography (Fig. 7A), a single, broad peak is obtained, reflecting charge heterogeneity of the oligosaccharides. Hydroxyapatite chromatography separates apoceruloplasmin and also Types I and II (Fig. 7B). In reversed-phase HPLC, an asymmetrical pattern is obtained, suggesting multiple components (Fig. 7C). Also, the 610-nm absorption disappears, indicating loss of the two blue copper ions. This is a reflection of the stringent conditions of reversed-phase chromatography, in which both the organic solvents and the acidic mobile phases tend to denature proteins[37]. However, the blue copper ions are retained in hydrophobic interaction chromatography, which is more gentle, and partial separation of the components, including apoceruloplasmin, occurs (Fig. 7D).

Leucine-rich glycoprotein

Leucine-rich glycoprotein consists of a single polypeptide, containing 312 residues, of which 66 are leucine, and it contains one GalN and four GlcN oligosaccharides[26] (Fig. 5). This protein is separated into three distinct peaks by anion-exchange chromatography (Fig. 8A), this despite the fact that this protein sample appeared homogeneous in sequence analysis and by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)[26]. However, leucine-rich glycoprotein was eluted as a single, slightly asymmetrical peak in reversed-phase chromatography (Fig. 8C) and in hydrophobic interaction chromatography (Fig. 8D), although the peak was clearly split in hydroxyapatite chromatography (Fig. 8B). We assume that the varying behavior of leucine-rich glycoprotein in the four modes of chromatography reflects two factors: (i) carbohydrate heterogeneity, and (ii) its amphipathic or bipolar character due to a repeating framework of leucine-rich domains[26].

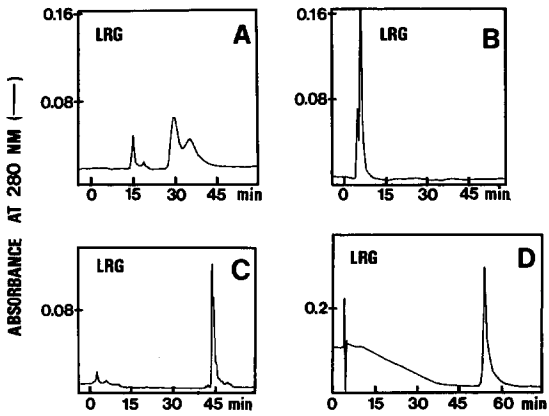

Fig. 8. Chromatographic characterization of human leucine-rich glycoprotein (LRG): (A) anion-exchange HPLC, (B) hydroxyapatite chromatography, (C) reversed-phase HPLC, (D) hydrophobic interaction chromatography. The columns and the conditions for chromatography were the same as in Fig. 7, except that in D water was used for elution after 40 min.

β_2-Glycoprotein I

β_2-glycoprotein I (β_2I) is a lipid-associated protein that has five GlcN oligosaccharides clustered in the middle of a single polypeptide chain, containing 326 amino acid residues[24] (Fig. 5). The carbohydrate structures are both biantennary and triantennary, and each type is probably heterogeneous. Furthermore, β_2I is a constituent of lipoproteins and binds lipid. It is no surprise, then, that β_2I glycoprotein exhibits heterogeneity in anion-exchange and hydrophobic interaction chromatography, as well as in hydroxyapatite chromatography (Fig. 9A, B and D). However, from a reversed-phase column, β_2I is eluted as a sharp major peak (Fig. 9C). The elution patterns are very different in the four types of chromatography. This multiglycosylated protein has a unique amino acid sequence with no evidence for polypeptide structural variants, and it appears homogeneous in SDS-PAGE[24]. Thus, the aberrant behavior in the different types of chromatography may be attributed either to microheterogeneity of the oligosaccharide structures or to binding of lipid. Probably because of removal of lipid by the organic solvent, β_2I was eluted as a single peak in reversed-phase chromatography (Fig. 9C), but two forms (one with and one without lipid) were separated by hydrophobic interaction chromatography in which no organic solvent was used (Fig. 9D). To ascertain the effect of carbohydrate on chromatographic behavior, we have begun a collaborative program to study preparations of β_2I in which a series of carbohydrate groups have been removed successively by glycosidases.

Immunoglobulin D

Immunoglobulin D presents special problems for HPLC, because it contains three heterogeneous GlcN oligosaccharides, one of which is of the high-mannose type and two are complex, biantennary structures[31]; furthermore, one of these is missing in a carbohydrate variant (Fig. 5). In addition, IgD contains five or more

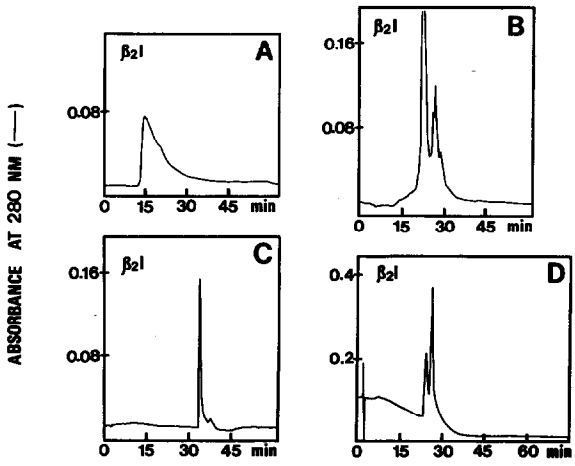

Fig. 9. Chromatographic characterization of human β_2-glycoprotein I (β_2I): (A) anion-exchange HPLC, (B) hydroxyapatite chromatography, (C) reversed-phase HPLC, (D) hydrophobic interaction chromatography. The columns and the conditions for chromatography were the same as in Fig. 7.

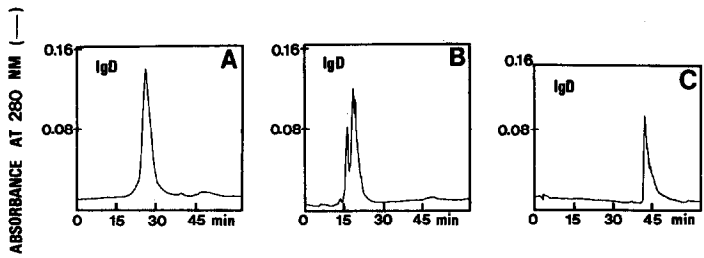

Fig. 10. Chromatographic characterization of human immunoglobulin D (IgD): (A) anion-exchange HPLC, (B) hydroxyapatite chromatography, (C) reversed-phase HPLC. The columns and conditions for chromatography were the same as in Fig. 7.

closely spaced GalN oligosaccharides that are heterogeneous in structure[17–20]. Also, IgD is rapidly degraded by proteolytic enzymes[22] with the result that Fab, Fc, and other fragments are present in most preparations. The combination of these factors results in quite different chromatograms for the same IgD preparation when columns having different modes of separations are used (Fig. 10A, B and C).

Hemopexin

Another example of different behavior in different types of chromatography is given by hemopexin. This multiglycosylated heme-colored protein has a GalN oligosaccharide at the amino terminus and five GlcN oligosaccharides spread throughout the polypeptide chain, containing 439 residues (Fig. 5). It also binds one heme tightly and stoichiometrically, but not covalently, and has a unique amino acid sequence with no evidence for polypeptide polymorphism or molecular weight heterogeneity[25]. Hemopexin is eluted as a broad split peak from the ion-exchange column with a coincidence of the 413-nm absorption due to the heme and the 280-nm absorption (Fig. 11A). The two peaks eluted from hydroxyapatite (Fig. 11B) may represent the oxidized and reduced forms of heme, because the second peak shifted to the first peak with time. Hemopexin was eluted as a rather sharp peak in reversed-phase chromatography, and it continued to bind heme (Fig. 11C). Thus, it did not appear to be denatured, unlike ceruloplasmin, which lost the absorption due to blue copper ions (Fig. 7C).

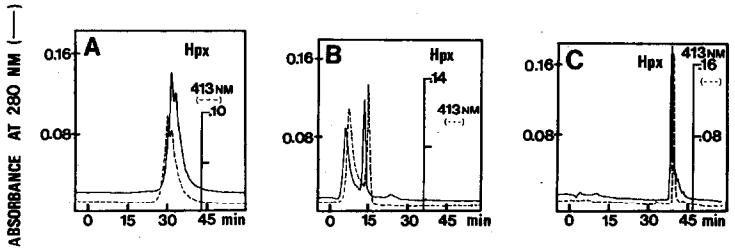

Fig. 11. Chromatographic characterization of human hemopexin: (A) anion-exchange HPLC, (B) hydroxyapatite chromatography, (C) reversed-phase HPLC. The columns and the conditions for chromatography were the same as in Fig. 7. The absorbance at 413 nm monitors the heme of hemopexin.

CONCLUSIONS

In summary, carbohydrate microheterogeneity is an important factor affecting the chromatographic behavior of glycoproteins. Even singly glycosylated polypeptides, such as the γ heavy chain of IgG, exhibit a multiplicity of oligosaccharide forms that affect both protein conformation and chromatographic properties. Because of the structural complexity of the carbohydrate of multiply glycosylated proteins, the effect of carbohydrate on chromatographic behavior is best studied by use of monoglycosylated glycopeptides of known amino acid sequence from different proteins. In reversed-phase chromatography the behavior of monoglycosylated glycopeptides is rather similar to that of the non-glycopeptides. However, multiply glycosylated peptides tend to be eluted earlier than expected from the hydrophobicity, calculated only from their amino acid content. Although the effect of the carbohydrate is smaller than expected, reversed-phase HPLC can be used to separate two peptides that have the same amino acid sequence but differ by the presence or absence of carbohydrate.

Microheterogeneity of the oligosaccharides presents a problem for the chromatographic purification and characterization of glycoproteins. Several conclusions were reached from a comparative study of the application of four types of columns to five plasma proteins. (1) Carbohydrate charge heterogeneity is reflected in ion-exchange chromatography. Multiply glycosylated proteins are eluted as broad, smeared, or split peaks. (2) HPLC with spherical hydroxyapatite columns is useful for the purification of glycoproteins, although the separation principles are not well known. Two forms of ceruloplasmin, differing by a single oligosaccharide, were separated by this method, as well as apoceruloplasmin. (3) The effect of carbohydrate in reversed-phase chromatography is relatively small, but the organic solvents and acidic conditions tend to denature some proteins,. The peaks tend to be narrow, but irregular in shape. (4) Hydrophobic interaction chromatography is more gentle, because no organic solvent is used. Carbohydrate did not seen to have much effect, but this method separated two forms of β_2I-glycoprotein that probably differed in the binding of lipid. (5) In general, reversed-phase and hydrophobic interaction chromatography seem to be most suitable for minimizing the effect due to microheterogeneity of carbohydrate. The recent commercial availability of a concanavalin A affinity column for HPLC offers a new approach to separation and structural characterization of glycopeptides and glycoproteins that is complementary to the approach described here.

ACKNOWLEDGEMENTS

We thank Dr. T. Isobe and Dr. N. Ishioka for helpful discussion, Mr. T. Ishikawa of Toa Nenryo K. K. for providing us with a spherical hydroxyapatite column, and Dr. Y. Kato of Toyo Soda Manufacturing Co. Ltd. for providing us with a TSK Phenyl 5 PW column. This work was supported by grants from the National Institutes of Health (NIH grants DK 19221 and CA 08497) and from the American Cancer Society (IM-2K).

REFERENCES

1 S. C. Churms (Editor), *CRC Handbook of Chromatography, Vol. I, Carbohydrates*, CRC Press, Boca Raton, FL, 1982.
2 J. Montreuil, *Adv. Carbohydr. Chem. Biochem.*, 37 (1980) 157.
3 F. W. Putnam, in F. W. Putnam (Editor), *The Plasma Proteins*, Vol. IV, Academic Press, Orlando, FL, 2nd ed., 1984, Ch. 2, p. 45.
4 N. Takahashi and F. W. Putnam, in K. M. Gooding and F. Regnier (Editors), *HPLC-Biological Macromolecules: Methods and Applications*, Marcel Dekker, New York, 1988, in press.
5 J. U. Baenziger, in F. W. Putnam (Editor), *The Plasma Proteins*, Vol. IV, Academic Press, Orlando, FL, 2nd ed., 1984, Ch. 5, p. 271.
6 R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.*, 54 (1985) 631.
7 S. C. Hubbard and R. J. Ivatt, *Annu. Rev. Biochem.*, 50 (1981) 555.
8 A. Kobata, in V. Ginsburg and P. W. Robins (Editors), *Biology of Carbohydrates*, Vol. 2, Wiley-Interscience, New York, 1984, p. 87.
9 J. F. G. Vliegenthart, L. Dorland and H. Halbeek, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209.
10 R. B. Trimble and P. H. Atkinson, *J. Biol. Chem.*, 261 (1986) 9815.
11 T. Araki, F. Gejyo, K. Takagaki, H. Haupt, H. G. Schwick, W. Bürgi, T. Marti, E. Rickli, R. Brossmer, P. H. Atkinson, F. W. Putnam and K. Schmid, *Proc. Natl. Acad. Sci. U.S.A.*, 85 (1988) 679.
12 R. B. Parekh, R. A. Dwek, B. J. Sutton, D. L. Fernandes, A. Leung, D. Stanworth, T. W. Rademacher, T. Mizouchi, T. Taniguchi, K. Matsuta, F. Takeuchi, Y. Nagano, T. Miyamoto and A. Kobata, *Nature (London)*, 316 (1985) 452.
13 I. Mononen and E. Karjalainen, *Biochim. Biophys. Acta*, 788 (1984) 364.
14 E. Bause and G. Legler, *Biochem. J.*, 195 (1981) 639.
15 H. Loebermann, R. Tokuoka, J. Deisenhofer and R. Huber, *J. Mol. Biol.*, 177 (1984) 531.
16 D. R. Davies and H. Metzger, *Annu. Rev. Immunol.*, 1 (1983) 87.
17 F. W. Putnam, N. Takahashi, D. Tetaert, L.-C. Lin and B. Debuire, *Ann. NY Acad. Sci.*, 399 (1982) 41.
18 S. J. Mellis and J. U. Baenziger, *J. Biol. Chem.*, 258 (1983) 11 557.
19 D. Tetaert, N. Takahashi and F. W. Putnam, *Anal. Biochem.*, 123 (1982) 430.
20 N. Takahashi, Y. Takahashi, T. L. Ortel, J. N. Lozier, N. Ishioka and F. W. Putnam, *J. Chromatogr.*, 317 (1984) 11.
21 N. Takahashi, N. Ishioka, Y. Takahashi and F. W. Putnam, *J. Chromatogr.*, 326 (1985) 407.
22 N. Ishioka, N. Takahashi and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 84 (1987) 61.
23 N. Takahashi, T. L. Ortel and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 390.
24 J. Lozier, N. Takahashi and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 3640.
25 N. Takahashi, Y. Takahashi and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 73.
26 N. Takahashi, Y. Takahashi and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 1906.
27 N. Ishioka, N. Takahashi and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 83 (1986) 2363.
28 N. Takahashi, Y. Takahashi and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 83 (1986) 8019.
29 S. J. Mellis and J. U. Baenziger, *J. Biol. Chem.*, 258 (1983) 11 546.
30 W. Buergi, M. Endo, K. Schmid, H. G. Schwick, H. van Halbeek, J. F. G. Vliegenthart, J. Lozier, N. Takahashi and F. W. Putnam, in M. A. Chester, D. Heinegard, A. Lundblad and S. Svensson (Editors), *Glycoconjugates, Proc. 7th Int. Symp.*, Rahms, Lund, 1983, p. 198.
31 F. W. Putnam, N. Takahashi and N. Ishioka, in F. Milgrom (Editor), *Antibodies: Protective, Destructive, and Regulatory Role*, Karger, Basel, 1984, p. 26.
32 F. W. Putnam, in F. W. Putnam (Editor), *The Plasma Proteins*, Vol. V, Academic Press, Orlando, FL, 2nd ed., 1987, Ch. 2, p. 49.
33 M. B. White, A. L. Shen, C. J. Word, P. W. Tucker and F. R. Blattner, *Science (Washington, D.C.)*, 228 (1985) 733.
34 L. Pollack and P. H. Atkinson, *J. Cell. Biol.*, 97 (1983) 293.
35 T. Sasagawa, T. Okuyama and D. C. Teller, *J. Chromatogr.*, 240 (1982) 329.
36 T. Osawa and T. Tsuji, *Annu. Rev. Biochem.*, 56 (1987) 21.
37 F. E. Regnier, *Science (Washington, D.C.)*, 238 (1987) 319.