The Macromolecular Properties of Blood-Group-Specific Glycoproteins

CHARACTERIZATION OF A SERIES OF FRACTIONS OBTAINED BY SOLVENT FRACTIONATION

By J. MICHAEL CREETH, K. RAMAKRISHNAN BHASKAR,
ALASTAIR S. R. DONALD and WALTER T. J. MORGAN

The Lister Institute of Preventive Medicine, Chelsea Bridge Road, London SW1W 8RH, U.K.

(Received 15 March 1974)

1. The glycoprotein components of a human ovarian-cyst fluid were isolated by a solvent [95% (w/w) phenol]-extraction procedure; the phenol-insoluble water-soluble glycoprotein was further fractionated by $(\text{NH}_4)_2\text{SO}_4$ and by ethanol to yield eight fractions. 2. The fractions were analysed in terms of amino acids, fucose, galactose, *N*-acetylglucosamine, *N*-acetylgalactosamine and sialic acid. Variations occurred, particularly in the proportion of peptide; these were partly correlated with varying extent of serological activity. 3. The fractions were characterized physicochemically in terms of buoyant density and degree of spreading in a density gradient, sedimentation velocity and molecular weight; their partial specific volumes and specific refraction increments were also determined. 4. The fractions showed wide variations in their sedimentation-velocity and density-gradient patterns, and gave evidence of pauci-dispersity in density. The fraction regarded as the most typical blood-group-specific glycoprotein sedimented as a single rapidly spreading peak and was of high molecular weight. 5. Significant correlations were observed between the physical properties of the glycoprotein fractions and the amount of their peptide component. The buoyant densities and sedimentation coefficients varied in a manner that suggested the existence of two families of glycoproteins. 6. It is suggested that variability in the extent of glycosylation, or in the degree of cross-linking, might account for the two families of glycoproteins, and that the extent of cross-linkage might also be a factor determining the solubility of these glycoproteins in hot saturated $(\text{NH}_4)_2\text{SO}_4$.

Earlier studies of the macromolecular properties of the blood-group-specific glycoproteins obtained from human ovarian cysts (Creeth & Knight, 1967, 1968) were concentrated on a single subfraction which was sufficiently homogeneous for the physical measurements to be easily interpreted. This fraction was obtained by the standard solvent-fractionation procedure (e.g. Morgan, 1967), which entails an initial separation from protein through the action of 95% (w/w) phenol, and subsequent fractionation with hot saturated $(\text{NH}_4)_2\text{SO}_4$ followed by ethanol. The particular substance studied, denoted glycoprotein 376, was obtained as a precipitate in the ethanol-concentration range 50–56% (v/v). Fractions obtained at both lower and higher ethanol concentrations were of very similar analytical composition and were equally active serologically: most, however, showed more than one peak in sedimentation-velocity experiments, and accordingly were regarded at that time as unsuitable for further physicochemical investigation. Less extensive studies were made of several preparations from other cyst fluids (see, e.g., Gibbons *et al.*, 1970) and these confirmed that the characteristics shown by glycoprotein 376 were shared by other similarly fractionated substances; again, however, the fractions characterized were only a small proportion of the total active glycoprotein of

the respective cyst fluids. Since their serological activity is the primary interest in these glycoproteins, it seemed desirable to extend the physicochemical characterization to as wide a range of active fractions as possible, so that any variation in properties could be correlated with analytical composition and degree of serological activity. With the present understanding of the macromolecular properties of these substances, the heterogeneity expected from some fractions is no longer a barrier to interpretation. Accordingly, the present paper is concerned with a series of eight fractions obtained from a single ovarian cyst, these fractions comprising the whole of the phenol-insoluble water-soluble component. The physical methods used are those of sedimentation velocity, intermediate-speed sedimentation equilibrium and distribution in a gradient of a dense electrolyte, these having proved both informative and economical.

Experimental

Materials

Reagents were of analytical grade, where obtainable, and were used without further purification. Galactose and fucose, used as analytical standards, were obtained from British Drug Houses Ltd.

(Poole, Dorset, U.K.); *N*-acetylneuraminic acid was obtained from Koch-Light Ltd. (Colnbrook, Bucks., U.K.). A buffer solution composed of 33 mM-NaCl, 16.7 mM- Na_2HPO_4 and 16.7 mM- NaH_2PO_4 , containing also 1 mM- NaN_3 , giving pH 6.8, *I* 0.10M, was used for all the physical measurements. Human serum albumin and immunoglobulin G were preparations obtained from the Lister Institute (Elstree, Herts., U.K.); each contained approx. 90% of the designated protein.

The ovarian-cyst fluid (no. 603) constituting the source of the glycoprotein was obtained from a 'non-secretor' person and was blood-group Le^a-specific and possessed neither A, B, H nor Le^b specificity. The fractionation scheme was based on that described by Morgan (1967); it is summarized, with the distributions obtained, in Scheme 1. The precipitates were redissolved in distilled water, the solutions were dialysed and the dialysis residues were

centrifuged at 30000g (r_{av} . 10.7cm) for 60min to remove traces of insoluble material before being freeze-dried.

Fraction 0-45, and to a minor extent fraction 0-43, did not dissolve completely at 4mg/ml. These fractions were, however, fully self-dispersing in water or buffer at this concentration, giving stable opalescent solutions, but the light-absorbing material was removable by low-speed centrifugation. Since ultracentrifuge measurements necessarily refer to the fully soluble portions, solutions for analysis were also clarified by centrifugation (approx. 18000g; r_{av} . 7.3cm; 3h), values of the concentration being found from the measured refraction increment, Δn , and the value of the specific refraction increment appropriate to the peptide content of the fraction. In addition, analytical values were obtained for the whole fractions, comprising both soluble and 'insoluble' portions. The latter term, which is used below

Scheme 1. Extraction and fractionation scheme for glycoproteins from cyst no. 603

without qualification, implies merely insolubility under the conditions used. Subsequent experience has shown that the 0-45 fraction will dissolve almost completely at 2.5mg/ml on prolonged standing.

Analytical methods

Galactose was determined by the phenol-H₂SO₄ method of Dubois *et al.* (1956), and fucose by the method of Gibbons (1955). The galactose values were corrected for the contribution from fucose. Sialic acid was determined as *N*-acetylneuraminic acid by the thiobarbituric acid method of Warren (1959) after liberation by heating in 0.05M-H₂SO₄ for 1 h at 80°C. Amino acids and amino sugars were determined with an amino acid analyser, following the procedures of Donald (1973).

Electrophoresis

The fractions were subjected to electrophoresis in the phosphate buffer specified on a modified cellulose-acetate support (Cellolog; Reeve Angel and Co., London E.C.4, U.K.); experiments were performed in duplicate, a serum sample being used as control on each strip. One strip was stained with Amido Black, and the other with periodic acid-Schiff reagent, by the procedures recommended by the manufacturer.

Partial specific volumes

These were determined on the fully soluble fractions, from measurements of the density of solutions containing approx. 10mg of glycoprotein/ml. An Anton Paar (Graz, Austria) densimeter, model DMA 02C, was used; this instrument has a precision approaching 1μg/ml. Solutions were prepared by dissolving the glycoprotein, dried by heating to 60°C over P₂O₅ *in vacuo*, directly in buffer solution: a micro-balance reading to 1μg was used and the weights of glycoprotein and buffer were corrected for air buoyancy. To conserve material, it was assumed that concentration-dependence effects were negligible [as was found by Creeth & Knight (1967) with glycoprotein 376] and accordingly that the apparent specific volumes ϕ , determined from eqn. (1), were identical with the partial specific volumes required:

$$\phi = \frac{1}{g_2} \left(\frac{1}{\rho} - \frac{g_1}{\rho_s} \right) \quad (1)$$

In this equation, ρ is the solution density, ρ_s that of the solvent, and g_1 and g_2 the weight fractions of the solvent and solute (glycoprotein) respectively. Specific refraction increments for the glycoprotein fractions were determined on the solutions used for density measurements; a differential refractometer (Polymer Consultants Ltd., Nottingham, U.K.), modified to

give a Rayleigh interference pattern in the image plane, was used. The precision of the instrument is about 5×10^{-6} in refractive index.

Ultracentrifuge methods

Solutions for ultracentrifugal characterization were prepared by dissolving undried glycoprotein in the phosphate-chloride buffer at a nominal concentration of 4.4mg/ml and clarified if necessary (see above). Measurement of Δn and use of the measured specific refraction increment then gave concentration values of adequate precision.

Sedimentation-velocity experiments were made on these solutions, by established procedures (e.g. Creeth, 1964); a Beckman model E analytical ultracentrifuge was used, with 30mm cells, usually of the single-sector type.

For analytical density-gradient experiments, the glycoprotein solutions were diluted to 1.4mg/ml in the same buffer; Cs₂SO₄ was then added to a weight fraction appropriate to the density required, the relationship given by Wake & Baldwin (1962) being used. The solutions were centrifuged to equilibrium at 42 040 rev./min, in 12mm double-sector cells, the experimental procedure described by Creeth & Denborough (1970) being followed.

In one instance, that of fraction 60-75, the Centri-scan 75 ultracentrifuge (MSE Ltd., Crawley, Sussex, U.K.) was used for a density-gradient experiment. This machine has both schlieren and u.v.-absorption optics; since the former system is roughly equally sensitive (at 546nm) to proteins and carbohydrates, whereas the latter (at 280nm) detects only protein, the instrument is particularly adapted for detecting protein in the presence of glycoprotein. The output of both systems is obtained on an X-Y recorder.

For sedimentation equilibrium, the glycoprotein stock solutions were diluted to 0.4mg/ml and dialysed against approx. 200-fold excess of buffer for 2-3 days: these solutions (now averaging about 0.3mg/ml) were then loaded into 30mm double-sector interference-type cells, 0.26ml being used to give a solution column of about 3mm. Heavy fluorocarbon oil (FC43) was used to define the lower limit of the solution column; the two compartments were loaded identically, the dialysis buffer being used for the reference compartment. The symmetrical double-slit diaphragm was used.

The equilibrium speed chosen was that defined by the condition $A = 0.6$, where $A = M(1 - \bar{v}\rho)\omega^2/2RT$, the symbols having their usual significance. A preliminary value of M was used, based on the sedimentation coefficient measured for the particular fraction and the known relationship between s and M , as modified for this type of glycoprotein (Creeth & Knight, 1968). The overspeeding technique of Hexner *et al.* (1961) was used to decrease the time required to

reach equilibrium, the overspeed being selected to be appropriate to the period of 16h; this entails a value for the ratio overspeed/operating speed of about 1.2. The attainment of equilibrium was verified by performing a complete set of measurements on the fringe pattern, the simpler method of establishing constancy of the concentration increment across the cell being inappropriate to this method (Teller *et al.*, 1969).

The speed defined by $A = 0.6$ is a compromise between that used earlier, 0.3 (Creeth & Knight, 1967), and the higher value of approx 0.9 used by Tanford *et al.* (1964). The value 0.3 gives minimal fractionation, so that extrapolation to the cell base is easy; however, concentration increments across the cell are small so that much higher initial concentrations must be used. At these concentrations, non-ideality effects are significant, and several experiments must be carried out to establish a single value of M_w^0 . At $A = 0.9$ fractionation effects with our materials are severe, and hence extrapolation to the base is very hazardous. At $A = 0.6$ a concentration of 0.3mg/ml in the 30mm cell gives a concentration increment of six fringes across the cell; the maximum concentration is sufficiently low that non-ideality effects (Creeth & Knight, 1968) may be ignored, and a single experiment suffices to define the molecular weight.

Calculations

Sedimentation coefficients were calculated and corrected to the '25, w' basis by standard methods; however, measurements of the optical recovery (Creeth & Knight, 1967) were not feasible in the conditions used. Accordingly, the mean concentration of the experiment was taken as the product of the

initial concentration and the mean dilution factor. The effects of minor differences in concentration were removed by correcting all s values to a standard concentration of 3.5mg/ml by the following procedure. Since the concentration-dependence of s follows the reciprocal form $s = s^0/(1 + Kc)$, $ds/dc = -s^2(K/s^0)$. It is assumed that K is directly proportional to s^0 , when the value found for glycoprotein 376 ($K/s^0 = 1.07 \times 10^{-3}$ litre $\cdot$ g $^{-1}$ $\cdot$ s $^{-1}$; Creeth & Knight, 1967) can be applied. Over small concentration differences, $\Delta s = \Delta c \cdot ds/dc$, enabling the small correction Δs to be obtained for each experiment from the known s and the mean concentration, $\bar{c}$, in the experiment: $\Delta c = (\bar{c} - 3.5)$ mg/ml. The maximum value of the correction so applied was 0.4S; the method was validated experimentally for much larger corrections.

In analytical density-gradient experiments, the calculations followed those described earlier (Creeth & Denborough, 1970), modified appropriately for processing by desk-top computer.

The values of the buoyant density, ρ^0 , are uncorrected for pressure effects; the selective solvation parameter, Γ' , was determined by combining the ρ^0 and $\bar{v}$ values for the individual fraction; the apparent molecular weights in the gradient, M^{app} , were determined from the value for the solvated species (found directly from the experiment) and the appropriate factor $(1 + \Gamma')$; calculations of M^{app} were made for both lobes of the patterns where possible, but where distortions occurred, indicating the presence of discrete further components, only one lobe could be so analysed.

In the calculation of molecular weights from intermediate-speed sedimentation equilibrium it was necessary to treat both c^0 , the initial concentration, and c_m , the concentration at the meniscus, as unknown

Table 1. *Serological activity, carbohydrate and total peptide composition of glycoprotein fractions from cyst no. 603*

Carbohydrate and total peptide compositions are expressed as g of component/100g of fraction. Values not determined are indicated by —. Relative activity is expressed in terms of the activity of fractions 0–45 and 43–50, both of which inhibited agglutination of Le^a cells by an anti-Le^a antiserum at a dilution of 3.2×10^6 from an initial solution of 5mg/ml.

Fraction (and no.)	Relative activity	Fucose	Galactose	N-Acetyl- glucos- amine	N-Acetyl- galactos- amine	Sialic acid	Peptide	Total
AmS sol	(1) 0.25	12.6	28.2	25.9	4.1	6.1	7.1	84
0–45 total	(2) 1	9.6	24.8	22.1	5.9	—	27.2	—
0–45 Sol*	(2s) —	11.6	30.0	27.0	5.9	7.4	11.6	93
45–50	(3) 0.5	11.1	28.4	25.1	6.9	7.9	13.0	92
0–43 total	(4) 1	7.2	22.2	20.1	4.9	—	23.9	—
0–43 Sol†	(4s) —	7.9	25.0	—	—	7.1	22.3	—
43–50	(5) 1	11.8	28.0	26.1	5.8	7.9	11.1	91
50–60	(6) 0.25	8.8	25.6	19.1	8.8	7.8	20.2	90
60–75	(7) 0.06	5.1	17.9	13.8	4.2	7.8	45.8	95
75+	(8) 0.00	—	1.7	1.0	0.2	—	98.0	101

* Approx. 40% was soluble.

† Approx. 90% was soluble.

Table 2. Amino acid composition of glycoprotein fractions from cyst no. 603
Values of individual amino acids are expressed as μmol of amino acid/ $100\mu\text{mol}$ of total amino acids; totals are expressed as g of amino acids/ 100 g of glycoprotein.

Fraction	Asp	Thr	Ser	Glu	Pro	Gly	Ala	$\frac{1}{2}\text{Cys}$	Val	Met	Ileu	Leu	Tyr	Phe	His	Lys	Arg	Total peptide
AmS sol	(1)	4.8	20.8	15.2	3.9	17.0	7.0	13.7	0.0	4.7	0.3	1.5	1.9	0.6	4.1	0.9	3.3	7.1
0-45 Total	(2)	8.1	13.1	10.8	9.3	9.2	6.8	8.3	2.0	6.4	1.0	3.1	7.4	3.5	2.3	2.6	3.9	27.2
0-45 Sol (~40%)	(2s)	5.9	20.4	15.2	6.7	11.6	7.5	9.3	1.2	3.8	0.1	2.1	3.9	2.2	2.8	2.3	4.4	11.6
45-50	(3)	4.7	21.9	15.1	5.8	13.8	5.6	9.4	1.1	4.7	0.4	2.2	3.8	2.1	2.8	1.7	3.7	13.0
0-43 Total	(4)	13.6	12.0	11.1	11.6	8.5	4.6	7.8	1.5	6.2	1.1	1.9	6.2	3.0	3.0	2.8	3.1	23.9
0-43 Sol (~90%)	(4s)	14.8	12.0	11.8	11.9	7.3	5.1	8.1	tr	6.2	tr	2.1	6.4	2.9	2.8	3.8	2.9	22.3
43-50	(5)	4.8	20.7	14.8	5.0	15.4	6.3	10.9	0.7	5.6	0.3	1.8	3.8	0.9	2.7	1.5	3.2	11.1
50-60	(6)	5.9	22.8	15.5	7.4	10.9	5.2	8.4	2.0	5.0	0.4	2.0	3.9	1.3	1.8	1.9	3.3	20.2
60-75	(7)	11.0	9.8	8.7	14.3	8.4	5.3	7.7	2.9	6.5	1.1	2.3	6.9	2.6	2.6	2.8	3.7	45.8
75+	(8)	11.1	5.4	4.6	17.5	4.9	2.7	10.5	2.7	7.8	0.4	1.6	11.1	5.3	2.1	6.1	3.4	98.0

(Creeth & Knight, 1968). This course was adopted because very low c^0 values cannot be defined by independent measurement. Under the speed condition chosen, the equilibrium distribution in the upper third of the cell follows approximately the ideal law for a single solute; accordingly the method of J. M. Creeth and J. C. Holt (unpublished work; see also Gratzer *et al.*, 1972) was used to define c_a . Logarithmic plots of the true concentration c versus r^2 were then extrapolated by a least-squares fit to a parabola over the bottom quarter of the cell to define c_b , the concentration at the lower limit of the cell. M_w^0 and c^0 were then obtained from the relationships

$$c^0 = 2 \int_a^b r c dr / (b^2 - a^2)$$

(2)

$$M_w^0 = [(c_b - c_a)/c^0 (b^2 - a^2)][2RT/(1 - \bar{v}\rho)\omega^2]$$

(3)

The integral was evaluated by a trapezoidal approximation at regular intervals of r^2 ; thirty measurements were performed. Even with the very low speeds used, the contribution of the last trapezium to the integral, which depends on the extrapolated value $c(b)$, was about 20% of the total. The extrapolation procedure is justified empirically by the observed closeness of fit in the measurable regions. M_w^0 is the weight-average value over the whole distribution. Molecular-weight measurement was not attempted if the sedimentation-velocity pattern gave evidence of more than one discrete component.

Results

The serological activity of the glycoprotein fractions, and their analytical composition in terms of the individual carbohydrates and the total peptide, are given in Table 1. The detailed amino acid compositions are summarized in Table 2. All the physico-chemical results are collected in Table 3, and the sedimentation-velocity and density-gradient patterns are shown in Fig. 1.

The ultracentrifuge patterns of all fractions except AmS sol indicate the presence of more than one component in either the velocity or the density-gradient experiment. Nevertheless, except for fraction 0-43, the velocity patterns indicate polydispersity rather than gross heterogeneity. The presence of incompletely resolved components in the density-gradient patterns is evidence of pauci-dispersity in density. This is an important observation, which is discussed below. However, because the density differences are numerically small, it does not seriously hinder the interpretation of the physical measurements; all the values in Table 3 can be regarded as approximate weight-average quantities. Similarly, although the particular heterogeneity of each sample must be kept in mind, the analytical values (which are

Table 3. *Physicochemical characterization of glycoprotein fractions from cyst no. 603*

The $s_{25,w}$ values are corrected to common concentration of 3.5 mg/ml. Abbreviations: M_w^0 , weight-average molecular weight from sedimentation equilibrium; M^{app} , apparent molecular weight from density-gradient experiment; Γ' , selective solvation parameter.

Fraction	Peptide content (%)	$10^4 \times \Delta n / \Delta c$ (ml/mg)	$\bar{v}$ (ml/g)	ρ^0 (g/ml)	Γ' (g/g)	$s_{25,w}$ (S)	$10^{-3} \times M^{app}$		
							$10^{-3} \times M_w^0$	+ve lobe	-ve lobe
AmS sol	7.1	1.56	0.608	1.351	0.507	8.40	840	96	160
0-45 Sol	11.6	—	—	1.354	—	11.3	2000	—	173
45-50	13.0	1.56	0.620	1.335	0.514	8.94	1870	71	—
0-43 Sol	22.3	—	—	1.333	—	8.06	—	—	—
43-50	11.1	1.56	0.611	1.345	0.516	9.29	1520	92	—
50-60	20.2	1.59	0.629	1.315	0.548	6.89	810	174	—
60-75	45.8	1.73	0.656	1.283	0.562	3.48	210	147	—
75+	98.0	1.85	0.731	1.243	0.378	4.90	170	131	—

themselves weight averages) refer to substances which are reasonably well-defined physically.

The immunodiffusion results are shown in Plate 1; they demonstrate the presence of a serum protein component as a trace contaminant in all fractions except AmS sol and 43-50. The result shown in Plate 1(b) proves that the contaminant in the active fractions is albumin. It should be emphasized that the proportion of the albumin contaminant is very low: 1% of a protein contaminant is detectable by analytical density-gradient ultracentrifugation, and the patterns (Fig. 1a) obtained with the Spinco ultracentrifuge for all fractions up to 50-60 show no such effect. The pattern for fraction 60-75 is less convincing because of the smaller difference in buoyant density of the main component from the value characteristic of proteins; however, in this case, the results obtained with the Centriscan ultracentrifuge are unambiguous. As shown in Fig. 1(c), the schlieren pattern closely resembles that obtained with the Spinco ultracentrifuge, whereas the pattern obtained with absorption optics at 280 nm shows a slightly asymmetric zone of mean buoyant density identical with that found from the schlieren pattern. Thus it must be concluded that fraction 60-75 is also essentially devoid of free protein and accordingly that the trace amount of protein observed in the immunodiffusion plates cannot have a significant effect on the analytical values for all fractions up to and including 60-75.

In the electrophoresis experiments, no components were detectable in the strip stained with Amido Black, except for fraction 75+. The glycoproteins stained with periodic acid-Schiff reagent showed ill-defined broad zones, generally of mobility between those of β - and γ -globulins. However, the fraction 60-75, which is of particular interest in view of its high peptide content, gave a comparatively well-defined zone of mobility between α_1 - and α_2 -globulins, when stained with periodic acid-Schiff reagent; this zone

was not detected in the strip stained with Amido Black. The fraction 75+ showed albumin and pre-albumin when stained with Amido Black, the pre-albumin being the only component visible when stained for glycoprotein.

The electrophoresis results, when taken with the results of the analytical density-gradient experiments, show that all fractions must be regarded as glycoproteins. However, fraction 75+ contains so little carbohydrate that its properties are essentially those of a globular protein, and accordingly it must be regarded separately from the other fractions.

The distribution given in Scheme 1 shows that fractions AmS sol, 0-45 (total), 43-50 and 50-60 are the major components, together constituting 93% of the recovered material. The analytical values given in Table 1 for these fractions are generally similar to those reported previously for typical Le^a substances (Watkins, 1972; Donald, 1973), the values for the five sugar components lying within the expected ranges. Although the variation in peptide content among these fractions is large, this is not uncommon in blood-group-specific glycoproteins (Donald, 1973). Inclusion of the remaining active fractions (i.e. excluding only fraction 75+) does not alter the general picture, except to emphasize the range of variation of the peptide content; a fraction that is nearly 50% peptide still has all five sugars and retains significant serological activity.

Because the analytical values for fucose, galactose and sialic acid depend on colorimetric methods, they must include appreciable experimental errors; moreover, the total peptide is obtained by addition of the values for the individual amino acids, among which small losses must be expected. General experience shows that complete additivity is rarely obtained, as the values for the totals in Table 1 exemplify. Accordingly, the values must be interpreted with caution. However, allowance for the probable error does not lead to any change in

Fig. 1. Ultracentrifuge patterns for the glycoprotein fractions

(a) For each of the fractions designated, the sedimentation-velocity pattern is in the upper row and the density-gradient pattern in the lower row. Velocity patterns were obtained at roughly similar times, at a speed of 50740 rev./min; initial concentrations were approx. 4 mg/ml. Density-gradient patterns were obtained after approx. 30 h at 42040 rev./min in Cs_2SO_4 solution. (b) Density-gradient patterns for the insoluble component of fraction 0-45, after resuspension in (i) Cs_2SO_4 of initial density 1.308 g/ml, (ii) CsBr of initial density 1.448 g/ml. (c) Density-gradient patterns for fraction 60-75 obtained with the Centriscan ultracentrifuge. Trace I (the single peak) was obtained with double-beam absorption optics at 280 nm; the twin-lobed pattern (II) was obtained with single-beam schlieren optics at 546 nm. The conditions of the experiment were identical with those in Fig. 1(a) except for the initial concentration and speed, which were respectively 0.5 mg/ml and 44000 rev./min. The positions of meniscus and cell base are denoted M and B respectively.

conclusion: no other carbohydrate component has been detected in blood-group-specific glycoproteins of high activity (Watkins, 1972).

In glycoproteins prepared by this method, the fractions precipitating in the range 45–50% ethanol usually possess the highest serological activity and most characteristic composition, particularly with regard to the detailed amino acid distribution. In agreement with this finding, the amino acid analyses given in Table 2 show that the minor fraction 45–50 and the essentially parallel major fraction 43–50 possess the profile characteristic of all blood-group-specific glycoproteins, irrespective of specificity (Pusztai & Morgan, 1963; Watkins, 1972). Thus the sum of the values for serine, threonine and proline exceeds 50% of the total amino acids, the values for aspartic acid and glutamic acid are low (totalling less than 12%), the basic amino acids are in yet smaller proportion, and the aromatic and sulphur-containing amino acids are only barely represented. Fraction 43–50 comprises 25% of the total recovered cyst glycoprotein and is the major fully soluble component; fraction 45–50 is essentially identical in amino acid profile, but possesses a slightly higher peptide content.

The remaining fractions, though broadly similar, show differences in detail. Fraction AmS sol is exceptional in its solubility characteristics, and in not being subjected to fractionation by ethanol: though its carbohydrate composition is similar to that of 43–50 its amino acid profile shows considerable differences; the total peptide content is also lower, as is its serological activity. Among the ethanol fractions, 0–45 is much the largest, constituting 50% of the total glycoprotein: however, more than half is insoluble under the conditions used, and significant analytical differences are implied between the soluble and insoluble components. Although direct analysis of the insoluble component was not possible, the results for fraction 0–45 (total) show that the whole fraction has a much higher proportion of peptide than fraction 43–50 (27% compared with 11%) and, moreover, that the peptide is richer in the acidic amino acids (aspartic acid+glutamic acid = 17% in fraction 0–45; 10% in fraction 43–50) and poorer in the hydroxy amino acids and proline (serine+threonine+proline 33% and 51% respectively). Since the soluble part of fraction 0–45 thus resembles fraction 43–50 in amino acid proportion and profile, it follows that the insoluble portion must possess a more markedly different profile than the values quoted suggest. Fraction 0–43, amounting only to 1% of the total, is presumably a carry-over from fraction 0–45 and the evident similarities in behaviour need no further comment.

As the precipitating concentration of ethanol rises above 50% and the proportion of peptide increases, the amino acid profile changes character, the high

proportion of serine+threonine+proline decreasing and the proportions of aspartic acid+glutamic acid rising. The profile for fraction 60–75, for example, is clearly markedly different from that for fraction 43–50; for fraction 75+, no relationship to the active fractions is apparent, but the values show several resemblances to those for human serum albumin (Heimburger *et al.*, 1964).

Where comparison is possible, the values of the physicochemical parameters in Table 3 show general agreement with those found previously for serologically active materials; for glycoprotein 376/50–56, for example, which is the closest parallel to glycoprotein 603/43–50, the values are: amino acid 15%, dn/dc 1.64, $\bar{v}$ 0.633, ρ^0 1.318, Γ' 0.509, $s_{25,w}$ 8.04S, M_w^0 1060000, M^{app} 133000, all values being in the units used in Table 3 (Creeth & Knight, 1967, 1968; Creeth & Denborough, 1970).

Many trends in the data are evident, which are discussed below in relation to the analytical variations; here it is necessary only to point out a few features of particular relevance. (i) The values of the specific refraction increment and partial specific volume show the expected behaviour, running from values close to those of carbohydrates (approx. 1.46×10^{-7} and 0.59 ml/g respectively; Gosting & Morris, 1949) for the fractions low in peptide to those characteristic of proteins (approx. 1.85×10^{-7} and 0.74 ml/g respectively; see, e.g., Charlwood, 1957) for fraction 75+. (ii) The buoyant densities exhibit similar behaviour; a high-molecular-weight dextran (Pharmacia Ltd., Uppsala, Sweden) was found to give ρ^0 1.483 g/ml in Cs_2SO_4 , whereas proteins have ρ^0 1.24 g/ml (Ifft & Vinograd, 1966). (iii) The buoyant density of the insoluble component of fraction 0–45, after resuspension in Cs_2SO_4 [Fig. 1b(i)] was approx. 1.320 g/ml; that is, much lower than the value for the soluble component reported in Table 3, and approximately in keeping with its high peptide content. As shown in Fig. 1b(ii) the insoluble component is not dispersed by CsBr, a reagent which is effective in solubilizing some secretions insoluble in other caesium salts. (iv) For the most active fractions the values of the apparent molecular weights in the salt gradient are approximately an order of magnitude lower than the true M_w^0 values found from conventional sedimentation equilibrium, as was found for all other cyst glycoproteins investigated (Creeth & Denborough, 1970). This reflects the effects of heterogeneity in density arising from variations in the proportion of carbohydrate. As the proportion of peptide rises, the values of M^{app} approach those found for M_w^0 .

Discussion

Because of the large difference in buoyant density between proteins and carbohydrates (1.24 g/ml and

(a)

(b)

EXPLANATION OF PLATE I

Immunodiffusion plates

Cyst 603 fractions were in outer wells, and anti-whole human serum in centre wells. (a) Well 1, fraction 0-45; 2, fraction 45-50; 3, fraction 0-43; 4, fraction 43-50; 5, fraction 50-60; 6, fraction 60-75. (b) Well 1, fraction 0-45; 2, human serum albumin; 3, fraction 75+; 4, human immunoglobulin G; 5, fraction AmS sol; 6, albumin. For the fractions, 20 μ l of solution containing 10 mg/ml of glycoprotein was used; for the albumin and globulin wells, 20 μ l of solution containing 1 mg/ml of protein was used. The centre wells contained 30 μ l of antiserum 1:1 diluted.

1.48 g/ml in Cs_2SO_4 ; 1.29 g/ml and 1.60 g/ml in CsCl), it has been apparent from the earliest studies (e.g. Franek & Dunstone, 1966, 1967; Dunstone, 1969; Creeth & Denborough, 1970) that density-gradient methods should be particularly useful with glycoproteins. Small differences in composition or structure should be reflected in changes of buoyant density, a quantity which is readily measurable to 0.001 g/ml. Although Cs_2SO_4 has a lower resolving power than CsCl (Ifft *et al.*, 1961), it was selected because the higher gradient for a given speed enables materials dispersed in density to be banded in a single experiment. With more homogeneous materials, CsCl could be used with some advantage, although considerably higher speeds would be necessary.

In the present study, where the composition of the glycoproteins in the series varies over quite wide limits, several correlations with the buoyant density are evident from the Tables and other less predictable relationships emerge.

Graphical presentation of some of the results enables the correlations to be seen more clearly. Figs. 2(a) and 2(b) show the main physical parameters as a function of the proportion of peptide. Attention has been restricted to the serologically active constituents; to avoid confusion straight lines or smooth curves have been drawn through the various sets of data.

Fig. 2(a) illustrates the relatively narrow band within which lie all buoyant densities at a given peptide content. It is apparent also that the fractions AmS sol, 45-50, 43-50 and 50-60 define the lower

straight line, whereas the other fractions lie quite close to the upper line. Because of the convergence of the lines at low peptide content, the differences at the left-hand edge of the Figure are less significant.

Similarly, the partial specific volumes lie fairly closely about a line: the lack of values for the incompletely soluble fractions precludes detection of any dual dependence analogous to that of the buoyant densities. The molecular weights of the ethanol fractions lie on a smooth curve, which becomes nearly vertical at low peptide content; however, fraction AmS sol is exceptional.

Fig. 2(b) demonstrates the close correlation of sedimentation coefficient and peptide content. Although a fairly narrow band defines the behaviour of all six ethanol fractions (fraction AmS sol is anomalous), it is evident that the points may be most closely fitted by two nearly parallel lines. The lower line relates to three of the four fractions which defined the lower line in the buoyant-density relationships, and the fractions on the upper line are similarly those defining the upper line in buoyant density. Since the buoyant density and the sedimentation coefficient are entirely independent measurements, but are both determined to some extent by particle density, it seems likely that two families of glycoprotein molecules exist, both having variable peptide content, but which also have distinctive features, causing them to possess different but characteristic relationships between the proportion of peptide and the physicochemical parameters. The existence of two families of glycoproteins, suggested by the density-gradient

Fig. 2. Physicochemical parameters of glycoprotein fractions as a function of proportion of peptide

Fraction numbers identifying points are as given in Table 1. (a) Molecular weights (■), buoyant densities (●) and partial specific volumes (▲). (b) Sedimentation coefficients (●) and selective solvation parameters (■).

Table 4. *Classification of the glycoprotein fractions from cyst no. 603 based on analytical variations*

Fraction		Peptide content (g of amino acids/100g of glycoprotein)	Amino acid content ($\mu\text{mol}/$ 100 μmol of total amino acids)		Molar ratios	
			Ser+Thr+Pro	Asp+Glu	GalNAc/Ser+Thr	GlcNAc/GalNAc
AmS sol	(1)	7.1	53.0	8.7	0.84	6.3
0-45 Sol	(2s)	11.6	47.2	12.6	0.77	4.6
45-50	(3)	13.0	50.8	10.5	0.77	3.6
43-50	(5)	11.1	50.9	9.8	0.78	4.5
50-60	(6)	20.2	49.2	13.3	0.65	2.2
0-45 Total	(2)	27.2	33.1	17.4	0.51	3.7
0-43 Total	(4)	23.9	31.6	25.2	0.50	4.1
60-75	(7)	45.8	26.9	25.3	0.28	3.3

patterns observed with several of the fractions, is also strongly supported by the behaviour of the original cyst fluid in density-gradient analyses (Creeth & Denborough, 1970); as in many other cases investigated (Denborough & Creeth, 1970), two distinct glycoprotein components were recognizable.

If the features distinguishing the two families were structural in origin, they would be expected to be revealed by trends in the analytical values. The relevant values are summarized in Table 4, where the fractions have been placed in two groups, largely on the basis of their physical properties. It is apparent that the analytical variations support this classification, except that fraction 0-45 (sol) fits more closely into the first analytical group than the second. For (serine+threonine+proline), application of the *t* test shows the differences of the means of the two groups to be highly significant: $0.99 < P < 0.995$. Similar calculations for (aspartic acid+glutamic acid) gave $0.975 < P < 0.99$.

The values for *N*-acetylgalactosamine are of particular interest, because this component probably occurs in Le^a substances only in the glycosidic linkage with the hydroxy amino acids (see Watkins, 1972); accordingly the molar ratio *N*-acetylgalactosamine/(serine+threonine) can be taken as a measure of the extent of glycosylation. It is clear that the fractions fall into two groups in terms of this ratio. Although the values for the molar ratios must be treated with caution, the correlation between the proportion of carbohydrate and the extent of glycosylation of hydroxy amino acids is firmly established. The relatively low values for the peptide-rich fractions are particularly noteworthy; the values for the most active fractions are similar to those reported by Donald (1973) for other H- and Le^a-active materials. If the two families were to differ consistently in extent of glycosylation, it would follow that, at a fixed proportion of carbohydrate, the more extensively oligosaccharide would necessarily have a shorter average oligosaccharide chain length. The molar ratio of *N*-acetylglucosamine to *N*-acetylgalactosamine is a

measure of chain length, for the reason given above. The values for this ratio also fall into distinct patterns for the two classes. In one class, the ratio falls steadily as the proportion of peptide increases, whereas in the second class, namely fractions 0-45, 0-43 and 60-75, the ratio is essentially constant over a wide range of peptide content. The analytical data thus provide some support for the proposal that the two families differ in their extent of glycosylation of hydroxy amino acids. The separation brought about by ethanol fractionation is quite incomplete, the individual fractions merely being enriched or depleted, as shown by the density-gradient patterns.

Although this interpretation is consistent with most of the observations, a second possibility is suggested by the work of Donald (1973), in terms of variability in extent of cross-linking. Donald (1973) confirmed earlier suggestions (Dunstone & Morgan, 1965; Kristiansen & Porath, 1968) that two distinct types of peptide region exist in blood-group-specific glycoproteins, by showing that a significant proportion of the peptide is removable from most preparations, without loss of carbohydrate, by the action of Pronase. The peptide so removed, which is relatively rich in aspartic acid, glutamic acid and cystine, could well provide cross-links between simpler glycoprotein units. Table 4 shows that the two classes do differ markedly in this respect, fractions 0-45, 0-43 and 60-75 being considerably richer in aspartic acid and glutamic acid than the fractions in the other class; accordingly this series may indeed be more extensively cross-linked than the series AmS sol, 45-50, 43-50 and 50-60.

Donald's (1973) findings may also provide an explanation for the exceptional properties of the fraction AmS sol. This fraction gives the most symmetrical density-gradient pattern, is carbohydrate rich but relatively low in molecular weight; moreover, the relation between the sedimentation coefficient and the molecular weight shows that the conformation is significantly less expanded than that of the other fractions. The property of solubility in hot

saturated $(\text{NH}_4)_2\text{SO}_4$ is not understood: B-specific preparations examined so far are fully soluble, a fact which does not appear in this case to be correlated with the proportion of peptide or a more compact conformation (cf. the properties of B substance 376/50–56 summarized above). Donald (1973) observed that a number of carbohydrate-rich residues obtained after Pronase treatment, which were fully active serologically, had similar molecular weights (approx. 7×10^5) even though some were derived from completely insoluble preparations. Moreover, these Pronase-treated materials have now been found to be fully soluble in hot saturated $(\text{NH}_4)_2\text{SO}_4$ when tested by the procedure described by Pusztai & Morgan (1961). It is suggested accordingly that the exceptional properties of the $(\text{NH}_4)_2\text{SO}_4$ -soluble glycoprotein arise because it shares with the Pronase-treated materials a relative paucity of the cross-linking peptide, or alternatively, if the peptide is present in fraction 603 AmS sol, that it fails to act as a cross-link. The latter suggestion is more in keeping with the amino acid compositions in Table 2.

Verification of these suggestions as to the role of a cross-linking peptide must await the results of studies of the action of Pronase on the same series of glycoprotein fractions.

The implications of the other correlations may now be summarized.

(i) Although $\bar{v}$ and ρ^0 show the expected inverse dependency, the derived quantity Γ' , the selective solvation parameter (Fig. 2b), shows that the buoyant density falls with increasing amounts of peptide more steeply than can be accounted for by the change in $\bar{v}$, i.e. the solvation by water increases also, for the series up to fraction 60–75. The value of Γ' for the inactive fraction 75+ is characteristic of proteins in Cs_2SO_4 (Ifft & Vinograd, 1966).

(ii) The sedimentation coefficients and the molecular weights demonstrate that solubility in ethanol decreases as the size of the molecule increases. Although this follows general expectations, in this case the least ethanol-soluble fractions are those richest in carbohydrate. Although fraction 0–45 appears to be an exception to this, the observation is still valid if only the soluble component is considered. These findings confirm, and give quantitative status to, the earlier observation (Creeth & Denborough, 1970) that the densest molecules are also those of highest molecular weight.

(iii) The observation that the insoluble component of fraction 0–45 is rich in peptide, and particularly in the content of acidic amino acids, is fully consistent with the results of Dunstone & Morgan (1965) and Dunstone (1969), where the correlation was firmly established between insolubility and high content of peptide (see also Donald, 1973). It is noteworthy that the insoluble component makes up 30% of the recovered glycoprotein of the cyst.

In conclusion several correlations have been established between analytical composition of blood-group-specific substances and their physical and macromolecular properties. It is hoped that these correlations, although mostly established for glycoprotein fractions from a single cyst fluid, will be of use in the general field of glycoproteins.

We are very grateful to Professor W. M. Watkins, F.R.S., for the serological determinations, and for much advice and encouragement. We are grateful also to Mr. J. R. Horton and Mr. H. S. Murray, who provided skilled technical assistance, and to the Medical Research Council for a supporting grant. It is a pleasure to acknowledge the support of the Wellcome Trust in providing the Wang 720B computer.

References

- Charlwood, P. A. (1957) *J. Amer. Chem. Soc.* **79**, 776–781
- Creeth, J. M. (1964) *Proc. Roy. Soc. Ser. A* **282**, 403–421
- Creeth, J. M. & Denborough, M. A. (1970) *Biochem. J.* **117**, 879–891
- Creeth, J. M. & Knight, C. G. (1967) *Biochem. J.* **105**, 1135–1145
- Creeth, J. M. & Knight, C. G. (1968) *Spec. Publ. Chem. Soc.* **23**, 303–313
- Denborough, M. A. & Creeth, J. M. (1970) *Clin. Chim. Acta* **30**, 447–451
- Donald, A. S. R. (1973) *Biochim. Biophys. Acta* **317**, 420–436
- Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956) *Anal. Chem.* **28**, 350–356
- Dunstone, J. R. (1969) *Eur. J. Biochem.* **9**, 128–132
- Dunstone, J. R. & Morgan, W. T. J. (1965) *Biochim. Biophys. Acta* **101**, 300–314
- Franek, M. D. & Dunstone, J. R. (1966) *Biochim. Biophys. Acta* **127**, 213–222
- Franek, M. D. & Dunstone, J. R. (1967) *J. Biol. Chem.* **242**, 3460–3467
- Gibbons, M. W. (1955) *Analyst (London)* **80**, 268–276
- Gibbons, R. A., Creeth, J. M. & Denborough, M. A. (1970) in *Blood and Tissue Antigens* (Aminoff, D., ed.), pp. 307–324, Academic Press, London and New York
- Gosting, L. J. & Morris, M. S. (1949) *J. Amer. Chem. Soc.* **71**, 1998–2006
- Gratzer, W. B., Creeth, J. M. & Beaven, G. H. (1972) *Eur. J. Biochem.* **31**, 505–509
- Heimburger, N., Heide, K., Haupt, H. & Schultze, H. E. (1964) *Clin. Chim. Acta* **10**, 293–307
- Hexner, P. E., Radford, L. E. & Beams, J. W. (1961) *Proc. Nat. Acad. Sci. U.S.A.* **47**, 1848–1852
- Ifft, J. B. & Vinograd, J. (1966) *J. Phys. Chem.* **70**, 2814–2822
- Ifft, J. B., Voet, D. H. & Vinograd, J. (1961) *J. Phys. Chem.* **65**, 1138–1145
- Kristiansen, T. & Porath, J. (1968) *Biochim. Biophys. Acta* **158**, 351–357
- Morgan, W. T. J. (1967) in *Methods in Immunology and Immunochemistry* (Williams, C. A. & Chase, M. W., eds.), vol. 1, pp. 75–81, Academic Press, London and New York
- Pusztai, A. & Morgan, W. T. J. (1961) *Biochem. J.* **80**, 107–121

- Pusztai, A. & Morgan, W. T. J. (1963) *Biochem. J.* **88**, 546–555
- Tanford, C., Marler, E., Jury, E. & Davidson, E. A. (1964) *J. Biol. Chem.* **239**, 4034–4040
- Teller, D. C., Horbett, J. A., Richards, E. G. & Schachman, H. K. (1969) *Ann. N. Y. Acad. Sci.* **164**, 66–101
- Wake, R. G. & Baldwin, R. L. (1962) *J. Mol. Biol.* **5**, 201–216
- Warren, L. (1959) *J. Biol. Chem.* **234**, 1971–1975
- Watkins, W. M. (1972) in *Glycoproteins: Their Composition Structure and Function* (Gottschalk, A., ed.), 2nd edn., pp. 830–891, Elsevier Publishing Co., Amsterdam