ADVANCES IN CARBOHYDRATE CHEMISTRY AND BIOCHEMISTRY, VOL. 47

THE MACROSTRUCTURE OF MUCUS GLYCOPROTEINS IN SOLUTION

By Stephen E. Harding

Department of Applied Biochemistry and Food Science, University of Nottingham, Sutton Bonington LE12 5RD, England

I. Introduction 345
II. Composition 347
III. Primary and Secondary Structure: The Basic Unit of the Mucus Glycoprotein .. 349
IV. Tertiary Structure: Assembly of Basic Units 352
1. Thiol Reduction: "Subunits" 352
2. Branched Models for Mucin Structure 353
3. Mucin Molecular Weights 356
4. Linear Models for Mucin Structure 367
V. The Gross Conformation of Mucus Glycoproteins in Solution 370
Mucins as Polyelectrolytes 373
VI. Mucin Heterogeneity 374
1. Polydispersity and Self-Association Behavior 377
2. Interactions with Other Macromolecules 380
VII. Summary and Prospects 380

I. Introduction

The importance of mucus glycoproteins in health and disease is undisputed: their biological function is closely related to their conformation, and yet they provide a macromolecular system that is very difficult to analyze. This article describes, in terms of both structural integrity (susceptibility to degradation phenomena) and physical behavior (very high thermodynamic non-ideality, discrete and quasi-continuous polydispersity, and possible self-association phenomena), how these difficulties have been addressed, focussing particularly on the suitability and difficulties associated with certain physical techniques. The various models for mucus glycoprotein "macrostructure" (namely, assembly, gross conformation, and heterogeneity) in solution—which have been based on various interpretations of the data provided by these techniques—are discussed, and the most likely model is assessed.

Over the past few years, considerable interest has arisen, across a wide spectrum of scientific disciplines, concerning the structure and behavior of mucus glycoproteins or "mucins." The importance of their role in health and disease is indisputable,[1-4] and yet they provide a heterogeneous macromolecular system very difficult to analyze.[5] For convenience, the term "heterogeneity" is used here in its widest sense to describe any system in which the solute species do not have a single value of molecular weight, no matter what the origin of this variation may be. As a result, our knowledge of these molecules is some two decades behind that of other more "fashionable" macromolecules, such as proteins or nucleic acids, wherein the subtleties of individual structure are now quite well understood.

The principal difficulty in attempting any form of structural or physicochemical analysis of mucins arises from three fundamental properties (see, for example, Refs. 6–10): *1*, they are highly *non-ideal* in the thermodynamic sense; *2*, they are *polydisperse* (that is, they consist of components of different molecular weight and partial specific volume that are not in chemical equilibrium with each other); and *3*, they *may be self-associating* in solution (although there is growing evidence to suggest that they are not). Largely because of these difficulties, but also because of sample variability through enzymic and possible mechanical degradation phenomena, there has been some level of disagreement in the literature as to the size, subunit composition, mode of assembly, and gross conformation of the native macromolecule. This was highlighted in a penetrating series of papers which appeared[10a] in 1984. The purpose of the present article is to assess critically the literature, and formulate what may be the most likely model for a mucus glycoprotein in solution.

Knowledge of the structure and behavior of mucus glycoproteins in solution is crucial for a proper understanding of their behavior *in situ*, where they often exist in a more concentrated form. It is these molecules which generally dictate the physical properties of mucus (namely, high viscosity, viscoelastic

(1) F. Avery-Jones, *Br. Med. Bull.* 34 (1978) 1–16.
(2) A. Allen, *Trends Biochem. Sci.*, 8 (1983) 169–173.
(3) A. Allen, in L. R. Johnson (Ed.), *Physiology of the Gastrointestinal Tract*, Raven Press, New York, 1981, pp. 617–639.
(4) H. R. P. Miller, J. F. Huntley, and G. R. Wallace, *Immunology*, 44 (1981) 419–429.
(5) I. Carlstedt, J. K. Sheehan, A. P. Corfield, and J. T. Gallagher, *Essays Biochem.*, 20 (1985) 40–76.
(6) S. E. Harding and J. M. Creeth, *IRCS Med. Sci.*, 10 (1982) 474–475.
(7) J. M. Creeth and C. G. Knight, *Biochem. J.*, 105 (1967) 1135–1145.
(8) J. M. Creeth and C. G. Knight, *Chem. Soc. Spec. Publ.*, 23 (1968) 303–313.
(9) S. E. Harding, *Biochem. J.*, *219* (1984) 1061–1064.
(10) S. E. Harding, *Biophys. J.*, 47 (1985) 247–250.
(10a) *Biochem. Soc. Trans.*, 12 (1984) 612–621.

and gel characteristics[1-3,11]). Mucins from a wide variety of sources seem to have the same basic physical properties and hence, it is reasonable to assume that they have the same basic structure in solution[12,13] but vary in their molecular size.[13-15]

It is not the purpose of this article to discuss in detail the gelation properties of these macromolecules, as consideration of this can be found elsewhere (see, for example, Ref. 12). The glycoproteins found in submandibular and submaxillary secretions will also not be considered. Although these substances are also referred to as mucins, they have somewhat different characteristics (namely, a considerably lower proportion of carbohydrate, shorter carbohydrate chains of different structure and composition) and properties (such as a lower viscosity).[16] A study has however[17] shown that they have a greater overall homology with other mucins in terms of assembly and conformation than was perhaps originally presumed.

The suitability of certain physical techniques (such as low-speed sedimentation equilibrium in the analytical ultracentrifuge, and electron microscopy) will be assessed and comment will be made regarding the inherent difficulties of others, such as light-scattering, calibrated gel chromatography, and free-boundary diffusion in the ultracentrifuge, interpretations from which have led to some of the disagreements in the literature. Particular emphasis will be laid on the utility of low-speed, sedimentation-equilibrium procedures in the analytical ultracentrifuge for characterizing the molecular size and heterogeneity of mucus glycoproteins, and some developments that make the technique particularly suited for characterizing these molecules will be described. However, before the macrostructure is considered, it may be helpful to review briefly some well established facts about mucin composition and primary structure.[18]

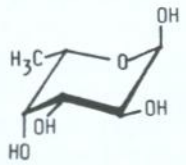
II. Composition

Native mucus secretions normally contain ~ 1% of salts and other dialyzable components, 0.5 – 1% of proteins, a similar proportion of carbohydrate-

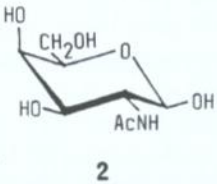
(11) J. M. Creeth, *Br. Med. Bull.*, 34 (1978) 17–24.
(12) A. Silberberg and F. A. Meyer, in E. N. Chantler, J. B. Elder, and M. Elstein (Eds.), *Mucus in Health & Disease*, Vol. II, Plenum, New York, 1982, pp. 115–133.
(13) I. Carlstedt and J. K. Sheehan, *Biochem. Soc. Trans.*, 12 (1984) 615–617.
(14) F. A. Meyer, *Biochim. Biophys. Acta*, 493 (1977) 272–282.
(15) J. K. Sheehan and I. Carlstedt, *Ciba Found. Symp.*, 109 (1984) 157–172.
(16) H. D. Hill, J. A. Reynolds, and R. L. Hill, *J. Biol. Chem.*, 252 (1977) 3791–3798.
(17) R. Shogren, A. M. Jamieson, J. Blackwell, and N. Jentoft, *Biopolymers*, 25 (1986) 1505–1517; see also, R. L. Shogren, A. M. Jamieson, J. Blackwell, P. W. Cheng, D. G. Dearbon, and T. F. Boat, *ibid.*, 12 (1983) 1657–1675.
(18) The present use of the terms "primary," "secondary," and "tertiary" structure for mucins does not necessarily correspond to their usage as applied to proteins.

rich glycoprotein, and ~95% of water.[11] Although comprising less than 1% of the total mucus secretion, it is this glycoprotein or mucin component which gives the mucus its characteristic high viscosity and viscoelastic characteristics.[11] The mucin component is normally extracted by using a two- (or more)stage density-gradient ultracentrifugation in cesium salts,[19] the glycoprotein component having a buoyant density (in CsCl) of ~1.5 g/mL as compared with ~1.3 g/mL for proteins, and ~1.7 g/mL for nucleic acids. Detailed extraction protocols have been given[15,20,21] in which the importance of meticulous inclusion of protease inhibitors has been clearly demonstrated.

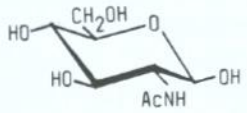
A mucus glycoprotein is composed, typically, of ~80% of carbohydrate which, for humans, is restricted to 5 monosaccharides: L-fucose (L-Fuc, **1**), *N*-acetylgalactosamine (GalNAc, **2**), *N*-acetylglucosamine (GlcNAc, **3**), galactose (Gal, **4**) and *N*-acetylneuraminic acid (NeuAc, **5**).[2,11,22]



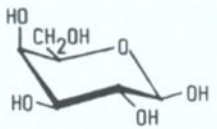
1



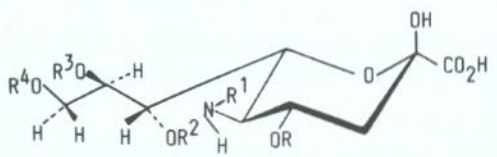
2



3



4



5

(19) J. M. Creeth, K. R. Bhaskar, J. R. Horton, I. Das, M. T. Lopez-Vidriero, and L. Reid, *Biochem. J.*, 167 (1977) 557–569.

(20) I. Carlstedt, H. Lindgren, J. K. Sheehan, U. Ulmsten, and L. Wingerup, *Biochem. J.*, 211 (1983) 13–22.

(21) I. Carlstedt, J. K. Sheehan, U. Ulmsten, and L. Wingerup, in Ref. 12, pp. 273–274.

(22) W. Pigman, in M. I. Horowitz and W. Pigman (Eds.), *Glycoconjugates*, Vol 1, Academic Press, New York, 1977, p. 132.

TABLE I

Amino Acid Composition[23] (Mol/100 Mol) of 2 Mucus Glycoproteins from an Ovarian Cyst ("603" and "485")

Amino acid	603	485
Asp	4.9	2.8
Thr	19.0	27.1
Ser	16.3	18.1
Glu	4.7	4.1
Pro	15.1	16.0
Gly	8.5	5.8
Ala	12.7	9.6
Cys	nd[a]	nd[a]
Val	4.3	3.7
Met	0.3	0.4
Ile	1.5	1.8
Leu	2.1	1.9
Tyr	0.7	0.5
Phe	1.0	1.0
His	3.2	2.8
Lys	1.9	1.3
Arg	3.2	2.6
Total peptide content (%)	7.6	12.0

[a] n.d., not determined.

The protein moiety is considered to consist of a single polypeptide chain about which the carbohydrate is built. Approximately one in every three residues is either L-serine or L-threonine, and the O-3 atoms of these provide the sites for glycosidic linkage. Understandably, therefore, amino acid composition data (see Table I) reveal a large proportion of L-serine and L-threonine. The significance of the large percentage of L-proline will be discussed in the following Section.

III. Primary and Secondary Structure: The Basic Unit of the Mucus Glycoprotein

The molecular weights of mucus glycoproteins range from $\sim 0.5 \times 10^6$ to 16.0×10^6, and it is now widely accepted that the mucins of higher molecular weight are made up[12,22,24,25] of multiples of a basic unit having a molecular

(23) J. M. Creeth, B. Cooper, A. S. R. Donald and J. R. Clamp, *Biochem. J.*, 211 (1983) 323–332.

(24) S. E. Harding, J. M. Creeth, and A. J. Rowe, *Proc. Int. Conf. Glycoconjugates, 7th*, Olsson-Reklambyra, Sweden, 1983, pp. 558–559.

(25) A. Silberberg, *Biorheology*, 24(1987) 605–614.

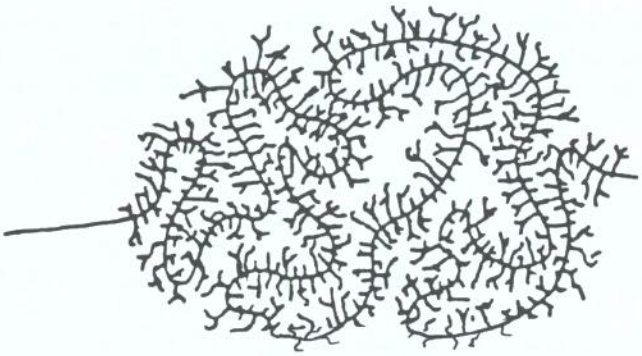


FIG. 1.—Schematic Mucin Basic Unit.[22,24] [The continuous line represents the polypeptide, and the attached chains, the oligosaccharides. Although two regions devoid of carbohydrate are shown, there may be only one.]

weight of $(0.4-0.6) \times 10^6$: protease digestion (and thiol reduction, but see later) of a wide range of mucins generally produces materials having molecular weights of this order.[12,25-27]

The basic unit[12,22,24] of a mucin is a single polypeptide chain which consists of two (or three) distinct regions, namely, one heavily glycosylated central core and one (or two) end peptide segments which are rich in cysteine and acidic groups but virtually devoid of carbohydrate (see Fig. 1). The central glycosylated region is rich in serine, threonine, and proline, with many multi-branched oligosaccharides ranging in length from 5 to 30 residues.

Insofar as "secondary structure" is concerned, there seems to be no evidence of regularly folded structures normally associated with globular proteins (α helices, β sheets, and the like). The polypeptide is, however, presumed, for three reasons, to adopt a loosely coiled structure. Firstly, it accounts for the high levels of proline in these macromolecules (see Fig. 2). A prolyl residue is most commonly found in the *trans* configuration in polypeptides: from minimum-energy considerations,[28] there are two allowed conformations, a compact form ($\psi = -55°$) and an extended form ($\psi = +145°$). The relative ease with which an isolated prolyl residue can adopt the compact form makes it ideal for producing bends and turns in the polypeptide backbone.[28] An alternative explanation for the large proportions of proline could be its use in the formation of helical and fibrous structures. However, there is no evidence for helical structures (in mucins) corresponding to polymers or copolymers of proline as found in, for example, collagen.

(26) M. Scawen and A. Allen, *Biochem. Soc. Trans.*, 3 (1975) 1107-1109.
(27) J. M. Creeth, unpublished results.
(28) P. R. Schimmel and P. J. Flory, *J. Mol. Biol.*, 34 (1968) 105-120.

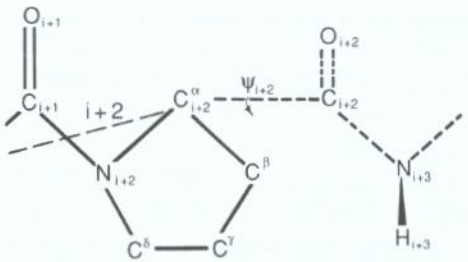


FIG. 2.—Schematic Diagram of a Portion of a Polypeptide Chain Containing an Isolated Proline Residue (from Refs. 28 and 29). [An isolated proline residue can produce a "kink" or "turn" in a polypeptide chain, if the C–C$^{\alpha}$ bond angle (ψ) is ~ −55°.]

A second reason supporting a loosely coiled structure for the basic unit is that such a structure would be more efficient at engulfing and immobilizing local solvent than an extended form.[24,29]

The third line of support for a coiled domain is from hydrodynamic data on T-domains.[30] These are formed by protease (namely, trypsin) digestion of the native mucin and are essentially equivalent to a basic unit but are lacking in naked peptide. Values for the ratio of the sedimentation concentration regression coefficient, k_s, to the intrinsic viscosity, $[\eta]$, are close to the value

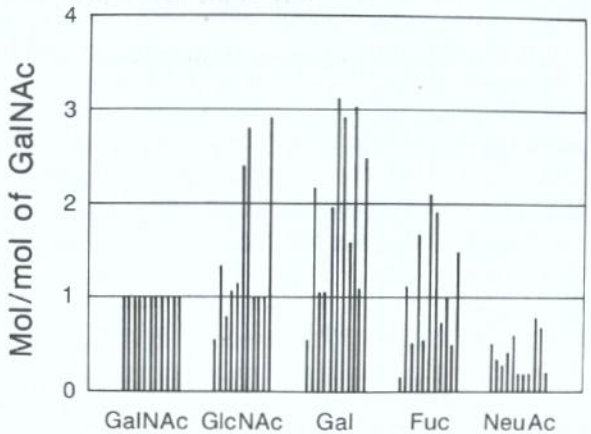


FIG. 3.—Variability[12] of Carbohydrate Side-Chain Composition for Eleven Mucins (Relative to GalNAc).

(29) C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry*, Part I, Freeman, San Francisco, 1980, p. 270.
(30) S. E. Harding, A. J. Rowe, and J. M. Creeth, *Biochem. J.*, 209 (1983) 893–896.

expected for a spheroidal or randomly coiled molecule (see, for example, Refs. 31 and 32).

Although amino acid composition data are relatively invariant from mucin to mucin, much greater variability seems to exist as far as the carbohydrate moiety is concerned[12]: Fig. 3 compares composition data for a number of different mucins normalized to the relative amounts of GalNAc present in α linkage. The first sugar in the chain is always GalNAc, and the chains are often terminated by L-fucose or NeuAc, with at least one branch per chain on average. It is now well established, from, for example, the results of gas–liquid chromatography (g.l.c.) and n.m.r. spectroscopy, that considerable microheterogeneity, both in length and complexity, exists among the side chains.[33–35] However, as Silberberg and Meyer[12] pointed out, mucus function may not necessarily depend upon a specific sequence in the oligosaccharide side-chains, as the protein moiety contains the more-specific features needed to build up a supramolecular structure, and thus produce the requisite physical properties.

IV. Tertiary Structure: Assembly of Basic Units

How are the basic units arranged to form the macrostructure of the larger mucins? Evidently, because of the nature of the degradation products of pronase digestion, the linkage is by way of the naked peptide regions.

1. Thiol Reduction: "Subunits"

It was originally suggested[26] that the links between the basic units are through intermolecular disulfide bridging between cysteine residues in the naked end regions. Reduction of mucins by thiols produced the M_r ~500,000 forms (see, for example, Refs. 26 and 36). However, Creeth[27] observed a variety of forms in the $0.5-2.0 \times 10^6$ region, and subsequent observations on cervical and other mucins by Carlstedt and Sheehan[13,15,31] yielded, in the presence of guanidine hydrochloride (Gu·HCl), species having forms $M_r \sim 2 \times 10^6$ which they referred to as

(31) J. K. Sheehan and I. Carlstedt, *Biochem. J.*, 217 (1984) 93–101.
(32) J. M. Creeth and C. G. Knight, *Biochim. Biophys. Acta*, 102 (1965) 549–558.
(33) H. Van Halbeek, L. Dorland, J. F. G. Vliegenthart, W. E. Hull, G. Lamblin, M. Lhermitte, A. Boersma, and P. Roussel, *Eur. J. Biochem.*, 127 (1982) 7–20.
(34) H. Van Halbeek, L. Dorland, J. F. G. Vliegenthart, J. Montreuil, B. Fournet, and K. Schmid, *J. Biol. Chem.*, 256 (1981) 5588–5590.
(35) G. Lamblin, M. Lhermitte, A. Klein, P. Roussel, H. Van Halbeek, and J. F. G. Vliegenthart, *Biochem. Soc. Trans.*, 12 (1984) 599–600.
(36) A. S. Mall, D. A. Hutton, R. M. Coan, L. A. Sellers, and A. Allen, *Biochem. Soc. Trans.*, 16 (1988) 585–586.

TABLE II
Terminology for Mucus Glycoprotein Components

Component	Description
Basic unit	basic mucin building block, $M_r \sim 5 \times 10^5$; glycosylated central-core region plus 1 or 2 regions of naked peptide
T-domain	product remaining after treatment of mucin with proteases; equivalent to a basic unit minus naked peptide
Subunit	product remaining after reduction of mucin by thiols, whatever form this may take

"subunits." Table II distinguishes between "basic unit" and "subunit"; henceforth in this article, the term subunit refers to the macromolecular entity produced by reduction of the native mucin by thiols, whatever the value of M_r for this entity may take. Subsequent action of trypsin produced forms having $M_r \sim 0.3-0.4 \times 10^6$ that are referred to as T-domains.[13,15,31] These workers argued that the earlier, low M_r values for the subunit could have been a result of proteolytic or mechanical degradation phenomena occurring during and after the extraction process, thus strengthening the case for the meticulous inclusion of protease inhibitors. It would appear, therefore, that the linkage between basic units (equivalent to T-domains plus naked peptide) is not exclusively through disulfide bridge attachment, but rather by peptide, or some other, linkage. Mall and coworkers[36] demonstrated, however, that pig gastric mucin, extracted in the presence of inhibitors but analyzed in the absence of Gu · HCl, gives a value of ~ 500,000 for the molecular weight of the subunit. It has been claimed[36,37] that the presence of Gu · HCl produces irreversible aggregation phenomena leading to anomalously high values ($\sim 2 \times 10^6$) for the subunit, although this observation conflicts with data for whole bronchial mucins[38] (see Section IV,2). Silberberg[25] demonstrated that the linkages between basic units or subunits could be lectin-like: intramolecular disulfide bridging could stabilize the naked peptide into a particular conformation which forms a binding site for a specific, but unusual, sugar sequence on the side chain of the adjacent unit.

Whether the units are linked into branched or linear arrays has been the subject of much debate which will now be considered.

2. Branched Models for Mucin Structure

One of the first proposals as to the form of this tertiary structure, at least for one particular mucin, was a branched structure, referred to as a "star" or

(37) A. Allen, A. Bell, M. Mantle, and J. P. Pearson, in Ref. 12, pp. 115–133.
(38) S. E. Harding and J. M. Creeth, *Biochim. Biophys. Acta*, 746 (1983) 114–119.

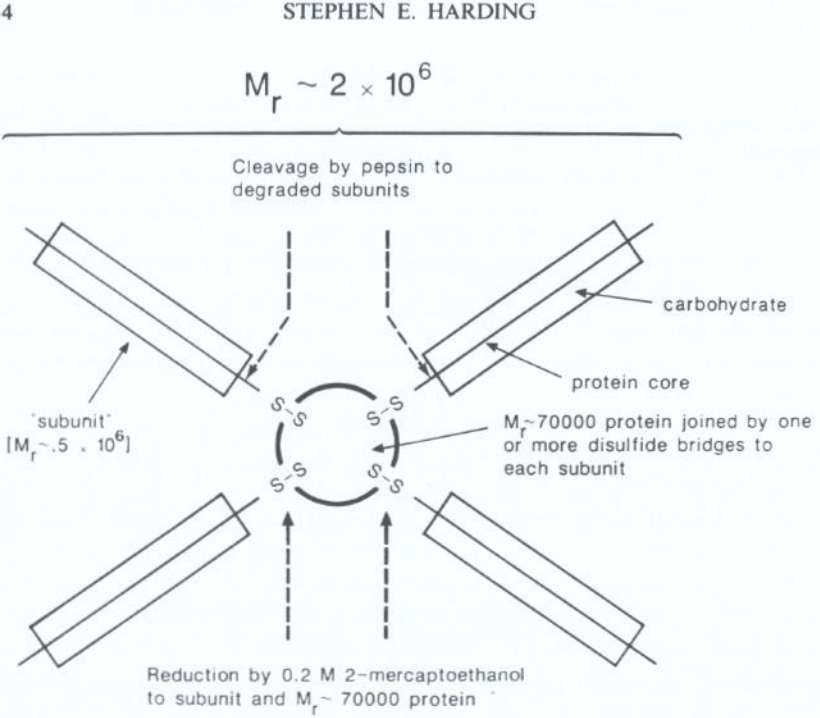


FIG. 4.—Branched Model for Pig Gastric Mucin.[37]

"windmill" form[2,3,37] (see Fig. 4). This model was based on several years of exhaustive work on pig-gastric mucin (p.g.m.) by Allen, Pain, and coworkers[26,37,39-41] in which a spheroidal model for the gross conformation was successfully predicted,[40] in agreement with others.[14,15,24,30,31] Their model for the assembly was supported by *(i)* the susceptibility of the mucin to attack by proteases or thiols[41]; *(ii)* estimates of molecular weights for the native and reduced pig-gastric mucin[2,26]; *(iii)* end-group amino acid analysis of the components[37]; and *(iv)* the discovery of a 70,000 molecular weight, link protein.[41] A similar protein ($M_r \sim 110{,}000$) was discovered by the Forstners and coworkers[42] for mucus glycoprotein from small intestines: these workers also considered where this protein could be located in both branched and linear models.

Using the value of $\sim 2 \times 10^6$ obtained for the molecular weight of the native molecule, and 0.5×10^6 for the reduced form, it was concluded that

(39) M. Mantle, D. Mantle, and A. Allen, *Biochem. J.*, 195 (1981) 267–275.
(40) A. Allen, R. H. Pain, and T. Robson, *Nature*, 264 (1976) 88–89.
(41) J. P. Pearson, A. Allen, and C. W. Venables, *Gastroenterology*, 78 (1980) 709–715.
(42) M. Mantle, M. Potier, G. G. Forstner, and J. F. Forstner, *Biochim. Biophys. Acta*, 881 (1986) 248–257.

there must be four basic units. Treatment with 2-mercaptoethanol yielded the forms having a molecular weight of $\sim 0.5 \times 10^6$; also released was a protein of molecular weight ~70,000, as judged from sodium dodecyl sulfate-gel electrophoresis. A branched or "star" model was proposed as a likely form of assembly,[26] and this was later modified to accommodate the M_r = 70,000 protein in the center.[40] This model has been used by other workers.[43] From results on bovine-cervical, human-ear, and bronchial mucus, Meyer and Silberberg[44] suggested very tentatively that this four-unit form may be a rather general building-block for other mucins.

Although, as a first estimate, this model provided a good fit to the data then available, subsequent analyses suggested that the branched form is unlikely, for this[13,45] and other mucins. Crucial to the model for p.g.m. is the assumption of a four basic-unit form and a molecular weight of $\sim 2 \times 10^6$. Molecular weight values for p.g.m. vary widely, depending on the technique used and the preparation procedure. For example, for p.g.m. extracted by using 6 *M* Gu·HCl and proteinase inhibitors, Creeth and Cooper,[45] using low-speed sedimentation equilibrium, obtained a molecular weight of $\sim 9 \times 10^6$, not 2×10^6.

Carlstedt and Sheehan[13] obtained values for the M_r of p.g.m. as high as 45×10^6 (in the presence of 6 *M* Gu·HCl) by using light-scattering, and they indicated that the earlier, lower value of 2×10^6 was a result of inadequate precautions against protease or mechanical degradation. On the other hand (and parallel to the discussions over the size of subunits; see Section IV,1), Allen and coworkers[36,46] argued that the *higher* values of the M_r for native p.g.m. and other mucins represented non-covalently bound aggregates, a result of some irreversible, anomalous association caused by the presence (or subsequent removal[47]) of 6 *M* Gu·HCl. However, this effect on p.g.m. has not been reproduced by others,[13] and no such aggregation (or dissociation) phenomenon has been observed for bronchial and ovarian-cyst mucins.[37]

The effect of possible mechanical disruption of covalent linkages during high-shear solubilization of p.g.m. mucins from the gel state (and especially without the presence of 6 *M* Gu·HCl) was also demonstrated by Carlstedt and Sheehan[13] for p.g.m. It has been pointed out,[48] however, that this is not a problem for other large glycoconjugates, such as proteoglycans.

(43) V. A. Bloomfield, *Biopolymers*, 22 (1983) 2141–2154.
(44) F. A. Meyer and A. Silberberg, *Ciba Found. Symp.*, 54 (1978) 203–218.
(45) J. M. Creeth and B. Cooper, *Biochem. Soc. Trans.*, 12 (1984) 618–621.
(46) A. Allen, D. A. Hutton, D. Mantle, and R. H. Pain, *Biochem. Soc. Trans.*, 12 (1984) 612–615.
(47) D. Snary, A. Allen, and R. H. Pain, *Biochem. J.*, 141 (1974) 641–646.
(48) J. E. Fitzgerald, G. G. R. Green, F. W. Stafford, J. P. Birchall, and J. P. Pearson, *Clin. Chim. Acta*, 169 (1987) 281–298.

Although difficulties in sample integrity would appear to be the most likely explanation for discrepancies in the measured size of mucins, problems associated with the methodology of the physical techniques employed may also have contributed to them. For example, the low values of molecular weight obtained previously[26,36,38,39] could possibly be explained by some difficulties in the particular method employed, difficulties manifested by correlating distributions of sedimentation coefficient[49] with distributions of molecular weight[45]: for flexible, linear polymers, M_r is not a linear function of the sedimentation coefficient, s, but rather,[50] $M_r \propto s^2$, so the mean value of s would not necessarily correspond to the mean M_r.

Absolute values for molecular weights were usually obtained from the Svedberg equation[51,52] by combining measurements of s with the (translational) diffusion coefficient, D, measured by free-boundary spreading in the analytical ultracentrifuge (see, for example, Ref. 52). This procedure has inherent difficulties when applied to these substances. Diffusion measurements on polydisperse materials are difficult to interpret, and the broad molecular-weight distribution of the slowly diffusing mucins makes it a possibility that high-molecular-weight material is sedimented out of solution during the long time-periods used, although no significant losses have thus far been observed[52a] in the majority of cases.

3. Mucin Molecular Weights

What, then, is a suitable method for determining the molecular weight of a mucin? Because molecular weight is such an important parameter in the evaluation of mucin macrostructure, some of the other procedures that have been used, and the difficulties and possible pitfalls encountered, will be considered here.

a. Light-Scattering.—In a series of papers on cervical[13,53–55] and other mucins (see, for example, Ref. 56), Carlstedt and Sheehan also used the Svedberg equation. However, they measured the (Z-average) diffusion coefficient, D_z, in a different way, by quasi-elastic light-scattering (q.l.s.). They also used total-intensity light-scattering, where the intensity scattered by a

(49) R. H. Pain, *Symp. Soc. Exp. Biol.*, 34 (1980) 359–376.
(50) C. F. Tanford, *Physical Chemistry of Macromolecules*, Wiley, New York, 1961, p. 382.
(51) T. Svedberg and K. O. Pedersen, *The Ultracentrifuge*, Oxford University Press, 1940.
(52) Ref. 50, p. 380.
(52a) R. H. Pain, personal communication.
(53) I. Carlstedt, H. Lindgren, and J. K. Sheehan, *Biochem. J.*, 213 (1983) 427–435.
(54) J. K. Sheehan and I. Carlstedt, *Biochem. J.*, 217 (1984) 93–101.
(55) J. K. Sheehan and I. Carlstedt, in Ref. 24, pp. 599–600.
(56) I. Carlstedt and J. K. Sheehan, in Ref. 24, pp. 580–581.

glycoprotein solution is measured as a function of the concentration and the angle; using a biaxial, extrapolation procedure to zero angle and zero concentration (Zimm plot), it is possible to obtain values for the weight-average molecular weight M_w and the root-mean-square radius, R_g.

Both of these light-scattering procedures have found very wide application to a variety of macromolecular systems, but it is fair to say that they are extremely difficult to apply to such heterogeneous systems as mucins, largely because of problems of dust and traces of large aggregates, particularly for measurements at low angles. In the case of q.l.s. to minimize these effects, an angle of 90° is often employed[57]; although this does not lead to any appreciable error for rigid, non-spherical particles, extrapolation to zero angle is normally necessary for the flexible mucins, because of the possible finite contribution to the observed autocorrelation data from rotational diffusion phenomena.[57] Unfortunately, at low angles, any supramolecular aggregates will seriously affect the extrapolation, and, as a result, molecular-weight values can be on the high side, depending on the extent of contamination.

The Zimm-plot technique, like q.l.s., also involves a difficult extrapolation to zero angle. Another difficulty is that its application assumes that the particles are Rayleigh–Gans–Debye scatterers[58] (namely, that there is no change of phase or other distortions of the incident radiation by the particle). For the larger mucins in particular, this may not be the case.

Apparent agreement between Zimm plots and the Svedberg equation (using D_z values measured by q.l.s.) can be misleading, in that the same effects producing high M_r values (and high R_g values) from the Zimm method would also contribute to lower D_z values (and, hence, higher M_r values from the Svedberg equation).

The dramatic influence of even small proportions of aggregates on the results from q.l.s. has been clearly demonstrated by, for example, Preston and coworkers[59] in related studies on proteoglycans. Another difficulty that is often not reported is the contribution to error caused by concentration measurement (of the unsolvated solute); concentrations can rarely be measured to better than 5%, and will contribute error in both the Zimm plot and the values for the refractive increment used for evaluation of the constant. If q.l.s. and the Svedberg equation are used, errors in concentration will also be manifested in the extrapolations of the diffusion and sedimentation coefficients.

(57) R. E. Godfrey, P. Johnson, and C. J. Stanley, in D. B. Sattelle, W. I. Lee, and B. R. Ware (Eds.), *Biomedical Applications of Laser Light Scattering*, Elsevier, Amsterdam, 1982, pp. 373–389.

(58) For larger particles, see, for example, S. H. Chen, M. Holz, and P. Tartaglia, *Appl. Opt.*, 16 (1977) 187–194.

(59) G. S. Harper, W. D. Comper, and B. N. Preston, *Biopolymers*, 24 (1985) 2165–2173.

Because of these constraints, light-scattering techniques should not, where possible, be the method of choice: were light-scattering to be used, confirmation of results by using an independent procedure would be desirable. Agreement with molecular weights from, for example, low-speed sedimentation equilibrium would also give greater confidence in other potentially useful parameters from light-scattering, such as the equivalent Stokes radius, r_H, from q.l.s., and also the root-mean-square radius, R_g, and the second thermodynamic virial coefficient, B, from Zimm plots.

It should also be pointed out that light-scattering may be the only absolute method applicable to species having molecular weights larger than $\sim 15 \times 10^6$ (the upper limit for accurate measurements from low-speed sedimentation equilibrium procedures, unless ultra-short solution columns are used).

Despite these difficulties, and only after the employment of meticulous preparative procedures,[15] reproducible results were obtained by Carlstedt and Sheehan for T-domains, subunits, and whole mucins, which appear to be in agreement with other data. For whole pig-gastric mucin, for example, they obtained a molecular weight of $\sim 45 \times 10^6$, some 20 times the value reported by Allen and coworkers.[36] Lower values ($15-40 \times 10^6$) have also been obtained,[54,55] presumably because of sample variability. Using low-speed sedimentation equilibrium, Creeth and Cooper[45] obtained a lower value ($\sim 9 \times 10^6$).

b. Relative Techniques.—Relative techniques for the determination of molecular weight, such as gel electrophoresis and gel-permeation chromatography (g.p.c.), also have their difficulties. They are referred to as "relative techniques" for macromolecular molecular weight analysis, because they require calibration using standards of known molecular weight. G.p.c. is useful for giving a qualitative indication of size distribution (after adequate correction for diffusion broadening) but, because of difficulties of obtaining standards of similar size and conformation, results can only be relative. An adequate calibration procedure is, however, now available that avoids the problem of inappropriate standards by combining g.p.c. with low-speed sedimentation equilibrium measurements.[60,61] An important development, not yet applied to mucins, is the availability of an instrument (Wyatt Technology, Santa Barbara, CA, U.S.A.) facilitating g.p.c. measurements on-line to a multi-angle, laser light-scattering detector. This serves two purposes; it *1*, provides an on-line "clarification" of macromolecular solutions prior to light-scattering, overcoming the principal difficulty already referred to in

(60) A. Ball, S. E. Harding, and J. R. Mitchell, *Int. J. Biol. Macromol.*, 10 (1988) 259–264.
(61) S. E. Harding, in G. O. Phillips, D. Wedlock, and P. Williams (Eds.), *Gums & Stabilisers in the Food Industry*, Vol IV, IRL Press, Oxford, 1988, pp. 15–23.

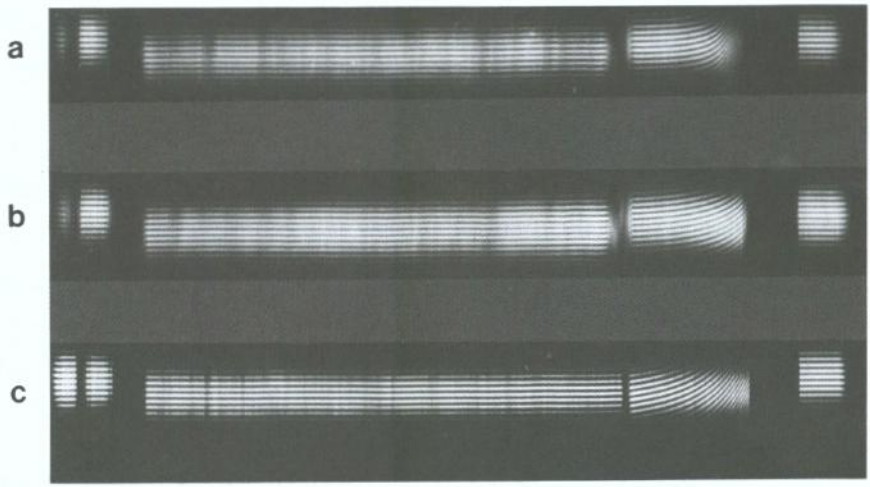


FIG. 5.—Rayleigh Interference Profiles from Low-speed Sedimentation Equilibrium. [(a) Native mucin (bronchial gp 376); (b) reduced gp 376; (c) immunoglobulin G. In (a), note the steep rise of the fringes near the cell base, but the finite slope at the meniscus.]

Section IV,3a, and *2*, facilitates a direct (and absolute) visualization of molecular-weight distributions.

With gel electrophoresis as a quantitative tool for size determination, the problem is more serious: unlike nucleic acids, mucins do not have a natural uniform charge : length ratio, and furthermore, unlike for unglycosylated proteins, sodium dodecyl sulfate does not bind uniformly. The technique appears to have some use, however, as a probe for possible mucin–protein interaction.[38]

c. Low-speed Sedimentation Equilibrium.—Arguably the most powerful technique for measuring mucin molecular weights (provided that sample molecular weights are $\lesssim 15 \times 10^6$) is the technique of low-speed sedimentation equilibrium using Rayleigh interference optics: the inherent effect of the ultracentrifugal field can be put to particular use in helping resolve the components of a heterogeneous system, provided that the effects of thermodynamic non-ideality can be properly taken into account. The low- or intermediate-[62–64] speed procedure is normally the method of choice, where the speed is sufficiently low to ensure adequate resolution of the fringes near the cell base.[62] Fig. 5 gives a comparison of Rayleigh interference patterns for

(62) J. M. Creeth and S. E. Harding, *J. Biochem. Biophys. Methods*, 7 (1982) 25–34.
(63) D. C. Teller, T. A. Horbett, E. G. Richards, and H. K. Schachman, *Ann. N.Y. Acad. Sci.*, 164 (1969) 66–101.
(64) H. K. Schachman, *Ultracentrifugation in Biochemistry*, Academic Press, New York, 1958.

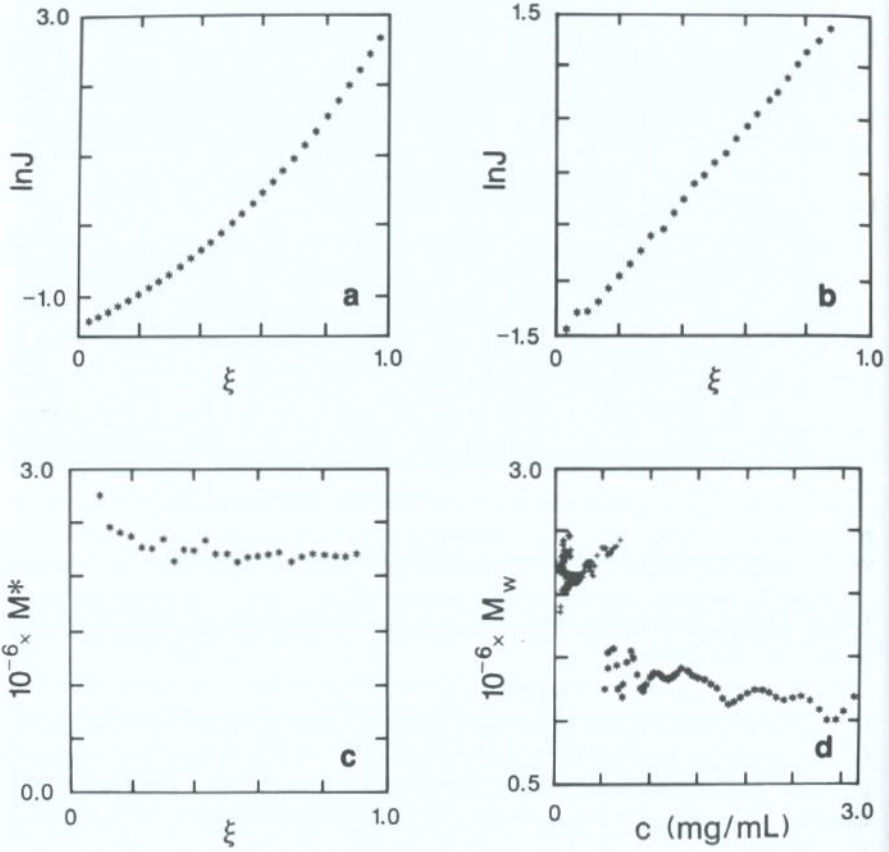


FIG. 6.—Low-speed Sedimentation Equilibrium of Mucus Glycoproteins. [(a) lnJ *vs.* ξ plot for the chronic bronchitic, bronchial mucin BM GRE, $M_r \sim 6 \times 10^6$ (Ref. 9); (b) lnJ *vs.* ξ plot for cystic fibrotic, bronchial mucin CF PHI, $M_r \sim 2 \times 10^6$ (Refs. 6, 10, 30); (c) M* *vs.* ξ plot for CF PHI; (d) Point weight average M_w *vs.* concentration (c), plots for CF PHI for two different loading concentrations: +, $c^o \sim 0.2$ mg/mL: *, $c^o \sim 2.0$ mg/mL (Ref. 9): non-overlap is diagnostic of polydispersity (presence of non-interacting components of different molecular weight).]

a native mucin, a reduced mucin, and immunoglobulin G at sedimentation equilibrium under these conditions: even before full data analysis, the steep rise at the cell base, but finite slope at the meniscus for the native mucin (see Fig. 5a), suggest the presence of considerable heterogeneity.

At equilibrium, the concentration at the air–solvent meniscus remains finite, but can be found without too much difficulty by mathematical manipulation of the fringe data.[62] A typical plot of ln J vs. ξ, where J is the absolute concentration (in fringe numbers) and $\xi = (r^2 - a^2)/(b^2 - a^2)$ [r being the radial displacement from the center of the rotor, and a and b the correspond-

ing positions of the meniscus and cell base, respectively] for a bronchial mucin (BM Gre, $M_r = 6.0 \times 10^6$) from a chronic bronchitis patient is given in Fig. 6a. The pronounced upward curvature is symptomatic of heterogeneity. In some cases, the effects of thermodynamic nonideality can obscure the effects of heterogeneity, to give a linear, pseudo-ideal plot, as shown in Fig. 6b for a bronchial mucin (CF PHI) from a cystic fibrosis patient.

Weight-average molecular weights for particular, whole-solute distributions, M_w^o can be readily obtained by using a particularly directly determinable point average[62] (the "star average," M*), *without* the requirement for an independent estimate for the initial concentration. M* is defined, for solute at a given radial position, r, by:

$$M^*(r) = \left\{ \frac{j(r)}{J_a (r^2 - a^2) + 2\int_a^r rj\, dr} \right\} \cdot \frac{2RT}{\omega^2 (1 - \bar{v}\rho)}, \qquad (1)$$

where j is the concentration in fringe numbers (relative to that at the meniscus), J_a is the meniscus fringe number, and the other parameters have their usual meaning. A plot of M* *vs.* ξ is given for CF PHI in Fig. 6c. M_w^o is obtained from the limiting value of M* at the cell base,[62] and several values for the M_w^o of mucins obtained in this way, together with other values using different techniques, are included in Table III.[13,14,30,35,40,47,56,65-69]

The partial specific volume, $\bar{v}$, (for mucins, normally in the range of 0.60 to 0.65 mL/g) has to be determined separately: $\bar{v}$ is usually obtained[70] from a plot of solution density *versus* concentration. The accuracy with which this value can be obtained (compare Section IV,3,a for refractive increment measurements) will largely be affected by errors in concentration measurement, although the final value can be checked by calculation by using the Traube rule (see, for example, Ref. 71), if the amino acid and carbohydrate composition of the mucin are known.

Other useful molecular-weight information is readily obtainable: for example, point weight average molecular weights, M_w, can be obtained by

(65) J. M. Creeth, B. Cooper, A. S. R. Donald, and J. R. Clamp, *Biochem. J.*, 211 (1983) 323-332.
(66) P. A. Feldhoff, V. P. Bhavanandan, and E. A. Davidson, *Biochemistry*, 18 (1979) 2430-2434.
(67) R. A. Gibbons and F. A. Glover, *Biochem. J.*, 73 (1959) 217-225.
(68) T. K. S. Mukkur, D. L. Watson, K. S. Saini, and A. K. Lascelles, *Biochem. J.*, 229 (1985) 419-428.
(69) J. P. Pearson, R. Kaura, W. Taylor, and A. Allen *Biochim. Biophys. Acta*, 706 (1982) 221-228.
(70) O. Kratky, H. Leopold, and H. Stabinger *Methods Enzymol., 27 (1973)* 98-110.
(71) R. A. Gibbons, in A. Gottschalk (Ed.), *Glycoproteins: Their Composition, Structure & Function*, Vol 5, Part A, Elsevier, Amsterdam, 1972, pp. 31-157.

TABLE III

Molecular Weights of Mucus Glycoproteins

Glycoprotein	Buffer	I[a]		$10^{-6} \times M_r$	n[b]	Technique	References
Ovarian cyst (human)							
603/AmSsol[c]	phosphate–chloride	0.10		0.56	1	low-speed sedimentation	37
			+*M* NaCl	0.57	1	equilibrium	
			+6 *M* Gu·HCl	0.55	1		
485[c]	sodium acetate	0.10		0.58	1		23
603/43–50[c]	phosphate–chloride	0.10		0.90	2		37
			+*M* NaCl	0.84	2		37
			+6 *M* Gu·HCl	0.85	2		37
603/43–50[c] (pronase-treated)	sodium acetate	0.10		0.47	1		
376				1.14	2		8
Bronchial (human)							
Cystic fibrosis "CFPHI"	phosphate–chloride	0.10		2.41	3–4		37
			+*M* NaCl	2.25	3–4		37
			+6 *M* Gu·HCl	2.18	3–4		37
						ultrashort-column sedimentation	
			+*M* NaCl	1.80	3–4	equilibrium	30
Chronic bronchitic "BM GRE"		0.10	+0.4 *M* $CaCl_2$	6.2	11–12	low-speed sedimentation equilibrium	9
			+5 mg/ fucose/mL	6.0	11–12		9
			+5 mg/ GlcNAc/mL	6.2	11–12		9
			+6 *M* Gu·HCl	5.5	11–12		9

Chronic bronchitic							
"BM GRE" (fractions 13–15)		0.10		5.3	10–11		45
			+6 *M* Gu·HCl	5.0	10–11		45
Chronic bronchitic "BM GRE" (fractions 13–15, thiol-reduced)				1.1	2		45
Chronic bronchitic	phosphate		+6 *M* Gu·HCl	18	40	light-scattering	13
Asthmatic (low-density fraction)	phosphate-buffered saline			5.6	10–11	high-speed sedimentation equilibrium	66
(high-density fraction)				4.6	9–10		66
Cervical (human)							
	phosphate–chloride		+6 *M* Gu·HCl	10.6	20	light-scattering	54
reduced				2.1	3–4		54
				1.6	3–4	high-speed sedimentation equilibrium	54
Trypsin-digested				0.29	1		54
				0.38	1	light-scattering	54
Bovine (estrus)	lithium citrate (citric acid)			4.16	8	intrinsic viscosity	67
Bovine (pregnancy)				3.93	8		67
Bovine	phosphate-buffered saline			16.4	35	light-scattering	14
Bovine (reduced)				5.2	10		14
				4.6	10	sedimentation–diffusion	14
Bovine (pronase-digested)				0.84	1	light-scattering	14
				0.50	1	sedimentation–diffusion	14
(pronase + reduced)				0.47	1	light-scattering	14
				0.29	1	sedimentation–diffusion	14
Gastric (pig)							
	phosphate–chloride	0.10		8.7	18	low-speed sedimentation equilibrium	45

(continued

TABLE III

Molecular Weights of Mucus Glycoproteins *(continued)*

Glycoprotein	Buffer	I[a]	$10^{-6} \times M_r$	n[b]	Technique	References
			2.1	4	sedimentation–diffusion	46,47
			2.3	4	low-speed sedimentation equilibrium	46,47
			45.0	70	light-scattering	15,56
			39		sedimentation–diffusion	56
reduced			0.5	1		46
papain-digested			0.5	1		46
reduced			2.5	4–5	light-scattering	56
Small intestine (human)						
native			1.7	3–4	c.g.c.[d]	38
papain-digested			0.24	1	c.g.c.[d]	38
native sheep			5.0	10–11	high-speed sedimentation equilibrium (absorption optics)	68
Gall bladder (human)						
papain-digested			0.5	1	c.g.c.[d]	69

[a] I = ionic strength. [b] n = approximate number of basic units, based on molecular weight of basic unit of 0.5×10^6. [c] Protease inhibitors not used in extraction procedure. [d] c.g.c. = calibrated gel chromatography.

using sliding strip fits to the log fringe concentration *versus* radial displacement squared plots,[9] and an example for two different loading concentrations, c^o, of CF PHI is given in Fig. 6d.

A potential pitfall if sedimentation equilibrium is used is a failure to allow properly for the effects of thermodynamic non-ideality, which tends to diminish measured molecular weights and to mask heterogeneity. The effects on M_w^o are normally minimized by using the lowest possible loading concentration (as low as 0.2 mg/mL if 30-mm path-length cells are employed). Alternatively, point weight averages can be extrapolated to zero concentration ($J = 0$), to yield a value independent of non-ideal or associative phenomena. The ideal value obtained in such a way may, however, be biased towards the lower end of the molecular weight distribution, but this bias can be minimized by using short solution columns and low speeds, and, in extreme cases, by extrapolating the value so obtained to zero gravitational field. Another way of coping with non-ideality is to combine weight average values with number, z-, and higher-order averages[72,73] if the precision in the data justifies this: it is possible to obtain number-average and Z-average whole cell average molecular weights (although these are less precise) and also point (M_n and M_z) and compound molecular weights (M_{y1}, M_{y2}):

$$M_{y1} = \left\{ \frac{2}{M_n} - \frac{1}{M_w} \right\}^{-1} \qquad (2)$$

$$M_{y2} = \frac{M_w^2}{M_Z}$$

M_{y2} and M_{y2} point averages are free from first-order, non-ideality effects: measured values are, however, generally reliable only if a laser-light source can be employed to generate the interference fringes, or if accurate on- or off-line data-capture procedures are available.[74–76]

High-speed or meniscus-depletion equilibrium techniques have been attempted by employing scanning absorption optics,[53–55] and by using Rayleigh interference.[66,77] Several values have been included for comparative purposes in Table III. In this method, the meniscus region is essentially depleted of sample (namely, zero concentration), and this simplifies the

(72) D. Roark and D. A. Yphantis *Ann. N.Y. Acad. Sci.*, 164 (1969) 245–278.
(73) D. C. Teller, *Methods Enzymol.*, 27 (1973) 346–441.
(74) S. E. Harding and A. J. Rowe, *Optics & Lasers in Engineering*, 8 (1988) 83–96.
(75) S. E. Harding and A. J. Rowe, in J. Tyrer and G. T. Reid (Eds.), *Automatic Fringe Analysis*, Open Tech. Press, Loughborough, U.K., 1987, pp. 187–200.
(76) S. E. Harding and A. J. Rowe, *Biochem. Soc. Trans.*, 15 (1987) 1046–1047.
(77) H. Woodward, B. Horsey, V. P. Bhavanadan, and E. A. Davidson, *Biochemistry*, 21 (1982) 694–701.

interpretation of the data if Rayleigh interference optics is used, because there is no need to measure meniscus concentrations. However, no real advantages are accrued by attempting meniscus depletion if absorption optics are used, because this treatment gives a direct record of meniscus absorbance. In addition, for such polydisperse materials as mucins, it is normally not possible to deplete the meniscus properly without losing optical registration at the cell base.[37,62] A common pitfall is to assume depletion conditions when this is clearly not valid.[78]

A further difficulty in attempting meniscus-depletion conditions with polydisperse materials is that the effective, thermodynamic, second virial coefficient, B_{eff}, can be greatly enhanced[8,79,80]:

$$B_{eff} = B\,(1 + \lambda^2 M_z^2/12 + \cdots), \qquad (3)$$

B being the "static" virial coefficient,[8] $\lambda = (1 - \bar{v}\rho)\omega^2\,(b^2 - a^2)/2RT$, ρ the solvent density, and ω the angular velocity. This "speed dependence" effect can be minimized by low speeds and short solution columns.[8,80] A problem peculiar to absorption optics (applicable to protein systems in general) is the difficulty in obtaining an accurate baseline due to, for example, anomalous absorption onto cell windows[81–83]: this difficulty is often ignored in the literature.

Rayleigh interference optics is by far the most suitable optical system. Furthermore, if a laser-light source is used, as opposed to the mercury-arc sources commercially supplied, fringe resolution can be greatly enhanced.[84–88] This, together with the possibility of using a commercial laser gel-scanner,[74] the Ultroscan (LKB Instruments, Bromma), for data capture, and a simple Fourier algorithm[74–76] for data analysis (enabling fringe displacements to be measured to an accuracy of 1/300th of a fringe, even with fringes generated by a mercury-arc light-source) opens up new possibilities

(78) J. Hinnie and A. Serafini-Fracassini, *Biopolymers*, 25 (1986) 1095–1107.
(79) H. Fujita, *Foundations of Ultracentrifuge Analysis*, Wiley, New York, 1975, Chap. 5.
(80) H. Suzuki, in B. Ranby (Ed.), *Physical Chemistry of Colloids and Macromolecules*, Blackwell Scientific, Oxford, 1987, pp. 101–109.
(81) S. P. Spragg, *The Physical Behaviour of Macromolecules with Biological Functions*, Wiley, New York, 1980, pp. 66–68.
(82) A. J. Rowe, *Protein Enzyme Biochem.*, 106 (1984) 1–37.
(83) P. H. Lloyd, *Optical Methods in Ultracentrifugation, Electrophoresis and Diffusion*, Oxford University Press, 1974, pp. 86–87.
(84) R. C. Williams, *Methods Enzymol.*, 48 (1978) 185–191.
(85) T. M. Laue and D. A. Yphantis, *Biophys. J.*, 25 (1979) 164.
(86) D. A. Yphantis, *Methods Enzymol.*, 61 (1979) 3–12.
(87) R. C. Williams, *Anal. Biochem.*, 48 (1972) 164–171.
(88) T. M. Laue, D. A. Yphantis, and D. G. Rhodes, *Anal. Biochem.*, 143 (1984) 103–112.

for detailed analysis of the nature of the heterogeneity and thermodynamic non-ideality.[10,89-93]

4. Linear Models for Mucin Structure

A model alternative to the branched assembly proposed earlier is one in which the basic units of the mucin are assembled in linear arrays. The first direct evidence for this structure came from electron microscopy.

Initial progress was made by Slayter and coworkers,[94-97] using the technique of rotational platinum-shadowing after air-drying bronchial and other mucin solutions onto mica. This procedure revealed a variety of conformations: in the main, linear structures were seen, with no evidence of branched forms. However, these linear structures generally lay in parallel arrays which could well have resulted from the high shearing and surface-tension forces on air-drying which may have uncoiled them from their true solution-conformation. On the other hand, this procedure did have the advantage of revealing the probable nature of the assembly (namely, linear). In a subsequent study on bronchial mucins, in which extraction procedures were used to minimize the effect of proteases, flexible, linear forms were again made visible, although, if negative staining with uranyl acetate was employed, the material appeared as aggregates.[98] Poorly defined images using negative staining were also reported for ovarian-cyst and cystic-fibrosis bronchial mucins.[99] Other workers also demonstrated linear forms on using both rotary shadowing[100,101] and staining procedures.[101,102]

(89) H. Kim, R. C. Deonier, and J. W. Williams, *Chem. Rev.*, 77 (1977) 659–690.
(90) P. W. Chun, S. J. Kim, J. D. Williams, W. T. Cope, L. H. Tang, and E. T. Adams, Jr., *Biopolymers*, 11 (1972) 197–214.
(91) M. L. Johnson, J. J. Correia, D. A. Yphantis, and H. R. Halvorson, *Biophys. J.*, 36 (1981) 575–588.
(92) S. H. Tindall and K. C. Aune, *Anal. Biochem.*, 120 (1982) 71–84.
(93) B. K. Milthorpe, L. W. Nichol, and P. D. Jeffrey, *Biophys. Chem.*, 3 (1975) 169.
(94) G. Lamblin, M. Lhermitte, P. P. Roussel, and H. S. Slayter, *Biochimie*, 61 (1979) 23–43.
(95) H. S. Slayter, A. G. Cooper, and M. C. Brown, *Biochemistry*, 13 (1974) 3365–3371.
(96) H. S. Slayter, in J. R. Harris (Ed.), *Electron Microscopy of Proteins*, Vol. 1, Academic Press, New York, 1981, pp. 197–254.
(97) P. Roussel, G. Lamblin, N. Houdret, M. Lhermitte, and H. S. Slayter, *Biochem. Soc. Trans.*, 12 (1984) 617–618.
(98) H. S. Slayter, G. Lamblin, A. LeTreut, C. Galabert, N. Houdret, P. Degand, and P. Roussel, *Eur. J. Biochem.*, 142 (1984) 209–218.
(99) S. E. Harding, A. J. Rowe, and J. M. Creeth, *Biochem. J.*, 209 (1983) 893–896.
(100) A. Mikkelsen, B. T. Stokke, B. E. Christensen, and A. Elgsaeter, *Biopolymers*, 24 (1985) 1683–1704.
(101) J. K. Sheehan, K. Oates, and I. Carlstedt, *Biochem. J.*, 239 (1986) 147–153.
(102) A. O. Jenssen, O. Harbitz, and O. Smidsrød, *Eur. J. Resp. Dis.*, 61 (1980) 71–76.

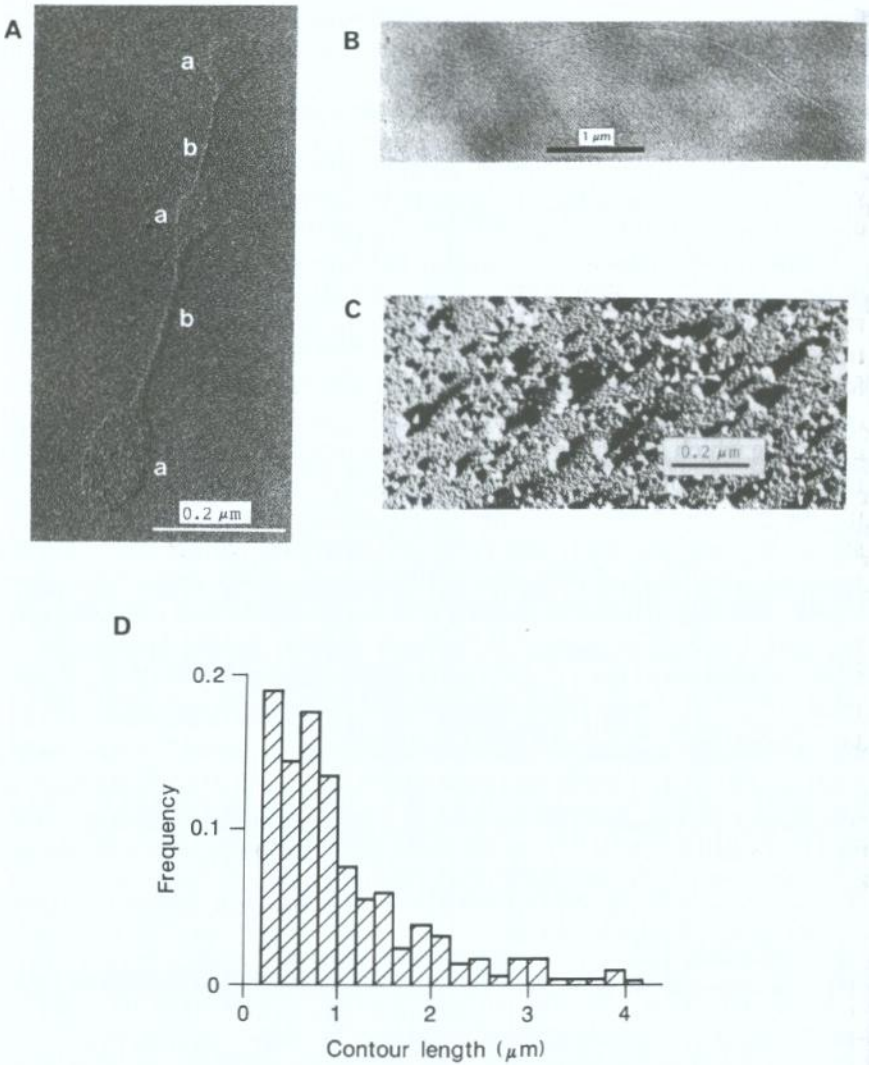


FIG. 7.—Electron Microscopy of Mucins. [(A) Unidirectionally platinum-shadowed CF PHI macromolecule, after air-drying onto mica (from Ref. 30). a, "low profile areas; b, linear extensions. (B) Unidirectionally shadowed pig gastric mucin (PGM) macromolecule after spreading on 6 *M* guanidine hydrochloride and air-drying.[101] (C) CF PHI, unidirectionally platinum-shadowed after fixation, and critical-point drying.[106] (D) Distribution of contour lengths for PGM.[101]]

Applications of similar, metal-replica preparation procedures to a bronchial mucin from a cystic fibrosis patient[99] also revealed images of linear assemblies (see Fig. 7A). However, instead of using rotary shadowing techniques, unidirectional shadowing was used. Low-profile areas, with linear extensions, were seen; the low-profile areas could clearly be derived from highly swollen particles of low asymmetry which flattened onto the mica upon drying; such structures would not be as clearly visible on using rotary shadowing,[98] although similar, "puddled" structures have been observed.[94]

The linear extensions linking these "low-profile" areas do not appear to be flattened, possibly because they contain less carbohydrate, and hence are less swollen through solvation prior to drying. It has been tentatively suggested that the "naked" regions linking the basic units/T-domains could be located within these "low glycosylation" extensions.[24] The form shown in Fig. 7A would therefore correspond to a mucin molecule consisting of three basic-unit forms. It should be stressed, however, that single, nonglycosylated, polypeptide chains having no apparent secondary structure would not be visible with the resolution used here. On the other hand, there could be a small amount of residual glycosylation in this region, insufficient, at least in patches, to protect from protease attack, but sufficient to allow visibilization by unidirectional shadowing. This suggestion is, however, highly speculative, and may be disproved or confirmed by using stains specific for carbohydrate-free regions.[103]

Much of the earlier work on mucins using electron microscopy[94-97] was on relatively small macromolecules ($M_r \sim 2 \times 10^6$). A comprehensive study by Sheehan and coworkers[101] on larger bronchial, cervical, and gastric mucins, using a variety of preparative procedures, appears to leave little doubt that the general mode of assembly for mucins from a variety of sources must be linear. Nodular forms in the flexible linear chains were often, but not always, made visible. Indeed, Fig. 7B[101] shows an image of a p.g.m. molecule using a preparation procedure similar to that for Fig. 7A, but after spreading on 6 *M* Gu · HCl; no evidence for any nodular or low-profile areas is apparent. These workers,[101] in agreement with others,[98,99] also observed considerable polydispersity in all their samples; for example, Fig. 7D[101] gives the distribution of contour lengths measured for p.g.m., which is of the same log-normal form as the predicted molecular-weight distribution reported by Creeth and Cooper[45] for this mucin from low-speed sedimentation equilibrium (see Section VI,1 and Figs. 10 and 12).

The original postulate of a linear model for basic unit/T-domain and subunit assembly in solution[53-56] was based on light-scattering studies on the

(103) J. K. Sheehan and I. Carlstedt, *Proc Int. Symp. Glycoconjugates, 9th,* 1989 (in press).

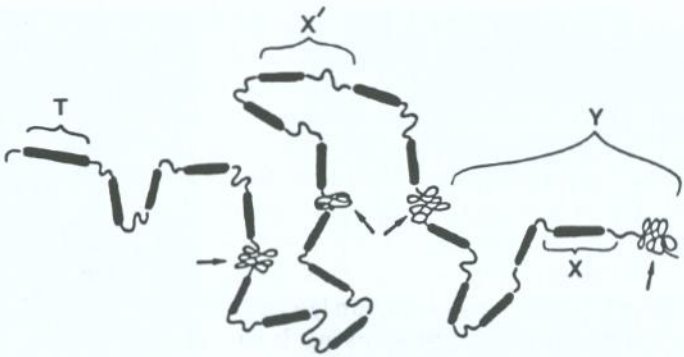


FIG. 8.—Linear Model for Cervical Mucin, Consisting of Basic Units or "T-Domains" Linked by Naked Peptide Regions, Some by Disulfide Bridging. [Arrowed regions are susceptible to thiol attack. From Ref. 53. Key: T, T-domain; X or X′, "basic unit"; and Y, "subunit."]

high-molecular-weight cervical mucin (see Fig. 8); for example, the Burchard ρ parameter[104] was used, and this had been proposed as a measure of the extent of branching in a polymer; it is obtained as the ratio of R_g to the Stokes radius r_H (from q.l.s. diffusion-coefficient measurements): the values obtained[54] for whole mucins and subunits (1.8 and 1.6, respectively) appear consistent only with a linear assembly. A similar, light-scattering study performed on submaxillary mucins (which have side-chain compositions and structures different from those of other mucins) gave very similar results,[17] suggesting that there is a close homology of these with other mucins.

V. The Gross Conformation of Mucus Glycoproteins in Solution

What domains do these mucin assemblies described herein occupy in solution? How do they interact with each other and with other molecules to form the mucosal barrier? In answer to the first question, there is growing evidence to suggest that they occupy enormously expanded, and thus, at low concentration ($\lesssim 1\%$), readily overlapping, spheroidal domains in solution.

Although the electron-microscopic technique of air-drying onto mica prior to platinum shadowing has yielded valuable information about the mode of assembly or secondary structure, it can yield misleading information about the gross conformation in solution, because of, for example, the very high shearing and surface-tension forces encountered by these highly solvated structures on air-drying onto mica. For example, it could be inferred from some earlier studies[94–96] that, in the native form, bronchial

(104) W. Burchard, M. Schmidt, and W. H. Stockmayer, *Macromolecules*, 13 (1980) 1265–1272.

mucins behave as semi-flexible rods. However, parallel arrays of such rods are often seen, strongly suggestive of alignment through shearing.

An attempt has been made to circumvent this problem by using glycerol prior to drying,[105] giving images of many complex conformations, both globular and linear. However, the effect on the native conformation of replacing the large proportion of solvent normally associated with mucins with concentrated glycerol remains open to question. Hallett and co-workers[106] used, instead, a critical-point drying procedure with prior fixation using glutaraldehyde. Images of a cystic-fibrosis, bronchial glycoprotein using such a procedure (see Fig. 7C[106]) yielded large, roughly spherical particles of dimensions very reasonably consistent with their prediction. Using a value of $\sim 2 \times 10^6$ for the molecular weight, a value for the swelling due to solvation, $v_s/\bar{v}$ [where v_s is the solvated specific volume and $\bar{v}$ is the partial (anhydrous) specific volume] of ~ 100 times was obtained.[106] Only with less-complete fixation was any evidence of elongated structures apparent. Again, however, questions could be raised as to how "native" the conformation remains after fixation.

The strongest evidence for a solvated, loosely coiled domain comes from hydrodynamics and related techniques: *1.* Estimation of the thermodynamic, second virial coefficient by using the technique of ultrashort-column sedimentation equilibrium under conditions where charge effects have been largely suppressed gives a value that is consistent with a molecular expansion of ~ 100 in solution.[8,30] *2.* Values for both the intrinsic viscosity, $[\eta]$, and the concentration-dependence regression-coefficient of sedimentation, k_s (corrected for radial dilution effects and solution density; see, for example, Ref. 107) are very high.[8,30,38,71,106] High $[\eta]$ or high k_s can arise from asymmetry or expansion, but the ratio $k_s/[\eta]$ has been shown by several workers to be a function of shape alone.[107–111] The calculated value of $\sim 1.4-1.6$ for ovarian-cyst[7] and cervical mucins[54] appears to be consistent with a coiled, spherical domain. Sheehan and Carlstedt[54] obtained similar values for subunits and T-domains. *3.* Double log plots of R_g, D_z, s, and $[\eta]$ *vs.* molecular weight, M_r, for whole mucins, subunits, and T-domains gave slopes consistent with a coiled structure[53,54] (see Fig. 9).

(105) M. C. Rose, W. A. Voter, C. F. Brown, and B. Kaufman, *Biochem. J.*, 222 (1984) 371–377.
(106) P. Hallett, A. J. Rowe, and S. E. Harding, *Biochem. Soc. Trans.*, 12 (1984) 878–879.
(107) S. E. Harding and P. Johnson, *Biochem. J.*, 231 (1985) 543–548.
(108) M. Wales and K. E. van Holde, *J. Polym. Sci.*, 14 (1954) 81–86.
(109) P. Y. Cheng and H. Schachman, *J. Polym. Sci.*, 16 (1955) 19–30.
(110) J. M. Creeth and C. G. Knight, *Biochim. Biophys. Acta*, 102 (1965) 549–558.
(111) A. J. Rowe, *Biopolymers*, 16 (1977) 2595–2611.

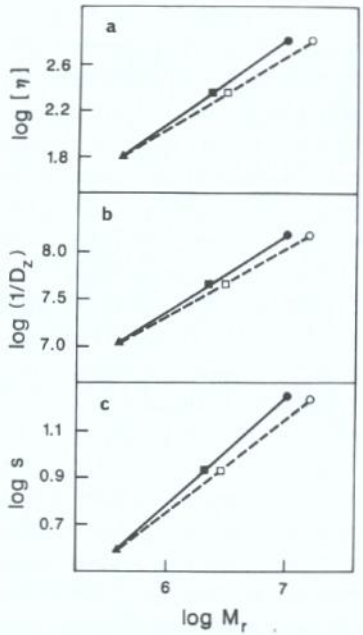


FIG. 9.—"Double Log Plots" of (a) Intrinsic Viscosity, (b) the Reciprocal of the Diffusion Coefficient, and (c) Sedimentation Coefficient Data *versus* Molecular Weight for Human Cervical Mucins.[54] [Key: ● and ○, "whole" mucins; ■ and □, "subunits"; and ▲, T-domains. Molecular weights determined from Zimm plots (filled symbols) or the Svedberg equation using QLS (open symbols). Values for the slopes are in all cases consistent with a random-coil model and not with a rigid sphere or a rod.]

Besides being enormously expanded and spheroidal, the mucins also appear very flexible: evidence for this derives from the ease with which these molecules are deformed with shear, as shown in streaming-birefringence experiments,[30,112] from the concentration dependence of viscosity,[7] and the temperature-dependence of sedimentation coefficients.[7] Mikkelsen and coworkers[100] confirmed these observations by using the dependence on field strength of electric birefringence relaxation-phenomena. Evidence for local flexibility in the carbohydrate side-chains has come from, for example, n.m.r.-spectral studies.[33,34,113,114]

Finally, the existence of mucins in solution as highly expanded, spheroidal domains is also consistent with their biological function as space-filling

(112) R. A. Gibbons and F. A. Glover, *Biochem. J.*, 73 (1959) 217–225.
(113) M. Barrett-Bee, G. Bedford, and P. Loftus, *Biosci. Rep.*, 2 (1987) 257–263.
(114) M. Barrett-Bee, G. Bedford, and P. Loftus, in Ref. 12, pp. 107–111.

water immobilizers[1-3]: at relatively low concentrations (below ~1%), the domains normally overlap,[115] corresponding to the onset of gelation.[11,115]

Mucins as Polyelectrolytes

One important feature of mucins is their behavior as polyelectrolytes; for example, this has been postulated as a possible control-mechanism in the periodic changes in cervical mucus.[116] This behavior is a result of the relatively large proportion of acidic sugar residues (NeuAc) normally present.

In viscosity studies (at a concentration of ~2.0 mg/mL), these effects can normally be suppressed at ionic strengths ~1.0 *M*, as demonstrated for bronchial and ovarian-cyst glycoproteins by Harding and Creeth[37] using either NaCl or Gu·HCl. No deleterious effects on the glycoprotein by these compounds (up to I = 6.0 *M* for Gu·HCl) were observed, consistent with the observations of Snary and coworkers.[47,117] The effects of NaCl and Gu·HCl were indistinguishable, with only a modest conformational change in the range I = 0.1–1.0 *M*, a typical polyelectrolyte effect.[118] The effects (including those of 6 *M* Gu·HCl) were found to be fully reversible.[37]

These observations imply a degree of flexibility in the macrostructure, a conclusion in agreement with other physical properties indicating a loose, highly solvated, greatly expanded structure. It was also observed[37] that a relatively low ionic strength (I = 0.1 *M*) is sufficient to suppress the primary salt effect, so that correct molecular-weight information can be obtained at I = 0.1 *M* if low concentrations are employed in sedimentation-equilibrium measurements. However, those physical measurements that are especially sensitive to non-ideality should be performed at ionic strengths not less than 0.5 *M*, unless extrapolations to infinite dilution can be employed.[37]

Tam and Verdugo[116] observed that changes in pH over the range of 6.5–8.0 sharply affect the swelling of cervical mucus, and suggested that changes in the polyelectrolyte nature of the gelatinous, mucin framework (for a Donnan equilibrium) are responsible. It was later pointed out, however,[37] that changes in the mucin moiety are not likely to be directly responsible, as only the small number of histidine residues (and not the NeuAc) could possibly be titrated in this range. The protein content of the secretion probably cannot be ignored in this instance.

(115) J. M. Creeth, *Mod. Prob. Paed.*, 19 (1988) 34–45.
(116) P. Y. Tam and P. Verdugo, *Nature*, 292 (1981) 340–342.
(117) D. Snary, A. Allen, and R. H. Pain, *Eur. J. Biochem.*, 24 (1971) 183–189.
(118) L. H. Kent, B. R. Record, and R. G. Wallace, *Phil. Trans. R. Soc., Ser. A.*, 250 (1957) 1–43.

VI. Mucin Heterogeneity

Mucins are highly heterogeneous substances (see Table IV), on at least two counts. *1.* They are polydisperse in a quasi-continuous sense (arising from variability in the carbohydrate composition.[11,12,115,119]). *2.* They are polydisperse in a discrete sense, arising from variability in the numbers of basic units.[10,12]

Furthermore, they are known to interact in solution with other components of mucus secretions, such as lysozyme,[120,121] and, in addition, they were originally considered to self-associate in solution,[6,122] although subsequent evidence suggested otherwise.[9,45] A further complication is that they are extremely non-ideal in the thermodynamic sense, arising from the very high solvent-association. Table V[7,8,30,123–125] compares the nonideality parameter, BM_r, where B is the thermodynamic second virial coefficient, for mucins with that for other macromolecules. Finally, besides being heterogeneous with respect to molecular weight,[126] they are heterogeneous with respect to density.[127]

One of the most powerful techniques for the investigation of macromolecular heterogeneity is the analytical ultracentrifuge, in particular, sedimenta-

TABLE IV

Terminology for Mucus Glycoprotein Heterogeneity

Term	Definition
Heterogeneity	Any macromolecular system where the solute species do not have a single value for the molecular weight, no matter what the origin of the variation may be. Can include polydispersity or self-association behavior.
Polydispersity	The presence of non-interacting components in a macromolecular system of different molecular weight or density, or both.
Primary polydispersity	Quasi-continuous distribution of mucin molecular weights arising from variability in carbohydrate side-chain composition.
Secondary polydispersity	Discrete distribution of mucin molecular weights arising from variation in numbers of basic units.

(119) P. Roussel, P. Degand, G. Lamblin, A. Laine, and J. J. Lafitte, *Lung,* 154 (1978) 241–260.
(120) J. M. Creeth, J. L. Bridge, and J. R. Horton, *Biochem. J.,* 181 (1979) 717–724.
(121) A. O. Jenssen, Ph.D. Thesis, Univ. Trondheim, Norway, 1980.
(122) J. M. Creeth, *Biochem. Soc. Trans.,* 8 (1980) 520–521.
(123) C. F. Tanford, Ref. 50, p. 196.
(124) S. E. Harding, *Int. J. Biol. Macromol.,* 3 (1981) 341–342.
(125) C. H. Emes and A. J. Rowe, *Biochim. Biophys. Acta,* 537 (1978) 110–124.
(126) R. A. Gibbons, *Br. Med. Bull.,* 34 (1978) 34–38.
(127) J. M. Creeth and J. R. Horton, *Biochem. J.,* 161 (1977) 449–463.

TABLE V

Comparative Thermodynamic Non-ideality of Macromolecular Solutions

Particle	BM_r (mL·g^{-1})	References
Spherical unhydrated protein[a]	2.9	123
Hemoglobin	4.8	124
Myosin	52.6	125
GP 376	300	7,8
CF PHI	300	30

[a] Based on a partial specific volume of 0.73 mL·g^{-1}.

tion equilibrium, especially when it is used in conjunction with such other techniques as electron microscopy. Heterogeneity can be characterized by combining whole-cell average molecular weights (M_n^0, M_w^0, and M_z^0) in the form of the Herdan relations[128,129]:

$$\frac{\sigma_n}{M_n^0} = \left[\left(\frac{M_w^0}{M_n^0}\right) - 1\right]^{\frac{1}{2}}; \frac{\sigma_w}{M_w^0} = \left[\left(\frac{M_z^0}{M_w^0}\right) - 1\right]^{\frac{1}{2}}, \qquad (4)$$

where σ_n and σ_w are respectively the number and weight-average standard deviations of the distribution, whatever form this may take, or by using ideal, molecular-weight moments see Eq. *2* and Ref. 72) and other diagnostic combinations of point average M_n, M_w, and M_n values.[73]

There have been many theoretical attempts to characterize heterogeneity by using quasi-elastic light-scattering (see, for example, Ref. 130), and size-distribution, software packages are available from commercial manufacturers.[131] The so-called 'polydispersity factor,' PF, namely, the Z-averaged, normalized variance of the distribution of diffusion coefficients,[132,133] has proved a useful guide to the homogeneity of a sample. However, because q.l.s. is so sensitive to declarification from dust and aggregates (see, for example, Refs. 57 and 59), a realistic application of the polydispersity factor and the more-sophisticated, size-distribution analyses[130,131] in order to obtain quantitative information about mucin heterogeneity appears to lie in the future. Furthermore, the analyses developed do not distinguish between self-association and genuine polydispersity, and do not, in general, take into

(128) J. M. Creeth and R. H. Pain, *Prog. Biophys. Mol. Biol.*, 17 (1967) 217–287.
(129) G. Herdan, *Nature*, 163 (1949) 139.
(130) S. W. Provencher, *Makromol. Chem.*, 180 (1979) 201–209.
(131) See, for example, Malvern Instruments Bulletin PB041 2-85 (1985).
(132) D. E. Koppel, *J. Chem. Phys.*, 11 (1972) 4814–4820.
(133) P. Pusey, in H. Z. Cummins and E. R. Pike (Eds.), *Photon Correlation and Light Beating Spectroscopy*, Plenum, New York, 1974, p. 387–428.

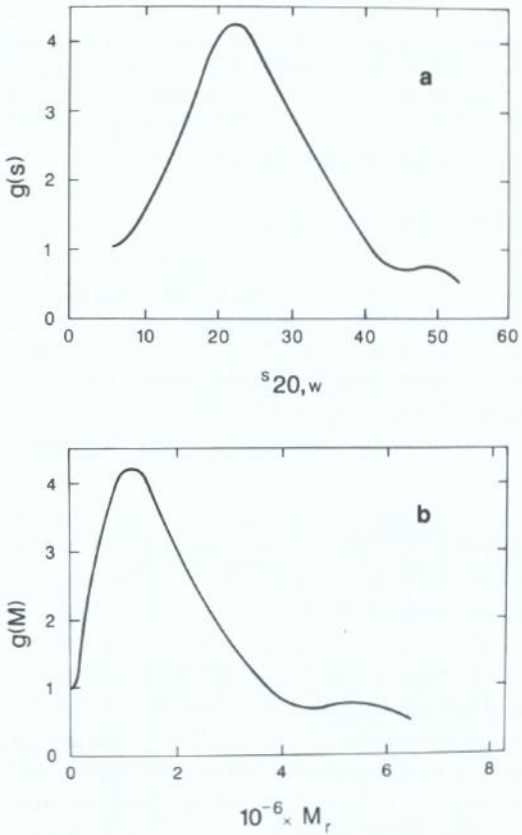


FIG. 10.—(a) Distribution of Sedimentation Coefficients for Pig Gastric Mucin (from Ref. 49). (b) Corresponding, Estimated Distribution of Molecular Weights, Using the Relation $s \propto M^{0.5}$.

consideration thermodynamic non-ideality; the approximation is also made that the diffusing species have similar conformations. Nevertheless, as a relative technique for demonstrating changes in a mucin sample (for example, detecting an interaction), q.l.s. would appear to be a useful tool.[134]

Another useful way of visibilizing heterogeneity is by using sedimentation velocity and using the approximation that all sedimenting species have similar conformations: for example, Pain and coworkers[46,49] used a procedure (and only after several additional approximations)[135,136] for calculating sedi-

(134) M. Smedley, PhD Thesis, Brighton Polytechnic, 1987.
(135) V. C. Hascall and S. W. Sadjera, *J. Biol. Chem.*, 244 (1969) 2384–2396.
(136) H. Fujita, *Foundations of Ultracentrifuge Analysis*, Wiley, New York, 1975, pp. 166–206.

mentation coefficient(s) distributions to demonstrate heterogeneity of pig gastric mucin (see Fig. 10a). To demonstrate what this could mean in terms of a size distribution, the assumption (valid for flexible polymers) is here used that $M_r \propto s^2$, and that an s value of 33×10^{-13} s is approximately equivalent to a molecular weight[46] of 2.5 million, to generate the equivalent molecular-weight distribution of Fig. 10b. The form of the latter distribution is similar to that of the distribution of contour lengths shown in Fig. 7D.

1. Polydispersity and Self-Association Behavior

It is normally impossible to distinguish between the effects of self-association and polydispersity (that is, non-interacting components of different M_w^0 or density, or both) from a single sedimentation-equilibrium experiment (see, for example, Ref. 137). However, it is known, both from the apparent nature of the biosynthetic process[12] and from combining the results of more than one sedimentation-equilibrium experiment using the diagnostic technique of non-overlap of plots of point weight average molecular weight *versus* concentration plots (see Fig. 6D) for different loading concentrations[9,72] that polydispersity in mucins is significant. Although it was originally considered that mucins might also be self-associating,[6,122] there is now increasing evidence that this is not significant, at least in dilute solution: this

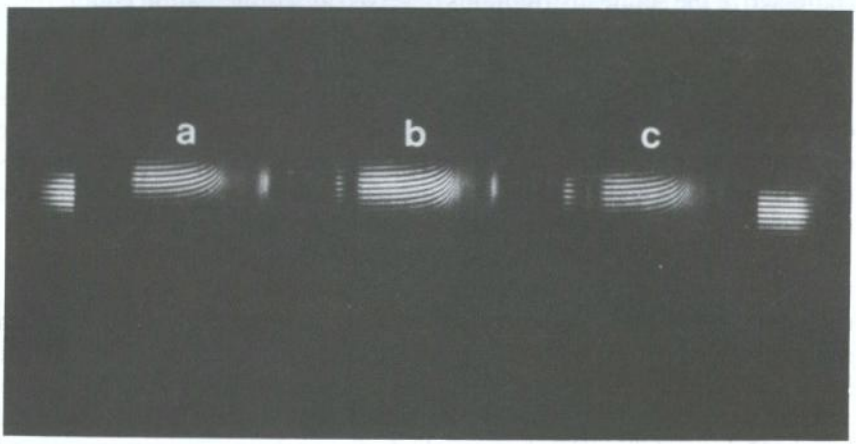


FIG. 11.—Rayleigh Equilibrium Interference Patterns[9] for BM GRE in (a) Phosphate–Chloride Buffer Containing 0.4 *M* CsCl; (b) Phosphate–Chloride Buffer + 5 mg of Fucose/mL; and (c) Phosphate–Chloride Buffer + 5 mg of GlcNAc /mL.

(137) J. M. Creeth and S. E. Harding, *Biochem. J.*, 205 (1982) 639–641.

conclusion was reached from the results of sedimentation-equilibrium experiments wherein potential sites of association were competitively blocked with inhibitors.[9,45]

There are two possible sites for self-association: on the glycosylated regions, there may be hydrophobic patches arising from possible localized clusters of L-fucose or GlcNAc residues. Such sites would be competitively blocked by having a swamping concentration of L-fucose or GlcNAc present in the solvent. However, in such an experiment using a multi-channel cell, there is little difference between the Rayleigh fringe distributions (see Fig. 11), or the calculated values for M_w^0 (all $\sim 6 \times 10^6$; see Table III), for a glycoprotein, from the bronchial secretion of a chronic bronchitis patient, in the presence of L-fucose or GlcNAc, compared with non-dissociating solvent. A similar experiment performed in 6 *M* Gu·HCl to block any hydrophobic sites on the naked peptide regions again yielded an M_w^0 of 6×10^6: there is therefore little evidence of self-association in this mucin. These observations were supported by similar experiments on a related bronchial mucin and pig gastric mucin,[45] and also lower-molecular-weight mucins.[37]

The observed heterogeneity of mucins in dilute solution would appear, therefore, to be due mainly to polydispersity, and not to self-association phenomena: it could be inferred from this conclusion that the gel-forming tendency of mucins is mainly achieved through molecular overlap of their huge, swollen domains, and not by virtue of association phenomena. To characterize the solute distributions observed at sedimentation equilibrium in terms of actual molecular-weight distributions is, however, not easy, largely because of the difficulties introduced by the presence of the high thermodynamic non-ideality. An indirect procedure for modelling the polydispersity of p.g.m. was used by Creeth and Cooper[45] by assuming the solute distribution to be the same as if it were an isodesmic, self-association process (see Fig. 12). This makes use of the principle that a reacting system cannot be distinguished, in a single experiment, from a non-reacting system that contains the same distribution of molecular weights. The form of the distribution agrees well with the distribution of contour lengths from electron microscopy[101] (see Fig. 7D), and also the form of Fig. 10b calculated from sedimentation-velocity data, although the actual values from Fig. 10b are lower (see Section IV,2).

A direct procedure for handling non-ideal, polydisperse distributions has now been derived,[10] and has been applied, after making certain assumptions, with some success to discrete distributions of molecular weights for bronchial mucins. Assuming that the observed polydispersity of a cystic fibrosis glycoprotein (CF PHI) was due entirely to variation in the numbers of mucin basic units, it was possible to obtain representations of the observed solute distributions by using a three-component, non-ideal fit (see Fig. 13), based

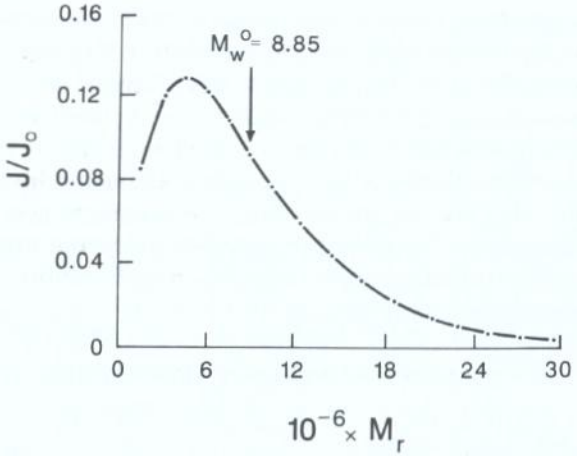


FIG. 12.—Molecular Weight Distribution for Pig Gastic Mucin,[45] Calculated by Using the Constants for a Hypothetical Isodesmic Association. [The value marked by an arrow corresponds to the weight-average molecular weight for the whole solute distribution, M_w^o.]

on the earlier observations from electron microscopy.[30] Because of the relatively large amount of computer time required to achieve such fits (even on a large, mainframe computer), it is not practical at the present time to perform (arguably, more representative) log-normal distribution fits, because of scaling problems, at present insurmountable, caused by the very large range in molecular weights. The algorithms are now available, however, and it is hoped that it would be possible to fit to such distributions as soon as the hardware is improved.

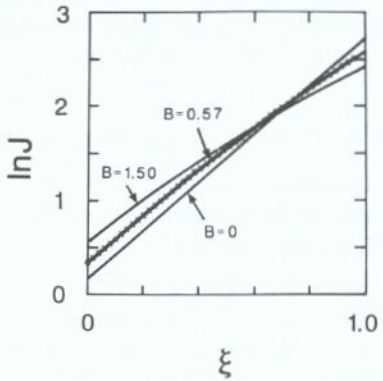


FIG. 13.—Plot of lnJ *vs.* ξ for Bronchial Mucin CF PHI.[10] [Lines fitted correspond to a three-component fit (of molecular weights, 1.2×10^6, 1.8×10^6, and 2.4×10^6) for various values of the thermodynamic, second virial coefficient, $B \times 10^{-4}$ mL. mol. g^{-2}].

Point average molecular-weight distributions also provide a very valuable insight into mucin polydispersity, although this, like the direct method, normally requires data of the highest quality (for example, obtained by using a laser light-source, and accurate data-capture and fringe-analysis procedures; compare Section IV,3). Again, as noted in Section IV,3 mention should be made of two methods, not yet applied to mucins, involving g.p.c. linked with an "absolute" molecular-weight technique to give molecular-weight distributions: *1.* "off-line" calibration of g.p.c. columns using low-speed sedimentation equilibrium[60,61] and *2.* "on-line" calibration of g.p.c. using multi-angle laser light-scattering.[137a]

2. Interactions with Other Macromolecules

Although the mucins (which constitute at least 30% of the non-water part of mucus) largely dictate the physical properties of whole mucus, their interactions with other components play an important role. For example, such interactions are considered to play an important cause of changes occurring in cervical mucus during the estral cycle.[37,116] Although it has been suggested that immunoglobulin IgA may be involved in the stabilization of mucus gels,[138,139] perhaps the clearest evidence for mucin–protein interaction is for bronchial mucin with lysozyme. This has been demonstrated by Creeth and coworkers[120] by using sedimentation-equilibrium techniques. Rheological techniques have also proved to constitute a powerful probe for such interactions: the viscometric data of Jenssen and coworkers[121,140] supported the observations of Creeth and associates,[120] and also suggested the interaction to be electrostatic in nature. A rheological investigation by Smedley[134] demonstrated a calcium ion-dependent interaction of alginate secreted from *Pseudomonas aeruginosa* with bronchial mucins. The use of various polymer systems, including alginates with intestinal mucins, for increasing the transit time of drug-carrier systems[141] is currently being explored.

VII. Summary and Prospects

Mucins from a wide range of sources show the same basic structural properties: a heavily glycosylated basic unit of $M_r \sim 0.5 \times 10^6$, linked by covalent (or other sufficiently long-lived) bonds into linear arrays that are polydisperse with respect to numbers of basic units (discrete polydispersity)

(137a) P. J. Wyatt, D. L. Hicks, C. Jackson, and G. K. Wyatt, *Am. Lab.*, June, 1988.
(138) J. R. Clamp, *Biochem. Soc. Trans.*, (1977) 1579–1581.
(139) E. Puchelle, J. M. Zahm, and R. Havey, *Bull. Physio-Pathol. Respir.*, 9 (1973) 237.
(140) O. Harbitz, A. O. Jenssen, and O. Smidsrød, *Eur. J. Resp. Dis.*, 65 (1984) 512–520.
(141) M. F. Anderson, S. E. Harding, and S. S. Davis, *Biochem. Soc. Trans.*, 1989 (in press).

and extent of glycosylation (quasi-continuous polydispersity). In solution, they occupy highly expanded and flexible, spheroidal domains (space-filling at concentrations $\lesssim 1\%$), and behave as classical polyelectrolytes. Self-association phenomena in solution does not appear to be significant, but bronchial mucins have been shown to interact with such other components of mucus as lysozyme and alginates.

Some of the questions that remain open, such as whether the apparently similar molecular architecture between mucus glycoprotein from different sources and species is purely fortuitous or has a definite genetic basis, will be answered once detailed information concerning the peptide core of the "basic unit," the gene or genes which code for it, and the post-translational processes responsible for complex glycosylation and sulfation is available. Progress is continually being made in all these areas, using direct methods for deglycosylating the peptide core,[142] and using mRNA probes.[143,144] For example, for the related case of submaxillary glycoproteins, evidence for tandemly repeated sequences of amino acid residues has been presented from studying apomucin cDNA.[144] We can reasonably look forward to the expectation that the growing information from these genetic probes into the basic unit structure, when coupled with detailed information from the physical probes concerning the gross conformation, polyelectrolyte behavior, and heterogeneity of the glycoprotein macrostructures, will provide us with a clear understanding of the structural–functional relationships of mucins in health and disease.

Acknowledgments

The author thanks Dr. J. M. Creeth, Dr. R. H. Pain, Professor A. Allen, Dr. J. K. Sheehan, and Dr. I. Carlstedt for their comments and suggestions, and other colleagues who have made helpful comments. The support of the Lister Institute of Preventive Medicine is greatly appreciated.

(142) N. Mian, K. Marks, J. G. Widdicombe, J. R. Davies, and P. Richardson, *Biochem. Soc. Trans.*, 14 (1986) 114–115.

(143) R. Nichols, D. M. Carlson, and J. E. Dixon, *Biochem. Biophys. Res. Commun.*, 149 (1987) 244–248.

(144) C. S. Timpte, A. E. Eckhardt, J. L. Abernathy, and R. L. Hill, *J. Biol. Chem.*, 263 (1988) 1081–1088.