Glycoconjugate Journal (1992) **9**:217–221

GLYCOPINION

Editor: RAYMOND A. DWEK

In this corner of glycoconjugate chemistry where scientists have traditionally shrugged their shoulders at the vast complexity of the large sticky molecules known as mucins, light is finally breaking through. The emerging facts pose new questions and make new technological demands; they also challenge us to develop new concepts which will enable us to identify and describe the significant features of the massively glycosylated protein networks formed by these molecules.

Throughout the animal kingdom, mucosal cells secrete mucus glycoproteins (mucins) which have wide ranging protective and lubricative functions. Mucins are giant molecules, their subunits commonly have molecular weights greater than one million daltons, while native molecules climb into the ten million daltons range, of which over 80% of the dry weight is oligosaccharide. The defensive role played by mucins at mucosal surfaces depends on their characteristic physicochemical properties. Gel formation and viscoelastic behaviour in extracellular secretions are due to the ability of mucins to aggregate, to form crosslinked networks and to bind large amounts of water. These properties result from the high carbohydrate content of the molecules which is largely present in the form of O-linked oligosaccharides. The basic principles of mucin carbohydrate composition have long been known and, with the advent of improved structural analysis, a considerable catalogue of mucin oligosaccharides has been compiled. Literally hundreds of structures have been identified even within one molecule, in cases where sufficient material and patience have triumphed.

In order to understand more about the structure, function and biosynthesis of mucins several questions need to be answered. These include:

- Is the function of the carbohydrate to provide bulk packing, or do the sugars contain specific information? Does it provide both functions?
- How much of the carbohydrate is necessary for a mucin to function?
- How can we develop current sequencing techniques to analyse these oligosaccharide structures and map their arrangement in mucin networks?
- Are fingerprinting techniques a way of highlighting significant changes in oligosaccharide structures in different mucosal secretions?
- How does the cell organize the synthesis of these vast molecules?
- What are the consequences of alterations in carbohydrates? Can specific alterations be linked to different functions or disease states?
- Can we design our own mucins on the basis of current knowledge?

In this article, Dr Tony Corfield of the Bristol Royal Infirmary considers some of these issues.

Mucus glycoproteins, super glycoforms: how to solve a sticky problem?

TONY CORFIELD
Department of Medicine Laboratories, Bristol Royal Infirmary, Bristol BS2 8HW, UK

Mucin carbohydrate diversity

Early glycopinion remained sceptical of the role of the vast amount and complexity of mucin oligosaccharides. The bulk of carbohydrate was clearly related to the chemical properties, but the diversity of structure had no explanation. Furthermore, important properties of molecular organization were found to be governed by the polypeptide. Reduction of disulfide bridges and proteolytic digestion led to loss of the viscoelastic properties vital for the mucosal defence functions [1–4]. The carbohydrate appeared to be less important, present only as a packaging to achieve the necessary physicochemical behaviour. However, in spite of this aversion to a wider role for the carbohydrate, some

0282-0080 © 1992 Chapman & Hall

features of oligosaccharide structure seemed to be overembellishments on a molecule 'simply' designed to form gels. Blood group antigens were identified on mucins from secretor positive individuals, and tissue specific patterns of sialylation and sulfation were detected [1, 4, 5]. The length and branching of oligosaccharides were also found to have a tissue specific pattern, giving a characteristic mucin oligosaccharide pattern for each animal mucosa [1, 4, 5].

In the light of these uncertainties, questions have persisted relating to the nature and function of the carbohydrate in these molecules. How much carbohydrate is 'bulk' and how much is specific 'information'? It is clear that these two features overlap. The complexity of oligosaccharide structure has a role not only in the bulk provision of carbohydrate for physicochemical properties, but also in many specific information transfer functions [4, 5]. Those identified include the interaction with bacterial cell surfaces through lectins or other receptors, and the expression of blood group activity [4, 5].

The overall charge of mucin molecules is also a feature which has been identified with normal and diseased states. Neutral and acidic mucins have been described, although these remain relative terms. The expression of neutral and acidic mucins is balanced in each tissue to provide a mucosal mucus layer suitable for the defensive function required at that site. Negative charge is found due to 'peripheral' residues of sialic acids and/or sulfate. Both sialylated and sulfated forms of secreted and membrane bound mucins are related to the normal state, malignancy and metastasis [1, 6, 7]. The sialic acids are of particular interest as they are found in mucins as a family of *O*-acetylated forms: they are expressed in man on a tissue specific basis [1], they are present in short, tumour related oligosaccharides which correlate with malignant and metastatic potential [6, 7], and they are also found in cancer mucins with the sialyl-dimeric Lewis X adhesion ligand. This last example mirrors the situation on the human neutrophil surface, where the ligand is bound by the major endothelial cell surface adhesion molecule, endothelial-leukocyte adhesion molecule 1 (ELAM1).

At present we possess a great deal of new structural knowledge about only a part of a few mucins. We must now ask whether we can achieve an efficient structural analysis for mucins in general, and whether it is possible to detect the organization of individual oligosaccharides within a single macromolecule. On top of this, we will want to follow structural changes which are linked to normal variation in development and differentiation and discriminate these from the changes occurring in disease [6, 7].

Solving the structural problem

The immediate problem is to sort out the multitude of structures present in these molecules (Figs 1 and 2). The increasing number of cases implicating oligosaccharide chains in highly specific biological interactions underlines the importance of a structural approach to this work [4]. It is significant that for many years a criterion of mucin purity was the absence of mannose–indicative of *N*-linked oligosaccharides. Molecular biology nucleotide sequence work has come to the rescue of those who have consistently found trace amounts of mannose in purified mucin preparations, by identifying Asn-X-Ser/Thr glycosylation sites in the polypeptide [4]. In spite of this, there is still no structural information for any *N*-linked oligosaccharide chain from a purified mucin. The problem does not rest there, because the enormous complexity of the *O*-linked oligosaccharide complement for each mucin requires a mammoth input of glycoanalysis to solve the structures [4, 5]. At present this remains a daunting task in glycobiology. There are still difficulties in the quantitative release of carbohydrate from mucins in a form that can be analysed structurally. Following this, there are problems in separating the very large numbers of charged and neutral oligosaccharides into individual components [4]. Finally, peripheral modifications such as sialylation, sulfation and *O*-acetylation may be lost during the isolation procedures. There is no doubt that the new wave of instrumentation and techniques for oligosaccharide analysis has brought with it a dramatic improvement in our ability to 'see' carbohydrate sequence and shape in glycoconjugates. The mucins present extra problems of quantity and detail which tax these methods to the limit and are yet to be satisfactorily addressed.

This is only a start to understanding the organization of oligosaccharides within a mucin polypeptide. There is some evidence from glycoproteins containing single *O*-glycosylation sites that the process of glycosylation is affected by the position of the serine or threonine residue in the polypeptide. Modification of the amino acid sequence including the linkage serine or threonine residue has revealed features of polypeptide structure which influence the initial glycosylation step. There does not appear to be a set of rules such as those which exist for Asn-X-Ser/Thr *N*-glycosylation sites. As the mucins present a structural problem simply identifying the location of individual oligosaccharides within the polypeptide, this type of sequence analysis is yet to be established successfully.

Mucins as superglycoforms

Current interest in glycoforms has highlighted the biological importance of oligosaccharide variation under defined conditions of glycosylation. Recombinant glycoproteins are currently a talking point in this respect, and there are also examples where changes in oligosaccharide structure occur under normal physiological conditions and correlate with altered glycoprotein function. Amongst the best examples of such glycoforms are tissue plasminogen activator, IgG and alpha1-acid glycoprotein. This is the beginning of a

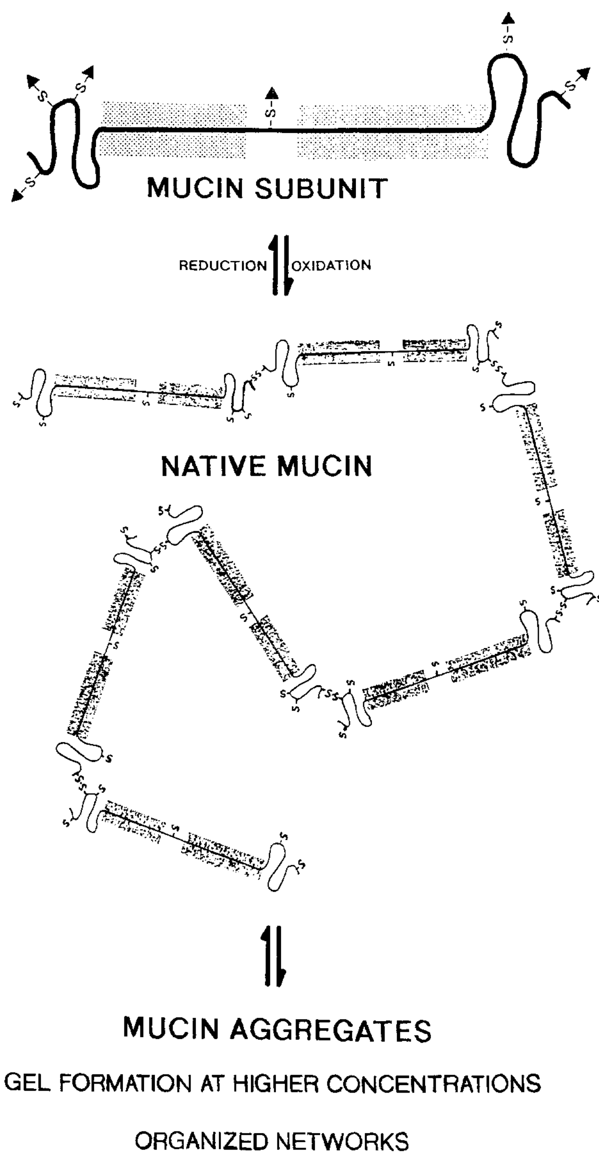

Figure 1. Organization of mucin structure.

particularly exciting field for glycobiologists. So far the majority of the information is for *N*-linked oligosaccharide chains. Application of this kind of analysis to mucin *O*-glycosylation in physiologically interesting situations means that we must deal with mucins as superglycoforms, and this is still a distant hope at present. However, the rewards are potentially very great, as the mucins are markers for normal development in mammalian mucosae [2, 3, 5], their carbohydrate structure changes with mucosal cell differentiation [2, 3, 5], and they also carry disease related carbohydrate epitopes [6, 7]. Thus, there is considerable interest in obtaining detailed structural and organizational information to 'map' tissue specific changes in health and disease.

How do cells produce mucins?

A final area of interest concerns the biosynthesis of these molecules. The mucosal cell must organize and regulate the synthesis of 'bulk' and 'informational' oligosaccharide structures within one molecule, on the skeleton of a polypeptide backbone. The cloning of mucin polypeptides has shown that a number of genes, the MUC genes, code for these products [4]. The detail gained from cloned peptide sequences has confirmed the existence of highly glycosylated regions, coded for in the form of tandem repeat structures with very high serine and threonine contents [4]. Other, 'globular' peptide regions contain a few sites of *O*-linked oligosaccharide attachment but, more importantly, Asn-X-Ser/Thr sites for *N*-glycosylation and the cysteine residues which allow disulfide bridge formation and thus crosslinking of the basic subunits. The same mucin polypeptides have been found expressed in different tissues and in various combinations [4]. It is likely that more MUC genes will be found, but these polypeptides function as skeletons for variable glycosylation on a tissue specific basis. It is the carbohydrate pattern which gives a mucin its tissue 'identity'.

The subcellular organization of mucin glycosylation has been poorly understood due to our incomplete knowledge of mucin structure. Part of this puzzle is now fitting into place (Fig. 3). The small number of *N*-linked oligosaccharides may have a role in the 'sorting' of mucin polypeptides into suitable compartments for the major *O*-linked oligosaccharide glycosylation process. Subsequent *O*-glycosylation follows by the stepwise action of glycosyltransferases, but there is no block synthesis as occurs in the *N*-linked oligosaccharides [8]. A feature of oligosaccharide biosynthesis is its reliance on the substrate specificity of the glycosyltransferases to reproduce important informational structure with high fidelity [8]. This process is not under the direct genetic control governing polypeptide sequence. The glycosyltransferases are the gene products but their specificity may be regulated by competition with other glycosyltransferases and also through limitations of substrate availability, i.e., suitable acceptor molecules and nucleotide sugar donors [8].

An important step in regulation of mucin oligosaccharide structure comes with the formation of the core unit attached to the peptide. There are only five or six common core units found in mucin oligosaccharides [1, 4] and the glycosyltransferases responsible for the synthesis of these structures are assumed to interact to conserve tissue specific oligosaccharide core structures [8]. The nature of the overall oligosaccharide chain depends on the interaction of these initial core glycosyltransferases with those responsible for formation of long 'backbone' repeat structures, branches or terminating residues resulting in short chains [1, 6–8]. The substrate specificity of these enzymes is now understood in many cases but their organization into functional groups

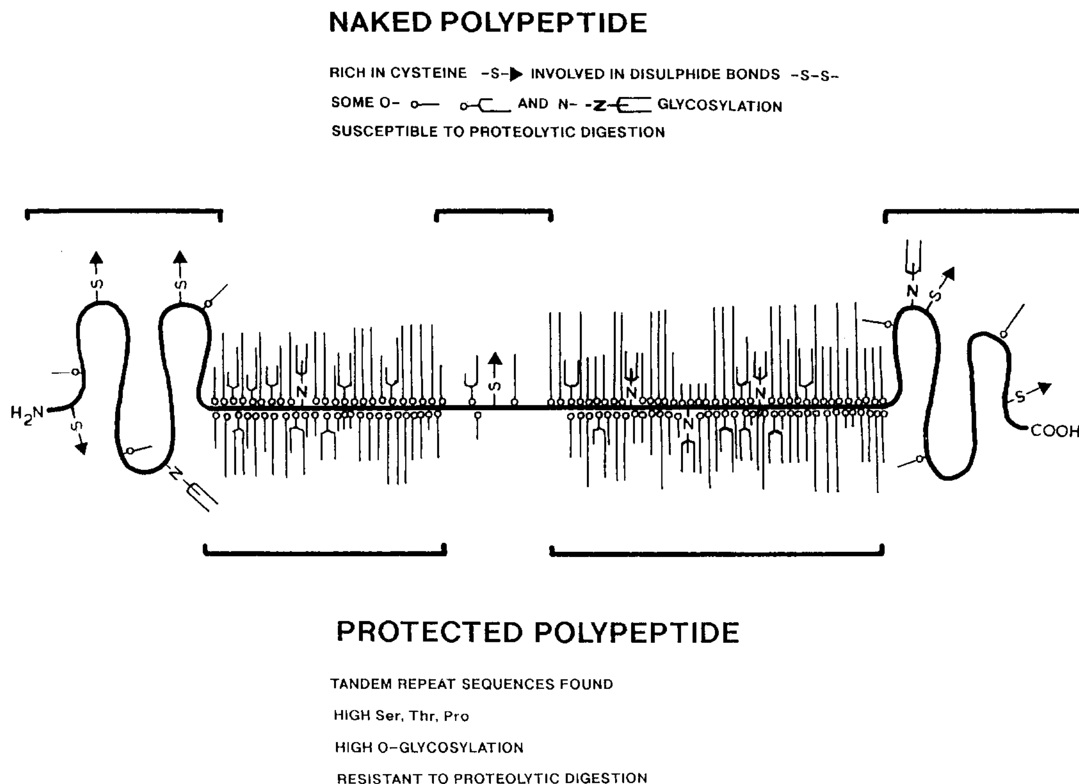

Figure 2. Diagrammatic representation of mucin subunit structure.

(multiglycosyltransferase complexes) is an old idea with little experimental proof. Many of the 'backbone' and peripheral features are shared with other glycoconjugates and the 'sharing' of multiglycosyltransferase complexes is still speculation rather than hard fact [8]. Answers to questions concerning the organization of glycosylation are likely to come from the cloning of individual glycosyltransferases, but must await data from sufficient enzymes to make a real impact on the assessment of the organization of total mucin synthesis.

Designer mucins and the future

The mucins are indeed a sticky problem. However, it is worth asking what we can do to answer some of the more immediate questions where mucin structure has gone wrong. Abberant mucin may occur in the form of inappropriate physicochemical properties, too much or too little secretion. The production of synthetic tears in dry eye diseases, strengthening of mucin gels in stomach ulcer and inflammatory bowel diseases and gels suited for transport on mucociliary mucosae in respiratory and genital tract ailments are examples where help is required. Designing mucins is thus a desirable aim. Have we progressed far enough to design our own mucins? Our improved understanding of mucin structure–function relationships indicates the type of mucin required in the tissue affected. Tandem repeat peptide sequences provide matrices for glycosylation and suggest the level of packing required for the oligosaccharide chains. The core, backbone and peripheral units may be constructed from recombinant multiglycosyltransferase complexes of cloned enzymes, and the whole molecule may be constructed in an efficient bioreactor which supplies and cycles substrates and products. Such bioreactors are now being used by glycochemists for the synthesis of glycoconjugates using enzyme specificity to create the required sequences. This 'designer' approach will allow us to find out how much bulk carbohydrate is necessary to create the right physicochemical properties and show how to produce mucins which have the right properties for each tissue situation. In the face of the enormous structural analytical problems facing 'classical mucinologists' studying these superglycoforms, this way forward must merit serious attention.

The challenges facing the study of mucin superglycoforms are indeed as big as the molecules themselves. The options open to us may lead into improved and novel structural analytical techniques or come from the dissection of the cellular machinery responsible for their biosynthesis. Designer mucins is a new idea which may throw up answers before either of the other options. It is certain that these sticky molecules warrant such attention, and that their secrets are now starting to be unravelled.

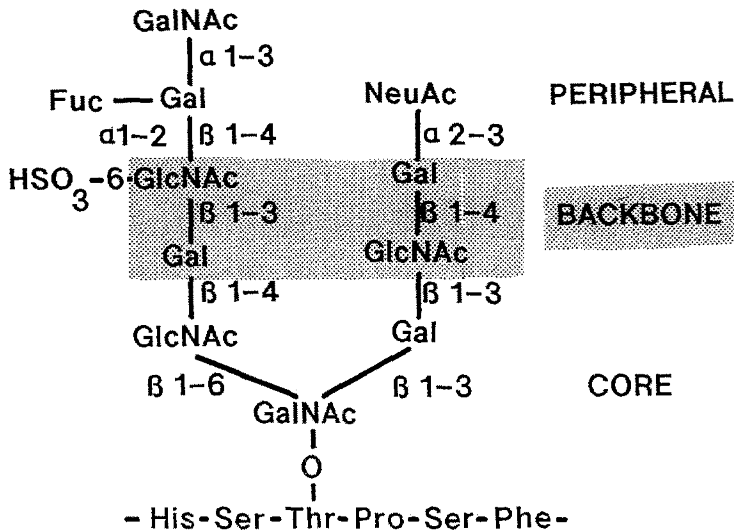

Figure 3. *O*-Linked oligosaccharide features in mucins.

References

1. Carlstedt I, Sheehan JK, Corfield AP, Gallagher JT (1985) *Essays Biochem* **20**:40–76.
2. Allen A (1989) in *Handbook of Physiology—The Gastrointestinal System* III (Forte JG ed.) Chapter 19, pp. 359–82. American Physiological Society.
3. Neutra MR, Forstner JF (1987) in *Physiology of the Gastrointestinal Tract* (Johnson LR ed.) 2nd edition, Vol 2, pp. 975–1009.
4. Various authors (1991) *Am Rev. Respiratory Diseases* **144**: Part 2, Supplement S1–S56.
5. Feizi T, Childs RA (1987) *Biochem J* **245**:1–11.
6. Feller WF, Henslee JG, Kinders RJ, Manderino GL, Tomita JT, Rittenhouse HG (1990) *Immunol Ser* **53**:631–72.
7. Irimura T, Matsushita Y, Hoff SD, Yamori T, Nakamori S, Frazier ML, Giacco GG, Cleary KR, Ota DM (1991) *Seminars Cancer Biol* **2**:129–39.
8. Schachter H, Brockhausen I (1989) *Symp Soc Exp Biol* **43**:1–26.