thank also Mr. S. Kurono for his help in technical assistance.

(1) Y. Kato and Y. Numajiri, *J. Chromatog.*, **562**, 81 (1991).

S.20 MUCINS AND MUCIN GENES

S20.1
Molecular Biology of Tracheobronchial Mucin Glycoprotein

M. Verma and E. A. Davidson
Department of Biochemistry and Molecular Biology, Georgetown University, Washington, D.C. 20007, USA.

A lambda gt11 cDNA library prepared from canine tracheal epithelial cells was screened with polyclonal anti-apo-canine tracheal mucin antibodies with the aim of obtaining the deduced amino acid sequence of the mucin core protein. Antibody positive clones of varying lengths were purified and used for nucleotide sequencing. Based on this information, synthetic oligonucleotide primers were constructed and both ends (5′ and 3′) of the DNA determined. The complete sequence was 3.7 kb which included an open reading frame with coding capacity for 1119 amino acids, two translation initiation ATG codons in context with Kozak consensus sequences, one polyadenylation site and a polyA tail. The protein was rich in thr, ser, pro, gly and ala and poor in tyr, phe and trp. Although tandem repeats of amino acids were absent in the deduced sequence, motifs TPTPTP and TTTTPV appeared 13 and 19 times respectively. The carboxyl terminal region contained a cysteine-rich domain as has been reported for several other mucins (bovine submaxillary, rat intestinal etc.). The significance of this domain is not clear at present. Three potential *N*-glycosylation sites were present and the amino acid sequence showed homology with human intestinal mucins and with partial data reported for human tracheal mucin. Structural analysis indicated that the *N*-terminal domain is more flexible. Primary transcripts appear polydisperse when total RNA is examined but exhibit discrete bands when polyA positive RNA is probed. We have established an immortalized line of canine tracheal epithelial cells that express the mucin gene and will form the basis for regulatory and other studies.

S20.2
Initiation of *O*-Glycosylation of Proteins

H. S. Conradt and E. Grabenhorst
Dept. of Cell Biology and Genetics, GBF, D-3300 Braunschweig, Federal Republic of Germany.

Our studies on a number of human secretory proteins suggest that a structural motif is recognized by the enzyme initiating, *O*-glycosylation, since all proteins secreted from different recombinant host cell lines analyzed so far bear *O*-glycans at the same amino acid that is occupied by carbohydrate in their natural counterpart. In all cases > 90% of *O*-glycans have the Galβ1-3GalNAc core structure with one or two NeuAc (BHK, CHO, Ltk$^-$, Hela and C 127 cells; C 127 cells synthesize a small proportion of sulfated *O*-linked structures). Comparison of some 120 *O*-glycosylation sites in mammalian proteins do not reveal any primary sequence similarity which may be recognized by the polypeptide: α1-*O*-GalNAc transferase.

By site directed mutagenesis of several cDNAs coding for human glycoproteins we could show that *O*-glycosylation of a given protein at a given site is *specific* for Ser or Thr, irrespective of the host cell line used.

An artificial *O*-glycosylation site (..G-G-A-P-T-P-P-P-G-G...) when introduced at different locations of model proteins could be shown to be quantitatively *O*-glycosylated in all cell lines tested whereas the wild-type glycosylation site of the proteins are recognized by GalNAc-transferase only at their natural positions.

Computer modelling of proteins/peptides containing natural as well as artificial *O*-glycosylation acceptor sequences will be presented.

In vitro glycosylation studies, using small (12 AA) synthetic peptide substrates reveal that initiation of *O*-glycosylation occurs also at sites which are not recognized when present on proteins expressed from recombinant cells. However, proteins and larger synthetic peptides (26-29 AA) containing secondary/tertiary structural elements (CD and NMR) adjacent to *O*-glycosylation sites were glycosylated at the same hydroxy amino acid observed in the recombinant proteins.

Studies using expression of mutant proteins bearing *O*- as well as *N*-glycosylation sites indicate that initiation of *O*-glycosylation (GalNAc transfer in BHK cells) occurs after *N*-glycans become resistant to treatment with Endo H.

S20.3
Molecular Cloning and Partial Sequencing of the Apoprotein for the Insoluble Mucin of the Rat Small Intestine

G. C. Hansson, D. Baeckström, I. Carlstedt[2] and K. Klinga[3]
Department of Medical Biochemistry, University of Göteborg, Göteborg, Sweden; [2]Department of Physiological Chemistry, University of Lund, Lund, Sweden; [3]Department of Genetics, University of Göteborg, Göteborg, Sweden.

Glycopeptide A (M_r 650 kDa) from the major mucins of rat small intestine, insoluble also in 6 M guanidinium chloride, was isolated. This glycopeptide was deglycosylated with hydrogen fluoride, the apoprotein dialysed and antibodies raised in rabbits. The antibodies were used to screen plaques formed from a rat intestinal cDNA library in lambda ZAPII. Poitive clones were detected, cloned into Bluscript and sequenced. The amino acid sequence of one insert reveal both a tandem repeat region rich in especially Thr and a unique region. Both regions were typical for the features of mucins. Notherns blots showed a distinct band of about 16 kb. This is much larger than the usual smear found for other mucins, due to no degradation of the mRNA as also MUC2, MUC3 and MUC4 show distinct large bands at notherns in our hands. The gene has been localized to rat chromosome 2.

The amino acid sequence of this mucin combined with the physical and oligosaccharide data presented separately will provide detailed information of structure of this type of mucin.