Human Mucin Genes MUC2, MUC3, MUC4, MUC5AC, MUC5B, and MUC6 Express Stable and Extremely Large mRNAs and Exhibit a Variable Length Polymorphism

AN IMPROVED METHOD TO ANALYZE LARGE mRNAs*

(Received for publication, July 3, 1997, and in revised form, October 31, 1997)

Virginie Debailleul‡, Anne Laine‡, Guillemette Huet‡§, Pascal Mathon‡¶, Maud Collyn d’Hooghe, Jean Pierre Aubert‡§, and Nicole Porchet‡§**

From the ∥Unité INSERM U 377, Laboratoire Gérard Biserte and [Unité INSERM U 124, place de Verdun, 59045 Lille Cedex, France, the §Laboratoire de Biochimie et Biologie Moléculaire de l’Hôpital Claude Huriez, Centre Hospitalier Régional Universitaire de Lille and ¶Faculté de Médecine, Université de Lille II, 59037 Lille Cedex, France

Of the nine mucin genes that have been characterized, only MUC1 and MUC7 have been fully sequenced, and their transcripts can be detected as distinct bands of predicted size by Northern blot analysis. In contrast, the RNA patterns observed for each of the other MUC genes have usually shown a very high degree of polydispersity. This polydispersity has been believed to be one of the typical features of the mucin mRNAs, but until now, its origin has remained unexplained. In the work described in the present paper, we investigated two possible kinds of explanation for this phenomenon: namely that the extensive polydispersity results from a biological mechanism or that it is artificial in origin. The data obtained, as a result of improving the purification and blotting methods, allowed us to show that in all of the tissues analyzed, each of the genes, MUC2–6, expresses mRNAs that are stable and are of an unusually large size to be found in eukaryotes (14–24 kilobases). Moreover, allelic variations in length of these mucin transcripts were observed. We demonstrate that these variations are directly related to the variable number of tandem repeat polymorphisms seen at the DNA level.

Mucins are a heterogeneous family of O-linked glycoproteins expressed in epithelial cells. These molecules, which are secreted or membrane-associated glycoproteins, are believed to play an important role in the protection and function of the epithelial surfaces.

Eight human mucin genes have been characterized that have been given MUC gene symbols numbered 1, 2, 3, 4, 5AC, 5B, 6, and 7 (1). More recently, a cDNA isolated from a bronchus library by Shankar et al. has been called MUC8 (2). Of these genes, MUC1 and MUC7 have been fully sequenced (3, 5). MUC1 encodes PEM, an integral membrane mucin glycoprotein expressed on the apical surface of most simple secretory epithelia (3). Expression of this gene has been extensively studied in numerous cell lines and epithelial tissues. Two transcripts are usually detected by Northern blot analysis in most individuals when a tandem repeat probe is used, their sizes being estimated between 4 and 7 kb (6, 7). The presence of these two messenger RNAs is a result of allelic variation due to a variable number of tandem repeats (VNTR) per allele of a 20-amino acid motif (from 20 to 125 repeats) (3, 4). The other MUC genes are believed to encode secreted mucins. The full-length cDNA sequence of MUC7 has been published recently (5). The MUC7 gene encodes a small secreted mucin expressed by submandibular and sublingual glands. MUC7 mRNA is 2.4 kb in length. This messenger appears as a distinct and moderately polymorphic band, most individuals exhibiting a unique allelic form corresponding to six repeats in their repetitive domains (5, 8).

Sequencing of the genes MUC2–6 is at different stages of progress. The complete structure of MUC2 cDNA is known and published spanning 15,563 base pairs (9). The MUC2 tandem repeat domain is polymorphic, and variations in length of the MUC2 mRNA might be expected among individuals; however, the MUC2 message is not limited to two allelic forms of one transcript but is very often polydisperse by Northern blot analysis (10–12). The MUC5B gene is entirely cloned and spans approximately 40 kb. The central exon of this gene, encoding the whole tandem repeat domain, was completely sequenced recently and is 10,713 base pairs in length (13). MUC5B shows the same polydisperse signal on Northern blots as the MUC2 probe (14). The complete genomic organization of the 3’-region of MUC5B was published very recently (15). No other precise information is available for any of these genes on the genomic organization, the existence of alternative splicing, or the number and the length of their messenger RNAs. In our previous studies, as mentioned for MUC2, we reported smears or very wide bands on Northern blot analyses of the three mucin genes we cloned: MUC4 (16), MUC5AC (17), and MUC5B (14). This has also been reported for MUC3 and MUC6 by the group who cloned them (12, 18). A large number of investigators, using other partial mucin cDNA probes (11, 19–21) or comparing other partial mucin cDNA probes (11, 19–21) or comparing

* This work was supported by ‘‘ l’Association pour la Recherche sur le Cancer,’’ ‘‘ le Comité du Nord de la Ligue Nationale Contre le Cancer,’’ and the Centre Hospitalier Régional Universitaire de Lille (PHRC 1994 and Contrat 96/09.29/9595: Appel d’Offres de Recherche du CH et U 1996). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby regarded as ‘‘advertisement’’ in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed. Tel.: 33-3-20-29-88-58; Fax: 33-3-20-53-85-62; E-mail: jpa@lille.inserm.fr.

1 The abbreviations used are: kb, kilobase(s); apoB-100, apolipoprotein B-100; Act D, actinomycin D; VNTR, variable number of tandem repeats; MTX, methotrexate; SSPE, saline/sodium/ephosphat/EDTA; MOPS, 3-[N-morpholino]proanesulfonic acid.

This paper is available online at http://www.jbc.org

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1998 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 273, No. 2, Issue of January 9, pp. 881–890, 1998
Printed in U.S.A.
acteristics of the mucin mRNAs. Two kinds of hypotheses have been proposed to explain these polydisperse signals: those that propose a biological origin due for example to rapid turnover, the existence of several related genes encoding similar transcripts varying in length, or the occurrence of alternative splicing; and those that suggest an artifactual origin due to specific properties of the mucin transcripts such as their unusual length and thus sensitivity to mechanical damage or a particular conformation due to a large tandem repeat domain that perturbs the migration of the molecules in electric fields. To elucidate this phenomenon, we chose three different approaches. (i) We determined the half-life of secreted mucin transcripts (MUC2, MUC3, MUC5AC, and MUC5B) expressed by the HT29-MTX cell line cultured in the presence of actinomycin D with or without cycloheximide and compared it with that determined for MUC1. (ii) We studied the heterogeneity of mucin messenger RNAs isolated from polysome preparations from the HT29-MTX cell line and from human colonic mucosa and compared the patterns given by nuclear, cytoplasmic, and total RNA preparations by Northern blot analysis. (iii) We explored the hypothesis that conventional methods of mRNA purification are not adapted to the study of very large mucin messenger RNAs, and we developed a modified protocol for the extraction and purification of mucin transcripts with the aim of reducing mechanical degradation. Moreover, we improved the blotting method and investigated the efficiency of poly(A)+ selection.

The data obtained allowed us to conclude that all the tissues analyzed expressed muc mRNAs that are stable and of unusually large size to be found in eukaryotes (14–24 kb) and in most cases show allelic length variations. We demonstrate that these variations are directly related to the variable number of tandem repeat polymorphisms seen at the level of the DNA.

**EXPERIMENTAL PROCEDURES**

**Study of the Half-life of Mucin mRNAs Expressed by HT29-MTX Cell Line**—Parental HT29 cells selected by adaptation to 10−3 m methotrexate (HT29-MTX) were obtained from Drs T. Lesuffleur and A. Zweibaum (Unité INSERM 178, Villejuif, France) (37, 38). Cells were maintained in culture until day 16, before the addition of 10−3 M methotrexate and by dot blot using previously described technical procedures (18), (19), and (20). A 10,000× g pellet was washed with 0.1% diethylpyrocarbonate, and it was then autoclaved. 2-Mercaptoethanol was added to 100 mM before use. The cesium chloride cushion (5.7 M) was sterilized by filtration before use. All experiments were performed at 4°C. Total RNA was pelleted from the HT29-MTX cell line cultured in the presence of actinomycin D with or without cycloheximide and was then autoclaved. 2-Mercaptoethanol was added to 100 mM before use. The cesium chloride cushion (5.7 M) was prepared by dissolving 24 g of cesium chloride in 25 ml of 0.1 M EDTA, pH 7.5, treated with 0.1% diethylyprocarbonate, and then autoclaved.

**Isolation of Nuclear, Polysomal, and Cytoplasmic RNAs**—Large RNAs were extracted from different human tissues (bronchus, antrum, fundus, colon) using an improved method we developed, derived from the guanidinium isothiocyanate protocol (40) as follows.

Guanidinium isothiocyanate buffer (4 M) was prepared by dissolving 23.6 g of guanidinium isothiocyanate, 73.5 mg of sodium citrate, and 250 mg of sodium N-lauroylsarcosine in 50 ml of water, treated with 0.1% diethylyprocarbonate, and it was then autoclaved. 2-Mercaptoethanol was added to 100 mM before use. The cesium chloride cushion (5.7 M) was prepared by dissolving 24 g of cesium chloride in 25 ml of 0.1 M EDTA, pH 7.5, treated with 0.1% diethylyprocarbonate, and then autoclaved.

**Human Mucin mRNAs**—We adapted a protocol that combined two procedures published previously (44, 45), which we applied to tissues (−0.5 g) or to cells (106 cells). Solid tissue was taken from liquid nitrogen storage, whereas the HT29-MTX cells were harvested at day 16 and used immediately for polysomal isolation (two 75-cm2 flasks). To minimize RNAse contamination, all glassware was heated at 180°C for 2 h, and all buffers were prepared using sterile 0.1% diethylyprocarbonate-treated water and were autoclaved, except for the magnesium acetate stock solution, which was sterilized by filtration before use. All experiments were performed at 4°C. The biological material was homogenized in liquid nitrogen in 10 ml of 25-ml tubes containing 1 ml of 200 m M Tris-HCl, pH 7.6, 100 mM KCI, 5 mM magnesium acetate, 10 mM 2-mercaptoethanol, 2 mM dithiothreitol, 0.1 unit/μl RNasin, and 0.25 μM sucrose. The homogenate was gently thawed and disrupted with 10 strokes of a Potter-Elvehjem homogenizer. Cell lysis was monitored by using trypan blue dye. The mixture was centrifuged at 10,000 × g for 10 min at 4°C, and a pellet of nuclei was thus obtained. The supernatant was carefully and gently removed and layered over 3 ml of a 1 M sucrose solution containing 20 mM Tris-HCl, pH 7.6, 100 mM KCI, 5 mM magnesium acetate, 10 mM 2-mercaptoethanol, 2 mM dithiothreitol, 0.1 unit/μl RNasin. Ultracentrifugation was performed at 38,000 rpm in a Beckman SW41 rotor for 3 h at 4°C, resulting in a polysomal fraction (pellet) and a cytoplasmic fraction (supernatant). The three different fractions containing nuclear, polysomal, and cytoplasmic RNAs, respectively, were extracted in guanidinium isothiocyanate solution and stored at −80°C. Purification of RNA was performed from these three fractions as described above.

**Improved Method for the Isolation and Transfer of Large RNAs**—Large RNAs were extracted from different human tissues (bronchus, antrum, fundus, colon) using an improved method we developed, derived from the guanidinium isothiocyanate protocol (40) as follows.

Guanidinium isothiocyanate buffer (4 M) was prepared by dissolving 23.6 g of guanidinium isothiocyanate, 73.5 mg of sodium citrate, and 250 mg of sodium N-lauroylsarcosine in 50 ml of water, treated with 0.1% diethylyprocarbonate, and it was then autoclaved. 2-Mercaptoethanol was added to 100 mM before use. The cesium chloride cushion (5.7 M) was prepared by dissolving 24 g of cesium chloride in 25 ml of 0.1 M EDTA, pH 7.5, treated with 0.1% diethylyprocarbonate, and then autoclaved.

Tissue (optimal weight of 1 g) was ground to a fine powder in a mortar and pestle in liquid nitrogen and mixed with 10 ml of guanidinium isothiocyanate buffer, still in liquid nitrogen. The homogenate was then allowed to thaw gradually at room temperature, during which time the guanidinium isothiocyanate and 2-mercaptoethanol efficiently solubilized the tissue mixture. The homogenate obtained was then transferred to 10 ml of cesium chloride cushion followed by ultracentrifugation for 16 h at 29,500 rpm in a Beckman SW41 rotor. The supernatant was removed, and the pellet of total RNA was very carefully resuspended in 0.1 × Tris-EDTA buffer, pH 8.0 (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA, pH 8.0), 0.1% SDS by using wide mouth pipettes and avoiding all shear forces. RNA was purified by the addition of two volumes of chloroform/methyl alcohol (4:1) mixture; the tubes were mixed by gentle inversion and vortexing 20 times. The top aqueous phase, which contained the RNA, was carefully removed with wide mouth pipettes and transferred to a fresh tube, and the RNA was precipitated by adding 0.1 volume of 3 M sodium acetate, pH 5.5, and 2.5 volumes of ethanol, at −80°C for 15 h. After centrifugation at 10,000 × g for 30 min at 4°C, the pellet of RNA was washed with ice-cold 90% ethanol and then 100% ethanol, centrifuged as above, and dried in a vacuum desiccator. The RNA pellet was very carefully resuspended in diethylyprocarbonate-treated water and quantified by measuring the A260 of an aliquot, and the final preparation was stored at −80°C. Extreme care was taken at each step during the experiments to prevent any risk of mechanical degradation. Homogenization, the use of syringes, and vortexing were strictly avoided.

RNA samples (the optimal quantity was 10 μg) were denatured in loading buffer containing 50% formamide, 18% deionized formaldehyde, and 0.02 M MOPS; heated at 68°C for 10 min; and then fractionated through a 0.9% agarose denaturing gel (3×18×0.3 cm) containing 18% formaldehyde and 0.02 M MOPS. The gel running buffer was 0.02 M MOPS, pH 7.0. The electrophoresis was performed for 16 h at 30 V. Prior to transfer, the gel was soaked for exactly 20 min in 0.05 M NaOH (three changes of gel buffer for 5 min). The gel was then rinsed in RNase-free water and soaked for 45 min in 20× SSC (46). The RNA was transferred onto Hybond N+ membrane (Amersham Corp.) via vacuum blotting (1 h in 20× SSC) and immobilized by baking at 80°C in a vacuum for 30 min and by exposure to UV light for 4 min. Standard RNA markers were obtained from Life Technologies, Inc. and from Boehringer Mannheim (Germany). The apoB-100 DNA fragment cloned
of both the band and the smear detected with all of the probes, of contact with Act D, there was a general decrease in intensity. D (T), the hybridization signal appeared in most cases as a

In the first experiment, the treatment time with Act D. In the first experiment, the treatment time with Act D: 3, 6, 9, 16, and 24 h.

To confirm this result and to estimate the half-life of mucin mRNAs, other experiments were performed adding longer incubation times with Act D: 3, 6, 9, 16, and 24 h. MUC2, MUC3, MUC5AC, and MUC5B messenger RNAs expressed in the HT29-MTX cell line were quantified by dot-blot analysis using 28 S rRNA as a control for the amount of total RNA analyzed in each assay. For each probe and for each incubation time with Act D, dot-blot analysis was performed four times, and the mean of these four values was calculated. The mucin mRNA/28 S rRNA hybridizations of the same blot are shown.

FIG. 1. Northern blot analysis of MUC5AC mRNA in HT29-MTX cells treated with Act D for 3 and 9 h before RNA purification. 20 µg of total RNA were loaded in each lane. MYC and 28 S rRNA hybridizations of the same blot are shown.

### RESULTS

**Study of the Half-life of Mucin mRNAs in Act D-treated HT29-MTX Cells**—To demonstrate the possible role a rapid turnover of mucin mRNA may have in leading to the polydisperse pattern shown on Northern blots, we blocked the synthesis of mRNA using actinomycin D (Act D), which is a transcription inhibitor. HT29-MTX cells were harvested at various times from 3 to 24 h after Act D treatment, and total RNA was extracted according to Ref. 40. The RNA transcripts in each sample were analyzed by Northern blotting and hybridized with probes for MUC2, MUC3, MUC5AC, and MUC5B. Several experiments were performed using different times of treatment with Act D. In the first experiment, the treatment time with Act D was 3 and 9 h. For RNA from cells cultured without Act D (T), the hybridization signal appeared in most cases as a band with a smear as described previously (51). After 3 and 9 h of contact with Act D, there was a general decrease in intensity of both the band and the smear detected with all of the probes, reflecting inhibition of transcription. MUC5AC is shown as an example in Fig. 1. The persistence of the hybridization signal of MUC2, MUC3, MUC5AC, and MUC5B mRNA at high levels in the presence of Act D at 9 h indicates that mucin transcripts in these cells are not particularly unstable and that the polydisperse pattern cannot be explained by a rapid turnover. The disappearance of MYC mRNA was a control of Act D efficiency (Fig. 1) (43).

To confirm this result and to estimate the half-life of mucin mRNAs, other experiments were performed adding longer incubation times with Act D: 3, 6, 9, 16, and 24 h. MUC2, MUC3, MUC5AC, and MUC5B messenger RNAs expressed in the
lar fractions (RNA from the nucleus, RNA specifically associated with polysomes, and RNA from cytoplasm), and we compared their patterns by Northern blot analysis. Analysis of these three fractions was aimed at discriminating subpopulations of mRNAs, namely newly translated RNA (associated with polysomes), pre-mRNA (found in the nucleus), and degraded RNA (found in the cytoplasm). The RNAs isolated from these three fractions and the corresponding total RNA were analyzed by Northern blotting (Fig. 3). The expression pattern of MUC2 was compared in the four preparations, the β-actin probe being used as a control of RNA integrity. The same polydisperse signal was observed in the total RNA fraction as well as in each of the three subfractions: nuclear, polysomal, and cytoplasmic. No variations in the distribution of mucin mRNAs in the smear were observed, indicating that the different fractions of RNA cannot be distinguished by length, i.e., RNAs that have the ability to be translated and are therefore physiologically important were no longer than the mRNA species in the cytoplasm, and the pre-mRNA in the nucleus was also not different. The same results were obtained with HT29-MTX cells (data not shown).

These results indicate that a biological mechanism is not responsible for the polydisperse pattern of mucin mRNAs.

Improved Method for the Isolation and Transfer of Large RNAs—Total RNA was prepared from human colon mucosa using the modifications we made to the original guanidinium isothiocyanate-ultracentrifugation method. These modifications mainly involved taking extreme care to prevent all risk of mechanical degradation by shear forces. The efficiency of this improved method for RNA isolation, compared with the original protocol, is illustrated in Fig. 4, A and B. Using our method, the Northern blot hybridized here with the MUC2 probe exhibits one distinct band (Fig. 4B). The size of this transcript is larger than 10 kb. The intensity of the smear was significantly decreased.

The efficiency of the transfer of large mRNAs was also improved; two different pretreatment protocols of the gel were compared. The first method, which had previously been used routinely, included three rinses in 10 × SSC (3 × 10 min) and one rinse in 20 × SSC (10 min). In the second method, the gel was soaked in 0.05 N NaOH (for a 3-mm-thick gel, the optimal signal was obtained after treatment for 20 min) prior to rinsing.
in RNase-free water and soaking for 45 min in 20× SSC (46). Aliquots (10 μg) of the same RNA sample prepared using our improved method (Fig. 4, B and C) were used. The intensity of the RNA band was significantly higher when the gel was soaked in 0.05 N NaOH before transfer. Transfer efficiency was increased approximately 10-fold. Moreover, only a very weak smear was observed.

Comparison with Another Large mRNA, ApoB-100—To confirm this result and to determine whether the degradation is an exclusive feature of mucin mRNA or affects other large messenger RNAs, we compared the patterns obtained with mucin probes and with the apoB-100 probe. The apoB-100 cDNA has been entirely sequenced; the apoB-100 transcript is 14.1 kb long (52–54). We compared two samples of RNA prepared from the same tissue specimen (small intestine) using the original guanidinium isothiocyanate ultracentrifugation protocol and our improved method. Each Northern blot was hybridized with MUC3 and apoB-100 probes. Using the guanidinium isothiocyanate ultracentrifugation protocol, a smear was obtained with the MUC3 probe as expected, and a polydisperse pattern was also obtained with the apoB-100 probe (Fig. 5A). Using our improved method (purification and transfer), distinct bands were obtained with both the MUC3 and apoB-100 probes (Fig. 5B).

Transcripts Expressed by the Mucin Genes MUC2 to MUC6—To determine the number of the transcripts expressed by each of the mucin genes, MUC2 to MUC6, preparations of RNA from human mucosae (bronchus, antrum, fundus, and colon) available from seven individuals were analyzed by Northern blotting using the improved RNA purification method and introducing the treatment with 0.05 N NaOH before the transfer. All of the mucin probes (MUC2 to MUC6) revealed either a single band or a double band where both bands had the same intensity. Moreover, the size of the band(s) can vary between individuals (Fig. 6). On the other hand, when the same specimens were extracted and purified using other methods (described under “Experimental Procedures”), smears were always obtained on Northern blots (data not shown).

The standard RNA molecular size markers used ranged from 0.24 to 9.49 kb. All MUC2–6 transcripts were much larger than 9.49 kb. To determine whether our standard curve could be meaningfully extrapolated, we used the apoB-100 probe, which gave a positive signal on Northern blots of colonic RNA, but this overestimated the size of the apoB-100 mRNA, giving a value of 22 rather than 14 kb, showing that the use of standard RNA molecular size markers is misleading in the case of very large RNAs. So, we used another standard curve using three points corresponding to apoB-100, 28 S rRNA, and β-actin signals. These points were not on a perfect straight line, so they were joined to form a nonlinear curve (Fig. 7), and where necessary (up to 14 kb), we used as an additional point
MUC5B, which has been shown to be 17.6 kb in length, is expressed in colon, and does not show common variation. Under these conditions, the sizes of mucin mRNAs were estimated from 14 to 24 kb. The results are shown in Table I.

Polymorphism Studies: Comparison between Southern Blot and Northern Blot Patterns—With mucin probes MUC2, MUC3, MUC4, MUCSAC, MUC5B, and MUC6, a single band or a double band was observed, depending on the individuals (Fig. 6). When two bands were present, they appeared with the same intensity, suggesting that two alleles are co-expressed. Moreover, RNA sizes vary among the individuals, except for MUC5B. To demonstrate that a double band corresponds to the expression of the two alleles and a single band to two RNAs with similar sizes, we compared the RNA sizes on Northern blot with the size of a fragment encompassing the genomic tandem repeat domain on Southern blot. It was possible to perform this study on the MUC2, MUC4, and MUC5B genes for which partial genomic organization is known. We selected restriction endonucleases with recognition sites on either side of the tandem repeat domains: Hinfl for MUC2 (55), EcoRI/PstI for MUC4 (56), and BglII for MUC5B (13). To evaluate the DNA polymorphism of the tandem repeat region, DNA digested with these four enzymes and from three out of the seven individuals were analyzed by electrophoresis and hybridized with the corresponding repetitive mucin probes: MUC2, MUC4, and MUC5B (Fig. 8A).

In the case of MUC4, four distinct alleles were identified for the three individuals tested, either at the DNA or at the RNA level. The allele sizes of the tandem repeat domain range from 9 to 19 kb, and the corresponding RNA sizes range from 16.5 to 24 kb. There is a notable similarity between Southern (Fig. 8A) and Northern blot (Fig. 8B) patterns; individual number 5, exhibiting an allele encompassing the lowest number of tan-

---

Fig. 6. Northern blot analysis of human mucin gene (MUC2 to MUC6) from mucosae of seven different individuals. Total RNA (10 μg) from bronchus (lane B), fundus (lane F), antrum (lanes A), and colon (lanes C) were prepared and transferred using our improved method. The blot was probed with the probes indicated. Autoradiography of the lane B hybridized with the MUC5B probe was performed for a shorter time.
level, and expression of alleles is co-dominant. Thus, the mucin gene polymorphism was reflected at the RNA level in individuals 3 and 4, where only one allele was present in the three individuals studied. The MUC5B gene had the distinction that for each person, the two patterns matched the two patterns. The apoB-100 gene, and the same similarity was observed between the two patterns. The greatest number of tandem repeats, had the largest transcript. Four genes containing divergent numbers of exact 39-base pair tandem repeats. On the contrary, other examples of polydisperse patterns have been attributed to artificial causes; messenger RNAs encoding human titin (23 kb) and dystrophin (14 kb) each show one band followed by a weak smear on Northern blot analyses. Initially, the presence of a polydisperse signal for mucin mRNAs was considered to be an artifact due to bad preparation of RNA, but rapidly this polydispersity became considered as an original and inherent feature of secreted mucins, because the hybridization of the same blots with a control probe (such as GAPDH or β-actin) showed a single band of the expected size, suggesting that the RNA was not degraded. Moreover, smears were also obtained using commercial mRNA preparations and Northern blots. The origin of this polydisperse pattern was unknown. Two kinds of hypotheses have been proposed in the literature to explain these polydisperse signals: those that propose a biological origin (such as rapid turnover, several related genes encoding similar transcripts varying in length, or alternative splicing); and those that suggest an artificial polydispersity due to specific properties of the mucin mRNAs (such as their unusual sensitivity to mechanical damage related to their large size or the occurrence of a particular conformation due to the large tandem repeat domains, which perturbs the migration of molecules in electric fields). In this work, different experiments were carried out to explore the hypothesis of a rapid turnover by studying the half-life of the mucin mRNAs in HT29-MTX cells treated with Act D with the aim of following the pathway of degradation.

**FIG. 7.** Graph indicating sizes of the hybridizing mRNA species. The standard curve was derived by using β-actin (2 kb, point 1), 28S rRNA (5 kb, point 2), apoB-100 (14.1 kb, point 3), and MUC5B (17.6 kb, point 4) as standards. The best formula corresponding to the nonlinear curve joining points 1–4 was as follows: Y = 31.05 exp(−0.273X), where Y represents size in kb and X represents migration distance in cm. This value was obtained using a Hewlett Packard 48GX calculator. A simpler method of drawing this curve is shown in this figure: a first straight line is obtained by joining points 1, 2, and 3. A second straight line is obtained by joining points 3 and 4. Point 5 (largest MUC4 mRNA size) is deduced from this second straight line.

**TABLE I**

Estimation of the sizes (kb) of the mucin mRNAs (MUC2 to MUC6)

| Genes   | 1 (bronchus) | 2 (fundus, antrum) | 3 (antrum) | 4 (colon) | 5 (colon) | 6 (colon) | 7 (colon) |
|---------|-------------|-------------------|------------|-----------|-----------|-----------|-----------|
| MUC2    | 16          | 16                | 16         | 16        | 15.5      |           |           |
| MUC3    | 14          | 15                | 17         | 17        |           |           |           |
| MUC4    | 17.5        | 18                | 18         | 18        |           |           |           |
| MUC5AC  | 17          | 17                | 18         |           |           |           |           |
| MUC5B   | 17.5        | 17                | 17.5       | 17.5      | 17.5      | 17.5      |           |
| MUC6    | 17.5        | 16.5              | 18         |           |           |           |           |

**DISCUSSION**

Since the initial observation in the case of the MUC2 gene, it has been a common characteristic of genes encoding secreted mucins (MUC2, MUC3, MUC4, MUC5AC, MUC5B, and MUC6) that the pattern observed by Northern blot analysis is in the form of a smear or very wide bands, suggesting heterogeneity of transcripts. In the literature, other examples of polydisperse patterns have been described that are of biological origin, for example the rat insulin-like growth factor I (57) and the apopolysialoglycoprotein of rainbow trout eggs (58). The insulin-like growth factor I gene is transcribed and processed in a complex manner to produce a large number of mature mRNAs (from 0.8 to 7.5 kb). These mRNAs result from multiple transcription start sites, the use of different polyadenylation sites, and alternative splicing. The multiple apopolysialoglycoprotein mRNAs (from 0.8 to 6 kb) are transcribed from multiple genes containing divergent numbers of exact 39-base pair tandem repeats. On the contrary, other examples of polydisperse patterns have been attributed to artificial causes; messenger RNAs encoding human titin (23 kb) and dystrophin (14 kb) each show one band followed by a weak smear on Northern blot analyses.

Promega was used, and the original protocol using selection by an oligo (dT)-cellulose column was also tested. Total RNA prepared using our improved method and corresponding poly(A)+ RNA was analyzed by Northern blot (Fig. 9). The yield of purification of MUC4 mRNA was estimated by comparing the intensity of the hybridization signals obtained with the β-actin probe. Whatever the protocol of purification, the yield of poly(A)+ selection was very low for mucin messengers when compared with smaller messengers such as the β-actin. Moreover, in some cases, total RNA showed two distinct bands, but in the corresponding poly(A)+ fraction, only the smallest transcript was selected.
Dissociation of the mRNAs and of estimating their half-life. Degradation of mucin transcripts proceeds by a slow and progressive decrease of all of the messenger RNAs, and the half-lives were all clearly long. Drugs such as Act D, which inhibit transcription, are commonly used to determine message half-life and this works well for messages with short half-lives. However, the long half-lives determined for the MUC genes (13 h for MUC2 and MUC3, 16 h for MUC5AC and MUC5B, and 17 h for MUC1) must be considered as cautious estimates, since prolonged treatment with Act D can inhibit metabolic processes other than transcription (62).

We also compared the half-life of a mucin gene exhibiting defined transcripts (MUC1) to the half-life of mucin genes showing polydisperse patterns of RNA and cultured the cells in the presence of Act D and cycloheximide. Since the mRNA decay showed similar kinetics in each case, we concluded that de novo synthesis of a regulatory protein was not involved in the generation of a polydisperse pattern of mucin messenger RNAs. These results indicate that in HT29-MTX cells the half-lives of the mRNAs encoding both membrane-associated and secreted mucins are very close to the average half-life of mRNA turnover, which is 10–20 h in eukaryotic cells (62). Thus, rapid turnover of mucin mRNAs could be ruled out as an explanation of the polydisperse pattern.

Isolated nuclear, polysomal, and cytoplasmic fractions exhibited the same polydisperse pattern of their mucin mRNAs, and it was not possible to discriminate, by size, the subpopulations of RNA that were translated, degraded, or immature. This result is in complete agreement with the study of the RNA turnover, i.e. the mucin transcripts are stable. We therefore decided to investigate an artifactual origin.

We chose to improve and adapt the original guanidinium isothiocyanate method for large RNA isolation. This technique is recognized for the purity and quality of the RNA obtained. The guanidinium isothiocyanate and 2-mercaptoethanol mixture combines the strong denaturing characteristic of guanidine with the chaotropic action of isothiocyanate and efficiently solubilizes tissue homogenates. Effective disruption can be obtained without the use of a homogenizer, which has otherwise been used routinely. Great care was taken to prevent mechanical degradation. Using this modified method, we were able to isolate large and intact mucin mRNAs from various human tissues (bronchus, fundus, antrum, and colon). Moreover, the yield of RNA signal was considerably better when, prior to transfer, the gel was soaked in NaOH. This treatment partially hydrolyzes the RNA and improves the efficiency of the transfer of large RNA and thus the detection of the signal. Hence, we demonstrated that the smear observed is due to partial degradation of very large mucin RNAs during purification and is exaggerated because of the better transfer of the products of degradation than the intact RNAs. We used the apoB-100 probe to confirm these results, because the apoB-100 gene is known to encode a large transcript, the cDNA being fully sequenced (14.1 kb), and it is expressed in the small intestine and

---

**Fig. 8. Comparison between the size of a fragment encompassing the genomic tandem repeat domain on Southern blot (A) and the mRNA sizes on Northern blot (B).** Genomic DNA from three individuals (numbers 5–7 in Fig. 6) was digested with the restriction enzymes EcoRI/PstI (for MUC4), HinfI (for MUC2), and BglII (for MUC5B). The Southern blots and Northern blots were hybridized with the probes indicated. The calculated sizes of the DNA fragments and the sizes of the corresponding RNA, hybridized with the repetitive mucin probe, are indicated by the arrows. The gel probed for MUC2 was electrophoresed for a longer time (24 h instead of 16 h in Fig. 6).

**Fig. 9. Selection of large mucin poly(A)+ RNA.** Total RNA (lane T) and poly(A)+ RNA (lane A) were prepared from human colon mucosa (individual 6 in Fig. 6) using our improved method and the PolyATtract mRNA Isolation system purchased from Promega. The blot was hybridized with the MUC4 probe. Autoradiography was performed for 2 days (A) and 10 days (B). RNA amounts were estimated by probing the same blot with a β-actin probe.
in the colon. The same kind of pattern was observed with the mucin and apoB-100 probes whatever the method used for RNA preparation. We concluded that degradation of RNA during purification is not a specific feature of mucin RNAs but occurs with other large messenger RNAs. We can therefore conclude that the demonstration of the integrity of the RNA preparation using a probe such as β-actin, GAPDH, or 28 S rRNA is misleading in the case of large RNAs, because these probes hybridize to short messages, for which the degradation problem is not encountered. Moreover, the use of standard RNA molecular size markers is not adapted to determine large sizes. We therefore recommend use of apoB-100 and MUC5B for these two purposes. Nevertheless, precise determination of sizes larger than 17.6 kb remains difficult because of the lack of RNA markers. Moreover, the use of large DNA fragments as size markers is inappropriate, because DNA and RNA molecules do not have the same behavior in denaturing gels. In the future, once it has been precisely sized after sequencing, it will also be possible to use the largest MUC4 allele of 24 kb in Fig. 6 and 7 (which is common) as a size standard (since it will be present in many of the samples and can be demonstrated by Southern blot analysis).4

We also studied the isolation of poly(A⁺) RNA, and we concluded that selection of poly(A⁺) is not recommended in the case of large mRNAs because of a very poor yield and because of additional risks of mechanical damage. Moreover, the preferential loss of the largest transcript observed in our experiment suggests that there is a risk of misinterpretation of the data.

To study the migration of large mRNAs that have large repetitive domains, we adapted a protocol using pulsed field gel electrophoresis for the separation of RNAs. With the aim of reducing mechanical damage of large RNAs, we attempted to carry out the electrophoresis directly from cells included in agarose plugs as performed for very large DNA fragments. Despite the fact that great amounts of RNasin were used, we did not succeed in separating large RNAs. Nevertheless, we were able to compare patterns on Northern blots performed with the same RNA preparation, using conventional electrophoresis and pulsed field gel electrophoresis. We observed that resolution of large mRNAs prepared according to our improved protocol was good using a contour-clamped homogenous electric field apparatus and electrophoretic conditions adapted to separation of molecules from 15 to 50 kb (68). However, the separation of similar size RNAs was not better (data not shown). Therefore, we recommend the use of conventional electrophoresis as described in this work.

In the case of MUC2, MUC4, and MUC5B, it was possible to compare the size of a fragment encompassing the genomic repetitive array with the RNA length, and we demonstrated that when size variations of transcripts among the individuals exist, they are directly related to the VNTR polymorphism that also exists at the DNA level. We speculate that the same observation will be made for MUC3, MUC5AC, and MUC6 genes, which also demonstrate variation in the size of transcripts in different individuals and thus putative VNTR variation at the level of the DNA but for which full restriction maps of the genes are not available. From this preliminary study, we can postulate that mucin genes exhibit a variable degree of VNTR polymorphism, the MUC4 gene appearing to be the most polymorphic (7.5 kb between the largest and the smallest allele), and on the contrary, the MUC5B gene seems not to be polymorphic. However, more samples will be required to confirm these results.

This work demonstrates clearly that mucin mRNAs are not polydisperse but are very large. This had been suggested only recently by Baeckström and colleagues (64). Human mucin transcripts are among the largest reported for eukaryotes. In the literature, two examples of proteins whose transcript sizes are larger than 20 kb have been described. An avian protein, the megadalan protein titin, is a single polypeptide chain of ~2800–3000 kDa that constitutes 8–10% of the myofibrillar proteins in striated muscle cells. Titin is encoded by a single RNA species of ~23 kb (59). The second protein is the human nebulin, a myofibrillar protein. Its full-length cDNA sequence is 20.7 kb long (65).

Our results are consistent with the message size necessary to synthesize large apomucin polypeptides. Recent studies have assessed the molecular mass of several mucin precursors immunoprecipitated with monoclonal antibodies and analyzed on reducing SDS-polyacrylamide gel electrophoresis (MUC2, MUC3, MUC4, MUC5AC, and MUC6 apomucins) (66). Mucins are synthesized as large N-glycosylated precursors (~500 kDa), which oligomerize and are converted by O-glycosylation into very large glycoproteins. The molecular masses found are consistent with the size of mucin mRNA estimated using our Northern blot analysis (Table I).

In Some Individuals, Mucin RNAs Are Detectable as a Double Band with Equal Intensity. We have shown that this finding reflects the expression of the two different alleles that could encode two mucin precursors with variable size that are probably synthesized in comparable amounts. The observation of co-expression of the two alleles is in accordance with other studies on mucin precursors. Evidence for co-expression of two alleles for rat gastric and colon apomucins with variation in molecular mass among animals has already been reported (67, 68).

Moreover, our data agree with the previous studies on MUC1. The MUC1 gene exhibits allelic variations due to the variation in the number of tandem repeats; the same polymorphism is demonstrated at the protein level, and the allele products are co-dominant (4).

In addition to the improvements in the methodology for isolation of intact large RNA usable for mucin mRNAs and all other large transcripts, this work provides an important step in understanding the role of mucin gene expression in mucin-associated pathologies. VNTR polymorphism results in the existence of different sizes of mucin core proteins bearing different numbers of potential O-glycosylation sites in different individuals. This work will now stimulate research into the biological implications of the length of the VNTR domain on mucin functional properties. Further studies will be required to demonstrate whether mucin gene polymorphism is implicated in the increase of susceptibility to any pathology, as proposed for the MUC1 gene in gastric cancer (69). Such studies are in progress in our laboratory.

Acknowledgments—We are grateful to Dr. Dallas Swallow for critical review of the manuscript. We thank Marie-Paule Delescaut for excellent technical assistance. We are grateful to Arnaud Porchet for analysis of the nonlinear curves.

REFERENCES
1. Gendler, S. J., and Spicer, A. P. (1995) Annu. Rev. Physiol. 57, 607–634
2. Shankar, V., Pichan, P., Eddy, R. L., Tonk, V., Nowak, N., Sait, S. N. J., Shows, T. B., Schultz, R. E., Gotway, G., Elkins, R. C., Gilmore, M. S., and Sachdev, G. P. (1997) Am. J. Respir. Cell Mol. Biol. 16, 232–241
3. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. N., and Wilson, D. (1999) J. Biol. Chem. 265, 15286–15293
4. Swallow, D. M., Gendler, S., Griffiths, B., Corney, G., Taylor-Papadimitriou, J., and Bramwell, E. (1987) Nature 328, 82–84
5. Bodek, I. A., Tsai, H., Biesbrock, A. R., and Levine, M. J. (1993) J. Biol. Chem. 268, 20563–20569
890 Human Mucin mRNAs

6. Gendler, S. J., Burchell, J., Duhig, T., Lampert, D., White, R., Parker, M., and "Taylors-Papadimitriou, J. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 5354–5358.
7. Litgenberg, M. J. L., Vos, H. L., Gennissen, A. M. C., and Hilken, J. (1990) J. Biol. Chem. 265, 5573–5578.
8. Bobek, L. A., Liu, J., Sait, S. N. J., Shows, T. B., Bobek, Y. A., and Levine, M. J. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 8316–8321.
9. Gum, J. R., Jr., Hicks, J. W., Toribara, N. W., Siddiki, B., and Kim, Y. S. (1994) J. Biol. Chem. 269, 2440–2446.
10. Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lampert, D. T. A., and Kim, Y. S. (1989) J. Biol. Chem. 264, 6480–6487.
11. Janey, B. H., Gallup, M. W., Yan, P. S., Gum, J. R., Jr., Kim, Y. S., and Basbaum, C. B. (1991) J. Clin. Invest. 87, 77–82.
12. Toribara, N. W., Robertson, A. M., Ho, S. B., Kuo, W. L., Gum, E., Jr., Byrd, J. C., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B., and Kim, Y. S. (1993) J. Biol. Chem. 268, 5879–5885.
13. Desseyn, J.-L., Desseyn, J. L., Aubert, J.-P., and Laine, A. (1997) J. Biol. Chem. 272, 3168–3178.
14. Dufossé, J., Porchet, A., Audie, J. P., Guyonnet Dufossé, V., Laine, A., Van-Seuningen, I., Degand, P., and Aubert, J. P. (1993) Biochem. J. 293, 329–337.
15. Desseyn, J.-L., Aubert, J.-P., Van Seuningen, I., Porchet, N., and Laine, A. (1997) J. Biol. Chem. 272, 16873–16883.
16. Porchet, N., Nguyen, V. C., Dufossé, J., Audie, J. P., Guyonnet Dufossé, V., Gross, M. C., Denis, C., Degand, P., Bernheim, A., and Aubert, J. P. (1991) Biochem. Biophys. Res. Commun. 175, 414–422.
17. Crépin, M., Porchet, N., Aubert, J. P., and Degand, P. (1996) Biochemistry 27, 471–484.
18. Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. L., Byrd, J. C., Lampert, D. T. A., Siddiki, B., and Kim, Y. S. (1990) Biochem. Biophys. Res. Commun. 171, 407–415.
19. Lesuffleur, T., Roche, F., Hill, A. S., Lacasa, M., Fox, M., Swallow, D. M., Zweibaum, A., and Real, F. X. (1996) J. Biol. Chem. 271, 13665–13673.
20. Mezzena, D., Charles, P., Desaki, E., Polyemoupolou, M. H., Martin, B. M., and Rose, M. C. (1994) J. Biol. Chem. 269, 12923–12932.
21. Klop, L. W. J., Van Rens, L., and Streus, G. J. (1995) Biochem. J. 308, 99–108.
22. Balague, C., Gambus, G., Carrato, C., Porchet, N., Aubert, J. P., Kim, Y. S., and Real, F. X. (1994) J. Cell Biol. 127, 1054–1061.
23. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Clin. Invest. 92, 7300–7305.
24. Sanyal, H., Emori, Y., Kawasaki, H., Kitajima, K., Inoue, S., Suzuki, K., and Inoue, Y. (1988) J. Biol. Chem. 263, 17678–17684.
25. Sanyal, H., Sater, G. R., Myers, A. M., Robson, R. M., and Huiatt, T. W. (1993) J. Biol. Chem. 268, 10431–10437.
26. Foyt, H. L., Leroith, D., and Roberts, C. T., Jr. (1991) J. Biol. Chem. 266, 7300–7305.
27. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
28. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
29. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
30. Foyt, H. L., Leroith, D., and Roberts, C. T., Jr. (1991) J. Biol. Chem. 266, 7300–7305.
31. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
32. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
33. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
34. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
35. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.
36. Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., and Kim, Y. S. (1993) J. Cell Biol. 127, 1054–1061.