ELSEVIER

Biochimica et Biophysica Acta 1326 (1997) 7–11

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Short sequence-paper

The carboxyl-terminal sequence of rat intestinal mucin RMuc3 contains a putative transmembrane region and two EGF-like motifs [1]

Ismat A. Khatri [a], Gordon G. Forstner [b], Janet F. Forstner [a,*]

[a] Division of Research Biochemistry, Research Institute, The Hospital for Sick Children and the University of Toronto, Toronto, Canada
[b] Division of Gastroenterology, Research Institute, The Hospital for Sick Children and the University of Toronto, Toronto, Canada

Received 27 January 1997; revised 11 March 1997; accepted 12 March 1997

Abstract

A 3′ RACE technique was used to establish the nucleotide sequence encoding the C-terminal 379 amino acids of rat intestinal Muc3. Unlike the C-terminus of Muc2 and many secretory mucins, Muc3 contains two EGF motifs and a putative transmembrane domain. The mRNA for rat Muc3 is 7.5–8.0 kb.

Keywords: Mucin; Intestine; EGF; Carboxyl terminus; Membrane domain; (Rat)

Reported sequences of cDNAs for mucus glycoproteins (mucins) are rarely full-length. This is in part due to the extraordinarily large size of mucin core peptides, and their long central stretch of tandem repeats (TRs), which make sequencing difficult. TRs are specific for each mucin, but even for the same organ, they show little interspecies conservation. Amino- and carboxyl-end regions tend to show more interspecies homology. For example, secretory mucins such as human MUC2 [1], rat Muc 2 [2], porcine submaxillary mucin [3], and frog integumentary mucin Fim-B.1 [4], have a similar distribution of cysteine residues at their C-terminal ends. In two cases, human MUC1 [5] and rat mammary sialomucin $ASGP_2$ [6], the C-terminal ends contain a putative transmembrane domain and the mucins have been shown biochemically to be membrane-associated.

The amino- and carboxyl-termini of rat Muc2 (previously called MLP) have now been sequenced [2,7,8]. There is a high degree of structural and sequence homology of rat Muc2 with the human intestinal mucin MUC2 at its N- and C-terminal ends.

Another rat intestinal cDNA clone called RMUC 176 was reported by Gum et al. [9], and it revealed a tandem repeat structure (consensus TTTPDV) and a unique cysteine-containing 92 amino acid sequence at its C-terminal end. A cDNA clone called M2-798 was also reported by our laboratory [10] and was found to encode the same consensus TR sequence, but a different (not cysteine-enriched) unique sequence of 82 amino acids at its C-terminal end. Both rat clones specify different regions of the same gene. Initially we assumed that this gene might be rat Muc2, but Northern blots using probes for the TRs and unique regions of clone M2-798 both hybridized to a transcript (7.5–8.0 kb) that was smaller than the

[*] Corresponding author: The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8. Fax: +1 416 8135022; E-mail: jfforst@sickkids.on.ca

[1] The nucleotide sequence data reported in this paper have been submitted to the GenBank/EMBL Data Bank under the accession number U76551.

PII S0005-2736(97)00063-1

mRNA for rat Muc2 (> 9.0 kb). We therefore suggested that a second rat intestinal mucin gene existed, and called it M2 [10].

Comparative chromosomal mapping studies [11] involving human and rodent DNA samples have revealed gene cluster homologies between rat Muc1, Muc2 and the mucin we called M2, with human MUC1, MUC2 and MUC3, respectively. This is based on the conservation of gene synteny for three different rat and human chromosomes. The name M2 is therefore now changed to rat Muc3.

Not surprisingly, there is no homology of the central tandem repeat regions of human MUC3 and rat Muc3 [9]. The C-terminal sequence of human MUC3 has been completed [12] but is not yet published, and the comparable region of rat Muc3 has not been reported. Therefore to provide data with which to judge possible sequence homologies between the two mucins, and to initiate functional studies, rat Muc3 sequencing experiments have been conducted and are the subject of this report.

Total RNA from rat small intestine was prepared [13], and first strand cDNA synthesis accomplished by reverse transcription (RT). The reaction was primed with an Adapter Primer (AP) and catalyzed by the Superscript II RNase H-reverse transcriptase (both from GIBCO BRL). The target cDNA was amplified by PCR using two different gene specific sense primers: primer S1 corresponded to nt 1068–1089 at the 5′ end of the unique (non tandem repeat) region of cDNA clone RMUC176 [9], and a nested primer, S2, corresponded to nt 1320–1341 at the 3′ end of the same cDNA. A dUMP-containing sequence was added to the 5′ ends of primers S1 and S2 to facilitate later cloning. The antisense primer UAP (BRL) also contained a dUMP sequence at the 5′ end. PCR utilized *Taq* polymerase (Perkin Elmer) and consisted of denaturation at 94°C for 2 min (1 cycle) followed by 30 cycles of denaturation at 94°C for 1 min, annealing and extension at 65°C for 1 min, with a final extension at 72°C for 5 min. PCR products were separated on a 1.5% agarose gel, and stained with ethidium bromide. The 3′ RACE product using primer S1 was 1.7 kb, while primer S2 (the nested primer) as expected, gave a product 300 bp shorter. The specific bands were excised from the gel and purified by genecleaning (GenClean II kit, BIO 101, Vista, CA).

The amplification products were cloned into vector pAMP1 (CLONEAMP pAMP1 system, GIBCO BRL, Gaithersburg, MD) and transformants confirmed by DNA sequencing. The 1.7 kb product was sequenced in both directions using SP6, T7 and subsequent overlapping primers (Biotechnology Services, the Hospital for Sick Children, Toronto). The sequence (Fig. 1) is 1690 bp long, with the first 273 bp in complete agreement with the corresponding sequence (nt 1068–1341) of clone RMUC176 [9]. The nested 3′ RACE product as expected, was shorter (1470 bp) and matched 100% with the sequence of the 3′ RACE

```
   1  CTGAACGGAGGGTACTGGTCAGGTGCGATGTGCGTCTGCCCCAATGGCTTCTCAGGGGAT
      L  N  G  G  Y  W  S  G  A  M  C  V  C  P  N  G  F  S  G  D     20

  61  CGCTGTCAGAACAGAGTCCCTGTGGTCGACTGCCAGAACGGAGGCACGTGGGATGGGCTC
      R  C  Q  N  R  V  P  V  V  D  C  Q  N  G  G  T  W  D  G  L     40

 121  AAATGTCAGTGTACCGGCCTCTTCTATGGCCCAAGGTGTGAGGAAGTGATGGAGAGCGTA
      K  C  Q  C  T  G  L  F  Y  G  P  R  C  E  E  V  M  E  S  V     60

 181  GAGATAAAGCCGACGGTCTCTGCATCTGTGGAAGTGAGTGTGACAGTAACCAGTCAAGAA
      E  I  K  P  T  V  S  A  S  V  E  V  S  V  T  V  T  S  Q  E     80

 241  TATAGTAATGAGCTACAGGACCGAAACTCCACAGAATTCCGTAACTTCAATGAGACATTC
      Y  S  N  E  L  Q  D  R  N  S  T  E  F  R  N  F  N  E  T  F    100

 301  ACAAAGCAGATGGCGATAATTTATGCTGGTATCCCGGAGTATGAAGGAGTTATCATCAAA
      T  K  Q  M  A  I  I  Y  A  G  I  P  E  Y  E  G  V  I  I  K    120

 361  AATCTGAGCAAAGGCAGTATCGTGGTGGATTATGATGTCATCCTGAAGGCCCAGTACACC
      N  L  S  K  G  S  I  V  V  D  Y  D  V  I  L  K  A  Q  Y  T    140

 421  CCAGGATTCGACAACACCTTAGATAACATCGTCAGTAACCTGGAGACAAAAATCAAGAAT
      P  G  F  D  N  T  L  D  N  I  V  S  N  L  E  T  K  I  K  N    160

 481  GCGACAACAGTTCAAGTACAAGATGCCAATAATACCTGTTCAGCCTTACTGTGTTTCAAC
      A  T  T  V  Q  V  Q  D  A  N  N  T  C  S  A  L  L  C  F  N    180

 541  TCGACTGCCACCCGGGTGCAAACCAACGTTACAGTAGTCAGTGACAATCTTGAGGAGGTG
      S  T  A  T  R  V  Q  T  N  V  T  V  V  S  D  N  L  E  E  V    200

 601  TGCAAGAAAGAGGCTGGAGAGGACTTTGCAAAGTATGTCACCCTGGGGCTCAAGGACAAT
       C  K  K  E  A  G  E  D  F  A  K  Y  V  T  L  G  L  K  D  N    220

 661  AAGTGGTACTGTGTCACGCCTTGTTCATCAGGCTACAGTACCTCCAAGAACTGCAGCTAT
      K  W  Y  C  V  T  P  C  S  S  G  Y  S  T  S  K  N  C  S  Y    240

 721  GGCAAATGTCAGTTGCAGCGAAGTGGACCCCGGTGCCTCTGCTTGTCCACGGATACTCAC
      G  K  C  Q  L  Q  R  S  G  P  R  C  L  C  L  S  T  D  T  H    260

 781  TGGTACAGCGGTGAAAACTGCGACTGGGGCACCCAGAAAAGTCTGGTGTATGGGCTCGTG
      W  Y  S  G  E  N  C  D  W  G  T  Q  K  S  L  V  Y  G  L  V    280

 841  GGAGCCGGGGTGGCAGTGCTGTTGGTGATTCTTGTGATCCTCGCTGTGTTCTCCGTCCAC
      G  A  G  V  A  V  L  L  V  I  L  V  I  L  A  V  F  S  V  H    300

 901  TACAGAAGAGATGCACAAAGGCAAGAGTCCAGAGTGTCTGAGATGTACAGGTGGGGTGAA
      Y  R  R  D  A  Q  R  Q  E  S  R  V  S  E  M  Y  R  W  G  E    320

 961  GAGGCAGGCAGAGCATCTCCCGGAACCTTCCACAACATTGGCTTTGAGCACAACGAAGAA
      E  A  G  R  A  S  P  G  T  F  H  N  I  G  F  E  H  N  E  E    340

1021  CAAGAAAACGATATTTCCTTGGACTCTGTGTACAGCAACTTCCAGCCCTCCCTAAGCCAC
      Q  E  N  D  I  S  L  D  S  V  Y  S  N  F  Q  P  S  L  S  H    360

1081  ATAAACCCTGAAGGAAAGATCCAGATCCAGAGACCCCAGGTGGCCATGACACCGTTGTAA
      I  N  P  E  G  K  I  Q  I  Q  R  P  Q  V  A  M  T  P  L  *    380

1141  ATACTAACACAAAAACGGTTTTGCACGGATCTGTGGAGCTGAAGACCGAGAAGTGTGTTT
1201  GACCAAGCATGGCGGAACACACCCTACGTTGAGATTTCCCCATGGAGACTTATTGGGAAA
1261  TAGACCCTGCAGTTAGCAAGAATGGATACGAGAGAGACGATGCTGGAAGCTCAGACAGTA
1321  TGGAAACTGCAGTGGCCATGTTATGACGGCAAGCGATGATCCCTACCGGAACCTCATCCC
1381  CAGAGATGCTCTCCTTCCCGTGGGCCCACATTCAGCACCAAGGACAGTGCCCACCTGCTG
1441  CCATTGCCATTGGTATAACATTAGATTGCATTTTACAAAACACACTTCTAGGAAAGCTTA
1501  CCTCCCATCTCCATGACAAGCACTGTTTTTAACACAGACTATCCATCTTTTTAAGGTCAT
1561  ACTTAGTTTTAAGTTTCTCTTTAACATATTAAAGCTTTTATTTTGTTGAATTCTAGTTTA
1621  CCTTATGTAAATAAAAGTATATGCTTATTTTCTAAAAAAAAAAAAAAAAAAAAAAAAAAA
1681  AAAAAAAAAA
```

Fig. 1. Nucleotide and deduced amino acid sequence of the 1690 bp 3′ RACE product. Nucleotides are numbered on the left, amino acids on the right. The eight predicted N-glycosylation sites are marked in bold, and the termination codon is marked with an asterisk. The putative transmembrane region is underlined. The polyadenylation signal AATAAA is in bold italics.

product at nt position 253–1690. There was a single open reading frame of 379 amino acids followed by an untranslated region of 550 nt, which includes the polyadenylation signal AATAAA, followed closely by a 37 nt polyA tail.

Unlike the tandem repeat region [9,10], the deduced amino acid composition of the C-terminal 379 residues consists of a relatively low content of serine plus threonine (15.5 mol%) and a much higher content of hydrophobic residues (31.8 mol% for combined leu, ile, met, val, phe, trp, and tyr). Cysteine comprises 4.49 mol%. Similarity searches of the Genbank (BLASTP or BLAST program [14]) did not reveal sequence homology (nucleic acid or amino acid) with any other published mucin sequences, including those for rat Muc2 and human MUC3, or other proteins.

Using the Wisconsin Sequence Analysis Package (Genetic Computer Group, Inc. Madison, Wisconsin), two EGF (epithelial growth factor) motifs were identified (residues 13–53 and 228–267), separated by 174 residues. As in many other EGF-containing proteins, the two EGF motifs each contain six cysteines. The well recognized EGF consensus alignment of cysteines is not perfectly conserved in the first EGF-like domain (residues 13–53), which implies that this domain may not be functional. However the second EGF motif (residues 228–267) has the same cysteine and glycine distribution that is particularly well conserved amongst functional EGF-bearing proteins (i.e. C-X-X-X-G-F/Y-X-G-X-X-C). The existence of EGF motifs in human intestinal mucin MUC3 has been reported in abstract form [12], but the only other mucin in which EGF-like sequences (EGF1 and EGF2) have been published is the ASGP2 membrane component of the cell surface sialomucin complex expressed on 13762 rat ascites mammary adenocarcinoma cells [6]. ASGP2 has been shown to activate the EGF receptor kinase from A431 cell membranes and to compete with EGF for binding [15]. The authors postulate that ASGP2 may act as a transmembrane growth factor. A potentially interesting observation is that the alignment of the second EGF motif of rat Muc3 is very close to the EGF2 of rat ASGP2.

Another interesting feature is that the region 200–267 contains a total of eight cysteines, as do both EGF-like domains in ASGP2. Residues 11–53, a region which includes the first EGF-like region of rat Muc3, shows seven cysteines, but there is another cysteine immediately N-terminal to the leucine located at position 1 (residue 356 of RMUC176 [9]). It may be reasonable therefore to postulate that the cysteine-rich regions in these two mucins adopt a four-loop structure rather than a typical three-loop EGF domain structure.

The rat Muc3 sequence reveals eight potential N-glycosylation sites (Fig. 1), two potential casein kinase (CK) II sites (residues 77 and 152) and two potential protein kinase C (PKC) phosphorylation sites (residues 234 and 271). Most of these modifications reside in the first 267 amino acids of the new C-terminal region. Although PKC has been shown to be involved in the regulation of mucin secretion from human colonic carcinoma cells [16], the PKC (and CK II) sites in the C-terminal sequence of rat Muc3 are unlikely to be functional, since they are in the putative extracellular domain.

Kyte and Doolittle [17] hydropathy analysis of the deduced 379 residue sequence (Fig. 2) reveals a putative transmembrane domain of 24 amino acids (residues 276–299). Between the transmembrane region and the stop codon at the C-terminus is a hydrophilic stretch (putative cytoplasmic tail) of 80 amino acids. These features suggest that rat Muc3, like the mucins human MUC1 [5,18] and rat ASGP2 [6], may be a membrane-associated mucin. The 8 amino acid distance from the second EGF-like do-

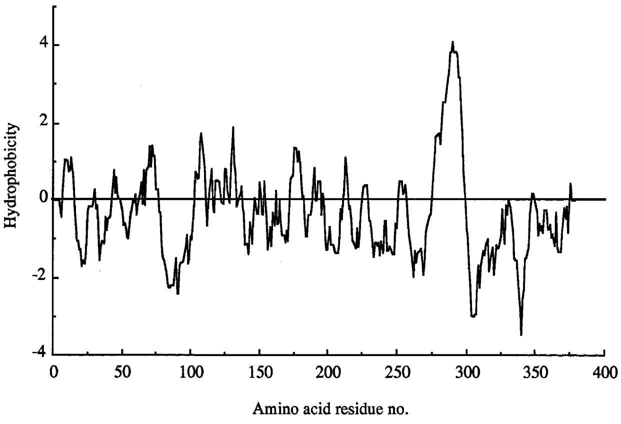

Fig. 2. Hydrophobicity plot of the 379 amino acid sequence of Fig. 1. The plot was derived from the algorithm of Kyte and Doolittle [17] using a window of 9 amino acids. Residues 276 to 299 represent the putative transmembrane domain.

main to the putative transmembrane domain sequence is essentially the same as that in ASGP2. This relationship between rat Muc3 and ASGP2 also strengthens the argument for rat Muc3 as a new membrane mucin. Human MUC3 is said to lack a putative transmembrane region [12].

Northern blot hybridization analyses were carried out on rat intestinal and colonic RNA [13,19] and a multiple tissue Northern blot (CLONTECH) containing RNA from rat heart, brain, spleen, lung, liver, skeletal muscle, kidney and testis. Probe A was a 1-kb cDNA fragment (clone M2-1000) described earlier [10], which consists entirely of tandem repeats of rat Muc3. Probe B consisted of the 3′ 1220 bp *ECo*R1, *Hin*dIII fragment (nt 275–1494) of the 1.7 kb 3′ RACE product presented in Fig. 1. A 1.8-kb human β-actin cDNA fragment (CLONTECH Laboratories Inc., Palo Alto, CA) was used as a control probe and confirmed that equivalent amounts of each RNA preparation were studied (not presented). Probes were labelled with [α-^{32}P]dCTP using the T7 Quick Prime kit (Amersham). Hybridizations were performed at 42°C for 18 h with 50% formamide, 5 × SSC, 2% blocking solution (Boehringer Manheim GmBH, Germany), 0.1% *N*-lauryl sarcosine, and 0.02% SDS. Blots were washed twice at 65°C with 2 × SSC and 0.1% SDS for 15 min, exposed to X-ray film and stored at −70°C. Exposure times were 2–3 h for probe A and 2 to 3 days for probe B. Both probes hybridized to rat intestinal mRNA at a position of 7.5–8.0 kb (Fig. 3). Another band appeared at ~ 4 kb, just below the 28S rRNA position. Although this band may be nonspecific, the possibility of the existence of two allelic forms of rat Muc3 cannot be ruled out. The only other tissue RNA to give a signal (7.5–8.0 kb) was rat colon (not presented). Preliminary tests with RNA of human colonic LS174T and CaCo-2 cells were negative, which probably reflects the known lack of human MUC3 in LS174T cells, and a very low level of MUC3 in CaCo-2 cells [20–23]. Further cross-hybridization assays and sequence comparisons will be necessary to judge homology of the human and rat genes.

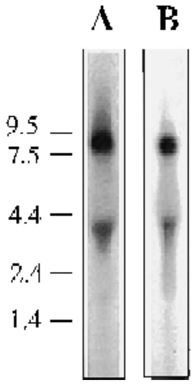

Fig. 3. Northern blots of rat intestinal RNA. Probe A is a tandem repeat probe for rat Muc3, and probe B is the 3′ RACE product of the present study.

In summary, we have determined the C-terminal sequence of rat Muc3. Unexpectedly, this intestinal mucin appears to have features that are more characteristic of a membrane-associated mucin than a true secretory mucin. The sequence is distinctly different from the C-terminus of the other known rat intestinal mucin, rat Muc2.

Financial support was provided from the MRC of Canada. The Hospital for Sick Children, Pharmacia and University of Toronto Biotechnology Centre performed oligonucleotide synthesis and DNA sequencing.

References

[1] J.R. Gum, J.W. Hicks, N.W. Toribara, E.-M. Rothe, R.E. Lagace, Y.S. Kim, J. Biol. Chem. 267 (1992) 21375–21383.

[2] G. Xu, L.J. Huan, I.A. Khatri, D. Wang, A. Bennick, R.E.F. Fahim, G.G. Forstner, J.F. Forstner, J. Biol. Chem. 267 (1992) 5401–5407.

[3] A.E. Eckhardt, C.S. Timpte, J.L. Abernethy, Y. Zhao, R.L. Hill, J. Biol. Chem. 266 (1991) 9678–9686.

[4] J.C. Probst, E.M. Gertzen, W. Hoffmann, Biochemistry 29 (1990) 6240–6244.

[5] S.J. Gendler, A.P. Spicer, E.N. Lalani et al., Am. Rev. Respir. Dis. 144 (1991) S42–S47.

[6] Z. Sheng, R. Wu, K.L. Carraway, N. Fregien, J. Biol. Chem. 267 (1992) 16341–16346.

[7] H. Ohmori, A.F. Dohrman, M. Gallup, T. Tsuda, H. Kai, J.R.J. Gum, Y.S. Kim, C.B. Basbaum, J. Biol. Chem. 269 (1994) 17833–17840.

[8] G.C. Hansson, D. Baeckstrom, I. Carlstedt, K. Klinga-Levan, Biochem. Biophys. Res. Commun. 198 (1994) 181–190.

[9] J.R. Gum, J.W. Hicks, R.E. Lagace, J.C. Byrd, N.W. Toribara, B. Siddiki, F.J. Fearney, D.T.A. Lamport, Y.S. Kim, J. Biol. Chem. 266 (1991) 22733–22738.

[10] I.A. Khatri, G.G. Forstner, J.F. Forstner, Biochem. J. 294 (1993) 391–399.

[11] K. Klinga-Levan, J.R. Gum, S.J. Gendler, Y. Kim, G.C. Hansson, Mamm. Genome 7 (1996) 248–250.

[12] Y.S. Kim, J.W. Hicks, J.J.L. Ho, D. Swallow, J.R. Gum, 3rd International Workshop on Carcinoma-Associated Mucins, Imperial Cancer Research Fund, 1994, A15.

[13] J.H. Han, C. Stratowa, W.J. Rutler, Biochemistry 26 (1987) 1617–1625.

[14] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403–410.

[15] K.L. Carraway, K.L.I. Carraway, R.A. Cerione, C.A.C. Carraway, FASEB J. 6 (1992) A47.

[16] G. Forstner, Annu. Rev. Physiol. 57 (1995) 585–605.

[17] J. Kyte, R.F. Doolittle, J. Mol. Biol. 157 (1982) 105–132.

[18] S.J. Gendler, C.A. Lancaster, J. Taylor-Papadimitriou, J. Biol. Chem. 265 (1990) 15286–15293.

[19] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning, a Laboratory Manual, Vol. 1, 2nd Edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[20] Y. Niv, J.C. Byrd, S.B. Ho, R. Dahiya, Y.S. Kim, Int. J. Cancer 50 (1992) 147–152.

[21] P.L. Devine, M.A. McGuckin, G.W. Birrell, R.H. Whitehead, G.P. Sachdev, P. Shield, B.G. Ward, Br. J. Cancer 67 (1993) 1182–1188.

[22] D.J. McCool, J.F. Forstner, G.G. Forstner, Biochem J. 302 (1994) 111–118.

[23] B.J.W. Van Klinken, E. Oussoren, J. Weenik, H.A. Buller, J. Dekker, A.W.C. Einerhand, Biochem. Soc. Trans. 23 (1995) 529S.