Molecular cloning and characterization of the gene encoding rat submandibular gland apomucin, Mucsmg*

EARL F. ALBONE1, FRED K. HAGEN1, CLAUDE SZPIRER2 and LAWRENCE A. TABAK1

1Departments of Dental Research and Biochemistry, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Ave., Box 611, Rochester, NY 14642, USA
2Laboratoire de Génétique, Université Libre de Bruxelles, Brussels, Belgium

Received 15 August 1995, revised 4 November 1995

Mucin glycoproteins are a major constituent of salivary secretions and play a primary role in the protection of the oral cavity. Rat submandibular glands (RSMG) synthesize and secrete a low molecular weight (114 kDa) mucin glycoprotein. We have isolated, partially sequenced, and characterized the gene which encodes the RSMG apomucin. The gene is encoded by three exons of 106 nt, 69 nt, and 991 nt, separated by introns of 921 nt and 12.5 kb. CAAT and TATA elements are present, at -68 and -26, respectively, in the 5' flanking sequence of the RSMG apomucin gene. The tandem repeat domain present in exon III consists of ten tandem repeats of 39 nt encoding the consensus sequence PTTDSTTPAPTTK. Sequence comparison and organization of the nucleic acid sequence encoding the tandem repeats of two alleles for this gene suggests that the apomucin gene has undergone recombinational events during its evolution. No significant sequence similarity was found with other mucin genes, or with other known salivary gland-specific genes. The gene was localized to rat chromosome 14 using somatic cell hybrids that segregate rat chromosomes. Since this, to our knowledge, represents the first RSMG mucin gene cloned, we have designated this gene Mucsmg.

Keywords: mucins, O-glycosylation, gene-expression

Abbreviations: RSMG, rat submandibular gland; RSM, rat salivary mucin; GRP, glutamine-glutamic-acid rich protein; nt, nucleotide; kb, kilobase

Introduction

Mucin-glycoproteins (mucins) are a principal organic constituent of the mucus secretions which coat the gastrointestinal, respiratory, and urogenital tracts. This slimy, viscoelastic coat aids in the protection of these exposed epithelial surfaces from microbial and physical insult [1]. Previous studies have shown that the mucin polypeptide backbone (apomucin) usually consists of tandem arrays of repeating amino acid sequence rich in threonine, serine, and proline, to which are attached numerous O-linked oligosaccharides. These O-linked side chains may constitute as much as 80% of the molecular mass of the molecule. Salivary apomucins vary greatly in size, with two general classes being identified. 'High' molecular weight salivary mucins are characterized by the presence of a cysteine-rich domain which forms multimeric complexes. In contrast, 'low' molecular weight salivary mucins lack this cysteine-rich domain and remain monomeric.

Rat submandibular glands (RSMG) secrete a low molecular weight (114 kDa) mucin which forms a major component of rat saliva. Conceptual translation of cDNAs encoding this apomucin [2] revealed three distinct regions; a basic N-terminus rich in the amino acids glutamine, proline, and tyrosine, but lacking in hydroxyamino acids, a threonine- and proline-rich tandem repeat segment (PTTDSTTPAPTTK)_{10-11} which showed allelic polymorphism in tandem repeat number, as has been seen for other mucins [3, 4], and a serine-, threonine-rich C-terminus. Although there is no signifi-
cant sequence similarity between the RSMG apomucin and human salivary mucin MUC7 [5], these mucins share a similar architecture. Therefore, it appears that RSMG mucin represents an analogue to the human salivary mucin, MUC7 [5], which is thought to promote the clearance of bacteria from the oral cavity [1].

Although cDNA cloning has revealed valuable information concerning the primary structure of mucins from a variety of sources, only the genes for the membrane-bound human tumour-associated epithelial mucin MUC1, and the mouse homologue, Muc 1 have been cloned in their entirety [3,6]. Partial gene structures for the human MUC2 [7] and the canine tracheo-bronchial mucin (CTM) [8] are also available. The promoter for MUC1 has been characterized, and elements which govern its tissue-specific expression have been identified [9–11]. Previously, by Southern blot analysis of rat genomic DNA, we estimated that the RSMG apomucin was encoded by a single copy gene [2]. In the present study we have isolated and characterized a gene, termed Mucsmg, which encodes the RSMG apomucin.

**Materials and methods**

**Isolation of genomic clones of Mucsmg gene**

To isolate genomic clones corresponding to RSMG apomucin, two screenings of a λDASH II rat genomic library (Stratagene) were performed. The library was plated out at a density of 5–8 × 10⁶ phage per 150 mm plate using LE392 as host (Stratagene) (10⁶ phage total). The plaques were lifted onto nitrocellulose filters (Schleicher and Schuell), denatured in 0.5 M NaOH, 1.5 M NaCl, neutralized in 1 M Tris (pH 8.0), 1.5 M NaCl, rinsed briefly in 2 × SSC, and baked for 2 h in vacuo. Screening was performed using the insert of cDNA clone pRSM-3, available from a previous study [2], which corresponds to position 123–1203 of the full length cDNA clone. Hybridization and washing were performed as previously described [2]. Four rounds of screening were performed, and two clones were isolated. Phage DNA was prepared by the method of Chisholm [12], digested with various restriction enzymes, and subjected to Southern blot analysis [13]. Results indicated that the clone gRSM-1 contained a single exon (exon III), but lacked the remaining 5' exonic sequence. Therefore, the library was replated at a density of 10⁵ phage per 230 mm × 230 mm plate (10⁶ phage total). Plaques were lifted onto 23 cm × 23 cm nylon filters (Schleicher and Schuell) and denatured, neutralized, and baked as described above. A 5' cDNA probe was generated by polymerase chain reaction (PCR) using the following primers: EA-1, TCTCTTCTCGAATTCTAACCGTAGC; PE-2, CGTAAA ATATGAAAGAAGAGCCAACAGG, whose 5' ends correspond to positions 1 and 174 of the published cDNA sequence [2]. Amplification was done by denaturation for 3 min at 94 °C, followed by 30 cycles of 94 °C, 15 s; 51 °C, 15 s; 72 °C, 30 s; followed lastly by a 72 °C

<table>
<thead>
<tr>
<th>Splice Acceptor (Y)₇₋₁₋₁NYAG/G</th>
<th>Exon consensus</th>
<th>Exon Exon Splice Donor Intron</th>
<th>(bp) Position</th>
<th>MAG/GTRAGT</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCTTCTTTCAG/G</td>
<td>I</td>
<td>106</td>
<td>+1</td>
<td>AAG/GTGAGT</td>
<td>921</td>
</tr>
<tr>
<td>TTTTTCTTTTTATTTCCACAG/C</td>
<td>II</td>
<td>69</td>
<td>+1028</td>
<td>ACG/GTAAGT</td>
<td>≈12,500</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>991</td>
<td>+14124</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** Overlap of genomic clones encoding the RSMG apomucin gene, Mucsmg. Exons are indicated by the solid rectangles. The repeat domain in exon III is indicated by an array of open rectangles. Eco RI sites are indicated by an 'R'. The sizes and positions of the exons (in nt) and introns are indicated in Table 1. Consensus and actual splice donor and acceptor sequences are also given. Invariant sequences at the splice junction are underlined.