

Characterization of a rat airway cDNA encoding a mucin-like protein

Tohru Tsuda,^{#+} Marianne Gallup,[#] Berthold Jany,^{#++} James
Gum,^{#*} Young Kim,^{&*} and Carol Basbaum,[#]

Departments of [#]Anatomy and [&]Medicine, University of California San Francisco,
and ^{*}Gastro-intestinal Research Unit, Veterans Administration Medical Center,
San Francisco, California 94143

Received July 20, 1993

Summary: The purpose of this study was to isolate airway mucin cDNAs for use in studies of mucin biosynthesis in rat models of human airway disease. To this end, we screened a rat airway cDNA library with the human intestinal mucin cDNA SMUC 41 and obtained 7 positive clones. Preliminary characterization of each of these led us to focus on the clone expressing the 390 bp cDNA RAM 7s. Evidence indicating that RAM 7s encodes part of a rat airway mucin gene is that RAM 7s: (a) hybridizes in plaque lifts to SMUC 41, (b) hybridizes in Northern blots to large, polydisperse transcripts, (c) has a sequence encoding threonine-rich tandem repeats and (d) shows appropriate tissue-specific expression of cognate mRNA. The repetitive peptide encoded by RAM 7s includes five copies of the consensus sequence TTTTITL. Because this sequence is different from those reported for two cDNAs previously isolated from rat intestinal libraries, we tentatively conclude that RAM 7s encodes part of a previously unidentified rat mucin gene. © 1993 Academic Press, Inc.

Mucin is a large glycoprotein (>10⁶ daltons) (1) consisting of ~80% carbohydrate and ~20% protein (2). Secreted by epithelial cells of the respiratory, gastrointestinal and urogenital tracts, mucin monomers polymerize extracellularly to form mucus gels. In the respiratory tract, air-borne impurities are trapped in the mucus, which is cleared from the airways by the action of epithelial cilia. This mucociliary clearance mechanism acts to purify the inspired air before it reaches the lungs.

⁺Present address: Division of Respiratory Disease, School of Medicine,
University of Occupational and Environmental Health, Kitakyushu 807, Japan.

⁺⁺Present address: Medizin Poliklinik der Universitat, Klinikstrasse 6-8,
D-8700 Wurzburg, Germany.

Although mucus acts protectively in healthy individuals, it can contribute to considerable morbidity and mortality when overproduced, such as in individuals with cystic fibrosis, chronic bronchitis and asthma. In these conditions, the amount of stored and secreted mucus increases, overwhelms the ciliary clearance mechanism, and leads to mucus inspissation and recurrent infection (3) (4) (5).

Changes similar to those observed in human disease have been observed in the airways of rats exposed to mycoplasma (6), endotoxin (7), SO₂ (8) and tobacco smoke (9). Controlled exposures of rats to these substances therefore provide model systems in which to investigate the molecular pathogenesis of mucus hypersecretion. Our previously reported data, obtained using the human mucin cDNA, SMUC 41 (10), showed that steady state mucin mRNA increased concurrently with the amount of goblet cell mucus in rat airways in which a bronchitis-like condition was induced by exposure to SO₂ and Sendai virus (11). This suggested that mucus overproduction might be regulated at the level of mucin mRNA transcription or stability. Further investigation of this possibility would be facilitated by the availability of rat airway mucin cDNAs.

In the present study, we have used the human intestinal mucin cDNA SMUC 41 (10) to screen a rat airway cDNA library. One of the cDNAs selected by this method was RAM (rat airway mucin) 7s. This cDNA appears to encode a mucin-like protein according to the following criteria: (a) hybridization in plaque lifts to SMUC 41, (b) hybridization in Northern blots to large, polydisperse transcripts, typical of mucin mRNAs (10, 12, 13, 14, 15) (c) sequence characterized by a large number of threonine codons (providing sites for O-linked glycosylation), (d) the presence of tandem repeats (10, 15, 16, 17, 18) and (e) appropriate tissue-specific expression of cognate mRNA. An abstract has appeared (19).

Materials and Methods

Construction of a rat airway cDNA library--To increase the abundance of mucin mRNA in the airway tissue used to construct the library, specific pathogen free (SPF) male Sprague-Dawley rats were exposed to sulfur dioxide (SO₂) at 400 ppm for three hours per day, five days per week for three weeks (11). The respiratory epithelium of treated rats was found to increase in height and to contain more goblet cells than that of control rats (11). RNA from the trachea of one treated rat was prepared by the guanidine thiocyanate and phenol method (21) and was used as a substrate for oligo (dT) and randomly primed cDNA synthesis. EcoRI linkers were ligated and DNA was inserted into the EcoRI site of lambda ZAP II (Stratagene, La Jolla, CA) using Gigapack Gold as a packaging kit (Stratagene). The unamplified library was plated at 30,000 plaques/150-mm x 10 plates.

Screening of the rat airway cDNA library with the human intestinal cDNA, SMUC 41--Plaques were lifted to duplicate nitrocellulose filters (Schleicher & Schuell, Keene, NH). 200 ng of SMUC 41 (10) was labeled with [³²P] dCTP

(Amersham Corp., Arlington Heights, IL) by random priming (BRL Random Priming kit) to a specific activity of $>10^9$ cpm/ μ g DNA. Hybridization was performed at 42°C for 14h in a solution containing 40% formamide, 0.8 M NaCl, 0.02 M PIPES (pH 6.5), 0.5% SDS and denatured salmon sperm DNA (20 μ g/ml). Nitrocellulose filters were then washed in 0.3 M NaCl and 0.03 M sodium citrate (2 x SSC) at room temperature and 0.5 x SSC, 1 % SDS at 63°C for 30 min. Plaque purification and colony hybridization were performed for seven positive clones.

Sequencing--Bluescript vector containing the insert of interest was prepared by *in vivo* excision. Nucleotide sequence was determined by the dideoxy chain termination method (20) using [35 S] dATP (Amersham Corp) and modified T7 DNA polymerase (United States Biochemical Corp., Cleveland, OH). Both strands were sequenced and sequence analysis was conducted using IBI MacVector 3.5 software.

Northern blots--Total RNA was extracted from rat airways and other rat organs by the guanidine thiocyanate and phenol method (21). For each sample, 10 μ g of RNA was separated by electrophoresis on denaturing 1 % agarose-formaldehyde gels and transferred to nylon membranes (GeneScreen, New England Nuclear, Boston, MA) by capillary blotting. cDNAs were labeled with [32 P] dCTP (Amersham Corp., Arlington Heights, IL) by the random priming method (BRL Random Priming kit) to a specific activity of $>10^9$ cpm/ μ g DNA. Hybridization was performed at 42°C for 14-18 h in a solution containing 50 % formamide, 5xSSC, 0.04 % bovine serum albumin, 0.04 % Ficoll, 1.0 % SDS and denatured salmon sperm DNA (>100 μ g/ml). Blots were washed in 2 x SSC at room temperature for 15 min x 2, 0.5 x SSC and 1 % SDS at 52°C for 15 min x 2.

Endotoxin exposure--Specific pathogen-free (SPF) female F344/N 12-14 week old rats from the breeding colony of the Inhalation Toxicology Research Institute, Albuquerque, N. M. were anesthetized with 5 % halothane in oxygen prior to receiving intratracheal instillations of 0.05 ml of saline containing 1 mg of endotoxin (lipopolysaccharide from *Escherichia coli* (*E. Coli* 0111:B4); Sigma Chemical Co., St. Louis, MO). Rats were sacrificed 48 h post-instillation.

SO₂ exposure--SPF male 12 wk old Sprague-Dawley rats were placed in plexiglass chambers and exposed to sulfur dioxide (SO₂) at 400 ppm for three hours per day, five days per week for one to three weeks (11). SO₂-concentrations were monitored in the chamber outflow by a colorimetric assay (22)

Southern blots-- Genomic DNA was prepared using proteinase K, RNase A, and phenol as described by Blin and Stafford (23) This material was digested with restriction enzymes Eco R1, BamH1, Sfi1 and NOT 1 and the fragments were separated by pulse field electrophoresis in 1% agarose gels containing 100 mM Tris, 100 mM boric acid, and 2 mM EDTA, pH 8.0. The gels were soaked for 30 min in 1.5 M NaCl, 0.5 M NaOH, neutralized in 1.5 M NaCl, 0.5M Tris pH 8.0, and then in 0.2 M Tris pH 7.5, 2 x SSC for the same period of time. DNA was transferred to nylon membranes and hybridization was performed at 42°C for 14-18 h in a solution containing 50 % formamide, 0.8 M NaCl, 0.02 M PIPES pH 6.5, 0.5 % SDS denatured salmon sperm DNA (>100 μ g/ml) and the random primer-labelled cDNA RAM 7s. Blots were washed in 2 x SSC at room temperature for 15 min x 2, 0.5 x SSC and 1 % SDS at 52°C for 15 min x 2, 0.2 x SSC and 0.5 % SDS at 58°C for 15 min.

Results

Screening of 3×10^5 plaques yielded 12 positive clones of which 7 were plaque purified. All seven were at least partly sequenced and their ability to hybridize with various tissue RNAs examined. Based on its ability to recognize large mRNA transcripts that were more abundant in the airways of rats with experimentally induced bronchitis than in control rats (11), we selected cDNA RAM 7s for further characterization.

Sequence determination--RAM 7s is 390 bp long (Fig 1A) and encodes many threonine residues (potential O-glycosylation sites) (Fig. 1B). An open reading frame spanning 274 bp is terminated by a stop codon and followed by a 3' untranslated region of 113 bp including a putative polyadenylation signal (AATAAA; nucleotides 306-311). The first two thirds of the cDNA encode seven variable 7-12 amino acid tandem repeats and five non-tandem copies of the consensus sequence TTTTITL. Codon usage for threonine is ACC in all positions. Sequence homology with SMUC 41 is approximately 60% in the first two thirds of the cDNA (Fig. 1B).

Nucleotide/amino acid homology search--Computer searches comparing RAM 7s with other registered proteins (GenBank) did not reveal homology greater than that seen between RAM 7s and SMUC 41. Despite limited amino acid homology, RAM 7s showed 85% nucleotide homology with a human thymidylate synthase gene intron in reverse orientation (RAM 7s from nucleotide 41 to 261 corresponded to the thymidylate synthase gene intron 4 from nucleotide 10388 to 10167.)

Northern blots--In Northern blots (Fig. 2), RAM 7s hybridized strongly to high molecular weight, moderately polydisperse transcripts in tracheal RNA from rats exposed to SO₂/Sendai virus (lane 3) or endotoxin (lanes 1 and 2), but not to tracheal RNA from specific pathogen-free rats (lane 4). RAM 7s also hybridized to mRNA from small and large intestine (lanes 10 and 11) but not to RNA from the heart, kidney, brain, liver or testis (lanes 5-9).

Because the 3' terminal 130 bp of RAM 7s has no sequence similarity to the 5' tandem repeat region, it initially seemed possible that the two domains had

Figure 1A. Nucleotide and deduced amino acid sequence of the 390 bp cDNA RAM 7s. Five copies of an eight amino acid, threonine-rich repeat consensus sequence are present (underlined). The BSTN1 site at nucleotide 271 is indicated. A putative polyadenylation signal (nucleotides 306-11) is in boldface type.

Figure 1B. Nucleotide sequence comparison between RAM 7s and the human intestinal mucin cDNA SMUC 41 (MUC 2).

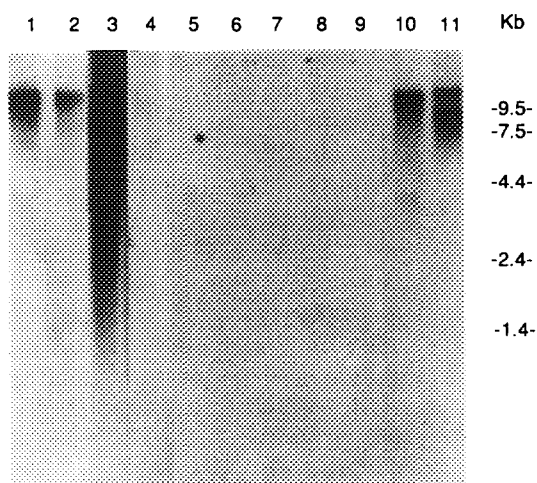


Figure 2. RNA from rat tissues hybridized with RAM 7s.

10 μ g of total RNA was loaded for each sample. Lanes 1 and 2: endotoxin-exposed airway; lane 3: SO₂/Sendai virus-exposed airway; lane 4: SPF airway; lane 5: SPF heart; lane 6: SPF kidney; lane 7: SPF brain; lane 8: SPF liver; lane 9: SPF testis; lane 10: SPF small intestine; lane 11: SPF large intestine.

combined artifactually during cloning. To investigate this, we compared the hybridization patterns produced in RNA (and genomic DNA, see below) blots using probes corresponding to the 5' tandem repeat region or the 3' non-repeat region of RAM 7s. The 5' and 3' end-specific probes were constructed by digestion of RAM 7s with BSTN1, which cleaves after nucleotide number 271, 12 nucleotides downstream of the last repeat. The 5' terminal 271 base fragment (5'-271) and the 119 bp 3' terminal fragment (3'-119) were purified by gel electrophoresis (24). As shown in Figure 3 A and B, both probes recognized the same polydisperse bands when used to probe RNA blots. The polydisperse band centered at ~7.0 kb is the characteristic signal observed in tracheal RNA from rats exposed for one week to SO₂ (11). After two and three weeks of SO₂, the polydisperse signal is centered at ~12 kb. That the repeat and non-repeat domains of RAM 7s showed the same hybridization pattern in these blots indicates that they are recognizing the same mRNA transcripts and are integral parts of the same gene.

Southern blots--Southern blot analysis of genomic DNA from the rat liver is shown in Figure 4. Rat genomic DNA digested with the hexanucleotide-specific restriction enzymes EcoRI and BamHI, and the rare cutters SfiI and NotI was electrophoresed on a pulse field gel. After blotting, the DNA was hybridized with

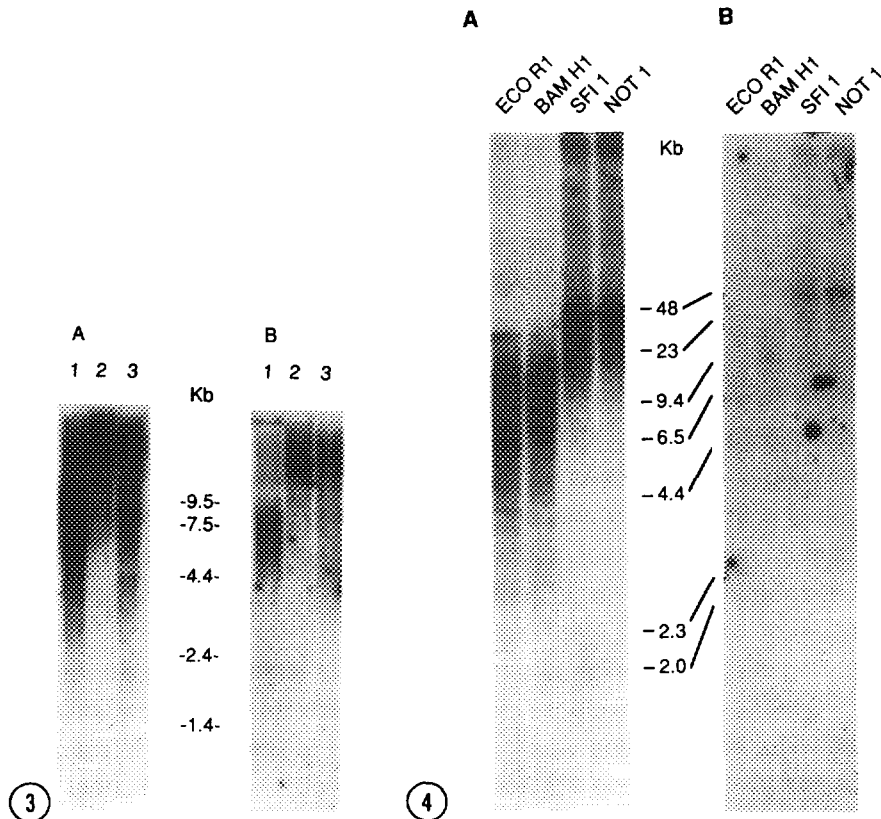


Figure 3. RNA from SO₂-exposed/Sendai virus infected rats hybridized with probes corresponding to either the 5' or 3' end of RAM 7s. 2 mg of total RNA was loaded for each sample. Panel A shows the hybridization pattern obtained using as probe the portion of RAM 7s (5'-271) encoding threonine-rich repeats. Panel B shows the hybridization pattern obtained using as probe the portion of Ram 7s (3'-119) encoding the non-repeat region.

Figure 4. Genomic DNA hybridized with probes corresponding to either the 5' or 3' end of RAM 7s. Panel A shows the hybridization pattern obtained using as probe the portion of RAM7s (5'-271) encoding threonine-rich repeats. Panel B shows the hybridization pattern obtained using as probe the portion of Ram 7s (3'-119) encoding the non-repeat region.

5'-271 (Fig. 4A) or 3'-119 (Fig. 4B). Whereas multiple bands were present after hybridization with the 5' repeat-containing fragment, only single bands were present after hybridization with the 3' non-repeat-containing fragment. The overlap in hybridization patterns obtained with the two fragments provides further

evidence that the 5' and 3' ends of this cDNA are integral parts of the same gene. The blots also indicate that while the threonine-rich repeat domain is common to a relatively large gene family, the non-repeat domain is unique in the rat genome.

Discussion

Our previous findings, showing that mucin steady state mRNA increased in rat tracheas in response to SO₂/Sendai virus exposure (11), suggested that one effect of pathogens and irritants on airway epithelial cells is to increase mucin biosynthesis. Further investigation of this possibility, including a determination of the potential involvement of mucin gene regulatory elements and nucleic acid binding proteins, requires isolation of rat-specific airway mucin cDNAs and genomic clones containing the gene promoter. The cDNA RAM 7s described in this report represents the first rat airway mucin cDNA to be isolated and should serve as a valuable tool for additional screening of rat cDNA and genomic libraries. Analysis of gene regulatory sequences and proteins binding to them under conditions of airway injury should help define the mechanisms controlling mucin gene expression in rat models of human airway disease.

Several lines of evidence indicate that cDNA RAM 7s is a rat airway mucin cDNA. The first of these was its hybridization to the human intestinal mucin cDNA SMUC-41 in plaque lifts from a rat airway cDNA library. Further evidence was provided by the tissue distribution of the recognized mRNA. RAM 7s recognized high molecular mRNA from mucin-secreting tissues in the airways and intestine, but not non-mucin secreting tissues in the heart, kidney, brain or liver or testis (Fig. 2). A related finding was that the mRNA recognized by RAM 7s was greatly increased in the tracheas of rats in which a bronchitis-like condition was induced by exposure to SO₂ and Sendai virus (11).

The large molecular size of mucin mRNAs (Fig. 2) is expected due to the large size of mucin glycoproteins. The size of apomucins from human bronchial mucus has been estimated by gel electrophoresis to range between 200 and 400 kD (25). Similar estimates have been made for rat gastric mucin (26). Assuming the mass of an average amino acid to be approximately 110 daltons, the apomucins analyzed by electrophoresis consist of between 1800 and 3600 amino acids. The mRNAs encoding these apomucins would have minimum sizes ranging between 5 and 11 kb to account for the coding sequences alone, with additional length corresponding to 5' and 3' untranslated regions. The 10-12 kb maximum size of the mRNA recognized by RAM 7s is in agreement with these estimates.

In addition to recognizing mRNAs of large size, mucin cDNAs frequently recognize transcripts of heterogeneous size (10, 14, 17, 27). The reasons for this polydispersity are not understood, nor is it understood why certain mucin

cDNAs (e.g. (28)) recognize non-polydisperse mRNAs. RAM 7s recognizes a major band of approximately 11 kb as well as a zone of lower molecular weight material immediately below. It is evident that this polydispersity, like that associated with other mucin mRNAs, does not occur as a general by-product of RNA degradation because the same RNA samples showing mucin mRNA polydispersity show discrete bands when hybridized with probes encoding other proteins such as β actin or GAPDH (data not shown).

The final line of evidence indicating that RAM 7s is a mucin cDNA was sequence information demonstrating the presence of variable tandem repeats encoding numerous threonine residues, potential sites for O-glycosylation. Tandem repeats enriched in threonine and/or serine codons are a characteristic feature of mucin cDNAs (e.g. 10, 14, 15, 28) although the specific nucleotide sequences comprising these repeats show surprisingly little homology among mucin genes. Indeed, the homology between the threonine-rich repeats in RAM 7s and SMUC-41, the cDNA with which RAM 7s was isolated by library screening, is only ~60%.

It is clear that at least several mucin genes are expressed in man (10, 14, 16, 17, 29). In the rat, which has been less extensively studied, at least two mucin genes, displaying different chromosomal localization, have been described. One of these, the RMUC gene cognate (28) encodes a repetitive peptide with consensus sequence TTTPDV, while the other, the MLP gene cognate (15), encodes a repetitive peptide with the consensus sequence PSTPSTPPPST. RAM 7s, encoding the consensus sequence TTTTIITI, thus bears no similarity to either of the previously described rat mucin genes. Additional sequence information and chromosomal mapping are required, however, to determine whether RAM 7s corresponds to a portion of one of the previously described, or alternatively a novel, rat mucin gene.

Acknowledgments

We thank Drs. Jack Harkema and Jon Hotchkiss for supplying RNA from the tracheas of endotoxin-exposed rats and Drs. Hisamitsu Omori and David Steiger for performing Northern blots with this RNA . This work was supported by NIH RO1-HL43762 and PPGHL24136 and the Department of Veterans' Affairs Medical Research Service. Dr. Tsuda was supported by a postdoctoral award from the Cystic Fibrosis Foundation.

References

1. Sheehan, J. K., Thornton, D. J., Somerville, M. and Carlstedt, I., (1991) *Am Rev Resp Dis Suppl* 144, S4-S9.
2. Schacter, H. and Williams, D., (1981) *Mucus in Health and Disease II* 3-28.

3. Aikawa, T., Shimura, S., Sasaki, H., Ebina, M. and Takishima, T., (1992) *Chest* 101, 916-21.
4. Aikawa, T., Shimura, S., Sasaki, H., Takishima, T., Yaegashi, H. and Takahashi, T., (1989) *American Review of Respiratory Diseases* 140, 477-482.
5. Lopez-Vidriero, M. T. and Reid, L., (1977) *Mod. Probl. Paediat.* 19, 120-128.
6. McDonald, D., Schoeb, T. and Lindsey, J. R., (1991) *J. Clin. Invest.* 87, 787-799.
7. Harkema, J. R. and Hotchkiss, J. A., (1991) *Experimental Lung Research* 17, 743-761.
8. Lamb, D. and Reid, L., (1968) *Journal of Pathology & Bacteriology* 96, 97-111.
9. Jeffery, P. K. and Reid, L. M., (1981) *Journal of Pathology* 133, 341-359.
10. Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A. and Kim, Y. S., (1989) *The Journal of Biological Chemistry* 264, 6480-6487.
11. Jany, B., Gallup, M., Tsuda, T. and Basbaum, C., (1991) *Biochemical and Biophysical Research Communications* 181, 1-8.
12. Bhargava, A. K., Weitach, J. T., Davidson, E. A. and Bhavanandan, V. P., (1990) *Proc. Natl. Acad. Sci. USA* 87, 6798-6802.
13. Eckhardt, A. E., Timpote, C. S., Abernethy, J. L., Zhao, Y. and Hill, R. L., (1991) *The Journal of Biological Chemistry* 266(15), 9678-9686.
14. Porchet, N., Van Cong, N., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A. and Aubert, J. P., (1991) *Biochemical and Biophysical Research Communications* 175(2), 414-422.
15. Xu, G., Huan, L.-J., Khatri, I. A., Wang, D., Bennick, A., Fahim, R. E. F., Forstner, G. G. and Forstner, J. F., (1992) *J. Biol. Chem.* 267, 5401-5407.
16. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, K., Lalani, E. N. and Wilson, D., (1990) *The Journal of Biological Chemistry* 265, 15286-15293.
17. Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. L., Byrd, J. C., Lamport, D. T. A., Siddiki, B. and Kim, Y. S., (1990) *Biochemical and Biophysical Research Communications* 171, 407-415.
18. Probst, J. C., Hauser, F., Joba, W. and Hoffmann, W., (1992) *J. Biol. Chem.* 267, No. 9, 6310-6.
19. Tsuda, T., Jany, B., Gallup, M. and Basbaum, C., (1991) *Am. Rev. Resp. Dis.* 143, A141.
20. Sanger, F., Nicklen, S. and Coulson, A., (1977) *PNAS USA* 74, 5463-5467.

21. Chomczynski, P. and Sachi, N., (1987) *Anal. Biochem.* 162, 156-159.
22. U.S. Department of Health, Education and Welfare Publication (1974) 74-128.
23. Blin, N. and Stafford, D., (1976) *Nucleic Acids Res.* 3, 2303.
24. Sambrook, J., Fritsch, E. F. and Maniatis, T., (1989) 2nd ed. Molecular Cloning, a Laboratory Manual.
25. Marianne, T., Perini, J.-M., Lafitte, J.-J., Houdret, N., Pruvot, F.-R., Lamblin, G., Slayter, H. S. and Roussel, P., (1987) *Biochemical Journal* 248, 189-195.
26. Dekker, J., Van Beurden-Lamers, W. and Strous, G., (1989) *J. Biol. Chem.* 18, 10431-7.
27. Jany, B., Gallup, M., Yan, P., Gum, J., Kim, Y. and Basbaum, C., (1991) *J. Clin. Invest* 87, 77-82.
28. Gum, J. R., Jr., Hicks, J. W., Lagace, R. E., Byrd, J. C., Toribara, N. W., Siddiki, B., Fearney, F. J., Lamport, D. T. A. and Kim, Y. S., (1991) *J. Biol. Chem.* 266(33), 22733-8.
29. Aubert, J.-P., Porchet, N., Crepin, M., Duterque-Coquillaud, M., Vergnes, G., Mazzuca, M., Debuire, B., Petiprez, D. and Degand, P., (1991) *American Journal of Respiratory Cell and Molecular Biology* 5, 178-185.