



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Identification of a human ortholog of the mouse Dcpp gene locus, encoding a novel member of the CSP-1/Dcpp salivary protein family

**Citation for published version:**

Mullins, JJ, Mullins, LJ, Dunbar, DR, Brammar, WJ, Gross, KW & Morley, S 2006, 'Identification of a human ortholog of the mouse Dcpp gene locus, encoding a novel member of the CSP-1/Dcpp salivary protein family' *Physiological Genomics*, vol. 28, no. 1, pp. 129-140. DOI: 10.1152/physiolgenomics.00153.2006

**Digital Object Identifier (DOI):**

[10.1152/physiolgenomics.00153.2006](https://doi.org/10.1152/physiolgenomics.00153.2006)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Physiological Genomics*

**Publisher Rights Statement:**

open access article

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Identification of a human ortholog of the mouse *Dcpp* gene locus, encoding a novel member of the CSP-1/Dcpp salivary protein family

John J. Mullins, Linda J. Mullins, Donald R. Dunbar, William J. Brammar, Kenneth W. Gross and Steven D. Morley

*Physiol. Genomics* 28:129-140, 2007. First published 5 September 2006;  
doi: 10.1152/physiolgenomics.00153.2006

## You might find this additional info useful...

---

This article cites 43 articles, 20 of which you can access for free at:  
<http://physiolgenomics.physiology.org/content/28/1/129.full#ref-list-1>

Updated information and services including high resolution figures, can be found at:  
<http://physiolgenomics.physiology.org/content/28/1/129.full>

Additional material and information about *Physiological Genomics* can be found at:  
<http://www.the-aps.org/publications/physiolgenomics>

---

This information is current as of September 19, 2013.

*Physiological Genomics* publishes results of a wide variety of studies from human and from informative model systems with techniques linking genes and pathways to physiology, from prokaryotes to eukaryotes. It is published 24 times a year (twice monthly) by the American Physiological Society, 9650 Rockville Pike, Bethesda MD 20814-3991. Copyright © 2007 the American Physiological Society. ISSN: 1531-2267. Visit our website at <http://www.the-aps.org/>.

# Identification of a human ortholog of the mouse *Dcpp* gene locus, encoding a novel member of the CSP-1/Dcpp salivary protein family

John J. Mullins,<sup>1</sup> Linda J. Mullins,<sup>1</sup> Donald R. Dunbar,<sup>1</sup>  
William J. Brammar,<sup>2</sup> Kenneth W. Gross,<sup>3</sup> and Steven D. Morley<sup>1</sup>

<sup>1</sup>Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh Medical School, Edinburgh; <sup>2</sup>Department of Biochemistry, University of Leicester, Leicester, United Kingdom; and <sup>3</sup>Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, New York

Submitted 18 July 2006; accepted in final form 13 August 2006

**Mullins JJ, Mullins LJ, Dunbar DR, Brammar WJ, Gross KW, Morley SD.** Identification of a human ortholog of the mouse *Dcpp* gene locus, encoding a novel member of the CSP-1/Dcpp salivary protein family. *Physiol Genomics* 28: 129–140, 2006. First published September 5, 2006; doi:10.1152/physiolgenomics.00153.2006.—Salivary fluid, the collective product of numerous major and minor salivary glands, contains a range of secretory proteins that play key defensive, digestive, and gustatory roles in the oral cavity. To understand the distinct protein “signature” contributed by individual salivary glands to salivary secretions, we studied a family of proteins shown by *in vitro* mRNA translation to be abundantly expressed in mouse sublingual glands. Molecular cloning, Southern blotting, and restriction fragment length polymorphism analyses showed these to represent one known and two novel members of the common salivary protein (CSP-1)/Demilune cell and parotid protein (Dcpp) salivary protein family, the genes for which are closely linked in the T-complex region of mouse chromosome 17. Bioinformatic analysis identified a putative human CSP-1/Dcpp ortholog, HRPE773, expressed predominantly in human salivary tissue, that shows 31% amino acid identity and 45% amino acid similarity to the mouse *Dcpp* query sequence. The corresponding human gene displays a similar structure to the mouse *Dcpp* genes and is located on human chromosome 16 in a region known to be syntenic with the T-complex region of mouse chromosome 17. The predicted mouse and human proteins both display classical NH<sub>2</sub>-terminal signal sequences, putative jacalin-related lectin domains, and potential N-linked glycosylation sites, suggesting secretion via sublingual saliva into the oral cavity where they may display antimicrobial activity or provide a defensive coating to enamel. Identification of a human CSP-1/Dcpp ortholog therefore provides a key tool for investigation of salivary protein function in human oral health and disease.

mouse sublingual gland; demilune cell parotid protein; jacalin-related lectin domain; genomics; syntenic chromosome region; common salivary protein

SALIVARY FLUID FULFILLS CRITICAL roles in lubrication and controlling the microflora of the oral and esophageal epithelia, in maintaining oral pH, limiting dental decay, and in softening and initiating the digestion of food so that it can be tasted and swallowed (24, 25, 44). Decreased salivary secretion as a consequence of age, therapeutic drug treatment, or salivary gland damage resulting from disease or radiotherapy frequently leads to a dry mouth (xerostomia) and recurrent oral infection

and inflammation, coupled with increased risk of dental erosion, dental caries, periodontal disease, and impaired digestion. Left untreated, this combination of symptoms may lead to malnutrition and poor long-term patient outcome, with consequent economic impact in both healthcare and social settings. Several major protein classes have been identified in salivary fluid, including the so-called proline-rich proteins, making up >70% of the secreted protein in human saliva (3), mucins, the major high-molecular-weight glycoprotein component of mucus (43, 44), enzymes initiating digestion (25), and hydrophilic molecule transporters, which may concentrate and deliver taste compounds to the gustatory system (19, 20). The functions of other salivary proteins, including the glutamine-rich and histidine-rich proteins and tyrosine-rich statherin (12, 31), remain unclear, as do the roles of enzymes, such as renin and the kallikreins, and growth factors, such as epidermal growth factor and nerve growth factor, which are also synthesized in the salivary glands and released either via a regulated pathway into the salivary ducts or by constitutive secretion into the blood stream (5, 6, 36). Consequently, if effective treatments and therapies are to be developed, it will first be necessary to understand more about the mechanisms underpinning salivary gland-specific gene expression, the contribution of specific glands to the final salivary fluid composition, and the specific functions of the proteins present therein.

Saliva is collectively the product of three major pairs of salivary glands, the submandibular, sublingual, and parotid glands and several hundred smaller glands, such as von Ebner's glands found within the tongue and the palatine, buccal, and labial glands located in the submucosa of the oral cavity (8, 19, 29, 35). The major salivary glands contribute distinct, but overlapping, protein “signatures” to the final salivary fluid profile and probably share common developmental origins, though undergoing divergent morphogenic and cellular differentiation programs that give rise to distinct secretory and nonsecretory exocrine cell phenotypes (11). Salivary gland secretory exocrine cells can be further subdivided morphologically and by their production or otherwise of salivary mucins into mucous (nonserous) and serous cell types. Serous cells contribute nonmucin components to the saliva, for example, the rat parotid secretory protein, which is related to the neonatally expressed rat submandibular gland protein SMG-A (30) and mouse sublingual demilune protein P20 (7). Sublingual demilune protein P20, the mouse equivalent of rat common salivary protein 1 (CSP-1) (15), is now formally defined as Demilune cell and parotid protein (Dcpp) and is secreted principally from the sublingual gland and in lesser quantities from the submandibular and parotid glands. Thus, in addition

Article published online before print. See web site for date of publication (<http://physiolgenomics.physiology.org>).

Address for reprint requests and other correspondence: J. J. Mullins, Rm. W3.31, Centre for Cardiovascular Science, Queen's Medical Research Inst., Univ. of Edinburgh Medical School, 47 Little France Crescent, Edinburgh, EH16 4TJ, UK (e-mail: j.mullins@ed.ac.uk).

to their developmental and sexually dimorphic expression patterns (4, 11), salivary gland proteins are also expressed gland specifically. Collectively, therefore, the salivary glands offer a valuable experimental system in which to investigate the mechanisms underlying developmental and cell-specific gene expression in secretory tissues and perhaps also the evolution of secretory protein function. Furthermore, an understanding of salivary gland biology will be essential to the development of gene therapy-based treatments of salivary or endocrine gland insufficiency (6, 47) or to exploit salivary glands for the expression of heterologous proteins in transgenic animals (28).

As a step toward understanding the basis of salivary gland-specific gene expression, we analyzed the expression of a family of mouse salivary proteins, provisionally termed SPT proteins because of their abundant expression in sublingual glands and lesser expression in parotid glands and tongue. Molecular and bioinformatic analysis identified the mouse salivary proteins as one known and two additional novel members of the CSP-1/Dcpp gene family and demonstrated the presence of three closely linked *Dcpp* genes, forming a locus, in the T-complex region of mouse chromosome 17. Protein sequence similarity searching identified HRPE773 protein as a putative human ortholog of the CSP-1/Dcpp family, showing 31% amino acid identity and 45% amino acid similarity to the SPT-2 query sequence. The corresponding human gene displays a similar structure to that of the mouse *Dcpp* genes and is located on the region of human chromosome 16 that is syntenic with the T-complex region of mouse chromosome 17. The possibility that the *Dcpp* proteins are mammalian lectins possessing antimicrobial properties is discussed.

#### EXPERIMENTAL PROCEDURES

**Animals and tissue collection.** All animal work was carried out in accordance with the provisions of the Animals (Scientific Procedures) Act (UK) 1986 and the National Institutes of Health *Guidelines for the Care and Use of Laboratory Animals*. Mice were killed by cervical dislocation, and following collection, tissues were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  to await extraction of RNA.

**Poly(A)<sup>+</sup> mRNA isolation and in vitro translation.** For most in vitro translation and Northern blotting experiments, total RNA was prepared from pooled male and female adult (11 wk old) DBA/2 mouse submandibular, sublingual, or parotid glands (see Fig. 1A) and a range

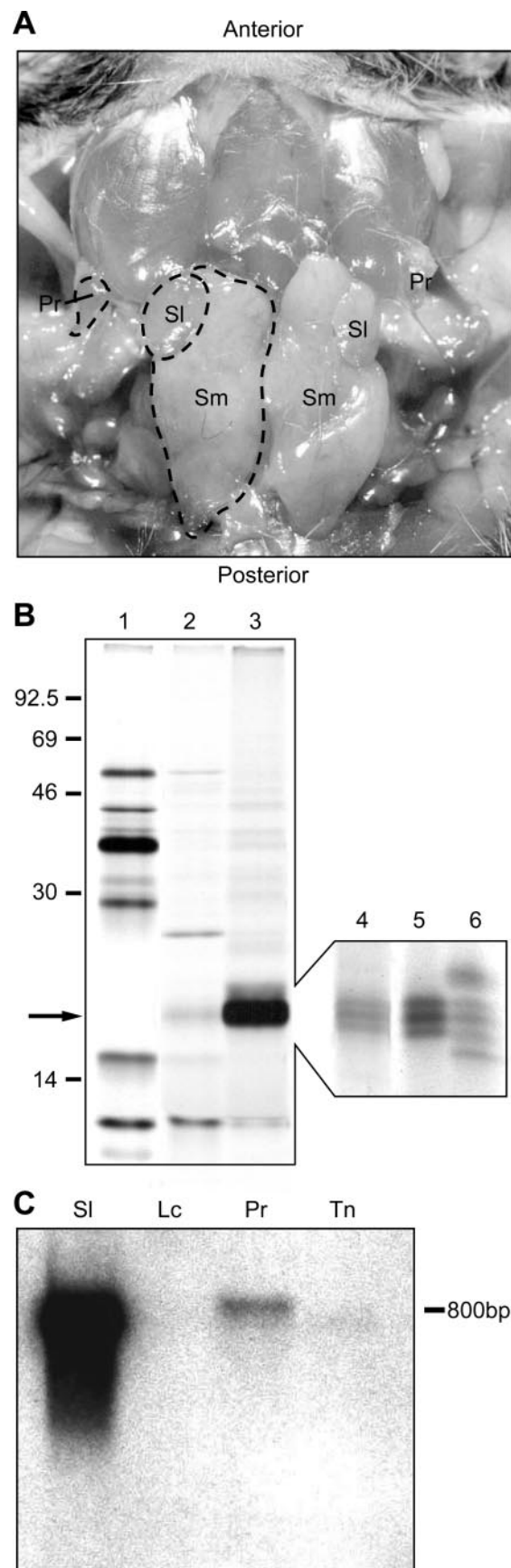


Fig. 1. In vitro translation and expression analysis of mouse salivary gland mRNAs. **A:** location of the major salivary glands in the neck region of the mouse; Sm, submandibular; SI, sublingual; Pr, parotid. "Anterior" and "posterior" indicate the orientation of the specimen. **B:** in vitro translation of mouse salivary gland poly(A)<sup>+</sup> mRNAs, analyzed on a 12.5% (wt/vol) SDS-PAGE gel: submandibular gland (lane 1), parotid gland (lane 2), sublingual gland (lane 3) and analyzed on a 17.5% (wt/vol) SDS-PAGE gel: sublingual gland (lane 4), cDNA clone pSMG-181 hybrid-selected sublingual gland poly(A)<sup>+</sup> mRNA (lane 5), and cDNA clone pSMG-181 hybrid-selected sublingual gland poly(A)<sup>+</sup> mRNA, reaction supplemented with dog pancreatic microsomes (lane 6). Protein standard size markers (kDa) are indicated at the left, and the major 16- to 17-kDa sublingual in vitro translation products are indicated by an arrow. **C:** Northern blot analysis of 10  $\mu\text{g}$  of poly(A)<sup>+</sup> RNA isolated respectively from pooled sublingual gland (SI), lachrymal (Lc), parotid (Pr), or tongue (Tn) from four 3- to 4-mo-old male DBA/2 littermates, hybridized with the random primed <sup>32</sup>P-labeled cDNA insert fragment from cDNA clone pSMG-181, and washed last with  $1\times$  SSC/0.1% (wt/vol) SDS. Low-level expression was also detected in submandibular gland (data not shown). Other tissues screened, but in which no signal was detected, were pancreas, seminiferous tubule, coagulant gland, preputial gland, epididymus, hardarian gland, adrenal gland, brain, liver, spleen, heart, and kidney.

of other tissues. For experiments examining possible sexually dimorphic expression, total RNA was prepared separately from groups of intact or castrated males or untreated or testosterone-implanted female mice. Poly(A)<sup>+</sup> RNA, purified by oligo-dT chromatography, was translated in nuclease-treated rabbit reticulocyte lysate (26) containing 20 μM each amino acid, excepting methionine, 1 mM ATP, 0.2M GTP, 80 mM potassium chloride, 2 mM magnesium acetate, 10 mM Tris·HCl pH 7.6, 2 mM glucose, 7 mg/ml creatinine phosphate, 50 μg/ml calf liver tRNA, 1 mCi/ml [<sup>35</sup>S]methionine (specific activity 1,160 Ci/mmol, Amersham), and 50 μg/ml creatine phosphokinase. In vitro translation products were resolved by SDS-PAGE. For in vitro processing experiments, in vitro translation reactions were supplemented with 4 OD<sub>260</sub> units/ml of dog pancreatic microsomes (New England Nuclear). Specific sublingual mRNA classes were isolated by hybrid selection (26) from a total sublingual gland poly(A)<sup>+</sup> RNA population, using plasmid DNA derived from individual cDNA clones isolated from a previously constructed DBA/2 mouse pooled adult male and female submandibular-sublingual gland cDNA library (48) immobilized on nitrocellulose filters. In brief, 10 μg of selected cDNA plasmids were denatured individually in 1 M NaCl, 0.5 M NaOH, and 10 mM EDTA at 65°C for 20 min, spotted on to separate nitrocellulose filter discs (Schleicher & Schüll, 0.45 μm pore size), dried, and rinsed briefly in 3× SSC (1× SSC = 0.15 M sodium chloride/0.015 sodium citrate, pH 7.0), baked at 80°C for 5 h, and rinsed finally with water at 80°C to remove any unbound plasmid DNA. Filters were prehybridized in 50% (wt/vol) formamide, 0.4 M NaCl, 4 mM EDTA, 0.6 mg/ml calf liver tRNA (Boehringer), 10 mM Na-PIPES buffer, pH 6.4, at 41°C for 3 h and then hybridized for 14 h at 41°C in prehybridization solution containing 30 μg/ml mouse DBA/2 sublingual gland poly(A)<sup>+</sup> mRNA. Filters were then washed successively, twice in 1× SSC, 0.5% (wt/vol) SDS for 3 min at room temperature, three times in 0.1× SSC, 0.5% (wt/vol) SDS for 3 min at room temperature, twice in 0.1× SSC, 0.5% (wt/vol) SDS for 3 min at 50°C, and twice in 10 mM Tris·HCl, pH 7.5, containing 1 mM EDTA, for 3 min at 50°C. Filter-bound RNA was eluted by two successive incubations in 150 μl of water for 2 min at 80°C and collected by ethanol precipitation of pooled eluates in the presence of 4 μg of calf liver tRNA as carrier. Half of this material was subjected to in vitro translation, as described previously.

**Double amino acid labeling of sublingual gland in vitro translation products.** Double amino acid labeling of major sublingual gland in vitro translation products and human placental and bovine pituitary mRNA controls was carried out essentially as described above, but using amino acid-depleted rabbit reticulocyte lysates (Amersham) to avoid dilution of tritiated amino acid labels and the adjustment to 125 mM potassium chloride in translation reactions to maximize translational efficiency. Double amino acid labeling was achieved by including 1 mCi/ml [<sup>35</sup>S]methionine (specific activity 1,160 Ci/mmol, Amersham) and a quantity of either [<sup>3</sup>H]alanine or [<sup>3</sup>H]serine empirically determined to result in the incorporation of approximately half the number of [<sup>3</sup>H] to [<sup>35</sup>S] counts, in the absence of their nonlabeled homologs. In vitro translation products were resolved on a 17.5% (wt/vol) SDS-PAGE gel and transferred electrophoretically to nitrocellulose filters (4 h, 60 V/0.2 A). Blots were air-dried, and translation products were located by autoradiography. Individual products were cut out from the blot, and incorporated counts were quantitated with a dual-channel (0–12 keV, 12–156 keV) scintillation counter. Counting efficiency was determined by counting known amounts of [<sup>3</sup>H]alanine and [<sup>35</sup>S]methionine spotted on to nitrocellulose strips. To allow for the differing specific activities of the [<sup>3</sup>H]- and [<sup>35</sup>S]-labeled amino acids, a correction factor was calculated by dividing real ratios of alanine-methionine and serine-methionine predicted from the known amino acid composition of human prolactogen (40, 41), by the experimentally determined ratios. These correction factors were validated by application to the alanine-methionine and serine-methionine ratios determined experimentally for bovine prolactin, another protein of known amino acid composition (38), giving corrected ratios

of 1.1 and 2.3, respectively, that correspond well with the real ratios of 1.25 and 2.5. Finally, these correction factors were applied to the experimentally determined alanine-methionine and serine-methionine ratios for the major sublingual gland in vitro translation products, and the resulting corrected ratios were compared with those determined for the previously characterized sublingual gland mucin (37).

**Northern and Southern blot analysis.** Poly(A)<sup>+</sup> RNA (10 μg) and size markers (5 μg total RNA) were denatured with glyoxal and resolved on a 2% (wt/vol) agarose gel containing 10 mM sodium phosphate, pH 7.0, buffer. RNA was then capillary-transferred to nitrocellulose membranes in 20× SSC, deglyoxylated, and cross-linked to the membrane by baking blots at 80°C under vacuum for 2 h (26). Similarly, 10 μg of genomic DNA, prepared from DBA/2 mouse tail biopsies, were digested to completion with selected restriction enzymes, fractionated on a 0.8% (wt/vol) agarose gel, transferred, and immobilized as above. Blots were hybridized under standard conditions with random-primed <sup>32</sup>P-labeled DNA probes (13) and washed at final SSC concentrations, as specified in figure legends, for 30 min at 65°C and autoradiographed finally for 48 h at –70°C.

**cDNA and genomic library screening.** cDNA prepared from 5 μg of DBA/2 mouse sublingual gland poly(A)<sup>+</sup> RNA by the method of Gubler and Hoffman (17) was inserted via adapters into the *Eco*RI site of the plasmid vector Bluescript SK(+) (Stratagene) and transformed into *Escherichia coli* DH5. Ampicillin-resistant colonies were arrayed on nitrocellulose filters and screened by standard procedures (26) [last wash 2× SSC/0.1% (wt/vol) SDS], using the random-primed <sup>32</sup>P-labeled cDNA insert from a partial cDNA clone pSMG-181 (JJ Mullins, LJ Mullins, and WJ Brammar, unpublished observation; see RESULTS for further details). cDNA inserts from colonies registering positive hybridization signals were sized by double digestion with *Bam*HI/*Hind*III restriction enzymes, whose recognition sequences flank the *Eco*RI site in the Bluescript SK(+) polylinker. A *Mus hortulanus* lambda phage genomic library (1) was screened by standard procedures using a SPT-1 partial cDNA clone (SPT-42 corresponding to bp 99–469 of the SPT-1 nucleotide sequence), and the isolated phage were mapped by *Bam*HI, *Sal*I, and *Hind*III partial restriction digestion.

**DNA sequencing.** cDNA inserts in Bluescript SK(+) (Stratagene) were either subcloned into M13mp8/9 or sequenced directly using the T7 Sequenase version 2.0 kit (USB). Genomic DNA fragments from positive lambda phage were subcloned into Bluescript SK(+) and sequenced using an ABI Automatic Sequencer and four-color fluorescent dye chemistry. Primer extension sequencing of the 5'-end of sublingual gland mRNA used a 33-base primer (5'-GTTTTCCAAC-TCTGGACCATGATAATCTTG TG-3') corresponding to bp 138–106 of the SPT-1 cDNA sequence.

**Pulsed-field gel electrophoresis.** High-molecular-weight DNA was prepared by embedding a suspension of DBA/2 or C57BL/6 spleen cells in agarose plugs and then lysing the cells using 1 mg/ml proteinase K and 1% (wt/vol) sarcosyl. Following in situ digestion of the DNA with rare-cutting restriction enzymes, DNA was separated using pulsed-field gel electrophoresis conditions that achieved separation between 50 and 400 kb, transferred to nitrocellulose membranes, and immobilized as described previously and hybridized using the random-primed <sup>32</sup>P-labeled SPT-42 cDNA insert.

**Recombinant inbred analysis.** Inbred strains of laboratory C57BL/6J and DBA/2J and BXD RI strains (derived from C57BL/6J and DBA/2J) (45) were obtained from the Jackson Laboratories (Bar Harbor, ME). Tail clip DNA was screened by Southern blotting, using the random-primed <sup>32</sup>P-labeled SPT-42 cDNA insert to identify a restriction fragment length polymorphism (RFLP) between the inbred strains C57BL/6J and DBA/2J. Strain distribution patterns were compared with published patterns (16) with statistical analysis being carried out according to the method of Silver (42).

**Bioinformatics analyses.** Nucleic acid and amino acid sequences were searched by basic local alignment search tool (BLAST) (2) with default parameters against several databases maintained by National

Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/BLAST/>) including nr and nr EST division. The genome build searched was NCBI32. Gene co-ordinates are based on the mouse genomic sequence AC110262.12. Sequences were aligned by CLUSTAL W (9) using default parameters: weight matrix Gonnet, gap open = 10.0, gap extension = 0.2, gap separation distance = 4.0. Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) was used to look for functional domains. Genome annotations of the mouse T-complex region of mouse chromosome 17 and the syntenic region of human chromosome 16 were abstracted from the Ensembl (<http://www.ensembl.org/>) project using the Ensembl V38 annotation of the mouse and human genome builds (18).

## RESULTS

*In vitro translation analysis of mouse sublingual gland mRNAs.* In vitro translations of poly(A)<sup>+</sup> mRNA prepared from pooled male and female adult DBA/2 mouse submandibular, parotid, or sublingual glands (Fig. 1A), carefully dissected free of surrounding tissue, revealed a variety of products of varying size (Fig. 1B, lanes 1 and 2). Of these, the prominent submandibular products with apparent molecular masses of 44 kDa, 36 kDa, and 30 kDa have been identified previously as representing preprorenin (32), a secreted acidic glycoprotein, termed Spot protein or salivary protein-1 (12, 48), and several members of the esteropeptidase (kallikrein) family (27). However, the striking major sublingual product of 16–17 kDa (Fig. 1B, lane 3), also present at lower levels in parotid gland but apparently absent from submandibular gland mRNA in vitro translations, has yet to be characterized. Further analysis of sublingual gland mRNA in vitro translations on a 17.5% (wt/vol) SDS-PAGE gel resolved the broad 16- to 17-kDa product, seen in Fig. 1B, lane 3, into three distinct bands of approximately equal intensity (Fig. 1B, lane 4), indicating the existence of three polypeptide species, perhaps differing in length by only a few amino acids. In vitro translation of sublingual gland poly(A)<sup>+</sup> mRNA from intact or castrated males or untreated or testosterone-implanted female mice failed to show any sexual dimorphism of the 16- to 17-kDa products (data not shown), suggesting that genes encoding these proteins are not subject to androgen-mediated regulation such as that seen for salivary gland renin, EGF, NGF, and kallikrein gene expression (5). Use of a number of plasmid cDNA clones from a previously constructed DBA/2 mouse pooled submandibular-sublingual gland cDNA library to hybrid-select specific sublingual poly(A)<sup>+</sup> mRNA classes identified a single plasmid, pSMG-181, which selected mRNAs encoding three in vitro translation products of identical size to the three major 16- to 17-kDa peptides produced by direct in vitro translation of sublingual gland poly(A)<sup>+</sup> RNA (Fig. 1B, lanes 4 and 5). The ability of a single cDNA clone to hybrid-select mRNAs encoding three distinct 16- to 17-kDa products supports the notion of a family of closely related abundant sublingual gland proteins. Addition of dog pancreatic microsomes to hybrid-selected mRNA in vitro translations yielded several additional translation products, suggesting that processing of a signal sequence and glycosylation may be occurring (Fig. 1B, lane 6). Double labeling of sublingual gland poly(A)<sup>+</sup> mRNA in vitro translation products with either [<sup>3</sup>H]Ala + [<sup>35</sup>S]Met or [<sup>3</sup>H]Ser + [<sup>35</sup>S]Met (Table 1) showed that the amino acid compositions of the major 16- to 17-kDa sublingual gland peptides are quite distinct from that of the

Table 1. Alanine-methionine and serine-methionine ratios for the major sublingual gland in vitro translation products

Sample	<sup>3</sup> H(Ala) dpm	<sup>35</sup> S(Met) dpm	Experimental Ratio	Corrected Ratio
HPPL	2,166	4,307	0.50	(1.43)
BP	11,600	30,528	0.38	1.1
SL-A	2,272	4,673	0.49	1.4
SL-B	4,632	8,752	0.53	1.5
SL-C	4,728	8,394	0.56	1.6
SL-D	2,564	4,502	0.57	1.6
SL-mucin				11.0

Sample	<sup>3</sup> H(Ser) dpm	<sup>35</sup> S(Met) dpm	Experimental Ratio	Corrected Ratio
HPPL	1,821	3,420	0.53	(2.86)
BP	18,398	43,451	0.42	2.3
SL-A	1,858	3,399	0.55	3.0
SL-B	3,544	5,650	0.63	3.4
SL-C	4,076	5,848	0.70	3.8
SL-D	1,906	2,999	0.64	3.5
SL-mucin				12.75

Sublingual gland (SL) major in vitro translation products and human placental preprolactogen (HPPL) and bovine prolactin (BP) (proteins of known size and amino acid composition expressed respectively in human placenta and bovine pituitary), were double-labeled, as described in EXPERIMENTAL PROCEDURES, by inclusion of [<sup>35</sup>S]methionine and either [<sup>3</sup>H]alanine or [<sup>3</sup>H]serine in in vitro translation reactions. Experimentally determined labeling ratios were corrected for the different specific activities of the amino acid labels, as described in EXPERIMENTAL PROCEDURES, by using a correction factor derived from comparison of the experimentally determined labelling ratios for HPPL with the real values calculated from its known amino acid composition. The validity of the correction factor was first established by correcting labeling ratios determined experimentally for BP, which were shown to compare closely with the real values.

previously characterized sublingual gland mucin (37), 60% of which consists of only three amino acids: threonine, serine, and alanine. The small size of the major sublingual mRNA in vitro translation products argues further against their identification as mucins (43). Nevertheless, the similarity of the amino acid labeling ratio for all three major sublingual gland in vitro translation products suggests that they may be members of a closely related protein family.

*Molecular cloning of mRNAs for the 16- to 17-kDa mouse sublingual gland proteins.* Oligo dT-primed reverse transcription of total sublingual RNA revealed the presence of a prominent cDNA product of 600–650 bp (data not shown), corresponding to an mRNA of 750–850 bases, assuming a poly(A) tail of 150–200 residues. Northern blotting of a panel of exocrine gland poly(A)<sup>+</sup> mRNAs revealed that mRNAs hybridizing to cDNA clone pSMG-181 were extremely abundant in the sublingual gland (Fig. 1C), suggesting that mRNAs corresponding to the 16- to 17-kDa sublingual peptides may represent the predominant sublingual transcription product. pSMG-181-hybridizing mRNAs were also present at lower levels in parotid gland, tongue, and submandibular gland, but undetectable in all other exocrine tissues examined (Fig. 1C). Interestingly, while hybridization signals from sublingual gland and tongue correspond to polyadenylated mRNAs of 750–850 bases, the parotid mRNA is ~50–100 bases longer, suggesting the use of either an alternative transcription start site or polyadenylation signal. While the prominent 16- to 17-kDa sublingual and parotid gland in vitro translation products (Fig. 1B, lanes 2 and 3) are consistent with the Northern hybridization signals seen for these tissues, failure to observe a

similar submandibular *in vitro* translation product (Fig. 1B, lane 1) indicates either that these proteins are present at lower levels in submandibular gland or, less likely, that the observed Northern signal may arise from a small contamination of submandibular tissue used to prepare poly(A)<sup>+</sup> mRNA for Northern blotting with sublingual tissue.

Since clone pSMG-181 contained an incomplete open reading frame, a DBA/2 mouse sublingual gland cDNA library was screened to isolate full-length clones. Of clones screened, 22.5% gave positive hybridization signals, confirming that pSMG-181-related sequences are highly represented in the total sublingual gland mRNA population. Nine pSMG-181-hybridizing clones, having cDNA inserts of >600 bp in length, were sequenced and resolved into two distinct but closely related families, provisionally termed SPT-1 and SPT-2, differing by ~7% at the nucleic acid level and 10% at the amino acid level (Fig. 2, A and B). Primer extension sequencing of the extreme 5'-region of DBA/2 sublingual gland poly(A)<sup>+</sup> RNA yielded a single unambiguous sequence (corresponding to bp 1–105 of the SPT-1/2 cDNA sequence, Fig. 3A), demonstrating that SPT-1 and SPT-2 share identical sequences between the extension primer and the transcription start site.<sup>1</sup> Both SPT-1 and SPT-2 cDNAs exhibit single open reading frames (Fig. 3A), predicting SPT protein moieties of 170 and 171 amino acids, respectively, each exhibiting classical signal sequences, a putative Jacalin-related lectin (JRL) domain, and potential sites for glycosylation (Fig. 2B), suggesting that the proteins are secreted into sublingual saliva. Both cDNA clone classes display two potential ATG “start” codons, the second of which is contained within a partial match to the Kozak translation initiation consensus motif (21). This suggests that translation may initiate from this second ATG, the consequence of which would be a shorter NH<sub>2</sub>-terminal “signal” sequence that, however, would still retain a secretory motif. These characteristics are shared by a further cDNA sequence class, termed SPT-3, deduced from datamining of the mouse genome (see later), which also predicts an open reading frame of 170 amino acids (Fig. 2B). In addition to a single amino acid deletion, the SPT-2 and SPT-3 coding sequences predict, respectively, 16 and 23 amino acid substitutions compared with SPT-1, while SPT-2 and SPT-3 coding sequences differ from each other by 18 amino acids, most of which are isofunctional in nature. Base changes between the SPT cDNAs and therefore amino acid substitutions between the conceptual SPT proteins are concentrated toward the middle of the predicted open reading frames, within the putative JRL domain, while the NH<sub>2</sub>- and COOH-terminal ends display a lesser degree of amino acid substitution.

*Southern blot hybridization identifies three genes for SPT proteins.* Southern analyses of either *Pst*I or *Bam*HI-digested DBA/2 genomic DNA, using SPT-54, a short centrally located 90-bp cDNA clone corresponding to bp 291–380 of the SPT-1 nucleotide sequence, as probe and moderately stringent washing [last wash 1× SSC/0.1% (wt/vol) SDS], revealed three hybridizing fragments (Fig. 3A). This suggests either the presence of multiple *Pst*I and *Bam*HI restriction sites within introns spanned by the short SPT-1 cDNA probe or, more likely, the detection of

multiple related genes. The single hybridizing fragment that remains following stringent washing [last wash 0.05× SSC/0.1% (wt/vol) SDS] most likely represents a gene that is completely homologous to the cDNA probe used, while the other hybridizing fragments probably correspond to closely related genes in the DBA/2 mouse genome from which the probe is selectively displaced at very high washing stringencies. This is supported by the detection of the same group of three fragments in *Pst*I, *Hind*III, or *Bam*HI DBA/2 genomic DNA digests, or combinations thereof, probed with short oligonucleotides corresponding to the common 5'- or 3'-regions of the three SPT cDNA sequences (Fig. 3B).

*The genes encoding the mouse SPT proteins are closely linked and located on mouse chromosome 17.* Partial restriction digestion and Southern analysis of several overlapping lambda genomic clones, spanning ~50 kb of the mouse genome, revealed the presence of three distinct SPT-1 cDNA-hybridizing sequences (Fig. 4A), suggesting that the genes encoding the 16- to 17-kDa SPT proteins are closely linked in the mouse genome. Subcloning and sequencing of the SPT-1-hybridizing *Bam*HI fragment from lambda phage clone 3 (Fig. 4A) revealed that it contained three short introns separating four exons, the first of which is untranslated, with an open reading frame commencing in the second exon (data not shown). The predicted transcribed sequence corresponded to the SPT-2 cDNA sequence and is preceded by a recognizable transcription initiation motif and TATA box.

Linkage of the three genes predicted to encode the 16- to 17-kDa SPT proteins was confirmed for DBA/2 and C57BL/6 mouse strains, using pulsed-field gel electrophoresis, which showed that in both of these strains the three genes reside within a single *Bss*HIII fragment of ~150 kb (Fig. 4B, lanes 2 and 3) and in the case of strains DBA/2 only, an ~350-kb *Not*I fragment (Fig. 4B, lane 5). Meanwhile, the genes in C57BL/6 DNA could be separated by *Hpa*I digestion, yielding a doublet at ~25 kb and an additional smaller fragment of ~14 kb (Fig. 4B, lane 4) and *Not*I digestion, resulting in an ~400-kb fragment and a doublet at ~350 kb (Fig. 4B, lane 6).

RFLP analysis identified a *Pvu*II RFLP between the inbred strains C57BL/6J and DBA/2J (Fig. 5A, compare lanes 1 and 2), which allowed two of the *SPT* genes to be mapped on the mouse genome using BXD recombinant inbred lines, by comparison with known markers (45) (the third gene was invariable between the two progenitor strains and therefore could not be mapped by this strategy). Of 24 lines screened, only one recombinant strain (BXD 25) was observed (Table 2), placing the RFLP identified by the SPT-1 probe distal to the two markers T66D and RP17 and proximal to the marker HBA-PS4 in the T-complex region of mouse chromosome 17. This assignment was confirmed by searching of the mouse genome, using SPT-1 cDNA as query sequence, which demonstrated the presence of three closely related genes linked along 40 kbp of the mouse chromosome 17 genomic sequence (AC110262.12, *Mus musculus* clone RP23–291L10; Fig. 5B). The four exon-three intron structure of all three SPT protein genes was confirmed by comparison of genomic sequences with cDNA clones SPT-1 and 2, which also allowed two of the genes to be numbered on the basis of cDNA sequence similarity (Fig. 5B). An additional cDNA sequence variant, termed SPT-3 (Fig. 2A), was then deduced from the third unassigned *SPT* gene homolog located at co-ordinate 132,568–134,549, assuming conservation of intron-exon boundaries. Mapping of *Pst*I,

<sup>1</sup> The nucleotide sequences reported in this paper have been deposited in the DDBJ/GenBank/EBI Data Bank with accession numbers DQ237936 (SPT-1), DQ237937 (SPT-2). These sequences have been scanned against the DDBJ/GenBank/EBI Data Bank, and relevant sequences with significant relatedness are identified in the text with their accession numbers.

```

A  SPT-1  AGACACCCAAGGCGAATCCTCTCAGCCAGACTTTGAGATGTTCCAGCTGGAGGCCATGCT 60
      SPT-2  AGACACCCAAGGCGAATCCTCTCAGCCAGACTTTGAGATGTTCCAGCTGGAGGCCATGCT 60
      SPT-3  AGACACCCAAGGCGAGTCTCTCAGCCAGACTTTGAGATGTTCCAGCTGGAGGCCATGCT 60
      *****

      SPT-1  GCCCTTACTGATTCTTGCCTTCTTGGGGACCCCCACGGTTTTGACACAAGATTATCATGG 120
      SPT-2  GCCCTTACTGATTCTTGCCTTCTTGGGGACCCCCACGGTTTTGACACAAGATTATCATGG 120
      SPT-3  GCCCTTACTGATTCTTGCCTTCTTGGGGACCCCCACGGTTTTGACACAAGATTATCATGG 120
      * *****

      SPT-1  TCCAGAAGTTGAAAACATTCCCTGCACTAGTGCACCTGAAGGAAAAATATAACAAGTAT 180
      SPT-2  TCCAGAAGTTGAAAACATTCCCTGCACTAGTGCACCTGAAGGAAAAATATAAAGTAT 180
      SPT-3  TCCAGAAGTTGAAAACATTCCCTGCACTAGCACCCTGAAGAAAAATATAACAAGTAT 180
      *****

      SPT-1  CCGGGTATTTCTTCAGGGTAGATCGATTGTTGGAATTCAGTTTAACTACAACAATGA--- 237
      SPT-2  CCGGGTATTTCTTCAGGGTAGATCGATTGTTGGAATTCAGTTCACCTATAATGACAACGA 240
      SPT-3  CCGGGTATTTCTTCAGGCTAGATCGATTGTTGGAATTCAGTTCAACTATGATGACAA--- 237
      *****

      SPT-1  GGACGGTCAAGTGATGGCTCTACTGCTGGGAAGGTGATGGTAGCAAGACTGAATAACGA 297
      SPT-2  GGACGGTCAAGTCTATGGTCTACTGCGCGAAGGTGATGGTAGCCAGACTGTCTAAGGA 300
      SPT-3  GGACGGTCAAGTGATGGCTCAACTGCGCGAAGGAAAGTGGTAGCCAGACTGTCTAAGGA 297
      *****

      SPT-1  AGAGTCCATCATTGCGGCCGAGGGCACATACTCTCCTTCGGCTTTAACTCAGATAATCTT 357
      SPT-2  AGAGCACATCATTGGGGTCGAAGGCACATATAATCCTCTGCTTTAACTCAGATAATCTT 360
      SPT-3  AGAGCACATCACTGGGTCGAAGGCTCATACTCTCCTCTGCTTTAACTCAGATAGTCTT 357
      **** *****

      SPT-1  TACGACCAACCAGCCTCGCCAGCTAAATGGTGGGGTATTATGTTGGGCAGTTCTGAGTACTC 417
      SPT-2  TACGACCAACCAGCCTCGCCAGCTAAATGGTGGGGTATCACGTGGGTAATTATCAGTACTC 420
      SPT-3  TACAACCACCCAGCCGCGCAGCTAAATGGTGGGGTATCACGTGGGTATTTATCAGTACTC 417
      *** *****

      SPT-1  TAGCTTCCCTGACGATCCAGCCATGTGCTCAAGGGAGCTTGTGTCTCTTGGAGGGCAGG 477
      SPT-2  TAGTACCCTGATGATCCAGCCTTTGTGCTCAAGGGAGCTTGTGTCTCTTGGAGGGCAGG 480
      SPT-3  TAGTACCCTGATGATCCAGCTTTGTGCTCAAGGGAGCTTGTGTCTCTTGGAGGGCAGG 477
      *****

      SPT-1  TGGCATAAAGAGTATCTTGTGTTTGTGGGGAACTGAAAACAGTTCATGTGTGAAAATATGG 537
      SPT-2  TGGCATAAAGAGTATCTTGTGTTTGTGGGGAACTGAAAACAGTTCATGTGTGAAAATATGG 540
      SPT-3  TGGCATAAAGAGTATCTTGTGTTTGTGGGGAACTGAAAACAGTTCATGTGTGAAAATATGG 537
      *****

      SPT-1  CCACTCAGGCTGAAGTGGTCTGTGTCAGGTACTTCCAAACTGAACCCATGAATAAATACA 597
      SPT-2  CCACTCAGGCTGAAGTGGTCTGTGTCAGGTACTTCCAAACTGAACCCATGAATAAATACA 600
      SPT-3  CCACTCAGGCTGAAGTGGTCTGTGTCAGGTACTTCCAAACTGAACCCATGAATAAATACA 597
      *****

      SPT-1  ACATCTGCATG 608
      SPT-2  ACATGTGCATG 611
      SPT-3  ACATCTGCATG 608
      **** *****

B  SPT-1  MFQLEAMLPLLILAFGLTPTVLTQDYHGPEVGKHSCTSAPEGKNIITSIRVFLQGRSIVGI 60
      SPT-2  MFQLEAMLPLLILAFGLTPTVLTQDYHGPEVGKHSCTSAPEGKNIITSIRVFLQGRSIVGI 60
      SPT-3  MFQLEAMLALLILAFGLTPTVLTQDYHGPEVGKHSCTSTPEEKNIITSIRVFLQARSIVGI 60
      *****

      SPT-1  QFNYN-NEDGQVYGSTAGKVMVARLNNEESIIAAEGTYSPSALTQIIIFTNQPRQLMVGY 119
      SPT-2  QLTYNDNEDGQVYGSTAGKVMVARLSKEEHIIGVGETYNPSALTQIIIFTNQPRQLMVGY 120
      SPT-3  QFNVD-DKDGQVYGSTAGKEMVARLSKEEHITGVEGYSYPSALTQIVFTTTPRQLMVGY 119
      *.:*.:*:*****.***:*..*:*.******:***.*****

      SPT-1  YVGSSEYSSFDDPSHVLKGCACVSWRAGGIKSIILFLWGTEENSSCVKYHSG 170
      SPT-2  HVGNYQYSSYPDDPSLVLKGCACVSWRAGGIKSIILFLWGSEENSSCVKYHSG 171
      SPT-3  HVGIYQYSSYPDDPSFVLKGCACVSWRAGGIKSIILFLWGSEENSSCVKYHSG 170
      :** :**:* ****.*****:*****
  
```

Fig. 2. Comparison of mouse sublingual-pancreatic gland-tongue protein (SPT) cDNA and deduced amino acid sequences. *A*: CLUSTAL W alignment of sublingual cDNA clone class 1 and 2 (SPT-1 and 2) nucleotide sequences and an additional homologous sequence, termed SPT-3, deduced from a 3rd *SPT* gene homolog located on mouse chromosome 17 (AC110262.12 mouse genomic clone co-ordinate 132,568–134,549; see text for further details), using default parameters. Nucleotide identities are indicated by an asterisk. Predicted ATG “start” and TGA “stop” codons are shaded. *B*: comparison of deduced amino acid sequences of SPT-1, 2, and 3. Deduced amino acid sequences were aligned using CLUSTAL W with default parameters. Amino acid identities in all sequences in the alignment are indicated by an asterisk, while conserved substitutions are shown by a colon, and semiconserved substitutions by a dot. Putative signal sequences (SPT-1 amino acids 1–23), Jacalin-like lectin domain (SPT-1 amino acids 44–133), and glycosylation sites (SPT-1 amino acids 161–163) are shaded.



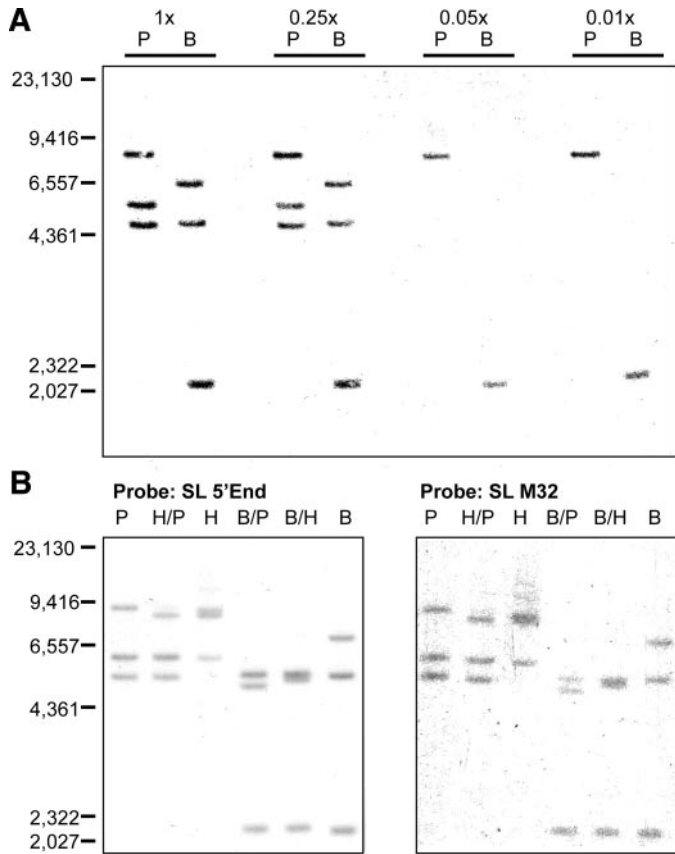


Fig. 3. Southern blotting analysis indicates the presence of 3 closely related *SPT* genes in the mouse genome. *A*: 10  $\mu$ g of DBA/2 mouse tail DNA were digested to completion with either *Pst*I (P) or *Bam*HI (B) and subjected to Southern blotting using SPT-54, a short centrally located 90-bp cDNA clone corresponding to bp 291–380 of the SPT-1 nucleotide sequence, as probe. Membranes were washed with SSC [20 $\times$  SSC = 3 M sodium chloride/0.3 M sodium citrate, pH 7.0]/0.1% (wt/vol) SDS] concentrations as indicated above the lanes on the blot at 65°C. *B*: 10  $\mu$ g of tail DNA from individual DBA/2 mice, digested to completion with either *Bam*HI (B), *Bam*HI/*Hind*III (B/H), *Bam*HI/*Pst*I (B/P), *Hind*III (H), *Hind*III/*Pst*I (H/P) or *Pst*I (P), were subjected to Southern blotting and hybridized to oligonucleotide probes corresponding to the 5'-end (probe SPT 5'-end = bp 1–20 of the SPT-2 nucleotide sequence) or the 3'-end (probe SPT-32 = bp 533–601 of the SPT-2 nucleotide sequence) of the sublingual cDNA sequences. Membranes were washed last with 1 $\times$  SSC/0.1% (wt/vol) SDS at 65°C. Both of these probes correspond to regions of sequence identity for SPT-1, SPT-2, and SPT-3 cDNA sequences. *Eco*RI/*Hind*III lambda DNA standard size markers (bp) are indicated at the left.

*Bam*HI, *Hpa*I, *Bss*HI, and *Not*I restriction sites within the 40-kbp genomic sequence region containing the *SPT* genes gave predicted *SPT* protein gene-bearing fragment sizes that, in nearly all cases, corresponded to those derived from the Southern blotting and pulsed-field gel analyses, reported in Figs. 3 and 4. Presumably, the few differences can be reconciled by strain-dependent sequence variation between the *M. musculus* clone RP23–291L10 DNA used to construct the chromosome 17 genomic sequence entry in the mouse genome database and the mouse strains employed in this study.

Database mining of the mouse genome identifies *SPT* proteins as members of the CSP-1/Dcpp protein family. BLAST searching of the GenBank nonredundant database revealed sequence similarities between the *SPT* cDNAs and several database entries for the CSP-1/Dcpp salivary protein family, including just a single base difference between DBA/2 SPT-1

and the Swiss Webster Dcpp cDNA sequence (accession no. S76879) (7), suggesting that these cDNA clones represent the same gene product. SPT-2 and SPT-3 cDNA sequences differ by 7% and 8%, respectively, from the Swiss Webster Dcpp cDNA sequence, suggesting that they correspond to distinct but closely related gene products. Furthermore, there were several additional close matches to IMAGE database sequences from mouse salivary gland and mammary tumor cDNA libraries, generated by the Mammalian Gene Collection Program Team (accession nos. BC045522, BC024627, BC090399, BC005655, BC059096) (23). A dbEST sequence search identified ESTs corresponding to all four exons of each of the three sublingual protein genes, suggesting that all of the

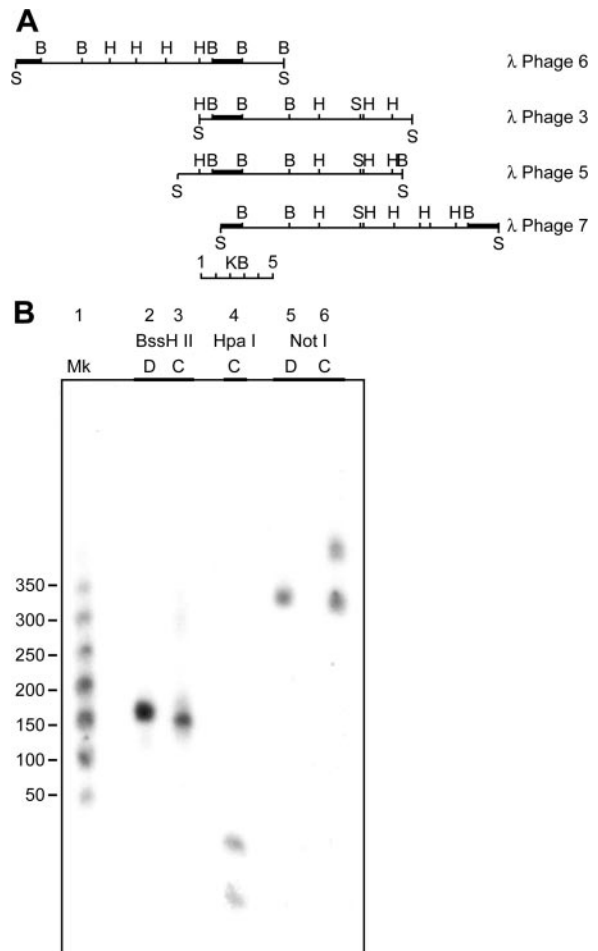


Fig. 4. Genes corresponding to the 3 *SPT* proteins are closely linked in the mouse genome. *A*: DNAs from several lambda genomic clones, selected by plaque hybridization using a SPT-1 partial cDNA clone (SPT-42 corresponding to bp 99–469 of the SPT-1 nucleotide sequence) as probe, were subjected to partial restriction digestion mapping, using either *Bam*HI (B), *Hind*III (H), *Sal*I (S), or combinations thereof, and partial restriction maps were assembled for each clone. This then allowed lambda clones 3, 5, 6, and 7 to be aligned in an overlapping array. The location of *SPT* cDNA-hybridizing fragments (shown as filled boxes) were determined for individual clones by Southern hybridization of complete lambda clone digests, using oligonucleotides corresponding to the 5'-end and 3' of the SPT-2 cDNA sequence, as described previously. *B*: DBA/2 DNA (*D*: lanes 2 and 5) or C57BL/6 DNA (*C*: lanes 3, 4, and 6) was digested with *Bss*HI, *Hpa*I, and *Not*I rare-cutting restriction enzymes, separated by pulsed-field electrophoresis, Southern blotted, and hybridized with the complete SPT-2 cDNA insert (bp 1–611). Lane 1, 50-kb DNA ladder marker (Mk) track.

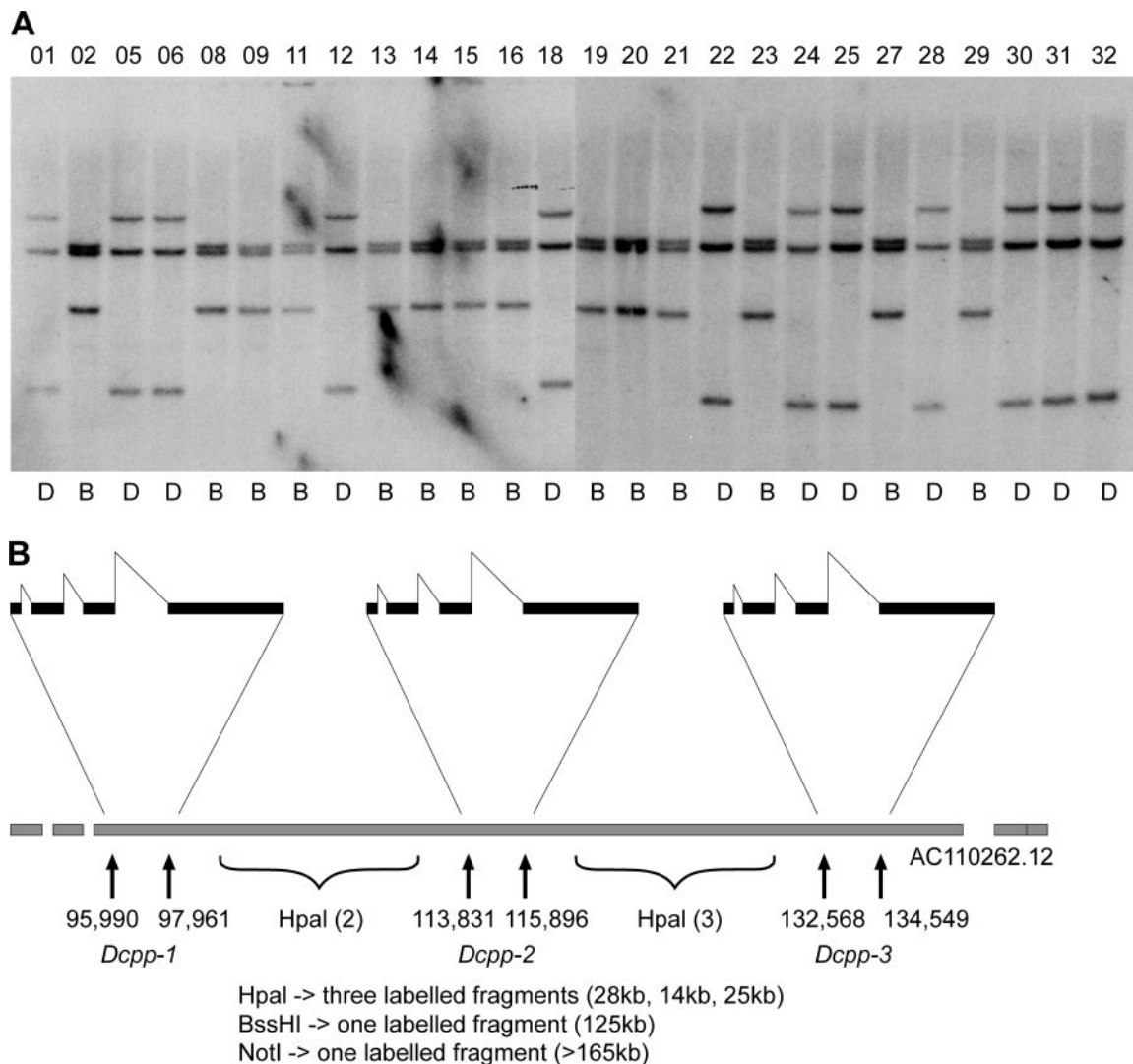


Fig. 5. Segregation of alleles for *SPT* genes and other chromosome 17 loci among BXD recombinant inbred (RI) strains. **A**: alleles for the *SPT* genes were determined by Southern blotting of *PvuII*-digested DNA from BXD RI strains, using an *SPT-1* partial cDNA clone (*SPT-42* corresponding to bp 99–469 of the *SPT-1* nucleotide sequence) as probe. Numbers represent the BXD RI strain number analyzed, whereas B and D are generic symbols for inherited alleles from C57BL/6J and DBA/2J, respectively, assigned on the basis of *PvuII* restriction pattern. **B**: location of *SPT* gene locus on *Mus musculus* chromosome 17 (AC110262.12, *M. musculus* clone RP23–291L10) detected by BLAST screening of the mouse genome with the *SPT-2* cDNA sequence. Intron-exon structure was determined by comparing *SPT* cDNA sequences with their genomic counterparts. Numbers in brackets indicate the number of *HpaI* restriction sites in the intergenic DNA, and the sizes of the *HpaI*, *BssHI*, and *NotI* sublingual protein gene-bearing fragments predicted from the *M. musculus* chromosome 17 sequence.

genes are transcriptionally active. This included reports of restricted EST expression patterns in late-gestational mouse embryo, probably reflecting maturation of the salivary glands and in adult skin, possibly suggesting expression in sebaceous or other glandular structures. Interrogation of microarray data associated with the GenBank sequence entry for P20 sublingual gland *Dcpp* (7) indicates possible expression in tissues other than salivary glands, including lung, liver (long oligo arrays), and brain (mouse 9k arrays, UCLA). This does not accord, however, with the expression data obtained in this study and presumably represents low-level expression below the threshold of sensitivity of Northern blotting. Nevertheless, preliminary QRT-PCR data from this laboratory indicate that *Dcpp* mRNA is also transcribed at moderate levels in mouse prostate, trachea, and oviduct (S Campbell, MR Nicol and SD

Morley; unpublished observations), the latter observations being corroborated by recent studies in the literature (22, 33).

**Identification of a putative human *Dcpp* protein ortholog.** A BLAST search of the DNA Data Bank of Japan (DDBJ)/GenBank/European Bioinformatics Institute (EBI) Data Bank nucleotide and protein databases for potential mammalian *Dcpp* protein orthologs, using the *SPT-2* cDNA nucleotide and predicted amino acid sequences as query sequences and default search parameters, revealed significant homology only to rat CSP-1 (GenBank accession no. NM\_133622), a previously recognized member of the CSP-1/*Dcpp* protein multigene family (7). Rat CSP-1 shows 43% amino acid identity and 63% amino acid similarity with the predicted *SPT-2* amino acid sequence (Fig. 6A). Nucleotide homology is concentrated in the 5'-regions of the mouse and rat cDNA sequences corresponding to the NH<sub>2</sub>-terminal signal se-

Table 2. *SPT* gene BxD allele pattern among BXD recombinant inbred strains in the T region of mouse chromosome 17

Chr. 17 BXD RI Strain	0	0	0	0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3
Chr. 17	1	2	5	6	8	9	1	2	3	4	5	6	8	9	0	1	2	3	4	5	7	8	9	0	1	2	
1 LDS1	B	B	D	D	D	B	B	D	B	B	D	D	B	D	D	D	D	B	B	D	B	D	D	B	B	B	
							X																				
2 ODC-10	B	B	D	D	D	B	D	D	B	B	D	D	B	D	D	D	D	B	B	D	B	D	D	B	B	B	
	X	X			X		X				X	X		X	X	X			X	X			X	X	X	X	
3 T119	D	D	D	D	B	B	B	D	B	B	B	B	B	B	B	B	D	B	D	B	B	D	B	D	D	D	
		X										X															
4 T66D	D	B	D	D	B	B	B	D	B	B	B	B	D	B	B	B	D	B	D	B	B	D	B	D	D	D	
5 RP17	D	B	D	D	B	B	B	D	B	B	B	B	D	B	B	B	D	B	D	B	B	D	B	D	D	D	
<b>SL BxD</b>	<b>D</b>	<b>B</b>	<b>D</b>	<b>D</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>D</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>D</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>D</b>	<b>B</b>	<b>D</b>	<b>D</b>	<b>B</b>	<b>D</b>	<b>B</b>	<b>D</b>	<b>D</b>	<b>D</b>	
6 HBA-PS4	D	B	D	D	B	B	B	D	B	B	B	B	D	B	B	D	D	B	D	D	D	B	D	B	D	D	
						X	X					X										X					
7 H2S	D	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	D	B	D	D	
							X															X					
8 PIM-1	D	B	D	D	B	D	D	B	B	B	B	D	D	B	B	D	D	B	D	D	D	B	D	B	D	D	
						X			X								X	X				X					
9 UPG-1	D	B	D	D	B	B	D	B	D	B	B	D	D	B	B	B	B	B	D	D	D	D	B	D	D	D	
			X	X		X			X								X	X		X							
10 HGPT2	D	B	B	B	B	D	D	B	D	D	B	D	D	B	B	B	D	D	D	B	D	D	B	D	D	D	
					X		X				X	X		X	X				X				X	X			
11 U1-9	D	B	B	B	D	D	B	B	D	B	D	B	D	D	D	B	D	D	D	D	D	D	D	D	B	D	D

Previously determined distribution for T66D, RP17, and HBA-P54 alleles among BXD recombinant inbred (RI) strains in the T region of chromosome 17. *SPT* gene BxD allele pattern, determined by *Pvu*I1 restriction digestion (Fig. 6), is indicated in bold. A crossover between two loci is indicated by an X.

quence (mouse SPT-2 cDNA bp 38–103), which are 97% homologous, leading to almost identical mouse SPT and rat CSP-1 signal peptides, while the respective 3'-untranslated regions downstream of the stop codon (mouse SPT-2 cDNA bp 562–6011) show 92% sequence identity, with lesser but significant homology within the mature protein-coding sequence (data not shown). The initial failure to identify corresponding human sequences by BLAST searching was surprising but may owe something to the known rapid divergent evolution of salivary protein family members both within and between species (12, 30). However, protein sequence similarity searching identified a putative human SPT protein ortholog HRPE773 (GenBank AAK89380.1), which represents a secreted protein of unknown function (10). The predicted HRPE773 protein displays 31% amino acid identity and 45% amino acid similarity to the predicted SPT-2 amino acid sequence and in common with mouse Dcpp proteins, exhibits a classical secretory signal sequence, a putative JRL domain, and a potential glycosylation site (Fig. 6A). The corresponding human cDNA sequence, AY359021.1, also shows striking sequence similarity with the 5'-NH<sub>2</sub>-terminal signal sequence-encoding regions and 3'-untranslated regions of both the mouse Dcpp and rat CSP-1 cDNA sequences. In particular, the human cDNA AY359021.1, mouse Dcpp, and rat CSP-1 cDNA sequences share 25 out of 34 base identities in the 3'-untranslated region surrounding the polyadenylation signals (mouse SPT-2 cDNA bp 577–609). Furthermore, the corresponding human gene located on human chromosome 16 (AC005361.1, *Homo sapiens* chromosome 16, cosmid clone 352F10, base pairs 9479–7345), displays a similar structure to the mouse Dcpp genes, comprising four exons, separated by three introns, with the ATG translation start codon also being located in the second exon (Fig. 6B). Significantly, the gene for HRPE773 is located at 16p13.3 on the region of human chromosome 16 known to be syntenic with the T-complex region of mouse chromosome 17, demonstrated

in this study to contain the three Dcpp genes (Fig. 6C). Examination of gene expression data associated with IMAGE clone 3879368, which contains the human cDNA sequence AY359021.1, reveals maximal expression levels in pooled human salivary glands, lower expression in trachea and prostate, and otherwise no significant expression among the rest of a panel of 79 physiologically normal human tissues examined (large-scale analysis of the human transcriptome, GNF1b, accession: GDS594; Gene Expression Omnibus). This suggests that the putative human ortholog, HRPE773, may share a similar tissue expression profile to CSP-1/Dcpp family members. The three Dcpp genes on mouse chromosome 17, the CSP-1 gene on rat chromosome 10, and the human ortholog on human chromosome 16 all lie in protease and secretory protein gene-rich regions of their respective genomes, each being flanked by genes for an orthologous group of serine proteases, including protease, serine 21 (Prss 21 *M. musculus*; PRSS 21 *H. sapiens*); protease, serine 22 (Prss 22 *M. musculus*; PRSS 22 *H. sapiens*); and protease, serine 33 (Prss 33 *M. musculus*; PRSS 33 *H. sapiens*); and genes for proteins with other functions, e.g., transcription elongation factor B polypeptide 2 (TCEB2), and serine/arginine repetitive matrix 2 (SRRM2) (Fig. 6C). This is clearly consistent with the demonstration that the equivalent regions of rat chromosome 10, human chromosome 16, and mouse chromosome 17 are syntenic and strongly supports the notion that HRPE773 is the human ortholog of the CSP-1/Dcpp family.

## DISCUSSION

The present study describes a novel family of 16- to 17-kDa mouse salivary proteins, identified initially by their abundant expression in mouse sublingual glands, comprising one known and two additional novel members of the CSP-1/Dcpp salivary

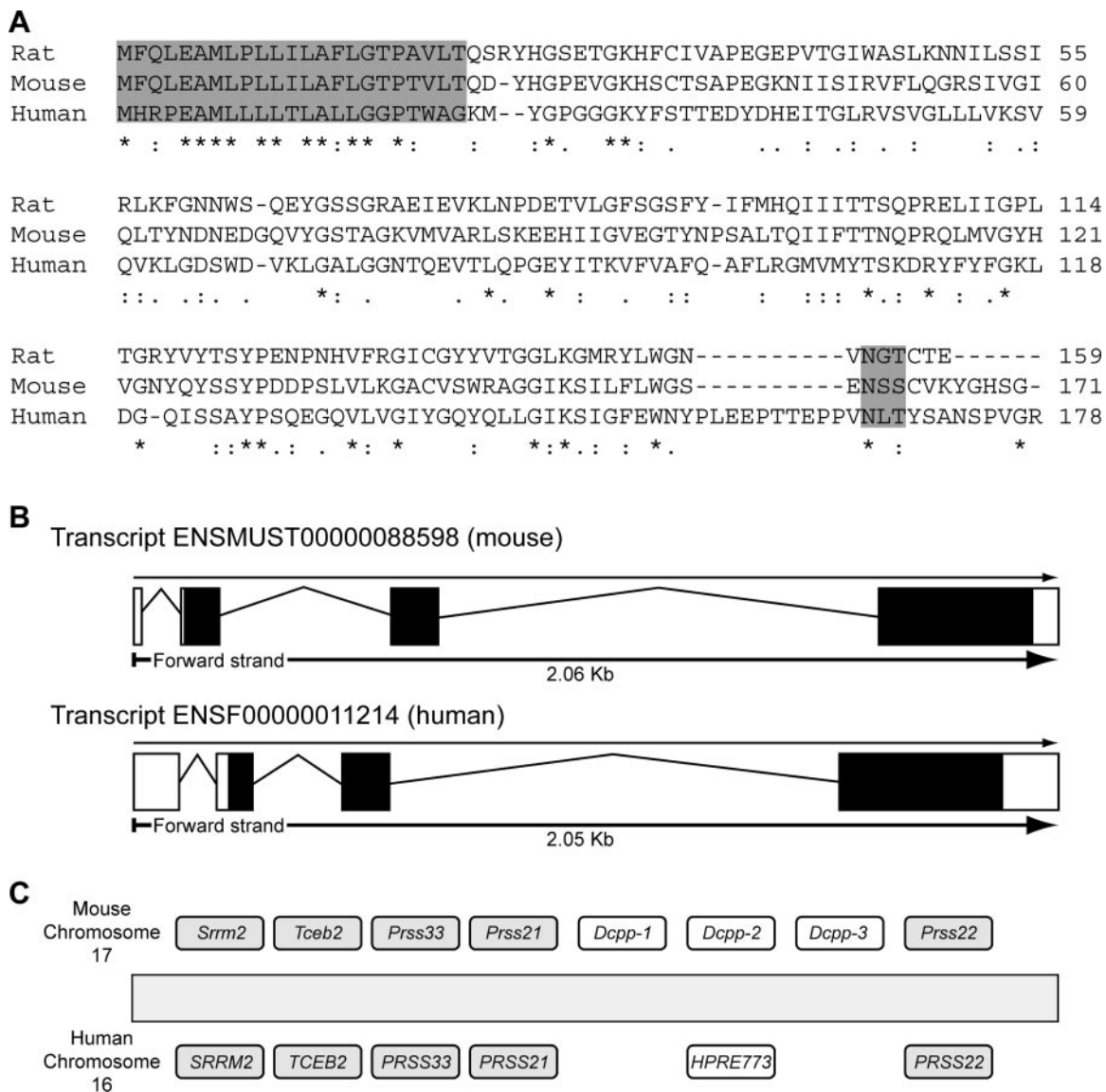


Fig. 6. Multiple sequence alignment and gene arrangements for CSP-1/Dcpp orthologs. **A**: mouse SPT-2, NP-5980306.1-rat (common salivary protein 1), and AAQ89380.1-human predicted amino acid sequences aligned by CLUSTAL W (1.82) multiple sequence alignment. Amino acid identities in all sequences in the alignment are indicated by an asterisk, whereas conserved substitutions are shown by a colon, and semiconserved substitutions by a dot. Putative signal peptides and glycosylation sites are shaded. **B**: mapping of mRNA transcripts for the mouse and putative human CSP-1/Dcpp orthologs on to their corresponding gene sequences reveals similar intron-exon structures. Boxes represent *Dcpp* exons, with black-shaded areas corresponding to open reading frames, while intervening intronic regions are represented by angled lines between exons. **C**: simplified comparison of the *SPT* locus in the T-complex region of mouse chromosome 17, with the location of the gene for the putative human CSP-1/Dcpp ortholog, HRPE773, on the syntenic region of human chromosome 16. Data are abstracted from Ensembl V38 annotation of the mouse and human genome builds, and named genes characterized in both mouse and human are indicated respectively by the Mouse Gene Informatics-approved gene symbol for mouse and HUGO Gene Nomenclature Committee-approved gene symbol for human. Note that detailed annotation of the syntenic region of rat chromosome 10 containing the gene for CSP-1 has yet to be included in the published rat genome compilation, thus precluding further analysis of CSP-1 gene structure here.

protein family. Secondly, this work establishes the previously unrecognized presence of three closely linked *Dcpp* genes, forming a locus, in the T-complex region of mouse chromosome 17. Finally, protein sequence similarity searching has identified HRPE773, a protein of previously unknown identity that is also expressed predominantly in human salivary tissues, as a putative human CSP-1/Dcpp ortholog. The similarities between HRPE773 and the *Dcpp* family are emphasized by the fact that the predicted mouse and human proteins both display classical NH<sub>2</sub>-terminal signal sequences, putative JRL domains, and potential N-linked glycosylation sites, suggesting that they may both be secreted into

sublingual saliva and thence into the oral cavity (10). The corresponding human gene, encoding HRPE773, displays a similar four exon-three intron structure to the mouse *Dcpp* genes and is located on the region of human chromosome 16 known to be syntenic with the T-complex region of mouse chromosome 17 and the region of rat chromosome 10 that contains the gene for rat CSP-1. Crucially, the three *Dcpp* genes on mouse chromosome 17 and the putative human ortholog on human chromosome 16 are flanked respectively by the mouse and human orthologs of several genes including a family of serine proteases (Prss 21, Prss 22, Prss 33 and PRSS 21, PRSS 22, PRSS 33) and also by genes for

TCEB2 and SRRM2. The presence of a series of orthologous genes located both proximally and distally to the genes for mouse Dcpp and human HRPE773 provides strong additional evidence that HRPE773 is the human ortholog of the CSP-1/Dcpp family.

A protein sequence similarity search was used successfully in this study to identify a putative human ortholog of the CSP-1/Dcpp family, which had not been detected previously. Surprisingly, an initial BLAST search with a *Dccp* query sequence failed to identify a corresponding sequence in the human genome. This is consistent, however, with the observation that salivary protein gene sequences show a characteristically high evolutionary divergence within coding regions, even between closely related species, sometimes to the point where molecular probes from orthologous genes of different mammalian species will not cross-hybridize in Southern blotting experiments, although protein sequence may show greater conservation (12, 30). Nevertheless, there is often a high degree of conservation of the NH<sub>2</sub>-terminal signal sequence coding sequences and 3'-untranslated regions of the corresponding cDNAs (30), perhaps indicating the presence of "optimal" motifs enabling high level expression and efficient secretion of these proteins into saliva. Precisely these homologies are observed between cDNAs for the mouse Dcpp proteins, rat CSP-1, and HRPE773, adding further weight to the identification of the latter as a human CSP-1/Dcpp ortholog.

In common with the mouse CSP-1/Dcpp family, HRPE773 is expressed principally in the salivary glands but also at lower levels in trachea and prostate. Tracheal expression may perhaps be accounted for by the hundreds of small glands that lubricate the tracheal surface epithelium, while the prostate is rich in secretory products. Preliminary data suggest that Dcpp is also expressed in mouse trachea and prostate, whereas both the human and mouse orthologs may be present in oviduct (S Campbell, MR Nicol, and SD Morley; unpublished observations). The similar tissue expression profiles displayed by HRPE773 and Dcpp emphasize their probable relatedness. Furthermore, conservation of six glycine residues between the putative human, mouse, and rat CSP-1/Dcpp family members suggests that these proteins may share similar secondary structures. Together, this evidence makes it highly probable that HRPE773 does represent a human counterpart of the CSP-1/Dcpp protein family.

The present study extends the observations both of Bekhor et al. (7), who reported only a single submandibular CSP-1/Dcpp mRNA class, and the Ensembl V38 (18) annotation of the mouse genome, which currently recognizes only a single Dcpp gene on mouse chromosome 17. Thus, three distinct CSP-1/Dcpp-related mRNAs are shown here to be expressed at high levels in the mouse sublingual gland and are encoded by three closely linked *Dcpp* genes located in the T-complex region of mouse chromosome 17. Sequence alignment with the BLAST algorithm revealed just a single base difference between DBA/2 SPT-1 and the Swiss Webster *Dccp* cDNA sequence (accession no. S76879) (7), suggesting that these cDNA clones represent the same gene product. We propose therefore that the gene corresponding to the SPT-1/Dcpp mRNA sequence should be allocated the systematic name *Dcpp-1* and that the genes corresponding to SPT-2 and SPT-3 mRNAs identified in this study should be termed *Dcpp-2* and *Dcpp-3*, respectively. This is convenient because this numerical assignment of the genes then corresponds to the positional annotation of individ-

ual genes on *M. musculus* chromosome 17 (AC110262.12, *M. musculus* clone RP23–291L10; see Fig. 5).

The presence of three linked CSP-1/Dcpp-related genes in the mouse contrasts with the human, as demonstrated in this study, and rat (11, 15), where apparently only a single CSP-1/Dcpp ortholog is present. Linkage of the mouse *Dcpp* genes on chromosome 17 was first demonstrated by us using classical Southern blotting and RFLP analyses combined with pulsed-field gel electrophoresis and then confirmed by bioinformatic interrogation of the mouse genome sequence. These classical approaches proved critical in establishing the presence of three closely linked *Dccp* genes in the mouse genome, because, as noted previously, published resources document the existence of only one mouse Dcpp protein species to date, while mouse genome sequencing and EST resources annotate just a single *Dcpp* gene, corresponding to SPT-1. Presumably previous bioinformatic database analyses have been confounded by the close linkage and high degree of similarity between the three mouse *Dcpp* genes and their corresponding proteins.

Some indicators of the function of the CSP-1/Dcpp family members may be deduced from their secretory nature and the presence of a putative JRL domain in both the mouse and human orthologs. Jacalin is the prototype of the JRL family of proteins, one of the seven major groups of lectin carbohydrate-binding proteins (or glycoproteins), found ubiquitously throughout the plant and animal kingdoms (34), and even include salivary snake venom toxins (14). The presence of a JRL domain in a protein is normally considered to be diagnostic for lectin activity, so it therefore seems possible that mammalian Dcpp proteins may play similar roles to plant lectins, for example in cell agglutination or by displaying antimicrobial activity in the mammalian oral cavity and respiratory and reproductive tracts. As for other salivary protein classes (12, 46), amino acid substitutions between the three SPT proteins are concentrated toward the middle of the predicted open reading frames, in the putative JRL domain. Consequently, mouse lineage-specific gene duplication, followed by sequence divergence, could provide a mechanism for evolution of carbohydrate binding specificities in response to species-specific requirements of the mouse, similar to the copy number variation and sequence polymorphisms observed, for example, in the genes of beta defensins (39).

Together, the observations reported in this study provide novel comparative functional genomic data for major salivary gland gene products, while identification of a human CSP-1/Dcpp ortholog provides a key tool toward the investigation of salivary protein function in the context of human oral health.

#### ACKNOWLEDGMENTS

We thank David Fettes for expert assistance with the automatic four-color fluorescent dye chemistry DNA sequencing and Corinne McLeod and Ted Pinner for assistance with the graphics. J. J. Mullins holds a Wellcome Trust Principal Research Fellowship.

#### GRANTS

We acknowledge funding from the UK Biotechnology and Biological Sciences Research Council, the National Institute of General Medical Sciences (Grant GM-30248), and the Wellcome Trust, including the Wellcome Trust Functional Genomics Initiative.

## REFERENCES

1. Abel KJ, Howles PN, Gross KW. DNA insertions distinguish the duplicated renin genes of DBA/2 and *M. Hortulanus* mice. *Mamm Genome* 2: 32–40, 1992.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 215: 403–410, 1990.
3. Ann DK, Carlson DM. The structure and organization of a proline-rich protein gene of a mouse multigene family. *J Biol Chem* 260: 15863–15872, 1985.
4. Ball WD, Mirels L, Hand AR. Psp and Smgb: a model for developmental and functional regulation in the rat major salivary glands. *Biochem Soc Trans* 31: 777–780, 2003.
5. Barka T. Biologically-active polypeptides in submandibular glands. *J Histochem Cytochem* 28: 836–859, 1980.
6. Baum BJ, Voutetakis A, Wang JH. Salivary glands: novel target sites for gene therapeutics. *Trends Mol Med* 10: 585–590, 2004.
7. Bekhor I, Wen Y, Shi ST, Hsieh CH, Denny PA, Denny PC. cDNA cloning, sequencing and *in-situ* localization of a transcript specific to both sublingual demilune cells and parotid intercalated duct cells in mouse salivary glands. *Arch Oral Biol* 39: 1011–1022, 1994.
8. Botts S, Jokinen M, Gaillard ET, Elwell M, Mann PC. Salivary, harderian and lacrimal glands. In: *Pathology of the Mouse* (1st ed.), edited by Maronpot RR. Vienna, IL: Cache River, 1999, p. 49–79.
9. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31: 3497–3500, 2003.
10. Clark HF, Gurney AL, Abaya E, Baker K, Baldwin D, Brush J, Chen J, Chow B, Chui C, Crowley C, Currell B, Deuel B, Dowd P, Eaton D, Foster J, Grimaldi C, Gu Q, Hass PE, Heldens S, Huang A, Kim HS, Klimowski L, Jin Y, Johnson S, Lee J, Lewis L, Liao D, Mark M, Robbie E, Sanchez C, Schoenfeld J, Seshagiri S, Simmons L, Singh J, Smith V, Stinson J, Vagts A, Vandlen R, Watanabe C, Wieand D, Woods K, Xie MH, Yansura D, Yi S, Yu G, Yuan J, Zhang M, Zhang Z, Goddard A, Wood WI, Godowski P, Gray A. The Secreted Protein Discovery Initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. *Genome Res* 13: 2265–2270, 2003.
11. Denny PC, Ball WD, Redman RS. Salivary glands: a paradigm for diversity of gland development. *Crit Rev Oral Biol Med* 8: 51–75, 1997.
12. Dickinson DP, Mirels L, Tabak LA, Gross KW. Rapid evolution of variants in a rodent multigene family encoding salivary proteins. *Mol Biol Evol* 6: 80–102, 1989.
13. Feinberg AP, Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132: 6–13, 1983.
14. Fry BG. From genome to “venome”: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res* 15: 403–420, 2005.
15. Girard LR, Castle AM, Hand AR, Castle JD, Mirels L. Characterization of common salivary protein-1, a product of rat submandibular, sublingual, and parotid glands. *J Biol Chem* 268: 26592–26601, 1993.
16. Green MC. *Catalogue of Mutant Genes and Polymorphic Loci*. Stuttgart, Germany: Gustav Fischer, 1981.
17. Gubler U, Hoffman BJ. A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263–269, 1983.
18. Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinski F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Birney E. Ensembl 2005. *Nucleic Acids Res* 33: D447–D453, 2005.
19. Kock K, Blaker M, Schmale H. Postnatal development of Von Ebner’s glands - accumulation of a protein of the lipocalin superfamily in taste papillae of rat tongue. *Cell Tissue Res* 267: 313–320, 1992.
20. Kock K, Morley SD, Mullins JJ, Schmale H. Denatonium bitter-tasting among transgenic mice expressing rat Von Ebner’s gland protein. *Physiol Behav* 56: 1173–1177, 1994.
21. Kozak M. An analysis of 5’-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* 15: 8125–8148, 1987.
22. Lee KF, Xu JS, Lee YL, Yeung WS. Demilune cell and parotid protein from murine oviductal epithelium stimulates preimplantation embryo development. *Endocrinology* 47: 79–87, 2006.
23. Mammalian Gene Collection Program Team. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci USA* 99: 16899–16903, 2002.
24. Mandel ID. Sialochemistry in diseases and clinical situations affecting salivary-glands. *Crit Rev Clin Lab Sci* 12: 321–366, 1980.
25. Mandel ID. The functions of saliva. *J Dent Res* 66: 623–627, 1987.
26. Maniatis T, Fritsch EF, Sambrook J. *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1989.
27. Mason AJ, Evans BA, Cox DR, Shine J, Richards RI. Structure of mouse kallikrein gene family suggests a role in specific processing of biologically active peptides. *Nature* 303: 300–307, 1983.
28. Mikkelsen TR, Brandt J, Larsen HJ, Larsen BB, Poulsen K, Ingerslev J, Din N, Hjorth JP. Tissue-specific expression in the salivary glands of transgenic mice. *Nucleic Acids Res* 20: 2249–2255, 1992.
29. Milne RW, Dawes C. Relative contributions of different salivary glands to blood-group activity of whole saliva in humans. *Vox Sang* 25: 298–307, 1973.
30. Mirels L, Ball WD. Neonatal rat submandibular-gland protein SMG-A and parotid secretory protein are alternatively regulated members of a salivary protein multigene family. *J Biol Chem* 267: 2679–2687, 1992.
31. Mirels L, Bedi GS, Dickinson DP, Gross KW, Tabak LA. Molecular characterization of glutamic-acid glutamine-rich secretory proteins from rat submandibular glands. *J Biol Chem* 262: 7289–7297, 1987.
32. Mullins JJ, Burt DW, Windass JD, McTurk P, George H, Brammar WJ. Molecular cloning of two distinct renin genes from the DBA/2 mouse. *EMBO J* 1: 1461–1466, 1982.
33. Okubo T, Hogan BLM. Hyperactive Wnt signalling changes the developmental potential of embryonic lung endoderm. *J Biol* 3: 1–17, 2004.
34. Raval S, Gowda SB, Singh DD, Chandra NR. A database analysis of jacalin-like lectins: sequence- structure-function relationships. *Glycobiology* 14: 1247–1263, 2004.
35. Romer AS, Parsons TS. *The Vertebrate Body*. Philadelphia, PA: W. B. Saunders, 1977.
36. Rougeot C, Rosinski-Chupin I, Mathison R, Rougeon F. Rodent submandibular gland peptide hormones and other biologically active peptides. *Peptides* 21: 443–455, 2000.
37. Roukema PA, Oderkerk CH, Salkinga-Salonen MS. Murine sublingual and submandibular mucins - their isolation and characterization. *Biochim Biophys Acta* 428: 432–440, 1976.
38. Sasavage NL, Nilson JH, Horowitz S, Rottman FM. Nucleotide sequence of bovine prolactin messenger RNA - evidence for sequence polymorphism. *J Biol Chem* 257: 678–681, 1982.
39. Semple CA, Maxwell A, Gautier P, Kilanowski FM, Eastwood H, Barran PE, Dorin JR. The complexity of selection at the major primate beta-defensin locus. *BMC Evol Biol* 5: 32–45, 2005.
40. Sherwood LM, Burstein Y, Schechter I. Primary structure of the NH<sub>2</sub>-terminal extra piece of the precursor to human placental-lactogen. *Proc Natl Acad Sci USA* 76: 3819–3823, 1979.
41. Shine J, Seeburg PH, Martial JA, Baxter JD, Goodman HM. Construction and analysis of recombinant DNA for human chorionic somatomammotropin. *Nature* 270: 494–499, 1977.
42. Silver J. Confidence limits for estimates of gene linkage based on analysis of recombinant inbred strains. *J Hered* 76: 436–440, 1985.
43. Strous GJ, Dekker J. Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol* 27: 57–92, 1992.
44. Tabak LA. In defense of the oral cavity - structure, biosynthesis and function of salivary mucins. *Annu Rev Physiol* 57: 547–564, 1995.
45. Taylor BA. *Recombinant Inbred Strains*. Stuttgart, Germany: Gustav Fischer, 1981.
46. Tronikleroux D, Senoralepose M, Rougeon F. 3 novel Smr1-related cDNAs characterized in the submaxillary gland of mice show extensive evolutionary divergence in the protein-coding region. *Gene* 142: 175–182, 1994.
47. Voutetakis A, Bossis I, Kok MR, Zhang W, Wang J, Cotrim AP, Zheng C, Chiorini JA, Nieman LK, Baum BJ. Salivary glands as a potential gene transfer target for gene therapeutics of some monogenic endocrine disorders. *J Endocrinol* 185: 363–372, 2005.
48. Windass JD, Mullins JJ, Beecroft LJ, George H, Meacock PA, Williams BRG, Brammar WJ. Molecular cloning of cDNAs from androgen-independent messenger RNA species of DBA/2 mouse submaxillary glands. *Nucleic Acids Res* 12: 1361–1376, 1984.