Characterization of ostrich (*Struthio camelus*) β-microseminoprotein (MSP): Identification of homologous sequences in EST databases and analysis of their evolution during speciation

CLAUDE LAZURE,[1] MICHÈLE VILLEMURE,[1] DANY GAUTHIER,[1] RYNO J. NAUDÉ,[2] AND MAJAMBU MBIKAY[3]

[1]Laboratory of Structure and Metabolism of Neuropeptides, Institut de recherches cliniques de Montréal (affiliated with the University of Montréal), Montréal, Québec H2W 1R7, Canada
[2]Department of Biochemistry and Microbiology, University of Port Elizabeth, Port Elizabeth 6000, Republic of South Africa
[3]Ottawa Health Research Institute at Ottawa Hospital (University of Ottawa), Ottawa, Ontario K1Y 4K9, Canada

Abstract

β-Microseminoprotein, alternatively called prostatic secretory protein of 94 amino acids, is a hydrophilic, unglycosylated, small protein rich in conserved half-cystine residues. Originally found in human seminal plasma and prostatic fluids, its presence was later shown in numerous secretions and its homologs were described in many vertebrate species. These studies showed that this protein had rapidly evolved, but they failed to unambiguously identify its biological role. Here, we show that a protein isolated from ostrich pituitary gland is closely related to a similar one isolated from chicken serum and that the two are structurally related to the mammalian β-microseminoprotein. The complete 90–amino acid sequence of the ostrich molecule was established through a combination of automated Edman degradation and matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometric procedures, including postsource decay (PSD) and ladder sequencing analyses. This study documents for the first time that β-microseminoprotein is present in aves. It is also the first report of a C-terminal amidated form for a member of this protein family and the first in which the disulfide linkages are established. Database searches using the herein-described amino acid sequence allowed identification of related proteins in numerous species such as cow, African clawed frog, zebrafish, and Japanese flounder. These small proteins show a strikingly high rate of amino acid substitutions, especially across phyla boundaries. Noticeably, no β-microseminoprotein–related gene could be found in the recently completed fruit fly genome, indicating that if such a gene exists in arthropods, it must have extensively diverged from the vertebrate ones.

Keywords: Microseminoprotein; prostate secretory protein; protein evolution; *Struthio camelus*; protein sequencing; disulfide bonds

β-Microseminoprotein (MSP), also called the prostate secretory protein of 94 amino acids (PSP$_{94}$), is a small unglycosylated protein derived from a 114–amino acid preprotein. It was initially reported as an abundant protein of the human seminal plasma (1 mg/mL) showing inhibin-like activity, that is, it is capable of inhibiting pituitary secretion of follicle-stimulating hormone (Thakur et al. 1981). However, other studies failed to confirm this activity (Kohan et al. 1986; Gordon et al. 1987). A variety of roles and potential uses of this protein, some not mutually exclusive, have been

proposed (Table 1), resulting in its designation by other names such as immunoglobulin-binding factor (Liang et al. 1991) and prostatic inhibin peptide (Garde et al. 1993b). MSP is primarily expressed in the prostate and is found in prostatic secretions of various mammals. However, immunohistochemical and Northern blot analyses have shown its presence in other tissues and biological fluids. Indeed, studies have shown it to be present in nonreproductive organs such as the gastrointestinal tract (especially in gastric mucosa, where it is expressed in mucin-producing cells [M-cells] and endocrine-type cells [E-cells]; Weiber et al. 1999; Ulvsbäck et al. 1989) and the respiratory tract (including tracheal, bronchial, and lung tissues; Ulvsbäck et al. 1989; Weiber et al. 1990). Even within a single species, the MSP sequence has been observed in cDNA libraries from a variety of tissues. For example, the UniGene site at National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/UniGene) lists 398 expressed sequence tag (EST) sequences of human MSP arising from expression in blood, brain, lung, testis, and prostate. Finally, MSP was shown not to be restricted to male tissues because the protein was isolated from porcine ovarian lutein cells (Tanaka et al. 1995), and its gene expression was recently reported in female reproductive tissues, breast, and endometrial cancer cell lines (Baijal-Gupta et al. 2000). Its single gene in human has been mapped on chromosome 10q11.2 (Ulvsbäck et al. 1991; Sasaki et al. 1996), outside the previously identified LOH (loss of heterozygosity) regions in prostate can-

cer cells, thereby casting a doubt on its putative role as a tumor suppressor gene. However, in one report, PSP$_{94}$ has been shown to inhibit the growth of a human prostatic carcinoma cell line in culture and in athymic mice (Garde et al. 1999). Also, an abstract at the most recent European Peptide Symposium described results showing that synthetic deca-peptides corresponding to the extreme MSP C-terminal region show growth inhibitory properties on prostate tumor cells (Kele et al. 2000).

Despite the lack of a well-established biological function, MSP has attracted scientific interest as a unique example of a rapidly evolving protein (Nolet et al. 1991; Fernlund et al. 1996; Mäkinen et al. 1999). Indeed, the primary structure of MSP shows a remarkably low level of conservation in amino acids among the species studied, often resulting in great variation of physico-chemical properties. Such a great number of amino acid substitutions has for years rendered the identification of MSP using immunological or hybridization techniques difficult in more distant species. Indeed, comparison of known sequences from human (Mbikay et al. 1987), rhesus monkey (Nolet et al. 1991), baboon (Xuan et al. 1997), cotton-top tamarin (Mäkinen et al. 1999), pig (Fernlund et al. 1994; Tanaka et al. 1995), rat (Fernlund et al. 1996), and mice (Xuan et al. 1999) reveals that apart from the 10 completely conserved half-cystine residues, these proteins share very few other conserved residues. This is further confirmed in sequence identities. For example, between the human sequence and that of rhesus monkey,

Table 1. *Putative biological roles and uses of β-microseminoprotein (MSP) (1981–2000)*

Proposed biological role and/or use	References
Inhibitor of FSH secretion (inhibin)[a]	Thakur et al. 1981
Sperm-coating protein	Tsuda et al. 1982
Barrier to interspecies fertility[b]	Brooks et al. 1986
	Doctor et al. 1986
Marker of prostate gland hyperplasy and neoplasy[c]	Dubé et al. 1987a
	Abrahamsson et al. 1988
Factor modifying mucus properties[d]	Weiber et al. 1990
Protective function for mucus (enzyme inhibitor and/or antibacterial agent)[d]	Weiber et al. 1990
Immunoglobulin Binding Factor (IgBF)[e]	Liang et al. 1991
	Garde et al. 1993a,b
Hormone-refractory prostate tumor growth inhibitor[f]	Lokeshwar et al. 1993
	Mundel and Sheth 1993
Sperm motility inhibiting factor (Na$^+$, K$^+$-ATPase inhibitor)	Chao et al. 1996
Binding to proteins found on LNCaP and PC-3 cells	Yang et al. 1998a,b
Tumor marker for gastric carcinoid disease	Weiber et al. 1999

[a] This role could not be confirmed through in vitro biological assays (Kohan et al. 1986; Gordon et al. 1987).
[b] This role though also ascribed to MSP (Mbikay et al. 1988) was associated vide infra, with molecules of evolutionary divergent structures found in urogenital tissues.
[c] The usefulness of MSP as a cancer marker is disputed, considering the demonstrated lack of correlation (Von der Kammer et al. 1993) and its demonstrated presence in numerous nonurogenital tissues (Weiber et al. 1990).
[d] Both of these roles are still hypothetical as no data to substantiate them is currently available.
[e] Involvement of MSP as a member of an Ig-binding protein family was further proposed (Kamada et al. 1998).
[f] Further studies showed that MSP appears able to induce apoptosis of prostate tumor cancer cells (Garde et al. 1999).

porcine, and rat, the identity is of 79%, 51%, and 45%, respectively. Furthermore, analysis of the cloned cDNA and genomic sequences also revealed that most amino acid substitutions did not occur following the classical "wobble hypothesis" by mutations at the third base of codons, but rather by mutations at the first and second bases (Fernlund et al. 1996; Mäkinen et al. 1999). In addition, comparative analysis of the complete genomic sequences indicates that exons evolve at double the rate of introns (Mäkinen et al. 1999). In some species, further MSP diversity is generated through gene duplications and alternate transcription. In cotton-top tamarin, for example, there are three known functional genes specifying closely related MSP sequences (Mäkinen et al. 1999). In human tissues, two different MSP mRNAs resulting from alternate splicing have been identified. The major one encodes the prostatic form of 94 amino acids; the minor one, a frame-shift variant of 57 amino acids that is mostly expressed in urogenital tissues (Xuan et al. 1995).

Studying the sequence, expression pattern, and tissue distribution of homologous or orthologous MSP genes could contribute to a better understanding of the molecular evolution of this protein and to eventual elucidation of its biological role. This paper describes for the first time the primary structure of a MSP-like molecule isolated from pituitary extracts of a ratite, the ostrich (*Struthio camelus*). Database searches led to the identification of related sequences in several species including bovids, batracians, and fish.

Results

From 1987 to 1992, we purified and characterized numerous ostrich adenohypophyseal hormones and/or proteins, including corticotropin (Naudé and Oelofsen 1977), both neurophysins (Lazure et al. 1989, 1990), a chromogranin A–related fragment (Lazure et al. 1987), and the N-terminal proopiomelanocortin fragment (Naudé et al. 1993). In 1987, we determined a 33–amino acid N-terminal sequence of a protein that proved, in its native form, remarkably resistant to cleavage by proteases. The sequence showed no significant resemblance to any protein known at the time. For many years, we searched protein and DNA databases for related sequences to no avail. More recently, we identified a unique partial sequence that displayed >50% sequence identity to the ostrich's sequence. This sequence corresponded to a chicken blood plasma protein that was antigenically cross-reactive with chicken β2-microglobulin but completely dissimilar in sequence to the latter (Warr 1990). With the emergence of fast and reliable means to determine nucleic acid sequences, there has been a tremendous increase in available cDNA sequences known as ESTs. Using the TBLASTN algorithm, we screened EST databases for sequences related to the chicken protein and identified two

EST hits: one from liver (GenBank accession No. AW198358), the other from activated T cell libraries (Gen-Bank accession No. AI979828; Tirunagaru et al. 2000). These ESTs clearly encoded a MSP-related protein based on the distribution and number of half-cystines as well as the size of the protein. To gain some insight into evolution of this intriguing molecule, we decided to also determine the complete amino acid sequence of the ostrich protein.

Determination of the ostrich MSP primary sequence and disulfide linkages

Though it is difficult to estimate the amount of protein extracted and purified from ostrich pituitary, it was nevertheless possible to recover ~1.1 mg of relatively pure MSP in the original reverse-phase high-performance liquid chromatography (RP-HPLC) fraction 6D-2 (see Materials and Methods). Following reduction and alkylation, we determined the first 49 N-terminal residues in a single sequenator run (Fig. 1). This determination not only confirmed the assignment of the first 33 amino acids we made >10 years ago, but it also allowed (1) positive identification of seven half-cystine residues occupying positions identical to those in mammalian MSPs (positions 2, 16, 34, 37, 39, 46, and 47, respectively) and (2) the localization at position 44 of the Met residue, predicted on the basis of amino acid composition of ostrich MSP (data not shown).

Subsequent sequences and mass analyses of the various peptides resulting from either endoproteinase AspN digestion (Fig. 2A) or cyanogen bromide (CNBr) cleavage (Fig. 2B) allowed us to determine the complete sequence (Fig. 3). It is worth noting that chemical cleavage through CNBr was
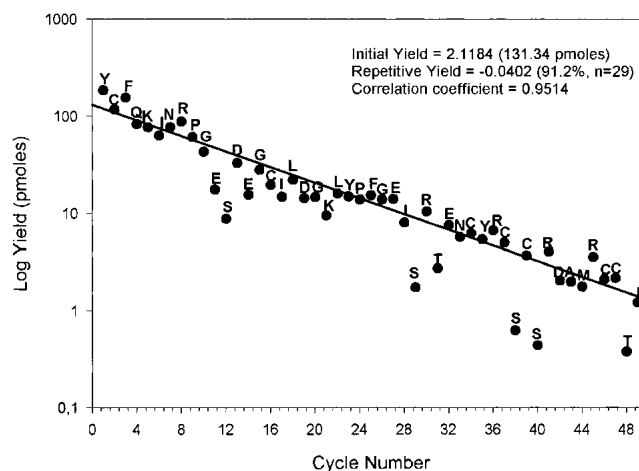
Fig. 1. Automatic N-terminal sequence of reduced and carboxamidated ostrich β-microseminoprotein. Quantitative yields of phenylthiohydantoin (PTH) amino acids normalized to a PTH internal standard are illustrated as a function of residue numbers. The slope and intercept were obtained by a linear regression analysis on 29 selected stable PTH amino acids.
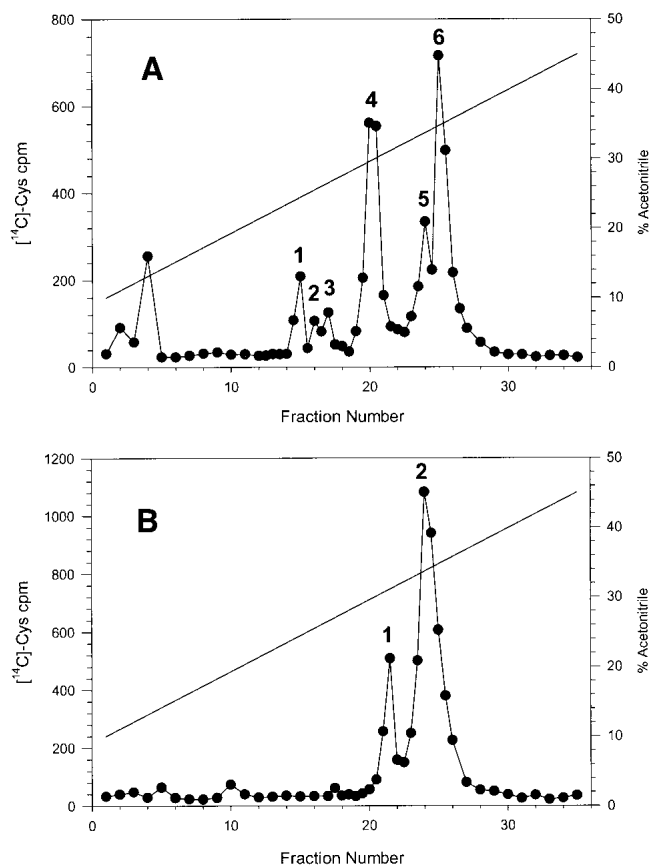
Fig. 2. Narrow-bore reverse-phase high-performance liquid chromatography separation of peptides, resulting from enzymatic digestion with endoproteinase AspN (*A*) and chemical cleavage with cyanogen bromide (*B*) of reduced and carboxamidated ostrich β-microseminoprotein. Elution was performed as described in the Materials and Methods section and using a TFA/acetonitrile gradient as shown. Fractions of 100 μL were manually collected, and the amount of radioactivity (in 5 μL aliquots) was determined by scintillation counting.

far from quantitative, as can be seen from the ratio of radioactivity in each peak. Indeed, ideally the ratio should have been unity as each fragment contains the same number (five) of half-cystine residues. This discrepancy is most probably because of extensive oxidation of the single Met residue over the years of storage, making it unreactive to CNBr. Nevertheless, sequencing of the material present in the early eluting peak (peak 1 in Fig. 2B) yielded positions 45–75 (repetitive yield, 87.9%; initial yield, 37 pmoles with a correlation coefficient of 0.9160 [n = 15]; data not shown). Peak 2 (Fig. 2B) contained both the CNBr-unreactive material as well as the N-terminal fragment 1–44. Isolation and chemical characterization of peptides AspN1, AspN3, and AspN4 arising from the endoproteinase AspN provided an independent confirmation of residues occupying positions 1–12 (peak 1 in Fig. 2A), 19–41 (peak 4), and 42–65 (also present in peak 4), respectively. The C-terminal sequence was determined from a peptide encompassing po-

sitions 73–90 (peak 3) and hence containing a Lys-Asp-Asp-Pro sequence that was cleaved very poorly by the AspN enzyme. This peptide, which contained AspN5, AspN6, and AspN7 together, also proved difficult to sequence because of a consistently poor repetitive yield (~75% to 77%; data not shown). Together, each position was sequenced on two to four occasions, with the exception of the Ser occupying position 69. Similarly, because of the rapidly decreasing repetitive yield, it was difficult to unambiguously deduce the amino acid sequence following position 87.

The C-terminal sequences of peptide AspN3 and AspN4 were confirmed through analysis of the peptides derived following carboxypeptidase-P digestion and ladder sequencing by MALDI-TOF (matrix-assisted laser desorption ionization–time of flight) as shown in Figure 4A and 4B. The deduced sequence for AspN3 corresponded to C(CAM)-Y-R-C(CAM)-S-C(CAM)-S-R; for AspN4, it corresponded to V-G-Y-N-K-E-K-C(CAM)-K-V-V-F-N-K-E-S-C(CAM)-N-Y and confirmed the Ser at position 69 and the Tyr at position 72. On the other hand, a similar analysis could not be conducted with the two C-terminal peptides 73–90 (AspN5/6/7) and 78–90 (AspN7, present in peak 2), as these proved resistant to the action of carboxypeptidase-P or -Y. A possible explanation for this resistance was that their C terminus is amidated. The AspN5/6/7 and AspN7 peptides, which yielded ions of m/z of 2175.4 and 1605.1, respectively (theoretical m/z, 2174.1 and 1604.8), were analyzed using fragmentation of the parent ion induced through laser excitation with the postsource decay (PSD) mode. Results deduced from analysis of the y-series (from y_3 to y_{11}) and b-series (from b_2 to b_{11}) ions confirmed the sequence C(CAM)-F-V-Y-S-R-V (data not shown). Furthermore, following enzymatic digestion by Arg-C protease of a fraction containing fragment 78–90; 80–90 (m/z, 1371.0; likely to result from acid cleavage of the Asp-Pro bond); and 83–90 (m/z, 1058.7; likely to result from cleavage by AspN N-terminal to a Glu residue as sometimes noticed; see discussion forum at http://www.abrf.org), all masses of the resulting peptides were decreased by 98.2 mass units, and these peptides were no longer resistant to the action of carboxypeptidases, hence confirming that the Val is amidated (data not shown). Such analyses confirmed the previously deduced sequence and further identified positions 87–88 as being Tyr-Ser and 89–90 as being Arg-Val. Identification of these residues is in full agreement with the observed masses of each respective parent ion.

In our effort to determine the disulfide linkages of ostrich MSP, one major difficulty was its resistance in its native state to trypsin and other proteolytic enzymes (data not shown). Indeed, even under prolonged digestion with a high ratio of enzyme to substrate, numerous potential sites of cleavage were not efficiently cleaved. This rendered identification of fragments difficult by the very sensitive
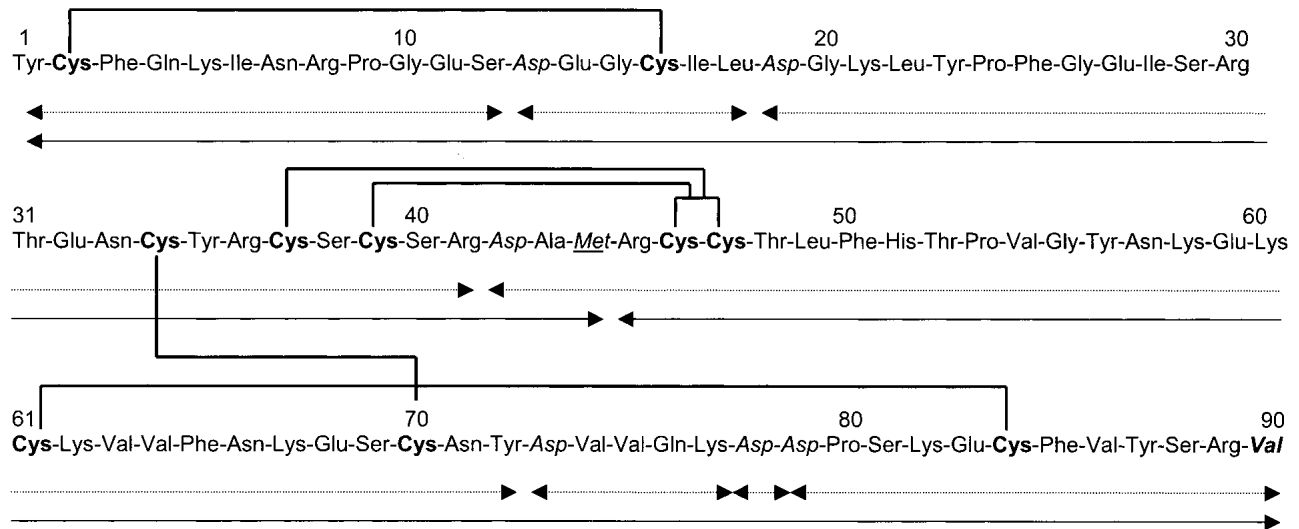
Fig. 3. The complete amino acid sequence of ostrich β-microseminoprotein. The peptides resulting from cyanogen bromide cleavage are shown with full arrows, whereas those resulting from cleavage using endoproteinase AspN are shown with broken arrows. Half-cystine residues are shown in bold lettering, whereas Asp and Met residues are shown in italics and underlined italics, respectively. The proposed disulfide pairing is indicated by the thick lines joining the half-cystine residues, whereas the proposed amidated C-terminal Val residue is shown in bold italics.

MALDI-TOF analysis. Nevertheless, as indicated in Table 2, with a single exception, we were able to isolate each disulfide-linked peptide and, on reduction and alkylation, to identify the peptides involved in the linkage. The exception was the peptide pair involving peptide 37–41 and peptide 46–58, linked by two disulfide bridges. In this case, it was not possible to determine the linkages between Cys-37 and Cys-41 to the Cys residues composing the pair 46–47 (see Fig. 3). This linkage may eventually be determined by de novo chemical synthesis, nuclear magnetic resonance or X-ray crystallography studies.

Comparison of the ostrich MSP sequence to mammalian, chicken, and EST-derived sequences

Table 3 provides the pairwise percentages of overall identity between the various MSP-related proteins (signal peptides excluded). These identities are illustrated in a multiple alignment of their amino acid sequences shown in Figure 5A. Except between baboon and rhesus monkey, they are generally <80%. Expectedly, they are greater within orders than between orders. Thus, although the ostrich and chicken sequences show 57% overall similarity, the similarity of either one to the related sequences in other species generally decreases <40%. Within orders, one can easily identify continuous stretches of relatively high similarities and identities between species. Such stretches are observed between rat and mouse, human and other primates, chicken and ostrich, and flounder and zebrafish (boxed in Fig. 5A). Such an

intraspecific conservation of amino acid clusters has been previously pointed out for rodent and primate MSP sequences (Fernlund et al. 1996; Mäkinen et al. 1999). A consensus sequence was difficult to derive from the alignment of all the sequences as, apart from the half-cystine residues, there is very little amino acid conservation across species. This observation explains the great difficulty previously encountered in finding homologs and/or orthologs in species other than primates.

We searched in the various databases for related proteins using the TBLASTN v2.0.12 algorithm, and as a query, either the full ostrich sequence determined in this study, the previously found chicken EST (GenBank, accession Nos. AI97928 and AW198358), or a hypothetical sequence containing the most conserved residues in all available oligonucleotide or amino acid sequences. The search allowed the identification of five candidate MSP-related sequences. The first sequence identified in the nonhuman and nonmouse EST bank was a partial bovine sequence (accession No. AW336761; E value of 2×10^{-7}), encompassing the first 42 N-terminal residues of the mature protein preceded by a 20–amino acid signal peptide. This sequence was present in a cDNA library of pooled lymph node, ovary, fat, hypothalamic, and pituitary tissues. It encodes a segment of the mature MSP that is 100% identical to the porcine homolog and highly similar to other mammalian MSP sequences (data not shown). The second EST sequence identified in the search yielded an open reading frame encoding a protein of 111 amino acids, including a 20–amino acid putative
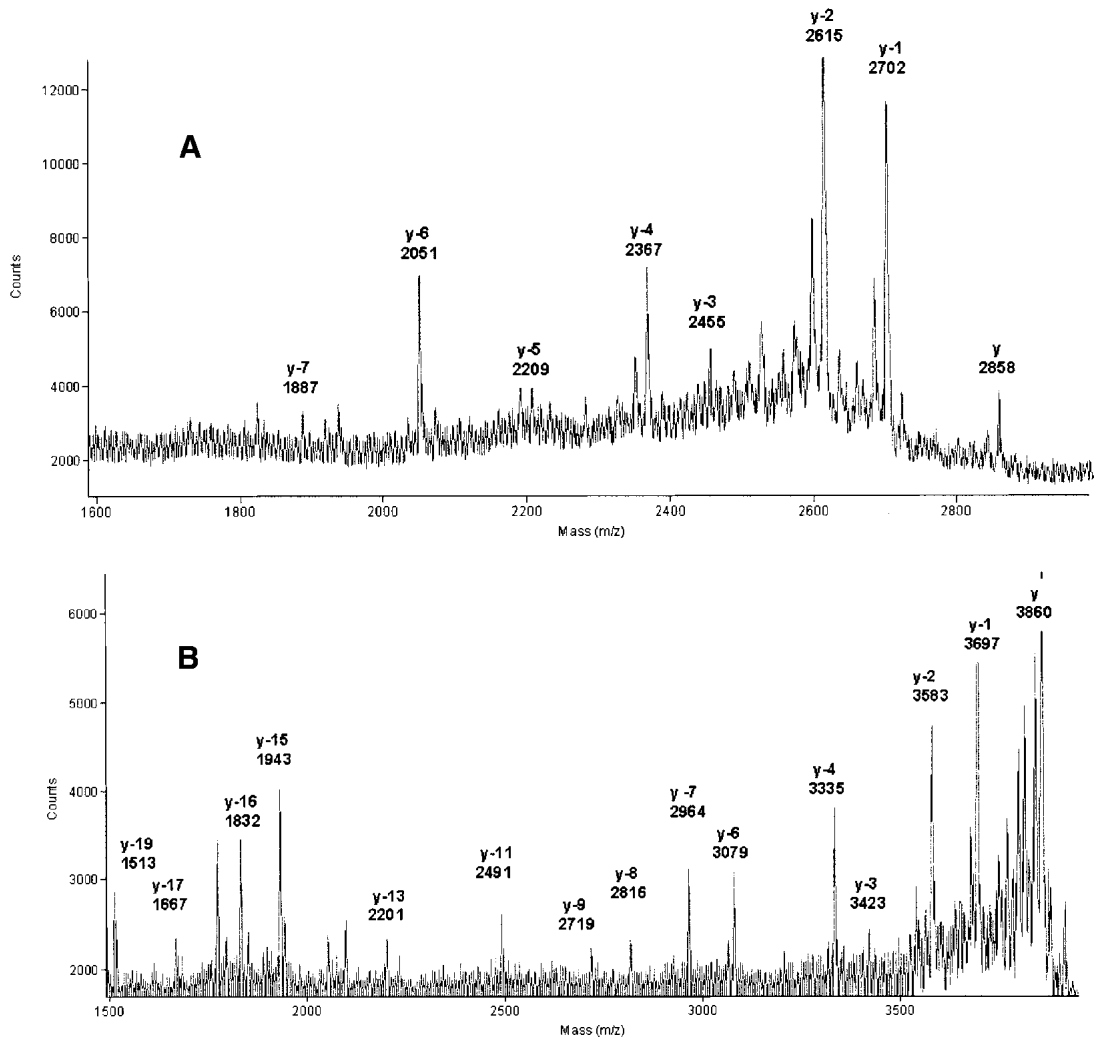
Fig. 4. (*A*) C-terminal ladder sequencing of peptide AspN3 following carboxypeptidase-P digestion. The reverse-phase high-performance liquid chromatography isolated, reduced, and carboxamidated AspN3 peptide was digested with carboxypeptidase-P for 30 sec as described in Materials and Methods, and the resulting mass spectrogram was obtained. The y ion corresponds to the complete peptide 19–41 (theoretical m/z, 2859), and the y-n label denotes the loss of n amino acids from the C terminus. (*B*) C-terminal ladder sequencing of peptide AspN4 following carboxypeptidase-P digestion. The RP-HPLC isolated reduced and carboxamidated AspN4 peptide was digested with carboxypeptidase-P for 60 sec as described in Materials and Methods, and the resulting mass spectrogram was obtained. The y ion corresponds to the complete peptide 42–72 (theoretical m/z, 3857), and the y-n label denotes the loss of n amino acids from the C terminus.

signal peptide (accession No. AW641318; E-value of 2×10^{-15}). It was isolated from a *Xenopus* egg cDNA library. Its closest relatives are the chicken and ostrich MSPs (see Table 3 and Fig. 5A). The third EST sequence was from zebrafish (accession No. AI497271; E value of 8×10^{-5}); it encoded a 90–amino acid mature protein preceded by a 20-residue signal sequence. Apart from the conserved 10 half-cystines, this sequence was unique as it bears little identity with MSP from other species. The last two EST sequences identified were found in a cDNA library of Japanese flounder liver and spleen. One of them (herein referred to as flounder-1, accession No. C23089; E-value of

3×10^{-6}) encodes an 87–amino acid mature protein containing 10 half-cystine residues and preceded by a 19-residue signal peptide; the other one (herein referred to as flounder-2, accession No. C23023; E-value of 0.81) encodes a 97–amino acid mature protein containing 12 half-cystines preceded by a 19-residue signal sequence. Despite the little similarity between the flounder-2 sequence and other MSPs (see Fig. 5A), it was nevertheless included in this family because of its significant overall identity (54%) with flounder-1 sequence (Table 3). These sequences are very probably related to MSPs despite the extensive non-conservation of residues with mammalian species. To verify

Table 2. *Proposed disulfide linkages of ostrich β-microseminoprotein (MALDI-TOF)*

Native tryptic peptide $(M + H)^+$		Reduced and alkylated peptide pairs $(M + H)^+$				Tryptic peptides identified (positions)
Observed	Computed	Observed	Computed	Observed	Computed	
2389.5	2392.1	741.6	746.3	1760.7	1760.8	(1–5) + (6–21)
1967.5	1970.8	843.6	843.3	1242.7	1241.5	(31–36) + (68–77)
1950.4	1953.9	—	565.3	1503.0	1503.7	(59–62) + (78–89)
1693.3	1696.8	—	307.1	1503.0	1503.7	(61–62) + (78–89)
2037.6	2038.9	671.3	670.2	1599.6	1597.7	(37–41) + (46–58)

this relatedness, we computer analyzed the phylogenic arrangement of all sequences (signal peptide excluded) shown in Table 3. In the resulting evolutionary tree (Fig. 5B), sequences are clearly grouped into specific branches for fish, amphibians, birds, rodents, ungulates, and primates. Based on this tree, cotton-top tamarin sequences diverged early from those of other primates, consistent with the percentages of primary structure identities among them (see Table 3).

Interestingly, unlike their mammalian counterparts, non-mammalian MSP sequences (the chicken one excepted) contain a Gly residue at either the penultimate or the C-terminal position (see Fig. 5A), indicating that like ostrich MSP, they might be terminally amidated through the activity of the peptidylglycine α-amidating monooxygenase (PAM). In cases in which the Gly residue occupies the penultimate position, as in the flounder and zebrafish sequences, this residue could be easily uncovered by removal of the C-terminal Lys residue through the action of carboxypeptidase-E or -H. Both PAM and carboxypeptidase-E are processing enzymes normally found in the secretory granules of the regulated pathway of secretion.

Discussion

This study reveals that a peptide long thought to function in mammalian reproductive tissues not only is expressed in a variety of other tissues, many of them without any role in reproduction, but also is present in many other species. MSP-related sequences are now known for representatives of the following classes and orders: artiodactyls (cow and porcine), primates (human, apes, and monkeys), rodents (rat and mouse), birds (chicken and ostrich), amphibians (*Xenopus*), and fish (zebrafish and flounder). The avian MSP closely resembles the *Xenopus* one, in agreement with the now favored view of evolutionary divergence of amniote vertebrates, which places birds closer to reptilians (such as crocodiles) and amphibians than to mammals (Hedges 1994; Hedges and Poling 1999). MSP could thus be considered a valid marker of speciation. The finding of related sequences in metatheria (marsupials), reptiles, and insects may allow further verification of this proposal. Intriguingly, no sequence that, by overall size and half-cystine distribution, could unambiguously be related to MSP was identified among back-translated proteins deduced from the now com-

Table 3. *Sequence identity between reported β-microseminoprotein*

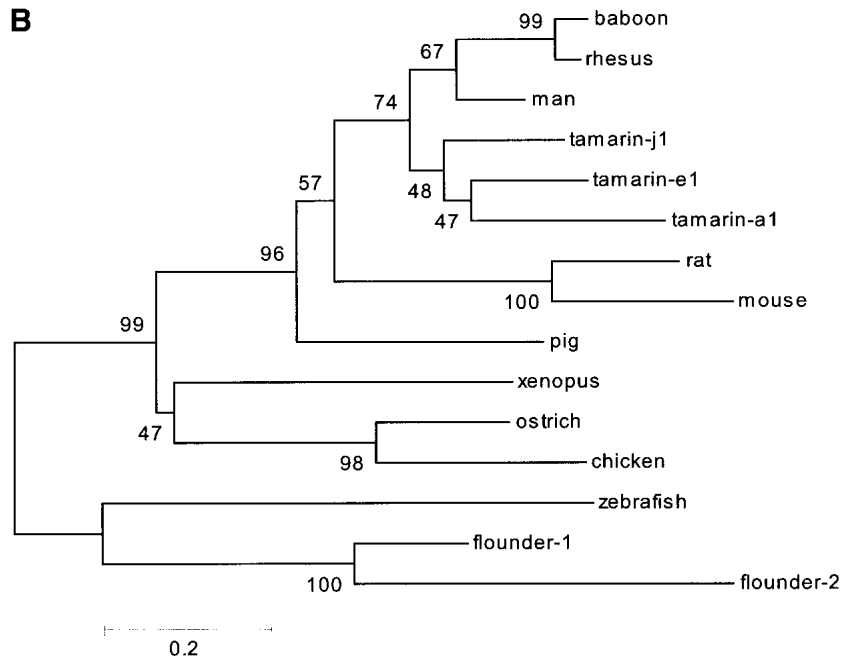| Species	B	Rh	Ta-e1	Ta-j1	Ta-a1	P	R	M	O	C	X	F-1	F-2	Z
H	79	79	71	70	60	51	45	44	39	38	33	28	21	24
B		94	67	66	51	51	49	44	39	37	35	29	20	21
Rh			67	67	52	53	50	46	40	40	36	28	19	22
Ta-e1				69	64	49	44	40	36	37	34	26	21	25
Ta-j1					64	52	43	38	40	38	39	30	22	23
Ta-a1						49	42	34	40	41	35	30	23	26
P							43	40	33	32	36	27	24	25
R								62	33	31	28	25	19	22
M									33	33	28	23	23	21
O										57	40	35	23	29
C											43	32	22	31
X												37	24	26
F-1													54	37
F-2														29

The sequence identities in percentages (to the nearest full integer) were determined following pair-wise comparison of full-length sequences of β-microseminoproteins (signal peptides excluded) using the Align algorithm at http://www2.igh.cnrs.fr/bin/align.guess.cgi.
B, Baboon; Rh, Rhesus monkey; Ta, cotton-tail tamarin; P, pig; R, rat; M, murine; O, ostrich; C, chicken; X, African clawed-frog; F, Japanese flounder; Z, zebrafish; and H, human.

A

```
man         SCYFIPNEGVPGDSTRKCMDLK-GNKHPINSEWQTDNCETCTCYET-EISCCTL  52
baboon      SCSFMPNERFPGDSTRECTDLK-GNKHPINSKWQTDNCEACTCYET-EIICCTL  52
rhesus      SCSFIPNERFPGDSTRECTDLK-GNKHPINSKWKTDNCERCTCYKT-EIICCTL  52
tamarin-e1  SCYLMPNKMVPGDSTKECTDLK-GNKHPLNSRWKTECCDECDCLEK-EISCCSL  52
tamarin-j1  SCYIILNDMIPGDSTNECTDLK-GNKHPINSKWRTDNCDSCTCREK-EISCCTL  52
tamarin-a1  SCYVIRHKIVPGETIKECTDLK-GNKHPLDSRWRTEDCELCACRDI-EISCCSL  52
pig         QCYFIPNQSLK---PNECQDLK-GVSHPLNSVWKTKDCEECTCGQN-AISCCNT  49
rat         ACSIQRLKRLPNEKSDECTDVD-GGKHVLNTYWQ-KNCEWCFCEKT-AITCCTK  51
mouse       VCSIENREIFPNQMSDDCMDAD-GNKHFLNTPWK-KNCTWCSCDKT-SITCCTN  51
ostrich     YCFQKINRPGESDESDEDCILD-GKLYPFGEISRTENCYRCSCSRD-AMRCCTL  52
chicken     FCFSKLFKPGEAEKG---CMLD-GVLYPFGEIPRTENCFRCSCSKN-EMHCCSL  49
xenopus     FCFT----ELPKMGQTEGCLYK-DKLHRLGSKFRTNDCMDCTCHMDGKMACCQA  49
flounder-1  DCFQ----ILTKPGMTHCQDHDDGTWHEVGSSWRNSRCYDCSCEG-----CCAA  41
flounder-2  DCFQ----ILTKPGMTHCQDHDDGTWHEVGSSWRNSRCYDCSCEG-----CCAA  45
zebrafish   YCFF----KLQKEGAKYCEDGDDQTRHELGSTWINSKCLRCICSST-EMECCDT  49

CONSENSUS   .C...............c.d...g..h.....w....C..C.C......CC..
```

```
man         VST--PVGYDKDNCQRIFKKED-CKYIVVEKKDPKKTCSVS--EWII------  94
baboon      IAT--PVGYDKNKCQRIFKKEE-CKYIVVEKKNPKKTCPID--QWIL------  94
rhesus      IAT--PVGYDKKKCQRIFKKED-CKYIVVEKKNPKKTCPID--QWIL------  94
tamarin-e1  VAI--PVGYDQDNCQKIFKQED-CKYIVVEKKDPNKTCEVT--QWI-------  93
tamarin-j1  VST--PVGYDTHKCQKIFNKED-CRISVVEKNDPSKTCEVN--AWIM------  94
tamarin-a1  VST--PVGYDRHNCRKIFNKET-CKISVVEKTDPNRPCGVS--GWIS------  94
pig         AAI--PTGYDTNKCQKILNKKT-CTYTVVEKKDPGKTCDVT--GWVL------  91
rat         TLI--PVSYDKKRCQRQFHSEN-CTYSVVERTNPGKTCPVN--GWTI------  93
mouse       ATR--PLSYDKDNCDVQFHPEN-CTYSVVDRKNPGKTCRVD--SWTM------  93
ostrich     FHT--PVGYNKEKCKVVFNKET-CNNDVVQKDDPSKECFVS--YRV-------  93
chicken     YHT--PVNYDKETVKVIFNKKLNCDYEVVQK-HPSKPCSGY--ARV-------  90
xenopus     YAT--PVEYDKEKCIAVFNRKA-CHYRVVEKRNRSKECEVL--AMVG------  91
flounder-1  FS---TPKRIPAECVSVFD-SNECVYRVFKKDDPTVLCPIL--AAVGK-----  87
flounder2   FQPLNVSRLNVCQCSTQMNVCTECSRRMTRQYCVQFWLQLASDAVIGPGGIKG  98
zebrafish   MG---RAIIKTEGCTVKYD-YSTCKFDVFHPEDPNIKCEYG---AVGK-----  90

CONSENSUS   .....p...yd...C...f.....C...Vv.k..p.k.C...............
```

B

```
                                  99 ┌─ baboon
                               67 ┌──┴── rhesus
                            74 ┌──┴──── man
                         ┌──┴──────── tamarin-j1
                      57 │     48 ┌── tamarin-e1
                   96 ┌──┤        └──┤ 47
                      │  │           └── tamarin-a1
                99 ┌──┤  │        ┌── rat
                   │  │  └────────┤ 100
                   │  │           └── mouse
                   │  │  ┌──────── pig
                   │  └──┤
                   │  47 ├──────── xenopus
                   │     │ ┌─────── ostrich
                   │     └─┤ 98
                   │       └─────── chicken
                ───┤──────────────── zebrafish
                   │       ┌──────── flounder-1
                   └───────┤ 100
                           └──────── flounder-2

        ├─────┤
         0.2
```

plete *Drosophila* genome. Many of the retrieved sequences showed partial alignment of half-cystines and little else. Most often, the aligned half-cystines belong to domains of proteins far larger than most MSPs. This could indicate that this protein, and hence its function, appeared following divergence of vertebrates from invertebrates. On the other hand, it is possible that current algorithms cannot find related sequences in too distant species, leaving open the possibility that MSP-like proteins might be present in insects too.

Inasmuch as expression and/or abundance of mRNAs cannot be taken as conclusive evidence that a protein is present in a given tissue (for example, see Anderson and Seilhamer 1997), the presence of MSP in reproductive and nonreproductive tissues can conclusively be established only through its purification or by immunocytochemistry. Where these criteria have been met, MSP is generally associated with secretory cells, of both endocrine and exocrine types. As mentioned in the Introduction, MSP has been found in E-cells of the gastric mucosa (Ulvsbäck et al. 1989). Our identification of this protein in ostrich pituitary extracts further supports this association. Furthermore, MSP is found in significant amounts in chicken serum (0.06 g/L; Warr 1990). Thus, birds may possess an active endocrine pathway of MSP secretion into plasma. The pituitary may be one of the secreting endocrine organs. Indeed, using primers corresponding to the reported chicken MSP cDNA, we have been able to detect its mRNA by reverse transcription–polymerase chain reaction (RT-PCR) in total RNA from the brain of this bird (M. Mbikay and C. Lazure, unpubl.).

In mammals, MSP appears to be most abundant in seminal plasma (in humans, its concentration in this fluid is as high as 0.9 to 2.2 g/L). It is also elevated in numerous other body fluids, especially in nasal (0.009 to 0.011 g/L) and tracheal (0.013 to 1.4 g/L) secretions (Weiber et al. 1990).

It is detectable in the plasma of asymptomatic men at a concentration of ~0.02 mg/L and at a slightly elevated level in patients with prostate cancer (0.12 mg/L; Dubé et al. 1987b). Its serum level, however, is of no diagnostic or prognostic value for this type of cancer (Von der Kammer et al. 1993). The biological function of MSP remains unknown. Whatever that function is, it is uncertain that it is the same among species. In this context, it would prove interesting and possibly revealing to examine the cellular and tissular distribution of MSP in birds and in other distant species such as amphibians and fish. Some preliminary immunofluorescence data with the chicken MSP tend to favor its relation with bursa and thymus cells (Warr 1990).

Structural studies on MSP are scarce. Disulfide pairing in all of them has not been determined. Their sequences bear no significant resemblance to proteins of known function. These facts make it difficult to even speculate about their biological functions. Indeed, the only structural data known so far have indicated that the secondary structure of chicken MSP shows negligible α-helicity but elevated β-sheet (68%) content. This property is also shown by β_2-microglobulin, which is antigenically cross-reactive with MSPs (Warr 1990). A high β-sheet content is observed in the epithelin/granulin protein family, another class of small proteins with a high number of half-cystines (for review, see Bateman and Bennett 1998). In the latter case, the overall structure is composed of four stacked β-sheet structures that are highly stabilized by a central axial rod of disulfide bridges. On the other hand, there is no a priori reason to consider the relationship of MSPs to epithelin/granulin to be stronger than possible relationships to other classes of small proteins highly rich in half-cystines, such as the defensin family of endogenous antibiotic peptides (for review, see Lehrer et al. 1991) or the disintegrin family of integrin inhibitory peptides from snake venoms (for review, see McLane et al. 1998). Interestingly, apart from the obvious

Fig. 5. (*A*) Multiple amino acid sequence alignment of mature β-microseminoproteins. Sequences retrieved from databases were automatically aligned using the CLUSTAL W algorithm (http://www.ebi.ac.uk/clustalw/) and the Multalin v5.4.1 software (http://www.toulouse.inra.fr). Gaps are indicated with dashes. Residues conserved in 14 of 15 sequences (>90%) are bolded and written in white over a black background; those conserved in >11 of 15 sequences (>70%) are bolded and shaded. Continuous stretches of five residues or more that are highly conserved among members of the same order (e.g., primates, rodents, aves, and fish) are boxed. The β-microseminoprotein amino acid sequences were either obtained from previously reported sequences—human (*Homo sapiens*; accession No. AJ13356; Mbikay et al. 1987), rhesus monkey (*Macacca mulatta*; accession No. M92161; Nolet et al. 1991), baboon (*Papio hamadryas anubis*; accession No. U49786; Xuan et al. 1997), cotton-top tamarin (*Saguinus oedipus*; accession Nos. mspE1, AJ010154; mspA1, AJ010158; and mspJ1, AJ010156; Mäkinen et al. 1999), porcine (*Sus scrofa*; accession No. S41663; Fernlund et al. 1994; Tanaka et al. 1995), rat (*Rattus norvegicus*; accession No. U65486; Fernlund et al. 1996), murine (*Mus musculus*; accession No. J89840; Xuan et al. 1999), ostrich (*Struthio camelus*; this study), and chicken (*Gallus gallus*; Warr 1990)—or back-translated from the reported cDNA sequences in the nonmouse and nonhuman expressed sequence tag entries maintained at the NCBI: chicken (accession Nos. AI97928 and AW198358), African clawed frog (*Xenopus laevis*; accession No. AW641318), zebrafish (*Danio rerio*; accession No. AI497271), and the two Japanese flounder sequences (*Paralichthys olivaceus*; accession Nos. C23089 and C23023). The bovine (*Bos taurus*; accession No. AW336761) was not included in the alignment because its full sequence has not been reported. (*B*) The phylogenetic tree of β-microseminoproteins. Based on the alignment shown in *A*, the phylogenetic tree analysis was conducted using neighbor-joining analysis (Poisson correction) with the MEGA v2.0 software (http://www.megasoftware.net). The confidence values obtained using a bootstrapping statistical analysis method are indicated on nodes; values of ≥95% are considered statistically significant. The scale bar indicates the d value, which represents the number of amino acid substitutions per site.

presence of an elevated content in half-cystine residues, all of these three classes of peptides possess biological activities proposed at one time or another for MSP (see Table 1). Together with the present demonstration of MSP-related molecules in distant species, these relationships, as tenuous as they are, might eventually lead finally to a clear definition of the biological activity of MSP >20 years after its initial discovery.

Materials and methods

Purification of MSP from ostrich pituitaries

The protein was obtained following extraction of 1060 g of adenohypophyses using the classical acid-acetone extraction procedure previously described (Naudé and Oelofsen 1977). Briefly, the resulting powder was submitted to NaCl-fractionation, and the corticotropin-containing fraction D was recovered and further purified through chromatography on a carboxymethyl-cellulose column using an ammonium acetate gradient for elution. The resulting fraction 6 (see Fig. 1 in Naudé and Oelofsen 1977) was purified on a Sephadex G-100 column, and the last eluting fraction 6D was collected. The pooled fraction 6D was further purified by RP-HPLC and eluted using a *n*-propanol gradient. Fraction 6D-2, eluting between 15% and 20%, *n*-propanol was collected, lyophilized, and kept at −20°C. Earlier protein characterization studies were performed using this fraction as starting material.

Chemical modifications and CNBr cleavage

The dried RP-HPLC fraction was dissolved in 0.4 M Tris-HCl buffer (pH 8.4) containing 8 M urea and 1 mM EDTA. The protein was reduced and alkylated using a two-step procedure: it was first incubated for 30 min in the presence of 0.1 mM DTT before treatment with iodo-[1-^{14}C]-acetamide (5 μCi, 59 mCi/mmol; Amersham Pharmacia Biotech), and it was then fully reduced and alkylated using a 5× excess of unlabelled iodoacetamide over DTT. The reduced and alkylated protein was desalted using a PD-10 column (Biorad) using 0.1 M ammonium bicarbonate as eluent; the resulting labeled fraction was dried and kept at −20°C until used. Cleavage of 100 μg of the reduced and carboxamidated protein with CNBr was conducted in 70% acetic acid overnight at room temperature in the dark; the resulting fragments were purified by RP-HPLC as described below.

Enzymatic digestion

The reduced and carboxamidated protein (100 μg) was dissolved in 100 μL of 0.1 M ammonium bicarbonate, and 1 μg of endoproteinase Asp-N (Roche Diagnostics) was added. Incubation was overnight at 37°C, and the resulting fragments were purified by RP-HPLC as described below. Digestion of the C-terminal peptide by Arg-C protease (Roche Diagnostics) was accomplished using 0.5 μg of enzyme, reconstituted according to the manufacturer's protocol in 100 μL of 100 mM Tris-HCl containing 10 mM CaCl$_2$ (pH 7.6) overnight at 37°C. Digestion of the native ostrich MSP (1 μg/μL) was performed with trypsin (Miles-Seravac) using a 1:50 (w/w) ratio in 0.1 M pyridine-acetate buffer (pH 6.5) for 24 h at 37°C. All enzymatic reactions were stopped by acidification using 0.1% CF$_3$COOH.

Peptide separation by RP-HPLC

Peptides resulting from the fragmentation mixtures were separated and purified using a μRPC C2/C18 narrow-bore column (SC 2.1/10; Amersham Pharmacia Biotech) maintained at 30°C. They were eluted using a 0.1% (v/v) TFA-acetonitrile system with linear gradient of either 0 to 100% acetonitrile in 35 min or 10% to 45% acetonitrile in 35 min at a flow rate of 200 μL/min. The eluting peaks were monitored through ultraviolet-absorbance at 225 nm and collected manually in 100-μL fractions, and radioactivity in each fraction was measured using a scintillation counter. All separations on HPLC were performed on an Applied Biosystems model 120A. A modification to the original configuration included addition of a manual rotating valve, allowing collection of the eluate immediately after the detector flow cell through a narrow bore PEEK tubing long enough to prevent degassing.

Isolation of the disulfide-linked peptides was accomplished on a CSC-Exsil A300 C$_{18}$ column (25 × 0.46 cm; Chromatography Sciences Co.). The buffer system consisted of an aqueous 0.1% (v/v) CF$_3$COOH solution and an organic phase of acetonitrile containing 0.1% (v/v) CF$_3$COOH. Elution was performed with a linear gradient from 0% to 60% organic phase in 60 min following a 5-min isocratic step at 0% at a flow rate of 1.0 mL/min. The eluate was monitored at 225 nm, and 500 μL fractions were collected. A Varian 9010 HPLC connected to a Varian 9050 ultraviolet-detector was used. Aliquots of each fraction were screened by MALDI-TOF as described below. Selected fractions were reduced and alkylated as described above before mass analysis.

Edman degradation, amino acid, and mass spectrometry analysis

Automated Edman degradation of the native RP-HPLC purified protein was performed as described previously (Lazure et al. 1989) using a gas phase sequenator from Applied Biosystems (model 470A). The resulting phenylthiohydantoin (PTH) amino acids were analyzed directly on an Applied Biosystems (model 120A) PTH-analyser. Automated Edman degradation of the reduced and carboxamidated protein and fragments thereof was performed using an Applied Biosystems (model 477A) sequenator operated in gas phase mode using *N*-methylpiperidine as coupling buffer and directly linked to the PTH-analyser.

Amino acid analyses were performed in duplicate after hydrolysis of dried samples in 5.7 N HCl in vacuo for 18 to 24 h at 110°C on a Beckman autoanalyser (model 6300) with a postcolumn ninhydrin detection system and coupled to a Varian DS604 integrator/plotter.

MALDI-TOF analysis of RP-HPLC–purified fractions was performed on a Voyager DE-Pro MALDI-TOF instrument (PerSeptive Biosystems) equipped with a 337-nm nitrogen laser and a delayed-extraction ion source. Positively charged ions were analyzed in linear, reflector, and PSD mode. Spectra were obtained as the sums of ions recovered after 64 to 128 laser shots using an acceleration voltage of 20 kV. In PSD mode, spectra were obtained using ion reflector mirror ratio from 1.0 to 0.3 in decreasing steps, as recommended by the manufacturer. The matrix solution was freshly prepared using a saturated solution of α-cyano-4-hydroxycinnamic acid (HCCA; Sigma-Aldrich) in 50% (v/v) acetonitrile containing 0.3% TFA and was mixed in a 1:1 ratio with samples in 0.1% TFA. For ladder sequencing, 0.3 μL peptide solution (1 pmole/μL in 0.1% TFA) was mixed with an equal volume of freshly prepared carboxypeptidase-P or -Y (0.1 μg/μL in 50 mM sodium citrate buffer at pH 6.0; Roche Diagnostics) and incubated at room temperature in a humid chamber for varying amounts of

time (from 3 to 60 sec). At the end of incubation, 0.6 μL of freshly prepared matrix solution was added and cocrystallisation left to occur at ambient temperature. For determination of disulfide bridges, peptides, following reduction and alkylation, were deposited on the MALDI-TOF plate using C_{18} ZipTip (Waters Corporation) prewetted with 2×10 μL 50% acetonitrile in water and equilibrated with 2×10 μL 0.1% CF_3COOH. After aspiration of the sample, bound peptides were eluted using 10 μL of HCCA solution (10 mg/mL in 50% acetonitrile), and 2 μL were spotted directly on the sample plate.

Sequence analysis

Amino acid sequences and/or nucleotide sequences were analysed using the Genetic Computer Group computer program package. Automated searches in the various databases were performed through the entrez search and retrieval system available at the NCBI site (http://www.ncbi.nlm.nih.gov) using the various BLAST algorithms operating in batch mode. Sequence alignment was accomplished using the Multalin v5.4.1 software (http://www.toulouse.inra.fr; Corpet 1988). Phylogenetic and molecular evolutionary analyses were conducted using MEGA (Kumar et al. 1994) version 2.0 (S. Kumar, K. Tamura, I.B. Jakobsen, and M. Nei, in prep.) down-loaded from the MEGA site (http://www.megasoftware.net). Peptide mass analyses were performed using the computer programs maintained at the University of California-San Francisco site (http://www.donatello.ucsf.edu).

The sequence data have been deposited with the Protein Identification Resource (accession No. A59385; National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC).

References

Abrahamsson, P.A., Lilja, H., Falkmer, S., and Wodström, L.B. 1988. Immunohistochemical distribution of the three predominant secretory proteins in the parenchyma of hyperplastic and neoplastic prostate glands. *Prostate* **12:** 39–46.

Anderson, L. and Seilhamer, J. 1997. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18:** 533–537.

Baijal-Gupta, M., Clarke, M.W., Finkelman, M.A., McLachlin, C.M., and Han, V.K.M. 2000. Prostatic secretory protein (PSP94) expression in human female reproductive tissues, breast and in endometrial cancer cell lines. *J. Endocrinol.* **165:** 425–433.

Bateman, A. and Bennett, H.P.J. 1998 Granulins: The structure and function of an emerging family of growth factors. *J. Endocrinol.* **158:** 145–151.

Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P., and Tiver, K.K. 1986.

Molecular cloning of the cDNA for two major androgen-dependent secretory proteins of 18.5 kDa synthesized by the rat epididymis. *J. Biol. Chem.* **261:** 4956–4961.

Chao, C.F., Chiou, S.T., Jeng, H., and Chang, W.C. 1996. The porcine sperm motility inhibitor is identical to β-microseminoprotein and is a competitive inhibitor of Na⁺, K⁺-ATPase. *Biochem. Biophys. Res. Comm.* **218:** 623–628.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16:** 10881–10890.

Doctor, V.M., Sheth, A.R., Sinha, M.M., Arbatti, N.J., Aaveri, J.P., and Sheth N.A. 1986. Studies on immunocytochemical localization of inhibin-like material in the human prostatic tissue: Comparison of its distribution in normal, benign and malignant prostates. *Br. J. Cancer* **53:** 547–554.

Dubé, J.Y., Frenette, G., Paquin, R., Chapdelaine P., Tremblay, J., Tremblay, R.R., Lazure, C., Seidah, N., and Chrétien, M. 1987a. Isolation from human seminal plasma of an abundant 16-kDa protein originating from the prostate, its identification with a 94-residue peptide originally described as β-inhibin. *J. Androl.* **8:** 182–189.

Dubé, J.Y., Pelletier, G., Gagnon, P., and Tremblay, R.R. 1987b. Immunohistochemical localization of a prostatic secretory protein of 94 amino acids in normal prostatic tissue, in primary prostatic tumors and in their metastases. *J. Urol.* **138:** 883–887.

Fernlund, P., Granberg, L., and Roepstorff, P. 1994. Amino acid sequence of β-microseminoprotein from porcine seminal plasma. *Arch. Biochem. Biophys.* **309:** 70–76.

Fernlund, P., Granberg, L., and Larsson, I. 1996. Cloning of β- microseminoprotein of the rat: A rapidly evolving mucosal surface protein. *Arch. Biochem. Biophys.* **334:** 73–82.

Garde, S.V., Sheth, A., Porter, A.T., and Pienta, K.J. 1993a. A comparative study on expression of prostatic inhibin peptide, prostate acid phosphatase and prostate specific antigen in androgen independent human and rat prostate carcinoma cell lines. *Cancer Lett.* **70:** 159–166.

———. 1993b. Effect of prostatic inhibin peptide (PIP) on prostate cancer cell growth in vitro and in vivo. *Prostate* **22:** 225–233.

Garde, S.V., Basrur, V.S., Li, L., Finkelman, M.A., Krishan, A., Wellham, L., Ben-Josef, E., Haddad, M., Taylor, J.D., Porter, A.T., and Tang, D.G. 1999. Prostate secretory protein (PSP94) suppresses the growth of androgen-independent prostate cancer cell line (PC-3) and xenografts by inducing apoptosis. *Prostate* **38:** 118–125.

Gordon, W.L., Liu, W.K., Akiyama, K., Tsuda, R., Hara, M., Schmid, K., and Ward, D.N. 1987. Beta-microseminoprotein (β-MSP) is not an inhibin. *Biol. Reprod.* **36:** 829–835.

Hedges, S.B. 1994. Molecular evidence for the origin of birds. *Proc. Natl. Acad. Sci.* **91:** 261–264.

Hedges, S.B. and Poling, L.L. 1999. A molecular phylogeny of reptiles. *Science* **283:** 998–1001.

Kamada, M., Mori, H., Maeda, N., Yamamoto, S., Kunimi, K. Takikawa, M., Maegawa, M., Aono, T., Futaki, S., and Koide, S.S. 1998. β-Microseminoprotein/prostatic secretory protein is a member of immunoglobulin binding factor family. *Biochem. Biophys. Acta* **1388:** 101–110.

Kele, P., Krishnan, A., and Leblanc, R.M. 2000. Decapeptide that can cure prostatic cancer? 26th European Peptide Symposium. *J. Pept. Sci.* **6:** S172. Abstract.

Kohan, S., Fröysa, B., Cederlund, E., Fairwell, T., Lerner, R., Johansson, J., Khan, S., Ritzen, M., Jörnvall, H., Cekan, S., and Diczfalusy, E. 1986. Peptides of postulated inhibin activity: Lack of in vitro inhibin activity of a 94-residue peptide isolated from human seminal plasma and of a synthetic replicate of its C-terminal 28-residue segment. *FEBS Lett.* **199:** 242–248.

Kumar, S., Tamura, K., and Nei, M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci.* **10:** 189–191.

Lazure, C., Saayman, H.S., Naudé, R.J., Oelofsen, W., and Chrétien, M. 1987. Complete amino acid sequence of a VLDV-type neurophysin from ostrich differs markedly from known mammalian neurophysins. *Int. J. Pept. Protein Res.* **30:** 634–645

———. 1989. Ostrich MSEL-neurophysin belongs to the class of two-domains "big" neurophysin as indicated by complete amino acid sequence of the neurophysin/copeptin. *Int. J. Pept. Protein. Res.* **33:** 46–58.

Lazure, C., Paquet, L., Litthauer, D., Naudé, R.J., Oelofsen, W., and Chrétien, M. 1990. The ostrich pituitary contains a major peptide homologous to mammalian chromagranin-A (1–76). *Peptides* **11:** 79–87.

Lehrer, R.I., Ganz, T., and Selsted, M.E. 1991. Defensins: Endogenous antibiotic peptides of animal cells. *Cell* **64:** 229–230.

Liang, Z.G., Kamada, M., and Koide, S.S. 1991. Structural identity of imunoglobulin binding factor and prostatic secretory protein of human seminal plasma. *Biochem. Biophys. Res. Comm.* **180:** 356–359.

Lokeshwar, B.L., Hurkadli, K.S., Sheth, A.R., and Block, N.L. 1993. Human

prostatic inhibin suppresses tumor growth and inhibits clonogenic cell survival of a model prostatic adenocarcinoma, the Dunning R3327G rat tumor. *Cancer Res.* **53:** 4855–4859.

Mäkinen, M., Valtonen-André, C., and Lundwall, Å. 1999. New World, but not Old World, monkeys carry several genes encoding β-microseminoprotein. *Eur. J. Biochem.* **264:** 407–414.

Mbikay, M., Nolet, S., Fournier, S., Benjannet, S., Chapdelaine, P., Paradis, G., Dubé, J.Y., Tremblay, R., Lazure, C., Seidah, N.G., and Chrétien, M. 1987. Molecular cloning of the cDNA for a 94-amino acid seminal plasma protein secreted by the human prostate. *DNA* **6:** 23–29.

Mbikay, M., Linard, C.G., Sirois, F., Lazure, C., Seidah, N.G., and Chrétien, M. 1988. Tissue-specific expression of the prostatic secretory protein PSP$_{94}$ in cyanomolgus monkey (*Macaca fasicularis*). *Cell. Mol. Biol.* **34:** 387–398.

McLane, M.A., Marcinkiewicz, C., Vijay-Kumar, S., Wierzbicka-Patynowski, I., and Niewiarowski, S. 1998. Viper venom disintegrins and related molecules. *Proc. Soc. Exp. Biol. Med.* **195:** 168–171.

Mundel, S.D. and Sheth, N.A. 1993. Suppression of DNA synthesis and induction of apoptosis in rat prostate by human seminal plasma inhibin (HSPI). *Cell Biol. Int.* **17:** 587–594.

Naudé, R.J. and Oelofsen, W 1977. The isolation and characterization of corticotropin from the pituitary gland of the ostrich *Struthio camelus. Biochem. J.* **165:** 519–523.

Naudé, R.J., Litthauer, D., Oelofsen, W., Chrétien, M., and Lazure, C. 1993. The production of the ostrich NH$_2$-terminal POMC fragment requires cleavage at an unique signal peptidase site. *Peptides* **14:** 519–529.

Nolet, S., St-Louis, D., Mbikay, M., and Chrétien, M. 1991. Rapid evolution of prostatic protein PSP$_{94}$ suggested by sequence divergence between rhesus monkey and human cDNAs. *Genomics* **9:** 775–777.

Sasaki, T., Matsumoto, N., Jinno, Y., Niikawa, N., Sakai, H., Kanetake, H., and Saito, Y. 1996. Assignment of the human beta-microseminoprotein gene (MSMB) to chromosome 10q11.2. *Cytogenet. Cell Genet.* **72:** 177–178.

Tanaka, T., Itahana, K., Andoh, N., Takeya, T., and Sato, E. 1995. Expression of prostatic secretory protein (PSP)-like protein in porcine corpus luteum: Isolation and characterization of a new gene encoding PSP$_{94}$-like protein. *Mol. Reprod. Dev.* **42:** 149–156.

Thakur, A.N., Vaze, A.Y., Dattatreyamunthy, B., and Sheth, A.R. 1981. Isolation and characterization of inhibin from human seminal plasma. *Indian J. Exp. Biol.* **19:** 307–313.

Tirunagaru, V.G., Sofer, L., Cui, J., and Burnside, J. 2000. An expressed sequence tag database of T-cell–enriched activated chicken splenocytes: Sequence analysis of 5251 clones. *Genomics* **66:** 144–151.

Tsuda, R., Inove, T., and Hara M. 1982. A seminal plasma specific antigen of prostate gland: Forensic immunological studies of body fluids and secretion. *Jpn., J. Legal Med.* **36:** 703–709.

Ulvsbäck, M., Lindström, C., Weiber, H., Abrahamsson, P.A., Lilja H., and Lundwall, Å. 1989. Molecular cloning of a small prostate protein, known as β-microsemenoprotein, PSP$_{94}$ or β-inhibin, and demonstration of transcripts in non-genital tissues. *Biochem. Biophys. Res. Comm.* **164:** 1310–1315.

Ulvsbäck, M., Spurr, N.K., and Lundwall, Å 1991. Assignment of the human gene for β-microseminoprotein (MSMB) to chromosome 10 and demonstration of related genes in other vertebrates. *Genomics* **11:** 920–924.

Von der Kammer, H., Jurincic-Winkler, C., Horlbeck, R., Klippel, K.F., Pixberg, H.U., and Scheit, K.H. 1993. The potential use of prostatic secretory protein of 94 amino acids (PSP$_{94}$) as a serum marker for prostatic tumor. *Urol. Res.* **21:** 227–233.

Warr, G.W. 1990. A 12 kDa protein in chicken serum antigenically cross-reactive with, but unrelated to, beta$_2$-microglobulin. *Dev. Comp. Immunol.* **14:** 247–253

Weiber, H., Andersson, C., Murne, A., Rannevik, G., Lindstöm, C., Lilja, H., and Fernlund, P. 1990. β-Microseminoprotein is not a prostate-specific protein: Its identification in mucous glands and secretions. *Am. J. Pathol.* **137:** 593–603.

Weiber, H., Borch, K., Sundler, F., and Fernlund, P. 1999. Beta-microseminoprotein in gastric carcinoids: A marker of tumour progression. *Digestion* **60:** 440–448.

Xuan, J.W., Chin, J.L., Guo, Y., Chambers, A.F., Finkelman, M.A., and Clarke M.W. 1995. Alternative splicing of PSP$_{94}$ (prostatic secretory protein of 94 amino acids) mRNA in prostate tissue. *Oncogene* **11:** 1041–1047.

Xuan, J.W., Wu, D., Guo, Y., Garde, S., Shum, D.T., Mbikay, M., Zhong, R., and Chin, J.L. 1997. Molecular cloning and gene expression analysis of PSP$_{94}$ (prostate secretory protein of 94 amino acids) in primates. *DNA Cell Biol.* **16:** 627–638.

Xuan, J.W., Kwong, J., Chan, F.L., Ricci, M., Imasato, Y., Sakai, H., Fong, G.H., Panchal, C., and Chin, J.L. 1999. cDNA, genomic cloning and gene expression analysis of mouse PSP94 (prostate secretory protein of 94 amino acids). *DNA Cell Biol.* **18:** 11–26.

Yang, J.P., Baijal-Gupta, M., Garde, S.V., Fraser, J.E., Finkelman, M.A., and Clarke, M.W. 1998a. Identification of binding proteins for PSP$_{94}$ in human prostate adenocarcinoma cell lines LNCaP and PC-3. *Prostate* **35:** 11–17.

Yang, J.P., Finkelman, M.A., and Clarke, M.W. 1998b. Detection of PSP$_{94}$ and its specific binding sites in the prostate adenocarcinoma cell line LNCaP. *J. Urol.* **160:** 2240–2244.