Cloning, expression and map assignment of chicken prosaposin

Norihiro AZUMA^{*1}, Hee-Chan SEO^{*2}, Øystein LIE[†], Qiang FU^{*3}, Robert M. GOULD[‡], Masao HIRAIWA^{*}, David W. BURT[§], Ian R. PATON[§], David R. MORRICE[§], John S. O'BRIEN^{*} and Yasuo KISHIMOTO^{*}

^{*}University of California, San Diego, Department of Neurosciences, Center for Molecular Genetics, 0634J, La Jolla, CA 92093, U.S.A., [†]Norwegian College of Veterinary Medicine, Department of Morphology, Genetics, and Aquatic Biology, N-0033 Oslo, Norway, [‡]New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY 10314, U.S.A., and [§]Division of Molecular Biology, Roslin Institute, Edinburgh EH25 9PS, U.K.

Prosaposin is the precursor of four small glycoproteins, saposins A–D, that activate lysosomal sphingolipid hydrolysis. A full-length cDNA encoding prosaposin from chicken brain was isolated by PCR. The deduced amino acid sequence predicted that, similarly to human and other mammalian species studied, chicken prosaposin contains 518 residues, including four domains that correspond to saposins A–D. There was 59% identity and 76% similarity of human and chicken prosaposin amino acid sequences. The basic three-dimensional structures of these saposins is predicted to be similar on the basis of the conservation

of six cysteine residues and an N-glycosylation site. Identity of amino acid sequences was higher among saposins A, B and D than in saposin C. The predicted amino acid sequence of saposin B matched exactly that of purified chicken saposin B protein. The chicken prosaposin gene was mapped to a single locus, *PSAP*, in chicken linkage group E11C10 and is closely linked to the *ACTA2* locus. This confirms the homology between chicken and human prosaposins and defines a new conserved segment with human chromosome 10q21-q24.

INTRODUCTION

Sphingolipid degradation occurs in lysosomes. Probably because of the low solubility of sphingolipids in the aqueous cell environment, many degradation steps require protein cofactors [1–3]. Five cofactors, often referred to as 'sphingolipid activator proteins', are known, and all have been well characterized. These are saposins A, B, C and D, and G_{M2} activator protein. The four saposins are encoded in discrete domains of a single precursor protein, prosaposin. They are generated post-translationally by partial proteolysis. The human prosaposin gene is located at a single locus on chromosome 10. The four saposins are structurally very similar to each other. All are about 80-amino-acids long and contain six highly conserved cysteine residues and a single glycosylation site. Human G_{M2} activator protein is not structurally related to saposins. It is the product of a separate human gene located on chromosome 5.

Both saposins A and C stimulate the activity of glucosylceramide β -glucosidase and galactosylceramide β -galactosidase [4,5]. Saposin B has a far broader specificity that also includes stimulation of G_{M1} -ganglioside β -galactosidase, cerebroside sulphate sulphatase and globotriaosylceramide α -galactosidase [6,7]. Saposin D stimulates sphingomyelinase [8,9] and ceramidase [10]. From their specificities in enzyme-activity-stimulation assays and their lack of lipid-binding specificity [11,12], it is believed that specificity is dependent on their interactions with specific hydrolases. In addition to their effect on lysosomal enzyme activity, saposins appear to work as non-specific binding and transport proteins for these lipids [11,13]. Recently we have shown that prosaposin can also act as a neurotrophic factor [14]

and identified the amino acid sequence responsible for this activity, which resides in the saposin C domain [15].

As mentioned above, saposins share a high degree of structural similarity with each other. In addition, several other proteins that contain small domains similar to the saposins have been reported [16]. Comparison of prosaposin sequences from different species may increase our understanding not only of biogenesis of prosaposin but also structure–function relationship of these saposin proteins. To date, human [17–19], rat [20] and mouse [21] prosaposin cDNAs have been characterized. Saposin B protein was isolated and sequenced from pig kidney [22], and saposin C proteins were characterized in guinea-pig liver [23] and bovine spleen [24].

In order to gain a better insight into structure–function relationships and the gene conservation of the prosaposins in different species, it is important to examine genes from non-mammalian sources, where conserved features are more likely to be of functional significance. In the present study we have isolated and characterized a full-length cDNA encoding prosaposin from chicken brain and also identified the saposin B protein in chicken liver. The conservation of the gene and structure of the vertebrate prosaposins are discussed on the basis of a comparison of amino acid sequences and gene-mapping data.

MATERIALS AND METHODS

Isolation of chicken saposin B

Fresh chicken liver (406 g) was homogenized in 800 ml of water. To the homogenate, 4 litres of chloroform/methanol (2:1, v/v)

Abbreviations used: RT, reverse transcription; RACE, rapid amplification of cDNA ends; ORF, open reading frame.

¹ Present address and address for correspondence: Department of Applied Biochemistry, Utsunomiya University, Utsunomiya 321, Japan.

² Present address: Molecular Embryology Section, IBMB HIB-Biobuilding, N-5020, University of Bergen, Bergen, Norway.

³ Present address: Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599, U.S.A.

The nucleotide sequences presented have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number AB003471.

Figure 1 DE-52 ion-exchange cellulose column chromatography of chicken G_{M1} activator

The G_{M1} -activator fraction obtained by reverse-phase HPLC on a C4 column was separated by an ion-exchange column of DE-52 cellulose (18 mm $\times$ 180 mm). The column was eluted with an increasing concentration of sodium chloride (0–1 M; continuous diagonal line) in 10 mM sodium phosphate buffer, pH 7.0 and 5 ml fractions were collected. A 10 μ l portion of each fraction was used for the assay of the stimulation activity towards chicken G_{M1} ganglioside β -galactosidase (60 min).

were added and the mixture was stirred for 1 h at room temperature. The mixture was centrifuged at 16300 g for 1 h. The top aqueous-methanolic layer was dialysed against de-ionized distilled water. After freeze-drying it was suspended in 50 ml of water and the cofactor for G_{M1} ganglioside β -galactosidase was purified by reverse-phase HPLC on a Protein C4 column (10 mm $\times$ 250 mm; Vydac, Hesperia, CA, U.S.A.) and DE-52 anion-exchange cellulose column chromatography (18 mm $\times$ 180 mm; Whatman, Maidstone, Kent, U.K.) as described previously [8,25]. The fractions containing the stimulation activity were finally purified by re-chromatography on a Vydac Protein C4 column (4.6 mm $\times$ 250 mm).

Protein sequencing was performed as described previously [26] after Western blotting.

Stimulation assay

The stimulation assay for G_{M1} ganglioside β -galactosidase activity was performed as described previously [5] using [3 H] G_{M1} ganglioside and p -aminophenyl thio- β -D-galactoside-affinity-purified chicken β -galactosidases [27].

Cloning of chicken prosaposin

A full-length cDNA encoding chicken prosaposin was cloned using the Marathon cDNA Amplification Kit (Clontech, Palo Alto, CA, U.S.A.) with ExTaq Polymerase (Takara, Shiga, Japan). To design gene-specific primers, chicken brain polyadenylated mRNA (Clontech) was reverse-transcribed using a cDNA Cycle Kit (Invitrogen, San Diego, CA, U.S.A.). We chose three non-degenerate primers from areas of human saposin B and the junctional portion between the A and B domains, which are highly conserved among human, rat and mouse sequences. These were Primer E [CAGCTGGAGTCCAATAAGATCCC (forward, 457–479)], Primer H [TCCTTGGGTTGCATGT-

GCATCA (reverse, 767–788)] and Primer I [TTCACCTCAT-CACAGAACCCAAC (reverse, 802–824)], the numbering being based on the human prosaposin sequence [17] from the beginning of the initiation codon. PCR was performed using 100 μ M dNTP and 2.5 units of Taq polymerase in the presence of 1.5 mM $MgCl_2$ in 50 μ l. The conditions for PCR were: 30 cycles of denaturation at 95 $^{\circ}C$, 1 min; annealing at 45 $^{\circ}C$, 1 min; and extension at 72 $^{\circ}C$, 1.5 min; after initial denaturation at 95 $^{\circ}C$, 5 min. Single bands 300 and 370 bp long were obtained from the reverse transcription (RT)-PCR reactions with primer combination E-H and E-I respectively. These bands were purified with Magic PCR Prep (Promega, Madison, WI, U.S.A.), subcloned using pT7Blue(R) T-Vector (Novagen, Madison, WI, U.S.A.), and sequenced as described previously [28]. The sequence was identified as a chicken prosaposin by comparison with other saposin sequences and confirmed by the N-terminal amino acid sequencing of chicken saposin B (see below).

On the basis of the chicken saposin B nucleotide sequence, two sequence specific primers for rapid amplification of cDNA ends (RACE) were prepared: 5'-CATGTGCATCATCATCTGGA (S104, reverse; 764–783 of chicken prosaposin) for 5'-RACE and 5'-TTCTCTGAACTGACGTCT (S109, forward; 502–519) for 3'-RACE. Two fragments of about 850 bp and about 3000 bp were obtained by 5'-RACE and 3'-RACE respectively, overlapping by 283 nucleotides. Fusion of these fragments by PCR yielded a single fragment of about 3.5 kb, which was gel-purified and subcloned (see above). A total of five clones were full-length cDNA as judged by their sizes and sequencing of 5'- and 3'-untranslated regions. Sequencing was performed on both strands with gene-specific primers.

Chicken reference backcross

Genetic loci were mapped on the East Lansing reference backcross [29,30]. A partially inbred red-jungle-fowl male (UCD-001) was crossed with a highly inbred White Leghorn female (UCD-003) and the F_1 males (UCD-001 $\times$ UCD003) were crossed with the UCD-003 females to produce a backcross population.

Identification of nucleotide polymorphism in the prosaposin gene in the East Lansing reference population

Intron/exon junctions in the chicken prosaposin gene were predicted from those characterized in the human gene [19]. The forward and reverse primers S118 (1417–1443; AAAGTAGG-AGTCTGTGGAGCA) and S123 (1578–1597; GGCAAACCTT-TCAGTACTGGA) were used to amplify a 1200 bp genomic region spanning intron 13 from the parents of the East Lansing backcross, UCD-003 female and (UCD-001 $\times$ UCD-003) F_1 male [29,30]. PCR reactions contained 100 ng of target DNA, 200 μ M dNTP, 1 μ M primers, 50 mM KCl, 20 mM Tris/HCl, pH 8.4, 1.5 mM $MgCl_2$ and 2.5 units of Taq polymerase in a final volume of 100 μ l. The profile for thermal cycling was: denaturation at 94 $^{\circ}C$, 1 min; annealing at 60 $^{\circ}C$, 2 min; extension at 72 $^{\circ}C$, 3 min; for 35 cycles. The PCR products obtained from both parents were cloned and sequenced. LASERGENE (DNA-STAR) was used for sequence assembly and comparison of sequences isolated from the parents of the East Lansing backcross. From this comparison a three-base-pair difference was identified (see the Results section for details).

PCR assay for prosaposin gene polymorphism

Primers flanking this three-base-pair difference were designed and used in a PCR assay to detect this prosaposin gene

Figure 2 Purification of the chicken liver G_{M1} -activator proteins by reverse-phase HPLC

(A) The two major G_{M1} activator preparations (peaks I and II) obtained by DE-52 ion-exchange chromatography were further purified by reverse-phase HPLC using a Protein C4 column (4.6 mm $\times$ 250 mm). Fractions of volume 2 ml each were collected from 26 min after start of the programme. After freeze-drying the samples were dissolved in 1 ml of distilled water and 20 μ l of each fraction was used for assay of the stimulation activity toward chicken G_{M1} ganglioside β -galactosidase. (B) SDS/PAGE of the purified G_{M1} activators. A 5 μ g portion of each purified preparation (indicated by the horizontal bars in (A)) was subjected to Tricine/SDS/PAGE (16% gel) and the gel was stained with Coomassie Brilliant Blue. The size markers used were myoglobulins (16.9, 14.4, 10.6, 8.2 and 6.2 kDa; Sigma Chemical Co.). Lane a, peak II; lane b, peak I.

polymorphism (see the Results section for details). These were Primer-606 (6-FAM-TACCTTTCGTTTCCAATCG and Primer-607 (AATAGGAGGGAATAGGGGAGG). One of the primers was labelled with the fluorescent dye 6-FAM (blue) to enable detection on an ABI 373A DNA sequencer (Perkin-Elmer, Foster City, CA, U.S.A.). PCR reactions contained 75 ng of target DNA, 200 μ M dNTP, 0.5 μ M primers, 50 mM KCl, 20 mM Tris/HCl, pH 8.4, 3.0 mM $MgCl_2$ and 1 unit of Taq polymerase in 15 μ l. The profile for thermal cycling was: denaturation at 94 $^{\circ}$ C, 1 min; annealing at 62.5 $^{\circ}$ C, 2 min; extension at 72 $^{\circ}$ C, 3 min; for 23 cycles. After PCR, 1 μ l of the fluorescent PCR products were mixed with 0.3 μ l of TAMRA 350 internal standard (red dye) (Perkin-Elmer) and 2.7 μ l of formamide loading buffer, denatured by heating at 94 $^{\circ}$ C for 1 min, and 2 μ l was loaded on to an ABI 373A DNA sequencer with a 6% polyacrylamide gel. Genescan 672 software was used to determine the size of the PCR products, and Genotyper software was used to call alleles prior to linkage analysis.

Chromosomal mapping of the chicken prosaposin locus *PSAP*

The *PSAP* locus was mapped by genetic linkage analysis of the East Lansing backcross [29,30]. A total of 52 DNAs from this cross have been typed for over 700 loci. Data were analysed using the Map Manager program [31] and map distances (centimorgans, cM) between markers were calculated using the Kosambi function [32]. Mapping information has been submitted to the chicken genome database (<http://www.ri.bbsrc.ac.uk/>).

RESULTS

Purification of saposin B from chicken liver

The crude chloroform/methanol extraction from chicken liver was first fractionated by reverse-phase HPLC on a Protein C4 column, and stimulation of activity by each fraction toward G_{M1} ganglioside β -galactosidase was determined. The G_{M1} activator fraction was resolved into three peaks on DE-52 ion-exchange column (Figure 1). Two major forms of G_{M1} activator (peaks I and II) were further purified by re-chromatography on a C4 column (Figure 2A). Peak II gave a single protein band of 9.5 kDa on SDS/PAGE (Figure 2B, lane a), whereas peak I contained high-molecular-mass proteins in addition to the 9.5 kDa protein. N-terminal-amino acid-sequence analysis up to the tenth residue revealed that the 9.5 kDa bands observed in peak I and II had the identical sequence EDVXQDXIRL. Given that the unidentified amino acid residues were cysteine, this sequence was homologous to human saposin B, indicating that the purified activator protein was chicken saposin B. The recovery in peak II was 2.46 mg from 406 g of liver.

Cloning of prosaposin

cDNA cloning indicated that chicken prosaposin mRNA is approx. 3500 nucleotides long. Sequence analysis, together with deduction of the amino acid sequence by translation, showed that the open reading frame (ORF) consisted of 1554 nucleotides (Figure 3). The 3'-end of the cDNA contained a typical poly(A) tail and a poly(A) recognition consensus sequence AATAAA 22 bp upstream of poly(A) tail.

The chicken prosaposin ORF consists of 518 amino acids and embraces domains for four saposins that correspond to the human saposins A–D. There were 59% identity between the amino acid sequences of these prosaposins, and the similarity was 76%. The deduced amino acid sequence of saposin B domain shared the same N-terminal amino acid sequence with

cgcaggaccATGGCGAGGCGGTCTTTGACCTCCTTGGCCCTTGGCGGCCCTTG GCC	60
M A R R L L T L L G L L A A A V A	17
AGCCCTGTTTTGTGGCAGAAGGACTGTGCAAAAGGTCGAAGTCTGGTGT CAGAGTCTT	120
S P V L W Q K D C A K G P E V W C Q S L	37
CGGACAGCATCGCAGTGTGGAGCGGTGAACACTGCCAGCAGAATGTCTGGAGCAAGCCT	180
R T A S Q C G A V K H C Q Q N V W S K P	57
GCTGTAACAGCATCCCCTGTGATTGTGCAGGAACTTGAACAGTGGTTGGCAAGGTT	240
A V N S I P C D L C K E L V T V V G K V	77
TTGAAGGCAATGGCACAGAGGATGAGATCCGCTCTTACTTGGAAAAAATGCGAGTTT	300
L K D N G T E D E I R S Y L E K T C E F	97
CCTTCTGACC AAGGCTGGCCTCTGAGTGCAAAGAAATTTGGACTCTCTACCTAC CAGTT	360
L P D Q G L A S E C K E I V D S Y L P V	117
ATCATGGACATGATCAAAGAAGAATTTGATAAACCTGAAGTTGATGTAGTGCCTCTCA	420
I M D M I K E E F D K P E V V C S A L S	137
TTGTGCCAGTCCCTTCAGAAGCATCTTGCAGCAATGAAACTTCAGAAACAGCTCCAGTCC	480
L C Q S L Q K H L A A M K L Q K Q L Q S	157
AATAAGATACCAAGAGCTGGACTTCTCTGAAGTACGCTCCGTT CATGGCTAATGTGCCT	540
N K I P E L D F S E L T S P F M A N V P	177
CTACTGCTTTACCTCAAGACAACCCAGCAGAAGTCAAAGGCAACTGAGGATGTATGC	600
L L L Y P Q D K P K Q K S K A T E D V C	197
CAAGACTGCATT CGGCTGGTTACTGATGTT CAGGAAGCTGTAAAGGCAAAATGCAACTTTT	660
Q D C I R L V T D V Q E A V R T N A T	217
GTGAAGT CCTTAGTTGCTCATGCCAAGGAGGAATGTGACCGCCTCGGACCTGGAATGTCT	720
V K S L V A H A K E E C D R L G G S	237
GATATGTCCAAGAGCTATATCTGAATATTTCTGACTTGGCCATCCAGATGATGAC	780
D M C K A S Y I S E Y S D L A I Q M M M H	257
ATGAAAGATCAGCAACCAAGGACATTTGTGCCATGGTTGGGTTCTGTCTCTGTGAAA	840
M K D Q Q P K D I C A M V G F C P S V K	277
TCTGTTCCCCCTCAGACTCTGGTGCCAGCTCAAGTGTT CATGAAGTAAAAATGGAACCT	900
S V P L Q T L V P A Q V V H E V K M E T	297
GTGGAGAAAGCCACAGTTCAAGAGAAGACTTTTTCTGTGTGAAATCTGTGAGACCATG	960
V E K A T V Q E K T F S V C E I C E T M	317
GTGAAAGAAGTGACAGGCCTGTTGGAGAGCAACAGACAGGAGGAGATTTGTCATGAA	1020
V K E V T G L L E S N K T E E E I V H E	337
ATGGAGTCTGCTGCTACCTGCTCCCAGCAAGTGTCAAAGCACAGTGAAGGACTTCATA	1080
M E V V C Y L L P A S V K D Q C K D F I	357
GAGGTCTATGGCCAGGCTTTGATTGACATGCTCTGGAAGCAACAATCTGAGGCTGTG	1140
E V Y G Q A L I D M L L E A T N P E A V	377
TGTGTTATGCTGAAATCGTGTGCAGCCAAGCCTCCACAGCAGCCAGTTGTGGTGAAA	1200
C V M L K S C A A N K P P Q Q P V V V K	397
CCCCAGGTGGCTTCTGTGATATCTGCAAGATGATAGTCTTATGCAGACAAAGAACTT	1260
P A G G F C D I C K M I V A Y A D K E L	417
GAGAAAGATGCCACAACAACCGAGATTGAGGCTTTACTTGA AAAAGTCTGTCACTTCTCTG	1320
E K N A T T T E I E A L L E K V C H F L	437
CCAGAGTCTGT CAGTGATCAGTGTGTT CAGTTTGTGGAACAGTATGAACCTGTGGTTGTG	1380
P E S V S D Q C V Q F V E Q Y E P V V V	457
CAACTCTTGGCAGAGATGATGGATCCCACTTTTGTTTGC ACTAACTAGGAGTCTGTGGA	1440
Q L L A E M M D P T F V C T K L G V C G	477
GCAGCTAAAAAGCCTCTCTAGGGGATGATGCTTGTGTTTGGGGTCCAGGTTACTGGTGT	1500
A A K K P L L G D D A C V W G P G Y W C	497
AAGAATATGGAGACTGCTGCTCAGTGCAACGCTGTGGATCACTGCAGACGT CATGTGTGG	1560
K N M E T A A Q C N A V D H C R R H V W	517
AACTAGaagaagtattttccagtaactgaaagtttgccatggctcgtaaaaacgtgggctg	1620
N *	519
aggctggagagctgtatactttaaatgtccctcctctctgcctcattaacaattgaccttc	1680
aagttcctttatgtgaactcttctgtagcagtgctgctgttggaaggatcatagcttcatt	1740
gttgacttaacaagatatcttctgactgtacagtttaactcataatgctcctaanaatag	1800
tcaagtgttgtgatagattttaattggtaggctgaagtggttttcaggctgctgggaagatg	1859

the saposin B protein isolated above. Figure 4 shows the comparison of the amino acid sequence of chicken prosaposin with those of human rat and mouse. Like the human equivalent, all chicken saposin domains are expected to have 82–84 amino acids, assuming that four mature saposins exist.

Insertion of the three-amino-acid sequence Lys-Asp-Gln into the saposin B domain is commonly seen in human [33] and mouse [24] saposin B, and was predominant in chicken cDNAs; 19 out of 20 clones we analysed had this insertion, although it was not confirmed at the protein level.

The 60-amino-acid-long N-terminal chain (upstream of A domain), the 51-amino-acid interdomain chain between domains A and B, and the 37-amino-acid C-terminal chain (downstream of the D domain) of chicken prosaposin were both highly conserved and hydrophobic. Cysteine-rich domains, similar to the a-type domain of surfactant-associated protein ('SP-B') precursor in both N- and C-end chains [34], were also found in both end chains of chicken prosaposin. On the other hand, the 35- and the 12-amino-acid interdomain chains between domains B and C, and C and D, respectively displayed low similarity with those of mammalian species. In both chains, amino acids closer to domains B and D were more conserved than those closer to domain C. The 31-amino-acid insertion between domains C and D found in rat and mouse [20,21], but not in human, was not found in chicken. A threonine residue found immediately before the a-type domain in the C-end chain [34] of all other species was altered in chicken to aspartic acid.

Identification of prosaposin gene polymorphism in the East Lansing reference backcross

PCR amplification of intron 13 in the chicken prosaposin gene, and sequence analysis of the cloned products from both parents of the East Lansing reference backcross, revealed a three-base-pair insertion at position 160 in the jungle-fowl sequence (Figure 5). A PCR assay using primers flanking this 3 bp insertion/deletion was developed to detect the jungle-fowl-specific allele for the prosaposin locus *PSAP*.

The prosaposin (*PSAP*) locus maps to chicken linkage group E11C10

PSAP variants were observed using the PCR assay on genomic DNA from the East Lansing jungle fowl (hereafter called 'JF') $\times$ White Leghorn (hereafter called 'WL') backcross [29,30]. A single 184 bp allele was obtained with the UCD-003 (WL) female grandparent. The (UCD-001 $\times$ UCD-003) F₁ male produced two alleles: WL-specific 184 bp and JF-specific (UCD-001) 187 bp DNA fragments. This PCR assay was used to genotype the progeny of the JF $\times$ WL backcross. Of the 52 backcross progeny, 29 inherited the JF allele. Inheritance of *PSAP* alleles in the progeny of the backcross were compared with the inheritance of previously typed markers, and the gene for prosaposin, *PSAP*, was mapped to a single locus on linkage group E11C10 at position 58.3 cM. A summary of the results is shown in Table 1.

DISCUSSION

The present investigation unambiguously identified a gene encoding chicken prosaposin and revealed that the predicted protein is highly similar to those characterized in mammals. Like that

Figure 3 Chicken prosaposin cDNA sequence obtained by RT-PCR cloning from chicken brain and its deduced amino acid sequence

				A-type	
Chicken	1:	MARRLLTLLGLLAAVASPVWLQKDCAKGPEVWCQSLRTASQCGAVKHCQQNVWSKPAVN	60		
Human	1:	MY-ALFLLASLLGAALAGPVLGLKECTRGSAVWCQNVKTASDCGAVKHCLQTVWNKPTVK	59		
Mouse	1:	MY-ALALFASLLATALTSPVQDPKTCSSGSAVLCRDVKTAVDCGAVKHCQQMVWSKPTAK	59		
Rat	1:	MY-ALALLASLLVTALTSPVQDPKICSSGSAVVCRDVKTAVDCRAVKHCQQMVWSKPTAK	59		
				B	
Chicken	61:	SIPCDLCKELVTVVGKVLKDNNGTEDEIRSYLEKTCEFLPDQGLASECKEIVDSYLPVIMD	120		
Human	60:	SLPCDICKDVVTAAGDMLKDNATEEEILVYLEKTCDWLPKPNMSASCKEIVDSYLPVILD	119		
Mouse	60:	SLPCDICKTVVTEAGNLLKDNATQEEILHYLEKTCWEIHDSSLASCKEVVDSYLPVILD	119		
Rat	60:	SLPCDICKTVVTEAGNLLKDNATEEEILHYLEKTCAWIHDSSLASCKEVVDSYLPVILD	119		
				C	
Chicken	121:	MIKEEFDKPEVVCSSALSLCQSLQKHLAAMKLQKQLQSNKIPELDFSELTSPFMANVPLLL	180		
Human	120:	IIKGEMSRPGEVCSALNLCESLQKHLAELNHQKQLESNKIPELDMTEVVAPFMANIPLLL	179		
Mouse	120:	MIKGEMSNPGEVCSALNLCQSLQEYLAE-QNQKQLESNKIPEVDMARVVAPFMSNIPLLL	178		
Rat	120:	MIKGEMSNPGEVCSALNLCQSLQEYLAE-QNQRQLESNKIPEVDLARVVAPFMSNIPLLL	178		
				D	
Chicken	181:	YPQDKPK--QSKATEDVCQDCIRLVTDVQEAVRTNATFVKSLVAHAKEECDRLGPGMSD	238		
Human	180:	YPQDGPRSKPQPKDNGDVCQDCIQMVTDIQTAVRTNSTFVQALVEHVKEECDRLGPGMAD	239		
Mouse	179:	YPQDHPRSQPQKANEDVCQDCMKLVSDVQTAVKTNSFFIQGFVDHVKEDCDRLGPGVSD	238		
Rat	179:	YPQDRPRSQPQKANEDVCQDCMKLVTDIQTAVRTNSFFVQGLVDHVKEDCDRLGPGVSD	238		
				E	
Chicken	239:	MCKSYISEYSDLAIQMMMHMKDQPKDICAMVGFCPSVKSVPQLTLVPA-QVV-HEV-KM	295		
Human	240:	ICKNYISQYSEIAIQMMMHM---QPKEICALVGFCDEVKEMPMQTLVPAKVASKNVIPAL	296		
Mouse	239:	ICKNYVDQYSEVCVQMLMHMQDQPKKEICVLAGFCNEVKRVPMKTLVPATETIKNILPAL	298		
Rat	239:	ICKNYVDQYSEVAVQMMMHM---QPKEICVMVGFCDEVKRVPMRTLVPAATEIAKNILPAL	295		
				F	
Chicken	296:	ETVE-KATVQEK--TFSVCEICETMVKEVTGLLESNKTEEEIVHEMEVVCYLLPASVKDQ	352		
Human	297:	ELVEPIKKHEVPAKSDVYCEVCEFLVKEVTKLIDNNKTEKEILDADFDMCKSKLPKSLSEE	356		
Mouse	299:	EMDPYEQNLVQAHNVLCTQCQFVMNKFSELIVNNATEELLVKGLSNACGVLDPARTK	358		
Rat	296:	ELTDPYEQDVIQAQNVIFCQVCQLVMRKLSELIINNATEELLIKGLSKACSLLPAPASTK	355		
				G	
Chicken	353:	CKDFIEVYGQALIDMLLEATNPEAVCVMLKSCAANK-----PP	390		
Human	357:	CQEVVDTYGSSILSILLEEVSPELVCSMLHLCSGTR-----L--P-----AL-----	396		
Mouse	359:	CQEVVGTFGPSLLDIFIHEVNPSSLCGVIGLCAARPELVEALEQPAPAIVSALLKEPTPP	418		
Rat	356:	CQEVLVTFGPSLLDVLMEVNPNFLCGVISLCSANPNLVGTLEQPAIAIVSALPKPEAPP	415		
				H	
Chicken	391:	QQPVVVK-----P--AGGFCDICKMIVAYADKELEKNATTTEIEALLEKVCHFLPE	439		
Human	397:	---TV--H--V-TQ--P-KDGGFCEVCKKLVGYLDRNLEKNSTKQEILAALEKGCFLPD	445		
Mouse	419:	KQPAQPKQSALPAHVPPQKNGGFCEVCKKLVLVYLEHNLEKNSTKEEILAALEKGCFLPD	478		
Rat	416:	KQPEEPKQSALRAHVPPQKNGGFCEVCKKLVIYLEHNLEKNSTKEEILAALEKGCFLPD	475		
				I	
Chicken	440:	SVSDQCVQFVEQYEPVVVQLLAEMMDPTFVCTKLGVCGAACKPLLGDACVWGPYWCQN	499		
Human	446:	PYQKQCDQFVAEYEPVLEILVEVMDPSFVCLKIGACPSAHKPLLGTKEKCIWGPSYWCQN	505		
Mouse	479:	PYQKQCDQFVAEYEPVLEILVEVMDPGFVCSKIGVCPSAYKLLLGTEKCVWGPSYWCQN	538		
Rat	476:	PYQKQCDQFVAEYEPVLEILVEVMDPSFVCSKIGVCPSAYKLLLGTEKCVWGPYWCQN	535		
				A-type	
Chicken	500:	METAAQCAVDHCRRHVWN	518		
Human	506:	TETAAQCAVEHCKRHVWN	524		
Mouse	539:	METAARCAVDHCKRHVWN	557		
Rat	536:	SETAARCAVDHCKRHVWN	554		

Figure 4 Alignment of the amino acid sequence of chicken prosaposin with those of human [17], rat [20] and mouse [21]

Residues conserved in all four species are shaded. Highly conserved cysteine residues and the potential N-glycosylation sites are marked with asterisks and arrowheads respectively. Saposin domains A to D are indicated and are based on the human saposin domains. A-type indicates the a-type domain of surfactant-associated-protein precursor [34].

Figure 5 Nucleotide sequence of the intron-13 polymorphism used to map the chicken prosaposin gene

The three-base-pair insertion/deletion polymorphism is underlined twice at position 1582. The numbering of the amino acid and nucleotide sequence is taken from the cDNA sequence. Primer sequences for PCR (S118, P#606, P#607) are underlined with a thick horizontal arrow.

Table 1 Mapping of the prosaposin (PSAP) and α -smooth-muscle actin (ACTA2) genes defines a conserved segment between chicken and man

The genome databases used were as follows: Chicken, <http://www.ri.bbsrc.ac.uk/>; human, <http://gdbwww.gdb.org>; Mouse, <http://www.informatics.jax.org>. The cytogenetic location of loci on human chromosome 10 are shown. The location of the mouse PSAP locus is given in cM from the centromere of chromosome 10. The location of the mouse ACTA2 locus is unknown, but is predicted to map to chromosome 7, based on conservation of synteny (S).

Gene	Mapping		
	Chicken	Human	Mouse
ACTA2	E11C10 (46.7)	10q22-q24	7 (S)
PSAP	E11C10 (50.6)	10q21.1	10 (24.0)

from human, rat and mouse, the cDNA isolated from chicken brain encodes a protein with four structurally similar domains: saposins A, B, C and D. In addition, we have isolated the saposin B peptide from chicken liver and show that the N-terminal amino acid sequence was identical with that deduced from the nucleotide sequence. Chicken saposin B had isoforms similar to those found in human, and these isoforms have been ascribed to alternative splicing of exon 8 [18]; the isoform with a three-amino-acid insertion was the most abundant (19/20 cDNAs isolated). Recently it has been demonstrated that the three human saposin B isoforms expressed in baby-hamster kidney cells are equally stable and have the same sphingolipid activator function [35]. However, saposin B protein with insertions has not been isolated from any tissue so far [36,37].

Using an RNA probe based on the nucleotide sequence we confirmed that prosaposin mRNA was expressed in the chicken embryo and post-hatched-chicken brain (results not shown), which supports our previous suggestion that prosaposin may play an important role in brain development [14].

Functionally important amino acids, including cysteine and proline residues, and potential glycosylation sites are conserved in birds and mammals, suggesting that the mature-chicken saposins may have a similar structure, and most likely the same function, as the mammalian saposins. Among the four saposins, chicken saposin A displayed the highest degree of sequence identity, followed by saposins B and D. Each saposin had a particularly high degree of interspecies similarity of hydrophobic amino acids. When identical and similar amino acids were aligned together, saposin A had all 30 hydrophobic amino acids in identical positions. The percentage of hydrophobic amino acids was 36% in A, 37% in B, 28% in C and 39% in D. This high degree of conservation of hydrophobic amino acids suggests that they play an important role in saposin structure and function. This is supported by the hydrophilicity/hydrophobicity plot [38] of chicken prosaposin which is very similar to that of human prosaposin. The plot of secondary-structure predictions according to Chou and Fasman [39] suggested that the saposin B domain was the most conserved, whereas the C and D regions are significantly different as a result of the replacement of β -turn-potential amino acids by others. Recently Qi et al. [40] demonstrated that the neurotrophic and acid β -glucosidase activation properties of human saposin C are associated with the N- and C-terminal 50% of the molecule respectively. Furthermore, the β -turn structure of human saposin C spanning residues 47–62 is required for the activation of β -glucosidase and the Gln⁴⁸ is responsible for the formation of this β -turn. In chicken saposin C, Gln⁴⁸ is replaced by lysine which may prevent β -turn formation. According to the Chou and Fasman prediction, there may be another β -turn in the neurotrophic region of human saposin C spanning Leu¹⁸–Leu²⁹ [15], in which Asp²⁰ is responsible for the β -turn formation. In chicken saposin C this residue is replaced by glutamic acid, which may prevent the formation of the β -turn structure. Furthermore, the consensus amino acid residue Asn²¹ among human, rat and mouse is replaced by serine in chicken saposin C. It has been demonstrated

that a synthetic peptide with a mutation in this residue (Asn → Asp) was inactive in stimulating neurite outgrowth and also in preventing cell death in NS20Y cells and in SK-N-MC cells [15]. Although we have not tested chicken saposin C for neurotrophic activity, it is likely that this peptide lacks this function.

The genes for *PSAP* and *ACTA2* map to human chromosome 10q21-q24 [41–43] and chicken linkage group E11C10. The homology in this genomic region provides further support for the orthology of the chicken and human prosaposin genes. Other examples of conserved synteny between the human and chicken genomes have been reported recently, and include *IGF1-GAPD* [44], *PPAT-PGM2-ALB-GC* [45], *HBB-PGR*, *ACTB-EGFR*, *CA2-CALB1-MYC*, *ESR-FYN-MYB* [30], and *CD1-NRAMP1-CD2-VIL1-CD28-NDUSF1* [46]. These examples of conserved synteny provide further evidence for the high degree of genome conservation found between avian and mammalian genomes [30].

Roman et al. [19] used earlier findings on the structure and genomic organization of the human prosaposin gene to propose that the prosaposin gene was derived from an ancestral gene by two duplication events and at least one gene rearrangement. Given that the mammalian and chicken prosaposin gene share this common structure, this ancient duplication event must have occurred before the emergence of distinct avian and mammalian lineages, and the ancestral gene may have formed prior to the emergence of vertebrates. It is likely that the ancestral gene product functioned at least as a cofactor for lysosomal enzymes and that its structure and function have adapted and differentiated partially with the development of the nervous system.

We have demonstrated homology between the avian and mammalian prosaposins at the gene and genomic level. Although the basic protein structure of prosaposin seems to be conserved, there are some significant differences. The basic function as a cofactor for lysosomal hydrolysis of sphingolipids has been conserved, probably in all vertebrates. However, comparison of critical amino acid residues predicts that the neurotrophic activity present in the saposin C domain in mammals is probably absent in chicken. Whether this is a 'loss of function' in the chicken or a 'gain of function' in mammals requires further analysis of other non-mammalian saposins.

We thank Dr. L. B. Crittenden (USDA Agricultural Research Service, East Lansing, MI, U.S.A.) for providing DNA samples of the reference mapping. This work was supported in part by National Institutes of Health grants NS13559 (to Y.K.), NS08682 (to J.S.O.), the Biotechnology and Biological Sciences Research Council, (U.K.) (to D.B.) and the Ministry of Agriculture, Fisheries and Food (U.K.) (to D.B.).

REFERENCES

- O'Brien, J. S. and Kishimoto, Y. (1991) FASEB J. **5**, 301–308
- Kishimoto, Y., Hiraiwa, M. and O'Brien, J. S. (1992) J. Lipid Res. **33**, 1255–1268
- Sandhoff, K., Conzelmann, E., Neufeld, E., Koback, M. M. and Suzuki, K. (1989) in Metabolic Basis of Inherited Disease, 6th edn, (Scriver, C. R., Beaudet, A. L., Sly, W. S. and Valle, D., eds.), pp. 1807–1839, McGraw-Hill, New York
- Wenger, D. A., Sattler, M. and Roth, S. (1982) Biochim. Biophys. Acta **712**, 639–649
- Morimoto, S., Kishimoto, Y., Tomich, J., Weiler, S., Ohashi, T., Barranger, J. A., Kretz, K. A. and O'Brien, J. S. (1990) J. Biol. Chem. **265**, 1933–1937
- Li, S.-C., Sonnino, S., Tettamanti, G. and Li, Y.-T. (1988) J. Biol. Chem. **263**, 6588–6591
- Li, S.-C., Kihara, H., Serizawa, S., Li, Y.-T., Fluharty, A. L., Mayes, J. S. and Shapiro, L. J. (1985) J. Biol. Chem. **260**, 1867–1871
- Morimoto, S., Martin, B., Kishimoto, Y. and O'Brien, J. S. (1988) Biochem. Biophys. Res. Commun. **158**, 403–410
- Tayama, M., Soeda, S., Kishimoto, Y., Martin, B. M., Callahan, J. W., Hiraiwa, M. and O'Brien, J. S. (1993) Biochem. J. **290**, 401–404
- Azuma, N., O'Brien, J. S., Moser, H. M. and Kishimoto, Y. (1994) Arch. Biochem. Biophys. **311**, 354–357
- Hiraiwa, M., Soeda, S., Kishimoto, Y. and O'Brien, J. S. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 11254–11258
- Soeda, S., Hiraiwa, M., O'Brien, J. S. and Kishimoto, Y. (1993) J. Biol. Chem. **268**, 18519–18523
- Vogel, A., Schwarzmann, G. and Sandhoff, K. (1991) Eur. J. Biochem. **200**, 591–597
- O'Brien, J. S., Carson, G. S., Seo, H. C., Hiraiwa, M. and Kishimoto, Y. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 9593–9596
- O'Brien, J. S., Carson, G. S., Seo, H. C., Hiraiwa, M., Weiler, S., Tomich, J. M., Barranger, J. A., Kahn, M., Azuma, N. and Kishimoto, Y. (1995) FASEB J. **9**, 681–685
- Munford, R. S., Sheppard, P. O. and O'Hara, P. J. (1995) J. Lipid Res. **36**, 1653–1663
- O'Brien, J. S., Kretz, K. A., Dewji, N. N., Wenger, D. A., Esch, F. and Fluharty, A. L. (1988) Science **241**, 1098–1101
- Holtschmidt, H., Sandhoff, K., Fürst, W., Kwon, H. Y., Schnabel, D. and Suzuki, K. (1991) FEBS Lett. **280**, 267–270
- Roman, E. G., Scheinker, V. and Grabowski, G. A. (1992) Genomics **13**, 312–318
- Collard, M., Sylvester, S., Tsuruta, J. and Griswold, M. (1988) Biochemistry **27**, 4557–4564
- Tsuda, M., Sakiyama, T., Endo, H. and Kitagawa, T. (1992) Biochem. Biophys. Res. Commun. **184**, 1266–1272
- Fluharty, A. L., Katona, Z., Meek, W. E., Frei, K. and Fowler, A. V. (1992) Biochem. Med. Metab. Biol. **47**, 66–85
- Sano, A., Radin, N. S., Johnson, L. L. and Tarr, G. E. (1989) J. Biol. Chem. **263**, 19597–19601
- Sano, A., Mizuno, T., Kondoh, K., Hineno, T., Ueno, S., Kakimoto, Y. and Morita, N. (1992) Biochim. Biophys. Acta **1120**, 75–80
- Morimoto, S., Yamamoto, Y., O'Brien, J. S. and Kishimoto, Y. (1990) Anal. Biochem. **190**, 154–157
- Hiraiwa, M., O'Brien, J. S., Kishimoto, Y., Galdzicka, M., Fluharty, A. L., Ginns, E. I. and Martin, B. M. (1993) Arch. Biochem. Biophys. **304**, 110–116
- Yamamoto, Y., Hake, C. A., Martin, B. M., Keith, K. A., Ahern-Rindel, A. J., Naylor, S. L., Mudd, M. and O'Brien, J. S. (1990) DNA Cell Biol. **9**, 119–127
- Seo, H. C., Yang, M., Tonlorenzi, R., Willems, P. J., Kim, A. H., Filocamo, M., Gatti, R., DiCioccio, R. A. and O'Brien, J. S. (1994) Hum. Mol. Genet. **3**, 2065–2066
- Crittenden, L. B., Provencher, L., Santangelo, L., Levin, I., Abplanalp, H., Briles, R. W., Briles, W. E. and Dodgson, J. B. (1993) Poultry Sci. **72**, 334–348
- Burt, D. W., Bumstead, N., Bitgood, J. J., Ponce de Leon, F. A. and Crittenden, L. B. (1995) Trends Genet. **11**, 190–194
- Manly, K. F. (1993) Mamm. Genome **4**, 303–313
- Kosambi, D. D. (1994) Ann. Eugenics **12**, 172–175
- Nakano, T., Sandhoff, K., Stümper, J., Christonanou, H. and Suzuki, K. (1989) J. Biochem. (Tokyo) **105**, 152–154
- Pathy, L. (1991) J. Biol. Chem. **266**, 6035–6037
- Henseler, M., Klein, A., Glombitza, G. J., Suzuki, K. and Sandhoff, K. (1996) J. Biol. Chem. **271**, 8416–8423
- Fluharty, A. L. (1995) Trends Genet. **7**, 167–189
- Fürst, W., Schubert, J., Machleidt, W., Meyer, H. E. and Sandhoff, K. (1990) Eur. J. Biochem. **192**, 709–714
- Hopp, T. P. and Woods, K. R. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 3824–3828
- Chou, P. Y. and Fasman, G. D. (1978) Annu. Rev. Biochem. **47**, 251–276
- Qi, X., Qin, W., Sun, Y., Kondoh, K. and Grabowski, G. A. (1996) J. Biol. Chem. **271**, 6874–6880
- Fujibayashi, S., Kao, F. T., Jones, C., Morse, H., Law, M. and Wenger, D. A. (1985) Am. J. Hum. Genet. **37**, 741–748
- Inui, K., Kao, F. T., Fujibayashi, S., Jones, C., Morse, H. G., Law, M. L. and Wenger, D. A. (1985) Hum. Genet. **69**, 197–200
- Bar-Am, I., Avivi, L. and Horowitz, M. (1996) Cytogenet. Cell Genet. **72**, 316–318
- Klein, S., Morrice, D. R., Sang, H., Crittenden, L. B. and Burt, D. W. (1996) J. Hered. **87**, 10–14
- Palmer, D. K. and Jones, C. (1986) J. Hered. **77**, 106–108
- Hu, J., Bumstead, N., Burke, D., Ponce de Leon, F. A., Skamene, E., Gros, P. and Malo, D. (1995) Mamm. Genome **6**, 809–815