J Mol Evol (1998) 46:361–369

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Analysis of Three Lipocalin-Like Proteins Present in the Milk of *Trichosurus vulpecula* (Phalangeridae, Marsupialia)

Christine P. Piotte, Airlie K. Hunter, Craig J. Marshall, Murray R. Grigor*

Department of Biochemistry and Centre for Gene Research, University of Otago, Dunedin, New Zealand

Received: 28 October 1996 / Accepted: 19 May 1997

Abstract. Three proteins have been identified in the milk of the common brush tail possum, *Trichosurus vulpecula* that from sequence analysis are members of the lipocalin family. They include β-lactoglobulin, which appears to have two forms; a homologue to the late-lactation protein found in tammar, *Macropus eugenii;* milk; and a novel protein termed *trichosurin.* Whereas β-lactoglobulin and trichosurin are both expressed throughout lactation, the late-lactation protein is not detected in samples taken before days 100–110 of lactation. The cDNAs encoding each of these proteins have been isolated from cDNA libraries prepared using possum mammary mRNA and sequenced. Phylogenetic analysis showed that the *T. vulpecula* β-lactoglobulin, along with two other macropod β-lactoglobulins, forms a subclass of β-lactoglobulins distinct from those for eutherian mammals; both marsupial late-lactation proteins appear to have similarities to a family of odorant-binding proteins, whereas trichosurin has similarities to the major urinary proteins of rodents.

Key words: Marsupial — Milk — Lipocalin — β-lactoglobulin — Late-lactation protein — Trichosurin — Phylogeny

Introduction

In metatherians (marsupials), much of the reproductive effort is directed toward lactation rather than gestation. The marsupial young is born at a very altricial stage and has to complete its development inside the mother's pouch. Lactation has two distinct phases. The initial period lasting approximately 120 days in *Trichosurus vulpecula* defines the early phase of lactation (Pilton and Sharman 1962), which has some characteristics of an extended eutherian colostral phase and is functionally equivalent to an external end of gestation. During the second phase, referred as late lactation, the young gains weight and switches progressively to a vegetarian diet and independency. Dramatic quantitative and qualitative variations of milk composition that occur between the two phases are well documented in *T. vulpecula* (Gross and Bolliger 1959; Cowan 1989; Crisp et al. 1989; Grigor et al. 1991; Piotte and Grigor 1996). For the most part the marsupial milk proteins appear to be similar to those found in eutherian milk although, to date, only a few marsupial milk genes have been cloned.

We have been involved in a systematic study of the major milk proteins and their expression in *T. vulpecula,* the common Australian brush-tailed possum, with the goal of identifying proteins expressed specifically in either early or late lactation. Recently we reported the identification, isolation, and cloning of a small glycoprotein expressed only in early lactation (Piotte and Grigor 1996). Earlier we had reported that transferrin was expressed primarily in late lactation (Grigor et al. 1991). Other proteins, such as α-lactalbumin and lysozyme are expressed throughout lactation (Piotte et al. 1997) as are

* *Present address:* AgResearch Dairy Science Group, Ruakura Research Centre, P.O. Box 3123, Hamilton, New Zealand
Correspondence to: M.R. Grigor; e-mail: grigorm@agresearch.cri.nz

two caseins, identified by their similarity to the wallaby proteins, as α- and β-casein. In this paper, we report the cDNA cloning of three other milk protein genes from *T. vulpecula*. The derived protein sequences indicate that all three belong to the lipocalin family. Two were identified as β-lactoglobulin and late-lactation protein (Nicholas et al. 1987). The third, however, is a novel milk protein, referred to as *trichosurin* by ourselves, which is related to the rodent major urinary proteins. This is the first time that three lipocalin-like members have been found in the milk of a single species.

The lipocalin family contains some 40 nonredundant secretory ligand-binding proteins that, despite having little sequence homology, have similar three-dimensional structures consisting of an eight-stranded antiparallel β barrel (Papiz et al. 1986; Flower 1996). Although it is known that many may transport small hydrophobic molecules (Pervaitz and Brew 1985; Flower 1996), in most cases, their physiological function remains to be established.

β-Lactoglobulin is the major whey protein of ruminant milk. It is also found in the milk of nonruminant eutherians and has been identified in the milk of two macropod marsupials, *Macropus giganteus*, the eastern grey kangaroo (Godovac-Zimmermann and Shaw 1987), and *Macropus eugenii*, the tammar wallaby (Collet et al. 1991). β-Lactoglobulin is thought to transport retinol and/or fatty acids (Godovac-Zimmermann 1988; Flower 1996).

Late-lactation protein has only been detected in marsupial milk to date. It is expressed at a high level during the late-lactation phase by *M. eugenii* mammary gland, accounting for 25% of total milk proteins (Nicholas et al. 1987). Very strong linkage disequilibrium has been shown between alleles of β-lactoglobulin and late-lactation protein in *M. eugenii* (Woodley et al. 1993).

Methods

Tissue Samples. Milk samples were collected as previously described (Piotte and Grigor 1996) and kept frozen at −70°C until further analy-

Table 1. List of the protein sequences used in this work, their GenBank locus name and accession number, and the code name used in this study

Database name	Accession number	Species	Type	Name
BTLGB	X14712	Bovine	BLG[a]	BLG cow
LACB_BUBAR	P02755	Buffalo	BLG	BLG buffalo
LACB_FELCA	P33687	Cat	BLG I	BLG I cat
LACA_FELCA	P21664	Cat	BLG II	BLG II cat
LACC_FELCA	P33688	Cat	BLG III	BLG III cat
LACA_CANFA	P33685	Dog	BLG	BLG dog
LACC_CANFA	P33686	Dog	BLG III	BLG III dog
LACB_EQUAS	P13613	Donkey	BLG I	BLG I donkey
LACA_EQUAS	P19647	Donkey	BLG II	BLG II donkey
CHBETLACA	Z19569	Goat	BLG	BLG goat
LACB_HORSE	P02758	Horse	BLG I	BLG I horse
LACA_HORSE	P07380	Horse	BLG II	BLG II horse
LACB_MACGI	P11944	Kangaroo	BLG	BLG roo
SSBLACMR	X54976/S42852	Pig	BLG	BLG pig
TVU34289	U34289	Possum	BLG	BLG possum
OALGLBR	X04520	Sheep	BLG I	BLG I sheep
MEBLGL	X15212	Wallaby	BLG	BLG wallaby
MUP1_MOUSE	P11588	Mouse	MUP	MUP 1
MUPM_MOUSE	P04939	Mouse	MUP	MUP M
MUP2_MOUSE	P11589	Mouse	MUP	MUP 2
MUP4_MOUSE	P11590	Mouse	MUP	MUP 4
MUP5_MOUSE	P11591	Mouse	MUP	MUP 5
MUP6_MOUSE	P02762	Mouse	MUP	MUP 6
MUP8_MOUSE	P04938	Mouse	MUP	MUP 8
MUP_RAT	P02761	Rat	MUP	MUP rat
OBP_BOVIN	P07435	Bovine	OBP	OBP cow
RATTY2G12A	M76734	Rat	OBP	OBP rat
TVU34287	U34287	Possum	LLP	LLP possum
MELLP	X15213	Wallaby	LLP	LLP wallaby
PP14_HUMAN	P09466	Human	PP14	PP human
VEGP_HUMAN	P31025	Human	VEGP	VEGP human
VEGP_RAT	P20289	Rat	VEG	VEGP rat
	U40376	Possum	Trichosurin	TRICH possum

[a] Abbreviations: BLG, β-lactoglobulin; LLP, late lactation protein; MUP, major urinary protein; OBP, odorant binding protein; VEGP, von Ebner's gland protein

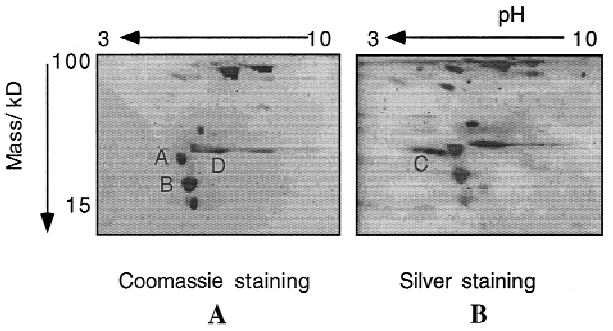

Fig. 1. Two-dimensional electrophoresis of *T. vulpecula* whey. Whey samples from a possum lactating for 110 days (switch period) were analyzed by two-dimensional electrophoresis as described in the text and stained with Coomassie blue **(A)** and silver **(B).** The spots labeled *A* and *B* had N-terminal sequences corresponding to β-lactoglobulin, *spot C* was identified as late lactation protein, and *spot D* represents a novel protein termed *trichosurin.*

sis. Mammary tissue samples were collected from captured feral *T. vulpecula,* held at AgResearch, Invermay Research Centre, Mosgiel, New Zealand, and frozen in liquid nitrogen for RNA analysis.

Milk Analysis. Samples of whey were analyzed using two-dimensional polyacrylamide gel electrophoresis and electroblotted onto PVDF (Problott) membranes. Individual proteins were subjected to N-terminal amino acid determination by Edman degradation using an ABI gas-phase sequencer with on-line PTH analyser (Applied Biosystems). This analysis was performed by the Protein Microchemistry Facility, Department of Biochemistry, University of Otago, Dunedin, New Zealand.

RNA Analysis. Total RNA was extracted from *T. vulpecula* mammary tissues according to Chomczynski and Sacchi (1987). For Northern blot transfer, RNA (~10 μg) was denatured in 2.2 M formaldehyde and separated by electrophoresis in 1% agarose gels according to Sambrook et al. (1992). RNA was electroblotted onto Hybond-N+ membrane (Amersham). Membranes were prehybridized at 65°C in 6 × SSC (20 × SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0), 6 × Denhardt's (50 × Denhardt's: 5% Ficoll, 5% polyvinyl-pyrrolidone, 5% bovine serum albumin), and 0.5% SDS. Hybridization was carried out in 20 ml of fresh buffer with [^{32}P]-dCTP-labeled cDNA probe for 16 h. Membranes were washed at 65°C in 2 × SSC, 0.1% SDS, and 1 × SSC, 0.1% SDS for 2 × 30 min each. Washed membranes were exposed to Cronex X-ray films at −80°C with intensifying screen and the film was subsequently developed in an Allpro-100 film developer.

cDNA Libraries and Screening Procedure. Early and late-lactation polyA+ RNA fractions were purified from total RNA extracted from early and late-lactation *T. vulpecula* mammary glands, respectively, using an Oligotex-dT mRNA Midi Kit (Qiagen). cDNA libraries were prepared with the Zap-cDNA Synthesis Kit (Stratagene). The screening was performed according to Sambrook et al. (1992). β-Lactoglobulin and late-lactation protein clones were isolated using [^{32}P]-dCTP-labeled *M. eugenii* β-lactoglobulin and late-lactation protein cDNA probes (Collet et al., 1989, 1991). Trichosurin clones were isolated by random selection of phagemid plaques that gave a strong signal when hybridized with α[^{32}P]-dCTP-labeled total *T. vulpecula* cDNA probe but did not hybridize with other possum milk protein cDNAs.

Sequence Analysis. Several β-lactoglobulin, late-lactation protein, and trichosurin cDNA clones containing inserts truncated in the non-translated 5′ region were subjected to DNA sequencing in both directions using an ABI 373 automated DNA sequencer (Applied Biosystems). Sequences obtained were analyzed using the software package of Genetics Computer Group, Version 8 (Madison, Wisconsin, USA).

Phylogenetic Analysis. Protein sequences (Table 1) were aligned by the programs AMPS (Barton and Sternberg 1987) and HOMED (Stockwell 1988). No manual alignment of the sequences was found to be necessary. Aligned sequences were analyzed by the programs from the PHYLIP package of phylogenetic software (Felsenstein 1988). The program SEQBOOT was used to generate 250 bootstrapped alignments which were then analyzed by maximum parsimony methods. Maximum parsimony analyses were done on bootstrapped sequence alignments using the program PROTPARS. The program CONSENSE was used to generate a consensus tree, which was further processed as previously described (Winefield et al. 1995) by DRAWTREE and DRAWGRAM to produce the final trees. All phylogenetic reconstruction methods make assumptions that may not be true of the data set under consideration. For example, each nucleotide or amino acid of sequence data is assumed to evolve independently, an assumption that is almost certainly not true for parts of most sequences. The ability of different phylogenetic approaches to reconstruct a known phylogeny has been examined (Hillis et al. 1992, 1994). Their findings suggest that despite limitations in data sets, maximum parsimony approaches perform slightly better than distance methods and that both approaches are capable of acceptable phylogeny reconstruction.

Results

Protein Analysis

Two-dimensional analysis of *T. vulpecula* whey displayed several proteins which were subsequently subjected to N-terminal sequence analysis. Two spots labeled A and B on Fig. 1 gave the sequence IENIYSK which enabled their identification as β-lactoglobulin by comparison with the *M. eugenii* sequence. A third spot (C), which was detectable only in late-lactation samples after silver staining, was not present in sufficient quantities to derive an N-terminal sequence but, based on its pattern of expression, molecular mass, and isoelectric point, it was identified as a possum homologue of the *M. eugenii* late-lactation protein. All samples analyzed showed the presence of a diffuse spot corresponding to a major protein (D). N-terminal sequence analysis of this protein gave the following sequence—LQPECSR. A search of the databases showed that this protein was not similar to any known milk protein. We have termed this protein *trichosurin.* In contrast to the late-lactation protein, both β-lactoglobulin species and the trichosurin were found in milk samples taken throughout lactation (data not shown).

Cloning and Sequencing T. vulpecula *Genes*

Screening a combination of early (70 days) and late (145 days)-lactation cDNA libraries with the *M. eugenii* β-lactoglobulin and late-lactation protein cDNAs led to the isolation of multiple clones for each gene which were subsequently sequenced to give the entire coding se-

A

```
1                                                          60
ATGAAATTCCTTCTGCTTACTGTTGGATTGACGAGCATTTGTGCCATCCAGGCCATTGAA
 M  K  F  L  L  L  T  V  G  L  T  S  I  C  A  I  Q  A  I  E
61                                                         120
AACATCTACTCCAAGGAAGAGTTAGTGGTTGAAAAGCTTATAGGGCCATGGTATCGCGTA
 N  I  Y  S  K  E  E  L  V  V  E  K  L  I  G  P  W  Y  R  V
121                                                        180
GAAGAAGCAAAGGCCATGGAATTCTCAATTCCCTTGTTTGACATGAACATTAAGGAAGTG
 E  E  A  K  A  M  E  F  S  I  P  L  F  D  M  N  I  K  E  V
181                                                        240
AACCGGACTCCTGAAGGCAATCTGGAACTGATTGTGTTAGAGCAGACTGACAGTTGTGTT
 N  R  T  P  E  G  N  L  E  L  I  V  L  E  Q  T  D  S  C  V
241                                                        300
GAGAAGAAGTTTCTTCTAAAGAAAACTGAGAAGCCTGCTGAGTTTGAGATCTACATCCCT
 E  K  K  F  L  L  K  K  T  E  K  P  A  E  F  E  I  Y  I  P
301                                                        360
TCAGAGTCTGCTTCATACACACTCTCTGTGATGGAAACTGATTATGACAACTACATCCTC
 S  E  S  A  S  Y  T  L  S  V  M  E  T  D  Y  D  N  Y  I  L
361                                                        420
GGTTGCTTGGAGAACGTCAACTACCGGGAGAAAATGGCATGCGCCCACTATGAGAGAAGA
 G  C  L  E  N  V  N  Y  R  E  K  M  A  C  A  H  Y  E  R  R
421                                                        480
ATAGAGGAGAACAAAGGAATGGAAGAATTCAAGAAAATTGTCAGGACTTTGACCATACCC
 I  E  E  N  K  G  M  E  E  F  K  K  I  V  R  T  L  T  I  P
481                                                525
TATACAATGATTGAAGCCCAAACAAGAGAAATGTGCCGTGTTTAG
 Y  T  M  I  E  A  Q  T  R  E  M  C  R  V  *
```

B

```
1                                                          60
ATGAAGGTCCTGTTCTTCACCATTGCACTAAGCCTGTTCTCTATCCTCCATGCTGATGAT
 M  K  V  L  F  F  T  I  A  L  S  L  F  S  I  L  H  A  D  D
61                                                         120
GTTGCATTTTCTGCTTTTACACCATCGGAGGGCACATACTATGTACAAGTCATTGCAGTA
 V  A  F  S  A  F  T  P  S  E  G  T  Y  Y  V  Q  V  I  A  V
121                                                        180
GACAAAGAATTTCCTGAGGAGGAGATACCTAGGGACATGTCACCTTTGACCATCATGTAC
 D  K  E  F  P  E  E  E  I  P  R  D  M  S  P  L  T  I  M  Y
181                                                        240
CTTGATGATGGCAAAATGGAGGCCAGATTTACCATGAAGAAGGATGACAACTGTGAAGAG
 L  D  D  G  K  M  E  A  R  F  T  M  K  K  D  D  N  C  E  E
241                                                        300
ATTAATATAATGTTGGAAAAAACAGCTGACCCCAGGAAAATTACCATGAACAGACGCCTA
 I  N  I  M  L  E  K  T  A  D  P  R  K  I  T  M  N  R  R  L
301                                                        360
CGCTACACCTGTGCTGCAGTAAGAACATCCAAGCAGAAACATTGGATTCTAGTCTGTCCA
 R  Y  T  C  A  A  V  R  T  S  K  Q  K  H  W  I  L  V  C  P
361                                                        420
AGAGAGTTTCAAGGCGAGACAATCAGAATGGCTAAACTTGTGGGTCCAAATACAGACAAG
 R  E  F  Q  G  E  T  I  R  M  A  K  L  V  G  P  N  T  D  K
421                                                        480
AACCCAAAGGCCCTGGAGGATTTTTATAGGTTTATCTACAGAGAACGATTTGATAAAAGA
 N  P  K  A  L  E  D  F  Y  R  F  I  Y  R  E  R  F  D  K  R
481                                                   531
AGAATTATCACCCCCAAACAAACTGAGGCCTGTGCCCCAGAACATGCTTAA
 R  I  I  T  P  K  Q  T  E  A  C  A  P  E  H  A  *
```

Fig. 2. DNA sequences for the coding regions of *T. vulpecula* β-lactoglobulin **(A),** late lactation protein **(B),** and trichosurin **(C).** Translated sequences are also included and signal sequences are *underlined.*

quence. Several trichosurin cDNA clones were obtained by screening the early lactation library with a cDNA made from total mammary mRNA and taking strongly hybridizing clones that did not produce hybridization signals with a mixture of cDNAs for *T. vulpecula* α-lactalbumin, β-lactoglobulin, early lactation protein, and α- and β-casein. These clones were subjected to DNA sequencing and those which gave a sequence that translated to give the N-terminal sequence for trichosurin detected above were then completely sequenced in both directions. The DNA sequences and inferred protein sequences for the three proteins are shown in Fig. 2.

The three genes, β-lactoglobulin, late-lactation protein, and trichosurin, had coding sequences of 522, 528, and 540 bp, respectively, and encoded polypeptides of 176, 176, and 180 residues. Each contained a signal sequence 18 residues long for β-lactoglobulin and late-lactation protein and 15 residues long for trichosurin. When the derived protein sequences were compared, it was apparent that there are regions in each protein with higher levels of similarity with one or the other protein (Fig. 3A). For example, late-lactation protein and trichosurin show considerable similarity between residues 50 and 60 (numbering as in Fig. 3A); late-lactation protein

C

```
1                                                                   60
ATGAAGCTCCTGCTGCTGAGTATGGGCTTGGCCCTGGTCTGTGGCCTCCAACCTGAGTGT
 M  K  L  L  L  L  S  M  G  L  A  L  V  C  G  L  Q  P  E  C
61                                                                 120
AGCAGATCAGAGGAAGACCTGTCAGATGAAAAGGAACGAAAATGGGAACAGCTTTCTAGA
 S  R  S  E  E  D  L  S  D  E  K  E  R  K  W  E  Q  L  S  R
121                                                                180
CACTGGCACACTGTTGTATTGGCCTCAAGTGACAGATCTCTGATTGAAGAGGAAGGTCCC
 H  W  H  T  V  V  L  A  S  S  D  R  S  L  I  E  E  E  G  P
181                                                                240
TTTAGGAATTTTATCCAAAATATCACCGTAGAGAGTGGAAACTTGAATGGATTCTTTCTG
 F  R  N  F  I  Q  N  I  T  V  E  S  G  N  L  N  G  F  F  L
241                                                                300
ACAAGGAAAAATGGCCAGTGCATTCCATTATATCTTACTGCTTTCAAGACTGAGGAAGCA
 T  R  K  N  G  Q  C  I  P  L  Y  L  T  A  F  K  T  E  E  A
301                                                                360
CGTCAGTTTAAATTGAACTATTATGGAACTAATGATGTCTACTATGGAAGTTCTAAGCCA
 R  Q  F  K  L  N  Y  Y  G  T  N  D  V  Y  Y  G  S  S  K  P
361                                                                420
AATGAATATGCCAAATTCATCTTCTATAACTACCATGATGGGAAAGTGAACGTTGTGGCA
 N  E  Y  A  K  F  I  F  Y  N  Y  H  D  G  K  V  N  V  V  A
421                                                                480
AACCTCTTTGGCCGGACTCCAAATCTAAGCAATGAAATCAAGAAAAGATTTGAGGAAGAT
 N  L  F  G  R  T  P  N  L  S  N  E  I  K  K  R  F  E  E  D
481                                                                540
TTTATGAACAGAGGATTTAGGAGGGAAAACATTTTAGATATATCTGAAGTTGATCATTGC
 F  M  N  R  G  F  R  R  E  N  I  L  D  I  S  E  V  D  H  C
  543
TAG
 *
```

Fig. 2. Continued.

and β-lactoglobulin show similarity between residues 155 to 160; and whereas there are conserved cysteines in all three proteins at residues 88 and 185, late-lactation protein and β-lactoglobulin but not trichosurin contain a cysteine at residue 131. The cysteines identified at positions 88 and 185 (Fig. 3A) correspond to cysteines known to form disulphide bonding in ruminant β-lactoglobulin near the base of the putative retinol binding site (Papiz et al. 1986). Two other cysteines were also detected in the marsupial β-lactoglobulin sequences at residues 104 and 116 (Fig. 3B numbering) and these also correspond to cysteines involved in disulphide bonding between β strands G and H in ruminant β-lactoglobulin (Papiz et al. 1986). By contrast, the two late-lactation protein sequences shown both have a pair of cysteines at residues 85 and 101 (Fig. 3C numbering). Assuming that the late-lactation protein molecules have a similar β-barrel structure to that of β-lactoglobulin, these cysteines would be sufficiently close to form a disulphide bond between beta strands F and G. Trichosurin has only a single pair of cysteines.

Overall the three marsupial β-lactoglobulins were very similar to the *T. vulpecula* β-lactoglobulin sequence, showing 73 and 74% identity (86 and 87% similarity) to each of the macropod β-lactoglobulins (Fig. 3B). By contrast, when compared with a eutherian β-lactoglobulin from the cow, there was only 29% identity (53% similarity). *T. vulpecula* and *M. eugenii* late-lactation proteins were also very similar, sharing 74% identical amino acids (82% similarity) (Fig. 3C). The nearest homologues detected were a bovine odorant binding protein and the human von Ebner's gland protein, where the level of identity with the *T. vulpecula* late-lactation protein was 18 and 32%, respectively (41 and 54% similarity). No milk protein homologue for trichosurin could be detected, and the most similar protein was a rat major urinary protein where there was 33% identity (52% similarity) (Fig. 3D).

Phylogenetic analyses of these members of the lipocalin family produced a tree similar to that of Sansom et al. (1994). The branching patterns found in the trees (Fig. 4) are, for the most part, well supported. In addition to a maximum parsimony analysis, distance trees were also constructed using the program NEIGHBOR. The results from this analysis were essentially the same as those from the maximum parsimony analysis and are not shown. Three main branches are evident: the β-lactoglobulin family, a group of lipocalins including odorant binding proteins, homologues of von Ebner's gland protein and the marsupial late-lactation proteins, and a final group related to the rodent urinary proteins that includes the *T. vulpecula* trichosurin. The marsupial β-lactoglobulins form a strongly supported group adjacent to but distinct from other mammalian β-lactoglobulins. The part of the tree containing the late-lactation proteins and the odorant binding proteins showed the least support for the reported branching pattern, reflecting considerable rearrangement of the members of this group.

Expression of T. vulpecula *Lipocalins*

Northern hybridization of RNA extracted from *T. vulpecula* mammary glands at different stages of lactation

(A)

```
                1                                                    50
      LLP       MKvLFFTIaL sLfsiLhAD. ...DvaFSa. ....FtpseG tYYvqViA.v
trichosurin     MKLLLLsmGL aLVCgLQpEc srsEEdLSde kErkWEqLsr hWhtVVlAss
β-lactoglobulin MKFLLLTVGL tsICaiQA.. ..iEniyS.k eElvvEkLiG pWYrVeeA..
Consensus       **-*-----* ---------- -------*-- ---------- -------*--

               51                                                   100
     LLP        DKeFpEEEiP rDMspltImy lDDGkMEarF tMkKdDnCeE iniMLeKTaD
trichosurin     DrsLiEEEgP FrnfIqnIt. vEsGNLngFF ltrKngqCIp lyLtafKTEE
β-lactoglobulin .KaMefsipl FDMnIkeVnr tpEGNLE.Li vLeqtDsCVE kkFLLkKTEk
Consensus       ---------- ---------- ---*------ -------*-- ------**--

               101                                                  150
     LLP        PRkitMNrrl rYtcaaV..r tSKqkhWILv CpreFqgeTI rMAklvgpnT
trichosurin     aRqFkLN... yYgtndVyyg sSKpneYakf iFyNYhdgkV nvvanlFgRT
β-lactoglobulin PaeFeiyips esasytlsvm etdydnYILg CLeNvnyrek .MAcahYeRr
Consensus       ---------- ---------- ---------- ---------- ----------

               151                                        190
     LLP        dknPKaLEDF YrfIYreR.. FdkrrIiTpk QtEaCapeha
trichosurin     pnlsneIkkr FeedFmnRg. FrrenIldIs EvDhC
β-lactoglobulin IeenKgMEEF kkiVrtltip YtmieaqTr. Em..Crv
Consensus       ---------- ---------- ---------- ----*-----
```

(B)

```
              1                                                    50
M.giganteus   VENIRSKNDL GVEKFVGSWY LREAAKTMEF SIPLFDMDIK EVNLTPEGNL
  M.eugenii   VENIRSKNDL GVEKFVGSWY LREAAKTMEF SIPLFDMDIK EVNLTPEGNL
T.vulpecula   IENIySKeEL vVEKLIGpWY rvEeAKaMEF SIPLFDMnIK EVNrTPEGNL
 Consensus    -***-**--* -***--*-** --*-**-*** ********-** ***-******

              51                                                   100
M.giganteus   ELVLLEKTDR CVEKKLLLKK TkKPTEFEIY ISSES.SYTF cVMETDYDSY
  M.eugenii   ELVLLEKTDR CVEKKLLLKK TqKPTEFEIY ISSESASYTF SVMETDYDSY
T.vulpecula   ELIvLEqTDs CVEKKFLLKK TeKPaEFEIY IpSESASYTL SVMETDYDnY
  Consensus   **--**-**- *****-**** *-**-***** *-***-***- -*******-*

              101                                                  150
M.giganteus   FLFCLYNISD REKMACAHYV RRIEENKGMN EFKKILRTLA MPYTVIEVRT
  M.eugenii   FLFCLYNISD REKMACAHYV RRIEENKGMN EFKKILRTLA MPYTVIEVRT
T.vulpecula   iLgCLeNVny REKMACAHYe RRIEENKGMe EFKKIvRTLt iPYTmIEaqT
  Consensus   -*-**-*--- *********- *********- *****-***- -***-****-*

              151
M.giganteus   RDMCHV
  M.eugenii   RDMCHV
T.vulpecula   REMCrV
  Consensus   *-**-*
```

Fig. 3. Sequence comparisons of *T. vulpecula* lipocalins **(A),** marsupial β-lactoglobulins **(B),** late-lactation proteins **(C),** and *T. vulpecula* triochosurin with rodent major urinary proteins **(D).** Alignments were performed using the Pileup program of GCG.

revealed the presence of unique transcripts for each of approximately 1.2 kb in size. None of the three lipocalin probes hybridized with RNA from *T. vulpecula* liver or rat liver and mammary gland (data not shown). Both β-lactoglobulin and trichosurin were expressed at high levels throughout lactation, whereas late-lactation protein transcripts could only be detected in samples obtained at the switch period between the two phases of lactation or during the late-phase lactation. (Fig. 5).

Discussion

Our results have shown that the *T. vulpecula* milk contains three proteins of similar size which appear to be members of the lipocalin family. Flower (1996) has subdivided the lipocalins into the ‘‘kernal’’ and ‘‘outlier’’ lipocalins based on the presence of three motifs. Motif 1, which is characteristic of all lipocalins, contains GXWY/W/F located at the beginning of β strand A and is found in *T. vulpecula* β-lactoglobulin as GPWY, and in late-lactation protein as GTYY. In trichosurin, however, the corresponding sequence is RHWH (Fig. 3A). Three separate trichosurin clones gave the same DNA sequence coding for arginine (R) in place of the glycine (G) that is characteristic of motif 1. Motif 2 is found only in the kernal lipocalins and is detected only in *T. vulpecula* β-lactoglobulin as TDYDN. The absence of this motif from trichosurin contrasts with its presence in the major urinary proteins (residues 103–106, Fig. 3D). The third motif, which is also confined to the kernal lipocalins, is found in both *T. vulpecula* β-lactoglobulin and trichosurin as YER and FGR, respectively, but not in the late-lactation protein, confirming the classification of the late-lactation proteins and the odorant binding proteins as outlier lipocalins. Despite the observation that trichosurin does not have archetypal motifs 1 and 2, several lines of evidence, including the sequence similarity to the

(C)

```
            1                                                  50
T.vulpecula DDvAFSaFtP SEGTYYVQVI AVDKEFPEEE IPRDmSPLTI mYLddGKMEA
  M.eugenii DDfAFScFkP SEGTYYVQVI AVDKEFPEDE IPRDiSPLTI tYLnnGKMEA
  Consensus **-***-*-* ********** *********-* ****-***** -***-*****

            51                                                 100
T.vulpecula rFTmKKDdNC EEINimLEKt aDPRKITmnR rLrytCaaVR TSkqKhWILv
  M.eugenii kFTvKKDnNC EEINltLEKi dEPRKITttR hLhhiCdtVR TSeeKyWILs
  Consensus -**-***-** ****--***- --******--* -*---*--** ***-*-***-

            101                                                150
T.vulpecula CpREFQGetI RmAkLVGPNT DkNPKALEDF YRFIyRERFd kRRIITPkQT
  M.eugenii CvREFQGtqI ReAeLVGPNT DeNPKALEDF YRFInRERFv eRRIITPrQT
  Consensus *-*******-* *-*-****** *-******** ****-****- -*******-**

            151
T.vulpecula EACapEhA
  M.eugenii EACtsEnA
  Consensus ***--*-*
```

(D)

```
            1                                                  50
  {musmup1} HAEEASSTgR NFNVEK.... ..INGEWhtI ILASDKREKI EDnGnfRlFl
  {musmup4} HAEEAtSkgq NLNVEK.... ..INGEWFSI lLASDKREKI EEhGSMRVFV
 {musmup15} HAEEsSSmeR NFNVEq.... ..IsGyWFSI aeASyeREKI EEhGSMRaFV
   {ratmup} HAEEASSTrg NLdVaK.... ..lNGDWFSI VvASnKREKI EEnGSMRVFm
T.vulpecula lqpEcSrsee dLsdEKerkw eqlsrhWhtV VLASsdRslI EEeGpfRnFI
  Consensus ---*------ ---------- ------*--- --**--*--* *--*--*-*-

            51                                                 100
  {musmup1} EqIhVLENSL VLKFHTVrde ECSELsMVAD KTEKAGEYsV tYDGFNTFTI
  {musmup4} EhlhVLENSL aFKFHTVIdG ECSEiFLVAD KTEKAGEYsV mYDGFNTFTI
 {musmup15} EnItVLENSL VFKFHlIVNe ECtEMtaIgE qTEKAGiYym NYDGFNTFsI
   {ratmup} qhIdVLENSL gFKFrikeNG ECrELYLVAy KTpedGEYfV eYDGgNTFTI
T.vulpecula qnItVesgnL ngfFlTrkNG qCipLYLtAf KTEeArqFkl NYyGtNdvyy
  Consensus --*-*----* ---*------ -*-------- -*-------- -*-*-*----

            101                                                150
  {musmup1} pKTDYDNFlM aHLINeKDGe TFQLMgLYGR ePDLSSDIKE rFAqLCEkHG
  {musmup4} LKTDYDNYIM FHLINeKDGK TFQLMELYGR kaDLnSDIKE KFvKLCEEHG
 {musmup15} LKTDYDNYIM iHLINkKDGK TFQLMELYGR ePDLSlDIKE KFAKLCEEHG
   {ratmup} LKTDYDrYVM FHLINfKnGe TFQLMvLYGR tkDLSSDIKE KFAKLCEaHG
T.vulpecula gsskpneYak FiFyNyhDGK vnvvanLFGR tPnLSnEIKk rFeedfmnrG
  Consensus ---------- ----*---*- ------*-** ---*---**- -*-------*

            151
  {musmup1} IlRENIIDLs naNRCLQARE
  {musmup4} IIkENIIDLT ktNRCLkARE
 {musmup15} IIRENIIDLT nvNRCLeARE
   {ratmup} ItRDNIIDLT ktdRCLQARg
T.vulpecula frRENIlDis evdhC.....
  Consensus ----**-*-- ----*-----
```

Fig. 3. Continued.

major urinary proteins, the presence of the cysteine pair likely to be involved in disulphide bonding, and the observation that it appears to exist as a dimer (Hunter, Piotte, and Grigor, unpublished observations), all argue for its inclusion in the family of lipocalins.

The milk of *T. vulpecula* appears to contain two forms of β-lactoglobulin, although, to date, the differences between them have not been established. Only one transcript size for β-lactoglobulin could be detected and the cloned gene encoded a mature peptide with a predicted molecular mass of 18, 401, which would correspond to the smaller form of β-lactoglobulin detected in Fig. 1. It is possible that the second spot, which gave the same β-lactoglobulin N-terminal sequence, has been modified post-translationally although, to date, no post-translational modifications of any β-lactoglobulin are known that cause such a change in relative mass. β-Lactoglobulin is known to exist as a dimer but the methods used in our analysis would be expected to dissociate any multimeric forms of the protein. Both late-lactation protein and trichosurin, however, migrate with a molecular mass that is considerably greater than that predicted for the mature peptides, consistent with some form of post-translational modifications. Several other lipocalins are known to be glycosylated (Flower 1996).

The β-lactoglobulins from the marsupials formed a group distinct from other mammalian β-lactoglobulins, consistent with the early evolutionary division of these groups. These data support the idea that β-lactoglobulin was an ancestral protein which has retained a similar function in both marsupials and eutherian mammals.

The two late-lactation proteins from marsupials form

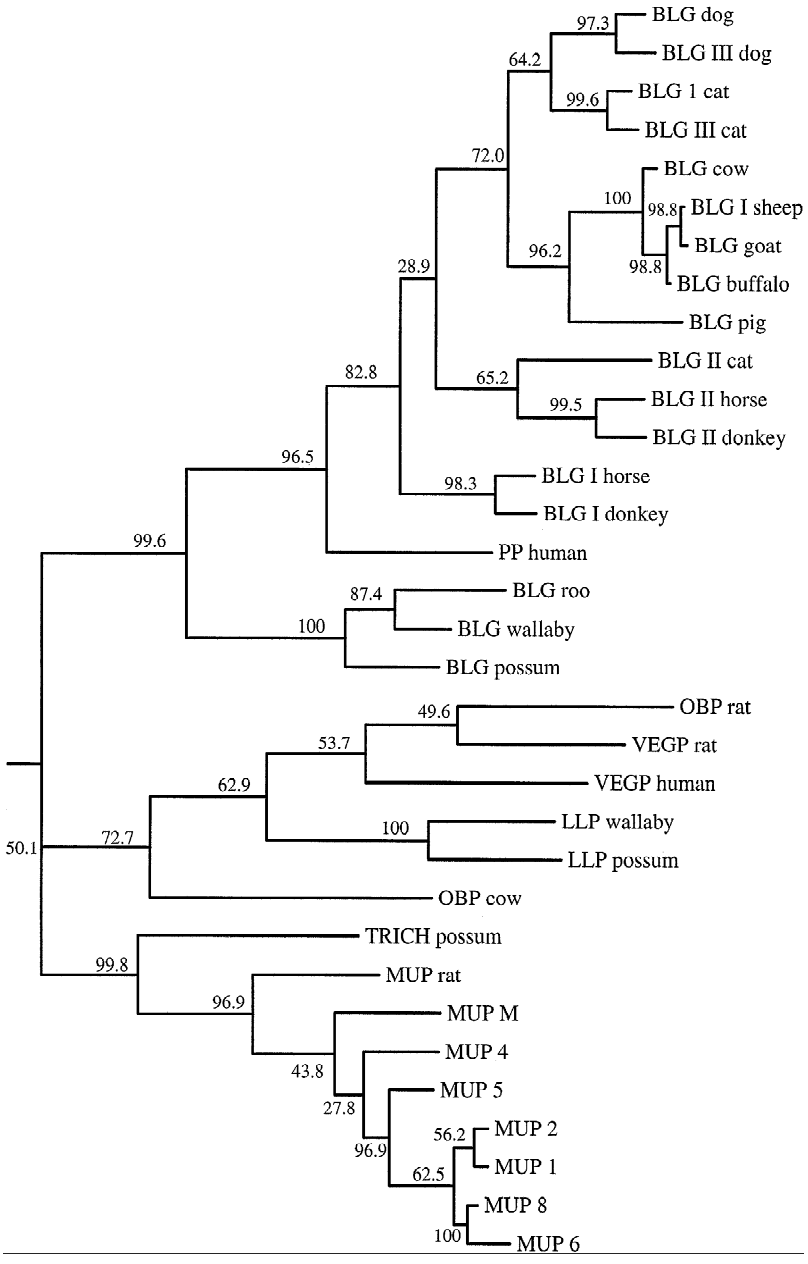

Fig. 4. Phylogenetic analysis of *T. vulpecula* lipocalins. Translated sequences of *T. vulpecula* β-lactoglobulin, late-lactation protein, and trichosurin were analyzed alongside those of other lipocalin molecules (see Table 1 for coding) by maximum parsimony as described in the text. The consensus tree produced by our analysis is shown as a cladogram. *Numbers at each branch* indicate the percentage of times that particular branching pattern or clade was found in the bootstrapped trees produced by the phylogenetic analysis. Relative evolutionary distances are proportional to distances along the ''x-axis'' whereas spacing along the ''y-axis'' is strictly for clarity of display and has no evolutionary significance.

a related pair of proteins which are in turn related to the odorant binding proteins and proteins from von Ebner's glands (Collet and Joseph 1993). The function of these proteins is less clear than is the function of β-lactoglobulin, but is presumably related to the binding of small molecules characteristic of many lipocalins. However, it is notable that whereas late-lactation protein has been reported to be a major protein in the milk of *M. eugenii* in the late phase of lactation (Nicholas et al. 1987), the *T. vulpecula* homologue appears to be expressed at much lower levels, as it could only be detected in gels after silver staining. Nevertheless, as in *M. eugenii*, no RNA expression could be detected in samples taken up to the switch period between the two phases of lactation. Studies of the promoter regions for the genes encoding these three proteins are currently under way in our laboratory.

Our analysis showed that trichosurin from *T. vulpecula* is more closely related to the mouse urinary proteins than to other lipocalins. Although the proteins are found in somewhat different fluids, it seems possible that they have a similar function given their phylogenetic similarity. However, small changes in lipocalin sequence may result in significant changes in ligand binding (Flower 1996), and the depth of the branch suggests some divergence within these proteins which may reflect a change of ligand.

Several lines of evidence argue that it is probable that all three genes in the possum have evolved from a single precursor gene. Published data also supports the close

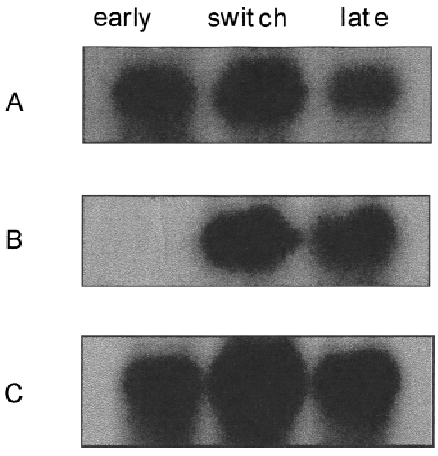

Fig. 5. Northern analysis of total mammary RNA from *T. vulpecula* taken at different stages of lactation. RNA was extracted from mammary tissue obtained from *T. vulpecula* lactating for 50 (early), 105 (switch), and 195 days (late), electrophoresed, blotted, and hybridized with radiolabeled cDNAs for *T. vulpecula* β-lactoglobulin **(A),** late-lactation protein **(B),** and trichosurin **(C)** as described in the text.

linkage of late-lactation protein and β-lactoglobulin genes in the tammar (Woodley et al. 1993) whereas recent unpublished results from our laboratory suggest that parts of the late-lactation protein and trichosurin genes can be detected on a single λ genomic clone containing an approximately 20-kb insert (Demmer, Ross and Grigor, unpublished). While eutherians do not appear to have either a late-lactation protein or a trichosurin equivalent in their milk, the presence of related proteins in other fluids suggests that the gene duplications required occurred before the marsupial eutherian split. It is also notable that β-lactoglobulin is absent from the milks of several eutherian species, in particular that of primates and rodents. This implies that during mammalian evolution, specific, but as yet unknown, factors have been responsible for the recruitment of particular genes for mammary function.

Acknowledgments. This work was supported by New Zealand FRST contract UOO-404. *Macropus eugenii* cDNA clones were kindly provided by C. Collet, CSIRO, Canberra, Australia.

References

Barton GJ, Sternberg MJE (1987) A strategy for the rapid multiple alignment of protein sequences. J Mol Biol 198:327–337

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Collet C, Joseph R (1993) A novel member of the lipocalin superfamily: tammar wallaby late-lactation protein. Biochim Biophys Acta 1167:219–222

Collet C, Joseph R, Nicholas K (1989) Molecular cloning and characterization of a novel marsupial milk protein gene. Biochem Biophys Res Commun 164:1380–1383

Collet C, Joseph R, Nicholas K (1991) A marsupial β-lactoglobulin gene: characterization and prolactin-dependent expression. J Mol Endocrinol 6:9–16

Cowan PE (1989) Changes in milk composition during lactation in the common brushtail possum, *Trichosurus vulpecula* (Marsupialia: Phalangeridae). Reprod Fertil Dev 1:323–333

Crisp EA, Cowan PE, Messer M (1989) Changes in milk carbohydrates during lactation in the common brushtail possum, *Trichosurus vulpecula* (Marsupialia: Phalangeridae). Reprod Fertil Dev 1:309–314

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565

Flower DR (1996) The lipocalin protein family: structure and function. Biochem J 318:1–14

Godovac-Zimmermann J (1988) The structural motif of β-lactoglobulin and retinol binding protein: a basic framework for binding and transport of small hydrophobic molecules? Trends Biochem Sci 13:64–66

Godovac-Zimmermann J, Shaw D (1987) The primary structure, binding site and possible function of β-lactoglobulin from eastern-grey kangaroo (*Macropus giganteus*) Biochem Hoppe-Seyler 368:879–886

Grigor MR, Bennett BL, Carne A, Cowan PE (1991) Whey proteins of the common brushtail possum (*Trichosurus vulpecula*). Comp Biochem Physiol 98B:451–459

Gross R, Bolliger A (1959) Composition of milk of the marsupial *Trichosurus vulpecula.* Am J Dis Child 98:768–775

Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ (1992) Experimental phylogenetics: generation of known phylogeny. Science 255:589–592

Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenies. Science 264:671–677

Nicholas KR, Messer M, Elliott C, Maher F, Shaw D (1987) A novel whey protein synthesized only in late lactation in the mammary gland from the tammar (*Macropus eugenii*). Biochem J 241:899–904

Papiz MZ, Sawyer L, Eliopoulos EE, North ACT, Findlay JCB, Sivaprasadarao R, Jones TA, Newcomer ME, Kraulis PJ (1986) The structure of β-lactoglobulin and its similarity to retinol-binding protein. Nature:324:383–385

Pervaitz S, Brew K (1985) Homology of β-lactoglobulin, serum retinol-binding protein and protein HC. Science 228:335–337

Pilton PE, Sharman GB (1962) Reproduction in the marsupial, *Trichosurus vulpecula.* J Endocrinol 25:119–136

Piotte CP, Grigor MR (1996) A novel marsupial whey protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors. Arch Biochem Biophys 330:59–64

Piotte CP, Marshall CJ, Hubbard MJ, Collett C, Grigor MR (1997) Lysozyme and α-lactalbumin from the milk of a marsupial, the common brushtailed possum (*Trichosurus vulpecula*). Biochem Biophys Acta (in press)

Sambrook J, Fritsch EF, Maniatis T (1992) Molecular cloning: a laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, New York

Sansom CE, North ACT, Sawyer L (1994) Structural analysis and classification of lipocalin and related proteins using a profile-search method. Biochim Biophys Acta 1208:247–255

Stockwell PA (1988) HOMED: a homologous sequence editor. Trends Biochem Sci 13:322–324

Winefield CS, Farnden KJF, Reynolds PHS, Marshall CJ (1995) Evolutionary analysis of aspartate aminotransferases. J Mol Evol 40: 455–463

Woodley GL, Gooley AA, Collet C, Cooper DW (1993) Origin of late lactation protein from β-lactoglobulin in the tammar wallaby. J Hered 84:460–465