SCR1

```
#Mmus.MUP4 : ----------------------HAEEATSKGQNLNVEK-----ING-EWFSILLASDKREKIE---EHGSMRVFVE--------HIHVLEN- : 52
#Mmus.MUP5 : ------------------------EEASSERQNFNVEK-----ING-KWFSILLASDKREKIE---EHGTMRVFVE--------HIDVLEN- : 50
#Mmus.MUP1 : ----------------------HAEEASSTGRNFNVEK-----ING-EWHTIILASDKREKIE---DNGNFRLFLE--------QIHVLEN- : 52
#Mmus.MUP6 : ----------------------HAEEASSTGRNFNVEK-----ING-EWHTIILASDKREKIE---DNGNFRLFLE--------QIHVLEN- : 52
#Rnor.a2g3 : ----------------------HAEEASFERGNLDVDK-----LNG-DWFSIVVASDKREKIE---ENGSMRVFVQ--------HIDVLEN- : 52
#Rnor.a2g4 : ----------------------HAEEANSERGNLDVDK-----LNG-DWFSIVVASNKREKIE---ENGSMRVFMQ--------HIDVLEM- : 52
#Rnor.a2g1 : ----------------------HAEEASSTRGNLDVDK-----LNG-DWFSIVVASDKREKIE---ENASMRVFMQ--------HIDVLEN- : 52
#Rnor.a2g2 : ----------------------HAEEASSTRGNLDVAK-----LNG-DWFSIVVASNKREKIE---ENGSMRVFMQ--------HIDVLEN- : 52
#Cfam.f2p  : ------------------------QEGNHEEPQGGLEE-----LSG-RWHSVALASNK--SDL-IKPWGHFRVFIH--------SMSAKDG- : 50
#Tvul.Lip  : ----------------LQPECSRSEEDLSDEKERKWEQ-----LSR-HWHTVVLASSD--RSL-IEEEGPFRNFIQ--------NITVESG- : 58
#Ecab.p19p : -------------------RRPHALHMGPGDPNFDEKL-----VKG-KWFSVALASNE--PKF-IAKDTDMKFFIH--------KIQVTPE- : 55
#Btau.OBP  : --------------------------AQEEEAEDNLSE-----LSG-PWRTVYIGSTNPEKIQ---ENGPFRTYFR--------ELVFDDE- : 48
#Sscr.OBP  : ---------------------------QEPQPEDDPFE-----LSG-KWITSYIGSSDLEKIG---ENAPFQVFMR--------SIEFDDK- : 47
#Mmus.Pbas : -------------------------------VMSLKKK-----IDG-PWQTIYLAASTMEKIN---EGSPLRTYFR--------HICVGRR- : 43
#Rnor.Pbas : ----------------------------MMTDKNLKKK-----IEG-NWRTVYLAASSVEKIN---EGSPLRTYFR--------RIECGKR- : 46
#Btau.alle : --------------------------AQETPAEIDPSK-----IPG-EWRIIYAAADNKDKIV---EGGPLRNYYR--------RIECIND- : 48
#Rnor.OBP1 : ---------------------------HHENLDISPSE-----VNG-DWRTLYIVADNVEKVA---EGGSLRAYFQ--------HMECGDE- : 47
#Ccri.Aphr : ---------------------------------QDFAE-----LQG-KWYTIVIAADNLEKIE---EGGPLRFYFR--------HIDCYKN- : 41
#Mmus.Aphr : -----------------------------------HAA-----MEG-PWKTVAIAADRVDKIE---RGGELRIYCR--------SLTCEKE- : 39
#Rnor.NGAL : ------------QDSTQNLIPAPPLISVPLQPGFWTER-----FQG-RWFVVGLAANA----V-QKERQSRFTMYS--------TIYELQE- : 60
#Mmus.NGAL : ------------QDSTQNLIPAPSLLTVPLQPDFRSDQ-----FRG-RWYVVGLAGNA----V-QKKTEGSFTMYS--------TIYELQE- : 60
#Hsap.NGAL : ------------QDSTSDLIPAPPLSKVPLQQNFQDNQ-----FQG-KWYVVGLAGNA----I-LREDKDPQKMYA--------TIYELKE- : 60
#Rnor.PGDS : ------------------------QGHDTVQPNFQQDK-----FLG-RWYSAGLASNS---SW-FREKKELLFMCQ--------TVVAPST- : 49
#Mmus.PGDS : --------------------QTPAQGHDTVQPNFQQDK-----FLG-RWYSAGLASNS---SW-FREKKAVLYMCK--------TVVAPST- : 53
#Cfam.PGDS : --------------------QTSAQAQVSLQPNFQQDK-----FLG-RWFTSGLASNS---SW-FREKKNVLSMCM--------SVVAPTA- : 53
#Ecab.PGDS : --------------------QTRAQAQPSLQPNFQQDK-----FLG-RWFTSGLASNS---SW-FREKKKVLSMCT--------SVVAPTA- : 53
#Bbub.BLB  : ------------------------IIVTQTMKGLDIQK-----VAG-TWYSLAMAASD--ISL-LDAQSAPLRVYV--------EELKPTP- : 50
#Btau.BLB  : ------------------------LIVTQTMKGLDIQK-----VAG-TWYSLAMAASD--ISL-LDAQSAPLRVYV--------EELKPTP- : 50
#Chir.BLB  : ------------------------IIVTQTMKGLDIQK-----VAG-TWYSLAMAASD--ISL-LDAQSAPLRVYV--------EELKPTP- : 50
#Oari.BLB  : ------------------------IIVTQTMKGLDIQK-----VAG-TWHSLAMAASD--ISL-LDAQSAPLRVYV--------EELKPTP- : 50
#Hsap.a1GP : ------------------QIPLCANLVPVPITNATLDQ-----ITG-KWFYIASAFRNEEYNK-SVQEIQATFFYF--------TPNKTED- : 58
#Hsap.a1GP : ------------------QIPLCANLVPVPITNATLDR-----ITG-KWFYIASAFRNEEYNK-SVQEIQATFFYF--------TPNKTED- : 58
#Ocun.a1GP : ------------------QDPACANFSTSPITNATLDQ-----LSH-KWFFTASAFRNPKYKQ-LVQHTQAAFFYF--------TAIKEED- : 58
#Mmus.a1G3 : -----------------QNPEHA-INIGDPITNETLSW-----LSG-KWFLIAVADSDPDYRQ-EIQKVQTIFFYL--------TLNKIND- : 58
#Rnor.a1GP : -----------------QNPEPANITLGIPITNETLKW-----LSD-KWFYMGAAFRDPVFKQ-AVQTIQTEYFYL--------TPNLIND- : 59
#Mmus.a1GP : -----------------QNPEHVNITIGDPITNETLSW-----LSD-KWFFIGAAVLNPDYRQ-EIQKTQMVFFNL--------TPNLIND- : 59
#Mcar.a1G2 : -----------------QNPEHVNITIGEPITNETLSW-----LSD-KWFFIGAAVLNPDYRQ-EIQKMQMVFFNI--------TPNLIND- : 59
#Mmus.a1G2 : -----------------QNPEHANFTIGEPITNETLSW-----LSD-KWFFMGAAFRKLEYRQ-AIQTMQSEFFYL--------TTNLIND- : 59
#Mcar.a1GP : -----------------QNPEHANITIGEPITNETLGW-----LSD-KWFFMGAAFRKLEYRQ-AIQMMQTEFFYL--------TTNLIND- : 59
#Maur.A1mg : -------------------DPVPALPDIQVQENFNESR-----IYG-KWFNLAVGSTCPWLSR-IKNKMSMSTLVL--------REGATGA- : 57
#Rnor.A1mg : -------------------DNVPTLPDIQVQENFNEAR-----IYG-KWFNLAVGSTCPWLRR-IKNKMSVSTLVL--------QEGATEA- : 57
#Mung.A1mg : -------------------DDAATLPDIQVQENFIESL-----IYG-KWFNLAVGSTCPWLRR-IKDKMSMSTLVL--------QEGATEA- : 57
#Mmus.A1mg : -------------------DPASTLPDIQVQENFSESR-----IYG-KWYNLAVGSTCPWLSR-IKDKMSVQTLVL--------QEGATET- : 57
#Hsap.C8GC : ------------QKPQRPRRPASPISTIQPKANFDAQQ-----FAG-TWLLVAVGSACRFLQ---EQGHRAEATTL--------HVAPQGT- : 62
#Ocun.C8GC : ------------RWAQKPRGAPSAISAIQPKANFDAQQ-----FAG-TWLLAAVGSACHFLQ---EQGHRAEATAL--------HVAPQGA- : 62
#Rnor.OBP2 : ----------------------------QEAPPDDQED-----FSG-KWYTKATVCDR----N-HTDGKRPMKVFP--------MTVTALE- : 44
#Mmus.VNSP : --------------------------QDSSFLAFNNGN-----FSG-KWFLKALVSED-------DIPINKVSPML--------ILVLNNG- : 44
#Mmus.VNSP : -----------------------LQTYDDLPFISEEDK-----LSG-VWFIKATVSQRRE----VEGETLVAFPIK--------FTCPEEG- : 50
#Rnor.VEG2 : --------------------------AQAFPTTEENQD-----VSG-TWYLKAAAWDKEIFTPDKKFGSVSVTPMK--------IKTLEGG- : 51
#Rnor.VEG1 : --------------------------AQAFPTTEENQD-----VSG-TWYLKAAAWDKEIFTPDKKFGSVSVTPMK--------IKTLEGG- : 51
#Hsap.VEG  : -------------------------AHHLLASDEEIQD-----VSG-TWYLKAMTVDREFPE----MNLESVTPMT--------LTTLEGG- : 48
#Sscr.VEG  : -------------------------AQEFPAVGDPLQD-----LLG-RWYLKAMTSDPEIP----GKKPESVTPLI--------LKALEGG- : 48
#Cfam.f1p  : --------------------------QDTPALGKDTVA-----VSG-KWYLKAMTADQEVP-----EKPDSVTPMI--------LKAQKGG- : 46
#Rapp.HBP1 : -------------DKPVWADEAANGEHQDAWKHLQKLV------EE-NYDLIKATYKNDPVWG--NDFTCVGTAAQ--------NLNEDEKN : 62
#Rapp.HBP2 : -------------NQPDWADEAANGAHQDAWKSLKADV------EN-VYYMVKATYKNDPVWG--NDFTCVGVMAN--------DVNEDEKS : 62
#Rapp.HBP3 : --------------NPTWANEAKLGSYQDAWKSLQQDQ------NK-RYYLAQATQTTDGVWG--EEFTCVSVTAE--------KIGKKKLN : 61
#Osav.TSGP : ------------------GPDGCVGSTEAKVAVFGEGGNAGSPTIG-YSYLVKTTYPD--------EHACVYILPP--------YGTADASG : 57
#Osav.TSGP : --------------------------ANDVWNVLKGSD-------S-KFLMVKRTYERG--------ANKCVYMKRT--------SMDESSH- : 42
#Omou.MB    : --------------------GCSVSDPLDALKAFKDGA-------G-TFLLQKSTDPQ---------ARDCLKGTP--------NGNRDGN- : 46
#Osav.TSGP : --------------------DCPTGFPTDAYVAFNEGQ-------G-AYILVKSTDLD---------ARDCLKGSA--------TGKKEGN- : 46
#Osav.TSGP : --------------------DCPTGFPTEAYVAFNEGK-------G-AYILVRSTNLN---------ARDCLKGEA--------TGKKEGN- : 46
#Rpro.PAI1 : -----------FCANPPKMPTGCKDLNSKAVKDFKYND----FFKD-KWILTHAERVTHP------DACETFTVNG--------NK------ : 56
#Rpro.PAI2 : ---------------TTTMPKECLELKGDIKPGFDANQ----FFTG-DWYWTHARDPKHP------KLCQKYQATS--------DLR----- : 53
#Tpal.PAL1 : --------------------EECELMP--PGDNFDLEK----YFSIPHVYVTHSRNGPKE------QVCREYKTTK--------NSDGTTTT : 52
#Tpal.PAL2 : --------------------EECELMP--PGDNFDLEK----YFSIPHVYVTHSRNGPKE------QVCREYKTTK--------NSDG-TTT : 51
#Rpro.NP2  : ---------------------DCSTNIS-PKQGLDKAK----YFSG-KWYVTHFLDKDPQ-VT--DQYCSSFTPRE--------SDG---TV : 51
#Rpro.NP3  : ---------------------DCSTNIS-PKKGLDKAK----YFSG-TWYVTHYLDKDPQ-VT--DPYCSSFTPKE--------SGG---TV : 51
#Rpro.NP1  : ---------------------KCTKNAL-AQTGFNKDK----YFNGDVWYVTDYLDLEPDDVP--KRYCAALAAGT--------ASG---KL : 53
#Rpro.NP4  : ---------------------ACTKNAI-AQTGFNKDK----YFNGDVWYVTDYLDLEPDDVP--KRYCAALAAGT--------ASG---KL : 53
#Bger.All4 : --------------------NEDCFRHESLVPNLDYER-----FRG-SWIIAAGTSEA----------LTQYKCWI--------DRFSYDD- : 47
#Dmel.Nlip : --------------AGTDAYGFGRCPNYPSMPKFNMSR-----VLG-HWYEVERSFYL-PEIA-SGCTTFQFEPYNKGEQSKFSNFKLAVA- : 69
#Hgam.CRC2 : -------------DGIPSFVTAGKCASVANQDNFDLRR-----YAG-RWYQTHIIENA-YQPV-TRCIHSNYEYST--------NDYGFKV- : 62
#Hgam.CRC1 : --------DKIPDFVVPGKCASVDRNKLWAEQTPNRNS-----YAG-VWYQFALTNNP-YQLI-EKCVRNEYSFDG---------KQFVIK- : 66
#Same.Laz  : ------------------AQETMGCADRSAINDFNATL-----YMG-KWYEYAKNGSMPYEEG-GVCVTAEYSMSS--------HNITVVN- : 58
#Dmel.Dlip : ----------------AQVPFPGKCPDVKLLDTFDAEA-----YMG-VWYEYAAYPFA-FEIG-KKCIYANYSLID--------NSTVSVV- : 59
#Hsap.ApoD : -----------------QAFHLGKCPNPPVQENFDVNK-----YLG-RWYEIEKIPTT-FENG-R-CIQANYSLME--------NGKIKVL- : 57
#Mmus.ApoD : -----------------QNFHLGKCPSPPVQENFDVKK-----YLG-RWYEIEKIPAS-FEKG-N-CIQANYSLME--------NGNIEVL- : 57
#Ccob.ApoD : -----------------QAFHFGKCPNPPVQENFDLNK-----YLG-RWYEIEKIPVS-FEKG-N-CIQANYSLKE--------NGRVKVL- : 57
#Ocun.ApoD : -----------------QAFHLGRCPTPPVQENFDVHK-----YLG-RWYEIEKIPVS-FEKG-N-CIQANYSLME--------NGNIKVL- : 57
#Pbra.Bbp  : -----------------NVYHDGACPEVKPVDNFDWSN-----YHG-KWWEVAKYPNS-VEKY-GKCGWAEYTPEG--------KSVKVSN- : 58
#Gmel.Gall : -------------------VHEGKCPDFKPVDNFNLTA-----YQG-VWYEISKTPNDAEKNG-K-CGQAEYKLEG--------EVVKVKN- : 56
#Msex.IcyA : ----------------GDIFYPGYCPEVKPVDDFDLSA-----FAG-AWHEIAKLPLE-NENE-GKCTVAEYKYDG--------KKASVYN- : 59
#Msex.IcyB : ----------------GDIFYPGYCPDVKPVDDFDLSA-----FAG-AWHEIAKLPLE-NENQ-GKCTIAEYKYDG--------KKASVYN- : 59
#Ggal.Purp : --------------------QTCAVDSFSVKDNFDPKR-----YAG-KWYALAKKDPE-GLFL-QDNISAEYTVEE--------DGTMTAS- : 55
#Hsap.RBP  : -------------------ERDCRVSSFRVKENFDKAR-----FSG-TWYAMAKKDPE-GLFL-QDNIVAEFSVDE--------TGQMSAT- : 56
#Ecab.RBP  : -------------------ERDCRVSSFRVKENFDKAR-----FSG-TWYAMAKKDPE-GLFL-QDNIVAEFSVDE--------YGQMSAT- : 56
#Sscr.RBP  : -------------------ERDCRVSSFRVKENFDKAR-----FSG-TWYAMAKKDPE-GLFL-QDNIVAEFSVDE--------NGHMSAT- : 56
#Btau.RBP  : -------------------ERDCRVSSFRVKENFDKAR-----FAG-TWYAMAKKDPE-GLFL-QDNIVAEFSVDE--------NGHMSAT- : 56
#Vcho.Lpro : --------------------CLGMPESVKPVSDFELNN-----YLG-KWYEVARLDHS-FERG-LSQVTAEYRVRN--------DGGISVL- : 55
#Ecol.OML  : ------------------CSSPTPPRGVTVVNNFDAKR-----YLG-TWYEIARFDHR-FERG-LEKVTATYSLRD--------DGGLNVI- : 57
#Cfre.OML  : ------------------CSSPTPPKGVTVVNNFDAKR-----YLG-TWYEIARFDHR-FERG-LDKVTATYSLRD--------DGGINVI- : 57
```

- VIII Mouse urinary proteins
- Ix α2u-globins
- XI Miscellaneous lipocalinsI
- X Chemoreception lipocalins I
- V Neutrophil gelatinase lipocalins Prostaglandin D synthases
- IV β-Lactoglobins
- XII α1-Acid glycoproteins
- VI α1-microglobins
- VII Complement 8γ
- XIII Chemoreception lipocalins II
- II Hard tick histamine-binding proteins
- II Soft tick TSGP's and platelet aggregation inhibitors
- II Triatomine bug platelet aggregation inhibitors
- II Triatomine bug nitrophorins
- II Allergen IV
- II nerve lipocalin
- II Crustacyanin
- II Nerve lipocalins
- II Metazoan apolipoprotein D
- II Insecticyanins
- III Retinol-binding proteins
- I Eubacterial and dictyostelid lipocalins

Fig. 7.9 (Page 1)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

```
                                                                                     SCR2
#Mmus.MUP4 : --SLAFKFHTVIDG--------ECSEIFLVADKTEK----AGEYSVMYDGFN----------TFTILKTDYD--NYIMFHLINEK-DG---- : 113
#Mmus.MUP5 : --SLAFKFHTVIDE--------ECTEIYLVADKTEK----AGEYSVTYDGFN----------TFTILKTDYD--NYIMFHLINKK-DE---- : 111
#Mmus.MUP1 : --SLVLKFHTVRDE--------ECSELSMVADKTEK----AGEYSVTYDGFN----------TFTIPKTDYD--NFLMAHLINEK-DG---- : 113
#Mmus.MUP6 : --SLVLKFHTVRDE--------ECSELSMVADKTEK----AGEYSVTYDGFN----------TFTIPKTDYD--NFLMAHLINEK-DG---- : 113
#Rnor.a2g3 : --SLGFTFRIKENG--------VCTEFSLVADKTAK----DGEYFVEYDGEN----------TFTILKTDYD--NYVMFHLVNVN-NG---- : 113
#Rnor.a2g4 : --SLGFKLCIKENG--------ECRKLYSVAYKTPK----IGEYFLEYDGGN----------TFTILKTDYE--RYVMFHLVNVN-NG---- : 113
#Rnor.a2g1 : --SLGFKFRIKENG--------ECRELYLVAYKTPE----DGEYFVEYDGGN----------TFTILKTDYD--RYVMFHLINFK-NG---- : 113
#Rnor.a2g2 : --SLGFKFRIKENG--------ECRELYLVAYKTPE----DGEYFVEYDGGN----------TFTILKTDYD--RYVMFHLINFK-NG---- : 113
#Cfam.f2p  : --NLHGDILIPQDG--------QCEKVSLTAFKTAT----SNKFDLEYWGHN----------DLYLAEVDPK--SYLILYMINQY-ND---- : 111
#Tvul.Lip  : --NLNGFFLTRKNG--------QCIPLYLTAFKTEE----ARQFKLNYYGTN----------DVYYGSSKPN--EYAKFIFYNYH-DG---- : 119
#Ecab.p19p : --SLQFHFHRKVRG--------MCVPTMMTAHKTKK------KFQYTVNHSG--------HKTIFLEKVDPK--HFVIFCAHSMK-HG---- : 116
#Btau.OBP  : --KGTVDFYFSVKR--------DGKWKNVHVKATKQ---DDGTYVADYEGQN----------VFKIVSLSRT--HLVAHNINVDK-HG---- : 110
#Sscr.OBP  : --ESKVYLNFFSKE--------NGICEEFSLIGTKQ----EGNTYDVNYAGN---------NKFVVSYASET--ALIISNINVDE-EG---- : 109
#Mmus.Pbas : --SNQVYLYFFIKK--------GTKCQLYKVIGRKK----QEVYYAQYEGS-----------IAFMLKMVNE--KILLFHYFNKN-RR---- : 103
#Rnor.Pbas : --CNRINLYFYIKK--------GAKCQQFKIVGRRS----QDVYYAKYEGS-----------TAFMLKTVNE--KILLFDYFNRN-RR---- : 106
#Btau.alle : --CESLSITFYLKD--------QGTCLLLTEVAKRQ---EGYVYVLEFYGTN-----------TLEVIHVSE--NMLVTYVENYD-GE---- : 109
#Rnor.OBP1 : --CQELKIIFNVKL--------DSECQTHTVVGQKHE---DGRYTTDYSGRN-----------YFHVLKKTD--DIIFFHNVNVD-ES---- : 108
#Ccri.Aphr : --CSEMEITFYVIT--------NNQCSKTTVIGYLKG---NGTYQTQFEGN-----------NIFQPLYITSD--KIFFTNKNMDR-AG---- : 103
#Mmus.Aphr : --CKEMKVTFYVNE--------NGQCSLTTITGYLQE--DGKTYKTQFQGN-----------NRYKLVDESPE--NLTFYSENVDR-AD---- : 102
#Rnor.NGAL : --DNSYNVTSILVR--------GQGCRYWIRTFVPS--SRPGQFTLGNIHSYP----QIQSYDVQVADTDYD--QFAMVFFQKTS-EN---- : 129
#Mmus.NGAL : --NNSYNVTSILVR--------DQDQGCRYWIRTFVPSSRAGQFTLGNMHRYP----QVQSYNVQVATTDYN--QFAMVFFRKTS-EN---- : 131
#Hsap.NGAL : --DKSYNVTSVLFR--------KKKCDYWIRTFVPG--CQPGEFTLGNIKSYP----GLTSYLVRVVSTNYN--QHAMVFFKKVS-QN---- : 129
#Rnor.PGDS : --EGGLNLTSTFLR--------KNQCETKVMVLQPA--GVPGQYTYNSPHWG-------SFHSLSVVETDYD--EYAFLFSKGTK-GP---- : 115
#Mmus.PGDS : --EGGLNLTSTFLR--------KNQCETKIMVLQPA--GAPGHYTYSSPHSG-------SIHSVSVVEANYD--EYALLFSRGTK-GP---- : 119
#Cfam.PGDS : --DGGLNLTSTFLR--------KDQCETRTLLLRPA--GTPGCYSYTSPHWG-------STHDVWVVATNYE--EYALLYTAGSK-GL---- : 119
#Ecab.PGDS : --DGGFNLTSTFLR--------KDQCETRTLLLQPA--GPPGCYSYTSPHWG-------MVHEVSVVETDYE--EYALLYTHAES-TKGL-- : 121
#Bbub.BLB  : --EGDLEILLQKWE--------NGECAQKKIIAEKT--KIPAVFKIDALNE-----------NKVLVLDTDYK--KYLLFCMENSA-E----- : 112
#Btau.BLB  : --EGDLEILLQKWE--------NGECAQKKIIAEKT--KIPAVFKIDALNE-----------NKVLVLDTDYK--KYLLFCMENSA-E----- : 112
#Chir.BLB  : --EGNLEILLQKWE--------NGECAQKKIIAEKT--KIPAVFKIDALNE-----------NKVLVLDTDYK--KYLLFCMENSA-E----- : 112
#Oari.BLB  : --EGNLEILLQKWE--------NGECAQKKIIAEKT--KIPAVFKIDALNE-----------NKVLVLDTDYK--KYLLFCMENSA-E----- : 112
#Hsap.a1GP : --TIFLREYQTRQD--------QCIYNTTYLNVQRE----NGTISRYVGGQE--------HFAHLLILRDTK--TYMLAFDVNDE-KN---- : 121
#Hsap.a1GP : --TIFLREYQTRQN--------QCFYNSSYLNVQRE----NGTVSRYEGGRE--------HVAHLLFLRDTK--TLMFGSYLDDE-KN---- : 121
#Ocun.a1GP : --TLLLREYITTNN--------TCFYNSSIVRVQRE----NGTLSKHDGIRN--------SVADLLLLRDPG--SFLLVFFAGKE-QD---- : 121
#Mmus.a1G3 : --TMELREYHTKDD--------HCVYNSNLLGFQRE----NGTLFKYEGEVE--------NFSHLRVLEKHG--AIMLFFDLKDE-KK---- : 121
#Rnor.a1GP : --TIELREFQTTDD--------QCVYNFTHLGVQRE----NGTLSKCAGAVK--------IFAHLIVLKKHG--TFMLAFNLTDE-N----- : 121
#Mmus.a1GP : --TMELREYHTIDD--------HCVYNSTHLGIQRE----NGTLSKYVGGVK--------IFADLIVLKMHG--AFMLAFDLKDE-KK---- : 122
#Mcar.a1G2 : --TMELREYHTIDD--------HCVYNSTHLGIQRE----NGTLSKYVGGVKI---------FADLIVLRKHG--AFMLAFDLKDE-KK---- : 122
#Mmus.a1G2 : --TIELRESQTIGD--------QCVYNSTHLGFQRE----NGTFSKYEGGVE---------TFAHLIVLRKHG--AFMLAFDLKDE-KK---- : 122
#Mcar.a1GP : --TIELRESQTIGD--------QCVYNSTHLGFQRE----NGTFSKYEGGVET---------FAHLIVLRKHG--AFMLAFDLNDE-KK---- : 122
#Maur.A1mg : --EISTTSTRWRRG--------VCEEVSGTYEKTD----MDGKFLYHKSKWN-------VTLESYVVHTNYD--EYAIFLTKKFS-RH---- : 121
#Rnor.A1mg : --EISVTSTQWRKG--------VCEEISGVYQKTD----IDGKFLYHKSKWN-------ATLESYVVHTNYD--EYAIFLTKKFS-HR---- : 121
#Mung.A1mg : --EISMTSTRWRRG--------VCEEISGAYEKTD----IDGKFLYHKSKWN-------ITLETYVVHTNYD--EYAIFLTKKFS-HY---- : 121
#Mmus.A1mg : --EISMTSTRWRRG--------VCEEITGAYQKTD----IDGKFLYHKSKWN-------ITLESYVVHTNYD--EYAIFLTKKSS-HH---- : 121
#Hsap.C8GC : --AMAVSTFRKLDG--------ICWQVRQLYGDTG----VLGRFLLQARGAR-------GAVHVVVAETDYQ--SFAVLYLERAG------- : 124
#Ocun.C8GC : --AMAVSTFRKLDG--------ICWQVSQRYGATG----VPGRFLLPARGPR-------GAVHVVAAETDYH--SFAVLYLERAR------- : 124
#Rnor.OBP2 : --GGDLEVRITFRG--------KGHCHLRRITMHKTD--EPGKYTTFKGKK-----------TFYTKEIPVK--DHYIFYIKGQR-HG---- : 106
#Mmus.VNSP : --DIELSITHMIYD--------QCLEVTTILEKTD----VPGQYLAFEGKT-----------HLQVQLSSVK--GHYMLYCDGEI-EG---- : 104
#Mmus.VNSP : --TLELRHTLASKG--------ECINVGIRLQRTEE----PGQYSAFWGHT-----------LFYIYDLPVK--DHYIIYCESHP-FQ---- : 110
#Rnor.VEG2 : --NLQVKFTVLISG--------RCQEMSTVLEKTDE----PGKYTAYSGKQ----------VFTVYSIPSAVE--DHYIFYYEGKI-HR---- : 113
#Rnor.VEG1 : --NLQVKFTVLIAG--------RCKEMSTVLEKTDEP----AKYTAYSGKQ----------VFTLYIIPSSVE--DHYIFYYEGKI-HR---- : 113
#Hsap.VEG  : --NLEAKVTMLISG--------RCQEVKAVLEKTD------EPGKYTADGGK----------HVAYIIRSHVK--DHYIFYCEGEL-HG---- : 108
#Sscr.VEG  : --DLEAQITFLIDG--------QCQDVTLVLKKTN------QPFTFTAYDGK----------RVVYILPSKVK--DHYILYCEGEL-DG---- : 108
#Cfam.f1p  : --NLEAKITMLTNG--------QCQNITVVLHKTSE----PGKYTAYEGQR-----------VVFIQPSPVR--DHYILYCEGEL-HG---- : 106
#Rapp.HBP1 : VEAWFMFMNNADTVY-------QHTFEKATPDKMYG---YNKENAITYQTE--------DGQVLTDVLAFSD-DNCYVIYALGP--DGSGA- : 132
#Rapp.HBP2 : IQAEFLFMNNADTNM-------QFATEKVTAVKMYG---YNRENAFRYETE--------DGQVFTDVIAYSD-DNCDVIYVPGT--DGNEE- : 132
#Rapp.HBP3 : --ATILYKNKHLTDL-------KESHETITVWKAYD---YTTENGIKYETQGT------RTQTFEDVFVFSDYKNCDVIFVPKE--RGSDE- : 132
#Osav.TSGP : --RYPYRMGYKDSND-------QWVKLDGKIKTEGS---KIIDNDPEYGDT--------VTTVLYTHLGG----GCDVTLFEGQ--KGQSKV : 123
#Osav.TSGP : --TLEVLMGYSKAGTTTDF--VEPSKYTVTATSEGA---STYNMMTVRRGPAS------HGVKFELVYSDDQ--GCNILQMKTSPFPG---- : 115
#Omou.MB   : --TLPVTMTYKDDS--------KWVSLNWMFTLEGA---NIVATLEGKRKQ--------RGELVYDVQSH----DCHITKLSS----G---- : 105
#Osav.TSGP : --KVPVMMTFKNEG--------QWVSLPWTFTLDGP---KVTATD-GQRTL--------KREVVYDVASH----HCHVEKLAS----G---- : 104
#Osav.TSGP : --TLPVMMTFKDEG--------KWVSLPWTFTLDGP---KVTATH-GQRTL--------KGEVVYDVPSH----HCHIEKLES----G---- : 104
#Rpro.PAI1 : ---ITFS-LGG-------------KEVSCTLVKVEG--AKFTKFNCELQG-------KKFTAYLSVLATDYK--NYALVYRCGSH------- : 113
#Rpro.PAI2 : ---LKFNGNSG-------------SDVTCQGAKVIGK-EGFYSFQCTTSG-------VTFTSFMAVVETDYN--NYALLYRCGRYG------ : 113
#Tpal.PAL1 : T-LVTSDYKTGGKP---------YHSELKCTNTPKSGG-KGQFSVECEVPNGN--GGKKKIHVETSVIATDYK--NYALLQSCTKT------- : 123
#Tpal.PAL2 : T-LVTSDYKTGGKP---------YHSELKCTNTPKSGV-KGQFSVECEVPNGN--GGKKKIHVETSVIATDYK--NYALLQSCTKT------- : 122
#Rpro.NP2  : K-EALYHYNANKKT---------SFYNIGEGKLESSGL-QYTAKYKTVDKKKAVLKEADEKNSYTLTVLEADD--SSALVHICLREG------ : 125
#Rpro.NP3  : K-EALYHFNSKKKT---------SFYNIGEGKLGSPGV-QYTAKYNTVDKKRKEIEPADPKDSYTLTVLEADD--SSALVHICLREG------ : 125
#Rpro.NP1  : K-EALYHYDPKTQD---------TFYDVSELQEESPG-KYTANFKKVEKNGNVKVDVTSGNYYTFTVMYADD--SSALIHTCLHKG------ : 126
#Rpro.NP4  : K-EALYHYDPKTQD---------TFYDVSELQVESLG-KYTANFKKVDKNGNVKVAVTAGNYYTFTVMYADD--SSALIHTCLHKG------ : 126
#Bger.All4 : --ALVSKYTDSQGK---------NRTTIRGRTKFEG-----NKFTIDYNDKGK------AFSAPYSVLATDYQ--NYAIVEGCPAA-AN---- : 111
#Dmel.Nlip : --IKNINRITGNPNVNIGYATPENSRSSIMDFKFTT------RFPFPDVIARL----LPGSGKYQVLYTDYE--NFAILWSCGSI-GS---- : 142
#Hgam.CRC2 : --TTAGFNPNDEYL--------KIDFKVYPTKEFP-----AAHMLIDAPSVF-------AAPYEVIETDYE--TYSCVYSCITT-DN---- : 124
#Hgam.CRC1 : --STGIAYDGNLLK--------RNGKLYPNPFGEP-------HLSIDYENS-------FAAPLVILETDYS--NYACLYSCIDY-N----- : 125
#Same.Laz  : --SMKCNTTHEVNT--------TTGWAEFASELHT-----DGKLSVHFPNSPS--------VGNYWILSTDYD--NYSIVWSCVKR-P--D-- : 121
#Dmel.Dlip : --NAAINRFTGQPS--------NVTGQAKVLGPGQ--------LAVAFYPTQP-----LTKANYLVLGTDYE--SYAVVYSCTSV-TP---- : 121
#Hsap.ApoD : --NQELRADGTVNQ--------IEGEATPVNLTEP------AKLEVKFSWFMP-------SAPYWILATDYE--NYALVYSCTCI-IQ---- : 119
#Mmus.ApoD : --NKELSPDGTMNQ--------VKGEAKQSNVSEP------AKLEVQFFPLMP-------PAPYWILATDYE--NYALVYSCTTF-FW---- : 119
#Ccob.ApoD : --NQELRPDGTVNQ--------IEGEATHSNITEP------AKLGVKFFQLMP-------SAPYWVLATDYD--NYALVYSCTNI-IW---- : 119
#Ocun.ApoD : --NQELRPDGTVNQ--------IEGQATQSNLTEP------AKLGVKFFQLMP-------TAPYWVLATDYE--NYALVYSCTTI-IW---- : 119
#Pbra.Bbp  : --YHVIHGKEYFIE--------GTAYPVGDSKIGK------IYHKLTYGGVTK-------ENVFNVLSTDNK--NYIIGYYCKYD-E--D-- : 120
#Gmel.Gall : --SHVVDGVQKYVE--------GTAKFAEDANKSA-----KLLVTLTYGAVNR-------ESPLNVIATDYQ--NYAIAYTCKYD-EK---- : 119
#Msex.IcyA : --SFVINGVKEYME--------GDLEIAPDAKLTK-----QGKYVMTFKFGPR-----VVVQVPWVLATDYK--NYAINYNCNYH-P--D-- : 124
#Msex.IcyB : --SFVVNGVKEYME--------GDLEIAPDAKYTK-----QGKYVMTFKFGQR-----VVNLVPWVLATDYK--NYAINYNCNYH-P--D-- : 124
#Ggal.Purp : --SKGRVKLFGFWV--------ICADMAAQYTVPDP-TTPAKMYMTYQGLAS--YLSSGGDNYWVIDTDYQ--NYAITYACRSL-KE-D-- : 127
#Hsap.RBP  : --AKGRVRLLNNWD--------VCADMVGTFTDTE----DPAKFKMKYWGVAS--FLQKGNDDHWIVDTDYD--TYAVQYSCRLL-NL-D-- : 126
#Ecab.RBP  : --AKGRVRLLNNWD--------VCADMVGTFTDTE----DPAKFKMKYWGVAS--FLQKGNDDHWIIDTDYD--TYAVQYSCRLL-NL-D-- : 126
#Sscr.RBP  : --AKGRVRLLNNWD--------VCADMVGTFTDTE----DPAKFKMKYWGVAS--FLQKGNDDHWIIDTDYD--TYAAQYSCRLQ-NL-D-- : 126
#Btau.RBP  : --AKGRVRLLNNWD--------VCADMVGTFTDTE----DPAKFKMKYWGVAS--FLQKGNDDHWIIDTDYE--TFAVQYSCRLL-NL-D-- : 126
#Vcho.Lpro : --NRGYSEEKGEWK--------EAEGKAYFVNGST-------DGYLKVSFFGP------FYGSYVVFELDRE--NYSYAFVSGP-------- : 114
#Ecol.OML  : --NKGYNPDRGMWQ--------QSEGKAYFTGAPT-------RAALKVSFFGP------FYGGYNVIALDRE---YRHALVCGP-------- : 115
#Cfre.OML  : --NKGYNPDREMWQ--------KTEGKAYFTGDPS-------TAALKVSFFGP------FYGGYNVIALDRE---YRHALVCGP-------- : 115
```

VIII Mouse urinary proteins
Ix α2u-globins
XI Miscellaneous lipocalinsI
X Chemoreception lipocalins I
V Neutrophil gelatinase lipocalins Prostaglandin D synthases
IV β-Lactoglobins
XII α1-Acid glycoproteins
VI α1-microglobins
VII Complement 8γ
XIII Chemoreception lipocalins II
II Hard tick histamine-binding proteins
II Soft tick TSGP's and platelet aggregation inhibitors
II Triatomine bug platelet aggregation inhibitors
II Triatomine bug nitrophorins
II Allergen IV
II nerve lipocalin
II Crustacyanin
II Nerve lipocalins
II Metazoan apolipoprotein D
II Insecticyanins
III Retinol-binding proteins
I Eubacterial and dictyostelid lipocalins

Fig. 7.9 (Page 2)

```
                               SCR3
#Mmus.MUP4 : --KTFQLMELYG-RKADLNSDIKEKFVKLCEEHGIIKENIIDLTKTNRCLKARE---------------------------- : 164
#Mmus.MUP5 : --ENFQLMELFG-REPDLSSDIKEKFAKLCEEHGIVRENIIDLSNANRCLQARE---------------------------- : 162
#Mmus.MUP1 : --ETFQLMGLYG-REPDLSSDIKERFAQLCEKHGILRENIIDLSNANRCLQARE---------------------------- : 164
#Mmus.MUP6 : --ETFQLMGLYG-REPDLMSDIKERFAQLCEEHGILRENIIDLSNANRCLQARE---------------------------- : 164
#Rnor.a2g3 : --ETFQLMELYG-RTKDLSSDIKEKFAKLCVAHGITRDNIIDLTKTDRCLQARG---------------------------- : 164
#Rnor.a2g4 : --EAFQLMELYG-RTKDLSSDIKEKFAKLCEAHGITRDNIIDLTKTDRCLQARG---------------------------- : 164
#Rnor.a2g1 : --ETFQAMVLYG-RTKDLSSDIKEKFAKLCEAHGITRDNIIDLTKTDHCLQARG---------------------------- : 164
#Rnor.a2g2 : --ETFQLMVLYG-RTKDLSSDIKEKFAKLCEAHGITRDNIIDLTKTDRCLQARG---------------------------- : 164
#Cfam.f2p  : --DTSLVAHLMV-RDLSRQQDFLPAFESVCEDIGLHKDQIVVLSDDDRCQGSRD---------------------------- : 162
#Tvul.Lip  : --KVNVVANLFG-RTPNLSNEIKKRFEEDFMNRGFRRENILDISEVDHC--------------------------------- : 165
#Ecab.p19p : --KETVVVTLFS-RTPTVSPDVMWMFKKYCKTHGIHTSNIVDLTQTDRCLHARH---------------------------- : 167
#Btau.OBP  : ---QTTELTGLFVKLNVEDEDLEKFWKLTEDKGIDKKNVVNFLENEDHPHPE------------------------------ : 159
#Sscr.OBP  : ---DKTIMTGLLGKGTDIEDQDLEKFKEVTRENGIPEENIVNIIERDDCPA------------------------------- : 157
#Mmus.Pbas : --NDVTRVAGVLAKG-KLNKEEMTEFMNLVEEMGIEEENVQRIMDTDNCPSKIRISITD----------------------- : 159
#Rnor.Pbas : -NDVTRVAGVLA-KGRQLTKDEMTEYMNFVEEMGIEDENVQRVMDTDTCPNKIRIR-------------------------- : 160
#Btau.alle : --RITKMTEGLA-KGTSFTPEELEKYQQLNSERGVPNENIENLIKTDNCPP------------------------------- : 157
#Rnor.OBP1 : --GRRQCDLVAG-KREDLNKAQKQELRKLAEEYNIPNENTQHLVPTDTCNQ------------------------------- : 156
#Ccri.Aphr : --QETNMIVVAG-KGNALTPEENEILVQFAHEKKIPVENILNILATDTCPE------------------------------- : 151
#Mmus.Aphr : --RKTKLLFILG-HG-PLTSEQKEKFAELAEEKGIPAGNIREVLITDYCPE------------------------------- : 149
#Rnor.NGAL : --KQYFKVTLYG-RTKGLSDELKERFVSFAKSLGLKDNNIVFSVPTDQCIDN------------------------------ : 178
#Mmus.NGAL : --KQYFKITLYG-RTKELSPELKERFTRFAKSLGLKDDNIIFSVPTDQCIDN------------------------------ : 180
#Hsap.NGAL : --REYFKITLYG-RTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDG------------------------------ : 178
#Rnor.PGDS : -GQDFRMATLYS-RAQLLKEELKEKFITFSKDQGLTEEDIVFLPQPDKCIQE------------------------------ : 165
#Mmus.PGDS : -GQDFRMATLYS-RTQTLKDELKEKFTTFSKAQGLTEEDIVFLPQPDKCIQE------------------------------ : 169
#Cfam.PGDS : -GQDFHMATLYS-RTQTPKAEIKEKFSTFAKTQGFTEDAIVFLPQTDKCMEENK---------------------------- : 171
#Ecab.PGDS : GGQDFRMATLYS-RVQSPRPEVKEKFSTFAKAQGFTEDAIVFLPQTDKCMEEHN---------------------------- : 174
#Bbub.BLB  : -PEQSLACQCLV-RTPEVDDEALEKFDKALKALPMHIRLSFNPTQLEEQCHV------------------------------ : 162
#Btau.BLB  : -PEQSLACQCLV-RTPEVDDEALEKFDKALKALPMHIRLSFNPTQLEEQCHI------------------------------ : 162
#Chir.BLB  : -PEQSLACQCLV-RTPEVDKEALEKFDKALKALPMHIRLAFNPTQLEGQCHV------------------------------ : 162
#Oari.BLB  : -PEQSLACQCLV-RTPEVDNEALEKFDKALKALPMHIRLAFNPTQLEGQCHV------------------------------ : 162
#Hsap.a1GP : -----WGLSVYA-DKPETTKEQLGEFYEALDCLRIPKSDVVYTDWKKDKCEPLEKQHEKERKQEEGES-------------- : 183
#Hsap.a1GP : -----WGLSFYA-DKPETTKEQLGEFYEALDCLCIPRSDVMYTDWKKDKCEPLEKQHEKERKQEEGES-------------- : 183
#Ocun.a1GP : -----KGMSFYT-DKPKASPEQLEEFYEALTCLGMNKTEVVYTDWTKDLCEPLEKQHEEERKKEKAES-------------- : 183
#Mmus.a1G3 : -----RGLSLSA-RRPDIPPELREVFQKAVTHVGMDESEIIFVDWKKDRCSEQEKKHLELEKETKKDPEESQA--------- : 188
#Rnor.a1GP : -----RGLSFYA-KKPDLSPELRKIFQQAVKDVGMDESEIVFVDWTKDKCSEQQKQQLELEKETKKETKKDP---------- : 187
#Mmus.a1GP : -----RGLSLNA-KRPDITPELREVFQKAVTHVGMDESEIIFVDWKKDRCSQQEKQQLELEKETKKDPEEGQA--------- : 189
#Mcar.a1G2 : -----RGLSLNA-KRPDITPELREVFQKAVKHVGMDESEIIFVDWKKDKCGQQEKKQLELEKETKKDPEEGQA--------- : 189
#Mmus.a1G2 : -----RGLSLYA-KRPDITPELREVFQKAVTHVGMDESEIIFVDWKKDRCGQQEKKQLELGKETKKDPEEGQA--------- : 189
#Mcar.a1GP : -----RGLSLYA-KRPDMTLELREVFQKAVKHVGMDESEIIFVDWKKDKCGQQEKKQLELGKETKKDPEEGQA--------- : 189
#Maur.A1mg : -HGPTITAKLYG-REPQLRDSLLQEFREVALSVGIPENSIVFMEDRGECVPGDLEQKSTSLLR------------------- : 182
#Rnor.A1mg : -HGPTITAKLYG-REPQLRDSLLQEFREVALSVGIPENSIVFMADRGECVPGDREVESTSFAR------------------- : 182
#Mung.A1mg : -HGPTITLKLYG-REPKLRDSLLLEFREVALSMGIPENSIVFMADKGECVPGDQEVKPSPHLR------------------- : 182
#Mmus.A1mg : -HGLTITAKLYG-REPQLRDSLLQEFKDVALNVGISENSIIFMPDRGECVPGDREVEPTSIAR------------------- : 182
#Hsap.C8GC : ----QLSVKLYA-RSLPVSDSVLSGFEQRVQEAHLTEDQIFYFPKYGFCEAADQFHVLDEVRR------------------- : 182
#Ocun.C8GC : ----QLSVKLYV-RSLPVSDSVLGAFEQRVAQANLTQDQVLFFPTYGFCEAADQFHILDEVRR------------------- : 182
#Rnor.OBP2 : --KSYLKGKLVG-RDSKDNPEAMEEFKKFVKSKGFREENITVPELLDECVPGSD---------------------------- : 157
#Mmus.VNSP : --MRFLMTQLIG-RDPQENLEALEEFKVFTQIKGLVAENLVILEQMEKCEPESFYELPSRPSE------------------- : 164
#Mmus.VNSP : --KISQFGYLIG-KYPEENQDTLEVFKEFIQHKGFLQEKIGVPEQRDRCIPIHDSAHQDHKC-------------------- : 169
#Rnor.VEG2 : --HHFQIAKLVG-RNPEINQEALEDFQNAVRAGGLNPDNIFTFIPKQSETCPLGSN-------------------------- : 166
#Rnor.VEG1 : --HHFQIAKLVG-RDPEINQEALEDFQSVVRAGGLNPDNIFTFIPKQSETCPLGSN-------------------------- : 166
#Hsap.VEG  : --KPVRGVKLVG-RDPKNNLEALEDFEKAAGARGLSTESILIPRQSETCSPGSD---------------------------- : 159
#Sscr.VEG  : --QEVRMAKLVG-RDPENNPEALEEFKEVARAKGLNPDIVRPQQSETCSPGGN----------------------------- : 158
#Cfam.f1p  : --RQIRMAKLLG-RDPEQSQEALEDFREFSRAKGLNQEILELAQSETCSPGGQ----------------------------- : 156
#Rapp.HBP1 : ----GYELWATD--YTDVP---ASCLEKFNEYAAGLP--VRDVYT-SDCLPE------------------------------ : 172
#Rapp.HBP2 : ----GYELWTTD--YDNIP---ANCLNKFNEYAVGRE--TRDVFT-SACLE------------------------------- : 171
#Rapp.HBP3 : ---GDYELWVSEDKIDKIP---DCCKFTMAYFAQQQEKTVRNVYTDSSCKPAPAQN-------------------------- : 182
#Osav.TSGP : -QGPFLELWYHS--GASEE-SMRCCEEEFRKNLKEGT--AVRKVN-KNCDYGDVA--------------------------- : 171
#Osav.TSGP : ----KCELWAPEGKAKNVE---SSCSGKFKELCGDAV---ETPYA-EGCRVP------------------------------ : 156
#Omou.MB   : ----VYQQWQSN--GSADDKDIKCCDEKFKELTSGID--YTKPQE-KGCETSAK---------------------------- : 150
#Osav.TSGP : ----AYEMWMLE--AGGLEVDIECCNKKYDELTSGQV--VIRPQD-KDC--------------------------------- : 144
#Osav.TSGP : ----AYDMWMLE--AGGLEVDIECCNKKYDELTSGQV--VIRPQD-KDC--------------------------------- : 144
#Rpro.PAI1 : ---ESPTKDNFLVAQRRKQSTFPSALESQVSKVGFGLKK-DSFKKFNC---------------------------------- : 157
#Rpro.PAI2 : ---SSAVEDNFLVFNRQSSGGIPGGLTTKLSQL--DLTP-TSFTKLGCT--------------------------------- : 156
#Tpal.PAL1 : ---ESGIADDVLLLQTKKEGVDP-GVTSVLKSVNWSLDDWFSRSKVNCDNMK------------------------------ : 171
#Tpal.PAL2 : ---ESGIADDVLLLQTKKEGVDP-GVTSVLKSVNWSLDDWFSRSKVNCDNMK------------------------------ : 170
#Rpro.NP2  : ---SKDLGDLYTVLTHQKDAEPSAKVKSAVTQAGLQLSQFVGTKDLGCQYDD-QFTSL------------------------ : 179
#Rpro.NP3  : ---PKDLGDLYTVLSHQKTGEPSATVKNAVAQAGLKLNDFVDTKTLSCTYDD-QFTSM------------------------ : 179
#Rpro.NP1  : ---NKDLGDLYAVLNRNKDTNAGDKVKGAVTAASLKFSDFISTKDNKCEYDNVSLKSLLTK--------------------- : 184
#Rpro.NP4  : ---NKDLGDLYAVLNRNKDAAAGDKVKSAVSAATLEFSKFISTKENNCAYDNDSLKSLLTK--------------------- : 184
#Bger.All4 : GHVIYVQIRFSV-RRFHPKLGDKEMIQHYTLDQVNQHKKAIEEDLKHFNLKYEDLHSTCH---------------------- : 170
#Dmel.Nlip : -LGHSDQIWILG-RDRDFEVDIRSKVYDVLKRL-SLDPERLIISKNKQCPEAL----------------------------- : 192
#Hgam.CRC2 : --YKSEFAFVFS-RTPQTSGPAVEKTAAVFNKNGVEFSK-FVPVSHTAECVYRA---------------------------- : 174
#Hgam.CRC1 : FGYHSDFSFIFS-RSANLADQYVKKCEAAFKNINVDTTR-FVKTVQGSSCPYDTQKTL------------------------ : 181
#Same.Laz  : SAASTEISWILL-RSRNSSNMTLERVEDELKNLQLDLNK-YTKTEQSAKYCA------------------------------ : 171
#Dmel.Dlip : -LANFKIVWILT-RQREPSAEAVDAARKILEDNDVSQAF-LIDTVQKNCPRLDGNGTGLAGEDGLDVDDFVSTTVPNAIEKA : 200
#Hsap.ApoD : -LFHVDFAWILA-RNPNLPPETVDSLKNILTSNNIDVKK-MTVTDQVNCPKLS----------------------------- : 169
#Mmus.ApoD : -LFHVDFFWILG-RNPYLPPETITYLKDILTSNGIDIEK-MTTTDQANCPDFL----------------------------- : 169
#Ccob.ApoD : -LFHVDHIWILG-RNRYLPQETVTYLKDILTSNNIDIEK-MTLTDQANCPDFL----------------------------- : 169
#Ocun.ApoD : -LFHMDHVWILG-RNRYLPPETVTYLKDILTANNIDIEK-MTVTDQVNCPEF------------------------------ : 168
#Pbra.Bbp  : KKGHQDFVWVLS-RSKVLTGEAKTAVENYLIGSPVVDSQ-KLYYSDFSEAACKVNN-------------------------- : 174
#Gmel.Gall : SKSHNDSIWILS-RAKKLEGDAKTAVDNYLKEHAKEIDA-SKLVQTDFSEEACKFTSTSAVTEPQTKKQ------------- : 186
#Msex.IcyA : KKAHSIHAWVLS-RNKVLEGNTKEVVDNVLKTFSHLIDA-SKFMSNEFSEAACQYSTTYSLTGPDRH--------------- : 189
#Msex.IcyB : KKAHSIHAWILS-KSKVLEGNTKEVVDNVLKTFSHLIDA-SKFISNDFSEAACQYSTTYSLTGPDRH--------------- : 189
#Ggal.Purp : GSCDDGYSLIFS-RNPRGLPPAIQRIVRQKQEEICMSGQ-FQPVLQSGAC-------------------------------- : 175
#Hsap.RBP  : GTCADSYSFVFS-RDPNGLPPEAQKIVRQRQEELCLARQ-YRLIVHNGYCDGRSERNLL----------------------- : 183
#Ecab.RBP  : GTCADSYSFVFA-RDPNGFPPEVQRIVRRRQEELCLARQ-YRLISHNGYCDGKSDRNLL----------------------- : 183
#Sscr.RBP  : GTCADSYSFVFA-RDPHGFSPEVQKIVRQRQEELCLARQ-YRLITHNGYCDGKSERNIL----------------------- : 183
#Btau.RBP  : GTCADSYSFVFA-RDPSGFSPEVQKIVRQRQEELCLARQ-YRLIPHNGYC-------------------------------- : 174
#Vcho.Lpro : ---NTEYLWLLS-RTPTVERGILDKFIEMSKERGFDTNR-LIYVQLQ----------------------------------- : 156
#Ecol.OML  : ---DRDYLWILS-RTPTISDEVKQEMLAVATREGFDVSK-FIWVQQPGS--------------------------------- : 159
#Cfre.OML  : ---DRDYLWILS-RTPTISDEMKQQMLAIATREGFEVNK-LIWVKQPGA--------------------------------- : 159
```

VIII Mouse urinary proteins

Ix α2u-globins

XI Miscellaneous lipocalinsI

X Chemoreception lipocalins I

V Neutrophil gelatinase lipocalins Prostaglandin D synthases

IV β-Lactoglobins

XII α1-Acid glycoproteins

VI α1-microglobins

VII Complement 8γ

XIII Chemoreception lipocalins II

II Hard tick histamine-binding proteins

II Soft tick TSGP's and platelet aggregation inhibitors

II Triatomine bug platelet aggregation inhibitors

II Triatomine bug nitrophorins

II Allergen IV

II nerve lipocalin

II Crustacyanin

II Nerve lipocalins

II Metazoan apolipoprotein D

II Insecticyanins

III Retinol-binding proteins

I Eubacterial and dictyostelid lipocalins

Fig. 7.9 (Page 3)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Although the global tree obtained for the lipocalins using neighbor-joining, differ from that obtained with maximum likelihood, the main monophyletic clades were still present (Fig. 7.10). The groupings within individual monophyletic clades were also essentially the same as found previously (Ganfornina *et al.* 2000; Gutiérrez *et al.* 2000). The lipocalins were still grouped into ancient lipocalins (clades I-III) which still formed a superclade although at cut-off value of 45% confidence, most arthropod groups from clade II were collapsed onto the main branch. This does however, give a better approximation of the true minimum evolutionary tree and shows that the sequences from clade II is problematic. The modern lipocalins (IV-XIII) grouped into a superclade of which the internal relationships differed from previous analysis. Clades VIII-XI, VI-VII and IV-V still grouped as supergroups. The positions of clades XIII (grouping with VI and VII) and XII (grouping with IV) differered from previous analysis where XIII grouped with clades V-IV and XII grouped in a separate clade with IX, X, XI and VIII (Fig. 7. 3). A lipocalin from cockroach (Bger.All4) previously grouped between the ancient and modern superclades, separated from the arthropod clade II (Gutiérrez *et al.* 2000). It was suggested that this is an artefact that reflects the inability of categorizing highly divergent sequences, although it does show sequence similarity to the nitrophorins. Of interrest is that both Bger.All4 and the triatomine bug lipocalins (nitrophorins as well as platelet aggregation inhibitors grouped into a monophyletic clade) as well as the tick lipocalins were placed between the ancient and modern lipocalins. Tick lipocalins grouped as a monophyletic clade, confirming that they are homologous. Internally, the tick lipocalins grouped into two monophyletic sub-clades composed of the hard tick (HBP) and the soft tick (TSGPs and moubatin) lipocalins.

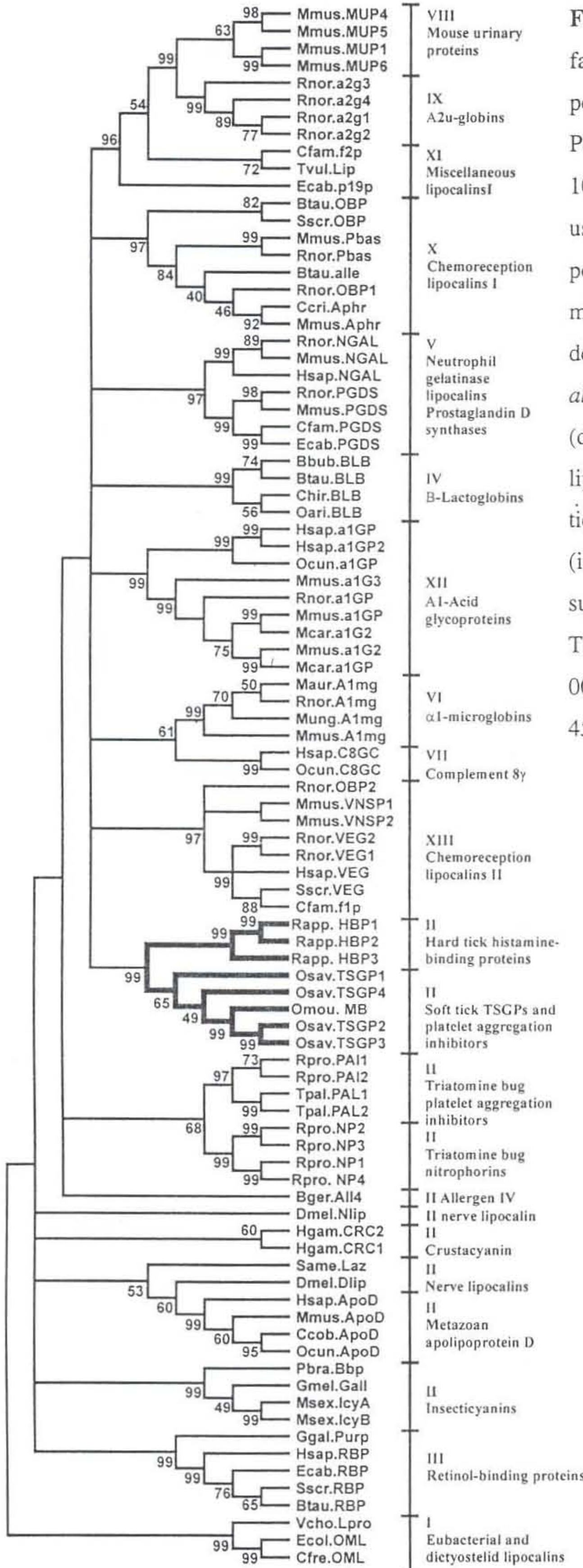

Fig. 7.10: Phylogenetic analysis of the lipocalins family. Neighbor-joining analysis were performed based on amino acid differences. Percentage confidence levels indicated are for 10000 bootstraps. The tree obtained was rooted using the eubacterial lipocalins. All gapped positions were ignored. Indicated is the different monophyletic clades (I-XIII) previously described (Ganfornina *et al.* 2000; Gutiérrez *et al.* 2000). Also present is a new clade (designated part of clade II) into which the lipocalins from the triatomine bugs group. The tick lipocalins group into a monophyletic clade (indicated in bold) that are divided into separate subgroups for the soft and hard tick lipocalins. The percentage confidence is indicated for 10 000 bootstraps and branches were collapsed at 45% confidence value.

7.3.6 Absence of TSGP2-4 in SGE from *O. moubata*

Phylogenetic analysis indicates that TSGP2 and TSGP3 group below moubatin. The high sequence similarity between moubatin and TSGP2-3 suggest that these are fairly recent gene duplication events. The possibility was raised that these events occurred after the divergence *O. moubata* and *O. savignyi*. If so, then it could be expected that TSGP2-3 will be absent from salivary gland extracts (SGE) of *O. moubata*. This possibility was investigated by a comparison of SGE using two-dimensional electrophoresis. There are several differences in protein expression patterns between salivary gland extract from *O. moubata* and *O. savignyi* (Fig. 7.11). Most notably is the different patterns observed for the TSGPs. While the same abundance of acidic TSGPs are observed in the SGE of *O. moubata*, their molecular masses and pI's differ from those of *O. savignyi*. A highly abundant basic protein (TSGP4) which corresponds to the basic toxin is also absent from the *O. moubata* proteome.

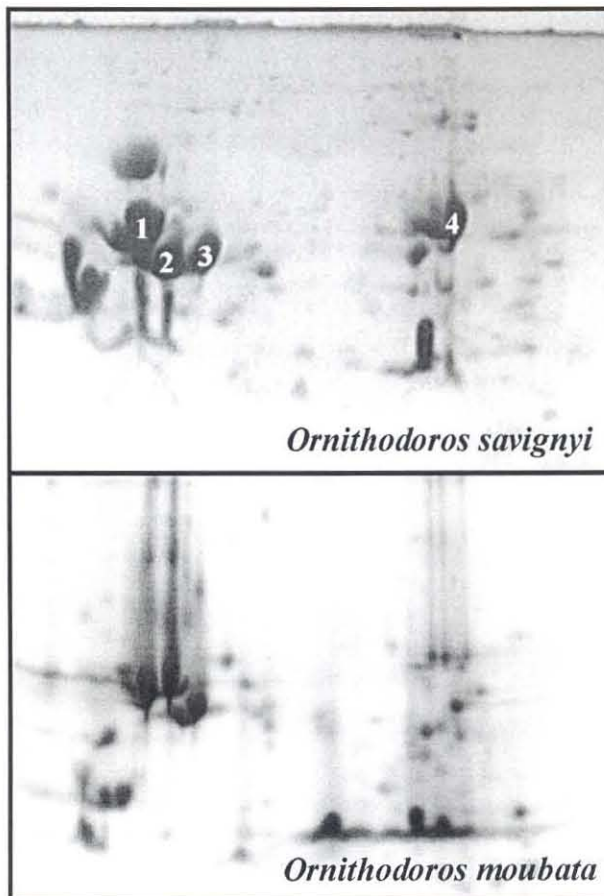

Fig. 7.11: A comparison of the SGE proteomes from *O. moubata* and *O. savignyi*. Acidic pI's are on the left and the anode at the bottom. The TSGPs previously identified are numbered 1-4.

To investigate this further, a comparative western blot analysis of SGE of *O. moubata* and *O. savignyi*, was performed using the polyclonal anti-sera directed against the different TSGPs (Fig. 7.12). Coomassie Brilliant Blue staining of SGE from *O. moubata* and *O. savignyi* shows that the protein concentrations in SGE from *O. moubata* was higher than SGE from *O. savignyi*. A very low dilution factor (500x) was also used with all anti-sera. Under such conditions, the presence of any toxins or proteins with a related sequence should have shown up clearly if present in the SGE from *O. moubata* in equivalent concentrations. Cross-reactivity of a-TSGP1 with a similar antigen in the SGE of *O. moubata* is evident. It is also clear that the signal for the *O. moubata* antigen is higher than that for *O. savignyi*, which fits with the higher protein concentration loaded on the gel. The detected protein is most probably the highly abundant antigen (20A1) previously detected in *O. moubata* (Baranda *et al.* 2000) that also shows N-terminal similarities to TSGP1. In contrast, both anti-sera specific for TSGP2 and TSGP3 gave no cross-reactivity with SGE from *O. moubata*, indicating the absence of both these proteins. There is however, cross-reactivity between TSGP2 and TSGP3 as was shown previously, which results in a broad band within which two distinct bands can be observed. Anti-sera against TSGP4 indicate no cross-reactivity for a strong band present in the SGE from *O. savignyi*, although there seems to be weak cross-reactivity with a lower molecular mass entity. These results do however also suggest that TSGP4 is absent in the SGE of *O. moubata*. Immunodetection is a validated method by which similarities between proteins can be compared. The results obtained clearly indicate the absence of any toxic entities and suggest a reason for the inability of *O. moubata* to produce toxicoses in their hosts.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

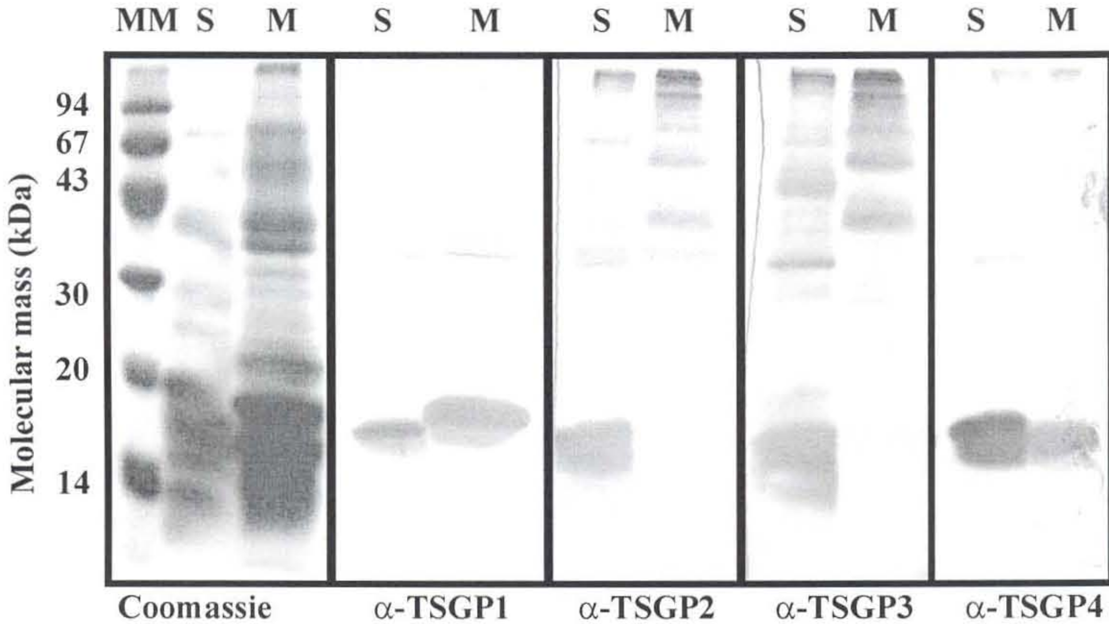

Fig. 7.12: Western blot analysis of TSGPs in SGE of *O. moubata* and *O. savignyi*. Indicated are molecular mass markers (MM), Coomassie Brilliant Blue staining of SGE of *O. savignyi* (S) and *O. moubata* (M). Cross-reactivity between anti-sera raised against TSGP1 can be observed, while there is no cross-reactivity for TSGP2 and TSGP3 anti-sera. A low molecular mass protein cross-reacts with TSGP4 anti-sera, although the main antigen is absent in SGE of *O. moubata*.

7.3.7 Molecular modeling of the TSGPs

The structure of HBP2 has been determined at high resolution of 1.25 Å (Paesen *et al.* 1999). Phylogenetic analysis indicates that the tick lipocalins are homologous, although sequence identity (13-15%) and similarity (27-34%) between the HBPs and TSGPs are quite low. Even with such low identity it might be possible to obtain structural models at low resolution (Fig. 7.13). The models obtained clearly indicate that all TSGPs exhibit the lipocalin fold although the models obtained gave very low RMSD values (4-6Å). Fitting the models obtained onto the structure of HBP2 shows that the main deviations occur in the mobile loop areas, while the β-barrel and α-helix structures are conserved.

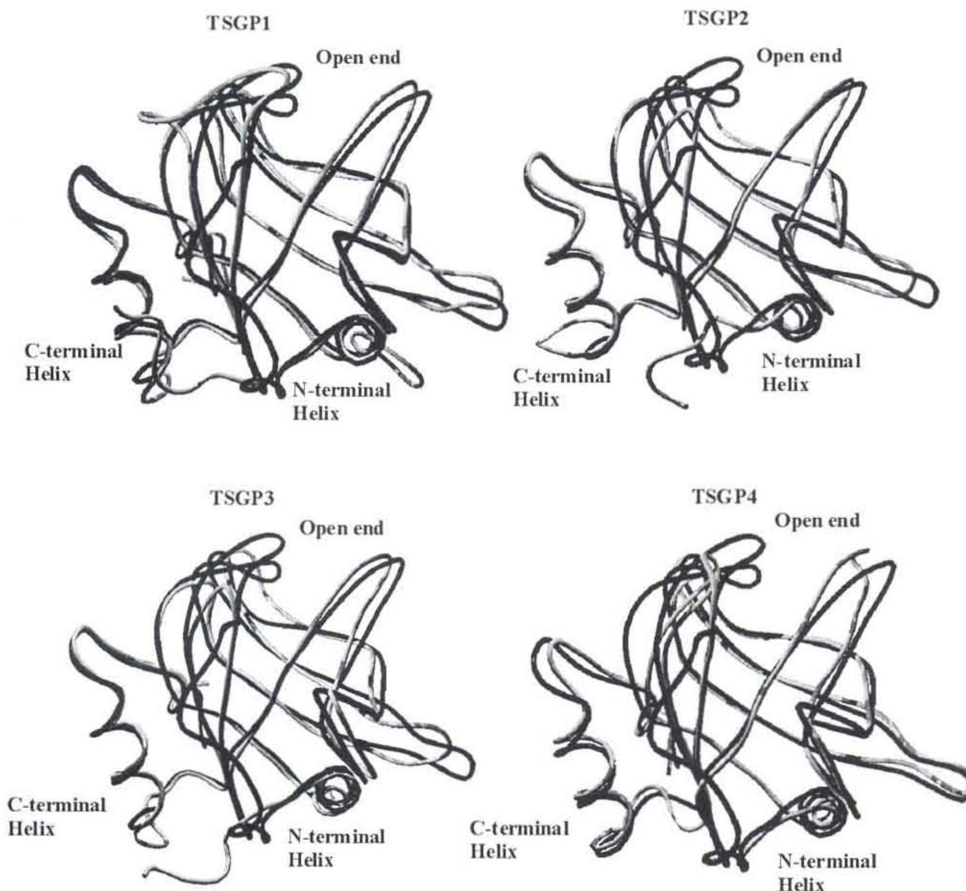

Fig. 7.13: Structural models obtained for the TSGPs fitted onto the structure of HBP2. HBP2 is indicated as the black ribbon tracing and the respective TSGPs as the dark gray ribbon tracing. Also indicated is the open end of the β-barrel and the N-and C-terminal α-helixes.

7.3.8 Analysis of the TSGP structures

Ramachandran plots for the TSGP structural models indicate that most of the psi/phi angles for the majority of the residues are in most favoured positions (Fig. 7.14). This centers around 72-76% for the TSGPs, with TSGP1 (73.4%), TSGP2 (74.8%), TSGP3 (71.1%) and TSGP4 (76.3%). The rest of the residues are predominantly in additionally allowed regions, with TSGP1 (20.9%), TSGP2 (20.3%), TSGP3 (23.1%) and TSGP4 (17.8%). There are 2.2%, 4.9%, 2.5% and 4.4% residues in generously allowed regions for TSGP1-4, respectively.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

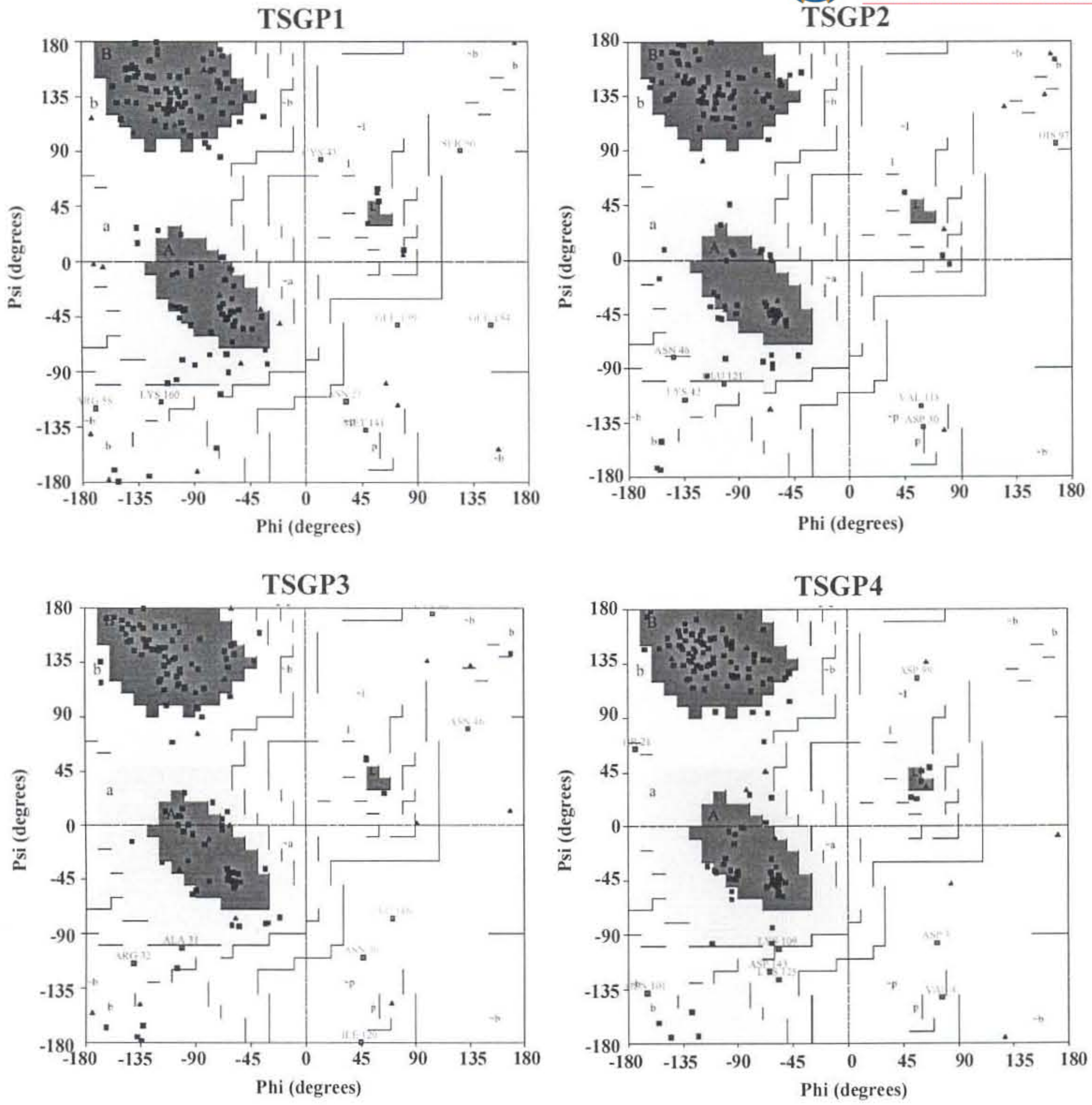

Fig. 7.14: Ramachandran plots for the TSGPs. Residues are indicated by triangles in most favoured (dark gray), additionally allowed (gray), generously allowed (light gray) and disallowed (white) regions.

7.3.9 Implications of the TSGP structures

The TSGP models show the lipocalin β-barrel and in each case three intact disulphide bonds (Fig. 7.15). Of interest is that the disulphide bonds in all the structures are localized to one side of the barrel. This could explain the lability observed for the toxins, as only one side of the barrel on the side of the C-terminal α-helix is stabilized.

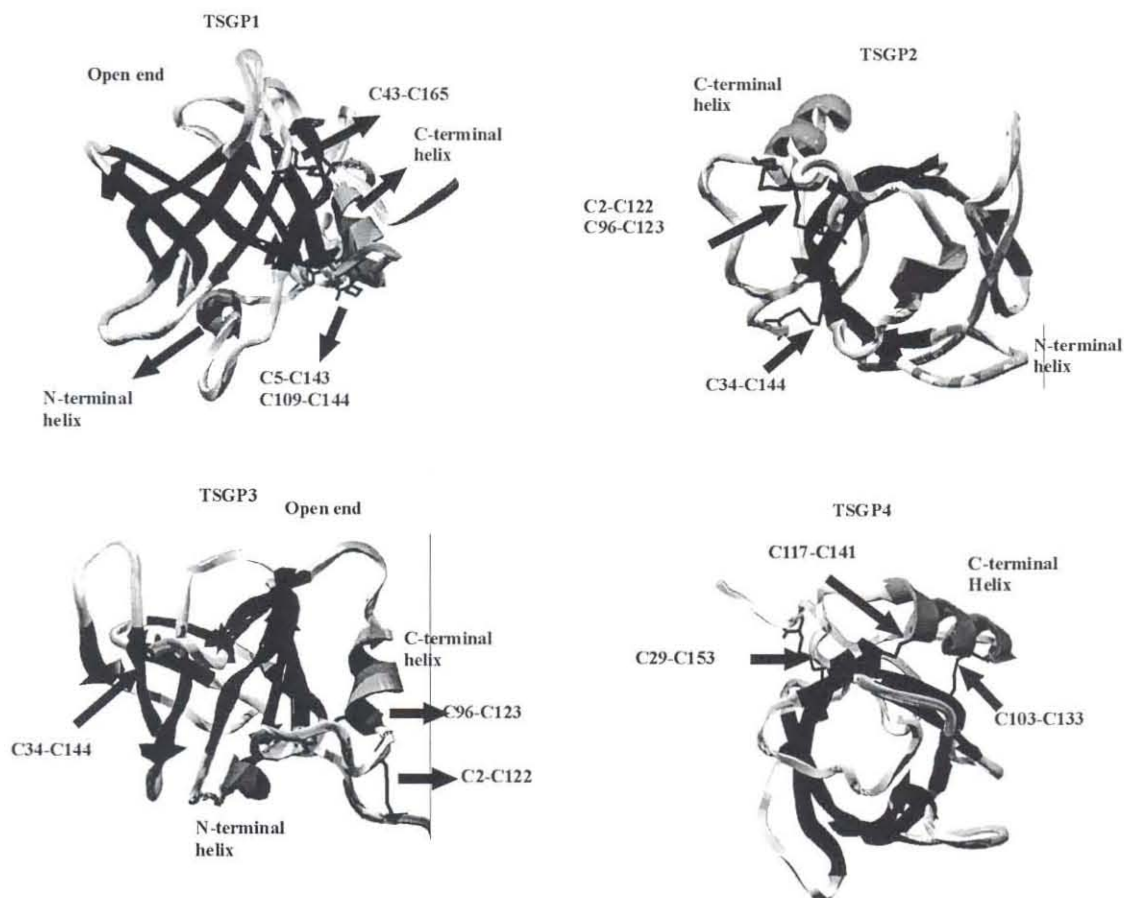

Fig. 7.15: The structures of the TSGPs and their intact disulphide bonds. The β-barrel is shown in black, helixes in dark gray and loops in light gray. The positions of the intact disulphide bonds are indicated. TSGP1 is viewed from the side and looking into the N-terminal 3_{10}-helix that closes the b-barrel. TSGP2 is viewed directly into the barrel from the closed end, so that the N-terminal 3_{10}-helix is at the top. TSGP3 is viewed from the side so that the disulphide bonds can clearly be seen. TSGP4 is viewed directly into the barrel so that the closed end is at the top.

7.3.10 Are toxins platelet aggregation inhibitors?

Sequencing and phylogenetic analysis have shown that TSGP2 and TSGP3 are closely related to moubatin, the inhibitor specific for collagen-induced platelet aggregation. Other lipocalins from the blood sucking bugs *R. prolixus* and *T. pallidipenis* have been shown to be specific inhibitors of the intrinsic Xase complex (nitrophorin-2), ADP-induced platelet aggregation (RPAI), thrombin (triabin) and collagen-induced platelet aggregation (pallidipin). This raised the question whether the TSGPs are involved in the regulation of the hemostatic system. Partial purification of a collagen-specific platelet aggregation inhibitory activity (designated savignygen) using RPHPLC and AEHPLC indicated no ADP- or collagen-induced platelet aggregation inhibitory activity associated with the toxins (Fig. 7.16a). ADP- and collagen-induced platelet aggregatory inhibitory activity was observed after RPHPLC in a distinct peak of which the first part (29-31) corresponded with savignygrin activity (Chapter 2), while the last part of the peak (31-33 minutes) corresponded to a 17 kDa protein that showed cross-reactivity with polyclonal antibodies directed against TSGP2 (Fig. 7.16a and Fig. 7.16b). This fraction was partially purified and showed specific inhibition of collagen-induced platelet aggregation, but not ADP-induced platelet aggregation (Fig. 7.16d). Of interest is that both this protein as well as moubatin elutes at ~20-25% acetonitrile, while the toxins elute at 40-45% acetonitrile (Waxman and Connolly, 1993). The N-terminal amino acid sequence obtained for the 17 kDa protein was: **AQDKCSEVRN** (results not shown).

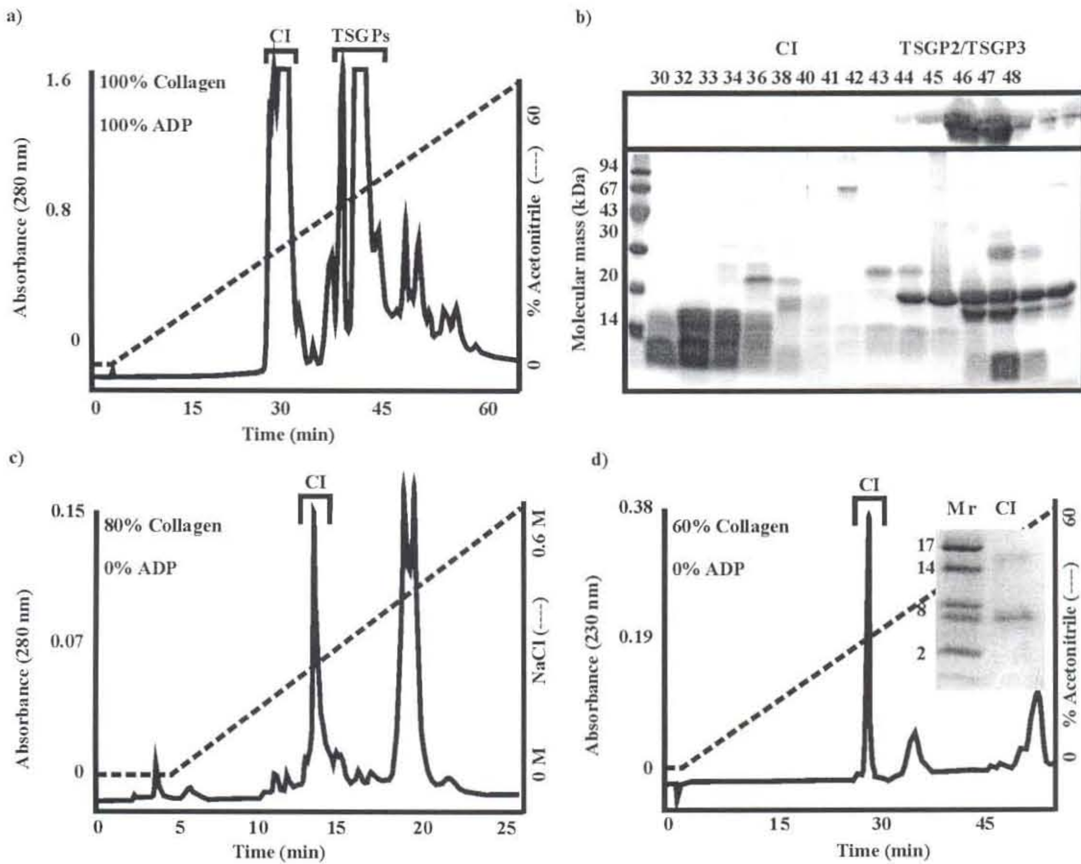

Fig. 7.16: Identification of savignygen, an inhibitor specific for collagen-induced platelet aggregation. (a) RPHPLC of SGE. Indicated are the fractions that inhibited ADP- and collagen-specific platelet aggregation (CI). The TSGPs that did not inhibit platelet aggregation are also indicated. (b) Western-blot analysis of RPHPLC fractions using a polyclonal antibody directed against TSGP2. Included is the corresponding SDS-PAGE electropherogram of the fractions. Numbers corresponds to fractions (in minutes) obtained from the RPHPLC. (c) AEHPLC of the inhibitor fractions (38-40 minutes) after RPHPLC. (d) Rechromatography of the inhibitory fraction obtained after AEHPLC. Indicated is percentage inhibition for ADP- and collagen-specific platelet aggregation inhibition, as well as a tricine SDS-PAGE analysis of the obtained fraction (insert). N-terminal sequence analysis of this fraction gave the sequence: AQDKCSEVNR.

It was also shown previously that no inhibitory activity of thrombin or fXa is associated with the TSGPs (Gaspar *et al.* 1995; Gaspar *et al.* 1996; Nienaber, Gaspar and Neitz, 1999). Using a photodiode array detector to collect wavelength spectra during RPHPLC showed that no prosthetic heme groups are associated with the TSGPs, as is found for the nitrophorins (Montford, Weichsel and Andersen, 2000). Histamine-agarose affinity

chromatography used to purify the HBPs failed to bind any TSGPs (results not shown) (Paesen *et al.* 1999).

Discussion

Evolution of tick lipocalins

The absence of histamine-binding proteins in soft ticks suggests that a lipocalin-like protein was present in the ancestral tick, which after speciation duplicated and evolved new functions. In the case of hard ticks that feed for extensive peroids of time, regulators of inflammation were evolved, while soft ticks that feed rapidly do not need anti-inflammatory mediators. Instead, inhibitors of platelet aggregation as in the case of TAI, moubatin and savignygen evolved to cope with the host's hemostatic system. Recent evidence indicates that ticks originated in the late Cretaceous (~120 MYA) and had already speciated into the main tick families by approximately 94 MYA (Klompen and Grimaldi, 2001). Using this estimate, the divergence of these two lipocalin families can be dated to this period. The selective stress of adaptation to a blood-feeding environment during this period of tick evolution, could account for the high degree of divergence of tick lipocalins relative to those of other arthropod lipocalins.

7.5.2 TSGP1

The placing of TSGP1 basal to the rest of the soft tick inhibitors suggests that it was present before the divergence of *O. savignyi* and *O. moubata*. This fits with the presence of a highly cross-reactive protein in *O. moubata* as well as N-terminal sequence similarity to a highly abundant antigenic protein 20A1 (Chapter 7).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

7.5.3 TSGP2 and TSGP3

Phylogenetic analysis indicate that TSGP2 and TSGP3 are grouped below moubatin. The absence of TSGP2/TSGP3 in SGE of *O. moubata*, the high sequence similarity between moubatin and TSGP2-3, as well as between TSGP2 and TSGP3 suggest that these are fairly recent duplication events that occurred after the divergence of *O. moubata* and *O. savignyi*.

7.5.4 Toxicity vs. non-toxicity: TSGP2 and TSGP3

It was previously shown that TSGP2 and TSGP3 have identical elution times during RPHPLC, that indicate similar global hydrophobicity, corroborated by their similar amino acid sequences (Chapter 7). In contrast, TSGP2 and TSGP3 separate on HIHPLC, indicating differences in their surface hydrophobicity and hence surface conformation. This is supported by the differences observed in the surface accessibility of the modeled structures of TSGP2 and TSGP3. The high identity of TSGP2 to TSGP3 and the apparent lack of toxicity for TSGP3 indicate that the toxic activity is probably localized to a very small region of different local surface conformation. A likely candidate is on loop 5, where TSGP2 has an arginine, while both TSGP3 and moubatin have glycines. It seems improbable that two sequences sharing such high sequence identity, would exhibit different biological activities. However, different activities have also been indicated for the nitrophorins. Nitrophorin 2 exhibits anti-coagulant activity, which is absent in nitrophorins 1, 3 and 4 that share a high percentage of amino acid sequence similarity (Zhang *et al.* 1998; Montford, Weichsel and Andersen, 2000). Another example is *Coffea arabica* methyltransferase (CaMXMT) which is an active enzyme in coffee plants, while a set of paralogues CaMTL (*C. arabica* methyltransferase-like) share high identity (80-84%) but show no activity (Ogawa *et al.* 2001).

7.5.5 TSGP4

Western blot analysis has indicated that TSGP2-4 are absent in *O. moubata*, although a degree of cross-reactivity was observed in both *O. savignyi* and *O. moubata* using anti-sera raised against TSGP4. This might indicate a related non-toxic protein present in both species. Phylogenetic analysis places TSGP4 before moubatin, which suggests that

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

TSGP4 might be present in both tick species. Caution should however, be exercised as these are paralogous genes and no knowledge exists about their last common ancestor. It could even be speculated that TSGP4 duplicated after divergence of the two tick species from a shared ancestral gene, indicated by cross-reactivity by western-blot analysis.

7.5.6 Evolution of tick lipocalins and its implication for tick toxicoses

Gain of toxic function after divergence of *O. moubata* and *O. savignyi* have important implications for the origins of toxicity in ticks. Recent aquisition of genes coding for toxins through gene duplication discounts a common ancient origin for all tick toxins and suggests the probability of multiple unrelated origins for toxins. Protein and sequence data indicate specific differences between the TSGP toxins and paralysis toxins that shows that their molecular nature is different (Chapter 6).

7.5.7 Tick lipocalins as toxins

That lipocalins can be deleterious is not surprising. Many common allergens involved in allergic reactions are lipocalins (Mäntyjärvi, Rautiainen and Virtanen, 2000). Multiple molecular recognition properties and binding to cell surface receptors, are being considered to be general properties of lipocalins (Flower, 2000). It is possible that some exogenous lipocalins, derived from blood feeding organisms can recognize specific receptors and impede their function. In fact, lipocalins that inhibit collagen-induced platelet aggregation, probably interact with the collagen platelet receptor (Waxman and Connolly, 1993).

7.5.8 Structure of TSGPs

The structure of HBP2 has two N-terminal α-helixes that close off the β-barrel in contrast to other lipocalins which posses a N-terminal 3_{10}-helix. Multiple sequence alignment indicate that the other tick lipocalins, probably also diverge from other lipocalins in this respect, although TSGP4 probably lacks the first short α-helix. It was previously shown that all cysteines of the toxins are involved in disulphide bonds (Chapter 6). Four cysteines conserved in al tick lipocalins are involved in disulphide bonds in the HBPs and serve to pin the C-terminal region to the side of the barrel (Paesen *et al.* 1999). Cys119-

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Cys148 pins the start of the C-terminal α-helix to the start of the β-G sheet, while Cys48-Cys168 links the C-terminal end with the base of the Ω-loop (Fig. 7.15). Assuming that this disulphide bond pattern are consistent for other tick lipocalins, the remaining disulphide bonds present in the toxins can be inferred from the overall topology of the lipocalin fold of HBP2 (Fig. 7.17). In TSGP4 Cys117-Cys140 also pins the end of the C-terminal α-helix to the start of the β-H sheet. The majority of the TSGP4 structure is thus not stabilized by disulphide bonds and could explain the lability of this toxin. The same is true for TSGP2 and TSGP3, where the extended N-terminal helix is disulphide bonded to the C-terminal helix (Cys2-Cys121). This probably stabilizes the position of the N-and C-terminal helixes relative to one another, but again stabilization of the majority of the structure by disulphide bonds is absent. Of interest is that whereas TSGP2 and TSGP3 have intra-protein disulphide bonds, as exemplified by ESMS and electrophoresis under reducing and non-reducing conditions, HBP3 lacks the N-terminal cysteine and forms a disulphide-linked dimer. It has been shown that the core structure of HBP2 is extremely rigid. In contrast the surface residues were found to be very mobile (Paesen *et al.* 1999). This corresponds to the loop areas, which are also the least conserved in the tick lipocalin family and generally in all lipocalins (Skerra, 2000). Such mobile structures could explain the lability of toxic activity (Chapter 6).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

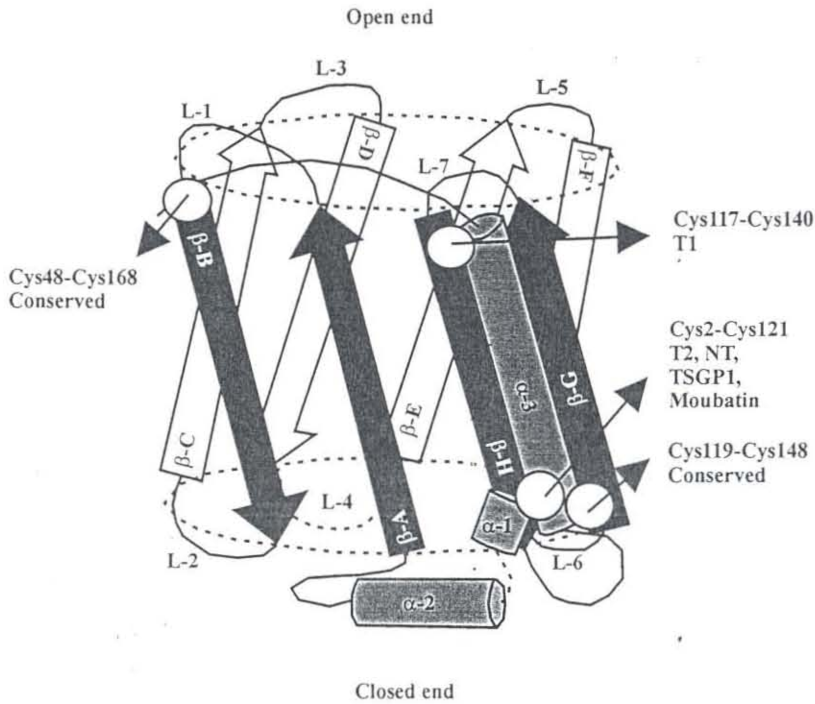

Fig. 7.17: The proposed disulphide bond pattern of the tick toxins. A schematic representation of the topology of HBP2 indicates two N-terminal α-helixes (α-1, α-2) followed by an 8 stranded β-barrel (β-A to β-H), which ends in a C-terminal α-helix α-3) that lies parallel to β-G and β-H. The conserved disulphide bonds of HPB2 is indicated with white circles (Cys48-Cys168; Cys119-Cys148) and shows that the C-terminal end is bound to the start of β-B and the start of the C-terminal helix to the start of β-G. Assuming that this disulphide bond pattern is conserved across the tick family, it follows that for TSGP4 the remaining disulphide bond is between the end of α-3 and the beginning of β-H and for TSGP1, TSGP2, TSGP3, and moubatin at the start of α-1 and the second cysteine at the beginning of α-3. The proposed disulphide bond patterns fits in well with the existing topology known for HBP2 (Paesen *et al.* 1999).

7.5.9 Implications of molecular crowding for the evolution of tick lipocalins

Molecular crowding as mechanism of granule formation has several implications for protein evolution. Molecular crowding as a phenomenon depends on high protein concentrations, while aggregation is also dependent on protein structure. Specific sequence signals possibly do not play such an important role in aggregation. This implies that residues important for the maintenance of the lipocalin fold would be conserved. It has been shown that as little as 8% of all amino acid residues are involved in the maintenance of protein structure (Rost, 1997). For the lipocalins this means that the rest

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

of the residues are under no selective pressure and can undergo genetic drift under neutrality theory. This could be one of the reasons why lipocalins is such a diverse protein family. It is also the ideal environment for the evolution of new protein function and could account for the fact that even orthologous proteins could be highly diverged. This is counter-intuitive to the general idea of the conservation of orthologous function, but would be the case if protein function is not coupled to specific sequence epitopes, but rather to a general protein fold.

7.5.10 The tick lipocalin family

The TSGPs are part of what is becoming an extended family of tick lipocalins. It is clear that these different paralogs evolved by gene duplication events. It is foreseen that more lipocalins with diverse functions will be found in different tick species. While lipocalins have a conserved fold, there are several significant differences between those of ticks and the rest of the lipocalin family. Vaccines targeted at these specific features, might thus be viable and it is possible that a wide range of cross-protection might be accomplished with these proteins.