Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 12809–12816, November 1997
Biochemistry

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 30, 1996.

The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase subfamily consisting of many trophoblast-expressed genes

(cattle/implantation/mutational rate/multiple gene family/placenta)

SANCAI XIE*†, JONATHAN GREEN‡, JAMES B. BIXBY†, BOZENA SZAFRANSKA†§, JAMES C. DEMARTINI¶, STEVEN HECHT¶‖, AND R. MICHAEL ROBERTS*†**

Departments of Veterinary *Pathobiology, †Animal Sciences and ‡Biochemistry, University of Missouri, Columbia, MO 65211; and ¶Department of Pathology, College of Veterinary Medicine and Biomedical Sciences, Fort Collins, CO 80523

Contributed by R. Michael Roberts, September 16, 1997

ABSTRACT **The pregnancy-associated glycoproteins (PAGs) are structurally related to the pepsins, thought to be restricted to the hooved (ungulate) mammals and characterized by being expressed specifically in the outer epithelial cell layer (chorion/trophectoderm) of the placenta. At least some PAGs are catalytically inactive as proteinases, although each appears to possess a cleft capable of binding peptides. By cloning expressed genes from ovine and bovine placental cDNA libraries, by Southern genomic blotting, by screening genomic libraries, and by using PCR to amplify portions of PAG genes from genomic DNA, we estimate that cattle, sheep, and most probably all ruminant Artiodactyla possess many, possibly 100 or more, PAG genes, many of which are placentally expressed. The PAGs are highly diverse in sequence, with regions of hypervariability confined largely to surface-exposed loops. Nonsynonymous (replacement) mutations in the regions of the genes coding for these hypervariable loop segments have accumulated at a higher rate than synonymous (silent) mutations. Construction of distance phylograms, based on comparisons of PAG and related aspartic proteinase amino acid sequences, suggests that much diversification of the PAG genes occurred after the divergence of the Artiodactyla and Perissodactyla, but that at least one gene is represented outside the hooved species. The results also suggest that positive selection of duplicated genes has acted to provide considerable functional diversity among the PAGs, whose presence at the interface between the placenta and endometrium and in the maternal circulation indicates involvement in fetal–maternal interactions.**

The placenta is the hallmark of the eutherian mammal, yet, rather than being the most anatomically conserved mammalian organ, it is arguably the most diverse (1). Placentation ranges from the invasive hemochorial type, as in the human, where the trophoblast surface is in direct contact with maternal blood, to the epitheliochorial (e.g., pig) where the uterine epithelium is not eroded (2). Not only is placental structure highly variable, the polypeptide hormones the placenta produces also vary between species (1, 3). For example, no group of mammals other than higher primates possesses a chorionic gonadotrophin homologous to hCG for luteal support in early pregnancy, and only the ruminant ungulates are known to produce type I interferon as an antilyteolytic hormone (4).

Placentation in ruminants, such as cattle and sheep, is superficial, relatively noninvasive, and known as synepitheliochorial cotyledonary (5). "Synepitheliochorial" describes the fetal–maternal syncytium formed by the fusion of trophoblast binucleate cells and uterine epithelial cells, whereas "cotyledonary" describes the gross structure of the placenta and specifically the tufts of villous trophoblast (cotyledons) that insinuate themselves into the crypts of the maternal caruncles. These regions of interdigitated and partially fused fetal cotyledonary and maternal caruncles are the placentomes and are the main sites for nutrient and gas exchange in the placenta. The binucleate cells, which compose about 20% of the surface epithelium (trophectoderm), migrate and fuse with maternal uterine epithelial cells and deliver their secretory products directly to the maternal system. Among the products are the placental lactogens (6) and the pregnancy-associated glycoproteins (PAGs) (7), the subject of this paper.

The PAGs, also known under a variety of other names including pregnancy-specific protein B (8), were discovered in attempts to develop pregnancy tests for livestock (9–11). Rabbits were injected with extracts of placental cotyledons, and antibodies not directed against placental antigens were removed by adsorption with tissue extracts from nonpregnant animals. The resulting antisera provided the basis of an accurate pregnancy test for cattle and sheep as early as 1 month postinsemination.

Xie *et al.* (12) used an antiserum directed against purified PAGs from cattle and from sheep to screen cDNA libraries from late placental tissue. The full-length cDNAs shared 86% nucleotide sequence identities with each other and a surprising 60% sequence identity to pepsinogens. The PAGs had mutations in and around their active sites that would render them inactive as proteinases (12, 13). The similarities to pepsin A (≈50% amino acid identity) and chymosin (≈45% amino acid identity) in primary structure has allowed atomic models of ovPAG1 and boPAG1 to be built (13). Both molecules have the bilobed structure typical of all known eukaryotic aspartic proteinases and possess a cleft between the two lobes capable

© 1997 by The National Academy of Sciences 0027-8424/97/9412809-8$2.00/0
PNAS is available online at http://www.pnas.org.

Abbreviations: PAG, pregnancy-associated glycoprotein; ovPAG, ovine PAG; boPAG, bovine PAG; RT-PCR, reverse transcription–PCR.

§Present address: Department of Animal Physiology, University of Agriculture and Technology, 10-718 Olsztyn, Poland.
‖Present address: Department of Biology, Northeast Louisiana University, Monroe, LA 71211.
**To whom reprint requests should be addressed. e-mail: michael_roberts@muccmail.missouri.edu.

of accommodating peptides up to 7 aa long. Modeling strongly suggested that both ovPAG1 and boPAG1 can bind the pepsin inhibitor pepstatin, a prediction that has been validated (J.G. and R.M.R., unpublished results).

Even in the initial studies (8, 10, 12) it was clear that the PAGs were heterogenous in molecular weight and charge, and as more isoforms have been purified it has become evident that they differ in their amino-terminal sequences (14, 15). Further library screening has revealed additional transcripts in ruminants (16, 17) and the existence of PAGs in nonruminant species such as the pig (18) and the horse (13, 19). This paper addresses the heterogeneity of the PAGs in cattle and sheep, the interrelationship of these molecules with PAGs from other species, and their evolutionary origin.

MATERIALS AND METHODS

Cloning of Ovine PAG (ovPAG) Transcripts from Early Conceptuses and Day 100 Placenta. Two procedures were used to clone ovPAG transcripts from day 13 conceptuses and day 100 placenta. First, a day 13 ovine conceptus cDNA library was screened with a mixture of bovine, ovine and porcine PAG1 and PAG2 cDNA, and equine PAG1 cDNA labeled with ^{32}P by random primer reactions as described (17, 18). The positive clones were isolated and partially sequenced.

Novel PAG transcripts from day 100 ovine placenta were cloned by reverse transcription (RT) and PCR procedures (15). Cellular RNA extracted from a day 100 placenta was first reverse transcribed into cDNA and then amplified by PCR with a pair of well-conserved primers (ovPAG forward, 5′-AGGAAGAAGCATGAAGTGGCT-3′; ovPAG reverse, 5′-TAAATGCTTGGACTTGTTCAGGAA-3′). The RT-PCR products were cloned into TA cloning vectors (Invitrogen). All the novel ovPAG cDNA were fully sequenced in both directions.

Cloning of boPAG Transcripts from Early Bovine Placenta (boPAG). Three procedures were used to clone boPAG transcripts from day 19 and 25 placentae. First, RNA from six (Simmental × Hereford) placentas at day 25 of pregnancy was used to construct a cDNA library in λZAPII (CLONTECH). The library was screened with a mixed probe of ^{32}P-labeled bovine, ovine, and porcine PAG1 and 2, and equine PAG cDNA (12, 16–19). The positive clones were isolated and analyzed for the size of inserts by PCR and restriction endonuclease digestion. Sixteen clones of the anticipated length were partially sequenced. The second screening identified boPAG transcripts that reacted with an anti-boPAG1 antiserum (10, 12). Duplicate filter screening was used to increase the frequency of isolation of full-length clones. The first filter was allowed to react with antiserum to identify immunopositive clones (12), whereas the second filter was hybridized with a ^{32}P-labeled probe corresponding to exons 1 and 2 of boPAG1, ovPAG1, and ovPAG2. The clones positive on both filters were purified and partially sequenced.

PAG transcripts from a day 19 trophoblast of a Holstein cow were cloned by RT and PCR procedures. Cellular RNA, extracted from day 19 trophoblast tissue dissected free of embryo proper and yolk sac, was first reverse transcribed into cDNA then amplified by PCR with a pair of well-conserved primers (boPAGexp5′, 5′-CCCAAGCTTATGAAGTGGCT-TGTGCTCCT-3′, and boPAGexp3′, 5′-GGGAAGCTTAC-TTGTCATCGTCGTCCTTGTAGTCGGTACCCACCTGT-GCCAGGCCAATCCTGTCATTTC-3′). The RT-PCR products were cloned into TA cloning vectors (Invitrogen). All the novel boPAG cDNAs were fully sequenced.

Cloning of ovPAG Gene Fragments. DNA isolated from sheep white blood cells (buffy coat) was used as a template to amplify PAG genes at regions corresponding to part of exon 3, intron 3, and part of exon 4 (see ref. 17) by a PCR procedure with two sets of primers. The first set was derived from regions well conserved in both ovPAG1 and boPAG1 (bp314e3f, 5′-TGGGTAACAT-CACCATTGGAA-3′; bp314e4r, 5′-TTGTTGTTCATGACA-CAGTTC-3′). The second set had sequences identical to ovPAG2 in corresponding regions (bp314e3f, 5′-TGGGTAA-CATCACCATTGGAA-3′; op13e4r, 5′-CTTGCCTATGA-CATGTTGG-3′). The amplified ovPAG gene fragments were then cloned into TA cloning vectors and partially sequenced. If a clone represented a distinct ovPAG gene, it was fully sequenced (20).

Southern Blot Analysis. Genomic DNA (5 μg) was digested to completion with a series of restriction endonucleases. The DNA fragments were resolved on a 0.8% (wt/vol) agarose gel by electrophoresis and blotted onto nylon membranes (20). Specific DNA probes (803 bp) from the boPAG1 and boPAG2 genes (a part of intron 6, exon 7, intron 7, exon 8, and a part of intron 8) (17) were labeled by using the PCR procedure in the presence of [α-^{32}P]dATP (20). Hybridization and subsequent washing of the membrane was performed under conditions described previously (17). The same blots were first hybridized with the boPAG1-specific probe and subsequently with the boPAG2 probe.

Genomic DNA from various mammals was digested with *Eco*RI and separated by gel electrophoresis as described (21). Southern blots of these gels were probed under standard conditions with the boPAG1 probe described above (17).

Evolutionary Analyses. The amino acid sequences of PAG and selected aspartic proteins were first assembled into multiple sequence alignments by the PILEUP program or the Wisconsin GCG Package, version 9.0 (Madison, WI). A distance matrix was then created (program DISTANCES) and a phylogenetic tree constructed (GROWTREE) by a neighbor-joining procedure (22).

To estimate the number of synonymous and nonsynonymous substitutions within coding regions of the PAG genes, the program NEWDIVERGE was used to make codon-by-codon comparisons (23). Kimura's two-parameter method (24) corrected for multiple hits and differences in substitution rates for transitions and transversions.

RESULTS AND DISCUSSION

Cloning of ovPAG from Ovine Placenta. Two ovPAG cDNA have been previously identified. The first (ovPAG1) was obtained by screening a day 100 placental library with an antiserum raised against an antigen purified from late placental tissue (12). The ovPAG2 cDNA was isolated from a cDNA library prepared from day 13 whole conceptuses by homology screening with a boPAG1 probe under relatively nonstringent conditions (13, 25). In the present study, mRNA from day 100 placental tissue was reverse transcribed and subsequently amplified by PCR. Because the primers used well-conserved PAG gene sequences located just 5′ to the ATG start codon and 3′ to the stop codon, all the cDNA identified contained uninterrupted ORFs.

A total of 66 cDNA were partially sequenced, of which 27 corresponded to the previously identified ovPAG2. However, 39 previously uncharacterized cDNAs were discovered and fully sequenced; they fell into seven groups, subsequently named ovPAG3 to ovPAG9, based on the order of their discovery (Fig. 1). To limit the amount of sequencing performed PAG cDNAs that differed by $<5\%$ at the nucleotide level in the initial sequencing reactions are not listed, even though they may represent distinct, presumably recently duplicated, genes. Only two cDNA (ovPAG3 and ovPAG7) encoded polypeptides corresponding at their amino termini to proteins that had previously been purified (15). The remainder represented novel PAG.

OvPAG4 and ovPAG7 are the most related of the cDNA, sharing 95% nucleotide and 90% inferred amino acid sequence identities. The most distantly related are ovPAG2 and ovPAG4

```
                                                                       ↓ (1)
Consensus  MKWLVLLGLV AFSECIVKIP LRRVKTMRKT LSGKNMLNNF LKEHPYRLSQ ISFR.SNLTI  60  (3)

ovPAG4     .......... ......F... .......... ........DV ........P. ....G...I.  60
ovPAG7     .......... ......F.V. ...L...... ...N.....V ...Y....P. ....D..V..  60
ovPAG3     ......F... ........V. ...L...... ...N.I...V ...Y....P. ....D..V..  60
ovPAG6     .......... SI........ .......... ..E....... ....A..... ....G.....  60
ovPAG1     .......... .......... ........N. ....K...S. ....A..... ....A.....  60
ovPAG5     ....SV.... RV.....E.. .MK.NV.QE. .NE.SL.... ..KY....A. N.T.N.KMAF  60
ovPAG9     .......... ........M. .........N P.....M... ....A..... ....GL...-  59
ovPAG2     .......W.. .L.....IM. .TKT....EI .RE..L.... .E.QAN.M.D D.ASDPK.ST  60
ovPAG8     .......A.. ......I... .......SN. A......... ..K....... ....G....T  60

Consensus  HPLRN..D.. YVGNITIGTP PQEFQVVFDT GSSDLWVPS. FC.S.T---C S.H.RFRH.Q  120 (60)

ovPAG4     .....IR.TF .......... ......L... ...V.....V L.N.S..... .I.V....L.  117
ovPAG7     L....MK.IF .......... .P......S T.........I .WN.S..... .TLV..K.R.  117
ovPAG3     L....MK.IF .......... .......... A.........I ..N.S..... .TRV....R.  117
ovPAG6     .....TK.LV .L........ ......S... ..........D ..AIEA.... .L.T....L.  117
ovPAG1     .....IM.ML .......... .......... ....L...I N.L.P.KRP. .KQDK.K.H.  120
ovPAG5     V.M..FL.LA ..P.M...RG GEQ.R..... .........N ..T.PA.... YS.IT.KYWE  117
ovPAG9     ...S.YI.ML .T........ ..K..I.... .........N ..T.PA.... .TQA...YR.  116
ovPAG2     .....AL.MA ....V..... .K..R..... .........I K.I.PA.... YT.IT.D.HK  117
ovPAG8     ...M.IW.LL .L........ ......L..R .........L L.N.S..... AK.VM...RL  117

Consensus  SSTFR.TNKT F.I.YG.G.M KGVVAHDTVR IGDLVSTDQP FGLS..E.GF ...PFDGVLG  180 (120)

ovPAG4     .....T.... .W.T..A.T. .......... ......I... ....MA.Y.. HGRR......  177
ovPAG7     .....T.... .W.T..A.T. .......... ......I... ....MA.Y.. MDRR...I..  177
ovPAG3     .....LA... .G.M..A.K. ....V..... .......... ....VV.S.. EHRQ......  177
ovPAG6     .....P..R. .S.T..C.TV ....V..... .......... ....TA.HVS RCT.......  177
ovPAG1     .....F..D. .R.YF.S.T. R.F....... .......... ...IFL.SWL D-I....I..  179
ovPAG5     ...Y.H.T.P .E.A..S.RI ..HL.Y..IQ ..N....... ....LV.Y.. NGL....I..  177
ovPAG9     .....L..R. .C.T..S.GL ..I....... ..N....... ...TLK.Y.. ENI.......  176
ovPAG2     .....L.RRP .R.L..S.M. N..L.Y.... ..N....... ....LQQF.. DNA.......  177
ovPAG8     ...Y.L.... .M.F.RV.KI E...VR.... ......A..T ....IA.T.. ENTTL..I..  177

Consensus  LNYP..S..G .IPIFDKLKN EGAISEPVFA FYLSKD.QEG SVVMFGGVDH .YYKGELNWV  240 (180)

ovPAG4     ....RQ.CCR PT........ Q......... ......E... .......... R.........  237
ovPAG7     ....RQ.CSK TKW......S Q......... ......E... .......... T.......SL  237
ovPAG3     ....NL.FSK T......... .......... ......E... .......... S....K....  237
ovPAG6     ....SI.FWS T......... .......... ......G... .......... R.........  237
ovPAG1     ....KI.FS. A......... ...F...... ...N..K... .......... R.........  239
ovPAG5     ....N-.IL. A.....N..K Q......I.. .....GTVN. ..L.L....K A........I  236
ovPAG9     ....NI.FS. AV........ ....L..... ..F...E... .......... R..E......  236
ovPAG2     .S..SLAVP. T........Q Q......I.. ....TNKEN. ..L.L..... S.H..K...I  237
ovPAG8     .S..NT.CF. A......... .......... ....N----- ---------- ----------  212

Consensus  PL..AGDW.I .VDRISM.R. VIACS.GC.A LVDTG.S.I. GP.RLVDNIQ KLIGA.P.G.  300 (240)

ovPAG4     ..VK.D..T. Q......R.E .....D..D. .L...A.F.H ..G..I.D.. ....SE.RDL  297
ovPAG7     ..VK.D..S. H......R.E .....D..D. .L...A.F.H ..G..I.D.. ....SEQRDF  297
ovPAG3     ..IK....SV R..S.T.K.E .....D..R. .....S.H.Q ..G..I..V. ....TM.Q.S  297
ovPAG6     ..IP..N.MV HM...YIE.N .....A..K. V....AAF.E ..KSQ...M. .F...R.R.S  297
ovPAG1     ..IHP.E.S. PL.....R.K .....G..E. ..G..T.L.L ..RTV.E... .H...TQQCF  299
ovPAG5     ..IRV...R. R..H...KGK L.G..G..E. .....P.L.N ..T...T... R....M.L.P  296
ovPAG9     ..TK....IV RL....IG.K .....GD.E. V....TAF.G ..RK..KK.. .....RRR.N  296
ovPAG2     .VSQTKS.L. T......NGK ..G.EH..E. .....T.L.H ..AGP.T... .F.H.V.YDS  297
ovPAG8     ---------. TMK.-----E .....D..R. .....S.H.Q ..G.....V. .H..TM.Q.S  258

Consensus  KHYVSCSAVN TLPSIIFTIN GINYPVPAQA YILK.S.G.C YT.F...RVR .STE.WVLGD  360 (300)

ovPAG4     ...I...... .......... .......... ....G.T.H. ..A.RAK... T...S.....  357
ovPAG7     ...I...... .......... .......... ....G.T.H. ..A.RAK... T...S.....  357
ovPAG3     M...P..... .......... .......... ....GFT.H. ..A.RAK... T...S.....  357
ovPAG6     .Y..P..V.. .......R.. S......GR. ....NHR.R. ..T.KENQWS P...I.I...  357
ovPAG1     EYF......Y A....V.... .......P.. .LV.D.R.Q. .SP.QVN.AN P.A.N.I...  359
ovPAG5     .......TI. I......... ....TL.... ....T.Q.G. F.T.KGA..F KHR.T.....  356
ovPAG9     .--------- ---------- ---------- ----DDA.R. S.T.QEN... L...S.L..N  323
ovPAG2     EYM...PVIS I..PV..... S.D.S...E. ..Q.I.NSL. LSS.HGDN-- --.DK.I...  353
ovPAG8     M...P..... .......... .......GR. ....DFR.R. ..N.KKKIA. TF..Y....E  318

Consensus  VFLRLYFSVF DRGNDRIGLA PAV                                          383 (323)

ovPAG4     .......... .......... ..M                                          380
ovPAG7     .......... .......... ..M                                          380
ovPAG3     .......... ...H...... R.G                                          380
ovPAG6     .......... ...H...... R..                                          380
ovPAG1     ....R..... .......... R..                                          382
ovPAG5     I........Y ..EHN..A.. Q.R                                          379
ovPAG9     ....V.L... .......... ...                                          346
ovPAG2     .........Y ....N..... ...                                          376
ovPAG8     P......... .......... ...                                          341
```

FIG. 1. The aligned amino acid sequences of different ovPAGs. Each structure was inferred from the sequences of its cDNA. The likely signal sequence is underlined, and a known site of propeptide sequence cleavage (ISF ↓ RG/DS) for certain PAGs is shown (vertical arrow). Many additional sequences, some from cDNA not containing entire ORF and others differing less than 5% in nucleotide sequence from those shown, are known but not shown. Numbering at end of rows is by amino acid residue starting with Met-1. Numbers in parentheses show the equivalent residue of pepsin. Boxes indicate the conserved sequences around the catalytic aspartic acid residues (Asp-32 and Asp-215). GenBank accession numbers for ovPAG1–ovPAG9 are M73962, U30251, and U94789–U94795, respectively.

(74% nucleotide; 58% amino acid). OvPAG8 and ovPAG9 appear to be alternatively transcribed, short forms. Based on the structures of the boPAG1 gene (17) and the boPAG2 and poPAG2 genes (B.S., J.G., S.X., and R.M.R., unpublished results), whose exon organizations are identical, the ovPAG8 cDNA lacked the entire sequence of exon 6, whereas ovPAG9 lacked exon 8. Full-length transcripts have not yet been isolated for either of these ovine genes, nor is it known if these transcripts give rise to PAG protein.

The inferred (26) signal sequences for each of the ovPAG are 15 aa and terminate in a completely conserved cysteine residue. Precisely where propeptide cleavage occurs is unclear for the majority of the PAGs shown in Fig. 1. One motif that is used in ovPAG9 (15), boPAG1 (10, 12), and most likely ovPAG1 (12) is the sequence ISF ↓ RG/DSN, with cleavage occurring between the Phe and Arg residues. Six of the PAGs in Fig. 1 contain either this or a closely similar motif. Nevertheless, numbering of amino acids poses difficulties because of the variability in position 1 of the proteins that have been purified (14, 15). As a consequence, pepsin, which has been used as the basis for construction of three-dimensional models of ovPAG1 and boPAG1, provides a convenient basis for comparative amino acid numbering (13).

A comparison of the residues surrounding the catalytic aspartic acid residues (32 and 215; pepsin numbering), which are boxed, shows that the sequences are relatively well conserved. The substitution of Gly for Asp-215 in ovPAG1 is likely to be inactivating (12, 27), as is the replacement of Thr-35 by Arg in ovPAG8-related molecules (28). If any of the other substitutions would impede either catalysis or access of substrate to the catalytic center is unclear. Modeling studies (13) have indicated that a number of PAGs with perfectly conserved sequences around Asp-32 and Asp-215 may be active because several other subsites in the binding cleft are affected by unfavorable amino acid substitutions.

The number of sites for potential N-glycosylation ranges from two in ovPAG2 to seven in ovPAG8 (Fig. 2) and may account for some of the variability in apparent mass of the PAG family (12, 29). The three Asn likely to be glycosylated in ovPAG1 (Asn-1, Asn-17, Asn-68) are all surface-located.

Cloning of boPAG cDNA from Bovine Trophoblast. BoPAG1 (12), boPAG2 (16), and boPAG3 (previously boPAG variant) (17) were each cloned from a day 260 bovine placental cDNA library by immunoscreening. Here, two procedures were used for obtaining additional cDNA. A day 25 bovine cDNA library, prepared from trophoblast (pre-placenta) mRNA, was screened with a mixture of PAG probes. Of 40,000 plaques screened, 3.4% hybridized strongly to one or more of the probes in the mixture. In the second procedure, RNA from a single day 19 conceptus was subjected to RT-PCR with PAG-specific oligonucleotides, as described above for isolating ovPAG cDNA. After subcloning randomly chosen positive clones, inserts of the anticipated ≈1,300-bp length were partially sequenced from their ends. Those that were judged novel were sequenced completely. A total of 15 clones from the day 25 library provided seven new PAG cDNA (boPAG4–boPAG11) (Fig. 3). Four of the remainder closely resembled the already characterized boPAG2 (16). The day 19 RNA provided one additional clone (boPAG12) as well as additional clones for boPAG2, boPAG8, boPAG9, and boPAG11.

BoPAG7 cDNA lacked exon 6, but, because the intron phases of introns 5 and 6 were unaltered, there was no shift in the reading frame of exon 7. Two forms of boPAG4 have been isolated. One, the short form, which is not shown in Fig. 4, lacked exon 7. In this case, the intron phase was disturbed and the ORFs of exons 8 and 9 destroyed. The significance, if any, of such alternatively spliced forms is unclear.

The two most related PAGs in Fig. 3 are boPAG1 and boPAG3, which share 91% nucleotide and 86% amino acid

N-Glycosylation Sites (Pepsin Numbering)	PAG																				
	ovPAG1	ovPAG2	ovPAG3	ovPAG4	ovPAG5	ovPAG6	ovPAG7	ovPAG8	ovPAG9	boPAG1	boPAG2	boPAG3	boPAG4	boPAG5	boPAG6	boPAG7	boPAG8	boPAG9	boPAG10	boPAG11	boPAG12
-28									*												
-21																			*		
-13																			*		
-8					*						*										
-1	*		*			*	*	*	*	*		*	*	*	*	*		*			
17	*	*	*	*	*	*	*	*	*	*	*	*			*	*		*		*	*
46			*	*			*	*					*		*	*					
60					*						*								*		*
68	*		*	*		*	*	*		*		*			*	*		*			
74																			*		
106																*					
112								*													
127			*		*			*	*	*		*		*	*		*				
155								*								*					
159																	*		*		
161		*																			
196											*										
200											*										
202																					*
292											*										
317																					*

FIG. 2. A tabular representation of the potential sites for glycosylation of the ovPAG and boPAG molecules shown in Figs. 1 and 3. Numbering is according to the sequence of pepsin.

```
                                                                          ↓ (1)
Consensus  MKWLVLLGLV AFSECIVKIP LR..KTMR.T LS.KN.LNNF LKE.AYRLSQ IS.RGSNLT-   60  (2)

boPAG 8    .......... .L........ .TKM...QEA IRE.QL.ED. .D.QPHS... H.DPDKKFS.   59
boPAG 2    .......... .L.....IL. .KKM..L.E. .RE..L.... .E.Q.....K N---D.KI..   56
boPAG12    .......... .L.....IL. ..KM..L.E. .RE..L.... .E.R.....K K---D.KI..   56
boPAG11    .....F.... .......IML .TKT....EI WRE.KL..S. .E.Q.N.M.D D.ASDPK.S.   59
boPAG 9    ...I...... .......... ..QV....K. ..G..M.K.. ...HP..... ..F.......   59
boPAG10    ....G..... .L...M.I.. ..QM....E. .RERHL.T.. SE.HP.N... KAANDQ.IIY   60
boPAG 4    .......... .......... ..RV...TK. ..G..M.... V..H...... ..F.......   59
boPAG 6    .......... .......... ..RV....NA I.G..T...I ...H....P. ..F.......   59
boPAG 1    .......... .......... ..RL....NV V.G..M.... ...H..S... ..F.......   59
boPAG 3    .DD....... .......... ..RV....N. V.G..I...I ...HV..... ..F.......   59
boPAG 5    .........L TS.....IL. .TKV....K. ..E..M.... ...Q...... ..S....I..   59
boPAG 7    .......... .......... ..RV....K. ..G..M.... ...DP....H ..F.......   59

Consensus  .HPLRNI.D. .YVGNITIGT PPQEFQVVFD TGSSDLWVPS I.C.S.AC.T H.RFRH..SS  120  (62)

boPAG 8    S.Q.K.FQNA V.F.T..... .......N.. .......... VD.Q.PS.SK .K..DPQK.T  119
boPAG 2    I.....YL.T A......... .....R.... ...AN....C .T.T.P..Y. .KT.NPQN..  116
boPAG12    I...K.YL.M A......... .....R.... ...A...... .S.V.P..Y. .KT.NLHN..  116
boPAG11    T.....AL.M A......... ..K..R.... .......... .K.I.P..H. .IT.D.HK..  119
boPAG 9    I......MNL V......... .......... .......... FCTM-P..SA PVW..QLQ..  118
boPAG10    H....SYK.F S.I...N... .......L.. ....S..... .Y.Q.SS.YK .NS.VPCN..  120
boPAG 4    I......R.F F......... .......I.. ....E..... .F.N.ST.SK .D....LE..  119
boPAG 6    -......R.L F......... .......I.. ........A. .F.N.SS.AA .V....HQ..  118
boPAG 1    T......K.L V.M....... .......... .A........ DF.T.P..S. .V....LQ..  119
boPAG 3    T......K.L I......... .......... .....F.... DF.T.R..S. .V....LQ..  119
boPAG 5    I......M.M V...K..... .......... ....E..... VF.P.S..S. .I....LE..  119
boPAG 7    I......R.I F......... .......I.. .......... .D.N.TS.A. .V....LQ..  119

Consensus  TFR.TNKTF. I.YGSG.MKG .L..DTVRIG .LVSTDQPFG LS.EE-.GF. ...-.DGVLG  180 (120)

boPAG 8    ..QPL.QKIE LV....T... V.GS..IQ.. N..IVN.I.. ..QNQSS.VL EQVPY..I..  179
boPAG 2    S..EVGSPIT .F....IIQ. F.GS...... N...PE.S.. ..L...Y..D SLP.F..I..  174
boPAG12    S.GQ.HQPIS .S..P.IIQ. F.GS...... N...LK.S.. ..Q...Y..D GAP.F.....  174
boPAG11    ...L.RRP.H .L....M.N. V.AY...... K......... ..LQQ.F..D NAP.F.....  177
boPAG 9    ..QP.....T .T....S... F.AY...... D......... ..VV..Y.LE GRN.Y.....  176
boPAG10    ..KA...I.N TN.TATSI.. Y.VY...... N...VA.... ..LK..F..D DVP.F..I..  178
boPAG 4    ...LSRR..S .T....RIEA LVVH...... D......Q.. .CL...S..E GMR.F.....  177
boPAG 6    ...P.....R .T....R... VVVH...... D......... .CLKD.S..K GIP.F..I..  176
boPAG 1    ...L.....R .T....R... VVVH...... N......... ..I...Y..E GRI.Y.....  177
boPAG 3    ...L.....R .T....R... VVAH...... D......... ..V...Y..E GRAYY.....  178
boPAG 5    .SGL.Q...S .T....ST.. F.AY...... D.L....E.. ..M...H..E DLPF-.....  177
boPAG 7    ...P.....R .I....R.N. VIAY...... D......... ..V...Y..A HKRF-..I..  177

Consensus  L.YP..S..G AIPIFDNLKN QGAISEPVFA FYLS....EG SVVMFGGVDH .YYKGELNW.  240 (180)

boPAG 8    .A..SLAIQ. TT.V...... REV....... ....SRPENI .T........ T.H..K.Q.I  239
boPAG 2    .AF.AMGIED T.......WS H..F...... ...NTNKP.. .......... R........I  234
boPAG12    .A..SI.IK. I.......WS ...F...... ...NTCQP.. .......... R........I  234
boPAG11    .S..SLAVP. T.....K..Q .......I.. ....TRKEN. ..L.L..... S.H..K...I  237
boPAG 9    .N..NI.FS. .......... .......... ....KNKQ.. .......... Q........I  236
boPAG10    .G..RRTIT. .N......WK ..V....... ....SQKEN. ........NR A........V  238
boPAG 4    .S.TNI.PS. .....YK... E......... ....KDER.. .......A.. R........I  237
boPAG 6    .S..NKTFS. .F....K... E......... ....KDKQ.. .......... R........V  236
boPAG 1    .N..NI.FS. ......K... .R........ ....KDER.. .......... R..E.....V  237
boPAG 3    .N..NI.FS. .......... .......... IL..KDEQ.. .......... R..E.....V  238
boPAG 5    .N..DM.FIT T......... ...F...... ...GKVK--. .......... T........V  235
boPAG 7    .N.WNL.WSK .M....K... E......... ....NI---- ---------- ----------  213

Consensus  P..Q.G.W.. .MDRISM.R. VIACS.GC.A LVDTGTS.I. GP.RLV.NI. KLI.A.P-.G  300 (239)

boPAG 8    .VT.ARF.QV A.SSMT.NGN .VG..Q..Q. V......LLV ..TH..TD.L ...NPN.ILN  299
boPAG 2    .VS.TSH.QI S.NN...NGT .T...C..E. .L.....M.Y ..TK..T..H ..MN.RL.EN  293
boPAG12    .VS.TRY.QI S.N....NGN .T...R..Q. .L.....M.H ..T..IT..H ..MN.RH.Q.  293
boPAG11    .VS.TKS.LI TV.....NGR ..G.EH..E. .......L.H ..A.P.T..Q .F.H.M..Y.  296
boPAG 9    .LIEA.E.RV H......K.T .....D..E. ..H....H.E ..G...N..H R..RTR..FD  295
boPAG10    .VS.V.S.HI NI.S...NGT .V..KR..Q. SWIR.-RLSA W.K.I.SK.Q ...H.R..ID  296
boPAG 4    .LMKA.D.SV H......K.K .....G..K. .....S.D.V ..ST..N..W ...G.T..Q.  296
boPAG 6    .LI.V.D.FV H...TT.K.K .....D..K. .......D.V ..ST..N..W ...R.R..L.  295
boPAG 1    .LI.A.D.SV H.....IE.K I....D..K. .......D.V ..R...N..H R..G.I..R.  296
boPAG 3    .LIEA.D.II H......K.K I....GS.E. I......A.E ..RK..NK.H ...G.R..RH  297
boPAG 5    .LI.A.E.SL H......K.K .....G..E. FY.....L.L ..R...N..Q ...G.T..Q.  294
boPAG 7    ---------- -----T.N.E .....E..A. .....S.N.Q ..G..ID..Q RI.G.T..R.  257

Consensus  SE..VSC.AV ..LPSIIFTI NGI.YPVP.. AYI.KDSRG. C...F..... ---.-S..ET  360 (295)

boPAG 8    D.QML..D.I NS..TLLL.. ...V....PD Y..QRF.ERI .FIS.QGGTE ILKNLGTS..  359
boPAG 2    ..YV...D.. KT..PV..N. ...D..LRPQ ...I.IQN-S .RSV.QGGTE ...N..SLN.  348
boPAG12    ..YV...D.. KT..PV..N. ...D..L.PQ ...T.AQN-F .LSI.HGGTE ...T..SP..  348
boPAG11    ..YM.L.PVI SI..PV.... ...D.S..RE ...Q.I.NSL .LST.HGD-D ...T.---DQ  348
boPAG 9    .KHY...F.T KY....T.I. ...K..MTAR ...F.....R .YSA.KENTV ...R.TSR..  351
boPAG10    R.HV...Q.I GT..PAV... ...D....AQ ...QSL.G-Y .FSN.LVRPQ ...RVNES..  352
boPAG 4    ..HY...S.. NS........ KSNN.R..GQ ...L.....R .FTA.KGHQQ ...S..ST.M  352
boPAG 6    PQYF...S.. NT........ ...N.RL.AR ...H.....R .YTA.KEHRF ...S..PI..  351
boPAG 1    ..HY.P.SE. NT....V... ...N....GR ...L..D..R .YTT.QENRV ...S..ST..  352
boPAG 3    .KYYI..S.. NT........ ...N..C.GR ..VL.....R .YSM.QENKV ...S..ST..  353
boPAG 5    ..HYI..F.. IS........ ...NI...AR ...H.....H .YPT.KENTV ...ST.-T..  350
boPAG 7    .KYY...S.. NI........ ..VN....PR ...L.....H .YTT.KEKRV ...RR.-T.S  313

Consensus  WILGDVFLRL YFSVFDRGND RIGLA.AV                                    388 (323)

boPAG 8    .......... ....Y....N .....P.A                                    387
boPAG 2    .....I...Q ........K.R .....P..                                   376
boPAG12    ....G....Q ........R.. S....-Q.                                   375
boPAG11    .......... ....Y....N .....P..                                    376
boPAG 9    .....A...R .......... .....R..                                    379
boPAG10    .......... .........N .....P..                                    380
boPAG 4    .......... .......RK. .....TK.                                    380
boPAG 6    .L.......R .......... .....R..                                    379
boPAG 1    .Y........ .......... .....R..                                    380
boPAG 3    .........V .......... .....R..                                    381
boPAG 5    .......... .......... .....QV-                                    377
boPAG 7    .V..E..... .......... .....RR.                                    341
```

FIG. 3. The aligned amino acid sequences of different boPAGs. See legend to Fig. 1 for details. GenBank accession numbers for boPAG1–boPAG12 are M73961, L06151, L06153, and AF020506–AF020514, respectively.

sequence identities. The most distant are boPAG4 and boPAG8 (67% nucleotide, 49% amino acid). Again they resemble pepsin more than they do each other.

Of the 12 PAGs identified by cDNA cloning (Fig. 3), only 2 have had their amino termini sequenced (15). BoPAG1, purified by conventional column chromatography, has Arg-54 (Arg-3, pepsin number) as its first residue (10, 12). BoPAG6 was isolated from the medium after culturing day 100 explants of bovine placenta and purified by means of graded pH and salt elution from a pepstatin affinity column (J.G. and R.M.R., unpublished results), its amino terminus also begins at Arg-54. Therefore, the ISF↓RG/DSN sequence again provides a conserved site for propeptide cleavage.

The boPAG, like their ovine relatives, are most likely variably glycosylated (Fig. 2). The number of potentially substituted Asn range from one in boPAG5 to six in boPAG2 and boPAG7. BoPAG10 has five such sites, but two of them may be in the propeptide (Asn-21 and Asn-13, pepsin numbering). Of the four Asn shown for boPAG1 N − 1 and N17 are almost certainly substituted, as they give no signals during amino acid sequencing (10, 12).

FIG. 4. Modeling of ovPAG1. (*A*) A ribbon diagram of ovPAG1 showing regions of amino acid hypervariability (red) that do not carry sites for N-glycosylation. All these regions are surface loops. Amino acids are numbered in yellow (ovPAG1 numbering). (*B*) A segment with inwardly projecting side chains colored in green (288–290) links two hypervariable loop domains in red (281–287 and 291–296). The flap (pink) on the amino-terminal lobe and the hypervariable region 242–247 are also visible. This view of the two loops and the intervening conserved segment shows the orientation of amino acid side chains, which face outward in the loops and inward in the intervening segment. The diagrams were prepared by Ben Dunn (University of Florida) based on coordinates of Guruprasad *et al.* (13).

Hypervariable Regions. A comparison of PAG sequences both within a single species (Figs. 1 and 3) and between species (13) reveals segments of primary structure that are hypervariable and others that are relatively constant. The signal sequences and the carboxyl termini, for example, are highly conserved even when the PAG are aligned with pepsinogen (ref. 13 and data not shown). There are also other constant regions. The sequence between residues 71 and 100 (13–42, pepsin numbering) shows very few substitutions and includes the first catalytic domain. In general, the conserved regions are ones that are internal and probably structurally important for retaining the overall three-dimensional fold of the molecule (13, 28). By contrast, the hypervariable regions are exposed and generally correspond with surface loops (13).

Many of the hypervariable surface regions carry potential sites for N-glycosylation. For example, the regions containing N46, N68, and N127 (pepsin numbering) are among the most variable in the entire series of PAG shown in Figs. 1 and 3. Not all loop regions contain such sites, although other post translational modifications such as O-glycosylation and phosphorylation (29) cannot be ruled out. Five such regions are illustrated in red in Fig. 4*A* for ovPAG1. The 281–296 region, adjoining the conserved carboxyl terminus and magnified in Fig. 4*B*, contains two exposed loops separated by three amino acids (in green) with inward-facing side chains. Region 242–

247 is preceded by a relatively variable helical domain and projects out prominently from the surface. In the case of regions 189–198 and 111–119, both loops are flanked by conserved β-strands (Fig. 4*A*). These data indicate that the hypervariable regions represent surface domains where amino acid substitutions could occur with little threat to the structural integrity of the molecules but which might provide a diversified repertoire of PAGs with different properties.

Southern Genome Blotting of Bovine DNA. The structure of one PAG gene (boPAG1) has been published (17). Two other genes, boPAG2 and poPAG2 (J.G., B.S., and R.M.R., unpublished data) have also been structurally characterized. All three genes so far cloned contained nine exons and eight introns organized in a manner similar to that of other mammalian aspartic proteinase genes (17).

Southern genomic blots of bovine DNA were performed with probes corresponding to a segment of the boPAG1 and boPAG2 genes encompassing part of intron 6, exon 7, intron 7, and exon 8 the proximal end and the proximal end of intron 8. Restriction enzymes were chosen that did not cleave either probe. Conditions of hybridization were such that the PAG1 probe did not bind the PAG2 gene and *vice versa*; nor would there be hybridization to genes for other known aspartic proteinases. Despite this planned selectivity, it is clear from Fig. 5 that many of the DNA fragments bound both probes, either because boPAG1- and boPAG2-related genes were contained in the fragments or because at least one gene recognized by both probes was present. Some of the faster migrating bands hybridized to only one of the probes. The data confirmed that the bovine genome contains many different PAG genes, but indicated also that Southern genomic blotting is too imprecise to provide even an estimate of total gene number.

Southern Genomic Blotting of DNA from Other Mammals. Multiple PAG genes were detectable in all species of the Bovidae family examined (Fig. 6). Signals were especially strong in the species closely related to *Bos taurus* within the subfamilies Bovinae (e.g., *Bos frontalis gaurus*, gaur; *Bos grunniens*, yak; *Syncerus caffer*, Cape buffalo) and Caprinae (e.g., *Ovis aires*, domestic sheep; *Ovis dalli*; Dall sheep; *Capra falconeri*, Markhor goat; *Nemorhaedus goral*, goral; *Budorcas taxicolor*, takin). Gazelle and antelope species in other related subfamilies, including the impala, gnu, duiker, and nyala, also gave strong signals.

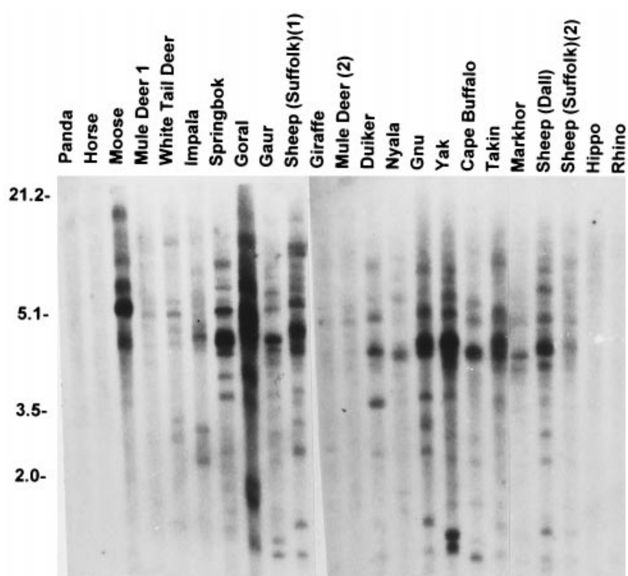

FIG. 6. Southern genomic blotting of DNA from some selected ruminant and nonruminant ungulate species and from a member of the Carnivora (Panda). DNA was digested with *Eco*RI and probed with a boPAG1 probe. DNA size markers are on the left. Some samples of DNA (e.g., Suffolk sheep and mule deer) were analyzed twice.

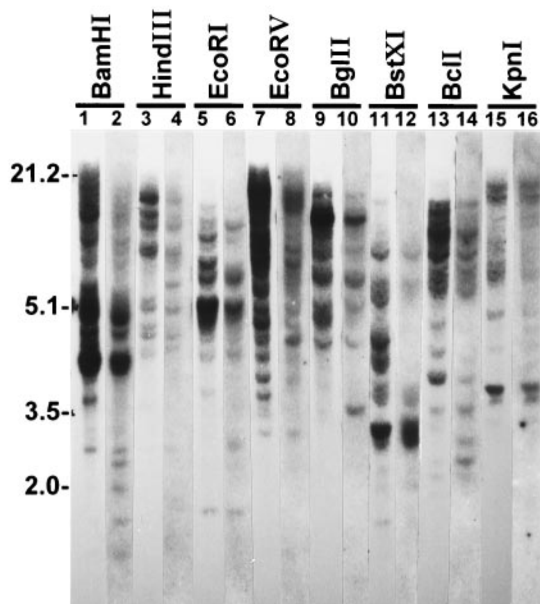

FIG. 5. Southern genomic blotting of a bovine (*Bos taurus*) DNA probed with boPAG probes. The restriction enzymes used are shown above the individual lanes. A single blot was hybridized first with a specific boPAG1 probe designed not to hybridize to boPAG2 DNA; they were subsequently hybridized with a specific boPAG2 probe. Each autoradiograph was photographed and the individual lanes cut from the photographic print and aligned to show side-by-side comparisons. Odd-numbered lanes were probed for boPAG1-related genes, and even-numbered lanes for boPAG2-related genes. Size markers are on the left.

In general hybridization, although detectable, was weaker to DNA of members of the Cervidae family, including the whitetail deer and mule deer, than to DNA from Bovidae. Unexpectedly, moose (*Alces alces*) gave a relatively strong signal. The giraffe (family Giraffidae) provided the weakest signal of the true pecoran ruminants, possibly reflecting its early divergence (31). Hybridization to DNA from the Nile hippopotamus was barely detectable with the boPAG1 probe employed. Because the hippo (family Hippotamidae; suborder Suiformes) is related to the domestic pig (*Sus scrofa*), a species with multiple PAGs (18), this result indicates the considerable divergence of the genes within the Artiodactyla order over the 55–65 million years of its existence.

DNA of more distantly related hoofed species, including the horse and rhinoceros, which are members of the Perissodactyla order, and of all nonhoofed species examined, including panda (shown in Fig. 7), mountain lion, tiger, and blackbear (data not shown) all of which are within the Carnivora order, gave no cross-hybridization. It is unclear whether these groups of mammals that failed to provide hybridization signals lack genes homologous to PAG or if evolutionary distance of the genes is so great that the probes were incapable of binding the DNA. Horse DNA, for example, failed to bind the probe, despite the fact that a PAG-related transcript expressed in the outer epithelial layer of the horse placenta has been cloned (13, 19). The nucleotide sequence of this transcript is sufficiently different (less than 65% identity) that its gene would not be detectable with the boPAG1 cDNA probe under the conditions of hybridization selected.

Estimation of PAG Gene Number from Screening a Genomic Library. As reported earlier (17), 120×10^6 phage plaques from a bovine GEM-11 genomic library were probed with a boPAG1 cDNA under conditions that eliminated binding to boPAG2 DNA. Thirty-four (0.028%) strongly pos-

```
Consensus   CACCYCYTSA GGAATTCCAG GTTVTCTTYG ACACAGSYTC AKCTKWHTTG TKGGTGSCST    60

OPGB17      ....C.C.C. .......... ...G....T. ......GC.. .T..GAC... .T....C.C.    60
OPGB14      ....C.C.C. .......... ...G....T. ......GC.. .T..TTT... .G....C.C.    60
OPGB25      ....C.C.C. .......... ...C....T. ......GC.. .T..TTT... .G....C.C.    60
OPGB26      ....C.C.G. .......... ...G....T. ......GC.. .T..GAT... .G....C.C.    60
OPGB21      ....C.C.C. .......... ...A....T. ......GC.. .T..GAC... .G....C.C.    60
OPGB22      ....C.C.C. .......... ...A....T. ......GC.. .T..GAC... .G....C.C.    60
OPGB 6      ....C.C.C. .......... ...G....T. ......GC.. .G..GTT... .G....C.C.    60
OPGB16      ....T.C.C. .......... ...G....T. ......CT.. .T..GAC... .G....C.C.    60
OPGB 1      ....T.T.C. .......... ...G....T. ......GC.. .T..GAA... .G....C.C.    60
OPGB23      ....C.C.C. .......... ...G....C. ......GC.. .T..GAC... .G....C.C.    60
OPGB24      ....C.C.C. .......... ...G....T. ......GC.. .T..GAA... .G....G.G.    60

Consensus   CYRWCWWTTG HVHBAKYBMA RCCAAGAGAC YCTGKDGTSA GTABRAGAHA MMYYVRTACC   120

OPGB17      .CAT.AA... CCTC.GCCC. A......... C...TA..G. ...G-...C. CCCT-G....   118
OPGB14      .TGT.TT... CAAG.GCTC. A--------- C...TT..G. ...TG-..C. CCCCC-....   109
OPGB25      .TGT.TT... CAAG.GCTC. A--------- C...TT..G. ...TG-..C. CCCCC-....   109
OPGB26      .TGA.TT... CACC.GTCC. G--------- C...TT..G. ...GA-..A. CCCTC-....   109
OPGB21      .TGT.TT... AAAC.GCTC. A--------- C...TT..C. ...CA-..-. CCCCC-....   108
OPGB22      .CGA.TT... CACC.GTCC. G--------- C...TT..G. ...CA-..C. ACCCA-....   109
OPGB 6      .CGT.TT... CAAC.GCTC. A--------- C...TT..G. ...CA-..C. CCCCA-....   109
OPGB16      .CAT.TT... CAAC.GCTC. A--------- C...TT..G. ...CA-..C. CATCC-....   109
OPGB 1      .CAT.TT... CAAC.GCTC. G--------- C...GT..G. ...CA...C. ACCCAG....   111
OPGB23      .CAA.TT... TGCC.TTGA. G--------- C...TT..G. ...CA-..T. CCTT-G....   109
OPGB24      .CGA.TT... CACT.GTCA. G--------- T...TG..G. ...C-...-- -CCCGA....   107

Consensus   BRRABBRTCY TTCACYKRCC YTGYYRCYYT RTTYHSRHCC KTGRCAMMTG RTGACAYYCA   180

OPGB17      CAG.CTA..C .....TGG.. T..CCA.CC. G..TACGA.. T..G..CC.. A.....CT..   178
OPGB14      GGG.CCA..T .....CTG.. C..CCA.CC. G..CCCAC.. T..G..CC.. A.....CT..   169
OPGB25      GGG.CCA..T .....CTG.. C..CCA.CC. G..CCCAC.. T..G..CC.. A.....CT..   169
OPGB26      TAG.CCG..C .....TTG.. C..CCA.CC. A..TCCAC.. G..G..CA.. A.....TT..   169
OPGB21      CAG.GCA..C .....TTG.. C..CTG.CC. G..TCGAC.. T..G..AC.. G.....CT..   168
OPGB22      CAG.GCA..C .....TTG.. C..CTG.CC. G..TCGAC.. T..G..AC.. G.....CT..   169
OPGB 6      CGA.TGA..T .....TTG.. C..CCA.CC. G..TTCAC.. T..G..CC.. A.....CT..   169
OPGB16      CGG.CCA..T .....TTA.. C..CCA.CC. G..TCCAC.. T..G..CC.. A.....CC..   169
OPGB 1      CAG.CCA..T .....TTG.. C..CCA.CT. A..TCCAC.. T..A..CC.. A.....CT..   171
OPGB23      CGG.CCA..C .....TTG.. C..CCA.TC. G..TCCAT.. T..G..CC.. A.....CT..   169
OPGB24      CGG.CCA..C .....TTG.. C..TCA.CC. G..TTCAT.. T..G..CC.. A.....CT..   167

Consensus   YCWCTTKTDT CTVCMRSYAM ACRNRHTARG TTYARAYAHC DYSWKTCKTC YAYCYTCYGG   240

OPGB17      T.T...G.T. ..G.AGGT.A ..AAGA..A. ..C.A.C.T. ACCAG..T.. C.C.T..C..   238
OPGB14      T.T...G.G. ..A.AGCT.A ..GCAT..G. ..C.G.C.T. TTCAG..T.. C.C.C..C..   229
OPGB25      T.T...G.G. ..A.AGCT.A ..GCAT..G. ..C.G.C.T. TTCAG..T.. C.C.C..C..   229
OPGB26      T.T...G.G. ..A.AGCT.C ..GTGT..G. ..C.G.C.A. ATCAG..T.. C.C.T..C..   229
OPGB21      T.T...G.G. ..G.CACC.C ..AGGT..G. ..C.A.C.T. TTCAT..G.. C.C.T..T..   228
OPGB22      T.T...G.G. ..G.AGCT.C ..AGGC..G. ..C.G.T.T. GTCAG..T.. C.C.T..C..   229
OPGB 6      T.T...G.G. ..G.AGCT.C ..ACGT..G. ..C.G.C.C. TTGAG..T.. C.T.T..C..   229
OPGB16      T.T...G.G. ..G.AGCT.C ..GTGT..G. ..C.G.C.T. GCCAG..T.. C.C.T..C..   229
OPGB 1      C.T...T.A. ..G.AGCT.C ..ATGT..G. ..T.G.C.T. TTCAG..T.. C.C.T..T..   231
OPGB23      T.T...G.G. ..G.AGCT.C ..AGGC..G. ..C.G.T.T. GTCAG..T.. C.C.T..C..   229
OPGB24      T.A...G.G. ..C.AGCT.C ..ACAA..G. ..C.G.T.T. GTCTG..T.. T.C.T..C..   227

Consensus   HYTRYMVRWR RVAYMTKCDG NWTMDHSTWY RRWKYYRGRR BMMTKRRARG ADG          293

OPGB17      TT.ACCAATG AC.CC.T.A. AA.CTAC.TT GGTTCTG.AA CAA.GAG.A. .T-          290
OPGB14      CT.ACAGATA AG.TC.G.A. GA.AACC.AC GGATCCA.AA GAA.TAA.G. .A-          281
OPGB25      CT.ACAGATA AG.TC.G.A. GA.AACC.AC GGATCCA.GA GAA.TAA.G. .A-          281
OPGB26      TC.ACCCGAA AG.CC.T.A. CA.CAAC.AT GGATCTG.GA CAA.GAG.-. .A.          281
OPGB21      CT.ATCAATA GA.CC.T.T. TA.CACC.AT GGATCTG.GG GCC.GAA.G. .G-          280
OPGB22      CT.ACCAATA GA.CC.T.T. TA.CACC.AT GGATCCG.GA GAA.TAA.G. .G-          281
OPGB 6      CC.ACCAATA AG.CA.T.T. GA.CAAC.AT GGAGCTG.GA GAA.TGA.G. .G-          281
OPGB16      CT.GCCAATA AG.CA.T.G. GA.CATG.AT GGAGTTG.GA TAA.TGA.G. .A-          281
OPGB 1      CC.ACCAATA AG.CC.T.A. CA.CACC.AT GGATCCG.GA GAA.TAA.G. .G-          283
OPGB23      CT.ACCAATA GA.CC.T.T. TA.CACC.AT AGATCTG.GG GCC.GAA.G. .G-          281
OPGB24      AT.ACCAATA AG.CC.T.A. CT.CGCC.AC GAATCTG.GA GAA.GAA.G. .G-          279
```

FIG. 7. Alignment of the nucleotide sequences of cloned ovPAG gene fragments. Sheep genomic DNA was amplified with a pair of primers (data not shown) that were identical to ovPAG1 DNA sequences in exon 3 and exon 4. Amplified gene fragments were cloned, sequenced, and aligned. The boxed areas show intron 3 flanked by exons 3 and 4. Gaps (-) were introduced to give optimal alignments. Completely conserved nucleotides are shown as dots below the consensus sequence (top).

itive plaques were identified, many of which contained portions of two different genes. If it is assumed that the bovine genome consists of $\approx 3 \times 10^9$ nucleotide pairs and that the library is representative, around 840,000 bp are likely to encode PAG1-related genes. If each PAG gene is between 10 and 20 kb in length, there may be between 80 and 40 such genes. A similar screening with a PAG2 probe under stringent conditions provided an estimated number of 30–60 PAG2-related genes.

Cloning of ovPAG Genomic Fragments by PCR. To confirm the presence of multiple related genes in the sheep, primers with specificities directed toward the ovPAG1 gene but that fail to hybridize to an ovPAG2 gene were used to amplify a selected sequence segment encompassing part of exon 3 and exon 4 by standard PCR. The resulting DNA, which included all of intron 3, was then subcloned. Colonies were randomly selected and, provided they possessed inserts of approximately 320bp, their amplicons were sequenced. Out of 29 individually cloned PAG sequences, 14 distinct genes that had identities ranging from about 80–90% with each other were identified. Eleven of these were fully sequenced in both directions (Fig. 7). Exon 3 was less variable than either intron 3 or exon 4 and encodes the first catalytic domain. Intron 3 was well conserved in length (96 to 100 bp) and somewhat less variable than exon 4 which followed it. Of the genes so identified by PCR, only one (*oPGB17*) corresponded with any of the cDNAs shown in Fig. 1 (ovPAG1). The nucleotide sequences at the 5′ donor and 3′ acceptor sites of all clones conformed to the GT-AG rule, with the exception of *oPGB21* which had GT-CA. Sampling of randomly selected clones in this manner should, in theory, allow an assessment to be made of the total number of genes, as long as each gene that bound the oligonucleotides and could be amplified had an equal chance of being in the sample. Unfortunately the PCR procedure showed clear selectivity. Seven of the 29 clones, for example, were *oPGB22* and five were *oPGB14*. Nevertheless, as only one clone matched any of the cDNAs in Fig. 1, the estimate of 40–80 PAG1 genes in the bovine made by screening genomic libraries may not be exaggerated.

A similar procedure was carried out with primers designed to amplify ovPAG2-related sequences. Here 13 clones were picked, and five different genes were identified (data not shown). Nine of the clones represented one unique gene (*oPGA3*). Two of the five genes (*oPGA12*, *oPGA5*) corresponded to ovPAG2 and ovPAG5 (Fig. 1); the remainder were unique. Because the sample size was small and the selection process again apparently not random, it was not possible to estimate total gene number, but again the indications were that the number would be large.

Distance Relationships. Fig. 8 is an unrooted GROWTREE phylogram based on the degree of divergence of amino acid sequences among the PAGs that have so far been cloned and the commoner mammalian aspartic proteinases to which the

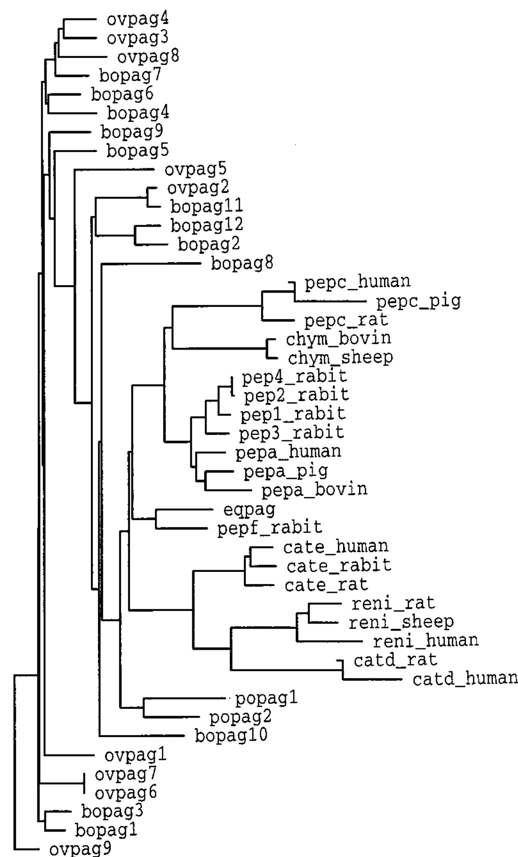

FIG. 8. A GROWTREE phylogram based on amino acid sequences showing the relationship of all known cloned PAGs to the commoner mammalian aspartic proteinases whose sequences are known. The tree was constructed by the Wisconsin GCG programs DISTANCES and GROWTREE. The lengths of the branches are proportional to the degree of amino acid diversity within pairs of proteins. Occasional negative branch lengths result because the neighbor-joining algorithm always attempts to create an additive tree, even when this is not possible. Analyses were carried out over the first 220 amino acids so as to include PAG sequences with missing exons. Protein data bank symbols are as follows: pep, pepsin; chym, chymosin; cate, cathepsin E, catd, cathepsin D; reni, renin; rabit, rabbit; bovin, bovine.

PAGs are structurally related. Because it is likely that there have been unequal rates of substitution for the different molecules in the data set (see below) and a constant molecular clock cannot be assumed, Fig. 8 probably does not provide an accurate phylogeny.

The active aspartic proteinases, which include the various pepsins, chymosins, cathepsin E and D, and renin, are clustered in the central branches of the tree. Included among them is equine PAG1 (eqPAG1), which is paired with rabbit pepsinogen F. EqPAG1 is an active proteinase after propeptide excision (19) and may therefore be the horse homolog of pepsin F. Unfortunately little is known about pepsinogen F; it has been cloned from the stomach of a neonatal rabbit (31), but its overall expression pattern in the fetus has not been studied. Nor has pepsinogen F been described in any other species, although, based on Fig. 8, it seems likely to be widespread. porcine PAG1 (poPAG1) and poPAG2 occupy an intermediate position between the enzymatically functional aspartic proteinases and the PAGs from cattle and sheep. Of the latter, boPAG8, boPAG10, and ovPAG5 are the three most distant and possibly most ancient gene products so far identified. Most closely related to them are ovPAG2 and boPAG2, boPAG11, and boPAG12. Together these seven PAGs are of particular interest because, as determined by *in situ* hybridization analysis their genes are expressed in both the mononucleated as well as the larger invasive binucleated cells of the outer trophectoderm layer of the placenta (S.X., B. Bao, and R.M.R., unpublished results).

The remaining PAG genes, ovPAG1, ovPAG3, ovPAG4, ovPAG6, ovPAG7, ovPAG8, and ovPAG9 and boPAG1, boPAG3, boPAG4, boPAG5, boPAG6, boPAG7, and boPAG9, which have diverged less than the grouping above, have strictly binucleate cell-specific expression. Because binucleate cells are a typical feature of the trophectoderm of the synepitheliochorial placentas of the pecoran ruminants (5), it is tempting to speculate that the PAG1-related genes diverged relatively recently.

If the entire PAG gene family arose by a series of relatively recent duplications during the diversification of the Artiodactyla, the expected lengths of the branches leading to the individual PAG might be anticipated to be relatively short. Instead many are long, far exceeding the distance between human, rabbit, and rat cathepsin E (Fig. 8) whose divergence encompasses more than 100 million years of evolutionary time. There seem to be two alternative explanations. One is that the recent origin theory is wrong and that duplication of PAGs occurred early in the diversification of mammals. The second is that the genes duplicated late but accumulated mutations at a high rate. Early diversification seems unlikely in view of the fact that large numbers of aspartic proteinase gene family members have not been described in either rodents or man despite considerable efforts to clone them (32). Our data for the horse (Perissodactyla) and cat (Carnivora) indicate only a limited number (and possibly only a single) expressed PAG gene in each species (J.G., S. Gan, and R.M.R., unpublished results). Therefore, we favor a late and rapid diversification of the PAG within the Artiodactyla. In this regard, the relatedness of ovPAG2 and boPAG11 (94% at the amino acid level) suggests they are functional homologs. These genes are the most closely related of all the PAGs shown in Fig. 8, despite a species separation of around 18 million years (30).

Synonymous vs. Nonsynonymous Nucleotide Substitutions in PAGs. An analysis (22, 23) of the nucleotide substitutions within the protein-coding regions of the PAG genes reveals that the ratio of synonymous (silent) mutations per synonymous site (Ks) to nonsynonymous (replacement) mutations per nonsynonymous site (Ka) in pairwise comparisons among all PAGs averages 1.18 ± 0.27 (mean ± SD). A closer examination indicates that within highly conserved regions the Ks to Ka ratio is high, whereas it is low in the hypervariable loop-encoding regions. For example, the Ks to Ka ratio averages 3.07 ± 1.08 for the highly conserved 29 codons encoding the buried carboxyl end of the molecules. By contrast, the value for the preceding 21 codons, which are hypervariable and encode the two loops (291–296 and 281–287) shown in Fig. 5*B*, is 0.53 ± 0.18. Thus, mutations that alter amino acids have accumulated faster than silent mutations.

Mutations that lead to amino acid changes are much more likely to be deleterious and therefore to be eliminated than synonymous changes. For this reason Ks/Ka ratios are generally greater than 2.0 (33). The PAGs appear exceptional in this respect, with the data suggesting that their high variability has occurred as the result of positive selection. Other related aspartic proteinases, such as ovine and bovine chymosins, enzymes whose coding regions are 95% identical in sequence (34, 35) despite 18 million years of separation (30), exhibit a Ks to Ka ratio of 2.47, a value more than twice as high as the average PAG pair. The only PAG pair that exhibits a comparable value to the chymosins is ovPAG2 and boPAG11 (ratio, 2.92) proteins whose relatedness has been commented upon earlier (Fig. 8) and which may be functional homologs. Equine PAG and rabbit pepsinogen F, both active enzymes, provide a value of 2.61. Conceivably these genes have also acquired a function that is less able to tolerate changes in the surface loop regions than PAGs in general.

The Evolution of Multigene Family and the Functions of PAGs. The evolution of multigene families has been the subject of several recent reviews (36–38). All agree that most duplicated genes are likely either to be quickly lost or accumulated as pseudogenes, as a result of "purifying" Darwinian selection, unless they acquire a novel function. By this argument it must be assumed that individual PAGs are not only functional molecules, but that each has a subtly different role. Hughes (37) has argued that weak bifunctionality must be acquired before gene duplication and that, once duplicated, genes become separated by a burst of amino acid replacements that allows a specific function to become fixed and enhanced. These mutations are likely to be acquired by a combination of nonsynonymous point mutations and by gene conversion events that can probably occur readily between closely linked, structurally similar genes (39). Genetic drift and natural selection will ensure the retention of those mutations that are not deleterious. At present it is not possible to estimate what kinds of mutational changes contributed most to PAG diversity.

Fryxell (38) has argued that the retention of a duplicated gene will, in general, require the presence of a pre-existing or similarly evolving family of complementary molecules with which the products of the duplicated genes can interact. Among the best known rapidly evolving gene families are Igs, T cell receptors and major histocompatibility complex antigens, the cytochrome p_{450} system, and the odorant receptors. In each of these cases, diversification is linked to a more exacting capacity to bind particular ligands. For the PAGs, it is tempting to speculate that their function relates to their peptide-binding capabilities, although a function involving some structural feature other than the cleft, such as the propeptide or carbohydrate, cannot be ruled out. Even though the regions around the two catalytic aspartyl residues are generally conserved in all aspartic proteinases (28, 40), substitutions elsewhere can markedly influence what peptides gain access to the catalytic center, clearly evident when the exceedingly narrow substrate specificity of renin is compared with that of pepsin A. The reorganization of the combining site of an antibody against a nitrophenyl phosphate hapten as it evolved from its germ-line precursor led to a 30,000-fold greater affinity for ligand and involved only a handful of amino acids, many of which were in a surface location and none of which made direct contact with the ligand (41). Small additive changes in the packing of loops provided a combining site able

to lock in the hapten with much greater efficiency. Similar events could presumably modify the peptide-binding cleft of PAGs and provide molecules with a considerable range of specificities.

The placenta is a site at which considerable genetic experimentation takes place, possibly because the relationship between the mother and the conceptus, rather than being strictly nurturing, has elements of conflict (1). It seems that the PAG gene family evolved in step with the superficial epitheliochorial and synepitheliochorial placentation that occurs in most ungulate species (2, 5). Although their function remains unknown, it is unlikely that the PAGs are simply curiosities in view of their long-term evolutionary survival and conspicuous presence at the interface between maternal and fetal tissues.

We thank Drs. Jordan Tang, Ben Dunn, and Tom Blundell for helpful discussions on PAG diversity; Dr. Ben Dunn and the Center for Structural Biology, University of Florida, for providing Fig. 5; and Drs. David Bullock, William Folk, Shirley Tilghman, and David Haig for providing comments on the manuscript. The assistance of Kelli Valdez and Livija Hakami in performing sequencing is gratefully acknowledged. This work was supported by National Institutes of Health Grant R37 HD29483 and U.S. Department of Argriculture Grant 9601842.

1. Haig, D. (1993) *Q. Rev. Biol.* **68,** 495–532.
2. Amoroso, E. C. (1952) in *Marshall's Physiology of Reproduction*, ed. Parkes, A. S. (Little Brown, Boston), Vol. 2, pp. 127–311.
3. Roberts, R. M., Xie, S. & Mathialagan, N. (1996) *Biol. Reprod.* **54,** 294–302.
4. Roberts, R. M., Liu, L. & Alexenko, A. (1997) *Prog. Nucleic Acid Res. Mol. Biol.* **56,** 287–326.
5. Wooding, F. B. P. (1992) *Placenta* **13,** 101–113.
6. Wooding, F. B. P. (1981) *J. Reprod. Fertil.* **62,** 15–19.
7. Zoli, A. P., Demez, P., Beckers, J. F., Reznik, M. & Beckers, A. (1992) *Biol. Reprod.* **46,** 623–629.
8. Butler, J. E., Hamilton, W. C., Sasser, R. G., Ruder, C. A., Hass, G. M. & Williams, R. J. (1982) *Biol. Reprod.* **26,** 925–933.
9. Sasser, R. G., Ruder, C. A., Ivani, K. A., Butler, J. E. & Hamilton, W. C. (1986) *Biol. Reprod.* **35,** 936–942.
10. Zoli, A. P., Beckers, J. F., Woutters-Ballman, P., Closset, J., Falmagne, P. & Ectors, F. (1991) *Biol. Reprod.* **45,** 1–10.
11. Zoli, A. P., Guilbault, L. A., Delabaut, P., Ortiz, W. B. & Beckers, J. F. (1992) *Biol. Reprod.* **46,** 83–92.
12. Xie, S., Low, B. G., Nagel, R. J., Kramer, K. K., Anthony, R. V., Zoli, A. P., Beckers, J. F. & Roberts, R. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10247–10251.
13. Guruprasad, K., Blundell, T. L., Xie, S., Green, J., Szafranska, B., Nagel, R. J., McDowell, K., Baker, C. B. & Roberts, R. M. (1996) *Protein Eng.* **9,** 849–856.
14. Atkinson, Y. H., Gogolin-Evans, K. J., Hounsell, E. F., Davies, M. J., Brandon, M. R. & Seamark, R. F. (1993) *J. Biol. Chem.* **268,** 26679–26685.
15. Xie, S., Green, J., Bao, B., Beckers, J. F., Valdez, K. E., Hakami, L. & Roberts, R. M. (1997) *Biol. Reprod.*, in press.
16. Xie, S., Low, B. G., Nagel, R. J., Beckers, J. F. & Roberts, R. M. (1994) *Biol. Reprod.* **51,** 1145–1153.
17. Xie, S., Green, J., Beckers, J. F. & Roberts, R. M. (1995) *Gene* **159,** 193–197.
18. Szafranska, B., Xie, S., Green, J. & Roberts, R. M. (1995) *Biol. Reprod.* **53,** 21–28.
19. Green, J., Xie, S. & Roberts, R. M. (1998) in *Handbook of Proteolytic Enzymes*, eds. Barrett, A. J., Woessner, F. & Rawlings, N. (Academic, London), in press.
20. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
21. Hecht, S. J., Stedman, K. E., Carlson, J. D. & DeMartini, J. C. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 3297–3303.
22. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York), pp 293–298.
23. Li, W. H. (1993) *J. Mol. Evol.* **36,** 96–99.
24. Kimura, M. (1980) *J. Mol. Evol.* **16,** 111–120.
25. Nagel, R. J., Xie, S. & Roberts, R. M. (1993) *Biol. Reprod.* **48** (Suppl. 1), abstr. 39.
26. VonHeijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
27. Lin, X., Wong, R. N. S. & Tang, J. (1989) *J. Biol. Chem.* **264,** 4482–4489.
28. Davis, D. R. (1990) *Annu. Rev. Biophys. Chem.* **19,** 189–215.
29. Xie, S., Nagel, R. F., Green, J., Beckers, J. F. & Roberts, R. M. (1996) *Biol. Reprod.* **54,** 122–129.
30. Miyamoto, M. M., Kraus, F., Laipis, P. J., Tanhauser, S. M. & Webb, S. D. (1993) in *Mammalian Phylogeny,* eds. Szalcy, F. S., Novacek, M. J. & McKenna, C. (Springer, New York), pp. 268–281.
31. Kageyama, T., Tanabe, K. & Koiwai, O. (1990) *J. Biol. Chem.* **265,** 17031–17038.
32. Birch, N. P. & Loh, Y. P. (1991) *Biochem. Biophys. Res. Commun.* **177,** 920–926.
33. Ohta, T. (1992) *Annu. Rev. Ecol. Syst.* **23,** 263–286.
34. Moir, D., Mao, J. I., Schumm, J. W., Vovis, G. F., Alford, B. L. & Taunton-Rigby, A. (1982) *Gene* **19,** 127–138.
35. Pungercar, J., Strukelj, B., Gubanasek, F., Turk, V. & Kregar, I. (1990) *Nucleic Acids Res.* **18,** 4602.
36. Ohta, T. (1995) *J. Mol. Evol.* **33,** 34–41.
37. Hughes, A. L. (1994) *Proc. R. Soc. London B* **256,** 119–124.
38. Fryxell, K. J. (1996) *Trends Genet.* **12,** 364–369.
39. Ohta, T. (1995) *J. Mol. Evol.* **41,** 115–119.
40. Takahashi, K., Tanji, M., Yakabe, E., Hirasawa, A., Athauda, S. B. & Kageyama, T. (1995) *Adv. Exp. Med. Biol.* **362,** 53–65.
41. Wedemayer, G. J., Patten, P. A., Wang, L. H., Schultz, P. G. & Stevens, R. C. (1997) *Science* **276,** 1665–1669.