Protein Engineering vol.9 no.10 pp 849–856. 1996

Comparative modelling and analysis of amino acid substitutions suggests that the family of pregnancy-associated glycoproteins includes both active and inactive aspartic proteinases

Kunchur Guruprasad[1], Tom L.Blundell[1,5], Sancai Xie[2], Jonathan Green[3], Bozena Szafranska[2], Robert J.Nagel[2], Karen McDowell[4], C.Ben Baker[4] and R.Michael Roberts[2,3]

[1]Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, University of London, London WCIE 7HX, UK, Departments of [2]Animal Sciences and [3]Biochemistry, University of Missouri, 158 Animal Sciences Research Center, Columbia, MO 65211 and [4]Department of Veterinary Science, University of Kentucky, Lexington, KY 40546, USA

[5]To whom correspondence should be addressed

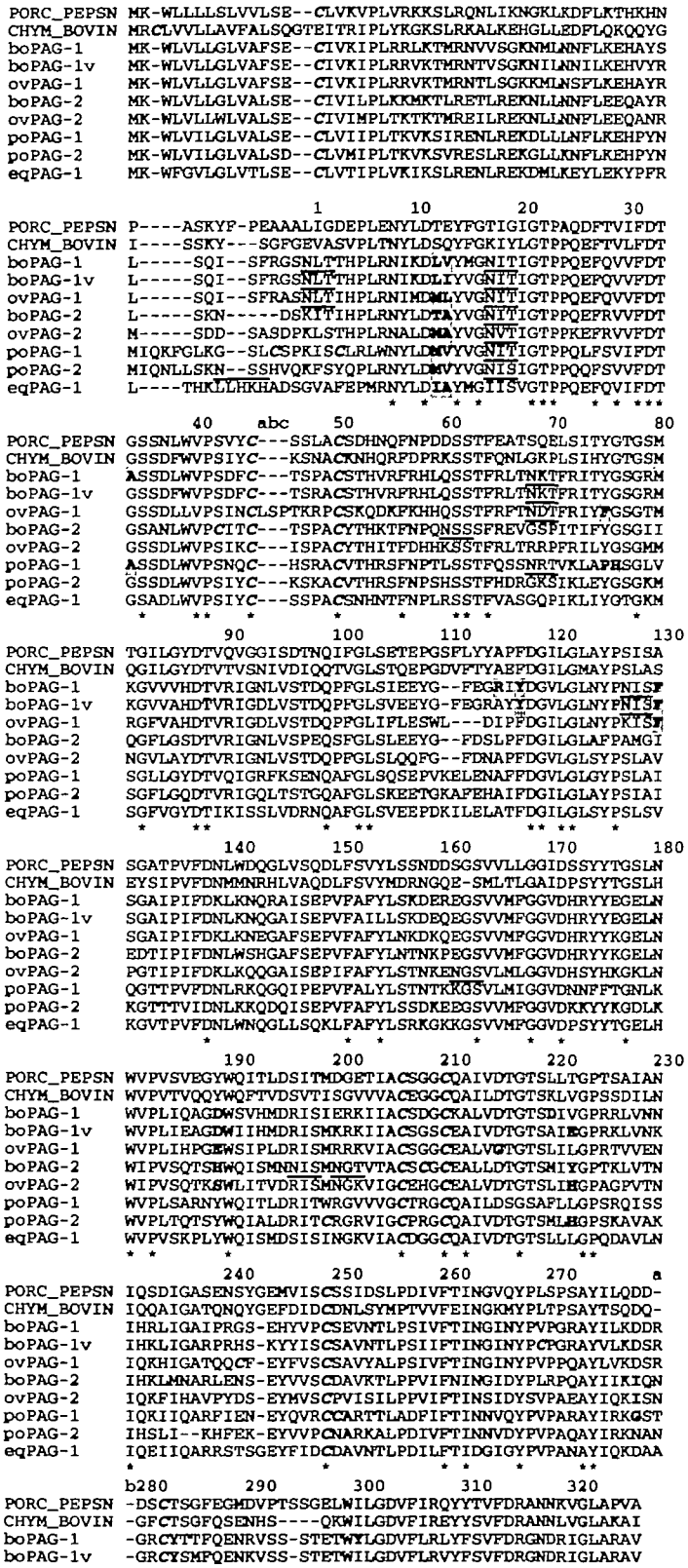
The pregnancy-associated glycoproteins (PAGs) are secretory products synthesized by the outer epithelial cell layer (chorion) of the placentas of various ungulate species. The amino acid sequences of eight PAGs have been inferred from cloned cDNA of cattle and sheep, as well as of the non-ruminant pig and horse. We compare the PAG sequences and present results of the three-dimensional models of boPAG-1 and ovPAG-1 that were constructed on the basis of the crystal structures of homologous porcine pepsin and bovine chymosin using a rule-based comparative modelling approach. Further, we compare peptide binding subsites defined by interactions with pepstatin and a decapeptide inhibitor (CH-66) modelled on the basis of crystal structures of other aspartic proteinases. We have extended our analysis of the peptide binding subsites to the other PAG molecules of known sequence by aligning the PAG sequences to the structural template derived from the pepsin family and by making use of the three-dimensional models of the boPAG-1 and ovPAG-1. The residues that are likely to affect peptide binding in the boPAG-1, ovPAG-1 and other PAG molecules have been identified. Sequence comparisons reveal that all PAG molecules may have evolved from a pepsin-like progenitor molecule with the equine PAG most closely related to the pepsins. The presence of substitutions at the S_1 and other subsites relative to pepsin make it unlikely that either bovine, ovine or the porcine PAG-1 have catalytic activity. Only two of the eight PAGs examined (porcine PAG-2 and equine PAG-1) retain features of active aspartic proteinases with pepsin-like activity. Our results indicate that in the PAGs so far characterized the peptide binding specificities differ significantly from each other and from pepsin, despite their high sequence identities. Analysis of the various peptide binding subsites demonstrates why both bovine and ovine PAG-1 are capable of binding pepstatin. The strong negative charge in the binding cleft of boPAG-1 and ovPAG-1 indicates a preference for lysine- or arginine-rich peptides. PAGs represent a family where the possible peptide binding function may be retained through their binding specificities, but where the catalytic activity may be lost in some cases, such as the boPAG-1, ovPAG-1 and the poPAG-1.

Keywords: comparative protein modelling/peptide binding subsites/pregnancy-associated glycoproteins/site-directed mutagenesis

Introduction

The aspartic proteinases (EC 3.4.23) are one of four main classes of endopeptidase and are generally characterized by a highly conserved bi-lobed structure, the presence of two aspartate residues near the active center (Asp32 and Asp215 by pepsin numbering) and inhibition by pepstatin (Tang and Wong, 1987; Davies, 1990). Proteolytic specificity can range from extremely broad as in the case of pepsin, whose primary function is digestion of food, to very narrow in the case of renin, whose only known role is to cleave the hormone precursor angiotensinogen into its active form. Recently, two members of the aspartic proteinase gene family have been described that are catalytically inactive because of key mutations in the amino acids close to the active site (Xie *et al.*, 1991). One of them, ovine pregnancy-associated glycoprotein-1 (ovPAG-1), lacks one of the two catalytic aspartic acid residues (Asp215; pepsin numbering) that participate in catalysis and is substituted by glycine. The other, boPAG-1, has an alanine in place of a normally invariant glycine (Gly34) residue whose presence would cause displacement of the catalytic water molecule from its normal position between the two catalytic aspartic residues (Asp32 and Asp215). Despite these anomalies, bo- and ovPAG-1 bind pepstatin (Xie *et al.*, 1991) and show ~50% sequence identity to pepsin. Their co-localization with placental lactogen in the hormone-producing placental binucleate cells of the placenta (Warren *et al.*, 1990; Zoli *et al.*, 1992a; Roberts *et al.*, 1994) and their presence in serum of pregnant cows (Sasser *et al.*, 1986; Humblot *et al.*, 1988; Zoli *et al.*, 1992b), ewes (Ruder *et al.*, 1988) and other ruminants (Houston *et al.*, 1986; Wood *et al.*, 1986; Haigh *et al.*, 1991) have suggested that they may play a role in fetal–maternal communication, possibly by a mechanism involving their substrate-binding clefts. Subsequently several additional members of this gene family in both cattle (Xie *et al.*, 1994, 1996) and sheep (Nagel *et al.*, 1993) have been identified and shown to be expressed largely if not exclusively in placenta. Southern genomic blotting of bovine (Xie *et al.*, 1996), ovine and porcine DNA (unpublished results) indicates that multiple genes for PAG-related molecules exist in the different species.

Sequence alignment with animal aspartic proteinases has shown that the PAG molecules are most similar to the pepsins. This resemblance extends throughout the full-length of the molecules and has allowed us to construct three-dimensional models for boPAG-1 and ovPAG-1 based on the crystal structures of pepsin and chymosin. The objectives were to explore interactions with peptides within the binding site cleft of boPAG-1 and ovPAG-1 that might provide clues to their specificity, and to compare the peptide binding subsites of boPAG-1 and ovPAG-1 with pepsin and other PAGs, including those recently cloned from pigs and horses. Earlier studies have clearly demonstrated the value of comparative modelling techniques in exploring substrate interactions of aspartic proteinases (Scarborough *et al.*, 1993; Guruprasad *et al.*, 1994; Naik *et al.*, 1995; Scarborough and Dunn, 1994).



Materials and methods

Cloning and sequencing of cDNA

The cDNAs were isolated from placental libraries of cattle, sheep, pigs and horses by standard procedures which have been described elsewhere (Xie *et al.*, 1991, 1994, 1996; Nagel *et al.*, 1993; Green *et al.*, 1994; Szafranska *et al.*, 1995). The data are all derived from full-length cDNA clones which were sequenced fully in both directions by the dideoxy procedure. Amino acid sequences were aligned by using the program MALIGN (Johnson *et al.*, 1993) (Figure 1). A phylogenetic tree was constructed on the basis of evolutionary distances derived from pairwise sequence identities by using the method of Fitsch and Margoliash (1967).

Modelling

Three-dimensional models of boPAG-1 and ovPAG-1 were constructed using the rule-based comparative modelling approach encoded in the COMPOSER suite of computer programs (Sutcliffe *et al.*, 1987a,b). The models were based on the three-dimensional structures of porcine pepsin (PDB code 5PEP at 2.3 Å resolution) (Cooper *et al.*, 1990) and bovine chymosin (PDB code 4CMS at 2.2 Å) (Newman *et al.*, 1991) with which boPAG-1 and ovPAG-1 have >40% sequence identity (Figure 2) and the disulphide connectivities in PAG were assumed to be identical with those in pepsin and chymosin. The structurally conserved regions (SCRs) used by COMPOSER to model the framework and the structurally variable regions (SVRs) are shown in Table I. The models were then refined by using energy minimization techniques in the SYBYL software (Tripos). The PROCHECK suite of programs (Laskowski *et al.*, 1993) was used to evaluate the stereochemical quality of the models. Nearly 84% of the residues in the boPAG-1 model and about 87% of residues in the ovPAG-1 model were in the 'most favoured' regions of the Ramachandran map.

Modelling peptides/inhibitor in the binding subsites

In order to understand how an inhibitor/peptide might bind in the binding subsites, two inhibitors, which have been analysed as crystal structure complexes with other mammalian aspartic proteinases, were modelled separately into the binding-site cleft of boPAG-1 and ovPAG-1. Thus, the three-dimensional structures of mouse renin complexed with a decapeptide inhibitor CH-66 (PDB Code: 1SMR at 2.0 Å) (DeAlwis *et al.*, 1994) and human cathepsin D complexed with the pepstatin inhibitor (PDB Code: 1LYB at 2.5 Å) (Baldwin *et al.*, 1993) were used as guides to model the respective inhibitors in boPAG-1 and ovPAG-1. CH-66 was chosen because it is the longest inhibitor defined by X-ray analysis (as a complex with mouse renin) and was valuable in defining the sites S_6 to S_4'. Pepstatin was chosen in order to understand the basis of its apparent ability to bind to boPAG-1 (S.Xie and R.M.Roberts,

Fig. 1. Alignment of the pregnancy-associated glycoprotein (PAG) sequences with those of porcine pepsin (SWISSPROT code: PORC_PEPSN) and bovine chymosin (CHYM_BOVIN) using the program MALIGN (Johnson *et al.*, 1993). Abbreviations: bo, bovine; ov, ovine; po, porcine; eq, equine. BoPAG-1v is the variant of boPAG-1. The numbering above PORC_PEPSN corresponds to the numbering in porcine pepsin as in the PDB code 5PEP and is corrected after residue number 230. Residues in PAG, which are close to the binding subsites and different from porcine pepsin, are highlighted and shaded. Identical residues in all the sequences are marked with an asterisk. The potential glycosylation sites are underlined. The half cystines are shown in bold italics.

	Pepsin	Chymosin	boPAG-1	boPAG-1v	boPAG-2	ovPAG-1	ovPAG-2	poPAG-1	poPAG-2	eqPAG
Pepsin	100									
Chymosin	59.5	100								
boPAG-1	49.5	42.5	100							
boPAG-1v	50.8	42.9	86.1	100						
boPAG-2	50.8	45.6	57.8	58.8	100					
ovPAG-1	49.4	42.3	70.6	71.6	58.5	100				
ovPAG-2	50.5	45.9	60.4	60.2	63.4	60.4	100			
po-PAG-1	48.6	43.5	48.8	50.5	48.5	47.4	52.5	100		
poPAG-2	52.9	44.3	56.2	55.1	56.7	54.2	57.4	61.8	100	
eqPAG	58.6	52.3	54.9	55.2	55.4	55.5	54.6	55.3	58.5	100

Fig. 2. Phylogenetic tree for PAGs and related aspartic proteinases constructed by the method of Fitch and Margoliash (1967). The distances derived from the pairwise percentage sequence identities using the multiple sequence alignment program MALIGN (Johnson *et al.*, 1993) were used for the tree construction.

unpublished results) and therefore to support future experimental studies on ligand binding. The inhibitor modelling was achieved by superimposing the model of boPAG-1 and ovPAG-1 on the crystal structures of the enzyme–inhibitor complexes and transferring the coordinates of the inhibitors onto the models. Small adjustments that were required to optimize the fit were carried out by FRODO (Jones and Thirup, 1986) and further refinement was achieved by using the energy minimization procedure in SYBYL. Interactions within the binding subsites between the PAG and inhibitor/peptide atoms were identified by a 4.0 Å cut-off value. The residues in PAG that satisfy this criterion are listed in Tables II and III. The coordinates of the PAG models are deposited in the Brookhaven Protein Data Bank (codes for the models provided by the authors: BPAG1.PDB, bovine PAG-1 and OPAG1.PDB, ovine PAG-1).

Results and discussion

Sequence comparisons

An alignment of the amino acid sequences of porcine pepsin, bovine chymosin and all of the PAG molecules currently available is shown in Figure 1. The numbering throughout corresponds to that of porcine pepsin (PDB code 5PEP). Partial sequences are available for several additional PAG from cattle and sheep, most likely reflecting the large numbers of genes that have been detected by Southern genomic blotting (Xie *et al.*, 1996). The numbering system of PAG reflects the order in which PAG cDNA were cloned from each of the species. Fortuitously, there is limited immunological cross-reactivity between boPAG-1, ovPAG-1 and poPAG-1, while none exists between PAG-1 and PAG-2 molecules even within a species (Xie *et al.*, 1991; Szafranska *et al.*, 1995). The boPAG-1v (v = variant) cDNA was identified by library screening with an anti-boPAG-1 antiserum and was sufficiently similar to boPAG-1 (91.5% identity at the nucleotide level) (Xie *et al.*, 1996) that it was not given a separate number. The sequences of all the PAG, except boPAG-1 and ovPAG-1, have been inferred entirely from their cloned cDNA. Limited sequencing of the N-termini of chromatographically purified boPAG-1 (Xie *et al.*, 1991) and ovPAG-1 (S.Xie and R.M.Roberts, unpublished results) has shown that the mature protein sequence of both begins at Arg4. As remarked upon previously, boPAG-1 and ovPAG-1 are almost certainly without enzymatic activity (Xie *et al.*, 1991). The former because of a Gly→Ala

Table I. Selection of fragments for structurally variable regions (SVRs) and structurally conserved regions (SCRs) in the models

	boPAG-1 model	
SVR	SCR	PDB code of protein used in the model
Arg4–Asn1		3APP:Ala1
	Leu1–Ile73	5PEP
Thr74–Ser77		5PEP:Leu71
	Gly78–Glu107	5PEP
Tyr108–Phe112		2PRK:Ser132
	Glu113–Phe130	5PEP
Ser131–Gly132		5PEP:Ile128
	Ala133–Ser156	4CMS
Lys157–Arg160		5PEP:Tyr154
	Glu161–Arg240	5PEP
Gly241–Ser242		3APP:Val21
	Glu244–Lys277	5PEP
Asp278–Gly280		1SMR:A/Ala237
	Arg281–Glu288	5PEP
Asn289–Glu297		5HVP:A/Gly73
	Thr298-Val326	4CMS
	ovPAG-1 model	
SVR	SCR	PDB code of protein used in the model
Arg4–Asn1		3APP:Ala1
	Leu1–Cys45	5PEP
Leu45a–Pro49		3APP:Ser128
	Cys50–Tyr74	5PEP
Phe75–Ser77		4CMS:Ser72
	Gly78–Glu107	5PEP
Ser108–Pro116		2HFL:L/Ala9
	Phe117–Phe130	5PEP
Ser131–Ala133		5PEP:Ile128
	Ile134–Lys157	5PEP
Asp158–Lys159		5PEP:Leu155
	Gln160–Gln240	5PEP
Cys241–Phe242		3APP:Val21
	Glu244–Lys277	5PEP
Asp278–Arg279		1LYB:B/Gln212
	Gly280–Asn289	4CMS
Arg290–Ala296		4HVP:A/Thr74
	Glu297–Val326	4CMS

Residues are numbered according to porcine pepsin (PDB code 5PEP) as in Figure 1. The PDB codes for modelling the SCRs and the PDB code, chain identifier, starting residue and residue number for modelling the SVRs are shown.

Table II. Binding subsite interactions in the models of boPAG-1 with peptides in the active site[a]

Pockets	CH-66[b] ($P_6 \rightarrow P_4'$)	Pepstatin[c] ($P_4 \rightarrow P_3'$)
S_6	**L12**	–
S_5	–	–
S_4	S219, **D220**, Q287	**L12**, S219
S_3	S77, **R115**, G217, T218, S219	**R115**, **Y117**, G217, T218, S219
S_2	**Y75**, G76, S77, T218, **V222**, N289, **Y300**	G76, S77, G217, T218
S_1	D32, **A34**, **Y75**, S77, **Y117**, L213, **D215**, G217, T218	V30, D32, **A34**, **Y75**, G76, S77, **R115**, **Y117**, **D215**, G217, T218
S_1'	**A34**, **D189**, L213, **Y300**	–
S_2'	**A34**, I128, F130	**A34**, F130
S_3'	F130	**T74**, **V291**, **Y300**
S_4'	–	–

Residues in the binding subsites that differ in the boPAG-1 and ovPAG-1 models are highlighted in bold.
[a] A distance cut-off (4.0 Å) was used to define the binding pockets; residues are numbered according to porcine pepsin (Protein Data Bank; Code 5PEP) (Bernstein *et al.*, 1977; Abola *et al.*, 1987).
[b] CH-66-trimethylacetyl-His-Pro-Phe-His-Leu-ψ(CH(OH)CH_2)-Leu-Tyr-Tyr-Ser-NH_2.
[c] Pepstatin: Iva-Val-Val-Sta-Ala-Sta.

Table III. Binding subsite interactions in the models of ovPAG-1 with peptides in the active site[a]

Pockets	CH-66[b] ($P_6 \rightarrow P_4'$)	Pepstatin[c] ($P_4 \rightarrow P_3'$)
S_6	**M12**,Y283	–
S_5	–	–
S_4	S219, **L220**, Q287	S219, **L220**
S_3	L13, S77, G217, T218, S219	**L13**, S77, G217, T218, S219
S_2	**F75**, G76, S77, T218, **L222**, **I300**	G76, S77, T218, S219
S_1	D32, **G34**, **F75**, S77, G217, T218	V30, D32, **G34**, **F75**, G76, S77, G217, T218
S_1'	**G34**, **E189**, L213, **I300**	–
S_2'	**G34**, I128, **E189**, F130	**G34**
S_3'	F130	**Y74**, **E189**, **I300**
S_4'	–	–

Residues in the binding subsites that differ in the boPAG-1 and ovPAG-1 models are highlighted in bold.
[a] A distance cut-off (4.0 Å) was used to define the binding pockets; residues are numbered according to porcine pepsin (Protein Data Bank; Code 5PEP) (Bernstein *et al.*, 1977; Abola *et al.*, 1987).
[b] CH-66-trimethylacetyl-His-Pro-Phe-His-Leu-ψ(CH(OH)CH_2)-Leu-Tyr-Tyr-Ser-NH_2
[c] Pepstatin: Iva-Val-Val-Sta-Ala-Sta.

substitution at position 34, the latter because one of the catalytic aspartic acid residues (Asp215) is replaced by Gly. Similar reasoning suggests poPAG-1, with an Ala instead of a Gly at position 34, is also likely to be inactive. The remaining PAG possess consensus sequences around Asp32 and Asp215 found in active aspartic proteinases, so that catalytic activity cannot be ruled out on this basis.

Figure 2 reflects the considerable degree of sequence identity existing between pepsin, chymosin and the PAGs. Sequence differences are not highly concentrated but are distributed throughout the full lengths of the molecules (Figure 1). The similarity of boPAG-1 and ovPAG-1 to pepsin suggests that all three molecules are closely related and would possess a similar three-dimensional fold. In order to analyse their relatedness more rigorously, a phylogenic tree was constructed on the basis of amino acid sequence differences between the

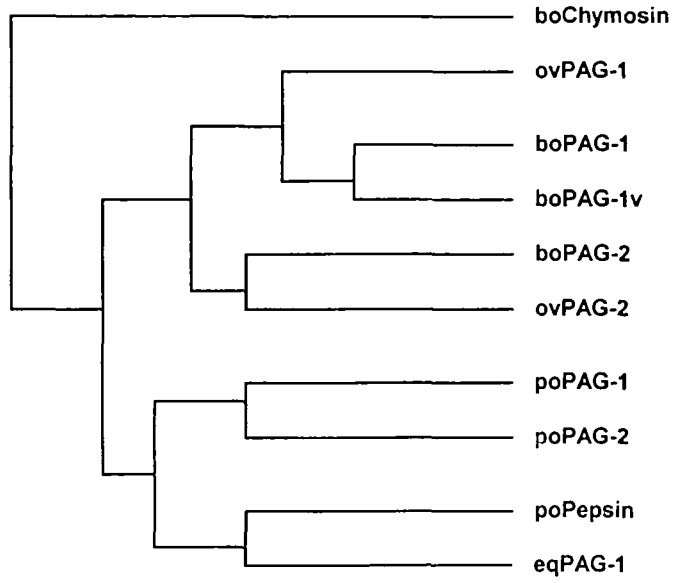


Fig. 3. Pairwise percentage identities of pepsin, chymosin and the PAG sequences.

Table IV. Hydrogen bonds involved in peptide binding defined by FRODO

Model	boPAG-1		Pepstatin	
	Atom	Residue	Atom	Residue
boPAG-1 model with pepstatin	OG	Ser219	Carbonyl oxygen	Iva (P_4)
	N	Ser219	Carbonyl oxygen	Val (P_3)
	OD_1	Asp32	OS	Sta (P_1)
	OD_2	Asp215	OS	Sta (P_1)
	boPAG-1		CH-66	
	Atom	Residue	Atom	Residue
boPAG-1 model with CH-66	N	Gly76	Carbonyl oxygen	His (P_2)
	ND_2	Asn289	NE_2	His (P_2)
	OD_1	Asp32	OH	Leu (P_1)
	OD_2	Asp215	OH	Leu (P_1)
	ovPAG-1		Pepstatin	
	Atom	Residue	Atom	Residue
ovPAG-1 model with pepstatin	OG	Ser219	Carbonyl oxygen	Iva (P_4)
	N	Ser219	Carbonyl oxygen	Val (P_3)
	N	Gly76	Carbonyl oxygen	Val (P_2)
	OD_1	Asp32	OS	Sta (P_1)
	ovPAG-1		CH-66	
	Atom	Residue	Atom	Residue
ovPAG-1 model with CH-66	N	Gly76	Carbonyl oxygen	His (P_2)
	OD_1	Asp32	OH	Leu (P_1)

PAG, bovine chymosin and porcine pepsin (Figure 3). This analysis again emphasizes the relatedness of the PAG-1 of cattle and sheep and the distinctiveness of the PAG-1 and PAG-2 groupings. However, the porcine PAG-1 and -2 are more related to each other than they are to the comparably numbered PAG of cattle and sheep. This observation is consistent with the relatively early divergence within the

Artiodactyls of the Suina and Ruminantia suborders between 55 and 60 million years ago (Van Soest, 1982). The even earlier split of the horses (order Perissodactyla) from the Artiodactyla order appears to be reflected in the fact that equine PAG most clearly resembles porcine pepsin and is the least related to the ruminant PAG.

The three-dimensional models

A schematic representation of the three-dimensional model of boPAG-1 (Figure 4A) shows the overall arrangement of helices and strands and pepstatin in the binding subsites. The overall three-dimensional fold of ovPAG-1 (not shown) is very similar to that of boPAG-1. In ovPAG-1 (figure not shown), an extra half-cystine (Cys241) is exposed to the solvent and may be involved in dimerization through the formation of an intermolecular disulphide bridge.

Comparison of the models of ovPAG-1 (324 amino acids) with boPAG-1 (327 amino acids) shows that for 314 topologically equivalent C^{α}-atoms in the two models, the root mean square deviation (r.m.s.d.) is 0.46 Å (the five residues between Cys45 and Cys50 in ovPAG-1 which are distant from the binding site cleft were not modelled as an equivalent fragment to model this loop was not available from either within or outside the homologous family).

Comparison of the models with porcine pepsin and bovine chymosin showed that in each case the models were closer to porcine pepsin; boPAG-1 and porcine pepsin have an r.m.s.d. of 0.45 Å for 308 of 327 C^{α}-atoms and ovPAG-1 and porcine pepsin have an r.m.s.d. of 0.39 Å for 302 of 324 C^{α}-atoms, whereas, with bovine chymosin, boPAG-1 has an r.m.s.d. of 0.77 Å for 300 topologically equivalent C^{α}-atoms and ovPAG-1 has an r.m.s.d. of 0.64 Å for 288 topologically equivalent C^{α}-atoms.

Binding subsites

The binding subsites for the two inhibitors (pepstatin and CH-66) complexed with boPAG-1 and ovPAG-1 are listed in Tables II and III, respectively. We have compared the binding sites of porcine pepsin and PAG in order to identify possible differences in specificity (Figure 1 and Tables II and III). The most obvious difference is in the charge distribution, which is shown for boPAG-1 in Figure 4(B) calculated by GRASP (Nichols and Honig, 1991). The electronegative charge at the binding sites is mainly due to the presence of Asp220 (S_4 subsite) and Asp189 (S_1' subsite) in addition to the catalytic Asp32 and Asp215 of boPAG-1. It suggests a preference for basic peptides in boPAG-1 and ov-PAG-1, whereas, in boPAG-2 and ovPAG-2 the equivalent residues are polar but not negatively charged.

There are also differences of hydrophobicity. For example, at the S_6 pocket the hydrophobic groups (leucine, isoleucine and methionine) in ovPAG-1, ovPAG-2, poPAG-1, poPAG-2 and eqPAG-1 imply preference for a hydrophobic residue at residue P_6, whereas boPAG-2, like pepsin, has a threonine (see Figure 1). At the S_4 pocket the specificity may be determined by the nature of residue 220 which is leucine in pepsin and cathepsin D. This residue together with Leu12 interacts with the side chain of Iva of pepstatin. The presence of hydrophobic groups in boPAG-2 (Met220), poPAG-1 (Phe220) and poPAG-2 (Met220) should favour pepstatin binding, but the presence of Asp220 in boPAG-1 is likely to lead to weaker binding. S_3 is a deep pocket that varies in hydrophobicity in different aspartic proteinases. PAGs may have an increased preference for hydrophobic residues, as residues such as 12, 13 and 117 tend to be hydrophobic. In contrast, mammalian aspartic proteinases have either glutamic acid or glutamine at position 13.

Large hydrophobic S_1 and S_1' pockets are characteristic of most aspartic proteinases and these are largely retained in PAGs. An exception is the tyrosine at position 189, which is conserved in all mammalian aspartic proteinases except mouse renin (Ser189) and human renin (Val189). Non-aromatic residues are found in PAGs, for example Asp189 (in boPAG-1 and boPAG-1v), Glu189 (in ovPAG-1), Ser189 (ovPAG-2) or His189 (boPAG-2). These are likely to influence peptide binding specificities. In particular, the aspartic acid and glutamic acid substitutions at position 189 would probably reduce binding and, therefore, proteolytic activity except at low pH.

Phe130, which is conserved between boPAG-1, boPAG-1v and ovPAG-1, may also be important for specificity. In other PAGs it is replaced by isoleucine or valine. In the CH-66 model of boPAG-1, the side chain of Phe130 clashes with the residue at P_2' (Tyr) and P_3' (Tyr), possibly precluding tight peptide binding. In contrast, pepstatin binding would be less affected owing to the much smaller residue at P_2' (alanine).

Only one binding site may be affected by deletions in the main chain in PAGs. For example, residues 115 and 117 line the S_1 subsite. Deletions of one to three residues close to this region occur in the bovine and ovine PAGs.

The hydrogen bond interactions proposed between PAG and the two inhibitors are shown in Table IV. In general, it should be possible to form the same main chain to main chain bonds between PAGs and ligands as found in normal aspartic proteinases. In ovPAG-1 the hydrogen bond with side chain of Asp215 is not possible owing to the mutation (Asp215–Gly). In addition, in the case of poPAG-1 the hydrogen bond between the carbonyl oxygen of P_4 residue and the side chain of Ser219 is not possible owing to substitution by alanine. In boPAG-1, the side chain of Asn289 is likely to interact with the peptide via a side chain hydrogen bond with the NE_2 atom of His (P_2). This may not be possible in ovPAG-1 owing to Pro285 (unique to ovPAG-1) that is likely to alter the conformation of the backbone, leading to differences in the side chain orientation of Asn289 in the bovine and ovine PAG-1 molecules.

These observations are in general consistent with binding data which show that pepstatin interacts with various PAGs. Stereo-views illustrating the extended binding nature of pepstatin in boPAG-1 and ovPAG-1 are shown in Figure 4(C) and (D), respectively. ovPAG-1 can be successfully trapped on affinity columns, along with other PAGs secreted simultaneously, from cultured ovine placental explants (R.J.Nagel and R.M.Roberts, unpublished results). The boPAG-2 and poPAG-2, after being expressed in the methanolic yeast *Pischia pastoris*, can be purified in a single-step procedure by affinity chromatography on a pepstatin-agarose column. The binding buffer is 50 mM sodium acetate buffer, pH 5.0, containing 200 mM NaCl, 2 mM EDTA, 0.2 mM phenylmethylsulphonyl fluoride (PMSF), 0.02% NaN_3. The column is washed with 10 volumes of 1 M NaCl. The boPAG-2 is eluted with 20 mM Tris–HCl, pH 8.0, containing 1 M NaCl plus EDTA, PMSF and NaN_3. poPAG-2 elution requires more stringent conditions involving pH 9.0 buffer (J.Green, S.Xie and R.M.Roberts, unpublished results). The proteolytic activity of eqPAG-1 (prepared in a refolded recombinant product in *Escherichia*

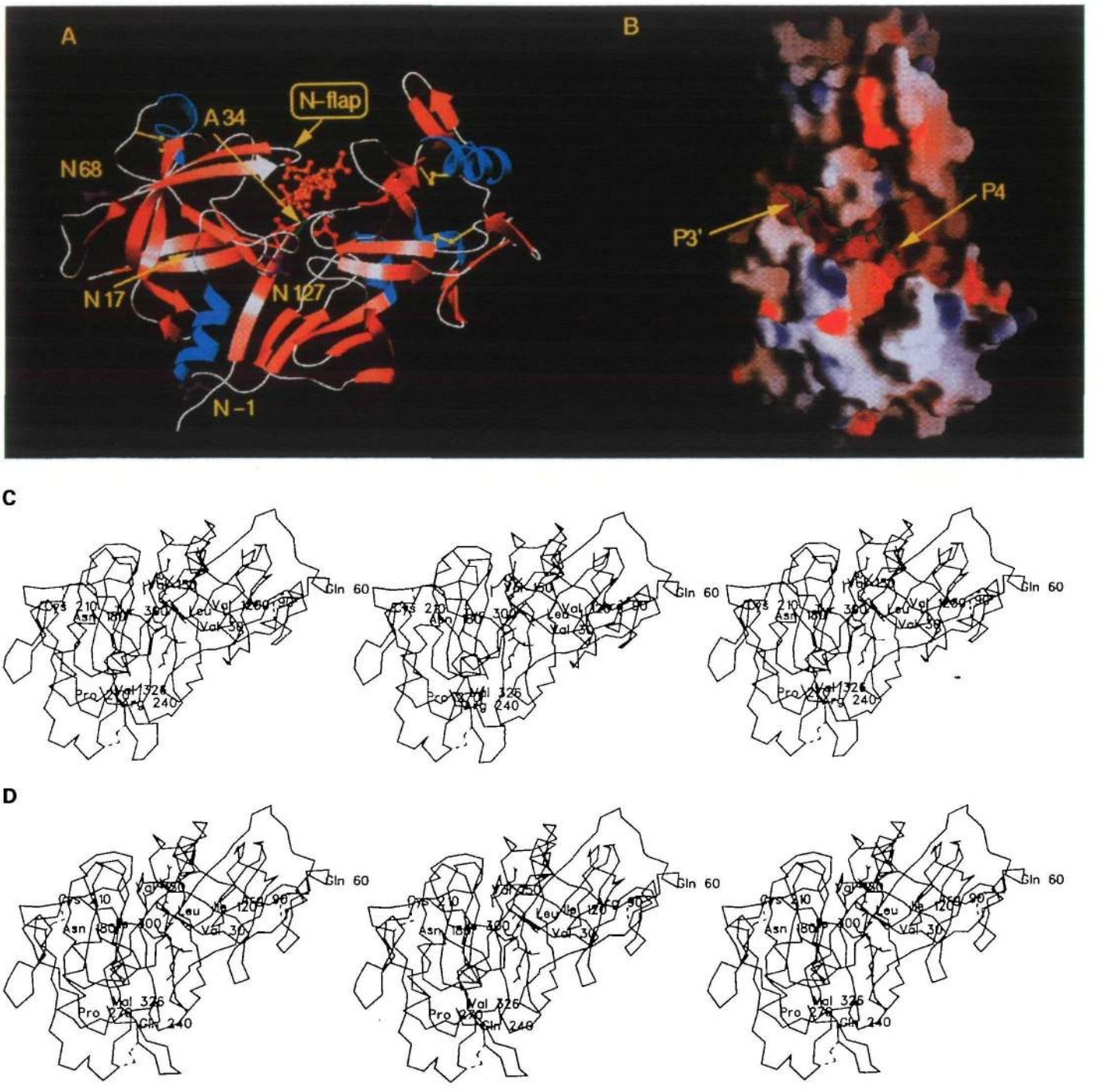


Fig. 4. (**A**) Schematic representation of the three-dimensional fold of boPAG-1. Helices (pale blue) are shown as coils and sheets (pink) as flat ribbons. The pepstatin inhibitor (brown) is shown in the binding site. The side chain of Ala34 (green) that displaces the catalytic water molecule in the aspartic proteinases is shown. The C^{β}-atom of Ala34 is ~2.01Å from the OS atom of Sta (P_1–P_1') in pepstatin. The side chains of the aspartates (Asp32 and Asp215) are shown in red and those of the potential glycosylation sites are shown in purple. (**B**) Pepstatin modelled in the binding cleft of boPAG-1. The molecular surface showing the charge distribution is drawn by using GRASP (Nichols and Honig, 1991). Color code: negative charges (red), positive charges (blue). The binding cleft appears to be highly electronegatively charged. (**C**) Stereo-view representation of boPAG-1 model with bound pepstatin showing the extended nature of peptide binding in the cleft. The direction of view is perpendicular to the chain of pepstatin with P_4 at the lower end and P_3' at the top. Every thirtieth residue in the model is labelled according to the numbering in porcine pepsin (PDB code 5PEP). (**D**) Stereo-view representation of ovPAG-1 model with bound pepstatin.

coli) towards its propeptide, is eliminated by a threefold molar excess of pepstatin (J.Green and R.M.Roberts, unpublished results).

Positions and effects of glycosylation

Asparagine residues that can potentially be *N*-glycosylated on the different PAGs, i.e. present in the context Asn–X–Thr/Ser,

are shown in Figure 1. The anomalously high molecular weight of the PAG (~70 000) (Xie *et al.*, 1991, 1994) may be due at least in part to glycosylation. Carbohydrate on certain ovPAG molecules has also been reported to contribute to their antigenicity (Atkinson *et al.*, 1993). The models indicate that the side chains of Asn-1, Asn68 and Asn127 of boPAG-1 are exposed to solvent, while that of Asn17 may be partially buried (Figure 4A) owing to the close proximity of the side chains of Arg7, Glu26 and Gln28 (not shown in the figure). Edman sequencing of ovPAG-1 has provided no signal on cycles 4 and 21 (positions −1 and 17, pepsin numbering) (S.Xie and R.M.Roberts, unpublished results). Therefore, both asparagines are likely to be glycosylated. In general these glycosylation sites are unlikely to inhibit substrate binding; the exception might be the N-linked carbohydrate chain in boPAG-2 and in eqPAG-1 at Asn294 which might interfere with peptide binding near the prime site subsites.

Are any of the PAG proteolytically active?

There are several residues that may affect catalytic activity. The substitution of Asp215 with glycine in ovPAG-1 provided the original basis for predicting that it would be proteolytically inactive, while the projection of the methyl group of Ala34 in boPAG-1 into the S_1 pocket would displace a water molecule that is required for catalysis (Xie *et al.*, 1991). The C^{β}-atom of Ala34 in boPAG-1 is within 2.0 Å of the OH atom of P_1 in the CH-66 inhibitor model. A similar argument holds for the likely inactivity of poPAG-1, which also has alanine at position 34 (Figure 1). Interactions with the 'flap' at position 75 are also important for activity as the ring hydrogens come close to the tetrahedral intermediate. Substitutions at position 75 in ovPAG-1 (Phe75) and in poPAG-1 (Pro75) (Figure 1) may reduce activity and also influence peptide binding in these two PAG, as will the replacement of the conserved Gly76 with histidine in poPAG-1.

Only poPAG-2 and eqPAG-1 have the complete consensus set of amino acids around the catalytic aspartic acid residues. Neither possesses unfavourable substitutions at residues 115, 117, 189, 289 and 300 and neither has deletions between Glu107 and Asp118. Both have relatively conservative substitutions for phenylalanine (isoleucine and valine, respectively) at position 130. The only major difference from pepsin is at position 12, where eqPAG-1 has isoleucine and poPAG-2 has a methionine. These observations are consistent with the observation that the unglycosylated recombinant eqPAG-1 synthesized in *E.coli* is active and can remove its own pro-peptide (J.Green and R.M.Roberts, unpublished results). Experiments are in progress to determine whether poPAG-2 also has retained proteolytic activity.

Conclusions

Sequence analyses have strongly suggested that PAGs most closely resemble pepsin. This work supports the conclusion that their three-dimensional folds are also likely to be very similar to that of pepsin. Although neither the biological role nor substrate specificity of PAG molecules are yet known, we have suggested those residues that are most likely to influence the binding specificity in PAG molecules and have thus provided the basis for further biological and biochemical experiments. Of the known PAG, only poPAG-2 and eqPAG-1 are likely to have potential proteolytic activity and in the case of eqPAG-1 this has now been confirmed. If the functions of the PAGs are to bind small peptides in their clefts, then they will exhibit dissimilar specificities. Bovine and ovine PAG-1 may have a general affinity for binding peptides with basic residues. The PAG represent a family where the peptide binding function may have been retained but the catalytic activity, at least in boPAG-1, ovPAG-1 and poPAG-1, may have been lost owing to the critical mutations within the binding subsites.

Acknowledgements

This research was funded by grant HD 29483 from the National Institutes of Health. Dr Kunchur Guruprasad was supported by grants from the UK Biotechnology and Biological Sciences Research Council and from Parke Davis Pharmaceutical Research, Ann Arbor, MI, USA. The nucleotide sequences in this paper have been deposited in the GenBank/EMBL Data Bank with Accession Numbers M73961, M73962, L06151, L06153, L34360, L34361 (submitted: entry code not yet received) and L38511. The three-dimensional coordinates for the boPAG-1 and ovPAG-1 models have been deposited (submitted: entry codes not yet received) in the Brookhaven Protein Data Bank (PDB codes provided by the authors: BPAG1.PDB, bovine PAG-1; OPAG1.PDB, ovine PAG-1).

References

Abola,E.E , Bernstein,F.C., Bryant,S.H., Koetzle,T.F. and Weng,J. (1987) In Allen,F.H., Bergerhoff,G. and Sievers,R. (eds), *Crystallographic Databases—Information Content, Software Systems, Scientific Applications.* Data Commission of the International Union of Crystallography, Bonn, pp. 107–132.

Atkinson,Y.H., Gogolin-Ewens,K.J., Hounsell,E.F., Davies,M.J., Brandon,M.R. and Seamark,R.F. (1993) *J. Biol. Chem.*, **268**, 26679–26685.

Baldwin,E.T. *et al.* (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6796–6800.

Bernstein, F.C. *et al.* (1977) *J. Mol. Biol.*, **112**, 535–542.

Cooper,J.B., Khan,G., Taylor,G., Tickle,I.J. and Blundell,T.L. (1990) *J. Mol. Biol.*, **214**, 199–222.

Davies,D.R. (1990) *Annu. Rev. Biophys. Chem.*, **19**, 189–215.

DeAlwis,C.G. *et al.* (1994) *J. Mol. Biol* , **236**, 342–360.

Fitsch,W.M. and Margoliash,E. (1967) *Science*, **155**, 279–284.

Green,J., Xie,S., Newman,A., Szafranska,B., Roberts,R.M., Baker,C.B. and McDowell,K. (1994) *Biol. Reproduct.*, **50** (Suppl. 1), 152 (Abstract).

Guruprasad,K., Tormäkangas,K., Kervinen,J. and Blundell,T.L. (1994) *FEBS Lett.*, **352**, 131–136.

Haigh,J.C., Gates,C., Ruder,A and Sasser,R. (1991) *Theriogenology*, **36**, 749-754.

Houston,D.B., Robbins,C.T., Ruder,C.A. and Sasser,R.G. (1986) *J. Wildl Manage.*, **50**, 740–742

Humblot,P , Camous,S., Martal,J , Charlery,J., Jeanguyot,N., Thibier,M. and Sasser,R.G. (1988) *Theriogenology*, **30**, 257–268.

Johnson,M.S., Overington,J.P. and Blundell,T.L. (1993) *J. Mol. Biol.*, **231**, 735–752.

Jones,T.A. and Thirup,S. (1986) *EMBO J.*, **5**, 819–822.

Laskowski,R.A., MacArthur,M W., Moss,D.S. and Thornton,J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Nagel,R.J., Xie,S. and Roberts,R M. (1993) *Biol. Reproduct.*, **48** (Suppl 1), 139 (abstract).

Naik,C.R., Guruprasad,K., Batley,B., Rapundalo,S., Hill,J., Blundell,T., Kay,J. and Dunn,B.M. (1995) *Proteins: Struct. Funct. Genet.*, **22**, 168–181.

Newman,M., Safro,M., Frazao,C., Khan,G., Zdanov,A., Tickle,I.J., Blundell,T.L. and Andreeva,N. (1991) *J. Mol. Biol.*, **221**, 1295–1309.

Nichols,A. and Honig,B. (1991) *J. Comput. Chem.*, **12**, 435–445.

Roberts,R.M., Kramer,K.K., Xie,S., Duffy,J., Nagel,R.J. and Wooding,F.B.P. (1994) In Mastroianni,L., Coelingh Bennink,H.J.T., Suzuki,S. and Vemer, H.M.G. (eds), *Gamete and Embryo Quality.* Parthenon Publishing, London, pp. 209–223.

Ruder,C.A., Stellflug,J.N., Dahmen,J.J. and Sasser,R.G. (1988) *Theriogenology*, **29**, 905–912.

Sasser,R.G., Ruder,C.A., Ivani,K.A., Butler,J.E. and Hamilton,W.C. (1986) *Biol. Reproduct.*, **35**, 936–942.

Scarborough,P.E. and Dunn,B.M. (1994) *Protein Engng*, **7**, 495–502.

Scarborough,P.E., Guruprasad,K., Topham,C., Richo,G.R., Conner,G.E., Blundell,T.L. and Dunn,B.M. (1993) *Protein Sci.*, **2**, 264–276.

Sutcliffe,M.J., Haneef,I., Carney,D. and Blundell,T.L. (1987a) *Protein Engng*, **1**, 377–384.

Sutcliffe,M.J., Hayes,F.R.F. and Blundell,T.L. (1987b) *Protein Engng*, **1**, 385–392.

Szafranska,B., Xie,S., Green,J. and Roberts,R.M. (1995) *Biol. Reproduct.*, **53**, 21-28.

Tang,J. and Wong,R.N S. (1987) *J. Cell. Biochem.*, **33**, 53–63.
Van Soest,E (1982) *Nutritional Ecology of the Ruminant* O&B Books, New York.
Warren,W.C., Liang,R., Krivi,G.G., Siegel,N.R. and Anthony,R.V. (1990) *J. Endocrinol.*, **126**, 141–149.
Wood,A.K., Short,R.E., Darling,A.-E., Dusek,G.L., Sasser.R.G. and Ruder,C.A. (1986) *J Wildl. Manage.*, **50**, 684–687.
Xie,S., Low,B.G., Kramer,K.K., Nagel,R.J., Anthony,R.V., Zoli,A P., Beckers,J.-F. and Roberts,R.M. (1991) *Proc. Natl Acad. Sci USA*, **88**, 10247–10251.
Xie,S., Low,B.G., Nagel,R.J., Beckers,J.-F and Roberts,R.M. (1994) *Biol Reproduct.*, **51**, 1145–1153.
Xie,S., Green,J., Beckers,J.-F. and Roberts,R.M. (1996) *Gene*, in press.
Zoli,A.P., Demez,P., Beckers,J.F., Reznik,M. and Beckers,A. (1992a) *Biol. Reproduct.*, **46**, 623–629.
Zoli,A.P., Guilbault,L.A., Delabaut,P., Ortiz,W.B. and Beckers,J.F. (1992b) *Biol. Reproduct.*, **46**, 83–92.

Received December 4, 1995; revised April 10, 1996, accepted April 18, 1996