Amelogenins

Sequence homologies in enamel-matrix proteins from three mammalian species

A. G. FINCHAM,*‡ A. B. BELCOURT,*§ J. D. TERMINE,* W. T. BUTLER† and W. C. COTHRAN†

**Skeletal Matrix Biochemistry Section, Laboratory of Biological Structure, National Institute of Dental Research, National Institutes of Health, Bethesda, MD, U.S.A., and †Institute of Dental Research, School of Dentistry, University of Alabama in Birmingham, AL, U.S.A.*

(Received 24 November 1982/Accepted 3 December 1982)

Partial amino acid sequences for selected amelogenin polypeptides isolated from the developing enamel of cow, pig and human foetuses are reported. It was found that there was an identity of sequence for the initial 28 residues of the polypeptides analysed, irrespective of their origin or size. A tyrosine-rich polypeptide was shown to be the *N*-terminal fragment of the principal higher-molecular-weight amelogenins, although a leucine-rich polypeptide of similar size was not identified in any other amelogenin structure. The findings demonstrate a striking degree of sequence conservation for the amelogenin proteins of the extracellular enamel matrix and support the concept of a discrete fragmentation of an initial 30 000 Da amelogenin molecule during the mineralization of the enamel.

Two general classes of proteins have been described for the extracellular matrix of developing dental enamel: the acidic enamelines, which, although representing only a small proportion of the immature matrix protein, survive at least in part in the mature enamel, and the more abundant proline-rich amelogenins, which are progressively lost during the mineralization of the foetal enamel (Termine *et al.*, 1980; Fincham *et al.*, 1982*a,c*). The biological role of these enamel-matrix proteins is presently unknown, although a functional participation in the structural organization and calcification of the dental enamel has been suggested (Eastoe, 1979; Fearnhead, 1979; Nylen, 1979).

The purification and characterization of the complex of proteins comprising the immature enamel matrix has presented considerable technical difficulties (Fincham, 1971, 1979; Mechanic, 1971; Eggert *et al.*, 1973; Fukae & Shimizu, 1974). More

recently, Termine *et al.*, (1980) have described dissociative extraction and isolation procedures for the bovine amelogenins. A description of the sizes and compositions of the principal molecular species has been presented (Fincham *et al.*, 1981*b*) and complete sequence data for two lower-molecular-weight (5000–6000) amelogenin polypeptides have been reported (Fincham *et al.*, 1981*b*). Apart from these studies of bovine amelogenins, Fukae *et al.* (1980) have reported partial sequences for porcine amelogenins, and compositional data for foetal human amelogenins have recently been obtained (Fincham *et al.*, 1982*b*).

These and other related studies have made it clear that during the mineralization of the enamel there is a gradual degradation of amelogenin molecules of some 30 000–40 000 Da and the appearance of a complex of lower-molecular-size species (Belcourt *et al.*, 1982; Fincham *et al.*, 1982*c*). Among the latter, two major forms have been described, the leucine-rich (LRAP) and the tyrosine-rich (TRAP) amelogenin polypeptides. Sequence data for these two polypeptides have shown that the initial 33 residues are identical for both forms, with enrichment in leucine or tyrosine being expressed only in the C-terminal regions of the 45–46-residue peptides. These sequences were shown to differ significantly from earlier, incomplete, bovine enamel sequence data (Seyer & Glimcher, 1977; Papas *et al.*, 1977),

Abbreviations used: LRAP, leucine-rich amelogenin polypeptide; TRAP, tyrosine-rich amelogenin polypeptide; SDS, sodium dodecyl sulphate.

‡ Present address: Medical Learning Resources Unit, Faculty of Medicine, University of the West Indies, Mona, Kingston 7, Jamaica, West Indies.

§ Present address: Group de Recherches INSERM U 157, Centre de Recherches Odontologiques, Faculté de Chirurgie Dentaire, Université Louis Pasteur, 1 Place de l'Hôpital, 67000-Strasbourg, France.

but were closely related to the partial sequence data reported for porcine amelogenins (Fukae *et al.*, 1979, 1980) and to that reported for the bovine amelogenin polypeptide partially sequenced by Zalut *et al.* (1980).

In the present study we report the sequence data for the *N*-terminal segments of several of the higher-molecular-weight bovine amelogenins, together with data for TRAP-like molecules isolated from both pig and human developing enamel. These studies reveal some striking homologies between the species studied.

Materials and methods

Enamel-matrix isolation

Bovine and porcine enamel-matrix proteins were isolated from fetuses 5–7 months *in utero* after guanidinium chloride extraction as described by Termine *et al.* (1980). Human enamel-matrix proteins were obtained by dilute-acetic acid extraction of enamel scrapings from foetal deciduous teeth removed 28 to 30 weeks *in utero* as described elsewhere (Fincham *et al.*, 1982b).

Gel-exclusion chromatography

The primary fractionation of the matrix proteins was carried out on tandem columns of Bio-Gel P30 (Bio-Rad Laboratories, Richmond, CA, U.S.A.) eluted with 0.1M-formic acid at 10°C as previously described (Fincham *et al.*, 1982a). The fractions isolated from the Bio-Gel P30 columns were further purified and characterized as to molecular size by chromatography on calibrated columns of Sephacryl S-200 (Pharmacia Fine Chemicals, Piscataway, NY, U.S.A.) in 4M-guanidinium chloride (Termine *et al.*, 1980).

Preparative and analytical electrophoresis

Analytical polyacrylamide-gel electrophoresis was conducted with both the SDS/urea and Tris/borate/urea gel systems as described previously (Termine *et al.*, 1980; Fincham *et al.*, 1972). LRAP material and the principal electrophoretic component of the 27000 Da bovine amelogenin complex ('Component-8'; Fincham, 1979) were isolated by preparative electrophoresis in the Tris/borate/urea system as described by Fincham *et al.* (1981b).

Amino acid analyses

Amino acid compositions were determined on a Durrum D-500 amino acid analyser as previously described (Termine *et al.*, 1980).

Sequence determinations

Sequences were determined by automated Edman degradation with a model-890C Beckman automatic sequencer operated at 56°C. The procedure

utilized the 'Slow Protein-Quadrol Program' (no. 042772; Beckman Instruments, 1978) with 0.5M-Quadrol and 2.0mg of Polybrene (Tarr *et al.*, 1978). Amino acid phenylthiohydantoin derivatives were identified by high-pressure liquid chromatography (Bhown *et al.*, 1981) and also by t.l.c. (Inagami & Murakami, 1972). Arginine and histidine phenylthiohydantoin derivatives were identified in the water layer with the phenanthrenequinone spot test and the Pauly diazo reagent respectively. Repetitive yields of amino acid phenylthiohydantoin derivatives, calculated by using the data from Leu-3 and Leu-20, were in general 94–99%.

Results

The chromatographic profiles for the Bio-Gel P30 and Sephacryl S-200 fractionations were presented previously (Fincham *et al.*, 1981a, 1982a,b). Several purified column fractions from bovine, human and porcine enamel matrix were each subjected to automatic Edman degradation. Each of these protein fractions appeared as a single band on SDS/urea/polyacrylamide-gel electrophoresis and as a single chromatographic peak when subfractionated on the Sephacryl S-200/guanidinium chloride columns. The data for the determination of the sequences of bovine TRAP and LRAP were reported previously (Fincham *et al.*, 1981b). A summary of the terminology for these amelogenins, together with the number of residues sequenced, is given in Table 1.

Table 1. *Chromatographic origin and characteristics of amelogenin proteins used in the present study*

For details, reference should be made to previous reports cited in the text.

Species	Fraction designation	Molecular weight*	Electrophoretic composition (no. of bands)†	Residues sequenced
Bovine	S-200B	27 000	Multiple	45
	S-200B'	21 000	4–6	10
	Component-8	27 000	1	53
	P30-D	12 500	2–4	41
	TRAP	5000	1	45
Human	LRAP	7000	1	46
	P30-I	5000	1	33
	P30-J	5000	1	30
Porcine	P30-G	5000	1	41

* Determined by Sephacryl S-200 chromatography in guanidinium chloride (Belew *et al.*, 1978).

† In the Tris/borate/urea system (Fincham *et al.*, 1972). All preparations gave a single band in the SDS/urea/polyacrylamide-gel system (Termine *et al.*, 1980).

Table 2. Amino acid compositions of amelogenins

Amino acid	Amelogenin ...	Composition (residues/molecule)*								
		Component-8	S-200B	S-200B'	D	TRAP†	LRAP†	I	J	G
Aspartic acid		6	6	5	4	1	2	1	1	2
Threonine		6	5	5	4	1	1	1	1	1
Serine		8	8	5	6	3	2	3	3	2
Glutamic acid		47	49	44	16	3	3	3	3	3
Proline		67	65	56	25	9	12	10	8	7
Glycine		9	9	4	7	5	2	3	5	5
Alanine		6	5	6	3	0	1	0	0	0
Valine		9	10	9	7	1	1	1	1	1
Methionine		12	10	9	5	3	3	1	2	3
Isoleucine		9	9	7	4	2	2	2	2	2
Leucine		24	24	20	8	3	6	4	3	3
Tyrosine		8	8	1	6	6	3	6	5	6
Phenylalanine		4	4	3	2	1	1	1	1	1
Histidine		20	20	15	11	3	3	3	2	3
Lysine		2	2	1	2	1	1	1	1	1
Arginine		2	2	1	2	1	1	1	1	1
Tryptophan		n.d.‡	n.d.	n.d.	n.d.	2	1	n.d.	n.d.	n.d.
Amide§		n.d.	47	n.d.	n.d.	2	2	n.d.	n.d.	n.d.

* Based on Sephacryl S-200 chromatography.

† Fincham *et al.* (1981b).

‡ Abbreviation used: n.d., not determined.

§ Fincham *et al.* (1981a).

It was noted that the fraction 'Component-8' (Table 1) represented the principal electrophoretic component seen in the multiple-banded pattern obtained when the bovine Sephacryl S-200B fraction was analysed in the Tris/borate/urea system. Both the Sephacryl S-200B fraction and Component-8 appeared as a single band in the SDS/urea polyacrylamide-gel system. Similarly, the bovine Bio-Gel P30-D fraction was seen to be homogeneous by SDS/urea/polyacrylamide-gel electrophoresis, but to comprise two to four bands when seen in the Tris/borate/urea system.

The bovine sample S-200B' and the human samples I and J were sequenced as far as residues 10, 33 and 30 respectively, but insufficient material or heterogeneity (in the case of the S-200B' sample) prevented further analysis of these fractions. In sequencing the S-200B' fraction, alanine and proline were detected as minor residues in cycles 1 and 4 respectively. All other cycles showed the major residues reported, with less than a 10% level of contamination being observed. Nevertheless, the limited data obtained from these samples provides a wider insight into the patterns of amelogenin sequences.

The sequences found for the several amelogenins are presented in Fig. 1, together with the data for porcine amelogenins of molecular size 21 000–27 000 Da, as reported by Fukae *et al.* (1980). The

amino acid compositions for these amelogenins, calculated as residues per molecule based on the Sephacryl S-200/guanidinium chloride column molecular-weight estimates are presented in Table 2.

Discussion

An examination of the data of Fig. 1 shows that, apart from some uncertainty in determinations for residues 6 and 9 (probably due to difficulties in determination of the histidine phenylthiohydantoin derivative), the *N*-terminal sequences of all of the amelogenins, with the notable exception of the S-200B' fraction, are identical up to Gln-27. The bovine sample LRAP shows an identity with the other bovine fractions as far as Pro-33 and then exhibits a sequence not seen in any other amelogenin so far analysed. Porcine fraction G is identical with bovine TRAP as far as determined (Pro-41), with the notable substitution of asparagine for serine at residue 28 and uncertainty regarding residues 35 and 36. The TRAP sequence appears to be the common *N*-terminal of all of the bovine amelogenins analysed, with the exceptions of LRAP and S-200B'.

The S-200B' fraction (M_r 21 000) is about 5000 Da smaller than the bovine S-200B fraction (or its major electrophoretic entity, 'Component-8'). The overall amino acid composition of the S-200B'

	5	10	15	20	25
Comp-8*	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
S-200B	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
LRAP†	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
TRAP†	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
D	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
Pig-G	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
Human-I	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
Human-J	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
Pig‡	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Gln-				
	Met-Pro-Leu-Pro-His-Pro-Gly-His-Pro-Gly-Tyr-Ile-Asn-Phe-Ser-Tyr-Glu-Val-Leu-Thr-Pro-Leu-Lys-Trp-Tyr-Glu-				
	30	35	40	45	50
Comp-8	Ser-Met-Ile-Arg-His-Pro-Tyr-Ser-Pro-Tyr-Gly-Tyr-Glu-Pro-Met-Gly-Gly-Trp-Leu-Trp-Gln-Gln-Ile-Pro-Val-				
S-200B	Xaa-Met-Ile-Arg-His-Pro-Tyr-Xaa-Xaa-Tyr-Gly-Tyr-Glu-Pro-Met-Gly-Gly-Trp-Leu-				
LRAP	Ser-Met-Ile-Arg-His-Pro- <u>Pro-Leu</u> † <u>Pro</u> † <u>Pro-Met-Leu-Pro-Asp-Leu-Pro-Leu-Glu-Ala</u>				
TRAP	Ser-Met-Ile-Arg-His-Pro-Tyr-Ser-Pro-Tyr-Gly-Tyr-Glu-Pro-Met-Gly-Gly-Trp				
D	Xaa-Met-Ile-Xaa-Xaa-Pro-Tyr-Xaa-Xaa-Tyr-Gly-Tyr-Glu-Pro-				
Pig-G	<u>Asn</u> † <u>Met-Ile-Arg-His-Pro-Tyr</u> † <u>Thr-Ser</u> † <u>Tyr-Gly</u> † <u>Thr</u> † <u>Glu-Pro-Met-Gly-Gly-Xaa-Leu</u> † <u>His-His-Glu</u> † <u>Ile-Ile-Pro-Val-Val-</u>				
Human-I	Ser-Ile-Xaa-Pro-Pro-Tyr†				
Human-J	Ser-Ile-Met†				
Pig†	Asp-Met-Ile-Xaa-His-Pro-Tyr† <u>Thr-Ser</u> † <u>Tyr-Gly</u> † <u>Thr</u> † <u>Glu-Pro-Met-Gly-Gly-Xaa-Leu</u> † <u>His-His-Glu</u> † <u>Ile-Ile-Pro-Val-Val-</u>				
S-200B†					Leu-Xaa-Gln-Gln-Ile-Ile-Pro-Val-Val-Met-
					5 10

* Abbreviation used: Comp-8, Component-8.

† Fincham et al. (1981b).

‡ Fukae et al. (1980).

Fig. 1. Amino acid sequences for amelogenin polypeptides

The N-terminal sequences for amelogenin proteins isolated from bovine, human and porcine developing dental enamel are shown. Refer to the text and Table 1 for the origin of the individual polypeptides. Sections of sequence which differ from that of the major bovine amelogenin ('Component-8') are indicated by the boxed areas.

fraction is approximated by the difference S-200B minus TRAP. The further observation, that the *N*-terminus of S-200B' appears to begin with Leu-46 of the Component-8 fraction, suggests that the S-200B' is the residue of the 27000 Da amelogenin after specific cleavage of the Trp-45-Leu-46 bond. Confirmation of this suggestion must await further studies.

The longest sequence determined, 53 residues of the bovine Component-8 material, shows this fraction to have the 45-residue TRAP sequence as its *N*-terminus. However, the 46-residue LRAP sequence, distinctive beyond residue 33, was not identified in any other amelogenin; the origin of this major low-molecular-weight amelogenin is unclear, although the presence of the *N*-terminal common 33-residue section suggests that this polypeptide may also be derived from the *N*-terminus of some larger, and as yet unidentified, amelogenin molecule. Alternatively, the LRAP sequence may arise from a different part of the 27000 Da amelogenin structure. Further, it should be noted that an equivalent leucine-rich amelogenin has been isolated from the enamel matrix of other mammalian species (Fincham *et al.*, 1982a).

Comparison of sequences between the three species studied (Fig. 1), together with the data of Fukae *et al.* (1980), shows a striking degree of conservation of structure, with the first 27 residues being identical [although Fukae *et al.* (1980) give residues 14 and 27 as aspartic acid and glutamic acid respectively rather than the amides determined in the present study]. Residue 28 was found to be asparagine in the porcine fraction G, as compared with serine in the bovine and human sequences. The most striking difference of sequence was seen in the human amelogenins, where an inversion of residues 29 and 30 from Met-Ile in the pig and cow to Ile-Met was noted. Further, in the human fraction I the continuation appeared different from that in the pig or cow, with a -Pro-Pro-Tyr- sequence being found in positions 31 to 33.

As has been noted above, the S-200B fraction (M_r 27000) appears as a single band when electrophoresed in SDS-containing systems, but as multiple bands in non-SDS-containing electrophoretic systems, such as the Tris/borate/urea gels used in the present study. Against these observations it is important to note that the *N*-terminal sequence found for the S-200B fraction is identical with that of Component-8. These observations suggest that the multiple electrophoretic components of the S-200B fraction have common *N*-terminal sequences and that the cause of the electrophoretic heterogeneity may be in small variations at the *C*-terminal ends of these polypeptides, or in variations in charged side-chain groupings, as yet undetermined.

The findings of the present study support the view

that the lower-molecular-weight amelogenins arise by a discrete fragmentation of the higher-molecular-weight molecules (Chrispens *et al.*, 1979; Fukae *et al.*, 1980; Fincham *et al.*, 1981a, 1982a,c). Although it has been suggested that this degradation is enzymically mediated (Shimizu *et al.*, 1979), the nature and specificity of such an 'amelogeninase' awaits further clarification. However, inspection of the sequence data presented here, together with previous findings (Fincham *et al.*, 1981b), indicate possible cleavage sites for the bovine amelogenins as being Trp-45-Leu-46 and Gly-43-Gly-44, and in the LRAP sequence, Leu-42-Pro-43.

The high level of conservation of the amelogenin sequence between species suggests a discrete biological function, and although the hydrophobic nature of the *N*-terminal section shows some similarity to 'signal' peptide sequences (Kreil, 1981), the precise role of this sequence still in covalent linkage with the 27000 Da amelogenin of the extracellular matrix has still to be established.

References

- Beckman Instruments (1978) in *The Model 890C Sequencer*, pp. 4-1-4-20, Beckman Instruments Inc., Palo Alto, CA, U.S.A.
- Belcourt, A. B., Fincham, A. G. & Termine, J. D. (1982) *Calcif. Tissue Int.* in the press
- Belew, M., Fohlman, J. & Janson, J. C. (1978) *FEBS Lett.* **91**, 302-304
- Bhown, A., Mole, J. E. & Bennett, J. C. (1981) *Anal. Biochem.* **110**, 355-359
- Chrispens, J., Weliky, B., Bringas, P. & Slavkin, H. C. (1979) *J. Dent. Res.* **58**(B), 988-990
- Eastoe, J. E. (1979) *J. Dent. Res.* **58**(B), 753-763
- Eggert, F. M., Allen, G. A. & Burgess, R. C. (1973) *Biochem. J.* **131**, 471-484
- Fearnhead, R. W. (1979) *J. Dent. Res.* **58**(B), 909-916
- Fincham, A. G. (1971) in *Tooth Enamel II* (Stack, M. V. & Fearnhead, R. W., eds.), pp. 79-87, Wright Bros., Bristol
- Fincham, A. G. (1979) *Calcif. Tissue Int.* **27**, 65-73
- Fincham, A. G., Burkland, G. & Shapiro, I. M. (1972) *Calcif. Tissue Res.* **9**, 247-259
- Fincham, A. G., Belcourt, A. B. & Termine, J. D. (1981a) in *The Chemistry and Biology of Mineralized Connective Tissues* (Veis, A., ed.), pp. 523-529, Elsevier/North-Holland, New York
- Fincham, A. G., Belcourt, A. B., Termine, J. D., Butler, W. T. & Cothran, W. C. (1981b) *Biosci. Rep.* **1**, 771-778
- Fincham, A. G., Belcourt, A. B., Lyaruu, D. M. & Termine, J. D. (1982a) *Calcif. Tissue Int.* **34**, 182-189
- Fincham, A. G., Belcourt, A. B. & Termine, J. D. (1982b) *J. Dent. Res.* in the press
- Fincham, A. G., Belcourt, A. B. & Termine, J. D. (1982c) *Caries Res.* **16**, 64-71

- Fukae, M. & Shimizu, M. (1974) *Arch. Oral. Biol.* **19**, 381–386
- Fukae, M., Tanabe, T., Ijiri, H. & Shimizu, M. (1979) *J. Dent. Res.* **58(B)**, 1000–1001
- Fukae, M., Tanabe, T., Ijiri, H., & Shimizu, M. (1980) *Tsurumi Shigaku (Tsurumi Univ. Dent. J.)* **6**, 87–94
- Inagami, T. & Murakami, K. (1972) *Anal. Biochem.* **47**, 501–504
- Kreil, G. (1981) *Annu. Rev. Biochem.* **50**, 317–348
- Mechanic, G. L. (1971) in *Tooth Enamel II* (Stack, M. V. & Fearnhead, R. W., eds.), pp. 88–92, Wright Bros., Bristol
- Nylen, M. U. (1979) *J. Dent. Res.* **58(B)**, 922–926
- Papas, A., Seyer, J. M. & Glimcher, M. J. (1977) *FEBS Lett.* **79**, 276–280
- Seyer, J. M. & Glimcher, M. J. (1977) *Biochim. Biophys. Acta* **493**, 441–451
- Shimizu, M., Tanabe, T. & Fukae, M. (1979) *J. Dent. Res.* **58(B)**, 782–788
- Tarr, G. E., Beecher, J. F., Bell, M. & McKean, D. J. (1978) *Anal. Biochem.* **84**, 622–627
- Termine, J. D., Belcourt, A. B., Christner, P. J., Conn, K. M. & Nylen, M. U. (1980) *J. Biol. Chem.* **255**, 9760–9768
- Zalut, C., Henzel, W. J. & Harris, H. W. (1980) *J. Biochem. Biophys. Methods* **3**, 11–30