Type II intermediate-filament proteins from wool

The amino acid sequence of component 5 and comparison with component 7c

Lindsay G. SPARROW,* Christine P. ROBINSON, Joanne CAINE,† David T. W. McMAHON‡ and Phillip M. STRIKE

C.S.I.R.O. Division of Wool Technology, Parkville, Vic. 3052, Australia

Component 5 is one of the four type II intermediate-filament proteins found in the hard keratin wool. It was isolated as the *S*-carboxymethyl derivative from Merino wool and its amino acid sequence was determined by manual and automatic sequencing of peptides produced by chemical and enzymic cleavage. Component 5 is an *N*-terminally blocked molecule of 503 residues and M_r (not including the blocking group) of 56600. The blocking group has not been identified. The amino acid sequence of component 5 shows 77% sequence identity with that of component 7c, another type II wool intermediate-filament protein [Sparrow, Robinson, McMahon & Rubira (1989) *Biochem. J.* **261**, 1015–1022]. The sequence similarity extends from the *N*-termini of the two molecules to residue 459 (component 5 sequence); however, there is no recognizable sequence similarity in the remaining *C*-terminal 43 amino acid residues. Details of procedures used in determining the sequence of component 5 have been deposited as a Supplementary Publication SUP 50168 (80 pages) at the British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire, LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1992) **281**, 5. The information comprises: (1) details of chemical and enzymic methods used for cleavage of component 5, peptide CN1, the peptide mixture CN2/3 and various other peptides, (2) details of the procedures used for the fractionation and purification of peptides from (1), including Figures showing the elution profiles from the chromatographic steps used, and (3) details of the method used to determine the *C*-terminal sequence of component 5.

INTRODUCTION

The most complex of the five classes of intermediate-filament proteins are the type I and type II proteins, which interact together to form keratin filaments (Steinert & Roop, 1988). Keratin filaments are found in tissues of epithelial origin and are often divided into two types according to their cystine content. Thus the 'soft' or epidermal keratins, which are cystine-poor, are found in skin and other epithelia, and the cystine-rich 'hard' keratins occur in tissues such as hair and nail. There are approximately 20 different epidermal keratin proteins, though only a limited subset of these is present in any one cell type (Moll *et al.*, 1982). Eight intermediate-filament proteins are found in the hard keratins, wool and hair; four of these are type I intermediate-filament proteins and four are type II intermediate-filament proteins (Heid *et al.*, 1986). For wool, the type I proteins are designated components 8a, 8b, 8c-1 and 8c-2, and the type II proteins are components 5, 7a, 7b and 7c (Crewther *et al.*, 1980).

The intermediate filaments of wool are embedded in an amorphous matrix to which they are cross-linked by disulphide bonds (Fraser *et al.*, 1972). The intermediate-filament proteins can be readily isolated after reduction of these disulphide bonds in disaggregating solutions followed by alkylation of the resulting thiol groups with iodoacetate. Reduced carboxymethylated wool intermediate-filament proteins can be partially fractionated by ion-exchange chromatography in buffers containing 8 M-urea (Crewther *et al.*, 1976). By this means two type I proteins (components 8a and 8c-1) and two type II proteins (components 5 and 7c) have been obtained as single species.

The amino acid sequence of component 8c-1 has been determined by direct protein sequencing (Dowling *et al.*, 1986), and that of a second wool type I intermediate-filament protein (of

M_r 47600) has been deduced from the corresponding genomic clone (Wilson *et al.*, 1988). cDNA clones coding for two different mouse hair type I intermediate-filament proteins have been isolated and characterized by Bertolino *et al.* (1990), who recognized several important relationships among the hair type I intermediate-filament proteins.

As yet the sequence of only one hard keratin type II intermediate-filament protein has been reported, that of component 7c from wool (Sparrow *et al.*, 1989). The present study reports the amino acid sequence of a second type II intermediate-filament protein from wool, component 5, thus allowing a comparison of the two type II components, 5 and 7c. Some of the component 5 sequence has been presented in a previous study (Dowling *et al.*, 1979).

EXPERIMENTAL

General methods

Component 5 was prepared by the method of Crewther *et al.* (1976). Methods used for cleavage of component 5, for enzymic digestions, peptide fractionations, amino acid analysis and sequence determination were as previously described (Sparrow *et al.*, 1989), except for the methods listed below.

CNBr digestion of component 5 and separation of the CNBr-derived peptides

Component 5 was cleaved at methionine residues by CNBr and the resulting peptide mixture was fractionated by gel-filtration chromatography in either buffered 8 M-urea or 70% (v/v) formic acid. Where necessary rechromatography of fractions was carried out by gel filtration, anion-exchange or reverse-phase chromatography.

* To whom correspondence should be addressed.

† Present address: Centre for Protein and Enzyme Technology, LaTrobe University, Bundoora, Vic. 3083, Australia.

‡ Present address: Scot's Presbyterian Church, Godfrey Street, Boort, Vic. 3537, Australia.

Limited acid cleavage of peptide L22

Peptide L22 was dissolved in 0.1 % (v/v) trifluoroacetic acid containing 16 % (v/v) acetonitrile and heated under vacuum at 108 °C for 2.5 h. The hydrolysis mixture was fractionated by reverse-phase h.p.l.c.

C-Terminal analysis of component 5

Component 5 was digested with carboxypeptidase Y in Mes buffer containing SDS. Samples were removed after various incubation times and the enzymic reaction was stopped by the addition of trifluoroacetic acid. The samples were then subjected to amino acid analysis.

RESULTS

The complete amino acid sequence of component 5 is given in Fig. 1, which also shows the peptides used in establishing this sequence as well as the previously published sequence data (Dowling *et al.*, 1979). Each of these peptides was purified and for most of them both amino acid sequence and composition were determined.

Cleavage of component 5 by CNBr gave rise to a mixture of six peptides, CN1, CN2, CN3, CN4, CN5 and CN6, from which CN1, CN4, CN5 and CN6 were obtained as purified species. The remaining two peptides, CN2 and CN3, could not be separated chromatographically and were subjected to sequence analysis as a mixture. The amino acid sequences of peptides CN4 and CN6 have been reported previously (Dowling *et al.*, 1979). For the remaining CNBr-cleavage peptides the sequence determination is summarized below.

Peptide CN1 (residues 1–279)

Peptide CN1, like component 5 itself, yielded no amino acid sequence information on direct sequencing in the spinning-cup Sequenator, indicating that the *N*-terminus is blocked. The peptide, H1 (residues 1–112), obtained by mild acid cleavage of peptide CN1 or of component 5 was isolated by gel-filtration chromatography and reverse-phase h.p.l.c. Its amino acid sequence was obtained by using data from peptides derived by cleavage of peptide H1 with trypsin (peptides T1–T11), chymotrypsin (peptides C1–C12) and thermolysin (peptides Th1–Th9) and of peptide CN1 with thrombin (peptides Throm1–Throm3) and the lysine-specific enzymes Endoproteinase Lys-C and the proteinase from *Achromobacter lyticus* (peptides L1, L2 and EL1). The order of the first two amino acid residues of peptide H1 has not been determined, nor has the nature of the group that blocked direct *N*-terminal sequence determinations of peptide H1, peptide CN1 and component 5.

Sequence analysis of the mixture obtained from acid cleavage of component 5 gave a unique sequence starting at Pro-113 and extending for 25 residues. The remaining sequence of peptide CN1 was obtained from peptides derived from digestion of peptide CN1 with trypsin after citraconylation (peptides T12–T14) and the proteinase from *A. lyticus* (peptides L3–L12). In order to obtain the complete sequence of peptide L8 it was further digested with trypsin to give peptides L8T1, L8T2 and L8T4. The overlap at the acid-cleavage point Asp-112–Pro-113 was found in the sequence of the large thrombin peptide Throm3. The remaining overlaps needed to complete the sequence of peptide CN1 were found in peptides derived from digestion of peptide CN1 with the *Staphylococcus aureus* V8 proteinase (peptides S15–S20) with the exception of those needed to link peptides L5 and L6 and peptides L10 and L11. These four peptides were aligned by comparison with the sequence of the related protein component 7c (see below) and with the consensus

sequences for type II intermediate-filament proteins presented by Conway & Parry (1988).

Peptides CN2 (residues 321–398) and CN3 (residues 410–503)

These peptides were not separated readily and determination of the amino acid sequences was carried out on the mixture. The preparation used for this determination also contained significant amounts of peptide CN4, so that peptides derived from this already known sequence were found in enzyme digests.

From the tryptic digest of peptides CN2/CN3 most of the peptides (peptides T15–T35) from Arg-283 to the *C*-terminus were isolated. Peptides T15 and T27, which were isolated in poor yield, both spanned methionine residues, showing that cleavage by CNBr at methionine was incomplete. A similar observation was made during the sequence determination of component 7c (Sparrow *et al.*, 1989). Most of the overlaps needed to order the tryptic-digest peptides were found in peptides from digestion of the CN2/CN3 mixture with the *S. aureus* V8 proteinase (peptides S1–S14). This digest also included peptides from peptide CN4, though these were not further characterized. Two useful peptides in this series, S13 and S14, arose from cleavage after serine residues rather than glutamic acid residues as normally expected for this enzyme.

Cleavage of peptides CN2/CN3 at lysine with the *A. lyticus* proteinase yielded peptides L15–L22, though only for the large *C*-terminal peptide L22 (residues 413–503) was full characterization necessary. The complete sequence of peptide L22 was deduced from the sequences of peptides produced by further digestion with chymotrypsin (L22C1–L22C5) and thermolysin (L22Th) and by hydrolysis with acid under mild conditions (L22H). Peptides L9–L14 were isolated in low yield from this digest. They are derived from peptide CN1 and probably arise through contamination of the mixture CN2/CN3 with traces of peptide CN1 or more likely the *C*-terminal acid cleavage peptide of CN1 produced during the CNBr digestion. Digestion with the *A. lyticus* proteinase also yielded a number of peptides, not shown in Fig. 1, resulting from cleavage at residues other than lysine. A number of these anomalous cleavages were at alanine residues, e.g. at positions 364, 365, 375, 452, 462 and 467. There was also evidence for cleavage after serine residues (490 and 496), after arginine residues (304 and 498) and after a glutamine residue at position 431.

C-Terminus of component 5

The *C*-terminal sequence was shown to be -Phe-Ala by digestion of component 5 with carboxypeptidase Y. This evidence was supported by the isolation of a peptide Phe-Ala (T35) from the tryptic digest of peptides CN2/CN3.

Order of the CNBr-derived peptides

The six CNBr-derived peptides were ordered by using the sequence similarity between component 5 and component 7c (Dowling *et al.*, 1979) (see below). Two methionine-containing peptides, T15 and T27, were isolated from the tryptic digest of peptides CN2/CN3; their sequences confirmed the proposed order for the CNBr-derived peptides, CN5–CN4 and CN6–CN3 respectively.

Discrepancies between the sequence presented in Fig. 1 and the published sequence data (Dowling *et al.*, 1979) were found at positions 32, 124, 134, 306, 318 and 319. These are further discussed in Supplementary Publication SUP 50168.

DISCUSSION

Aspects of the sequence analysis

During the course of this work two different lysine-specific

Fig. 1. Amino acid sequence of component 5

Designation of peptides : —→, sequence only; —, sequence and composition; ----, composition only; ·····, previously published (Dowling *et al.*, 1979) (no further experimental detail is given for peptides from the previous publication); T, tryptic; C, chymotryptic; Th, thermolytic; S, *S. aureus* V8 proteinase; L, lysine-specific proteinase; EL, anomalous cleavage by lysine-specific proteinase; Throm, thrombin; H, acid cleavage; CN, CNBr. Peptides from these digests are numbered in order of their occurrence in the complete sequence.

proteinases, Endoproteinase Lys-C (from *Lysobacter enzymogenes*) and the proteinase from *A. lyticus* (Masaki *et al.*, 1978), were used. Both proteinases, though being largely lysine-specific, showed evidence for cleavage at residues other than lysine. Digests with both proteinases yielded a number of peptides resulting from cleavage after arginine residues, and, in addition, the *A. lyticus* proteinase cleaved after some alanine and serine residues. Data for only one of the peptides resulting from these anomalous cleavages (EL1) were used in establishing the sequence of component 5.

Peptides covering the entire sequence shown in Fig. 1, and most of the necessary overlapping peptides, were isolated either in this or the previous study (Dowling *et al.*, 1979). Four overlaps were assigned by using the sequence similarity of component 5 with component 7c (see below) and with a consensus sequence for type II intermediate-filament proteins. These assignments are justified on the grounds that all four sequences are within the central rod-like region of the intermediate-filament molecule. These regions are highly conserved in sequence and are of invariant size (Steinert & Roop, 1988). Sequencing of the *N*- and *C*-terminal segments, which are not highly conserved in intermediate-filament proteins, was facilitated by the ability to isolate large peptides covering these two segments. The peptide, H1, covering the *N*-terminal 112 residues was obtained by mild acid cleavage of the bond Asp-112-Pro-113, and the *C*-terminal 91 residues were obtained as the single peptide, L22, from cleavage at Lys-412 with the *A. lyticus* proteinase. The sequences

of peptides from a number of different enzyme digests were needed to assign sequences unambiguously to peptides H1 and L22.

The order of the first two residues of component 5 has not been established, nor has the nature of the *N*-terminal blocking group. The sequence in this region of the molecule was deduced from the amino acid compositions of the blocked chymotryptic-digest peptide C1 and the blocked *A. lyticus*-proteinase-digest peptide L1 together with the sequences of the tryptic-digest peptide T1 and the thrombin-derived peptide Throm1.

In Table 1 the amino acid compositions of component 5 and the two CNBr-cleavage peptides purified and sequenced in this study, CN1 and CN5, are compared with values calculated from the sequence shown in Fig. 1. The M_r calculated from this sequence is 56600 (not including the blocking group); this is in good agreement with the value of 57000 determined by low-speed sedimentation equilibrium (Woods, 1979) and by gel-filtration chromatography (Blagrove & Frenkel, 1977).

Analysis of the sequence

Analysis of the sequence of component 5 shows that it fits readily into the pattern characteristic of intermediate-filament proteins (Steinert & Roop, 1988) in having a central rod-like helical domain that is divided by three non-helical linking segments, L1, L12 and L2, into four helical tracts, 1A, 1B, 2A and 2B; each of the helical tracts shows the heptad repeat characteristic of potential coiled-coil molecular structures (Crick,

Table 1. Amino acid composition of component 5 and some peptides derived by cleavage with CNBr

Values in parentheses were calculated from the sequence shown in Fig. 1. Tryptophan was determined separately. Values for peptide CN1 are experimental data from Dowling *et al.* (1979). Abbreviations: CmCys, *S*-carboxymethylcysteine; Hse, homoserine.

Amino acid	Amino acid composition (mol of residue/mol)		
	Component 5	Peptide CN1	Peptide CN5
CmCys	20.6 (21)	11.7 (12)	
Asp	40.7 (39)	23.4 (22)	4.0 (4)
Thr	21.1 (21)	10.6 (11)	
Ser	45.8 (52)	27.1 (30)	1.1 (1)
Glu	71.9 (71)	37.1 (36)	
Pro	13.6 (11)	8.9 (8)	
Gly	37.7 (35)	20.9 (19)	
Ala	47.3 (47)	21.2 (20)	
Val	31.7 (35)	19.3 (21)	
Met or Hse	4.5 (5)	0.8 (1)	1.1 (1)
Ile	21.1 (22)	10.9 (12)	
Leu	43.8 (43)	26.2 (25)	1.2 (1)
Tyr	17.6 (16)	11.4 (11)	
Phe	14.6 (13)	11.4 (11)	
Lys	26.7 (26)	14.5 (15)	
His	6.0 (5)	3.6 (4)	
Arg	38.2 (39)	20.0 (20)	1.3 (1)
Trp	2.5 (2)	0.8 (1)	

1953). The structure of component 5 suggested by this analysis is shown schematically in Fig. 2, which also depicts two segments, H1 and H2, that are highly conserved in a number of intermediate-filament classes (Steinert & Parry, 1985).

An earlier study showed that component 5 shares regions of sequence similarity with component 7c, another type II wool intermediate-filament protein (Dowling *et al.*, 1979). Comparison of the entire sequences of these two proteins now shows that the sequence similarity extends almost from the *N*-termini of the two molecules through to residue 459 (component 5 sequence). This is illustrated in Table 2 part (a), in which the sequences of the two proteins are compared, segment by segment, and in Fig. 3, where

Fig. 2. Schematic representation of the structure of component 5

The scale is in amino acid residues. For an explanation of the symbols see the text.

the sequences of the *N*-terminal non-helical regions are compared. To maximize the sequence similarity of the *N*-terminal segments (Fig. 3) it is necessary to introduce a number of gaps in the component 7c sequence and one in the component 5 sequence. In the *C*-terminal segments of the molecules (beyond residue 460 of component 5) no sequence similarity is found. Like component 7c, component 5 also shows sequence similarities with other type II intermediate-filament proteins when compared with a consensus sequence for type II intermediate-filament proteins (Conway & Parry, 1988) (results not shown). The high degree of sequence similarity between component 5 and other type II intermediate-filament proteins, particularly component 7c, was used to deduce the order of a number of peptides during the sequence determination. Thus it was not considered necessary to isolate overlapping peptides for a number of positions in the central region of component 5, where the sequence similarity is highest.

The amino acid sequences of four hard keratin type I intermediate-filament proteins are known, two from wool and two from mouse hair. A segment-by-segment comparison of the two wool proteins, component 8c-1 (Dowling *et al.*, 1986) and an unnamed 47600-*M_r* protein (Wilson *et al.*, 1988), is shown in Table 2 part (b), and the two mouse hair proteins, MHKA-1 and MHKA-2 (Bertolino *et al.*, 1990), are compared in Table 2 part (c).

Table 2. Segment-to-segment comparison of amino acid sequences of intermediate-filament proteins from hard keratins

For an explanation of the symbols refer to Fig. 2. For each pair, the first two rows give the positions of the various segments within the proteins and the third row gives the number of sequence identities between the two proteins; where gaps were introduced to maximize sequence identity these are scored as non-identity. Note that a segment H2 is not recognized in the type I molecules (Conway & Parry, 1988). (a) Type II proteins from wool; (b) type I proteins from wool; (c) type I proteins from mouse hair.

	N	H1	1A	L1	1B	L12	2A	L2	2B	H2	C	Total
(a)												
Component 5	1-91	92-122	123-157	158-167	168-268	269-285	286-304	305-312	313-433	434-447	448-503	1-503
Component 7c	1-78	79-109	110-144	145-154	155-255	256-272	273-291	292-299	300-420	421-434	435-491	1-491
Sequence identities	58/91	28/31	31/35	8/10	86/101	16/17	18/19	8/8	113/121	13/14	8/56	387/503
(b)												
Component 8c-1	1-48	49-55	56-90	91-101	102-202	203-218	219-237	238-245	246-366	-	367-412	1-412
47600 <i>M_r</i> protein	1-48	49-55	56-90	91-101	102-202	203-218	219-237	238-245	246-366	-	367-403	1-403
Sequence identities	46/48	6/7	34/35	11/11	93/101	16/16	18/19	8/8	117/121	-	28/46	377/412
(c)												
MHKA-1	1-48	49-55	56-90	91-101	102-202	203-218	219-237	238-245	246-366	-	367-416	1-416
MHKA-2	1-48	49-55	56-90	91-101	102-202	203-218	219-237	238-245	246-366	-	367-392	1-392
Sequence identities	44/48	7/7	35/35	11/11	89/101	16/16	18/19	8/8	117/121	-	14/25*	

* Based on the shorter sequence.

		10
Component 5	X-(Ser.Cys)-Arg- Ser -Tyr-Arg-Ile- Ser -Pro-Gly-Tyr- Ser -Val-	
Component 7c	X-(Gly.Cys)-Phe- Ser -Thr-Val-Gly- Ser -Gly-Phe-Gly- Ser -----	
		20
Component 5	Thr- Arg -Thr-Phe- Ser -Ser-Cys- Ser -Ala-Val-Ala-Pro-Lys-Thr-	
Component 7c	---- Arg -Ala-Phe- Ser -Cys-Val- Ser -Ala-Cys-Gly-Pro-Arg-Pro-	
		30
Component 5	Gly-Ser- Arg -Cys-Cys-Ile-Ser-Ala-Ala-Pro-Tyr-Arg-Gly-Val-	40
Component 7c	Gly----- Arg -Cys-Cys-Ile-Thr-Ala-Ala-Pro-Tyr-Arg-Gly-Ile-	
		50
Component 5	Ser -Cys-Tyr-Arg-Gly-Leu-Thr-Gly-----Phe-Gly- Ser -Arg- Ser -	
Component 7c	Ser -Cys-Tyr-Arg-Gly-Leu-Thr-Gly-Gly-Phe-Gly- Ser -Arg- Ser -	
		60
Component 5	Val-Ser-Ala-Leu-Gly-Ser- Cys -Gly-Pro-Arg-Ile-Ala-Val-Ser-	
Component 7c	Val-----Cys-Gly-----	
		70
Component 5	Gly-Phe- Arg -Ala-Gly-Ser-Cys-Gly-Arg- Ser -Phe-Gly-Tyr-Arg-	80
Component 7c	Gly-Phe- Arg -Ala-Gly-Ser-Cys-Gly-Arg- Ser -Phe-Gly-Tyr-Arg-	
		90
Component 5	Ser -Gly-Gly-Val-Gly-Gly-Leu- Ser -Pro-	
Component 7c	Ser -Gly-Gly-Val-Gly-Gly-Pro- Ser -Pro-	

Fig. 3. Comparison of the amino acid sequences of the *N*-terminal non-helical segments of type II wool intermediate-filament proteins, components 5 and 7c

Bold-face letters denote amino acid residues identical in both sequences. Numbers refer to the position in the sequence of component 5.

Major differences within the type I protein pairs are seen only in the *C*-terminal non-helical segments, which are different in size but show significant sequence similarity (greater than 50% identity). The wool type II proteins have *C*-terminal non-helical segments almost identical in length (56 and 57 residues) but quite different in amino acid sequence (14% identity) and *N*-terminal non-helical segments different in length (91 and 78 residues) but similar in sequence (64% identity).

The data presented above, showing a different pattern of variation between the non-helical *N*- and *C*-terminal segments of the type I and type II intermediate-filament proteins of hard keratins, suggests a different functional role in the filament for these two types of intermediate-filament proteins. Bertolino *et al.* (1990) have also suggested that distinct functional requirements are satisfied by the multiplicity of hair keratins on the basis of their demonstration that there is differential expression of individual type I intermediate-filament proteins in mouse hair.

We thank Dr. C. M. Roxburgh for performing the amino acid analyses and Mr. L. M. Dowling for many helpful discussions during the course of this work.

REFERENCES

- Bertolino, A. P., Checkla, D. M., Heitner, S., Freedberg, I. M. & Yu, D.-W. (1990) *J. Invest. Dermatol.* **94**, 297–303
- Blagrove, R. B. & Frenkel, M. J. (1977) *J. Chromatogr.* **132**, 399–404
- Conway, J. F. & Parry, D. A. D. (1988) *Int. J. Biol. Macromol.* **10**, 79–88
- Crewther, W. G., Dowling, L. M., Gough, K. H., Inglis, A. S., McKern, N. M., Sparrow, L. G. & Woods, E. F. (1976) *Proc. Int. Wool Text. Res. Conf.* 5th 1975, vol. 2, pp. 233–242
- Crewther, W. G., Dowling, L. M., Gough, K. H., Marshall, R. C. & Sparrow, L. G. (1980) in *Fibrous Proteins: Scientific, Industrial and Medical Aspects* (Parry, D. A. D. & Creamer, L. K., eds.), vol. 2, pp. 151–159, Academic Press, London and New York
- Crick, F. H. C. (1953) *Acta Crystallogr.* **6**, 689–697
- Dowling, L. M., Gough, K. H., Inglis, A. S. & Sparrow, L. G. (1979) *Aust. J. Biol. Sci.* **32**, 437–442
- Dowling, L. M., Crewther, W. G. & Inglis, A. S. (1986) *Biochem. J.* **236**, 695–703
- Fraser, R. D., B., MacRae, T. P. & Rogers, G. E. (1972) in *Keratins: Their Composition, Structure and Biosynthesis* (Kugelmass, I. N., ed.), Charles C. Thomas, Springfield
- Heid, H. W., Werner, E. & Franke, W. W. (1986) *Differentiation* **32**, 101–119
- Masaki, T., Nakamura, K., Isono, M. & Soejima, M. (1978) *Agric. Biol. Chem.* **42**, 1443–1445
- Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. (1982) *Cell* **31**, 11–24
- Sparrow, L. G., Robinson, C. P., McMahon, D. T. W. & Rubira, M. R. (1989) *Biochem. J.* **261**, 1015–1022
- Steinert, P. M. & Parry, D. A. D. (1985) *Annu. Rev. Cell Biol.* **1**, 41–65
- Steinert, P. M. & Roop, D. R. (1988) *Annu. Rev. Biochem.* **57**, 593–625
- Wilson, B. W., Edwards, K. J., Sleigh, M. J., Byrne, C. R. & Ward, K. A. (1988) *Gene* **73**, 21–31
- Woods, E. F. (1979) *Aust. J. Biol. Sci.* **32**, 423–435

Received 24 June 1991; accepted 9 July 1991