The complement component C4 of mammals

Alister W. DODDS* and S. K. Alex LAW

MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Human complement component C4 is coded by tandem genes located in the HLA class III region. The products of the two genes, C4A and C4B, are different in their activity. This difference is due to a degree of 'substrate' specificity in the covalent binding reactions of the two isotypes. Mouse also has a duplicated locus, but only one gene produces active C4, while the other codes for the closely related sex-limited protein (Slp). In order to gain some insight into the evolutionary history of the duplicated C4 locus, we have purified C4 from a number of other mammalian species, and tested their binding specificities. Like man, chimpanzee and rhesus monkey appear to produce two C4 types with reactivities similar to C4A and C4B. Rat, guinea pig, whale, rabbit, dog and pig each expresses C4 with a single binding specificity, which is C4B-like. Sheep and cattle express two C4 types, one C4B-like, the other C4A-like, in their binding properties. These results suggest that more than one locus may be present in these species. If this is so, then the duplication of the C4 locus is either very ancient, having occurred before the divergence of the modern mammals, or there have been three separate duplication events in the lines leading to the primates, rodents and ungulates.

INTRODUCTION

C4 is a component of the C3 and C5 convertases of the classical complement pathway (Reid & Porter, 1981; Müller-Eberhard, 1988). Part of its function is to localize the damage caused by complement to the immediate vicinity of the site of activation. This is accomplished by the presence, within the newly activated molecule, of a binding site which has an extremely short half-life. This binding site in the native molecule takes the form of a thioester bond between a Cys and a Gln residue located within the α -chain (Law, 1983; Tack, 1985). Upon activation of C4 to C4b by C1s, the bond becomes exposed and is reactive. Much of the nascent C4b reacts with water and is inactivated. However, a proportion of the C4b reacts with structures on the activating surface and becomes covalently bound (Campbell *et al.*, 1980; Law *et al.*, 1980).

In human there are two isotypes of C4 which are coded by closely linked genes located within the HLA Class III region (Campbell *et al.*, 1988). These two proteins have α -chains of apparently different molecular mass when run on SDS/PAGE with low levels of cross-linker in the gels (Roos *et al.*, 1982). They are also different in their reactivity with different chemical structures. C4A is more reactive than C4B with amino (Law *et al.*, 1984a; Isenman & Young, 1984) and thiol (Sim *et al.*, 1989) groups, while C4B binds more rapidly than C4A to hydroxyl groups (Law *et al.*, 1984a; Isenman & Young, 1984). These differences can be seen very clearly in the rates of reaction of the two C4 types with the small molecules glycine and glycerol, following activation by C1s (Dodds *et al.*, 1985). This difference in reactivity with different structures is reflected in the specific haemolytic activities of the two C4 isotypes. C4B can react more efficiently than can C4A with the erythrocyte surface, which is rich in hydroxyl groups, and is therefore more haemolytically active. We have recently proposed that the difference in

reactivity between the two human C4 isotypes is due largely to a single amino acid substitution at position 1106 in the primary structure. In C4B this is a His residue while an Asp residue is found in C4A (Dodds & Law, 1988). Recently, using site-directed mutagenesis, Carroll *et al.* (1989) were able to express C4 molecules with different residues in the isotypic region. They demonstrated that the Asp/His difference at position 1106 could indeed account for the C4A and C4B levels of specific haemolytic activity, although substitutions at other positions could also affect, to a lesser extent, the reaction.

The mouse also has two closely linked genes located in the H2 Class III region which are C4-like (Campbell *et al.*, 1988). Only one of these genes produces an active C4 protein. The other codes for the closely related sex-limited protein (Slp) (Passmore & Shreffler, 1970), which has no known function (Sepich *et al.*, 1987). Both of these proteins have a His residue at the position proposed to be involved in conferring binding specificity (Sepich *et al.*, 1985; Nonaka *et al.*, 1985; Ogata & Sepich, 1985). The binding specificity of mouse C4 is C4B-like (Dodds & Law, 1988). The overall sequence identity between human and mouse C4 is about 77% (Nonaka *et al.*, 1985). Human C4A and C4B are over 99% identical (Belt *et al.*, 1984, 1985) and mouse C4 and Slp show approximately 94% identity (Ogata & Sepich, 1985). These sequence data suggest that the two duplications were separate events which happened after the divergence of the two species.

Here we present data on the covalent binding specificities of C4 from a number of other mammalian species. These indicate that, like man, chimpanzee and rhesus monkey appear to express two C4 types with binding specificities similar to C4A and C4B. Of the other species studied, rat, guinea pig, rabbit, dog, pig and whale are similar to mouse in that they produce C4 with only a single binding specificity, which is C4B-like in its

Abbreviations used: DFP, di-isopropyl fluorophosphate; EACA, ϵ -aminocaproic acid (6-aminohexanoic acid); PMSF, phenylmethanesulphonyl fluoride; Slp, sex-limited protein.

* To whom correspondence and reprint requests should be addressed.

reactivity. Cattle and sheep however were found to have two C4 types whose binding properties closely resemble those of human C4A and C4B.

METHODS

Plasma samples

Whole blood was collected into 100 mM-EDTA, pH 7.4, to give a final EDTA concentration of 10 mM. Cells were separated by centrifugation and the plasma was either used immediately or stored at -70°C .

Chimpanzee plasma was a gift from Dr. Gunnar Baatrup, Odense, Denmark. Rhesus monkey plasma was from Dr. Carolyn Giles, Hammersmith Hospital, London. Plasma from the fin whale *Balaenoptera physalus* was a gift from Dr. Rami Spilliaert, Reykjavik, Iceland. Dog plasma was from Professor R. L. Dawkins, Perth, Australia. Guinea pig, rat and rabbit plasma were from colonies in the Department of Biochemistry, University of Oxford. Bovine plasma (Gloucester) was supplied by Mr. Alan Hayward, Swinford Farm, Eynsham, Oxford. Sheep plasma was from Ms. Libby Henson of the Cotswold Farm Park, Guiting Power, Gloucestershire.

Purification of proteins

The allotypes C4A3 and C4B1 of human C4 were purified from the plasma of an individual with the phenotype C4A3B1 by affinity chromatography using a monoclonal antibody (LOO3) which has different affinities for the two C4 isotypes (Dodds *et al.*, 1985; Hsiung *et al.*, 1987). The proteins were concentrated, and trace contaminants removed by anion-exchange chromatography on Mono Q (Pharmacia) (Dodds & Law, 1988). Chimpanzee C4 was purified by the method used for human C4.

C4 from the other species was not fully purified, but was obtained functionally pure as follows. All columns were run at room temperature, but fractions were collected on to ice. Plasma (5–10 ml), fresh or stored at -70°C , was made 2.5 mM with di-isopropyl fluorophosphate (DFP). In the case of bovine plasma, it was made 25 mM with benzamidine in addition to 2.5 mM with DFP before being centrifuged at 5000 *g* for 10 min to remove debris. The plasma was loaded at 2 ml/min on to a column (1 cm diam. $\times$ 10 cm) of Q Sepharose Fast Flow (Pharmacia) equilibrated with 10 mM-Tris/50 mM- ϵ -aminocaproic acid (EACA)/5 mM-EDTA/0.2 mM-phenylmethanesulphonyl fluoride (PMSF)/0.02 % sodium azide/100 mM-NaCl, pH 7.5. Bound protein was eluted at 2 ml/min with a 50 ml linear gradient from starting buffer to a final buffer of the same composition but with 500 mM-NaCl; 2 ml fractions were collected. The elution position of the C4 protein was determined by SDS/PAGE (Roos *et al.*, 1982) of the fractions, and in some instances also by haemolytic assay of the fractions (Law *et al.*, 1980). The C4-containing fractions were pooled, diluted with 0.5 vol. of water and made 2.5 mM with DFP. The C4 pool was loaded at 1 ml/min on to a column (0.5 cm diam. $\times$ 5 cm) of Mono Q (Pharmacia) equilibrated with the same buffer used for Q Sepharose. Bound proteins were eluted at 1 ml/min with a 20 ml linear gradient to 500 mM-NaCl in the same buffer; 0.5 ml fractions were collected. The C4-containing fractions were assayed by SDS/PAGE, made 2.5 mM with DFP and dialysed into 10 mM-sodium phosphate/140 mM-NaCl/1 mM-EDTA, pH 7.4, and stored at 4°C .

Human C1s was purified by the method of Gigli *et al.* (1976).

Binding reactions

[2- ^3H]Glycine (15 Ci/mmol), [U- ^{14}C]glycine (113 mCi/mmol) and [^3H]methylamine (28.9 Ci/mmol) were purchased from Amersham International and [2- ^3H]glycerol (200 mCi/mmol) was from NEN Research Products. The concentration of active C4 was determined by the reaction of [^3H]methylamine (200 mCi/mmol) with the intact thiolester bond (Law *et al.*, 1984b). The covalent binding of the proteins to [^3H]glycine (at 2.5 and 0.05 mM, 200 mCi/mmol) and [^3H]glycerol (at 10 mM, 200 mCi/mmol) was determined in 10 mM-sodium phosphate/140 mM-NaCl/1 mM-EDTA, pH 7.4, using human C1s to activate the C4.

Binding efficiency (BE), defined as the fraction of active C4 (determined by [^3H]methylamine binding) which bound to radioactive small molecules on activation by C1s, was determined for glycine and glycerol. Reaction rate ratio k'/k_0 , where k_0 is the first-order reaction rate of activated C4b with water and k' is the second-order reaction rate of activated C4b with the radiolabelled small molecule, was calculated by the equation:

$$k'/k_0 = \frac{\text{BE}}{[\text{S}](1 - \text{BE})}$$

where [S] is the concentration of the radioactive small molecule (Law *et al.*, 1984b).

Double-labelling experiments were performed by activating C4 with C1s in the presence of [^3H]glycerol (10 mM, 200 mCi/mmol) and [^{14}C]glycine (0.05 mM, 113 mCi/mmol). The samples were reduced, run on SDS/PAGE (Roos *et al.*, 1982) and stained with Coomassie Blue. The gel lanes were cut into six slices of approx. 1 mm each, centred on the α' -chain of the C4. The individual slices were incubated at 60°C for 2 h with 2 ml of NCS tissue solubilizer (Amersham International). Toluene scintillant (0.5 % PPO/0.03 % dimethyl-POP) was added and the radioactivity determined, in an LKB 1211 Minibeta scintillation counter, in spectral windows 008–060, which contained 73 % of the total ^3H counts and 9 % of the total ^{14}C counts, and 090–165, which contained 0.1 % of total ^3H counts and 64 % of total ^{14}C counts (Dodds *et al.*, 1986). The total ^3H and ^{14}C in each lane, and the percentage of the total in each slice, were calculated accordingly.

RESULTS

Purification of chimpanzee C4 by affinity chromatography

DFP-treated chimpanzee plasma (5 ml) was loaded on to the LOO3 column at neutral pH. The bound C4 was eluted with a 40 ml pH gradient from pH 8.5 to pH 11.5. Chimpanzee C4 was eluted as a single peak at approx. pH 9.0, intermediate in position between those normally occupied by human C4A and C4B. SDS/PAGE analysis indicated the presence of an α -chain doublet in all fractions, with no separation of the two types across the chromatographic peak. The C4 was concentrated and trace contaminants were removed by chromatography on Mono Q.

Fig. 1. Elution of guinea pig C4 from Q Sepharose

(a) Guinea pig plasma was loaded on to a column (1 cm $\times$ 10 cm) of Q Sepharose Fast Flow equilibrated with 10 mM-Tris/50 mM-EACA/5 mM-EDTA/0.2 mM-PMSF/0.02% sodium azide/100 mM-NaCl, pH 7.5, and eluted with a 50 ml linear gradient to 500 mM-NaCl in the same buffer. (b) The numbered fractions were analysed by SDS/PAGE and the C4-containing fractions were pooled as shown.

Purification of C4 from other species by ion-exchange chromatography

C4 was purified from human, rhesus monkey, rat, guinea pig, rabbit, dog, pig, fin whale, sheep and bovine plasma by ion-exchange chromatography on Q Sepharose and Mono Q. C4 is one of the most negatively charged plasma proteins, and in all cases was one of the last proteins to be eluted from the anion-exchange columns. As all of the preparations were very similar, only the preparation of guinea pig C4 will be described

Fig. 2. Elution of guinea pig C4 from Mono Q

(a) The pooled guinea pig C4 from Q Sepharose was diluted with 0.5 vol. of water and loaded on to a column (0.5 cm $\times$ 5 cm) of Mono Q equilibrated with 10 mM-Tris/50 mM-EACA/5 mM-EDTA/0.2 mM-PMSF/0.02% sodium azide/100 mM-NaCl, pH 7.5, and eluted with a 20 ml linear gradient to 500 mM-NaCl in the same buffer. (b) The numbered fractions were analysed by SDS/PAGE and C4-containing fractions were pooled as shown.

in detail; variations found with the other species will be discussed in later sections.

Fresh guinea pig EDTA plasma (5 ml) was run on a Q Sepharose column as described in the Methods section. The elution profile is shown in Fig. 1(a) and SDS/PAGE analysis of the fractions in Fig. 1(b). The bulk of the plasma proteins either did not bind to the column or else were eluted early in the salt gradient. Guinea pig C4 was eluted in the final peak in fractions 9–18. The C4-containing pool (volume of 20 ml) was diluted with 10 ml of water and made 2.5 mM with DFP. The diluted sample was loaded on to a Mono Q column and eluted by salt gradient as described in the Methods section. The elution profile is shown in Fig. 2(a) and SDS/PAGE analysis of the fractions in Fig. 2(b). C4 was eluted as a single peak

Table 1. Elution positions and reactivity with small 'substrates' of mammalian C4 molecules

C4 type	Elution position from Mono Q (mM-NaCl)	Reaction rate (k'/k_0)	
		Glycine	Glycerol
Human C4A	490	13400	1.3
Human C4B	490	119	15.5
Chimpanzee	490	5319	9.5
Rhesus monkey	440	4090	12.3
Mouse*	—	136	26.0
Rat	440	133	19.0
Guinea pig	450	255	22.5
Rabbit	440	100	17.6
Dog	430	280	18.2
Pig	430	235	20.0
Fin whale I	390	652	11.1
Fin whale II	470	575	13.6
Sheep	480	21600	4.1
Bovine C4A	440	17700	2.0
Bovine C4B	480	126	17.6

* From Dodds & Law (1988)

between 430 and 470 mM-NaCl with a maximum at 450 mM-NaCl.

C4 from the other species studied was eluted from the Q Sepharose and Mono Q columns in positions similar to that observed with guinea pig C4. The salt con-

centration at which each C4 type was eluted from the Mono Q column is shown in Table 1.

SDS/PAGE analysis of some of the C4 types studied is shown in Fig. 3. Tracks 1 and 2 show human C4A and C4B prepared, by ion-exchange chromatography on Q Sepharose and Mono Q, from the plasma of individuals who express only a single C4 type. The apparent difference in the molecular mass of the α -chains can be seen. In both cases one major contaminant, running near to the top of the gel, and a number of minor contaminants are evident. The major contaminant is probably inter- α -trypsin inhibitor. Track 3 shows rhesus monkey C4b which contains a number of contaminants which run between the β - and γ -chains. The C4 used for this gel was an old sample which had become activated on storage. Track 4 shows rat C4. Tracks 5 and 6 show two forms of fin whale C4, both of which are functionally active but which could be separated by ion-exchange chromatography. Track 7 is guinea pig C4b which has become activated during storage. Track 8 shows partially activated sheep C4. Four bands are visible in the α -chain region; the upper pair of bands are the α -chains of unactivated C4 and the lower pair of bands are the α' -chains of C4b. There is considerably more of the higher molecular mass C4A-like C4 than of the C4B-like C4. The approximate apparent molecular masses of the α , β and γ chains of the various C4 types studied are shown in Table 2.

Properties of the C4 purified from different species

Each of the C4 types studied was tested by methylamine incorporation for the presence of an intact thiolester bond. All were found to be active by this criterion and also by their ability to be cleaved by human C1s and to bind to the small 'substrate' molecules glycine and glycerol. The reaction rates with glycine and glycerol of the various C4 types tested are summarized in Table 1.

Primate C4

Chimpanzee C4, when analysed by SDS/PAGE, showed the presence of two α -chains as in human C4 (see Fig. 4b). The reaction rates of the chimpanzee C4 with glycine and glycerol were determined and found to be

Fig. 3. SDS/PAGE analysis of partially purified C4

C4 was purified by ion-exchange chromatography on Q Sepharose and Mono Q. Samples were reduced and analysed by SDS/PAGE. Lanes: 1, human C4A; 2, human C4B; 3, rhesus monkey C4b; 4, rat C4; 5, fin whale C4-I; 6, fin whale C4-II; 7, guinea pig C4b; 8, sheep C4 and C4b.

Table 2. Apparent molecular masses of mammalian C4 polypeptide chains

C4 type	Apparent molecular mass (kDa)		
	α -Chain	β -Chain	γ -Chain
Human C4A	96	75	30
Human C4B	94	75	30
Chimpanzee	94/96	77	30
Rhesus monkey	95	75	29
Mouse	91	68	34
Rat	90	69	28
Guinea pig	92/94	67	25
Rabbit	92	78	27
Dog	90	74	26
Pig	83	78	26
Fin whale I	95	76	27
Fin whale II	85	76	27
Sheep	91/93	70	26
Bovine C4A	88	71	25
Bovine C4B	92	71	25

similar to those which would have been obtained with a mixture of human C4A and C4B (Table 1). In order to test whether the two α -chains observed on SDS/PAGE reacted differently with glycine and glycerol, chimpanzee C4 was activated in the presence of [^{14}C]glycine and

[^3H]glycerol. After SDS/PAGE the gel was sliced in the region around the α '-chains and counted in windows specific for the two isotopes. The result is shown in Fig. 4: panel (a) shows a mixture of human C4A and C4B, and panel (b) shows chimpanzee C4. In both cases there is a separation of the ^3H and ^{14}C . The [^{14}C]glycine radioactivity migrates with the apparently higher molecular mass C4A α '-chain while the [^3H]glycerol is associated with the lower molecular mass C4B α '-chain. This indicates that, like man, chimpanzee has two C4 types with different glycine and glycerol binding specificities. Unlike human C4, however, the chimpanzee C4 isotopes do not show differential binding to the LOO3 monoclonal antibody.

The binding reactions of rhesus monkey C4 with glycine and glycerol were similar to those of the chimpanzee C4, and intermediate between human C4A and C4B (Table 1). SDS/PAGE analysis of rhesus monkey C4 showed only one band for the α -chain (Fig. 3, lane 3) and a double labelling experiment failed to demonstrate any separation of the glycine and glycerol binding components (Fig. 4c). Whilst it appears that rhesus monkey has only one C4 with intermediate reactivity to glycine and glycerol, based on its relation to man and chimpanzee, it is more likely that the α '-chain of its C4 isotopes are not resolved by SDS/PAGE. The residues that account for the difference in apparent molecular mass of the α -chains of human C4A and C4B also reside

Fig. 4. Binding of [^3H]glycerol and [^{14}C]glycine to C4

C4 samples were activated by CIs in the presence of [^3H]glycerol and [^{14}C]glycine. The samples were reduced and run on SDS/PAGE (Roos *et al.*, 1982). After staining with Coomassie Blue the gel was sliced into six approximately 1 mm slices around the α '-chain of the activated C4. The individual slices were counted for ^3H (●) and ^{14}C (○) radioactivity. (a) Human C4A and C4B; (b) chimpanzee C4; (c) rhesus monkey C4; (d) guinea pig C4; (e) sheep C4; (f) bovine C4.

Fig. 5. SDS/PAGE analysis of bovine C4 from Mono Q

Bovine C4 was partially purified on Q Sepharose and then chromatographed on Mono Q. Three consecutive fractions were reduced and analysed by SDS/PAGE with and without treatment with human CIs. Lanes 1 and 2, bovine C4A; lanes 3 and 4, a mixture of bovine C4A and C4B, predominantly C4B; lanes 5 and 6, bovine C4B. Lanes 1, 3 and 5 show the native protein, and lanes 2, 4 and 6 show protein treated with CIs.

in the isotopic region (Dodds *et al.*, 1986; Yu *et al.*, 1986) and Carroll *et al.* (1989) demonstrated that this difference could be reproduced by a proline/leucine exchange at position 1101. It is therefore possible that the two putative C4 isotypes of the rhesus monkey have the same residue at this particular position.

Species which have C4 with a single binding specificity

C4 purified from rat, guinea pig, rabbit, dog, pig and fin whale plasma showed only a single binding specificity. In all cases except the fin whale, this was found to be very similar to that of mouse C4 and human C4B (Table 1). Fin whale C4 showed a somewhat higher rate of reaction with glycine than did the other C4 types studied, though this was considerably lower than that seen with human C4A. Dog C4 is known to be moderately polymorphic (Kay *et al.*, 1985). C4 was therefore isolated from four individual animals, each with a different C4 phenotype. All were found to be similar in their binding characteristics.

In some of the species studied more than one form of the α -chain was apparent on SDS/PAGE. C4 was purified from the plasma of three individual fin whales;

in all cases two forms of C4 were observed which had α -chains of different molecular mass (Fig. 3, lanes 5 and 6). These two forms were well separated on the Mono Q column. Whale C4-I (high molecular mass α -chain) was eluted from the Mono Q at 390 mM-NaCl. Whale C4-II (low molecular mass α -chain) was eluted at 470 mM-NaCl. Both forms incorporated [3 H]methylamine into their α -chains and both were cleavable by human CIs. Both forms were similar in their binding specificities with glycine and glycerol. Guinea pig C4 showed two α -chains on SDS/PAGE (Fig. 2b). However, a double-labelling experiment failed to show preferential binding of glycine or glycerol to either form (Fig. 4d).

Non-primates with more than one C4 type

Sheep C4 was purified from 10 ml of plasma. The final preparation of C4 contained what appears to be four α -chain-related products (Fig. 3, lane 8). The upper pair of α -chain bands incorporated [3 H]methylamine and could be cleaved by CIs, causing them to migrate in the position of the lower pair. The lower pair of bands were inactive by both of these criteria, and so are assumed to be native C4 α -chains. The upper pair of bands which appear to be native C4 are not present in equal amounts. Considerably more of the high molecular mass form of C4 was present. Attempts were made to analyse the amount of each band present by gel scanning, but the two forms were too poorly separated to allow quantification. When the binding properties of the sheep C4 were studied it was found that the sheep C4 bound very effectively to glycine but rather poorly to glycerol (Table 1). When a double-labelling experiment was performed (Fig. 4e), the [14 C]glycine bound preferentially to the high molecular mass form of the α -chain while [3 H]glycerol was associated with the lower molecular mass form. It would appear therefore that sheep have two forms of C4, one like human C4A, the other like human C4B in their binding properties.

Bovine C4 was purified from 10 ml of fresh bovine plasma which was treated with EDTA, DFP and 25 mM-benzamidine. C4 was eluted from the Mono Q column in two distinct peaks. The first C4 peak was eluted at approx. 440 mM-NaCl, the second at 480 mM-NaCl. SDS/PAGE analysis of the two C4 types (Fig. 5) indicated that the first form eluted (lane 1) had a lower molecular mass α -chain than the second (lane 5). Lane 3 shows the intermediate fraction which contains both types. Treatment of both C4 types with human CIs led to cleavage of the α -chains to α' -chains (lanes 2, 4 and 6). Both C4 types also incorporated [3 H]methylamine, indicating that both contained intact thiolester bonds. Binding of the two C4 types to small substrates indicated that the first form eluted from the Mono Q column is C4A-like, while the second form is C4B-like (Table 1). In contrast to the C4 of human, chimpanzee and sheep, whose glycine-reactive forms (C4A) have a higher molecular mass α -chain, the cattle C4 with C4A-like activity has a lower molecular mass α -chain. This feature is also demonstrated in a double labelling experiment (Fig. 4f) performed on the fraction that contained both C4 types (Fig. 5, lane 3).

DISCUSSION

The two different isotypes of human C4, C4A and C4B, differ by only four amino acid residues at positions 1101, 1102, 1105 and 1106 in their primary structures. The six

residues in C4A at positions 1101–1106 are PCPVL and those in C4B are LSPVIH (Belt *et al.*, 1984, 1985). Yet these differences confer on them different covalent binding reactivities with amino and hydroxyl groups. This difference leads to a difference in the haemolytic activity of the two isotypes. Detailed analysis of the organization of the genes within the HLA Class III region has revealed that the genes for C4A and C4B are arranged in tandem (Campbell *et al.*, 1988), and are largely similar in their restriction maps (Yu & Campbell, 1987). It is postulated that the two genes arose by a recent duplication event, followed by their divergence by mutation and possibly gene conversion.

The only other animal whose C4 gene organization has been studied in similar detail is the mouse. Although the C4 gene of the mouse has also undergone a duplication, the end products in this case are quite different from those found in human. Firstly, one of the gene products, sex-limited protein (Slp), is inactive in the complement system, even though it is around 94% identical in amino acid sequence to murine C4. Secondly, in the region where the two human isotypes differ in sequence mouse C4 and Slp are identical. Furthermore, this sequence is a hybrid of the two human sequences with the six corresponding residues PCPVIH (Nonaka *et al.*, 1985; Ogata & Sepich, 1985; Sepich *et al.*, 1985).

Work on other animals is much more limited, and has been mainly at the level of traditional genetic mapping or protein typing by electrophoresis. In the case of the chimpanzee it has been shown that there are two genes present which have products equivalent to human C4A and C4B (Granados *et al.*, 1987). There appears to be only a single locus encoding C4 for the dog (Kay & Dawkins, 1984) and for the guinea pig (Burge *et al.*, 1980). Two forms of C4 have been identified for sheep (Groth *et al.*, 1988) and cattle (Groth *et al.*, 1987). The presence of both forms in a number of breeds in both species suggested that the two C4 types are not allelic but the products of two loci.

In this study, C4 from a number of animals were purified from their plasma. They are identified to be C4 by: (i) their elution from Q Sepharose and Mono Q columns amongst the most negatively charged proteins; (ii) their cleavability by human C1s; (iii) their ability to incorporate methylamine; and (iv) their three-chain structure by SDS/PAGE. By these criteria, the proteins described are C4. However, they may not be all of the C4 types present in the animals studied. C4 with a very different elution pattern from the ion exchange columns, for example, could have escaped our detection. In our study on the mouse C4, for example, the presence of Slp, which could be distinguished from mouse C4 by the molecular masses of the α , β and γ chains in SDS gels (Roos *et al.*, 1978), is not evident. The covalent binding activities of C4 to glycine and glycerol were studied. There were a number of reasons for doing this. Firstly, although the evidence from the primary structures of mouse and human C4 suggested that gene duplications in these two species were separate events which occurred after they had diverged, there was a possibility that the duplication was an ancient one and that the sequences of the duplicated genes had been prevented from diverging by a mechanism such as gene conversion. Secondly, we have proposed that the major residue responsible for conferring substrate specificity in C4 is amino acid 1106, which is His in C4B and Asp in C4A. We have shown

that mouse C4, which has a His residue at the relevant position, is C4B-like in its binding properties (Dodds & Law, 1988). It would therefore be useful to search for C4A-like activity in other animals and to determine their corresponding residues in comparison with human C4A. Furthermore, it may help our understanding of the thiolester-mediated binding reaction to identify C4 from other animals which have binding specificities different from both C4A and C4B.

Most of the mammals in the present study appear to be similar to the mouse in that they express C4 with a single binding specificity. It is not possible to say whether they have either (i) a duplicated locus with only one gene coding for an active protein, like the mouse; (ii) more than one gene, some of which code for gene products with indistinguishable reactivity; or (iii) only one single gene coding for one active protein. In a number of these cases there is evidence for more than one population of C4 molecules in the plasma. This was most pronounced in the case of the fin whale, where C4-I and C4-II were completely separated by ion-exchange chromatography, and where the two C4 types had α -chains of markedly different molecular mass. Guinea pig C4 also possessed two distinct α -chains which could be separated by SDS/PAGE. However in both cases there is no detectable difference between the reactivities of the two molecules. From the present evidence it is impossible to say whether these differences are due to the presence of more than one gene, or whether they are due to differences in post-translational modification. Two possibilities for the latter are differential attachment of carbohydrate and differences in the processing of the single chain pro-C4 molecule to the three-chain form. In the cases of mouse and human C4 it is known that different molecular mass forms of the α -chain can be generated by the incomplete removal of a 26-amino-acid peptide from the C-terminus of the α -chain (Chan & Atkinson, 1983; Law & Gagnon, 1985).

The binding specificities of rat, guinea pig, rabbit, dog and pig C4 were very similar to those of human C4B and mouse C4. In the case of the fin whale the binding specificity of the two C4 types present is different from those seen in C4B-like C4. The reaction rate with glycine is 5–6-fold faster than that seen with human C4B and mouse C4, while the reactivity with glycerol is at the low end of the range observed for other C4B-like proteins. It is possible that the whale C4 has a specificity-defining residue other than the His or Asp seen in human and mouse C4. Alternatively, other residues in three-dimensional proximity to the thiolester bond may be altered.

Two separate groups of animals appear to be similar to man in having two distinct C4 types with different binding characteristics. These are the primates and the sheep/cattle branch of the even-toed ungulates, although the pig, which is also an even-toed ungulate, and the whale, which is thought to have evolved from an ancestor of this group (Novacek *et al.*, 1988), appear to have C4 with a single binding specificity. There is good genetic evidence that the chimpanzee has two C4 loci (Granados *et al.*, 1987), and it is reasonable to assume that this is also the case in the rhesus monkey. Sheep and cattle are known to have two forms of C4 and this has been demonstrated in a number of breeds of both species (Groth *et al.*, 1987, 1988). It is therefore likely that the two C4 types of different binding specificities in sheep and cattle are isotypes.

It is premature to build any model on the evolution of the C4 genes in mammals. Structural information is required both at the coding level to account for the different binding specificities, especially for the C4A-like activity found in sheep and cattle, and at the genomic level to clarify the number of C4 and C4-like genes in various species.

We would like to thank Gunnar Battrup, Garth Cooper, James Davies, Roger Dawkins, Steve Dodsworth, Carolyn Giles, Alan Hayward, Libby Henson, John Hickman, Asta Palsdottir, John Penfold, Rami Spilliaert and Julia Vass for their help in obtaining the plasma samples used in these experiments

REFERENCES

- Belt, K. T., Carroll, M. C. & Porter, R. R. (1984) *Cell* **36**, 907–914
- Belt, K. T., Yu, C. Y., Carroll, M. C. & Porter, R. R. (1985) *Immunogenetics* **21**, 173–180
- Burge, J., Nicholson-Weller, A. & Austen, F. (1980) *Mol. Immunol.* **18**, 47–54
- Campbell, R. D., Dodds, A. W. & Porter, R. R. (1980) *Biochem. J.* **189**, 67–80
- Campbell, R. D., Law, S. K. A., Reid, K. B. M. & Sim, R. B. (1988) *Annu. Rev. Immunol.* **6**, 161–195
- Carroll, M. C., Fathallah, D. M., Bergamaschini, L., Alicot, E. M. & Isenman, D. E. (1989) *FASEB J.* **3**, A367
- Chan, A. & Atkinson, J. P. (1983) *J. Clin. Invest.* **72**, 1639–1649
- Dodds, A. W. & Law, S. K. A. (1988) *Complement* **5**, 89–97
- Dodds, A. W., Law, S. K. A. & Porter, R. R. (1985) *EMBO J.* **4**, 2239–2244
- Dodds, A. W., Law, S. K. A. & Porter, R. R. (1986) *Immunogenetics* **24**, 279–285
- Gigli, I., Porter, R. R. & Sim, R. B. (1976) *Biochem. J.* **157**, 541–548
- Granados J., Awdeh, Z. L., Chen, J. H., Giles, C. M., Balner, H., Yunis, E. J. & Alper, C. A. (1987) *Immunogenetics* **26**, 344–350
- Groth, D. M., Wetherall, J. D., Umotong, B. A., Sparrow, P., Lee, I. R. & Carrick, M. J. (1987) *Complement* **4**, 1–11
- Groth, D. M., Wetherall, J. D., Taylor, L., Sparrow, P. R. & Lee, I. R. (1988) *Mol. Immunol.* **25**, 577–584
- Hsiung, L. M., Mason, D. Y. & Dodds, A. W. (1987) *Mol. Immunol.* **24**, 91–96
- Isenman, D. E. & Young, J. R. (1984) *J. Immunol.* **132**, 3019–3027
- Kay, P. H. & Dawkins, R. L. (1984) *Tissue Antigens* **23**, 151–155
- Kay, P. H., Dawkins, R. L. & Penhale, J. W. (1985) *Immunogenetics* **21**, 313–319
- Law, S. K. A. (1983) *Ann. N.Y. Acad. Sci.* **421**, 246–258
- Law, S. K. A. & Gagnon, J. (1985) *Biosci. Rep.* **5**, 913–921
- Law, S. K. A., Lichtenberg, N. A. & Levine, R. P. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7194–7198
- Law, S. K. A., Dodds, A. W. & Porter, R. R. (1984a) *EMBO J.* **3**, 1819–1823
- Law, S. K. A., Minich, T. M. & Levine, R. P. (1984b) *Biochemistry* **23**, 3267–3272
- Müller-Eberhard, H. J. (1988) *Annu. Rev. Biochem.* **57**, 321–347
- Nonaka, M., Nakayama, K., Yeul, Y. D. & Takahashi, M. (1985) *J. Biol. Chem.* **260**, 10936–10943
- Novacek, M. J., Wyss, A. R. & McKenna, M. C. (1988) in *The Phylogeny and Classification of the Tetrapods. Vol. 2: Mammals* (Benton, M. J., ed.), pp. 31–77, Clarendon Press, Oxford
- Ogata, R. T. & Sepich, D. S. (1985) *J. Immunol.* **135**, 4239–4244
- Passmore, H. C. & Shreffler, D. C. (1970) *Biochem. Genet.* **4**, 351–365
- Reid, K. B. M. & Porter, R. R. (1981) *Annu. Rev. Biochem.* **50**, 433–464
- Roos, M. H., Atkinson, J. P. & Shreffler, D. C. (1978) *J. Immunol.* **121**, 1106–1115
- Roos, M. H., Mollenhauer, E., Dement, P. & Rittner, C. H. (1982) *Nature (London)* **298**, 854–856
- Sepich, D. S., Noonan, D. J. & Ogata, R. T. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5895–5899
- Sepich, D. S., Rosa, P. A. & Ogata, R. T. (1987) *J. Biol. Chem.* **262**, 4935–4938
- Sim, E., Dodds, A. W. & Goldin, A. (1989) *Biochem. J.* **259**, 415–419
- Tack, B. F. (1985) in *Complement* (Müller-Eberhard, H. & Miescher, J., eds.), pp. 49–72, Springer, New York
- Yu, C. Y. & Campbell, R. D. (1987) *Immunogenetics* **25**, 383–390
- Yu, C. Y., Belt, K. T., Giles, C. M., Campbell, R. D. & Porter, R. R. (1986) *EMBO J.* **5**, 2873–2881

Received 7 June 1989/24 August 1989; accepted 31 August 1989