The EMBO Journal vol.5 no.11 pp.2873-2881, 1986

# Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity

C.Yung Yu, K.Tertia Belt, Carolyn M.Giles[1], R.Duncan Campbell and Rodney R.Porter*

MRC Immunochemistry Unit, Department of Biochemistry, Oxford University, South Parks Road, Oxford OX1 3QU and [1]Department of Immunology, Royal Postgraduate Medical School, Hammersmith Hospital, London W12 0HF, UK

*Deceased

Communicated by A.F.Williams

**The human complement components C4A and C4B are highly homologous proteins, but they show markedly different, class-specific, chemical reactivities. They also differ serologically in that C4A generally expresses the Rodgers (Rg) blood group antigens while C4B generally expresses the Chido (Ch) blood group antigens. C4A 1 and C4B 5 are exceptional variants which possess their class-specific chemical reactivities, but express essentially the reversed antigenicities. The genes encoding the typical Rg-positive C4A 3a and Ch-positive C4B 3 allotypes and the interesting variants C4A 1 and C4B 5 have been cloned. Characterization of the cloned DNA has revealed that the genes encoding the A 3a, A 1 and B 3 allotypes are 22 kb long, but that encoding B 5 is only 16 kb long. Comparison of derived amino acid sequences of the polymorphic C4d fragment has shown that C4A and C4B can be defined by only four isotypic amino acid differences at position 1101-1106. Over this region C4A has the sequence PCPVLD while C4B has the sequence LSPVIH, and this presumably is the cause of their different chemical reactivities. Moreover, the probable locations of the two Rg and the six Ch antigenic determinants have been deduced. Our structural data on the *C4A* and *C4B* polymorphism pattern suggests a gene conversion-like mechanism is operating in mixing the generally discrete serological phenotypes between C4A and C4B.**

*Key words:* 6-7 kb intron/C4d exon structure/isotypic residues/gene conversion/Rg and Ch determinants

## Introduction

Human complement component C4 plays a key role in the activation of the classical complement pathway, especially in the formation of the complex C3 and C5 convertases (for reviews see Reid and Porter, 1981; Porter, 1984). There are two isotypes, C4A and C4B, which are encoded by tandem loci ~10 kb apart located in the major histocompatibility complex on chromosome 6 (Tiesberg *et al.*, 1976; Carroll *et al.*, 1984a; reviewed by Porter, 1985; Carroll *et al.*, 1985a). As shown by derived amino acid sequences from complete and/or partial cDNA and genomic sequences, the two protein molecules are very similar and there is <1% sequence variation between the 1725 amino acid pro-C4 molecules (Belt *et al.*, 1984, 1985). C4 is synthesized as a single-chain precursor molecule (mol. wt 200 000) but post-translationally processed to the three-chain disulphide-linked structure (Hall and Colten, 1977; Janatova, 1986).

In the middle of the $\alpha$ chain is a characteristic thiolester bond which is also shared by complement component C3 and the protease inhibitor, $\alpha_2$ macroglobulin (reviewed by Tack, 1985). After activation of C4, the thiolester bond becomes labile and readily cleaved to give a reactive carbonyl that will form a covalent ester or amide bond with nearby antigen or IgG surfaces (Law *et al.*, 1980; Campbell *et al.*, 1980). Despite the high sequence homology, C4A and C4B exhibit a remarkable difference in thiolester reactivity. Recent analyses by two laboratories (Isenman and Young, 1984, 1986; Law *et al.*, 1984a; Dodds *et al.*, 1986) have independently led to the similar conclusion that C4A binds preferentially to amino groups of peptide antigens and C4B more efficiently binds to hydroxyl groups of carbohydrate antigens. This intrinsic difference in covalent binding affinity results in a 3- to 4-fold higher binding of C4B to sensitized sheep red cells and explains its well-known higher activity in the classical haemolytic assay (Awdeh and Alper, 1980). The $\alpha$ chain of C4A has a lower electrophoretic mobility on SDS-PAGE than C4B resulting in an apparent mol. wt difference of ~2000 daltons (i.e. C4A-96 000 versus C4B-94 000; Roos *et al.*, 1982). The human blood group antigens Rodgers (Rg) and Chido (Ch) were shown to correspond to C4A and C4B respectively (O'Neill *et al.*, 1978). The apparent 2000 dalton mol. wt difference of the $\alpha$ chains and Rg/Ch antigenic determinants resides in the proteolytic degradation fragment, C4d (Tilley *et al.*, 1978; Lundwall *et al.*, 1982; Chan and Atkinson, 1985).

C4A and C4B are very complex both at genomic and at protein levels. This may have important implications for the origin of many HLA-related autoimmune diseases (Porter, 1983; Dawkins *et al.*, 1983; Fielder *et al.*, 1983). *C4* genes of different individuals can differ both in number and in size (Carroll *et al.*, 1985b; reviewed by Carroll *et al.*, 1985a). The C4A and C4B proteins are highly polymorphic and this has been demonstrated by differences in electrophoretic mobilities of neuraminidase-treated plasma (Mauff *et al.*, 1983), serological typing (Roos *et al.*, 1984; Giles *et al.*, 1984; Giles, 1987), and direct DNA sequencing (Belt *et al.*, 1984, 1985; and this work). A combination of all these approaches suggests that there are altogether >40 alleles in the two classes.

By means of many polyspecific anti-Rg and anti-Ch antisera obtained from transfused patients who lacked either C4A or C4B, respectively, it has been defined that there are two Rodgers (Rg: 1,2) and six Chido (Ch: 1,2,3,4,5,6) antigenic determinants (Giles, 1985a,b; Giles, 1987). Although located at the C4d region, there is not a complete association of Rg determinants with C4A and Ch determinants with C4B. Despite maintaining the class-specific differential thiolester reactivities and apparent $\alpha$ chain mol. wt (Rittner *et al.*, 1984; Roos *et al.*, 1984; Dodds *et al.*, 1986), some A1 allotypes do not possess either of the two Rg determinants, but express four Ch determinants (i.e. Rg: -1,-2; Ch: 1,-2,3,-4,5,6). Moreover, a minor proportion of the B 5 allotypes have been shown to express Rg1 (Rg2 has not been determined), but only contain two of the six Ch determinants (i.e. Ch: -1,-2,-3,4,-5,6) (Rittner *et al.*, 1984; Roos *et al.*, 1984; Giles *et al.*, 1984; Giles, 1987). It was not clear

whether the Rg/Ch determinants and the differential thiolester reactivities are correlated with one another.

In order to correlate the structure and function of the two C4 isotypes particularly with respect to the structural basis of the differential thiolester reactivity, the antigenicity of the Rg/Ch blood groups, and the polymorphism of C4A and C4B, we have constructed genomic libraries from DNA of three individuals with interesting Rg/Ch phenotypes. From these libraries genomic clones of the complete $Rg^+$ A 3a (A 3a is a sub-type of C4A 3) and the complete $Ch^+$ B 3 were isolated. In addition we have obtained clones for the A 1 and B 5 allotypes, described above, which possess their own class-specific properties, but essentially the reverse antigenicity. The *C4* genes from these clones were characterized and their polymorphic C4d regions sequenced. This work allows many important conclusions to be drawn on the genetics and polymorphism of the C4A and C4B isotypes and advances our understanding of the structure and function of two very similar but unique proteins which probably originated from the same ancestral gene by gene duplication.

## Results

### *Cloning, mapping and assignment of C4 genes*

Hybridization of full-length and 5′ *C4* cDNA probes to restriction digests of cloned *C4* genes has shown that the *C4* genes vary in size. The mapping of a *C4A* gene in the cosmid Cos 3A3 revealed that the gene is 22 kb in length and has a 6–7 kb intron located near the 5′ end of the gene (Carroll *et al.*, 1985a; Carroll *et al.*, 1986). In contrast two different *C4B 1* alleles, *B 1a* from Cos 1E2 and *B 1b* from Cos KEM-1 (Carroll *et al.*, 1985a,b; Belt *et al.*, 1985) do not possess this intron and are only 16 kb in size. A combination of Southern blot analysis on subcloned DNA using a full length cDNA probe, and direct DNA sequencing of the *C4A* gene from Cos 3A3 has enabled us to construct an accurate restriction map of the gene (Figure 1a). The 6–7 kb intron is located ~2–2.5 kb from the 5′ end of the gene. Its presence can be detected by Southern blot analysis of *Bam*HI or *Kpn*I restricted genomic DNA, using a 476 bp 5′ cDNA ($P_A$) and a 927 bp *Bam*HI genomic fragment specific for the C4d region of the gene ($P_B$), respectively. The long *C4* genes can be identified by the presence of the 4.8 kb *Bam*HI and the 7.5 kb *Kpn*I fragments. A short *C4* gene is characterized by the presence of the 3.3 kb *Bam*HI and the 8.5 kb *Kpn*I fragments (Carroll *et al.*, 1985a,b). The genomic DNA of the three individuals whose DNA was used to construct genomic libraries was examined by this approach. In the individual AW (*C4A 3; C4B 3/C4A 3; C4B QO*), only the 4.8 kb *Bam*HI and 7.5 kb *Kpn*I fragments were detected (Figure 2, lanes 1 and 4). As the *B QO* allele in this individual was known to be deleted (Carroll *et al.*, 1985a,b; C.Y.Yu, unpublished results) and only a single *C4B* gene is present, it was concluded that the gene encoding *C4B 3* is long. Four *C4B* clones were characterized from the AW genomic library. Fine mapping of cloned DNA confirmed that the *C4B 3* gene possesses the 6–7 kb intron and is therefore 22 kb in length.

In the individual AD (*C4A 1; C4B QO/C4A 3; C4B QO*) only the 4.8 kb *Bam*HI and 7.5 kb *Kpn*I fragments were detected (Figure 2, lanes 3 and 6). Therefore, the *C4* genes present in this case are all long. It has been suggested that the *C4B QO* allele of the HLA haplotype *A 3 B 47 DR 7 C4A 1; C4B QO* is deleted (White *et al.*, 1985; Carroll *et al.*, 1985a; Schneider *et al.*, 1985). On the other hand, the *C4B QO* allele of the HLA haplotype *A 2 B 44 DR 6 C4A 3; C4B QO* is a long *C4* gene

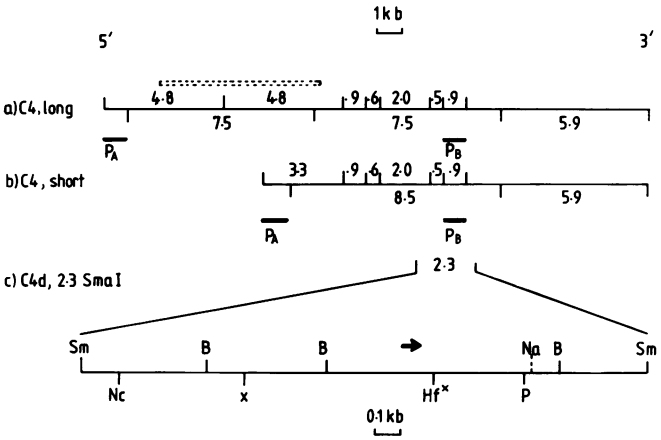


**Fig. 1.** Restriction maps of (**a**) long, (**b**) short human C4 genes and (**c**) the polymorphic C4d region. (**a**) and (**b**) Restriction maps of a long *C4* gene (22 kb) were derived from Cos 3A3 and that of a short *C4* gene from λJM-2a. Open dotted box indicates the approximate position of the 6–7 kb intron. $P_A$ and $P_B$ are the 5′-specific *C4* cDNA probe and C4d-specific genomic probe, respectively. For each map, the top represents the order and size of *Bam*HI fragments, and the bottom the *Kpn*I fragments. When $P_A$ and $P_B$ are used, long *C4* genes have 4.8 kb *Bam*HI and 7.5 kb *Kpn*I fragments, and short *C4* genes have 3.3 kb *Bam*HI and 8.5 kb *Kpn*I fragments. (**c**) The 2.3 kb *Sma*I fragment encodes >90% of C4d. Sm = *Sma*I, B = *Bam*HI, Nc = *Nco*I, X = *Xho*I, P = *Pvu*II, Na = *Nar*I (present in some C4 genes only; refer to text), and $Hf^x$ = *Hin*fI (only the single site in the 927 bp *Bam*HI fragment is shown). → depicts the position of the sequencing primer near to the isotypic region.

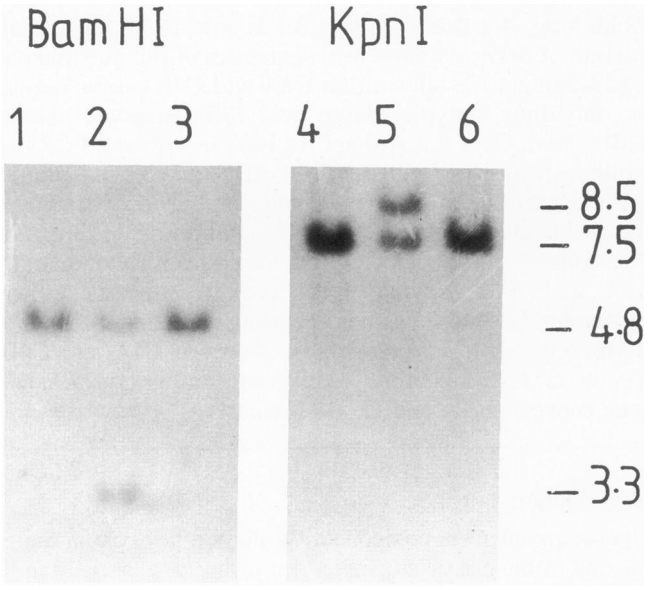


**Fig. 2.** Genomic Southern analysis of AW, JM and AD DNA. **Lanes 1** and **4**, AW (*C4A 3; C4B 3/C4A 3; C4B QO*); **lanes 2** and **5**, JM (*C4A 4; C4B 5/C4A 3; C4B QO*); **lanes 3** and **6**, AD (*C4A 1, C4B QO/C4A 3; C4B QO*). **Lanes 1–3** show DNA restricted with *Bam*HI and probed with the 5′-specific probe, $P_A$. **Lanes 4–6** show DNA restricted with *Kpn*I and probed with the C4d-specific probe, $P_B$. Long *C4* genes are characterized by 4.8 kb *Bam*HI and 7.5 kb *Kpn*I fragments. Short *C4* genes are characterized by 3.3 kb *Bam*HI and 8.5 kb *Kpn*I fragments.

and has been cloned (C.Y.Yu, N.R.Rodrigues and R.D.Campbell, unpublished).

Four *C4A* genomic clones were identified and characterized from the AD library. In order to differentiate between the *A 3*

**Table I.** Haplotypes of individuals from whom λ EMBL-3 genomic libraries were constructed

| Individuals | HLA | | | | C4 | | Deduced Rg/Ch types | References |
|---|---|---|---|---|---|---|---|---|
| 1. A.W. | A2 | Cw3 | B15 | DR1 | A3 | B3 | Rg: 1,2; Ch: 1,2,3,4,5,6 | Carroll *et al.*, 1985a; this work |
| | A2 | – | B7 | DR2 | A3 | BQO | Rg: 1,2; Ch: −1,−2,−3,−4,−5,−6 | |
| 2. J.M. | A3 | | B35 | DR2(?) | A3 | BQO | Rg: 1,2; Ch: −1,−2,−3,−4,−5,−6 | Giles *et al.*, 1984; Hing *et al.*, |
| | A24 | | B18 | DR2(?) | A4 | B5 | Rg: 1,2; Ch: −1,−2,−3,4,−5,6 | 1986; Giles, 1987 |
| 3. A.D. | A3 | C6 | B47 | DR7 | A1 | BQO[a] | Rg: −1,−2; Ch: 1,−2,3,−4,5,6 | Bentley *et al.*, 1985 |
| | A2 | C5 | B44 | DR6 | A3a | BQO | Rg: 1,2; Ch: −1,−2,−3,−4,−5,−6 | |

[a]This was previously typed as *C4A QO; C4B 3.1* but further studies show that it is *C4A 1; C4B QO.*

and *A 1* alleles a *Bam*HI/*Nar*I double digest of the cloned DNA was carried out. This digest was chosen because all *C4A* alleles so far characterized contained a single *Nar*I site in the 927 bp *Bam*HI fragment (Figure 1c) whereas the *C4B* alleles did not. It was anticipated that as C4A 1 had been serologically typed Ch positive (i.e. contained the C4B antigenic determinants, Rittner *et al.*, 1984) the *C4A 1* allele may also not contain the *Nar*I site. The 927 bp *Bam*HI fragment at the C4d region of three of the clones was cleaved to 810 and 117 bp fragments by *Nar*I. The nucleotide sequence of the *Bam*HI fragment from one of these clones, λAD-2α, was found to be identical to the sequence of the gene in Cos 3A3 previously assigned as coding for C4A 3a (Belt *et al.*, 1985). The *Bam*HI fragment in the fourth clone, λAD-1α, remained intact after the *Nar*I digest, and therefore corresponds to the *A 1* allele.

In the individual JM (*C4A 4; C4B 5/C4A 3; C4B QO*) both 22 and 16 kb *C4* genes were detected, as shown by the presence of the 4.8 and 3.3 kb *Bam*HI fragments (Figure 2, lane 2) as well as the 7.5 and 8.5 kb *Kpn*I fragments (Figure 2, lane 5). Further genomic Southern blot analysis showed that there were no long *C4B* genes present in the individual JM and the polymorphic C4d region of the short *C4* genes appeared identical in many marker restriction digests (data not shown). Two *C4B* clones were characterized from the JM library. One of these clones, λJM-2a, includes a 17 kb insert that encodes a complete 16 kb *C4B* gene. The other clone, λJM-6a, starts at the middle of the *C4B* gene and extends 3′ to the 21-hydroxylase B gene. Fine restriction mapping showed that the polymorphic C4d regions of these clones are identical. This implies that the *C4B QO* allele of JM is either deleted or is a short *C4B* gene which is identical to the *B 5* allele at the C4d polymorphic region. Further studies favour the former proposal that the *C4B QO* allele of the HLA *A 3 B 35* haplotype is deleted (data not shown). The clones λJM-2a and λJM-6a are therefore assigned as coding for the C4B 5 allotype.

*Sequence analysis of nine allotypes of C4 at the C4d region*

Comparison of nucleotide sequences from a full length and two partial *C4* cDNA clones corresponding to *A 4, A 3* and *B 2*, and limited genomic sequence of the *B 1a, B 1b* and *B 2* alleles (Belt *et al.*, 1984, 1985) had already shown that there are only 15 nucleotide changes present in the coding sequence, which result in 13 amino acid differences. When additional data from amino acid sequencing of C4 proteins is included (D.N.Chakravarti, R.D.Campbell and J.Gagnon, unpublished), 15 amino acid differences are apparent, 12 of which are clustered in the section of the α-chain C-terminal to the thiolester. This very small number of changes is apparently responsible for the substantial differences in the reactivity of the thiolester bond, and for the charge and antigenic variation in different allotypes. How these structural and functional differences were related to changes in amino acid sequence was not clear. In order to define which amino acid changes were responsible for these differences, we set out to determine the primary structure of the C4d polymorphic region from four informative alleles, *A 3a, A 1, B 3* and *B 5* (Table I). Parallel functional studies had already shown that all allotypes of each class possess their own characteristic differential thiolester reactivity (Dodds *et al.*, 1986). A 3a was deduced to be Rg: 1,2; Ch: −1,−2,−3,−4,−5,−6 and B 3 was Rg: −1,−2; Ch: 1,2,3,4,5,6. The structure of these two isotypes can therefore give an account of the amino acid sequences that are related to the Rg/Ch antigenic determinants in addition to those that determine the thiolester reactivity. A 1 is an exceptional C4A variant because it does not possess either of the Rg determinants but acquires four of the Chido determinants (Rittner *et al.*, 1984; Giles, 1987). It was typed Rg: −1,−2; Ch: 1,−2,3,−4,5,6. B 5 is essentially the reverse to A 1. It was typed Ch: −1,−2,−3,4,−5,6 and was Rg positive (Roos *et al.*, 1984; Hing *et al.*, 1986; Giles *et al.*, 1984; Giles, 1987). By comparing the structural data of A 1 and B 5 with A 3 and B 3, the amino acid sequences involved in determining the thiolester reactivity and Rg/Ch antigenicity can be distinguished.

The variation of nucleotide sequences and their resulting amino acid changes, if any, of *A 3a, B 3, A 1, B5* and five other published alleles, *A 3b, A 4, B 1a, B 1b* and *B 2* (Belt *et al.*, 1985) are depicted in Figure 3a and b respectively.

*C4A 3a*

As mentioned before, the *C4A* genes from Cos 3A3 and λAD-2α were identical and were assigned *C4A 3a*. They differ by one nucleotide from a second *C4A 3* allele, encoded in pAT-42, which is therefore assigned *C4A 3b*. The *A 3a* allele in Cos 3A3 and λAD-2α encode an Ala residue at position 1267. However, due to a G→T transversion at the equivalent position in pAT-42 the *A 3b* allele encodes a Ser residue (Belt *et al.*, 1985).

The complete nucleotide sequence of the C4d region of the *A 3a* allele has been determined by the M13 sequencing method using randomly sonicated DNA. Comparison of the genomic DNA sequence with the cDNA sequence showed that the 380 amino acid C4d fragment is split by seven introns into eight exons that altogether span 2050 bp. The exons are termed D1−D8 and they encode 30, 70, 26, 52, 39, 57, 78 and 56 amino acid residues respectively (Figures 3 and 5b). The C4bp-factor I cleavage sites are encoded in exons D1 and D8, respectively, as shown in Figure 3a. The residues involved in the thiolester bond are encoded near the middle of exon D2. With reference to our previous results (Belt *et al.*, 1985), exons D1, D2, D5 and D8 appear highly conserved and no polymorphic sites have been detected so far. More than 90% of the C4d coding sequence is encompassed in a unique 2.3 kb *Sma*I fragment. Within this fragment is a 927 bp *Bam*HI and a neighbouring 410 bp

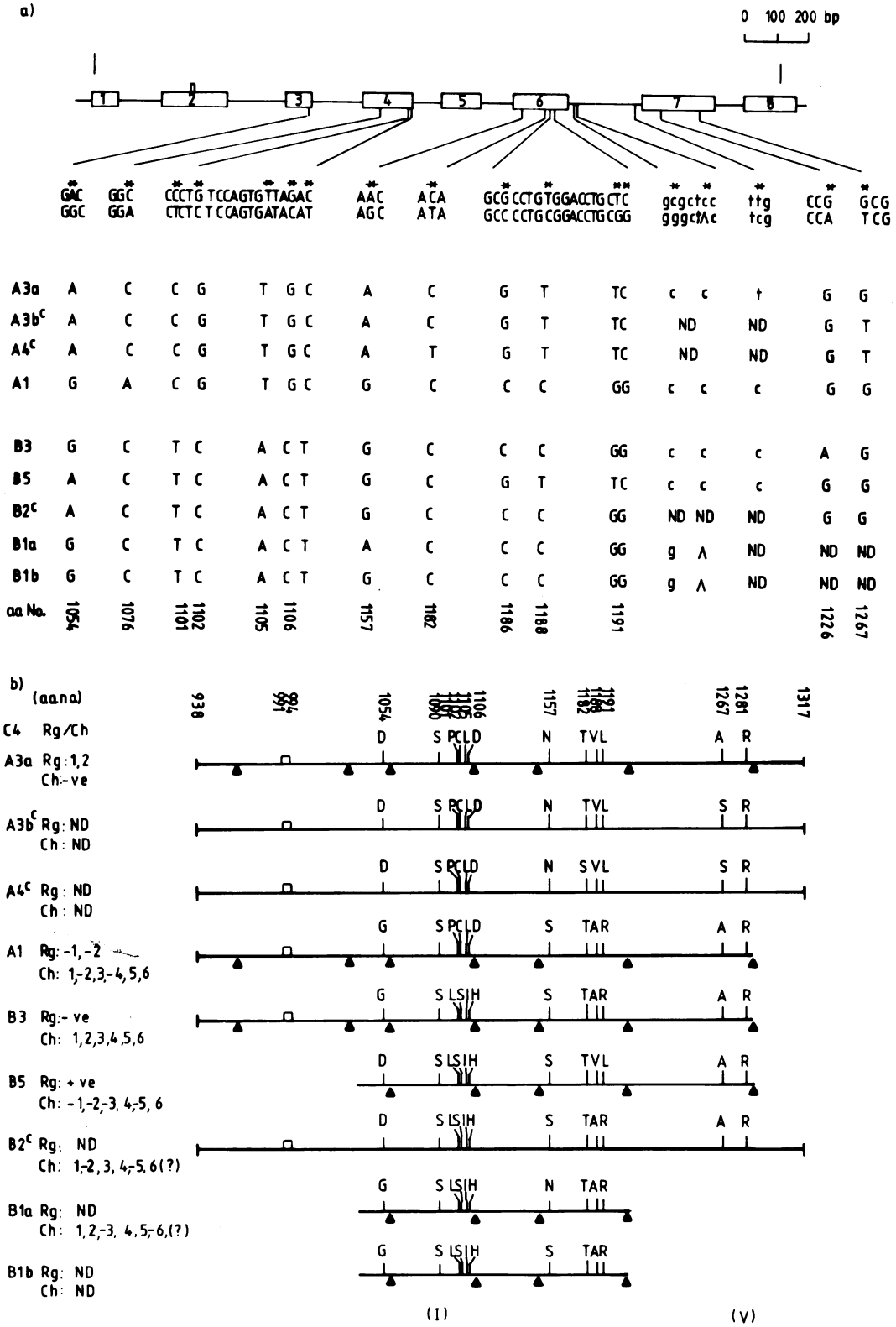


**Fig. 3.** Summary of differences in (**a**) nucleotide sequences and (**b**) derived amino acid residues at the C4d region for nine C4 alleles/allotypes. (**a**) C4d exon−intron structure is shown, position and nature of DNA sequence variations are indicated. Exons are numbered 1−8 in open boxes. Varied nucleotides are marked by asterisks and their presence in various alleles tabulated. Coding sequences are in capital letters and non-coding sequences in small letters. Notice the mutually exclusive variation pattern of DNA coding for amino acids (aa) 1054, 1101−1106, 1157 and 1186−1191. The C4bp-factor I cleavage sites are encoded by the first and the eighth exons, marked by vertical strokes. Λ = deletion, 凵 = DNA coding for thiolester residues, ND = not determined. (**b**) Derived amino acid changes for (a) are shown. In addition Rg/Ch types are included (those of B 1a and B 2 were from deduction only) (Giles *et al.*, 1984; Giles, 1987). Complete C4d fragment is from amino acids 938−1317 (numbering is for the continuous sequence of pro-C4). Residues in brackets (I and V) were found by amino acid sequencing, but were not detected in the DNA sequences. Superscript c denotes data from cDNA sequence (Belt *et al.*, 1984, 1985). Data on B 1a and B 1b were from Belt *et al.* (1985). 凵 = thiolester site; ▲ shows position of introns in genomic-derived sequences. ⊢ = boundary of C4d region, ND = not determined.

*Bam*HI−*Sma*I subfragment (Figure 1c) which together cover all the reported polymorphic sites of the C4 α chain.

*C4B 3*

The clone λAW-3B encodes the C4B 3 allotype which was serologically typed as fully Rg negative and fully Ch positive. The complete DNA sequence of the 2.3 kb *Sma*I fragment was determined. Comparing *B 3* with *A 3a*, a total of 13 nucleotide changes were detected. These resulted in eight amino acid changes at four regions, i.e. 1054, 1101−1106, 1157 and 1188−1191, of which B 3 has the sequence G-LSPVIH-S-ADLR and A 3a has the sequence D-PCPVLD-N-VDLL. Since these eight amino acid residues are the only detected changes at the C4d region between the two typical C4 isotypes, they must be related to thiolester reactivity and to the Rg/Ch antigenic determinants. These eight amino acid changes can also explain that, in general, C4A is slightly acidic and C4B is slightly basic. C4B is associated with two basic amino acid residues (His 1106 and Arg 1191) and C4A with two acidic residues (Asp 1106 and Asp 1191).

Interestingly, the B 3 sequence G-LSPVIH-S-ADLR is also common to B 1b (Belt *et al.*, 1985). The faster mobility of B 3 than B 1 on agarose typing gel (Mauff *et al.*, 1983) is therefore due to the presence of some other polymorphic sites on B 3, presumably outside the C4d region. Two other C4B allotypes, B 1a and B 2, differ from B 3 at single positions. At 1157, B 1a has the C4A 3-related Asn residue instead of the C4B 3-related Ser residue. At 1054, B 2 has the other C4A 3-related Asp residue instead of the C4B 3-related Gly residue (Figure 3b).

*C4A 1*

The A 1 allotype encoded in the clone λAD-1α was serologically typed Rg: −1,−2 and Ch: 1,−2,3,−4,5,6. It is an extremely interesting C4A variant not only because it expressed four of the Chido determinants and neither of the Rodgers determinants generally associated with C4A, but it is also one of the slowest moving C4A allotypes in the agarose typing gel (Rittner *et al.*, 1984; Roos *et al.*, 1984; Mauff *et al.*, 1983). The DNA sequence of the 2.3 kb *Sma*I fragment of the *A 1* allele from clone λAD-1α was completely determined. The derived amino acid sequence at the C4d region revealed that it is almost identical to that of the B 3 allotype except at position 1101−1106. At this region, the A 1 retains the C4A 3-related sequence, PCPVLD, but the four amino acid residues at the other three polymorphic sites are all C4B 3-like. In other words, the C4A 1 has G-PCPVLD-S-ADLR compared with D-PCPVLD-N-VDLL for A 3a, and G-LSPVIH-S-ADLR for B 3. Since A 1 is a typical C4A allotype with respect to its thiolester reactivity (Dodds *et al.*, 1986), this suggests that the four amino acid differences at position 1101−1106, i.e. PC--LD in C4A versus LS--IH in C4B, are solely responsible for the differences in thiolester reactivity observed between the C4A and C4B isotypes. Compared with B 3, the A 1 allotype does not possess the Ch2 and Ch4 antigenic determinants. This infers that the C4B isotypic amino acid residues are involved in expressing the Ch2 and Ch4 epitopes because they are the only difference between A 1 and B 3 at the C4d region. However, the C4A isotypic residues, PC--LD, are not related to either of the two Rg determinants because C4A 1 is Rg negative. The amino acid variations at the other three polymorphic regions, i.e. at position 1054, 1157 and 1188−1191, also appear related to the expression of Rg/Ch antigenic determinants.

*C4B 5*

Historically, B 5 was the first C4B allotype that was demonstrated as having both Rg and Ch determinants expressed on the same molecule (Roos *et al.*, 1984). To differentiate from another allotype with the same electrophoretic mobility, this variant was also designated as C4B $5Rg^+$. The B 5 allotype encoded in clone λJM-2a was serologically typed as Rg positive and Ch: −1,−2,−3,4,−5,6. The DNA sequence of the 927 bp *Bam*HI fragment and its neighbouring 410 bp *Bam*HI−*Sma*I fragment were determined. The derived amino acid sequence at the C4d polymorphic region reveals that the B 5 allotype retains the B 3-related sequences at two regions, 1101−1106 and 1157. However, the sequences at the other two regions, 1054 and 1188−1191 are both C4A 3-like, i.e. B 5 has the sequence D-LSPVIH-S-VDLL. This again suggests that the amino acid sequence at position 1101−1106, LS--IH, is C4B isotypic and is involved in determining the C4B class-specific properties. The B 5 allotype expressed Ch4 and Ch6. Its two B 3-related sequences, LSPVIH 1101−1106 and S 1157, are therefore related to these two Chido antigenic determinants.

*Comparison of intron sequences of C4d*

One of the very striking features of the DNA polymorphism of *C4* alleles is that their intron sequences at the C4d region are highly conserved. From the data available, the seven introns in the 2.3 kb *Sma*I fragment of *B 3* and *A 1* are totally identical, and the *A 3a* differs only by a C→T transition in the intron 3′ to exon D6 (see Figure 3a). In addition the five introns of the *B 5* allele sequenced are identical to those in *B 3*. This reinforces our previous report (Belt *et al.*, 1985), in which we compared the introns in the 927 bp *Bam*HI fragments among the *A 3a, B 2, B 1a* and *B 1b* alleles. Two differences in the intron 3′ to exon D6 of the *B 2, B 1a* and *B 1b* alleles compared with other alleles sequenced have been defined. One is a G→C transversion. The second is the deletion of a C nucleotide (Figure 3a and Belt *et al.*, 1985).

## Discussion

Here we report four significant findings on the genetics and polymorphism of human complement components C4A and C4B: (i) *C4B* genes of different gene size (i.e. 22 versus 16 kb) were cloned; (ii) C4A and C4B can be defined by four isotypic amino acid residues at position 1101−1106, which are likely to be responsible for their different thiolester reactivities; (iii)the probable locations of the two Rg and six Ch antigenic determinants have been deduced; and (iv) the pattern of DNA polymorphism of *C4A* and *C4B* at the C4d region seems to be the result of gene conversion events. For this study, three genomic libraries were constructed and four interesting *C4* genes representing two alleles from each class were isolated and partially characterized.

Variation in the number of expressed *C4* genes has been noted for 10 years (Tiesberg *et al.*, 1976), but differences in gene size at the same locus is not a common phenomenon. It is now clear that the *C4B* genes can be either 22 or 16 kb in size, as exemplified by our cloned *B 3* and *B 5* alleles respectively. Southern blot analysis has also indicated that a portion of the *B 1* alleles (such as those linked to *A 3*) are also 22 kb in size (A.Palsdottir, unpublished results). Three other *C4B* genes from cosmid clones coding for B 2, B 1a and B 1b, respectively, were found to be 16 kb in size only (Carroll *et al.*, 1985a). No 16 kb *C4A* genes have been detected so far. Detailed restriction mapping and

Southern blot analysis suggested that a long *C4* gene is due to a large intron lying ~2–2.5 kb 3′ to the transcription start site. The presence of the very similar *C4A* and *C4B* genes in tandem might have caused some mispairing of the non-allelic genes between homologous chromosomes during meiosis with unequal crossing over (Carroll *et al.*, 1985b). The widespread distribution of two homologous *C4B* genes with a size difference of ~6–7 kb would increase the chance of mispairing between the long *C4B* gene and the long *C4A* gene. This is of particular interest because of the very high frequency (~10–15%) of the non-expression of *C4A* or *C4B* alleles in the population, and deletion–duplication of *C4* genes is not unusual (Carroll *et al.*, 1985b; Hauptmann *et al.*, 1986). Obviously, it is crucial to determine the nature of the 6–7 kb intron and to see whether it is involved in influencing the recombinational events of C4 genes and their extensive polymorphism. This 6–7 kb intron could be a member of the long interspersed sequences, LINE or L1 (Singer, 1982).

Our structural data on the complete $Rg^+$ A 3a and complete $Ch^+$ B 3 shows that there are eight amino acid changes at four discrete regions between a common C4A and a common C4B allotype. At positions 1054 (I), 1101–1106 (II), 1157 (III) and 1188–1191 (IV), A 3a is characterized by the sequence D-PCPVLD-N-VDLL while B 3 is characterized by the sequence G-LSPVIH-S-ADLR (Figure 4). These are the only detectable changes in the polymorphic C4d region and therefore account for the many differences between the C4A and C4B isotypes. These include the differential reactivities of the labile thiolester, the apparent $\alpha$ chain mol. wt. difference of 2000 daltons, and the Rg/Ch antigenic determinants.

Structural information from other less common variants of each isotype, C4A 1 and C4B 5, contributes to our understanding of the molecular basis of the isotypic properties and on the location and nature of the Rg/Ch antigenic determinants. The A 1 and B 5 proteins possess their class-specific properties, but essentially reversed antigenicities. A 1 is characterized by the sequence G-PCPVLD-S-ADLR and B 5 is characterized by the sequence D-LSPVIH-S-VDLL. A combination of the structural data on A 3a, A 1, B 3, B 5 and five other published C4 allotypes (Belt *et al.*, 1984, 1985) conclusively suggests that the amino acid changes at regions (I), (II) and (IV) are only related to Rg/Ch antigenic differences and there are only four amino acid changes common to all C4A and to all C4B allotypes (Figure 4). The C4A isotype can be defined by $\overset{**}{P}CPV\overset{**}{L}D$, and C4B defined by $\overset{**}{L}SPV\overset{**}{I}H$ at position 1101–1106 (i.e. region II). Apparently, these four isotypic residues are the root of the many class-specific properties. The previous suggestion that the VDLL/ADLR at 1188–1191 (Belt *et al.*, 1984, 1985; Law *et al.*, 1984a) might be C4A/C4B isotypic residues can now be ruled out.

Thus it appears that the C4 isotypic region is solely responsible in modulating the thiolester reactivity. The putative thiolester and the isotypic residues are separated by 106 amino acids. A hydropathy plot of C4d (Figure 5a) shows that the isotypic residues are located at a hydrophilic region separated from the thiolester by two hydrophobic zones. The thiolester residues appear at the middle of a hydrophobic pocket. It is likely that the proposed conformational change during C4 activation by $C\bar{1}$ (Isenman and Kells, 1982) would bring the isotypic and thiolester residues close together for certain interactions. The negatively charged Asp 1106 of C4A might increase the nucleophilicity of the nearby amino/amine groups towards the thiolester carbonyl (Law *et al.*, 1984a). This may provide an explanation for the higher covalent binding affinity of C4A to peptide antigens. By the same analogy, the positively charged His 1106 of C4B might participate in hydrogen-bonding charge shift in hydroxyl substrates, resulting in more efficient binding to hydroxyl groups of carbohydrate antigens. This postulation leads to one fundamental question on the mechanism of the breakage of the thiolester bond and its reactivity. Instead of the generally suggested conformational change exposing a reactive thiolester which undergoes rapid reaction with any available nucleophile, a situation may occur in which a conformational change rearranges the structure of the isotypic and the thiolester regions, forming a transient state that can catalyse the nucleophilic attack by the amino/hydroxyl groups on the thiolester carbonyl.

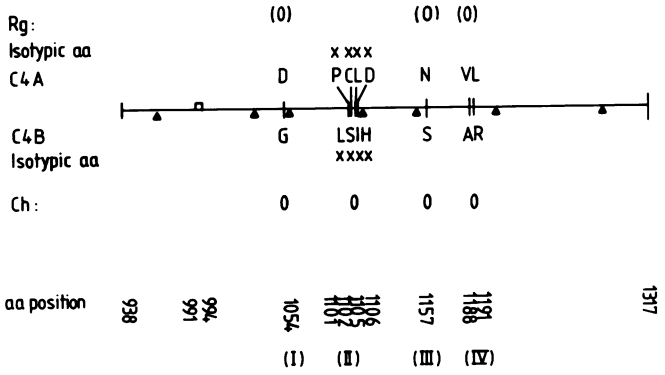


**Fig. 4.** Isotypic residues and possible location of Rg/Ch antigenic determinants. Isotypic residues of C4A and of C4B are marked by crosses. The probable location of Rg/Ch determinants are indicated by '0'. The C4A isotypic residues are not related to the Rg determinants. The two Rg epitopes lie in two of the regions I, III and IV on C4A. The six Ch determinants are formed by conformational and probably sequential epitopes from the four polymorphic sites on C4B. ▲ shows position of introns in genomic-derived sequences. ⊥⊥ = thiolester site.

Besides the Asp/His 1106, the other three isotypic residues and the sequence nearby (Figure 5b) might be important in modulating the thiolester reactivity as well. It is possible that the sulphydryl group of Cys 1102 in C4A may interact with the carboxyl of Asp 1106 or with the sulphydryl of the thiolester Cys 991 at some stage. There is no other Cys residue at the C4d region of either isotype. The Pro 1101 of C4A would possibly cause considerable conformational change at the isotypic region. Together with Pro 1103, Pro 1101 may impose significant constraints on the local structure. It has been suggested that Pro 1101 and Cys 1102 may be responsible for the anomalous slower mobility of the C4A $\alpha$ chain in the modified SDS–polyacrylamide gel (Chan and Atkinson, 1985).

There is only one functional *C4* gene in mouse, which has to cope with both peptide and carbohydrate antigens. The mouse C4 protein appears to be a hybrid of human C4A and C4B, with respect to the isotypic residues, i.e. PCPVIH (Levi-Strauss *et al.*, 1985; Sepich *et al.*, 1985; Nonaka *et al.*, 1985). The first two residues are human C4A specific, but the last two are C4B specific. It will be of interest to know if the reactivity of other thiolester-containing proteins such as the mouse C4, the complement component C3 and the protease inhibitor $\alpha_2$ macroglobulin are also modulated by the area analogous to the C4 isotypic region. A comparison of the amino acid sequences of these thiolester-containing proteins at the area analogous to the human C4 isotypic region is shown in Table II. Since mouse C4, human C3 and mouse C3 all possess the human C4B-specific squence, IH 1105–1106, they might have high covalent binding

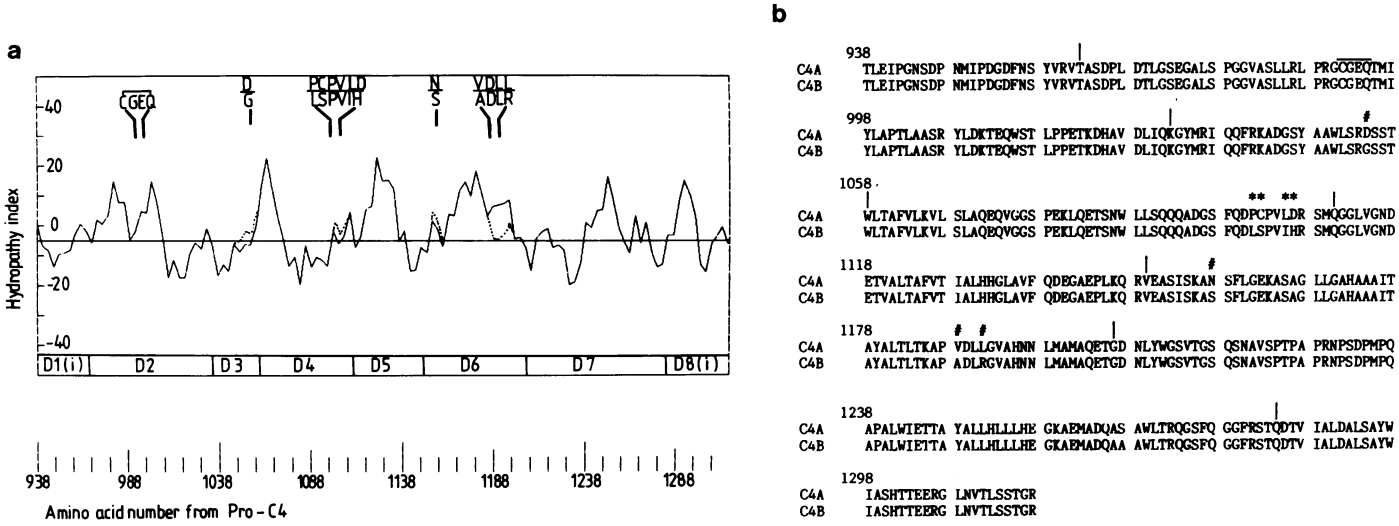


**Fig. 5.** Comparison of (**a**) hydropathic profiles and (**b**) derived amino acid sequences of C4d fragments between complete $Rg^+$ C4A and complete $Ch^+$ C4B. (**a**) The computer program HYDROPLOT, based upon the algorithm of Kyte and Doolittle (1982), was employed with a window of 11 amino acids. C4A profile is shown and the differences between C4A and C4B is in dotted lines. Position of the thiolester residues, and polymorphic residues between $Rg^+$ C4A and $Ch^+$ C4B are indicated. The region encoded in exons D1−D8 is represented in boxes. The C4d boundaries are encoded by exons D1 and D8 (see text). (i) = incomplete exon. (**b**) The corresponding amino acid sequence for (a). The complete $Rg^+$ C4A is based on data from C4A 3a. Complete $Ch^+$ C4B is based on data from C4B 3 for the sequence encoded by exons D1−D7 and C4B 2 for that of exon D8. Positions of the exon boundaries are marked by vertical strokes. The thiolester site is overlined. The isotypic residues for C4A and C4B are marked by asterisks. The probable location of Rg/Ch determinants are marked by #, and * (for C4B).

**Table II.** Comparison of amino acid sequences at the thiolester sites and at the region analogous to the human C4 isotypic site of some thiolester-containing proteins

| | Thiolester site | Region analogous to the human C4 isotypic site | References |
|---|---|---|---|
| | | * * * * | |
| Human C4A | P R G C G E Q T M I | F Q D P C P V L D R S M | Campbell *et al.*, 1981; Hellman *et al.*, 1984; |
| Human C4B | P R G C G E Q T M I | F Q D L S P V I H R S M | Belt *et al.*, 1984, 1985; this work |
| Murine C4 | P Q G C A E Q T M I | F H D P C P V I H R A M | Levi-Strauss *et al.*, 1985; Sepich *et al.*, 1985; Nonaka *et al.*, 1985 |
| Murine Slp[a] | P R S C A E Q T M I | F H D P C P V I H R A M | Ogata and Sepich, 1985; Nonaka *et al.*, 1986 |
| Human C3 | P S G C G E Q N M I | F Q E D A P V I H Q E M | De Bruijn and Fey, 1985 |
| Murine C3 | P A G C G E Q N M I | F Q E D G P V I H Q E M | Wetsel *et al.*, 1984 |
| Human $\alpha_2$M | P Y G C G E Q N M V | F R S S G S L L N N A I | Kan *et al.*, 1985 |

[a]Haemolytically inactive.

affinity for hydroxyl groups and this has indeed been shown for human C3 (Law *et al.*, 1984b).

Analysis of the sequence and serological data among A 3a, A 1 and B 5 sheds light on the location of the six Ch antigenic determinants. There is not a direct one-to-one correlation between amino acid substitution at each of the four polymorphic regions and antigenic determinants. At the C4d region, A 1 has the identical amino acid sequence to that of B 3 except at the isotypic residues and A 1 expressed all Ch determinants except Ch2 and Ch4. This suggests that C4B isotypic residues are involved in the expression of these two epitopes. The conversion of G to D at 1054, and ADLR to VDLL at 1188−1191 results in the loss of the Ch1, 2, 3 and 5 epitopes on the B 5 molecule. Therefore, the C4B 3-related G at 1054 and ADLR at 1188−1191 seem related to Ch1, 2, 3 and 5. It is apparent that the structural component of each of the four discrete polymorphic sites is involved in the formation of more than one epitope. On the other hand, the presence of G 1054, S 1157 and ADLR 1188−1191 on A 1 only forms Ch1, 3, 5 and 6, and the presence of LSPVIH 1101−1106 and S 1157 on B 5 only forms Ch4 and Ch6. This would suggest the six Chido antigenic determinants are formed by overlapping epitopes, some of which are conformational (Benjamin *et al.*, 1984). It seems that each of the four discrete polymorphic sites does not only form a possibly sequential epitope, but will also interact with the structural component of another epitope to give an additional conformational antigenic determinant. To this end, Ch4 and Ch6 could be sequential epitopes and Ch2 a conformational epitope. Using a similar approach, the probable location of the two Rg determinants can be derived by comparing sequence and serological data between A 3a and A 1. They may fall within D 1054, N 1157 and VDLL 1188−1191, although their exact correlation is yet unclear. Contrary to the situation in C4B, the C4A isotypic residues seem non-immunogenic and are not related to either of the two Rg determinants, as exhibited by the A 1 allotype. This would again reflect the conformational constraints imposed by the C4A isotypic amino acid residues.

Our structural studies on nine alleles from *C4A* or *C4B* show that there is a mosaic polymorphism pattern which is due to the shuffling of four discrete regions, or markers, within 1 kb of

DNA at the C4d region. This is also manifested by the mixing of the two generally discrete phenotypes, i.e. C4A with two Rg and C4B with six Ch determinants. The interchange of the discrete sequences at regions I, III and particularly IV, will result in the expression of the reverse antigenicity. Of equal possibility, though not shown in this paper, the interchange of the sequence at region II, i.e. the isotypic region, will convert the C4 protein expressed to the opposite class. On the other hand, it is interesting to note that although there are some 13 changes in nucleotides at the coding region between the two isotypes at C4d there are only one or two changes in the intron sequences. Moreover, the overall G/C content for the 2050 bp of DNA coding for C4d is relatively high, i.e. 60%. One of the most plausible explanations for this mosaic polymorphism pattern is by some gene conversion-like events (Petes and Fink, 1982) between the repeated *C4A* and *C4B* genes. The high G/C content at the C4d region would favour this event to occur (Weiss *et al.*, 1983a,b). Alternatively, it could be caused by many independent, double crossing over events but this seems unlikely. As discussed before, it will be very interesting to determine whether the extensive C4 polymorphism and the unusually high frequency of recombination events are correlated. Knowledge of this would be very helpful to our understanding of the recombination mechanisms in human genetics.

## Materials and methods

*Preparation and analysis of genomic DNA*

Genomic DNA was isolated from whole blood (J.M.) or from EBV-transformed cell lines (AD and AW) according to Bell *et al.* (1981). 5–10 μg of DNA was digested to completion with 20 U of *Bam*HI or *Kpn*I (Amersham) and electrophoresed on a 0.8% agarose gel and transferred to nitrocellulose filters as described by Southern (1975). The blots were hybridized to nick-translated (Rigby *et al.*, 1977) probes $P_A$ and $P_B$ (Figure 1). $P_A$ is a 476 bp 5′-specific *C4* cDNA probe (Belt *et al.*, 1984) and $P_B$ is a 927 bp *C4d* genomic probe (Belt *et al.*, 1985). The blots were washed, dried and autoradiographed as described by Maniatis *et al.* (1982).

*Construction of genomic libraries*

Three libraries, AW, JM and AD, were constructed. For each library, 200–300 μg of genomic DNA was partially digested with *Mbo*I and size fractionated through a sucrose gradient. DNA fragments between 15 and 23 kb were ligated to *Bam*HI/*Eco*RI-digested λ EMBL-3 vector and packaged as described by Frischauf *et al.* (1983). $1-2 \times 10^6$ recombinants were plated on *Escherichia coli* strain Q359. Filters were screened using a full-length *C4* cDNA (Belt *et al.*, 1984) probe and a C4d-specific genomic probe, $P_B$ (Belt *et al.*, 1985) as described by Maniatis *et al.* (1982). Rescreening of positives was done on *E. coli* strain Q358. C4 clones were separated into two groups by further screening using [$\gamma$-$^{32}$P]-ATP and T4 kinase-labelled *C4A*- or *C4B*-specific oligonucleotides corresponding to amino acids 1101–1106 (Belt *et al.*, 1984). DNA from λ clones was prepared according to Maniatis *et al.* (1982).

*Restriction mapping and sub-cloning*

Detailed and accurate restriction maps of a long *C4* gene (22 kb) were obtained from sub-cloned restriction fragments from Cos 3A3, after single and double digests and Southern blot analysis. Most of the sites were confirmed by DNA sequencing (C.Y.Yu, unpublished results). Restriction maps of a short *C4* gene (16 kb) were constructed based on the *C4B 5* clone, λJM-2a, obtained in this work. For the *C4* λ clones obtained from the AW, JM and AD libraries, a 2.3 kb *Sma*I and/or 927 bp *Bam*HI fragment corresponding to the polymorphic C4d region were sub-cloned into the plasmid vector pAT153/PvuII/8.

*Synthesis of oligonucleotide as sequencing primer*

A 20-mer oligonucleotide (5′ CAGGAGACATCTAACTGGCT 3′) was chemically synthesized by the phosphotriester method according to Sproat and Gait (1984). This oligonucleotide corresponds to the DNA sequence 30 nucleotides 5′ to the C4d isotypic region described in this paper.

*DNA sequencing*

Complete genomic DNA sequence coding for the C4d region from Cos 3A3 (Carroll *et al.*, 1984a) was obtained by Sanger's dideoxy termination method (Sanger *et al.*, 1977). Sub-cloned DNA from Cos 3A3 was randomly sonicated and cloned into M13mp9 (Messing and Vieira, 1982) as described previously (Belt *et al.*, 1985). DNA coding for the polymorphic C4d region from the 2.3 kb *Sma*I and 927 bp *Bam*HI fragments of λAW-3B, λJM-2a, λAD-1α and λAD-2α was determined by non-random cloning of specific DNA fragments into M13mp8 or mp9. Restriction enzymes used for generating specific fragments are shown in Figure 1c. In addition a chemically synthesized 20-mer oligonucleotide (indicated by an arrow, Figure 1c) was used as sequencing primer to confirm DNA sequence at the highly polymorphic region around amino acids 1101–1106. All DNA fragments were sequenced at least twice and mostly in both orientations.

*Rg and Ch typing*

Rg and Ch types of AW, JM and AD (Table I) were determined by haemagglutination inhibition techniques as published (Giles *et al.*, 1984; Giles, 1987). The B 3 allotype was shown by α chain and serological typing to be Rg: −1,−2; Ch: 1,2,3,4,5,6. The A 3a allotype (AD) was deduced from serological typing of the Q family (Bentley *et al.*, 1985) to express Rg: 1,2; Ch: −1,−2,−3,−4,−5,−6. The Rg1 was assigned to A 3a as Rg-1 is expected in A 1 (Rittner *et al.*, 1984) and Ch-6 was also assigned to A 3a as the majority of *C4A 3; C4B QO* haplotypes do not express Ch6 (Giles, 1987). B 5 (JM) is essentially the reverse of *C4A 1*. The B 5 allotype was deduced as Rg1 and Ch: −1,−2,−3,4−5,6 by means of α chain and serological testing (Hing *et al.*, 1986; Giles, 1987). The Rg2 on B 5 has not been determined. Ch6 is assigned to B 5 rather than to the *C4A 3; C4B QO* haplotype, as described.

## Acknowledgements

C.Yung Yu is a Commonwealth Scholar, K.Tertia Belt held an MRC Studentship. C.Y.Y. would like to thank the Unit Director, Dr Ken Reid, for his encouragement, particularly after the tragic death of the beloved Professor. We sincerely thank Ms Tracy Robson for Rg/Ch data on C4 α chains of the three blood donors, and Ms Angela Fielder for HLA and C4 phenotyping. We are indebted to Mike Carroll (Harvard University), Bob Sim, Edith Sim, David Bentley, Alex Law, Neil Barclay, Alister Dodds and Lai-chu Wu for advice and helpful discussions, to Ms Carolyn Brooks and Mr Frank Caddick for their help in preparing the manuscript, and to the blood donors for their co-operation. A minor portion of this work was orally presented at the XI International Complement Workshop, 3 November, 1985 in Miami, FL, and published in abstract form, *Complement*, **2**, 89–90.

## References

Awdeh,Z.L. and Alper,C.A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3576–3580.
Bell,G.I., Karam,J. and Rutter,W.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 5759–5763.
Belt,K.T., Carroll,M.C. and Porter,R.R. (1984) *Cell*, **36**, 907–914.
Belt,K.T., Yu,C.Y., Carroll,M.C. and Porter,R.R. (1985) *Immunogenetics*, **21**, 173–180.
Benjamin,D.C., Berzofsky,J.A., East,I.J., Gurd,F.R.N., Hannum,C., Leach,S.J., Margoliash,E., Michael,J.G., Miller,A., Prager,E.M., Reichlin,M., Sercarz,E.E., Smith-Gill,S.J., Todd,P.E. and Wilson,A.C. (1984) *Annu. Rev. Immunol.*, **2**, 67–101.
Bentley,D.R., Campbell,R.D. and Cross,S.J. (1985) *Immunogenetics*, **22**, 377–390.
Campbell,R.D., Dodds,A.W. and Porter,R.R. (1980) *Biochem. J.*, **189**, 67–80.
Campbell,R.D., Gagnon,J. and Porter,R.R. (1981) *Biochem. J.*, **199**, 359–370.
Carroll,M.C., Campbell,R.D., Bentley,D.R. and Porter,R.R. (1984a) *Nature*, **307**, 237–241.
Carroll,M.C., Belt,K.T., Palsdottir,A. and Porter,R.R. (1984b) *Phil. Trans. R. Soc. London, B.*, **306**, 379–388.
Carroll,M.C., Belt,K.T., Palsdottir,A. and Yu,C.Y. (1985a) *Immunol. Rev.*, **87**, 39–60.
Carroll,M.C., Palsdottir,A., Belt,K.T. and Porter,R.R. (1985b) *EMBO J.*, **4**, 2547–2552.
Carroll,M.C., Palsdottir,A., Belt,K.T. and Yu,C.Y. (1986) *Biochem. Soc. Symp.*, **51**, 29–36.
Chan,A.C. and Atkinson,J.R. (1985) *J. Immunol.*, **134**, 1790–1798.
Dawkins,R.L., Christiansen,F.T., Kay,P.H., Garlepp,M., McCluskey,J. Hollingsworth,P.N. and Zilko,P.J. (1983) *Immunol. Rev.*, **70**, 5–22.
De Bruijn,M.H.L. and Fey,G.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 708–712.
Dodds,A.W., Law,S.K. and Porter,R.R. (1986) *Immunogenetics*, **24**, in press.
Fielder,A.H.L., Walport,M.J., Batchelor,J.R., Rynes,R.I., Black,C.M., Dodi,I.A. and Hughes,G.R.V. (1983) *Br. Med. J.*, **286**, 425–428.
Frischauf,A.-M., Lehrach,H., Poustka,A. and Murrey,N.E. (1983) *J. Mol. Biol.*, **170**, 827–842.
Giles,C.M. (1985a) *Vox Sang.*, **48**, 160–166.
Giles,C.M. (1985b) *Vox Sang.*, **48**, 167–173.
Giles,C.M. (1987) *Human Immunol.*, in press.

Giles,C.M., Batchelor,J.R., Dodi,I.A., Fielder,A.H.L., Rittner,C., Mauff,G., Bender,K., Levene,C., Schreuder,G.M.T. and Well,L.J. (1984) *J. Immunogenet.*, **11**, 305–317.
Hall,R.E. and Colten,H.R. (1977) *J. Immunol.*, **118**, 1903–1905.
Hauptmann,G., Goetz,J., Uring-Lambert,B. and Grosshans,E. (1986) *Prog. Allergy*, in press.
Hellman,U., Eggertsen,G., Lundwall,A., Engstrom,A. and Sjoquist,J. (1984) *FEBS Lett.*, **170**, 254–258.
Hing,S.N., Giles,C.M., Fielder,A.H.L. and Batchelor,J.R. (1986) *Immunogenetics*, **23**, 151–155.
Isenman,D.E. and Kells,D.I.C. (1982) *Biochemistry*, **21**, 1109–1117.
Isenman,D.E. and Young,J.R. (1984) *J. Immunol.*, **132**, 3019–3027.
Isenman,D.E. and Young,J.R. (1986) *J. Immunol.*, **136**, 2542–2550.
Janatova,J. (1986) *Biochem. J.*, **233**, 819–825.
Kan,C.C., Solomon,E., Belt,K.T., Chain,A.C., Hiorns,L.R. and Fey,G. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2282–2286.
Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
Law,S.K., Lichetenberg,N.A., Holcombe,F.H. and Levine,R.P. (1980) *J. Immunol*, **125**, 634–639.
Law,S.K.A., Dodds,A.W. and Porter,R.R. (1984a) *EMBO J.*, **3**, 1819–1823.
Law,S.K.A., Minich,T.M. and Levine,R.P. (1984b) *Biochemistry*, **23**, 3267–3272.
Levi-Strauss,M., Tosi,M., Steinmetz,M., Klein,J. and Meo,T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1746–1750.
Lundwall,A., Hellman,U., Eggertsen,G. and Sjöquist,J. (1982) *Mol. Immunol.*, **19**, 1655–1665.
Maniatis,T., Frisch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Mauff,G., Alper,C.A., Awdeh,Z., Batchelor,J.R., Bertrams,T., Bruun-Petersen,G., Dawkins,R.L., Demant,P., Edwards,J., Grosse-Wilde,H., Hauptmann,G., Klouda,P., Lamm,L., Mollenhauer,E., Nerl,C., Olaisen,B., O'Neill,G., Rittner,C., Ross,M.H., Skanes,V., Teisberg,P. and Wells,L. (1983) *Immunobiology*, **164**, 184–191.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Nonaka,M., Nakayama,K., Yeul,Y.D. and Takahashi,M. (1985) *J. Biol. Chem.*, **20**, 10936–10943.
Nonaka,M., Nakayama,K., Yeul,Y.D. and Takahashi,M. (1986) *J. Immunol.*, **136**, 2989–2993.
Ogata,R.T. and Sepich,D.S. (1985) *J. Immunol.*, **135**, 4239–4244.
O'Neill,G.J., Yang,SY., Tegoli,J., Berger,R. and DuPont,B. (1978) *Nature*, **273**, 668–670.
Petes,T. and Fink,G.R. (1982) *Nature*, **300**, 216–217.
Porter,R.R. (1983) *Mol. Biol. Med.*, **1**, 161–168.
Porter,R.R. (1984) *CRC Crit. Rev. Biochem.*, **16**, 1–19.
Porter,R.R. (1985) *Immunol. Rev.*, **87**, 7–17.
Reid,K.B.M. and Porter,R.R. (1981) *Annu. Rev. Biochem.*, **50**, 433–464.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.
Rittner,C., Giles,C.M., Roos,M.H., Demant,P. and Mollenhauer,E. (1984) *Immunogenetics*, **19**, 321–333.
Roos,M.H., Mollenhauer,E., Demant,P. and Rittner,C. (1982) *Nature*, **298**, 854–855.
Roos,M.H., Giles,C.M., Demant,P., Mollenhauer,E. and Rittner,C. (1984) *J. Immunol.*, **133**, 2634–2640.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5468.
Schneider,P.M., Rittner,C., Carroll,M.C. and Colten,H.R. (1985) *Complement*, **2**, 71 (Abstract).
Sepich,D.S., Noonan,D.J. and Ogata,R.T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5895–5899.
Singer,M.F. (1982) *Cell*, **28**, 433–434.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Sproat,B.S. and Gait,M.J. (1984) In Gait,M.J. (ed.), *Oligonucleotide Synthesis*. IRL Press, Oxford, pp. 83–115.
Tack,B.F. (1985) In Müller-Eberhard,H.J. and Meischer,P.J. (eds), *Complement*. Springer-Verlag, Berlin, pp. 49–72.
Tiesberg,P., Akesson,I., Olaisen,B., Gedde-Dahl,T.,Jr and Thorsby,E. (1976) *Nature*, **264**, 253–254.
Tilley,C.A., Romans,D.G. and Crookston,M.C. (1978) *Nature*, **276**, 713–715.
Weiss,E.H., Mellor,A., Golden,L., Fahrner,K., Simpson,E., Hurst,J. and Flavell,R.A. (1983a) *Nature*, **301**, 671–673.
Weiss,E., Golden,L., Zakut,R., Mellor,A., Fahrner,K., Kvist,S. and Flavell,R.A. (1983b) *EMBO J.*, **2**, 453–462.
Wetsel,R.A., Lundwall,A., Davidson,F., Tack,B.F. and Fey,G.H. (1984) *J. Biol. Chem.*, **259**, 13857–13862.
White,P.C., Grossberger,D., Onufer,B.J., New,M.I., Dupont,B. and Strominger,J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1089–1093.

*Received on 21 July 1986*