Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Structure and Organization of the C4 Genes

M. C. Carroll, Tertia Belt, Astridur Palsdottir and R. R. Porter

Phil. Trans. R. Soc. Lond. B 1984 **306**, 379-388
doi: 10.1098/rstb.1984.0098

References Article cited in:
http://rstb.royalsocietypublishing.org/content/306/1129/379#related-urls

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click here

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: http://rstb.royalsocietypublishing.org/subscriptions

This journal is © 1984 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Phil. Trans. R. Soc. Lond. B **306**, 379–388 (1984)
Printed in Great Britain

Structure and organization of the *C4* genes

BY M. C. CARROLL, TERTIA BELT, ASTRIDUR PALSDOTTIR
AND R. R. PORTER, F.R.S.

*M.R.C. Immunochemistry Unit, Department of Biochemistry, South Parks Road, Oxford OX*1 3*QU, U.K.*

This 200000 M_r serum protein is coded for by at least two separate loci, *C4A* and *C4B*, which map in the HLA Class III region on chromosome 6 in man. Both loci are highly polymorphic with more than 30 alleles, including null alleles assigned to the two loci.

The complete nucleotide sequence of a full length C4A cDNA clone and a substantial part of a C4B cDNA clone has shown class differences which can be used to synthesize nucleotide probes specific for C4A and C4B.

Three *C4* loci of approximately 16 kilobases each spaced by 10 kilobases have been identified in DNA from one individual and aligned 30 kilobases from the *factor B* gene by overlapping cloned genomic fragments from a cosmid library.

Characterization of these genes by restriction mapping, nucleotide sequence analysis and hybridization with C4A and C4B specific synthetic oligonucleotides show that these genes are very similar.

The fourth component of human complement is synthesized in the liver and by macrophages as a single chain promolecule of approximately 200000 M_r, then processed before secretion into the serum in its native three chain structure, that is, α, 95000 M_r; β, 75000 M_r; and γ, 35000 M_r (Hall & Colten 1977; Roos *et al.* 1978).

The α chain which is important to the function of the C4 protein contains sites involved in: activation, covalent binding, inactivation, C4A and C4B structural differences, and allelic differences. Activation by $C\bar{1}\bar{s}$ releases a 7 kDa peptide, C4a, from the N-terminus of the alpha chain and releases a reactive acyl group in the C4d region of the α chain, which may form a covalent bond with adjacent immune complex or cell surfaces (Law *et al.* 1980; Campbell *et al.* 1980). This covalent bond, often with the Fd region of the Ig heavy chain (Campbell *et al.* 1980) serves to focus the C3 convertase activity at the site of activation. Coupling of C4 as well as C3 which also forms a covalent interaction with the immune complex (Law *et al.* 1979) may facilitate clearance of complexes by both disruption of aggregates and facilitating uptake through C3 and C4 specific cell surface receptors on red blood cells and monocytes (Bianco & Nussenzweig 1977; Schifferli *et al.* 1982). Further cleavage of the α chain by Factor I and C4bp releasing a 44000 M_r internal peptide, C4d, results in inactivation of C4.

The fourth component is coded for by two separate but closely linked loci, that is, *C4A* and *C4B* (O'Neill *et al.* 1978; Olaisen *et al.* 1979), which map in the HLA class III region along with *C2* and *FB* on chromosome 6 in man. The complement genes *C2*, *FB* and *C4A* and *C4B* are closely linked within this region of approximately 0.7 centimorgans (cM) between HLA B and D (Barnstable *et al.* 1979; Olaisen *et al.* 1983). This region of HLA also appears to be important in susceptibility to a number of autoimmune diseases such as rheumatoid arthritis

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.* STOR www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(Batchelor & Welsh 1982), insulin-dependent diabetes (McCluskey *et al.* 1983) and systemic lupus erythematosus (Fielder *et al.* 1983). Certain combinations of alleles at each of the loci within HLA referred to as extended haplotypes (Awdeh *et al.* 1983) appear to be conserved, that is, frequency of that haplotype is significantly greater than predicted given normal frequency of recombination. Thus phenotypic markers mapping to the class III region may be useful as markers in linkage studies of HLA and disease susceptibility.

While limited structural comparison of C4A and C4B proteins suggest similarity, differences have been detected mainly in the C4d region of the α chain by serology (Tilley *et al.* 1978), electrophoresis (Awdeh & Alper 1980) and by limited amino acid sequencing (Chakravarti *et al.* 1983). On the basis of charge separation by electrophoresis both loci are very polymorphic with 13 alleles at *C4A* and 22 alleles at *C4B* locus including null alleles at each locus (Mauff *et al.* 1983). Many of the charge differences appear to be localized to the C4d region of the α chain as shown by two-dimensional gel electrophoresis of separated chains (Mevag *et al.* 1981). Clustering of both class differences, that is, A and B differences, as well as allelic differences in the C4d region near the internal thioester, probably would affect the efficiency of covalent binding between C4 and the immune complex.

Variation in the number of *C4* genes expressed has been suggested from family studies where there is evidence for duplication of either the *A* or *B* loci (Raum *et al.* 1984). Thus there are haplotypes with one, two or three active *C4* genes.

To determine the complete C4 structure, specific structural differences between C4A and C4B proteins and between two *C4A* alleles, the cDNA of a complete C4A mRNA has been sequenced and compared to the partly complete derived amino acid sequence of a C4B clone and a second *C4A* allele. Nucleotide differences were detected and have been tentatively identified as class differences between A and B and allelic differences between the two *A* alleles (Belt *et al.* 1984). The *C4* genes have been mapped relative to each other and to the *C2* and *factor B* genes, by preparing a molecular map representing 120 kilobases of genomic DNA in the HLA class III region (Carroll *et al.* 1984).

Results and discussion

The structure of C4

To determine the structure of the C4 protein the C4 mRNA was cloned and the nucleotide sequence of a C4 cDNA clone was determined (Belt *et al.* 1984). The pro-C4 molecule is approximately 200000 M_r or an estimated 1700 amino acids with an mRNA size of greater than 5 kilobases. Therefore cDNA libraries were prepared from a 28 *S* fraction of total human liver RNA shown by Carroll & Porter (1983) to contain C4 message activity. To optimize for full length cDNA inserts, double strand DNA, prepared by the procedure of Wickens *et al.* (1978), was size-fractionated on sucrose gradients before blunt end cloning into the plasmid cloning vector pAT-153-PVU 11-8 (a gift from G. G. Brownlee). Two libraries, I and II, were prepared from size fractions estimated to contain DNA fragments of greater than 4 and 2–4 kilobases, respectively. Screening of both libraries using a C4-specific cDNA probe, pAlu-7 (Carroll & Porter 1983) identified approximately one positive clone per thousand.

One recombinant from library I, pAT-A with an insert size of 5.5 kilobases and two recombinants from library II, pAT-F and pAT-42 with inserts of 4.8 and 2.5 kilobases, respectively, were purified for nucleotide sequence analysis. The 5.5 kilobases DNA insert was

Downloaded from rstb.royalsocietypublishing.org

shown to contain the complete coding sequence of pro-C4 mRNA using the method of Maxam & Gilbert (1977) to sequence of the 5′ and 3′ ends. While the 3′ end coded for both a polyadenylation signal, that is, ATTAAA, and a poly-A tail, the 5′ end coded for the known N-terminal sequence of pro-C4 preceded by a possible leader sequence of 13 hydrophobic residues. The initiation codon for methionine was not seen as this region of the mRNA probably was lost in the loop-back procedure used in preparing the cDNA. For complete nucleotide sequence analysis, the total insert was sheared and blunt end fragments of 500–900 base pairs were cloned into the M13 vector mp-8 (Messing & Vieira 1982). Recombinants were selected at random and the insert sequence determined using the dideoxy procedure of Sanger (Sanger & Coulson 1978).

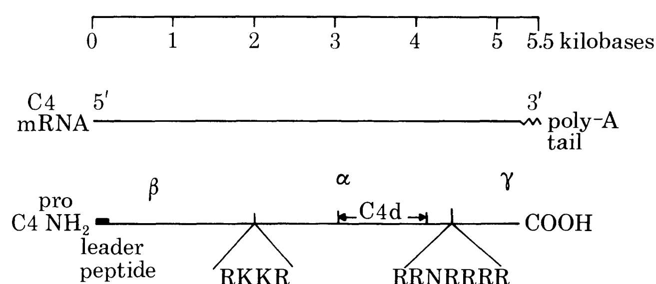


FIGURE 1. Diagram of C4 mRNA and pro C4 protein. The derived amino acid sequence of pro C4 was determined by nucleotide sequence analysis of 5.5 kilobase cDNA insert from recombinant plasmid pAT-A (see Belt *et al.* (1984) for sequence). Preceding the known N-terminus of the pro C4 protein was a stretch of 13 hydrophobic amino acids which probably represent a leader peptide. Pro C4 is split before secretion into three chains, α, β and γ. Processing of the single polypeptide chain requires excision of the basic tetrapeptide Arg-Lys-Lys-Arg between the β and α chains and a basic heptapeptide Arg-Arg-Asn-Arg-Arg-Arg-Arg between the α and γ chains.

Comparison of the derived amino acid sequence (Belt *et al.* 1984) with the regions of known sequence (Gigli *et al.* 1977; Moon *et al.* 1981; Press & Gagnon 1981; Chakravarti *et al.* 1983; D. N. Chakravarti, R. D. Campbell and J. Gagnon, unpublished; S. K. Law and J. Gagnon, unpublished) showed good agreement with several minor differences that may be due to polymorphism. The amino acid sequence of the N-terminal ends of each of three chains, that is, β, α and γ was known (Gigli *et al.* 1977) as well as the C terminus of both the β and γ chains (S. K. Law and J. Gagnon, unpublished). The C terminal sequence of the alpha chain was inferred from the pro-C4-derived amino acid sequence immediately preceding the N-terminus of the γ chain. The derived sequence shows a basic heptapeptide of Arg-Arg-Asn-Arg-Arg-Arg-Arg which agrees with the tetra-arginine peptide sequence reported for both human and mouse pro C4 (Whitehead *et al.* 1983; Ogata *et al.* 1983, respectively). The basic heptapeptide may be excised as has been proposed for a similar basic tetra-arginine of α chain in human pro C3 (Domdey *et al.* 1982). Likewise the basic peptide Arg-Lys-Lys-Arg between the known C terminus of the β chain and N terminus of the α chain must be excised during processing of the pro C4 molecule into its native three chain structure. The excision reaction may be catalysed by a similar endopeptidase of trypsin-like specificity followed by the exopeptidase carboxypeptidase B as has been shown to occur in the release of some hormones from precursor molecules (Lazure *et al.* 1983).

The derived amino acid sequence of the region containing the internal thioester, that is, Cys-Gly-Glu-Gln where the thiol group of cysteine reacts with the acyl group of glutamine,

Downloaded from rstb.royalsocietypublishing.org

shows a glutamine rather than a glutamic as has been proposed from protein sequence (Campbell *et al.* 1981). This agrees with similar derived sequence of pro C3 in the thioester region (Domdey *et al.* 1982).

Polymorphism of C4

Two forms of C4 isolated from pooled human serum have been identified by separation by ion exchange chromatography of large tryptic fragments, that is, 30000 and 28000 M_r (Lundwall *et al.* 1981). The larger fragment was shown serologically to come from C4A and the smaller from C4B. Amino acid sequencing showed that C4A contained the sequence Asp-*Pro-Cys*-Pro-Val-*Leu-Asp*-Arg whereas the C4B fragment had in the corresponding region the sequence of Asp-*Leu-Ser*-Pro-Val-*Ile-His*-Arg (Hellman *et al.* 1984).

Comparison with the derived amino acid sequences showed that the cDNA clones pAT-A and pAT-42 had the C4A sequence while pAT-F contained the C4B sequence between amino acid residues 1100 and 1107 (pro C4 numbering). Thus clones pAT-A and pAT-42 represent

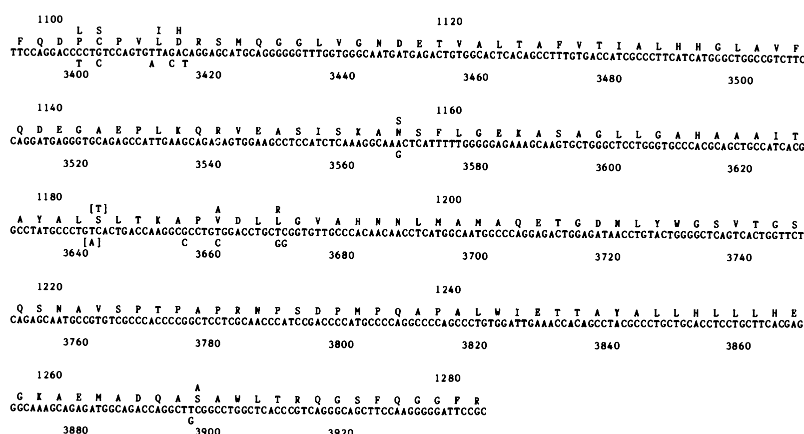


FIGURE 2. Polymorphisms in the nucleotide and derived amino acid sequence from the C4d section of the C4 alpha chain. Numbering is from the N-terminus of the pro molecule and gives the sequence of pAT-A, C4A. The alternative residues are those of the pAT-F, C4B, but the residue in brackets is also present in pAT-42, C4A, suggesting that this may be an allelic variation of C4A.

C4A and clone pAT-F represents C4B therefore the four differences seen in positions 1100–1107 probably represent C4A and C4B differences and tentatively have been referred to as class differences. These results may provide a structural basis for the C4d region serological determinants recognized by the anti-Rodgers, or C4A and anti-Chido, C4B antisera described by Tilley *et al.* (1978). Further the Asp-His change at position 1107 may contribute to the charge difference between the respective A and B proteins. The location of these differences within the same region as the thioester may be partly responsible for the differences in haemolytic activity between A and B proteins.

Synthetic oligonucleotide probes specific for the C4A and C4B nucleic acid sequence between position 1100 and 1107 have been prepared. As will be described later these probes have been used to identify *C4* genes on cloned fragments as either *C4A* or *C4B*.

In the C4d section four other differences between C4A and C4B-derived sequences have been determined at positions 1157, 1188, 1191, and 1267. However, it is not known if these differences are specific for A and B or if they are allelic as the sequence of Hellman *et al.* did

Downloaded from rstb.royalsocietypublishing.org

not extend through this region. Of the remaining derived sequence compared outside the C4d region, that is, 223 amino acids in the α chain and 291 amino acids in the γ chain, only one difference was seen, that is a Tyr–Asp in the γ chain.

One possible allelic difference was observed between the two C4A sequences at position 1182. The pAT-A derived sequence showed serine whereas both the second C4A sequence, that is,

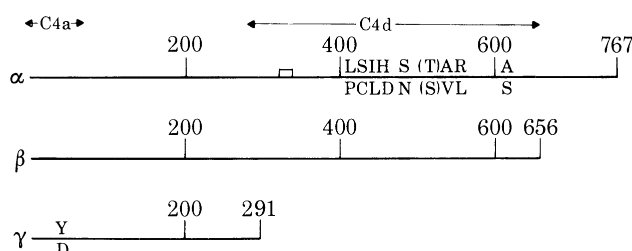


FIGURE 3. Polymorphisms in the derived amino acid sequence of the three chain C4 molecule. Comparison of the derived amino acid sequence of cDNA clones pAT-A, C4A, and pAT-F, C4B, at a total of 874 positions showed nine substitutions with the C4A sequence on the lower line. Eight of the substitutions were in the C4d section of the α chain wheras only one substitution was seen in the remaining 494 positions examined outside of C4d section, that is, 203 in α chain and 291 in γ chain.

pAT-42 and the C4B sequence showed a threonine. With this one exception the two C4A-derived sequences were identical. Since the cDNA library was prepared from liver tissue of a donor typed as heterozygous at both *C4A* and *C4B* loci, that is, *C4A*, 4, 3 and *C4B* 1, 2 (S.Cross, A. Palsdottir and J. Edwards, unpublished), the Ser–Thr difference probably represents an allelic variation between the *C4A* 4 and 3 proteins. It is interesting that an allele of *C4A*, that is pAT-42, was identical to an allele of *C4B*, that is, pAT-F at this site of *C4A* allelic variation. This sort of allelic difference, where an allele of A appears more similar at a given residue of B than A might be expected from the protein typing results where some alleles of C4A overlap in charge with some alleles of C4B (Mauff *et al.* 1983). Similarly, some antisera show cross reactions between different alleles of C4A and C4B (Rittner *et al.* 1983).

Typing by Southern analysis of genomic DNA and correlation of restriction fragment length polymorphism (r.f.l.p.) with specific protein variants would improve the resolution of C4 typing. This procedure has been used recently by Palsdottir *et al.* (1983) to detect the C4A 6 variant. Whitehead *et al.* (1984) have shown that additional variants not seen by protein typing methods may be identified using Southern analysis. However, this technique has the limitation that only nucleotide changes which affect known sites of cleavage by restriction endonucleases will be detected. None of the sequence differences between the three cDNA clones described here would be detected by the restriction enzyme available at present.

A more general procedure for detection of variants is based on the use of oligonucleotides complementary to sequences with known structural differences. Thus, oligonucleotides specific for the variant sequence, which may differ by only one nucleotide are synthesized and then hybridized to Southern blots of genomic DNA. This procedure has been used successfully in detecting allelic variants of haemoglobin (Orkin *et al.* 1983) and alpha-1-anti-trypsin (Kidd *et al.* 1983). Oligonucleotides specific for the C4A and C4B nucleotide sequences in the region of 1100 to 1107 have been used successfully to differentiate between the *C4A* and *C4B* genes in cloned genomic fragments. These probes should prove useful also for C4A and C4B typing of genomic DNA.

BIOLOGICAL SCIENCES THE ROYAL SOCIETY PHILOSOPHICAL TRANSACTIONS OF
BIOLOGICAL SCIENCES THE ROYAL SOCIETY PHILOSOPHICAL TRANSACTIONS OF

Downloaded from rstb.royalsocietypublishing.org

Several additional differences have been identified in the known amino acid sequence from pooled C4 and derived amino acid sequence of the three cDNA clones. One of these differences, that is, position 616, Cys–Ser, is in the β chain while the other three, that is, positions 1054 (Asp–Gly) 1090 (Ile–Ser) and 1281 (Arg–Val) are located in the C4d region. Of the total combined, known and 874 derived amino acid sequences compared, 13 out of the 14 variant positions were localized in the C4d region. Within this region, only the four variant positions between 1100 and 1107 are adjacent but in the tertiary structure the variant residues may be closer owing to the folding of the chain.

One striking feature of the comparison of the nucleotide sequence of C4A and C4B is the tenfold fewer silent changes compared to replacement changes. Only one silent substitution, that is, a nucleotide change which does not alter amino acid sequence, was seen compared to 11 replacement substitutions, that is, changes which alter the amino acid sequence, of the total 2622 nucleotides compared.

These results suggest there may be a mechanism for maintaining identity between the A and B sequences except in certain positions where differences may be of a biological advantage.

Molecular map

To determine the organization of *C4* genes a cosmid library of human genomic DNA prepared from the DNA of white blood cells by Grosveld *et al.* (1982) was screened using the *C4*-specific cDNA probes. The HLA type and complotype of the DNA donor, that is, HPFH, was: A3, 31; B14, 37; DR1, 2; C4A 2, 3; C4B 2, 1; FB S, C2 C. Although family typing was not available the haplotype *B14*, *DR1*, *C4A 2*, *C4B 2*, *FB S*, *C2 C* was inferred as this haplotype has been shown to exist as an extended haplotype in the population (Awdeh *et al.* 1983). This *B14*, *DR1* haplotype has also been shown in cases from informative families to have a duplication at the *C4B* locus, that is, three active *C4* genes *C4A 2*, *C4 B2*, *C4 B1* (Raum *et al.* 1984).

By using the C4 cDNA probe, three overlapping cosmid clones with genomic inserts of approximately 35–40 kilobases were isolated and characterized. A molecular map of approximately 120 kilobases was prepared by restriction mapping using a collection of 10 restriction endonucleases. Not all restriction sites are shown on the map. The map was extended another 40 kilobases when it was shown that cosmid 3A3 overlapped approximately 8 kilobases with cosmid 1a which contained the *factor B* and *C2* genes (Campbell & Porter 1983). Finding these five complement genes within 140 kilobases may explain why recombinants among the complement genes have not been observed. The *C2* and *factor B* genes while separated by less than 2 kilobases were 30 kilobases away from the *C4* genes which were separated from each other by approximately 10 kilobases. Analysis of the cosmid clones using the C4A and C4B class specific synthetic oligonucleotides showed that there was a single *C4A* gene and two *C4B* genes. While orientation of the complement genes with respect to the *HLA D* and *B* loci has not been determined, Olaisen *et al.* (1983) have proposed on the basis of existing haplotypes that the *C4* genes were closer to the *HLA B* locus than *factor B*.

Evidence that the molecular map was representative in the population was obtained by analysis of a series of Southern blots using uncloned genomic DNA of various individuals and hybridizing with either coding sequence probes or genomic probes prepared from non-coding regions. By demonstrating that the probes hybridized to common restriction fragments as predicted from the map, coding sequence probes from *C2* were linked to *factor B* and then to *C4A* through linking of flanking region probes E and F (Carroll *et al.* 1984)

Downloaded from rstb.royalsocietypublishing.org

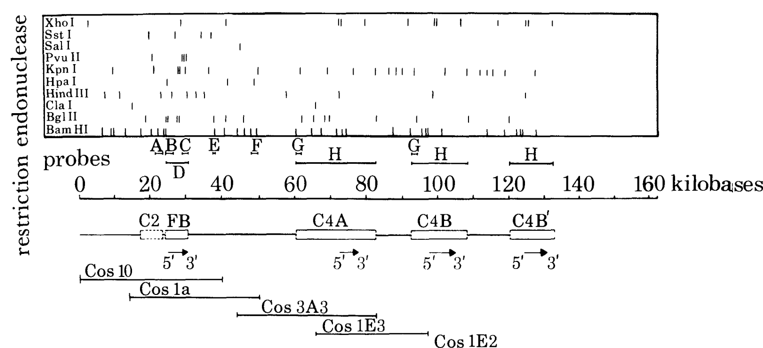


FIGURE 4. Molecular map of 160 kilobases of genomic DNA in the HLA class III region containing five complement genes.

Approximately 200000 colonies from the HPFH cosmid library prepared by Grovsveld *et al.* (1982) from genomic DNA extracted from white blood cells of a female donor, were screened with C4 and factor B cDNA probes. DNA inserts from positive clones were characterized further by restriction mapping and Southern analysis using probes A–H as described below. Six cosmid clones 10, 1a, 3A3, 1E3, and 1E2, were shown to overlap covering a distance of 140 kilobases. The precise limits of the *C2*, and three *C4* genes were not determined, however use of a complete coding sequence probe, probe H on the *C4* genes allowed approximate limits to be determined.

Probes used: A, 400 base pairs (b.p.) C2 cDNA-pC201 (Bentley & Porter 1984); B, 660 b.p. *Cla 1/Bam H1* fragment from 5′ end of full length factor B cDNA clone pFB3b; C, 500 b.p. *Xho I/Bam HI* fragment from 3′ end factor B cDNA clone pFB3b; D, 2.4 factor B cDNA clone – pFB3b (Morley & Campbell 1984); E, 600 b.p. *Sst I/Bam HI* fragment from Cos 1a (Campbell & Porter 1983); F, 1.6 kilobases *Bam HI* fragment from Cos 3A3; G, 400 b.p. *Bam HI/Kpn I* fragment from 5′ end of full length C4 cDNA clone pAT-A; H, 5.5 kilobases C4 cDNA clone (Belt *et al.* 1984).

The 3′ end of *C4A* and 5′ end of the adjacent *C4B* gene were linked in a similar manner by using a 2.1 *Kpn I* genomic fragment probe isolated from the flanking region between the *A* and *B* genes (not shown on map).

Comparison of the restriction maps of the two *C4B* genes showed that the maps were similar. This is not surprising as they are likely to be very similar in sequence as has been found for the two *C4A* alleles. This similarity of restriction maps would explain the failure to detect a third *C4* gene by using the coding sequence probes by Southern blots of genomic DNA.

There were several restriction fragment differences between the *C4A* and *C4B* genes of this individual which permit detection of two genes using cDNA probes on Southern blots of uncloned DNA. For example, Southern analysis of uncloned genomic DNA from the library donor HPFH using the enzymes *Bam H1* and *Kpn I* and hybridizing with the 5′ end cDNA probe *G* shows two forms.

The 12 kilobase *Kpn I* and 4.6 *Bam H1* fragments represent the *C4A* gene while the 3.5 *Kpn I* and 3.3 *Bam H1* fragments represent the *C4B* gene.

In addition to the restriction fragment length difference between the *A* and *B* genes, the *A* gene appears to be approximately 7 kilobases longer, that is, 23 Kb and 16 Kb respectively. The 5′ ends of the genes were estimated using the probe G which represents the N-terminal end of the pro C4 protein.

The two proposed haplotypes for the donor HPFH are illustrated in figure 6. The *B37 DR2* haplotype represents the expected two *C4* loci, that is, *A* and *B*, chromosome as most haplotypes have only two active *C4* genes. However, since an overlapping cosmid from HPFH library with only one *C4B* gene has not been identified so far only the three *C4* gene haplotype map is known. Alternatively, the *B37* haplotype may have a single active *A3* gene and null at the *B* locus since the complement typing would not identify the number of *B1* genes expressed.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

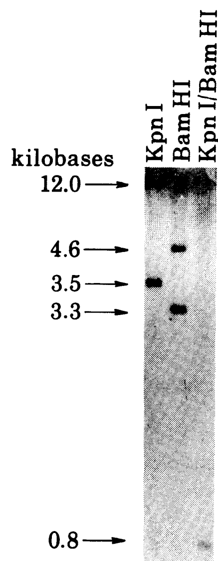


FIGURE 5. Two common forms of *C4* gene in genomic DNA. Southern analysis of *Kpn I*, *Bam HI*, or *Kpn I/BAM HI*, digest of genomic DNA from individual HPFH hybridized with a 400 b.p. cDNA probe, G, specific for 5′ coding sequence of a full-length clone, pAT-A (see figure 4 for position of probe). The 12 kilobases *Kpn I* and 4.6 *Bam HI* bands are characteristic of the *C4A* gene whereas the 3.5 *Kpn I* and 3.3 *Bam HI* fragments are characteristic of the two *C4B* genes.

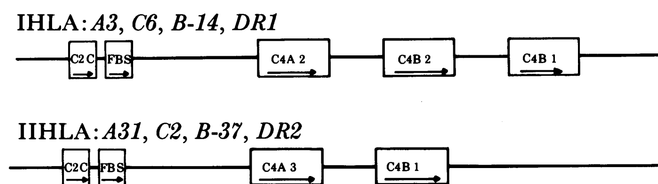


FIGURE 6. Proposed haplotype of HPFH. Chromosome I carries the *B14*, *DR1* haplotype and has a duplication of the *C4B* locus resulting in three *C4* genes, that is, *C4A 2*, *C4B 2*, *1*. The homologous chromosome, II, carries the *B37*, *DR2* haplotype and has two *C4* genes, that is, *C4A 3* and *C4B 1*.

The *B14*, *DR1* haplotype shown with a duplicated *B* locus agrees with previous reports that this extended haplotype may have one *C4A* and two *C4B* genes (Raum *et al.* 1984). Unequal crossing over would explain both duplication and deletion of *C4* loci as has been suggested for α globin genes (Dozy *et al.* 1979; Orkin *et al.* 1979). By this mechanism, misalignment during meiosis between homologous chromosomes containing *C4 A* and *B* gene each could result in an exchange of one locus, for example, the *B* locus, leaving one chromosome with three genes and the donor with only one. Results from the molecular map would be consistent with this type of mechanism. Haplotypes with a null allele, that is only one active *C4* gene, would represent the donor chromosome. Preliminary evidence from Southern analysis of individuals typed as homozygous for the *C4A* null allele suggests that in many cases one of the genes is missing (M. C. Carroll, A. Palsdottir and R. R. Porter, unpublished). However, it is unlikely that all *C4* null alleles are due to gene deletion.

Molecular mapping of the class III complement region in other individuals will provide a basis for further comparison of haplotypes with a null allele, and with two or three active genes.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

We thank Dr Ulf Hellman and colleagues for communicating the C4 amino acid sequence results before publication. We thank Professor J. Edwards for donation of blood samples and Ms Susan Cross for C4 and factor B typing.

References

Awdeh, Z. L. & Alper, C. A. 1980 Inherited structural polymorphism of the fourth component of human complement. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3576–3580.

Awdeh, Z. L., Raum, D., Yunis, E. J. & Alper, C. A. 1983 Extended HLA complement allele haplotypes: Evidence for T/t-like complex in man. *Proc. natn. Acad. Sci. U.S.A.* **80**, 259–263.

Barnstable, C. J., Jones, E. A. & Bodmer, W. F. 1979 In *Defense and recognition.* IIB. *Cellular aspects. Int. Rev. Sci. Ser. Biochem.* (ed. E. J. Lennox), **22**, 151–224. Baltimore: University Park Press.

Batchelor, J. R. & Welsh, K. I. 1982 Association of HLA antigens with disease. In *Clinical aspects of immunology* (ed. P. J. Lachmann and D. K. Peters). Oxford: Blackwell Scientific Publications.

Belt, K. T., Carroll, M. C. & Porter, R. R. 1984 The structural basis of the multiple forms of human complement component C4. *Cell* **36**, 907–914.

Bentley, D. R. & Porter, R. R. 1984 Isolation of cDNA clones for human complement component C2. *Proc. natn. Acad. Sci. U.S.A.* (In the press.) **81**, 1212–1215.

Bianco, C. & Nussenzweig, V. 1977 Complement receptors. *Contemp. Topics molec. Immunol.* **6**, 145–176.

Campbell, R. D., Dodds, A. W. & Porter, R. R. 1980 The binding of human complement component C4 to antibody-antigen aggregates. *Biochem. J.* **189**, 67–80.

Campbell, R. D., Gagnon, J. & Porter, R. R. 1981 Amino acid sequence around the thiol and reactive acyl groups of human complement component C4. *Biochem. J.* **199**, 359–370.

Campbell, R. D. & Porter, R. R. 1983 Molecular cloning and characterisation of the gene coding for human complement protein factor B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4464–4468.

Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. 1984 A molecular map of the major histocompatibility complex class III region of man linking the complement genes *C4*, *C2* and *Factor B*. *Nature, Lond.* **307**, 237–241.

Carroll, M. C. & Porter, R. R. 1983 Cloning of a human complement component *C4* gene. *Proc. natn. Acad. Sci. U.S.A.* **80**, 264–267.

Chakravarti, D. N., Campbell, R. D. & Gagnon, J. 1983 Amino acid sequence of a polymorphic segment from C4d of human complement component C4. *FEBS Lett.* **154**, 387–390.

Domdey, H., Wiebauer, K., Kazmaier, M., Muller, V., Odink, K. & Fey, G. 1982 Characterization of the mRNA and cloned cDNA specifying the third component of mouse complement. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7619–7623.

Dozy, A. M., Kan, Y. W., Embury, S. H., Mentzer, W. C., Wang, W. C., Lubin, B., Davis, J. R. & Koenig, H. M. 1979 α-Globin gene organization in blacks precludes the severe form of α-thalassaemia *Nature, Lond.* **280**, 605–607.

Fielder, A. H. L., Walport, M. J., Batchelor, J. R., Rynes, R. I., Black, C. M., Dodi, I. A. & Hughes, G. R. U. 1983 A family study of the MHC of patients with SLE. Null alleles of *C4A* and *C4B* may determine disease susceptibility. *Br. med. J.* **28 b**, 425–428.

Gigli, I., von Zabern, I. & Porter, R. R. 1977 The isolation and structure of C4, the fourth component of human complement. *Biochem. J.* **165**, 439–446.

Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. & Flavell, R. A. 1982 The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucl. Acids Res.* **10**, 6715–6731.

Hall, R. E. & Colten, H. R. 1977 Molecular size and subunit structure of the fourth component of guinea pig complement. *J. Immunol.* **118**, 1903–1905.

Hellman, U., Eggertsen, G., Lundwall, A., Engstrom, A. & Sjoquist, J. 1984 Primary sequence differences between Chido and Rodgers variants of tryptic C4d of the human complement system. *FEBS Lett.* (In the press.)

Kidd, V. J., Wallace, R. B., Itakura, K. & Woo, S. L. 1983 Human α-l antitrypsin deficiency detection by direct analysis of the mutation in the gene. *Nature, Lond.* **304**, 230–234.

Law, S. K., Lichtenberg, N. A., Holcombe, F. H. & Levine, R. P. 1980 Interaction between the labile binding sites of the fourth (C4) and fifth (C5) human complement proteins and erythrocyte cell membranes. *J. Immunol.* **125**, 634–639.

Law, S. K., Lichtenberg, N. A. & Levine, R. P. 1979 Evidence for an ester linkage between the labile binding site of C3b and receptive surfaces. *J. Immunol.* **123**, 1388–1394.

Lazure, C., Seidah, N. G., Pelaprat, D. & Chretin, M. 1983 Proteases and post translational processing of prohormones; a review. *Canad. J. Biochem. Cell Biol.* **61**, 501–515.

Lundwall, A., Hellman, U., Eggersten, G. & Sjoquist, J. 1981 Isolation of tryptic fragments of human C4 expressing Chido and Rodgers antigens. *Molec. Immunol.* **19**, 1655–1665.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

McCluskey, J., McCann, V. J., Kay, P. H., Zilko, P. J., Christiansen, F. T., O'Neill, G. J. & Dawkins, R. L. 1983 HLA and complement allotypes in type 1 (insulin-dependent) diabetes. *Diabetologia* **24**, 162–165.

Mauff, G., Alper, C. A., Awdeh, Z., Batchelor, J. R., Bertrams, T., Braun-Petersen, G., Dawkins, R. L., Demant, P., Edwards, J., Grosse-Wild, H., Hauptmann, G., Klonda, P., Lamm, L., Mullenhauer, E., Nerl, C., Olaisen, B., O'Neill, G. O., Rittner, C., Roos, M. H., Skanes, V., Teisberg, P. & Wells, L. 1983 Statement on the nomenclature of human C4 allotypes. *Immunobiology* **164**, 184–191.

Maxam, A. M. & Gilbert, W. 1977 A new method for sequencing DNA. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564.

Messing, J. & Vieira, J. 1982 A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**, 269–276.

Mevag, B., Olaisen, B. & Teisberg, P. 1981 Electrophoretic polymorphism of human C4 is due to charge differences in the α chain presumably in the C4α fragment. *Scand. J. Immunol.* **14**, 303–307.

Moon, K. E., Gorski, J. P. & Hugli, T. E. 1981 Complete primary structure of human C4a anaphylatoxin. *J. biol. Chem.* **256**, 8685–8692.

Morley, B. J. & Campbell, R. D. 1984 Internal homologies of the Ba fragment from human complement component factor B, a class III MHC antigen. *EMBO J.* **3**, 153–157.

Ogata, R. T., Schreffler, D. C., Sepich, D. S. & Lilly, S. P. 1983 cDNA clone spanning the α–γ subunit junction in the precursor of the murine fourth component of complement. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5061–5065.

Olaisen, B., Teisberg, R., Jonassen, R., Thorsby, E. & Gedde-Dahl, T. 1983 Gene order and gene distances in the HLA regions studied by the haplotype method. *A. Hum. Genet.* **47**, 285–292.

Olaisen, B., Teisberg, R., Nordhagen, R., Michaelson, T. & Gedde-Dahl, T. 1979 Human complement C4 is duplicated on some chromosomes. *Nature, Lond.* **279**, 736–737.

O'Neill, G. J., Yang, S. Y. & Dupont, B. 1978 Two HLA linked loci controlling the fourth component of human complement. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5165–5169.

Orkin, S. H., Markham, A. F. & Kazazian, H. H. 1983 Direct detection of the common mediterranean β thalassemia gene with synthetic DNA probes. *J. clin. Invest.* **71**, 775–779.

Orkin, S. H., Old, J., Lazarus, H., Altay, C., Gurgey, A., Weatherall, D. J. & Nathan, D. G. 1979 The molecular basis of α thalassemias: Frequent occurrence of dysfunctional α loci among non-Asians with HG H disease. *Cell* **17**, 33–42.

Palsdottir, A., Cross, S. J., Edwards, J. H. & Carroll, M. C. 1983 Correlation between a DNA restriction fragment length polymorphism and the C4 A6 protein. *Nature, Lond.* **306**, 615–616.

Press, E. M. & Gagnon, J. 1981 Human complement component C4. *Biochem. J.* **199**, 351–357.

Raum, D., Awdeh, Z., Anderson, J., Strong, L., Granados, J., Pevan, L., Giblett, E., Yunis, E. J. & Alper, C. A. 1984 Human C4 haplotypes with duplicated *C4A* or *C4B*. *Am. J. Hum. Genet.* (In the press.)

Rittner, C. L., Tippett, P., Giles, C. M., Mollenhauer, E., Berger, R., Nordhagen, R., Buskjoer, L., Petersen, G. B., Lamm, L. & Roos, M. H. 1983 An international reference typing for Ch and Rg determinants on rare human C4 allotypes. *Vox. Sang.* **46**, 224–225.

Roos, M. H., Atkinson, J. P. & Schreffler, D. C. 1978 Molecular size and characterization of the Ss and Slp (C4) proteins of the mouse H-2 complex. *J. Immunol.* **121**, 1106–1115.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977 DNA sequencing with chain terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5468.

Schifferli, J. A., Woo, J. H. & Peters, D. K. 1982 Complement mediated inhibition of immune precipitation: I. Role of the classical and alternative pathways. *Clin. exp. Immunol.* **45**, 555–562.

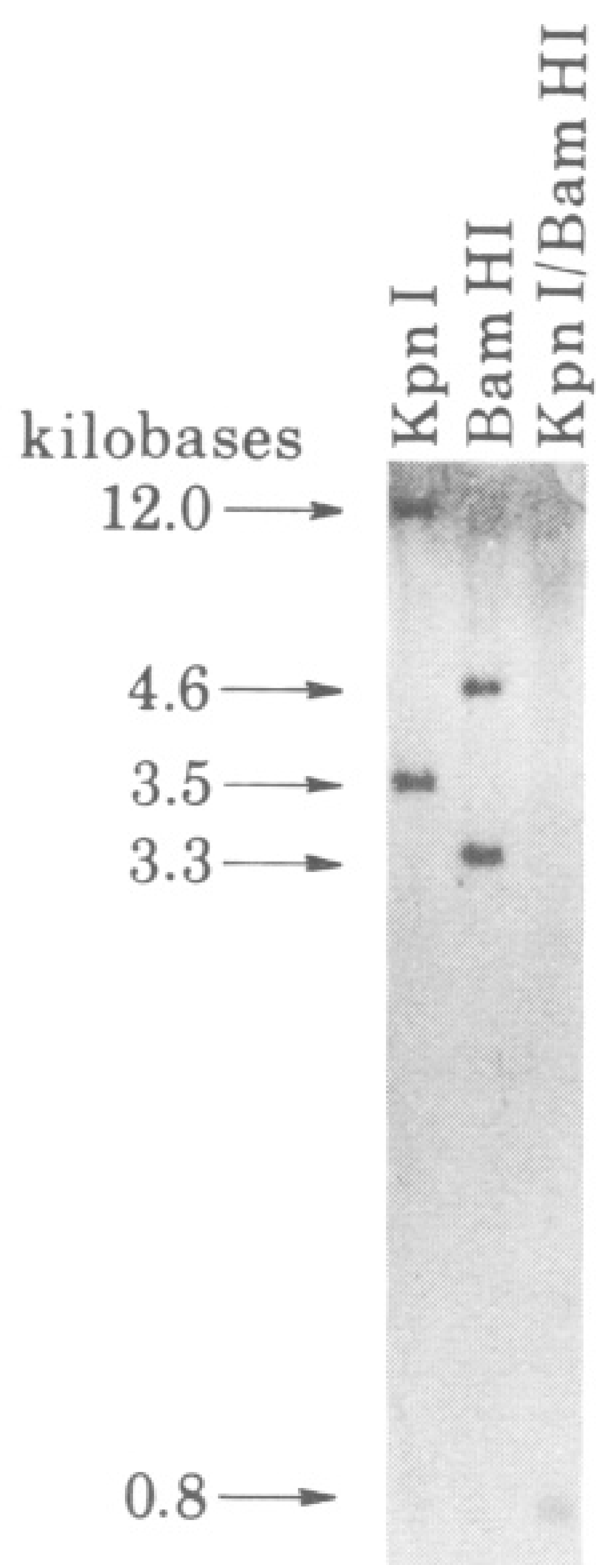
Tilley, C. A., Romans, D. G. & Crookson, M. C. 1978 Localization of Chido and Rodgers determinants to the C4d fragment of C4. *Nature, Lond.* **276**, 713–715.

Whitehead, A. S., Goldberger, G., Woods, D. E., Markham, A. F. & Colten, H. R. 1983 Use of a cDNA clone for the fourth component of human complement (C4) for analysis of a genetic deficiency of C4 in guinea pig. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5387–5391.

Whitehead, A. S., Woods, D. E., Fleischmek, E., Chin, J. E., Yunis, E., Katz, A. J., Gerald, P. S., Alper, C. A. & Colten, H. R. 1984 DNA polymorphisms of the *C4* genes: A new marker for analysis of the major histocompatibility complex. *New Engl. J. Med.* **310**, 88–91.

Wickens, M. P., Buell, G. N. & Schimke, R. T. 1978 Synthesis of double-stranded DNA complementary to lysozyme, ovomucoid, and ovalbumin mRNAs. *J. mol. Biol.* **253**, 2483–2495.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

FIGURE 5. Two common forms of *C4* gene in genomic DNA. Southern analysis of *Kpn I*, *Bam HI*, or *Kpn I/BAM HI*, digest of genomic DNA from individual HPFH hybridized with a 400 b.p. cDNA probe, G, specific for 5′ coding sequence of a full-length clone, pAT-A (see figure 4 for position of probe). The 12 kilobases *Kpn I* and 4.6 *Bam HI* bands are characteristic of the *C4A* gene whereas the 3.5 *Kpn I* and 3.3 *Bam HI* fragments are characteristic of the two *C4B* genes.