

## Isolation of a Complementary DNA Clone for the Human Complement Protein C2 and Its Use in the Identification of a Restriction Fragment Length Polymorphism

Derek E. Woods, Michael D. Edge, and Harvey R. Colten  
*Division of Cell Biology, Department of Medicine, Children's Hospital, Ina Sue Perlmutter Cystic Fibrosis Research Center, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115; Imperial Chemical Industries, Pharmaceuticals Division, Mereside, Alderly Park, Macclesfield, Cheshire, England*

**A**bstract. Complementary DNA (cDNA) clones corresponding to the major histocompatibility (MHC) class III antigen, complement protein C2, have been isolated from human liver cDNA libraries with the use of a complex mixture of synthetic oligonucleotides (17 mer) that contains 576 different oligonucleotide sequences. The C2 cDNA were used to identify a DNA restriction enzyme fragment length polymorphism that provides a genetic marker within the MHC that was not detectable at the protein level. An extensive search for genomic polymorphisms using a cDNA clone for another MHC class III gene, factor B, failed to reveal any DNA variants. The genomic variants detected with the C2 cDNA probe provide an additional genetic marker for analysis of MHC-linked diseases.

### Introduction

The genes for three of the serum complement proteins, the second (C2) and fourth (C4) components and factor B, have been localized to the major histocompatibility complex (MHC) in humans, mice, and guinea pigs (1) and have been designated class III MHC antigens. Human C4 is a three-chain glycoprotein ~200,000  $M_r$  coded for by two loci, C4A and C4B on the short arm of chromosome 6 (2). The products of these loci differ structurally (3) and functionally (4) and several protein polymorphic variants have been identified for each locus (4). Factors B and C2 are single chain glycoproteins of 95,000 (5) and 102,000

$M_r$  (6), respectively, which function in the alternative and classical pathways of complement activation. Several genetic variants of factor B protein have been identified by differences in mobility on agarose electrophoresis. These have been useful in analysis of populations and families with MHC-linked disorders such as insulin dependent juvenile onset diabetes mellitus (7).

C2 protein variants have been recognized by isoelectric focusing of serum samples (8) in polyacrylamide gels and detected functionally by overlaying with an agarose gel containing antibody-sensitized sheep erythrocytes and either C2-deficient human serum or dilute normal human serum. These reveal a common variant, C2C, a basic variant in 4–5% of unrelated individuals designated C2B, and rarely an acidic variant (1). Two such acidic C2 variants, designated C2 A1 and C2 A2 have been recognized in <1% of the population. Individually, the class III gene products display less polymorphism than the class I and class II proteins.

Recently, cDNA clones corresponding to human and mouse class I and II histocompatibility antigens (HLA and H2) (9–13) and the class III proteins factor B (14, 15) and C4 (16–18) have been isolated and characterized. The availability of cDNA probes for genes within the MHC have permitted an analysis of the fine structure of this region (13, 19). Additional structural data derived from nucleotide sequencing will lead to an understanding of the molecular basis of polymorphic variants important in regulation of the immune response.

DNA polymorphisms of the C4 genes have been identified by two groups (20, 21). These C4 DNA polymorphisms were detected among individuals who were homozygous as defined by C4 protein variants. This permitted subdivision of the C4 variants already identified and therefore generated additional markers within the MHC. These results prompted efforts to isolate human C2 cDNA clones that could be used together with factor B cDNA (14) probes to generate additional markers for analysis of MHC-linked diseases. Genomic polymorphisms at a given locus are detected by digestion of DNA samples with restriction endonucleases that recognize specific nucleotide se-

Received for publication 20 January 1984 and in revised form 5 April 1984.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/84/08/0634/05 \$1.00

Volume 74, August 1984, 634–638

quences. The resulting DNA fragments are size-separated electrophoretically and transferred to nitrocellulose filters. Radio-labeled DNA probes are then used to detect fragments that contain complementary sequences (22).

In this report, we describe the isolation of cDNA clones for human C2. The C2 clone and a factor B clone (14) have been used to investigate polymorphisms of these genes at the DNA level.

## Methods

**Isolation of C2 cDNA clones.** A 17-nucleotide long oligonucleotide mixture (Fig. 1) was synthesized by a solid-phase phosphotriester method (14). In position of nucleotide 7 and 13 for the glutamine (Gln) codons, 25% cytosine (C) was added as well as the guanosine (G) in order to obtain representation of the codon for glutamic acid (Glu). The oligonucleotide mixture was kinase labeled with [<sup>32</sup>P]dATP (23) and used to screen a human liver cDNA library, as described by Woods et al. (14).

Approximately 40,000 clones were screened. The final posthybridization washes were at 45°C in 6 × SSC (0.15 M NaCl, 0.015 M Na citrate, and 0.05% sodium pyrophosphate).<sup>1</sup> Positive clones were colony purified and rescreened as described above. The five positive clones isolated were used as probes in Northern blot analysis (24) against human liver mRNA to determine if they hybridized to a mRNA of approximately the correct size. One of the clones, pC2-10 (~500 base pairs (bp)), was partially sequenced by using the Maxam and Gilbert sequencing procedure (23). pC2-10 was used to rescreen 200,000 clones of the human adult liver cDNA library (14) and 15 additional C2 clones were identified. The largest of these was pC2-6 (900 bp). This insert was used to screen a human fetal liver cDNA library (25) and a C2 clone (pC2-F2) of 1.6 kilobases (kb) was isolated.

**Restriction enzyme analysis of human DNA.** Restriction enzyme digestions were carried out and DNA samples were analyzed by the Southern blot technique (22). The insert from the C2 clone (pC2-6) and the factor B clone (pBfA28) (14) were nick translated (26) and used as probes.

**Complement protein typing.** Blood for genetic typing of BF, C2, and C4 was collected into 1 mg/ml of sodium ethylenediaminetetra-acetate (Na<sub>2</sub> EDTA). Plasma was obtained by centrifugation and immediately frozen and stored at -80°C. The specimens were thawed immediately before analysis.

Plasma samples were subjected to agarose gel electrophoresis. Factor B patterns were visualized by immunofixation with goat anti-human B (Atlantic Antibodies, Scarborough, ME) as described previously (27). For C2 typing, plasma samples were subjected to isoelectric focusing in polyacrylamide gels (8) and C2 patterns were developed with an agarose gel overlay containing antibody-sensitized sheep erythrocytes and fresh normal human serum diluted 1:90 instead of C2-deficient human serum as originally described.

For C4 typing, plasma samples were desialated and detected by electrophoresis of desialated plasma in agarose gels and then immunofixed with goat anti-human C4 antiserum (Atlantic Antibodies) (4).

**Complement genetic nomenclature.** The nomenclature for genetic polymorphism of the complement proteins conforms to the International

System for Human Gene Nomenclature (1979) (28) as revised at the 4th International Workshop for the Genetics of Complement, Boston, July 1982. Alleles of the complotypes are written in the order BF, C2, C4A, and C4B although the true order of the genes is C2, BF, C4A, and C4B (29).

## Results

The C2 oligonucleotide mixture shown in Fig. 1 was kinase labeled and used as a probe to screen 40,000 clones of the human adult liver cDNA library. The nitrocellulose filters were washed in 6 × SSC at 45°C and autoradiographed. Under these conditions, five weakly hybridizing clones were identified. Analysis of these clones by hybridization to a Northern blot of human liver mRNA showed that only two clones hybridized to mRNA large enough to code for C2 i.e., ~2.7 kb. One of these clones, pC2-10 (~500 bp), was partially sequenced by using Maxam and Gilbert DNA sequencing; the results are shown in Fig. 2. This sequence represents the 5'-terminus of the clone, which corresponds to the *amino*-terminus of the C2a fragment (30) (the 70,000-D *carboxy*-terminal fragment of C2 generated by the C1 enzyme during C2 activation) plus the nucleotides for five amino acids of the *carboxy*-terminus of the C2b fragment. Fig. 2 also shows a comparison of the protein sequence generated from the cDNA sequence (a), the published sequence for the *amino*-terminus of human C2a (30) (b), and the published sequence of guinea pig C2a (30) (c). The amino acid sequence derived from the cDNA sequence is homologous to the published amino acid sequences for both human and guinea pig C2, thus identifying pC2-10 as a C2 cDNA clone. The amino acid sequence derived from the cDNA clone differs from the published human C2a protein sequence at residue 6 (Arg for Ser) and 10 (Leu for Lys). Differences between the derived human C2a sequences and the published sequence for guinea pig C2a (30) were evident at positions 18, 20, 24, 25, 27 and 28 (Fig. 2).

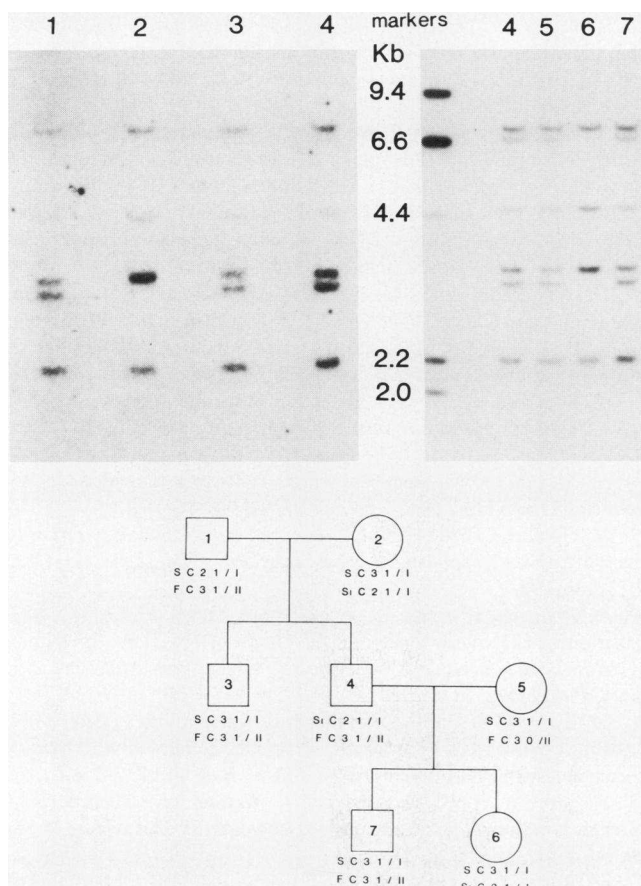
DNA isolated from leukocytes of six unrelated individuals

						Protein Sequence (30)
Lys	Ile	Gln	Ile	Gln	Ser	
1	2	3	4	5	6	
AAG	AUA	CAA	AUA	CAA	UC	mRNA Sequence
A	C(G)G		C(G)G	AG		
	U		U			
TTC	TAT	GTT	TAT	GTT	AG	Oligonucleotide Sequence
T	G(C)C		G(C)C	TC		
A			A			

**Figure 1.** Amino acid sequence of human C2a used to construct the synthetic oligonucleotide mixture, which was then used to isolate C2 cDNA clones. The derived potential mRNA sequences and the oligonucleotide sequence are shown.

1. Abbreviations used in this paper: SSC, 0.15 M NaCl, 0.015 M Na citrate, and 0.05% sodium pyrophosphate.





**Figure 4.** Inheritance pattern of protein complotypes and C2 DNA variants. Complotypes given below each family member BF—, C2—, C4A—, C4B—/DNA pattern. Individual number 4 was repeated to align the two Southern blots.

mode of inheritance of the C2 genomic patterns. Ascertainment of the protein complotypes of each family member (Fig. 4) reveals that in this family the C2 DNA pattern II is associated with the F protein variant of factor B. However, individuals with the F variant do not always show the type II DNA pattern since two other unrelated individuals with this factor B type have been identified and were homozygous for the DNA pattern I.

## Discussion

The second component of complement is synthesized in liver (31) and mononuclear phagocytes (32) at extremely low levels, (0.1% of total protein), hence C2-specific mRNA is of relatively low abundance in these tissues. The availability of a human liver cDNA library (14) containing ~200,000 recombinants allowed isolation of a C2 cDNA with the use of a synthetic oligonucleotide mixture. Synthetic oligonucleotides have been

used for isolation of many low abundant cDNA clones. Only limited published amino acid sequence data for human C2 were available (30). This sequence necessitated synthesis of an extremely complex mixture of oligonucleotides in order to include the authentic coding sequence (Fig. 1). In the regions corresponding to amino acids at positions 3 and 5 (Glu) we constructed a mixture that contained 75% glutamine codons and 25% glutamic acid codons in order to account for possible protein sequencing errors. The resulting oligonucleotide mixture contained 576 different sequences. A difference was found between the amino acid at position 6 of the published sequence and that specified by the cDNA clone (pC2-10). This difference led to a nucleotide sequence error in the synthetic oligonucleotide probe used to isolate the clone. Despite this error and the complexity of the oligonucleotide mixture, a weak hybridization signal was obtained in the initial screening of the cDNA library. The differences between the published and derived amino acid sequence at positions 6 and 10 could be the result of errors in protein sequencing or may represent genetic variants.

Several interesting differences between the derived human amino acid sequence and the published sequence for guinea pig C2a were noted. For instance, a Cys at position 18 in the human sequence is Ala in the guinea pig, which may result in a higher order structural difference between the guinea pig and human C2 proteins. Such a substitution might account for the previously recognized functional differences between guinea pig and human C2 (33).

Overall structural and functional similarities between C2 and factor B led to speculation that these were homologous and therefore may have arisen as a result of gene duplication. Experiments designed to detect cross hybridization even at low stringency failed to demonstrate a high degree of homology between these two genes. Thus, it was possible to separately analyze genomic polymorphisms for each gene. The fact that no genomic variants of factor B were detected plus the observation that mouse and human factor B genes display 83% nucleotide sequence homology (34) suggests that this region of the genome has been relatively highly conserved when compared with other genes within the MHC.

A new marker for C2 polymorphism has been identified at the DNA level by using the Southern blotting technique. The availability of this marker coupled with C4 variants (protein and DNA) and C2 and factor B protein polymorphisms has increased the potential for application of complement genetics to the study of human diseases and for tissue typing.

## Acknowledgments

We thank Dr. Chester Alper for performing the protein complotyping, Jane Ipsen and Elizabeth Davidson for technical assistance, and Ms. Helen Hourihan for her secretarial assistance.

This work was supported by U. S. Public Health Service grants HD 17461, AI 20032, HL 32361, AI 18612, and by a grant from the Charles H. Hood Foundation.

## References

- Alper, C. A. 1981. Complement and the MHC. In *The role of the major histocompatibility complex in immunobiology*. M. E. Dorf, editor. Garland Publishing, Inc., New York. 173-220.
- O'Neill, G. J., S. Y. Yang, and B. Dupont. 1978. Two HLA-linked loci controlling the fourth component of human complement. *Proc. Natl. Acad. Sci. USA*. 75:5165-5169.
- Roos, M. H., E. Mollenhauer, P. Demant, and C. Ritter. 1982. A molecular basis for the two locus model of human complement component C4. *Nature (Lond.)*. 298:854-856.
- Awdeh, Z. L., and C. A. Alper. 1980. Inherited structural polymorphism of the fourth component of human complement. *Proc. Natl. Acad. Sci. USA*. 77:3576-3580.
- Gotze, O., and H. J. Eberhard. 1976. The alternative pathway of complement activation. *Adv. Immunol.* 24:1-35.
- Kerr, M. A., and R. R. Porter. 1978. The purification and properties of the second component of human complement. *Biochem. J.* 171:99-107.
- Raum, D., Z. Awdeh, and C. A. Alper. 1981. BF types and the mode of inheritance of insulin dependent diabetes mellitus (IDDM). *Immunogenetics*. 12:59-74.
- Alper, C. A. 1976. Inherited structural polymorphism in human C2. Evidence for genetic linkage between C2 and BF. *J. Exp. Med.* 144:1111-1115.
- Lee, J. S., J. Trowsdale, and W. F. Bodmer. 1982. cDNA clones coding for the heavy chain of HLA-DR antigen. *Proc. Natl. Acad. Sci. USA*. 79:545-549.
- Malissen, M., B. Malissen, and B. R. Jordan. 1982. Exon/intron organization and complete nucleotide sequence of an HLA gene. *Proc. Natl. Acad. Sci. USA*. 79:893-897.
- Auffray, C., A. J. Korman, M. Roux-Dosseto, R. Bono, and J. L. Strominger. 1982. cDNA clone for the heavy chain of the human B cell alloantigen DC1. Strong sequence homology to the HLA-DR heavy chain. *Proc. Natl. Acad. Sci. USA*. 79:6337-6341.
- Barbosa, J. A., M. E. Kamarck, P. A. Biro, S. M. Weissman, and F. H. Ruddle. 1982. Identification of human genomic clones coding the major histocompatibility antigens HLA-A2 and HLA-B7 by DNA-mediated gene transfer. *Proc. Natl. Acad. Sci. USA*. 79:6327-6331.
- Hood, L., M. Steinmetz, and B. Malissen. 1983. Genes of the major histocompatibility complex of the mouse. *Ann. Rev. Immunol.* 1:529-568.
- Woods, D. E., A. F. Markham, A. T. Ricker, G. Goldberger, and H. R. Colten. 1982. Isolation of cDNA clones for the human complement protein factor B, a class III major histocompatibility complex gene product. *Proc. Natl. Acad. Sci. USA*. 79:5661-5665.
- Campbell, R. D., and R. R. Porter. 1983. Molecular cloning and characterization of the gene coding for human complement protein factor B. *Proc. Natl. Acad. Sci. USA*. 80:4464-4468.
- Whitehead, A. S., G. Goldberger, D. E. Woods, A. F. Markham, and H. R. Colten. 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. Natl. Acad. Sci. USA*. 80:5387-5391.
- Carroll, M. C., and R. R. Porter. 1983. Cloning of a human complement component C4 gene. *Proc. Natl. Acad. Sci. USA*. 80:264-267.
- Ogata, R., D. Shreffler, D. Sepich, and S. Lilly. 1983. cDNA clone spanning the alpha-gamma subunit junction in the precursor of the murine fourth complement component (C4). *Proc. Natl. Acad. Sci. USA*. 80:5061-5065.
- Chaplin, D. D., D. E. Woods, A. S. Whitehead, G. Goldberger, H. R. Colten, and J. G. Seidman. 1983. Molecular map of the murine S region. *Proc. Natl. Acad. Sci. USA*. 80:6947-6951.
- Palsdottir, A., S. J. Cross, J. H. Edwards, and M. C. Carroll. 1983. Correlation between a DNA restriction fragment polymorphism and C4A6 protein. *Nature (Lond.)*. 306:615-616.
- Whitehead, A. S., D. E. Woods, E. Fleischnick, J. E. Chin, E. J. Yunis, A. J. Katz, P. S. Gerald, C. A. Alper, and H. R. Colten. 1984. DNA polymorphism of the C4 genes. A new marker for analysis of the major histocompatibility complex. *N. Engl. J. Med.* 310:88-91.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
- Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA*. 74:560-564.
- Goldberg, D. A. 1980. Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA*. 77:5794-5798.
- Michaelson, A. M., A. F. Markham, and S. H. Orkin. 1983. Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-encoded phosphoglycerate kinase. *Proc. Natl. Acad. Sci. USA*. 80:472-476.
- Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237-251.
- Alper, C. A., T. Boenisch, and L. Watson. 1972. Genetic polymorphism in human glycine-rich beta-glycoprotein. *J. Exp. Med.* 135:68-80.
- Shows, T. B., C. A. Alper, D. Bootsma, et al. 1979. International system for human gene nomenclature. *Cytogenet. Cell Genet.* 25:96-116.
- Carroll, M. C., R. D. Campbell, D. Bentley, and R. R. Porter. 1984. A molecular map of the major histocompatibility class III region of man linking the complement genes C4, C2 and factor B. *Nature (Lond.)*. 307:237-241.
- Kerr, M. A., and J. Gagnon. 1982. The purification and properties of the second component of guinea pig complement. *Biochem. J.* 205:59-67.
- Morris, K. M., D. P. Aden, B. B. Knowles, and H. R. Colten. 1982. Complement biosynthesis by the human hepatoma derived cell line HepG2. *J. Clin. Invest.* 70:906-913.
- Einstein, L. P., E. E. Schneeberger, and H. R. Colten. 1976. Synthesis of the second component of complement by long-term primary cultures of human monocytes. *J. Exp. Med.* 143:114-126.
- Borsos, T., H. J. Rapp, and H. R. Colten. 1970. Immune hemolysis and the functional properties of the second (C2) and fourth (C4) components of complement. I. Functional differences among C4 sites on cell surfaces. *J. Immunol.* 105:1439-1446.
- Sackstein, R., H. R. Colten, and D. E. Woods. 1983. Phylogenetic conservation of a class III major histocompatibility complex antigen, factor B: Isolation of nucleotide sequencing of mouse factor B cDNA clones. *J. Biol. Chem.* 258:14693-14697.