Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1212–1215, February 1984
Immunology

Isolation of cDNA clones for human complement component C2

(HLA class III antigen/serine protease/cDNA cloning/RNA transfer blot analysis/Southern blot analysis)

DAVID R. BENTLEY AND RODNEY R. PORTER

Medical Research Council Immunochemistry Unit, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, United Kingdom

Contributed by R. R. Porter, November 7, 1983

ABSTRACT **Two cDNA clones for complement component C2 have been isolated from a high-complexity human liver cDNA library by using a mixture of 64 synthetic oligonucleotides as a probe. The 400-base-pair insert of pC201 codes for a region containing the active site serine residue and the secondary substrate binding pocket of the serine protease. This part of C2 is 34% homologous to the corresponding region of the related serine protease factor B and additional similarity is evident from a number of conservative amino acid replacements in this region. The insert of pC201 was used as a specific probe in RNA transfer analysis to determine the size of the C2 mRNA as ≈2.9 kilobases. Southern blot analysis of genomic DNA of unrelated individuals identified a single C2 locus and showed no cross-hybridization with the factor B locus.**

Complement component C2 is a class III gene product of the major histocompatibility complex [human leukocyte antigen (HLA) region] in man (1–3). The C2 locus is polymorphic (2, 4, 5) and is closely linked to the loci that code for the other class III proteins, the complement components factor B and C4, all of which have been mapped in between the HLA-D and HLA-B regions (3).

The C2 gene product is a novel type of serine protease that occurs in the serum in zymogen form as a single chain glycoprotein of M_r 102,000 (6, 7). When activated by a single cleavage event, the catalytic subunit (M_r 60,000) in association with activated C4 forms the C3 convertase of the classical pathway of the complement system (8). In structure, function, and mechanism of activation, C2 resembles factor B, which together with C3b forms the C3 convertase of the alternative pathway (8). Most of the amino acid sequence (9, 10) and gene structure (11) of factor B has been determined. Partial amino acid sequence analysis (ref. 12; J. Gagnon, personal communication) has confirmed the extensive structural homology between C2 and factor B.

C2 is synthesized in the liver (13) and also in macrophages (14). It is present in low levels in human serum (≈15 mg/liter) (8, 15) compared to most of the other complement proteins—e.g., factor B (150 mg/liter) (16), C4 (400 mg/liter) (15), and C3 (1200 mg/liter) (15). It has been possible to isolate limited amounts of C2 protein for sequence analysis (17) but the yields obtained are low because the protein is unusually susceptible to proteolysis during isolation (8, 18).

To isolate the gene for C2 and to determine the extent of the homology between C2 and factor B and to map precisely the C2 locus relative to the C4 and factor B loci in the HLA region, cDNA clones corresponding to C2 mRNA were isolated and characterized for use as specific hybridization probes.

Because of the low levels of C2 protein present in the blood, it was estimated that the level of C2 mRNA in the liver might be as low as 0.01%. Therefore, it was necessary to construct a cDNA library of high complexity from human liver RNA. C2 cDNA clones were identified by using a mixture of sixty-four 17-long oligodeoxyribonucleotides as a hybridization probe and were characterized by nucleotide sequence analysis.

MATERIALS AND METHODS

Synthesis of Oligodeoxyribonucleotides. A mixture of sixty-four 17-long oligonucleotides of sequence

$$3'\ \text{C-T-}\tfrac{A}{G}\text{-G-T-}\tfrac{A}{G}\text{-C-T-}\tfrac{T}{C}\text{-T-T-}\tfrac{A}{G}\text{-C-T-}\tfrac{T}{C}\text{-}\tfrac{A}{G}\text{-A}\ 5'$$

was synthesized against the known C2 protein sequence Asp-His-Glu-Asn-Glu-Leu by using the solid-phase phosphotriester technique (19, 20). The mixture was radioactively labeled by using T4 polynucleotide kinase and [γ-^{32}P]rATP to a specific activity of 5×10^8 cpm/μg of DNA for use as a hybridization probe.

Isolation and Purification of RNA. Total RNA was extracted from ≈10 g of human liver by using the guanidine thiocyanate procedure of Chirgwin *et al.* (21). The RNA was fractionated on a 15–30% sucrose gradient, and all fractions containing RNA of >500 nucleotides in length were pooled. The RNA was further purified by oligo(dT)-cellulose chromatography (22).

Construction and Screening of cDNA Library. First- and second-strand DNA syntheses were based on standard procedures (23, 24) and used 40 μg of purified RNA. The double-stranded DNA was desalted on a 1.5-ml Sephacryl S-400 column, partially digested with *Hae* III, and then ligated into the *Pvu* II site of the plasmid vector pAT153/Pvu II/8 (25). The products of the ligation were transformed into competent *Escherichia coli* K-12 strain MC1061 (26). The library was amplified by growth for 2 hr in 2× concentrated TY broth containing 100 μg of ampicillin per ml before storage at −70°C.

Approximately 300,000 colonies of the fragment library were screened on Whatman 541 filters following the method of Gergen *et al.* (27). The hybridization temperature was 37°C, and the filters were washed at 46°C.

RNA Transfer Blot Analysis. RNA was electrophoresed in 1.5% agarose/37% formaldehyde/20 mM 3-(*N*-morpholino) propanesulfonic acid (pH 7.0) as described (28). The gel was prewashed and blotted onto Gene Screen overnight as described by the manufacturer (New England Nuclear). The 400-base-pair (bp) insert of pC201 (see *Results and Discussion*) was purified by elution from a 4% polyacrylamide thin gel (29), nick-translated (30) to a specific activity of ≈1×10^8 cpm/μg of DNA, and hybridized in 50% formamide buffer supplemented with dextran sulfate (10%).

Southern Blot Analysis. Restriction digests of genomic DNA were electrophoresed in 0.7% agarose, and the DNA was blotted onto nitrocellulose (31, 32). The filter was

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HLA, human leukocyte antigen; C2, C3, and C4, second, third, and fourth components of complement; bp, base pair(s); kb, kilobase pair(s).

prewashed and hybridized with the nick-translated insert of pC201 under the same conditions used for the RNA transfer blot analysis.

RESULTS AND DISCUSSION

Isolation and Characterization of C2 cDNA Clones. The fragment library was constructed from human liver RNA and contained $>4 \times 10^6$ transformants, of which >95% were recombinants. The average size of the inserts was 300 bp. By using the oligonucleotide mixture as a probe, 300,000 colonies of the library were screened. Two of the positively hybridizing clones, pC201 and pC204, contained inserts of ≈400 bp. Nucleotide sequence analysis by the method of Maxam and Gilbert (29) showed that both contained a C2 coding sequence that agreed with the known C2 amino acid sequence (12) in the region of the active site serine residue. Both clones also contained a common sequence that matched one of the oligonucleotides in the probe mixture.

The nucleotide sequence of the pC201 insert is shown in Fig. 1. The amino acid sequence of this part of C2, inferred from the nucleotide data, agrees with the results of amino acid sequence analysis (regions previously subjected to sequence analysis at the protein level are underlined in Fig. 1), except that the DNA sequence of both clones indicates the presence of a valine residue at position 96, instead of a leucine residue. The sequence around the active site serine residue of C2 as determined above was compared with the corresponding region of factor B (see Fig. 2). The overall homology between the two proteins in this region is 34%. A further 12.5% of the residues are conservative replacements between them. The position and number of all five cysteine residues in this region is also conserved. The extensive homology found between C2 and factor B in this part of the

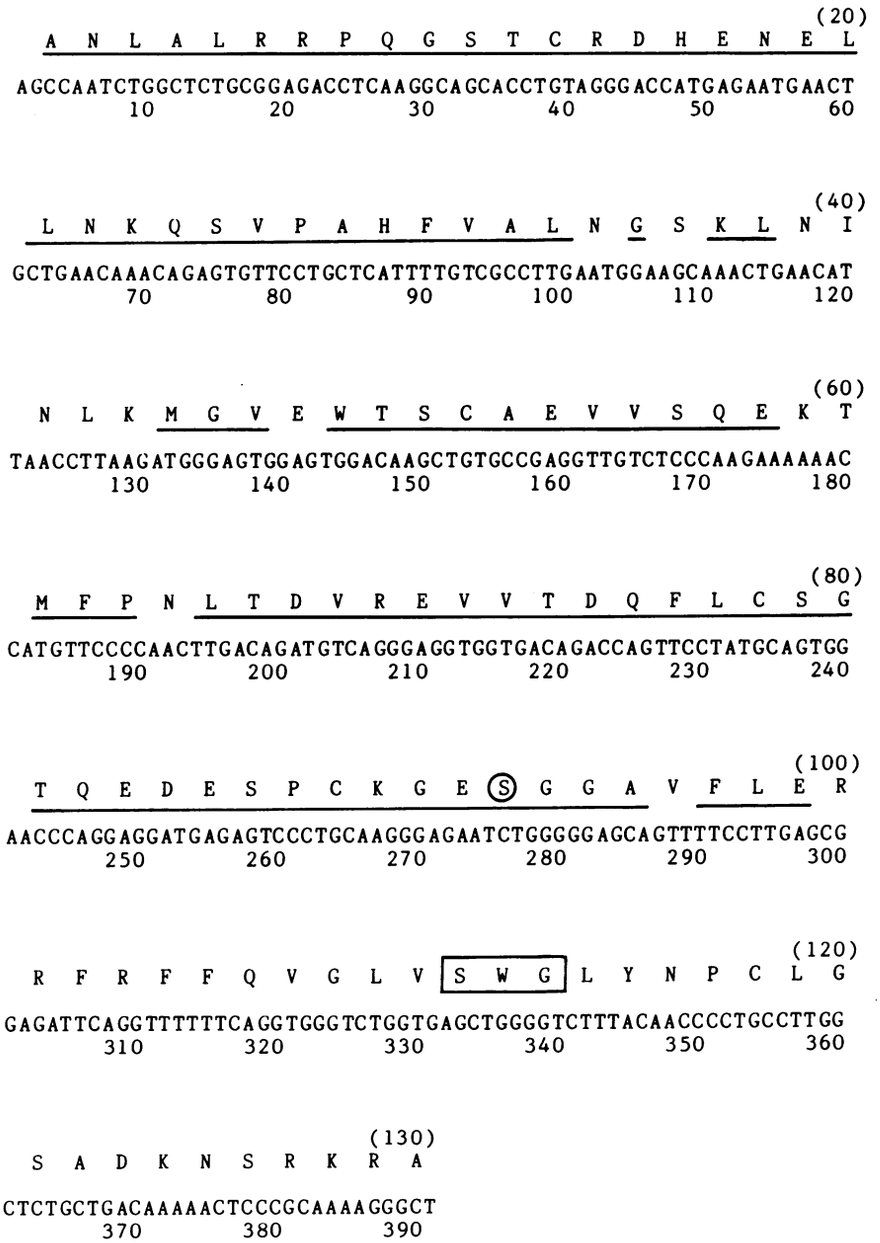

FIG. 1. Nucleotide sequence of the insert of pC201. The translated sequence is shown in the one-letter code above the DNA sequence. Numbers below the sequence are of the nucleotides; numbers above in brackets are of the amino acids. The active site serine residue is circled and the residues that form the secondary substrate binding pocket are boxed. Amino acid residues underlined agree with the previously determined protein sequence (12).

```
                    10                  20                  30                  40                  50
C2:  A N L A L R R P Q G S T C R D H E N E L L N K Q S V P A H F V A L N G S K L - - - N I N L K M G V E W T S C A E V V S Q
FB:  T T R A L R L P P T T T C Q Q Q K E E L L P A Q D I K A L F V S E E E K K L T R K E V Y I K N G D K K G S C - E R D A Q

                    60                  70                  80                  90                  100                 110
C2:  E K T M F P N L T D V R E V V T D Q F L C S G - - - T Q E D E S P C K G E S G G A V F L E R R F R F F Q V G L V S W G L
FB:  Y A P G Y D K V K D I S E V V T P R F L C T G G V S P Y A D P N T C R G D S G G P L I V H K R S R F I Q V G V I S W G V

                    120                 130
C2:  Y N P C L G S A D K N S R K R A
FB:  V D V C K N Q K R Q K Q V P A H
```

FIG. 2. Comparison of the amino acid sequence encoded by the insert of pC201 with the corresponding region of factor B. The region of factor B is from residue 344 to residue 478 of the B_b subunit (9). The alignment has been adjusted by the insertion of several gaps to maximize the homology. Identical amino acids in the two sequences are shown boxed; conservative replacements are underlined. Cysteine residues occur at positions 13, 51, 78, 88, and 118. The active site serine residue is at position 92.

active site (residue 92 in Fig. 2) and the secondary substrate binding pocket (residues 111–113 in Fig. 2) correlates with the functional similarity of the two serine proteases in C3 activation (8, 33).

RNA Transfer Blot Analysis. The insert of pC201 was used as a hybridization probe in an RNA transfer blot analysis of total, poly(A)$^+$, or poly(A)$^-$ human liver RNA (Fig. 3). A single diffuse band with an approximate size of 2.9 kilobase pairs (kb) was observed. Therefore, the C2 mRNA is slightly larger than the factor B mRNA (2.6 kb) as determined by RNA transfer blot analysis (34). This is expected as the C2

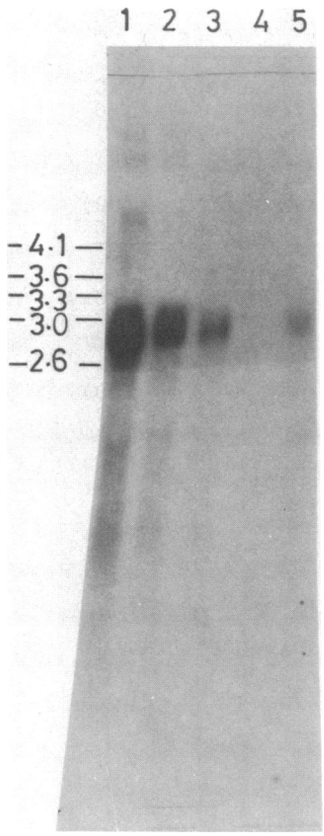

FIG. 3. RNA transfer blot analysis of human liver RNA by using the insert of pC201 as a C2-specific hybridization probe. Lanes 1–3, 20, 10, and 5 μg of poly(A)$^+$ RNA; lane 4, 10 μg of poly(A)$^-$ RNA (i.e., RNA that did not bind to the column during oligo(dT)-cellulose chromatography); lane 5, 10 μg of total RNA. Positions of denatured DNA frements of known sizes are shown; sizes are in kb.

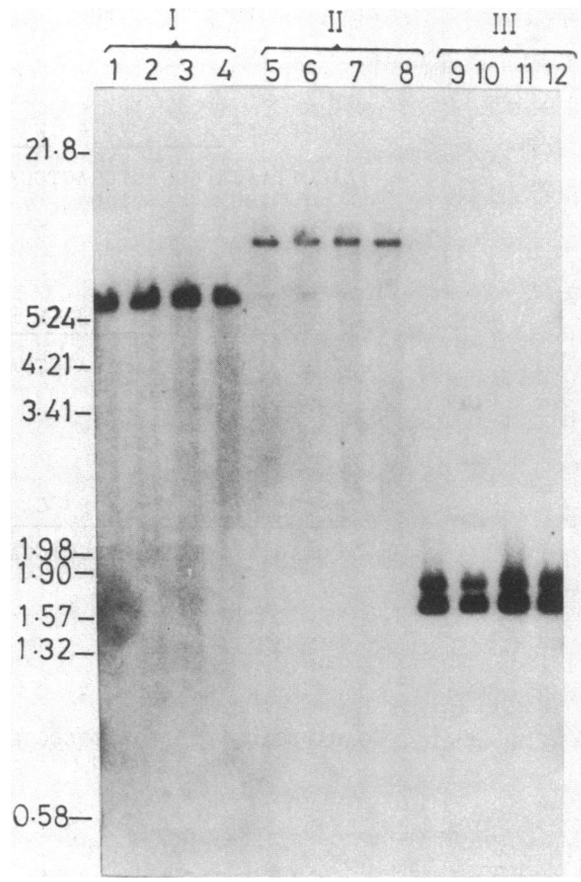

FIG. 4. Southern blot analysis of the genomic DNA from four individuals by using the insert of pC201 as a probe. DNA was digested with *Bgl* II (group I), *Sst* I (group II), or with *Bam*HI (group III). Known sizes in kb of DNA marker fragments electrophoresed on the same gel are as shown. Faint bands in lane 10 are the result of partial digestion of the genomic DNA. The lower intensity of the bands in group II are probably the result of less efficient transfer of large fragments of DNA during the blotting procedure.

mRNA codes for a protein product of M_r = 102,000, which is larger than the factor B protein product of M_r = 90,000.

Southern Blot Analysis. The pC201 probe was used in a Southern blot analysis of the genomic DNA from four unrelated individuals (Fig. 4). Each DNA sample was digested with *Bgl* II, *Sst* I, or *Bam*HI. The results showed that the pC201 probe hybridizes to a single 5.7-kb *Bgl* II fragment (lanes 1–4), a single 7.5-kb *Sst* I fragment (lanes 5–8), and two fragments of 1.6 and 1.7 kb in the *Bam*HI digest (lanes 9–12). It has previously been shown that the region of the factor B gene encoding the active site serine residue is contained within a 14-kb genomic *Bam*HI fragment (35). The pC201 probe does not hybridize to this fragment (see Fig. 4) despite the apparent homology between the sequences of the two proteins in this region. These results indicate that the pC201 probe is specific for C2.

Further characterization of the C2 locus has been done by using the pC201 probe in Southern analysis of cosmid clones, and the precise location of the C2 locus relative to the C4 and factor B genes in the class III HLA region has been determined (35). Further investigation is necessary to extend the genetic analysis to the study of individuals who are deficient in the expression of C2 and also to determine the extent and nature of C2 polymorphism by analyzing DNA from individuals who have been C2-typed.

Special thanks are due to Dr. A. K. Connolly for collaboration in the construction and screening of the cDNA library and for helpful discussion. We also thank Professor G. G. Brownlee and K. Gould for providing facilities for oligonucleotide synthesis and for valuable advice and Drs. K. H. Choo, R. D. Campbell, M. C. Carroll, K. B. M. Reid, and I. M. Jones for discussion. D.R.B. is a Beit Memorial Junior Research Fellow.

1. Mortensen, J. P., Buskjaer, L. & Lamm, L. U. (1980) *Immunology* **39,** 541–549.
2. Weitkamp, L. R. & Lamm, L. U. (1982) *Cytogenet. Cell Genet.* **32,** 130–143.
3. Olaisen, B., Teisberg, P., Jonassen, K., Gedde-Dahl, T., Moen, T. & Thorsby, E. (1981) *Hum. Immunol.* **2,** 247–254.
4. Barnstaple, C. J., Jones, E. A. & Bodmer, W. F. (1979) in *Defense and Recognition*, ed. Lennox, E. S. (University Park Press, Baltimore), Vol. 2A, pp. 151–227.
5. Meo, T., Atkinson, J., Bernoco, M., Bernoco, D. & Cepellini, R. (1976) *Eur. J. Immunol.* **6,** 916–919.
6. Polley, M. J. & Muller-Eberhard, H. J. (1968) *J. Exp. Med.* **128,** 533–551.
7. Kerr, M. A. & Porter, R. R. (1978) *Biochem. J.* **171,** 99–107.
8. Reid, K. B. M. & Porter, R. R. (1981) *Annu. Rev. Biochem.* **50,** 433–464.
9. Christie, D. L. & Gagnon, J. (1983) *Biochem. J.* **209,** 61–70.
10. Gagnon, J. & Christie, D. L. (1983) *Biochem. J.* **209,** 51–60.
11. Campbell, R. D. & Porter, R. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4464–4468.
12. Porter, R. R. (1984) *Crit. Rev. Biochem.*, in press.
13. Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colten, M. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5661–5665.
14. Colten, M. R., Ooi, Y. M. & Edelson, P. J. (1979) *Ann. N.Y. Acad. Sci.* **332,** 482–490.
15. Porter, R. R. & Reid, K. B. M. (1979) *Adv. Protein Chem.* **33,** 1–71.
16. Curman, B., Sanberg-Trägardh, L. & Pederson, P. A. (1977) *Biochemistry* **16,** 5368–5375.
17. Parkes, C., Gagnon, J. & Kerr, M. A. (1983) *Biochem. J.* **213,** 201–209.
18. Kerr, M. A. (1979) *Biochem. J.* **183,** 615–622.
19. Gait, M. J., Popov, S. G., Singh, M. & Titmas, R. C. (1980) *Nucleic Acids Res.* **7,** 243–258.
20. Duckworth, M. L., Gait, M. J., Goelet, P., Morg, G. F., Singh, M. & Titmas, R. C. (1981) *Nucleic Acids Res.* **9,** 1691–1706.
21. Chirgwin, J. M., Przbyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
22. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
23. Buell, G. N., Wickens, M. P., Payvar, F. & Schimke, R. T. (1978) *J. Biol. Chem.* **253,** 2471–2482.
24. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253,** 2483–2495.
25. Choo, K. H., Gould, K. G., Rees, D. J. G. & Brownlee, G. G. (1982) *Nature (London)* **299,** 178–180.
26. Casadaban, M. J. & Cohen, S. N. (1980) *J. Mol. Biol.* **138,** 179–207.
27. Gergen, J. P., Stern, R. H. & Websink, P. C. (1979) *Nucleic Acids Res.* **7,** 2115–2136.
28. Lehrach, H., Diamond, D., Wezney, J. M. & Boldtker, H. (1977) *Biochemistry* **16,** 4743–4751.
29. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
30. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
31. Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3683–3687.
32. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
33. Cooper, N. R. (1975) *Biochemistry* **14,** 4245–4251.
34. Morley B. M. & Campbell R. D. (1984) *EMBO J.*, in press.
35. Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. (1984) *Nature (London)*, in press.