## Use of a cDNA clone for the fourth component of human complement (C4) for analysis of a genetic deficiency of C4 in guinea pig

(major histocompatibility complex class III antigen/oligonucleotide/cDNA/complement deficiency)

# ALEXANDER S. WHITEHEAD<sup>\*</sup>, GABRIEL GOLDBERGER<sup>\*</sup>, DEREK E. WOODS<sup>\*</sup>, ALEXANDER F. MARKHAM<sup>†</sup>, AND HARVEY R. COLTEN<sup>\*</sup>

\*Divisions of Cell Biology, Department of Medicine, Children's Hospital, Ina Sue Perlmutter Cystic Fibrosis Research Center, and the Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115; and <sup>†</sup>Imperial Chemical Industries, Pharmaceuticals Division, Mereside, Alderley Park, Macclesfield, Cheshire, England

Communicated by Herman N. Eisen, May 16, 1983

ABSTRACT A cDNA clone for the fourth complement component (C4), pC4AL1, has been isolated from a human adult liver cDNA library by using a synthetic oligonucleotide mixture containing all 384 possible sequences coding for residues 14-21 of the C4  $\gamma$ -chain amino acid sequence. This clone spans the entire C4 y-chain coding sequence and includes a short 3' untranslated region, a poly(A) recognition site, and 16 nucleotides of the poly(A) tail. The 5' end of the clone begins 18 nucleotides upstream from the amino terminus of the C4  $\gamma$  chain and codes for Arg-Asn-Arg-Arg-Arg-Arg, a highly charged proteolytic cleavage site involved in the processing of pro-C4 to native C4. Liver mRNA preparations from C4-deficient guinea pigs were incapable of directing synthesis of pro-C4 or C4 peptides in cell-free translation experiments. Southern blot analysis using pC4AL1 as a hybridization probe of C4-deficient guinea pig DNA established that the deficiency is not the result of deletion of the entire C4 gene. RNA blot analysis using pC4AL1 as a hybridization probe of normal guinea pig liver mRNA revealed a C4 mRNA of 5.0 kilobases (kb). No such mRNA species was observed in C4-deficient guinea pig liver mRNA; however, a 7.0-kb RNA was detected, indicating the presence of a C4 precursor RNA. These results suggest that the basis of C4 deficiency in the guinea pig is a post-transcriptional defect in the processing of C4 precursor RNA to mature C4 mRNA.

The complement system consists of more than 20 plasma proteins that interact in a highly specific manner to mediate several functions associated with host defense against infection (1). The structural genes for three of the complement proteins (the second component, C2; the fourth component, C4; and factor B) have been localized to the major histocompatibility complex (MHC) in man, mouse, and guinea pig (2) and have been designated class III MHC genes.

C4 is a serum glycoprotein of approximately 200 kilodaltons (kDa) that is comprised of disulfide-linked polypeptide chains of 93 kDa ( $\alpha$  chain), 73 kDa ( $\beta$  chain), and 33 kDa ( $\gamma$  chain) (3). C4 in man, mouse, and guinea pig is synthesized as a single-chain precursor molecule, pro-C4 (4–6), in the order NH<sub>2</sub>- $\beta$ - $\alpha$ - $\gamma$ -COOH (7–9). Pro-C4 undergoes intracellular processing prior to secretion as native C4. The major sites of C4 biosynthesis are the liver and extrahepatic macrophages (reviewed in ref. 10).

It has been shown that two C4 loci (C4A and C4B, corresponding to the erythrocyte antigens Rodgers and Chido, respectively) are present in man (11). Allelic variants, including null alleles, have been described for each locus (12), and several C4-deficient patients with no detectable C4A and C4B gene products have been reported (13). In mouse, variants of C4 have been identified and linked to the S region of the mouse MHC (14), and the large differences observed in C4 concentration in the sera of different mouse strains are genetically determined by this region (15). In guinea pig, a genetically determined deficiency of C4 (16) has been extensively studied (for review, see ref. 17). C4-deficient guinea pigs are normal in all respects except for failure to synthesize and secrete C4. No pro-C4 is detected intracellularly nor can C4 be detected in culture media by either immunochemical or functional assays (18). However, polysome-bound immunoprecipitable C4 peptides were apparently generated in an endogenous cell-free translating system derived from C4-deficient guinea pig liver. In these experiments, no intact pro-C4 was detected (19).

The availability of cDNA probes for C4 and the other class III antigens will provide the means to analyze the structure and expression of these MHC-linked genes. Recently, cDNA clones corresponding to the Bb fragment of human factor B (20) and the C4d portion of the human C4  $\alpha$  chain (21) have been isolated by using specific synthetic oligonucleotide sequences as hybridization probes. In this report, we describe the isolation from an adult human liver library of a cDNA clone encompassing the human C4  $\gamma$  chain and the use of this clone in the analysis of the basis of C4 deficiency in the guinea pig.

#### MATERIALS AND METHODS

Human Adult Liver cDNA Library. The preparation of the human liver cDNA library has been described (20). Briefly, the library consists of 230,000 recombinant clones derived from the insertion of double-stranded G·C-tailed cDNA into the plasmid pKT218 and subsequent transformation of the *Escherichia coli* K-12 host strain MC1061. The mRNA from which the doublestranded cDNA had been synthesized was from the liver of a white male cadaver donor (HLA A1,2; Cw3,w6; B15,w39; complotype Bf FS; C2C, C4A4,4; C4B2, QO). In vitro translation and immunoprecipitation had previously established the presence of mRNA species capable of directing synthesis of pro-C4 in this preparation (20).

Preparation of the Synthetic Oligonucleotide Mixture. A mixture of 23-nucleotide-long oligonucleotides that would code for the C4  $\gamma$ -chain amino acid sequence between residues 14 and 21 (1) was synthesized by a solid-phase phosphotriester

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: C2, C3, and C4, second, third, and fourth components of complement; MHC, major histocompatibility complex; kDa, kilodalton(s); kb, kilobase(s).

method using a library of dimer anions (20).

Identification of C4 cDNA Clones. Clones were screened on 82-mm nitrocellulose filters with the synthetic oligonucleotide mixture by using a modification of the Grunstein-Hogness procedure (22). The oligonucleotide mixture was 5' labeled by using  $[\gamma^{-32}P]$ ATP and T4 polynucleotide kinase (Bethesda Research Laboratories) (23) and hybridized at 37°C for 16 hr to the filter-bound DNA at a concentration of 1  $\mu g/20$  ml in 0.9 M NaCl/90 mM Na citrate/1× Denhardt's solution (24)/0.05% Na pyrophosphate/tRNA (100  $\mu g/m$ l). Filters were washed extensively at 20°C with 0.9 M NaCl/90 mM Na citrate/0.05% Na pyrophosphate and at 50°C for 10 min in the same buffer. Hybridization signals were visualized by autoradiography.

Isolation and Analysis of C4 cDNA. Plasmid DNA was isolated from bacteria by the cleared-lysate method (25). Insert cDNA was excised from the vector by digestion with *Pst* I and purified by agarose or polyacrylamide gel electrophoresis. DNA sequence analysis was carried out as described by Maxam and Gilbert (23).

Southern and RNA Blot Analysis. The sizes of restriction endonuclease-digested DNA fragments and specific RNA species were estimated by Southern (26) and RNA (27, 28) blot analysis. Specific sequences were identified by hybridization under standard conditions with the appropriate radiolabeled cDNA probes, prepared by nick-translation (29), followed by washing under the conditions described for each experiment and visualization by autoradiography.

**RNA Isolation, Cell-Free Translation, and Immunoprecipitation.** Human and guinea pig liver RNAs were isolated within 10 min of removal of the organs. The tissues were finely minced and homogenized in 0.25 M sucrose/0.14 M NaCl/0.01 M Tris·HCl, pH 7.5/1.5 mM MgCl<sub>2</sub> at 4°C in a glass Dounce homogenizer. The homogenate was centrifuged at 16,000 × g for 5 min at 4°C and the postmitochondrial supernatant was made 1% in NaDodSO<sub>4</sub>. Total cytoplasmic RNA was isolated by phenol/ chloroform extraction and ethanol precipitation as described by Perry *et al.* (30), and poly(A)<sup>+</sup> mRNA was purified by affinity chromatography on oligo(dT)-cellulose (31). The mRNA was translated in a reticulocyte lysate cell-free system (32) using [<sup>35</sup>S]-methionine as radioactive precursor, and the specific translation products were immunoprecipitated as described (33), subjected to NaDodSO<sub>4</sub>/polyacrylamide gel electrophoresis (34), and visualized by autoradiography.

### RESULTS

Identification of C4-Specific cDNA Clones. A synthetic oligonucleotide mixture comprising all possible 23-nucleotide-long DNA sequences that would code for the amino acid sequence of the C4  $\gamma$  chain between residues 14 and 21 (1) was generated (Fig. 1). This region contains amino acid residues for which there is a high degree of codon ambiguity; therefore, it was necessary to synthesize 384 different oligonucleotides to ensure the presence of the correct sequence.

To identify liver cDNA clones containing C4 sequences, the synthetic oligonucleotide mixture was radiolabeled and used to screen approximately 50,000 clones. One hundred sixty-five clones hybridized specifically to the oligonucleotide mixture. After colony purification, plasmid DNA was isolated from 16 of the positive clones (25). The insert from the largest of the clones, pC4AL1 (approximately 950 base pairs), was shown by Southern blot hybridization experiments to cross-hybridize with 11 of the other clones and was selected for further study.

Partial Nucleotide Sequence Analysis of pC4AL1. The 5' end of the clone begins 18 nucleotides upstream from the triplet coding for the amino-terminal glutamate of the C4  $\gamma$  chain (Fig. 1). This 18-nucleotide segment codes for Arg-Asn-Arg-Arg-Arg-Arg, a highly positively charged amino acid sequence. The amino acid sequence derived from the next 63 nucleotides corresponds to that of the published amino-terminal segment of the C4  $\gamma$  chain (1) except for the presence of arginine instead of glycine at position 21. The 3' end of pC4AL1 contains a poly(A) recognition site, A-T-T-A-AA, and 16 nucleotides corresponding to the poly(A) tail of the mRNA (data not shown), indicating





that pC4AL1 spans the 3' untranslated region, the entire C4  $\gamma$ -chain coding sequence, and 18 nucleotides corresponding to sequence in the primary translation product amino terminal to the  $\gamma$  chain.

Analysis of Cell-Free Translation Products from Normal and C4-Deficient Guinea Pig Liver mRNA. Immunoprecipitation of primary translation products from normal guinea pig liver mRNA (Fig. 2, lane B) reveals the presence of pro-C4 whereas no pro-C4 was detected after translation of C4-deficient guinea pig mRNA (lane C). Both preparations direct synthesis of equivalent levels of total protein, factor B, and pro-C3 (data not shown). Synthesis of pro-C3 in both C4-deficient and normal mRNA indicates that both preparations contain functional mRNA of a similar size to that directing synthesis of pro-C4. The low molecular weight radiolabeled bands detected in both C4-deficient and normal guinea pig liver mRNA translations were also detected in control immunoprecipitations with antiserum to ovalbumin (data not shown). Similar results were obtained using mRNA isolated from three additional normal and three additional C4-deficient guinea pig livers (data not shown).

**RNA Blot Analysis of Human and Guinea Pig RNA with pC4AL1.** Specific hybridization of pC4AL1 to a C4 mRNA species with an apparent size of 5.0 kilobases (kb) was detected in samples isolated from human liver (Fig. 3, lane B). A similarsized RNA species was detected in  $poly(A)^+$  mRNA from normal guinea pig liver (lane E). When  $poly(A)^+$  mRNA from C4deficient guinea pig liver was subjected to similar analysis (lanes







FIG. 3. RNA blot analysis of human and guinea pig C4 mRNA. RNA samples were size fractionated by agarose gel electrophoresis and transferred to nitrocellulose filters (27, 28). Samples were hybridized by standard procedures (35) with the pC4AL1 probe that had been radiolabeled by nick-translation (29). After hybridization, filters were washed with 150 mM NaCl/15 mM Na citrate/0.1% NaDodSO<sub>4</sub> at 50°C for 45 min; autoradiography was for 16 hr. Lanes: A, control, U937 total RNA (20  $\mu$ g) (U937 is a human monocyte cell line that does not synthesize C4); B, human liver total RNA (50  $\mu$ g); C and D, C4-deficient guinea pig poly(A)<sup>+</sup> mRNA (20  $\mu$ g); E, normal guinea pig poly(A)<sup>+</sup> mRNA (20  $\mu$ g); F and G, UV photographs of RNA in lanes D and E prior to transfer to nitrocellulose. Size estimations were made by corunning radiolabeled *Hind*III-digested  $\lambda$  phage DNA. Exposure time for autoradiography of lanes A–E was 18 hr except for lane C, which was a 90hr exposure of lane D.

C and D), pC4AL1 hybridized to a higher molecular weight RNA species (of apparent size 7.0 kb) but did not detect the 5kb species identified in the normal guinea pig liver  $poly(A)^+$ mRNA preparation. The higher molecular weight species seen in the C4-deficient RNA was detected as a faint band after 18 hr exposure (lane D) but was not detected even after the longer exposure time (90 hr) in normal guinea pig RNA (data not shown). Lanes F and G show the RNA from lanes D and E under UV illumination prior to RNA blotting and indicate that similar amounts of poly(A)<sup>+</sup> mRNA were present in each. Furthermore, subsequent hybridization of the same RNA blot with a factor B-specific cDNA probe (20) showed that the C4-deficient and normal guinea pig mRNA preparations contained similar amounts of factor B mRNA (data not shown). Identical results were obtained with  $poly(A)^+$  mRNA isolated from another set of C4-deficient and normal guinea pig livers.

Southern Blot Analysis of Human and Guinea Pig DNA with pC4AL1. Specific hybridization between pC4AL1 and an approximately 13-kb DNA fragment was detected in human DNA after digestion with *Bam*HI (Fig. 4, lane A). Similar analysis of *Bam*HI-digested normal and deficient guinea pig DNA revealed hybridization of pC4AL1 to fragments of approximately 18.5 and 5.6 kb in the former but hybridization only to the larger band in the C4 deficient (lanes B and C).



FIG. 4. Southern blot analysis of human and guinea pig DNA. DNA samples were prepared from human peripheral blood lymphocytes and guinea pig liver (36) and digested with BamHI, size fractionated by agarose gel electrophoresis, and transferred to nitrocellulose filters (26). Samples were hybridized by standard procedures (37) with the pC4AL1 probe that had been radiolabeled by nick translation (29). After hybridization, filters were washed in 150 mM NaCl/15 mM Na citrate/ 0.1% NaDodSO4 at 50°C for 45 min prior to autoradiography. Lanes: A, human DNA (7 µg); B, normal guinea pig DNA (10 µg); Č, Č4-deficient guinea pig DNA (10  $\mu$ g). The autoradiograph was exposed for 5 days. Size estimations were made by corunning radiolabeled HindIII-digested  $\lambda$  phage DNA.

#### DISCUSSION

A cDNA clone, pC4AL1, has been isolated from a human adult liver cDNA library by using a synthetic oligonucleotide mixture complementary to the mRNA sequence that would code for residues 14–21 of the published C4  $\gamma$ -chain amino acid sequence. The mixture contained 384 different 23-nucleotide-long oligonucleotides to ensure complementarity. Definitive identification of pC4AL1 was based on the presence of a nucleotide sequence corresponding to the first 21 amino acids of the C4  $\gamma$  chain. Of a total of 16 colony-purified clones, 12, including pC4AL1, contained cDNA inserts that cross-hybridized with one another under high stringency conditions, indicating that most of the clones contained C4 cDNA sequences in spite of the complexity of the oligonucleotide mixture used for the initial screening. It should be noted that the nucleotide sequence of pC4AL1 indicates that, for the C4 species specified by this clone, there is an arginine rather than a glycine at residue 21. The consequent mismatch of the second nucleotide from the 5' end of the appropriate oligonucleotide (Fig. 1) did not, however, prevent identification of cDNA clones bearing sequences complementary to this region of C4.

The amino acid sequence derived from the pC4AL1 nucleotide sequence specifies glutamine, valine, and isoleucine at residues 7, 13, and 19 rather than histidine, alanine, and cysteine (given as alternatives in the published sequence) at residues 7, 13, and 19. These differences may reflect the allelic variant of C4 represented by pC4AL1 or technical considerations in amino acid sequence analysis.

The 18 nucleotides at the 5' end of pC4AL1 code for the six amino acid residues adjacent to the amino terminus of the C4  $\gamma$  chain. These residues (Arg-Asn-Arg-Arg-Arg-Arg) define a highly charged region of the pro-C4 molecule that may represent an exposed proteolytic cleavage site. Plasmin cleavage of pro-C4 has been described (8) and this region may be the carboxyl terminus of a short section between the  $\alpha$  and  $\gamma$  chains that is excised during processing to native C4. A similar argi-

nine-rich cleavage site has been identified from nucleotide sequence analysis of the junction of the  $\beta$  and  $\alpha$  chains of the precursor of the third component of mouse complement (38), the conversion of which also can be mediated by plasmin (39). The availability of larger C4 cDNA clones and identification of the carboxyl terminus of the mature C4  $\alpha$  chain will permit a more complete description of this region.

Analysis of the in vitro translation products of C4-deficient guinea pig liver  $poly(A)^+$  mRNA indicated the total absence of both pro-C4 and C4 peptides. RNA blot analysis using pC4AL1 as the probe established that mature C4 mRNA was not present in the  $poly(A)^+$  RNA of C4-deficient guinea pigs; however, the presence of a higher molecular weight species was detected (Fig. 3). The existence of this putative C4 precursor RNA suggests that the basis of C4 deficiency in the guinea pig is a defect of post-transcriptional processing. Taken together, these data suggest that the previous finding of polysome-bound C4 peptides in the endogenous translation of C4-deficient mRNA (19) was artifactual. However, as pC4AL1 is a reagent that has only C4 y-chain specificity, the presence of truncated mRNA terminating before the  $\gamma$ -chain sequence or of highly unstable mature mRNA cannot be excluded. Southern blot analysis of normal and C4-deficient guinea pig DNA shows that both carry similar BamHI-digested fragments bearing sequences complementary to the C4  $\gamma$  chain and that the deficient state is not merely the result of a deletion in this portion of the C4 gene. The presence of a smaller BamHI fragment in the normal guinea pig only may represent an allelic nucleotide sequence variant of the C4 gene present in the normal but not in the homozygous C4-deficient animal. A partial deletion on one side of a BamHI restriction site within the C4  $\gamma$ -chain DNA sequence of the deficient gene is not excluded. However, absence of the entire C4  $\gamma$ -chain coding sequence is excluded, based on the high molecular weight transcription product revealed in RNA blots. Finally, pC4AL1 and other C4 cDNA clones will be useful for the analysis of C4 deficiency and C4 polymorphism in man.

We thank Dr. Uma S. Tantravahi for her advice on the RNA blotting technique, Miriam Rits for her technical assistance, and Ms. Helen Hourihan for her secretarial assistance. This work was supported by U.S. Public Health Service Grants AI15033, HD17461, AI18612, and HL22487 and by a grant from the Charles H. Hood Foundation. A.S.W. is supported by a Helen Hay Whitney Foundation Fellowship (F488).

- Porter, R. R. & Reid, K. B. M. (1979) Adv. Protein Chem. 33, 1-1. 71.
- Alper, C. A. (1981) in The Role of the Major Histocompatibility 2. Complex in Immunobiology, ed. Dorf, M. E. (Garland, New York),
- pp. 173-220. Schreiber, R. D. & Muller-Eberhard, H. J. (1974) J. Exp. Med. 140, 3. 1324 - 1335
- Gigli, I. (1978) Nature (London) 272, 836-837. 4.
- Parker, K. L., Roos, M. H. & Shreffler, D. C. (1979) Proc. Natl. Acad. Sci. USA 76, 5853-5857. 5.
- Hall, R. E. & Colten, H. R. (1977) Proc. Natl. Acad. Sci. USA 74, 6. 1707-1710.
- Goldberger, G., Abraham, G. N., Williams, J. & Colten, H. R. 7. (1980) J. Biol. Chem. 255, 7071-7074.
- Goldberger, G. & Colten, H. R. (1980) Nature (London) 286, 514-8.
- Parker, K. L., Capra, J. D. & Shreffler, D. C. (1980) Proc. Natl. Acad. Sci. USA 77, 4275-4278. 9.
- Fey, G. & Colten, H. R. (1981) Fed. Proc. Fed. Am. Soc. Exp. Biol. 10. **40, 2099–2104**
- O'Neill, G. J., Yang, S. Y. & Dupont, B. (1978) Proc. Natl. Acad. 11. Sci. USA 75, 5165-5169.
- 12.
- Awdeh, H. S. & Alper, C. A. (1980) *Immunobiology* **158**, 35–40. Rosen, F. S. (1981) in *Hematology of Infancy and Childhood*, eds. Nathan, D. G. & Oski, F. A. (Saunders, Philadelphia), Vol. 2, pp. 13. 866-886

- Ferreira, A., Takahashi, M. & Nussenzweig, V. (1977) J. Exp. Med. 14. 146, 1001-1018.
- 15. Shreffler, D. C. & Owen, R. D. (1963) Genetics 48, 9-25.
- Ellman, L., Green, I. & Frank, M. M. (1970) Science 170, 74-75. 16.
- 17. Colten, H. R. (1982) Mol. Immunol. 19, 1279-1285.
- 18. Colten, H. R. & Frank, M. M. (1972) Immunology 22, 991-999.
- Hall, R. E. & Colten, H. R. (1978) Science 199, 69-70. 19.
- Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colten, H. R. (1982) Proc. Natl. Acad. Sci. USA 79, 5661–5665. 20.
- 21. Carroll, M. C. & Porter, R. R. (1983) Proc. Natl. Acad. Sci. USA 80, 264-267
- 22. Grunstein, M. & Hogness, D. S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
- Maxam, A. M. & Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 23. 560-564.
- 24. Denhardt, D. R. (1966) Biochem. Biophys. Res. Commun. 23, 641-646.
- 25. Clewell, D. B. & Helinski, D. (1969) Proc. Natl. Acad. Sci. USA **62**, 1159–1166.
- 26. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
- 27. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) Biochemistry 16, 4743-4751.

- Proc. Natl. Acad. Sci. USA 80 (1983) 5391
- Goldberg, D. A. (1980) Proc. Natl. Acad. Sci. USA 77, 5794-5798. 28
- 29. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. 113, 237-251.
- Perry, R. P., LaTorre, J., Kelley, D. E. & Greenberg, J. R. (1972) 30. Biochim. Biophys. Acta 262, 220-226.
- Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-31. 1412.
- 32. Pelham, H. R. B. & Jackson, R. J. (1976) Eur. J. Biochem. 67, 247-256.
- Rothenberg, E. & Boyse, E. A. (1979) J. Exp. Med. 150, 777-791. Laemmli, U. K. (1970) Nature (London) 227, 680-685. 33.
- 34.
- 35. Wahl, G. M., Stern, M. & Stark, G. R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.
- Gross-Bellard, M., Oudet, P. & Chambon, P. (1977) Eur. J. Biochem. 36, 32-38. 36.
- 37.
- Jeffreys, A. J. & Flavell, R. A. (1977) Cell 12, 429-439. Domdey, H., Wiebauer, K., Kazmaier, M., Muller, V., Odink, K. & Fey, G. (1982) Proc. Natl. Acad. Sci. USA 79, 7619-7623. 38.
- Goldberger, G., Thomas, M. L., Tack, B. F., Williams, J., Col-39. ten, H. R. & Abraham, G. N. (1981) J. Biol. Chem. 24, 12617-12619.