AMINO ACID SEQUENCES OF MOUSE COMPLEMENT C3 DERIVED FROM NUCLEOTIDE SEQUENCES OF CLONED cDNA[a]

George H. Fey[b]

Scripps Clinic and Research Foundation
Department of Immunology
Research Institute of Scripps Clinic
La Jolla, California 92037

Karin Wiebauer[c]

Department of Human Genetics
University of Michigan Medical School
Ann Arbor, Michigan 48109

Horst Domdey[c]

Biology Department
California Institute of Technology
Pasadena, California 91125

INTRODUCTION

The complement system consists of a group of plasma proteins that play an important role in the defense of higher vertebrates against microbial infections.[1,2] In addition, complement and complement-derived peptides participate in inflammatory reactions and in the regulation of antibody synthesis by lymphocytes.[3-6] C3,[d] the third component of complement, is an essential element in the functioning of both the classical and the alternative pathways of complement activation. One of the most striking characteristics of C3 is its capacity to interact specifically with a number of other proteins, and all the net effects of activation of both the classical and the alternative pathways of complement depend upon these interactions of C3 and its derivatives:

—As a substrate C3 interacts with the C3-convertases of the classical and alternative pathways, which cleave it into the inflammatory peptide C3a (M_r = 8300) and the activated protein C3b (M_r = 171,000).

—As a constituent C3b participates in the assembly of the C5-convertases of both pathways and the alternative-pathway C3-convertase.

[a]This is publication number 2929IMM from the Department of Immunology, Scripps Clinic and Research Foundation, La Jolla, Calif. This work was supported by Grant 1 ROI AI 19651-01 from the National Institutes of Health, the National Institute of Allergy and Infectious Diseases, and by a grant from the Swiss National Science Foundation.

[b]Author to whom correspondence should be addressed.

[c]Drs. Domdey and Wiebauer are recipients of postdoctoral fellowships from the Deutsche Forschungsgemeinschaft and the European Molecular Biology Organization, respectively.

[d]**Abbreviations:** C3, C4, third and fourth components of complement; C3a, C3b, C3α, C3β, fragments and subunits of C3 abbreviated according to the Bulletin of the World Health Organization, 1968, **39:** 935; pro C3, precursor polypeptide of C3; cDNA, complementary DNA; mRNA, messenger RNA.

—C3b is the target of the control proteins β1H (H) and C3b-inactivator (I) of the alternative pathway.

—C3 and derived peptides interact as ligands with a variety of specific receptor proteins on the surface of mast cells, platelets, erythrocytes, monocytes/macrophages, and lymphocytes.[7-10]

In addition, C3b binds in a nonspecific manner to bacterial and cellular surfaces and to immune complexes through the activation of its internal thioester bond.[11-15]

One of the main tasks in the biochemical analysis of C3 is to map these interactive sites on the amino acid sequence of the polypeptide. This requires knowledge of the primary structure of C3. The protein is available in large amounts and can be purified, but the complete amino acid sequence has not been determined. Sequencing could be performed exclusively on the protein level; however, the analysis would benefit from confirmation by sequencing of complementary DNA (cDNA), which reflects the sequence of messenger RNA. This article describes the partial nucleotide sequence analysis of cloned mouse liver C3 cDNA and the derived amino acid sequences of a few selected functional domains of the polypeptide, in particular of the internal thioester site.

Materials and Methods

Double-stranded cDNA was prepared from size-fractionated mouse liver messenger RNA (mRNA) using reverse transcriptase. After addition of homopolymeric extensions with terminal transferase, it was inserted into the pst I restriction site of the plasmid pBR 322 and propagated in *Escherichia coli*, strain K12-HB101, using standard procedures. The cloning and the identification of C3 cDNA clones by hybrid-promoted translation have been described in Reference 16. DNA sequencing, the preparation of a genomic DNA library from strain A mouse liver DNA in the phage replacement vector λ 1059, and the isolation and characterization of C3 clones were performed as described.[17-19]

Results

The analysis of genomic DNA and cDNA sequences revealed several pieces of information about the C3 molecule and its biosynthetic pathways:

(a) Both the C3α and C3β subunits are contained in the sequence of a single-chain precursor protein: pro C3.

(b) The order of the subunits in the precursor is: NH_2-β-α-COOH.

(c) The initial translation product of C3 messenger RNA is a molecule, tentatively named prepro C3, carrying an extension peptide at its amino terminus, a so-called signal or leader peptide.[20] The complete amino acid sequence of the C3 signal peptide has been derived from cDNA and genomic DNA sequences: Met-Gly-Pro-Ala-Ser-Gly-Ser-Gln-Leu-Leu-Val-Leu-Leu-Leu-Leu-Leu-Ala-Ser-Ser-Pro-Leu-Ala-Leu-Gly. In its length, amino acid composition, and sequence this predicted peptide shows properties that are frequently found in other signal peptides, including a hydrophobic core region of eight amino acid residues.[18]

(d) The transition region of the C3β and C3α subunits in the precursor pro C3 has been analyzed. Unfortunately, no mouse C3 amino acid sequences are known which could be compared with the predictions from cDNA sequence data. At present comparison can only be made with the known partial amino acid sequences of human

C3.[21] From this comparison it was suggested that the proteolytic processing of pro C3 into the C3α and C3β subunits is not a single peptide bond hydrolysis, but a multiple-step event. In the transition region the precursor contains four consecutive arginine residues which are not found in the mature subunits and which must be removed during the maturation process.

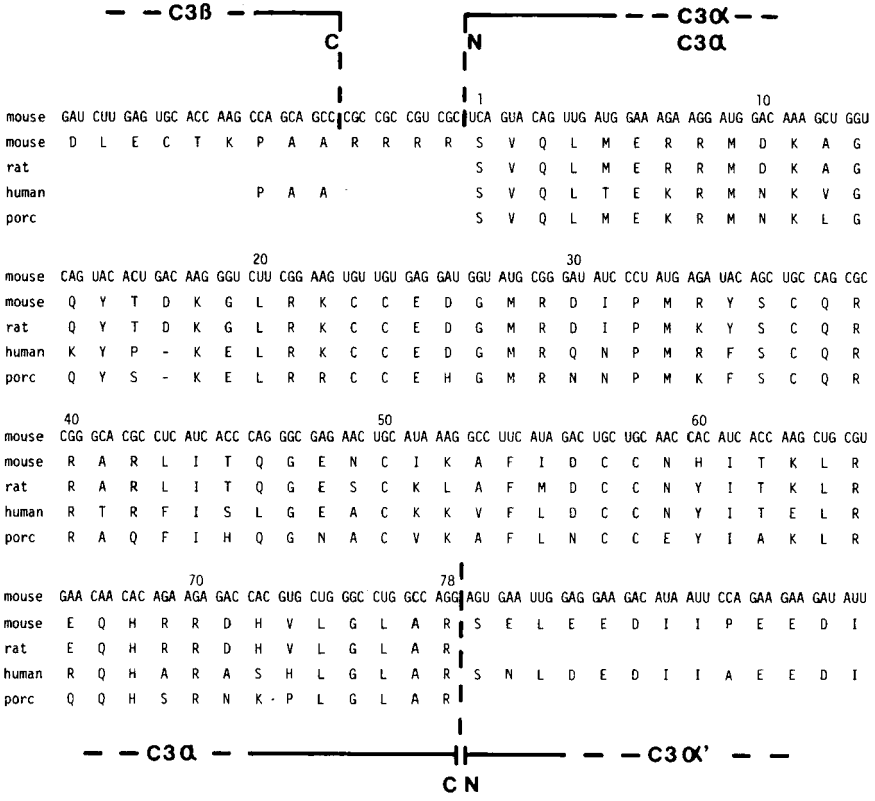

FIGURE 1. Amino acid sequences of C3a and its flanking regions in the precursor pro C3. Top: Positions of the COOH terminus (C) of the C3β subunit and the NH_2 terminus (N) of the C3α subunit and the C3a fragment. Middle: First row shows the mouse C3 mRNA sequence as derived from the cDNA sequence; second row gives the mouse C3 amino acid sequence as predicted from the mRNA (cDNA) sequence; third through fifth rows show amino acid sequences of the rat, human, and porcine C3a anaphylatoxins.[22-24] Bottom: Positions of the COOH terminus (C) of C3a and the NH_2 terminus (N) of C3α'. Reprinted from Domdey *et al.*, Proc. Natl. Acad. Sci. USA **79:** 7619–7623, 1982, with permission.

(e) The entire amino acid sequence of the inflammatory peptide C3a of mice has been determined as well as the following 13 amino acid residues of the NH_2 terminus of the C3α'-chain (FIG. 1). The mouse anaphylatoxin peptide C3a shows strong sequence conservation with rat (72/78 residues conserved), human (51/77), and porcine (50/77) C3a.[17]

(f) The cleavage between C3a and C3α′ by the C3-convertases is a single peptide bond hydrolysis. All the amino acids found in the precursor at this position become part of the active C3a and C3b fragments.[17, 21]

(g) Significant conservation of the thioester regions of mouse and human C3 was observed (35/41 residues, one uncertain residue) (see FIG. 2). All those amino acid residues that are shared in this region by human C3 and C4, and which are therefore believed to be important,[15, 25, 26] are also found in mouse C3 (the boxed residues in FIG. 2). The second glutamyl residue in the center of the thioester region, which participates in the formation of the thioester bond, was found to be encoded as a glutamine.

DISCUSSION

The findings summarized here confirm and extend earlier reports about the existence of the precursor pro C3 polypeptide and the order of the subunits within pro

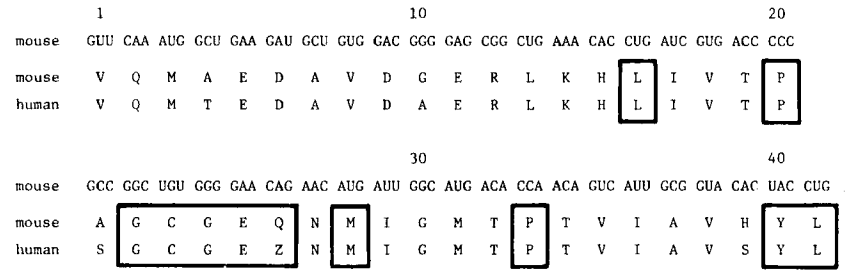

FIGURE 2. Comparison of the amino acid sequences encompassing the internal thioester site from murine (predicted from the nucleotide sequence) and human C3.[15] Upper row: Mouse mRNA sequence determined from cDNA sequence. Middle row: Mouse amino acid sequence predicted from cDNA sequence. Bottom row: Human amino acid sequence. The numbering of amino acid residues is according to Thomas *et al.*[15] The boxes indicate conserved amino acid residues that are identical in human C3, C4, and α_2-macroglobulin.[22, 23]

C3.[27–29] Since C3 is a secreted glycoprotein, it was expected to be initially synthesized with an amino-terminal signal peptide. However, a signal peptide had not previously been detected at the protein level. It may have escaped detection because most likely it is cleaved off during or soon after translation. Therefore the molecule tentatively named here prepro C3 may never exist in the cell as a full-length entity containing both the signal peptide and the entire pro C3 sequence. However, because the predicted signal peptide would have a number of properties that are characteristic for signal peptides, it appears highly probable that it indeed exists. This prediction needs to be confirmed at the protein level. Similarly, the existence of the four arginine residues in the transition region between C3β and C3α needs to be substantiated at the polypeptide level.

The cloned cDNA sequences coding for the mouse C3a anaphylatoxin are now precisely known. This offers the possibility of inserting these sequences into an expression vector and of producing mouse C3a in bacteria or in a eucaryotic expression system. Moreover, specifically designed variants of the peptide could probably be generated by means of genetic engineering. Peptide fragments that mimic the active

portion of C3a have recently been synthesized based on the human C3a sequence and shown to carry a substantial fraction of the biologic activities of native C3a.[4, 30] Since the inflammatory and immune regulatory functions of C3a need to be studied in much greater detail and because most of these functions can best be studied in mouse model systems, mouse C3a and variants thereof may become a valuable tool for these investigations.

Whenever protein sequencing is performed on human C3 and C4 proteins purified from plasma, the thioester bond is hydrolyzed. As a consequence, the residue that originally had participated in the formation of this bond appears as a glutamyl residue. As the formation and hydrolysis of the thioester bond are modifications of the primary amino acid sequence, it is difficult to determine the original nature of this residue from protein sequence data. The cDNA sequence implies a glutamine residue at this position in the original mouse C3 translation product. Several models have been formulated for the biosynthesis of this biologically important bond. One of them proposes an enzymatic conversion of a glutamyl residue into a pyroglutamyl residue as a first step.[31] If the initial residue is indeed a glutamine, then this model needs to be revised.

Two central questions need to be answered about the biosynthesis of the thioester bond:

(a) What are the primary amino acid sequence requirements for the formation of this bond?

(b) Can the biosynthesis of this bond proceed spontaneously or are enzymatic steps involved?

Cloned C3 cDNA sequences can now be used for studying both of these questions. In a first round of experiments, cloned cDNA sequences coding for the few core amino acids of the thioester region could be inserted into an expression vector and introduced into a eucaryotic expression system (tissue culture cells). The presence of the thioester bond in the resulting peptide products could be assayed for by incorporation of radioactive methylamine. If a bond is formed, the coding sequences can be modified by site-directed mutagenesis *in vitro* and amino acid residues that are essential for the formation of the bond can be identified. If no bond is formed in the first round, the coding sequences inserted into the vector could be progressively extended in the second round to include coding sequences for further amino-terminal or carboxy-terminal flanking residues of the core region or both. As soon as formation of the thioester bond occurs, the essential amino acids can again be identified by genetic manipulations of the coding sequences as described above.

The use of bacterial expression systems could be advantageous for studying whether formation of the thioester bond requires modifying enzymes. Bacteria do not possess some of the enzyme systems that are involved in post-translational modifications of eucaryotic proteins, such as the eucaryotic glycosylating enzymes. They may also not possess the enzymes required for the formation of the thioester bond. Therefore, the same coding sequence for a peptide containing the thioester region on a vector could be brought to expression in both a eucaryotic and a bacterial expression system. If the peptide produced in the eucaryotic system possesses the bond and the bacterial peptide does not, this would suggest an involvement of enzymatic steps and render a spontaneous synthesis unlikely. If, on the other hand, the bacterial product also has the bond, then this could indicate either spontaneous synthesis or enzymatic synthesis. The problem of identifying the enzymes involved would be greatly simplified, however, because bacteria are genetically less complex. The availability of cloned C3 sequences now allows us to approach these problems.

ACKNOWLEDGMENTS

The research described here was performed at the Swiss Institute for Experimental Cancer Research, Epalinges, Switzerland. We thank our colleagues from the Swiss Cancer Research Institute for their support.

REFERENCES

1. MÜLLER-EBERHARD, H. J. & R. D. SCHREIBER. 1980. Adv. Immunol. **23:** 1–53.
2. REID, K. B. M. & R. R. PORTER. 1981. Annu. Rev. Biochem. **50:** 433–464.
3. HUGLI, T. E. & H. J. MÜLLER-EBERHARD. 1979. Adv. Immunol. **26:** 1–53.
4. HUGLI, T. E. 1981. CRC Crit. Rev. Immunol. **1:** 321–366.
5. WEIGLE, W. O., E. L. MORGAN, M. G. GOODMAN, D. CHENOWETH & T. E. HUGLI. 1982. Fed. Proc. **41:** 3099–3103.
6. HOBBS, M. V., R. L. FELDBUSH, B. W. NEEDLEMAN & J. M. WEILER. 1982. J. Immunol. **128:** 1470–1475.
7. FEARON, D. T. 1980. J. Exp. Med. **152:** 20–30.
8. LAMBRIS, J. D., N. J. DOBSON & G. D. ROSS. 1981. Proc. Natl. Acad. Sci. USA **78:** 1828–1832.
9. PERLMAN, H., P. PERLMAN, R. D. SCHREIBER & H. J. MÜLLER-EBERHARD. 1981. J. Exp. Med. **153:** 1592–1603.
10. DIERICH, M. P., O. SCHEINER, H. H. MUSSEL, K. P. HAMMAN, R. E. SCHOPF & T. SCHULZ. 1982. Mol. Immunol. **19:** 1255–1265.
11. LAW, S. K. & R. P. LEVINE. 1977. Proc. Natl. Acad. Sci. USA **74:** 2701–2705.
12. LAW, S. K., D. T. FEARON & R. P. LEVINE. 1979. J. Immunol. **122:** 759–765.
13. TACK, B. F., R. A. HARRISON, J. JANATOVA, M. L. THOMAS & J. W. PRAHL. 1980. Proc. Natl. Acad. Sci. USA **77:** 5764–5768.
14. HOSTETTER, M. K., M. L. THOMAS, F. S. ROSEN & B. F. TACK. 1982. Nature **298:** 72–75.
15. THOMAS, M. L., J. JANATOVA, W. R. GRAY & B. F. TACK. 1982. Proc. Natl. Acad. Sci. USA **79:** 1054–1058.
16. ODINK, K. G., G. FEY, K. WIEBAUER & H. DIGGELMAN. 1981. J. Biol. Chem. **256:** 1453–1458.
17. DOMDEY, H., K. WIEBAUER, M. KAZMAIER, V. MULLER, K. ODINK & G. FEY. 1982. Proc. Natl. Acad. Sci. USA **79:** 7619–7623.
18. WIEBAUER, K., H. DOMDEY, H. DIGGELMAN & G. FEY. 1982. Proc. Natl. Acad. Sci. USA **79:** 7077–7081.
19. WHITEHEAD, A. S., E. SOLOMON, S. CHAMBERS, W. F. BODMER, S. POVEY & G. FEY. 1982. Proc. Natl. Acad. Sci. USA **79:** 5021–5026.
20. BLOBEL, G. & B. DOBBERSTEIN. 1975. J. Cell Biol. **67:** 852–862.
21. TACK, B. F., S. C. MORRIS & J. W. PRAHL. 1979. Biochemistry **18:** 1497–1503.
22. JACOBS, J. W., J. S. RUBIN, T. E. HUGLI, R. A. BOGARDT, I. K. MARIZ, J. S. DANIELS, W. H. DAUGHADAY & R. A. BRADSHAW. 1978. Biochemistry **17:** 5031–5038.
23. HUGLI, T. E. 1975. J. Biol. Chem. **250:** 8293–8301.
24. CORBIN, N. C. & T. E. HUGLI. 1976. J. Immunol. **117:** 990–995.
25. HARRISON, R. A., M. L. THOMAS & B. F. TACK. 1981. Proc. Natl. Acad. Sci. USA **78:** 7388–7392.
26. CAMPBELL, R. D., J. GAGNON & R. R. PORTER. 1981. Biochem. J. **199:** 359–370.
27. BRADE, V., R. E. HALL & H. R. COLTEN. 1977. J. Exp. Med. **146:** 759–765.
28. HALL, R. E. & H. R. COLTEN. 1977. Proc. Natl. Acad. Sci. USA **74:** 1707–1710.
29. GOLDBERGER, G., M. L. THOMAS, B. F. TACK, J. WILLIAMS, H. R. COLTEN & G. N. ABRAHAM. 1981. J. Biol. Chem. **256:** 12617–12619.
30. HUGLI, T. E. & B. W. ERICKSON. 1977. Proc. Natl. Acad. Sci. USA **74:** 1826–1830.
31. KHAN, S. A. & B. W. ERICKSON. 1982. J. Biol. Chem. **257:** 11864–11867.