Nucleotide Sequence of Complementary DNA and Derived Amino Acid Sequence of Murine Complement Protein C3

G. H. Fey, A. Lundwall, R. A. Wetsel, B. F. Tack, M. H. L. de Bruijn and H. Domdey

References	Article cited in: **http://rstb.royalsocietypublishing.org/content/306/1129/333#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**

Nucleotide sequence of complementary DNA and derived amino acid sequence of murine complement protein C3

G. H. FEY[1], Å. LUNDWALL[1], R. A. WETSEL[1], B. F. TACK[1], M. H. L. DE BRUIJN[1]
AND H. DOMDEY[2]

[1] *Department of Immunology, Research Institute of Scripps Clinic, La Jolla, California* 92037, *U.S.A.*

[2] *Division of Biology, California Institute of Technology, Pasadena, California* 91125, *U.S.A.*

The nucleotide sequences coding for murine complement component C3 have been determined from a cloned genomic DNA fragment and several overlapping cloned complementary DNA fragments. The amino acid sequence of the protein was deduced. The mature β and α subunits contain 642 and 993 amino acids respectively. Including a 24 amino acid signal peptide and four arginines in the β–α transition region, which are probably not contained in the mature protein, the unglycosylated single chain precursor protein preproC3 would have a molecular mass of 186484 Da and consist of 1663 amino acid residues. The C3 messenger RNA would be composed of a 56 ± 2 nucleotide long 5' non-translated region, 4992 nucleotides of coding sequence, and a 3' non-translated region of 39 nucleotides, excluding the poly A tail. The β chain contains only three cysteine residues, the α chain 24, ten of which are clustered in the carboxy terminal stretch of 175 amino acids. Two potential carbohydrate attachment sites are predicted for the α chain, none for the β chain. From a comparison with human C3 cDNA sequence (of which over 80% has been determined) an extensive overall sequence homology was observed. Human and murine preproC3 would be of very similar length and share several noteworthy properties: the same order of the subunits in the precursor, the same basic residue multiplet in the β–α transition region, and a glutamine residue in the thioester region. The equivalent position of the known factor I cleavage sites in human C3α could be located in the murine C3 α chain and the size and sequence of the resulting peptide were deduced. A comparison of the amino acid sequences of murine C3 and human alpha$_2$-macroglobulin is given. Several areas of strong sequence homology are observed, and we conclude that the two genes must have evolved from a common ancestor.

INTRODUCTION

Among the complement components C3 has received particular attention for several reasons: (i) it is the most abundant complement protein in mammalian serum and therefore, as the evolutionary argument goes, probably the one which is needed in highest amounts; (ii) it plays a crucial role in both the classical and the alternative pathways of complement activation (total C3 deficiencies in humans are frequently fatal); (iii) it is one of the members of the complement family about which the most detailed functional knowledge has been accumulated (Müller-Eberhard *et al.* 1980; Reid *et al.* 1981). This polypeptide is of interest because of its multiple functions. After cleavage by the classical and the alternative pathway C3 convertases it gives rise to the small peptide C3a and the large fragment C3b. C3b binds to some of the earlier acting complement components and to microbial particle surfaces. This happens mostly in the presence of specific antibodies, but can also occur in their absence. C3b participates in the

[55]

formation of the C5 convertases of the classical and alternative pathways by specific interactions with other constituents of these enzyme complexes. A very important function of C3b, separate from its participation in the events leading to bacteriolysis, is its opsonizing capacity. It can be covalently attached to the surfaces of invading microorganisms and lead to their enhanced uptake and clearance by phagocytic cells. Other peptide fragments generated from C3 by specific proteolytic cleavage also exhibit distinct functions. Anaphylatoxin C3a binds to receptors on mast cells and leads to release of histamine and other components stored in granules (Hugli 1981). It is also a regulator of the humoral immune response (Weigle *et al.* 1982). Fragment C3d-K, generated by the concerted actions of the proteases factor I, cofactor H and kallikrein, acts as an inhibitor of lymphocyte mitogenesis and as a leukocyte-mobilizing substance (Meuth *et al.* 1983). Fragment C3d-g, generated by the combined action of factors I and H, is entirely contained within C3d-K (Ross *et al.* 1982; Lachmann *et al.* 1982; Davis *et al.* 1983). Trypsin and plasmin further cleave C3d-g into C3g and C3d, which lack the leucocyte-mobilizing activity (Ross *et al.* 1982; Meuth *et al.* 1983). C3d and other C3 derived peptide fragments can bind to specific cell surface receptors on distinct and functionally differentiated cell types, and thereby mediate their effects (Schreiber 1984; Fearon 1983). To understand further the functions of these active peptides and in particular their interactions with receptors, it should be helpful to know their primary amino acid sequences as well as the primary structure of the C3 polypeptide. It would allow investigators to synthesize these peptides and variants thereof and to study their actions in greater detail.

In this paper we report the sequence of murine C3 determined by nucleotide sequence analysis of cloned genomic and cloned cDNA fragments. Additionally, we have compared murine C3 with available human C3 cDNA data and with the recently published amino acid sequence of human alpha$_2$-macroglobulin (Sottrup-Jensen *et al.* 1984a).

Experimental procedures

Recombinant plasmids pMLC3/4 and pMLC3/7 contain inserts of murine liver C3 cDNA in the vector pBR322 and were isolated and characterized as described (Domdey *et al.* 1982). The origin and properties of the recombinant phage λ MC3/KW4 carrying the 5'-half of the murine C3 gene have previously been reported (Wiebauer *et al.* 1982). Plasmid and phage DNA were prepared from large scale cultures by standard methods (Maniatis *et al.* 1982). Restriction enzyme fragments were recovered after electrophoresis in low melting temperature agarose gels. The fragments chosen for sequencing were: the 1956 base pair HindIII-HindIII fragment, carrying the first three exons of the gene, from λ MC3/KW4; the 1665 base pair EcoRI-PstI and 242 base pair StuI-StuI fragments from pMLC3/7 for the β chain coding sequences, and the 1990 base pair BglII-PvuI, 1239 base pair PstI-PstI and 393 base pair SalI-HindIII fragments from pMLC3/4 for the α chain coding sequences. The respective map locations of these fragments are given elsewhere (Wetsel *et al.* 1984; Lundwall *et al.* 1984). Sequencing was according to standard protocols (Bankier *et al.* 1983; Biggin *et al.* 1983). Briefly, these fragments were self-ligated to oligomerize them and then sonicated. Subfragment ends were repaired with T4 DNA polymerase and size selected by agarose gel electrophoresis. The size fraction of 300–600 base pairs was ligated randomly into the SmaI site of the phage vector M13mp8. The recombinant DNA was used to transfect *Escherichia coli* JM101-TG1 cells (T. Gibson, unpublished). Single-stranded phage DNA ('template DNA') was prepared from plaque-

isolated small scale cultures and sequenced using the dideoxynucleotide chain termination method (Sanger *et al.* 1977; Messing *et al.* 1982). The sequence data were analysed with the computer programs of Staden (1982*a*, *b*).

Approximately 44 000 characters of primary sequence data were generated with the dideoxy-technique in addition to the partial sequence data generated with the Maxam and Gilbert technique which have previously been published (Domdey *et al.* 1982; Wiebauer *et al.* 1982; Fey *et al.* 1983). Each nucleotide of the cDNA sequence given in figure 1 was sequenced on the average over six times with the dideoxy-technique, and over 95 % of the sequence have been covered by two or more independent gel readings on both of the DNA strands.

Results

Murine C3 coding sequences

A genomic DNA fragment from clone λ MC3/KW4, carrying the 5′-end of the mouse C3 gene and the C3 cDNA inserts of the recombinant plasmids pMLC3/4 and pMLC3/7 comprised all the sequences that are retained in the mature C3 mRNA. Suitable subfragments of these clones were prepared and sequenced as described in the experimental methods section. The resulting sequence is given in figure 1 is a composite.

The cDNA sequence given here reflects all the nucleotides of the mature C3 mRNA, and was derived from a combination of genomic DNA and cDNA and cDNA sequences together with other experimental data reported previously (Wiebauer *et al.* 1982; Domdey *et al.* 1982). According to these data the C3 mRNA would be composed of a 56 ± 2 nucleotide 5′ non-translated region, a 4992 nucleotide coding sequence (including the termination codon) and a 39 nucleotide 3′ non-translated region. Altogether the mRNA would be 5087 nucleotides long, excluding the poly A tail.

The derived amino acid sequence shows a single chain preproC3 precursor polypeptide of 1663 amino acids, molecular mass 186 484 Da (non-glycosylated). The signal peptide is 24 amino acids long (molecular mass 2322 Da), the mature β chain 642 amino acids (molecular mass 70 641 Da non-glycosylated) and the mature α chain 993 amino acids (molecular mass 112 933 non-glycosylated). Two potential carbohydrate attachment sites were located in the α chain coding sequences: amino acid residues 939 (Asn-Lys-Thr) and 1617 (Asn-Thr-Ser) in figure 1, none were found in the β chain coding sequences. The triplets (Asn-Lys-Thr) and (Asn-Thr-Ser) are known as the attachment sites for the most common carbohydrate chains in human glycoproteins (Marshall 1972). The β chain contains only three cysteine residues, the α chain 24. At the carboxy terminal end, ten of these are clustered in a noticeably ordered pattern (figure 2).

Comparison of murine and human C3 cDNA sequences

A subclone of a human C3 genomic DNA clone has previously been prepared and was reported to contain no repetitive sequences (Whitehead *et al.* 1982). This clone was used as a probe to screen a human liver cDNA library which was prepared by Ms T. Belt and Dr M. C. Carroll under conditions optimized for long cDNA inserts. One of the clones found contains a cDNA insert of approximately 4300 base pairs, covering the coding portion for the carboxy terminal part of human preproC3, but lacking about 600 nucleotides from the 5′ end of human C3 mRNA. This insert has been sequenced with the dideoxy chain-termination
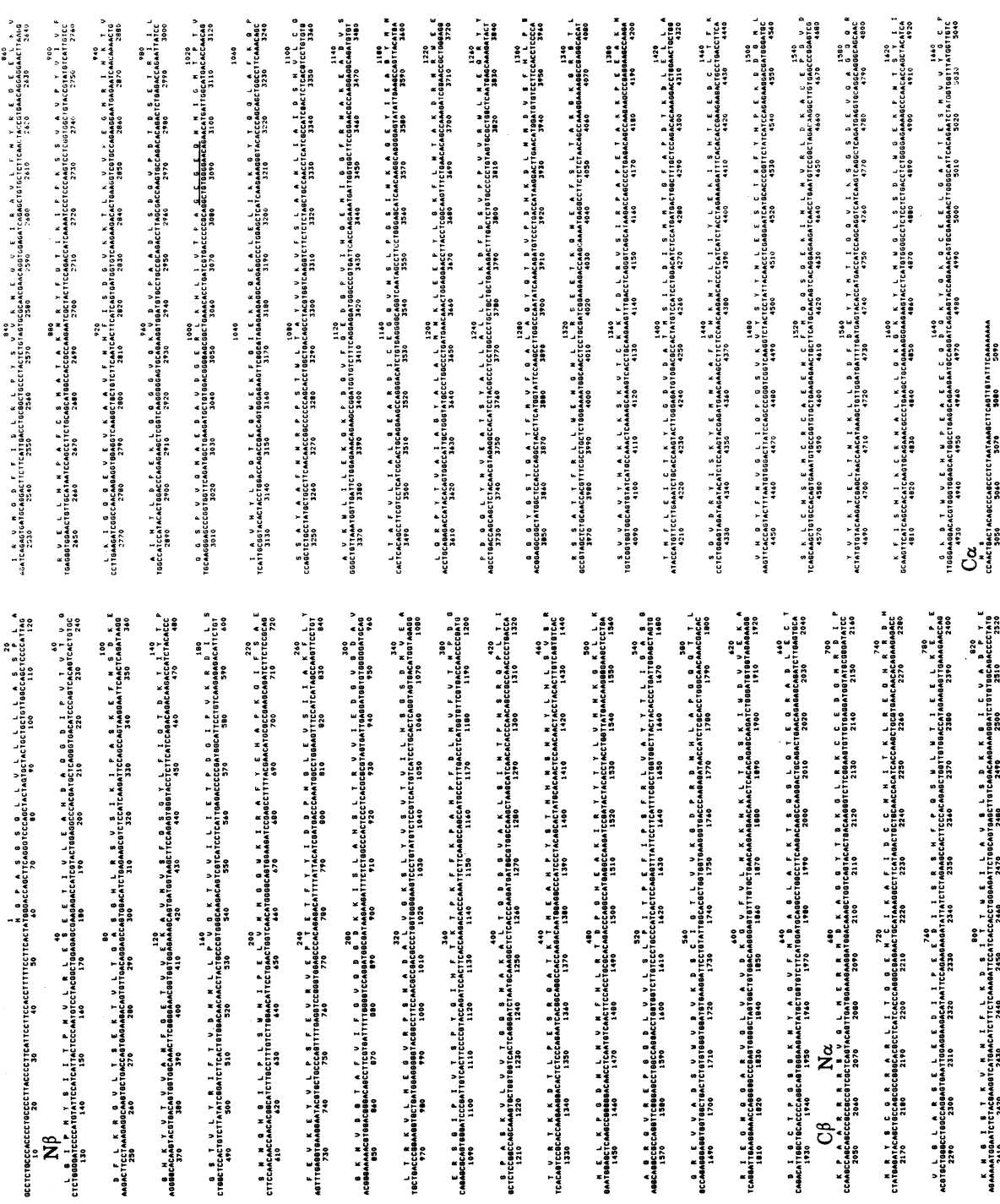
FIGURE 1. Nucleotide sequence of cDNA reflecting the murine C3 mRNA and derived amino acid sequence. The amino- and carboxy-termini of the mature β and α chains are labelled Nβ, Nα, Cβ and Cα, the thioester site is underlined amino acids 1019–1023. Numbering of the nucleotide sequence starts at the cap site. Poly A tail starts at 5088. Numbering of the amino acid sequence starts at the N-terminus of the signal peptide.
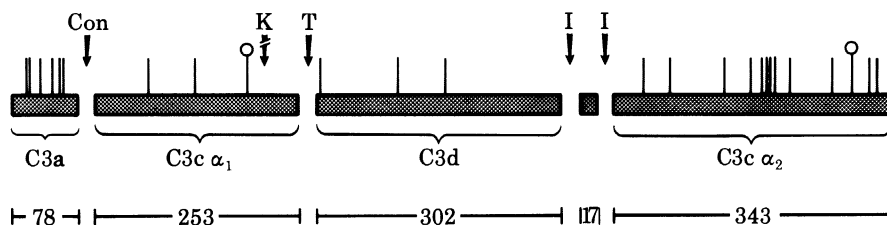
FIGURE 2. A proposed model for the murine C3 α chain, illustrating the position of cystein residues (vertical bars) and possible carbohydrate attachment sites ( ). The cleavage sites for C3 convertase (Con), and the proposed cleavage sites for trypsin (T) and factor I (I) are indicated by arrows. The site of kallikrein cleavage (K) which occurs in human C3, but may not occur in murine C3 is shown by a slashed arrow. The number of amino acids contained in the various α chain fragments are noted under the respective fragments. C3c is composed of both C3c α chain fragments C3c α_1, and C3c α_2 linked covalently to the β chain.
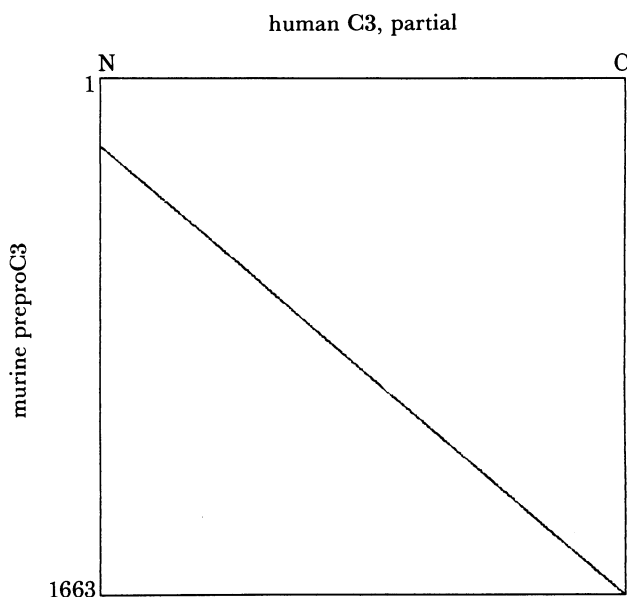

FIGURE 3. Diagonal matrix comparison of murine C3 amino acid sequence and that part of the human C3 amino acid sequence that is currently known. Horizontal axis: partial human C3 sequence, left-to-right corresponds to N-to-C orientation. Vertical axis: murine C3 amino acid sequence (preproC3). Top-to-bottom corresponds to N-to-C orientation. The correspondence line is not a diagonal, because the murine preproC3 sequence is complete and the human sequence is not. Sliding window was 25 amino acids, percentage score parameter was 280. The analysis was performed with the program DIAGON (Staden 1982 b).

technique (M. H. L. de Bruijn, unpublished data). The nucleotide sequence and derived amino acid sequence were compared with the corresponding murine C3 sequences using the diagonal matrix computer program DIAGON (Staden 1982 b). The result (figure 3) illustrates, that a significant sequence homology was observed on the amino acid level, evenly distributed along the whole sequence so far available. Human and murine C3 share the following properties: the order of the subunits is the same for both (N-β-α-C), the precursor proC3 contains four arginine residues in the β–α transition region (figure 4) and a glutamine residue is found in the thioester region of both molecules (figure 5).

Limited amino acid sequence data have been published for certain peptide fragments derived from human C3 (Tack *et al.* 1979; Tack *et al.* 1982; Thomas *et al.* 1982; Goldberger *et al.* 1981;

Meuth *et al.* 1983). By comparing those data with the human and murine C3 cDNA sequences, it was possible to locate the equivalent positions on the murine C3 sequence. For example the amino termini of two human α chain fragments generated by the action of factors I and H have been reported (Davis *et al.* 1983). Both of these cleavage events would occur between an arginine and a serine residue in the murine C3 α chain, after residues 1303 and 1320 in figure 1. Although no corresponding data have been reported for the murine C3 protein, it appears

(a)	mouse C3 β—α	...P A A R R R R S V Q...
(b)	human C3 β—α	...P A A ＋ R R R R ＋ S V Q...
(c)	human C4 α—γ	.... R N ＋ R R R R ＋ E A P...
(d)	mouse C4 α—γ	...R R S R R R R E A P...
(e)	human c.g.r.p.	...:...G R R R R

FIGURE 4. Basic residue multiplets as recognition or processing sites, or both, for enzymes involved in the maturation of mature C3 and C4 from their respective precursors. (*a*) Murine cDNA sequence from Domdey *et al.* 1982; (*b*) human cDNA sequence from de Bruijn, this symposium (*c*) human C4 cDNA sequence from Whitehead *et al.* (1983); (*d*) murine C4 cDNA sequence from Ogata *et al.* (1983); (*e*) human calcitonin gene-related peptide (c.g.r.p.) from Rosenfeld *et al.* (1983). Sequence between arrows known from protein sequence data not to be present in the mature polypeptide.

(a)	C3 human protein	..L I V T P S G C G E Z N M I G M T P..
(b)	C3 human cDNA	..L I V T P S G C G E Q N M I G M T P..
(c)	C3 murine cDNA	..L I V T P A G C G E Q N M I G M T P..
(d)	C4 human protein	..L L R L P R G C G E E T M I Y L´ A P..
(e)	A2MAC human protein	..L L Q M P Y G C G E E N M V L F A P..
		* * * * * *(*) * *

FIGURE 5. The internal thioester region in C3, C4 and alpha$_2$-macroglobulin. Asterisks indicate identical residues, an asterisk in parentheses indicates identity after hydrolysis of the thioester bond. (*a*) From Thomas *et al.* (1982); (*b*) from de Bruijn, this symposium (*c*) from Domdey *et al.* (1982); (*d*) from Campbell *et al.* (1981); (*e*) from Swenson & Howard (1980).

plausible that these positions would also qualify for cleavage of murine C3 by factor I, and that factor I may therefore be a trypsin-like protease. These positions are indicated in figure 2. Similarly, Thomas *et al.* (1982) have located a trypsin cleavage site eight residues upstream of the thioester site in human C3, which was used to define the amino terminus of C3d. Amino acid residue 1001 in murine preproC3 (figure 1), located at an equivalent position, is a lysine. It is considered very probable, that murine C3 would also be cleaved by trypsin at this position, and this site is therefore indicated in figure 2. An elastase cleavage site is present 13 residues upstream from this trypsin site in human C3 (Tack *et al.* 1982). The cleavage occurs between a valine and an alanine residue. The equivalent two murine residues (987 and 988 in figure 1) are both valines, and it is therefore not clear whether elastase would also cleave murine C3 at this position. Finally, a kallikrein cleavage site has been reported upstream of this elastase site for human C3 (Meuth *et al.* 1983). Together with factors I and H, kallikrein produces the C3d-K peptide, a C3d-related peptide, but larger (reported molecular masses 25 000–34 000 Da for C3d, 41 000 Da for C3d-K, Tack *et al.* (1982); Meuth *et al.* (1983)). The C3d-K, peptide

has two apparently independent functions: suppression of T cell proliferation and leucocyte mobilization; C3d possesses only the T cell proliferation-suppressing activity, but is unable to mobilize leucocytes. Partial amino acid sequence data for the amino termini of both peptides have been obtained, and they were found not to overlap. On this basis C3d-K was proposed to be larger than C3d by an amino terminal extension peptide of molecular mass 8000–10 000 Da, that would most likely carry the leucocyte-mobilizing activity (Meuth *et al.* 1983). The
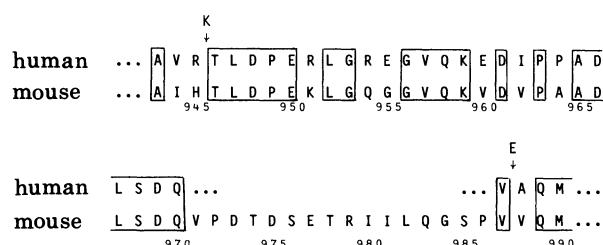

FIGURE 6. Structure of a potential leucocyte-mobilizing peptide from murine C3 α. The partial sequence data of human C3d-K of Meuth *et al.* (1983) and of C3d (Thomas *et al.* 1982) were compared with the murine cDNA-derived sequence data to produce this alignment. K: kallikrein cleavage site; E: elastase cleavage site.

corresponding sequences could be located on the murine and human C3 sequences presented here. The model proposed by Meuth *et al.* (1983) was found to be correct. Our data suggest that the kallikrein cleavage site is indeed located on the amino terminal side of the elastase site (figures 2 and 6). However, the difference between C3d-K and C3d (generated by elastase) for the murine peptide would be only 42 amino acids.

The difference peptide, possibly the equivalent of the previously described C3e fragment (Ghebrehiwet *et al.* 1979), would then have a molecular mass only half as big as the value previously estimated for the human fragment (Meuth *et al.* 1983). Although C3d-K may be an important breakdown fragment of human C3, because it occurs under natural conditions in sterile serum (Lachmann *et al.* 1982), it may not have a functional analogue derived from murine C3. Kallikrein cleavage of human C3 would be predicted from the data of Meuth *et al.* (1983) and us to occur after an arginine residue (figure 6), whereas the corresponding residue in murine C3 (amino acid 945 in figure 1 and figure 6) is a histidine. It is therefore not clear, whether a kallikrein-like protease would cleave murine C3 at this position (figure 2).

Comparison of murine C3 and human alpha$_2$-macroglobulin

The complete amino acid sequence of human alpha$_2$-macroglobulin has recently been published (Sottrup-Jensen *et al.* 1983). We have performed a detailed comparison with the murine C3 sequence and the full account of this will be published separately (Sottrup-Jensen *et al.* 1984). Figure 7 gives the result of a comparison using the diagonal matrix computer program DIAGON (Staden 1982b).

There are approximately 15 regions of significant homology (amino acid conservation of about 25%). The thioester region and one region in the amino terminal part of the C3 β chain (amino acid residues 110–120 in figure 1) show the highest degree of amino acid conservation (30–50%). Three selected areas of high homology are printed in a residue by residue comparison in figure 8.

From these data it must be concluded that the genes for C3-and alpha$_2$-macroglobulin were
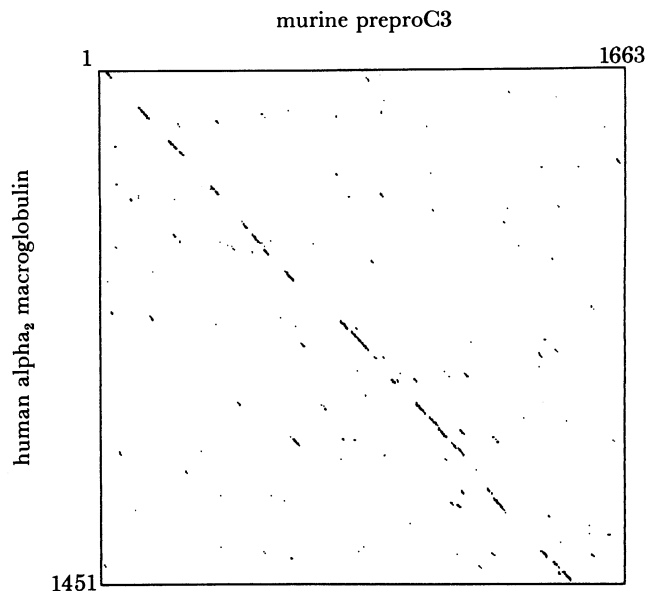
murine preproC3

FIGURE 7. Diagonal matrix comparison of murine C3 and human alpha$_2$-macroglobulin amino acid sequences. Horizontal axis: murine preproC3, 1693 amino acid residues, left-to-right orientation corresponds to N-to-C orientation. Vertical axis: human alpha$_2$-macroglobulin, 1451 amino acids (Sottrup-Jensen et al. 1983); top-to-bottom orientation corresponds to N-to-C orientation. Sliding window was 25 amino acids, percentage score parameter setting was 280. The comparison was performed with the program DIAGON (Staden 1982b).
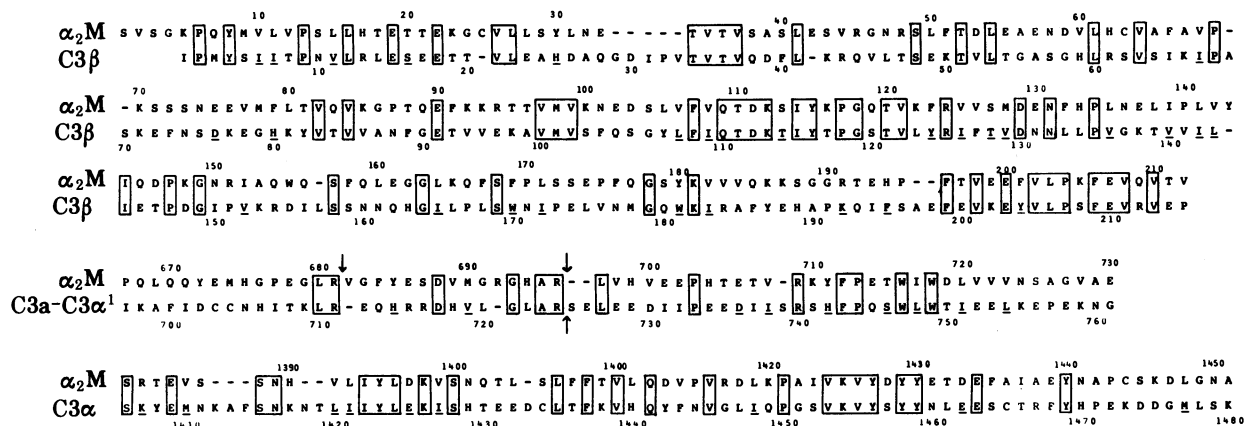

FIGURE 8. Homology regions between murine C3 and human alpha$_2$-macroglobulin (Sottrup-Jensen et al. 1984). After the strong homology regions had been located with the help of the DIAGON program (figure 7), visual inspection was used to optimize the alignment. Gaps or deletions were only introduced, if for each gap or deletion the overall alignment improved by at least two amino acid residues. Identical residues are boxed, chemical equivalent residues underlined. Numbering in the C3 sequence is different from figure 1 and starts with the first residue of the mature β chain. Arrows in the alpha$_2$-macroglobulin sequence indicate activation cleavage site (681) and possibly deleted activation site (residue 696) at the C-terminus of the remnants of an anaphylatoxin-like sequence. Arrow in the C3 sequence at this position indicates activation cleavage site for the C3 convertases.

derived from a common ancestral gene. It is striking to note, that C3 is longer than alpha$_2$-macroglobulin by 158 amino acid residues at its carboxy terminus, that the locations of the cysteine residues are entirely different, and that about half of the equivalent coding sequences for the C3a anaphylatoxin peptide are missing in alpha$_2$-macroglobulin. The arginine quartet at the C3 β–α junction would also be absent in alpha$_2$-macroglobulin. If this

basic residue multiplet in the transition region constitutes an essential recognition or cleavage sequence for processing, or both, as we propose, then this would explain why alpha$_2$-macroglobulin is a single-chain polypeptide.

A region equivalent to part of the C3 convertase cleavage site in C3 is preserved in alpha$_2$-macroglobulin: an alanine–arginine doublet is present in both proteins at this site (residues 695 and 696 in alpha$_2$-macroglobulin, 723 and 724 in prepro C3, figure 8, which corresponds to residues 747 and 748 in figure 1). However, the following two amino acids of C3 (serine and glutamic acid) have no equivalent in alpha$_2$-macroglobulin. From the pattern of sequence conservation between both molecules that is resumed downstream of this area in alpha$_2$-macroglobulin (figure 8) it is likely that the corresponding two amino acids have been deleted from the alpha$_2$ sequence during evolution. This observation would explain why alpha$_2$ is not a substrate for C3 convertase-like proteases which have absolute sequence specificity for activation cleavage at this unique site (Müller-Eberhard & Schreiber 1980). It is known that several different activation cleavages of alpha$_2$-macroglobulin can occur in the neighbourhood of this site (around positions 680 and 708 in the alpha$_2$-macroglobulin sequence, figure 8, Sottrup-Jensen *et al.* 1984*a*), at a spectrum of possible sites. Although alpha$_2$-macroglobulin would not give rise to an exact equivalent of the C3a anaphylatoxin, it is worthwhile to note that it contains a sequence homologous to the active carboxy terminus of C3a (LGLAR) (Hugli 1981).

DISCUSSION

Generally, it is not possible to deduce mRNA sequence from an analysis of genomic DNA sequence data only without further independently generated experimental information, because of the potential use of different transcription initiation sites and of alternative mRNA maturation pathways. Therefore, the determination of the murine C3 mRNA sequence reported here has relied not only on a combination of genomic and cDNA nucleotide sequence data, but also on two other sources of information: (i) on sequence generated by primer extension experiments, using specific cloned fragments as primer, as previously described (Domdey *et al.* 1982; Wiebauer *et al.* 1982) and (ii) on the knowledge of partial amino acid sequence data of peptide fragments from the amino terminal portion of the human and guinea-pig C3 β chains (see below). cDNA sequence data were derived here for the major part of the mRNA sequence, extending up to the EcoRI site (nucleotide 342 in figure 1). The EcoRI-HindIII fragment (nucleotides 342–428 in figure 1) was used, labelled at the HindIII end, for a first primer-extension experiment with murine liver mRNA as a template. The sequence analysis of the product resulted in another 100 nucleotides of sequence data. In addition to genomic sequence data, amino acid sequences for the amino termini of human and guinea-pig C3 β had been published (Goldberger *et al.* 1981; Thomas *et al.* 1982). Also, Lundwall *et al.* (1983) have determined partial sequences for 11 cyanogen bromide fragments of the human C3 β chain. Two of these, CN7a and CN3 matched very well with the mouse data derived from exon 2 of the gene, starting with YSIITPN ... (amino acid 28 in figure 1) and VLEAHDAQGD ... (amino acid 45 in figure 1). Therefore, the sequence could be considered correct until the amino terminus of the mature β chain. The fact that the sequence of the first exon overlapped by three amino acids with the amino terminus of the β chain (IPM amino acid 25 in figure 1) would probably be sufficient to determine the reading frame, and therefore the translation initiation codon of the signal peptide. However, to determine the cap site of

the mRNA in a formally correct manner, another primer-extension experiment has been performed (Wiebauer *et al.* 1982). Here a BamHI-HaeIII subfragment of the first exon was used as a primer on murine liver mRNA, and the resulting cDNA copy was subjected to size analysis on a sequencing gel, containing the corresponding genomic DNA sequence ladder as a size marker. This experiment allowed location of the cap site with a precision of ± 2 nucleotides at the position of nucleotide 1 in figure 1. From all of these different data together it can be concluded that the sequence reported here for the mature mRNA is probably correct. The size of the mRNA determined here agrees well with the previously reported value of 5100 ± 200 base pairs, determined using different techniques (Domdey *et al.* 1982).

One of the main practical uses of the C3 amino acid sequence will be to determine the sequence of certain peptide fragments. Often, only partial amino terminal and carboxy terminal sequence data can be generated for such peptides on the protein level. Once these can be located on the primary structure of the whole C3 molecule, the entire sequence of the peptides can be deduced. An example of this was shown here: the murine equivalent of the small fragment generated by factor I cleavage of the human C3 α chain could be deduced. It was found to consist of only 17 amino acid residues (figure 2). Also, the sequence of the murine equivalent of the human peptide fragment, which is located between a kallikrein and an elastase site in C3d-K, was derived. This peptide could now be synthesized and the question of whether this murine fragment or further subfragments of it also contain a leucocyte-mobilizing activity similar to that of the human peptide reported by Meuth *et al.* (1983), could be experimentally answered. The synthesis of peptides involved in the interaction of C3 with the various cell surface receptors will be another important objective, whose solution may benefit from the sequence information now available.

It can be expected that functionally essential areas of the C3 molecule will have a more strictly conserved amino acid sequence between the murine and human protein than others. Therefore, from studying strictly and less strictly conserved amino acid sequences in both proteins, we may be able to. generate additional predictions for the mapping of functional domains on the molecule. This expectation was already borne out in two instances reported here: the β–α transition region and the thioester region, including the glutamine residue, was found to be strictly identical. Similarly, a comparison of C3, C4, C5 and alpha$_2$-macroglobulin sequences would help to delineate common functional domains. Consequently, from an analysis of non-conserved regions, one might obtain indications for the location of functional domains specific for each one of these molecules. For example, the carboxy terminal 158 amino acid residues of C3 have no equivalent in alpha$_2$-macroglobulin, yet they contain a cluster of cysteine residues. They may therefore constitute a functional domain specific for C3. To test ultimately the idea that a certain region of a sequence constitutes an independently functioning domain, it would be necessary to synthesize a peptide reflecting this region and to demonstrate that it mediates the function in question; this may only be possible in a minority of cases. However, as pointed out above, alpha$_2$-macroglobulin shares some of the sequences of the functionally very important carboxy terminus of the C3a anaphylatoxin. It would be interesting to synthesize the peptide VGFYESDVMGRGHAR, located between the two arginine residues at amino acid residues 681 and 696 in figure 8, and to ask whether it mediates some form of anaphylatoxin function. This might lead to a better understanding of anaphylatoxins and possibly· to the discovery of corresponding active peptides derived physiologically from alpha$_2$-macroglobulin.

Finally, the observed relatedness of the genes for C3 and alpha$_2$-macroglobulin raises one point of importance to researchers interested in the control of gene expression. It should now be possible to isolate the corresponding genes, to compare the DNA sequences involved in the control of their expression and to ask the question, why C3 is an acute phase reagent in human inflammation, and why alpha$_2$-macroglobulin is not.

The sequence data reported here and the comparison of C3 with other related proteins have led us to design experiments aimed at understanding the inflammatory regulation of these genes. Also, these data may become helpful in the analysis of the wild-type human C3 gene and its defective alleles, which are responsible for inherited C3 deficiencies in humans.

REFERENCES

Bankier, A. T. & Barrell, B. G. 1983 Shotgun DNA sequencing. In *Techniques in nucleic acid biochemistry* (ed. R. A. Flavell), vol. B5, pp. 1–34. New York, Amsterdam: Elsevier.

Biggin, M. D., Gibson, T. J. & Hong, G. F. 1983 Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963–3965.

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

Davis, A. E. & Harrison, R. A. 1982 Structural characterization of factor I mediated cleavage of the third component of complement. *Biochemistry, Wash.* **21**, 5745–5749.

Davis, A. E., Harrison, R. A. & Lachmann, P. J. 1983 C3b inactivation: isolation and structural analysis of C3c, C3dg (α2D), and C3g. *Immunobiology* **164**, 231.

Domdey, H., Wiebauer, K., Kazmaier, M., Müller, V., Odink, K. & Fey, G. H. 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.

Fearon, D. T. 1983 The human C3b receptor. *Springer Semin. Immunopathol.* **6**, 159–172.

Fey, G. H., Domdey, H., Wiebauer, K., Whitehead, A. S. & Odink, K. 1983 Structure and expression of the C3 gene. *Springer Semin. Immunopathol.* **6**, 119–147.

Ghebrehiwet, B. & Müller-Eberhard, H. J. 1979 C3e: Acidic fragment of human C3 with leukocytosis-inducing activity. *J. Immunol.* **123**, 616–621.

Hugli, T. E. 1981 The structural basis for anaphylatoxin and chemotactic functions of C3a, C4a and C5a. *C.R.C. Crit. Rev. Immunol.* **1**, 321–366.

Lachmann, P. J., Pangburn, M. K. & Oldroyd, R. G. 1982 Breakdown of C3 after complement activation. *J. exp. Med.* **156**, 205–216.

Lundwall, Å., Hellman, U., Eggertsen, G. & Sjöquist, J. 1984 Chemical characterization of fragments from a cyanogen bromide digest of human C3 β chain. *FEBS Lett.* (Submitted).

Lundwall, Å., Wetsel, R. A., Domdey, H., Tack, B. F. & Fey, G. H. 1984 Structure of murine complement component C3. I. Nucleotide sequence of cloned complementary and genomic DNA coding for the β chain. *J. biol. Chem.* (Submitted.)

Maniatis, T., Fritsch, E. F. & Sambrook, J. 1982 *Molecular cloning. A laboratory manual.* Cold Spring Harbor Press.

Marshall, R. D. 1972 Glycoproteins. *Ann. Rev. Biochem.* **41**, 673–702.

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.

Meuth, J. L., Morgan, E. L., DiScipio, R. G. & Hugli, T. E. 1983 Suppression of T lymphocyte functions by human C3 fragments. *J. Immunol.* **130**, 2605–2611.

Müller-Eberhard, H. J. & Schreiber, R. D. 1980 Molecular biology and chemistry of the alternative pathway of complement. *Adv. Immunol.* **29**, 1–53.

Ogata, R. T., Schreffler, D. C., Sepich, D. S. & Lilly, S. P. 1983 A cDNA clone spanning the alpha-gamma subunit junction in the precursor of the murine fourth complement (C4). *Proc. natn. Acad. Sci. U.S.A.* **80**, 5061–5065.

Reid, K. B. M. & Porter, R. R. 1981 The proteolytic activation systems of complement. *A. Rev. Biochem.* **50**, 433–464.

Rosenfeld, M. G., Mermod, J. J., Amara, S. G., Swanson, L. W., Sawchenko, P. E., Rivier, J., Vale, W. W. & Evans, R. M. 1983 Production of a novel neuropeptide encoded by the calcitonin gene via tissue specific RNA processing. *Nature, Lond.* **304**, 129–135.

Ross, G. D., Lambris, J. D., Cain, J. A. & Newman, S. L. 1982 Generation of three different fragments of bound C3 with purified factor I or serum. *J. Immunol.* **129**, 2051–2060.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977 DNA sequencing with chain-terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

Schreiber, R. D. 1984 Complement receptors. *Springer Semin. Immunopathol.* **7** (In the press.)

Sottrup-Jensen, L., Stepanik, T. M., Wierzbicki, D. M., Jones, C. M., Lønblad, P. B., Kristensen, T., Petersen, T. E. & Magnusson, S. 1983 The primary structure of α_2-macroglobulin and localization of a factor XIIIa crosslinking site. *Ann. N.Y. Acad. Sci.* **421**, 41–60.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Lønblad, P. B., Jones, C. M., Wierzbicki, D. M., Magnusson, S., Domdey, H., Wetsel, R., Lundwall, Å, Tack, B. F. & Fey, G. H. 1984 Common evolutionary origin of alpha$_2$-macroglobulin and complement components C3 and C4. *Proc. natn. Acad. Sci. U.S.A.* (Submitted.)

Swenson, R. P. & Howard, J. B. 1980 Amino acid sequence of the tryptic peptide containing the alkylamine-reactive site from human alpha$_2$-macroglobulin. *J. biol. Chem.* **255**, 8087–8091.

Staden, R. 1982a Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucl. Acids Res.* **10**, 4731–4751.

Staden, R. 1982b An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucl. Acids Res.* **10**, 2951–2961.

Tack, B. F., Morris, S. C. & Prahl, J. W. 1979 Third component of human complement: Structural analysis of the polypeptide chains of C3 and C3b. *Biochemistry, Wash.* **18**, 1497–1503.

Tack, B. F., Janatova, J., Thomas, M. L., Harrison, R. A. & Hammer, C.H. 1982 The third, fourth, and fifth components of human complement: Isolation and biochemical properties. *Methods Enzymol.* **80**, 64–101.

Thomas, M. L., Janatova, J., Gray, W. R. & Tack, B. F. 1982 Third component of human complement: Localization of the internal thiolester bond. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1054–1058.

Weigle, W. O., Goodman, M. G., Morgan, E. L. & Hugli, T. E. 1983 Regulation of immune response by components of the complement cascade and their activated fragments. *Springer Semin. Immunopathol.* **6**, 173–194.

Wetsel, R. A., Lundwall, Å., Davidson, F., Gibson, T., Tack, B. F. & Fey, G. H. 1984 Structure of murine complement component C3. II. Nucleotide sequence of cloned complementary DNA coding for the α chain. *J. biol. Chem.* (Submitted.)

Whitehead, A. S., Solomon, E., Chambers, S., Bodmer, W. F., Povey, S. & Fey, G. H. 1982 Assignment of the structural gene for the third component of human complement to chromosome 19. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5021–5026.

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 genetic deficiency of C4 in guinea pig. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5387–5391.

Wiebauer, K., Domdey, H., Diggelmann, H. & Fey, G. H. 1982 Isolation and analysis of genomic DNA clones encoding the third component of mouse complement. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7077–7081.