Structural Features of the Human C3 Gene: Intron/Exon Organization, Transcriptional Start Site, and Promoter Region Sequence†

Dennis P. Vik,‡ Patrick Amiguet,‡,§ Graeme J. Moffat,‡ Marianne Fey,‡ Françoise Amiguet-Barras,‡ Rick A. Wetsel,‖ and Brian F. Tack*,‡

Department of Immunology, Scripps Clinic and Research Foundation, La Jolla, California 92037, and Department of Pediatrics, Children's Hospital, Washington University School of Medicine, St. Louis, Missouri 63110

Received August 7, 1990; Revised Manuscript Received October 15, 1990

ABSTRACT: The third component of human complement (C3) is a key molecule in the activation of the complement cascade. C3 cDNA fragments were used to identify seven cosmid clones that covered all but 1 kilobase pair (kb) of the C3 gene. The remainder of the gene was cloned by using the polymerase chain reaction. These clones were used to identify the intron/exon boundaries and to map the gene. The C3 gene is 42 kb in length and comprises 41 exons ranging in size from 52 to 213 base pairs (bp). The transcription start site was identified by primer extension, and approximately 1 kb of DNA upstream of this site was sequenced. Putative TATA and CAAT boxes were identified along with a number of regions that shared homology with known regulatory sequences. These include responsive elements for interferon-γ, interleukin-6, nuclear factor κB, estrogen, glucocorticoids and thyroid hormone. Several of these agents have been shown to affect C3 synthesis and mRNA levels. The sizes of the exons in C3 were compared to those of C4 and α_2-macroglobulin (α_2M). Thirty-nine of 41 exons in C4 were found to be of similar size to the analogous ones in C3, and two-thirds of those in α_2M were also similarly sized, supporting the hypothesis that these genes arose from a common ancestor.

The complement system is a group of plasma proteins involved in the humoral immune response. Upon activation by either of two pathways, termed the classical and alternative pathways, a cascade of proteolytic cleavage events takes place, the products of which mediate reactions involved in inflammation, opsonization of foreign particles, and the killing of foreign cells (Müller-Eberhard & Miescher, 1985). The third component, C3,[1] is of major importance in this system. It is involved in both pathways, is present in rather high concentrations in the plasma (1.3 mg/mL), and undergoes several proteolytic cleavage events, yielding a number of fragments that have vasoactive, opsonic, or inflammatory properties.

C3 is synthesized as a single-chain precursor of 185 kDa whose full-length cDNA has been cloned (de Bruijn & Fey, 1985). After the signal peptide is cleaved, a four amino acid sequence is excised from the C3 polypeptide, producing a molecule comprising an α- and a β-chain. Upon activation, the α-chain is further cleaved by C3 convertases into C3a, a 77 amino acid peptide with anaphylatoxic and inflammatory properties, and C3b, in which the remainder of the α-chain (α') remains covalently bound to the β-chain. The α'-chain of C3b can undergo further cleavage by complement factor I to yield the fragments iC3b and C3dg. One of the unusual features of the C3 molecule is the presence of an internal thiolester formed between Cys^{1010} and Gln^{1013} (Tack et al., 1980), a site also shared with the complement protein C4 (Campbell et al., 1981; Harrison et al., 1981) and the protease inhibitor α_2-macroglobulin (α_2M) (Howard et al., 1980; Sottrup-Jensen et al., 1981). When the α-chain of C3 is cleaved to form C3a and C3b, the thiolester bond is broken, producing a reactive carbonyl group capable of forming a covalent bond with an amino or a hydroxyl moiety (Law et al., 1981; Hostetter et al., 1982). The attachment of C3b to the surface of a particle or immune complex (Law & Levine, 1977) allows it to interact with complement receptor type 1 (CR1). As the C3b is cleaved first to iC3b and then to C3dg, it is able to interact with turn with CR3 and then with CR2 (Fearon & Ahearn, 1989) and CR5 (Vik & Fearon, 1985).

The major site of synthesis of C3 is the liver (Alper & Rosen, 1976), and the levels of C3 increase approximately 50% in the acute phase. Peripheral blood monocytes and the cell line U937, a histiocytic leukemia, also synthesize C3, and these cells, along with liver cell lines, have been used to study the effect on C3 levels of various stimulators. C3 synthesis increases substantially in response to lipopolysaccharide, IL-1, tumor necrosis factor, glucocorticoids, and possibly IL-6, whereas IFN tends to diminish the levels of C3 (Barnum et al., 1989b). C3 levels are also increased in the serum of male mice (Churchill et al., 1967) and in the uterus of rats in response to estradiol (Sundstrom et al., 1990).

Preliminary structure of the gene encoding C3 have been reported (Barnum et al., 1989a; Botto et al., 1989). We wished to extend these studies to obtain the complete genomic organization of this molecule in order to gain insight into structure/function relationships of C3, to examine the 5′ regulatory region to determine if there are potential cis regulatory elements through which the various stimulatory factors may be operating, and to compare the C3 gene with the genes

† This work was supported by U.S. Public Health Service Award AI 19222 to B.F.T. and AI 25011 and Research Career Development Award AI 00919 to R.A.W.

* Address correspondence to this author.

‡ Scripps Clinic and Research Foundation.

§ Present address: Service de Pediatrie, Laboratoire de Neurochimie, CHUV, Lausanne, Switzerland.

‖ Washington University School of Medicine.

[1] Abbreviations: α_2M, α_2-macroglobulin; C3 (C4), third (fourth) component of complement; CR1, complement receptor type 1; IFN, interferon; IL, interleukin; TNF, tumor necrosis factor; PCR, polymerase chain reaction.

0006-2960/91/0430-1080$02.50/0

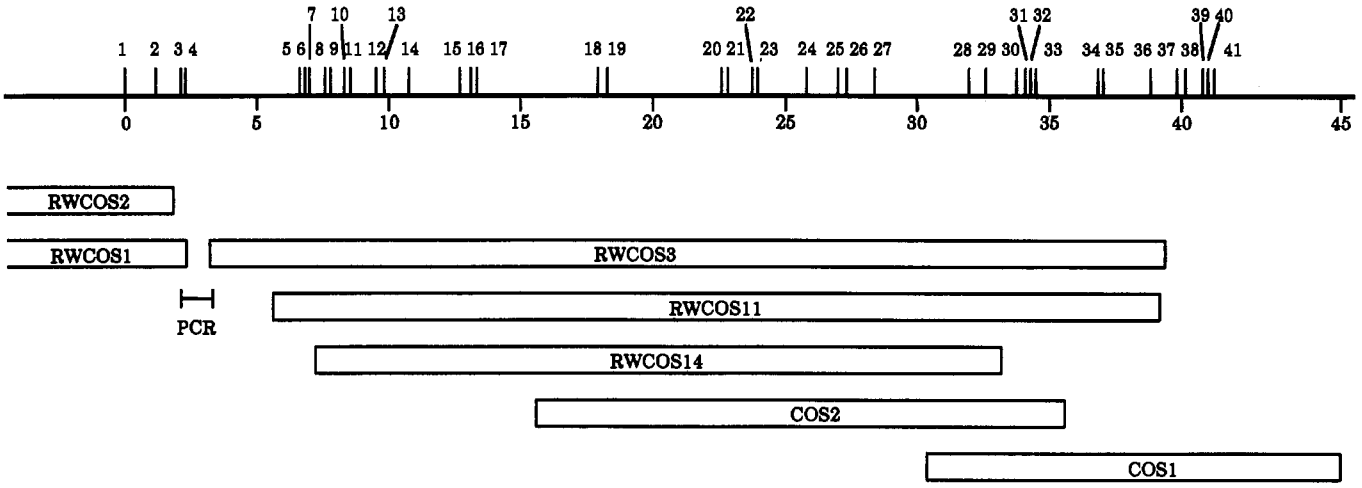

FIGURE 1: Map of the C3 gene. The positions of the 41 exons are indicated above the line. The positions of the seven cosmid clones and the region cloned with PCR are indicated below the line.

that code for C4 and α_2M to gain insight into the evolution of this family.

Experimental Procedures

Enzymes and Reagents. Restriction endonucleases and modifying enzymes were purchased from Promega Biotec, New England Biolabs, and Boehringer Mannheim Biochemicals. Radiolabeled nucleotides were purchased from Amersham. Nylon filters (MSI, Micron Separations Inc.) were purchased from Fisher Scientific.

Construction and Screening of the Cosmid Library. A cosmid library was constructed in the pTCF vector (Grosvelt et al., 1982) as previously described (Chaplin et al., 1983) with human genomic DNA. Filters were prehybridized and hybridized at 42 °C in 50% formamide buffer. The cosmid library was probed with ^{32}P-labeled C3 cDNA fragments (de Bruijn & Fey, 1985). Positive clones were colony-purified and cosmid DNA was isolated (Maniatis et al., 1982).

Sequencing of Intron/Exon Junctions. Cosmid DNA was sequenced according to the method of Bankier and Barrell (1983). Fragments of 300–600 bp in length, randomly generated by sonication, were ligated into the *Sma*I site of M13mp8 after end-repairing with T4 DNA polymerase. Exon-containing phage were identified by transferring a portion of the phase to MSI nylon filters, in duplicate, and probing the filters with the appropriate ^{32}P-labeled C3 cDNA fragments (de Bruijn & Fey, 1985). The fragments were labeled by the random-prime method (Feinberg & Vogelstein, 1983) with a kit from Boehringer Mannheim. Filters were prehybridized and hybridized at 42 °C in 50% formamide buffer. Phage that screened positive were then sequenced by the dideoxy chain-termination method (Sanger et al., 1977) with α-^{35}S-labeled deoxyadenosine 5′-thiotriphosphate (Biggin et al., 1983).

Southern Blot Analysis. A 1-μg aliquot of cosmid DNA was digested to completion with the appropriate restriction endonuclease and electrophoresed in a 0.8% agarose gel. The DNA was transferred to nylon filter paper, and the filters were prehybridized and then hybridized in 50% formamide buffer at 42 °C. The probes were ^{32}P-labeled M13 subclones derived from sequencing the intron/exon junctions. The phage was labeled by using standard sequencing reaction conditions (Bankier & Barrell, 1983) in the presence of [α-^{32}P]dCTP and the absence of dideoxynucleotides. After overnight hybridization, the filters were washed (0.2× SSC and 0.1% SDS at 65 °C for 30 min) and exposed to X-ray film with an intensifying screen for the appropriate time at -70 °C.

Primer Extension Analysis. Total RNA was isolated from HepG2 cells (a human hepatoblastoma cell line) by the guanidinium isothiocyanate extraction procedure (Chirgwin et al., 1979). A 30-mer oligonucleotide, 5′-CAGCAGGCTGGGACCTGAGGTGGGTCCCAT-3′, complementary to nucleotides 61–90 of the C3 cDNA sequence (de Bruijn & Fey, 1985) was end-labeled, by using [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase (New England Biolabs), to a specific activity of 4×10^7 cpm/μg. The labeled oligonucleotide (5 ng) and 25 μg of HepG2 total RNA were coprecipitated, resuspended in 30 μL of hybridization buffer (1 M NaCl, 0.1 M PIPES, pH 6.4, and 25 mM EDTA, pH 8.0), and annealed at varying temperatures (37, 42, 48, and 55 °C) for 16 h.

The hybrids were resuspended in 45 μL of extension buffer (80 mM Tris-HCl, pH 8.3, 10 mM $MgCl_2$, 0.5 mM DTT, 1.33 mM dNTPs, and 40 units of RNasin), and the primer was extended by using 50 units of AMV reverse transcriptase (Boehringer Mannheim) at 42 °C for 2 h. Following RNase A digestion and phenol/chloroform extraction, the extended products were analyzed on a 7% denaturing polyacrylamide gel with a sequenced M13 DNA template from the same region as a size marker.

Polymerase Chain Reaction. Two oligonucleotides were synthesized a 20-mer based on exon 3 sequence and containing the *Kpn*I recognition site and a 28-mer complementary to the sequence in intron 4 at the beginning of the RWCOS3 clone and creating a *Bam*HI site. An aliquot (50 or 100 pmol) of each primer was incubated with 0.25 or 1 μg of human genomic DNA, and the reaction was carried out for 30 cycles by using the GeneAmp DNA amplification kit and thermal cycler (Perkin Elmer Cetus, Norwalk, CT). The resulting band was ligated into the *Kpn*I and *Bam*HI sites of pUC19.

Results and Discussion

Human C3 cDNA was used to screen a genomic library, and five clones (RWCOS1-3, 11 and 14) were isolated in addition to two clones (COS1 and 2) previously described (Barnum et al., 1989a). These clones (see Figure 1) were used to determine the intron/exon junction sequences and to map the positions of the exons within the gene. In addition to the 24 exons previously described in the region coding for the α'-chain, 16 more exons were completely identified, including exons 1–3 and 5–17. RWCOS1 was found to contain the beginning of exon 4, but the clone ended in the middle of this

A

```
5050   TGACCACACC CCCATTCCCC CACTCCAGAT AAAGCTTCAG TTATATCTCa
       cgtgtctgga gttctttgcc aagagggaga ggctgaaatc cccagccgcc
       tcacctgcag ctcagctcca tcctacttga aacctcacct gttcccaccg
       cattttctcc tggttcgcct gctagtgtct gacttcttta gccaaggagc
       atggacctgc ctcacctgca cgtggcatgc ct
```

B

```
GATCAATATGAATATATTATACACACAGACACACACACAGACACACACACACACACAC  -817
ACAAACAATACAATTTAATATCCTAAGAGGATATTGACATTAGACAGGTACAAAAGCTCT  -757
AGAAATGAGGACTTTCCTCAGTGATGACTTTTTTCACCACCAAAGTCACTCAGGCATCCT  -697
GACAAGGGTAAGTGAGGGGAGCCTCCTTGGAAAATAAACTCACTTGGATAGTGAACTCCT  -637
GCACATACCTCAAAGCCCATCTGAAATGTCCCCTCCTACAGGAAGTTTTCCCTGACCCTC  -577
CAAGAAGCAGAGTTCTATTTCACTGGGGAAAACATTTCTTCTTCTTCTTTTTTTCCCTG  -517
CCCTGCACATGAGCTAGAAAACATTTCATGAAACTGGGAGTTTCTGTGCTGGGCTCTGTC  -457
CCTCCCCCATTCTACTTCCCCTCCCTCAGCATGGAAGCCTCTGGAAGTGGGGCTCTGACT  -397
CCCAGCCTACAGAGAGATTCCTAGGAAGTGTTCGACTGATAAACGCATGGCCAAAAGTGA  -337
ACTGGGGATGAGGTCCAAGACATCTGCGGTGGGGGGTTCTCCAGACCTTAGTGTTCTTCC  -277
ACTACAAAGTGGGTCCAACAGAGAAAGGTCTGTGTTCACCAGGTGGCCCTGACCCTGGGA  -217
GAGTCCAGGGCAGGGTGCAGCTGCATTCATGCTGCTGGGGAACATGCCCTCAGGTTACTC  -157
ACCCCATGGACATGTTGGCCCCAGGGACTGAAAAGCTTAGGAAATGGTATTGAGAAATCT  -97
GGGGCAGCCCCAAAAGGGGAGAGGCCATGGGGAGAAGGGGGGGCTGAGTGGGGGAAAGCA  -37
                                    ↓ ↓
GAGCCAGATAAAAAGCCAGCTCCAGCAGGCGCTGCTCACTCCTCCCCATCCTCTCCCTC   24
                                   +1
                                        M  G  P  T  S  L  L
TGTCCCTCTGTCCCTCTGACCCTGCACTGTCCCAGCACCATGGGACCCACCTCAGGTCCC   84

 L  L  L  L  L  L  L  T  H  L  P  L  A  L  G  S  P  M
AGCCTGCTGCTCCTGCTACTAACCCACCTCCCCCTGGCTCTGGGGAGTCCCATgtgagt
```

FIGURE 2: (A) Sequence of the 3′ region of the C3 gene. Sequence obtained from the cDNA clone is capitalized and that obtained from the genomic clone is in lowercase letters. The putative polyadenylation signal is underlined. (B) Sequence of the 5′ region. The two transcription start sites are indicated with arrows, and the 5′ one is designated +1. Putative CAAT and TATA boxes are underlined. The amino acid sequence of the first exon is shown, and the beginning of intron 1 is in lowercase letters.

exon. Approximately 3 kb of DNA at the 5′ end of RWCOS3 was fully sequenced and did not contain the fourth exon. On the basis of Southern blot analysis of this region using genomic DNA, the missing region was calculated to be approximately 1 kb. Consequently, the polymerase chain reaction was performed on genomic DNA, using oligonucleotides based on the sequence at the 3′ end of RWCOS1 and the 5′ end of RWCOS3. The resulting fragment of 1050 bp was cloned into pUC19 and sequenced to identify the 3′ end of the exon 4. Thus, the complete C3 gene is composed of 41 exons that range in size from 52 to 213 bp (Table I). Some of the introns were completely sequenced, and the size of the remaining ones were estimated on the basis of mapping the clones with various exon probes. The introns ranged in size from 82 bp to 4.4 kb, and the 5′ and 3′ splice donor and acceptor sequences followed the usual GT/AG pattern with the exception of intron 29, which begins with a GC instead of a GT, as previously noted (Barnum et al., 1989a). The exon sequences were identical with the published cDNA sequence except at three positions. The G's at positions 972 and 1896 of the cDNA were both A's in clone RWCOS14 but G's in clone RWCOS11, and thus they may represent allelic differences. Neither one of these changes would result in an amino acid substitution. The T at position 1001 of the cDNA was a C in both clones RWCOS11 and RWCOS14, which would lead to the substitution of proline for leucine.

The 5′ and 3′ regions of the gene were examined in more detail. The original cDNA clone used to identify the 3′ end

Table I: Structure of Intron–Exon Junctions

exon	location in cDNA[a]	length (bp)	splice acceptor	splice donor	intron size (bp)[b]	exon	location in cDNA[a]	length (bp)	splice acceptor	splice donor	intron size (bp)[b]
1	1–134	134		CAT/gtgagt	(1100)	22	2857–2923	67	gcgtag/CCG	GTG/gtgagt	126
2	135–327	193	cgacag/GTA	ATG/gtgagt	(760)	23	2924–3010	88	ctgcag/AAG	AAG/gtgaga	(1750)
3	328–493	166	gcctag/ATC	CAG/gtgagg	82	24	3011–3214	204	ctgcag/GGA	AGG/gtgggc	(1000)
4	494–564	71	cccag/TTC	GAG/gtgcag	(3500)	25	3215–3290	76	ctgcag/GGT	CTG/gtgagt	(250)
5	565–659	95	ccacag/AAC	CAA/gtatgt	102	26	3291–3450	160	ctccag/GCT	ATT/gtaaga	(900)
6	660–742	83	tcacag/CAT	ACG/gtaaga	83	27	3451–3549	99	tctcag/GGT	ACC/gtaagt	(3500)
7	743–833	91	cccag/TGC	CAG/gtgagg	(500)	28	3550–3706	157	tcatag/AGC	AAG/gtgagg	(470)
8	834–936	103	ccgtag/GTT	CCG/gtacca	91	29	3707–3870	164	ttgcag/ATA	CAG/gcaagt	(1000)
9	937–1063	127	tctcag/ATT	CAG/gtgagg	437	30	3871–4029	159	ttctag/GCC	GAG/gtacag	155
10	1064–1179	116	cctcag/GCA	ATG/gtgaga	98	31	4030–4089	60	ctgcag/ACC	TCG/gtaagg	124
11	1180–1329	150	tcacag/GTG	ACG/gtgcgt	(780)	32	4090–4180	91	taccag/GTG	CAG/gtaaaa	138
12	1330–1539	210	acccag/GTG	CTG/gtccgt	141	33	4181–4232	52	ttctag/AAA	CAG/gtaaga	(2300)
13	1540–1746	207	cccag/ATC	TCG/gtaagt	(700)	34	4233–4320	88	atccag/GTA	CAG/gtatga	111
14	1747–1905	159	ccacag/CTG	AAG/gtaagg	(1800)	35	4321–4410	90	ccccag/CTG	AAG/gtaagg	(1700)
15	1906–2035	130	ccgcag/ATC	CAG/gtgagg	262	36	4411–4516	106	cctcag/GTC	TGG/gtgagc	(900)
16	2036–2107	72	ctgcag/AAC	AAG/gtggga	192	37	4517–4606	90	ccgcag/AGG	AGG/gtgagt	196
17	2108–2305	198	ccgcag/TCG	GGA/gtaggt	(4400)	28	4607–4690	84	ctgcag/AGA	ATG/gtgagt	(600)
18	2306–2414	109	ctgcag/GTA	TGG/gtaagg	242	39	4691–4774	84	ccacag/TGT	CAG/gtcagg	85
19	2415–2501	87	atccag/AAT	AGG/gtgaga	(4200)	40	4775–4910	136	ctccag/GCT	CAA/gtgagt	128
20	2502–2643	142	gagcag/GAT	AAG/gtgggt	94	41	4911–5098	188	ctccag/CCT		
21	2644–2856	213	tggcag/GTG	GTG/gtgagt	(700)						

[a] Numbering as per de Bruijn and Fey (1985). [b] Sizes in parentheses are estimates based on Southern blotting.

of the C3 sequence contained 17 bp after the stop codon but did not have a poly(A) tail (de Bruijn & Fey, 1985), so the full length of the 3′ untranslated region was unknown. We had previously used a thiolester-specific probe to isolate a C3 cDNA clone that contained a poly(A) tail, and this region was sequenced and compared to the genomic sequence obtained from COS1 (Figure 2A). Starting with the TGA stop codon at position 5050 of the cDNA (de Bruijn & Fey, 1985), the two sequences were identical for 49 bp at which point the cDNA clone had a poly(A) tail. Inspection of the region upstream of this tail revealed the putative polyadenylation signal GATAAA 21 bp 5′ to the poly(A) tail, and there were no other potential signals observed a further 200 bp downstream. Thus, the 3′ untranslated region of the human C3 mRNA is rather short, only 46 bp. A similarly short region has been found for murine C3 where the poly(A) tail was found to be 39 bp downstream of the TGA stop codon (Wetsel et al., 1984).

Approximately 1 kb of DNA upstream of the ATG start codon was also sequenced (Figure 2B). In order to determine the transcription initiation site of C3, primer extension experiments were performed by using RNA from the HepG2 cell line and an oligonucleotide complementary to the C3 mRNA (Figure 3). Two bands were observed at positions corresponding to the A and T residues 61 and 63 nucleotides 5′ to the ATG start codon (Figure 2B), and the T at position 63 was designated +1. This T is 3 bp upstream of the published sequence of the cDNA clone (deBruijn & Fey, 1985). Possible TATA and CAAT boxes were observed at positions -28 and -87, respectively. Many factors such as interleukin-1 (Perlmutter et al., 1986), interferon-γ (Hamilton et al., 1987) and estrogen (Sundstrom et al., 1990) have been shown to affect C3 gene expression. Consequently, the upstream sequence was inspected for homologies with known cis-acting regulatory sequences. A number of potential enhancer sequence elements were identified and they are listed in Table II. These include sequences with 100% homology with the two types of IFN-γ responsive elements (Basta et al., 1987; Taylor & Grossberg, 1990), 12 regions of 100% homology with the IL-6 responsive element (Li et al., 1990), two regions that share 85% homology with the estrogen responsive element (Klein-Hitpass et al., 1986), two sequences with 75–88% homology with the glucocorticoid (core) responsive element

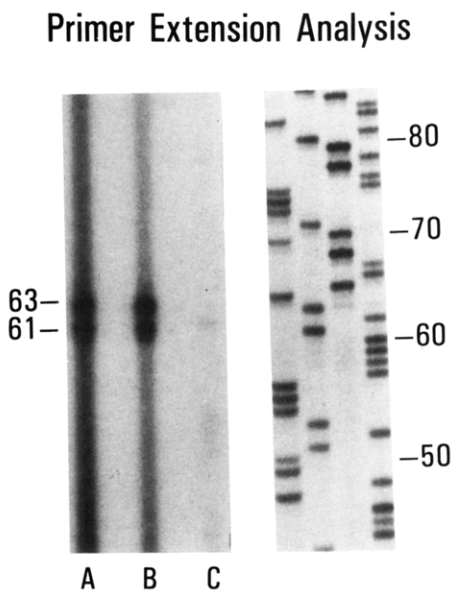

FIGURE 3: Primer extension analysis of C3 mRNA. A 30-mer complementary to positions 61–90 of the cDNA sequence was annealed to 25 μg of RNA from HepG2 cells at 48 °C (lane A) or 55 °C (lane B) or to tRNA at 37 °C (lane C), and a second strand was synthesized by using reverse transcriptase. M13 sequence was run as a size marker and is shown on the right with the positions marked.

(Karin et al., 1984; Payvar et al., 1983; Scheidereit et al., 1983), two sequences that share 82–88% homology with the thyroid hormone responsive element (Glass et al., 1988) and three sequences that are 83–94% homologous with the NF-κB responsive element (Sen & Baltimore, 1986). NF-κB is the trans-acting factor that has been shown to respond to a number of different stimuli, including IL-1 and TNF-α. Although some of these agents have been shown to affect C3 synthesis of mRNA levels, which, if any, of these elements are functional in vivo has not been determined.

The structure of the human C3 gene was compared to that of the gene structures of murine C4 (Ogata et al., 1989) and rat α_2M (Hattori et al., 1989) (Figure 4). C4, whose mRNA is 5372 bp in length, is also composed of 41 exons, whereas α_2M has 36 exons that comprise its message of 4598 bp. There is no correlation between intron sizes in any of the genes, which is evident when one compares the sizes of the genes: 42 kb

Table II: Homologies between Sequences in the 5′ Flanking Region of the C3 Gene and Known Regulatory Sequences[a]

IFNγ responsive element	GAAAN----------NGAAA	IL-6 responsive element	-106 TGAGAAA 100
	-474 GAAACTCCCAGTTTCATGAAA -493		-48 TGGGGGA -42
	-127 GAAAAGC------TTAGGAAA -113		
	-116 GAAATGGT----ATTGAGAAA -100	NF-κB responsive element	GGGGAC-TTTCC
			-750 GaGGAC-TTTCC -74-
IFN-γ responsive element (class II)	AGAAGBDAG		-604 GGGGACaTTTCa -615
	-574 AGAAG-CAG -567		+136 tGGGAC-TccCC +126
	-535 AGAAGAAAt -543		
		Estrogen responsive element	GGTCANNNTGACC
IL-6 responsive element	TGRRRRA		-234 GGTggCCCTGACC -223
	-737 TGAGGAA -743		+33 tGTCcCTCTGACC +45
	-721 TGAAAAA -728		
	-669 TGGAAAA -663	Glucocorticoid responsive element	AGABCAGB
	-553 TGGGGAA -547		-280 AGAACAcT -287
	-448 TGGGGGA -454		-262 ccAACAGA -255
	-429 TGAGGGA -435		
	-294 TGGAGAA -300	Thyroid hormone responsive element	TCAGGTNNNNNACCTGA
	-276 TGGAAGA -282		-167 TCAGGTTACTCACCccA -151
	-221 TGGGAGA -215		+76 TCAGGTCCCA-gCCTGc +91
	-181 TGGGGAA -175		

[a] N, any base; B, A or T; D, A or C; R, A or G; -, gap introduce to maximize homology; lowercase letters designate nonhomologous bases.

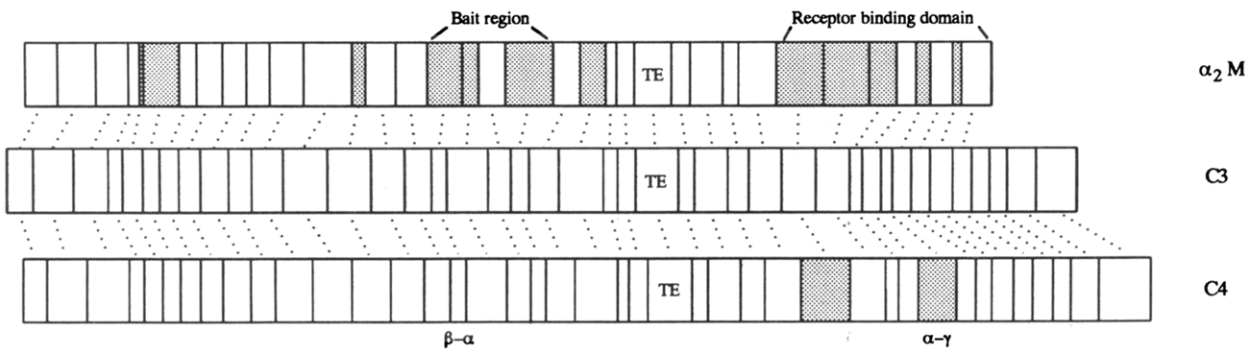

FIGURE 4: A comparison of exon sizes between C3, C4, and α_2M. The locations of the intron/exon boundaries are marked, and analogous exons are connected with dotted lines. The bait region and receptor-binding domain of α_2M and the thiolester (TE) region and junctions of the α- and β-chains and α- and γ-chains are designated. Exons varying in size by more than 30% with C3 exons are shaded.

for C3, 17 kb for C4, and 50 kb for α_2M. However, a similarity emerges when one compares the sizes of the exons in C3 to those of C4 and α_2M. The correlation between the sizes of exons in C3 and C4 was striking, with only 2 exons differing in size by more than 30% between the two sequences. One of these exons, number 33, codes for the junction region between the α- and γ-chains of C4, a region that is not present in C3.

The similarity of C3 with α_2M is less striking but still considerable, with two-thirds of the exons being of comparable size between the two sequences. One area where the two molecules differed significantly was immediately 5′ to the thiolester region. This is the portion of α_2M that codes for the bait region (Gehring et al., 1987), a specialized domain whose function is to bind various proteinases, leading to their inactivation. The 3′ region of the two molecules also showed a substantial degree of variation. This region in α_2M codes for the receptor-binding domain (Sottrup-Jensen et al., 1986), which allows for rapid clearance of the proteinase–α_2M complexes via receptor-mediated endocytosis. Neither of these functions exists in the C3 molecule, and thus it is not surprising that the two sequences have diverged in these regions. Nevertheless, despite the fact that these three genes are from three different species, it is evident that they are related. Although the primary sequence of the molecules may have changed and the locations of the exons within the genes appear to have shifted, many of the locations of the exon boundaries have remained the same. This is especially notable for C3 and C4. The areas that have changed the most are associated with specialized domains of the molecule.

Added in Proof

Upon resequencing the region upstream of the transcription start site, we found the Gs at positions −39 and −36 to be doublet GGs. After submission of our manuscript, Fong et al. (1990) published an article with results similar to ours.

Acknowledgments

This is Publication No. 6489-IMM from the Research Institute of Scripps Clinic. We thank Dr. Scott Barnum and Brenda Aguas for sequencing the 3′ end of the C3 cDNA, Lori Hicks-Graham for excellent technical assistance, Dr. Georg Fey for providing the C3 cDNA clones, and Bonnie Towle for excellent secretarial assistance.

Registry No. Complement C3, 80295-41-6.

REFERENCES

Alper, C. A., & Rosen, F. S. (1976) *Adv. Hum. Genet. 7*, 141–181.

Bankier, A. T., & Barrell, B. G. (1983) in *Techniques in Nucleic Acid Biochemistry* (Flavell, R. A., Ed.) Vol. 85-08, pp 1–34, Elsevier/North-Holland Scientific, Limerick, Ireland.

Barnum, S. R., Amiguet P., Amiguet-Barras, F., Fey, G., & Tack, B. F. (1989a) *J. Biol. Chem. 264*, 8471–8474.

Barnum, S. R., Fey, G., & Tack, B. F. (1989b) *Curr. Top. Microbiol. Immunol. 153*, 23–43.

Basta, P. V., Sherman, P. A., & Ting, J. P.-Y. (1987) *J. Immunol. 138*, 1275–1280.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Botto, M., Fong, K. Y., So, A. K., Poznansky, M., & Walport, M. J. (1989) *Complement Inflammation 6*, 315–316.

Campbell, R. D., Gagnon, J., & Porter R. R. (1981) *Biochem. J. 199*, 359–370.

Chaplin, D. D., Woods, D. E., Whitehead, A. S., Goldberger, G., Colten, H. S., & Seidman, J. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6947–6951.

Chirgwin, J. M., Prsbyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Churchill, W. H., Jr., Weintraub, R. M., Borsos, T., & Rapp, H. P. (1967) *J. Exp. Med. 125*, 657–672.

de Bruijn, M. H. L., & Fey, G. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 708–712.

Fearon, D. T., & Ahearn, J. M. (1989) *Curr. Top. Microbiol. Immunol. 153*, 83–98.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Fong et al. (1990) *Genomics 7*, 579–586.

Gehring, M. R., Shiels, B. R., Northemann, W., de Bruijn, M. H. L., Kan, C.-C., Chain, A. C., Noonan, D. J., & Fey, G. H. (1987) *J. Biol. Chem. 262*, 446–454.

Glass, C. K., Holloway, J. M., Devary, O. V., & Rosenfeld, M. G. (1988) *Cell 54*, 313–323.

Grosvelt, F. G., Lund. T., Murray, E. J., Mellor, A. L., Dahl, H. M., & Flavell, R. A. (1982) *Nucleic Acids Res. 10*, 6715–6732.

Hamilton, A. O., Jones L., Morrison, L., & Whaley, K. (1987) *Biochem. J. 242*, 809–815.

Harrison, R. A., Thomas, M. L., & Tack, B. F. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7388–7392.

Hattori, M., Kusakabe, S.-I., Ohgusu, H., Tsuchiya, Y., Ito, T., & Sakaki, Y. (1989) *Gene 77*, 333–340.

Hostetter, M. K., Thomas, M. L., Rosen, F. S., & Tack, B. F. (1982) *Nature 298*, 72–75.

Howard, J. B., Vermeulen, M., & Swenson, R. P. (1980) *J. Biol. Chem. 255*, 3820-3823.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westpahl, H. M., & Beato, M. (1984) *Nature 308*, 513–519.

Klein-Hitpass, L., Schorpp, M., Wagner, V., & Ryffel, G. V. (1986) *Cell 46*, 1053–1061.

Law, S. K., & Levine, R. P. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2701–2705.

Law, S. K., Minich, T. M., & Levine, R. P. (1981) *Biochemistry 20*, 7457–7463.

Li, S.-P., Liu, T.-Y., & Goldman, N. D. (1990) *J. Biol. Chem. 265*, 4136–4142.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Müller-Eberhard, H., & Miescher, P. A. (1985) in *Complement* (Müller-Eberhard, H., Ed.) Springer, Berlin.

Ogata, R., Rosa, P. A., & Zepf, N. E. (1989) *J. Biol. Chem. 264*, 16565–16572.

Payvar, F., DeFranco, D., Firestone, G. L. , Edgar B., Wrange, Ö., Okret, S., Gustafsson, J.-A., & Yamamoto, K. R. (1983) *Cell 35*, 381–392.

Perlmutter, D. H., Goldberger, G., Dinarello, C. A., Mizel, S. B., & Colten, H. R. (1986) *Science 232*, 850–852.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Scheidereit, C., Geisse, S., Westphal, H. M., & Beaton, M. (1983) *Nature 304*, 749–752.

Sen, R., & Baltimore, D. (1986) *Cell 46*, 705–716.

Sottrup-Jensen, L., Hansen, H. F., Mortensen, S. B., Peterson, T. E., & Magnusson, S. (1981) *FEBS Lett. 123*, 145–148.

Sottrup-Jensen, L., Gliemann, J., & Van Leuven, F. (1986) *FEBS Lett. 205*, 20–24.

Sundstrom, S. A., Komm, B. S., Xu, Q., Boundy, V., & Lyttle, C. R. (1990) *Endocrinology 126*, 1449–1456.

Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M., & Prahl, J. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5764–5768.

Taylor, J. L., & Grossberg, S. F. (1990) *Virus Res. 15*, 1–26.

Vik, D. P., & Fearon, D. T. (1985) *J. Immunol. 134*, 2571–2579.

Wetsel, R. A., Lundwall, Å., Davidson, F., Gibson, T., Tack, B. F., & Fey, G. H. (1984) *J. Biol. Chem. 259*, 13857–13862.