Complement C1r and C1s genes are duplicated in the mouse: differential expression generates alternative isomorphs in the liver and in the male reproductive system

Gérard GARNIER*, Antonella CIRCOLO†, Yuanyuan XU† and John E. VOLANAKIS*†[1]

*Biomedical Sciences Research Center 'A. Fleming', 34, Al. Fleming St., Vari 16672, Greece, and †Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, U.S.A.

C1r and C1s are the serine proteases that form the catalytic unit of the C1 complex, the first component of complement. In the present study, we found that the genes encoding murine C1r and C1s are duplicated. One set of these genes, referred to as *c1rA* and *c1sA*, are primarily expressed in the liver and are therefore the homologues of the human C1r and C1s genes. The other two genes, termed *c1rB* and *c1sB*, are expressed exclusively in male reproductive tissues, specifically the coagulating gland and the prostate. The predicted C1rB and C1sB proteins share 96 and 93% amino acid identity with C1rA and C1sA respectively. Most of the substitutions are clustered in the serine protease domains, suggesting differences in catalytic efficiencies and/or substrate specificities or alternatively adaptation to different physiological environments. The high homology of C1rB and C1sB with C1rA and C1sA in the non-catalytic regions indicates that they are probably capable of assembling the C1 complex. The expression of alternative genes encoding isomorphs of activating components of complement in male reproductive tissues raises the possibility of new mechanisms of complement activation in the male genital tract or of novel functions for complement proteases in reproduction.

Key words: coagulating gland, mouse chromosome 6, prostate, serine protease.

INTRODUCTION

C1r and C1s are the enzymes initiating the activation of the classical pathway of complement, a major recognition and effector humoral system of innate and adaptive immunity [1]. The Ca^{2+}-dependent C1s-C1r-C1r-C1s tetramer combines with the non-catalytic subcomponent C1q to form the C1 complex. C1q itself consists of six heterotrimeric collagen-like helices, which terminate in globular domains. It is generally accepted that classical pathway activation is initiated by conformational changes in the C1 complex, triggered by the interaction of the globular domains of C1q with immune complexes or pathogens, leading to autocatalytic activation of C1r. Subsequent activation of C1s by C1r allows cleavage of C4 and C2 and the assembly of the proteolytic complex C4b2a, the classical pathway C3 convertase.

Both the structure and functions of human C1r and C1s have been studied quite extensively. They are serine proteases (SPs) sharing high-sequence homology and identical modular organizations [1]: from N- to C-terminus, two CUB [(first identified in) C1r/C1s, sea urchin epidermal growth factor (uEGF) and human bone morphogenic protein-1] modules flanking an EGF-Ca^{2+} module, are followed by two complement control protein (CCP) modules, a short connecting segment (CS) and an SP domain of the chymotrypsin fold type. Proteolytic activation of C1r and C1s is achieved by cleavage of single Arg–Ile bonds between the CS and the SP domain. The assignment of catalytic functions and inter- and intra-molecular interactions within the C1 complex to structural correlates has been the focus of studies for decades [1], illustrated by a recent model of the C1 activation mechanism [2].

C1r and C1s are serum glycoproteins primarily produced by the liver [3]. Synthesis and secretion of the functional proteins by human hepatoma cells [4] and guinea-pig hepatocyte primary cultures [5] have been reported. In addition, their expression was demonstrated in monocytes, fibroblasts, endothelial and epithelial cells [3], myoblasts [6] and glioma cells [7].

Human C1r and C1s are encoded by single genes located in chromosome 12 region p13 [8], at a distance of 9.5 kb from each other, and transcriptionally orientated tail to tail [9]. The C1s gene comprises 12 exons spanning 13 kb of DNA [10,11]. The C1r gene has been characterized only partially [10]. Genetic deficiencies of C1r and C1s are rare and often associated with systemic lupus erythematosus-like syndromes [12]. In the past few years, complement-related pathologies and biological functions of individual components, receptors and regulators have been intensively studied *in vivo* using gene-targeting methodologies. However, C1r and C1s structure, function and genetics are poorly documented in animal models. The murine C1r cDNA [13] and the rat [14] and hamster [15] C1s cDNAs have been cloned and sequenced, and in the mouse, both genes were mapped to the telomeric end of chromosome 6 [16], but the C1r and C1s genes have not yet been characterized in animals and little is known about their expression *in vivo* [9,17].

Abbreviations used: CCP, complement control protein; CS, connecting segment; CUB, C1r/C1s, sea urchin epidermal growth factor (uEGF) and human bone morphogenetic protein-1; DAF, decay accelerating factor; EST, expressed sequence tag; LPS, lipopolysaccharide; MCP, membrane cofactor protein; ORF, open reading frame; RACE, rapid amplification of cDNA ends; RT, reverse transcriptase; SP, serine protease; UTR, untranslated region.

[1] To whom correspondence should be addressed (e-mail volanaki@uab.edu).

The mammalian male genital tract and spermatozoa express many complement regulatory proteins [18], including decay accelerating factor (DAF, CD55), membrane cofactor protein (MCP, CD46), CD59 and C4b-binding protein [19]. Most of them derive from other genes, promoters, mRNA splice forms or glycosylation processes rather than the isoforms in other tissues and fluids. These regulatory proteins are thought to play an important role in protecting the sperm from complement attack in the female tract and evidence has been presented for the direct involvement of some of them in reproductive functions [20]. In the seminal plasma, the glycosylphosphatidylinositol-anchored proteins DAF and CD59 are present both on extracellular organelles (prostasomes) and as membrane-free forms. These and the prostasome-bound transmembrane protein MCP are capable of migrating to cell membranes [21], thus increasing the resistance of heterologous cells to complement attack. In contrast with the regulators, no complement component isoforms or alternative complement activation mechanism has been described so far in the mammalian reproductive system.

In the present study, we report that the C1r and C1s genes are duplicated in the mouse: one set (*c1rA* and *c1sA*) is the murine homologue of the human genes, whereas the other set (*c1rB* and *c1sB*) is expressed exclusively in two male reproductive accessory glands. This is the first report of alternative genes encoding isomorphs of complement-activating components expressed in the male genital tract.

EXPERIMENTAL

Cloning, mapping and sequencing the murine C1r and C1s genes

A genomic clone, bac185, was isolated from a mouse 129 library of partially *Hin*dIII-digested DNA in the pBeloBAC 11 vector (BAC Mouse ES-129/SvJ library) by Genome Systems Inc. (St. Louis, MO, U.S.A.) using a murine C1s gene-specific PCR assay. The primer sequences (forward: 5′-CTAGGGGAGTATCGATGTGCTG-3′; reverse: 5′-GATGTGACAGCAACTGTTGATG-3′; 305 bp amplicon), located in regions corresponding to exon 11 and intron 11 of the human C1s gene (GenBank® contig accession no. NT_009731) [10] were designed on sequences of preliminary clones. These clones were generated by PCR using 129/Sv genomic DNA as template and the murine liver C1s cDNA sequence for primer design. All oligonucleotides were supplied by MWG AG Biotech (Edersberg, Germany).

The bac185 clone was mapped by Southern-blot analysis [22], using genomic probes and six restriction enzymes (*Hin*dIII, *Xho*I, *Kpn*I, *Sma*I, *Cla*I and *Sac*I) (Promega, Madison, WI, U.S.A.), alone and in all possible pair combinations. Additional mapping was performed with *Bam*HI and *Pst*I. All probes, except one, were generated by PCR amplification of bac185 or 129/Sv genomic DNA, using C1rA and C1sA cDNA sequence data (see below). For PCR design, putative exon–intron junctions were initially inferred from the human C1s gene structure [10,11]. The C1sA 5′-most probe was excised from the 7 kb *Hin*dIII subclone of BAC-185 spanning from intron 1 to exon 7. Probes were labelled with [^{32}P]dCTP (Amersham Pharmacia Biotech, Piscataway, NJ, U.S.A.), using the random primed DNA labelling kit (Boehringer Mannheim, Somerville, NJ, U.S.A.). Blotting was performed on nitrocellulose filters (Bio-Rad Laboratories, Hercules, CA, U.S.A.) and the images were visualized using a phosphor screen, the STORM 860 system and the ImageQuant software for Windows NT (Molecular Dynamics, Sunnivale, CA, U.S.A.).

Restriction fragments from bac185 encompassing the entire *c1sA* (from 5′ to 3′: 8 kb *Bam*HI–*Xho*I, 7 kb *Hin*dIII, 6.9 kb *Pst*I and 6.2 kb *Hin*dIII fragments) and *c1rA* (17.5 kb *Hin*dIII fragment) genes were subcloned into pBluescript II SK + (Stratagene, La Jolla, CA, U.S.A.) or pSP72 (Promega) plasmid vectors and sequenced on an ABI Prism 377 automated sequencer (Applied Biosystems, Foster City, CA, U.S.A.) and in part by MWG AG Biotech. The 5′ *c1rB* genomic sequence, including exons 1–3, was generated by PCR, whereas a 6.5 kb *Bam*HI fragment, encompassing the 3′-region, was a subclone of bac185 and it was partially sequenced from both ends.

cDNA cloning and sequencing

Reverse transcriptase (RT)–PCR was performed from total RNA by using the superscript preamplification system (Life Technologies, Gaithersburg, MD, U.S.A.) with oligo(dT) primers for first-strand DNA synthesis. The liver C1s cDNA (C1sA, Balb/c strain) was obtained as two overlapping RT–PCR clones spanning the entire open reading frame (ORF) ($-162/+669$ and $+439/+2623$; for all cDNA nucleotide numbering: $+1$ is the adenosine of the putative initiation codon). Primers were designed on the rat 5′-untranslated region (UTR) sequence (GenBank® accession no. D88250) [14] and two mouse expressed sequence tag (EST) sequences encompassing exons 4–6 (GenBank® accession no. AA798057) and exon 12 to poly(A) (GenBank® accession no. AI327365). The latter and a poly(A)-containing C1r EST (GenBank® accession no. AA871944) allowed the localization of the 3′-ends of the C1sA and C1rA genes respectively. The transcription start sites of both genes were determined by cap-specific rapid amplification of cDNA ends (RACE) of Balb/c liver RNA as described previously [23]. For RACE and RT–PCR clones, five and at least two colonies respectively, were entirely sequenced.

The C1rB cDNA was obtained from RNA of C57BL/6 male reproductive tissues (consisting of epididymis, vas deferens, prostate, coagulating glands and seminal vesicles) as three overlapping RT–PCR clones spanning the entire ORF ($-27/+359$, $+346/+1798$ and $+1484/+2139$). Primers were designed on C1rA cDNA, except for the C1rB-specific reverse primer for the largest clone, which was designed on C1rB genomic sequence data. All three clones were sequenced on two colonies each.

A C1sB cDNA clone extending to the 3′-UTR ($+1196/+2623$) was generated by RT–PCR of C57/BL6 male reproductive tissue RNA using primers designed on C1sA cDNA; two colonies were sequenced. An overlapping RT–PCR product, extending to the 5′-UTR ($-54/+1464$), was generated using a forward primer designed on C1sA and a reverse primer specific for C1sB, gel-purified and directly sequenced.

Mouse tissue mRNA analysis

RNA was extracted from tissues of (6–8)-week-old mice (The Jackson Laboratory, Bar Harbor, ME, U.S.A.) by guanidinium thiocyanate lysis and purified by CsCl gradient ultracentrifugation [24]. RNA (12 μg) from C57BL/6 tissues was subjected to Northern-blot analysis as described previously [25]. Equal loading and RNA quality were assessed by ethidium bromide staining. The C1r cDNA probe, spanning exon 1 to exon 9 ($-8/+1257$), was obtained from an RT–PCR product of Balb/c liver RNA, which was cloned in pBluescript SK+. The C1s cDNA probe, spanning exon 5 to 3′-UTR ($+421/+2640$, including the primer), was prepared from the C1sA cDNA clone as described above. C1qA and C1qC cDNA probes were prepared from previously reported cDNA clones [26]. For endotoxin-induced inflammation, 10 μg of lipopolysaccharide (LPS, *Escherichia coli* 0111:B; Sigma) was injected intraperitoneally 24 h before the mice were killed.

C1rA, C1rB, C1sA and C1sB mRNAs from C57BL/6 tissues were selectively analysed by isomorph-specific PCR of oligo(dT)-primed reverse transcripts. Aliquots (0.8 μl) of 20 μl reverse transcription reactions, programmed by 10 μg of total RNA, were used as templates in each 50 μl PCR. Common forward primers were used for C1rA and C1rB (5′-CTTCCGCTACATC-ACCAC-3′) and for C1sA and C1sB (5′-ATGGAACATGAA-GAAGGTG-3′). Reverse primers were specific for C1rA (5′-GCTAACTTATCTTCTGTTA-3′), C1rB (5′-GGTATCC-TATTTCCTGTTG-3′), C1sA (5′-TTGTCCTCACAGACAT-GGT-3′) and C1sB (5′-GCCAGTCTCAAAGCCGTGA-3′), predicting 663 bp (+1155/+1817), 660 bp (+1155/+1814), 307 bp (+1177/+1483) and 308 bp (+1177/+1484) products respectively. PCR mixes were subjected to 26 cycles of amplification. Reaction specificity was controlled by direct sequencing of the PCR products and by digestion at a polymorphic *Pvu*II site (at +1378) for C1r and C1s reactions respectively. In addition, all four reaction products (diluted to 10^{-6}) could only be reamplified if their specific reverse primers were used, even for 35 cycles. Preliminary cycling kinetics indicated that the reactions for C1rA and C1sA in the liver and for C1rB and C1sB in the coagulating gland reached a plateau between 24 and 26 cycles. Therefore for these two tissues, the corresponding specific mRNA input is slightly underestimated when compared with the other tissues.

Anti-murine C1s serum and Western-blot analysis

A 393 bp C1sA cDNA segment (+46/+438), encoding the first CUB and part of the EGF-Ca^{2+} module, was obtained by RT–PCR of Balb/c liver RNA and cloned into the *Bam*HI and *Hin*dIII sites of the PQE30 vector (Qiagen, Valencia, CA, U.S.A.). The construct was expressed in *E. coli* and the resulting His_6-tagged protein extracted from bacteria lysate and purified by Ni^{2+}-nitrilotriacetic acid metal-affinity chromatography using a QIAexpress system kit (Qiagen), in accordance with the manufacturer's recommendations. Rabbits were immunized with purified protein according to standard methods. The IgG fraction was purified from the rabbit serum by caprylic acid precipitation [27].

C57BL/6 tissues were Dounce-homogenized in 20 mM Tris, 140 mM NaCl, 1 % (v/v) Nonidet P40, 1 mM PMSF (pH 7.4) [28]. Secreted fluid from the coagulating glands and prostate was harvested by washing the organs in PBS, containing 10 mM EDTA and 1 mM PMSF. Protein content was estimated from the absorbance (A_{280}) value. Western blots were developed using the rabbit anti-C1s IgG, a sheep anti-rabbit IgG horseradish peroxidase-conjugated IgG (Sigma) and 4-chloro-1-naphthol (Sigma) as substrate.

RESULTS

The murine C1r and C1s genes are duplicated

Southern-blot analysis of Balb/c genomic DNA with C1r and C1s genomic probes showed at least two bands for most of the restriction digests used (Figure 1), indicating that both C1r and C1s gene sequences are duplicated. The possibility that double-band patterns could result from digestion within single target sequences was ruled out by restriction analysis of the corresponding PCR products (results not shown). Similar results as in Figure 1 were obtained with C57BL/6 and 129 genomic DNA and with other enzymes.

An approx. 180 kb genomic clone, bac185, was isolated from a mouse 129/SvJ BAC library by C1s-specific screening and mapped by Southern blotting with multiple C1r and C1s genomic probes. With the same probe/enzyme combinations as in Figure 1, the C1r restriction pattern of bac185 was identical with that of genomic Balb/c DNA, whereas the C1s probe hybridized with only one of the fragments detected in the genomic DNA. The bac185 clone therefore contained one of the two copies of C1s genes detected in genomic Balb/c DNA and both copies of C1r genes (Figure 2).

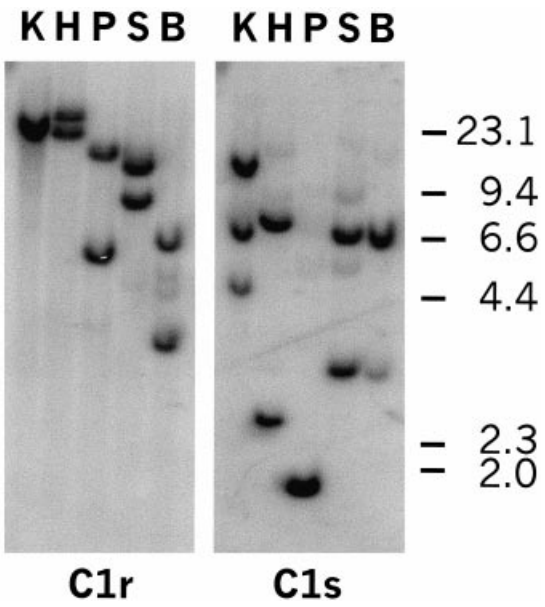

Figure 1 C1r and C1s genes are duplicated in the mouse

Balb/c liver genomic DNA (12 μg) were digested with restriction enzymes and subjected to Southern-blot analysis. The C1r and C1s probes were genomic segments encompassing 700 bp of *c1rA* exon 11 and 1100 bp of *c1sA* exons 5–6 respectively (depicted in Figure 2). K, *Kpn*I; H, *Hin*dIII; P, *Pst*I; S, *Sac*I; B, *Bam*HI. DNA size markers (in kb) are indicated on the right-hand-side margin.

Characterization of the C1sA, C1rA and C1rB genes

The contiguous restriction map of a region spanning approx. 110 kb of bac185 showed that the C1s gene (*c1sA*) and one of the C1r genes (*c1rA*) are at a distance of approx. 8 kb from each other, in opposite transcriptional orientation (Figure 2). Restriction fragment subcloning allowed sequencing of the entire *c1rA* and *c1sA* genes, except for 2 and 3 intronic gaps respectively. It also led to the identification of a C1r-related gene (*c1r-LP*) located 5′ to *c1rA* as described previously [29]. The C1r-LP gene is expressed mainly in the liver and encodes a serum protein of unknown function, structurally related to C1rA but missing the region from the EGF module to the beginning of the second CCP of C1rA. The remaining amino acid sequence shares 60 % residue identity with C1rA. No cross-hybridization was observed between *c1r-LP* and C1r or C1s genomic sequences. The second C1r gene (*c1rB*) located on the remaining 70 kb region of bac185 was also mapped and partially subcloned and sequenced.

Alignments of genomic and cDNA sequences (see below), completed with cap-specific 5′-RACE analysis of Balb/c liver RNA and information from poly(A)-containing EST sequences, allowed us to establish the complete exon boundary structure of both *c1rA* (Table 1) and *c1sA* (Table 2) genes. *c1rA* and *c1sA* comprise 11 and 12 exons respectively, spanning 11 and 12 kb of DNA respectively. Except for a single 5′-UTR exon and the

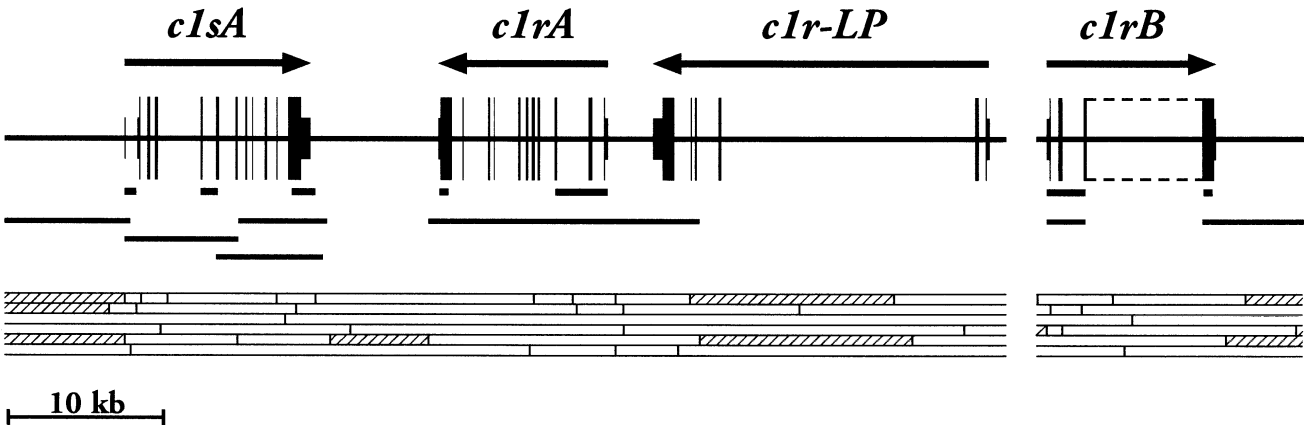

Figure 2 The murine C1r/C1s gene cluster

Shown are the 64 kb (left) and 17 kb (right) portions of the mapped approx. 110 and 70 kb regions of the approx. 180 kb bac185 genomic clone, containing the *c1sA*, *c1rA*, *c1r-LP* and *c1rB* genes. Because these two maps were not bridged together, orientation and position of *c1rB* in relation to the other genes (at an estimated distance of 30–90 kb) are unknown. Arrows denote the directions of transcription. Broken lines indicate the portion of *c1rB* of unknown exon–intron structure. Thick lines immediately underneath the genes show the location of the probes used for mapping, including the homologous target regions for the two C1r probes in *c1rA* and *c1rB*. Thinner lines show the plasmid subclones generated from bac185; from left to right: (*c1sA*) 8 kb *Bam*HI–*Xho*I, 7 kb *Hin*dIII, 6.9 kb *Pst*I and 6.2 kb *Hin*dIII; (*c1rA*) 17.5 kb *Hin*dIII; (*c1rB*) 3 kb PCR subclone and 6.5 kb *Bam*HI. The lowest part represents the restriction map for (from top to bottom) *Sac*I, *Kpn*I, *Cla*I, *Sma*I, *Hin*dIII and *Xho*I; hatched areas indicate uncharacterized segments. Genomic sequence data are deposited in GenBank® under the following accession nos: (*c1rA*) AF459008, AF459009, AF459010 and AF459011; (*c1rB*) AF459012, AF459013 and AF459014; (*c1sA*) AF459015, AF459016 and AF459017.

Table 1 Structure of murine C1rA exon–intron junctions

n	Exon (*n*)	Intron (*n*)	Exon (*n*+1)	Exon (bp)	Intron* (bp)	Amino acid†
		(5′-fl)…tgggaa	CTCAGATTTTT‡			
1	TTGAGAC AT	gtgagt…tgacag	G TGG CTC	184	851	Met^{1}
2	TAT GTT AAG	gtaggc…caacag	ATT TCT GCT	226	~ 2090	Lys^{76}
3	GCT GTA G	gtgagt…ttccag	AC CTT GAT	193	852	Asp^{141}
4	TGC CAG G	gtgagg…ttatag	CT GAG TGC	147	97	Ala^{190}
5	CAG CTC CAG	gtagag…catcag	ATC TAC GCT	197	208	Gln^{255}
6	ACT GAA A	gtgaga…catcag	CC ATC AAG	148	184	Thr^{305}
7	CTC ATG GAG	gtaaga…aaacag	GGA AAT CAG	122	~ 1640	Glu^{345}
8	TGC AAG A	gtaagt…tctaag	TC AAG AAC	79	290	Ile^{372}
9	ATG CGA G	gtagga…cctcag	GG ATA TAT	156	~ 1720	Gly^{424}
10	CTG CCA G	gtaaag…ctgcag	TG TGT GGG	75	506	Val^{449}
11	TCTGTAATCCC	atcaac…(3′-fl)§		922		

* Length of introns 2, 7 and 9 determined by restriction mapping.
† Amino acid corresponding to the junction codon (underlined in the sequence).
‡ First line shows 5′-flanking/5′-UTR junction, as determined by cap-specific RACE (5′-UTR, 182 bp).
§ Last line shows 3′-UTR/3′-flanking junction, as identified by poly(A)-containing ESTs (3′-UTR, 146 bp).

splice site interrupting the putative initiating codon, *c1rA* exon–intron junctions are all located at the same positions as those in *c1sA* and human *c1s* [10,11], in relation with predicted amino acid sequence (Figure 3). Sequencing of *c1rA* and *c1sA* included 2 and 1 kb respectively of 5′-flanking regions. Both sequences were devoid of TATA boxes and showed no significant homology with each other but substantial homology (more than 65% nucleotide identity for the first 200 bp) with their human counterparts (GenBank® accession nos. AC094008 and AC006512 respectively).

Although the position and orientation of the *c1rB* gene in relation with the *c1sA*/*c1rA*/*c1r-LP* gene cluster have not been determined, restriction mapping indicated that it spans 11 kb of DNA and is located within a distance of 30–90 kb from these genes. Partial characterization of the 3′-region of *c1rB* indicated that similar to *c1rA* and *c1sA*, the predicted SP domain is encoded by a single exon. The second C1s gene (*c1sB*) has not been isolated as it was absent not only from bac185 but also from a second BAC clone, which contained *c1sA*, *c1rA* and *c1r-LP* but not *c1rB*.

C1rA, C1rB, C1sA and C1sB cDNAs and predicted amino acid sequences

A full-length ORF cDNA sequence for murine C1r (C57BL/6 strain), spanning −166/+2250, was reported previously [13] (GenBank® accession no. AF148216). The sequence of this clone was identical with exonic sequences of the *c1rA* gene, except for two single basepair insertions in the 5′-UTR, 1 bp substitution in the ORF (predicting R101S amino acid change; amino acid numbering from the putative initiation methionine) and 1 bp substitution in the 3′-UTR. These differences were attributed to polymorphisms between the C57BL/6 and 129 strains. Therefore the previously reported C1r cDNA was assigned to the *c1rA* gene

Table 2 Structure of murine C1sA exon–intron junctions

n	Exon (*n*)	Intron (*n*)	Exon (*n*+1)	Exon (bp)	Intron* (bp)	Amino acid†
		(5′-fl)...gcaggg	GCTTAACTGCA‡			
1	AGCTCCTGAAG	gtaaca...tcaaag	GTGACACTGAG	64	779	
2	GGG ATG TG	gtaagt...ttccag	G TGT TTG	156	336	Trp^{2}
3	TCA GTG CAG	gtatat...tcttag	ATA ATC TCA	208	290	Gln^{71}
4	GCC ATA G	gtaagg...ttgcag	AC ATA AAT	178	~ 2750	Asp^{131}
5	TGT GGA G	gtgagt...ctctag	TC AAC TGT	126	829	Val^{173}
6	AGT TTA ACT	gtgcgt...ttccag	TTT GCT TCA	200	~ 980	Thr^{239}
7	GGA GAT C	gtgagt...tccaag	CC ATC TCC	154	744	Pro^{291}
8	GTT GTG GAG	gtaaag...ctccag	GGA CAT GTT	116	383	Glu^{329}
9	TGT CAA C	gtaggt...ttgtag	CT GTG TAT	79	493	Pro^{356}
10	GAA GGT G	gtaggt...ttctag	GG GAG TAT	129	658	Gly^{399}
11	ATT CCA G	gtaatc...tcctag	CC TGT GGA	75	800	Ala^{424}
12	AATTGAACATT	gctggt...(3′-fl)§		1396		

* Length of introns 4 and 6 determined by restriction mapping.
† Amino acids corresponding to the junction codon (underlined in the sequence).
‡ First line shows 5′-flanking/5′-UTR junction, as determined by cap-specific RACE (5′-UTR, 215 bp in two exons).
§ Last line shows 3′-UTR/3′-flanking junction, as identified by poly(A)-containing ESTs (3′-UTR, 602 bp).

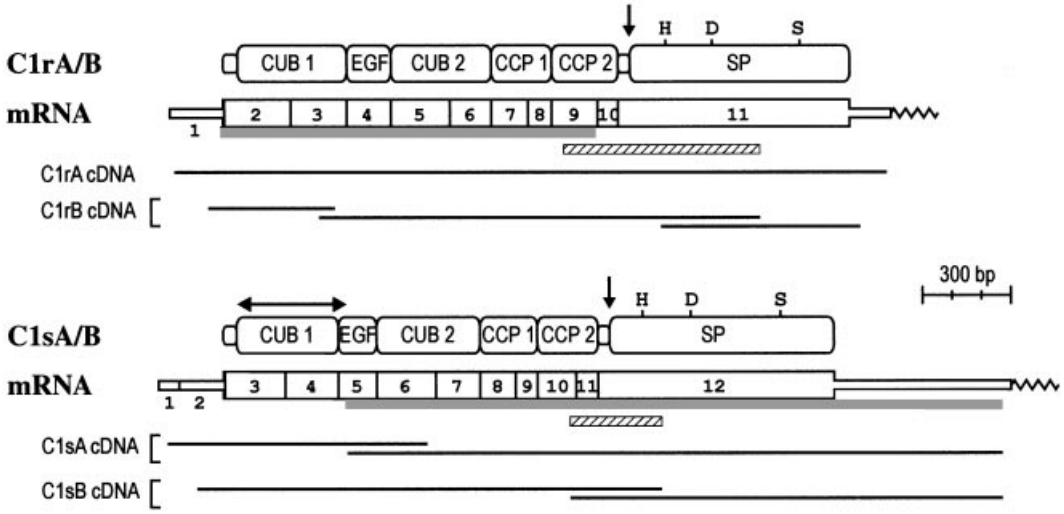

Figure 3 Schematic alignment of C1rA, C1rB, C1sA and C1sB cDNAs and predicted preproteins

The relationship of exons to modular structure. Narrower portions in the proteins represent the leader peptide and the CS. H, D, S, catalytic triad residues; vertical arrow, activating cleavage site; double-arrowhead line, portion of C1sA/B recognized by the antibody used for Western-blot analysis (Figure 7). mRNA 5′- and 3′-ends and splice junctions were determined for C1rA and C1sA only; shadowed areas, cDNA probes used for Northern-blot analysis (Figure 5); hatched areas, cDNA segments corresponding to the RT–PCR assays specific for each of the four mRNAs (Figure 6). Black lines, cDNA segments defining full-length ORF cDNA sequences used in the present study. C1rA cDNA is a previously reported clone (GenBank® accession no. AF148216) [13]. C1rB, C1sA and C1sB cDNA sequence data were generated in the present study and are deposited in GenBank® under accession nos AF459018, AF459019 and AF45920 respectively.

and will be referred to as C1rA cDNA. C1rB-specific RT–PCR, designed on *c1rB* genomic sequence data, consistently failed to generate any product from Balb/c or C57BL/6 liver RNA. However, C1rB-specific RT–PCR products were obtained using RNA from C57BL/6 pooled male reproductive tissues, which were found to contain high levels of C1r mRNA by Northern-blot analysis. Three overlapping C1rB cDNA clones encompassing the entire ORF (Figure 3) were thus generated from these tissues and sequenced. All three sequences were distinguishable from their corresponding regions of C1rA cDNA, whereas sequences of the 5′ and 3′ cDNA clones were identical with the corresponding regions of the *c1rB* gene. The C1rA cDNA sequence is 2449 bp long [from cap site to poly(A) site] and shares 97.8% nucleotide identity with C1rB, throughout the ORF. Predicted C1rA and C1rB amino acid sequences (707 and 706 residues respectively) share 95.5% residue identity and are about equally similar to human C1r [30] (80.9 and 79.9% identity respectively).

Murine C1s cDNA clones were generated by RT–PCR from Balb/c liver RNA and sequenced. The Balb/c liver C1s cDNA matched the exonic sequences of the 129/Sv *c1sA* gene, differing by 11 bp substitutions, eight of them located in the ORF (predicting $Gly^{76} \rightarrow Asp$, $Lys^{86} \rightarrow Arg$ and $Glu^{305} \rightarrow Gln$ substitutions), an 8 bp repeat insertion in the 5′-UTR and two minor insertion/deletions in the 3′-UTR, which were attributed to polymorphism between the 129/Sv and Balb/c strains. Therefore the Balb/c cDNA clone was assigned to the *c1sA* gene. Using C57BL/6 male reproductive tissue RNA as template and primers

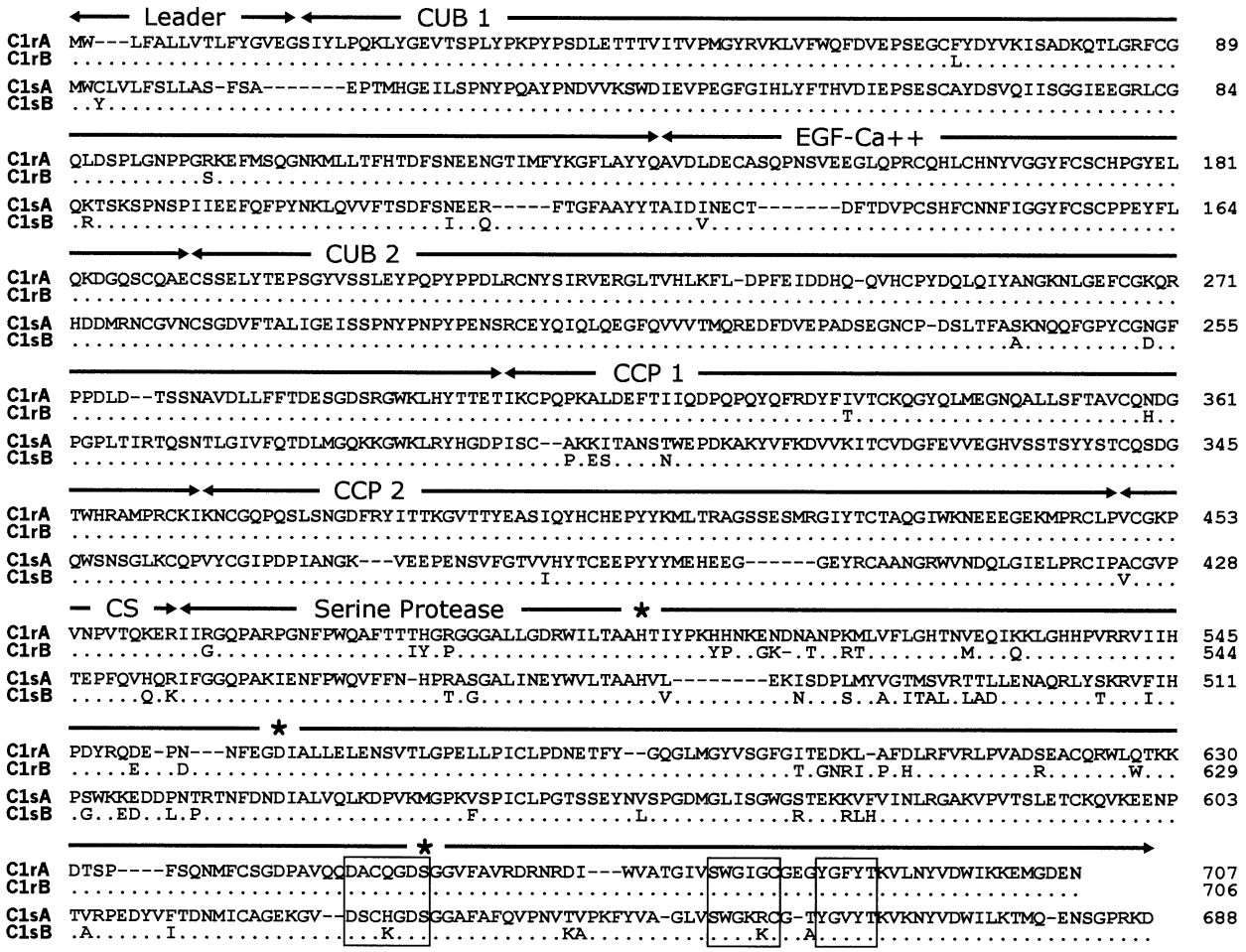

Figure 4 Amino acid sequence alignment of C1rA, C1rB, C1sA and C1sB

C1rB and C1sB amino acid residues are specified only when differing from C1rA or C1sA respectively (positions of identity indicated by dots); alignment gaps are indicated by dashes. Amino acid numbering starts at the putative initiation Met. Protein module boundaries, as inferred from human C1r and C1s, are indicated above the sequences. The His, Asp and Ser residues of the catalytic triad are indicated by asterisks. Sequences corresponding to the walls of the primary specificity pocket are enclosed in boxes. Alignment was performed using CLUSTAL method with PAM250 residue weight table. Calculated M_r of predicted preproteins: C1rA, 80151; C1rB, 79945; C1sA, 76865; C1sB, 76708.

designed on *c1sA* exon 10 and 3′-UTR, an RT–PCR clone was generated (Figure 3), whose sequence substantially differed from that of Balb/c liver C1sA cDNA (82 substitutions in 1428 bp, mostly in the region encoding the SP domain, predicting 36-amino-acid substitutions). Sequences were confirmed on two colonies and by direct sequencing of the PCR product. An overlapping RT–PCR product extending to the 5′-UTR was generated using this diverging sequence for reverse primer design and predicted 12 additional amino acid substitutions. That these differences resulted from polymorphism between the Balb/c and C57BL/6 strains was ruled out by a C57BL/6 EST (GenBank® accession no. AI386094), encompassing the 3′-end of the gene, which was identical with Balb/c cDNA and 129 C1sA genomic sequences. Although the *c1sB* gene has not been isolated, the C1s cDNA from male reproductive tissues does not appear to derive from the *c1sA* gene and will be referred to as C1sB cDNA. C1sA cDNA is 2882 bp long [from cap site to poly(A)] and shares 95.6% nucleotide identity with C1sB cDNA. C1sA- and C1sB-predicted amino acid sequences (688 residues each) show 93.0% residue identity with each other and 74.0 and 74.5% identity respectively, with human [31], 79.7 and 82.7% with hamster [15], and 81.0 and 84.3% with rat [14] C1s.

The modular structure characteristic of C1r and C1s and the mannan-binding lectin-associated SPs [32] is conserved in all four murine sequences (Figure 3). Amino acid alignment of the predicted murine preproteins (Figure 4) shows the differences between C1rA and C1rB, and C1sA and C1sB, in relation to their modular structure. C1rA and C1rB amino acid sequences differ by 28 residue substitutions and a single residue deletion/insertion, whereas C1sA and C1sB differ by a total of 48 residue substitutions. Most of these differences are located in the SP domains. However, the catalytic triad residues (His^{501}, Asp^{559} and Ser^{656}, in C1rA numbering), their surrounding sequences, and the residues forming the three walls of the primary substrate specificity pocket in SPs of known structure [33] are conserved between C1rA and C1rB and also between C1sA and C1sB, except for two substitutions, $His^{628} \rightarrow$ Lys and $Arg^{658} \rightarrow$ Lys, in C1sB; human C1s has Lys and Gln residues at the respective positions. The Asp residue at the bottom of the primary specificity pocket (position 189 in chymotrypsinogen), determining substrate specificity of SPs with trypsin-like specificity [34], is conserved in all four murine enzymes. At the activating cleavage site, the Arg-Ile-Ile sequence is conserved between C1rA and C1rB, whereas a Lys-Ile-Phe sequence in C1sB is substituted for

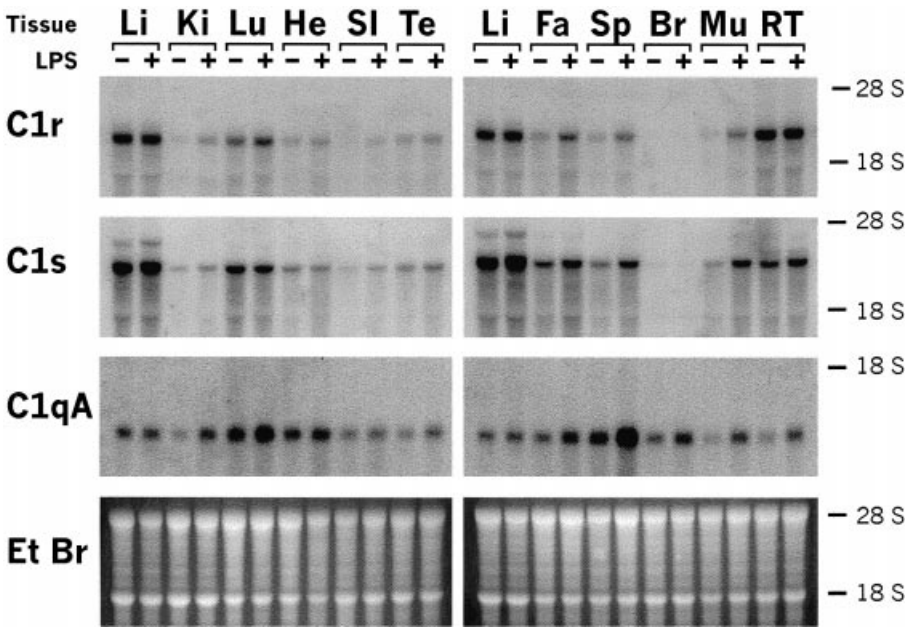

Figure 5 Tissue distribution of C1r, C1s and C1qA mRNAs

Constitutive and LPS-regulated expression as assessed by Northern-blot analysis of total RNA. Li, liver; Ki, kidney; Lu, lung; He, heart; Sl, small intestine; Te, testis; Fa, epididymal fat; Sp, spleen; Br, brain; Mu, skeletal muscle; RT, pooled male reproductive tissues, consisting of epididymis, vas deferens, seminal vesicles, coagulating glands and prostate. C1r and C1s probes are depicted in Figure 3. Left- and right-hand-side panel refers to two separate gels and filters; thus liver RNA was loaded in both gels for comparison. Ethidium bromide staining (lower panels) shows equal loading and quality of the RNA. Responsiveness of the mice to LPS was monitored by the SAA mRNA induction in the liver (not shown). mRNA quantification is shown in Table 3.

Table 3 Tissue distribution of C1r, C1s and C1qA mRNAs

Constitutive level is given as the percentage of hepatic level and values in parentheses indicate the fold increase after LPS injection. Li, liver; Ki, kidney; Lu, lung; He, heart; Sl, small intestine; Te, testis; Fa, epididymal fat; Sp, spleen; Br, brain; Mu, skeletal muscle; RT, pooled male reproductive tissues consisting of epididymis, vas deferens, seminal vesicles, coagulating glands and prostate.

mRNA	Li	Ki	Lu	He	Sl	Te	Fa	Sp	Br	Mu	RT
C1r (A+B)	100.0	2.6	40.8	10.0	1.2	10.7	17.0	13.0	0.3	5.0	148.0
	(0.9)	(5.7)	(1.5)	(1.5)	(2.5)	(1.5)	(3.0)	(3.0)		(7.0)	(0.9)
C1s (A+B)	100.0	2.9	55.2	13.2	2.5	9.5	25.0	12.9	0.3	5.7	31.0
	(1.1)	(3.1)	(0.9)	(0.9)	(2.0)	(1.7)	(2.0)	(3.0)		(5.8)	(1.5)
C1qA	100.0	28.3	219.0	136.0	57.0	36.0	93.0	254.0	83.0	23.0	45.0
	(1.0)	(4.3)	(1.6)	(1.3)	(1.3)	(2.2)	(1.9)	(3.5)	(2.2)	(6.0)	(2.7)

Arg-Ile-Phe in C1sA. The presence of a Lys instead of an Arg residue at the scissile bond makes C1sB unique among complement proteins cleaved during activation. All Cys residues are conserved in the four murine proteins, at positions equivalent to those in human C1r and C1s. In the non-catalytic modules, C1rB differs from C1rA by only four residue substitutions located in the first CUB and in the first CCP. One of the substitutions ($Arg^{101} \rightarrow Ser$) actually reflects a polymorphism affecting C1rA of C57BL/6 compared with the 129/Sv strain. C1sA differs from C1sB in the non-catalytic region by 15 residue substitutions. One substitution in the first CUB module ($Lys^{86} \rightarrow Arg$) reflects a polymorphism affecting Balb/c C1sA. C1r isomorphs have three potential N-glycosylation sites each (N124, N220 and N583/582), at positions equivalent to three of the four glycosylation sites of human C1r [32]. C1sA has five potential glycosylation sites (N174, N301, N558, N602 and N641), the first one being located at the same position as one of the two sites of human C1s [32]. Only three of the five sites are conserved in C1sB (N174, N558 and N602).

Tissue distribution of C1rA, C1rB, C1sA and C1sB mRNAs

Northern-blot analysis of C1r and C1s mRNAs in C57BL/6 tissues (Figure 5 and Table 3) showed single bands at approx. 2.5 and 3.2 kb respectively, consistent with the respective cDNA sequence lengths. C1r and C1s mRNAs have a very similar tissue distribution. Both messages were detected predominantly in the liver and to a lesser extent in all other tissues examined, except in the brain. Of particular interest is the high level of both mRNAs, especially C1r in male reproductive tissues. The overall tissue distribution of C1r and C1s mRNA significantly differs from that of C1qA (Figure 5), and of C1qC, which is very similar to that of C1qA (results not shown). LPS-induced inflammation

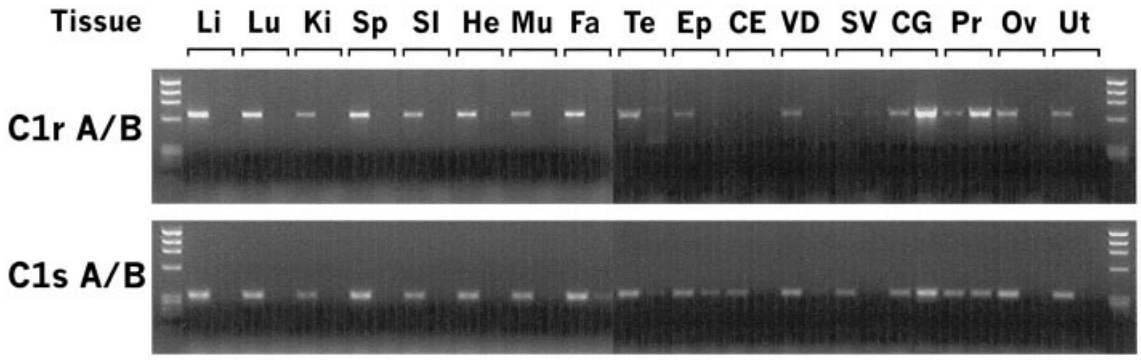

Figure 6 Tissue distribution of C1rA, C1rB, C1sA and C1sB mRNAs

For each tissue, (upper panel) C1rA and C1rB (lanes 1 and 2 respectively), and (lower panel) C1sA and C1sB (lanes 1 and 2 respectively) mRNAs were analysed by RT–PCR. PCR product sizes were approx. 650 bp for C1rA and C1rB and 310 bp for C1sA and C1sB (see Figure 3). Ep, corpus and caput epididymis; CE, cauda epididymis; VD, vas deferens; SV, seminal vesicle; CG, coagulating gland; Pr, prostate; Ov, ovary; Ut, female reproductive tissues, including uterus and oviduct. DNA size markers (ΦX 174, *Hae*III digest) were loaded in the first and last lanes of each gel. The same RNA samples as for the Northern-blot analysis (Figure 5) were analysed here for the first nine tissues from the left-hand side.

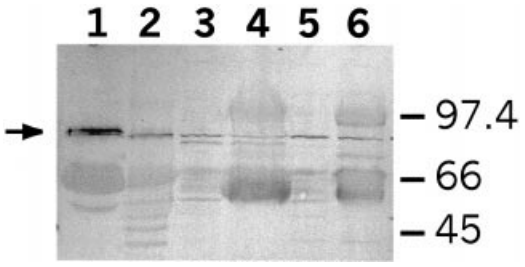

Figure 7 Western-blot analysis of C1s in serum, liver and reproductive tissues

Lane 1, serum (15 μl at 1:10 dilution); lane 2, liver lysate; lane 3, coagulating gland lysate; lane 4, coagulating gland lavage fluid; lane 5, prostate lysate; lane 6, prostate lavage fluid. In lanes 2–6, approx. 100 μg of protein/lane were loaded. Molecular mass markers ($\times 10^{-3}$) are shown on the right-hand side. Arrow indicates specific bands corresponding to C1s.

did not affect hepatic levels of C1r and C1s mRNAs and produced only modest increases (2- to 7-fold) in some extra-hepatic sites, i.e. in the kidney, small intestine, fat, spleen and skeletal muscle. Effectiveness of the LPS treatment was verified by monitoring induction of hepatic serum amyloid A mRNA. Similar results were obtained using Balb/c RNA.

Because Northern-blot analysis could not distinguish A from B isomorphs, semiquantitative RT–PCR assays specific for each isomorph were developed and used for C57BL/6 tissue RNA analysis. As shown in Figure 6, C1rB mRNA was only detected in two male reproductive accessory organs, mostly in the coagulating gland, and to a lesser extent in the prostate. These tissues also had some C1rA mRNA but were among the lowest expressing sites for this transcript. C1rA mRNA was detected in all other tissues analysed, including female and other male reproductive tissues, except in the cauda epididymis and in the seminal vesicles; the highest C1rA signal was found in the liver. Tissue distribution of C1sA and C1sB mRNAs was similar to that of C1rA and C1rB mRNAs respectively, except that some C1sB mRNA was also detected in the pooled caput and corpus epididymis and in trace amounts in epididymal fat and testes. In preliminary experiments, RT–PCR on the RNA from pooled C57BL/6 male reproductive tissues indicated that the C1r and C1s mRNAs detected by Northern blotting in this sample (Figure 5) were essentially C1rB and C1sB mRNAs respectively. In Balb/c, C1rB and C1sB mRNAs were also detected at high levels in the pooled male reproductive tissues but were undetectable in the liver (results not shown).

Analysis of C1s protein in the liver and male reproductive tissues

The presence of C1s in the coagulating gland and prostate lysates and secreted fluids in comparison with serum and liver lysate was assessed by Western-blot analysis (Figure 7). The polyclonal antibody used was raised against a recombinant polypeptide encompassing the first 131 amino acids of the putative mature C1sA protein, corresponding to the first CUB and 40% of the EGF-Ca^{2+} modules. C1sA and C1sB differ by only 3–5 residue substitutions in this region (two of them being polymorphic among strains) and are, therefore, expected to be both recognized by the antibody. An approx. 92 kDa protein was detected in the serum, whereas an 88 kDa band was present in the liver lysate, in the tissue lysates and secreted fluids of the coagulating gland and the prostate. A third faint band of approx. 84 kDa was present only in the coagulating gland. These results indicate the presence in the coagulating gland and in the prostate of protein species immunologically related to C1s, consistent with expression of C1sA and C1sB mRNA in these tissues (Figure 6).

DISCUSSION

In the present study, we demonstrate that the *c1r* and *c1s* genes are duplicated in the mouse. Three of these genes (*c1rA*, *c1sA* and *c1rB*) were localized within the same 180 kb region and characterized. The *c1rA* and *c1sA* genes are primarily expressed in the liver and are thus identified as the murine homologues of the human *c1r* and *c1s* genes. The expression of the other two genes, *c1rB* and *c1sB*, is restricted to the male reproductive system, wherein proteins antigenically related to C1s were detected. Full-length cDNA sequences for all four genes have been determined. Predicted amino acid sequences of C1rB and C1sB are very similar to their respective hepatic isoforms, with

most of the differences residing in surface loops of their SP domains.

Early genetic mapping studies provided evidence that human C1r and C1s are encoded by single genes [9,10,31]. Both proteins were shown to be synthesized and secreted as active enzymes by cells of hepatic origin [4,5]. Tissue distribution studies of C1r and C1s mRNA in the baboon [9] and of C1s in the hamster [17] identified the liver as the primary site of expression. Similarly, C1rA and C1sA mRNAs were found predominantly in the liver and to a lesser extent in other tissues. Murine C1rA and C1sA, C1qA and C1qC mRNAs were also present in fat with levels increasing during LPS-induced inflammation. Expression in fat tissues has not been reported before for C1r and C1s. *c1rB* and *c1sB* genes showed parallel expression, which was essentially restricted to the coagulating gland, the prostate and the epididymis. Their co-ordinated expression suggests a common mechanism for *c1rB* and *c1sB* gene regulation, which apparently differs from the mechanism regulating the *c1rA* and *c1sA* genes.

Expression of functional complement in the mammalian female reproductive tract is well established [18]. In the male reproductive system, however, previous studies have demonstrated expression of complement regulatory proteins [18,20,21], but not of C3 or other complement components [35]. The present finding of C1r and C1s isoforms, expressed exclusively in the male genital tract, is reminiscent of murine DAF [36] and CD59 [37], regulators of the C3 convertases and of the complement lytic pathway respectively. In humans and rats, *Daf* and *cd59* are single copy genes but in the mouse both are duplicated, and one copy of each (*Daf-TM* and *cd59b*) is specifically expressed in the testis. The expression of C1rB and C1sB by accessory glands but not by testis may reflect a role in different sites or physiological events when compared with that for DAF-TM and CD59B. Also in rodents, the single genes encoding MCP [38] and C4b-binding protein α [19] are predominantly expressed in the testis and in the epididymis respectively. Finally, a tissue-specific isoform of human MCP, arising from alternative splicing of the single gene, is found on the inner acrosomal membrane of the sperm and is reportedly involved in sperm–oocyte interaction [20]. The specific expression of these control proteins and isoforms in the male reproductive tissue suggested that complement may play a role in reproduction.

The murine isoforms of C1r and C1s have by far a higher homology with each other than with their human homologues, raising the possibility that they may not differ substantially in function. Comparisons of predicted primary structures of C1rA and C1sA with their respective isomorphs and their human homologues indicate that although most substitutions are located in their SP domains, structural features of functional importance are highly conserved. This includes the catalytic triad residues, the oxyanion hole, the walls of the primary specificity pocket and the non-specific substrate-binding site. Also highly conserved in both pairs of murine isoforms are the three groups of interacting residues, which confer a functionally important rigid intramolecular CCP2–SP interface in human C1r [2] and C1s [39]. Similarly, most of the residues involved in the CCP1–SP interaction and responsible for human C1r dimerization [2] have been conserved in both murine C1r isomorphs. In human C1r and C1s, the N-terminal CUB and EGF-Ca^{2+} domains were shown to be necessary and sufficient to promote the strong Ca^{2+}-dependent C1r–C1s interaction [40]. Throughout this region, the sequences of the murine isoforms are virtually identical and remarkably conserved with their human homologues. Included are the conserved Ca^{2+}-binding residues in the C1r EGF module [41], the consensus sequence for Ca^{2+} binding [1] in all four murine proteins and the EGF module of C1s [42], which was also found to promote C1r interaction with C1q. Taken together, the structural similarities of key functional elements between the murine C1r and C1s isoforms and their human homologues strongly suggest that both pairs of murine proteins are capable of forming Ca^{2+}-dependent heterotetramers similar to the human C1s-C1r-C1r-C1s, as well as C1 complexes with C1q. Expression of C1q in the male genital tract (Figure 5) is consistent with this interpretation. The biological relevance of duplicated genes in the mouse would then reside in the tissue-specific control of expression, the structural differences possibly reflecting functional optimization of the isoforms in the serum and in the seminal fluid. Accordingly, the single human C1r and C1s proteins may combine functional properties of both murine isoforms. Currently available sequence data of the human genome are still presenting major gaps in the *c1r/c1s* region of chromosome 12, including a part of *c1r* itself and the *c1r–c1s* intergenic region. Therefore the alternative possibility that human isoforms could be generated from single genes by alternative splicing of SP domain-encoding exons located in this region cannot be excluded. This would be analogous to the recently described case of two SPs of the complement lectin pathway of identical modular organization to C1r and C1s, the mannan-binding lectin-associated SPs 1 and 3 [43], which are encoded by a single gene, differing only in their SP domains.

In the light of the recently reported crystal structures of the catalytic domains of human C1r [2] and C1s [39], most of the differences in the amino acid sequence between C1rB and C1sB and their liver isoforms, are located in regions corresponding to surface loops surrounding the active site of the human SPs. These loops have a conserved position among SPs and their variability in length and sequence influences substrate specificity and catalytic efficiency [44]. The presence of multiple amino acid residue differences in these loops suggests differences in substrate specificity between the murine isoforms and raises the possibility of a role in mouse reproduction independent of complement activation. Non-complement functions have already been suggested for C1s as it has been shown to cleave non-complement proteins, such as the heavy chain of the MHC class I antigens [45] and insulin-like growth factor-binding protein 5 [46]. Whether C1sB has the same substrates as serum C1sA, i.e. the complement proteins C4 and C2, or some currently unknown substrates exclusively found in the male reproductive tract remains to be established.

This work was supported by United States Public Health Service (grant no. RO1 AR44505). We thank Dr Janming Wu and Dr Andrew W. Gibson for their invaluable help with the genomic sequences, Dr Maria Alexiou for kindly helping with the dissection of mouse reproductive tissues, and Dr Franz Petry and Dr Michael Loos for providing the C1qA and C1qC cDNA clones.

REFERENCES

1 Arlaud, G. J., Gaboriaud, C., Thielens, N. M., Rossi, V., Bersch, B., Hernandez, J. F. and Fontecilla-Camps, J. C. (2001) Structural biology of C1: dissection of a complex molecular machinery. Immunol. Rev. **180**, 136–145

2 Budayova-Spano, M., Lacroix, M., Thielens, N. M., Arlaud, G. J., Fontecilla-Camps, J. C. and Gaboriaud, C. (2002) The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. EMBO J. **21**, 231–239

3 Colten, H. R. and Strunk, R. C. (1993) Synthesis of complement components in liver and at extrahepatic sites. In Complement in Health and Disease, 2nd edn (Whaley, K, Loos, M. and Weiler, J. M., eds.), pp. 127–158, Kluwer Academic Publishers, Dordrecht, The Netherlands

4 Morris, K. M., Aden, D. P., Knowles, B. B. and Colten, H. R. (1982) Complement biosynthesis by the human hepatoma-derived cell line HepG2. J. Clin. Invest. **70**, 906–913

5 Ramadori, G., Heinz, G. H., Martin, H., Meyer zum Buschenfelde, K. H. and Loos, M. (1986) Biosynthesis of the subcomponents C1q, C1r and C1s of the first component of complement (C1) by guinea pig hepatocyte primary cultures. Eur. J. Immunol. **16**, 1137–1141

6 Legoedec, J., Gasque, P., Jeanne, J. F., Scotte, M. and Fontaine, M. (1997) Complement classical pathway expression by human skeletal myoblasts *in vitro*. Mol. Immunol. **34**, 735–741

7 Gasque, P., Ischenko, A., Legoedec, J., Mauger, C., Schouft, M. T. and Fontaine, M. (1993) Expression of the complement classical pathway by human glioma in culture. A model for complement expression by nerve cells. J. Biol. Chem. **268**, 25068–25074

8 Nguyen, V. C., Tosi, M., Gross, M. S., Cohen-Haguenauer, O., Jegou-Foubert, C., de Tand, M. F., Meo, T. and Frezal, J. (1988) Assignment of the complement serine protease genes C1r and C1s to chromosome 12 region 12p13. Hum. Genet. **78**, 363–368

9 Kusumoto, H., Hirosawa, S., Salier, J. P., Hagen, F. S. and Kurachi, K. (1988) Human genes for complement components C1r and C1s in a close tail-to-tail arrangement. Proc. Natl. Acad. Sci. U.S.A. **85**, 7307–7311

10 Tosi, M., Duponchel, C., Meo, T. and Couture-Tosi, E. (1989) Complement genes C1r and C1s feature an intronless serine protease domain closely related to haptoglobin. J. Mol. Biol. **208**, 709–714

11 Endo, Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H., Matsushita, M., Nonaka, M. and Fujita, T. (1998) Two lineages of mannose-binding lectin-associated serine protease (MASP) in vertebrates. J. Immunol. **161**, 4924–4930

12 Colten, H. R. (1993) Deficiencies of the first component of complement (C1): an update. Behring Inst. Mitt. **93**, 287–291

13 Byun, S. J., Bahk, Y. Y., Ryoo, Z. Y., Kim, K. E., Hwang, H. Y., Lee, J. W., Yi, J. Y. and Kim, T. Y. (2001) Identification of cDNA encoding a serine protease homologous to human complement C1r precursor from grafted mouse skin. J. Invest. Dermatol. **116**, 374–379

14 Sakai, H., Nakashima, S., Yoshimura, S., Nishimura, Y., Sakai, N. and Nozawa, Y. (1998) Molecular cloning of a cDNA encoding a serine protease homologous to complement C1s precursor from rat C6 glial cells and its expression during glial differentiation. Gene **209**, 87–94

15 Kinoshita, H., Sakiyama, H., Tokunaga, K., Imajoh-Ohmi, S., Hamada, Y., Isono, K. and Sakiyama, S. (1989) Complete primary structure of calcium-dependent serine proteinase capable of degrading extracellular matrix proteins. FEBS Lett. **250**, 411–415

16 Lawson, P. R. and Reid, K. B. (2000) A novel PCR-based technique using expressed sequence tags and gene homology for murine genetic mapping: localization of the complement genes. Int. Immunol. **12**, 231–240

17 Sakiyama, H., Yamaguchi, K., Chiba, K., Nagata, K., Taniyama, C., Matsumoto, M., Suzuki, G., Tanaka, T., Tomoasawa, T., Yasukawa, M. et al. (1991) Biochemical characterization and tissue distribution of hamster complement C1s. J. Immunol. **146**, 183–187

18 Oglesby, T. J. (1998) The complement system in reproduction. In The Human Complement System in Health and Disease (Volanakis, J. E. and Frank, M. M., eds.), pp. 355–373, Marcel Dekker, New York

19 Nonaka, M. I., Wang, G., Mori, T., Okada, H. and Nonaka, M. (2001) Novel androgen-dependent promoters direct expression of the C4b-binding protein α-chain gene in epididymis. J. Immunol. **166**, 4570–4577

20 Anderson, D. J., Abbott, A. F. and Jack, R. M. (1993) The role of complement component C3b and its receptors in sperm–oocyte interaction. Proc. Natl. Acad. Sci. U.S.A. **90**, 10051–10055

21 Rooney, I. A., Atkinson, J. P., Krul, E. S., Schonfeld, G., Polakoski, K., Saffitz, J. E. and Morgan, B. P. (1993) Physiologic relevance of the membrane attack complex inhibitory protein CD59 in human seminal plasma: CD59 is present on extracellular organelles (prostasomes), binds cell membranes and inhibits complement-mediated lysis. J. Exp. Med. **177**, 1409–1420

22 Brown, T. (1993) Analysis of DNA sequences by blotting and hybridization. In Current Protocols in Molecular Biology (Ausube, F. M., Brent, R., Kingston, R. E., Moore, D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 2.9.1–2.10.16, John Wiley & Sons, New York

23 Schaefer, B. C. (1995) Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal. Biochem. **227**, 255–273

24 Chirgwin, J. M., Przybyla, A. E., Mac Donald, R. J. and Rutter, W. J. (1979) Isolation of biologically active ribonucleic acid from source enriched in ribonucleases. Biochemistry **18**, 5294–5299

25 Sackstein, R. and Colten, H. R. (1984) Molecular regulation of MHC class III (C4 and factor B) gene expression in mouse peritoneal macrophages. J. Immunol. **133**, 1618–1626

26 Petry, F., Reid, K. B. and Loos, M. (1991) Gene expression of the A- and B-chain of mouse C1q in different tissues and the characterization of the recombinant A-chain. J. Immunol. **147**, 3988–3993

27 Steinbuch, M. and Audran, R. (1969) The isolation of IgG from mammalian sera with the aid of caprylic acid. Arch. Biochem. Biophys. **1314**, 279–284

28 Inoue, N., Fuki, A., Nomura, M., Matsumoto, M., Toyoshima, K. and Seya, T. (2001) A novel chicken membrane-associated complement regulatory protein: molecular cloning and functional characterization. J. Immunol. **166**, 424–443

29 Circolo, A., Garnier, G. and Volanakis, J. E. (2003) A novel murine complement-related gene encoding a C1r-like protein. Mol. Immunol., in the press

30 Leytus, S. P., Kurachi, K., Sakariassen, K. S. and Davie, E. W. (1986) Nucleotide sequence of the cDNA coding for human complement C1r. Biochemistry **25**, 4855–4863

31 Tosi, M., Duponchel, C., Meo, T. and Julier, C. (1987) Complete cDNA sequence of human complement Cls and close physical linkage of the homologous genes Cls and Clr. Biochemistry **26**, 8516–8524

32 Arlaud, G. J., Volanakis, J. E., Thielens, N. M., Narayana, S. V. L., Rossi, V. and Xu, Y. (1998) The atypical serine proteases of the complement system. Adv. Immunol. **69**, 249–307

33 Cohen, G. H., Silverton, E. W. and Davies, D. R. (1981) Refined crystal strucutre of γ-chymotrypsin at 1.9 Å resolution. J. Mol. Biol. **148**, 449–479

34 Graf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J. and Rutter, W. J. (1987) Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. Biochemistry **26**, 2616–2623

35 Bozas, S. E., Kirszbaum, L., Sparrow, R. L. and Walker, I. D. (1993) Several vascular complement inhibitors are present on human sperm. Biol. Reprod. **48**, 503–511

36 Spicer, A. P., Seldin, M. F. and Gendler, S. J. (1995) Molecular cloning and chromosomal localization of the mouse decay-accelerating factor genes. Duplicated genes encode glycosylphosphatidylinositol-anchored and transmembrane forms. J. Immunol. **155**, 3079–3091

37 Qian, Y. M., Qin, X., Miwa, T., Sun, X., Halperin, J. A. and Song, W. C. (2000) Identification and functional characterization of a new gene encoding the mouse terminal complement inhibitor CD59. J. Immunol. **165**, 2528–2534

38 Miwa, T., Nonaka, M., Okada, N., Wakana, S., Shiroishi, T. and Okada, H. (1998) Molecular cloning of rat and mouse membrane cofactor protein (MCP, CD46): preferential expression in testis and close linkage between the mouse Mcp and Cr2 genes on distal chromosome 1. Immunogenetics **48**, 363–371

39 Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G. J. and Fontecilla-Camps, J. C. (2000) Crystal structure of the catalytic domain of human complement c1s: a serine protease with a handle. EMBO J. **19**, 1755–1765

40 Thielens, N. M., Enrie, K., Lacroix, M., Jaquinod, M., Hernandez, J. F., Esser, A. F. and Arlaud, G. J. (1999) The N-terminal CUB-epidermal growth factor module pair of human complement protease C1r binds Ca^{2+} with high affinity and mediates Ca^{2+}-dependent interaction with C1s. J. Biol. Chem. **274**, 9149–9159

41 Bersch, B., Hernandez, J. F., Marion, D. and Arlaud, G. J. (1998) Solution structure of the epidermal growth factor (EGF)-like module of human complement protease C1r, an atypical member of the EGF family. Biochemistry **37**, 1204–1214

42 Busby, T. F. and Ingham, K. C. (1990) NH_2-terminal calcium-binding domain of human complement C1s- mediates the interaction of C1r- with C1q. Biochemistry **29**, 4613–4618

43 Dahl, M. R., Thiel, S., Matsushita, M., Fujita, T., Willis, A. C., Christensen, T., Vorup-Jensen, T. and Jensenius, J. C. (2001) MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway. Immunity **15**, 127–135

44 Perona, J. J. and Craik, C. S. (1997) Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. J. Biol. Chem. **272**, 29987–29990

45 Eriksson, H. and Nissen, M. H. (1990) Proteolysis of the heavy chain of major histocompatibility complex class I antigens by complement component C1s. Biochim. Biophys. Acta **1037**, 209–215

46 Busby, Jr, W. H., Nam, T. J., Moralez, A., Smith, C., Jennings, M. and Clemmons, D. R. (2000) The complement component C1s is the protease that accounts for cleavage of insulin-like growth factor-binding protein-5 in fibroblast medium. J. Biol. Chem. **275**, 37638–37644

Received 7 October 2002/17 December 2002; accepted 6 January 2003
Published as BJ Immediate Publication 6 January 2003, DOI 10.1042/BJ20021555