Molecular Analysis of Human Complement Component C5: Localization of the Structural Gene to Chromosome 9†

Rick A. Wetsel,*,‡,§ Richard S. Lemons,∥ Michelle M. Le Beau,⊥ Scott R. Barnum,‡ Deborah Noack,‡ and Brian F. Tack‡

Department of Immunology, Research Institute of Scripps Clinic, La Jolla, California 92037, Department of Tumor Cell Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38101, and Joint Section of Hematology/Oncology, University of Chicago, Chicago, Illinois 60637

Received July 27, 1987; Revised Manuscript Received October 27, 1987

ABSTRACT: A human C5 clone (pC5HG2) was isolated from a cDNA library constructed from Hep G2 mRNA. The DNA sequence showed that the pC5HG2 insert was comprised of 3309 base pairs of pro-C5 coding sequence and 404 base pairs of 3'-untranslated sequence. The derived amino acid sequence contained the entire coding sequence of the C5 α-chain, the β-α-chain junction region, and 100 amino acids (approximately 50%) of the β-chain. Protein sequences of four C5 tryptic peptides were aligned exactly to this sequence and demonstrated that C5 synthesized and secreted by Hep G2 cells is probably identical with plasma-derived C5. Coding sequence alignment of the human C5 sequences with those of murine C5 indicated that 80% of the nucleotides and 79% of the amino acids were placed identically in the two species. Amino acid sequence alignment of the homologous family members C3, C4, and α_2-macroglobulin with that of C5 demonstrated 27%, 25%, and 19% identity, respectively. As was found in murine C5, the corresponding thiol ester region of human C5 contained several conserved amino acids, but the critical cysteine and glutamine residues which give rise to the intramolecular thiol ester bond in C3, C4, and α_2-macroglobulin were absent in C5, having been replaced by serine and alanine, respectively. With the use of a panel of hamster-human somatic cell hybrids, the C5 gene was mapped to human chromosome 9. In situ chromosomal hybridization studies employing metaphase cells further localized the gene to bands 9q32-34, with the largest cluster of grains at 9q34.1.

Complement is a self-assembling biological system of plasma proteins which constitutes a major humoral defense system against bacterial and viral pathogens. The fifth component of complement (C5)[1] is of particular importance because, after activation, it participates in both inflammatory and cytolytic processes.

C5 is synthesized as an intracellular single-chain precursor, pro-C5, which is processed and secreted as a two-chain functionally active glycoprotein (Ooi & Colten, 1979; Patel & Minta, 1979). Native C5 is comprised of disulfide-linked α- and β-polypeptides with molecular weights of 115 000 ± 12 000 and 75 000 ± 8000, respectively (Nilsson et al., 1975; Tack et al., 1979). The native C5 molecule is cleaved during complement activation into two biologically active fragments, C5a and C5b. The larger macromolecular product, C5b, initiates the assembly of the C5b-9 complex which mediates membrane lysis of target cells [reviewed by Muller-Eberhard (1984)]. The smaller C5a fragment (C5a anaphylatoxin) is a cationic peptide derived from amino acid residues 1-74 of the C5 α-subunit (Fernandez & Hugli, 1978; Tack et al., 1979). C5a, the most potent of the complement-derived anaphylatoxins, expresses many important biological activities including (a) contraction of smooth muscle (Cochrane & Muller-Eberhard, 1968), (b) degranulation of mast cells (Johnson et al., 1975), (c) chemotaxis of polymorphonuclear neutrophils (Jensen et al., 1969), and secretion of granular enzymes from polymorphonuclear neutrophils (Goldstein & Weissman, 1974). Several physiologically relevant proteases also cleave C5, thereby generating C5a-like activities (Bronza et al., 1977; Hugli, 1977; Wetsel & Kolb, 1982, 1983).

Individuals affected by C5 deficiency lack bactericidal activity and have a severely impaired ability to induce chemotaxis (Miller & Nilsson, 1970; Rosenfeld et al., 1976; Nilsson et al., 1979). Genetic C5 deficiency is present in 39% of inbred mice strains, and sera of these deficient mice lack detectable C5 hemolytic activity and protein (Cinder & Dubiski, 1964; Cinder et al., 1964). It has been demonstrated that there are restriction fragment polymorphic differences within the C5-deficient structural gene which correlate with the protein deficiency (Wheat et al., 1987). The gene encoding mouse C5 has been mapped to the *Hc* locus on chromosome 2 (D'-Eustachio et al., 1986).

Extensive sequence studies of C3 (Lundwall et al., 1984; Wetsel et al., 1984; deBruijn & Fey, 1985), C4 (Belt et al.,

† Publication No. 4939-IMM from the Research Institute of Scripps Clinic. This work was supported by U.S. Public Health Service Grants AI19222 and AI22214 to B.F.T., GM35377 to R.S.L., and AI25011 to R.A.W., the University of Chicago Cancer Research Foundation (M.M.L.), the Washington University Biomedical Research Support Foundation (R.A.W.), and the American-Lebanese-Syrian Associated Charities of St. Jude Children's Research Hospital (R.S.L.). M.M.L. is a Special Fellow of the Leukemia Society of America. S.R.B. is the recipient of a National Research Service award (AI 07648) from the National Institutes of Health.

‡ Research Institute of Scripps Clinic.

§ Present address: Departments of Pediatrics and of Microbiology and Immunology, Washington University School of Medicine, St. Louis, MO 63110.

∥ St. Jude Children's Research Hospital.

⊥ University of Chicago.

[1] Abbreviations: C3, C4, and C5, third, fourth, and fifth components of complement; α_2M, α_2-macroglobulin; kb, kilobase(s); PBS, phosphate-buffered saline; DTT, dithiothreitol; HPLC, high-pressure liquid chromatography; HAT/O medium, hypoxanthine/aminopterin/thymidine/oubain medium.

0006-2960/88/0427-1474$01.50/0

1984, 1985; Sepich et al., 1985; Nonaka et al., 1985), C5 (Lundwall et al., 1985; Wetsel et al., 1987), and α_2M (Sottrup-Jensen et al., 1984b) have established that these proteins are homologous family members and share a common evolutionary origin. With the exception of C5, each of these proteins has a highly reactive internal β-cysteinyl-γ-glutamyl thiol ester bond (Janatova et al., 1980; Tack et al., 1980; Campbell et al., 1981; Harrison et al., 1981; Sottrup-Jensen et al., 1980). As demonstrated by sequencing of a mouse C5 cDNA, the cysteine and proximal glutamine which give rise to the intramolecular thiol ester bond in C3, C4, and α_2M are replaced by serine and alanine, respectively (Wetsel et al., 1987).

In this paper, we report the isolation and sequence of a large cDNA coding for approximately 80% of the human C5 promolecule. In addition, we report the isolation and partial characterization of a human C5 genomic fragment. Collectively, these sequence data allowed us to (1) compare the human and mouse C5 sequences, thereby evaluating the degree of phylogenetic conservation of this molecule's amino acid and mRNA structure, (2) compare this cDNA with that of an unusual C5 cDNA which contained Alu consensus sequence, and (3) exactly align the human C3, C4, C5, and α_2M protein sequences, thereby analyzing their evolutionary relationships. Furthermore, we employed the cDNA as a probe to determine the chromosomal location of the human C5 structural gene.

Materials and Methods

Isolation of the Human C5 cDNA Clone. A cDNA library of 150 000 recombinants was prepared from twice poly(A)-selected RNA employing the method of Okayama and Berg (1983). The RNA was isolated from cultured Hep G2 cells by the guanidinium isothiocyanate method (Davis et al., 1986). High-density screening of the library was performed by using duplicate filters and a ^{32}P nick-translated *Rsa*I restriction fragment derived from a partial human C5 clone (Lundwall et al., 1985). One clone hybridized in duplicate with the radiolabeled probe. The C5 positive recombinant (pC5HG2) was colony purified by subsequent rounds of screening, its plasmid purified by the alkaline lysis procedure (Maniatis et al., 1982), and its insert excised intact with the restriction enzyme *Xho*I. The C5 cDNA insert of approximately 4.0 kilobases was recovered from 1.0% agarose preparative gel by trough-elution (Maniatis et al., 1982).

Isolation of the Human C5 Genomic Clone. A human fetal liver genomic library, constructed in the λ Ch4A vector (Maniatis et al., 1978), was screened for clones containing C5 gene fragments. Of the 600 000 recombinants screened, 3 hybridized with the ^{32}P nick-translated human C5 *Rsa*I fragment. The positive recombinants were colony purified by two additional rounds of screening, and the phage particles containing the C5 genomic fragments were purified (Helms et al., 1985). The DNA (2 μg) isolated from each clone was digested with several restriction enzymes and characterized by Southern blot analysis. To map the arrangement of genomic fragments, Southern blots were probed with several ^{32}P nick-translated human C5 cDNA restriction fragments. A 1.3-kb genomic fragment which corresponded to the previously described (Lundwall et al., 1985) J-16 C5 clone's Alu sequence boundary was identified. This genomic fragment was isolated by trough-elution employing a 0.8% agarose preparative gel.

DNA Sequence Analysis. Both the cDNA and genomic sequences reported in this paper were obtained by the "shotgun" DNA sequencing strategy (Bankier & Barrell, 1983). Ten micrograms of the cDNA or genomic fragments was self-ligated, sheared randomly by sonication, end-repaired, and cloned into M13mp8. The M13 recombinants were sequenced by the dideoxy chain termination method (Sanger et al., 1977) employing ^{35}S label and gradient gels (Biggin et al., 1983). Each cDNA or genomic DNA strand was sequenced at least once, and on average, each cDNA and genomic base position was sequenced 4.8 and 3.2 times, respectively. All nucleotide data were complied and joined into a contiguous sequence by the Staden computer programs (Staden, 1982b).

Isolation and Sequence Determination of Tryptic Peptides of C5. Human C5 was purified by ion exchange (Hammer et al., 1981) and immunochemical (Wetsel et al., 1980) column chromatographic procedures as briefly outlined below. Partially purified C5, which was recovered after DEAE-Sephacel column chromatography (Hammer et al., 1981), was dialyzed against PBS and applied to a goat anti-human C5–Sepharose CL-4B column (Wetsel et al., 1980). The purified C5 was eluted from the column with 4 M guanidinium chloride. Twenty milligrams (100 nmol) of purified C5 was reduced with DTT and carboxymethylated with iodo[2-^{3}H]acetic acid (Sottrup-Jensen et al., 1984a) under fully denaturing conditions. After being desalted on a Sephadex G-25 column (equilibrated with 0.15 M ammonium bicarbonate) (pH 8.25), the protein was digested with 1% (w/w) *N*-tosylphenylalanine chloromethyl ketone treated bovine trypsin for 100 min at room temperature. Tryptic C5 peptides were then isolated by DEAE-Sephacel and reverse-phase high-pressure liquid chromatography (HPLC) as described in detail elsewhere (Kristensen et al., 1986). After amino acid analysis of 20 peptides employing a Beckman 6300 Amino acid analyzer (Beckman Instruments, Palo Alto, Ca), 4 peptides were selected and sequenced with a Beckman 890D sequencer. The 0.1 M Quadrol program (345801) was employed, and phenylthiohydantoin amino acids were analyzed by using an analytical Dupont Zorbax ODS C_{18} column (Zimmerman & Pisano, 1977).

Probe Preparation. The human C5 probe, employed in the somatic cell and in situ hybridizations, was prepared by digestion of 50 μg of pC5HG2 C5 insert with *Xmn*I. The 3398 (Figure 1) base pair fragment was isolated by trough-elution (Maniatis et al., 1982). The ABL probe was a 1-kb *Kpn*I restriction fragment of a human ABL genomic clone (gift of C. A. Westbrook, University of Chicago).

Somatic Cell Hybrids. A panel of 34 somatic cell hybrids was selected for mapping studies from a large number of hybrids created by poly(ethylene glycol)-mediated fusion of Syrian BHK-B1 or Chinese E36 hamster cells to human VA-2, A549, or IMR90 fibroblast cells. The hamster cells have mutant thymidine kinase or hypoxanthine phosphoribosyltransferase genes, respectively, while human cells cannot survive in 10 μM oubain, permitting the selection of hybrids in hypoxanthine/aminopterin/thymidine/oubain (HAT/O) medium. After initial selection, the hybrids were routinely cultured in HAT medium. The hybrids retain the entire rodent genome but have selectively lost different combinations of human chromosomes by segregation. The human chromosome composition of each hybrid was determined by screening for up to 34 gene enzyme systems (O'Brien et al., 1982), and in selected cases by complete cytogenetic analyses using trypsin–Giemsa banding (Lemons et al., 1978). High molecular weight DNA for Southern blots and cell homogenates for isozymes were prepared from the same passage of cells for each hybrid clone.

The high molecular weight DNA was digested to completion with *Eco*RI or *Bam*HI restriction endonucleases (New England Biolabs, Beverly, MA), separated by electrophoresis

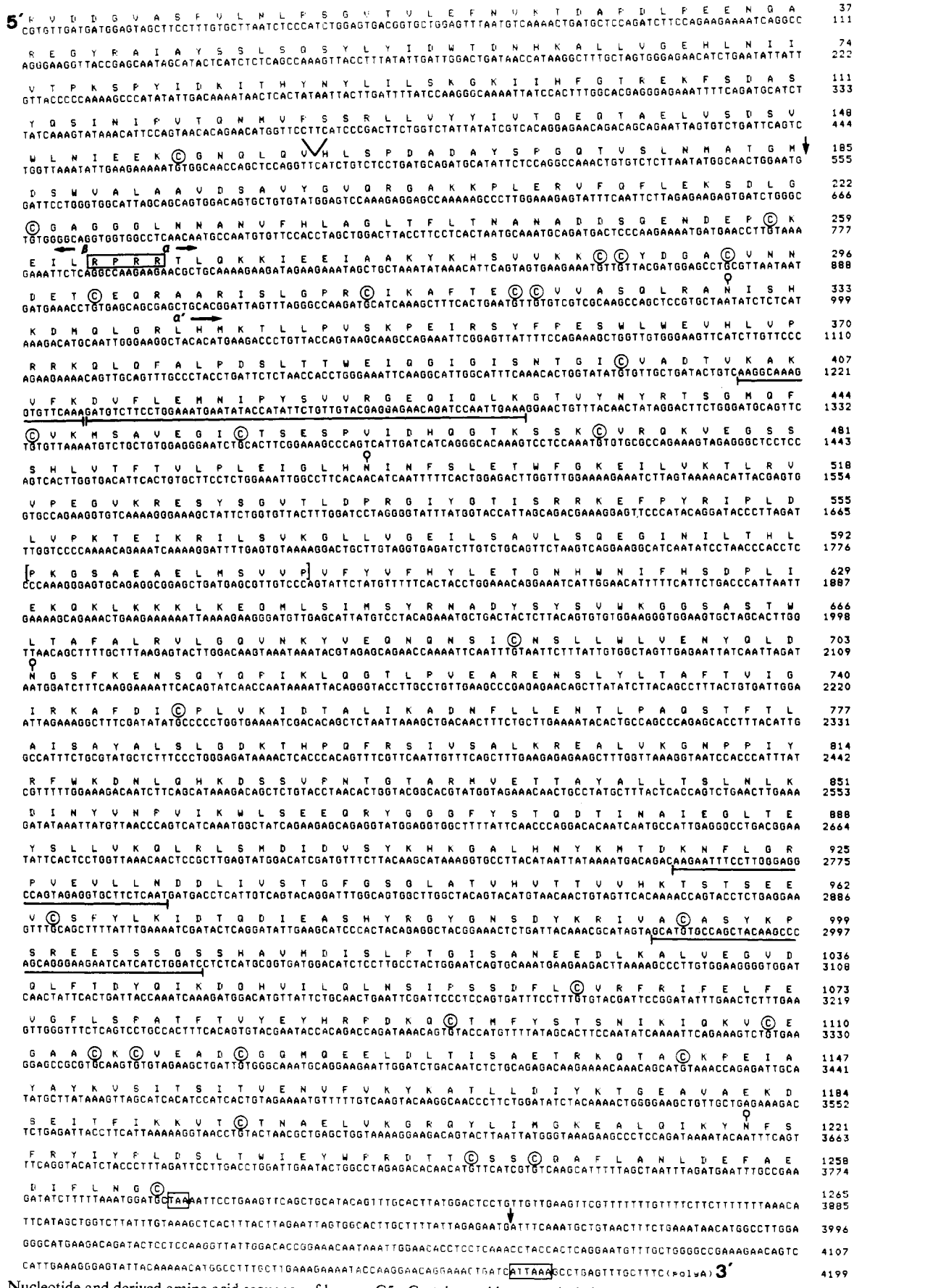


FIGURE 1: Nucleotide and derived amino acid sequence of human C5. Cysteine residues are circled. Potential N-linked carbohydrate attachment sites are indicated by the symbol (♀). The region which corresponds to the thiol ester domains of C3, C4, and α_2M is bracketed. Boxed sequences correspond to the β–α junction, the stop codon, and the putative polyadenylation recognition signal. The 5′ sequence which comes before the wedge symbol (V) was derived from the clone J-16; the remaining sequence was derived from the clone pC5HG2. The sequences corresponding to the human C5 tryptic peptides are depicted by lines above their amino acids. *Xmn*I cleavage sites are indicated by the downward arrows (↓).

through agarose gels, transferred by Southern blotting to nylon membranes, and hybridized with the C5 or ABL probes. The human probes were ^{32}P radiolabeled by the random oligonucleotide priming method (Feinberg & Vogelstein, 1984) to a specific activity of (1–3) × 10^9 dpm/μg. The C5 probe detects multiple human *Bam*HI or *Eco*RI restriction fragments distinct from those of hamster. The ABL probe detects a single *Eco*RI and two *Bam*HI fragments distinct from those of hamster.

In Situ Chromosomal Hybridization. Human metaphase cells prepared from phytohemagglutinin-stimulated peripheral blood lymphocytes were hybridized with a ^3H-labeled 3.4-kb *Xmn*I fragment of a C5 cDNA clone. Radiolabeled probes were prepared by nick translation of the purified fragment DNA with all four ^3H-labeled deoxynucleoside triphosphates to a specific activity of 1.5 × 10^8 dpm/μg. In situ hybridizations were performed as previously described (Le Beau et al., 1984). Metaphase cells were hybridized with 4.0 and 8.0 ng of probe/mL of hybridization mixture.

Results and Discussion

Isolation and Characterization of pC5HG2 cDNA. A cDNA library constructed from Hep G2 mRNA (Materials and Methods) was screened for clones containing C5 inserts. One C5 positive recombinant was obtained (pC5HG2) and, after restriction analysis, was determined to contain a cDNA insert of approximately 4.0 kb. The C5 cDNA was recovered intact after *Xho*I digestion.

Nucleotide and Derived Amino Acid Sequence of pC5HG2. The 4.0-kb C5 cDNA was fully sequenced by the "shotgun" dideoxy chain termination method (Bankier & Barrell, 1983). The pC5HG2 nucleotide and derived amino acid sequences are shown in Figure 1. Nucleotide sequence data indicated that the pC5HG2 insert was comprised of 3309 base pairs of pro-C5 coding sequence and 404 base pairs of 3′-nontranslated sequence. The nontranslated region began with the stop codon (TAA) and contained a putative polyadenylation signal (ATTAAA) located 16 base pairs upstream of a poly(A) tail of undetermined length. As was the case in earlier human (Lundwall et al., 1985) and murine (Wetsel et al., 1987) C5 cDNA sequences, the C5 cDNA obtained from Hep G2 cells displayed the β–α-chain orientation. The derived amino acid sequence indicated that the pC5HG2 cDNA contained the entire coding sequence for the C5 α-chain, the β–α-chain junction region, and 100 amino acids of the carboxy terminus of the β-chain. Additionally, the protein sequence indicated that the C5 molecule synthesized by Hep G2 cells is probably structurally identical with the C5 molecule obtained from human serum. This conclusion was based on two observations: (1) the sequences of four peptides, derived from trypsin digestion (Materials and Methods) of human C5 protein, were aligned exactly to the pC5HG2 amino acid sequence (see Figure 1); (2) the human C5a peptide sequence (Fernandez & Hugli, 1978) was identical with the corresponding C5a sequence derived from the Hep G2 C5 cDNA.

Carbohydrate has been reported to be contained within the C5a and C5α′ fragments (DiScipio et al., 1983). From our sequence data, four potential N-linked carbohydrate attachment sites were localized within the C5α chain. One was the previously determined site found within the C5a fragment at position 330. Three additional attachment sites were all localized within the C5 α-chain at positions 500, 704, and 1219 of Figure 1. Additionally, there were 26 cysteine residues found in the deduced human C5 α-chain amino acid sequence. Seven of these cysteines were clustered in the C5a fragment. Eleven cysteines were clustered in the final 203 carboxy-terminal amino acids. Of the 262 amino acids thus far determined for the human C5 β-chain, there were 3 cysteine residues found and no N-linked carbohydrate binding sites.

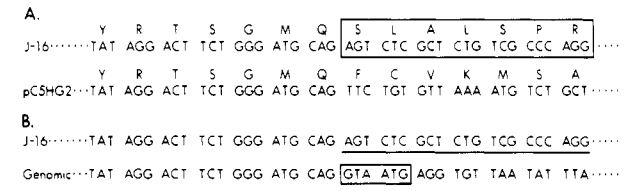


FIGURE 2: (A) Alignment of J-16 and pC5HG2 nucleotide and derived amino acid sequences. Nucleotides and amino acid residues at the point of divergence between the two clones are shown. The Alu sequence which gives rise to the divergence between the two clones is boxed. The amino acid sequence shown here for pC5HG2 corresponds to amino acids 437–450 of Figure 1. (B) Comparison of J-16 and C5 genomic nucleotide sequences. The Alu sequence in J-16 is underlined. The 5′ intron consensus sequence is boxed.

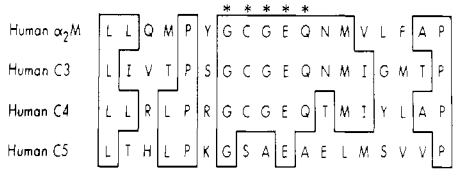
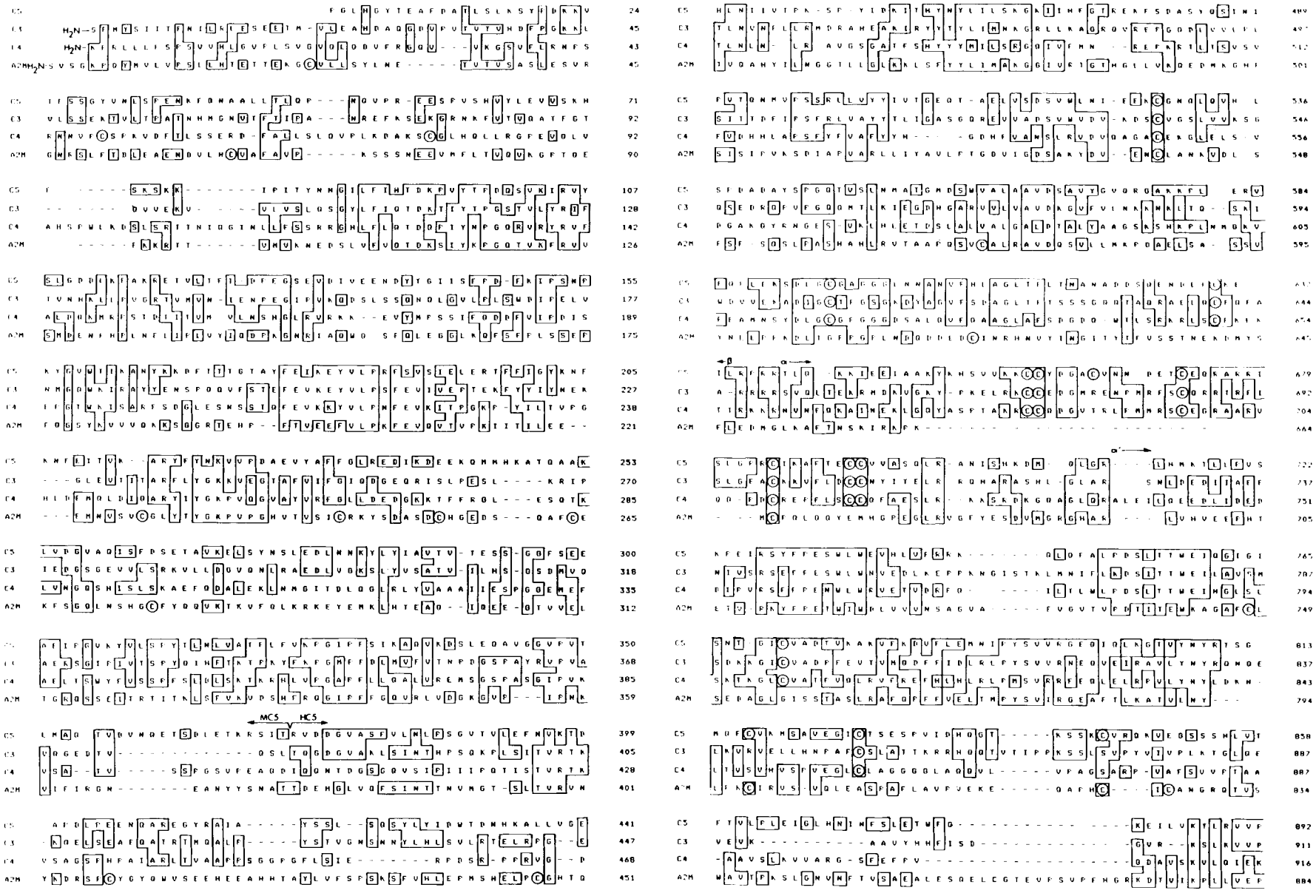


FIGURE 3: Thiol ester domains of human α_2M, C3, and C4 aligned with the corresponding region in human C5. Amino acids which are identical in at least two of the protein sequences are boxed. The central five amino acids highly conserved in thiol ester containing proteins are indicated by the asterisks.

Comparison of Human C5 Sequences Obtained from J-16 and pC5HG2 Clones. A partial human C5 cDNA clone (J-16), obtained from an acute-phase liver library, has been characterized previously (Lundwall et al., 1985). Although this clone contained approximately 1320 nucleotides of coding sequence for human C5, it also contained approximately 150 base pairs of human consensus Alu nucleotide sequence inserted immediately after the C5 coding sequence and right before the stop codon, TGA. Alignment of the J-16 and pC5HG2 cDNA sequences (Figure 2A) reveals that the actual C5 coding sequence of J-16 ended at the glutamine residue (immediately before the boxed residues in Figure 2A) rather than the serine three residues upstream as previously thought (Lundwall et al., 1985). Past the location where the C5 coding sequence of the J-16 clone ends, the cDNA sequences of the two clones completely diverge. That is to say, pC5HG2 contains no comparable Alu consensus sequence, and, in addition, the stop codons, 3′-untranslated sequences, and polyadenylation signals of these two clones are completely different.

A human C5 genomic clone has recently been isolated which contains a DNA sequence that corresponds to the C5 coding region preceding the Alu consensus sequence in the C5 J-16 clone. DNA sequence analysis revealed that at exactly the nucleotide position where the C5 coding sequence stops in the J-16 clone an intron sequence begins in the human C5 gene (Figure 2B). However, this intron does not contain the Alu consensus sequence present in the J-16 clone. This finding indicates that the Alu sequence and the following 3′-untranslated region present in the J-16 clone are not derived by incomplete splicing of the adjacent intron. Instead, the Alu and 3′ region of the J-16 clone must be obtained from a more distant intron in the C5 structural gene, or from another C5 gene or from an unrelated gene. Isolation of the complete C5 structural gene and subsequent Southern analysis and sequencing should indicate if this Alu sequence is C5 derived.

Comparison of the Murine and C5 Sequence. A large murine cDNA clone (5217 base pairs) which codes for pro-C5


α
β
α'
MC5
HC5

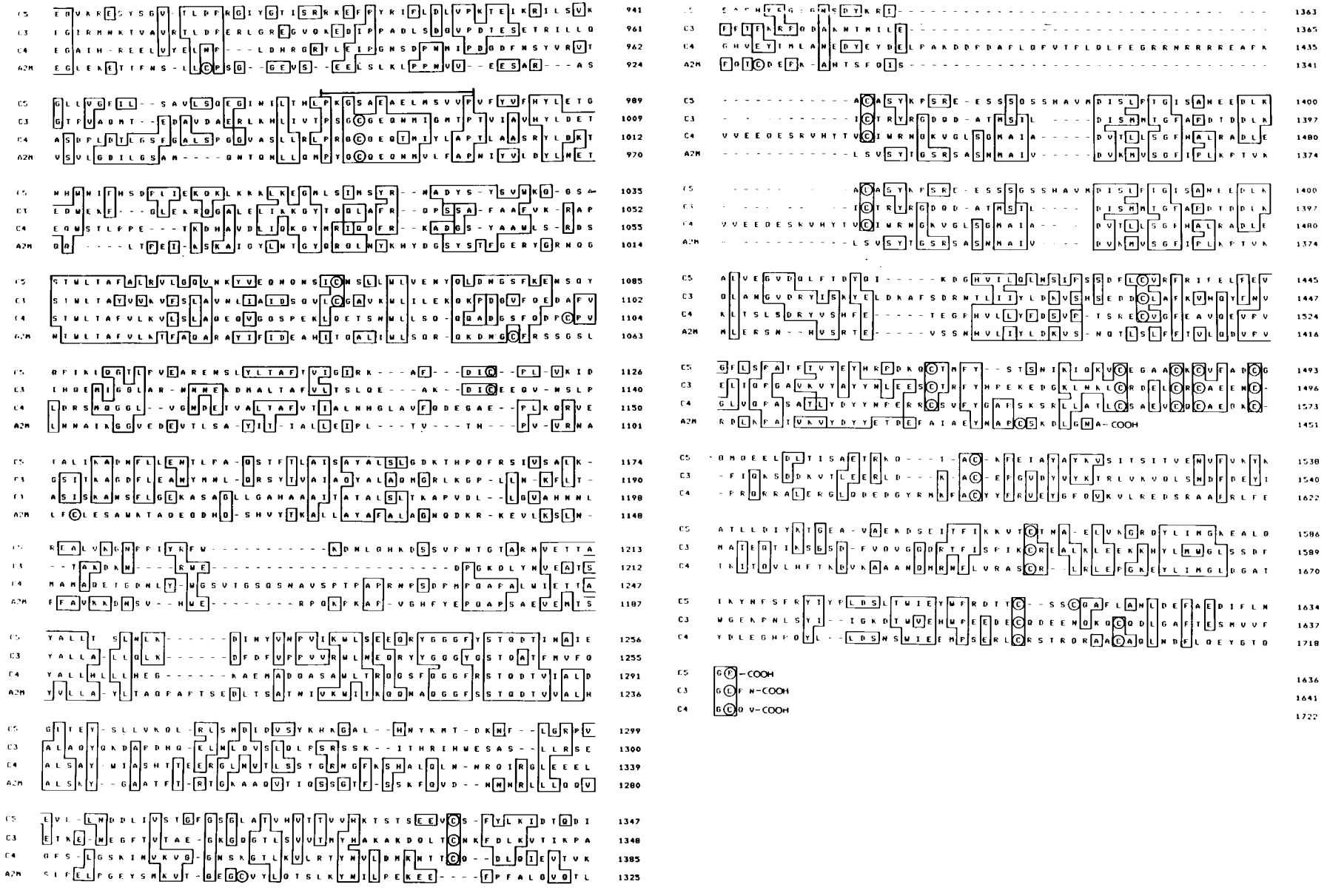


FIGURE 4: Alignment of the amino acid sequences of human C5, C3, C4, and α_2-macroglobulin. The amino acid sequences of pro-C3, pro-C4, α_2M, and pro-C5 were aligned fist by comparing 150 amino acid stretches using the computer program BESTFIT (University of Wisconsin) which aligns sequences on the basis of identical residues. The homologies were then maximized by hand by using the Dayhoff scoring matrix of similar amino acids (Staden, 1982a). Gaps of varying size were introduced to help strengthen the homologies. Where indicated, the amino-terminal sequence of pro-C5 was aligned by employing the mouse pro-C5 sequence. Identical residues in any pair of compared sequences are boxed. The thiol ester domains are indicated by the bold line above the sequences, and the cysteine residues are circled.

has recently been isolated and fully sequenced (Wetsel et al., 1987). Coding sequence alignment of the murine C5 clone with the human pC5HG2 C5 clone reveals 79% identity at the protein level and 80% identity at the nucleotide level. Of the 759 nucleotide changes in the pC5HG2 clone, 159 are silent substitutions. In other words, 21% of the substituted nucleotides which occur in the human C5 mRNA result in no amino acid substitutions in the human C5 protein sequence.

The mouse and human C5 amino acid sequences each contained 29 cysteine residues, 3 in each β-chain and 26 in each α-chain. All of the cysteines remain identically located in these two species except at two positions. The cysteine at position 293 of the human C5 sequence (Figure 1) is replaced by an arginine residue in the mouse sequence. Likewise, the tyrosine at position 700 of the human C5 sequence is substituted by a cysteine in the mouse sequence.

It has been postulated that a given protein will evolve at a rate determined by the functional constrains placed on its primary structure (Dickerson, 1971). For example, fibrinopeptides have no specific function other than being released as degradation fragments from fibrinogen during blood clotting. Therefore, there are no structural constraints put on these peptides, and consequently they exhibit the fastest known rate of evolution. On the basis of the 79% identity of its murine and human primary sequences, C5 would be expected to evolve at a moderate rate. As anticipated, C5 has evolved slower than, for example, albumin which has 68% amino acid identity between the mouse and human amino acid sequences (Dugaiczyk et al., 1982; Minghetti et al., 1985). However, C5 has evolved much faster than cytochrome *c* which has a 96% mouse/human primary sequence identity.

Interestingly, the portion of the mouse and human C5 sequences which exhibits the greatest degree of divergence is the C5a fragment. Even when the three additional amino acids contained at the amino terminus of the mouse C5a fragment are excluded from the calculations, the two C5a sequences share only 64% identically placed residues (36/74). This observation would suggest that only portions of the C5a sequence are necessary to mediate its inflammatory activities.

Evolutionary Relationships among C3, C4, C5, and α_2-Macroglobulin. It has been established from previous sequence comparisons that the serum protease inhibitor α_2-macroglobulin shares a common evolutionary origin with the complement components C3 and C4 (Sottrup-Jensen et al., 1985). All three proteins harbor a highly reactive β-cysteinyl-γ-glutamyl thiol ester enabling each protein to bind covalently to surfaces at amino and hydroxyl groups. Recently, sequence determination of a mouse C5 cDNA clone has demonstrated definitively that the fifth complement component lacks this thiol ester moiety but otherwise shares extensive homology with C3, C4, and α_2M (Wetsel et al., 1987). Alignment of the protein sequences of these family members indicated that the corresponding thiol ester region of C5 contained several of the conserved amino acids. However, the critical cysteine and glutamine residues, which bridge to form the thiol ester, had been replaced by serine and alanine, respectively. From our current study, these exact two amino acid substitutions were found in human C5 (Figure 3).

When the complete primary structures of human C3, C4, and α_2M are compared with the primary structure of C5 (Figure 4), it is clear that C5 shares, in addition to the corresponding thiol ester domain, extensive sequence homology with its family members. As expected, C5 displayed more sequence homology with C3 and C4 than with α_2M. Of the 1639 amino acids of C5 compared, 27% were identical with human C3, while 25% were identical with human C4. Of the 1478 amino acids (1–1478) of C5 compared to human α_2M, 19% of the residues were identical.

Table I: Synteny Test of the C5 Gene and Human Chromosomes in Rodent X Human Hybrid Clones[a]

human chromosome	C5 gene/human chromosome				
	+/+	+/−	−/+	−/−	asynteny
1	11	8	4	5	43
2	5	12	3	6	58
3	1	9	1	0	91
4	5	5	2	0	58
5	10	7	4	3	46
6	11	7	7	2	52
7	0	9	0	1	90
8	10	8	5	4	48
9	25	0	0	9	0
10	5	12	0	6	52
11	14	2	7	0	39
12	8	8	6	1	61
13	13	5	8	0	50
14	13	5	7	2	44
15	8	9	3	5	60
16	13	4	6	3	38
17	19	6	9	0	44
18	7	9	4	3	56
19	11	8	5	3	48
20	10	8	5	6	41
21	2	15	0	9	58
22	0	12	0	1	92
X	5	5	5	4	53

[a] Somatic cell hybrids were scored for the presence (+) or absence (−) of specific human chromosomes and for the presence or absence of human C5 coding sequences as described under Materials and Methods.

The complement component family members displayed high conservation in the distribution of their cysteine residues. As shown in Figure 4, there are 11 conserved cysteines clustered around the carboxy termini of the β-chains and the amino termini of the α-chains of these complement promolecules. In addition, there are 11 completely conserved cysteines in the carboxy termini of the α-chains of C5 and C3 and within the γ-chain of C4. In contrast, the distribution of the cysteines of α_2M is not conserved with the complement family members. Only two of the α_2M cysteines (positions 540 and 666) are identically placed within the C3, C4, and C5 primary structures. Interestingly, the least conserved area of α_2M resides within the anaphylatoxin sequences of the complement components. Thirty-five amino acids present in the C3a, C4a, and C5a domains have been deleted in the corresponding region of α_2M.

Chromosome Localization. Southern blot analyses of *Bam*HI- or *Eco*RI-digested DNA from the hybrid panel revealed 100% concordance between human C5 sequences and human chromosome 9. All other human chromosomes showed a discordancy of 38–92% with C5 sequences (Table I). As an internal control to confirm this assignment, the Southern blots were washed to background and rehybridized with the ABL probe previously assigned to human chromosome 9q34 (Heisterkamp et al., 1982). All 34 clones showed 100% concordance between C5 and ABL sequences and human chromosome 9.

The murine C5 gene has been mapped to the proximal portion of mouse chromosome 2 (D'Eustachio et al., 1986). Comparative mapping studies have revealed several loci on this portion of mouse chromosome 2, C5 now being identified as a member of this linkage group, whose human analogues are located on human chromosome 9 (Lalley & McKusick, 1985). Thus, the C5 gene is highly conserved as it displays

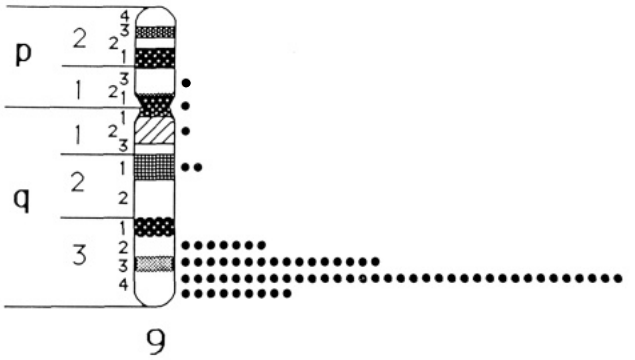


FIGURE 5: Distribution of labeled sites on chromosome 9 in 100 normal metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized with the C5 cDNA clone. The labeled sites observed in this hybridization were clustered at 9q32–34; the largest cluster of grains was located at 9q34.1.

both sequence and chromosome (linkage) homology between mouse and man.

To sublocalize the C5 gene on chromosome 9, we performed in situ chromosomal hybridization to normal human metaphase cells using a cDNA clone specific for this gene. This hybridization resulted in specific labeling of chromosome 9. The hybridization was repeated twice and gave identical results. In hybridizations performed with the cDNA probe, we observed specific labeling of the distal long arm of chromosome 9. Of 100 metaphase cells examined from this hybridization, 49 (49%) were labeled on region q3, bands q32–34 of one or both chromosome 9 homologues ($P < 0.0005$). The distribution of labeled sites on chromosome 9 is illustrated in Figure 5. A total of 72 grains were observed on this chromosome; of these, 67 (93%) were clustered at bands q32–34 and represented 40% (67/167) of all labeled sites. The largest cluster of grains was observed at 9q34.1.

Thus, the C5 gene is localized to chromosome 9, at bands q32–34. As discussed above, C3, C4, C5, and α_2M have amino acid sequence homology indicating that these proteins are homologous family members and share a common evolutionary origin. The genes encoding C3, C4, and α_2M have been localized to 19p13.2–p13.3 (Whitehead et al., 1982; Pericak-Vance et al., 1985), 6p21.3 (Carroll et al., 1984), and chromosome 12 (Kan et al., 1985), respectively, indicating that these genes are dispersed to different human chromosomes. Our localization of the C5 gene to 9q32–34 indicates that the C5 gene also maps to a distinct chromosomal region.

ACKNOWLEDGMENTS

We thank Drs. Torsten Kristensen and Lars Sottrup-Jensen for advice in sequencing the C5 peptides and Dr. Dan Noonan for assistance in construction of the Hep G2 cDNA library. In addition, we thank Sandra P. West and Rafael Espinosa, III, for excellent technical assistance and Barb Pellerito and Bonnie Towle for typing the manuscript.

REFERENCES

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

Belt, K. T., Carroll, M. C., & Porter, R. R. (1984) *Cell (Cambridge, Mass.) 36*, 907–914.

Belt, K. T., Yu, C. Y., Carroll, M. C., & Porter, R. R. (1985) *Immunogenetics 21*, 173–180.

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

Bronza, J., Senior, R. M., & Kreutzer, D. L. (1977) *J. Clin. Invest. 60*, 1280–1288.

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

Carroll, M. C., Campbell, R. D., Bentley, D. R., & Porter, R. R. (1984) *Nature (London) 307*, 237–241.

Cinder, B., & Dubiski, S. (1964) *Nature (London) 202*, 102–103.

Cinder, B., Dubiski, S., & Wardlaw, A. C. (1964) *J. Exp. Med. 120*, 897–924.

Cochrane, C. G., & Muller-Eberhard, H. J. (1968) *J. Exp. Med. 127*, 371–386.

Davis, L. G., Dibner, M. D., & Battey, J. F. (1986) in *Basic Methods in Molecular Biology* (Davis, L. G., Dibner, M. D., & Battey, J. F., Eds.) pp 1–388, Elsevier, New York.

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

D'Eustachio, P., Kristensen, T., Wetsel, R. A., Riblet, R., Taylor, B. A., & Tack, B. F. (1986) *J. Immunol. 137*, 3990–3995.

Dickerson, R. E. (1971) *J. Mol. Evol. 1*, 26–45.

DiScipio, R. G., Smith, C. A., Muller-Eberhard, H. J., & Hugli, T. E. (1983) *J. Biol. Chem. 258*, 10629–10636.

Duzaiczyk, A. S., Law, S. W., & Dennison, O. E. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 71–75.

Feinberg, A., & Vogelstein, B. (1984) *Anal. Biochem. 137*, 266–267.

Fernandez, H. N., & Hugli, T. E. (1978) *J. Biol. Chem. 253*, 6955–6962.

Goldstein, I. M., & Weissman, G. (1974) *J. Immunol. 113*, 1583–1588.

Hammer, C. H., Wirtz, G. H., Renfer, L., Gresham, H. D., & Tack, B. F. (1981) *J. Biol. Chem. 256*, 3995–4006.

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

Heisterkamp, N., Groffen, J., Stephenson, J. R., Spurr, N. K., Goodfellow, P. N., Solomon, E., Carritt, B., & Bodmer, W. F. (1982) *Nature (London) 299*, 747–767.

Helms, C., Grahm, M. V., Dutchik, J. E., & Olson, M. V. (1985) *DNA 4*, 39–49.

Hugli, T. E. (1977) in *Chemistry and Biology of Thrombin* (Lundblad, R. L., Fenton, J. W., & Mann, K. G., Eds.) pp 345–360, Ann Arbor Science, Ann Arbor, MI.

Janatova, J., Lorenz, P. E., Schechter, A. N., Prahl, J. W., & Tack, B. F. (1980) *Biochemistry 19*, 4471–4478.

Jensen, J. S., Snyderman, R., & Mergenhagen, S. E. (1969) in *Cellular and Humoral Mechanisms in Anaphylaxis and Allergy* (Movat, H. Z., Ed.) pp 265–278, S. Karger, Basel.

Johnson, A. R., Hugli, T. E., & Muller-Eberhard, H. J. (1975) *Immunology 28*, 1067–1080.

Kan, C. C., Solomon, E., Belt, K. T., Chain, A. C., Hiorns, L. R., & Fey, G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2282–2286.

Kristensen, T., Wetsel, R. A., & Tack, B. F. (1986) *J. Immunol. 136*, 3407–3411.

Lalley, P. A., & McKusick, V. A. (1985) *Cytogenet. Cell Genet. 40*, 536–566.

LeBeau, M. M., Westbrook, C. A., Diaz, M. O., & Rowley, J. D. (1984) *Nature (London) 312*, 70–71.

Lemons, R. S., Nash, W. G. O'Brien, S. J., Benveniste, R. E., & Sherr, C. J. (1978) *Cell (Cambridge, Mass.) 14*, 995–1005.

Lundwall, A., Wetsel, R. A., Domdey, H., Tack, B. F., & Fey, G. H. (1984) *J. Biol. Chem. 259*, 13851–13856.

Lundwall, A., Wetsel, R. A., Kristensen, T., Whitehead, A. S., Woods, D. E., Ogden, R. C., Colten, H. R., & Tack, B. F. (1985) *J. Biol. Chem. 260*, 2108–2112.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–701.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, a Laboratory Manual* (Maniatis, T., Fritsch, E. F., & Sambrook, J., Eds.) pp 1–545, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Miller, M. E., & Nilsson, U. R. (1970) *N. Engl. J. Med. 282*, 354–358.

Minghetti, P. P., Law, S. W., & Dugaicyzk, A. S. (1985) *Mol. Biol. Evol. 2*, 347–358.

Muller-Eberhard, H. J. (1984) *Springer Semin. Immunopathol. 7*, 93–141.

Nilsson, U. R., Mandle, R. J., & McConnel-Mapes, J. A. (1975) *J. Immunol. 114*, 815–822.

Nilsson, U. R., Miller, M. E., & Wyman, S. (1979) *J. Immunol. 112*, 1164–1176.

Nonaka, M., Nakayama, K., Yeul, Y. D., & Takahashi, M. (1985) *J. Biol. Chem. 260*, 10936–10943.

O'Brien, S. J., Simonson, J. M., & Eichelberger, M. (1982) in *Techniques in Somatic Cell Genetics* (Shay, J. W., Ed.) pp 342–370, Plenum, New York.

Okayama, H., & Berg, P. (1983) *Mol. Cell. Biol. 3*, 280–289.

Ooi, Y. M., & Colten, H. R. (1979) *Nature (London) 282*, 207–208.

Patel, F., & Minta, J. O. (1979) *J. Immunol. 123*, 2494–2498.

Pericak-Vance, M. A., Yamaoka, L., Assinder, R., Bartlett, R. J., Ross, D. A., Fey, G., Humphries, S., Williamson, R., Conneally, P. M., & Roses, A. D. (1985) *Cytogenet. Cell Genet. 40*, 721–722.

Rosenfeld, S. I., Baum, J., Steigbigil, R. T., & Leddy, J. P. (1976) *J. Clin. Invest. 57*, 1635–1643.

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

Sepich, D. S., Noonan, D. J., & Ogata, R. T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5895–5899.

Sottrup-Jensen, L., Petersen, T. E., & Magnusson, S. (1980) *FEBS Lett. 121*, 275–279.

Sottrup-Jensen, L., Stepanik, T. M., Jones, C. M., Longblad, P. B., Kristensen, T., & Wierzbicki, D. M. (1984a) *J. Biol. Chem. 259*, 8293–8298.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Wierzbicki, D. M., Jones, C. M., Magnusson, S., & Petersen, P. E. (1984b) *J. Biol. Chem. 259*, 8318–8327.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Longblad, P. B., Jones, C. M., Wierzbicki, D. M., Magnusson, S., Domdey, H., Wetsel, R. A., Lundwall, A., Tack, B. F., & Fey, G. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 9–13.

Staden, R. (1982a) *Nucleic Acids Res. 10*, 2951–2961.

Staden, R. (1982b) *Nucleic Acids Res. 10*, 4731–4751.

Tack, B. F., Morris, S. C., & Prahl, J. W. (1979) *Biochemistry 18*, 1490–1497.

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

Wetsel, R. A., & Kolb, W. P. (1982) *J. Immunol. 128*, 2209–2216.

Wetsel, R. A., & Kolb, W. P. (1983) *J. Exp. Med. 157*, 2029–2048

Wetsel, R. A., Jones, M. A., & Kolb, W. P. (1980) *J. Immunol. Methods 35*, 319–335.

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

Wetsel, R. A., Ogata, R. T., & Tack, B. F. (1987) *Biochemistry 26*, 737–743.

Wheat, W. H., Wetsel, R. A., Falus, A., Tack, B. F., & Strunk, R. C. (1987) *J. Exp. Med. 165*, 1731–1736.

Whitehead, A. S., Solomon, E., Chambers, S., Bodmer, W. F., Povey, S., & Fey, G. H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5021–5025.

Zimmerman, C. L., & Pisano, J. J. (1977) *Methods Enzymol. 47*, 45–55.