Primary Structure of the Fifth Component of Murine Complement†

Rick A. Wetsel,‡,§ Ronald T. Ogata,‖ and Brian F. Tack*,‡

Departments of Immunology and Molecular Biology, Scripps Clinic and Research Foundation, La Jolla, California 92037

Received August 1, 1986; Revised Manuscript Received October 10, 1986

ABSTRACT: A cDNA library was constructed by the method of Okayama and Berg, [Okayama, H., & Berg, P. (1983) *Mol. Cell. Biol. 3*, 280–289] employing size-selected (>28 S) poly(A+) liver RNA from the mouse strain B10.WR. A total of 150000 recombinants were screened with a partial human C5 cDNA probe; 16 C5-positive clones were identified, 1 of which contained an insert greater than 5.2 kilobase pairs in length. This cDNA insert was fully sequenced by the dideoxy method. The DNA sequence of this insert had an open reading frame of 4920 base pairs specifying a sequence of 1640 amino acid residues. The region corresponding to positions 372–812 exhibited high homology with the previously determined partial structure for human C5 of 438 amino acid residues. A four-residue basic sequence (Arg-Ser-Lys-Arg) was identified upstream of the amino-terminal Asn of C5a, thereby specifying a $\beta\alpha$-chain orientation for the promolecule form of murine C5. The 3′ end of this clone contained 351 base pairs of untranslated sequence. The presumed polyadenylation recognition site CATAAA was located 17 base pairs upstream of the poly(A) tail. Comparison of the derived murine C5 sequence with previously determined structures for murine C3 and C4 revealed regions of high sequence similarity, including the thiol ester region present in C3 and C4. The cysteine and proximal glutamine which give rise to the intramolecular thiol ester bond in C3 and C4 were absent in C5, having been replaced by serine and alanine, respectively. Three amino acids (Asn, Leu, and His) not present in human, porcine, or bovine C5a were further localized at the amino terminus of murine C5a.

The fifth component of complement (C5)[1] is an important participant in inflammatory and cytolytic processes. Sera from C5-deficient individuals lack bactericidal activity and have a severely impaired ability to induce chemotaxis (Miller & Nilsson, 1970; Rosenfeld et al., 1976; Nilsson et al., 1979). C5 is synthesized in hepatocytes (Patel & Minta, 1979) and macrophages (Ooi & Colten, 1979) as an intracellular single-chain precursor, pro-C5. Pro-C5 is processed and secreted as a two-chain glycoprotein comprised of disulfide-linked α- and β-polypeptides with molecular weights of 115000 + 12000 and 75000 + 8000, respectively (Nilsson & Mapes, 1973; Tack et al., 1979). It has been previously established from cDNA sequencing studies that human pro-C5 is synthesized with a $\beta\alpha$-chain orientation (Lundwall et al., 1985).

On complement activation, C5 is cleaved by the classical (Goldlust et al., 1974) or alternative (Schreiber et al., 1978) pathway enzyme complex (C4b2a3b or C3bBbP, respectively) into two important biologically active derivatives, C5a and C5b. The larger macromolecular product, C5b, initiates the assembly of the C5b–9 complex which mediates membrane lysis of target cells [reviewed in Müller-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). The amino acid sequences of human (Fernandez & Hugli, 1978), porcine (Gerard & Hugli, 1980), bovine (Gennaro et al., 1986), and rat (Cui et al., 1985) C5a have been determined. C5a, the most potent of the complement-derived anaphylatoxins, expresses several important biological activities [reviewed by Hugli (1984)] including (a) contraction of smooth muscle (Cochrane & Müller-Eberhard, 1968), (b) degranulation of mast cells (Johnson et al., 1975), (c) chemotaxis of polymorphonuclear neutrophils (Jensen et al., 1969), and (d) secretion of granular enzymes from polymorphonuclear neutrophils (Goldstein & Weissmann, 1974). A number of physiologically relevant proteolytic enzymes, e.g., trypsin, α-thrombin, and elastase, also cleave C5, thereby generating C5a-like activities (Bronza et al., 1977; Hugli, 1977; Wetsel & Kolb, 1982, 1983).

Comparison of the complete primary structures of C3 (Lundwall et al., 1984; Wetsel et al., 1984; deBruijn & Fey, 1985), C4 (Belt et al., 1984; Sepich et al., 1985; Nonaka et al., 1985), and α_2-macroglobulin (α_2M) (Sottrup-Jensen et al., 1984) together with the partial human C5 sequence (Lundwall et al., 1985) has established that all four proteins share extended regions of similar sequence, indicating that these proteins have a common evolutionary origin [Sottrup-Jensen et al., 1986; reviewed in Sottrup-Jensen (1986)]. Interestingly, earlier studies which indicated the presence of an internal β-cysteinyl-γ-glutamyl thiol ester bond in C3 (Janatova et al., 1980; Tack et al., 1980), C4 (Campbell et al., 1981; Harrison et al., 1981), and α_2M (Sottrup-Jensen et al., 1980; Salveson et al., 1981), also indicated the absence of this site in C5 (Law et al., 1980; Janatova & Tack, 1981; DiScipio, 1981). However, the corresponding region of the C5 molecule was not included in the above partial human C5 sequence (Lundwall et al., 1985).

Despite their evolution from a common ancestor, the genes for this family of proteins now reside for the most part on

† Publication No. 4389-IMM from the Research Institute of Scripps Clinic. This work was supported by U.S. Public Health Service Grants AI 19222 and AI 22214 to B.F.T. and GM 29831 to R.T.O. R.A.W. was supported by NIH Training Grant HL 07195.

* Correspondence should be addressed to this author.

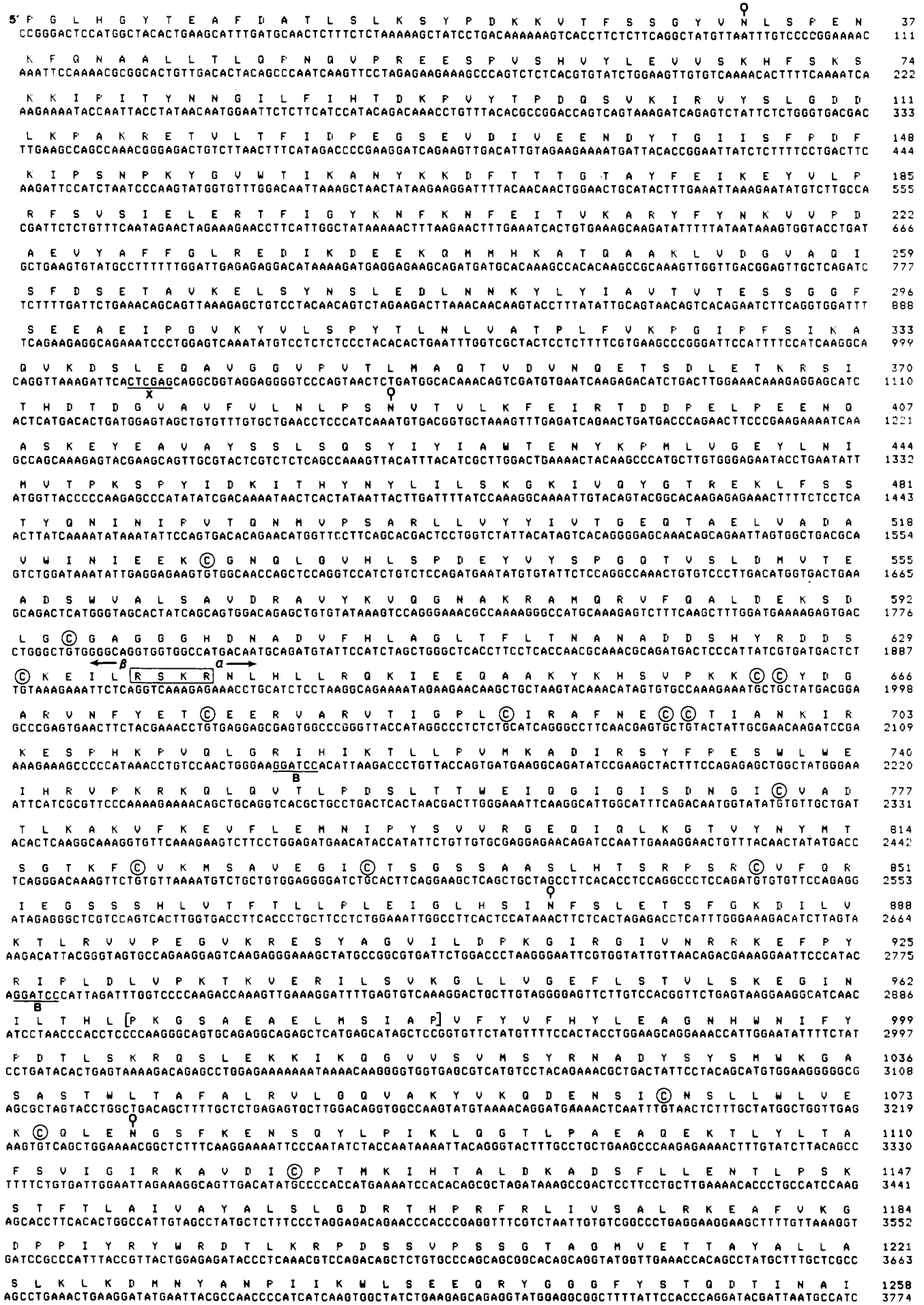
‡ Department of Immunology.

§ Present address: Department of Pediatrics, Children's Hospital, Washington University School of Medicine, St. Louis, MO 63178.

‖ Department of Molecular Biology.

[1] Abbreviations: C3, C4, and C5, third, fourth, and fifth components of complement; Slp, sex-limited protein; α_2M, α_2-macroglobulin; kb, kilobase(s).

0006-2960/87/0426-0737$01.50/0


5' P G L H G Y T E A F D A T L S L K S Y P D K K V T F S S G Y V N L S P E N 37
CCGGGACTCCATGGCTACACTGAAGCATTTGATGCAACTCTTTCTCTAAAAAGCTATCCTGACAAAAAAGTCACCTTCTCTTCAGGCTATGTTAATTTGTCCCCGGAAAAC 111
K F Q N A A L L T L Q P N Q V P R E E S P V S H V Y L E V V S K H F S K S 74
AAATTCCAAAACGCGGCACTGTTGACACTACAGCCCAATCAAGTTCCTAGAGAAGAAAGCCCAGTCTCTCACGTGTATCTGGAAGTTGTGTCAAAACACTTTTCAAAATCA 222
K K I P I T Y N N G I L F I H T D K P V Y T P D Q S V K I R V Y S L G D D 111
AAGAAAATACCAATTACCTATAACAATGGAATTCTCTTCATCCATACAGACAAACCTGTTTACACGCCGGACCAGTCAGTAAAGATCAGAGTCTATTCTCTGGGTGACGAC 333
L K P A K R E T V L T F I D P E G S E V D I V E E N D Y T G I I S F P D F 148
TTGAAGCCAGCCAAACGGGAGACTGTCTTAACTTTCATAGACCCCGAAGGATCAGAAGTTGACATTGTAGAAGAAAATGATTACACCGGAATTATCTCTTTTCCTGACTTC 444
K I P S N P K Y G V W T I K A N Y K K D F T T T G T A Y F E I K E Y V L P 185
AAGATTCCATCTAATCCCAAGTATGGTGTTTGGACAATTAAAGCTAACTATAAGAAGGATTTTACAACAACTGGAACTGCATACTTTGAAATTAAAGAATATGTCTTGCCA 555
R F S V S I E L E R T F I G Y K N F K N F E I T V K A R Y F Y N K V V P D 222
CGATTCTCTGTTTCAATAGAACTAGAAAGAACCTTCATTGGCTATAAAAACTTTAAGAACTTTGAAATCACTGTGAAAGCAAGATATTTTTATAATAAAGTGGTACCTGAT 666
A E V Y A F F G L R E D I K D E E K Q M M H K A T Q A A K L V D G V A Q I 259
GCTGAAGTGTATGCCTTTTTTGGATTGAGAGGAGACATAAAAGATGAGGAGAAGCAGATGATGCACAAAGCCACACAAGCCGCAAAGTTGGTTGACGGAGTTGCTCAGATC 777
S F D S E T A V K E L S Y N S L E D L N N K Y L Y I A V T V T E S S G G F 296
TCTTTTGATTCTGAAACAGCAGTTAAAGAGCTGTCCTACAACAGTCTAGAAGACTTAAACAACAAGTACCTTTATATTGCAGTAACAGTCACAGAATCTTCAGGTGGATTT 888
S E E A E I P G V K Y V L S P Y T L N L V A T P L F V K P G I P F S I K A 333
TCAGAAGAGGCAGAAATCCCTGGAGTCAAATATGTCCTCTCTCCCTACACACTGAATTTGGTCGCTACTCCTCTTTTCGTGAAGCCCGGGATTCCATTTTCCATCAAGGCA 999
Q V K D S L E Q A V G G V P V T L M A Q T V D V N Q E T S D L E T K R S I 370
CAGGTTAAAGATTCACTCGAGCAGGCGGTAGGAGGGGTCCCAGTAACTCTGATGGCACAAACAGTCGATGTGAATCAAGAGACATCTGACTTGGAAACAAAGAGGAGCATC 1110
T H D T D G V A V F V L N L P S N V T V L K F E I R T D D P E L P E E N Q 407
ACTCATGACACTGATGGAGTAGCTGTGTTTGTGCTGAACCTCCCATCAAATGTGACGGTGCTAAAGTTTGAGATCAGAACTGATGACCCAGAACTTCCCGAAGAAAATCAA 1221
A S K E Y E A V A Y S S L S Q S Y I Y I A W T E N Y K P M L V G E Y L N I 444
GCCAGCAAAGAGTACGAAGCAGTTGCGTACTCGTCTCTCAGCCAAAGTTACATTTACATCGCTTGGACTGAAAACTACAAGCCCATGCTTGTGGGAGAATACCTGAATATT 1332
M V T P K S P Y I D K I T H Y N Y L I L S K G K I V Q Y G T R E K L F S S 481
ATGGTTACCCCCAAGAGCCCATATATCGACAAAATAACTCACTATAATTACTTGATTTTATCCAAAGGCAAAATTGTACAGTACGGCACAAGAGAGAAACTTTTCTCCTCA 1443
T Y Q N I N I P V T Q N M V P S A R L L V Y Y I V T G E Q T A E L V A D A 518
ACTTATCAAAATATAAATATTCCAGTGACACAGAACATGGTTCCTTCAGCACGACTCCTGGTCTATTACATAGTCACAGGGGAGCAAACAGCAGAATTAGTGGCTGACGCA 1554
V W I N I E E K C G N Q L G V H L S P D E Y V Y S P G Q T V S L D M V T E 555
GTCTGGATAAATATTGAGGAGAAGTGTGGCAACCAGCTCCAGGTCCATCTGTCTCCAGATGAATATGTGTATTCTCCAGGCCAAACTGTGTCCCTTGACATGGTGACTGAA 1665
A D S W V A L S A V D R A V Y K V Q G N A K R A M Q R V F Q A L D E K S D 592
GCAGACTCATGGGTAGCACTATCAGCAGTGGACAGAGCTGTGTATAAAGTCCAGGGAAACGCCAAAAGGGCCATGCAAAGAGTCTTTCAAGCTTTGGATGAAAAGAGTGAC 1776
L G C G A G G G H D N A D V F H L A G L T F L T N A N A D D S H Y R D D S 629
CTGGGCTGTGGGGCAGGTGGTGGCCATGACAATGCAGATGTATTCCATCTAGCTGGGCTCACCTTCCTCACCAACGCAAACGCAGATGACTCCCATTATCGTGATGACTCT 1887
β α
C K E I L R S K R N L H L L R Q K I E E Q A A K Y K H S V P K K C C Y D G 666
TGTAAAGAAATTCTCAGGTCAAAGAGAAACCTGCATCTCCTAAGGCAGAAAATAGAAGAACAAGCTGCTAAGTACAAACATAGTGTGCCAAAGAAATGCTGCTATGACGGA 1998
A R V N F Y E T C E E R V A R V T I G P L C I R A F N E C C T I A N K I R 703
GCCCGAGTGAACTTCTACGAAACCTGTGAGGAGCGAGTGGCCCGGGTTACCATAGGCCCTCTCTGCATCAGGGCCTTCAACGAGTGCTGTACTATTGCGAACAAGATCCGA 2109
K E S P H K P V Q L G R I H I K T L L P V M K A D I R S Y F P E S W L W E 740
AAAGAAAGCCCCCATAAACCTGTCCAACTGGGAAGGATCCACATTAAGACCCTGTTACCAGTGATGAAGGCAGATATCCGAAGCTACTTTCCAGAGAGCTGGCTATGGGAA 2220
B
I H R V P K R K Q L Q V T L P D S L T T W E I Q G I G I S D N G I C V A D 777
ATTCATCGCGTTCCCAAAAGAAAACAGCTGCAGGTCACGCTGCCTGACTCACTAACGACTTGGGAAATTCAAGGCATTGGCATTTCAGACAATGGTATATGTGTTGCTGAT 2331
T L K A K V F K E V F L E M N I P Y S V V R G E Q I Q L K G T V Y N Y M T 814
ACACTCAAGGCAAAGGTGTTCAAAGAAGTCTTCCTGGAGATGAACATACCATATTCTGTTGTGCGAGGAGAACAGATCCAATTGAAAGGAACTGTTTACAACTATATGACC 2442
S G T K F C V K M S A V E G I C T S G S S A A S L H T S R P S R C V F Q R 851
TCAGGGACAAAGTTCTGTGTTAAAATGTCTGCTGTGGAGGGGATCTGCACTTCAGGAAGCTCAGCTGCTAGCCTTCACACCTCCAGGCCCTCCAGATGTGTGTTCCAGAGG 2553
I E G S S S H L V T F T L L P L E I G L H S I N F S L E T S F G K D I L V 888
ATAGAGGGCTCGTCCAGTCACTTGGTGACCTTCACCCTGCTTCCTCTGGAAATTGGCCTTCACTCCATAAACTTCTCACTAGAGACCTCATTTGGGAAAGACATCTTAGTA 2664
K T L R V V P E G V K R E S Y A G V I L D P K G I R G I V N R R K E F P Y 925
AAGACATTACGGGTAGTGCCAGAAGGAGTCAAGAGGGAAAGCTATGCCGGCGTGATTCTGGACCCTAAGGGAATTCGTGGTATTGTTAACAGACGAAAGGAATTCCCATAC 2775
R I P L D L V P K T K V E R I L S V K G L L V G E F L S T V L S K E G I N 962
AGGATCCCATTAGATTTGGTCCCCAAGACCAAAGTTGAAAGGATTTTGAGTGTCAAAGGACTGCTTGTAGGGGAGTTCTTGTCCACGGTTCTGAGTAAGGAAGGCATCAAC 2886
B
I L T H L [P K G S A E A E L M S I A P] V F Y V F H Y L E A G N H W N I F Y 999
ATCCTAACCCACCTCCCCAAGGGCAGTGCAGAGGCAGAGCTCATGAGCATAGCTCCGGTGTTCTATGTTTTCCACTACCTGGAAGCAGGAAACCATTGGAATATTTTCTAT 2997
P D T L S K R Q S L E K K I K Q G V V S V M S Y R N A D Y S Y S M W K G A 1036
CCTGATACACTGAGTAAAAGACAGAGCCTGGAGAAAAAAATAAAACAAGGGGTGGTGAGCGTCATGTCCTACAGAAACGCTGACTATTCCTACAGCATGTGGAAGGGGGCG 3108
S A S T W L T A F A L R V L G Q V A K Y V K Q D E N S I C N S L L W L V E 1073
AGCGCTAGTACCTGGCTGACAGCTTTTGCTCTGAGAGTGCTTGGACAGGTGGCCAAGTATGTAAAACAGGATGAAAACTCAATTTGTAACTCTTTGCTATGGCTGGTTGAG 3219
K C Q L E N G S F K E N S Q Y L P I K L Q G T L P A E A Q E K T L Y L T A 1110
AAGTGTCAGCTGGAAAACGGCTCTTTCAAGGAAAATTCCCAATATCTACCAATAAAATTACAGGGTACTTTGCCTGCTGAAGCCCAAGAGAAAACTTTGTATCTTACAGCC 3330
F S V I G I R K A V D I C P T M K I H T A L D K A D S F L L E N T L P S K 1147
TTTTCTGTGATTGGAATTAGAAAGGCAGTTGACATATGCCCCACCATGAAAATCCACACAGCGCTAGATAAAGCCGACTCCTTCCTGCTTGAAAACACCCTGCCATCCAAG 3441
S T F T L A I V A Y A L S L G D R T H P R F R L I V S A L R K E A F V K G 1184
AGCACCTTCACACTGGCCATTGTAGCCTATGCTCTTTCCCTAGGAGACAGAACCCACCCGAGGTTTCGTCTAATTGTGTCGGCCCTGAGGAAGGAAGCTTTTGTTAAAGGT 3552
D P P I Y R Y W R D T L K R P D S S V P S S G T A G M V E T T A Y Y A L L A 1221
GATCCGCCCATTTACCGTTACTGGAGAGATACCCTCAAACGTCCAGACAGCTCTGTGCCCAGCAGCGGCACAGCAGGTATGGTTGAAACCACAGCCTATGCTTTGCTCGCC 3663
S L K L K D M N Y A N P I I K W L S E E Q R Y G G G F Y S T Q D T I N A I 1258
AGCCTGAAACTGAAGGATATGAATTACGCCAACCCCATCATCAAGTGGCTATCTGAAGAGCAGAGGTATGGAGGCGGCTTTTATTCCACCCAGGATACGATTAATGCCATC 3774

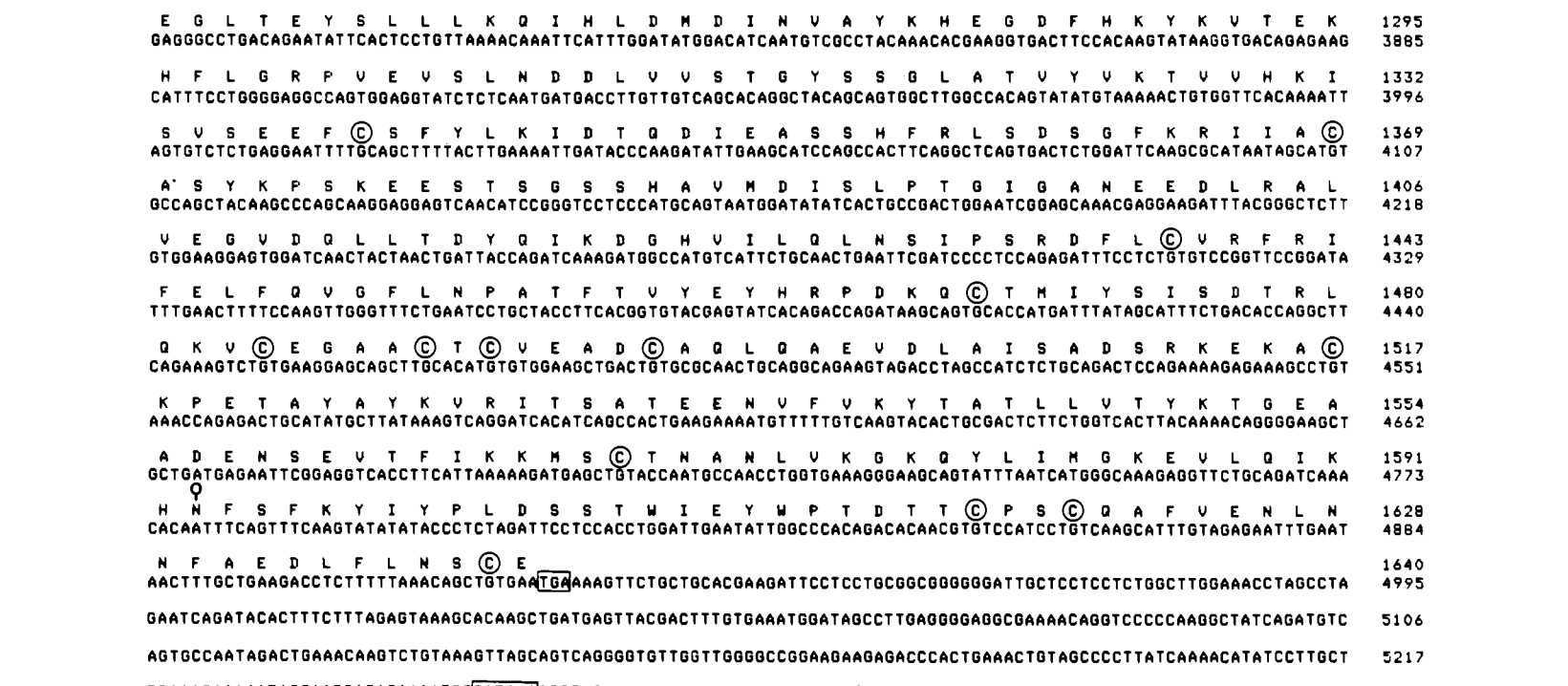


FIGURE 1: Nucleotide and derived amino acid sequence of pMC5/04. Cysteine residues are circled. Possible N-linked carbohydrate attachment sites are indicated by the symbol (♀). The restriction sites for the enzymes *Bam*HI (B) and *Xho*I (X) are indicated by lines underneath the nucleotide sequence. The region which corresponds to the thiol ester domains of C3, C4, and α_2M is bracketed. Boxed sequences correspond to the $\beta\alpha$ junction, the stop codon, and the putative polyadenylation recognition signal.

separate chromosomes. The structural loci for C4 are within the major histocompatibility complex on chromosome 6 in humans (Carroll et al., 1984) and chromosome 17 in the mouse (Passmore & Schreffler, 1970; Carroll & Capra, 1979). Human C3 and α_2M have been mapped to chromosomes 19 (Whitehead et al., 1982) and 12 (Kan et al., 1985), respectively; murine C3 has been localized outside the major histocompatibility complex on chromosome 17 (Penalva da Silva et al., 1978; Natsuume-Sakai, 1979). Recently, the structural gene for murine C5 has been mapped to the Hc locus on chromosome 2 (D'Eustachio et al., 1986).

Here we report the complete nucleotide sequence of a large cDNA clone encoding most, and possibly all, of murine C5. Our studies confirm earlier work which suggested that C5 was devoid of an internal thiol ester site. Nevertheless, C5 does share extensive sequence homology with C3, C4, and α_2M including the region corresponding to their thiol ester domains. Murine C5a also contains three additional amino acids at its amino terminus that are not present in human, porcine, or bovine C5a.

Materials and Methods

Construction and Screening of a cDNA Library. A cDNA library was constructed (Sepich et al., 1985) from size-selected (>28 S) poly(A+) B10.WR mouse liver RNA by the method of Okayama and Berg (1983). Before first-strand synthesis was performed, the template RNA (1 μg) was denatured in 5 mM methylmercury(II) hydroxide. Recombinant plasmids were transformed into *Escherichia coli* strain MC1061 (Casabadan & Cohen, 1980). A total of 150000 transformants were plated and screened (Hanahan & Meselson, 1980) with a ^{32}P nick-translated *Rsa*I restriction fragment derived from a partial human C5 clone (Lundwall et al., 1985). The 16 clones which hybridized with this probe were colony purified and their inserts excised by digestion with *Bam*HI, *Hpa*I, or *Xho*I. Insert sizes were estimated by subjecting the digested samples to 1% agarose gel electrophoresis. Inserts obtained for sequence analysis were isolated from 250 μg of digested plasmid by the trough elution method of Maniatis et al. (1982), employing a 1% agarose preparative gel.

Determination of DNA Sequence. Ten micrograms of cDNA was self-ligated, sheared randomly by sonication, end-repaired, and cloned into M13 mp8 (Bankier & Barrell, 1983). M13 recombinants were sequenced by the dideoxy chain termination method (Sanger et al., 1977, 1980) using ^{35}S label and gradient gels (Biggin et al., 1983). Each cDNA strand was sequenced at least once, and on average, each base position was sequenced 5.4 times. Some stretches of cDNA sequence were obtained by directed chemical sequencing methods (Maxam & Gilbert, 1980). All nucleotide data were compiled and joined into a contiguous sequence by the Staden computer programs (Staden, 1982b). Protein homologies were determined and graphically illustrated by employing the DIAGON computer program (Staden, 1982a). Hydrophobicity calculations were performed and plotted by utilizing the computer program of Kyte and Doolittle (1982).

Results

Isolation and Sequence Determination of Murine C5 cDNA. Of the 150000 recombinants screened, 16 clones were found to contain C5 cDNA inserts. The C5 cDNA of greatest length (>5.2 kb) was obtained from the clone pMC5/04. Because of internal *Xho*I, *Hpa*I, and *Bam*I restriction sites, this C5 insert could not be isolated as an intact fragment. Therefore, two fragments (1.0 and 4.3 kb) obtained after digestion with *Xho*I were isolated and sequenced by the shotgun–dideoxy method (Bankier & Barrell, 1983). The overlapping sequence between the two *Xho*I-generated fragments was obtained after isolation and sequencing of a 1.2 kb *Bam*I fragment derived from the 5′ end of the pMC5/04 insert (see Figure 1). Some areas of the C5 cDNA sequence proved difficult to obtain by the shotgun–dideoxy method. These regions were sequenced by utilizing Maxam and Gilbert sequencing strategies (Maxam & Gilbert, 1980).

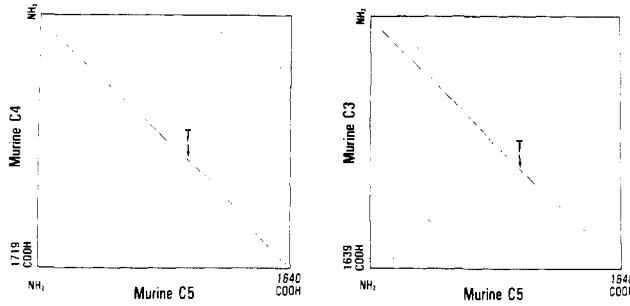


FIGURE 2: DIAGON analysis of the derived pro-C5 protein sequence compared to the sequences of pro-C4 (left panel) and pro-C3 (right panel). Sliding window was 25 amino acids; the percentage score parameter was 280. The arrows indicate the region in C5 compared to the corresponding thiol ester domains of C4 and C3.

Murine C3	P	A	G	C	G	E	Q	N	M	I	G	M	A	P
Murine C4	P	Q	G	C	A	E	Q	T	M	I	Y	L	A	P
Murine C5	P	K	G	S	A	E	A	E	L	M	S	I	A	P

FIGURE 3: Thiol ester domains of murine C3 and C4 aligned with the corresponding region in murine C5. Amino acids which are identical in all three proteins are boxed.

The murine C5 cDNA sequence and derived protein sequence are shown in Figure 1. The sequence spanned 5270 nucleotides including 350 untranslated residues at the 3′ end which preceded a poly(A) tail of undetermined length. A putative polyadenylation signal, CATAAA, was located 17 nucleotides upstream from the polyadenylation site. This differs from the AATAAA sequence previously determined for murine C3 (Wetsel et al., 1984; Lundwall et al., 1984), C4 (Sepich et al., 1985; Nonaka et al., 1985), and Slp (Ogata & Sepich, 1985; Nonaka et al., 1986). The deduced protein sequence of 1640 amino acids indicated a $\beta\alpha$-chain orientation with a four-residue basic sequence (RSKR) interposed between the β- and α-chains of the promolecule.

Analysis of the Murine C5 Protein Structure. Analysis of the murine C5 protein sequence data by DIAGON (Figure 2), an interactive graphics program for comparing and aligning protein and DNA sequences, illustrated extensive homology with C3, C4, and α_2M (data not shown). Furthermore, this program identified a region within the C5 sequence which was homologous to each of the thiol ester domains present in C3, C4, and α_2M. Alignment of the corresponding protein sequences of these family members (Figure 3) indicated that 9 of 14 residue positions were conserved within the thiol ester like domain of C5. However, the cysteine and glutamine residues, which bridge to form the thiol ester bond in C3, C4, and α_2M, have been replaced in C5 by serine and alanine, respectively. Interestingly, the prolines which flank the thiol ester region in C3, C4, and α_2M remain conserved within C5.

There were 29 cysteine residues found in the deduced murine C5 amino acid sequence, 3 in the β-chain and 26 in the α-chain. By comparison, there are 27 and 28 cysteines reported in murine C3 (Lundwall et al., 1984; Wetsel et al., 1984) and C4 (Sepich et al., 1985), respectively. A schematic representation of these cysteines and their locations within each protein sequence is illustrated in Figure 4. The positions of 22 of these cysteines are highly conserved among C3, C4, and C5. Of these conserved residues, 11 range in position from near the carboxy termini of the β-chains to approximately 180 amino acid residues into the amino termini of the α-chains. The remaining 11 conserved cysteines are clustered at the carboxy termini of the α-chains of C3 and C5 and within the γ-chain of C4.

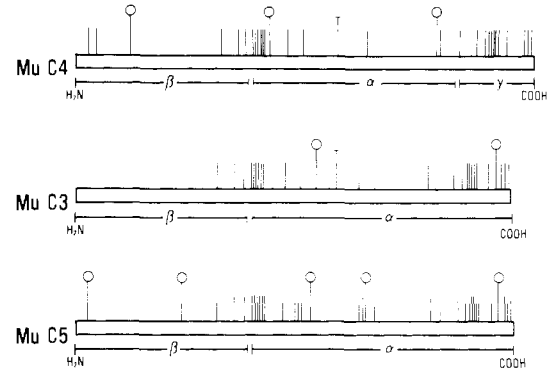


FIGURE 4: Schematic representation of cysteinyl residues and possible carbohydrate attachment sites in murine C3, C4, and C5. The cysteine which is critical to the formation of the thiol ester is indicated by a T.

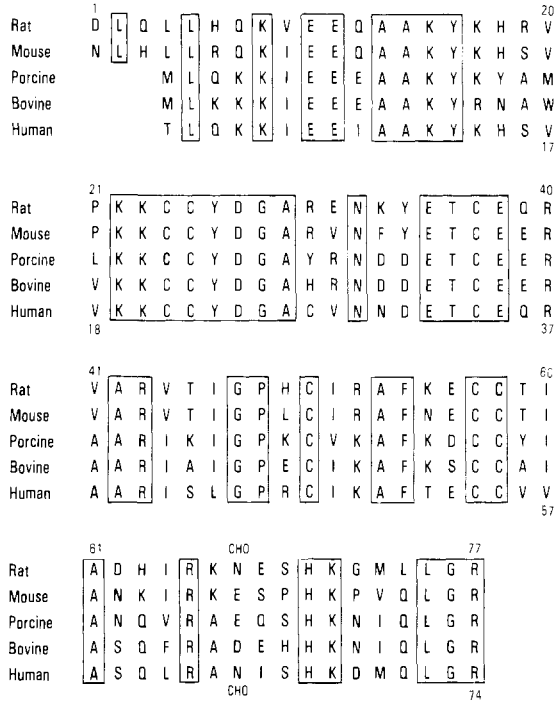


FIGURE 5: Comparison of rat, murine, porcine, bovine, and human C5a sequences. Residues which are identical in all four species are boxed.

There were five possible N-linked (Marshall, 1972) carbohydrate attachment sites localized in the murine C5 sequence; two were present in the β-chain (positions 32 and 387), and three were contained in the α-chain (positions 875, 1079, and 1593). When comparing murine C3, C4, and C5, neither the number of potential carbohydrate binding sites nor their locations were found to be conserved.

From the cDNA sequence data, it was discovered that mouse C5a would be predicted to contain, at its amino terminus, three additional amino acids (Asn, Leu, and His) not present in human (Fernandez & Hugli, 1978), porcine (Gerard & Hugli, 1980), or bovine (Gennaro et al., 1986) C5a (Figure 5). The presence of three additional amino acids (Asp, Leu, and Gln) at the amino terminus of rat C5a has also been reported (Cui et al., 1985). Unlike rat C5a, mouse C5a did not contain a potential carbohydrate attachment site located at residue 67.

Diagrammed in Figure 6 are hydropathy plots constructed from sequence data of the murine pro-C3 and pro-C5 molecules. It has been demonstrated that for globular proteins these plots provide a good indication of protein topology; i.e., interior

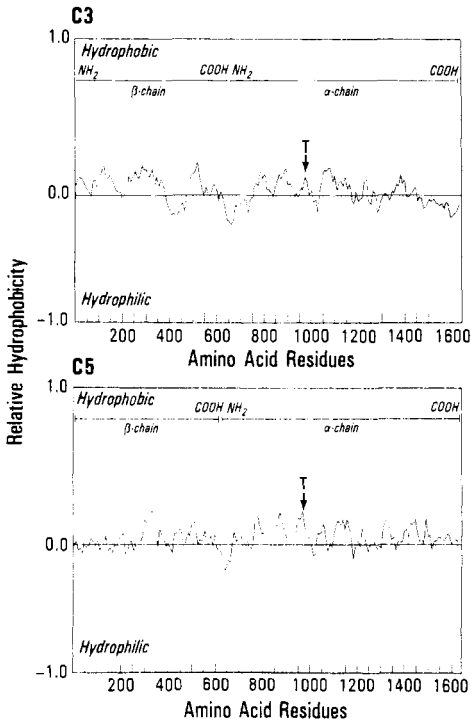


FIGURE 6: HYDROPLOT analysis of pro-C5 and pro-C3 protein sequences. The bar above the plot represents the corresponding protein structure of the promolecules. Regions which score above-average hydropathy are indicated by positive values. The thiol ester domain of C3 and the homologous region in C5 are depicted by the arrows.

regions score more hydrophobic than average (Kyte & Doolittle, 1982). As might be expected, hydropathy analyses of pro-C3 and pro-C5 yielded similar profiles. These plots predicted a general hydrophobic nature for both C3 and C5, with C5 exhibiting a few more hydrophobic regions. The most hydrophilic stretch of sequence contained within these promolecules was located around the β–α junction region. The four-residue β–α junctions, the carboxy termini of the β-chains, and anaphylatoxin sequences are abundant in charged amino acids. This stretch of approximately 90 residues is known from protein studies to be most susceptible to endopeptidase activity. In contrast, the thiol ester domain of C3 and the corresponding region in C5 were predicted to be localized in hydrophobic areas. Interestingly, the C5 domain which does not harbor the metastable β-cysteinyl-γ-glutamyl thiol ester moiety resided in a more hydrophobic environment than the corresponding C3 domain.

DISCUSSION

The mouse C5 sequence reported here spans 1640 amino acid residues. This is one amino acid more than was present in the complete mouse pro-C3 sequence (Wetsel et al., 1984; Lundwall et al., 1984). The N-terminus of the C5 β-chain is blocked and therefore cannot be directly sequenced (Tack et al., 1979; DiScipio et al., 1983). Thus, we are not presently able to determine if pMC5/04 contains the entire 5′ end of the pro-C5 coding sequence. On the basis of alignment with the murine C3 and C4 sequences (data not shown), we believe that these 1640 amino acids represent all but approximately 16 residues of the mouse pro-C5 structure. Our current studies involving the isolation and sequence determination of murine C5 genomic DNA may help to establish the primary sequence of the amino-terminal region of pro-C5. However, in the absence of direct protein sequence data for the amino terminus of the C5 β-chain, nucleotide sequence alone will not be sufficient to establish the correct amino-terminal sequence for the posttranslationally modified pro-C5 molecule.

Interestingly, the nearly ubiquitous AATAAA or ATTAAA polyadenylation sequence, located 10–30 nucleotides upstream of the poly(A) tail (Proudfoot & Brownlee, 1976; Birnstiel et al., 1985), was not present in the murine C5 cDNA sequence. Instead, murine C5 utilizes the somewhat rare CATAAA sequence (Wickens & Stephenson, 1984). This sequence was not an artifact of this particular clone since analysis of the 3′ end of an independent C5 clone (pMC5/17) revealed the same sequence.

Several investigators have previously indicated that the β-cysteinyl-γ-glutamyl thiol ester present in C3, C4, and α_2M was not present in C5 (Law et al., 1980; Janatova & Tack, 1981; DiScipio, 1981). The present sequence reveals that C5 lacks the cysteine and glutamine residues required for thiol ester formation but otherwise shares extensive homology with C3, C4, and α_2M including their thiol ester regions. This finding argues that the absence of a thiol ester bond in C5 is the result of single amino acid differences rather than the deletion or substitution of a larger structural element. It has been determined recently in our laboratory that human C5 also contains the cysteine to serine and glutamine to alanine amino acid substitutions within its corresponding thiol ester like region (Wetsel et al., 1985). In addition, the hen egg white proteinase inhibitor ovostatin, which has similar biochemical and functional properties to α_2M, apparently lacks the internal thiol ester site (Nagase & Harris, 1983; Nagase et al., 1983). It will be quite interesting to determine if ovostatin has undergone similar amino acid changes within its presumed thiol ester like domain as have occurred in mouse and human C5. Further isolation, characterization, and sequence analysis of genomic DNA corresponding to this family of homologous proteins should foster additonal insights regarding the evolution of the thiol ester.

As illustrated in Figure 5, mouse C5a was found to contain three additional amino acids not present in the human, porcine, and bovine C5a sequences previously determined. Presently, the function of these residues, which are located at the very amino terminus, is not known. However, Cui et al. (1985) have recently isolated and sequenced the C5a peptide derived from the rat. They found that the rat C5a also contained three additional amino acids at the amino terminus (Figure 5). Furthermore, these investigators discovered that the rat C5a was 100–1000-fold more active in guinea pig ileum contraction assays when compared to human C5a. By employing the C5a model of Greer (1985), these researchers postulated that the three residues at the N-terminal end of rat C5a may greatly enhance the functional activity of the molecule by stabilizing the helical conformation of the critical C-terminal region of the molecule. Since mouse C5a contains three additional amino acids located at the amino terminus, it is feasible that it will exhibit enhanced anaphylatoxic activity as well. Surprisingly, mouse C5a was determined to contain two prolines (positions 69 and 72) not present in any other C5a that has been sequenced. Currently, it is not known what effect these residues will have on the biological activity of the peptide. By employing synthetic peptides constructed from these C5a sequences and recombinant C5a analogues, it will be possible to address these intriguing structure–function relationships.

ACKNOWLEDGMENTS

We thank Deborah Noack and James M. Jacquet for their excellent technical assistance and Diane S. Sepich for advice in performing the Maxam and Gilbert sequencing. We thank Tony E. Hugli and L. X. Cui for making the rat C5a sequence available before its publication. In addition, we are grateful

to Joan Cleste for typing the manuscript, to Scott R. Barnum for proofreading the manuscript, and to Torsten Kristensen and Dan Noonan for helpful discussions.

Registry No. Complement C5, 80295-53-0.

REFERENCES

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

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

Birnstiel, M. D., Busslinger, M., & Strub, K. (1985) *Cell (Cambridge, Mass.) 41*, 349–359.

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

Carroll, M. C., & Capra, J. D. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4641–4645.

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

Casabadan, M., & Cohen, S. (1980) *J. Mol. Biol. 138*, 179–207.

Cui, L. X., Ferreri, K., & Hugli, T. E. (1985) *Complement 2*, 18.

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.

DiScipio, R. G. (1981) *Biochem. J. 199*, 497–504.

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

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

Gennaro, R., Simonic, T., Negri, A., Mottola, C., Secchi, C., Ronchi, S., & Romeo, D. (1986) *Eur. J. Biochem. 155*, 77–86.

Gerard, C., & Hugli, T. E. (1980) *J. Biol. Chem. 255*, 4710–4715.

Goldlust, M. B., Shin, H. S., Hammer, C. H., & Mayer, M. M. (1974) *J. Immunol. 113*, 998–1007.

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

Greer, J. (1985) *Science (Washington, D.C.) 228*, 1055–1060.

Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63–67.

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

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 Publishers, Ann Arbor, MI.

Hugli, T. E. (1984) *Springer Semin. Immunopathol. 7*, 193–219.

Janatova, J., & Tack, B. F. (1981) *Biochemistry 20*, 2394–2402.

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. E., Hiorns, L. R., & Fey, G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2282–2286.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Law, S. K., Lichtenberg, N. A., Holcomb, F. H., & Levine, R. P. (1980) *J. Immunol. 125*, 634–639.

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., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, a Laboratory Manual* (Maniatis, T., Fritsch, E. F., & Sambrook, J., Eds.) pp 164–167, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marshall, R. D. (1972) *Annu. Rev. Biochem. 41*, 673–702.

Maxam, A., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

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

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

Natsuume-Sakai, S., Hayakawa, J. I., Amono, S., & Takahashi, M. (1979) *J. Immunol. 123*, 947–948.

Negase, H., & Harris, E. D., Jr. (1983) *J. Biol. Chem. 258*, 7490–7498.

Negase, H., Harris, E. D., Jr., Woessner, J. F., & Brew, K. (1983) *J. Biol. Chem. 258*, 7481–7489.

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*, 109–10943.

Nonaka, M., Nakayama, K., Yeul, Y. D., & Takahashi, M. (1986) *J. Immunol. 136*, 2989–2993.

Ogata, R. T., & Sepich, D. S. (1985) *J. Immunol. 135*, 4239–4244.

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

Ooi, Y. M., & Colten, H. R. (1979) *J. Immunol. 123*, 2494–2498.

Passmore, H. C., & Schreffler, D. C. (1970) *Biochem. Genet. 4*, 351–365.

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

Penalva da Silva, F., Hoecker, G. F., Day, N. K., Vienne, K., & Rubinstein, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 963–967.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 163*, 211–214.

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

Salveson, G. S., Sayer, C. A., & Barret, A. J. (1981) *Biochem. J. 195*, 453–461.

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

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., & Roe, B. A. (1980) *J. Mol. Biol. 143*, 161–178.

Schreiber, R. D., Pangburn, M. K., Lesavre, P. H., & Muller-Eberhard, H. J. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3948–3952.

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

Sottrup-Jensen, L. (1986) *Plasma Proteins (2nd Ed.)* (in press).

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

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Wierzbicki, D. M., Jones, C. M., Magnusson, S., & Petersen, T. E. (1984) *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., Lundwall, A., Davidson, F., Gibson, T., Tack, B. F., & Fey, G. H. (1984) *J. Biol. Chem. 259*, 13857–13862.

Selective Radiolabeling of Cell Surface Proteins to a High Specific Activity†

James A. Thompson, Alice L. Lau, and Dennis D. Cunningham*
Department of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, California 92717
Received June 9, 1986; Revised Manuscript Received October 22, 1986

ABSTRACT: A procedure was developed for selective radiolabeling of membrane proteins on cells to higher specific activities than possible with available techniques. Cell surface amino groups were derivitized with ^{125}I-(hydroxyphenyl)propionyl groups via ^{125}I-sulfosuccinimidyl (hydroxyphenyl)propionate (^{125}I-sulfo-SHPP). This reagent preferentially labeled membrane proteins exposed at the cell surface of erythrocytes as assessed by the degree of radiolabel incorporation into erythrocyte ghost proteins and hemoglobin. Comparison with the lactoperoxidase–[^{125}I]iodide labeling technique revealed that ^{125}I-sulfo-SHPP labeled cell surface proteins to a much higher specific activity and hemoglobin to a much lower specific activity. Additionally, this reagent was used for selective radiolabeling of membrane proteins on the cytoplasmic face of the plasma membrane by blocking exofacial amino groups with uniodinated sulfo-SHPP, lysing the cells, and then incubating them with ^{125}I-sulfo-SHPP. Exclusive labeling of either side of the plasma membrane was demonstrated by the labeling of some marker proteins with well-defined spacial orientations on erythrocytes. Transmembrane proteins such as the epidermal growth factor receptor on cultured cells could also be labeled differentially from either side of the plasma membrane.

In many studies where it is important to identify surface membrane proteins in low abundance, it is critical to radiolabel them to a high specific activity under conditions that label cytoplasmic proteins only minimally (Moss & Cunningham, 1981; Phillips & Agin, 1974). Several procedures have been developed to label surface membrane proteins, although each has certain limitations. Metabolic labeling of cultured cells with [^{3}H]fucose (Atkinson & Summers, 1974) or [^{3}H]-glucosamine (Liau & Horowitz, 1982) results in the labeling of glycoproteins of both the plasma membrane and also intracellular membranes. In addition, these procedures take long periods of time and results in labeled proteins of relatively low specific activity. Other methods involve the oxidation of membrane glycoproteins with galactose oxidase (Baumann & Doyle, 1978) or periodate (Steck & Dawson, 1974) followed by reduction with sodium [^{3}H]borohydride. However, proteins labeled with tritium very poorly expose X-ray film during autoradiography because of the low energy of this isotope (Bonner & Laskey, 1974). The most common method for labeling cell surface proteins with a high-energy isotope involves the generation of [^{125}I]iodine by incubating cells with [^{125}I]iodide plus hydrogen peroxide (Phillips & Morrison, 1970; Morrison, 1974, 1980) or with [^{125}I]iodide plus lactoperoxidase and glucose oxidase (Hubbard & Cohn, 1975). However, this procedure does not label membrane proteins to a high specific activity; only 0.1% of the available tyrosines become labeled (Hubbard & Cohn, 1975). Chloroglycoluril has also been used to label membrane proteins on cells with ^{125}I; however, there is significant labeling of cytoplasmic proteins when conditions are adjusted to achieve very high specific activities of membrane proteins (Markwell & Fox, 1978). There are other membrane-impermeant labeling agents that can be iodinated, but they are low in reactivity with membrane proteins at a neutral pH (Cabantchik & Rothstein, 1974; Berg, 1969).

In the present study, a procedure was developed to radiolabel membrane proteins to a high specific activity with a high-energy isotope with little labeling of cytoplasmic proteins. This procedure utilizes a water-soluble form of the iodinated Bolton–Hunter reagent, ^{125}I-sulfosuccinimidyl (hydroxyphenyl)propionate (^{125}I-sulfo-SHPP).[1] This reagent retains the highly reactive amino-labeling properties of the Bolton–Hunter reagent (Bolton & Hunter, 1973) but is rendered membrane impermeable by sulfation. Experiments were conducted with erythrocytes to evaluate the degree of surface membrane vs. cytoplasmic labeling. Clean cytoplasmic and membrane fractions can easily be obtained from erythrocytes. Also, erythrocytes have several protein markers that have been well characterized that sharply delineate the cytoplasmic space as well as the inner and outer faces of the plasma membrane

† This research was supported by National Institutes of Health Research Grant CA 12306.
* Correspondence should be addressed to this author.

[1] Abbreviations: sulfo-SHPP, sulfosuccinimidyl (hydroxyphenyl)-propionate; HF, human foreskin; CHEF, Chinese hamster embryonic fibroblast; EDTA, (ethylenedinitrilo)tetraacetic acid; PBS, phosphate-buffered saline; D-PBS, Dulbecco's modified phosphate-buffered saline; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; SDS, sodium dodecyl sulfate; EGF, epidermal growth factor; SDS–PAGE, SDS–polyacrylamide gel electrophoresis; kDa, kilodalton(s); TPCK, tosylphenylalanine chloromethyl ketone.

0006-2960/87/0426-0743$01.50/0