Complete Amino Acid Sequence of a Mouse μ Chain: Homology among Heavy Chain Constant Region Domains†

Marilyn R. Kehry,* Jonathan S. Fuhrman, James W. Schilling,‡ John Rogers, Carol H. Sibley, and Leroy E. Hood

ABSTRACT: The complete amino acid sequence of the mouse μ chain secreted by the MOPC 104E myeloma tumor has been determined. There are four constant region domains in the μ chain and a 20-residue COOH-terminal segment that plays a role in the polymerization of pentameric immunoglobulin M molecules. There are six sites of carbohydrate attachment in the MOPC 104E μ chain. Three complex-type and two high-mannose oligosaccharides are located in the μ chain constant region. The general type and location of carbohydrate moieties in the μ chain constant region are completely conserved between mouse and human μ chains. Homology in the location of carbohydrate structures on different classes of heavy chains is discussed.

Immunoglobulin molecules are composed of two polypeptide chains, light (L)[1] and heavy (H) chains. Each heavy chain may be divided into a variable (V_H) or antigen-binding region and a constant (C_H) region that performs the various effector functions of the immunoglobulin molecule. Heavy chain constant regions may be further subdivided into homology regions or domains approximately 110 residues in length (Edelman et al., 1969) that correspond to individual gene segments separated by noncoding sequences in the germ-line genes (Kataoka et al., 1980; Tucker et al., 1979a,b; Early et al., 1979; Sakano et al., 1979a,b; Gough et al., 1980; Calame et al., 1980; Kawakami et al., 1980). The μ chain has four C_H domains that are designated $C_\mu 1$, $C_\mu 2$, $C_\mu 3$, and $C_\mu 4$ from the NH_2 to the COOH terminus. There is an additional COOH-terminal segment in the μ chain that is composed of 20 amino acid residues (Kehry et al., 1979).

IgM molecules are unique among the different classes of immunoglobulins in that they exist in two different polymeric forms that perform different effector functions in the immune system. First, the IgM molecule is a cell-surface receptor that functions in triggering the differentiation of precursor B lymphocytes to antibody-producing plasma cells (Warner, 1974). This membrane-bound form of the IgM molecule is a monomeric (μ_2L_2) immunoglobulin embedded in the plasma membrane of B lymphocytes (Vitetta et al., 1971). Second, antibody-producing plasma cells secrete IgM molecules that are assembled into pentamers $(\mu_2L_2)_5$ by the addition of a joining (J) chain (Della Corte & Parkhouse, 1973). These soluble IgM pentamers then circulate in the bloodstream where they function to bind and neutralize foreign antigens and to fix complement. IgM molecules also are unique in being the first immunoglobulins to appear in vertebrate evolution (Marchalonis, 1972) and the first to appear during the development of the immune system (Lawton et al., 1975). The above features make the IgM molecule an interesting protein to structurally characterize and to use for investigating the evolution of immunoglobulin heavy-chain classes.

We present here a complete structural analysis of a secreted μ chain from the mouse myeloma tumor MOPC 104E. A preliminary report of this structure has been presented elsewhere (Kehry et al., 1979). DNA sequence analyses on a cloned mouse μ gene have defined the locations of noncoding DNA sequences within the MOPC 104E C_μ region (Calame et al., 1980) that delineate the exact boundaries between C_μ region domains. We discuss the homologies in the sites and types of carbohydrates attached to the different classes of heavy chains.

Experimental Procedures

Isolation of MOPC 104E μ Chain. The IgM-secreting myeloma MOPC 104E (μ, λ), subsequently abbreviated 104E, was obtained from Dr. M. Potter and was passaged subcutaneously or intraperitoneally in $(BALB/c \times DBA/2)F_1$ mice. Proteins were precipitated from ascites fluid (40 mL) by two cycles of treatment with 50% saturated ammonium sulfate at 4 °C. The final precipitate was dissolved in 0.15 M sodium borate, 0.14 M NaCl, and 0.02% NaN_3 (BBS), and the pentameric IgM molecules were purified by gel filtration on a column (5 × 100 cm) of ACA 22 (LKB) equilibrated in BBS. IgM molecules were eluted just after the void volume. Fractions were pooled and precipitated with 50% saturated ammonium sulfate at 4 °C, and the precipitate was dissolved in 6 M Gdn-HCl, 0.25 M Tris-HCl (pH 8.5), and 0.14 M EDTA. Complete reduction and alkylation were carried out essentially as described previously (Konigsberg, 1972). When iodo[1-^{14}C]acetamide was employed for alkylation, the reaction was allowed to proceed for 30 min, followed by the addition of nonradioactive iodoacetamide to 0.025 M. After dilution of the guanidine concentration to 4 M by the addition of 0.4 M ammonium bicarbonate, heavy and light chains were separated by gel filtration on a column (5 × 100 cm) of ACA 34 (LKB) equilibrated in 3 M Gdn-HCl, 0.2 M ammonium bicarbonate, and 0.02% NaN_3.

† From the Division of Biology, California Institute of Technology, Pasadena, California 91125 (M.R.K., J.S.F., J.W.S., and L.E.H.), the Department of Genetics, University of Washington, Seattle, Washington 98195 (C.H.S.), and the Department of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, California 90024 (J.R.). *Received April 12, 1982.* This work was supported by National Institutes of Health Grants AI 16950, CA 20314, AI 13410, and CA 12800. C.H.S. was the recipient of Faculty Research Award FRA-178 from the American Cancer Society. M.R.K. was the recipient of a Gordon Ross Medical Foundation fellowship. A portion of this work has previously been published (Kehry et al., 1979).

* Address correspondence to this author at the Institute of Molecular Biology, University of Oregon, Eugene, OR 97403.

‡ Present address: Department of Biological Sciences, Stanford University, Stanford, CA 94305.

[1] Abbreviations: L, light chain, H, heavy chain; 104E, MOPC 104E myeloma; BBS, borate-buffered saline; EDTA, ethylenediaminetetraacetic acid; $NaDodSO_4$, sodium dodecyl sulfate; TPCK, L-1-(tosylamido)-2-phenylethyl chloromethyl ketone; PTH, phenylthiohydantoin; V_H, immunoglobulin heavy chain variable region; C_H, immunoglobulin heavy chain constant region; J, joining chain; IgM, immunoglobulin M; Gdn-HCl, guanidine hydrochloride; Tris, tris(hydroxymethyl)aminomethane; HPLC, high-performance liquid chromatography.

0006-2960/82/0421-5415$01.25/0

Isolation of Cyanogen Bromide Fragments. Purified μ chains (between 35 and 160 mg) in 3 M Gdn-HCl and 0.2 M ammonium bicarbonate were concentrated (Millipore molecular separator), acidified by gradual addition of 88% formic acid, and desalted on a column (2.5 × 40 cm) of Bio-Gel P-2 in 50% formic acid at 4 °C. Pooled μ chains were concentrated by partial lyophilization, the concentration of formic acid was adjusted to 70%, and the protein was cleaved by cyanogen bromide (50 mg/mL) in the dark with constant stirring (Gross, 1967; Kehry et al., 1979). The peptides were dissolved sequentially in 8 M Gdn-HCl, 3 M Gdn-HCl and 0.2 M ammonium bicarbonate, and finally 0.4 M ammonium bicarbonate. The solutions were combined, and the cyanogen bromide peptides were separated by gel filtration on a column (3.5 × 140 cm) of ACA 54 (LKB) equilibrated in 3 M Gdn-HCl, 0.2 ammonium bicarbonate, and 0.02% NaN_3. Aliquots were counted for radioactivity when the cysteine residues were ^{14}C labeled. For direct sequence determinations, pooled fragments were either neutralized and desalted on a column (2.5 × 40 cm) of Bio-Gel P-2 in 20% formic acid or dialyzed (Spectra/Por 3) against three changes of 5% formic acid at 4 °C, followed by lyophilization (see paragraph at end of paper regarding supplementary material).

Polyacrylamide Gel Electrophoresis. $NaDodSO_4$–polyacrylamide gel electrophoresis (Laemmli, 1970) was routinely used to monitor the purifications described above (Kehry et al., 1979). Guanidine was removed from the samples prior to electrophoresis by dialysis (Spectra/Por 3) against 5% formic acid, followed by lyophilization. Gels were stained with 0.025% Coomassie Brilliant Blue in 25% 2-propanol and 10% acetic acid and destained in 10% acetic acid.

Composite gels (1% agarose and 2.5% polyacrylamide) were run by a modification of the procedure of Peacock & Dingman (1968). Gels contained 0.5 M urea, 0.2 M sodium phosphate (pH 7.2), and 0.1% $NaDodSO_4$. The running buffer contained 0.5 M urea, 0.1% $NaDodSO_4$, and 0.75 M sodium phosphate (pH 7.2), and the sample buffer was composed of 2% $NaDodSO_4$, 10% glycerol, and 0.05 M sodium phosphate (pH 6.8). Gels were 10-cm long and were electrophoresed at 3.5 mA/gel for 16 h, stained for 1 h with 0.25% Coomassie Brilliant Blue in methanol–acetic acid–water, 5:1:5, and destained in the same solution without Coomassie Brilliant Blue, followed by equilibration in 10% acetic acid.

Tryptic, Chymotryptic, and Thermolysin Digestion. Cyanogen bromide fragments (2–20 mg) were dialyzed (Spectra/Por 3) exhaustively against 0.2 M ammonium bicarbonate, lyophilized, and dissolved in a small volume of 0.2 M ammonium bicarbonate. TPCK-trypsin (Worthington) or chymotrypsin (Sigma) was added in 0.001 N HCl, or thermolysin (Sigma) was added in 0.2 M ammonium bicarbonate (freshly prepared enzyme stock) at an enzyme:substrate ratio of 1:100. Enzyme addition was performed 3 times at 1-h intervals. Incubation was carried out at 37 °C for a total period of 2.5 h, and the digestion was terminated by lyophilization.

Peptides were separated in two dimensions as described previously (Katz et al., 1959). Ninhydrin-positive spots on analytical and preparative fingerprints were excised, eluted by two extractions into 0.5 M NH_4OH, and lyophilized.

Cleavage at Arginine or Tryptophan Residues. For specific cleavage at arginine residues, cyanogen bromide fragments (5–15 mg) were dialyzed exhaustively (Spectra/Por 3) against 5% formic acid at 4 °C, lyophilized, and dissolved in 12 mL of 6 M Gdn-HCl and 0.01 M Tris-HCl (pH 9.5). Amino groups on the protein were succinylated by gradual addition of finely powdered succinic anhydride (40-fold excess by weight over protein) in a pH stat (Klapper & Klotz, 1972). When reversible blockage of lysine was desired, the protein was reacted with citraconic anhydride or maleic anhydride under similar conditions. In some instances the protein was radiolabeled by reacting with [^{14}C]succinic anhydride. Reagents were removed by exhaustive dialysis against 0.1 M NH_4OH, followed by lyophilization. Trypsin digestion was performed as described in the previous section. In some cases, peptides were separated on a column of Sephadex G-75 or G-50 (Sigma) equilibrated in 0.2 M ammonium bicarbonate or ACA 54 (LKB) equilibrated in 3 M Gdn-HCl, 0.2 ammonium bicarbonate, and 0.02% NaN_3. The maleylation or citraconylation was reversed by acidifying (pH 3.5) the protein overnight. Trypsin cleavage at lysine residues was then performed as described in the previous section.

For specific cleavage at tryptophan residues (Ozols et al., 1977), cyanogen bromide fragments (5–10 mg) were dialyzed exhaustively (Spectra/Por 3) against 5% formic acid at 4 °C, succinylated, lyophilized, and dissolved in equal volumes of 88% formic acid and anhydrous heptafluorobutyric acid (Pierce). The reaction vessel was vented into water in a fume hood, and the tryptophan cleavage reaction was performed essentially as described by Ozols et al. (1977). In some cases, peptides were then separated on a column of Sephadex G-75 or G-50 (Sigma) equilibrated in 0.2 M ammonium bicarbonate. The tryptophan cleavage reaction also modified tyrosine residues in an unknown fashion that greatly altered the chromatographic behavior of the modified tyrosine phenylthiohydantoin (PTH) derivative obtained by sequence analysis. The elution position of the modified PTH-tyrosine was identified by HPLC analysis of sequence determinations on modified peptides with tyrosine residues at known positions.

Fc Digestion. IgM Fc fragments were produced by a modification of the method of Shimizu et al. (1975). Whole IgM in BBS (4.5 mg) was concentrated and desalted on a column (3 × 15 cm) of Sephadex G-25 (Sigma) equilibrated in freshly prepared 5 M urea, 0.1 M Tris-HCl (pH 8), 0.15 M NaCl, and 0.01% NaN_3. After concentration of the desalted IgM molecules, a stock solution of 2 mg/mL TPCK-trypsin (Worthington) in 0.1 M Tris-HCl (pH8.0), 0.1 M $CaCl_2$, and 0.15 M NaCl was added to give an enzyme:substrate ratio of 1:100 and incubated at 25 °C for 18 h. The efficiency of pentameric Fc production was monitored by removing aliquots at specific times after addition of enzyme. Proteins were then electrophoresed without reduction on composite 1% agarose and 2.5% polyacrylamide–$NaDodSO_4$ gels (Peacock & Dingman, 1968), or the proteins were reduced and analyzed on 10% polyacrylamide–$NaDodSO_4$ gels (Laemmli, 1970). For preparative isolation of Fc fragments with 150 mg of whole IgM, lima bean trypsin inhibitor (LBI) (LBI:trypsin, 1:3 w/w) was added to inhibit further digestion. The mixture was dialyzed against BBS at 4 °C to remove urea, and the pentameric Fc fragments were purified on a column of ACA 22 (LKB) equilibrated in BBS. The μ-chain Fc fragments (yield of 7 mg) were obtained from the pentameric Fc fragment after complete reduction and alkylation of disulfide bridges, followed by purification on a column of ACA 34 (LKB) equilibrated in 3 M Gdn-HCl, 0.2 M ammonium bicarbonate, and 0.02% NaN_3.

Mild Acid Cleavage. For cleavage predominantly at aspartic acid–proline peptide bonds, completely reduced and alkylated 104E μ chain (45 mg) was dissolved in 10% acetic acid and 7 M Gdn-HCl, adjusted to pH 2.5 with pyridine, and incubated at 45 °C for 108 h with occassional mixing. The

resulting fragments were dialyzed against 3 M Gdn-HCl and 0.2 M ammonium bicarbonate at 4 °C, concentrated, and separated by gel filtration on a column (2.5 × 100 cm) of ACA 34 (LKB) equilibrated in 3 M Gdn-HCl and 0.2 M ammonium bicarbonate.

Amino Acid Analysis. Peptide samples of 2–10 nmol were flushed with N_2 and evacuated 3 times, sealed under vacuum, and hydrolyzed as described previously (Kehry et al., 1979).

Sequence Determination. All large fragments and peptides were sequenced by loading between 2 and 50 nmol on a modified Beckman sequenator with an automatic PTH conversion system with trifluoroacetic acid (Hunkapiller & Hood, 1978, 1980; Wittmann-Liebold, 1973; Wittmann-Liebold et al., 1976). Samples were loaded into a spinning cup coated with 10 mg of polybrene (Aldrich) (containing 20 nmol of glycylglycine and prerun for five to seven cycles) in trifluoroacetic acid (Pierce) and 10–20% water. The CN3 fragment was loaded in dilute NH_4OH, and the small peptides generated by enzymatic cleavages were dissolved in water. The first step was coupled twice. Aliquots (5–30% of each cycle, depending upon the amount of protein loaded and the number of sequenator cycles) of the amino acid PTH derivatives were analyzed by high-performance liquid chromatography (Waters Associates, Du Pont) as described previously (Hunkapiller & Hood, 1978; Johnson et al., 1979).

DNA sequences were determined by the method of Maxam & Gilbert (1977), by using the G (alternative), A > C, T + C, and C reactions as described previously (Rogers et al., 1980). For homology comparisons domains were defined on the basis of intron locations in the μ- and γ-chain genes (Calame et al., 1980; Tucker et al., 1979a,b) and domain homology in the α and ε chains (Beale & Feinstein, 1976).

Results

Protein Sequence Strategy. Cyanogen bromide cleavage of the 104E μ chain produced nine peptides (CN1–CN9) that were separated by gel filtration (Figure 1). Determination of the NH_2-terminal sequence of each peptide showed that each peak gave a single amino acid sequence. Amino-terminal sequence homology of the 104E C_H region cyanogen bromide peptides to the human μ chain sequence (Putnam et al., 1973) permitted determination of a presumptive alignment of peptides (top portion of Figure 1). Additional sequence determinations completed the sequence of each fragment and in all but two cases directly overlapped the fragments. Fragment CN1 was not isolated but was found to migrate between fragments CN1-2 and CN2 on the ACA 54 column by chromatographing [^{3}H]leucine- and [^{3}H]tyrosine-labeled 104E μ chain cyanogen bromide fragments (not shown). Incomplete cleavage of the methionine–serine bonds at positions 20 and 568 by cyanogen bromide resulted in the isolation of two partial fragments, CN1-2 and CN8-9, respectively, as indicated in Figure 1.

The complete amino acid sequence of the 104E μ chain is shown in Figure 2, along with the data used to determine the sequence. The amino acid and carbohydrate composition of the 104E μ chain and the cyanogen bromide fragments as derived from amino acid analyses and the sequence data are given in Table I. There are three amino acid differences from the previously published report (Kehry et al., 1979) that were detected in repeated sequenator runs; position 130 (alanine to serine) was miscalled due to contaminating sequences, and residues 233 (histidine to glutamic acid) and 545 (glutamic acid to histidine) were identified by using different analytical methods for amino acid PTH separation (McMillan et al., 1977; Johnson et al., 1979). In addition, we have determined

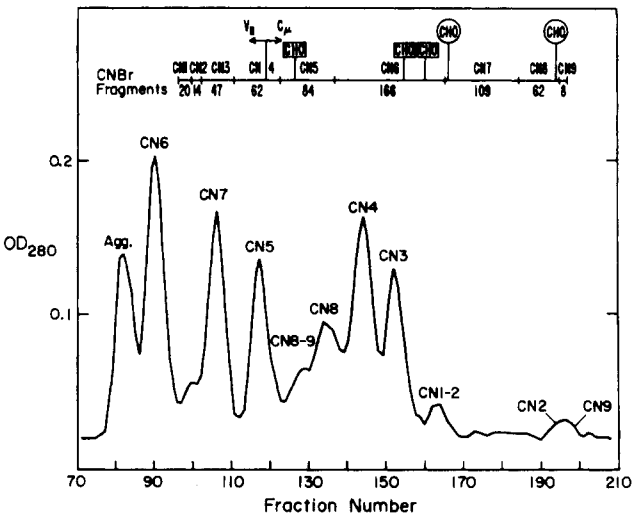


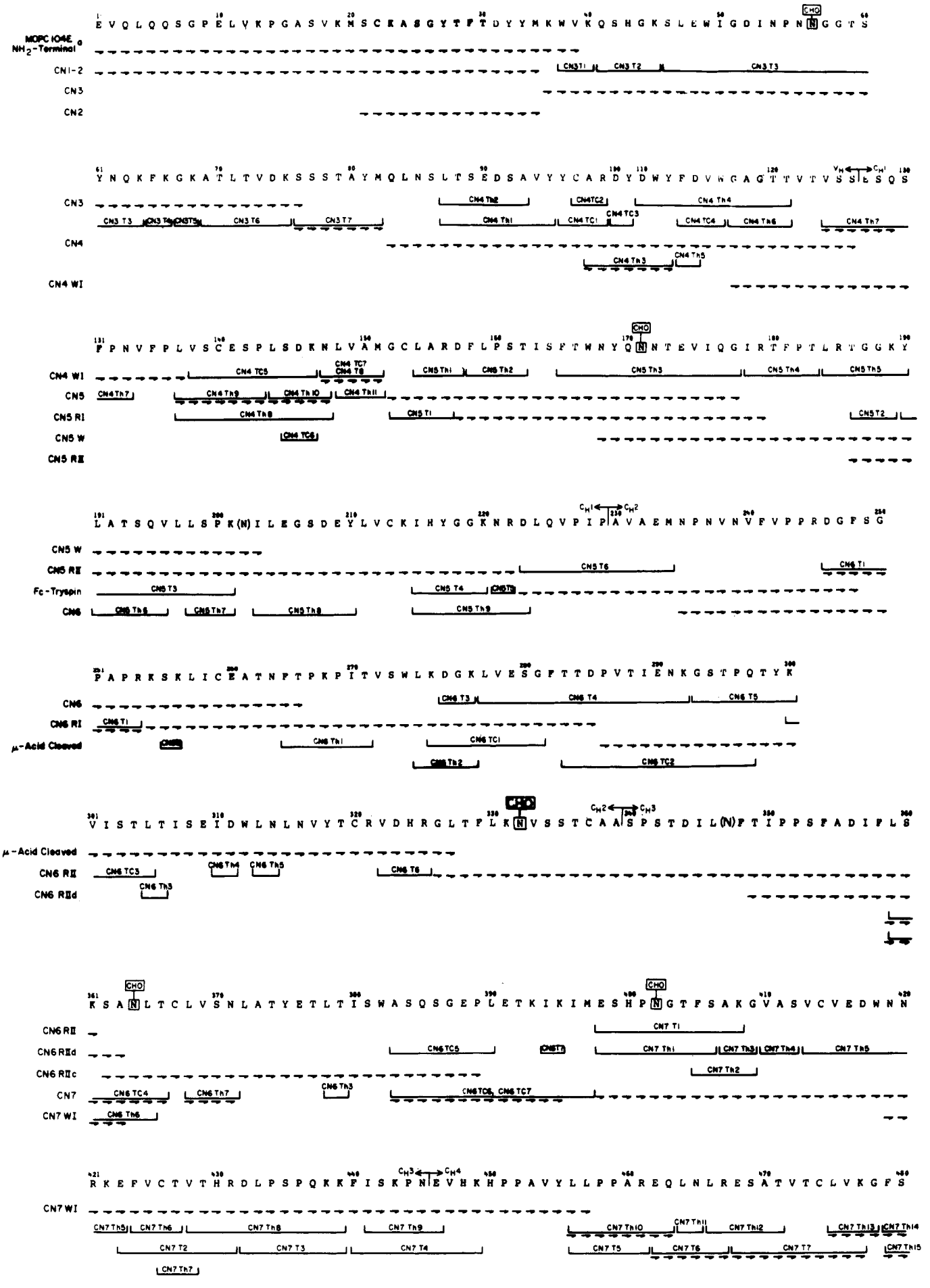
FIGURE 1: Separation of MOPC 104E cyanogen bromide fragments. Completely reduced and alkylated 104E μ chains (35 mg) were cleaved with cyanogen bromide and the fragments separated by gel filtraton on a column of ACA 54 (LKB) as described under Experimental Procedures. The fraction volume was 5 mL. The fragments are labeled (CN1–CN9) according to their position in the μ-chain sequence beginning at the NH_2 terminus. Agg. is the excluded column peak that consists of aggregated fragments and large uncleaved peptides. At the top of the figure is a schematic drawing of the 104E μ chain (to scale) showing the linear order and sizes of the cyanogen bromide fragments (CN1–CN9). Two incomplete cleavage products, CN1-2 and CN8-9, result from partial cleavage of the methionine residues at positions 20 and 568, respectively. CHO denotes the sites of carbohydrate attachment. The complex-type carbohydrates are indicated by boxes, and the high-mannose carbohydrates are indicated by circles.

the type of oligosaccharide attached to each of the six carbohydrate sites in the 104E μ chain.

Sequence of Cyanogen Bromide Fragments and Alignment. (*a*) *Sequence of CN1 and CN2 Fragments.* Amino-terminal sequence analysis of the first 38 residues of the intact 104E μ chain (Barstad et al., 1978) established the order of fragments CN1–CN3 (Figure 1). A sequenator run on the entire CN1-2 fragment verified its identity as residues 1–34. The sequence of CN2 as residues 21–34 also was completely determined in a single sequenator run (Figure 2).

(*b*) *Sequence of CN3 Fragment.* The CN3 fragment extends from $lysine_{34}$ to $methionine_{82}$ and is 47 residues in length. The NH_2-terminal sequence plus sequence determination of the peptide CN3 T7 provided the overlap data that established the complete sequence of CN3 (Figure 2). The CN1 and CN3 fragments are the only μ-chain cyanogen bromide fragments that contain no cysteine residues. The second hypervariable region is located in CN3 and includes a carbohydrate moiety attached to $asparagine_{57}$ (see below).

(*c*) *Sequence of CN4 Fragment.* The CN4 fragment spans the variable–constant region junction from $glutamine_{83}$ to $methionine_{151}$ and is 62 residues in length. The sequence overlap between CN3 and CN4 has not been rigorously established, but because this variable region sequence can be aligned without sequence gaps with those of other heavy chains (Kabat et al., 1976), the CN3 and CN4 fragments are probably contiguous. The NH_2-terminal sequence of CN4 was determined to the variable–constant region junction. The remaining sequence was obtained and confirmed by determining the sequences of the major peptide produced by tryptophan cleavage (Ozols et al., 1977) and the tryptic and thermolysin peptides CN4 Th3, CN4 Th7, CN4 Th9, CN4 Th10, and CN4 T8 (Figure 2). The one inserted residue in the mouse μ chain at the end of this fragment (position 148)


MOPC 104E
NH2-Terminal
CN1-2
CN3
CN2
CN3 T1
CN3 T2
CN3 T3
CN3 T7
CN3 T6
CN4 Th2
CN4 Th4
CN4TC2
CN4 TC3
CN4 Th1
CN4 TC1
CN4 TC4
CN4 Th6
CN4 Th7
CN4
CN4 Th3
CN4 Th5
CN4 WI
CN4 TC5
CN4 TC7
CN4 T8
CN5 Th1
CN5 Th2
CN5 Th3
CN5 Th4
CN5 Th5
CN5
CN4 Th7
CN4 Th9
CN4 Th 10
CN4 Th11
CN5 RI
CN4 Th8
CN5 T1
CN5 T2
CN5 W
CN4 TC6
CN5 RII
CN5 W
CN5 RII
CN5 T6
CN6 T1
Fc-Trypsin
CN5 T3
CN5 T4
CN6
CN6 Th6
CN6 Th7
CN6 Th8
CN5 Th9
CN6
CN6 T3
CN6 T4
CN6 T5
CN6 RI
CN6 T1
μ-Acid Cleaved
CN6 Th1
CN6 TC1
CN6 Th2
CN6 TC2
μ-Acid Cleaved
CN6 RII
CN6 TC3
CN6 Th4
CN6 Th5
CN6 T6
CN6 Th3
CN6 RIId
CN6 RII
CN7 T1
CN6 RIId
CN6TC5
CN7 Th1
CN7 Th3
CN7 Th4
CN7 Th5
CN6 RIIc
CN7 Th2
CN7
CN6 TC4
CN6 Th7
CN6 Th3
CN6TC6, CN6 TC7
CN7 WI
CN6 Th6
CN7 WI
CN7 Th5
CN7 Th6
CN7 Th8
CN7 Th9
CN7 T2
CN7 T3
CN7 T4
CN7 Th7
CN7 Th10
CN7 Th11
CN7 Th12
CN7 Th13
CN7 Th14
CN7 T5
CN7 T6
CN7 T7
CN7 Th15

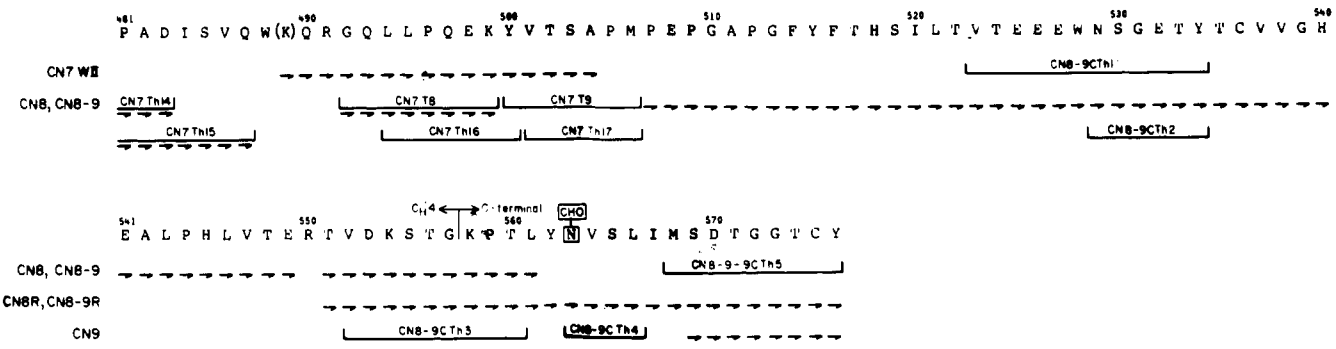


FIGURE 2: Complete sequence of MOPC 104E μ chain. The one-letter code for amino acids is given in Dayhoff (1976). Tryptic (T), chymotryptic (C), and thermolysin (Th) peptides are indicated. The sequences of peptides produced by cleavage at methionine (CN), arginine (R), and tryptophan (W) residues are also indicated in the left-hand margin. All peptides are identified by the cyanogen bromide fragment number from which they were derived. Residues sequenced automatically are indicated by forward half-arrows. CHO denotes the sites of carbohydrate attachment to asparagine residues. Weakly identified residues in the degradations are assigned by parenthese. The constant region domain boundaries are from DNA sequence determinations on a cloned BALB/c μ gene (Calame et al., 1980). For facilitation of homology comparisons, the numbering of the 104E μ chain corresponds to the amino acid sequence of the human μ chain OU (Putnam et al., 1973), and the total amino acid residues in each 104E μ chain fragment are not necessarily obtained by counting the assigned residue numbers (Kehry et al., 1979). Footnote *a* denotes the NH_2-terminal sequence from intact μ chain (Barstad et al., 1978).

Table I: Amino Acid Composition of CNBr Fragments from 104E μ Chain[a]

	CN1-2[b]	CN3	CN4	CN5	CN6[c]	CN7	CN8	CN9	whole μ chain[d]
CMCys[e]	0.3 (1)		2.1 (2)	2.5 (2)	2.8 (4)	2.1 (3)	1.3 (1)	0.7 (1)	14
Asx	1.1 (1)	6.0 (6)	7.7 (8)	7.7 (8)	16.8 (18)	8.3 (8)	3.8 (3)	0.9 (1)	
Asp[f]	(1)	(2)	(5)	(3)	(7)	(3)	(1)	(1)	23
Asn[f]		(4)	(3)	(5)	(11)	(5)	(2)		31
Thr	1.9 (2)	3.9 (4)	4.3 (4)	7.0 (7)	17.9 (20)	5.7 (6)	8.8 (9)	1.7 (2)	53
Ser	3.4 (4)	5.2 (6)	8.1 (10)	6.2 (5)	14.7 (18)	7.7 (9)	4.1 (4)	1.0 (1)	56
Glx	5.2 (5)	3.5 (3)	5.3 (5)	7.4 (8)	10.4 (9)	13.1 (13)	7.0 (7)	1.0 (1)	
Glu[f]	(2)	(1)	(3)	(4)	(7)	(7)	(7)		31
Gln[f]	(3)	(2)	(2)	(4)	(2)	(6)			19
Pro	1.9 (2)	1.2 (1)	3.5 (3)	5.5 (5)	12.4 (13)	13.4 (11)	4.6 (5)		40
Gly	3.3 (3)	5.0 (5)	2.9 (2)	7.1 (7)	8.5 (7)	4.7 (4)	6.2 (5)	2.6 (2)	35
Ala	2.2 (2)	2.2 (2)	4.1 (4)	4.3 (4)	8.5 (8)	7.1 (7)	2.3 (2)	0.4	30
Val	3.2 (3)	2.3 (2)	6.9 (7)	5.4 (5)	12.0 (11)	10.6 (11)	5.3 (6)		45
Met									8
Ile		2.0 (2)		5.3 (6)	10.4 (12)	2.2 (2)	1.8 (2)		24
Leu	2.1 (2)	2.2 (2)	5.0 (5)	9.2 (9)	14.9 (15)	8.9 (8)	4.8 (5)		46
Tyr	2.6 (3)	1.8 (2)	3.8 (4)	3.7 (4)	3.0 (3)	1.9 (2)	2.9 (3)	0.3 (1)	22
Phe	0.9 (1)	1.1 (1)	3.1 (3)	3.2 (3)	8.0 (8)	4.0 (4)	1.9 (2)		22
Trp[g]		(2)	(2)	(1)	(3)	(2)	(1)		11
His		0.8 (1)		1.8 (1)	1.5 (1)	3.6 (4)	2.6 (3)		10
Hse[h]	1.1 (2)	0.7 (1)	0.7 (1)	0.5 (1)	0.3 (1)	0.7 (1)	0.4 (1)		
Lys	3.1 (3)	6.9 (7)	1.7 (1)	4.3 (4)	10.4 (11)	7.3 (9)	1.8 (2)		37
Arg			1.1 (1)	3.0 (4)	4.5 (4)	4.5 (5)	1.0 (1)		15
Glc NH_2[i]		0.2		2.9	3.8	1.7	1.6		
CHO moiety[i]		(1)		(1)	(2)	(1)	(1)		6
total	34	47	62	84	166	109	62	8	572
residue no.[j]	1-34	35-82	83-151	152-234	235-397	398-506	507-568	569-576	
M_r[k]	3720	5180	6895	9640	18 118	12 230	6815	800	63 213

[a] Values reported are amino acid residues determined by amino acid analyses of CNBr fragments isolated as in Figure 1. Amino acids present at a level of less than 0.2 residue are omitted. Values in parentheses represent the number of residues present in the completed sequence. [b] Separate analyses of the CN1 and CN2 fragments were not done. [c] Due to the large size of the CN6 fragment and its proximity to the peak of aggregated and uncleaved material when isolated by gel filtration on ACA 54 (see Figure 1), the amino acid composition derived from hydrolysis is not accurate. [d] Values of composition of the whole μ chain are summarized from the completed sequence. [e] This was detected as the (carboxymethyl)cysteine derivative. Quantitation was not possible, but the peak was in every case consistent with the numbers determined from sequence data. [f] Values for Asx and Glx have been divided into the acid and amide assignments from sequence determinations. [g] Trp values are from sequence determinations. [h] This includes homoserine as well as homoserine lactone. [i] Number of carbohydrate attachment sites in each cyanogen bromide fragment is expressed. Locations: CN3, residue 57; CN5, residue 171; CN6 A, residue 332; CN6 B, residue 364; CN7, residue 402; CN8, residue 563. The presence of glucosamine was determined by a separate hydrolysis in 6 N HCl at 110 °C for 3 h. [j] The numbering does not correspond to the number of amino acids due to the introduction of gaps for sequence alignment. [k] Molecular weight is in daltons, calculated from the sequence of amino acids in the fragment; it does not include carbohydrate.

also has been confirmed independently by Milstein et al. (1975) through an analysis of the corresponding cysteine-containing peptide and by Kawakami et al. (1980) through a DNA sequence analysis of the mouse μ gene.

(*d*) *Sequence of CN5 Fragment.* The CN5 fragment extends from $glycine_{152}$ to $methionine_{234}$ and constitutes the major portion of the C_H1 domain. This 84-residue fragment also contains the first complex carbohydrate moiety in the constant region. Determination of the NH_2 terminal sequence plus the sequences of the two arginine fragments (CN5 RI and CN5 RII) and the fragment generated by cleavage at the tryptophan residue (CN5 W) completed the sequence of the CN5 frag-

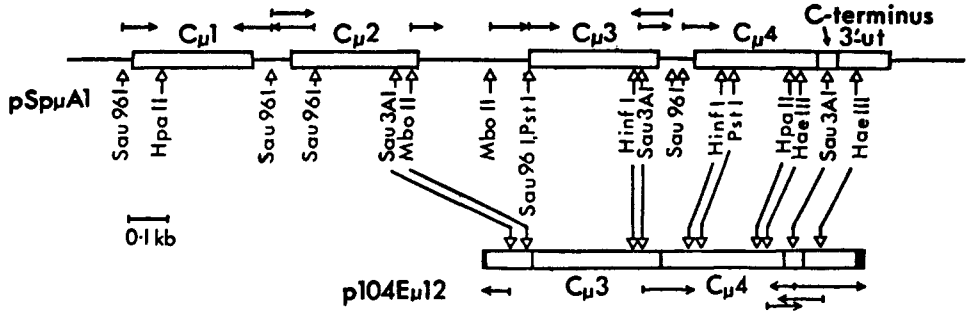


FIGURE 3: Summary of DNA sequences obtained from the C_μ gene. These maps are adapted from Calame et al. (1980). The restriction sites used for sequencing are shown by open arrow heads. Lines show intervening DNA sequences in the C_μ gene that are removed from the primary RNA transcript, and open boxes are regions that are present in the mature secreted μ-chain mRNA. Filled-in boxes at the 3′ and 5′ ends of p104Eμ12 represent a $(dA)_{15}$ segment from the poly(A) of the mRNA plus a synthetic $(dC)_{13}$ linker and the reconstituted plasmid *PstI* site. Horizontal arrows indicate sequenced regions. See Table II for the codon numbers. pSpμA1 is a subclone from genomic clone ChSpμ7; p104Eμ12 is a cDNA clone of secreted μ chain from MOPC 104E. The C_μ4 domain 3′ to the *Hpa*II site was also sequenced in cDNA clone p104Eμ6 (not shown). For additional details and nucleotide sequences, see Calame et al. (1980) and Rogers et al. (1980).

ment (Figure 2). The Fc fragment produced by trypsin digestion of the IgM molecule in the presence of 5 M urea (Shimizu et al., 1975) provided the COOH-terminal sequence of CN5 and the overlap between CN5 and CN6. The first 30 amino acids of the Fc fragment contain seven proline residues that are in the boundary region between C_H1 and C_H2. This proline-rich stretch in the μ chain bears a striking resemblance to the proline-rich hinge regions of γ and α heavy chains (Aldersberg, 1976).

(*e*) *Sequence of CN6 Fragment.* The CN6 fragment encompasses the entire C_H2 domain and half of the C_H3 domain in its 166 residues. Extending from asparagine$_{235}$ to methionine$_{397}$, CN6 contains two complex carbohydrate moieties, one in C_H2 and one in C_H3 (see below and Figure 2). The amino acid sequence of the first half of CN6 was established by Edman degradations of the NH_2 terminus and of the two arginine fragments (CN6 RI and CN6 RII). Cleavage of the one aspartic acid–proline peptide bond in the fully reduced and alkylated 104E μ chain by mild acid produced a fragment with an NH_2 terminus at proline$_{286}$. The sequence of this acid-derived fragment was then overlapped with the NH_2 terminus of the CN6 RII arginine fragment. The remaining COOH-terminal sequence of CN6 was determined by isolating CN6 RII, followed by demaleylating or decitraconylating the lysine residues and digesting the fragment with trypsin. Four predominant amino termini resulted from this procedure: CN6 RIIa, beginning at the NH_2 terminus of CN6 RII, glycine$_{326}$; CN6 RIIb, produced in low yield due to incomplete cleavage of the peptide bond between lysine$_{331}$ and glycosylated asparagine$_{332}$; CN6 RIIc, beginning at serine$_{362}$; and CN6 RIId, the result of a chymotryptic cleavage at phenylalanine$_{348}$. The sequence was confirmed and completed by the isolation of the thermolysin and trypsin plus chymotrypsin peptides, CN6 TC4, CN6 Th7, and CN6 TC6 (Figure 2). The overlap between CN6 and CN7 has not been determined, but the sequences of CN6 and CN7 are contiguous when aligned by homology with two human μ chains and a dog μ chain (Watanabe et al., 1973; Putnam et al., 1973; Wasserman & Capra, 1978). Therefore, we believe that no additional cyanogen bromide fragment is located between CN6 and CN7.

(*f*) *Sequence of CN7 Fragment.* The CN7 fragment, which is 109 residues in length (glutamic acid$_{398}$ to methionine$_{506}$), spans the remaining portion of C_H3 and the first two-thirds of the C_H4 domain (Figure 2). The one high-mannose carbohydrate moiety in CN7 is attached to asparagine$_{402}$ in the C_H3 domain, making C_H4 the only μ chain constant region domain that has no covalently attached oligosaccharide moiety. In addition to an NH_2-terminal sequence, the sequences of the two tryptophan fragments (CN7 WI and CN7 WII) and the sequences of the overlapping thermolysin and tryptic peptides, CN7 Th10, CN7 T6, CN7 T7, CN7 Th13, CN7 Th14, CN7 Th15, and CN7 T8, completed the covalent structure of the CN7 fragment (Figure 2).

(*g*) *Sequence of CN8, CN8-9, and CN9 Fragments.* NH_2-terminal sequenator runs on CN8 and CN8-9, beginning at proline$_{507}$, established the sequence through tyrosine$_{562}$. Cleavage at the single arginine residue in CN8-9 allowed determination of the remaining sequence of the COOH-terminal segment (to tyrosine$_{576}$). Sequence determination on the arginine fragment of CN8 indicated that at least two-thirds of this peak lacked residues 569–576 (Figure 2). Isolation and sequence analysis of the CN9 fragment from a cyanogen bromide digestion of ^{14}C-alkylated 104E μ chains confirmed the cyanogen bromide cleavage between methionine$_{568}$ and serine$_{569}$. The high-mannose carbohydrate attached to asparagine$_{563}$ in the COOH-terminal segment was identified from a chymotryptic and thermolysin peptide (CN8-9 CTh4). The sequence of a peptide that overlaps CN7 and CN8 has been determined by Milstein et al. (1975).

DNA Sequences. Recently, several C_μ gene clones have been obtained, both from BALB/c germ-line chromosomal DNA and from MOPC 104E μ chain messenger-complementary DNA (Calame et al., 1980; Kawakami et al., 1980). As is shown in Table II, our previous DNA sequence analysis of clones (Calame et al., 1980; Rogers et al., 1980), as well as additional sequences analyses, provides data covering approximately 50% of the 104E μ chain constant region (220 out of 450 residues). A schematic drawing of the C_μ gene illustrating various restriction sites and summarizing the DNA sequence analyses on all of the clones in shown in Figure 3. These DNA sequences confirm the protein sequence at all positions except for two discrepancies (asparagine$_{347}$ to threonine and threonine$_{379}$ to asparagine). Since the cloning of this gene involved no in vitro copying of the DNA, it is unlikely that the differences are due to a cloning artifact, unless a mutation has occurred during replication in *Escherichia coli*. These residues also seem to be clearly identified in the automatic sequenator runs. Thus the protein sequence presented here conflicts with the DNA data at these two positions.

Recently, Kawakami et al. (1980) have completed the nucleotide sequence of a BALB/c mouse C_μ germ-line gene. Their published sequence disagreed with our previous report (Kehry et al., 1979) at eight positions. The five differences that we detected by protein sequencing (residues 130, 233, and 545) and DNA sequencing (residues 347 and 379) mentioned above are all confirmed by the sequence presented by Ka-

Table II: Sequenced Portions of C_μ Gene[a]

codons (inclusive)	clone	reference
127-153	μA1	M. Kehry et al., unpublished results
218-250	μA1	M. Kehry et al., unpublished results
300-323	μ12	Calame et al., 1980
336-388	μA1	M. Kehry et al., unpublished results
432-466	μA1	M. Kehry et al., unpublished results
432-470	μ12	M. Kehry et al., unpublished results
531-556	μ6	Rogers et al., 1980
542-576	μ12	Calame et al., 1980

[a] Numbering is as for OU. For a description of the clones and maps showing sequencing strategies, see Figure 3, Calame et al. (1980), and Rogers et al. (1980). μA1 = SpμA1; this is a subclone from ChSpμ7, which contains chromosomal DNA from BALB/c sperm. μ6 = p104Eμ6, and μ12 = p104Eμ12; these contain complementary DNA from MOPC 104E mRNA. The sequences of codons 218-250 and 542-576 were obtained from two strands; the remaining sequences were obtained from only one of the two strands.

wakami et al. (1980). Of the remaining three differences, one (residue 225) is in agreement with both our protein and DNA sequence analyses. We have no DNA sequence information on the other two residues (202 and 489), and since they are weakly identified in the protein sequence analyses, they have been enclosed by parentheses in Figure 2. These sequence differences may represent mutations during the multiple passages of the 104E cells, a natural polymorphism for the C_μ gene, or technical errors in the protein or DNA sequencing.

Carbohydrate Moieties. Six carbohydrate moieties are attached to asparagine residues in the 104E μ chain. One site is located in the V_H region on asparagine$_{57}$ (Figure 2). The residues that are glycosylated were identified as described (Kehry et al., 1979).

In order to identify whether the carbohydrates attached to the 104E μ chain were high-mannose or complex type, we labeled galactose residues on intact 104E IgM molecules extrinsically with galactose oxidase and sodium boro[^{3}H]hydride after treatment with neuraminidase (Kehry et al., 1980). Only the complex carbohydrate moieties contain galactose residues and would be labeled by this procedure. The 104E μ chain cyanogen bromide fragments CN5 and CN6 were found to contain labeled galactose residues in a ratio of 1:2.5 CN5:CN6 (Kehry et al., 1980). Since amino acid sequence analyses show that the CN6 fragment contains two sites of carbohydrate attachment, the galactose labeling establishes that the first three constant region oligosaccharides attached to residue 171 in the CN5 fragment and residues 332 and 364 in the CN6 fragment are of the complex type. All of the C_μ region carbohydrate moieties were labeled biosynthetically by incubating 104E myeloma cells with medium containing [^{3}H]mannose (Kehry et al., 1980). Thus, the cyanogen bromide fragments CN7 and CN8 plus CN8-9 contain mannose but not galactose, and the carbohydrates attached to these fragments (at residues 405 and 563) are high-mannose oligosaccharides. The carbohydrate moiety attached to fragment CN3 at asparagine$_{57}$ in the V_H region contains some glucosamine by ninhydrin analysis but does not contain galactose or mannose as judged from the labeling studies (Kehry et al., 1980). This must be a unique small carbohydrate structure, not of the high-mannose or complex type.

Discussion

Complete μ-Chain Sequence. The complete amino acid sequence of the MOPC 104E μ chain has been determined (Figure 2). The C_μ region can be subdivided into four domains of approximately 110 residues whose boundaries have previously been delineated by comparisons of internal heavy-chain amino acid homology with the three-dimensional structure of the immunoglobulin molecule (Beale & Feinstein, 1976). In addition, a 20-residue COOH-terminal segment is located COOH terminally to the C_μ4 domain in the secreted μ chains.

Immunoglobulin heavy-chain domains also are known to be discrete coding elements at the level of the heavy-chain gene and are separated from one another by short intervening DNA sequences (Early et al., 1979; Sakano et al., 1979a,b; Honjo et al., 1979; Yamawaki-Kataoka et al., 1980; Calame et al., 1980; Tucker et al., 1979a,b). The coding region for the COOH-terminal segment of the secreted μ chain, though, is contiguous with that of the C_μ4 domain (Calame et al., 1980; Kawakami et al., 1980). The precise domain junctions that have been determined by DNA sequence analysis of BALB/c μ genomic and complementary DNA clones (Calame et al., 1980; Kawakami et al., 1980) are in remarkable agreement with those determined from the μ-chain amino acid sequence (Beale & Feinstein, 1976; Kehry et al., 1979). For the secreted μ chain, there are five domain boundaries that have been described (Beale & Feinstein, 1976), and four of these boundaries match those determined from the μ-gene DNA sequence or are only shifted by one amino acid (Calame et al., 1980). The remaining boundary between C_μ1 and C_μ2 in the μ-chain gene is shifted by just three amino acids from the boundary previously described from the protein structure.

Carbohydrate Moieties in the μ Chain. Two high mannose type and three complex-type carbohydrate moieties are attached to asparagine residues in the μ chain constant region, with only the C_μ4 domain lacking carbohydrate. Both types of oligosaccharide structures are composed of a core carbohydrate containing two *N*-acetylglucosamine and five mannose residues (Kornfeld & Kornfeld, 1976). The high-mannose structures are synthesized by the additions of glucosamine, galactose, fucose, and *N*-acetylneuraminic acid to the core structure (Kornfeld & Kornfeld, 1976). In addition, a carbohydrate structure that is not of the high-mannose or complex type is located in the 104E second hypervariable region.

As noted previously, four of the five C_μ-region carbohydrates are attached to identical positions in the mouse and human μ chains (Kehry et al., 1979). Interestingly, the DNA sequence discrepancy that suggests that position 379 is an asparagine residue would create an additional recognition sequence for glycosylation. In the mouse μ chain there is an oligosaccharide that is attached to asparagine$_{364}$ (Kehry et al., 1979). This oligosaccharide is in a different location in the human μ chain (asparagine$_{395}$). This change in carbohydrate location is the result of a change in amino acid sequence between mouse and human μ chains. Thus, if residue 379 of the mouse C_μ region is asparagine, the selective pressure in this region to form a recognition sequence for glycosylating enzymes is remarkably high.

The general type of carbohydrate moiety, complex or high mannose, is completely conserved between mouse and human μ chains (Shimizu et al., 1971). Proceeding from the C_μ1 to the COOH-terminal segment, one finds that the first three oligosaccharides attached to the μ chain are of the complex type, containing terminal galactose and presumably *N*-acetylneuraminic acid residues. The remaining two oligosaccharides, including the one in the COOH-terminal segment, are high-mannose carbohydrate structures composed of glucosamine and mannose. Apparently, the glycosylating enzymes involved in the synthesis of these distinct types of oligosaccharides efficiently discriminate among multiple recognition

sites on the same molecule. The complex and high-mannose oligosaccharides then presumably have different structural or functional roles in the IgM molecule that require a precise localization.

Carbohydrate structures attached to heavy chains of different classes, however, differ in type and location (Torano et al., 1977). For example, most human γ chains possess one oligosaccharide N-glycosidically linked to an asparagine residue in the C_H2 domain (Kornfeld et al., 1971; Beale & Feinstein, 1976). Human α chains differ remarkably in glycosylation even between the α_1 and α_2 subclasses (Torano et al., 1977; Tsuzukida et al., 1979). Human ϵ chains, like μ chains, contain only N-linked oligosaccharides, three of which are located in the C_H1 domain, one of which is in the C_H2 domain, and two of which are in the C_H3 domain (Bennich & von Bahr-Lindstrom, 1974). However, the one oligosaccharide in γ chains (in the $C_\gamma2$ domain) is in a homologous position to an N-linked carbohydrate in the C_H3 domain of μ and ϵ chains (Beale & Feinstein, 1976), suggesting that either these domains have a common evolutionary origin or that they perform similar functions. The one common feature shared by all heavy-chain classes is that the last domain ($C_\gamma3$, $C_\alpha3$, $C_\mu4$, or $C_\epsilon4$) lacks a carbohydrate moiety entirely. In the case of γ and ϵ chains, a carbohydrate moiety may interfere with Fc receptor–immunoglobulin interactions in this COOH-terminal domain.

One additional exception to the lack of homology in the location and type of heavy-chain carbohydrate structures is the high-mannose oligosaccharide situated in the COOH-terminal segment of both μ (asparagine$_{563}$) and α (asparagine$_{459}$) chains (Putnam, 1974). This carbohydrate is located on homologous residues in human α, human μ, and mouse μ chain COOH-terminal segments. We envision that the conserved COOH-terminal high-mannose oligosaccharide plays a structural or functional role in the IgM and IgA polymer. Moreover, the amino acid sequence homology between human α and μ chain COOH-terminal segments is a striking 70% (14/20 identical residues) compared to a 40% homology between the remainder of the C_μ and C_α regions. This homology is reasonable in view of the fact that IgM and IgA are the only immunoglobulin classes that form polymers with the J chain (Della Corte & Parkhouse, 1973; Mestecky & Schrohenloher, 1974; Mestecky et al., 1974; Mendez et al., 1973) and are the only heavy chains that possess the 20-residue COOH-terminal segment. This region must be very important in the polymerization process and in the maintenance of the polymer structure. The importance of the COOH-terminal segment in polymer formation is strengthened by the fact that membrane IgM molecules, which exist only as monomeric immunoglobulins, exhibit a different COOH-terminal segment and lack the carbohydrate moiety present in secreted μ and α chains (Kehry et al., 1980; Rogers et al., 1980; Early et al., 1980).

The composition of the COOH-terminal high-mannose oligosaccharide is different from that of the other oligosaccharides in the μ chain. Quantitation of [^{3}H]mannose incorporated into 104E μ chain cyanogen bromide fragments correlates well with the number of carbohydrate moieties attached to each fragment (Kehry et al., 1980). The CN7 and CN5 fragments, each with one carbohydrate, and the CN6 fragment, with two carbohydrates, are labeled with [^{3}H]-mannose in a 1:1:2.5 ratio (CN7:CN5:CN6). However, quantitation of [^{3}H]mannose-labeled CN8 plus CN8-9 fragments shows that only 0.5 carbohydrate moiety is attached to the COOH-terminal segment. Either this carbohydrate structure contains half as many mannose residues as a mature high-mannose oligosaccharide, or alternatively, only a portion of the μ chains are glycosylated at this position. We favor the latter explanation in view of our amino acid sequence data on the arginine fragment of CN8. In the automatic sequenator, the asparagine residue to which a carbohydrate moiety is attached is not extracted from the sequentor cup and therefore shows up as a blank cycle. In the sequence of the CN8 arginine fragment, asparagine$_{563}$ showed approximately 0.5 residue of asparagine, indicating that some of the 104E μ chains are unglycosylated in the COOH-terminal region. Moreover, the fractional mannose incorporation into the CN8 fragment is reproducibly seen in other secreted μ chains and in internal precursors to secreted μ chains (Kehry et al., 1980), suggesting that it may be significant at the structural and functional level.

Domain Homologies among Heavy Chains. (*a*) *μ-Chain Homologies.* When the mouse, dog, and human C_μ sequences are compared, there is an increasing homology from the $C_\mu1$ domain to the COOH-terminal segment (Kehry et al., 1979). We have previously shown that among mouse, human, and dog μ chains, the $C_\mu4$ domain is the most highly conserved of the four μ chain constant region domains (Kehry et al., 1979). Interestingly, the binding site for the first component of the complement cascade has been localized to a region in $C_\mu4$, residues 468–491 joined to residues 515–546 (Hurst et al., 1975). These two stretches of amino acids are at least 75% identical in the sequences of the mouse human, and dog μ chains. This complement-binding region makes a significant contribution to the high degree of $C_\mu4$ sequence conservation among the three species and presumably reflects the structural and/or functional constraints related to complement fixation that have been imposed upon this distinct region during species evolution.

(*b*) *Homologies among μ, γ, α, and ϵ Chains.* An interesting question in the evolution of the different heavy-chain classes is how additional domains have evolved (four C_ϵ and C_μ vs. three C_γ and C_α domains) or how regions have been deleted [no hinge region in μ or ϵ chains; lack of a $C_\gamma2$ region in mouse δ chains (Liu et al., 1980; Tucker et al., 1980)]. The intervening DNA sequences between domains in heavy-chain genes provide an excellent mechanism for rearranging, duplicating, or deleting individual domains. We therefore examined the homologies among all of the completely sequenced C_H domains in hope of identifying a primordial domain or pair of domains from which all heavy-chain genes evolved by gene duplication. These comparisons, however, were unsuccessful in identifying a primordial domain (T. Hunkapiller and L. Hood, unpublished observations). Interestingly, this result may reflect the evolutionary history of immunoglobulin gene segments, suggesting that, as has been observed for the γ-chain subclasses (Miyata et al., 1980), the C_H gene segments may have recombined with one another prior to the divergence of classes. Of particular importance is the fact that the most primitive μ-like chains found in lampreys and hagfish display a molecular weight characteristic of the four domain C_μ region of mammals (Raison et al., 1978). Emergence of a second class of H chains distinct from μ chains is first seen in a few higher bony fishes (Hildemann et al., 1981). Therefore, the C_μ region domains in the first IgM molecules produced by primitive vertebrates may have had ample opportunity to recombine with one another prior to the divergence of H-chain classes.

Acknowledgments

We gratefully acknowledge the technical assistance of V.

Farnsworth and P. Cartier, the teaching of sequence analysis by M. W. Hunkapiller, and the assistance of R. Miake in the early stages of this work.

Supplementary Material Available

Tables giving (a) amino acid compositions of peptides obtained by enzymatic digestion of μ-chain cyanogen bromide fragments and (b) sequences of small peptides and quantitative graphs of all sequenator runs on the μ-chain cyanogen bromide fragments, arginine peptides, and tryptophan peptides, which establishes the complete μ-chain sequence (34 pages). Ordering information is given on any current masthead page.

References

Aldersberg, J. B. (1976) *Ric. Clin. Lab. 6*, 191.

Barstad, P., Hubert, J., Hunkapiller, M., Goetze, A., Schilling, J., Black, B., Eaton, B., Richards, J., Weigert, M., & Hood, L. (1978) *Eur. J. Immunol. 8*, 497.

Beale, D., & Feinstein, A. (1976) *Q. Rev. Biophys. 9*, 135.

Bennich, H., & von Bahr-Lindstrom, H. (1974) *Progress in Immunology, Proceedings of the International Congress of Immunology, 2nd* (Brent, L., & Holborrow, J., Eds.) Vol. 1, p 49, North-Holland, Amsterdam.

Calame, K., Rogers, J., Early, P., Davis, M., Livant, D., Wall, R., & Hood, L. (1980) *Nature (London) 284*, 452.

Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure*, Vol. 5, pp 189–190, National Biomedical Research Foundation, Silver Spring, MD.

Della Corte, E., & Parkhouse, R. M. E. (1973) *Biochem. J. 136*, 597.

Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N., & Hood, L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 857.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., & Hood, L. (1980) *Cell (Cambridge, Mass.) 20*, 313.

Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U., & Waxdal, M. J. (1969) *Proc. Natl. Acad. Sci. U.S.A. 63*, 78.

Gough, N. M., Kemp, D. J., Tyler, B. M., Adams, J. M., & Cory, S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 554.

Gross, E. (1967) *Methods Enzymol. 11*, 238.

Hildemann, W. H., Clark, E. A., & Raison, R. L. (1981) *Comprehensive Immunogenetics*, Elsevier/North-Holland, New York, NY.

Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N., & Mano, Y. (1979) *Cell (Cambridge, Mass.) 18*, 558.

Hunkapiller, M. W., & Hood, L. E. (1978) *Biochemistry 17*, 2124.

Hunkapiller, M. W., & Hood, L. E. (1980) *Science (Washington, D.C.) 207*, 523.

Hurst, M. M., Valanakis, J. E., Stroud, R. M., & Bennett, J. C. (1975) *J. Exp. Med. 142*, 1322.

Johnson, N., Hunkapiller, M., & Hood, L. (1979) *Anal. Biochem. 100*, 335.

Kabat, E. A., Wu, T. T., & Bilofsky, H. (1976) *Variable Regions of Immunoglobulin Chains*, Medical Computer Systems, Cambridge, MA.

Kataoka, T., Kawakami, T., Takahashi, N., & Honjo, T. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 919.

Katz, A. M., Dreyer, W. J., & Anfinsen, C. B. (1959) *J. Biol. Chem. 234*, 2897.

Kawakami, T., Takahashi, N., & Honjo, T. (1980) *Nucleic Acids Res. 8*, 3933.

Kehry, M., Sibley, C., Fuhrman, J., Schilling, J., & Hood, L. E. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2932.

Kehry, M., Ewald, S., Douglas, R., Raschke, W., Sibley, C., Fambrough, D., & Hood, L. (1980) *Cell (Cambridge, Mass.) 21*, 393.

Klapper, M. H., & Klotz, E. M. (1972) *Methods Enzymol. 25*, 531.

Konigsberg, W. (1972) *Methods Enzymol. 25*, 185.

Kornfeld, R., & Kornfeld, S. (1976) *Annu. Rev. Biochem. 45*, 217.

Kornfeld, R., Keller, J., Baenziger, J., & Kornfeld, S. (1971) *J. Biol. Chem. 246*, 3259.

Laemmli, U. (1970) *Nature (London) 227*, 680.

Lawton, A. R., Kincade, P. W., & Cooper, M. D. (1975) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 34*, 33.

Liu, C.-P. Tucker, P. W., Mushinski, J. F., & Blattner, F. R. (1980) *Science (Washington, D.C.) 209*, 1348.

Marchalonis, J. J. (1972) *Nature (London) 236*, 84.

Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560.

McMillan, M., Cecka, J. M., Murphy, D. B., McDevitt, H. O., & Hood, L. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5135.

Mendez, E., Prelli, F., Frangione, B., & Franklin, E. C. (1973) *Biochem. Biophys. Res. Commun. 55*, 1291.

Mestecky, J., & Schrohenloher, R. E. (1974) *Nature (London) 249*, 650.

Mestecky, J., Schrohenloher, R. E., Kulhavey, R., Wright, G. P., & Tomana, M. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 544.

Milstein, C. P., Richardson, N. E., Deverson, E. F., & Feinstein, A. (1975) *Biochem. J. 151*, 615.

Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M., & Honjo, T. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2143.

Ozols, J., Gerard, C., & Stachelek, C. (1977) *J. Biol. Chem. 252*, 5986.

Peacock, A. C., & Dingman, C. W. (1968) *Biochemistry 7*, 668.

Putnam, F. W. (1974) *Progress in Immunology, Proceedings of the International Congress of Immunology, 2nd* (Brent, L., & Halborrow, J., Eds.) Vol. 1, p 25, North-Holland, Amsterdam.

Putnam, F. W., Florent, G., Paul, C., Shinoda, T., & Shimizu, A. (1973) *Science (Washington, D.C.) 182*, 287.

Raison, R. L., Hull, C. J., & Hildemann, W. H. (1978) *Dev. Comp. Immunol. 2*, 253.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., & Wall, R. (1980) *Cell (Cambridge, Mass.) 20*, 303.

Sakano, H., Hüppi, K., Heinrich, G., & Tonegawa, S. (1979a) *Nature (London) 280*, 288.

Sakano, H., Rogers, J. H. Hüppi, K., Brack, C., Trannecker, A., Maki, R., Wall, R., & Tonegawa, S. (1979b) *Nature (London) 277*, 627.

Shimizu, A., Putnam, F. W., & Paul, C. (1971) *Nature (London) 231*, 73.

Shimizu, A., Watanabe, S., Yamamura, Y., & Putnam, F. W. (1975) *Immunochemistry 11*, 719.

Torano, A., Tsuzukida, Y., Liu, Y. S. V., & Putnam, F. W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2301.

Tsuzukida, Y., Wang, C. C., & Putnam, F. W. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1104.

Tucker, P. W., Marcu, K. B., Slightom, J. L., & Blattner, F. R. (1979a) *Science (Washington, D.C.) 206*, 1299.

Tucker, P. W., Marcu, K. B., Newell, N., Richards, J., & Blattner, F. R. (1979b) *Science (Washington, D.C.) 206*, 1303.

Tucker, P. W., Liu, C.-P., Mushinski, J. F., & Blattner, F. R. (1980) *Science (Washington, D.C.) 209*, 1353.
Vitetta, E. S., Bauer, S., & Uhr, J. W. (1971) *J. Exp. Med. 134*, 242.
Warner, N. (1974) *Adv. Immunol. 19*, 67.
Wasserman, R. L., & Capra, J. D. (1978) *Science (Washington, D.C.) 200*, 1159.
Watanabe, S., Barnikol, H. U., Horn, J., Bertram, J., & Hilschmann, N. (1973) *Hoppe-Seyler's Z. Physiol. Chem. 354*, 1505.
Wittmann-Liebold, B. (1973) *Hoppe-Seyler's Z. Physiol. Chem. 354*, 1415.
Wittmann-Liebold, B., Graffunder, H., & Kohls, H. (1976) *Anal. Biochem. 75*, 621.
Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N., Obata, M., & Honjo, T. (1980) *Nature (London) 283*, 786.

Cloning of a Complementary Deoxyribonucleic Acid Encoding a Portion of Rat Intestinal Preapolipoprotein AIV Messenger Ribonucleic Acid†

Jeffrey I. Gordon,* Dennis P. Smith, David H. Alpers, and Arnold W. Strauss

ABSTRACT: Apolipoprotein AIV is one of the principal apolipoproteins synthesized by the rat small intestine. We have cloned a cDNA encoding a portion of preapolipoprotein AIV mRNA. A kinetically fractionated cDNA probe highly enriched for the abundant intestinal mucosal mRNA sequences was used to screen a library of recombinants containing cDNA generated from total intestinal epithelial mRNA. The abundant class of mRNA sequences was defined by hybridization analyses. This frequency class had an aggregate complexity of 5300 nucleotides and represented 25% of accumulated mRNA sequences. The mRNAs comprising this class were identified by in vitro translation and included preapolipoprotein AIV, preproapolipoprotein AI, intestinal fatty acid binding protein, and liver fatty acid binding protein. A cDNA-containing clone derived from preapolipoprotein AIV mRNA was identified among probe-positive recombinants. This cDNA was used to establish that apolipoprotein AIV mRNA has a mass of 550 000 daltons (equivalent to 1780 nucleotides) and represents 0.013% of total cellular RNA in the fasting state. Acute feeding with triglyceride-rich meals resulted in a 2-fold increase in preapolipoprotein AIV mRNA after 4 h.

Apolipoprotein (apo)[1] AIV is one of the major apolipoproteins synthesized by the intestine (Wu & Windmueller, 1978; Imaizumi et al., 1978). Wu and Windmueller used in vivo perfusion with labeled amino acids to estimate that 59% of rat plasma apo AIV was derived from the intestine, with the rest being synthesized by the liver (Wu & Windmueller, 1979). The physiologic function of apo AIV is not known (Swaney et al., 1977). It is one of the principal plasma apolipoproteins in rats. Although this apoprotein is found in VLDL, HDL, and chylomicrons, at least 50% appears to be unassociated with any of the plasma lipoprotein density classes (Fidge, 1980). Nonetheless, using charge shift electrophoresis, Utermann and co-workers have shown that apo AIV binds detergents through hydrophobic interactions in the same fashion that other amphipathic apolipoproteins do (Beisiegel & Utermann, 1979; Helenius & Simons, 1977).

There have been divergent observations on the effects of triacylglycerol feeding on rat intestinal apo AIV biosynthesis. Windmueller and colleagues were unable to show any effect on intestinal apo AIV synthesis when rats were deprived of exogenous or endogenous (biliary) lipid for 16 h to 12 days (Windmueller & Wu, 1981). On the other hand, Krause et al. studied mesenteric lymph apolipoprotein concentrations and transport rates after infusion of lipid into the duodenum and found that transport of all classes of apolipoproteins was increased but only apo AIV displayed an increase in mesenteric concentration (Krause et al., 1981).

Apo AIV is initially synthesized in the intestine as a preprotein. We have recently shown that the early proteolytic processing of apo AIV is different from the other major intestinal apolipoprotein, AI (Gordon et al., 1982a,b). The primary translation product of rat intestinal apo AI mRNA is a preproprotein that contains a highly usual prosegment. This segment does not terminate with a pair of basic amino acids as most propeptides do but rather has a C-terminal Gln-Gln sequence. On the other hand, the initial product of apo AIV mRNA translation contains a 20 amino acid N-terminal extension which behaves in a cotranslational assay system as a prepeptide. No prosegment could be defined. There is a remarkable degree of amino acid sequence homology between the prepeptides of apo AI and AIV, although the N termini of the mature proteins are quite dissimilar.

In order to define apo AIV protein structure and to study regulation of apo AIV gene expression, we needed a pure complementary DNA. We have successfully cloned such a

† From the Departments of Biological Chemistry (J.I.G., A.W.S., and D.P.S.), Medicine (J.I.G. and D.H.A.), and Pediatrics (A.W.S.), Washington University School of Medicine, St. Louis, Missouri 63110. *Received April 19, 1982.* This research was supported by Grants AM 20407 and AM 30292 from the National Institutes of Health. J.I.G. is a fellow of the John A. and George L. Hartford Foundation. A.W.S. is an Established Investigator of the American Heart Association.

[1] Abbreviations: apo, apolipoprotein; HDL, high-density lipoprotein; $NaDodSO_4$, sodium dodecyl sulfate; SSC, NaCl (0.15 M)–sodium citrate (15 mM, pH 7.0); cDNA, complementary deoxyribonucleic acid; Pipes, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid); Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; EDTA, ethylenediaminetetraacetic acid; AMV, avian myeloblastosis virus; Tc^r, tetracycline resistant; Ap^s, ampicillin sensitive.

0006-2960/82/0421-5424$01.25/0