Proc. Natl. Acad. Sci. USA
Vol. 77, No. 8, pp. 4909–4913, August 1980
Immunology

Complete amino acid sequence of a mouse immunoglobulin α chain (MOPC 511)

(protein structure/phosphocholine binding/immunoglobulin evolution/mutation in IgA 47A)

ELIZABETH A. ROBINSON AND ETTORE APPELLA

Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Communicated by Elvin A. Kabat, April 22, 1980

ABSTRACT **The complete amino acid sequence of the 432-residue heavy (α) chain of mouse myeloma MOPC 511 has been determined. The variable region of the α chain of IgA 511, a phosphocholine-binding protein, is highly homologous to that of the other phosphocholine-binding immunoglobulins. Comparison of the 511 α chain constant region with that of other mouse and human heavy chains shows that sequence divergences and deletions are more extensive within domain disulfide bridges than in other parts of a domain. The C_H3 domain disulfide bridge of the 511 α chain, for example, consists of only 28 amino acid residues compared to 60 residues for other chains and domains. Sequence divergences are also apparent at the C_H2/C_H3 domain boundary, an area where a number of frameshift mutations have occurred. One mutant, mouse IgA 47A, lacks the entire C_H3 domain. Comparison of the 511 α chain with the 47A α chain reveals two nonconservative amino acid changes at the COOH terminus of the 47A chain, Ser-Gln for Val-Thr in the 511 chain. These changes and the deletion of the C_H3 domain can be explained by a single genetic event—namely, a frameshift mutation followed by premature chain termination. The remainder of the 47A constant region, including the hinge region, is identical to the 511 α chain, except for two conservative changes in the C_H1 domain: serine-126 and threonine-197 in the 511 α chain are both replaced by alanine in the 47A chain.**

The predominant immunoglobulins produced by plasmacytomas in BALB/c mice are the IgA class (1), defined by its α heavy chain. The three-dimensional structure of a mouse IgA Fab fragment has been determined (2) but only partial amino acid sequence data for mouse IgA molecules are available. The complete amino acid sequences of the heavy chains of two other classes of mouse immunoglobulins—IgG (γ_1, γ_{2a}, and γ_{2b} chains) (3, 4) and IgM (μ chain) (5)—have been determined; however, sequence analyses of mouse α chains have focused mainly on their variable regions (6). Recently, the complete amino acid sequence of the α chain of the mouse variant IgA 47A was reported (7). Comparison with the sequences of human α chains showed that the entire C_H3 domain of the 47A chain was deleted. However, in the absence of complete sequence data for a normal mouse α chain, the mechanism by which the deletion in the 47A α chain was generated was not apparent.

Mouse myeloma MOPC 511 secretes a normal IgA molecule which is a phosphocholine-binding protein. The amino acid sequence of its light (κ) chain has been reported (8). We now report the amino acid sequence of the 511 α heavy chain. Based on the sequence of the 511 α chain, a mechanism for the deletion in the variant 47A α chain is proposed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

IgA 511 was purified from ascites fluid by affinity chromatography on phosphocholine-Sepharose columns (9). Purification of the partially reduced and carboxyamidomethylated chain, cleavage of the chain with cyanogen bromide, isolation of the cyanogen bromide fragments (CBs) before and after complete reduction and carboxymethylation, enzymatic digestions, and carbohydrate analyses were all performed exactly as described for the mouse IgA 47A heavy chain (7).

Succinylation was performed according to Klotz (10). Amidination with ethyl acetimidate (Aldrich) and deamidination with methylamine (Fisher) was performed according to DuBois *et al.* (11).

Automated spinning-cup Edman degradations were performed with an updated Beckman 890B sequencer. The 0.1 M Quadrol program 060275 was used with double coupling at step 1. In most cases, Polybrene (2 mg) was applied to the cup before application of the sample (12). Phenylthiohydantoin amino acids were identified by high-performance liquid chromatography (13) and by hydrolysis with HI (14).

RESULTS

Amino Acid Sequence. Cleavage at the seven methionine residues of the MOPC 511 α heavy chain with cyanogen bromide yielded eight fragments which accounted for the entire chain (Table 1). The NH_2-terminal sequence of the chain and of each fragment was determined by automated Edman degradation. CB2, -3, -5, and -7 were digested with trypsin, *Staphyococcus aureus* protease, thermolysin, or chymotrypsin, or a combination, and selected peptides were subjected to automated degradation to complete the sequences of these fragments. Tryptic digestion of succinylated, maleylated, or amidinated CB5, the 204-residue fragment, yielded five peptides which accounted for the entire fragment (Table 2). NH_2-terminal analysis of the 58-residue succinylated peptide TS3/CB5 (residues 311–368) established the sequence of the first 26 residues. Digestions of TS3/CB5 with *S. aureus* protease or thermolysin did not provide any additional sequence data. Attempts to demaleylate TM3/CB5 for subsequent tryptic digestion were not successful, even in the presence of 6 M urea. However, the amidinated peptide Am3/CB5 was successfully deamidinated. The peptide was then digested with trypsin and the COOH-terminal tryptic peptide (residues 336–368) was subjected to automated degradation to complete the sequence (Fig. 1).

Abbreviation: CB, cyanogen bromide fragment.

Table 1. Amino acid composition of the cyanogen bromide fragments*

Amino acid	CB1	CB2	CB3	CB4	CB5	CB6	CB7	CB8	Chain
Cys(CM)	0.5 (1)		2.1 (3)		8.2 (11)			0.4 (1)	13.1 (16)
Asp	1.2 (1)	4.0 (4)	8.9 (8)	2.9 (3)	16.1 (15)		3.0 (3)	1.2 (1)	34.7 (35)
Thr	1.9 (2)	3.1 (3)	7.4 (7)	2.5 (4)	15.6 (16)		2.7 (3)		35.7 (35)
Ser	3.8 (5)	6.0 (7)	6.6 (7)	3.2 (2)	19.8 (22)		1.9 (2)	1.0 (1)	50.5 (46)
Hse	0.6 (1)	1.1 (1)	0.8 (1)	0.9 (1)	1.1 (1)	0.7 (1)	0.6 (1)		6.7 (7)
Glu	3.0 (3)	5.4 (5)	2.7 (2)		24.8 (25)	1.2 (1)	1.9 (2)	1.1 (1)	44.8 (39)
Pro	1.1 (1)	1.0 (1)	6.0 (6)	2.1 (2)	17.9 (20)	1.2 (1)	0.7 (1)		34.7 (32)
Gly	5.5 (6)	3.3 (3)	6.5 (6)	5.0 (4)	16.5 (13)	1.1 (1)	1.0 (1)	2.1 (2)	37.7 (36)
Ala	1.2 (1)	3.6 (4)	7.4 (7)	2.4 (2)	11.6 (12)	1.0 (1)			27.4 (27)
Val	2.9 (3)	2.8 (3)	4.2 (4)	1.9 (2)	16.2 (18)	1.0 (1)	2.4 (3)		34.5 (34)
Ile		2.4 (3)	4.7 (5)	1.0 (1)	4.4 (4)		1.9 (2)	1.0 (1)	14.6 (16)
Leu	3.6 (4)	3.1 (3)	5.2 (5)	1.2 (1)	21.3 (22)	0.9 (1)	0.8 (1)		37.8 (37)
Tyr	1.0 (1)	2.6 (3)	6.4 (7)	0.8 (1)	3.6 (2)			0.8 (1)	16.5 (15)
Phe	2.3 (3)	1.2 (1)	2.0 (2)	1.0 (1)	4.2 (4)		1.0 (1)		13.1 (12)
His			0.7 (1)		3.4 (3)	0.8 (1)			6.0 (5)
Lys	1.0 (1)	2.6 (3)	0.6 (0)	2.1 (2)	8.0 (8)		1.8 (2)		16.3 (16)
Arg	1.0 (1)	4.1 (5)	2.4 (3)	1.0 (1)	4.9 (5)		1.0 (1)		15.3 (16)
Trp		ND (2)	ND (2)	ND (1)	ND (3)				ND (8)
GlcN				+			+		+
Total	34	51	76	28	204	8	23	8	432
Yield	28%	18%	32%	24%	51%	10%	14%	10%	

* Integral values are shown in parentheses; the total number of residues was determined from sequence analysis. ND, not determined.

The NH_2-terminal sequence of the chain was used to align CB1 and CB2. Two products of incomplete cleavage at methionine-161 (CB3-4) and -189 (CB4-5) permitted alignment of CB3, -4, and -5. COOH-terminal analysis of the intact chain, which yielded 0.7 residue of tyrosine, permitted the placement of CB8. Another product of incomplete cleavage of methionine-424 (CB7-8) ordered CB7 with respect to CB8. CB6 was placed between CB5 and -7 by homology with the sequence of the human Bur α_1 chain (15).

Oligosaccharide Assignments. The carbohydrate composition of the 511 α chain, mol/mol of heavy chain, was: sialic acid, 0.49; mannose, 7.68; fucose, 1.23; galactose, 3.85; *N*-acetylglucosamine, 6.57. This composition indicated the presence of two *N*-glycosidically linked oligosaccharide units in the chain. These were positioned at asparagine-162 and asparagine-419 by the characterization of small tryptic peptides and by blank cycles in automated sequencer runs. The unit linked to asparagine-162 was present in the mouse variant 47A α chain but not in the human Bur α_1 chain and the mouse 104E μ chain. The second unit, at asparagine-419, also was present in human α_1 and α_2 chains, the mouse 104E μ chain, and the mouse 173 γ_{2a} chain. Unlike the mouse 47A α chain and human α_1 chains, the 511 α chain contained no *O*-glycosidically linked units in the hinge region (residues 225–247). This conclusion is based

Table 2. Amino acid composition of tryptic peptides of succinylated CB5*

Amino acid	TS1	TS2	TS3	TS4	TS5	CB5
Cys(CM)	4.5 (5)	2.0 (3)	1.6 (2)		0.6 (1)	8.2 (11)
Asp	6.7 (7)	2.6 (3)	2.7 (3)	1.0 (1)	1.0 (1)	16.1 (15)
Thr	5.4 (5)	2.0 (2)	5.8 (8)		0.8 (1)	15.6 (16)
Ser	8.3 (10)	3.7 (5)	4.0 (5)		1.8 (2)	19.8 (22)
Hse					0.8 (1)	1.1 (1)
Glu	10.2 (11)	4.6 (5)	4.7 (5)	0.9 (1)	2.6 (3)	24.8 (25)
Pro	10.7 (10)	2.7 (3)	5.1 (6)	1.1 (1)		17.9 (20)
Gly	4.5 (5)	4.0 (4)	2.7 (3)		1.4 (1)	16.5 (13)
Ala	3.3 (3)	3.5 (4)	3.0 (3)	1.2 (1)	1.0 (1)	11.6 (12)
Val	5.1 (5)	4.7 (5)	4.3 (5)	1.9 (2)	0.8 (1)	16.2 (18)
Ile	2.7 (3)	0.2	0.9 (1)			4.4 (4)
Leu	9.9 (11)	1.1 (1)	6.8 (9)	1.2 (1)		21.3 (22)
Tyr	0.5	0.8 (1)	0.3		0.7 (1)	3.6 (2)
Phe	0.3	0.9 (1)	1.4 (2)	1.1 (1)		4.2 (4)
His	1.1 (1)		1.5 (2)			3.4 (3)
Lys	1.2 (1)	2.2 (3)	1.3 (2)	1.0 (1)	0.8 (1)	8.0 (8)
Arg	1.9 (2)	1.0 (1)	0.7 (1)	1.0 (1)		4.9 (5)
Trp	ND (0)	ND (1)	ND (1)	0.0 (0)	ND (1)	ND (3)
Total	79	42	58	10	15	204
Yield	47%	65%	88%	69%	68%	

* Integral values are given in parentheses; the total number of residues was determined from sequence analysis. ND, not determined.

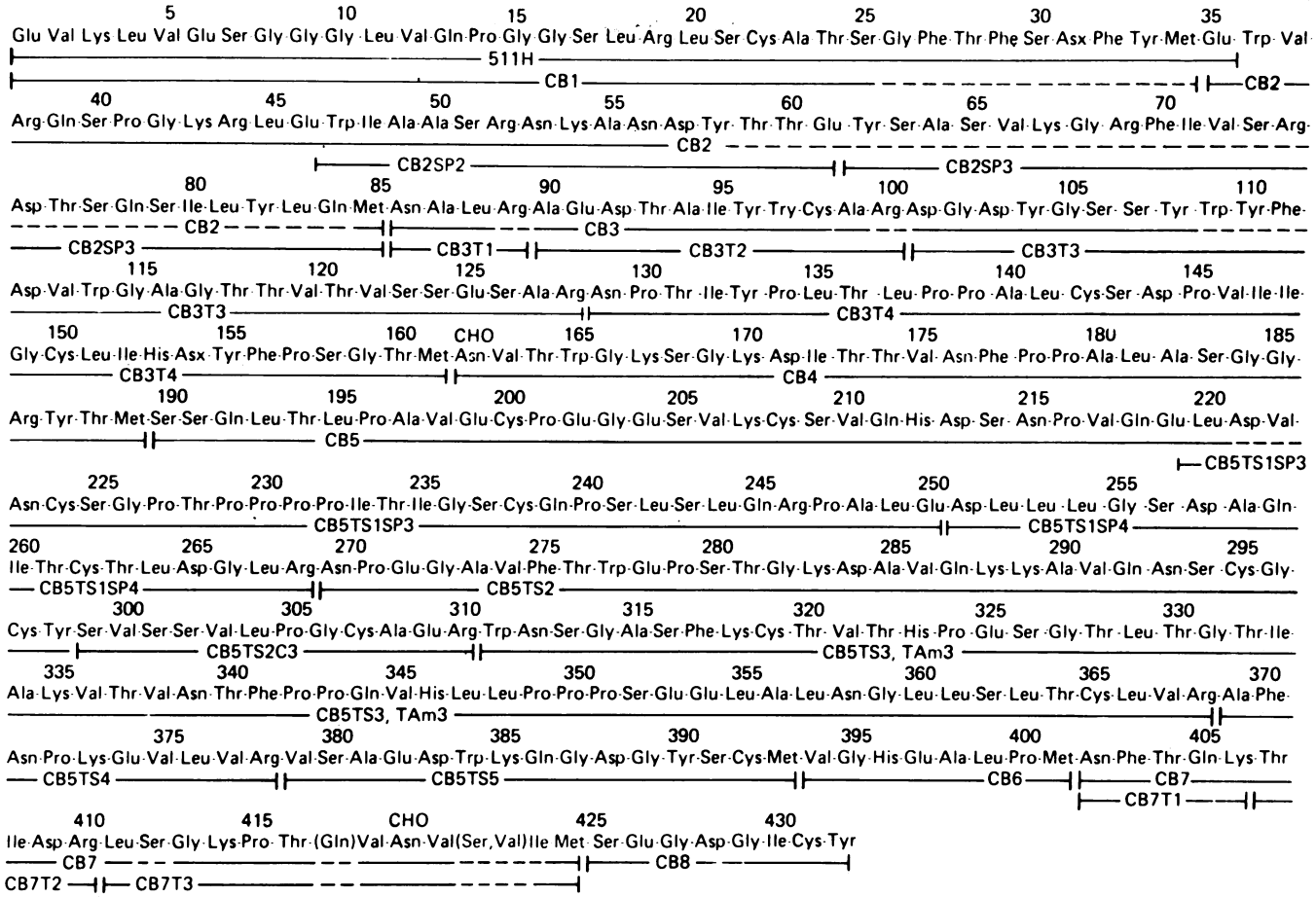

FIG. 1. The primary structure of the heavy (α) chain of mouse IgA 511 deduced from characterizations of the chain (H), the cyanogen bromide fragments (CB), and their tryptic (T), chymotryptic (C), thermolytic (Th), and *S. aureus* protease (SP) peptides. Tryptic peptides from chemically modified polypeptides are assigned a second letter: S, succinylated (TS); Am, amidinated (TAm).

upon carbohydrate analyses of both the chain (see above) and the peptide SP3/TS1/CB5 in which no *N*-acetylgalactosamine was found.

Mutation in the Mouse Variant IgA 47A. The complete amino acid sequence of the α chain of the mouse variant IgA 47A has shown that it lacks the entire C_H3 domain (7). Recent characterization of the mRNA coding for the 47A α chain has shown that the mutation is not due to posttranslational modification and that the mRNA is the same size as that coding for a normal mouse α chain (16). The sequence of the 511 α chain at the C_H2/C_H3 domain boundary now makes it possible to propose a mechanism for the deletion in the 47A α chain. Fig. 2 shows that the sequence Val-336-Thr-337 in the 511 α chain is replaced by the COOH-terminal sequence Ser-Gln in the 47A chain. In the predicted mRNA sequence of the 511 α chain, the deletion of a single nucleotide of the codon for valine-336 would alter the reading frame such that the next two obligatory nucleotide triplets would become UCA and CAG, which code for serine and glutamine, respectively. The codon following that for glutamine would then be UAA, a termination codon.

The remainder of the 47A α chain constant region, including the hinge region, is identical to that of the 511 α chain except for two amino acid changes in the C_H1 domain. Serine-126 and threonine-197 in the 511 α chain are both replaced by alanine in the 47A chain (Fig. 2). These changes are conservative in nature—i.e., they are attributable to single base changes.

DISCUSSION

The primary structure of the 511 α chain (Fig. 1) consists of the following parts: a variable region (residues 1–107); a J-piece, recently described for S107, another phosphocholine-binding protein (17) (residues 108–124); the C_H1 domain (residues 125–224); the hinge region (residues 225–247); the C_H2 domain (residues 248–337); and the C_H3 domain (residues 338–432). The sequence of the variable region is highly homologous to that of other phosphocholine-binding proteins (18). The few amino acid exchanges that have been found are distributed in both the complementarity regions and the framework regions. All but one require a single base change. Comparison of the sequence

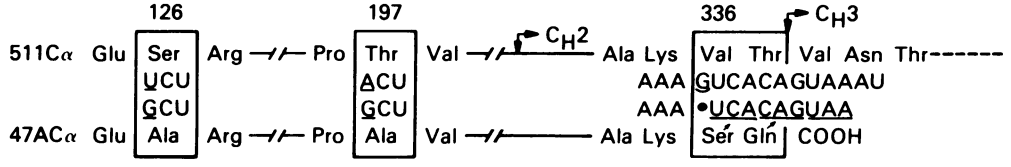

FIG. 2. Comparison of the sequence of the constant region of the mouse 511 α heavy chain with that of the mouse variant 47A α heavy chain. The numbering is that of the 511 α chain. The corresponding residues of the 47A α chain are 118–330. The sequences are identical except where indicated by boxed areas.

of the entire constant region of the 511 α chain with that of the human Bur α_1 chain, the mouse 104E μ chain, and the mouse 173 γ_{2a} chain shows 62%, 33%, and 25% homology, respectively, with the COOH-terminal domains being the most highly conserved (homologies of 78%, 49%, and 38%, respectively). This conservation of the COOH-terminal domains has also been noted for mouse, dog, and human μ chains (5).

Comparative analyses of the constant regions of these chains also show that sequence divergences and deletions are more extensive within domain disulfide bridges than in any other part of a domain. The C_H3 domain disulfide bridge of the 511 α chain (cysteine-365 and -392 by homology), for example, consists of only 28 amino acid residues compared to an average of 60 residues for other classes and domains (Fig. 3). In addition, in the "gap" between residues 378 and 379 of the 511 α chain, the sequences of the human α_1 chain and the mouse μ and γ_{2a} chains are more divergent than in any other part of the domain. It should also be noted that the γ_{2a} chain is "missing" five residues at the end of the 511 chain gap.

Another area of sequence divergence is at the boundary of the C_H2 and C_H3 domains (Fig. 3, top line). A number of mutations have also been described in this area. In the mouse MPC 11 IgG2b frameshift mutant M311, lysine-335 (511 α chain numbering) is replaced by COOH-terminal asparagine (19). The spontaneous mutant IF3 of MOPC 21 IgG1 is also a frameshift mutant involving the deletion of two nucleotides at glutamine-343 (20). Six spontaneous mouse IgA variants which are 3.9S, two-chain molecules have been described (21, 22). The α chains of three of them, 47A, 6C, and 844, terminate at residue 337. The human IgA1 variant Vo also appears to terminate at this point (23).

The sequence of the 511 α chain has shown that the mutation generating the deleted 47A α chain is probably a frameshift mutation followed by premature chain termination. The same mechanism appears to account for the mutations in the MPC 11 IgG2b variant M311 and the MOPC 21 IgG1 variant IF3. Nucleotide sequences of cloned genes of MPC 11 γ_{2b} (24) and MOPC 21 γ_1 (25) chains show that the boundary of the C_H2 and C_H3 domains of both chains is at the point indicated in Fig. 3. The proposed frameshift mutation and premature chain termination in the M311 chain clearly occurs within the C_H2 domain; the mutation in IF3 is within the C_H3 domain. If the C_H2/C_H3 domain boundary of the mouse α chain is at the same point as in the γ chains, the proposed frameshift mutation and chain termination requires the reading of nucleotides encoding both the C_H2 and C_H3 domains. In view of this, another mechanism for the mutation in the 47A α chain should be considered. If all or part of the intervening sequence separating the C_H2 and C_H3 domains were still present in the 47A mRNA, due to a mutation at or near a splicing site, the presence of such a sequence could also allow premature chain termination to occur. Determination of the sequence of the appropriate DNA fragments will test these models and also may provide an explanation for the seemingly high mutation rate at the C_H2/C_H3 domain boundary.

Two amino acid changes were found in the C_H1 domain of the 47A α chain in addition to the deletion. For the MOPC 21 IgG1 variants, no other sequence differences were detected in addition to the mutations nor have any been reported for other variants. The additional changes found for the 511 and 47A α chains may mean that IgA 511 and IgA 47A represent subclasses of mouse IgA that arose recently. The human IgA subclasses have 23 amino acid exchanges in the constant regions in addition to significant differences in the hinge regions (26). These more extensive differences may indicate that the human subclasses arose earlier in evolution. Another possibility, other than subclasses, is that the amino acid changes in the C_H1 domains of 511 and 47A α chains could be due to somatic mutations that

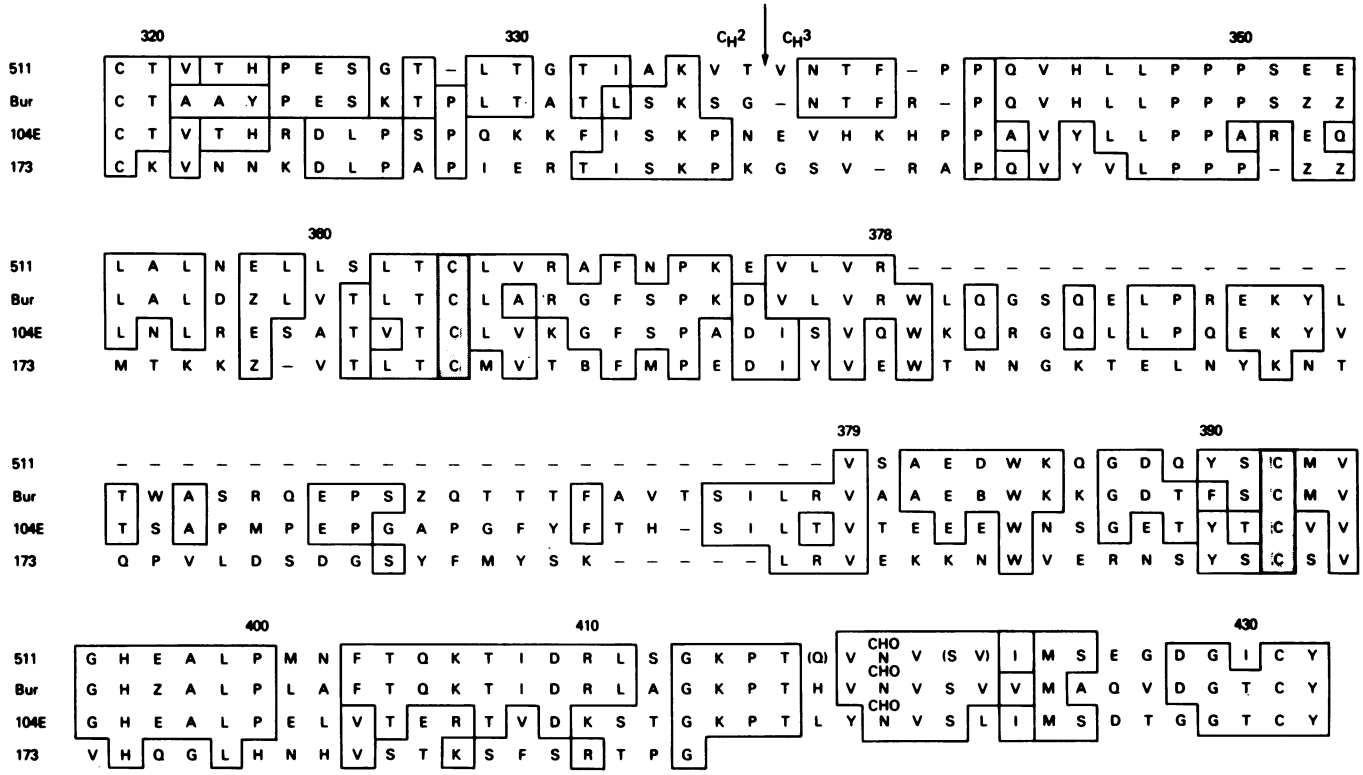

FIG. 3. Comparison of the COOH-terminal sequences of the mouse 511 α chain, the human Bur α_1 chain (15), the mouse 104E μ chain (5), and the mouse 173 γ_{2a} chain (3). The shaded cysteine (C) residues participate in the domain disulfide bridge. The C_H2/C_H3 domain boundary is the C_H3/C_H4 boundary of the 104E μ chain. The numbering is that of the 511 α chain.

arose during the generation or propagation of the tumors. Further studies on the other mouse IgA variants will be required to define such amino acid substitutions.

We thank Dr. David F. Smith for the carbohydrate analyses and Mr. Sam C. Morris for assistance in peptide isolations.

1. Potter, M. & Cancro, M. (1978) in *Cell Differentiation and Neoplasia*, ed. Saunders, G. F. (Raven, New York), pp. 145–161.
2. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4298–4302.
3. Milstein, C., Adetugbo, K., Cowan, N. J. & Secher, D. S. (1974) in *Progress in Immunology II*, eds. Brent, L. & Holborow, J. (North-Holland, Amsterdam), Vol. 1, pp. 157–168.
4. Fougereau, M., Bourgois, A., de-Preval, C., Rocca-Serra, J. & Schiff, C. (1976) *Ann. Immunol. (Paris)* **127C**, 607–631.
5. Kehry, M., Sibley, C., Fuhrman, J., Schilling, J. & Hood, L. E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2932–2936.
6. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) *Sequences of Immunoglobulin Chains* (Medical Computer Systems, Bolt, Beranek & Newman, Cambridge, MA).
7. Robinson, E. A. & Appella, E. (1979) *J. Biol. Chem.* **254**, 11418–11430.
8. Appella, E. (1980) *Mol. Immunol.*, in press.
9. Chesebro, B. & Metzger, H. (1972) *Biochemistry* **11**, 766–771.
10. Klotz, I. M. (1967) *Methods Enzymol.* **11**, 576–580.
11. DuBois, G. C., Robinson, E. A., Inman, J. K., Perham, R. N. & Appella, E. *J. Biol. Chem.*, in press.
12. Tarr, G. E., Beecher, J. F., Bell, M. & McKean, D. J. (1978) *Anal. Biochem.* **84**, 622–627.
13. Zimmerman, C. L., Appella, E. & Pisano, J. J. (1977) *Anal. Biochem.* **77**, 569–573.
14. Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. & Ballantyne, D. C. (1971) *Biochemistry* **10**, 4912–4921.
15. Putnam, F. W., Liu, Y.-S. V. & Low, T. L. K. (1979) *J. Biol. Chem.* **254**, 2865–2874.
16. Robinson, E. A., Ferrini, U., Seidman, J. G. & Appella, E. (1980) *J. Biol. Chem.* **255**, 4988–4991.
17. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980) *Cell* **19**, 981–992.
18. Padlan, E. A., Davies, D. R., Rudikoff, S. & Potter, M. (1976) *Immunochemistry* **13**, 945–949.
19. Kenter, A. L. & Birshtein, B. K. (1979) *Science* **206**, 1307–1308.
20. Adetugbo, K. & Milstein, C. (1978) *J. Mol. Biol.* **121**, 239–254.
21. Potter, M. & Kuff, E. L. (1964) *J. Mol. Biol.* **9**, 537–544.
22. Mushinski, J. F. & Dungan, M. M. (1976) *J. Immunol.* **117**, 1668–1675.
23. Despont, J. P. M., Abel, A., Grey, H. M. & Penn, G. M. (1974) *J. Immunol.* **112**, 1517–1525.
24. Tucker, P. W., Marcu, K. B., Newell, N., Richards, J. & Blattner, F. R. (1979) *Science* **206**, 1303–1306.
25. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T. K., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
26. Torano, A. & Putnam, F. W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 966–969.