Primary Structure of the Variable Regions of Two Canine Immunoglobulin Heavy Chains†

Richard L. Wasserman and J. Donald Capra*

ABSTRACT: The complete amino acid sequences of the variable regions of two canine immunoglobulin heavy chains have been determined by automated Edman degradation and found to be strongly homologous to the human V_HIII subgroup. The canine sequences were identical with each other at 76 of 113 residue positions. Twenty-three of the 37 differences are located within the four hypervariable regions previously defined by the sequences of several human V_HIII proteins. Forty-five of 77 framework residue positions are invariant in the seven human and two canine V_HIII proteins which have been completely sequenced. The canine proteins are 78% homologous to the framework of the human prototype. Phylogenetically associated residues before the first hypervariable region were confirmed and several potential phylogenetically associated residues were identified between the first and third hypervariable regions. This study represents the first complete amino acid sequences of V_H regions of spontaneously occurring, nonhuman homogeneous immunoglobulins. The data demonstrate a high degree of preservation of V_HIII structure in another species.

Immunoglobulin light and heavy chains are characterized by significant sequence variability in their amino-terminal or variable region domains (Hilschmann and Craig, 1965; Capra and Kehoe, 1974a). Sequence studies of the variable regions of human light and heavy chains have led to the definition of subgroups of variability of κ, λ, and heavy chains (Milstein, 1967; Niall and Edman, 1967; Köhler et al., 1970). Within subgroups of λ (Novotny et al., 1972) and heavy chains (Kehoe and Capra, 1972; Capra et al., 1973), particular amino acid substitutions at certain residue positions have been shown to be characteristic of a given phylogenetic grouping. These phylogenetically associated residues are most compatible with a pauci-gene model of antibody diversity.

Earlier sequence studies which have identified phylogenetically associated residues have primarily examined only the amino-terminal 30 residues (Capra et al., 1973). The primary structures of the heavy chain variable regions of two canine myeloma proteins were determined in order to examine whether the evolution of the amino-terminal 30 residues of the heavy chain variable regions is representative of the entire V-region framework. Complete sequences of the V_H regions of several human myeloma proteins of the V_HIII subgroup have been determined (Ponstingl et al., 1970; Watanabe et al., 1973; Capra and Kehoe, 1974a,b). Comparison of these sequences with the two canine sequences reported here confirms the high degree of evolutionary conservation seen in the amino-terminal 30 residues. Additionally, several potential phylogenetically associated residues were identified beyond the first hypervariable region.

Materials and Methods

Serum Donors. Sera containing canine paraproteins Gom and Moo were a gift from Dr. A. Hurvitz, Animal Medical Center, New York, N.Y. IgAκ Gom was obtained by plasmaphoresis of an Airdale terrier with multiple myeloma. IgMλ Moo was obtained by plasmaphoresis of a Scottish terrier with lymphosarcoma.

Preparation of Proteins. Both proteins were initially isolated by zone electrophoresis using Pevikon (Pevikon C-870, Mercer Chemical Corp.) in barbital buffer (pH 8.6), ionic strength 0.5 (Kunkel, 1954). The myeloma proteins were eluted from the Pevikon with 0.05 M sodium phosphate buffer, pH 7.85, and were further purified by ion-exchange chromatography on Sephadex DEAE-A50 equilibrated with 0.05 M sodium phosphate buffer, pH 7.85. After washing with starting buffer, the proteins were eluted with a linear gradient of 0.0 to 1.0 M NaCl in starting buffer. The proteins thus obtained migrated as a single band on microzone electrophoresis on cellulose-acetate membranes and formed a single arc on microimmunoelectrophoresis against rabbit anti-normal dog serum.

Preparation of Heavy Chains. The purified myeloma proteins were completely reduced in 1.0 M tris(hydroxymethyl)-aminomethane (Tris) (Sigma Chemical Corp.) (pH 8.2)-6.0 M Gdn·HCl-0.005 M EDTA using 0.1 M 2-mercaptoethanol or 2.6 mM dithiothreitol for 1 h. The protein concentrations were: Gom, 10-13 mg/mL; Moo, 20 mg/mL. The reduced chains were carboxymethylated with iodo[^{3}H]acetamide or iodo[^{14}C]acetic acid (0.3 μCi/mg of protein) for 5 min followed by unlabeled iodoacetamide or iodoacetic acid at 0.15 M for 20 min at room temperature. The reduced and carboxymethylated chains were separated on 2.5 × 120 cm columns packed with Sephadex G-100 equilibrated with 5.0 M Gdn·HCl.

Isolation of Cyanogen Bromide Peptides of α Chain Gom. Peptides derived from Gom and Moo were named according to the following scheme: peptides derived from cyanogen bromide digests were named CB and numbered sequentially from the amino terminus. Peptides derived from further digestions of cyanogen bromide peptides carry the name of the digested peptide followed by T for tryptic, CT for citraconylated tryptic, and CHY for chymotryptic digests and are numbered from the amino terminus of the digested peptide.

† From the Department of Microbiology, The University of Texas, Southwestern Medical School, Dallas, Texas 75235. *Received July 15, 1976.* This work was supported by grants from The National Science Foundation (BMS-74-0441) and the National Institutes of Health (AI-12127, AI-12796). Submitted to City University of New York, Mt. Sinai School of Medicine in partial fulfillment of the requirements for the degree of Doctor of Philosophy (R.L.W.).

[1] Abbreviations used: DEAE, diethylaminoethyl; Gdn·HCl, guanidine hydrochloride; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; Pth, phenylthiohydantoin.

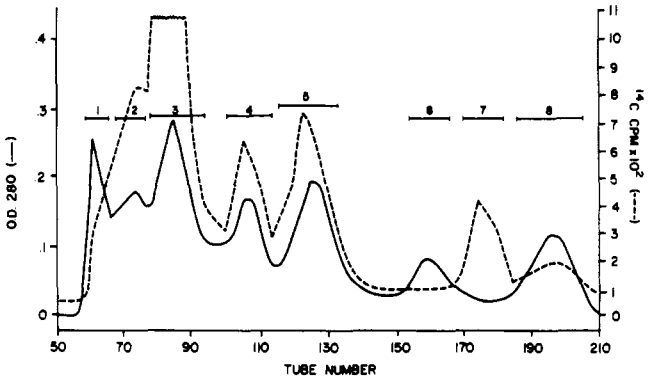

FIGURE 1: Chromatography of the cyanogen bromide peptides of α chain Gom. The digest was applied to a 5.0 × 150 cm column packed with Sephadex G-100 equilibrated with 5.0 M Gdn·HCl. Pools 5, 6, and 7 contained variable region peptides.

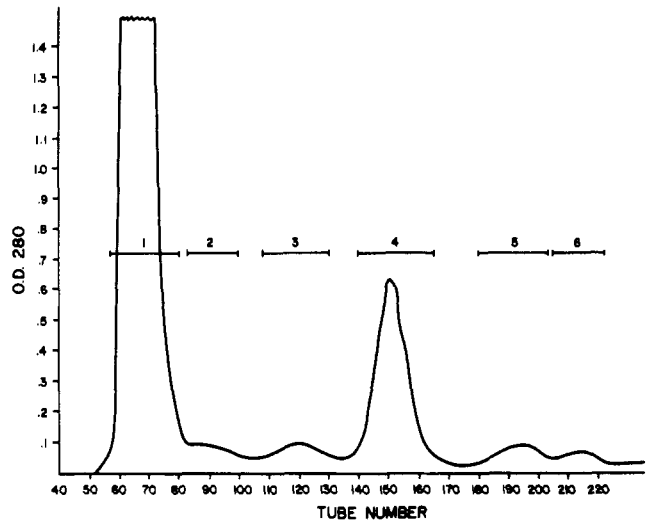

FIGURE 2: Chromatography of the cyanogen bromide peptides of IgAκ Gom. The digest was applied to a 5.0 × 150 cm column packed with Sephadex G-100 equilibrated with 5.0 M Gdn·HCl. Pool 4 contained CB2 and two constant region peptides.

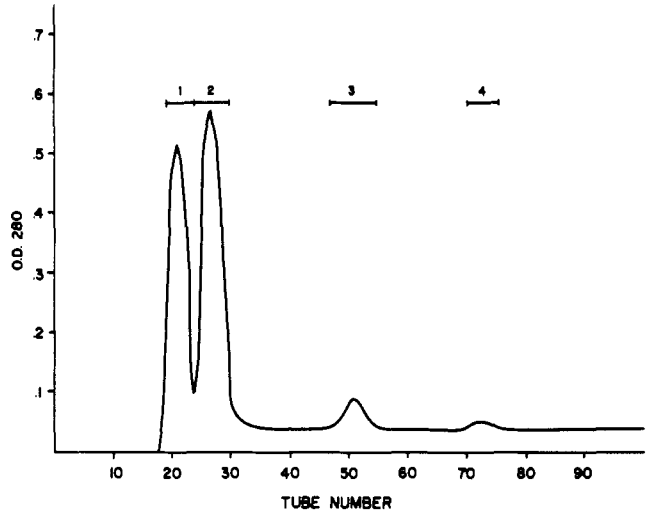

FIGURE 3: Chromatography of the cyanogen bromide peptides of IgMλ Moo. The digest was applied to a 2.5 × 120 cm column packed with Sephadex G-100 equilibrated with 5.0 M Gdn·HCl. Pools 2 and 3 contained variable region peptides.

Completely reduced and carboxymethylated heavy chains were dissolved in 70% formic acid and reacted with cyanogen bromide overnight at room temperature (protein:CNBr ratio 1:5). The digest was then mixed with water (1:10) and lyophilized. The resulting peptides were separated by molecular exclusion chromatography on a 5 × 150 cm column of Sephadex G-100 equilibrated with 5 M Gdn·HCl. The elution profile is shown in Figure 1. Each peak was pooled as indicated.

Pools 5, 6, and 7 were found to contain peptides homologous to human V_HIII peptides. Pools 5 (CB3) and 7 (CB1) required no further purification. Pool 6 contained three peptides, one of which was homologous to the peptide beginning at position 35 derived from human proteins; the other two peptides were not homologous to human variable region peptides. The variable region peptide (CB2) was isolated by precipitation in 0.01 M ammonium acetate, pH 5.0, as described by Rudikoff and Potter (1974). Since pool 6 of the heavy chain cyanogen bromide digest was not radioactive, the peptides eluting at that position contained no carboxymethylcysteine and were not linked to other parts of the α chain by disulfide bonds. Therefore, to obtain additional amounts of CB2, the intact Gom IgAκ dimer was digested with cyanogen bromide in formic acid as described above and the peptides were separated by molecular exclusion chromatography on Sephadex G-100 in 5 M Gdn·HCl as shown in Figure 2. Pool 4 of the digestion of the dimer was found to contain the same three peptides as pool 6 of the heavy chain digest and the variable region peptide, CB2, was isolated as described above.

Isolation of Cyanogen Bromide Peptides of μ Chain Moo. Intact pentamer was digested with cyanogen bromide as described above. The resulting peptides were initially separated on a 2.5 × 120 cm column of Sephadex G-100 equilibrated with 5.0 M Gdn·HCl. The elution profile is shown in Figure 3. Pool 2 of the digest was resolved into two components by Sephadex DEAE-A50 ion-exchange chromatography in 6 M urea–0.05 M sodium phosphate, pH 7.85. After washing and collecting the fall-through, a linear gradient of 0.0 to 1.0 M NaCl in the running buffer was applied to the column and a single peak eluted. Sequence analysis of the two separated groups of peptides showed them to be different: the material not retained by the ion exchanger contained two peptides homologous to human V_HIII peptides (CB1 and CB3); the peak eluted with sodium chloride contained only peptides homologous to constant region peptides.

The variable region peptides derived from pool 2 of the cyanogen bromide digest of the pentamer were reduced and carboxymethylated as described above. The resulting peptides were separated on a 2.5 × 120 cm column of Sephadex G-75 equilibrated with 5 M Gdn·HCl. The elution profile is shown in Figure 4. CB3 was found in the fourth pool which also contained two constant region peptides and was isolated from these constant region peptides on a 2.5 × 240 cm column of G-50 fine equilibrated with 0.5 M ammonia (Figure 5).

Isolation of Peptides Derived from Enzymatic Digests of Cyanogen Bromide Fragments. Tryptic digestion was performed in 1% NH_4HCO_3 buffer at 37 °C for 16 h at an enzyme-to-substrate ratio of 1:100. Trypsin which was L-1-tosylamido-2-phenylethyl chloromethyl ketone treated was obtained from Worthington Biochemical Corp. Gom fragment CB2 (1.0–2.0 μmol) was digested with trypsin and the peptides were separated by high-voltage paper electrophoresis at either pH 2.1 or 3.5.

The tryptic peptides of Gom CB3 were initially separated on a column of Sephadex G-50, fine (2.5 × 240 cm), equili-

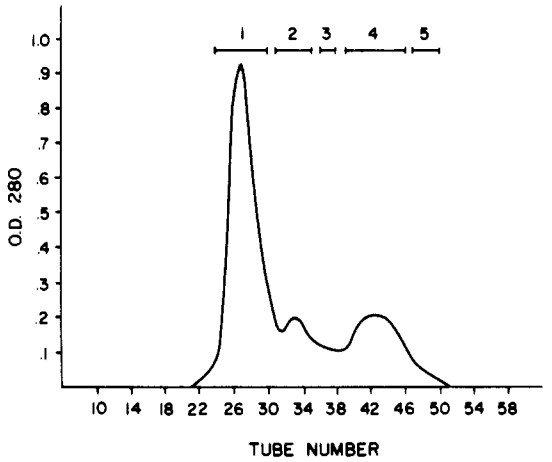

FIGURE 4: Chromatography of completely reduced and carboxymethylated cyanogen bromide fragments of IgMλ Moo. These fragments were derived from pool 2 of the chromatography of the cyanogen bromide digest of the pentamer (Figure 3) and were separated from other constant region peptides by ion-exchange chromatography as described in the text. The reduced and carboxymethylated peptides were applied to a 2.5 × 120 cm column packed with Sephadex G-75 equilibrated with 5.0 M Gdn·HCl. Pool 4 contains CB3 and constant region peptides.

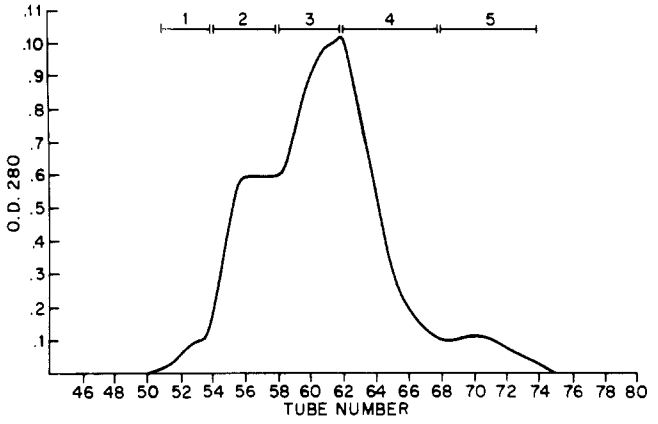

FIGURE 5: Chromatography of cyanogen bromide peptides of IgMλ Moo. Pool 4 of Figure 4 was applied to a 2.5 × 240 cm column packed with Sephadex G-50 fine equilibrated with 0.5 M ammonia. Pool 3 contains peptide CB3 only.

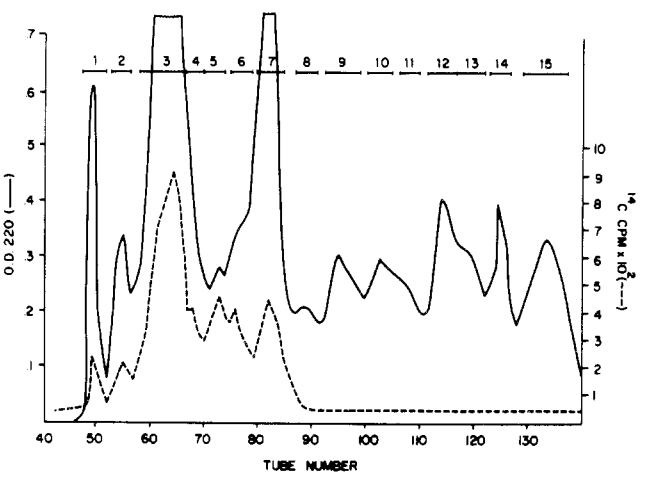

FIGURE 6: Chromatography of the tryptic peptides of fragment CB3 of α chain Gom. The digest was applied to a 2.5 × 240 cm column packed with Sephadex G-50 fine equilibrated with 0.5 M ammonia. Pools 12 and 13 contained peptide CB3T1 and constant region peptides. Pool 7 contained peptide CB3T2.

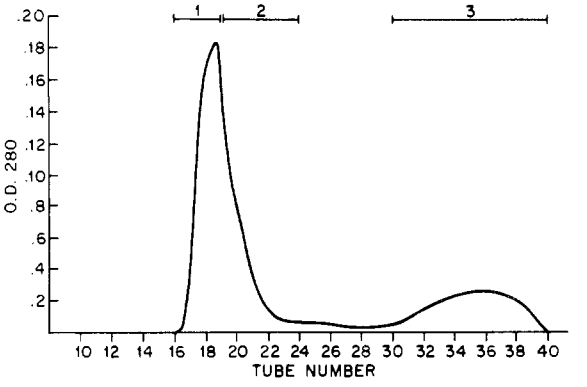

FIGURE 7: Chromatography of the tryptic peptides of fragment CB2 of Moo obtained by digestion of citraconylated CB2. The digest was applied to a 1.5 × 200 cm column packed with Sephadex G-25 fine equilibrated with 0.5 M ammonia. Pool 1 contains CB2CT2. Pool 2 contains CB2CT4. Pool 3 contains CB2CT1 and CB2CT3.

brated with 0.5 M NH_3 (Figure 6). The peptides were pooled as indicated and lyophilized. Peptides in pools 12 and 13 were further separated by high-voltage paper electrophoresis at pH 2.1.

The tryptic peptides of Moo CB2 were separated by high-voltage paper electrophoresis at pH 2.1 and 3.5. Fragment CB2 of Moo (1.0 μmol) which had been reacted with citraconic anhydride (45 μL for 30 min) in 1 mL of 6 M Gdn·HCl–1 M Tris (pH 9.0) and then desalted in 1% NH_4HCO_3 was digested with trypsin as described above. The digest was applied to a column of Sephadex G-25 fine (1.5 × 200 cm) equilibrated with 0.5 M NH_3 and two peaks were resolved and pooled as shown (Figure 7). The peptides in pool 3 were further separated by high-voltage paper electrophoresis at pH 2.1 and 3.5. Peptide CB2CT2 of Moo (0.3–0.8 μmol) was dissolved in 1% NH_4HCO_3 and digested with chymotrypsin (Worthington) at 37 °C for 2 h (enzyme–substrate 1:100 w/w) and the resultant peptides were separated by high-voltage paper electrophoresis at pH 2.1.

High-Voltage Paper Electrophoresis. Electrophoresis on Whatman No. 1 and Whatman No. 3 paper was carried out in pyridine–acetate buffers at pH 2.1 and 3.5 under Varsol (Du Pont) in a Savant apparatus (Crumpton and Wilkinson, 1965). Peptides were eluted from No. 3 paper with 0.5 M NH_3.

Amino Acid Analysis. Peptides were hydrolyzed in 6 M HCl, after evacuation in a nitrogen atmosphere, for 20 h at 110 °C. The hydrolysates were analyzed on a Durrum D-500 amino acid analyzer.

Automated Sequencing. The use of the automated protein–peptide sequencer in this laboratory has been previously described (Capra and Kunkel, 1970; Capra et al., 1973). Some peptides were treated with 4-sulfophenyl isothiocyanate (Pierce Chemical Co.) prior to sequencing to enhance retention in the sequencer reaction cup (Inman et al., 1972). Identification of Pth-amino acids was made by gas-liquid chromatography employing a Beckman GC-65 (Pisano and Bronzert, 1969), a modified (substituting BBOT [Packard] as the fluor) thin-layer chromatography system (Summers et al., 1973), scintillation counting in a Beckman LS-330 liquid scintillation counter (Capra et al., 1972), and amino acid analysis after Pth hydrolysis with hydriodic acid as previously described (Mole et al., 1975).

Results

The Sequence of the Amino-Terminal Cyanogen Bromide

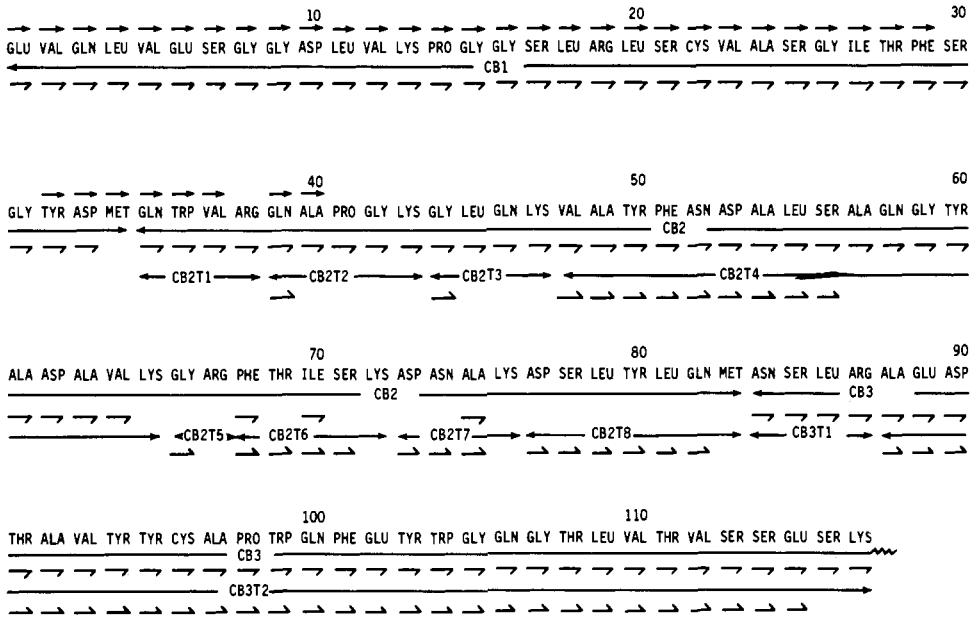

FIGURE 8: The sequence of the variable region of α chain Gom. Identifications made by sequencing the intact heavy chain (→), the cyanogen bromide fragments (⇢), and the tryptic peptides (—) are indicated.

Fragment of Gom α Chain. Three hundred nanomoles of heavy chain was subjected to 41 cycles of sequential Edman degradation. Unequivocal identification of all positions except 30, 31, and 38 was established. This experiment resulted in the identification of most of the amino-terminal CB1 peptide. The two unidentified residues were identified by sequencing the isolated CB1 peptide. Thirty-four cycles of Edman degradation were performed and all positions up to and including position 33 were identified. These two extended sequencer runs completed the sequence of the amino-terminal cyanogen bromide peptide and extended into CB2, thus providing an overlap as shown in Figure 8. The composition of CB1 is shown in Table I.

The Sequence of the Second Cyanogen Bromide Fragment of the Variable Region of α Chain Gom. CB2 isolated as described above was sequenced for 41 Edman degradation steps on the automated sequencer and identifications made of all positions up to and including step 30. In addition, positions 34, 36, and 41 were also identified. One micromole of CB2 was digested with trypsin as described. The tryptic peptides were initially separated by high-voltage electrophoresis on Whatman No. 3 paper at pH 2.1 using 3000 V for 30 min and eluted with 0.5 M NH_3. Peptides which migrated together at pH 2.1 were separated by a second electrophoresis at pH 3.5 using 3000 V for 20 min.

Except for determination of their amino termini, peptides CB2T2 and CB2T3 were not sequenced since their sequence had already been established by the amino-terminal run on the intact CB2 fragment. In order to confirm certain positions, peptide CB2T4 was sequenced for 19 cycles after reaction with 4-sulfophenyl isothiocyanate to enhance the retention of the peptide in the sequencer cup. Peptides CB2T5, CB2T6, CB2T7, and CB2T8 were subjected to 2, 5, 4, and 7 cycles of Edman degradation, respectively. Peptide CB2T7 was also reacted with 4-sulfophenyl isothiocyanate before a second sequencer run. Peptides CB2T1, CB2T2, CB2T3, CB2T4, and CB2T5 were ordered by fitting their sequences to the sequence derived from the intact CB2 peptide. The two-residue gap left between CB2T4 and CB2T6 was filled by CB2T5. CB2T8 was known to be the carboxy-terminal peptide because of the homoserine lactone which was found on amino acid analysis. Since all other peptides were accounted for, CB2T7 was therefore assigned the penultimate position. The alanine obtained in step 41 of the sequencer run on the whole CB2 fragment is also consistent with the placement of CB2T7 in this position. The complete sequence of CB2 is shown in Figure 8. The composition of CB2 and its tryptic peptides are shown in Table I.

The Sequence of the Third Variable Region Cyanogen Bromide Fragment of α Chain Gom. CB3 was subjected to 37 cycles of Edman degradation. Steps 1 through 33 and step 37 were identified. The sequence of this peptide from step 6 through step 14 is identical with the region 91–100 of all human V_HIII proteins and it is therefore reasonable to assign it as the third variable region CNBr peptide without a peptide to overlap residues 84–86.

Approximately 2 μmol of CB3 was digested with trypsin and the resultant peptides were applied to a 2.5 × 240 cm column packed with Sephadex G-50 fine equilibrated with 0.5 M ammonia. The elution profile is shown in Figure 6. Pools 12 and 13 contained peptide CB3T1 together with constant region peptides. Peptide CB3T1 was isolated by high-voltage electrophoresis on Whatman No. 3 paper at pH 2.1 using 3000 V for 30 min. The peptide was eluted with 0.5 M NH_3. Pool 7 contained peptide CB3T2 in high yield and other peptides in very low yield. The material obtained from pool 7 was subjected to 29 cycles of Edman degradation and steps 1 through 26 were identified. The sequence of CB3T2 was identical with that obtained from the intact CB3 peptide.

The sequence of CB3 to the end of CB3T2 is shown in Figure 8. The compositions of CB3T1 and CB3T2 as determined by sequence and amino acid analysis are shown in Table I. This peptide completes the sequence of the variable region and extends past the putative V/C bridge into the beginning of the $C_{\alpha 1}$ domain.

The Sequence of the Amino-Terminal Cyanogen Bromide Fragment of the Variable Region of the μ Chain Moo. Two hundred and eighty nanomoles of heavy chain was subjected to 39 cycles of automated Edman degradation and definitive identifications were made for all positions except 35. The

TABLE I: Amino Acid Composition of Gom Variable Region Peptides.

	CB1		CB2		CB2T1		CB2T2		CB2T3		CB2T4		CB2T5		CB2T6		CB2T7		CB2T8		CB3T1		CB3T2	
Amino acid	AAA[a]	SEQ[b]	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ
CM-Cys	0.9	1																					0.7	1
Asp	2.8	2	6.2	6							2.8	3					2.0	2	0.9	1	0.9	1	2.3	1
Thr	1.6	1	2.5	1											0.8	1			0.8	1			2.5	3
Ser	4.8	5	3.0	3							2.0	1			2.0	1	1.4				0.8	1	3.1	3
Glu	2.5	3	4.7	5		1	1.1	1	1.0	1	1.5	2					0.7		1.0	1			3.8	5
Pro	0.9	1	1.2	1			1.0	1															1.4	1
Gly	5.4	6	3.7	4			1.6	1	1.7	1	1.8	1	1.9	1			2.3						1.8	2
Ala	1.7	1	6.3	7			1.2	1			3.2	5					1.8	1					3.1	3
Val	2.4	4	2.3	3		1					2.0	2											2.7	2
Ile	1.3	1	1.6	1											0.9	1								
Leu	3.6	4	3.3	4					0.9	1	1.3	1					0.8		1.7	2	0.8	1	1.4	1
Tyr	1.3	1	3.8	3							1.6	2							0.8	1			2.1	3
Phe	1.4	1	2.1	2							1.1	1			1.4	1							0.8	1
Lys	1.3	1	3.4	5			0.7	1	1.0	1	0.8	1			1.0	1	1.5	1					1.0	1
Arg	1.1	1	2.2	2		1							1.0	1							0.7	1		
Hse	+	1	+	1															+	+				
Trp			ND	1		1	ND		ND		ND		ND		ND		ND		ND		ND		ND	2
Mobility 2.1							−0.43		0.44		−0.75		1.02		0.25		0.44		−0.75		0.32			
Mobility 3.5									0.43		0.12						0.12		0.00					

[a] Amino acid analysis. [b] Composition derived from the sequence. The sequence was determined as described in the text. Mobility relative to lysine = +1, aspartic acid = −1, neutral = 0. ND, not determined.

TABLE II: Amino Acid Composition of Moo CB2 and Peptides.

	CB2		CB2T1		CB2T2		CB2T3		CB2T4		CB2T5		CB2T6		CB2CT1		CB2CT2		CB2CT3		CB2CT4		CB2CT2-CHY3	
Amino acid	AAA[a]	SEQ[b]	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ
Asp	5.5	6				3					2.0	2	0.9	1			2.8	3			2.0	3	1.1	1
Thr	3.8	2				1					2.0	2	0.9	1			2.8	3			2.0	3	1.1	1
Ser	6.0	6	1.0	1		3			1.5	2	0.6				1.0	1	2.6	3	1.5	2				
Glu	5.0	5				4							0.7	1			3.0	4			1.3	1		
Pro	1.6	1				1											1.1	1						
Gly	4.2	4				3	4.6	1	0.7		0.5						2.8	3					1.5	1
Ala	4.0	4				3					1.3	1					2.0	3			1.1	1	1.9	2
Val	2.1	3	0.7	1		2									0.7	1	1.7	2					1.0	1
Ile	1.9	2				1			1.0	1							1.0	1	1.0	1				
Leu	3.5	3				1			0.5				2.0	2			1.5	1						
Tyr	3.0	3				2							1.2	1			1.4	2			1.0	1	0.6	1
Phe	1.5	1							1.0	1									1.0	1				
Lys	1.8	2				1					0.7	1					1.1	1			0.7	1	1.0	1
Arg	2.5	3	1.0	1			1.0	1	1.5	1					1.0	1	0.7	1	1.5	1			1.0	1
Hse	+	1											+								+			
Trp	ND	2	ND	1	ND	1	ND		ND		ND		ND		ND	1	ND	1	ND		ND			
Mobility 2.1					−0.37		0.43		−0.07		0.22		−0.37										0.48	
Mobility 3.5					0.02								−0.36											

[a] Amino acid analysis. [b] Composition derived from the sequence. The sequence was determined as described in the text. Mobility relative to lysine = +1, aspartic acid = −1, neutral = 0. ND, not determined.

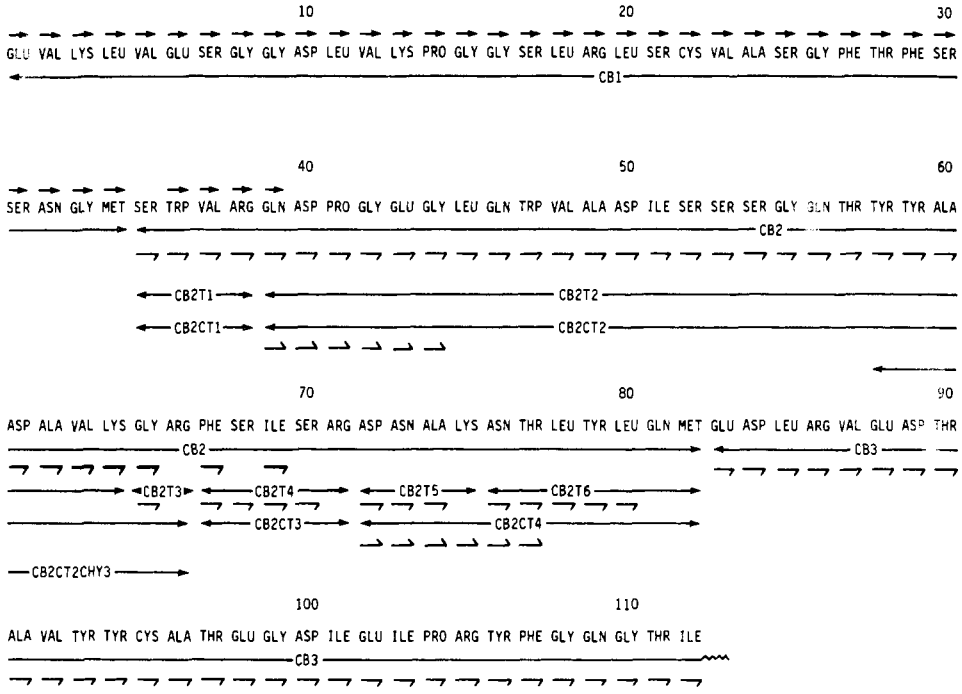

FIGURE 9: The sequence of the variable region of μ chain Moo. Identifications made by sequencing the intact heavy chain (→), the cyanogen bromide fragments (—), and the tryptic peptides (—) are indicated.

complete sequence of this cyanogen bromide peptide and the overlap into CB2 are shown in Figure 9. Since the sequence of the fragment was completely established by studies of the intact heavy chain, this peptide was not isolated.

The Sequence of the Second Cyanogen Bromide Fragment of the Variable Region of μ Chain Moo. CB2 was sequenced for 35 Edman degradation steps on the automated sequencer and identifications were made of all steps except 32 and 34. One micromole of CB2 was digested with trypsin and the resulting peptides were separated by high-voltage electrophoresis on paper at pH 2.1 and pH 3.5. Peptide CB2T1 was not sequenced since its sequence was determined on the sequencer run of the intact CB2 peptide. CB2T2 could not be isolated from paper. Peptides CB2T3, CB2T4, CB2T5, and CB2T6 were sequenced for 2, 4, 3, and 6 cycles, respectively. The sequence of each of these peptides was derived from the sequencer runs and the compositions of the peptides. The assignment of the amide to the penultimate residue of peptide CB2T6 was based on the mobility of the peptide on paper electrophoresis.

In order to confirm the order of peptides CB2T3, CB2T4, CB2T5, and CB2T6, which was initially made by homology, 1 μmol of CB2 was reacted with citraconic anhydride as described and then digested with trypsin. The peptide mixture was applied to a column of Sephadex G-25 fine, equilibrated with 0.5 M NH_3, and the eluate pooled as shown in Figure 7. Pool 1 contained CB2CT2 and pool 2 CB2CT4, both in high yield. CB2CT2 was deblocked with 50% acetic acid and sequenced for six cycles to establish its identification and purity. The placement of CB2T3 was confirmed by the composition of CB2CT2. CB2CT4 contained homoserine on amino acid analysis, placing it as the carboxy-terminal peptide. CB2CT4 was similarly deblocked and six cycles of sequential Edman degradation showed that CB2CT4 was made up of CB2T5 and CB2T6, placing CB2T5 as the penultimate tryptic peptide of CB2.

Although the sequence of CB2T2 was derived from the amino-terminal sequencer run, a peptide comprising the carboxy-terminal portion of CB2T2 was sought to confirm unequivocally the sequence. CB2CT2, comprising CB2T2 and CB2T3, was digested with chymotrypsin for 2 h in 1% ammonium bicarbonate and the resultant peptides were resolved by high-voltage electrophoresis. The composition of CB2CT2CHY3, the carboxy-terminal peptide of CB2CT2, agreed with the derived sequence.

	5	10	15	20
Canine Pool	E V Q L V E S G G D L V Q P G G S L R L S C V A			
Ga			A	
Le				V
Co				
Lo		A V		
St			A	I G
Gom			K	
Moo	K		K	

FIGURE 10: Amino-terminal sequences of canine heavy chains. The sequence of pooled heavy chain is shown at the top. For the myeloma proteins, only those residues that differ from the pool have been shown. Original data from Capra et al. (1973) and Kehoe and Capra (1972).

The sequence of CB2 is shown in Figure 9. The composition of this peptide and its fragments is shown in Table II.

The Sequence of the Third Cyanogen Bromide Fragment of the Variable Region of μ Chain Moo. The sequence of this portion of the variable region was determined by subjecting the peptide to 40 cycles of Edman degradation. All positions up to the V/C bridge were unambiguously identified as shown in Figure 9. Since this peptide contains substantial portions of the constant region, its composition is not germane and is not included.

The assignment of CB3 as the third cyanogen bromide peptide of the variable region was based on the striking simi-

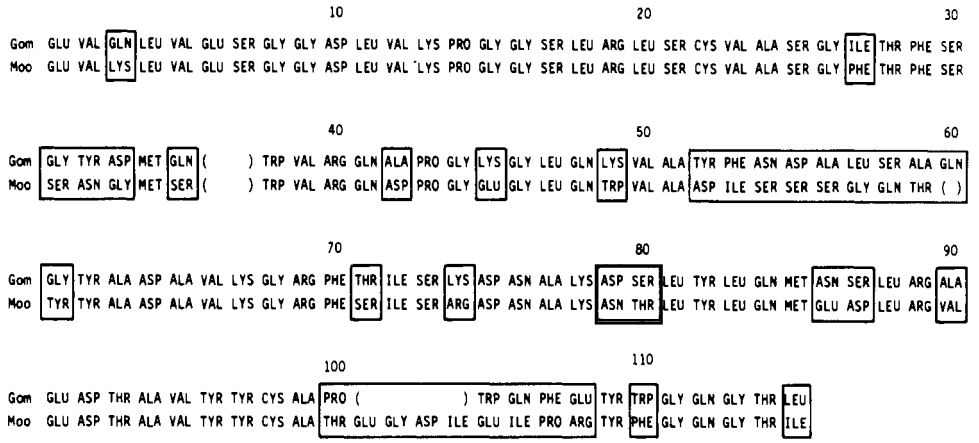

FIGURE 11: The complete amino acid sequence of two canine heavy chain variable regions. The differences between the two sequences are boxed. In human myeloma proteins hypervariable regions have been defined in the V_HIII subgroup from 31–35, 50–61, 83–87, and 100–108 in the numbering system shown here. The rest of the V region has been called the "framework."

larity of this peptide to human V_HIII proteins which have been sequenced.

Discussion

This study demonstrates that the primary structure of immunoglobulin heavy chain variable regions is highly conserved. Two unselected myeloma proteins produced spontaneously by a nonhuman species are strongly homologous to the several human V_HIII proteins which have been sequenced. The characteristic patterns of sequence similarity and difference derived from sequence studies of human proteins (Capra and Kehoe, 1974a) are evident in the two canine proteins studied. Phylogenetically associated residues, which have been described in the amino-terminal portion of heavy chains from several species and can be considered evolutionary markers (Capra et al., 1973), appear to be present in the variable region beyond the first hypervariable region. While the presence of phylogenetically associated residues beyond the first hypervariable region is not confirmed by these studies, this evidence suggests that hypotheses concerning mechanisms of variation and conservation operating on the variable region derived from examination of the portion amino terminal to the first hypervariable region apply to much of the variable region framework.

Figure 10 shows the sequence of the amino-terminal 24 residues of pooled canine heavy chains and the 7 homogeneous canine heavy chains that have been sequenced (Kehoe and Capra, 1972). There are a total of 10 out of 168 positions at which the sequences of the myeloma proteins differ from pooled sequence. Proteins Ga and St have the same substitution of alanine for glycine at position 15 and proteins Gom and Moo both have lysine instead of glytamine at position 13. The prototypic sequence, derived from all seven myelomas, is identical with the sequence of the pool. It is important to recognize that information derived from the sequence of myelomas statistically reflects the normal, and that the significance of such information is directly related to the number of myeloma proteins examined. For example, if only two proteins had been sequenced (Ga and St, or Gom and Moo), the prototypic sequence derived from myelomas would be different from normal pooled material.

The complete amino acid sequences of the variable regions of α chain Gom and μ chain Moo are shown in Figure 11. The differences in the two sequences have been boxed. The two dog proteins are identical at 76 of 113 positions. Of the 37 differences, 23 are in regions defined as hypervariable by sequence studies of human myeloma proteins (Capra and Kehoe, 1974a) and 11 are in framework residues. These degrees of differences in both the hypervariable regions and the framework portions of the variable region are quite similar to the degrees of differences found for any two unselected human myelomas of the V_HIII subgroup (Capra and Kehoe, 1974a). The locations of clustered differences in the two dog proteins are the same as the clusters of differences in human V_HIII proteins which define the hypervariable regions and, like the human V_HIII, differ from the human V_HII in the precise location of the second hypervariable region (V_HIII:52–61; V_HII:61–65).

In the framework portion of the variable region, all the differences reflect conservative, one base changes in the genetic code except the lysine–tryptophan interchange at position 49. Similar variation is seen in human myeloma proteins. Additionally, each canine sequence shows a high degree of homology to a prototypic framework sequence derived from the seven human myelomas (Gom 60 of 77 and Moo 59 of 77 framework residue identities) (Figure 12). The amino acids found at positions 10, 13, 23, 48, 65, 67, 75, 76, and 77 are the same in both dog proteins, but these particular amino acids have not been detected at the indicated residue positions in the human V_HIII proteins sequenced to date. Two of the differences, at positions 10 and 23, have been defined as phylogenetically associated residues based on the sequence of pooled dog heavy chains (Capra et al., 1973). These substitutions were also found in the five other dog proteins for which there are data available (Kehoe and Capra, 1972). Position 13, however, while identical in Gom and Moo, is different in the other five proteins and the pool, and is therefore not a phylogenetically associated residue. The other six positions (48, 65, 67, 75, 76, 77) *may* represent phylogenetically associated residues but definite assignment requires more data, either on pooled canine material or several other canine myeloma proteins.

It is interesting to note that there are no paired differences (i.e., where the dog proteins contain the same difference relative to the human) after position 91, the end of the third hypervariable region, and only 2 of 15 framework residues in this region have differences from the human. Thus, to date, neither subgroup specific (Kehoe and Capra, 1971) nor phylogenetically associated residues have been identified in the carboxy-terminal portion of the heavy chain variable region.

As has been previously noted, the only clusters of variability (less than 50% identity in a five residue sequence) in the two

canine variable regions are located within the previously defined hypervariable regions. Each canine hypervariable region contains at least one amino acid difference between the two proteins requiring a two nucleotide interchange in the DNA codon for that position. The greatest variability is seen in the first, second, and fourth heavy chain hypervariable regions which have been shown to contribute to the putative antibody combining site (Padlan et al., 1973; Poljak et al., 1973).

The first hypervariable region of both proteins contains the relatively conserved methionine at position 34. The significance of this finding, even when considered with similar findings in other species (Capra and Kehoe, 1975), remains unclear.

The second hypervariable regions of the two canine proteins exhibit the extensive variability characteristic of this region (Birshtein and Cebra, 1971) and contain several amino acid differences which necessitate two nucleotide interchanges in the DNA. In order to demonstrate the obvious homology between the two proteins after position 61, it is necessary to assume a deletion within the second hypervariable region of Moo. The finding of this deletion in the second hypervariable region confirms the fact that the size of the second hypervariable region is not subgroup specific.

Extreme variability and insertions or deletions are characteristic of the fourth hypervariable region. X-ray crystallographic studies (Padlan et al., 1973; Poljak et al., 1973) have shown that this hypervariable region forms a loop that is relatively distinct from the β-sheet framework of the heavy chain variable region. This hypervariable region probably plays a major role in defining the size of the antibody combining site. The variation in length seen between the two canine proteins is consistent with variations seen in other species and is consistent with the presumed function of this portion of the molecule.

The third hypervariable region is the least variable of the canine hypervariable regions. In the human proteins studied, this region is also less variable than the other hypervariable regions (Capra and Kehoe, 1974a). This hypervariable region does not contribute to the combining site of those proteins which have been studied by x-ray crystallography (Padlan et al., 1973; Poljak et al., 1973). Sequence studies on the pooled heavy chains of rabbits (Mole, 1975) have suggested that this hypervariable region may be an expression of genetic polymorphism. Such allotype or strain specific differences would be unrelated to the formation of the antibody combining site. The variability of the two canine proteins in this area is consistent with such an explanation since the proteins came from animals of two distinct breeds.

There is a high degree of sequence homology in the V_HIII subgroup among mammals. The sequences of both Gom and Moo frameworks are about 90% homologous to the human V_HIII protein Tie. This level of homology is greater than interspecies homology of any other portion of an immunoglobulin. Such conservation probably reflects the tertiary structural requirements necessary to support a functional antibody combining site. Despite structural requirements, both defined (i.e., by evolutionary relationships) and random variations, unrelated to the combining site, exist.

The distribution of phylogenetically associated residues over much of the framework portion of the variable region, suggested by the finding of six paired differences in the canine heavy chains beyond position 30, is significant. Phylogenetically associated residues reflect the forces of change to which the framework portion of the variable region is subject. The presence of phylogenetically associated residues in different portions of the framework is evidence that those portions of the framework are all controlled by the same genetic mechanism and influenced by evolutionary forces in the same manner. The absence of possible phylogenetically associated residues in the carboxy-terminal portion of the variable region suggests that a structural feature, either of the DNA coding for this portion of the molecule or of the protein itself, exerts a negative selection pressure on such variability. Confirmation of these findings on the distribution of phylogenetically associated residues in certain portions of the variable region requires the examination of pooled heavy chains. Based on the results of the sequences of the Gom and Moo heavy chain variable regions and the subgroup distribution of heavy chain variable regions, such studies would be most easily and most fruitfully undertaken on pooled canine immunoglobulin.

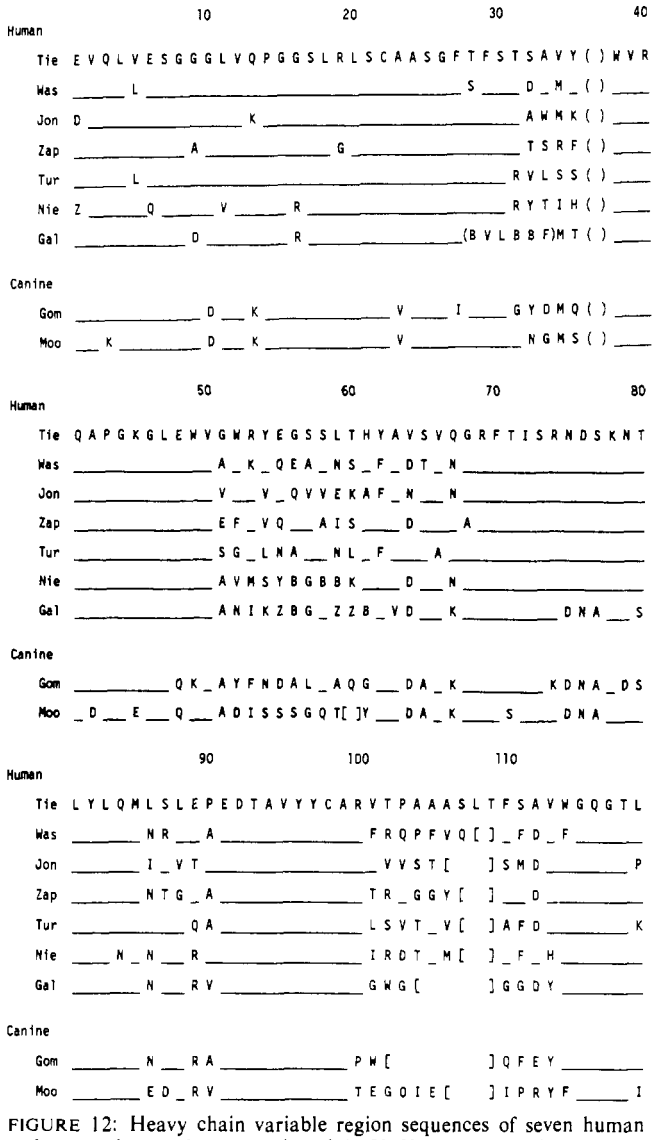

FIGURE 12: Heavy chain variable region sequences of seven human and two canine myeloma proteins of the V_HIII subgroup. The sequence of the human protein Tie is listed at the top. For the other proteins, only differences from Tie are shown.

The absence of phylogenetically associated residues after residue 91 and the confirmation of length variability in the second hypervariable region within the V_HIII subgroup require the introduction of additional mechanisms of genetic control into systems already burdened by theoretical explanations for which there are no biologic precedents.

The hypervariable region insertion model (Wu and Kabat, 1970; Capra and Kindt, 1975) deals satisfactorily with one of

these findings. Under this construct the framework is carried by one or a few germ-line genes while the many hypervariable regions are encoded by a large second group of genes, and the particular hypervariable regions necessary for a given specificity are introduced into the framework gene. Since in this formulation the hypervariable region sequences are carried as distinct genes, variations in the length of a given hypervariable region are easily introduced. This model does not account for the absence of phylogenetically associated residues in the carboxy-terminal portion of the variable region.

Thus, the findings reported herein place additional burdens on current theories of antibody diversity. Nonetheless, they establish that: (1) the high degree of sequence homology observed in previous studies of the amino-terminal portion of the heavy chain variable region extends throughout that domain. The structure of the heavy chain variable region appears to be more highly conserved in evolution than any other portion of the immunoglobulin molecule. (2) Phylogenetically associated residues are probably a feature of the framework portion of the heavy chain variable region up to, but not beyond, the third hypervariable region. (3) The primary structural design of the heavy chain variable region, alternating framework and hypervariable segments, is the same in dog and man as well as other species. These basic features necessary for the formation of an antibody combining site have been preserved in evolution at least since mammalian speciation.

Acknowledgments

We thank Margaret Cason and Jackie Rosenthal for their excellent technical assistance and Betty Bechtel for her secretarial services. We especially thank Dr. A. Hurvitz for the sera used in these studies.

References

Birshtein, B. K., and Cebra, J. J. (1971), *Biochemistry 10,* 4930.

Capra, J. D., and Kehoe, J. M. (1974a), *Proc. Natl. Acad. Sci. U.S.A. 71,* 845.

Capra, J. D., and Kehoe, J. M. (1974b), *Proc. Natl. Acad. Sci. U.S.A. 71,* 4032.

Capra, J. D., and Kehoe, J. M. (1975), *Adv. Immunol. 20,* 1.

Capra, J. D., Kehoe, J. M., Kotelchuck, D., Walter, R., and Breslow, E. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 431.

Capra, J. D., and Kindt, T. J. (1975), *Immunogenetics 1,* 417.

Capra, J. D., and Kunkel, H. G. (1970), *Proc. Natl. Acad. Sci. U.S.A. 67,* 87.

Capra, J. D., Wasserman, R. L., and Kehoe, J. M. (1973), *J. Exp. Med. 138,* 410.

Crumpton, M. J., and Wilkinson, J. M. (1965), *Biochem. J. 94,* 545.

Hilschmann, N., and Craig, L. C. (1965), *Proc. Natl. Acad. Sci. U.S.A. 53,* 1403.

Inman, J. K., Hannon, J. E., and Appella, E. (1972), *Biochem. Biophys. Res. Commun. 46,* 2075.

Kehoe, J. M., and Capra, J. D. (1971), *Proc. Natl. Acad. Sci. U.S.A. 68,* 2019.

Kehoe, J. M., and Capra, J. D. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 2052.

Köhler, H., Shimizu, A., Paul, C., Moore, V., and Putnam, F. W. (1970), *Nature (London) 227,* 1318.

Kunkel, H. G. (1954), *Method Biochem. Anal. 1,* 141-155.

Milstein, C. (1967), *Nature (London) 216,* 330.

Mole, L. E. (1975), *Biochem. J. 151,* 351.

Mole, L. E., Goodfriend, L., Lapkoff, C. B., Kehoe, J. M., and Capra, J. D. (1975), *Biochemistry 14,* 1216.

Niall, H. O., and Edman, P. (1967), *Nature (London) 216,* 262.

Novotny, J., Dolejs, L., and Franêk, F. (1972), *Eur. J. Biochem. 31,* 277.

Padlan, E. A., Segal, D. M., Spande, T. F., Davies, D. R., Rudikoff, S., and Potter, M. (1973), *Nature (London), New Biol. 245,* 165.

Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244,* 5597.

Poljak, R. J., Amzel, L. M., Avery, H. P., Chen, B. L., Phizackerley, R. P., and Saul, F. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 3305.

Ponstingl, H., Schwarz, J., Reichel, W., and Hilschmann, N. (1970), *Hoppe-Seyler's Z. Physiol. Chem. 351,* 1591.

Rudikoff, S., and Potter, M. (1974), *Biochemistry 13,* 4033.

Summers, M. R., Smythers, G. W., and Oroszlan, S. (1973), *Anal. Biochem. 53,* 624.

Watanabe, S., Barnikol, H. V., Horn, J., Bertram, J., and Hilschmann, N. (1973), *Hoppe-Seyler's Z. Physiol. Chem. 354,* 1505.

Wu, T. T., and Kabat, E. A. (1970), *J. Exp. Med. 132,* 211.