generated at the V/J junction, a point of high variability.

Acknowledgments

We acknowledge with pleasure the skilled technical assistance of Donna Atherton, Peggy Frank, Julie Winnette, and Martha Danhof. Fruitful discussions with Dr. L. Edward Cannon are gratefully acknowledged. We thank Kathy Able for typing this manuscript and Dr. Abe Clark for his careful reading of the final version.

References

Agnello, V., Arbetter, A., De Kasep, I., Powell, R., Tan, E., & Joslin, F. (1980) *J. Exp. Med. 151*, 1514-1527.

Andrews, D. W., & Capra, J. D. (1981) *Biochemistry* (following paper in this issue).

Bernard, O., Hozumi, N., & Tonegawa, S. (1978) *Cell (Cambridge, Mass.) 15*, 1133-1144.

Brient, B. W., & Nisonoff, A. (1970) *J. Exp. Med. 132*, 951-962.

Capra, J. D., & Kehoe, J. M. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4032-4036.

Capra, J. D., & Kehoe, J. M. (1975) *Adv. Immunol. 20*, 1-40.

Capra, J. D., & Nisonoff, A. (1979) *J. Immunol. 123*, 279-284.

Capra, J. D., Tung, A. S., & Nisonoff, A. (1975) *J. Immunol. 114*, 1548-1553.

Capra, J. D., Tung, A. S., & Nisonoff, A. (1977) *J. Immunol. 119*, 993-999.

Crumpton, M. J., & Wilkinson, J. M. (1965) *Biochem. J. 94*, 545-556.

Franklin, E. C., & Frangione, B. (1971) *J. Immunol. 107*, 1527-1540.

Jerne, N. (1974) *Ann. Immunol. (Paris) 125C*, 373-387.

Kabat, E., Wu, T. T., & Bilofsky, H. (1979) *Sequences of Immunoglobulin Chains*, U.S. Department of Health, Education and Welfare, U.S. Public Health Service, National Institutes of Health, Bethesda, MD.

Klapper, D. G., & Capra, J. D. (1976) *Ann. Immunol. (Paris) 127C*, 261-271.

Klapper, D. G., Wilde, C. E. III, & Capra, J. D. (1978) *Anal. Biochem. 85*, 126-131.

Kunkel, H. G., Agnello, V., Joslin, F. G., Winchester, R. J., & Capra, J. D. (1973) *J. Exp. Med. 137*, 331-342.

Kunkel, H. G., Winchester, R. J., Joslin, F. G., & Capra, J. D. (1974) *J. Exp. Med. 139*, 128-136.

Max, E. E., Seidman, J. G., & Leder, P. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3450-3454.

McCumber, L. J., Qadeer, M., & Capra, J. D. (1980) *Methods Pept. Protein Sequence Anal., Proc. Int. Conf., 3rd, 1979*, 165-172.

McLaughlin, C. L., & Solomon, A. (1972) *J. Biol. Chem. 247*, 5017-5025.

Pisano, J. J., Bronzert, T. J., & Brewer, H. B., Jr. (1972) *Anal. Biochem. 68*, 47-53.

Sakano, H., Huppi, K., Heinrich, G., & Tonegawa, S. (1979) *Nature (London) 280*, 288-294.

Seidman, J. G., Leder, A., Nau, M., Norman, B., & Leder, P. (1978) *(Washington, D.C.) 202*, 11-17.

Sher, A., & Cohn, M. (1972) *J. Immunol. 109*, 176-178.

Smithies, O., Gibson, D., Fanning, E. M., Goodfleisch, R. M., Gilman, J. G., & Ballantyne, D. L. (1971) *Biochemistry 10*, 4912-4921.

Williams, R. C., Kunkel, H. G., & Capra, J. D. (1968) *Science (Washington, D.C.) 161*, 379-380.

Amino Acid Sequence of the Variable Regions of Heavy Chains from Two Idiotypically Cross-Reactive Human IgM Anti-γ-globulins of the Wa Group†

David W. Andrews‡ and J. Donald Capra*

ABSTRACT: The amino acid sequences of the variable regions of the heavy chains derived from two idiotypically related human monoclonal rheumatoid factors are reported. The sequences were obtained through automated Edman degradations of the intact, pyroglutamate aminopeptidase cleaved heavy chains and peptides produced from digestion of the pentameric IgM with CNBr. The peptides generated from the CNBr reaction were further digested by trypsin and the *Staphylococcus aureus* V8 protease. Comparisons of these sequences and those of the light chains from these molecules [Andrews, D. W., & Capra, J. D. (1981) *Biochemistry* (preceding paper in this issue)] suggest that the idiotypic determinant(s) in this system may reside in the framework portions of these molecules or in their J segments.

In order to further our objective of defining a structural correlate of a serologically defined idiotype, we have sequenced the variable regions of the heavy chains of two idiotypically related rheumatoid factors. Previous structural analysis of two proteins from the Po cross-idiotypic group (Capra & Kehoe, 1974; Klapper & Capra, 1976) provided strong evidence that the CDRs of the light and heavy chains of these molecules generated the predominant idiotypic determinant in this system. Our initial studies on the light chains of two proteins from the Wa cross-idiotypic group (Andrews & Capra, 1981) did not confirm this view.

Since the idiotype in this system is dependent on both the heavy and light chains (Kunkel et al., 1973, 1974), the complete variable region sequences of both the heavy and light

† From the Department of Microbiology, The University of Texas Health Science Center at Dallas, Dallas, Texas 75235. *Received April 6, 1981.* Supported by National Institutes of Health Grant AI12127 and National Science Foundation Grant PCM 792348.

‡ Recipient of support from a training grant awarded to the Graduate Program in Immunology, The University of Texas Health Science Center at Dallas, by the National Cancer Institute, Department of Health, Education and Welfare (Grant No. CA09082). Present address: Harvard University, Biological Laboratories, Cambridge, MA 02138.

0006-2960/81/0420-5822$01.25/0

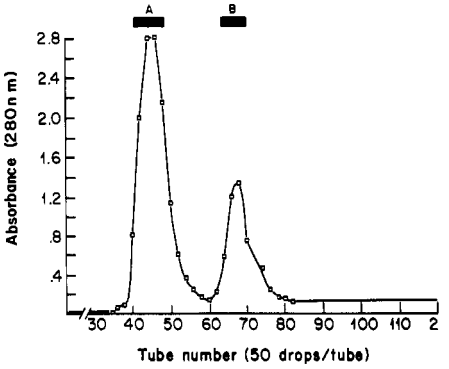

FIGURE 1: Gel filtration of the CNBr reaction mixture of Wol IgM through Ultrogel AcA34 in 5 M guanidine hydrochloride. Column dimensions were 2.5 × 100 cm, and the effluent was monitored by optical density at 280 nm. $NaDodSO_4$–polyacrylamide gel electrophoresis analysis [15% acrylamide; 1:18 ratio of bis(acrylamide) to acrylamide] of the completely reduced pool B revealed that the largest band was of 25 kilodaltons, consistent with it being light chain. Fractions were pooled as indicated.

chains are required in order to approach such a structural definition. This paper and the preceding one (Andrews & Capra, 1981) provide an opportunity for comparing the complete variable region amino acid sequences of two monoclonal rheumatoid factors which share a cross-reacting idiotype.

Experimental Procedures

The materials and methods were the same as described in the preceding paper (Andrews & Capra, 1981) with the additions indicated.

Cyanogen Bromide Reaction. Purified proteins or peptides (5–10 mg/mL) were dissolved in 70% HCOOH, and an amount of CNBr was added equal to a 100-fold molar excess over the methionine content of the reactant. The reactions were allowed to proceed at ambient temperature for 24 h, after which time they were diluted with H_2O and lyophilized.

Staphylococcus aureus V8 (SV8)[1] *Protease Digestion.* Proteins or peptides were dissolved in 50 mM NH_4HCO_3 (pH 8.0) to which an amount of the protease (Miles Laboratories) was added in a 1:50 ratio by weight. Reactions were allowed to proceed for 10–20 h at 37 °C, after which time they were lyophilized or applied to a column.

Pyroglutamate Aminopeptidase Digestion. Blocked proteins or peptides were dissolved in buffer, which consisted of 50 mM phosphate/10 mM EDTA/5 mM DTT/5% glycerol, pH 8.0, as described (Podell & Abraham, 1978). After flushing with N_2, 0.5 mg of pyroglutamate aminopeptidase (Boehringer-Mannheim) was added, and reactions were allowed to proceed for 9 h at 4 °C. Then, another 0.5 mg of enzyme was added, the vessel flushed with N_2 again, and reactions allowed to proceed for 15 h at ambient temperature. The resulting solutions were freed from salts by gel filtration or dialysis and then lyophilized.

$NaDodSO_4$–Polyacrylamide Gel Electrophoresis. The size and purity of proteins and peptides were assessed by using $NaDodSO_4$–polyacrylamide gel electrophoresis on slab gels at various concentrations of acrylamide and bis(acrylamide), according to the system devised by Laemmli (1970).

[1] Abbreviations used: SV8, *Staphylococcus aureus* V8; EDTA, ethylenediaminetetraacetic acid; TPCK, L-1-(tosylamido)-2-phenylethyl chloromethyl ketone; DTT, dithiothreitol; $NaDodSO_4$, sodium dodecyl sulfate; pGlu and pE, pyrrolidonecarboxylic acid residue; DNA, deoxyribonucleic acid.

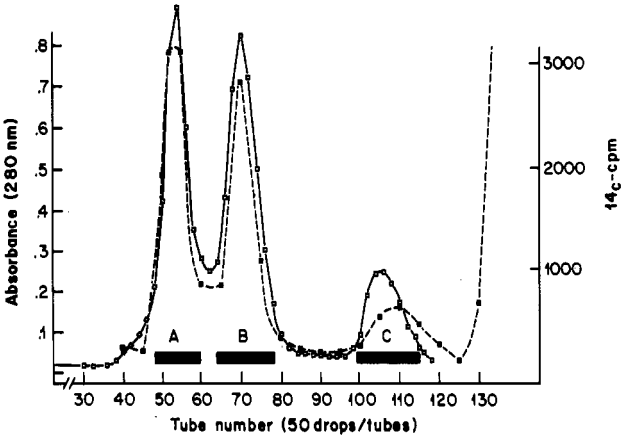

FIGURE 2: Gel filtration of completely reduced and alkylated pool B (Figure 1) through Sephadex G-100 in 5 M guanidine hydrochloride. Column dimensions were 2.5 × 100 cm, and the effluent was monitored by optical density at 280 nm (solid line) and liquid scintillation counting (dashed line) (20-μL aliquots). By $NaDodSO_4$–polyacrylamide gel electrophoresis, pool A has one band of 25 kilodaltons, pool B has two closely spaced bands of 16 kilodaltons, and pool C has one band of 8 kilodaltons. Fractions were pooled as indicated.

Results

The strategy used in generating peptides was the same for both cryoglobulins. The initial procedure involved reaction of the intact IgM pentamer with cyanogen bromide. Filtration of the resulting mixture through a column of Ultrogel AcA34 equilibrated in 5 M guanidine hydrochloride gave virtually identical elution profiles for both Wol and Sie cryoglobulins (for example, Figure 1). These included two major peaks, one excluded from the gel (pool A) and another eluting directly after it (pool B).

Analysis of the major peaks by $NaDodSO_4$–polyacrylamide gel electrophoresis revealed that pool B was Fab. While not conclusive, the gel pattern suggested that both μ chains contained a methionyl residue near position 80, as well as the expected constant region methionines at 204, 318, and 336 in the Fd portion of the μ chain. This disposition of methionines would result in peptides of 25, 16, 8, and 2 kilodaltons.

Each pool B from the previous CNBr reaction was completely reduced, alkylated, and then filtered through a column of Sephadex G-100 equilibrated in 5 M guanidine hydrochloride. The resulting elution profiles (for example, Figure 2) were consistent with the prior $NaDodSO_4$ gel patterns. Peaks were pooled as indicated and freed from salts by dialysis, followed by lyophilization. Pool A was light chain, pool B contained at least two polypeptides of ~16 kilodaltons, and pool C contained at least one peptide of ~8 kilodaltons.

By use of this single, predictable, reproducible chemical cleavage, the light chains and all peptides necessary for sequencing the variable regions of the heavy chains could be isolated.

Initial studies of both Wol and Sie μ chains involved isolation of intact heavy chains. This was accomplished via complete reduction and alkylation of the IgM pentamer, followed by filtration through a column of Sephadex G-100 in 5 M guanidine hydrochloride. After being freed of salts by dialysis and lyophilization, the heavy chains were generally quite insoluble. They could be solubilized, however, by citraconylation and could then be dialyzed into the appropriate buffer for treatment with pyroglutamate aminopeptidase, as described (Podell & Abraham, 1978). After dialysis and lyophilization, the resulting cleaved heavy chains were subjected to the automated amino acid sequencer. In this manner, the sequence of the initial 20 residues of both chains was obtained. The

Table I: Tryptic Peptides of Wol μ (1–81)

	TD-B1B (20–23)		TB-B3 (24–33)		TB-B2 (39–43)		TD-N1B (55–60)		TC-B1 (61–67)		TB-B4 (68–81)	
amino acid	AAA[a]	Seq[b]	AAA	Seq	AAA	Seq	AAA	Seq	AAA	Seq	AAA	Seq
S-CM-Cys[c]	0.4	1										
Asp			0.9	1			1.3	1	0.8	1	0.7	1
Thr			1.7	2								
Ser	1.4	1	0.8	1					1.2	1	2.7	3
Glu					1.0	1	0.7	1			0.8	1
Pro					0.9	1			ND[d]	1	0.8	1
Gly			2.0	2	0.8	1	1.2	1	1.3	1		
Ala					1.3	1					1.4	1
Val	1.1	1	1.4	1			0.8	1	1.8	2	1.9	2
Met												1[e]
Ile												
Leu											0.8	1
Tyr			0.8	1								
Phe			0.8	1			0.7	1			1.0	1
His											0.9	1
Lys	1.0	1	0.8	1	0.6	1	0.7	1			1.0	1
Arg												
Trp									ND	1		

[a] AAA = amino acid analysis. [b] Seq = determined from sequence analysis. [c] *S*-(Carboxymethyl)cysteine. [d] ND = not determined. [e] Determined as homoserine.

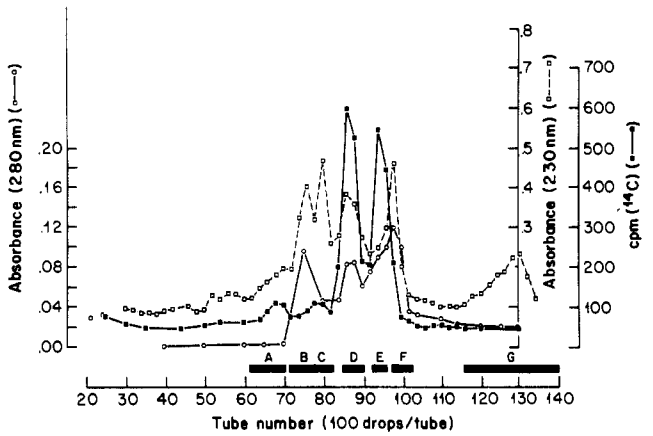

FIGURE 3: Gel filtration of the tryptic digest of Wol μ (1–81), derived from pool C (Figure 2), through Sephadex G-50SF in 1% ammonium bicarbonate. Column dimensions were 0.8 × 100 cm, and the effluent was monitored by optical density at 280 and 230 nm, as well as by liquid scintillation counting (20-μL aliquots). Fractions were pooled as indicated.

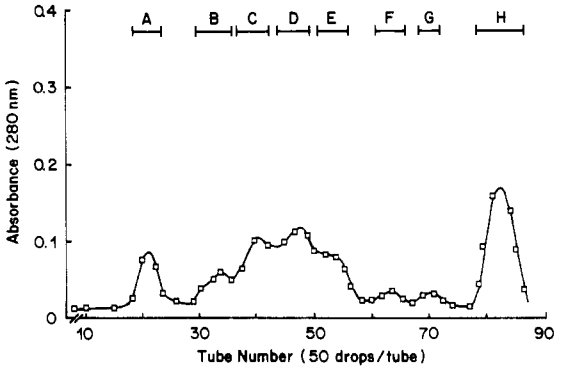

FIGURE 4: Gel filtration of the SV8 digest of Wol μ (1–81), derived from pool C (Figure 2), through Sephadex G-50SF in 1% ammonium bicarbonate. Column dimensions were 0.8 × 100 cm, and the effluent was monitored by optical density at 280 nm. Fractions were pooled as indicated.

remainder of the sequence of the variable regions was determined from peptides derived as described from CNBr cleavage of the intact IgM pentamer.

Wol Heavy Chain. (1) Residues 1–81. This portion of the sequence was determined from pool C, derived from the complete reduction and alkylation of the Fab fragment of Wol IgM (Figure 2). Two procedures were used to generate peptides from this fragment: tryptic digestion and digestion with the *S. aureus* V8 protease (SV8).

Tryptic digestion of 5 mg of pool C (Figure 2) and filtration of the reaction mixture through a column of Sephadex G-50SF in 1% NH_4HCO_3 gave the elution profile in Figure 3. All of the pools were subjected to analytical high-voltage paper electrophoresis at pH 3.5 and pH 6.5. Analysis of these data lead to all of the fractions, save pool G, being subjected to preparative high-voltage paper electrophoretic purification.

Six milligrams of pool C (Figure 2) was subjected to digestion with SV8 protease. Filtration of the reaction mixture through a column of Sephadex G-50SF equilibrated in 1% NH_4HCO_3 gave the elution profile in Figure 4.

The peptides from this digest were, in general, larger and less amenable to paper electrophoresis than those produced in the tryptic digest. Whether this was due to their size or charge characteristics, or more likely a combination of these

Table II: SVB Peptides of Wol μ (1–81)

	SV-F (32–46)		SV-G (47–58)	
amino acid	AAA[a]	Seq[b]	AAA	Seq
S-CM-Cys[c]				
Asp			1.1	1
Thr				
Ser				
Glu	2.3	2	2.4	2
Pro	1.3	1	0.9	1
Gly	2.8	3	1.8	2
Ala	1.1	1		
Val	1.0	1	1.1	1
Met				
Ile			0.8	1
Leu	1.9	2	0.9	1
Tyr	0.9	1		
Phe			0.9	1
His				
Lys	1.8	2		
Arg	ND[d]	1	ND	1
Trp	ND	1	ND	1

[a] AAA = amino acid analysis. [b] Seq = determined from sequence. [c] *S*-(Carboxymethyl)cysteine. [d] ND = not determined.

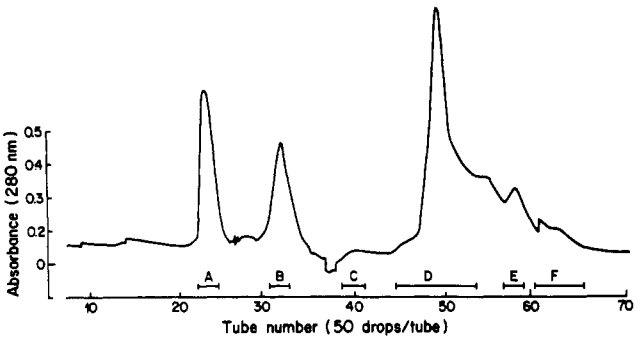

FIGURE 5: Gel filtration of the tryptic digest of pool B (Figure 2) through Sephadex G-50SF in 1% ammonium bicarbonate. Column dimensions were 0.8 × 100 cm, and the effluent was monitored by optical density at 280 nm. Fractions were pooled as indicated.

features, was not investigated. They were submitted for automated sequence analysis without further purification.

The peptides from these digests that were used for deducing the sequence of this portion of Wol heavy chain are illustrated in Tables I and II. The sequences derived from the peptides generated in these two digests, in concert with the data produced from the unblocked heavy chain, completed the sequence for residues 1–81 of the Wol chain.

(*2*) *Residues 82–123.* This part of the sequence was determined from pool B, derived from the complete reduction and alkylation of the Fab fragment of Wol IgM (Figure 2). This pool contained at least two peptides by the criterion of $NaDodSO_4$–polyacrylamide gel electrophoresis, and it was expected to contain the fragments spanning residues 82–204 and 205–318. Fortuitously, the amino-terminal residue of the latter fragment is glutamine, which cyclizes during the CNBr reaction, rendering it indifferent to the Edman reagents. Thus, the mixture can be sequenced with the confidence that only the variable region peptide is being degraded. In this way, the sequence of the initial 25 residues of this peptide was obtained.

Tryptic digestion of this mixture, followed by filtration through a column of Sephadex G-50SF in 1% NH_4HCO_3 gave the elution profile in Figure 5. Lyophilization of the excluded peak (pool A) from this filtration afforded a peptide extending from position 99 to position 168, free of contamination with any other peptides. Automated sequence analysis of this peptide completed the sequence of the variable region of Wol heavy chain. The complete sequence of the variable region of Wol heavy chain is shown in Figure 6.

Sie Heavy Chain. (*1*) *Residues 1–81.* The sequencing of this portion of Sie heavy chain was addressed in the same manner as was that of Wol heavy chain, with hydrolyses catalyzed by trypsin and SV8 protease. Five milligrams of the 1–81 fragment (isolated in a manner similar to that used for 1–81 of Wol; see Figure 2) was subjected to hydrolysis catalyzed by trypsin. The reaction mixture was filtered through a column of Sephadex G-50SF equilibrated in 1% NH_4HCO_3 giving the elution profile in Figure 7. These peptides were labeled T-A–H. The pools T-B, T-C, T-D, T-E, and T-F were subjected to preparative high-voltage electrophoresis. Peptides T-A, T-G, and T-H were judged substantially pure. The peptides from this digest that were used in deducing the sequence of this portion of Sie heavy chain are illustrated in Table III.

Six milligrams of the 1–81 fragment was subjected to hydrolysis catalyzed by SV8 protease. The reaction mixture was filtered through a column of Sephadex G-50SF, equilibrated in 1% NH_4NCO_3, giving the elution profile in Figure 8. Analysis of these pools by high-voltage paper electrophoresis

Table III: Tryptic Peptides of Sie μ (1–81)

	TD-B3 (44–53)		T-G (54–64)		TD-N1 (65–67)		T-A (68–81)	
amino acid	AAA[a]	Seq[b]	AAA	Seq	AAA	Seq	AAA	Seq
S-CM-Cys[c]								
Asp			0.8	1			1.2	1
Thr			1.2	1			1.1	1
Ser	1.3	1					1.9	2
Glu	0.8	1	1.1	1	0.9	1	1.1	1
Pro	0.9	1	1.3	1			1.1	1
Gly	2.0	2	1.3	1				
Ala	1.4	1					1.3	1
Val	1.0	1	0.9	1			1.6	2
Met								1[e]
Ile			0.8	1				
Leu	0.8	1					0.8	1
Tyr			0.9	1			1.3	1
Phe			1.3	1			0.8	1
His								
Lys	1.0	1	0.8	1			1.3	1
Arg	1.0	1			ND	1		
Trp	ND[d]	1	ND	1	ND	1		

[a] AAA = amino acid analysis. [b] Seq = determined from sequence. [c] *S*-(Carboxymethyl)cysteine. [d] ND = not determined. [e] Determined as homoserine.

Table IV: Insoluble Residue from SVB Digest of Sie μ (1–81)

	SV-I (11–46)	
amino acid	AAA[a]	Seq[b]
S-CM-Cys[c]	0.8	1
Asp		
Thr	3.8	4
Ser	5.4	5
Glu	2.4	2
Pro	ND[d]	2
Gly	6.7	6
Ala	1.1	1
Val	3.7	4
Met		
Ile	0.7	1
Leu	0.8	1
Tyr	0.9	1
Phe	0.9	1
His		
Lys	2.6	3
Arg	2.1	3
Trp	ND	1

[a] AAA = amino acid analysis. [b] Seq = determined from sequence. [c] *S*-(Carboxymethyl)cysteine. [d] ND = not determined.

showed that they were all substantially pure and needed no further purification. An insoluble residue from this reaction was filtered through a column of the same gel, equilibrated in 9% HCOOH. Pool A (SV-I) (Table IV) from this filtration was lyophilized and gave a homogeneous sequence, extending from position 11 to position 46 of the heavy chain. The sequence of this peptide, together with the previously described tryptic peptides, completed the sequences of residue 1–81 of Sie heavy chain.

(*2*) *Residues 82–123.* This portion of the Sie heavy chain was determined by employing the 82–204 and 205–318 mixture derived from the complete reduction and alkylation of the Fab fragment of Sie IgM. As was the case for this pool from Wol IgM, one peptide (205–318) becomes blocked to the Edman reaction during the cyanogen bromide reaction. This fortuitous circumstance allows the accurate analysis of a sequencer run on this mixture. An automated sequencer run

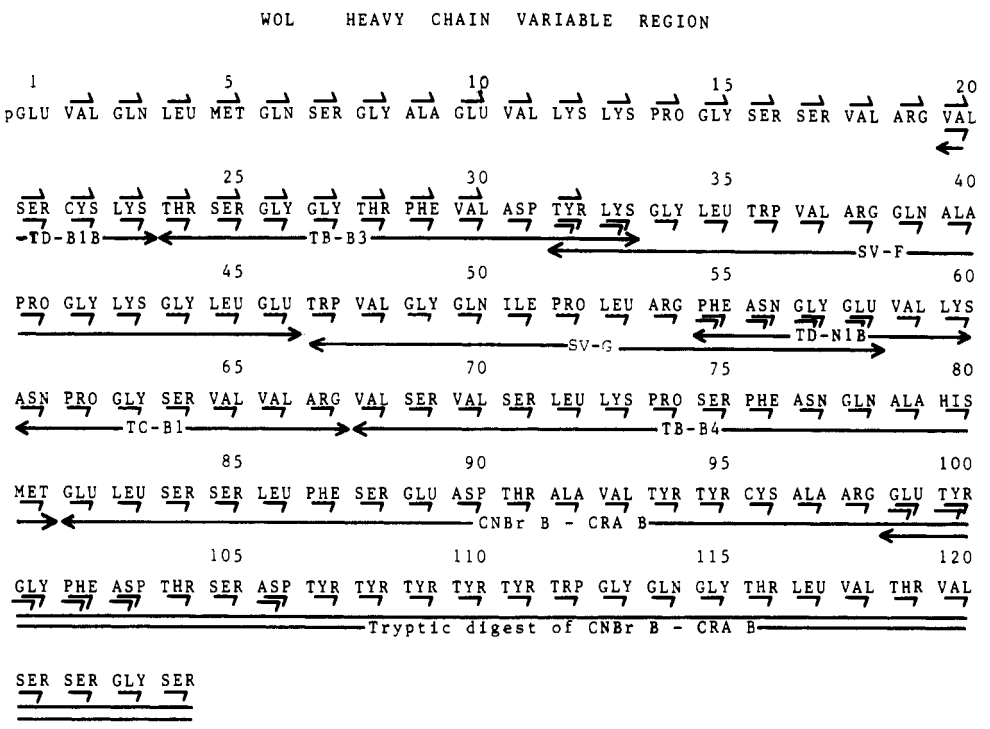

FIGURE 6: Amino acid sequence of the variable region of Wol heavy chain. Arrows above the sequence represent assignments made on a sequencer run of unblocked heavy chain. Arrows below the sequence are assignments made on individual sequencer runs on isolated peptides.

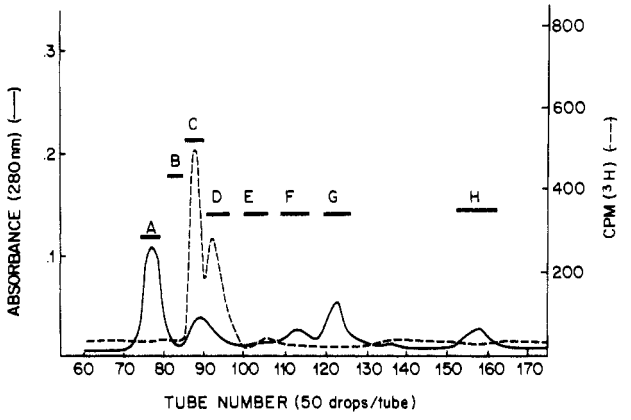

FIGURE 7: Gel filtration of the tryptic digest of Sie μ (1–81) through a column of Sephadex G-50SF in 1% ammonium bicarbonate. Column dimensions were 2.5 × 200 cm, and the effluent was monitored by optical density at 280 nm and by scintillation counting (20-μL aliquots). Fractions were pooled as indicated.

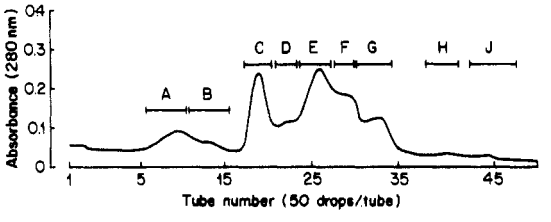

FIGURE 8: Gel filtration of the SV8 digest of Sie μ (1–81) through a column of Sephadex G-50SF in 1% ammonium bicarbonate. Column dimensions were 0.8 × 100 cm, and the effluent was monitored by optical density at 280 nm. Fractions were pooled as indicated.

of 40 steps was carried out, and identifications could be made through the beginning of the last complementarity-determining region.

The remainder of this portion of the variable region was obtained by hydrolysis of the peptide mixture catalyzed by trypsin. Filtration of the reaction mixture through a column of Sephadex G-50SF equilibrated in 1% NH_4HCO_3 gave an excluded peak that was lyophilized and subjected to automated sequencing. This methodology was somewhat complicated by the presence of a lysine residue at position 101, within the last complementarity-determining region. Indeed, the presence of another trypsin-sensitive residue within this region would have made this approach considerably less attractive. Fortunately, this was not the case, and the large peptide produced in this digest began at position 102 and extended to position 168. Automated sequence analysis of this peptide completed the sequence for the variable region of Sie heavy chain. The complete sequence of the variable region of Sie heavy chain is illustrated in Figure 9.

Discussion

With respect to the heavy chains (Figure 10), Sie and Wol share the similarities in framework constitution that one would expect in proteins of the same variable region subgroup. In the variable region, exclusive of the CDRs and the putative J segment, there are nine differences (positions 5, 21, 30, 43, 69, 80, 84, 85, 88, and 91) between Sie and Wol (88% homology). However, these heavy chains are difficult to classify in any of the V_H subgroups. The homology exhibited between these heavy chains and the V_HI prototype, Eu (Eu/Sie = 75%; Eu/Wol = 79%; see Figure 11), is considerably less than one would expect among proteins of the same V_H subgroup (usually around 90%). As a general rule, intrasubgroup homology is in the 80–90% range, while the intersubgroup homology (neglecting V_HI) is 40–70%. Sie and Cor (a V_HII protein) have 43% homology, and Wol and Cor have 45% homology. Thus, Sie and Wol are clearly within the intrasubgroup range, but their relationship to the V_HI prototype, Eu, is intermediate.

This issue is complicated by the fact that there is only one complete V_HI sequence, Eu, which was reported over 10 years ago (Cunningham et al., 1970). The only other V_HI sequence that is substantially complete is Di (Kohler et al., 1970; Florent et al., 1974), which has 78 of 123 residues assigned in its variable region, of which 64 are in the framework portions. Of these 64 framework residues, 22 are different between Di and Eu, and 20 are different between Sie and Di. Thus, we

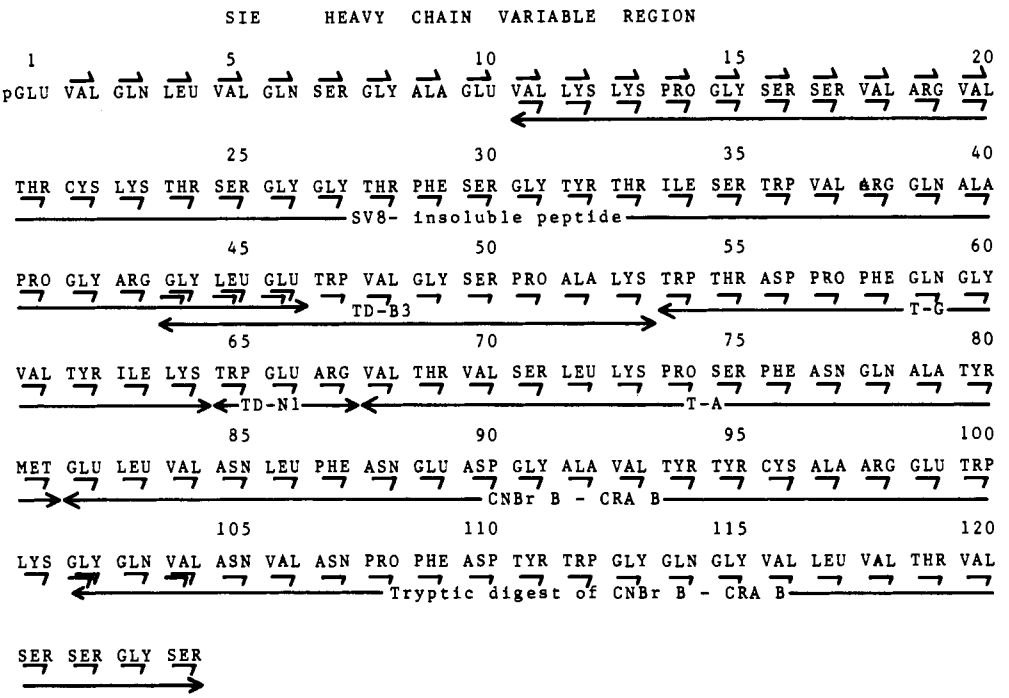

FIGURE 9: Amino acid sequence of the variable region of Sie heavy chain. Arrows above the sequence represent assignments made on a sequencer run of unblocked heavy chain. Arrows below the sequence are assignments made on individual sequencer runs of isolated peptides.

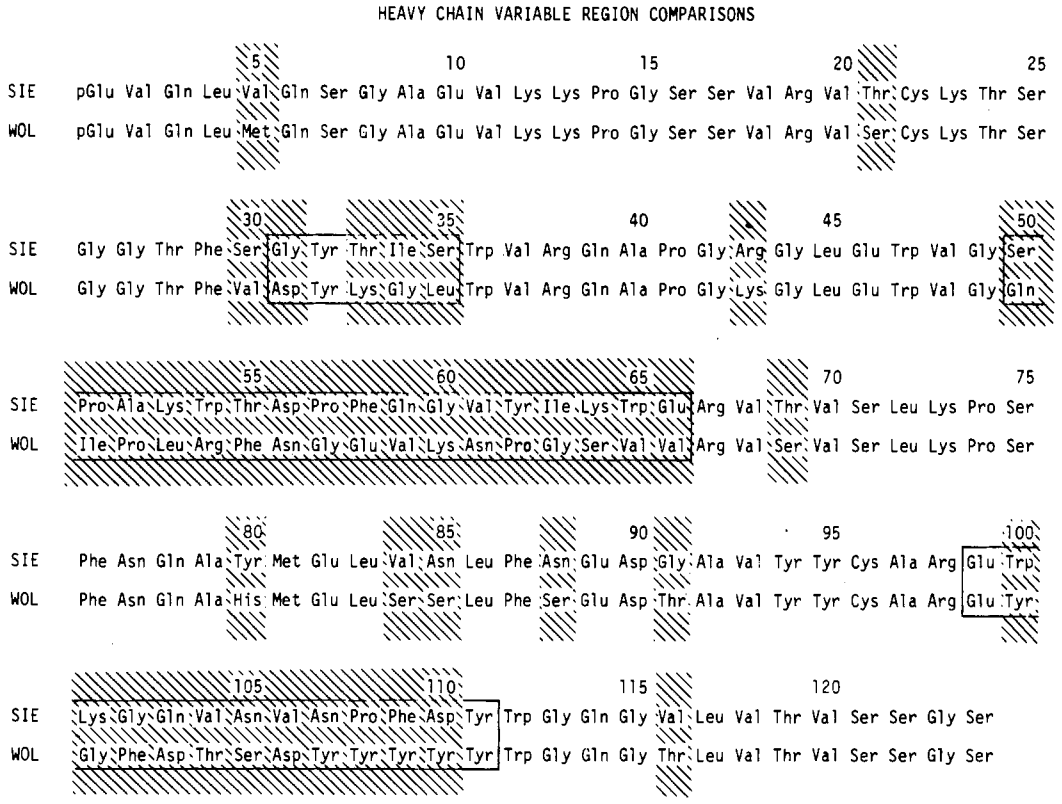

FIGURE 10: Amino acid sequences of the variable regions of the heavy chains of two IgM anti-γ-globulins, Sie and Wol, which share idiotypic determinants. The differences in the sequences are shaded, and the complementarity-determining regions are boxed.

believe an assignment of Sie and Wol to $V_H I$ (or any other extant subgroup) is premature.

Within the complementarity-determining regions, Sie and Wol heavy chains differ at 32 of 35 positions. These heavy-chain data make a connection between idiotype and complementarity-determining regions difficult to rationalize in this idiotypic system. There are several possible interpretations for these data.

Sie and Wol are not members of the same idiotypic set. This is the most obvious and trivial conclusion. However, it seems untenable. In addition to their original assignment to the Wa group of monoclonal rheumatoid factors, recent data obtained by using a different system confirm this assignment (V. Agnello, personal communication; Agnello et al., 1980).

The presumptive anti-idiotype actually recognizes patterns related to framework sequences. This is quite possible. The four chains of these two molecules are strikingly similar [98% homology in the light chains (Andrews & Capra, 1981) and 88% homology in the heavy chains, both with respect to framework residues].

HEAVY CHAIN VARIABLE REGION COMPARISONS

	10	20	30	40
SIE	pE V Q L V Q S G A E V K K P G S S V R V T C K T S G G T F S G Y T I S W V R Q A			
WOL	——— M ——— S ——— V D – K G L ———			
EU	——— K – S — A ——— R S A – I ———			

	50	60	70	80
SIE	P G R G L E W V G S P A K W T D P F Q G V Y I K W E R V T V S L K P S F N Q A Y			
WOL	— K ——— Q I P L R F N G E V K N P G S V V — S ——— H			
EU	— Q ——— M – G I V P M F G – P N Y A Q K F Q G —— I T A D E – T – T —			

	90	100	110	120
SIE	M E L V N L F N E D G A V Y Y C A R E W K G Q V N V N P F D Y W G Q G V L V T V S S G S			
WOL	—— S S — S — T ——— Y G F D T S D Y Y Y Y ——— T ———			
EU	—— S S – R S — T – F – F — G G Y G I Y S( )P E E Y N – G ———			

FIGURE 11: Comparison of the amino acid sequence of the heavy chain variable regions of Sie, Wol, and Eu (Cunningham et al., 1970), the prototype for the V_{HI} subgroup.

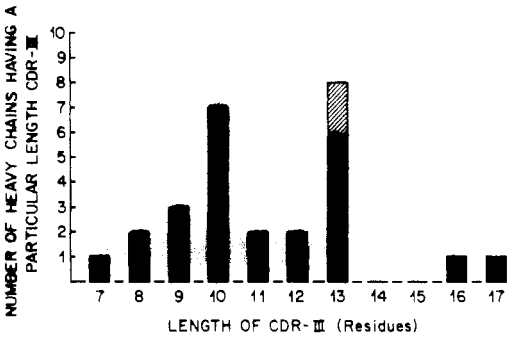

FIGURE 12: Histogram of the number of residues in the third complementarity-determining region of human heavy chains vs. the number of residues. The diagonal lines shown at length 13 indicate data derived from this report.

The idiotype may be correlated with the length of the third CDR in the heavy chain. This is obviously a gross oversimplication. And it is also prima facie disprovable since Lay, Pom, Wol, and Sie all have the same length of CDR III but are idiotypically distinct. However, the constitution and the length of this CDR may *together* generate an idiotypic determinant. Also, the possibility is worth considering from the point of view of specificity, if not idiotype.

The lengths of the first and second CDRs of human heavy chains show very little variation. However, there is a large variation in the lengths of CDR III among human heavy chains. This is graphically represented in Figure 12. The length of CDR III is critical in determining the size and shape of the antibody combining site [reviewed by Padlan (1977)] and may have a similar effect on the idiotypic determinants. The heavy chains of Wol, Sie, Pom, and Lay all have the same length of their third CDR. The probability of this having occurred by chance can be addressed by the following analysis.

The probability of picking two heavy chains at random that have CDR III of equal length is represented by

$$P = p(7)p(7/7) + p(8)p(8/8) + p(9)p(9/9) + p(10)p(10/10) + p(11)p(11/11) + p(12)p(12/12) + p(13)p(13/13) + p(14)p(14/14) + p(15)p(15/15) + p(16)p(16/16) + p(17)p(17/17)$$

where $p(n)$ is the probability of picking a heavy chain with a CDR III of length n residues, and $p(n/n)$ is the conditional probability of picking a heavy chain with a CDR III length n residues, *having already picked one of n residues.*

By use of the data of Kabat et al. (1979), this equation becomes

$$P = (1/25)(0/24) + (2/25)(1/24) + (3/25)(2/24) + (7/25)(6/24) + (2/25)(1/24) + (2/25)(1/24) + (6/25)(5/24) + 0 + 0\ (1/25)(0/24) + (1/25)(0/24)$$

$$P = (2 + 6 + 42 + 2 + 2 + 30)/[(25)(24)]$$

$$P = 84/6000$$

$$P = 0.14$$

The probability of picking four human heavy chains having their CDR III of equal length can be addressed in the same way

$$P = 0 + 0 + 0 + (7/25)(6/24)(5/23)(4/22) + 0 + 0 + (6/25)(5/24)(4/23)(3/22) + 0 + 0 + 0 + 0$$

$$P = (840 + 360)/[(25)(24)(23)(22)]$$

$$P = 1200/303600$$

$$P = 0.00395$$

Thus, the probability that Wol, Sie, Lay, and Pom all have the same length CDR III simply by chance is less than 1 in 250. It is unclear, however, what this reveals about these molecules. The fact that they are all rheumatoid factors *may* be significant. Five other human heavy chains have a CDR III of 13 residues. They are Cor, Tei, Tro, Cam, and Ga. None of the molecules from which these are derived are reported to have any demonstrable specificity or idiotypic classification.

The presumptive anti-idiotype is specific for determinants formed by the J segments of Wol and Sie heavy chains. Another possibility for a structural correlate to idiotype in heavy chains is the section analogous to the J segment in light chains. The movement of the DNA encoding the V segment (residues 1–96) to a position contiguous with one of several J segments (residues 97–109) has been extensively analyzed in the mouse λ chain system (Seidman et al., 1978; Bernard et al., 1978; Sakano et al., 1979; Max et al., 1979).

The situation in heavy chains is complicated by two aspects of their natural history of expression. The first is that different classes of heavy chain may express, in a sequential fashion, the same variable region (Nossal et al., 1964; Gally &

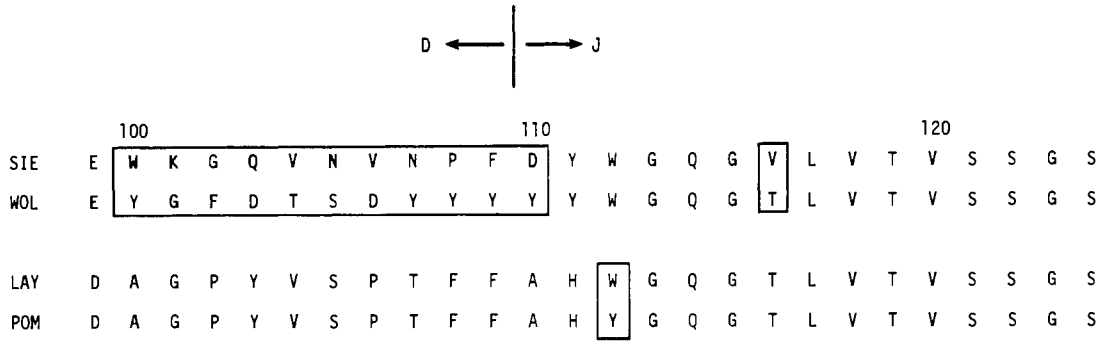

FIGURE 13: D/J_H segments of the heavy chains of four IgM anti-γ-globulins. Sie and Wol share idiotypic specificities which are different from those shared by Lay and Pom. The differences *within* each set are boxed.

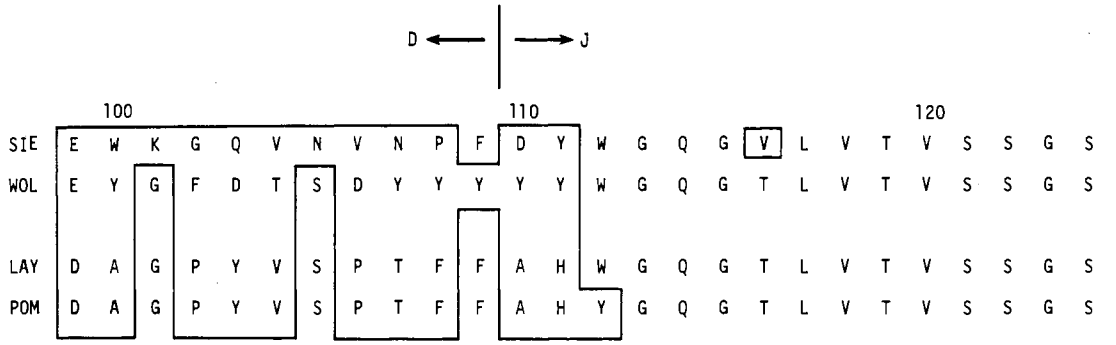

FIGURE 14: D/J_H segments of the heavy chains of four IgM anti-γ-globulins. Sie and Wol share idiotypic specificities which are different from those shared by Lay and Pom. The differences *between* the two sets are boxed.

Edelman, 1972). This so-called "switch" obviously requires genetic manipulations beyond those involved in light chains. Secondly, a lymphocyte can make two varieties of the same class of heavy chain: one to be inserted in the membrane and serve as a receptor; another to be secreted. These are structurally different and demand different coding and/or processing information.

Nevertheless, it appears that expression of the variable regions of heavy chains is analogous to that in light chains. DNA segments corresponding to variable region (position 1 to ~100) and J_H segments (approximately 104–123) have been sequenced from the mouse (Early et al., 1980; Bernard & Gough, 1980), and the presence of a third segment, to connect V_H and J_H, had to be invoked to explain this gap. In fact, its presence was inferred from sequence data (Capra & Kindt, 1975). There is a section of variability in mouse heavy chains at this point, flanked by areas of more conserved sequences. This segment was termed "D" (for diversity) by Early et al. (1980).

The human genes corresponding to these areas have not, as yet, been as completely studied, but early reports suggest a similar organization (Matthyssens & Rabbitts, 1980). Their presence has been inferred from protein sequences of the region 100–125 in human heavy chains (Lehman & Putnam, 1980; Rao et al., 1979). The boundaries of V_H, J_H, and D have not yet been well-defined, either in mice or humans, and a consensus will probably have to wait on the DNA sequences.

It is, however, clear that there exists in human heavy chains a segment abutting the third CDR that is relatively conserved and which assorts itself independently with respect to different V_H regions. This segment is analogous to the J_H segment in mouse immunoglobulin heavy chains. Whether this segment is responsible for the idiotype in some systems is the next of the possible conclusions from our sequence data.

The connection between J segments and idiotype seems best exemplified in the anti-dextran antibodies in the mouse (Schilling et al., 1980). In this idiotypic system, there are two kinds of idiotype: a common idiotype (Idx) shared by all dextran-binding mouse antibodies and an individually specific idiotype (Idl) which typifies a single antibody (IdIM104 and IdIJ558 are the two examples derived from myeloma proteins). The structural basis for this serological difference between these two molecules appears to be confined to the third CDR of the heavy chain, more specifically to the D segment. Thus, only molecules which are identical, or differ by a single conservative substitution in this section, fully express this set of determinants. Differences in the D and/or J segments abrogate this reactivity, while leaving the Idx unaffected. In this system, then, one form of idiotypy is associated with the third CDR and the J_H segment.

In the human system, such correlations are much more difficult to accumulate. Two conditions must be met in order to apply such a definition to the idiotypically defined sets among monoclonal rheumatoid factors. First, those molecules sharing idiotype must exhibit virtual identity in their D and/or J segments. Second, members of different idiotypic sets must differ in these areas.

These conditions are met in both sets of the monoclonal rheumatoid factors. With respect to heavy chains (Figure 13 and 14), Lay and Pom have identical third CDRs, and their J_H segments have but one difference, a tyrosine/tryptophan substitution at position 112. Sie and Wol are quite different in their D segments, having but one residue in common out of eleven, but their J_H segments differ at only two positions (position 110, Asp/Tyr; position 116, Val/Thr). The J_H segments of Sie and Wol differ from Lay and Pom at two, three, or four positions, depending on the comparison: Wol and Pom at 110, 111, and 112; Wol and Lay at 110 and 111; Sie and Pom at 110, 111, 112, and 116; Sie and Lay at 110, 111, and 116. These are all structurally nonconservative changes.

Thus, an argument can be made for the J segments of these heavy chains being the predominant idiotypic determinant in the monoclonal rheumatoid factor system. First, these areas

are quite similar between Wol and Sie and between Lay and Pom. Second, there are *differences* in these segments between the members of these two idiotypic sets that could account for their disparate antigenic reactivities. These areas of the heavy and light chains are brought into proximity in the native conformation of the immunoglobulin molecule, according to the published crystallographic structure (Poljak et al., 1973). They are also on the surface of the molecule, constituting an accessible and potent pattern. Taken together, these data suggest that the J regions are involved in the idiotypic classification of this system. Here, as in the light chains (Andrews & Capra, 1981), a potential idiotypic determinant is generated at a point in the polypeptide chain that is apparently designed to produce extreme variability.

The network theory of immune regulation (Jerne, 1974) requires that each recognition unit (whether on a B or T cell) express a unique set of determinants, called the idiotype. This set may be composed of one or more subsets, called idiotopes. The monoclonal human rheumatoid factors provide an example of two different sets of determinants that are both expressions of idiotypy. In the Po group, it is tempting to regard the CDRs of both the heavy and light chains as the predominant idiotope. This is clearly not the case in the Wa group, where the framework segments of the light chain and/or the J segments of the component chains are likely candidates. Further structural analysis on proteins from both these groups should clarify the issue of idiotypy in this system.

Acknowledgments

We acknowledge with pleasure the skilled technical assistance of Donna Atherton, Peggy Frank, Julie Winnette, and Martha Danhof. Fruitful discussions with Dr. L. Edward Cannon are gratefully acknowledged. We thank Kathy Able for typing this manuscript.

References

Agnello, V., Arbetter, A., Ibanez de Kasep, G., Powell, R., Tan, E. M., & Joslin, F. (1980) *J. Exp. Med. 151*, 1514–1527.

Andrews, D. W., & Capra, J. D. (1981) *Biochemistry* (preceding paper in this issue).

Bernard, O., & Gough, N. M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3630–3634.

Bernard, O., Hozumi, N., & Tonegawa, S. (1978) *Cell (Cambridge, Mass.) 15*, 1133–1144.

Capra, J. D., & Kehoe, J. M. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4032–4036.

Cunningham, B. A., Rutishauser, U., Gall, W. E., Gottlieb, P. D., Waxdal, M. J., & Edelman, G. M. (1970) *Biochemistry 9*, 3161–3170.

Early, P., Huans, H., Davis, M., Calame, K., & Hood, L. (1980) *Cell (Cambridge, Mass.) 19*, 981–992.

Florent, G., Lehman, D., & Putnam, F. W. (1974) *Biochemistry 13*, 2482–2498.

Gally, J. A., & Edelman, G. M. (1972) *Annu. Rev. Genet. 6*, 1–46.

Jerne, N. (1974) *Ann. Immunol. (Paris) 125C*, 373–387.

Kabat, E., Wu, T. T., & Bilofsky, H. (1979) *Immunoglobulin Sequences*, U.S. Department of Health, Education and Welfare, U.S. Public Health Service, National Institutes of Health, Bethesda, MD.

Klapper, D. G., & Capra, J. D. (1976) *Ann. Immunol. (Paris) 127C*, 261–271.

Kohler, H., Shimizu, A., Paul, C., Moore, V., & Putman, F. W. (1970) *Nature (London) 227*, 1318–1320.

Kunkel, H. G., Mannik, M., & Williams, R. C. (1963) *Science (Washington, D.C.) 140*, 1218–1219.

Kunkel, H. G., Agnello, V., Joslin, F. G., Winchester, R. J., & Capra, J. D. (1973) *J. Exp. Med. 137*, 331–342.

Kunkel, H. G., Winchester, R. J., Joslin, F. G., & Capra, J. D. (1974) *J. Exp. Med. 139*, 128–136.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Lehman, D. W., & Putnam, F. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3239–3243.

Matthyssens, G., & Rabbitts, T. H. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6561–6565.

Max, E. E., Seidman, J. G., & Leder, P. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3450–3454.

Nossal, G. J. V., Szenberg, A., Ada, G. L., & Austin, C. M. (1964) *J. Exp. Med. 119*, 485–502.

Padlan, E. A. (1977) *Q. Rev. Biophys. 10*, 35–65.

Podell, D. N., & Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun. 81*, 176–185.

Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P., & Saul, F. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3305–3311.

Rao, D. N., Rudikoff, S., Krutzsch, H., & Potter, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2890–2899.

Sakano, H., Huppi, K., Heinrich, G., & Tonegawa, S. (1979) *Nature (London) 280*, 288–294.

Schilling, J., Clevinger, B., Davie, J. M., & Hood, L. (1980) *Nature (London) 283*, 35–40.

Seidman, J. G., Leder, A., Nau, M., Norman, B., & Leder, P. (1978) *Science (Washington, D.C.) 202*, 11–17.