COMPLETE AMINO ACID SEQUENCES OF VARIABLE REGIONS OF TWO HUMAN IgM RHEUMATOID FACTORS, BOR AND KAS OF THE Wa IDIOTYPIC FAMILY, REVEAL RESTRICTED USE OF HEAVY AND LIGHT CHAIN VARIABLE AND JOINING REGION GENE SEGMENTS

By MARIANNA M. NEWKIRK,* RIZGAR A. MAGEED,‡ ROY JEFFERIS,‡ POJEN P. CHEN,§ AND J. DONALD CAPRA*

*From the *Department of Microbiology, the University of Texas Health Science Center, Dallas, Texas 75235; the ‡Department of Immunology, the University of Birmingham, Birmingham, B15 2TJ, England; and the §Department of Basic and Clinical Research, Scripps Clinic and Research Foundation, La Jolla, California 92037*

Human monoclonal rheumatoid factors (RFs)[1] can be divided by polyclonal antisera into three idiotypic families Wa, Po, and Bla (1–3). The predominant group is the Wa family, representing ~60% of the RFs isolated from the plasma of patients with mixed cryoglobulinemia (2). Also, patients with rheumatoid arthritis and other autoimmune diseases often have polyclonal antibodies that bear these idiotypes (4, 5). A number of investigators have analyzed the molecular nature of the Wa idiotype with major emphasis on the light chain variable region (6–9). The highly homologous amino acid sequences of several of these light chains as well as studies with antipeptide sera have suggested that there may be only a single VK gene segment used in the construction of these molecules. The VK gene 321 has been cloned and sequenced (10), and is thought to represent the germline VK gene segment that gives rise to the light chains of this family of autoantibodies. Whereas among Wa family members approximately fifteen light chain variable regions have been completely or nearly completely sequenced (8, 11, 12), before this study only two other Wa heavy chain variable regions had been sequenced (13).

Several studies have documented that RFs express several shared public idiotopes (14–16). However, because there was little or no sequence information available on the molecules used in such studies, an interpretation of the critical amino acids involved in the idiotypic determinants was not possible.

This paper presents the amino acid sequences of both the heavy and light chain variable regions of two additional members of the Wa family, BOR and KAS. In addition to the sequence information, the antigen binding specificities and expressed idiotopes of these and some previously sequenced Wa members

[1] *Abbreviations used in this paper:* CDR, complementarity-determining region; PTH, phenylthiohydantoin; RF, rheumatoid factor; TFA, trifluoroacetic acid.

has been determined. The data suggest that Wa family heavy chain variable regions appear to be restricted to the $V_H I$ subgroup and have a uniform-sized D region such that the third complementarity-determining region (CDR) of the heavy chain is about the same size. In addition, all Wa RFs sequenced to date use a $J_H 4$ gene segment. This striking restriction in V_H, D_H, and J_H structures suggests that these elements, along with the previously described light chain structures, may play a major role in determining the antigenic specificity and idiotypy of these molecules.

Materials and Methods

Preparation of RFs. Monoclonal IgM RFs were isolated from the washed cryoprecipitates (with cold PBS) from the plasma of patients BOR and KAS who had mixed cryoglobulinemia (the plasma was kindly donated by R. Wistar). Monoclonal IgM was separated from polyclonal IgG by gel filtration using an S300 column (170 × 2.5 cm, Pharmacia Fine Chemicals, Piscataway, NJ) in PBS, at room temperature. The IgM was >95% pure as analyzed by SDS-PAGE.

Preparation of Fabs of IgM RFs. Monoclonal IgM RFs were digested in the presence of cysteine (Sigma Chemical Co., St. Louis, MO) with mercuric papain (Worthington Biochemicals, Malvern, PA) at 1:100 for 16 h at room temperature, as previously described (17). The Fabs were first purified by gel filtration and then by ion-exchange chromatography (17).

Preparation of Fragments for Protein Sequencing. The purified Fab fragments were first chemically cleaved with an equal weight of CNBr in 70% formic acid for 24 h at room temperature. The mixture was diluted with water and then lyophilized. After CNBr cleavage, the peptides were resuspended in 0.2% trifluoroacetic acid (TFA), and small cleavage products were separated from large ones by reverse-phase chromatography using a C4 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% TFA in water to 0.1% TFA in 50% acetonitrile, 15% isopropanol for 90 min at a flow rate of 1 ml/min. The large fragments (which eluted at the end of the gradient) were pooled and subjected to complete reduction and alkylation in guanidine (CRA) as previously described (12). These peptides were separated by reverse-phase chromatography using the conditions described above, however, initially the C4 column was used to remove the salts used in the complete reduction and alkylation procedure. To optimally separate the peptides, we used reverse-phase chromatography with a C18 column (Aquapore RP300; Brownlee, Santa Clara, CA) with a linear gradient of 0.1% TFA in water to 0.1% TFA in 50% acetonitrile for 90 min at a flow rate of 1 ml/min. Peptides that were ⩾35 amino acids in length were further digested with trypsin either with or without citraconylation as the case required (12). These peptides were separated by reverse-phase chromatography using the C18 column as above.

Removal of NH_2-terminal Pyroglutamic Acid from Heavy Chain. Initial studies proved that the NH_2 terminus of the heavy chain was resistent to Edman degradation. Thus, the heavy chain NH_2-terminal peptide (after chemical cleavage with CNBr) was identified by compositional analysis using a Beckman 6300 amino acid analyzer (Beckman Instruments, Palo Alto, CA). Using pyroglutamic amino peptidase according to the method of Podill et al. (18), the cyclized NH_2-terminal amino acid was successfully cleaved. This CNBr peptide, after the aminopeptidase treatment, was separated from the intact enzyme by reverse-phase chromatography using the C18 column as above.

Protein Sequencing. Peptides were subjected to an automated protein sequencer using either the gas-phase model 470 Applied Biosystems sequencer (Foster City, CA) with the model 120 on-line HPLC phenylthiohydantoin (PTH) amino acid identification system, or a Beckman spinning cup model 890M sequencer. In the latter case, the PTH amino acids were identified using a NovaPac column in a Waters model 840 HPLC system (Millford, MA). To help identify the peptides before sequencing, compositional analyses of the peptides were done by amino acid analysis.

Idiotype and Idiotope Analysis.   The purified RFs were analyzed for expression of the Wa idiotype using three different polyclonal typing antisera raised to Wa RFs by hemagglutination (kindly performed by D. Posnett, Cornell University Medical Center, New York [16], and V. Agnello, Lahey Clinic, Burlington, MA [19]), and by ELISA. In the latter case, polystyrene plates were coated with BOR or KAS (10 μg/ml) in PBS by incubation at 37°C for 2 h, and dilutions of F(ab')$_2$ of polyclonal antisera in PBS/0.05% Tween 20 were added to the wells. Bound antisera were detected using peroxidase-conjugated sheep anti-rabbit Ig antisera. Analyses of expressed idiotopes on these molecules were performed using an ELISA with a panel of monoclonal antibodies that had been raised to a number of human mRFs. For the monoclonal antibodies G6, H1 (14), C7, B12 (14) G8, C8 (R. A. Mageed, manuscript in preparation) and 17-109 (20) (kindly donated by D. Carson, Scripps Clinic and Research Foundation, La Jolla, CA), the ELISA plates, coated as above, were incubated with monoclonal antibodies diluted (1:10^{-4}) in PBS/Tween. The bound mAbs were detected with peroxidase-conjugated sheep anti-mouse Ig antisera. For the monoclonal antibodies 86.3, 108.12, 102.2, and 128.4 (kindly donated by D. Posnett [16]) the ELISA plates were coated with BOR or KAS (10 μg/ml) in carbonate/bicarbonate buffer, pH 9.6, for 16 h at 4°C. The monoclonal antibodies that had been purified by protein A affinity chromatography (Pierce Chemical Co., Rockford, IL) were added at a concentration of 1 μg/ml in PBS and incubated for 2 h at 37°C. Bound antibodies were detected using peroxidase-conjugated goat anti-mouse Ig antisera.

Reactivity with Antisera Raised to SIE CDR Peptides.   The heavy and light chains from BOR and KAS RFs were separated on SDS-PAGE and immunoblotted with antisera to the peptides PSL2, PSL3, and PSH3 derived from the RF SIE and PPH2, and PPH3 derived from the RF POM as previously described (12, 21–23).

Antigenic Specificity of the mRFs.   The antigenic specificities of the mRFs for different subclasses of IgG were determined independently by ELISA and hemagglutination. ELISA plates were coated with myeloma proteins of the appropriate subclass of IgG (10 μg/ml) in carbonate/bicarbonate buffer, pH 9.6, for 16 h at 4°C. The RFs, either at 1 or 0.5 μg/ml (diluted in PBS) were incubated at room temperature for 4 h. Bound RFs were detected with peroxidase-conjugated goat anti-human IgM (Fc-specific) antisera. For the hemagglutination assay, sheep red blood cells were passively sensitized with the Ig (isotype or subclass) in question using chromic chloride. The starting concentration of RF was 110 μg/ml (BOR), 83 μg/ml (KAS), or 100 μg/ml (SIE).

Results

Expressed Idiotype and Idiotope.   The results of assays determining the nature of the idiotype and idiotopes expressed by BOR, KAS, SIE, and WOL are presented in Table I. As can be seen, both BOR and KAS are representatives of the Wa idiotypic family of human RFs. SIE and WOL had previously been determined to be members of this idiotypic family (11). From the results of ELISA with monoclonal antibodies that are thought to recognize determinants specific for human RFs it can be seen that BOR, KAS, and SIE show similar reactivities. The slight differences in the degree of the reactivity of these RFs with the antiidiotopes may either be due to the degree of purity of the RFs (the BOR and SIE preparations were >98% pure while the KAS preparation was ~95% pure), or the degree of denaturation of the IgM (the SIE RF preparation was >7 yr old). Based on the reactivity of these monoclonal antibodies with a number of other RFs (as yet unsequenced), BOR, KAS, and SIE appear to belong to a Wa subfamily. However, an epitope was detected on BOR and SIE by mAb G8 that was not detected on the KAS molecule, suggesting that the Wa family can be further subdivided.

Antigenic Specificity.   The subclasses of human IgG to which BOR, KAS, and

TABLE I

Expression of Idiotype/Idiotopes by Completely Sequenced Human RFs

Antisera/ antibody	Specificity	Idiotype/idiotope on RF			
		BOR	KAS	SIE	WOL*
Anti-GLO‡	Wa	++++	++++	ND	ND
Anti-McD‡	Wa	++++	++++	ND	ND
Anti-KO§	Wa	++++	++++	++++	ND
17-109§	Wa LC	+++	++	++	ND
H1	Wa HC (two RFs)	−	−	−	ND
B12	KIII	++++	+++	+++	ND
C7	KIII	++++	+++	+++	ND
G8	Conformational epitope	+++	−	+++	ND
G6	Wa HC	+++	++	++	ND
C8	IgML-RF	−	−	−	ND
86.3	RF-CRI	+	+	+	−
108.12	KIIIb	++	+	+	++++
102.2	KIIIb	+++	++	+	++
128.4	IgMK	+	+	−	−

Results are expressed as: −, negative; +, OD_{492} 0.05–0.5; ++, 0.51–1.0; +++, 1.01–1.5; ++++, 1.51–2.0.
* RIA results (16).
‡ Hemagglutination assay (12, 19).
§ ELISA.

TABLE II

Specificity of BOR, KAS, and SIE RFs for IgG Subclasses

Isotype/subclass of Ig	Human RF					
	BOR		KAS		SIE	
	ELISA	HA titer	ELISA	HA titer	ELISA	HA titer
IgG pool	ND	6	ND	8	ND	11
Fc(IgG1)	+++	11	+++	10	+++	22
IgG1	+++	5	+++	5	+++	9
IgG2	+	0	+++	2	++	0
IgG3	−	1	++++	4	++	0
IgG4	−	1	++	3	+	0
BJ	ND	0	ND	0	ND	0
IgM	ND	0	ND	0	ND	0
IgA	ND	0	ND	0	−	0

For ELISA, results are expressed as: −, negative; +, OD_{492} 0.05–0.5; ++, 0.51–1.0; +++, 1.01–1.5; ++++, 1.51–2.0. Hemagglutination (HA) titer log 2.

SIE bind are presented in Table II. All three RFs show similar binding to IgG1 as well as to Fc (IgG1). However, BOR binds only weakly if at all to the other human IgG subclasses, while KAS binds well to IgG3. SIE recognizes the IgG subclasses 1–3, and weakly 4, but binds best to IgG1. None of these RFs bound to either IgM, IgA, or light chain dimers of either κ or λ isotype. Slight discrepancies in the binding profiles were obtained by the two methods, possibly

```
        PSL2  PSL3                                        CDR I
                                      20                              40
BOR     ++    +      EIVLTQSPGT LSLSPGERAT LSCRASQSVSS SYLAWYQQKP
KAS     +     ++     D--------- ---------- ---C-----L-- T---------
SIE     ++    ++     ---------- ---------- ---C-----N --C-------
WOL     ++    -      ---------- ---------- ---C------- G--GC-----


                        CDR II
                                      60                      80
BOR     GQAPRLLIYG ASSRATGIPD RFSGSGSGTD FTLTISRLEP EDFVVYYC
KAS     ---------C ------CV-- ---------- ---------- ---A----
SIE     ---------C ------C--- ---------- ---------- D--A----
WOL     ---------C ------C--- ---------- ---------- ---A----
321     ---------C ------C--- ---------- ---------- ---A----


                    CDR III
                              100
BOR     QQ YGNSPQTFGQ GTKVEIKR      J1
KAS     -- --S--F--G ---------      J4
SIE     -- --S----- -S------       J1
WOL     -- --SLGR--- --------       J1
321     -- --S--
```
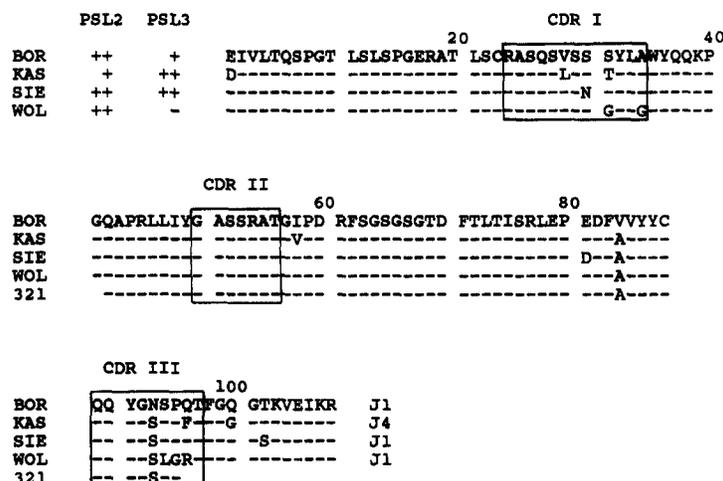
FIGURE 1. The complete amino acid sequences of the light chain variable regions of four monoclonal RFs of the Wa idiotypic family, along with the genomic KIIIb sequence (321). Amino acids are given in the single-letter code, numbered according to Kabat (24). Identical amino acids are represented by dashes, changes in amino acids are indicated. Boxed areas delineate the CDRs (CDR I–III). SIE and WOL sequences (11), 321 sequence (10). Reactivity to the antipeptide sera from immunoblots PSL2 refers to the peptide derived from SIE positions 49–61, PSL3 refers to the peptide from SIE positions 88–99. The results from immunoblots with the antipeptide sera for WOL as previously reported (9, 25).

due to the different myeloma proteins (IgG) that were used in different laboratories.

Amino Acid Sequences of Light Chain Variable Regions of BOR and KAS. The sequences derived from tryptic peptides of the variable regions of the light chains of BOR and KAS are presented in Fig. 1. NH_2-terminal sequence analysis on the first 38 amino acids of the intact BOR light chain and 35 amino acids for the KAS light chain aided in the alignment of the peptides. The remaining peptides were positioned according to the sequences of related light chain variable regions of the KIIIb subgroup. For comparison, the germline VK 321 sequence of Chen et al. (10) and two other Wa family RFs previously sequenced in our laboratory are presented (11).

Amino Acid Sequence of the Heavy Chain Variable Regions of BOR and KAS. The sequences derived from peptides of the variable regions of the heavy chains of BOR and KAS are presented in Fig. 2. For comparison, the other Wa RFs, SIE and WOL, (13), EU (not an RF) (27), and one representative of the translated DNA sequence of a rearranged gene (J. V. Ravetch and S. Korsmeyer, personal communication), all members of the human V_HI subgroup, are presented. While there are only five complete protein sequences in this subgroup, they can be easily subdivided into subsubgroups with BOR, KAS, and EU members of one subsubgroup designated V_HIa, all having >85% amino acid identity (Table III, positions 1–94), and SIE and WOL representatives of other subsubgroups of V_HI termed V_HIb, c, etc. Other sequences derived from DNA (both germline and rearranged) of members of this family (28–29) show a diminishing gradation of relatedness to the V_HIa subsubgroup (Table III).
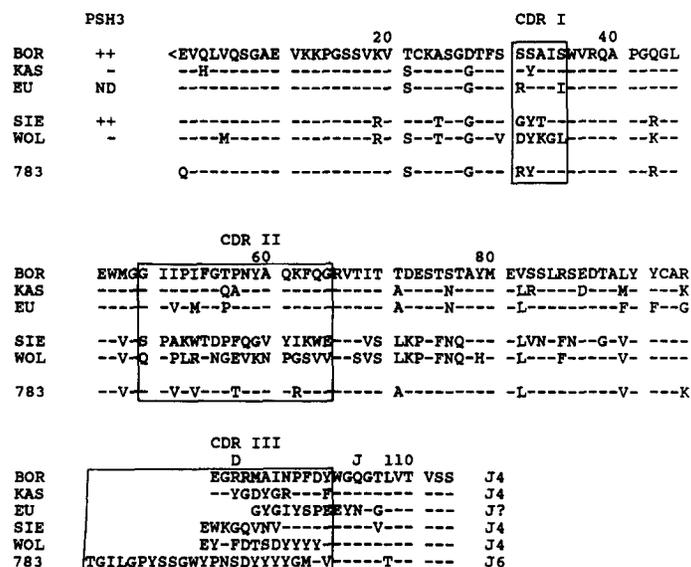
```
         PSH3                                          CDR I
                                20                            40
BOR      ++     <EVQLVQSGAE VKKPGSSVKV TCKASGDTFS  SSAISWVRQA PGQGL
KAS      -         --H------- ---------- S-----G---  -Y---f---- -----
EU       ND        ---------- ---------- S-----G---  R---I----- -----

SIE      ++        ---------- --------R- ---T--G---  GYT--f---- --R--
WOL      -         ----M----- --------R- S--T--G--V  DYKGL----- --K--

783             Q--------- ---------- S-----G---  RY---f---- --R--


                     CDR II
                        60                    80
BOR      EWMGG IIPIFGTPNYA QKFQGRVTIT TDESTSTAYM EVSSLRSEDTALY YCAR
KAS      ---f- ------QA--- ----f---- A----N---- -LR----D---M- ---K
EU       ---f- -V-M--P--- ----f---- A----N---- -L---------F- F--G

SIE      --V-S PAKWTDPFQGV YIKWB---VS LKP-FNQ--- -LVN-FN--G-V- ----
WOL      --V-Q -PLR-NGEVKN PGSVV--SVS LKP-FNQ-H- -L---F-----V- ----

783      --V-f -V-V---T--- -R--f---- A--------- -L---------V- ---K


                  CDR III
                  D            J  110
BOR      EGRRMAINPFDYWGQGTLVT VSS    J4
KAS         --YGDYGR---F-------- ---    J4
EU           GYGIYSPEEYN-G--- ---    J?
SIE         EWKGQVNV----f----V--- ---    J4
WOL         EY-FDTSDYYYY-------- ---    J4
783      TGILGPYSSGWYPNSDYYYYGM-V-----T-- ---    J6
```

FIGURE 2. The complete amino acid sequences of the heavy chain variable regions of four Wa family RFs BOR, KAS, SIE, and WOL, with sequences of other V$_H$I members, EU (protein, not an RF) and 783 (from a rearranged DNA sequence, specificity unknown) for comparison. Identical amino acids are represented by dashes, changes in amino acids are indicated. Boxed regions delineate the CDRs. EU (27), SIE, WOL (13), 783 (Ravetch and Korsmeyer, personal communication). The reactivity to anti-PSH3 sera from immunoblots. PSH3 refers to the SIE positions 95–102. WOL reactivity was previously reported (25).

TABLE III

Amino Acid Identity in V$_H$I Members (Residues 1–94)

	BOR	KAS	783	EU	HG3	266BL	SIE	WOL
BOR	100	86	86	87	78	67	60	60
KAS		100	86	87	76	68	58	61
783			100	88	77	71	62	61
EU				100	78	66	57	60
HG3					100	73	57	60
266BL						100	56	57
SIE							100	63
WOL								100

Residues are numbered according to Kabat (24); 783 (Ravetch and Korsmeyer, personal communication), EU (27), HG3 (28), 266BL (29), SIE, WOL (13). Numbers are percent homology. The V$_H$Ia subsubgroup has been boxed.

Reactivity with Antipeptide Sera. The results of studies determining the reactivity with BOR and KAS of the antipeptide sera raised to peptides derived from the heavy and light chain CDRs of SIE are presented in Figs. 1 and 2. Both BOR and KAS show reactivity with the light chain–specific peptide antisera (Fig. 1) that bind to the second or third CDRs of SIE (PSL2 and PSL3). The anti-PSL2 sera bound to the BOR light chain to a greater extent than to the KAS light chain, while the opposite pattern of reactivity was observed with the anti-PSL3 sera that bound best to the KAS light chain. The BOR but not the KAS heavy chain was recognized by the antipeptide sera that is directed to the CDR3 of the

SIE heavy chain (PSH3) (Fig. 2). This is the first time that any RF heavy chain other than SIE has been found to significantly react with this antisera (23, 25). There was no detectable binding of the antipeptide sera raised to the POM peptides (PPH2 and PPH3) to any of these RFs (results not shown).

Discussion

This study provides evidence that human RFs of the Wa idiotypic family have a conserved primary structure, suggesting that they may be derived from relatively few germline V_H, D_H, J_H, V_K, and J_K gene elements. Several lines of evidence had suggested previously that the variable region of the light chain of all the Wa family members appears to be derived from a single VKIIIb gene segment (10). As can be seen from Fig. 1, BOR and KAS V_L amino acid sequences are 93% identical, and the other RFs are >92% identical (positions 1–96). Although the mutation rate has not clearly been established for human V regions, it is probably similar to mouse V regions, which is estimated to be 10^{-3} mutations per cell base cell division (approximately one mutation in the coding region per three generations) (30). Considering that these RFs are from outbred, unrelated humans, and that the number of B cell generations is not known at the time the sera were harvested, it is interesting to note that the light chain sequences of SIE, BOR, KAS, and WOL (1–96) can be explained by a minimum of two to four nucleotide changes (silent changes are unknown) from the proposed ancestral KIIIb 321 gene.

Previously only J_K1 and J_K2 gene segments were seen in conjunction with the V_KIIIb region at a ratio of 3:2 in the Wa family. In studies of mouse RFs, a similar finding of high frequency usage of the proximal JK segments (J1 and 2) has been found (31). The KAS RF is the first described of this Wa family to use a J_K4 gene segment. BOR uses the more commonly used J_K1 gene segment.

The junctional amino acid (at position 96) between V_K and J_K has been shown in a number of systems to be highly variable and important for antigen binding. In the 13 completely sequenced light chains or Wa RFs (8, 11, 12) six different amino acids have been seen at this position, with Arg present most frequently (31%). The Gln at the VJ junction in BOR was previously seen in the RF SIE. While a number of genetic mechanisms can give rise to this amino acid, with the knowledge of the nucleotide sequence of the VKIIIb 321 germline gene that presumably gives rise to the light chains of these RFs (10), and the genomic sequence of the human J_K1 gene segment (26), it is likely that the junctional amino acid Gln in BOR and SIE is encoded by CAG, with the 3' end of the V gene segment contributing the C, an A added by one of the diversity mechanisms, and the last nucleotide of the codon coming from the G from the 5' end of the germline J_K1 Trp codon. In KAS, the junctional Phe codon is most likely to be encoded by TTC and can be explained by adding a T (possibly by a terminal transferase) after the Pro codon (CCT) in the germline V region, and then by intracodonic mechanisms that join the TC from the Leu codon (CTC) from the 5' end of the J_K4 germline gene element. This is the first time that a Phe has been seen at this highly variable junctional position.

Studies with antipeptide sera raised to SIE light chain CDRs reveal that the BOR light chain is recognized by the antipeptide sera to PSL3 even though there

is a substitution of Asn for Ser at position 93 in the light chain, whereas WOL, which has the Ser but multiple differences after the Ser, is not (25). From the combined results of sequence information and the reactivity of this antisera to a number of Wa RFs (8, 12) is was previously deduced that the region of the light chain recognized by the anti-PSL3 sera was amino acids 88–93. It had been found that a substitution of an Ala at position 93 reduced the reactivity of the antisera to PSL3 by 50% (8), and it appears that the Asn substitution also results in a similarly reduced binding. It is difficult to explain the lack of reactivity of the anti-PSL3 sera to WOL given these data. The KAS light chain is the first Wa light chain sequenced to have a substitution at position 59 (framework 3) within the region to which the antisera to the peptide PSL2 was raised. There appears to be a diminished reactivity of the anti-PSL2 sera to the KAS light chain, probably due to the Val for Ile substitution.

The similarities of the heavy chain V regions of the four Wa members sequenced to date show that BOR and KAS are more related to one another than are SIE and WOL (Fig. 2). In the mouse, families of heavy chain V_H gene segments have been defined as sequences that have >80% identity at the DNA level. In man, the subdivision of heavy chain V region proteins has been based on >75% identity in the framework regions. BOR, KAS, SIE, and WOL show ≥75% identity, and thus by definition are members of the V_HI subgroup. However, when the heavy chain V regions (positions 1–94) are compared, it can be seen that the second CDR varies considerably between BOR/KAS and SIE/WOL. The lack of amino acid identity in this region cannot be explained by translation in another reading frame when the DNA sequence of the related V_HI 783 rearranged gene is used. We propose a new designation, V_HIa, to include proteins that have >85% amino acid identity to EU in the V region (positions 1–94), and >75% amino acid identity in CDRII. BOR and KAS by this definition would be members of this sub-subgroup, excluding SIE and WOL. In the V_HIII subgroup of human immunoglobulin heavy chains, where there is considerably more amino acid sequence data available, the second CDR also appears to be a region that is highly variable (24) and has been seen to have a shifted reading frame in this region (32), and, in the extreme, to have been deleted (33). Perhaps, in man, a number of diversity mechanisms are active in this region. However, it is notable that in the Po idiotypic family of RFs, the CDRII of the V_H as well as the entire V_H are relatively invariant for the proteins sequenced to date (34), and may be important in the expression of the Po idiotype.

As only a limited number of complete sequences of V_HI subgroup members exist (five protein, three DNA) it would appear that the V_HI subgroup is used less frequently than the other known V_H subgroups. Interestingly, half of these known sequences are from Wa family RFs. However, since all of the proteins (in the V_HI subgroup) sequenced to date have cyclized NH_2 termini, thus rendering the protein resistant to Edman degradation, the limited number of protein sequences available could be due to an investigator bias. Serological studies have indicated that human V_HI regions are present in 16% of expressed immunoglobulins (V_HII, 34%; V_HIII 50%) (35). The V_HIII subgroup has been found by sequence analysis to be present on 20–75% of polyclonal normal Ig, depending

558   STRUCTURAL ANALYSIS OF HUMAN RHEUMATOID FACTORS

```
                V              D                    J
                92      95                  99            J            113
Genomic J-4                                 Y F D Y W G Q G T L V T V S S
BOR             C A R   E () G R R M A I N P  - - - - - - - - - - - - - - -
KAS             - - K   - - - Y G D Y G R -    - - F - - - - - - - - - - -
SIE             - - -   - W K G Q V N V - -    - - - - - - - V - - - - - -
WOL             - - -   - Y - F D T S D Y    - Y Y - - - - - - - - - - -
```

FIGURE 3.   The amino acid sequences of the heavy chain D and J regions of four Wa family human RFs with the genomic sequence of the J_H4 (45) for comparison.

on the isotype (36). From studies using V_H DNA probes in cosmid clones, the ratio of $V_H I/II/III$ appears to be 23:8:30 (37).

The organization of V region genes in the mouse appears to be in discrete families (38, 39), although some exceptions do exist (40). In human V_K and rabbit V_H, however, it appears that V region genes are not organized into discrete families but are interdigitated (41, 42). Some human κ pseudogenes, three of which have been sequenced (those contain major abnormalities that would prevent processing) are found on chromosomes 1, 15, and 22 (43), while the transcribable κ genes are encoded on chromosome 2. Similarly, in the human V_H gene organization there is considerable interdigitation of the three subgroups (37).

From our study it appears that the BOR and KAS V_H may be derived from one $V_H Ia$ gene (each would have a minimum of 10 mutations from a consensus $V_H Ia$ sequence) while SIE and WOL are likely to be single representatives of other $V_H I$ genes. Thus, a minimum of three genes (all restricted to the $V_H I$ subgroup) could encode RFs of the Wa idiotypic group. The other major idiotypic family, Po (2, 34) seems to be equally restricted, but to the $V_H III$ subgroup. We have recently sequenced two additional members of this latter family (an IgM and an IgG RF) that also have a $V_H III$ heavy chain, which will be the subject of another communication (44).

It appears that, although at least three related genes can encode the heavy chain V gene segments of Wa RFs, the critical features that determine both the antigenic specificity for the Fc portion of IgG (one or more subclasses) and the idiotypic expression are the nature of the heavy chain D and J gene segments (Fig. 3). All four of these Wa family members have a D region that is 9 or 10 amino acids long (9 is the average size of expressed human D gene segments). However, this very restricted limit in size (the variation seen in other human immunoglobulins of all families is 5–16) is probably meaningful in the context of the nature of the binding pocket. A similar conservation in the size of the CDRIII in the heavy chain of mouse RFs has also been observed (31). The other members of this human V_H subgroup have a greater variation in the number of amino acids in this region of the antibody molecule. The D regions of these RFs do not match any published sequence of human D regions (45, 46), however the KAS and EU D segments appear to be highly related. Interestingly, all of the Wa RFs have D segments that begin with a Glu (Fig. 3), and if the genomic sequence of these genes is similar to HG3 (Table III), this can be explained by adding either A or G to the existing GA (3' of the Arg codon in the germline) thus creating the triplet codon for Glu. This A or G could be supplied by the D gene segment or by another N segment addition mechanism. Three of the four RFs have D segments that end (at the junction of DJ) with Pro. An explanation

of this requires knowledge of the D gene segments at the DNA level as it is likely that the J gene segment contributes only the third nucleotide (which can be any base in the case of Pro). These Glu and Pro residues may be important in antigen specificity and/or the expression of idiotypy, as they occur more frequently than random chance would predict.

Perhaps the most striking structural feature of the Wa family RFs is that they all use the J_H4 gene segment (Fig. 3), confirming the previous observations of Andrews and Capra (47) and adding further support for their argument that this region may be crucial for the Wa crossreactive idiotype. Even though all four members of this family use the J_H4 gene segment, there are slight variations among the individuals. It can be seen that WOL likely uses a longer portion of this J_H gene element (the Phe to Tyr and the Asp to Tyr substitutions in WOL are both one-nucleotide changes) whereas BOR, KAS, and SIE all use the same size of J_H. Also, KAS has a Phe instead of Tyr in the J_H4 (a one-nucleotide change), that, coupled with the charged group (Arg) in the D section probably create a stretch of amino acids that is not recognized by the antipeptide sera that binds to the SIE and BOR CDRIIIs. Thus the common structural features of the Wa family of RFs are a heavy chain variable region consisting of a V_HI subgroup gene segment, a 9- and 10-amino-acid D region beginning with Glu and ending with Pro (in three of the four), and a J_H4 gene segment.

Although BOR and KAS as well as SIE and WOL are all members of the Wa idiotypic family, BOR and KAS are much more related to one another in the V_H region than to SIE or WOL. The data from studies using the panel of monoclonal antiidiotopes also support a close relationship shared by all three RFs (BOR, KAS, and SIE). Studies with the WOL molecule are incomplete due to the present unavailability of the molecule. It is likely that the conformational idiotope recognized by the mAb G8 is in the region of amino acids 94–100 of the light chain, with the junctional Gln pivotal, coupled with the stretch of amino acids in the heavy chain CDRIII that BOR and SIE share. Studies using anti-PSH3 sera confirm the latter, as SIE and BOR show common reactivity to this antisera, which detects CDRIII of the SIE heavy chain. When the amino acid sequences of SIE and BOR are compared in this region, there is a stretch of only five identical amino acids (Asn-Pro-Phe-Asp-Tyr) at the COOH terminus of this peptide that probably is responsible for the crossreactivity of the antipeptide sera. KAS neither shares the idiotope recognized by the mAb G8 nor is reactive with the anti-PSH3 sera, and has major amino acid differences from SIE and WOL in the CDRIIIs of both heavy and light chains.

There probably exists within the Wa idiotypic family a number of subfamilies defined by the fine specificity of antigen binding and expression of idiotopes. Until further sequence information on other Wa members becomes available one can not as yet determine their number and size. Also the exact binding site on the Fc portion of the IgG molecule has long been in doubt. It is likely that different members of this family bind different parts of Fcγ. It has been speculated that the binding site for polyclonal RFs is in the cleft between C$\gamma2$ and C$\gamma3$, as it is inhibitable by protein A (48). However, since KAS but not BOR binds IgG3, a subclass that does not bind protein A (49), a different portion of the Fc molecule may be recognized by this RF although this does not preclude

the binding site being in the $C\gamma2/C\gamma3$ region. The major differences in the KAS and BOR/SIE amino acid sequences are in the CDRIIIs of the heavy and light chains. These areas are likely candidates that could confer a difference in the antigen-binding specificity of the KAS molecule. An IgM monoclonal RF (not of the Wa idiotypic family) has been shown (50) to bind to a discontinuous epitope on the Fc of IgG, but is dependent on the presence of the His at position 435 (near the COOH terminal of the $C\gamma3$ domain).

A number of investigators have postulated that the light chain (possibly framework regions) is the common element in determining an RF, and while this may be true in the mouse (32) it is unlikely to be so in humans. Another human Ig (EV1-15) sequenced recently in our laboratory, which uses a KIIIb light chain, neither has RF activity nor expresses the Wa idiotype. The Wa idiotypic RFs appear to have in common the KIIIb light chain, a heavy chain V_HI, a D region that is uniform in size beginning with a Glu and ending with a Pro, and a J_H4. Evidence from three-dimensional studies on these molecules should help determine the nature of the antigenic binding site and what residues are critical for the expression of idiotypy. BOR and KAS have recently been crystallized and await further resolution of these issues (17).

It is interesting to compare the human Wa idiotypic family of human RFs to another well-characterized idiotypic family, that of the antiarsonate response in the A/J mouse. Extensive studies (51–53) have shown that these molecules use very specific heavy chain V-D-Js and light chain V-J to both bind antigen and express the major crossreactive idiotype (51–53). In fact, certain amino acids such as the Ser at position 99 in the heavy chain have been well characterized as to their essential role (in this case in antigen binding).

Thus, there is a growing consensus that RFs of both the Po and Wa idiotypic families are derived from only a limited number of germline gene segments. The presence of these idiotypes in both mixed cryoglobulinemia and in rheumatoid arthritis suggests either that there are common gene segments present in the majority of the population that are activated in these diseases by an as yet unknown mechanism, or that only certain individuals, those carrying these gene segments, are indeed susceptible to these diseases. The availability of genetic probes for these gene segments should facilitate an answer to these questions.

Summary

Evidence derived from the complete amino acid sequences of the variable regions of both the heavy and light chains of two members (BOR and KAS) of the Wa idiotypic family of human rheumatoid factors suggests that not only are the light chains of these molecules derived from possibly one variable region gene segment, but the heavy chain variable regions are all derived from the V_HI subgroup of human V region genes. These molecules exhibit a surprising conservation in the size of D region, and all use the J_H4 gene element. This restriction in use of V_L, V_H, D, and J_H suggests all of these elements may play a crucial role in either antigen binding and/or expression of the crossreactive idiotype.

Received for publication 28 January 1987 and in revised form 13 March 1987.

References

1. Kunkel, H. G., M. Mannik, and R. C. Williams. 1963. Individual antigenic specificity of isolated antibodies. *Science (Wash. DC).* 140:1218.
2. Kunkel, H. G., V. Agnello, F. G. Joslin, R. J. Winchester, and J. D. Capra. 1973. Cross-idiotypic specificity among monoclonal IgM proteins with anti-γ-globulin activity. *J. Exp. Med.* 137:331.
3. Agnello, V., A. Arbetter, G. I. de Kasep, R. Powell, E. M. Tan, and F. Joslin. 1980. Evidence for a subset of rheumatoid factors that cross-react with DNA-histone and have a distinct cross-idiotype. *J. Exp. Med.* 151:1514.
4. Forre, O., J. H. Dobloug, T. E. Michaelsen, and J. B. Natvig. 1979. Evidence of similar idiotypic determinants on different rheumatoid factor populations. *Scand. J. Immunol.* 9:281.
5. Gharavi, A. E., B. M. Patel, G. R. V. Hughes, and K. B. Elkon. 1985. Common IgA and IgG rheumatoid factor idiotypes in autoimmune diseases. *Ann. Rheum. Dis.* 44:144.
6. Kunkel, H. G., R. J. Winchester, F. G. Joslin, and J. D. Capra. 1974. Similarities in the light chains of anti-γ-globulins showing cross-idiotypic specificities. *J. Exp. Med.* 139:128.
7. Ledford, D. K., F. Goni, M. Pizzolato, E. C. Franklin, A. Solomon, and B. Frangione. 1983. Preferential association of KIIIb light chains with monoclonal human IgMK autoantibodies. *J. Immunol.* 131:1322.
8. Goni, F., P. P. Chen, B. Pons-Estel, D. A. Carson, and B. Frangione. 1985. Sequence similarities and cross-idiotypic specificity of L chains among human monoclonal IgMk with anti-γ-globulin activity. *J. Immunol.* 135:4073.
9. Chen, P. P., F. Goni, S. Fong, F. Jirik, J. H. Vaughan, B. Frangione, and D. A. Carson. 1985. The majority of human monoclonal IgM rheumatoid factors express a "primary structure-dependent" cross-reactive idiotype. *J. Immunol.* 134:3281.
10. Chen, P. P., K. Albrandt, N. K. Orida, V. Radoux, E. Y. Chen, R. Schrantz, F.-T. Liu, and D. A. Carson. 1986. Genetic basis for the cross-reactive idiotypes on the light chains of human IgM anti-IgG autoantibodies. *Proc. Natl. Acad. Sci. USA.* 83:8318.
11. Andrews, D. W., and J. D. Capra. 1981a. Amino acid sequence of the variable regions of light chains from two idiotypically cross-reactive human IgM anti-γ-globulins of the Wa group. *Biochemistry.* 20:5816.
12. Newkirk, M., P. P. Chen, D. Carson, D. Posnett, and J. D. Capra. 1986. Amino acid sequence of a light chain variable region of a human Rheumatoid Factor of the Wa idiotypic group, in part predicted by its reactivity with antipeptide antibodies. *Mol. Immunol.* 23:239.
13. Andrews, D. W., and J. D. Capra. 1981b. Amino acid sequence of the variable regions of heavy chains from two idiotypically cross-reactive human IgM anti-γ-globulins of the Wa group. *Biochemistry.* 20:5822.
14. Mageed, R. A., D. Dearlove, D. M. Goodall, and R. Jefferis. 1986. Immunogenic and antigenic epitopes of immunoglobulins XVII- Monoclonal antibodies reactive with common and restricted idiotopes to the heavy chain of human rheumatoid factors. *Rheumatol. Int.* 6:179.

15. Mageed, R. A., M. R. Walker, and R. Jefferis. 1986. Restricted light chain subgroup expression on human rheumatoid factor paraproteins determined by monoclonal antibodies. *Immunology.* 59:473.

16. Posnett, D. N., R. Wisniewolski, B. Pernis, and H. G. Kunkel. 1986. Dissection of the human antigammaglobulin idiotype system with monoclonal antibodies. *Scand. J. Immunol.* 23:169.

17. Newkirk, M. M., A. B. Edmundson, R. Wistar, and J. D. Capra. 1987. A new protocol to digest human IgM with papain results in homogeneous Fab preparations which can be routinely crystallized. *Hybridoma.* In press.

18. Podill, D. N., and G. N. Abraham. 1978. A technique for the removal of pyroglutamic acid from the amino terminus of proteins using calf liver pyroglutamate amino peptidase. *Biochem. Biophys. Res. Commun.* 81:176.

19. Agnello, V., and J. L. Barnes. 1986. Human rheumatoid factor crossidiotypes. I. WA and BLA are heat-labile conformational antigens requiring both heavy and light chains. *J. Exp. Med.* 164:1809.

20. Carson, D. A., and S. Fong. 1983. A common idiotope on human rheumatoid factors identified by a hybridoma antibody. *Mol. Immunol.* 20:1081.

21. Chen, P. P., R. A. Houghten, S. Fong, G. H. Rhodes, T. A. Gilberston, J. H. Vaughan, R. A. Lerner, and D. A. Carson. 1984. Antihypervariable region antibody induced by a defined peptide: an approach for studying the structural correlates of idiotypes. *Proc. Natl. Acad. Sci. USA.* 81:1784.

22. Chen, P. P., S. Fong, D. Normansell, R. A. Houghten, J. G. Karras, J. H. Vaughan, and D. A. Carson. 1984. Delineation of a cross-reactive idiotype of human autoantibodies with antibody against a synthetic peptide. *J. Exp. Med.* 159:1502.

23. Goldfien, R. D., P. P. Chen, S. Fong, and D. A. Carson. 1985. Synthetic peptides corresponding to third hypervariable region of human monoclonal IgM Rheumatoid Factor heavy chains define an immunodominant idiotype. *J. Exp. Med.* 162:756.

24. Kabat, E. A., T. T. Wu, H. Bilofsky, M. Reid-Miller, and H. Perry. 1983. Sequences of proteins of immunological interest. U. S. Government Printing Office, Bethesda, MD.

25. Chen, P. P., F. Goni, R. A. Houghten, S. Fong, R. Goldfien, J. H. Vaughan, B. Frangione, and D. A. Carson. 1985. Characterization of human rheumatoid factors with seven antiidiotypes induced by synthetic hypervariable region peptides. *J. Exp. Med.* 162:487.

26. Hieter, P. A., J. V. Maizel, Jr., and P. Leder. 1982. Evolution of human Immunoglobulin K J region genes. *J. Biol. Chem.* 257:1516.

27. Cunningham, B. A., U. Ruthishauser, W. E. Gall, P. D. Gottlieb, M. J. Waxdal, and G. M. Edelman. 1970. The covalent structure of a human γG-immunoglobulin. VII Amino acid sequence of heavy-chain cyanogen bromide fragments H1-H4. *Biochemistry.* 9:3161.

28. Rechavi, G., D. Ram, L. Glazer, R. Zakut, and D. Givol. 1983. Evolutionary aspects of immunoglobulin heavy chain variable region (V_H) gene subgroups. *Proc. Natl. Acad. Sci. USA.* 80:855.

29. Kenten, J. H., H. V. Molgaard, M. Houghton, R. B. Derbyshire, J. Viney, L. O. Bell, and H. J. Gould. 1982. Cloning and sequence determination of the gene for the human immunoglobulin ε chain expressed in a myeloma cell line. *Proc. Natl. Acad. Sci. USA.* 79:6661.

30. McKean, D., K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert. 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. USA.* 81:3180.

31. Shlomchik, M. J., D. A. Nemazee, V. L. Sato, J. Van Snick, D. A. Carson, and M. G.

Weigert. 1986. Variable region sequences of murine IgM anti-IgG monoclonal autoantibodies (Rheumatoid Factors). A structural explanation for the high frequency of IgM anti-IgG B cells. *J. Exp. Med.* 164:409.

32. Bakhshi, A., P. Guglielmi, U. Siebenlist, J. V. Ravetch, J. P. Jensen, and S. J. Korsmeyer. 1986. A DNA insertion/deletion necessitates an aberrant RNA splice accounting for a μ heavy chain disease protein. *Proc. Natl. Acad. Sci. USA.* 83:2689.

33. Takahashi, N., T. Noma, and T. Honjo. 1984. Rearranged immunoglobulin heavy chain variable region (V_H) pseudogene that deletes the second complementarity-determining region. *Proc. Natl. Acad. Sci. USA.* 81:5194.

34. Capra, J. D., and D. G. Klapper. 1976. Complete amino acid sequence of the variable domains of two human IgM anti-gamma globulins (Lay/Pom) with shared idiotypic specificities. *Scand. J. Immunol.* 5:677.

35. Forre, O., J. B. Natvig, and H. G. Kunkel. 1976. Serological detection of variable region (V_H) subgroups of Ig heavy chains. *J. Exp. Med.* 144:897.

36. Capra, J. D., and D. G. Klapper. 1976. Distribution and association of heavy and light chain variable region subgroups among human IgM immunoglobulins. *Ric. Clin. Lab.* 6(Suppl. 3):39.

37. Kodaira, M., T. Kinashi, I. Umemura, F. Matsuda, T. Noma, Y. Ono, and T. Honjo. 1986. Organization and evolution of variable region genes of the human immunoglobulin heavy chain. *J. Mol. Biol.* 190:529.

38. Brodeur, R., and R. Riblet. 1984. The immunoglobulin heavy chain variable region in the mouse I. 100 IgH-V genes comprise 7 families of homologous genes. *Eur. J. Immunol.* 14:922.

39. Brodeur, P., M. Thompson, and R. Riblet. 1984. The content and organization of mouse IgH-V families. *In* Regulation of the Immune System. H. Cantor, L. Chess, and C. Sercarz, editors. Alan R. Liss, Inc., New York. 445.

40. Rathbun, G. A., J. D. Capra, and P. W. Tucker. 1987. Organization of the murine V_H locus in the inbred strains. *Eur. Mol. Biol. Organ. (EMBO) J.* In press.

41. Pech, M., and H. G. Zachau. 1984. Immunoglobulin genes of different subgroups are interdigitated within the V_K locus. *Nucleic. Acids Res.* 12:9229.

42. Knight, K. L., R. C. Burnett, and J. M. McNichols. 1985. Organization and polymorphism of rabbit immunoglobulin heavy chain genes. *J. Immunol.* 134:1245.

43. Lotscher, E., K.-H. Grzeschik, H. G. Bauer, H.-D. Pohlenz, B. Straubinger, and H. G. Zachau. 1986. Dispersed human immunoglobulin K light-chain genes. *Nature (Lond.).* 320:456.

44. Newkirk, M. M., and J. D. Capra. 1987. Restricted usage of immunoglobulin variable region genes in human autoantibodies. *In* The Immunoglobulin Gene. T. Honjo, T. Rabbits, and F. W. Alt, editors. Academic Press, New York. In press.

45. Ravetch, J. V., U. Siebenlist, S. Korsmeyer, T. Waldmann, and P. Leder. 1981. Structure of the human immunoglobulin μ locus: Characterization of embryonic and rearranged J and D genes. *Cell.* 27:583.

46. Siebenlist, U., J. V. Ravetch, S. Korsmeyer, T. Waldmann, and P. Leder. 1981. Human immunoglobulin D segments encoded in tandem multigenic families. *Nature.* 294:631.

47. Andrews, D. W., and J. D. Capra. 1981. Complete amino acid sequence of variable domains from two monoclonal human anti-gamma globulins of the Wa cross-idiotypic group: suggestion that the J segments are involved in the structural correlate of the idiotype. *Proc. Natl. Acad. Sci. USA.* 78:3799.

48. Nardella, F. A., D. C. Teller, C. V. Barber, and M. Mannik. 1985. IgG rheumatoid factors and staphylococcal protein A bind to a common molecular site of IgG. *J. Exp. Med.* 162:1811.

49. Hjelm, H. 1975. Isolation of IgG3 from normal human sera and from a patient with multiple myeloma by using protein A–Sepharose 4 B. *Scand. J. Immunol.* 4:633.

50. Jefferis, R., M. I. Nik Jaafar, and M. Steinitz. 1984. Immunogenic and antigenic epitopes of immunoglobulins. VIII. A human monoclonal rheumatoid factor having specificity for a discontinuous epitope determined by histidine/arginine interchange as residue 435 of immunoglobulin G. *Immunol. Lett.* 7:191.

51. Siegelman, M., and J. D. Capra. 1981. Complete amino acid sequence of light chain variable regions derived from five monoclonal anti-*p*-azophenylarsonate antibodies differing with respect to a crossreactive idiotype. *Proc. Natl. Acad. Sci. USA.* 78:7679.

52. Slaughter, C. A., and J. D. Capra. 1983. Amino acid sequence diversity within the family of antibodies bearing the major antiarsonate crossreactive idiotype of the A strain mouse. *J. Exp. Med.* 158:1615.

53. Meek, K., D. Jeske, M. Slaoui, O. Leo, J. Urbain, and J. D. Capra. 1984. Complete amino acid sequence of heavy chain variable regions BALB/c mice expressing the major crossreactive idiotype of the A/J strain. *J. Exp. Med.* 160:1070.