

# Antigen-binding specificities of antibodies are primarily determined by seven residues of $V_H$

( $V_H$  subgroups/antigen-binding pocket/conserved residues in complementarity-determining regions 1 and 2)

SUSUMU OHNO, NOZOMU MORI, AND TAKESHI MATSUNAGA

Beckman Research Institute of the City of Hope, Duarte, CA 91010

Contributed by Susumu Ohno, January 2, 1985

**ABSTRACT** Although the antigen-binding pocket of all antibodies consists of  $V_L + V_H$  dimers (where  $V_L$  and  $V_H$  represent immunoglobulin light and heavy chain regions, respectively), subgroups of their  $V_H$  largely determine their antigen-binding specificities. This  $V_H$  subgroup dependence automatically relegates subsidiary roles to  $V_L$  as a whole and to the complementarity-determining region 3 (CDR-3) of  $V_H$  encoded by independent diversity ( $D$ ) and joining ( $J$ ) coding segments in determining antigen-binding specificities of individual antibodies. As a sequel to our previous paper, which emphasized the role conserved residues in CDR-1 and CDR-2 of  $V_H$  play in general shaping of the primordial antigen-binding cavity, here we propose that the three short clusters of amino acid sequences in CDR-1 and CDR-2 that are placed in the immediate vicinity of the tryptophan loop primarily determine subgroup-dependent antigen preference of individual  $V_H$ , therefore, antibodies. The three clusters are the 31st to 35th positions of CDR-1 and the 50th to 52nd and 58th to 60th positions of CDR-2. Of those, the 32nd, 34th, 51st, and 59th positions tend to be occupied by tyrosine, methionine, isoleucine, and tyrosine, respectively. Nevertheless, free amino acid substitutions at the remaining seven sites can generate  $20^7$  or  $1.28 \times 10^9$  varieties of amino acid sequence combinations. Some of these astronomically numerous sequence combinations no doubt contribute to the maintenance of the vast repertoire of antigen-combining diversity, which might be as large as  $10^7$ , whereas others serve to vary binding affinities toward the same antigen. Ironically, but not surprisingly, a single nonconservative amino acid substitution at one of these sites often suffices to change the antigen preference of  $V_H$  from one to another, whereas more substitutions affecting two or more clusters are apparently required to change the binding affinity toward the same antigen. In the case of mouse anti-*p*-azophenylarsenate antibodies, the principle of  $V_H$  subgroup dependence is violated, their  $V_H$  belonging to either subgroup 1 or 3. It appears that the mouse genome lacks anti-*p*-azophenylarsenate germ line  $V_H$ , residues of CDR-3 derived from one particular  $J_H$  coding segment coming to rescue to cope with this unnatural man-made antigen.

In our previous paper (1), we argued that, although the antigen-binding pocket of immunoglobulins is made of  $V_L + V_H$  dimers (where  $V_L$  and  $V_H$  represent immunoglobulin light and heavy chain regions, respectively), the antigen-binding specificities of individual antibodies are primarily determined by a small number of residues in the complementarity-determining regions 1 and 2 (CDR-1 and CDR-2) of  $V_H$  that are brought in the immediate vicinity of its tryptophan loop. The fact is that to which subgroup its  $V_H$  belongs largely dictates the antigen-binding specificity of an individual antibody, thus, directly contradicting the popularly held notion

that such a specificity is determined neither by individual  $V_L$  nor by individual  $V_H$ , but by a particular  $V_L + J_L \times V_H + D_H + J_H$  combination (where  $D$  = diversity and  $J$  = joining). This trend of  $V_H$  subgroup dependence already evident in the 1983 compilation of immunoglobulin  $V$  region sequences by Kabat *et al.* (2) became even more convincing in the newer classification of mouse  $V_H$  to 7 subgroups proposed by Dildrops (3). These subgroups are said to correspond to 7 clusters of germline  $V_H$  genes residing in chromosome 12 of the mouse (3). The surprisingly strict  $V_H$  subgroup dependence of the antibody's antigen-binding specificities revealed by Dildrops' newer classification is compiled in Fig. 1. Such  $V_H$  subgroup dependence automatically assigns the subsidiary role not only to  $V_L$  but also to CDR-3 of  $V_H$  in determining the antigen-binding specificity of individual antibodies.

During differentiation of the B-cell lineage,  $V_H + D_H + J_H$  fusion occurs first, followed by either  $V_{\kappa L} + J_{\kappa L}$  or  $V_{\lambda L} + J_{\lambda L}$  fusion (4, 5). Unless a subsequently chosen  $V_L$  can accommodate the first chosen  $V_H$  reasonably well, a given B-cell clone cannot become an effective antibody producer.  $V_L$  is clearly destined to play a subsidiary role to  $V_H$ . CDR-3 of  $V_H$  is derived from independent coding segments  $D$  and  $J$ , the sequence diversity within CDR-3 being attributable to the presence of several  $D_H$  as well as  $J_H$  segments in the genome (4, 5). Thus, the same germ line  $V_H$  amino acid sequence is often followed by divergent CDR-3 sequences. Inasmuch as CDR-3 divergence has nothing whatsoever to do with  $V_H$  subgroups, the subgroup-dependent antigen preference by germ line  $V_H$  coding segments automatically precludes the active role by CDR-3 in determination of antigen preference.

In this paper, we develop our proposal further and identify 7 critical positions in CDR-1 and CDR-2 of mouse  $V_H$  as the primordial determiners of antigen-combining specificities.

## Four Evolutionarily Conserved Anchoring Residues of CDR-1 and CDR-2 Positioned in the Immediate Vicinity of the $V_H$ Tryptophan Loop

The cardinal rule of molecular evolution is that so long as a particular function is assigned to a polypeptide chain, free amino acid substitutions are permissible only at functionally trivial parts (6). Thus, one would expect that many positions of CDR-1 and CDR-2 represent trivial positions irrelevant to  $V_H$ 's antigen preference. Furthermore, a lesson can be drawn from an enzyme's substrate-binding pocket, which is made only of conserved residues. Thus, we arrived at the conclusion that there must be evolutionarily conserved residues within CDR-1 and CDR-2 and that a few hypervariable positions in the immediate neighborhood of such conserved

Abbreviations: CDR-1, -2, and -3, complementarity-determining regions 1, 2, and 3; CRI (+/-), crossreactive idiotype (+/-); GAT, poly(Glu<sup>60</sup>Ala<sup>30</sup>Tyr<sup>10</sup>); FR-1, -2, and -3, framework regions 1, 2, and 3; NP, 4-hydroxy-3-nitrophenylacetyl; NIP, 5-iodo-4-hydroxy-3-nitrophenylacetyl;  $V_L$ , immunoglobulin light chain variable region;  $V_H$ , immunoglobulin heavy chain variable region;  $D$ , diversity;  $J$ , joining.

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

SUBGROUP	ANTIGEN SPECIFICITIES
1	ANTI-(4-HYDROXY-3-NITROPHENYL) ACETYL, ANTI-POLY (GLU <sup>60</sup> , ALA <sup>30</sup> , TYR <sup>10</sup> ), ANTI- $\alpha$ -1,3-DEXTRAN, ANTI-AZOPHENYLARSONATE (CRI <sup>+</sup> )
2	ANTI-2-PHENYLOXAZOLONE
3	ANTI-DINITROPHENYL, ANTI-AZOPHENYLARSONATE (CRI <sup>-</sup> )
4	ANTI- $\beta$ -1,6-D-GALACTAN, ANTI- $\beta$ -2,6-FRUCTOSAN
5	
6	ANTI- $\beta$ -2,1-FRUCTOSAN
7	ANTI-PHOSPHOCHOLINE

FIG. 1. Antigen-binding specificities of antibodies determined by  $V_H$  subgroups.

positions must decide antigen preference of individual  $V_H$ .

In Fig. 2, the inner  $\beta$ -sheet amino acid sequences of  $V_L$  (left) and  $V_H$  (right) of mouse antiphosphocholine myeloma MOPC511 (2) are arranged in the manner previously developed (1) that reasonably simulates the actual configuration of the  $V_L + V_H$  dimeric antigen-binding pocket as first revealed by the x-ray crystallographic study by Saul *et al.* (7). The floor of the dimeric antigen-binding pocket is formed by the contact between 5 residues each of  $V_L$  and  $V_H$ . In Fig. 2, these 10 residues are placed in square boxes and are identified by roman numerals. A pair of residues identified by the same roman numeral contacts each other—e.g., the 99th phenylalanine of  $V_L$  and the 103rd tryptophan of  $V_H$  contact each other; thus, both bear the identification numeral I. Accordingly, the absolutely invariant 35th tryptophan of  $V_L$  and 36th tryptophan of  $V_H$  emerge as the heart of antigen-binding pockets. Although a few hypervariable positions of CDR-3 of both  $V_L$  and  $V_H$  are positioned close to the above noted pair of invariant tryptophans, they would be excluded from the present consideration for the reason already given. Our view is that CDR-3 of  $V_H$  and of  $V_L$  directly concern themselves with antigen preference only when dealing with either exceptionally large antigenic determinants such as a hydroxy derivative of vitamin K<sub>1</sub>, which is a form of naphthoquinone (8), or odd, man-made antigens such as *p*-azophenylarsonate (9). Particularly instructive was the finding that a single amino acid substitution affecting CDR-3 (101st alanine being replaced by aspartic acid) of T15  $V_H$  of S107 antiphosphocholine antibody did not change a mutant's binding affinity to phosphocholine itself, only changing that to phosphocholine + carrier (10). Fig. 2 also reveals the reason why CDR-1 and CDR-2 of  $V_H$  rather than those of  $V_L$  primarily determine the antigen-binding specificity of antibodies. In the case of  $V_H$ , the first 9–12 residues, depending upon the occupancy of 52A, 52B, and 52C positions, of CDR-2 form an ascending loop immediately above the invariant 36th tryptophan. In sharp contrast, all 7 residues of the truncated CDR-2 of  $V_L$  are used as the connecting strand between inner and outer  $\beta$ -sheet structures. Since this connecting strand begins with the very invariant last 2 residues (48th isoleucine and 49th tyrosine) of framework region 2 (FR-2), all of the 7 CDR-2 residues of  $V_L$  are placed a little distance away from the heart of the antigen-binding pocket. Thus, the only hypervariable residues of  $V_L$  that may be involved in antigen discrimination are the last 3 residues of CDR-1 immediately above the invariant 35th tryptophan (Fig. 2).

As to CDR-1 and CDR-2 of  $V_H$ , we have detected 4 nearly invariant residues placed in the immediate vicinity of the 36th tryptophan. They are the 32nd tyrosine, 34th methionine, 51st isoleucine, and 59th tyrosine. Among  $V_H$  amino acid sequences of seven mammalian species, including man, tabulated by Kabat *et al.* (2), the above 4 residues occurred

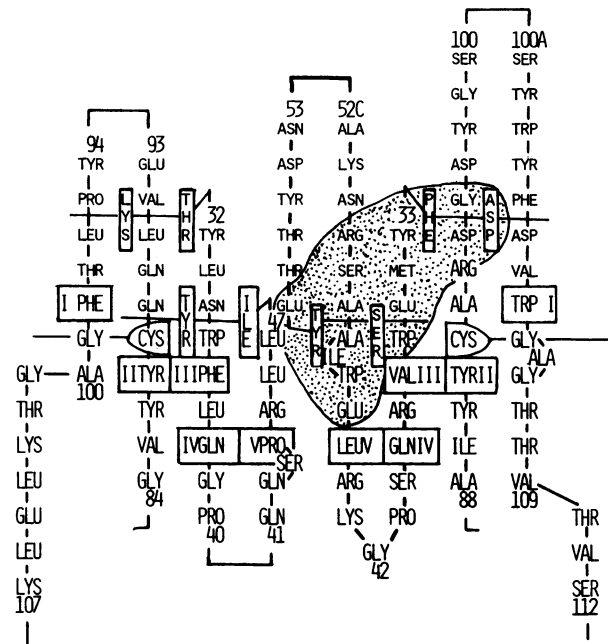


FIG. 2. The  $V_L + V_H$  dimeric antigen-binding pocket of mouse antiphosphocholine myeloma MOPC511 (2) formed by inner  $\beta$ -sheet structures of  $V_{AL}$  (left) and subgroup 7  $V_H$  (right) is shown as the collapsed three-dimensional structure flattened to two dimensions after disruption of two intrachain disulfide bridges. Residues of FR-2, FR-3, and FR-4 of both are shown in large capital letters; those of CDR-1, CDR-2, and CDR-3 are shown in small capital letters. Pertinent residues of connecting strands that bridge inner  $\beta$ -sheet structure with outer  $\beta$ -sheet structure (not shown) are shown as encased vertical strips. Residues at both ends of each  $\beta$ -strand are numbered. The floor of this dimeric antigen-binding pocket is made by contacts between 5 framework residues each of  $V_{AL}$  and  $V_H$ . These contacting residues are encased in square boxes and members of each contacting pair bear the identical roman numerals—e.g., the 98th phenylalanine of  $V_{AL}$  FR-4 and 103rd tryptophan of  $V_H$  FR-4 are labeled by roman numeral I; the 44th proline of  $V_{AL}$  FR-2 and 45th leucine of  $V_H$  FR-2 are identified by roman numeral V. The 88th cysteine of  $V_{AL}$  and 92nd cysteine of  $V_H$  are encased in bullet-shaped boxes. The intrachain disulfide bridge that each formed with the 23rd or 22nd cysteine of the outer  $\beta$ -sheet structure has been disrupted to avoid collapsing of outer  $\beta$ -sheet structure atop the inner  $\beta$ -sheet structure. In view of the  $V_H$  subgroup dependence of the antibody's antigen-binding specificities, 16 residues of  $V_H$  thought to be involved directly in discrimination of antigens are encircled in the shaded area. They are all 5 residues of CDR-1 (31st to 35th), the head and tail of FR-2 (36th tryptophan and 46th to 49th), and the head and tail of CDR-2 (50th to 52nd and 58th to 60th).

with the following frequencies: 32nd tyrosine (54.2%, its closest relative, phenylalanine, accounting for an additional 23.3%), 34th methionine (74.1%), 51st isoleucine (69.0%), and 59th tyrosine (86.6%). Perhaps more revealing was the finding on germ line  $V_H$  unadulterated by somatic mutations. Excluding 5 pseudogenes, 19 complete germ line  $V_H$  sequences of the mouse are currently known (4, 9, 11–16). Of those, if we regarded the 32nd phenylalanine as a true equivalent of the 32nd tyrosine, the conservation at all 4 positions was seen in 13. As to the remaining 6, there was not a single germ line  $V_H$  that sustained more than one violation. Excluding 1 pseudogene, 4 human germ line  $V_H$  sequences have been determined thus far (17, 18). The 32nd tyrosine, 34th methionine, 51st isoleucine, and 59th tyrosine were conserved *in toto* by all 4. Observing the top row, left and 2nd from left of Fig. 3, it should be noted that the 32nd tyrosine, 34th methionine, and 59th tyrosine were also conserved by one germ line  $V_H$  each derived from the South American alligator *Caiman* and from the primitive shark *Heterodontus*

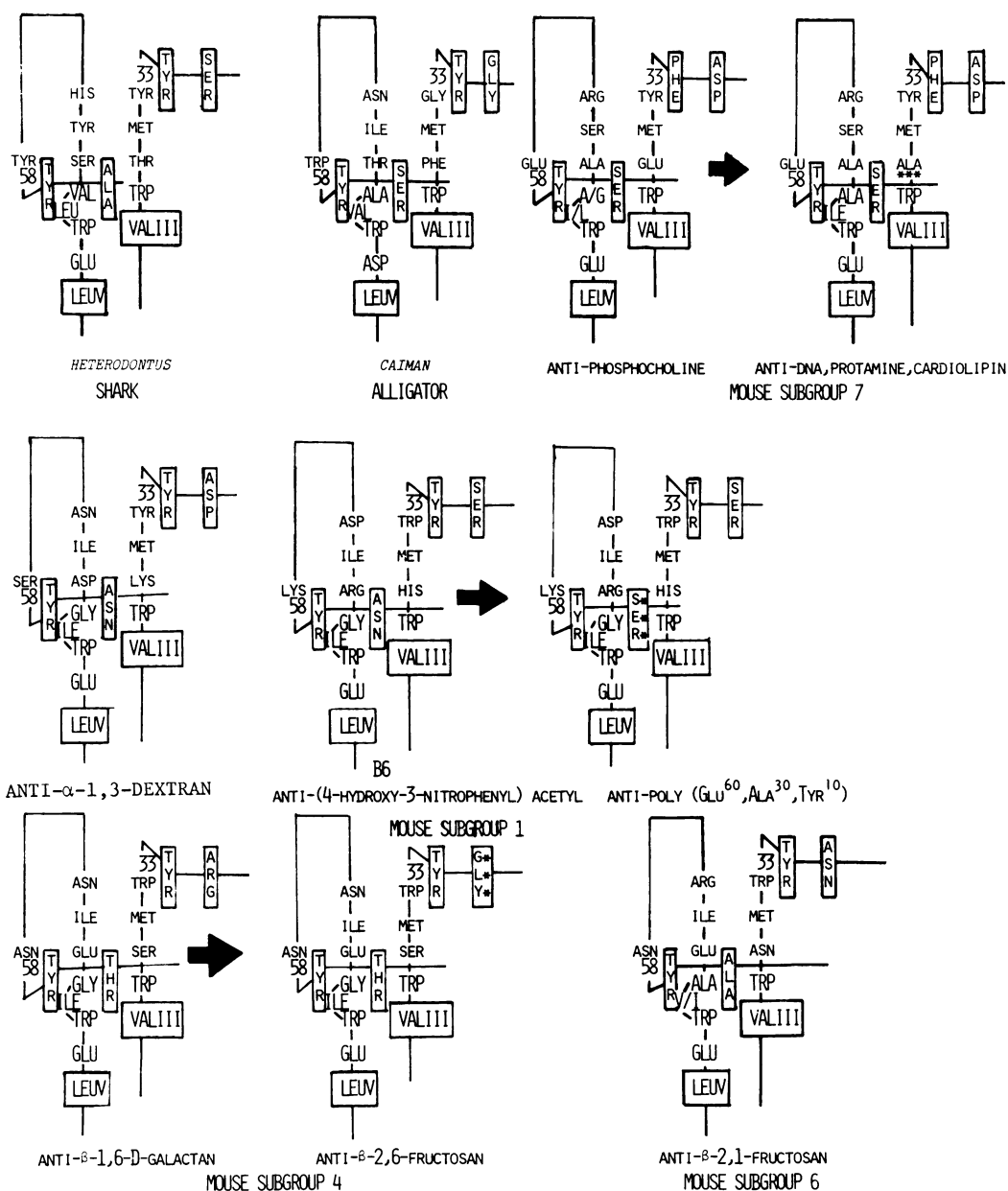


FIG. 3. Ten pertinent examples in three rows of the shaded area of V<sub>H</sub> identified in Fig. 2 primarily determining the antigen-binding specificities of V<sub>H</sub>, therefore, antibodies. In addition to 15 residues placed within the shaded area, the 37th VAL<sup>III</sup> and 45th LEUV are shown to signify the floor of the antigen-binding pocket. With regard to the 5-residue-long cluster of CDR-1, and two 3-residue-long clusters of CDR-2, all 10 are reasonably homologous with, but distinctly different from, each other. In fact, there are three examples of a single amino acid substitution within this small confine of V<sub>H</sub> changing the binding affinity of antibodies from one antigen to another: see the 3rd and 4th from the left of the top row, the 2nd and 3rd from the left of the middle row, and the 1st and 2nd from the left of the bottom row. Each pair is centered by a large arrow. In each of these three pairs, a single substituted amino acid is identified by asterisks. *Top row: left.* In one V<sub>H</sub> of the primitive shark *Heterodontus* (19), the 32nd tyrosine and 34th methionine of CDR-1 and the 59th tyrosine of CDR-2 were conserved, but not the 52nd isoleucine. *The 2nd from the left.* In one V<sub>H</sub> of the South American alligator *Caiman* (20), the 52nd isoleucine of CDR-2 too was conserved. *The 3rd and 4th from the left.* Two V<sub>H</sub> belonging to mouse V<sub>H</sub> subgroup 7 in Dildrop's new classification (3) differed by a single amino acid substitution at the 35th position of CDR-1. Yet, this substitution changed antiphosphocholine to anti-DNA, antiprotamine, and anticardioplin (21). This 3-cluster CDR-1, CDR-2 amino acid sequence combination shown at the 3rd from the left characterized 25 T15 antiphosphocholine V<sub>H</sub> sequences tabulated (2). *Middle row: left.* Mouse V<sub>H</sub> of anti- $\alpha$ -1,3-dextran, anti-4-hydroxy-3-nitrophenylacetyl (anti-NP), and anti-poly(Glu<sup>60</sup>Ala<sup>30</sup>Tyr<sup>10</sup>) (anti-GAT) belong to subgroup 1 as shown in Fig. 1. The 3-cluster amino acid sequence combination of CDR-1 and CDR-2 shown at the left characterized 19 of the 20 complete V<sub>H</sub> sequences of anti- $\alpha$ -1,3-dextran antibodies tabulated (2), whereas a single amino acid substitution at the 60th position of CDR-2 changed B6 strain-specific anti-NP V<sub>H</sub> shown at the 2nd from the left (12) to anti-GAT shown at the 3rd from the left (22). *Bottom row:* Within mouse V<sub>H</sub> subgroup 4 (Fig. 1), the 3-cluster amino acid sequence combination of CDR-1 and CDR-2 shown at the left characterized all 4 complete V<sub>H</sub> sequences of anti- $\beta$ -D-1,6-galactan antibodies tabulated (2), whereas one complete anti- $\beta$ -2,6-fructosan V<sub>H</sub> sequence shown at the 2nd from the left differed from the left only by a substitution of the 31st arginine with glycine (2). With regard to other positions not shown, four anti- $\beta$ -D-1,6-galactan V<sub>H</sub> sequences differed as much from each other as they differed from one anti- $\beta$ -2,6-fructosan V<sub>H</sub>. Mouse subgroup 6 is represented by the 3-cluster amino acid sequence combination shown at the 3rd from the left. This combination characterized all 8 complete V<sub>H</sub> amino acid sequences of anti- $\beta$ -2,1-fructosan antibodies tabulated (2).

(19, 20). The 51st isoleucine was also conserved by *Caiman* but not by *Heterodontus*.

The observed evolutionary conservation at these 4 presumably hypervariable positions (2) clearly indicated that the

above 4 CDR-1 and CDR-2 residues were as instrumental as the invariant 36th tryptophan and conserved 46th to 49th residues of FR-2 in shaping the general outline of primordial antigen-combining cavities. It thus followed that antigen-combining specificities of these primordial cavities were to be determined by truly hypervariable residues in the immediate neighborhood of the above 4 conserved CDR-1 and CDR-2 residues. In view of the above, all 5 CDR-1 residues (31st to 35th) and two tripeptidic stretches of CDR-2 residues (50th to 52nd and 58th to 60th) are placed within the shaded area of Fig. 2 that signifies the primordial antigen-combining cavity of  $V_H$ . Included in this shaded area are the absolutely invariant 36th tryptophan and 46th to 49th Glu-Trp-Ile-Ala of the tryptophan loop entirely comprised of FR-2.

**Hypervariabilities at the 31st, 33rd, and 35th Positions of CDR-1 and Those at the 50th, 52nd, 58th, and 60th Positions of CDR-2 Suffice to Define the Subgroup-Dependent Antigen Preference of  $V_H$**

At this stage, we asked the question of whether or not the 3-cluster amino acid sequence combination involving the above noted 11 CDR-1 and CDR-2 positions suffices to define the individual  $V_H$ 's subgroup-dependent antigen preference. More precisely, the same particular 3-cluster amino acid sequence combination should not be found among  $V_H$  directed against different antigens and whenever there is an instance of a single amino acid substitution changing  $V_H$ 's preference from one antigen to another, such a substitution should affect 1 of the 11 CDR-1 and CDR-2 positions. As far as the data available at the present are concerned, both expectations have been amply satisfied.

Within the subgroup 7  $V_H$  gene cluster in the genome of BALB/c inbred mice, there exists T15  $V_H$  (11) and all of the 25 antiphosphocholine antibodies of BALB/c listed by Kabat *et al.* (2) utilized this T15  $V_H$ . The 3-cluster amino acid sequence combination characteristic of this T15 is the 31st to 35th Asp-Phe-Tyr-Met-Glu, 50th to 52nd Ala-Ser-Arg, and 58th to 60th Glu-Tyr-Ser as shown at the top row, 3rd from the left, of Fig. 3. On the basis of their x-ray crystallographic study of mouse antiphosphocholine T15 myeloma MOPC603, Kazim and Attasi (23) have identified 5 of the 11 residues of the 3 clusters as combining to phosphocholine. They were the 33rd tyrosine and 35th glutamic acid of CDR-1 (1st cluster), the 52nd arginine of the 2nd cluster, as well as the 59th tyrosine and 60th serine of the 3rd cluster. Most revealing was the recent finding of a single amino acid substitution affecting T15  $V_H$  that changed the antiphosphocholine antibody to an autoantibody directed against DNA, protamine, and cardiolipin, as reported by Diamond and Scharff (21). As shown at the top row, right of Fig. 3, this substitution affected the 35th position: glutamic acid being replaced by alanine. A similar, albeit less complete, finding was also reported by Eilat *et al.* (24). Although the 52nd arginine combining to the phosphate group (23) was retained, a new cavity apparently became a little larger by glutamic acid to alanine substitution. Thus, a new antibody combined to larger phosphate group-containing compounds. In sharp contrast, many somatically sustained amino acid substitutions affecting T15 at sites other than the 3-cluster positions did not alter T15's antiphosphocholine binding specificity (2, 10, 11). Nevertheless, T15  $V_H$  did not appear to be the only germ line antiphosphocholine  $V_H$  in the mouse genome. The 3-cluster amino acid sequence combination of CBA/J 6G6 antiphosphocholine  $V_H$  was different from that of T15 at 4 of the 11 positions: the 31st to 35th was Asp-Tyr-Tyr-Met-Ser instead of Asp-Phe-Tyr-Met-Glu and the 50th to 52nd was Leu-Ile-Arg instead of Ala-Ser-Arg (2). This CBA/J 6G6 antiphosphocholine  $V_H$  was apparently derived from germ line V13  $V_H$  identified by Crews *et al.* in the neighborhood of T15  $V_H$  (11), for V13  $V_H$  was endowed with the same 3-cluster amino acid sequence combination as

CBA/J 6G6  $V_H$ . Comparison of T15 with CBA/J 6G6 was also very instructive in understanding how different 3-cluster amino acid sequence combinations may manage to recognize the same antigen. As shown in Figs. 2 and 3, the 35th position of CDR-1 and the 50th position of CDR-2 are in direct apposition of each other. In T15  $V_H$ , these 2 positions were occupied by glutamic acid and alanine, whereas in CBA/J 6G6  $V_H$ , serine and leucine replaced them. To accommodate the same antigen, the space between these 2 apposing residues had to remain similar. Accordingly, a nonconservative substitution at the 35th position called for a second substitution of a specific kind at the 50th position to preserve the similar spacing. It should be recalled that a single 35th glutamic acid to alanine substitution changed antiphosphocholine T15 to an autoantibody directed against larger phosphate group-containing compounds (21).

According to the more recent classification (3),  $V_H$  of anti- $\alpha$ -1,3-dextran, anti-NP, and anti-GAT antibodies belong to the subgroup 1 gene cluster of the mouse genome. Nineteen of the 20 complete  $V_H$  sequences of mouse anti- $\alpha$ -1,3-dextran antibodies listed by Kabat *et al.* (2) demonstrated the unique but identical 3-cluster amino acid sequence combination: the 31st to 35th Asp-Tyr-Met-Lys, 50th to 52nd Asp-Ile-Asn, and 58th to 60th Ser-Tyr-Asn (Fig. 3, middle row, left). The remaining one differed from the above by a single conservative substitution at the 34th position: methionine being replaced by valine (2). Although NP and 5-iodo-4-hydroxy-3-nitrophenylacetyl (NIP) are not naturally occurring antigens, the genome of C57BL/6 (B6) inbred mice harbors one strain-specific anti-NP  $V_H$  locus (12). When combined with  $V_{AL}$ , this  $V_H$  produces an antibody with a peculiar heterocliticity, the antibody provoked by NP demonstrating a higher affinity toward a derivative of NP (NIP) than to NP itself (12). The unique 3-cluster amino acid sequence combination of this peculiar B6  $V_H$  is shown at the middle row, center, of Fig. 3. Once again, a single amino acid substitution changed this anti-NP  $V_H$  to anti-GAT  $V_H$  (22). Subgroup 1  $V_H$  of an anti-GAT antibody raised in a B6  $\times$  DBA/2  $F_1$  mouse differed only by a single amino acid substitution from anti-NP B6  $V_H$ , serine replacing asparagine at the critical 60th position (Fig. 3, middle row, right). Ironically, 5 somatically sustained amino acid substitutions, 2 of which affect the critical 33rd (tryptophan to leucine) and 58th (lysine to threonine) positions, merely increased this B6 specific anti-NP  $V_H$ 's binding affinity to NIP 4-fold (12).

With regard to subgroup 4  $V_H$ , the 3-cluster amino acid sequence combination unique to, but identical in, all 4 complete anti- $\beta$ -D-1,6-galactan  $V_H$  (2) is shown at the bottom row, left, of Fig. 3. Interestingly, the germ line  $V_H$ 441 gene was anti- $\beta$ -D-1,6-galactan  $V_H$  and this germ line  $V_H$  was picked up by a nucleic acid probe derived from UPC10  $V_H$ , which was anti- $\beta$ -2,6-fructosan (14). Thus, anti- $\beta$ -D-1,6-galactan  $V_H$  and anti- $\beta$ -2,6-fructosan  $V_H$ , belonging to the same subgroup 4, differed again by a single amino acid substitution at the critical 31st position, arginine being replaced by glycine (Fig. 3, bottom row, middle). As to subgroup 6  $V_H$ , the 3-cluster amino acid sequence combination unique to, but identical in, all 8 anti- $\beta$ -2,1-fructosan  $V_H$  (2) is shown at the bottom row, right, of Fig. 3. Reflecting subgroup difference, this anti- $\beta$ -2,1-fructosan specific combination differed from both anti- $\beta$ -D-1,6-galactan and anti- $\beta$ -2,6-fructosan combinations at 4 of the 11 critical positions.

**The Mouse Genome May Not Contain a Single Anti-*p*-Azophenylarsonate  $V_H$**

Observing Fig. 1, it should be noted that the principle of  $V_H$  subgroup dependence on an antibody's antigen preference is violated by mouse anti-*p*-azophenylarsonate antibodies, for their  $V_H$  may belong to either subgroup 1 or subgroup 3,  $V_H$

subgroups in this instance determining crossreactive idiotypes (+/-) [CRI (+/-)] rather than antigen-binding specificities. According to Siekevitz *et al.* (9),  $V_H$  of CRI (+) anti-*p*-azophenylarsonate antibodies raised in A/J inbred mice was invariably derived from 1 of the 3 related subgroup 1 germ line  $V_H$ . Although this particular  $V_H$  was endowed with its own unique 3-cluster amino acid sequence combination, which was the 31st to 35th Ser-Tyr-Gly-Ile-Asn, 50th to 52nd Tyr-Ile-Asn, and 58th to 60th Lys-Tyr-Asn, this primordial antigen-combining cavity was apparently directed against an unknown antigen and not at all to a man-made *p*-azophenylarsonate, for they found that anti-*p*-azophenylarsonate activity was entirely dependent upon  $V_H$  CDR-3. Only when its CDR-3 region was contributed by the  $J_H2$  coding segment did this  $V_H$  become anti-*p*-azophenylarsonate, whereas the contribution of CDR-3 by the  $J_H4$  coding segment made CRI (+) antibodies that showed no binding affinity toward *p*-azophenylarsonate (9). Thus, we have gained an additional insight into phenomenal versatility of the adaptive immune system. Germ line  $V_H$  presumably directed against naturally occurring antigens may become specific to odd man-made antigens via their CDR-3 contributed by independent coding segments  $D_H$  and  $J_H$ .

### Evolutionary Consideration

At the present, it appears that the adaptive immune system is a unique property of vertebrates only. Inasmuch as the spontaneous mutation rate per base pair/year of viruses and bacteria is several orders of magnitude higher than that of vertebrate hosts—of the order of  $10^{-3}$  and  $10^{-4}$  versus  $10^{-9}$  (25)—to effectively combat these parasitic pathogens, the adaptive immune system had no choice but to generate antigen-combining sites ahead of the actual encounter with specific antigens. The recently demonstrated ability of the immune system to generate specific antibodies directed against all sorts of man-made nonbiological compounds is an ample testimony to this unique property of the adaptive immune system. However, there was a price to be paid for this audacity in anticipating the future, the actual encounter with new antigens having had no effect upon the repertoire of antigen-combining diversity that was already producing antibodies directed against those new antigens for an eon. In short, because of this audacity, the antigen-combining repertoire was placed beyond the reach of natural selection.

It follows that the adaptive immune system at its inception in lower vertebrates had to contain the germ from which the enormous repertoire of antigen-combining diversity comparable to that possessed by modern mammals could have sprung in a quick order. Indeed, the already quoted works by Litman's group on ancient reptile and shark indicated that germ line  $V_H$  gene pools as complex as those of mammals were already possessed by them (19, 20). Although there apparently was no *a priori* constraint upon the number of  $V_H$  as well as  $V_L$  gene copies the vertebrate genome could have acquired at the inception of the adaptive immune system, there had to be an acute shortage in a variety of available originals. Thus, the largely time-dependent diversification of antigen preferences among those gene copies had to be speeded up by some ingenious means. What if the ultimately ancestral  $V_H$  gene of an eon ago already encoded the 32nd tyrosine and 34th methionine of CDR-1 as well as the 51st isoleucine and 59th tyrosine of CDR-2? Those invariant residues and equally invariant 36th tryptophan, 46th to 49th glutamic acid, tryptophan, and isoleucine (or valine) and alanine (or glycine) of FR-2 would have defined the general

shape of the primordial antigen-binding cavity. Free amino acid substitutions at 7 positions in the immediate neighborhood of these invariant residues would quickly have increased the antigen-combining repertoire of  $V_H$ , therefore, antibodies. Random amino acid substitutions at 7 positions have the potential to generate  $20^7$  or  $128 \times 10^7$  amino acid sequences. It is granted that many of these combinations shall be directed toward the same antigen, but with varying degrees of binding affinities. Yet, even if 128 combinations on the average are directed against the same antigen, that leaves an antigen-combining repertoire of  $10^7$  which should suffice for the adaptive immune system of modern mammals. When contributions from independently encoded CDR-3 of  $V_H$  and those from  $V_L$  are added to the above, the adaptive immune system becomes truly versatile, being able to handle odd, man-made compounds to which the system has never been exposed.

1. Ohno, S., Matsunaga, T. & Lee, A. D. (1984) *Scand. J. Immunol.* **20**, 377–388.
2. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. (1983) *Sequences of proteins of immunological interest* (Natl. Inst. Health, Bethesda, MD).
3. Dildrop, R. (1984) *Immunol. Today* **5**, 85–86.
4. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature (London)* **286**, 676–683.
5. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980) *Cell* **19**, 981–992.
6. Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, New York).
7. Saul, F. A., Amzel, L. M. & Polijak, R. K. (1978) *J. Biol. Chem.* **253**, 585–597.
8. Twining, S. S. & Attasi, M. Z. (1978) *J. Biol. Chem.* **253**, 5259–5262.
9. Siekevitz, M., Huang, S. Y. & Geftter, M. L. (1983) *Eur. J. Immunol.* **13**, 123–132.
10. Cook, W., Rudikoff, S., Giusti, A. M. & Scharff, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1240–1244.
11. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981) *Cell* **25**, 59–66.
12. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewski, K. & Baltimore, D. (1981) *Cell* **24**, 625–637.
13. Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. (1981) *Cell* **27**, 573–581.
14. Olo, R., Auffray, C., Sikorav, J. L. & Rougeon, F. (1981) *Nucleic Acids Res.* **9**, 4099–4109.
15. Givol, D., Zakut, R., Effron, K., Rechevi, G., Ram, D. & Cohen, J. B. (1981) *Nature (London)* **292**, 426–430.
16. Cohen, J. B., Effron, K., Rechevi, G., Ben-Neriah, Y., Zakut, R. & Givol, D. (1982) *Nucleic Acids Res.* **10**, 3353–3370.
17. Rechevi, G., Bienz, B., Ram, D., Ben-Neriah, Y., Cohen, J. B., Zakut, R. & Givol, D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4405–4409.
18. Rechevi, G., Ram, D., Glazer, L., Zakut, R. & Givol, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 855–859.
19. Litman, G. W., Murphy, K., Berger, L., Litman, R., Hinds, K. & Erickson, B. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 844–848.
20. Litman, G. W., Berger, L., Murphy, K., Litman, R., Hinds, K., John, C. L. & Erickson, B. W. (1983) *Nature (London)* **303**, 349–352.
21. Diamond, B. & Scharff, M. D. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5841–5844.
22. Rocca-Serra, J., Tonnelle, C. & Forgereau, M. (1983) *Nature (London)* **304**, 353–355.
23. Kazim, A. L. & Attasi, Z. (1981) *Biochem. J.* **187**, 661–666.
24. Eilat, D., Hochberg, M., Pumphrey, J. & Rudikoff, S. (1984) *J. Immunol.* **133**, 489–494.
25. Gojobori, T. & Yokoyama, S. (1985) *Proc. Natl. Acad. Sci. USA* **82**, in press.