The human immunoglobulin heavy chain locus contains 39 functional human VH elements. All 39 VH elements (with their adjacent heptamer/nonamer signal) were tested for site-specific cleavage with purified human core RAG1 and RAG2, and HMG1 proteins in a 12/23-cleaved coupling reaction. Both nicking and hairpin formation were measured. The individual VH cleavage efficiencies vary over nearly a 30-fold range. These measurements will be useful in considering the factors affecting the generation of the immunoglobulin and T-cell receptor repertoires in the adult humans. Interestingly, when these cleavage efficiencies are summed for each of the VH families, the six VH family efficiencies correspond closely to the observed profile of unselected VH family usage in the peripheral B cells of normal adult humans. This correspondence raises the possibility that the dominant factor determining VH element utilization within the 1-megabase human genomic VH array is simply the individual RAG cleavage efficiencies.

The antigen receptor repertoire is a composite of many factors (1–3). One major factor is the efficiency with which the V, D, and J elements are cleaved by the recombinase complex. This complex contains RAG1 and RAG2 proteins (recombination activating genes), along with HMG1 (or HMG2). The RAG complex binds to the heptamer/nonamer signal sequences (also called RSS for recombination signal sequences) associated with each V, D, or J element. The RAG complex makes an initial nick adjacent to the heptamer of each signal and then generates a hairpin configuration at that site at the coding end terminus of the V, D, or J element (4). The hairpin formation at the coding end results in a blunt-ended double-strand break at the end of the signal (4). A single recombination event involves two elements, such as a V and a J, or a D and a J, or a V and a DJ. The two elements always have recombination signal sequences (RSS) that are different in the spacing between their heptamer and nonamer (12 or 23 base pairs), and this is known as the 12/23 rule (5). The 12/23 rule is enforced at the hairpin formation step by the RAG complex (6–8). The two coding ends (a D and a J, or a V and a DJ, for example) are joined by the nonhomologous DNA end joining repair pathway (2, 9) to create the variable domain exon that encodes a portion of the binding pocket for the antigen receptor (immunoglobulins (Ig) or T-cell receptor). The two signal ends are also joined together by nonhomologous DNA end joining to form a signal joint.

There is a consensus sequence for the heptamer (CACAGTG) and for the nonamer (ACAAAAACA). This consensus appears to be the optimal one for V(D)J recombination. However, the actual RSS associated with each V, D, and J element usually deviates considerably from the consensus. The variations affect recombination efficiency over several orders of magnitude (10, 11). In addition, the terminal two or three coding end nucleotides also influence the efficiency of nicking at the adjacent signal, and this coding end effect can influence the efficiency of recombination by an additional one to two orders of magnitude (12–15). Of the more than 10^9 possible combinations of heptamer and nonamer variations, only a small number (fewer than 100) of the possible variations have been tested (10, 11, 16). Hence, although some of the principles have been established concerning how signal and coding end sequence can influence V(D)J recombination, the recombination or cleavage efficiencies of the actual V, D, and J elements relative to one another have not been systematically determined for any of the human or murine loci. Therefore, the actual efficiencies cannot be deduced from the current literature, and direct experiments are required to determine the efficiencies that generate the repertoire of the antigen receptor loci.

Another factor that is known to influence V(D)J recombination is chromatin structure (17). There are six antigen receptor loci, and they do not undergo recombination simultaneously because of differences in chromatin structure. CpG methylation is one major factor that determines the accessibility of any vertebrate genetic locus (18–20), and the antigen receptor loci are no exception. CpG methylation is typically accompanied by histone deacetylation, which results in a tighter association between the nucleosome and the DNA wrapped around it. When RAG cleavage assays have been done on short DNA fragments that accommodate one nucleosome, cleavage is suppressed (21–24). Although chromatin structure clearly determines the RAG complex accessibility differences between the six antigen receptor loci, it has been uncertain whether such effects influence the differential V, D, or J element utilization within any one locus. Individual active antigen receptor loci that have been examined have acetylated histones throughout
the locus (25), consistent with much earlier data indicating that active antigen receptor loci are hypomethylated (26, 27).

In neonatal mice, it has been observed that VH segments in proximity to the JH cluster are used more frequently than the distal VH segments (28-31). This proximity preference across the murine VH array is either much less marked or not detectable in adult mice (28-30, 32). In fetal human, it has been unclear whether or not there is a small bias in favor of the most proximal VH (30). If there is a small bias in fetal human, it might suggest that in mouse and, to a lesser extent, in human, the locus opening (chromatin change) emanates from the intron enhancer during early development and that this proximity effect dissipates with progression to adulthood. This is especially obvious in humans where it is clear that, in adults, there is no proximity preference within the Ig heavy chain locus (30, 33–36).

Once the primary repertoire is generated as a result of recombination and overlapping chromatin effects, positive and negative selective forces shape the repertoire into its observed profile in the peripheral blood. Here, we have examined each of the VH elements in the six human VH families for their efficiency in the initial stages of V(DJ) recombination by using human core RAG complexes to cleave oligonucleotide fragments encompassing the signal and the adjacent coding end nucleotides. We find that the cleavage efficiency of each of the six families corresponds very closely to data for the nonproductive VH repertoire in the adult human adult is determined to a significant extent by the recombination cleavage efficiency.

**EXPERIMENTAL PROCEDURES**

**DNA Substrates**—All oligonucleotides were synthesized by Operon Technologies, Inc. (Alameda, CA). Cleavage substrates were made by annealing two complementary oligonucleotides to form a blunt-ended double-stranded DNA. Each cleavage substrate contains a 15-bp (base pair) coding flank and a 5-bp flank on the nonamer side. The sequences of each VH are listed in Fig. 1. The substrate with the 12-bp spacer RSS that was used for pairing with each VH element was derived from the left (upstream) side of the human DH4-4 substrate (also called DA4). It is made by annealing oligonucleotides KY106 (5'-GAG TGT TAC TGT AGT CAC ACA GTA GGA GGA CCC TTC ACA AAA AGC CCC TG-3') and KY107 (5'-CGG GCG CTT TTG GTG AAG GGT CCT CTT ACT GTG TTA CTA CAG TAA TTC CC-3'). The pre-nicked D3-4 substrate is composed of ML126, ML127, and KY107, which were annealed (see below) in molar ratios of 1:1:2. The sequences are as follows. ML126 (5'-GAGT TAC TGT AGT GCACA-3') and ML127 (5'-pACA GGA GGA CCC TTC ACA AAA AGC CCC TG-3').

We polycyclamide gel-purified the full-length form of each oligonucleotide. We then determined the concentration spectrophotometrically. Each cleavage substrate was made at the 5'-end of the coding flank with γ²32P-ATP (3000 Ci/mol) (PerkinElmer Life Sciences, Boston, MA) and T4 polynucleotide kinase (New England BioLabs, Beverly, MA) according to the manufacturer's instructions. Unincorporated radiolabeled DNA was removed using G-50 Sephadex (Amersham Biosciences, Inc., Piscataway, NJ) spin-column chromatography. It is important to note that labeled oligonucleotides were mixed with twice the molar amount of unlabeled complementary oligonucleotides in a buffer containing 10 mM Tris-hydrochloride, pH 8.0, and 100 mM NaCl (37). The mixture was heated at 95 °C for 5 min and allowed to cool down to room temperature for more than 1 h. The amount of unannealed labeled single-stranded oligonucleotide was less than 5% in all cases (and was typically at undetectable levels (less than 2%)).

**Protein Expression and Purification**—Human RAG expression plasmids (core RAG1; amino acids 383–1008 and full-length RAG2) were kind gifts of Dr. Sadro Santagata and Dr. Patricia Cortes. The core RAG2 (amino acids 1-383) expression vector was made by inserting the corresponding PCR fragment into the pEBG vector (38). The sequence was confirmed by dideoxynucleotide sequencing. Recombinant human RAG proteins were expressed as glutathione S-transferase (GST) fusion proteins by cotransfection of RAG1 (core) and RAG2 (core) expression vectors into the human epithelial cell line 293T. (Previous work demonstrates the similar kinetics of nicking and hairpin formation for GST versus maltose binding protein RAG fusion proteins (8).) The coexpressed human core RAG proteins were then purified with glutathione-agarose (Sigma Chemical Co., St. Louis, MO). C-terminal truncated mouse HMG1 was expressed in bacteria as a six-histidine-tagged protein and purified over a nickel-nitrilotriacetic acid column (Qiagen Inc., Valencia, CA). Protein concentration was determined against known concentrations of bovine serum albumin (fraction V) on a Coomassie Blue-stained gel with a densitometer (Model GS710, Bio-Rad, Hercules, CA) and quantified with Quantity One software (Bio-Rad, Hercules, CA).

**Coupled Cleavage Assay**—A 5-µl reaction mixture containing 10 fmol of a 32P-labeled VH substrate, 10 fmol of unlabelled D3-4 (DA4), and 25 mM MOPS, pH 7.0, 2.5 mM MgCl2, 30 mM potassium chloride, 30 mM potassium glutamate, 1 pmol of HMG1, and 20 ng of RAGs was incubated at 37 °C for 30 min. The reaction was stopped by the addition of 5 µl of formamide and immediately heated to 100 °C for 5 min before plugging into ice water. At least two independent cleavages on two independent gels were done for every VH. In addition, independent annealings were done on a subset of VH segments that deviated more than expected relative to the most similar VH elements. The independent annealings were indistinguishable.

Five femtomoles of each substrate of the two specified substrates was used (see Fig. 5, lanes 3 and 6). Ten femtomoles was used when only one substrate is involved. The difference in band intensity for equal molar amounts of substrate is simply due to labeling efficiency differences. Such labeling efficiency differences are not a complication, because equal molar amounts are used, and the conversion to nicked and hairpin products is expressed as a percentage of the substrate input in that reaction.

**Denaturing Polyacrylamide Gel Electrophoresis**—Reaction products were separated on 15% polyacrylamide gels containing 7 M urea in 1× Tris borate-EDTA buffer. Gels were visualized by autoradiography with a Molecular Dynamics PhosphorImager 445SI (Sunnyvale, CA) and quantified with ImageQuant software (version 5).

**Statistical Analysis**—Evaluations of the correlation between RAG nicking (or hairpin formation) and published values for nonproductive rearrangements were done as follows. The RAG nicking contribution (expressed as a percentage) for each VH family relative to the total was calculated as described in Table I. The nonproductive rearrangement frequencies are taken from Refs. 37 and 38. The RAG nicking percent-

**TABLE I**

|               | VH1 | VH2 | VH3 | VH4 | VH5 | VH6 |
|---------------|-----|-----|-----|-----|-----|-----|
| RAG nicking   | 9.8 | 5.7 | 48.9| 27.8| 5.2 | 2.6 |
| RAG hairpinning| 8.7 | 9.9 | 53.7| 20.5| 4.9 | 2.2 |
| Nonproductive rearrangementsa | 10 | 4.6 | 45 | 33 | 4.5 | 3.0 |
| Percentage if all VH used equallyb | 23.1 | 7.7 | 48.7 | 15.4 | 2.6 | 2.6 |

---

a The data for VH usage among nonproductive rearrangements in human peripheral B cells is from Table III of Ref. 49 and Table I of Ref. 50 by the Lipsky laboratory.

b The theoretical VH usage, if each VH has equal usage frequency, and the repertoire is merely a reflection of the family size.
age was plotted versus the nonproductive rearrangement frequency for each of the six families. The correlation coefficient was determined, and a probability value, \( p \), was calculated, as given in Table I.

### RESULTS

#### Experimental Strategy for Assessing the \( V_H \) Array in Paired 12/23 Cleavage Reactions

The sequence alignment of all 39 functional human \( V_H \) elements in the region relevant to V(D)J recombination (the heptamer and nonamer of the recombination signal sequence and the 15 bp of coding flank) illustrates that only five of the heptamer/nonamer signals conform to the CACAGTG/ACAAAAACC consensus (Fig. 1) (39). These five are \( V_H3-9, 3-43, 4-34, 4-39, \) and \( 4-59 \). The most common sequence of the \( V_H \) elements is shown in the top line, and this sequence differs from the optimal heptamer/nonamer in the fourth position of the nonamer. Five of the 39 \( V_H \) elements have heptamers that deviate from the optimal sequence, and 33 \( V_H \) elements have nonamers that deviate from the optimal sequence. The effects of these deviations on recombination efficiency cannot be reliably estimated from the current knowledge of the limited number of tested signal sequence variations.

The \( V_H \) elements have previously been grouped into six families based on the \( V_H \) coding sequence (39). Within each of the six \( V_H \) families, the 23-bp spacer region between the heptamer and the nonamer is highly homologous. The most distinct deviation is seen for the nine members of family 1, which, by comparison, have unique nonamer sequences and spacers relative to members from the other \( V_H \) families.

Previous work has demonstrated that the RAG cleavage efficiency of each substrate is determined by its similarity to the optimal signal sequence (40, 41) as well as for the coding end sequence adjacent to the signal (42). These biochemical cleavage efficiencies correspond very well with cellular V(D)J substrate quantitation, in those cases where equivalent substrates have been compared (12, 42). We determined the RAG...
cleavage efficiency of each human V_H element by using oligonucleotide-based DNA substrates, synthesized to correspond to the published sequences of human functional V_H elements. Each V_H substrate contains a 15-bp flank on the heptamer side (coding flank) and a 5-bp flank on the nonamer side. Previous studies have documented that these lengths of extension beyond the heptamer/nonamer signal are sufficient to recapitulate any surrounding sequence effects (42, 43). One strand of each double-stranded oligonucleotide was labeled such that the nicked product and the hairpin product could be distinguished from the substrate on a denaturing polyacrylamide gel (42, 44). Cleavage was carried out with purified human RAG proteins with Mg^{2+} as the divalent cation. A 12-bp RSS partner substrate is necessary for the 12/23 coupling at the hairpin formation step (8), and this must be the same partner for all of the V_H substrates to allow comparison. We chose D_H4-4 (also called DA4), which is efficiently cleaved by the human RAG complex (data not shown). However, it is important to note that the nicking step is independent of the presence or absence of a partner signal (44).

Each V_H substrate was assayed in replicate sets of experiments. To compare the different sets, an optimal 23-bp spacer substrate (KY36/37) was included in each set as a standard to normalize for any slight variations of cleavage efficiency. Human core RAG1 and RAG2 were used rather than the more commonly used murine core RAG1 and 2, because we were interested in establishing the relative cleavage efficiencies of the human V_H repertoire.

The RAG Cleavage Efficiency of Human V_H Targets Varies Markedly—The reaction time courses conformed to the expected kinetics for nicking and hairpin formation (41, 44, 45), and the products increased over the initial 50 min when incubated at 37 °C (Fig. 2). Cleavage efficiency was determined as the percentage of the substrate that is converted to the nicked or hairpin product at the 30-min time point. Although we used initial reaction rates in our previous study (44), the 30-min time point permits greater precision when comparing different substrates. The nicking and hairpin formation of a subset of the human V_H elements are illustrated in Fig. 3. Each V_H element was analyzed in two independent experiments in duplicate, and the degree of concordance between measurements was quite good, as reflected in the standard deviations (Fig. 1).

The nicking and hairpin formation efficiency of the V_H substrates varies markedly, depending on the sequence of the substrate. This is true even for members of the same family for those cases where there are differences in the signal sequence or coding end. The difference in nicking efficiency between the highest (V_H4-34) and lowest V_H (V_H1-58) is 28-fold. Three of the six V_H family 4 members (4-34, 4-39, and 4-59), the one family 5 member, and six of the 19 family 3 members have the highest cleavage efficiencies (Fig. 1). Not surprisingly, all five of the V_H elements with optimal heptamer/nonamer signals (V_H3-9, 3-43, 4-34, 4-39, 4-59) are among the highest efficiency RAG-nicking targets. The most common deviation of the V_H elements, the C in the fourth position of the nonamer, did not markedly reduce recombination efficiency as illustrated by the fact that V_H3-30, 3-53, and 3-66 are nicked at efficiencies that are only 1.1- to 2.6-fold lower than the V_H elements with optimal heptamer/nonamer signals.

V_H segments with substantial deviations in the heptamer or nonamer, such as V_H3-72, and most of the members in family 1, had lower cleavage efficiencies (Fig. 1). Some of the individual V_H substrates vary only slightly from the consensus, and yet have large reductions in cleavage efficiency. For example, V_H3-20 has a consensus heptamer and only a one-nucleotide deviation from the consensus nonamer, and yet it is relatively low in cleavage efficiency.

Conversely, large deviations from the consensus are not necessarily associated with large reductions. For example, V_H5-51 has three nucleotide deviations from the consensus nonamer, and yet it is cleaved efficiently by the human RAG complex. However, V_H1-58 also has three nucleotide deviations from the consensus nonamer (at positions different from those of V_H5-51), and yet its cleavage is 20-fold reduced relative to V_H5-51. These instances illustrate the lack of predictability when deviations from the consensus are present and illustrate that direct measurements of the cleavage or recombination efficiency are essential.

The one to three nucleotides of the coding end that directly adjacent to the heptamer have been previously demonstrated to affect V(D)J recombination (12–14), and this effect has been traced, at the biochemical level, to the nicking step (42). Here we see evidence of this when comparing V_H3-15 with V_H3-73 (also compare V_H4-34 with 4-59), where the only differences are not in the signal but in the coding end (Fig. 1).

Sequence variations in the spacer region or deep into the coding flank (more than three nucleotides from the nick site) generally show little effect on cleavage efficiency (compare V_H3-7 with 3-23, or 3-64 with 1-3, Fig. 1), consistent with previous findings that variations at these positions do not affect recombination efficiency in any large way (12, 15, 46–48). Nevertheless, limited effects of spacer sequence may be observable (e.g., compare V_H3-23 or 3-15 with 3-30).

The hairpin formation efficiency also varies over nearly a 28-fold range. In general, there are no marked disparities be-
between the nicking and hairpin formation results (see “Discussion”).

The Effect of Competition between Different V\textsubscript{H} Elements—At the genomic antigen receptor locus in the cell, the RAG proteins can bind at the signal sequence of any of a number of different V\textsubscript{H} elements; hence, there is potential competition between the elements. We were interested in whether the cleavage studies that we had done would be affected by competition or whether the V\textsubscript{H} substrates would be cleaved independently. We tested this by first determining the cleavage of V\textsubscript{H} substrates when individually paired with the D\textsubscript{H} element, as described above. We then tested two V\textsubscript{H} substrates with the D\textsubscript{H} element and followed the hairpin formation of these two V\textsubscript{H} substrates in the same reaction (Fig. 4). We were able to follow the hairpin formation of two V\textsubscript{H} substrates in the same reaction, because hairpins of different sequence have different gel mobility, despite having the same length.

In the first set of experiments, we chose two V\textsubscript{H} elements that are cleaved with comparable efficiency (Fig. 4, lanes 1–3). In the second set of experiments, we chose two V\textsubscript{H} elements that have a 5-fold difference in hairpin formation efficiency (Fig. 4, lanes 4–6). We find that the V\textsubscript{H} substrates are cleaved in this competitive study in a manner that was indistinguishable from their noncompetitive cleavage (Fig. 4, compare lane 3 with lanes 1 and 2; also compare lane 6 with lanes 4 and 5). Therefore, the V\textsubscript{H} substrates are cleaved independently, and in vitro competition does not affect their cleavage efficiency.

**DISCUSSION**

Analysis of RAG Cleavage Relative to the Observed Unselected Repertoire—We were interested in comparing the data observed in vitro with the repertoire observed in the human adult peripheral blood. The nonproductive rearrangements among B cells in the peripheral blood are the most appropriate comparison, because these represent recombination events prior to any immunological selection. Other investigators have done measurements of V\textsubscript{H} usage among nonproductive rearrangements at the human heavy chain locus by using single-cell PCR (49–51). The number of events is insufficient to evaluate each individual V\textsubscript{H} element, but comparisons can be done at the level of each family. To compare our cleavage data with those from single-cell PCR, we first calculate the cleavage efficiency of each individual V\textsubscript{H} relative to the sum of the 39 V\textsubscript{H} substrates.

![FIG. 3. In vitro RAG cleavage of V\textsubscript{H} substrates. Solid lines indicate DNA strands. The shaded triangle represents the 23 RSS. The star symbol indicates the position of the radioisotope label on each DNA molecule. S, substrate; N, nicked product; HP, hairpin product. Lanes: 1, (1-18); 2, (3-7); 3, (4-34); 4, (4-39); 5, (3-23); 6, (1-8); 7, (1-24); 8, (1-58); 9, (1-46); 10, (1-45); 11, (2-5); 12, (2-26); 13, KY36/37; 14, (5-51); 15, (4-59); 16, (3-72); 17, (4-31); 18, (3-20); 19, (3-64); 20, KY36/37 (a consensus heptamer/nonamer RSS substrate).](image)

![FIG. 4. Competition of different V\textsubscript{H} substrates in the same RAG cleavage reaction. Lanes 1–3, V\textsubscript{H} elements with comparable hairpin formation efficiency (lane 1, V\textsubscript{H}6-1 and lane 2, V\textsubscript{H}3-48) were either cleaved alone (lanes 1 and 2), or in a reaction that contains equal amounts of both (lane 3). Only half the amount of the labeled V\textsubscript{H}6-1 or V\textsubscript{H}3-48 is present in lane 3 as compared with lane 1 or lane 2 to preserve the same amount of labeled V\textsubscript{H} substrates in all three lanes. Lanes 4–6, similarly, V\textsubscript{H} elements differing 5-fold in hairpin formation efficiency (lane 4, V\textsubscript{H}3-9 and lane 5, V\textsubscript{H}3-49) were cleaved alone (lanes 4 and 5), or in a reaction that contains half the amount of each (lane 6). The ratio of the band intensity between a and b is 2.0 and is equal to the ratio between a' and b'. The ratio between c and d is 8.0 and is similar to the ratio between c' and d', which is 7.6. S designates the substrate, HP designates the hairpinned form, and N designates the nicked form.)
elements. We then add the members for each family. We find that this aggregate family nicking or hairpin formation efficiency matches the data on the usage frequency of nonproductively rearranged V_H elements surprisingly well (for nicking, \( p < 0.001 \); for hairpin formation, \( p < 0.01 \); Table 1).

The usage frequencies calculated from our cleavage data as well as from the \textit{in vivo} nonproductive rearrangement data (49, 50) are not merely a reflection of the V_H family size. For example, family 1 has nine members and family 4 has only six. Yet family 4 has higher cleavage values and \textit{in vivo} usage (Table 1). If all V_H elements were used with equal frequency and the repertoire were shaped based only on family size, one would expect family 1 to contribute 23.1\% and family 4 to contribute 15.4\% of the repertoire (Table 1). This is not the case for either our \textit{in vitro} measurements or the \textit{in vivo} nonproductive repertoire measurements (49, 50).

The similarity, at the family level, between our cleavage data and the observed peripheral unselected repertoire may provide a mechanistic understanding for the V_H usage frequencies \textit{in vivo}. These \textit{in vivo} frequencies may simply reflect how well the V_H elements are cleaved by the RAG complex. If so, this could be of considerable practical significance. For example, with a clearer understanding of the repertoire generation, deviations from the baseline repertoire of V_H usage might be more easily recognized and may be useful as a very early indication of monoclonal gammapathies due to B cell malignancies.

That the nicking efficiencies for the V_H families are very similar to the nonproductive repertoire in the peripheral blood suggests that the rate-limiting step for V(D)J recombination may be (a) the nicking step, (b) the preceding step in which the RAG complex binds the substrate, or (c) a combination of these two steps. If any of the subsequent steps (hairpin formation, hairpin opening, or any of the steps of nonhomologous end joining) were rate-limiting, then the nicking efficiencies would not be expected to so closely match the nonproductive peripheral repertoire.

\textbf{Biological Significance of a Predictable Initial Repertoire—} The similarity of the RAG cleavage efficiencies and the nonproductive pre-immune repertoire is interesting from the standpoint of factors that affect the V_H usage on human chromosome 14. The nonproductive pre-immune repertoire could conceivably be a reflection of more factors than simply RAG cleavage efficiencies, including such factors as chromatin structure (histone acetylation and CpG methylation) and local transcription. The fact that the \textit{in vitro} RAG cleavage efficiencies match those in the nonproductive pre-immune repertoire means that if factors other than RAG cleavage efficiency are anything other than negligible, then they may be offsetting each other.

The V_H families 2, 5, and 6 are the clearest in suggesting that the RAG cleavage efficiencies are the dominant factor in dictating the representation in the peripheral blood repertoire. Families 5 and 6 have only one member each, V_H5-51 and V_H6-61 are used in proportion to their RAG cleavage (Table 1), even though they are located 630 kb apart in the human genome. The three members of family 2 are spread across a similar distance of the V_H array, and each has very similar RAG cleavage efficiencies in our experimental system. We find that the representation of family 2 in the peripheral blood is also in proportion to its cleavage efficiency (Table 1). If RAG cleavage efficiency is a dominant factor determining V_H element recombination, it suggests that the chromatin structure across the 1 megabase V_H array does not vary dramatically, at least as it affects V(D)J recombination.

Our results are interesting in light of data from an independent line of work. When most of the human IgH array (35 functional V_H) was randomly integrated (via a yeast artificial chromosome) as a transgene into the germline of mice, the rearrangement of the V_H gene families was surprisingly similar to that seen in humans (52). These studies were done on productively rearranged alleles. Hence, immunological selection in the mouse \textit{versus} humans complicates the analysis. Nevertheless, in light of our work, the correspondence between the human unselected (nonproductive) repertoire and a transgene that is randomly integrated in a different species suggests that V_H cleavage efficiencies are a dominant factor in shaping the repertoire.

As mentioned earlier, in neonatal mice, it has been observed that V_H segments in proximity to the J_H cluster are used more frequently than the distal V_H segments (28–31). In adult humans, there is no data to suggest any proximity preference within the Ig heavy chain locus (28–30). Therefore, there is no contradiction between our suggestion that recombination signal strength could be a major factor and the murine fetal data on proximity as a major factor, because fetal mice and adult humans are quite different in this regard. This difference between mice and humans is not surprising. There are other examples where Ig repertoire diversification has been achieved by quite different mechanisms. For example, the diversification of IgH complementarities region 3 involves generation of a D\_\_\_ protein, whereas there is no D\_\_\_ in human, and quite different mechanisms are utilized to ensure diversification of this portion of the heavy chain (53).

\textbf{Evolution of the Human V_H Array and the Individual Recombination Efficiencies—} If the RAG cleavage efficiency is a major determinant for the frequency of usage of V_H elements, then it is reasonable to assume that the signal (and coding end sequences) evolved so as to optimize the level of each V_H in the repertoire as needed to handle the threat of prevailing pathogens. Hence, there may have been two levels of evolutionary pressure at the DNA sequence level, one being the sequence of the V and the other being the efficiency of the signal (and adjacent coding end) for cleavage (54). The sequence of the V_H coding region determines the range of antigens bound, whereas the sequence of the coding end and heptamer/nonamer determines the abundance of that V_H element in the steady-state repertoire. Insofar as that steady-state repertoire is the initial response to an invading microbe, the balance of V_H elements in that repertoire is important. It is intriguing that the relative ratios of family usage in this initial Ig heavy chain repertoire may be predictable from the relative RAG cleavage efficiencies in a manner that appears uncomplicated by other factors.

\textbf{Acknowledgments—} We thank Dr. Malcolm Pike (University of Southern California) for statistical advice. We thank K. Schwarz (Ulm, Germany) for suggesting the competition experiment. We thank C. Hsieh and Ryan Irvine for comments on the manuscript, and we thank members of our laboratory for helpful discussion.

\textbf{REFERENCES}

1. Fugmann, S. D., Lee, A. I., Shockett, P. E., Villes, J. J., and Schatz, D. G. (2000) \textit{Ann. Rev. Immunol.} 18, 495–527
2. Lieber, M. R. (1998) \textit{Ann. N.Y. Acad. Sci.} 849, 153–1332
3. Lewis, S. M. (1994) \textit{Adv. Immunol.} 56, 27–150
4. McBlane, J. F., Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gollet, M., and Oettinger, M. A. (1995) \textit{Cell} 83, 387–395
5. Tonegawa, S. (1983) \textit{Nature} 302, 575–581
6. vanGent, D. C., Ramsden, D. A., and Gollet, M. (1996) \textit{Cell} 85, 107–113
7. Eastman, Q. M., Lew, T. M. J., and Schatz, D. G. (1996) \textit{Nature} 380, 85–88
8. West, R. B., and Lieber, M. R. (1998) \textit{Mol. Cell. Biol.} 18, 6498–6415
9. Grawunder, U., West, R. B., and Lieber, M. R. (1998) \textit{Curr. Opin. Immunol.} 10, 172–180
10. Hesse, J. E., Lieber, M. R., Mizuki, K., and Gollet, M. (1989) \textit{Genes Dev.} 3, 1053–1067
11. Akira, S., Okazaki, K., and Sakano, H. (1987) \textit{Science} 238, 1134–1138
12. Gerstein, R. M., and Lieber, M. R. (1993) \textit{Genes Dev.} 7, 1459–1469
13. Bozholov, N. V., Wills, Z. P., and Weaver, D. T. (1995) \textit{Nucleic Acids Res.} 23, 1060–1067
14. Ezekiel, U. R., Engler, P., Stern, D., and Storb, U. (1995) \textit{Immunity} 2, 381–389
15. Ezekiel, U. R., Sun, T., Bozek, G., and Storb, U. (1997) \textit{Mol. Cell. Biol.} 17,
