Evolutionary aspects of immunoglobulin heavy chain variable region ($V_H$) gene subgroups

(DNA sequence/saltatory replication/pseudogenes)

GIDEON RECHAVI, DANIELA RAM, LILLIAN GLAZER, RINA ZAKUT, AND DAVID GIVOL

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel

Communicated by Elinor A. Kabat, October 29, 1982

ABSTRACT We isolated and determined the sequences of two human germ-line heavy chain variable region ($V_H$) genes and compared them with mouse $V_H$ genes. The results show that the human $V_HI$ subgroup is evolutionarily related to the mouse $V_HII$ subgroup. Evolutionary preservation of homologies in $V_H$ genes of the same subgroup includes not only the coding region but also intron size and homology in noncoding regions. This suggests that a $V_H$ gene subgroup constitutes a multigene family that undergoes concerted evolution. The homology between genes of the same subgroup in different species is greater than that between genes of different subgroups within a species. One of the $V_HII$ genes contains, in complementarity-determining region 2 (CDR2), a 13-base-pair previously shown to be in CDR2 of a $V_HIII$ gene and in a heavy chain diversity region gene, $D_H$ [Wu, T. T. & Kabat, E. A. (1982) Proc. Natl. Acad. Sci. USA 79, 5001–5002], suggesting the insertion of diversity region gene sequences into the $V_H$ gene. One of the human $V_H$ genes is a pseudogene because of a terminator, which, together with our previous results, shows that the $V_H$ gene repertoire contains 40% pseudogenes. In one of the $V_H$ genes, direct and inverted repeats at both 5' and 3' ends of the gene suggest a potential transposable element that encompasses the entire $V_H$ gene. It is possible that such a structure may facilitate saltatory replication and rapid expansion of $V_H$ gene families.

The variable region of the immunoglobulin heavy chain is encoded in the germ line in three separate DNA segments: $V_H$, $D_H$, and $J_H$ (1–3). Three of the framework regions (FRs) and two of the complementarity-determining regions (CDRs) are included in the coding region of the $V_H$ segment whereas recombination between variable (V), diversity (D), and joining (J) segments is necessary to form the expressed $V_H$ gene. It is likely that there are several hundred $V_H$ gene segments organized in tandem with spacers of 8–15 kilobases (4–6). The multitude of $V$ region amino acid sequences has been subdivided into subgroups that show a higher extent of homology and linkage between some amino acids along the chain (7–9). However, comparison of the amino acid sequences of mouse and human $V_H$ does not always allow the affiliation of $V$ regions from the different species to homologous subgroups (10); the DNA sequences of the $V_H$ genes may provide a better tool for comparison of homologous subgroups in various species. Hybridization with $V_H$ DNA probes showed a correlation between the extent of cross-hybridization and $V_H$ gene subgroups (11). It has been shown that homologous genes of the same subgroup ($V_HIII$) are physically linked and that genes of different subgroups are not interdispersed and can be ordered on the chromosome by deletion mapping (12, 13). This indicates that throughout evolution the $V_H$ gene subgroup is maintained as a multigene family. A $V_H$ multigene family (subgroup) preserves some characteristics in the noncoding segments that are different from those of other $V_H$ families (11) and comparison of $V_H$ genes from different species will increase our understanding of the structural features that characterize a subgroup in coding and noncoding regions.

We isolated human germ-line $V_H$ Genes by cross-hybridization with mouse $V_HII$ genes (6). The DNA sequence shows that the human genes are homologous to human $V_HI$ (subgroup I), as defined by Kabat et al. (14) and we will refer to the human genes analyzed here as human $V_HII$ genes. The homology between mouse and human $V_HII$ genes is preserved in the size (84 base pairs (bp)) and also in the sequence of the intron present at codon 4. Another $V_H$ gene subgroup ($V_HIII$) contains a longer intron (102 bp) in both mouse and human (11). The homology in both intron and framework codons between mouse and human genes of the same subgroup is greater than between $V_H$ genes of different subgroups in the same species. The sequence also suggests that DNA repeats at the 5' and 3' regions flanking the $V_H$ gene can form a large stem-and-loop structure that may facilitate saltatory replication (15).

MATERIALS AND METHODS

A human fetal liver gene library donated by T. Maniatis was screened by plaque hybridization to plasmid pCh104 containing a mouse $V_HII$ gene (6). Restriction endonuclease mapping of purified recombinant phage DNA was carried out as described (11). DNA sequences were determined mainly by cloning and analysis in the M13 mp8 vector (16, 17). Recombinant phage fragments that gave positive hybridization with pCh104 were

Abbreviations: $V_H$, $D_H$, and $J_H$, heavy chain variable, diversity, and joining regions; CDR, complementarity-determining region; FR, framework segment; bp, base pair(s).
electroeluted from agarose and ethanol precipitated. One or two micrograms of DNA was dissolved in 25 μl of 10 mM Tris-HCl/0.1 mM EDTA, pH 7.4, and sonicated in an Eppendorf tube, using a model W375 sonicator (375 W, 20 KHz; Heat System/Ultrasonics, Plainview, NY). Sonication was done at 4°C by four 5-sec bursts at settings of 4 (output control) and 50% (duty cycle).

The sonicated DNA was subjected to end repair with DNA polymerase I (Klenow fragment) in the presence of 0.1 mM dATP/2 μM [32P]dATP for 30 min at room temperature, followed by phenol extraction and ethanol precipitation. The DNA was dissolved in 5 μl of 10 mM Tris-HCl/1 mM EDTA, pH 7.4, and run on a 1.5% agarose minigel with size markers. The 300- to 500-bp fractions of the sonicated material were collected by cutting a small trough below the 300-bp marker and continuing the electrophoresis for nine consecutive 1-min periods at 20 mA. The nine fractions collected were pooled and ethanol was added. The precipitated DNA was dissolved in 10 μl of 10 mM Tris-HCl/0.1 mM EDTA, pH 7.4, and a portion was ligated to the Sma I site of M13 mp8 (10 ng) and used to transform JM101

**Fig. 2.** Comparison of DNA sequences of human and mouse V_{H} genes. The sequence of HG3 is translated to amino acids and compared with the VH amino acid sequence of human protein Eu (14). I and II, the direct and inverted repeats (see text) at nucleotides 136–157 and 1,008–1,016. In the flanking regions, small deletions (empty space) or insertions (above the dashes) were inserted to maximize the homology. Nucleotides identical to those in HG3 are marked by dashes, CDR regions are underlined, and recombination signals at the end of V_{H} are boxed. The sequence of mouse gene 108A (subgroup II) is from ref. 6 and the sequence of human protein Eu (subgroup I) is from ref. 14.
bacteria. Recombinant M13 mp8 (white plaques) was plated onto a lawn of JM101 bacterial host and, after overnight growth, transferred to nitrocellulose and hybridized to 32P-labeled pCh104. Positive plaques were picked into 1 ml of JM101 and grown for 5 hr, and the DNA was prepared and analyzed as described (18). In some cases, sequence analysis by the chemical degradation procedure (19) was also used.

RESULTS AND DISCUSSION

The Human $V_{H}$ Subgroup Is Homologous to the Mouse $V_{H}$ Subgroup. On the basis of the amino acid sequences, it is not always easy to relate subgroups from different species, and it was particularly difficult to discern the structural counterpart of the human $V_{H}$ subgroup in other species (10). It is shown here and in ref. 12 that the structure of the germ-line genes, including the noncoding regions, allows a better understanding of the evolutionary relationship of the $V_{H}$ subgroups and thus comparison among $V_{H}$ subgroups from different species. Several recombinant phages were isolated from a human gene library (20) by cross-hybridization with the mouse $V_{H}$ gene-containing pCh104 (6). The sequences of two of these genes (HG3 and H2A) were determined by a combination of the M13 cloning/analysis technique (7) and the chemical degradation procedure (19) as shown in Fig. 1. The sequence of >1,100 bp of a human $V_{H}$ gene (HG3) is shown and compared with that of the human H2A gene and the mouse 108A gene, described in ref. 6, in Fig. 2. The amino acid sequence deduced from the DNA sequence of HG3 is compared in Fig. 2 with that of a human myeloma protein Eu, belonging to the $V_{H}$I subgroup (11). The results show that protein Eu is highly homologous to the predicted amino acid sequence of HG3, differing by only 1 residue in the first 25. This suggests an evolutionary relationship between the mouse $V_{H}$II subgroup (108A gene) and the human $V_{H}$I subgroup. This is further supported by the homology in sequence and size of the intron at codon -4. We have shown previously that mouse $V_{H}$II genes contain an intron of 84 bp (6) whereas both mouse and human $V_{H}$III genes contain an intron of 102 bp (12, 21). The results in Fig. 2 show conservation of the intron size between mouse and human $V_{H}$II genes and provide further evidence for their evolutionary relationship.

Table 2. Sizes of mouse and human gene segments

<table>
<thead>
<tr>
<th>Signal peptide, codons</th>
<th>Intron, bp</th>
<th>$V_{H}$ codons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse $V_{H}$II</td>
<td>19</td>
<td>84</td>
</tr>
<tr>
<td>Human $V_{H}$II</td>
<td>19</td>
<td>84</td>
</tr>
<tr>
<td>Mouse $V_{H}$III</td>
<td>19</td>
<td>103, 159</td>
</tr>
<tr>
<td>Human $V_{H}$III</td>
<td>19</td>
<td>103</td>
</tr>
</tbody>
</table>

Data sources are as in Table 1. * +2 bp between codon 98 and the recombination signals.

Characteristics of $V_{H}$ Gene Subgroups. To evaluate the homology between $V_{H}$ genes of various subgroups, we compared the homology of every gene segment within and between subgroups and also between mouse and human genes (Table 1). The results show that, in the framework regions (FR1–FR3), the homology within a subgroup is significantly greater than that between subgroups. Moreover, the homology between mouse and human genes of the same subgroup is greater than that between $V_{H}$ genes of different subgroups in the same species. The homology in the intron also follows this rule and is even more significant with regard to the size of the intron (Table 2). This analysis indicates strong preservation of noncoding elements, in addition to the coding regions, in $V_{H}$ genes of the same subgroup.

We have shown previously that $V_{H}$ genes are organized in clusters of subgroups that are not interdispersed but show physical linkage of the cross-hybridizing genes (12). It is likely that each such cluster of $V_{H}$ genes is a multigene family that undergoes concerted evolution and preservation of the sequence characteristics of this subgroup (15). The data presented here extend this finding to the $V_{H}$II subgroup by comparing the similarity between mouse and human genes.

High Proportion of Pseudogenes. As shown in Fig. 2, codon 6 in H2A is a termination codon, TAG. Hence, H2A may be a pseudogene that cannot be expressed. It has been shown previously that the mouse $V_{H}$ gene repertoire is rich in pseudogenes (6, 23). From a larger collection of $V_{H}$ gene sequences (unpublished data), we can now estimate that 40% of the germ-line genes analyzed (seven mouse and four human $V_{H}$ genes) drift and become pseudogenes because of terminators or other single-base replacements. This restricts the use of the entire germ-line $V_{H}$ gene repertoire and shows that there is a conflict between gene diversity and gene expressibility in the immune system. It is possible that mechanisms such as gene conversion or recombination (24, 25) may operate to allow the animal to make use of the coding potential present in these pseudogenes, but this has not yet been shown.

A $D_{H}$-Like Sequence in CDR2 of a $V_{H}$II Gene. A significant portion of the CDR3 codons in the expressed $V_{H}$ gene is contributed by $D_{H}$ gene segments after V–D–J recombination (1–3). This recombination is controlled by highly preserved signals

---Fig. 3. Locations of an identical 13-bp segment in CDR2 of two $V_{H}$ genes and in $D_{H}$. Numbers above the codons show their positions in the $V_{H}$ gene as in Fig. 2 (upper line) or as in ref. 26 (middle line). The sequence of $D_{H}$ is from ref. 3.---
located 5', 3', and at both ends of the V_H, J_H, D_H, respectively. Hence, the recombination joins the coding segments V-D-J end to end, where D sequences are represented in CDR3. Recently, however, Wu and Kabat (26) showed an identity between a 14-bp sequence present in D1H2 and in the CDR2 of a human germ-line V_HIII gene (V_H26) and suggested a possible insertion of D sequences into CDR2, within the V_H gene. It is of interest that 13 bp of the sequence described by Wu and Kabat (26) are present in CDR2 of gene HG3 in the same position as in V_H26 (Fig. 3). This 13-bp sequence appears as part of the CDR2 of two different subgroups (V_HII and V_HIII) as well as in the D minigene, coding for CDR3. If this identity between CDR2 and D1H2 sequences is not coincidental, it suggests that introduction of D-like sequences into V_H genes, either by insertion (26) or by gene conversion (25), is not limited to only one subgroup and may be an additional factor in generating antibody diversity. An alternative possibility is that the D_H minigenes may have evolved from a segment corresponding to CDR2 of a primordial V_H gene. The significant homology among various D_H genes (3) supports this hypothesis.

A Potential Transposable Element in V_H Genes. Gene sequences conserved between species may be of importance and we have compared noncoding regions of human and mouse V_H genes. In the flanking region 5' to the V_H gene, we find a sequence of 20 bp (nucleotides 136–157, Fig. 2) that is highly conserved in all mouse and human V_HII genes analyzed (Fig. 4). This sequence begins 124 bp 5' to the initiator ATG codon and may have a role in the function of the V_H genes. In Fig. 2, this 20-bp long sequence is subdivided into segments marked I and II. It is of interest that in gene HG3 (Fig. 2) an almost perfect direct repeat of segment I is found in the 3'-flanking region of the V_H gene (nucleotides 1,008–1,016, I in Fig. 2) and is preceded by an inverted repeat of segment II (nucleotides 996–1,007, II in Fig. 2). Hence, segments I and II at both the 5' and 3'-flanking regions of V_H form a potential stem-and-loop structure that includes the entire V_H gene (Fig. 5). The repeats are reminiscent of the border regions in prokaryotic, viral, and eukaryotic transposable elements (27–29). A similar structure was also found by inspecting mouse V_HII gene sequences published previously (24). The two structures of human and mouse V_H genes depicted in Fig. 5 show remarkable sequence similarity at the stem-and-loop borders, although the sizes of the loops are slightly different. A stem-and-loop structure may function in generating tandemly duplicated sequences leading to amplification of V_H gene families by the salutary replication mechanism (15). This mechanism suggests that, because of the stem-and-loop structure, several rounds of replication occur at a single initiation site, generating an "onion skin" replicated bubble. After ligation of the ends of the duplicated DNA, these replicated structures may recombine in the chromosome to generate a tandem array of gene sequences vicinal to the original gene duplicated (see ref. 30 for further discussion). Such a mechanism was suggested recently to explain somatic gene amplification and generation of multigene families (29). These events may be very rare, as suggested by the similarity of restriction enzyme patterns of V_H genes in Southern blots of DNA from various human individuals (unpublished data). However, it is possible that, at times of genome instability (perhaps during periods of speciation), salutary replication may take place. The immunoglobulin V_H gene families can undergo rapid evolutionary expansion (12, 15) and salutary replication or unequal crossing-over have been suggested to explain this phenomena (15). Salutary replication is a more attractive mechanism to explain the concerted evolution of immunoglobulin V genes in species-specific residues.

Another example for the possible involvement of DNA repeats in transposing DNA segments into genes was recently discussed by Cooke and Baxter (31). They suggested that exon 1 of the prolactin, growth hormone, and chicken conalbumin genes was independently inserted into these genes as a mobile genetic element, as indicated by the direct repeats flanking this region in all three genes.

Regulatory sequences were recently located in immunoglobulin V_H genes within 100 nucleotides 5' to the V_H coding region (32). Hence, these sequences are included in the V_H gene, which may be replicated in the stem-and-loop structure shown in Fig. 5. The replication of V_H genes by a salutary replication mechanism may be facilitated by the presence of a transposable element-like structure. This may explain some of the mechanisms operating in generating the large repertoire of V_H genes and processes of rapid expansion of V_H genes during evolution.

We thank Mrs. S. Hazum and S. Wylder for excellent technical assistance, Dr. T. Maniatis for the human gene library, and Dr. S. Fields for his help with the M13 sequence analysis technique.