J. Mol. Biol. (1986) **190**, 529–541

Organization and Evolution of Variable Region Genes of the Human Immunoglobulin Heavy Chain

Mieko Kodaira[1]†, Tatsuo Kinashi[1], Isao Umemura[1], Fumihiko Matsuda[1] Takafumi Noma[1], Yasushi Ono[2] and Tasuku Honjo[1]

[1]*Department of Medical Chemistry*
Kyoto University Faculty of Medicine
Sakyo-ku, Kyoto 606, Japan

and

[2]*Department of Microbiology*
Nihon University School of Medicine
Itabashi, Tokyo 173, Japan

(Received 6 January 1986, and in revised form 6 March 1986)

We have isolated 23 different cosmid clones of the heavy-chain variable region genes (V_H) of human immunoglobulin. These clones encompass about 1000×10^3 base-pairs of DNA containing 61 V_H genes. Characterization of the 23 clones by Southern blot hybridization showed that V_H genes belonging to different families were physically linked in many regions. Cluster 71, which was analyzed in detail, comprised seven V_H segments arranged in the same orientation with different intervals. This clone contained internal homology regions, each carrying two V_H segments of different families. Comparison of the nucleotide sequences of V_H segments within each family showed that profiles of accumulation of mutations in framework (FR) and complementarity-determining (CDR) regions were different. CDR had more mutations at amino-acid-substituting positions than at silent positions, whereas FR had the reverse distribution of mutations. Five out of seven V_H segments of this cluster were pseudogenes containing various mutations. V_H pseudogenes were classified into two distinct groups; one with a few replacement mutations (conserved pseudogenes), and the other with rather extensive mutations (diverged pseudogenes). The possibility that conserved pseudogenes serve as a reservoir of V_H segments is discussed.

1. Introduction

Immunoglobulins are composed of the heavy and light chains, each of which consists of variable and constant regions. The V‡ regions are confined to the amino-terminal domains of both chains and comprise the antigen binding site. The repertoire of V region diversity is determined by the combination of three different genetic mechanisms: (1) the evolutionary increment of the numbers of the germline V, D and J segments; (2) somatic recombination between V, D and J segments; and (3) somatic point mutations (for reviews, see Honjo, 1983; Tonegawa, 1983; Honjo & Habu, 1985). To evaluate the relative contributions of these mechanisms we and others have set out to quantify each mechanism. We have shown that the murine J_{K1} and J_{K2} segments were used for V–J recombination four to five times more efficiently than the J_{K4} and J_{K5} segments, and that the J_{K3} segment was inactive for recombination (Nishi *et al.*, 1985). The rate of somatic mutation was estimated to be 10^{-3} per base division (McKean *et al.*, 1984).

Assessment of the exact number of V segments is important for evaluation of the germline immunoglobulin repertoire. Although the numbers of D and J segments were determined (Early *et al.*, 1980; Sakano *et al.*, 1980; Wood & Tonegawa, 1983), the germline repertoire of the V segment has not been fully evaluated. The total number of murine V_K segments was estimated to be between 90 and 300, by two independent methods (Cory *et al.*, 1981; Nishi *et al.*, 1985). Bentley & Rabbitts (1981) estimated that the number of human V_K segments was smaller. Recent reports indicate that the total

† Present address: Radiation Effects Research Foundation, Hiroshima, Japan.

‡ Abbreviations used: V, variable region of immunoglobulin; D, diversity segment; J, joining segment; H, heavy chains; CDR, complementarity-determining region; bp, base-pairs; kb, 10^3 base-pairs; FR, framework region.

0022-2836/86/160529-13 $03.00/0

number of V_H segments of mouse might be about 100 (Kemp *et al.*, 1981; Brodeur & Riblet, 1984). The most direct approach to estimating the number of V segments is to clone all the germline DNA-encoding V segments. Such an approach was initiated for the V_K segments by Zachau and his colleagues (Pech *et al.*, 1984).

Three hypervariable regions of the V region correspond to the complementarity-determining regions (CDR 1 to 3) for antigen binding (Wu & Kabat, 1970). The hypervariability of CDR-3 depends on the randomly paired recombination of V, D and J segments, whereas the hypervariability of CDR-1 and CDR-2 depends on the diversity encoded by V segments, and the somatic mutation introduced after completion of the V gene (Early *et al.*, 1980; Sakano *et al.*, 1980; Kim *et al.*, 1981; Kataoka *et al.*, 1981). The relative contributions of the two kinds of diversity source to the hypervariability of CDR-1 and CDR-2 are less well understood. Complete characterization of the germline V segments will be required to answer this question.

Elucidation of the V_H segment organization may also be important for understanding the developmental pathway of B lymphocytes. Recent analyses of Abelson-virus-transformed B cell lines suggest ordered and preferred selection of a V_H segment for V–D–J recombination during B cell differentiation (Yancopoulos *et al.*, 1984).

With these goals in mind, we set out to clone the whole human V_H locus. In this work, we report cloning of 23 linkage groups of cosmid clones containing human V_H segments. Extensive characterization of one of the clones containing seven V_H segments revealed several interesting features of human V_H organization and evolution.

2. Materials and Methods

(a) *Genomic Southern blots*

High molecular weight DNA from human placenta and peripheral blood cells was prepared as described (Blin & Stafford, 1976; Yaoita & Honjo, 1980). Restriction endonuclease digests of DNA were electrophoresed through a 0·7% agarose gel and transferred to nitrocellulose filters according to the method of Southern (1975). The filters were hybridized with appropriate nick-translated probes as described (Honjo *et al.*, 1979). Following hybridization at 65°C for 14 to 16 h, filters were washed twice under standard conditions (at 50°C in 0·1 M-NaCl, 10 mM-sodium citrate, 0·1% sodium dodecyl sulfate). To classify cloned V_H genes into 3 V_H families, filters were first washed under standard conditions followed by washing twice under stringent conditions (at 65°C in 10 mM-NaCl, 1 mM-sodium citrate, 0·1% sodium dodecyl sulfate).

(b) *Cloning of V_{HBV}*

V_{HBV} was isolated from a human Epstein-Barr virus-transformed cell line (TAPC301-CL4), which produces immunoglobulin G (IgG) against surface antigens of hepatitis type B (Y. Ono, unpublished results). Southern blot hybridization of restricted DNA of TAPC301-CL4 with human J_H (Takahashi *et al.*, 1980) showed a $\approx$28 kb fragment containing a rearranged V_H gene (unpublished results). The 28 kb *Eco*RI fragment hybridizing to the J_H probe was partially purified by agarose gel electrophoresis and further partially digested with *Eco*RI*. The *Eco*RI* digests were ligated with λgt*WES* arms to construct a phage library. $\lambda V_{HB\cdot26}$ was isolated by screening with the J_H probe from this library. $pV_{HB\cdot26}$ was constructed as a plasmid containing a rearranged V_H gene of $\lambda V_{HB\cdot26}$.

(c) *Construction and screening of the cosmid library*

Cosmid libraries were constructed from DNA of human placenta or human peripheral blood. DNA was partially digested with *Sau*3AI or *Taq*I, and ligated with the cosmid vector pJB8 as described by Grosveld *et al.* (1981) and Ish-Horowicz & Burke (1981). An average efficiency of 3×10^5 colonies per μg of insert DNA was obtained when *Escherichia coli* 490A was transfected. The cosmid libraries were screened with a mixture of 3 unrelated V_H probes as described (Hanahan & Meselson, 1980), with a slight modification. The filters were washed 3 times for 20 min each under standard washing conditions, and exposed overnight with an intensifying screen. Colonies hybridized were picked up and purified by repeating colony hybridization.

(d) *Determination of nucleotide sequences*

Plasmid DNA of the cosmid clones was isolated by the alkaline lysis method as described (Birnboim & Doly, 1979; Maniatis *et al.*, 1982). Restriction maps of cosmid clones were constructed mainly according to the conventional procedure, but some of them were constructed using the "Southern Cross™ Restriction Mapping System" (New England Nuclear Co.) according to the manufacturer's directions. Isolated restriction fragments were cloned into M13mp10, M13mp11 (Messing, 1983), pUC18 and pUC19 (Yanisch-Perron *et al.*, 1985). The chain termination method was used for sequencing M13 clones (Sanger *et al.*, 1980). The strategy used to sequence plasmid DNA by the dideoxy DNA sequencing method was as described (Hattori & Sakaki, 1986). Briefly, plasmid DNA (1 to 5 pmol) was denatured with alkali. The resultant single-stranded DNA was used as a template after annealing with either of 2 different primers hybridizing to the 2 strands. Sequence analysis was performed by the method of Sanger *et al.* (1980). Primers and sequence kits were purchased from Takara Shuzo Co. Ltd. (Kyoto) or Amersham Japan (Tokyo).

3. Results

(a) *V_H segments of three families interspersed in the human V_H locus*

The human V_H proteins have been classified into three different subgroups according to their amino acid sequence homology (Kabat *et al.*, 1983). We have isolated and characterized five human V_H probes (Fig. 1), which were also classified into the three subgroups as shown in Table 1. The nucleotide sequence homology among members of subgroup I was more than 75% whereas the homologies between different subgroups were less than 60% (Table 2). We refer to "subgroups" defined by the amino acid sequence homology and by the DNA sequence homology as "subgroup" and

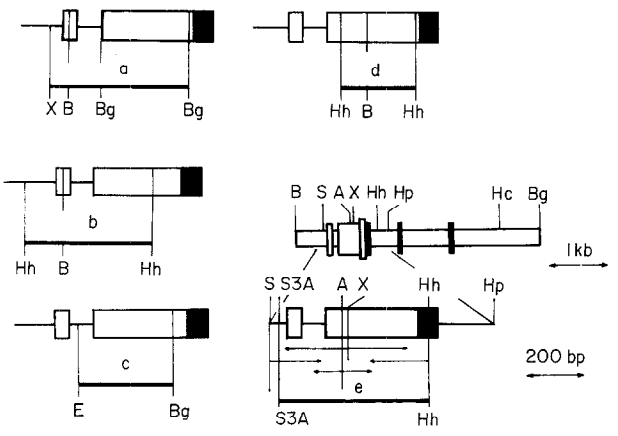

Figure 1. Restriction map of V_H probe. V_H probes were prepared as follows: V_{266BL} (a) was a 0·45 kb *Xba*I–*Bgl*II fragment of pHIGE-11 (Nishida *et al.*, 1982). This fragment was obtained by partial digestion with *Bgl*II. $V_{CE\text{-}114}$ (b) was a 0·45 kb *Hha*I fragment prepared by digesting the *Eco*RI–*Hin*dIII fragment of pCE-114 (Takahashi *et al.*, 1984). $V_{E3\text{-}D10}$ (c) was a 0·3 kb *Eco*RI–*Bgl*II fragment of $pV_{E3\text{-}D10}$ (Noma *et al.*, 1984). $V_{CE\text{-}1}$ (d) was a 0·25 kb *Hha*I fragment of $pV_{CE\text{-}1}$ (Takahashi *et al.*, 1984). $V_{HB\text{-}26}$ (e) was a 0·6 kb *Sau*3A–*Hha*I fragment of $pV_{HB\text{-}26}$ (see Materials and Methods). Nucleases are abbreviated: A, *Ava*II; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*cII; Hh, *Hha*I; Hp, *Hap*II; S, *Sac*I; S3A, *Sau*3AI; X, *Xba*I.

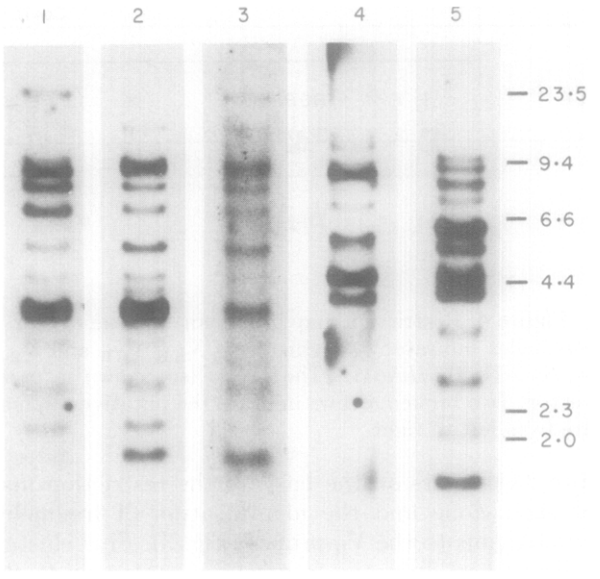

Figure 2. Southern hybridization of human placenta DNA by V_H probes. Southern blot filters of *Eco*RI digests of human placenta DNA were hybridized with V_H probes as indicated and washed under standard washing conditions. DNA size markers are indicated in kb. Each lane contained 2 μg DNA. Probes used for each lane are: lane 1, V_{266BL} ($V_{H\text{-}I}$); lane 2, $V_{CE\text{-}114}$ ($V_{H\text{-}I}$); lane 3, $V_{E3\text{-}D10}$ ($V_{H\text{-}I}$); lane 4, $V_{CE\text{-}1}$ ($V_{H\text{-}II}$); lane 5, V_{HBV} ($V_{H\text{-}III}$).

"family", respectively, in this paper. Subgroups I, II and III correspond to families I, II and III, respectively.

The V_H probe for each family hybridized to a distinct set of *Eco*RI fragments on Southern blots of human placenta DNA as shown in Figure 2. Similar results were obtained with *Bam*HI and *Hin*dIII digests (data not shown). In addition, three probes of the $V_{H\text{-}I}$ family, V_{266BL}, $V_{CE\text{-}114}$ and $V_{E3\text{-}D10}$, showed hybridization patterns that were almost identical, with variable relative intensities (Fig. 2). The results suggest that classification into subgroups I, II and III originally defined by the amino acid sequence is valid for the nucleotide sequence as well.

To investigate the germline organization of human V_H segments, cosmid libraries were constructed from DNA of human placenta and human peripheral blood. These cosmid libraries were screened with three unrelated V_H probes (V_{266BL}, $V_{CE\text{-}1}$ and V_{HBV}) and 120 recombinant clones were isolated. The restriction fragments hybridizing to the V_H probes in each clone were identified by Southern blot hybridization. All clones contained one or more restriction fragments hybridizing to any of the V_H probes. Isolated clones were grouped

Table 1
Sources of V_H probes

Subgroups (family)	Probes	DNA sources	References
I	V_{266BL}	266BL	Nishida *et al.* (1982) Kenten *et al.* (1982)
	$V_{CE\text{-}114}$	CESS	Takahashi *et al.* (1984)
	$V_{E3\text{-}D10}$	GUV-E3-D10	Noma *et al.* (1985)
II	$V_{CE\text{-}1}$	CESS	Takahashi *et al.* (1984)
III	V_{HBV}	TAPC301-CL4	This work

These probes were cloned as rearranged V_H genes from each cell line. Each probe was prepared as described in the legend to Fig. 1.

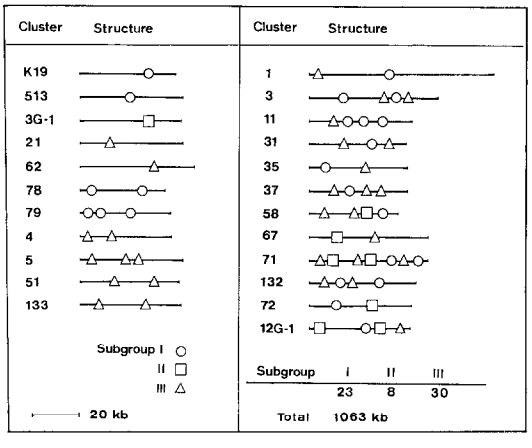

Figure 3. Distribution of V_H segments in cosmid clusters. Segments homologous to V_H probe were detected under the conditions described in Materials and Methods. The classification into the V_H family was done according to the relative intensity of hybridization to each V_H probe. Numbers of V_H segments were estimated by the numbers of *Eco*RI, *Hin*dIII and *Bam*HI fragments hybridizing to V_H probes, and the sum of each subgroup is shown at the bottom of the right-hand column. Location of V_H segments of each subgroup is shown by: (○) subgroup I; (□) subgroup II; (△) subgroup III.

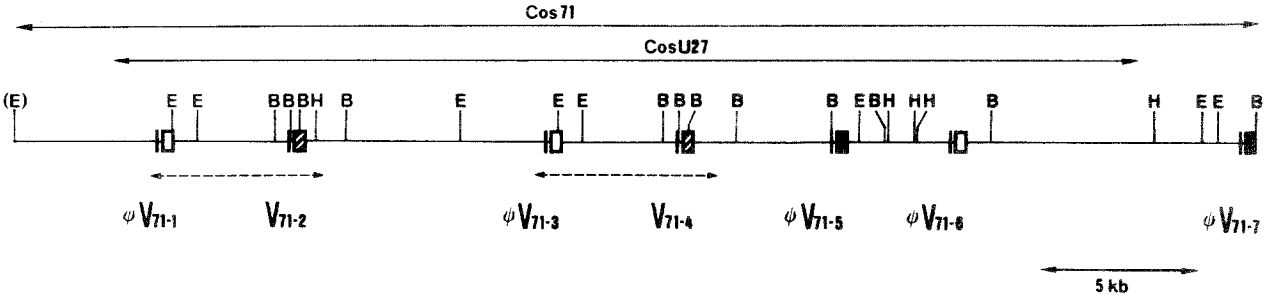

Figure 4. Restriction map of human V_H gene cluster 71. The map comprises 7 V_H genes. The filled, hatched and open rectangles indicate V_H segments of $V_{H\text{-}I}$, $V_{H\text{-}II}$ and $V_{H\text{-}III}$ families, respectively. Nucleases are abbreviated: E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII. The *Eco*RI site shown in parentheses indicates the artificial joining site to the cosmid (Cos) vector. Two broken arrows indicate the putative duplication regions. The orientations of all V_H segments in this cluster are from left to right.

into 23 clusters on the basis of the restriction map of each clone and the identification of fragments homologous to the V_H probes (Fig. 3). Five clusters isolated had only one fragment homologous to the V_H segment, but 18 clusters contained two or more fragments homologous to the V_H probes. Twelve clusters contained V_H segments of different families. This result suggests that V_H segments from three different families are interspersed in the human chromosome, although in mouse V_H segments of the same family are believed to be clustered (Kemp *et al.*, 1981; Rechavi *et al.*, 1982).

(b) *Cluster 71 containing seven V_H segments*

To analyze the human V_H segment organization in more detail, we focused on cluster 71, which consisted of two overlapping cosmid clones, Cos71 and CosU27 as shown in Figure 4. Southern blot hybridization with three different V_H probes (V_{266BL}, $V_{CE\text{-}1}$ and V_{HBV}) detected seven distinct DNA segments numbered in order on cluster 71 DNA from left to right. Seven V_H segments in cluster 71 were classified into three families by Southern blot hybridization; the $V_{H\text{-}I}$ family ($V_{71\text{-}5}$ and $V_{71\text{-}7}$), the $V_{H\text{-}II}$ family ($V_{71\text{-}2}$ and $V_{71\text{-}4}$) and the $V_{H\text{-}III}$ family ($V_{71\text{-}1}$, $V_{71\text{-}3}$ and $V_{71\text{-}6}$). These results clearly demonstrate that V_H segments of different families are interspersed in cluster 71.

To characterize further the V_H segments defined by hybridization experiments, we determined the DNA sequence according to the strategies shown in Figure 5. The coding regions were identified by comparison with the sequences of the probes used in

Table 2
DNA sequence homology

Families		$V_{H\text{-}I}$				$V_{H\text{-}II}$			$V_{H\text{-}III}$			
		V_{266BL}	$V_{E3\text{-}D10}$	$\psi V_{71\text{-}5}$	$\psi V_{71\text{-}7}$	$V_{CE\text{-}1}$	$V_{71\text{-}2}$	$V_{71\text{-}4}$	V_{HBV}	$\psi V_{71\text{-}1}$	$\psi V_{71\text{-}3}$	$\psi V_{71\text{-}6}$
$V_{H\text{-}I}$	V_{266BL}	100	81 (74)	78 (68)	79 (67)	50 (28)	60 (41)	59 (41)	56 (34)	59 (39)	60 (46)	53 (37)
	$V_{E3\text{-}D10}$		100	88 (78)	82 (73)	52 (36)	57 (45)	56 (45)	59 (39)	62 (46)	63 (50)	57 (42)
	$\psi V_{71\text{-}5}$			100	69 (63)	51 (31)	58 (44)	58 (44)	55 (31)	62 (43)	63 (45)	57 (23)
	$\psi V_{71\text{-}7}$				100	56 (41)	60 (45)	59 (45)	60 (45)	63 (43)	60 (45)	56 (37)
$V_{H\text{-}II}$	VCE-1					100	65 (58)	65 (58)	47 (32)	54 (38)	53 (39)	47 (38)
	$V_{71\text{-}2}$						100	99 (97)	56 (37)	61 (43)	60 (45)	51 (39)
	$V_{71\text{-}4}$							100	54 (35)	59 (44)	59 (46)	49 (39)
$V_{H\text{-}III}$	V_{HBV}								100	78 (66)	77 (67)	53 (31)
	$\psi V_{71\text{-}1}$									100	94 (87)	60 (40)
	$\psi V_{71\text{-}3}$										100	62 (43)
	$\psi V_{71\text{-}6}$											100

The V_H sequences were aligned for maximum homology. The DNA sequence (or amino acid sequence) homologies between V_H coding regions except for the leader sequence are given as percentage. The amino acid sequence homology is shown in parentheses.

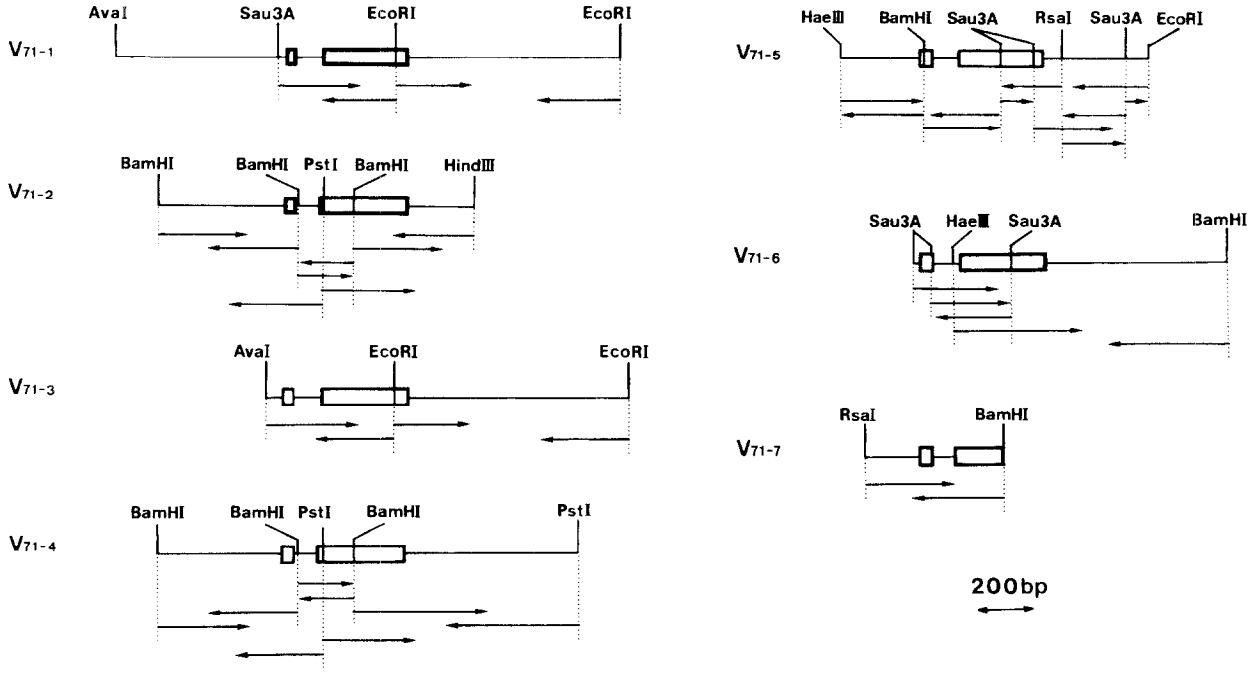

Figure 5. The strategy for sequence analysis of 7 V_H genes in cluster 71. Extents and directions of sequencing are shown by horizontal arrows. Wide rectangles indicate exons.

this work. All the V_H-hybridizing fragments contained the nucleotide sequences closely related to the V_H segments of the probes used (Fig. 6). The nucleotide sequences were homologous (more than 70%) between members of the same families, but much less homologous (less than 60%) between members of different families (Table 2).

Since a high extent of homology (more than 60%) in the amino acid sequence was expected within the same subgroup (Kabat *et al.*, 1983), the amino acid sequences deduced from the nucleotide sequences were also compared with those of their respective V_H probes (Fig. 6(d)). Significant homology was shown between the same family members (more than 60%) but not between members of different families (Table 1). The results again confirmed that the family was directly related to the subgroup and that classification of V_H segments by Southern hybridization was valid.

Although locations of introns are conserved in V_H segments, V_H segments of different families can be distinguished by their lengths of introns as described before (Rechavi *et al.*, 1982; Cohen & Givol, 1983). Two V_H segments ($V_{71\text{-}1}$ and $V_{71\text{-}3}$) contained long introns (103 bp) as expected for family III, whereas the other four V_H segments contained introns of about 80 bp as expected for families I and II.

All V_H segments in this cluster were orientated in the same direction. The 5′-flanking regions of immunoglobulin V genes have been suggested to play a crucial role in determining tissue specificity of V gene expression: a TATA consensus sequence and an octanucleotide sequence are believed to be essential for the correct and tissue-specific expression of V gene (Falkner & Zachau, 1984; Parslow *et al.*, 1984; Foster *et al.*, 1985; Grosschedl & Baltimore, 1985). In fact, the octanucleotide sequence, "ATGCAAAT"†, conserved in heavy-chain genes was found to lie about 110 bp upstream from the coding region in $V_{71\text{-}2}/V_{71\text{-}4}$ and $V_{71\text{-}5}/V_{71\text{-}7}$, although $V_{71\text{-}7}$ has an 11 bp deletion in the 3′-proximal end of the octanucleotide sequence (Fig. 6(a)). These locations of the octanucleotide sequences are consistent with the consensus structure of the 5′ ends of immunoglobulin genes suggested by Parslow *et al.* (1984). The octanucleotide sequences in $V_{CE\text{-}1}$ and V_{HB} lie about 30 bp downstream ($V_{CE\text{-}1}$) and 20 bp upstream (V_{HB}) from the locations of other octamers (Fig. 6(b) and (c)). The TATA consensus sequence was found in $V_{71\text{-}5}$ and $V_{71\text{-}7}$ (Fig. 6(a)), but was not so obvious in V_{HB}, $V_{CE\text{-}1}$, $V_{71\text{-}2}$ and $V_{71\text{-}4}$ (Fig. 6(b) and (c)).

Although the intergenic distances in cluster 71 ranged from 3 kb ($V_{71\text{-}5}/V_{71\text{-}6}$) to 8 kb ($V_{71\text{-}6}/V_{71\text{-}7}$), the average distance between V_H segments in the human genome was calculated to be about 17 kb, on the basis of the total number of the V_H segment and the DNA encompassed in recombinant cosmid clones. This distance seems comparable to those reported for V_H clusters in mouse (8 to 15 kb) (Kemp *et al.*, 1981) and about V_K clusters in human (5 to 25 kb) (Pech *et al.*, 1984).

(c) *Five of seven V_H segments in cluster 71 are pseudogenes*

The sequence analysis showed that five out of seven V_H segments in cluster 71 were pseudogenes (Fig. 6(a), (b) and (c)). This proportion is slightly more abundant than reported previously (40%) in

† Hyphens have been omitted from all sequences for clarity.

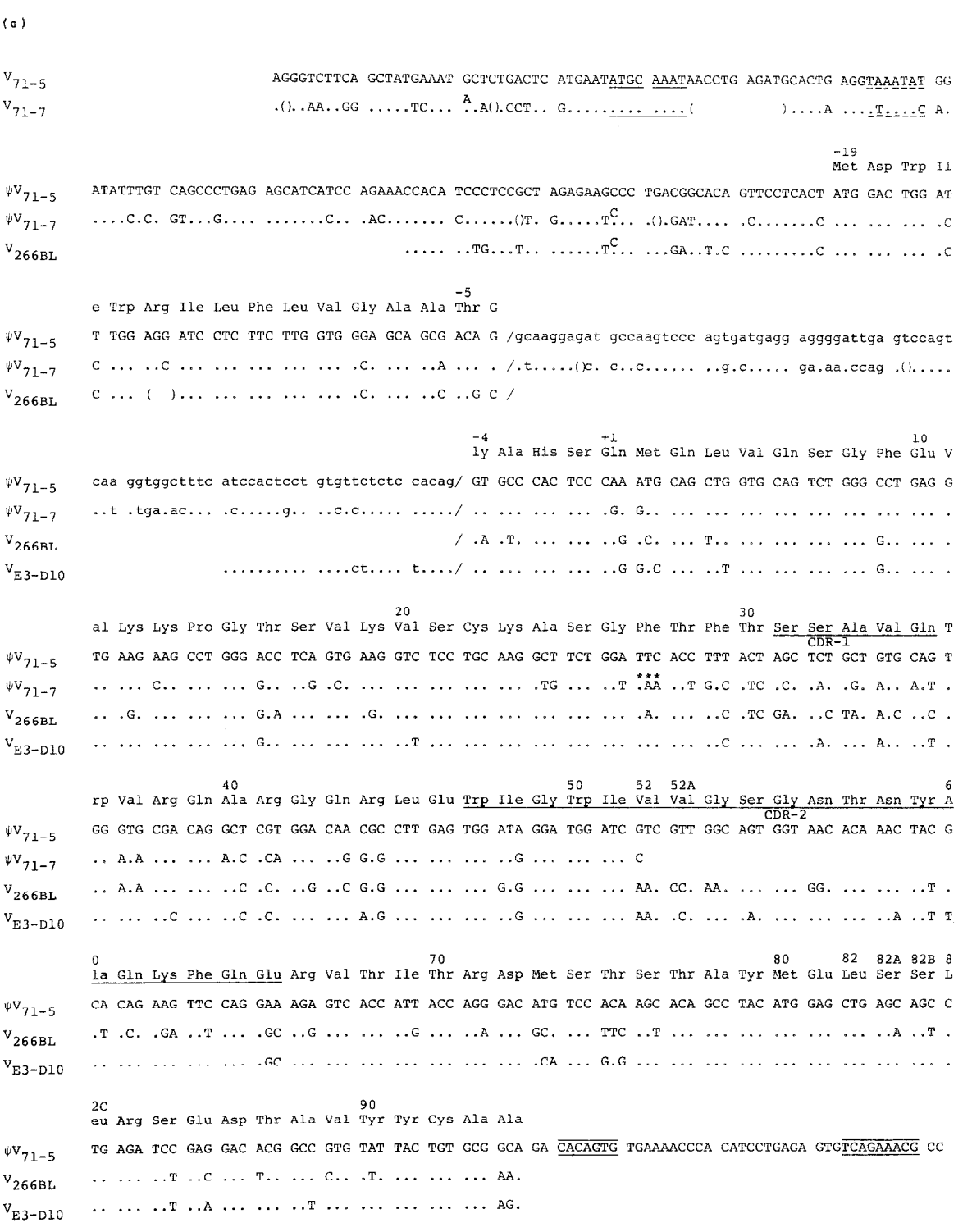

Figure 6. Comparison of human V_H sequences. The nucleotide sequences of 7 V_H segments in cluster 71 are compared with those of probes used in this work. Sequences are aligned to maximize homology with a minimal number of deletions (parentheses) or insertions (inserted bases are placed above the next base). Sequences that align with the protein coding sequence are given in triplets, and the introns are in lower cases. Asterisks indicate nonsense codons. Numbers above the codons refer to amino acid position, according to Kabat *et al.* (1983). Complementarity determining regions (CDR-I and II) and octanucleotide sequences conserved among all known V regions are underlined. The A+T-rich regions (a TATA sequence homologue) are indicated by dotted lines. The signals for V_H–D joining are boxed. Exon and intron junctions are shown by slashes. (a) $V_{H\,I}$ family. (b) $V_{H\text{-}II}$ family. (c) $V_{H\text{-}III}$ family. (d) The amino acid sequences of 7 V_H segments of cluster 71 are compared with those of probes used. The amino acid sequences were deduced from their DNA sequences shown in (a), (b) and (c). Although $V_{71\text{-}3}$ and V_{HBV} have frameshift mutations, putative translations in the original frame are shown. Asterisks indicate nonsense codons. Parentheses mean a deletion or an insertion in the triplet codon. Numbers and invariant codons are according to Kabat *et al.* (1983). CDR-I and II are overlined.

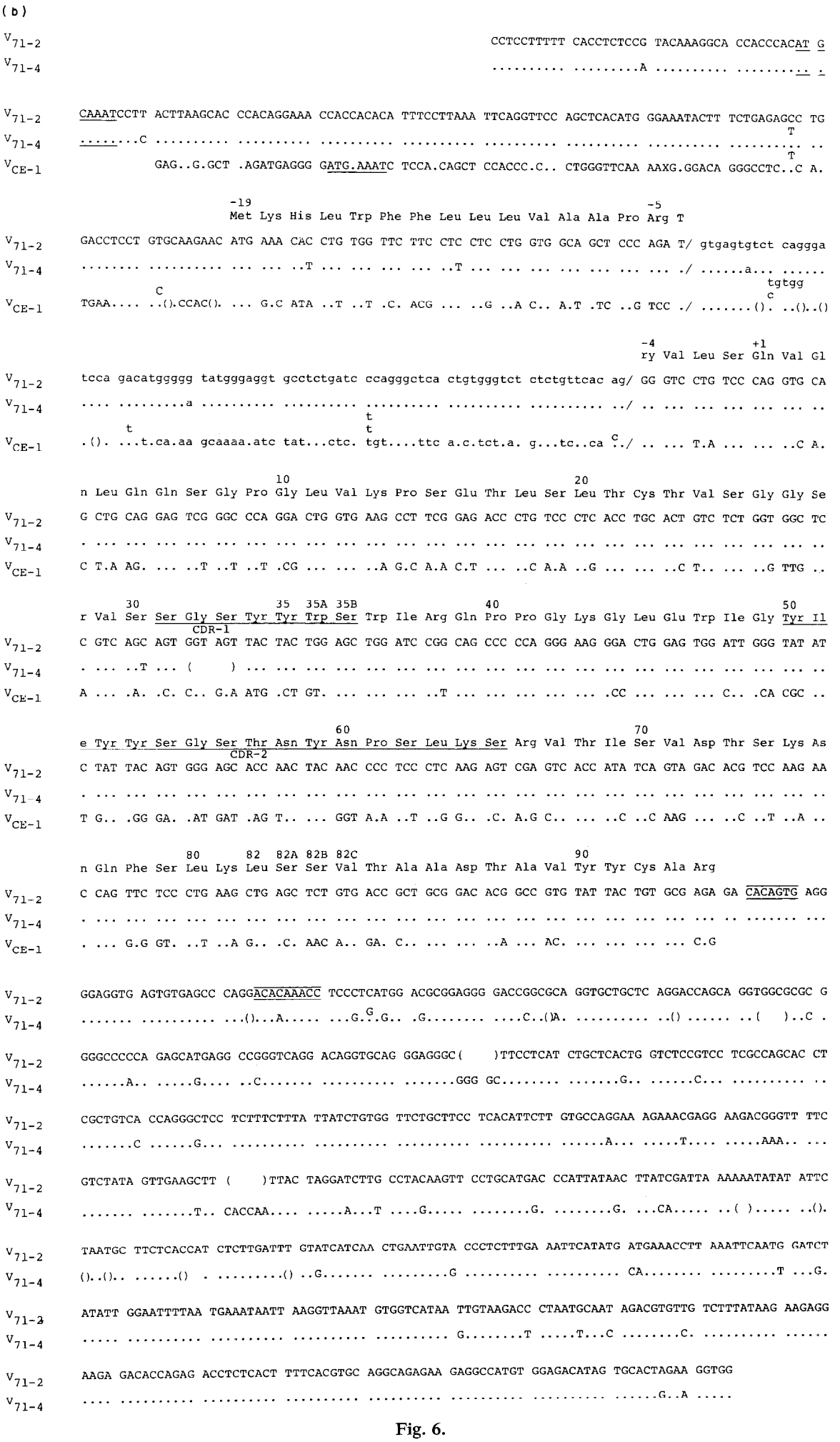
(b)
V71-2 CCTCCTTTTT CACCTCTCCG TACAAAGGCA CCACCCACAT G
V71-4 ....A ....
CAAATCCTT ACTTAAGCAC CCACAGGAAA CCACCACACA TTTCCTTAAA TTCAGGTTCC AGCTCACATG GGAAATACTT TCTGAGAGCC TG
V71-2
V71-4 .....C ....
V CE-1 GAG..G.GCT .AGATGAGGG GATG.AAATC TCCA.CAGCT CCACCC.C.. CTGGGTTCAA AAXG.GGACA GGGCCTC..C A.
-19 -5
Met Lys His Leu Trp Phe Phe Leu Leu Leu Val Ala Ala Pro Arg T
V71-2 GACCTCCT GTGCAAGAAC ATG AAA CAC CTG TGG TTC TTC CTC CTC CTG GTG GCA GCT CCC AGA T/ gtgagtgtct caggga
V71-4 ....T ... ...T ... .../ .....a... .....
V CE-1 TGAA.... ..().CCAC(). ... G.C ATA ..T ..T .C. ACG ... ..G ..A C.. A.T .TC ..G TCC ./ .....(). ..().()
-4 +1
ry Val Leu Ser Gln Val Gl
V71-2 tcca gacatggggg tatgggaggt gcctctgatc ccagggctca ctgtgggtct ctctgttcac ag/ GG GTC CTG TCC CAG GTG CA
V71-4 ....a ..../ .. ..
V CE-1 .(). ...t.ca.aa gcaaaa.atc tat...ctc. tgt....ttc a.c.tct.a. g...tc..ca ./ .. ... T.A ... .C A.
10 20
n Leu Gln Gln Ser Gly Pro Gly Leu Val Lys Pro Ser Glu Thr Leu Ser Leu Thr Cys Thr Val Ser Gly Gly Se
V71-2 G CTG CAG GAG TCG GGC CCA GGA CTG GTG AAG CCT TCG GAG ACC CTG TCC CTC ACC TGC ACT GTC TCT GGT GGC TC
V71-4 . ...
V CE-1 C T.A AG. ... ..T ..T ..T .CG ... .A G.C A.A C.T ... ..C A.A ..G ... .C T.. ... ..G TTG ..
30 35 35A 35B 40 50
r Val Ser Ser Gly Ser Tyr Tyr Trp Ser Trp Ile Arg Gln Pro Pro Gly Lys Gly Leu Glu Trp Ile Gly Tyr Il
CDR-1
V71-2 C GTC AGC AGT GGT AGT TAC TAC TGG AGC TGG ATC CGG CAG CCC CCA GGG AAG GGA CTG GAG TGG ATT GGG TAT AT
V71-4 . ... ..T ... ( ) ... ...
V CE-1 A ... .A. .C. C.. G.A ATG .CT GT. ... ...T ... ... .CC ... ... C.. .CA CGC ..
60 70
e Tyr Tyr Ser Gly Ser Thr Asn Tyr Asn Pro Ser Leu Lys Ser Arg Val Thr Ile Ser Val Asp Thr Ser Lys As
CDR-2
V71-2 C TAT TAC AGT GGG AGC ACC AAC TAC AAC CCC TCC CTC AAG AGT CGA GTC ACC ATA TCA GTA GAC ACG TCC AAG AA
V71-4 . ...
V CE-1 T G.. .GG GA. .AT GAT .AG T.. ... GGT A.A ..T ..G G.. .C. A.G C.. ... ..C ..C AAG ... ..C ..T ..A ..
80 82 82A 82B 82C 90
n Gln Phe Ser Leu Lys Leu Ser Ser Val Thr Ala Ala Asp Thr Ala Val Tyr Tyr Cys Ala Arg
V71-2 C CAG TTC TCC CTG AAG CTG AGC TCT GTG ACC GCT GCG GAC ACG GCC GTG TAT TAC TGT GCG AGA GA CACAGTG AGG
V71-4 . ...
V CE-1 . ... G.G GT. ..T ..A G.. .C. AAC A.. GA. C.. ... .A ... AC. ... ... C.G
V71-2 GGAGGTG AGTGTGAGCC CAGGACACAAACC TCCCTCATGG ACGCGGAGGG GACCGGCGCA GGTGCTGCTC AGGACCAGCA GGTGGCGCGC G
V71-4 ....()...A..... ...G..G.. .G..... ....C..()A. ....() ....( )..C .
V71-2 GGGCCCCCA GAGCATGAGG CCGGGTCAGG ACAGGTGCAG GGAGGGC( )TTCCTCAT CTGCTCACTG GTCTCCGTCC TCGCCAGCAC CT
V71-4 .....A.. .....G.... ..C..... ....GGG GC..... .....G.. .....C... ....
V71-2 CGCTGTCA CCAGGGCTCC TCTTTCTTTA TTATCTGTGG TTCTGCTTCC TCACATTCTT GTGCCAGGAA AGAAACGAGG AAGACGGGTT TTC
V71-4 .....C .....G... ....A... .....T.... .....AAA.. ...
V71-2 GTCTATA GTTGAAGCTT ( )TTAC TAGGATCTTG CCTACAAGTT CCTGCATGAC CCATTATAAC TTATCGATTA AAAAATATAT ATTC
V71-4 ....T.. CACCAA.... .....A...T ....G..... .....G. .....G. ...CA..... ..( )..... ..().
V71-2 TAATGC TTCTCACCAT CTCTTGATTT GTATCATCAA CTGAATTGTA CCCTCTTTGA AATTCATATG ATGAAACCTT AAATTCAATG GATCT
V71-4 ()..()..... .....() . .....() ..G..... .....G .... CA..... .....T ...G.
V71-2 ATATT GGAATTTTAA TGAAATAATT AAGGTTAAAT GTGGTCATAA TTGTAAGACC CTAATGCAAT AGACGTGTTG TCTTTATAAG AAGAGG
V71-4 .... G.....T .....T...C .....C. ....
V71-2 AAGA GACACCAGAG ACCTCTCACT TTTCACGTGC AGGCAGAGAA GAGGCCATGT GGAGACATAG TGCACTAGAA GGTGG
V71-4 ....G..A .....

Fig. 6.

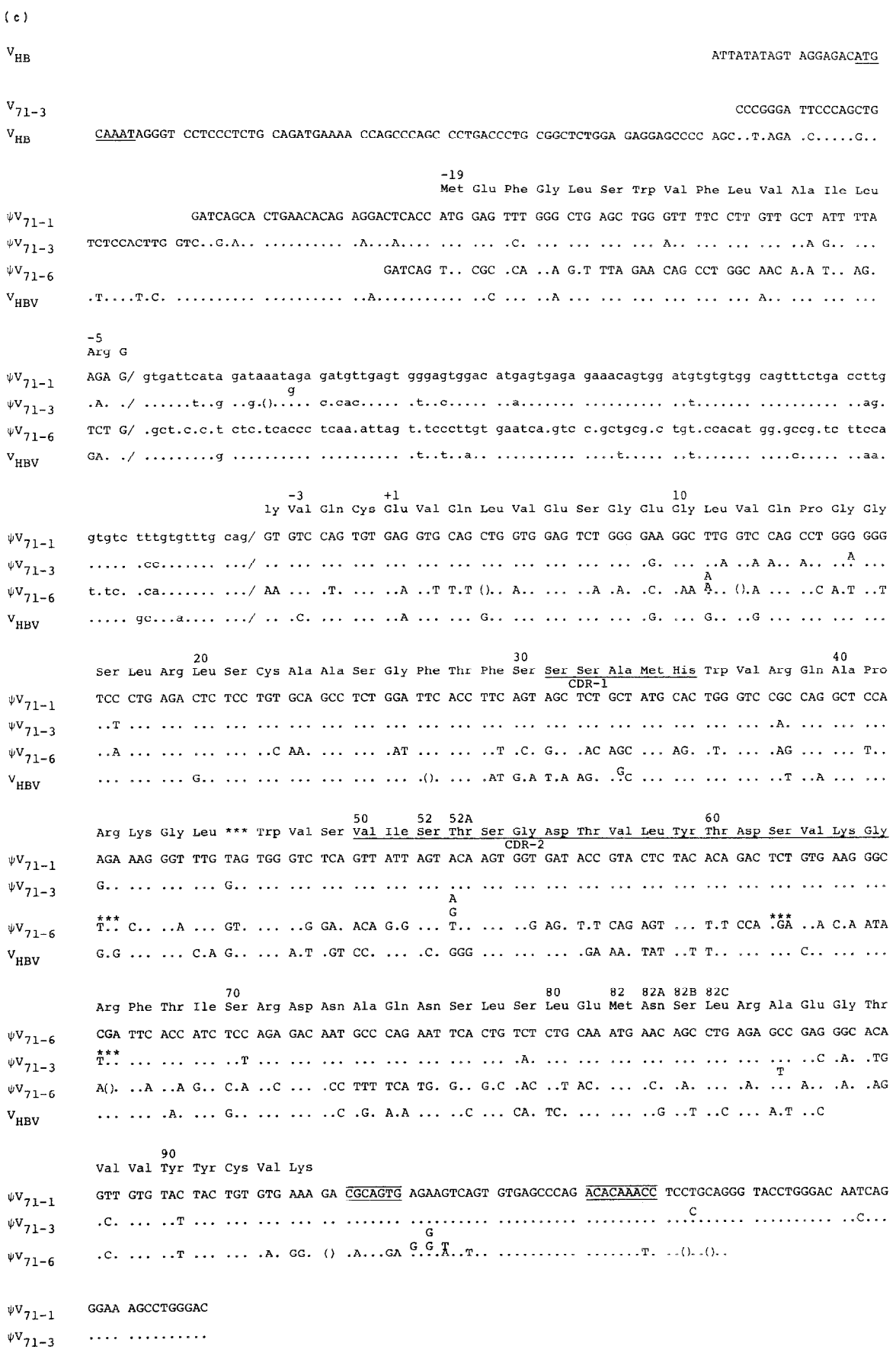
(c)
V_{HB} ATTATATAGT AGGAGACATG
V_{71-3} CCCGGGA TTCCCAGCTG
V_{HB} CAAATAGGGT CCTCCCTCTG CAGATGAAAA CCAGCCCAGC CCTGACCCTG CGGCTCTGGA GAGGAGCCCC AGC..T.AGA .C.....G..
-19
Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Ile Leu
ψV_{71-1} GATCAGCA CTGAACACAG AGGACTCACC ATG GAG TTT GGG CTG AGC TGG GTT TTC CTT GTT GCT ATT TTA
ψV_{71-3} TCTCCACTTG GTC..G.A.. ..... .A...A.... ... ... .C. ... ... A.. ... ...A G.. ...
ψV_{71-6} GATCAG T.. CGC .CA ..A G.T TTA GAA CAG CCT GGC AAC A.A T.. AG.
V_{HBV} .T....T.C. ....A..... ..C ... ..A ... ... A.. ... ...
-5
Arg G
ψV_{71-1} AGA G/ gtgattcata gataaataga gatgttgagt gggagtggac atgagtgaga gaaacagtgg atgtgtgtgg cagtttctga ccttg
ψV_{71-3} .A. ./ .....t..g ..g.()..... c.cac..... .t..c..... ..a..... ....t..... ....ag.
ψV_{71-6} TCT G/ .gct.c.c.t ctc.tcaccc tcaa.attag t.tcccttgt gaatca.gtc c.gctgcg.c tgt.ccacat gg.gccg.tc ttcca
V_{HBV} GA. ./ .....g .... .t..t..a.. ....t..... ..t..... ....c..... ..aa.
-3 +1 10
ly Val Gln Cys Glu Val Gln Leu Val Glu Ser Gly Glu Gly Leu Val Gln Pro Gly Gly
ψV_{71-1} gtgtc tttgtgtttg cag/ GT GTC CAG TGT GAG GTG CAG CTG GTG GAG TCT GGG GAA GGC TTG GTC CAG CCT GGG GGG
ψV_{71-3} ..... .cc..... .../ .. .. .G. ... ..A ..A A.. A.. ..A ...
ψV_{71-6} t.tc. .ca..... .../ AA ... .T. ... ..A ..T T.T ().. A.. ... ..A .A. .C. .AA A.. ().A ... ..C A.T ..T
V_{HBV} ..... gc...a.... .../ .. .C. ... .A ... ... G.. ... ... .G. ... G.. ..G ... ...
20 30 40
Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Ser Ala Met His Trp Val Arg Gln Ala Pro
CDR-1
ψV_{71-1} TCC CTG AGA CTC TCC TGT GCA GCC TCT GGA TTC ACC TTC AGT AGC TCT GCT ATG CAC TGG GTC CGC CAG GCT CCA
ψV_{71-3} ..T ... ... .A. ... ...
ψV_{71-6} ..A ... ...C AA. ... ... .AT ... .T .C. G.. .AC AGC ... AG. .T. ... .AG ... ... T..
V_{HBV} ... ... G.. ... ...(). ... .AT G.A T.A AG. ..C ... ...T ..A ... ...
50 52 52A 60
Arg Lys Gly Leu *** Trp Val Ser Val Ile Ser Thr Ser Gly Asp Thr Val Leu Tyr Thr Asp Ser Val Lys Gly
CDR-2
ψV_{71-1} AGA AAG GGT TTG TAG TGG GTC TCA GTT ATT AGT ACA AGT GGT GAT ACC GTA CTC TAC ACA GAC TCT GTG AAG GGC
ψV_{71-3} G.. ... ... G.. ... ...
ψV_{71-6} T.. C.. ..A ... GT. ... ..G GA. ACA G.G ... T.. ... ..G AG. T.T CAG AGT ... T.T CCA .GA ..A C.A ATA
V_{HBV} G.G ... ... C.A G.. ... A.T .GT CC. ... .C. GGG ... ... .GA AA. TAT ..T T.. ... ... C.. ... ...
70 80 82 82A 82B 82C
Arg Phe Thr Ile Ser Arg Asp Asn Ala Gln Asn Ser Leu Ser Leu Glu Met Asn Ser Leu Arg Ala Glu Gly Thr
ψV_{71-6} CGA TTC ACC ATC TCC AGA GAC AAT GCC CAG AAT TCA CTG TCT CTG CAA ATG AAC AGC CTG AGA GCC GAG GGC ACA
ψV_{71-3} T.. ... ...T ... ... .A. ... ...C .A. .TG
ψV_{71-6} A(). ..A ..A G.. C.A ..C ... .CC TTT TCA TG. G.. G.C .AC ..T AC. ... .C. .A. ... .A. ... A.. .A. .AG
V_{HBV} ... ... .A. ... G.. ... .C .G. A.A ... ..C ... CA. TC. ... .G ..T ..C ... A.T ..C
90
Val Val Tyr Tyr Cys Val Lys
ψV_{71-1} GTT GTG TAC TAC TGT GTG AAA GA CGCAGTG AGAAGTCAGT GTGAGCCCAG ACACAAACC TCCTGCAGGG TACCTGGGAC AATCAG
ψV_{71-3} .C. ... ..T ... ...C...
ψV_{71-6} .C. ... ..T ... ... .A. GG. () .A...GA ...A..T.. ....T. ..()..()..
ψV_{71-1} GGAA AGCCTGGGAC
ψV_{71-3} ....

Fig. 6.

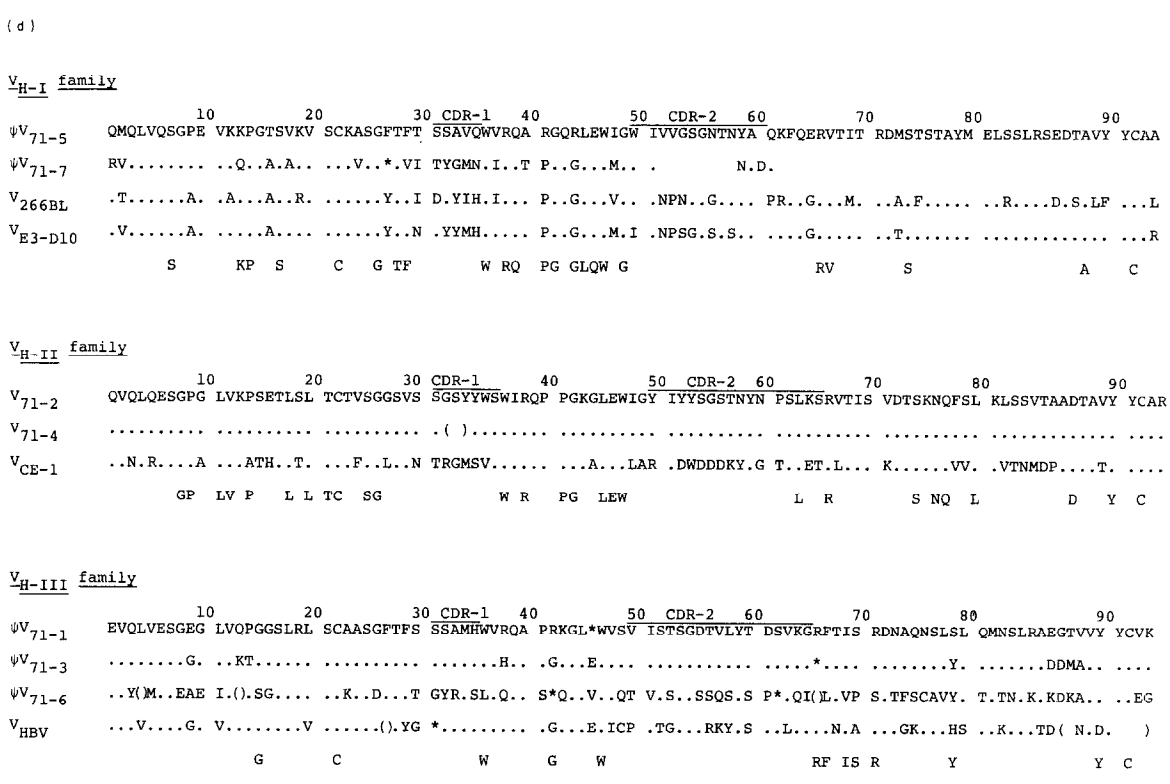

Fig. 6.

both mouse and man (Bothwell *et al.*, 1981; Huang *et al.*, 1981; Givol *et al.*, 1981; Rechavi *et al.*, 1982, 1983; Loh *et al.*, 1983; Cohen & Givol, 1983). The V_{71-1} and V_{71-7} segments were pseudogenes because they contained termination codons (V_{71-1} at position 46, V_{71-7} at position 27). Comparison of the V_{71-3} and V_{71-1} sequences indicated that V_{71-3} contained one base insertion between the second and third bases at codon position 15, resulting in a change of the reading frame. The V_{71-3} segment had another mutation at position 66, resulting in termination of the original reading frame. Both V_{71-1} and V_{71-3} segments had a point mutation at the same position in the heptamer signal for V_H–D–J_H recombination. V_{71-5} had an abnormal splicing signal AG–GC at the 5′ end of the intron and two base substitutions in the nonamer signals for recombination. These four pseudogenes in cluster 71 carried only a few point mutations, including one base insertion, and still maintained the extensive homology with the active V_H segment.

In contrast, we found a diverged V_H pseudogene, V_{71-6}, in the human V_H locus. Since V_{71-6} hybridized weakly with the V_H probe of the V_{H-III} family, the V_{71-6} sequence was aligned with those of the V_{H-III} family by introduction of insertions and deletions. Little homology was found except for the 5′ half of the coding exon (Fig. 6(c)). V_{71-6} segments accumulated a relatively large number of deleterious mutations throughout the segment. These mutations not only altered the translation reading frame of the segment, but also created termination codons in every frame. The deleterious mutations in five V_H pseudogenes are summarized in Table 3.

4. Discussion

(a) *Mutations in germline V_H segments*

Comparison of the V_H segments within each family indicated that the extents of the mutations varied among various regions of the active V_H segments. Mutations were classified into those replacing amino acid change (replacement) and those giving no change (silent). Generally, the extent of diversity is greater in the CDR of the active V_H segments than in the FR (Table 4). However, the proportion of silent substitutions in the total substitutions is lower in the CDR than in the FR. According to the codon table the proportion of silent substitutions among the total substitutions is about 0·3. The silent substitution in the FR of the active V_H segments is always higher than 30% of the total substitution. Lower mutation frequency in the replacement position of the FR indicates that they are under some selection pressure at the protein level. A similar observation was reported in murine V_H segments (Baltimore & Loh, 1983; Cohen & Givol, 1983).

On the other hand, the silent substitution in the CDR of the active V_H segments is always slightly lower than 30% of the total substitution. This may be because the CDR is completely free from selective constraint. Alternatively, mutations in the replacement position of the CDR might be positively selected. A similar phenomenon was also

Table 3
Deleterious mutations in 5 V_H pseudogenes of cluster 71

Gene	Mutation	Position (codon no.)	Affected function
$\psi V_{71\text{-}1}$	GAG(Glu) → TAG(ter)	46	Terminator
	CACAGTG → CGCAGTG	Recombination signal	V_HDJ_H recombination
$\psi V_{71\text{-}3}$	1 bp insertion	35	Frameshift
	CGA(Arg) → TGA	66	Terminator
	CACAGTG → CGCAGTG	Recombination signal	V_HDJ_H recombination
$\psi V_{71\text{-}5}$	AG/GT → AG/GC	5′ Exon/intron border	RNA splicing
	ACAAAAACC → TCAGAAACG	Recombination signal	V_HDJ_H recombination
$\psi V_{71\text{-}6}$	ATG(Met) → TTG(Leu)	−19	Translation initiation
	AG/GT → TG/GG	5′ Exon/intron	RNA splicing
	1 bp deletion	4	Frameshift
	2 bp insertion	10–11	Frameshift
	1 bp deletion	12	Frameshift
	AGA(Arg) → TGA(ter)	42	Terminator
	2 bp insertion	52–52A	Frameshift
	TCT(Ser) → TGA(ter)	62	Terminator
	1 bp deletion	66	Frameshift
	1 bp insertion	84	Frameshift
	CACAGTG → CACAGGA	Recombination signal	V_HDJ_H recombination
$\psi V_{71\text{-}7}$	TAC(Tyr) → TAA(ter)	27	Terminator

The deleterious mutations of $V_{71\text{-}6}$ were determined in the putative alignment as in Fig. 4(a). ter, a stop codon.

observed in murine V_H segments (Baltimore & Loh, 1983).

In contrast, the proportions of the silent substitution in the CDR and FR are almost identical (about 30%) in $V_{71\text{-}6}$, the most diverged pseudogene, which we assume has escaped from either positive or negative selection. Since the relative distributions of mutations in the replacement and silent positions of the active V_H segments are different from those of the most diverged $V_{71\text{-}6}$ that may not be under any selection pressure, there may have been some positive selection for the replacement substitution in CDRs of the active V_H segments.

Table 4
Percentage substitution in V_H gene segments

		DNA sequence differences								
		% Substitution/base			% Silent substitutions/ total substitutions			Amino acid difference (%)		
Family	V_H segments compared	FR	CDR	Total	FR	CDR	Total	FR	CDR	Total
$V_{H\text{-}I}$	$\psi V_{71\text{-}5}/\psi V_{71\text{-}7}$	(17·5)	(36·4)	(21·4)	(44·7)	(12·5)	(36·4)	(31·8)	(71·4)	(37·3)
	$\psi V_{71\text{-}5}/V_{266BL}$	16·7	45·5	21·6	42·1	26·9	35·9	26·3	50·0	31·6
	$\psi V_{71\text{-}7}/V_{266BL}$	(19·7)	(31·8)	(21·4)	(40·9)	(9·1)	(31·7)	(27·3)	(71·4)	(33·3)
	$\psi V_{71\text{-}5}/V_{E3\text{-}D10}$	10·0	17·3	11·8	59·1	15·4	42·9	13·2	54·5	22·4
	$\psi V_{71\text{-}7}/V_{E3\text{-}D10}$	(19·5)	(12·9)	(18·8)	(44·0)	(0)	(37·9)	(22·7)	(57·1)	(27·5)
	$V_{266BL}/V_{E3\text{-}D10}$	16·2	31·8	19·4	37·1	18·2	29·8	19·7	45·5	25·5
$V_{H\text{-}II}$	$V_{71\text{-}2}/V_{71\text{-}4}$	0·4	2·9	1·0	100	0	33·3	0	13·0	3·0
	$V_{71\text{-}2}/V_{CE\text{-}1}$	30·4	53·6	35·4	44·1	27·0	38·1	31·6	78·3	42·4
	$V_{71\text{-}4}/V_{CE\text{-}1}$	30·9	50·7	35·0	43·5	28·6	38·5	31·6	78·3	42·4
$V_{H\text{-}III}$	$\psi V_{71\text{-}1}/\psi V_{71\text{-}3}$	8·8	0·0	6·4	31·6	—	31·6	17·1	0	13·3
	$\psi V_{71\text{-}1}/\psi V_{71\text{-}6}$	34·3	58·8	39·5	30·4	32·5	31·1	56·6	72·7	60·2
	$\psi V_{71\text{-}3}/\psi V_{71\text{-}6}$	32·0	58·8	38·4	25·0	32·5	27·6	52·6	72·7	57·1
	$\psi V_{71\text{-}1}/V_{HBV}$	(19·9)	34·8	(22·4)	(44·7)	22·7	(36·7)	(31·3)	40·9	(33·7)
	$\psi V_{71\text{-}3}/V_{HBV}$	(20·9)	31·8	(23·4)	(43·9)	22·7	(36·5)	(29·9)	40·9	(32·6)
	$\psi V_{71\text{-}6}/V_{HBV}$	(41·9)	63·8	(47·1)	(28·2)	31·8	(29·4)	(64·2)	77·3	(67·4)

The percentage substitution was calculated as a comparison between the pair indicated. The silent substitution is the % ratio of silent substitutions to total replacements. "Total" means a % substitution in the V_H coding region, except the leader sequence. The parentheses indicate an incomplete comparison.

(b) *The significance of pseudogenes in the human V_H locus*

We and others have indicated that the V_H locus is rich (up to 40%) in pseudogenes (Bothwell *et al.*, 1981; Huang *et al.*, 1981; Givol *et al.*, 1981; Rechavi *et al.*, 1982, 1983; Cohen & Givol, 1983; Loh *et al.*, 1983; Takahashi *et al.*, 1984). A remarkable feature of V_H pseudogenes is that many of them are not very different from active V_H segments (Bothwell *et al.*, 1981; Huang *et al.*, 1981; Givol *et al.*, 1981). Five out of seven V_H segments in cluster 71 are pseudogenes (about 70%), and four out of five genes carry only one or two point mutations. It is worth noting that the relative distributions of the silent mutation in the FR and CDR in these "conserved pseudogenes" are very similar to those of the active V_H segments, as if some selection constraints operated in these pseudogenes. A similar observation was made in the mouse V_H segment (Schiff *et al.*, 1985).

Cohen & Givol (1983) described one highly drifted V_H gene ($\psi V_H 124$) in mouse. We have identified two diverged pseudogenes; the $V_{71\text{-}6}$ segment, and the $V_{CE\text{-}114}$ segment, which has two large deletions and two stop codons (Takahashi *et al.*, 1984). The V_H pseudogenes seem to be classified into two distinct groups; those with a few mutations (conserved) and those with drastic changes (diverged). This suggests that there might be a threshold for divergence that could accept the correction mechanism. Once mutations in a V_H pseudogene exceed this threshold, the V_H pseudogene would diverge rapidly like many other pseudogenes. The "diverged pseudogene" ($V_{71\text{-}6}$) has apparently random distribution of mutations at replacement and silent positions in both the FR and CDR, in agreement with the assumption that the diverged pseudogene is free from selection and correction.

Since it was hard to imagine the functional selection pressure on the pseudogenes, involvement of some correction mechanisms such as gene conversion or unequal crossing-over was proposed (Baltimore, 1981; Loh *et al.*, 1983; Cohen & Givol, 1983). More recently, somatic gene conversion was suggested to take place (Reynaud *et al.*, 1985). We propose another possibility, that somatic mutations could revive conserved pseudogenes with terminator mutations. If this could happen, conserved V_H pseudogenes would not be very different from active V_H segments and would be under selection pressure at the protein level. To explain conservation of the $V_{71\text{-}1}$, $V_{71\text{-}3}$, $V_{71\text{-}5}$ and $V_{71\text{-}7}$ segments by this mechanism we have to assume that $V_{71\text{-}1}$, $V_{71\text{-}3}$, $V_{71\text{-}5}$ and $V_{71\text{-}7}$ are able to undergo V–D–J recombination, as somatic mutation takes place only in the rearranged V_H gene (Gorski *et al.*, 1983; Gearhart & Bogenhagen, 1983). Although pseudogenes are generally considered evolutionary relics of once intact genes, the maintenance of the "conserved" V_H pseudogenes suggests that they may still have some role in the V_H gene repertoire.

(c) *Internal duplication in cluster 71*

Examination of the restriction map of cluster 71 showed the presence of the tandem homology units, each containing two V_H segments of the $V_{H\text{-}III}$ and $V_{H\text{-}II}$ families: $V_{71\text{-}1}/V_{71\text{-}2}$ and $V_{71\text{-}3}/V_{71\text{-}4}$ (Fig. 4). The homology units extend to 1 to 5 kb to the right of the $V_{71\text{-}2}$ and $V_{71\text{-}4}$ segments, on the basis of the fine restriction mapping and partial sequence determination (Fig. 6(b)). The left end of the homology unit was not clear. DNA sequence homology of each pair ($V_{71\text{-}1}/V_{71\text{-}3}$ or $V_{71\text{-}2}/V_{71\text{-}4}$) is up to 90%, which is much higher than the homology to other published V_H segments belonging to the same V_H family. We can estimate the divergence time of these V_H segments by the molecular clock of the silent substitution in the coding region as described by Sakoyama *et al.* (1986). The duplication event was calculated to have taken place 15 million years ago and 10 million years ago on the basis of the homology of $V_{71\text{-}1}/V_{71\text{-}3}$ and of $V_{71\text{-}2}/V_{71\text{-}4}$, respectively. We assumed that the pseudogenes diverged 1·9 times faster than active genes.

Comparison between $V_{71\text{-}2}$ and $V_{71\text{-}4}$, which were similar in sequence, revealed that 6 bp was deleted from CDR-1 of $V_{71\text{-}4}$. This deletion in CDR-1 may have some influence on the V_H structure and may contribute to expressing a new specificity for antigen recognition. $V_{71\text{-}1}$ and $V_{71\text{-}3}$ had the same mutation at the recombination heptamer signal (Table 3), suggesting that the duplication event occurred after this mutation.

(d) *Did the V_H loci of mouse and man evolve in different manners?*

Studies on the organization of the mouse V_H segment have suggested that a family of related V_H segments constitutes a physically linked cluster separated from another V_H family and that the two different families are not intermingled in the phage clones identified to date (Kemp *et al.*, 1981; Givol *et al.*, 1981; Brodeur & Riblet, 1984). It is, therefore, reasonable to assume that the first step in the evolution of the murine V_H locus may have been the formation of a primordial V_H segment for each family. Subsequent amplification of each V_H segment may have lead to the formation of clusters of closely related V_H segments.

On the other hand, V_H segments of different families are completely interdigitated in the human V_H locus. There are two possibilities to explain the marked difference in the V_H segment organization in man and mouse. In one case, V_H loci of the two species evolved independently. Since most distant V_H segments in mouse or human evolved before mammalian radiation (about 75 million years), this could have happened during a process of recent expansion and contraction of the V_H loci as demonstrated among various strains of mice (Kataoka *et al.*, 1982). In the other case, the murine type of organization of the V_H segment was

converted to the human type of organization. To make the reverse conversion is very difficult.

Either case requires rather rapid and drastic genetic events such as series of duplication, recombination and translocation. We have already obtained molecular genetic evidence that the human immunoglobulin H-chain locus underwent drastic rearrangements recently: (1) the human C_H locus has the region containing two C_γ, one C_ε and one C_α genes duplicated (Flanagan & Rabbits, 1982; Nishida *et al.*, 1982); and (2) cluster 71 contains duplicated regions containing two V_H segments. Pech & Zachau (1984) showed that the human V_K locus contained "mixed V_K segment clusters" like the human V_H locus. They assumed that the construction of a "mixed cluster" required transposition-like events before the final duplication of the gene cluster.

Two classes of relatively conserved sequences have been reported near mouse V_H segments (Kemp *et al.*, 1981): one is the sequence immediately flanking the V_H segments and the other is the repeated element, both of which can be expected to facilitate recombination. Sequences found in the 5′- and 3′-flanking regions of V_K segments can form a stem-and-loop structure including a V_K segment *per se* (Pech *et al.*, 1984). These sequences were proposed to be responsible for the gene-conversion or transposition-like events and hence to modify the germline V_H or V_K repertoire (Kemp *et al.*, 1981; Pech *et al.*, 1984). However, it is hard to determine whether they are cause or result. No such sequence was identified in V_H segments studied in this work.

We are grateful to Ms K. Hirano for preparation of the manuscript and to Ms K. Harado and Ms Y. Ishida for excellent technical assistance. This work was supported by Ministry of Education, Science and Culture of Japan.

References

Baltimore, D. (1981). *Cell*, **24**, 592–594.

Baltimore, D. & Loh, D. Y. (1983). In *Progress in Immunology V* (Yamamura, Y. & Tada, T., eds), pp. 115–121, Academic Press, New York.

Bentley, D. L. & Rabbitts, T. H. (1981). *Cell*, **24**, 613–623.

Birnboim, H. C. & Doly, J. (1979). *Nucl. Acids Res.* **7**, 1513–1523.

Blin, N. & Stafford, D. W. (1976). *Nucl. Acids Res.* **3**, 2303–2308.

Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981). *Cell*, **24**, 635–637.

Brodeur, P. & Riblet, R. (1984). *Eur. J. Immunol.* **14**, 922–930.

Cohen, J. B. & Givol, D. (1983). *EMBO J.* **2**, 1795–1800.

Cory, S., Tyler, B. M. & Adams, J. M. (1981). *J. Mol. Appl. Genet.* **1**, 103–116.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980). *Cell*, **19**, 981–992.

Falkner, F. G. & Zachau, H. G. (1984). *Nature (London)*, **310**, 71–74.

Flanagen, J. G. & Rabbitts, T. H. (1982). *Nature (London)*, **300**, 709–713.

Foster, J., Stafford, J. & Queen, C. (1985). *Nature (London)*, **315**, 423–425.

Gearhart, P. J. & Bogenhagen, D. G. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 3439–3443.

Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. & Cohen, J. B. (1981). *Nature (London)*, **292**, 426–430.

Gorski, J., Rollini, P. & Mach, B. (1983). *Science*, **220**, 1179–1181.

Grosschedl, R. & Baltimore, D. (1985). *Cell*, **41**, 885–897.

Grosveld, F. G., Dahl, H. H. M., de Boer, E. & Flavell, R. A. (1981). *Gene*, **13**, 227–237.

Hanahan, D. & Meselson, M. (1980). *Gene*, **10**, 63–67.

Hattori, M. & Sakaki, Y. (1986). *Anal. Biochem.* **152**, 232–238.

Honjo, T. (1983). *Annu. Rev. Immunol.* **1**, 499–528.

Honjo, T. & Habu, S. (1985). *Annu. Rev. Biochem.* **54**, 803–830.

Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979). *Cell*, **18**, 559–568.

Huang, H., Crews, S. & Hood, L. (1981). *J. Mol. Appl. Genet.* **1**, 93–101.

Ish-Horowicz, D. & Burke, J. F. (1981). *Nucl. Acids Res.* **9**, 2989–2998.

Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. (1983). *Sequences of Proteins of Immunological Interest*, NIH publication, Washington, D.C.

Kataoka, T., Miyata, T. & Honjo, T. (1981). *Cell*, **23**, 357–368.

Kataoka, T., Nikaido, T., Miyata, T., Moriwaki, K. & Honjo, T. (1982). *J. Biol. Chem.* **257**, 277–285.

Kemp, D. J., Tyler, B., Bernard, O., Gough, N., Gerondakis, S., Adams, J. M. & Cory, S. (1981). *J. Mol. Appl. Genet.* **1**, 245–261.

Kenten, J. H., Molgoard, H. V., Houghteon, M., Dehbyshire, R. B., Viney, J., Bell, L. O. & Gould, H. J. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 6661–6665.

Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. (1981). *Cell*, **27**, 573–581.

Loh, D. Y., Bothwell, A. L. M., White-Scharf, M. E., Imanishi-Kari, T. & Baltimore, D. (1983). *Cell*, **33**, 85–93.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W. & Weigert, M. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 3180–3184.

Messing, J. (1983). *Methods Enzymol.* **101**, 20–78.

Nishi, M., Kataoka, T. & Honjo, T. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 6399–6403.

Nishida, Y., Miki, T., Hisajima, H. & Honjo, T. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 3833–3837.

Noma, Y., Yaoita, Y., Matsunami, N., Rosen, A., Klein, G. & Honjo, T. (1984). *Mol. Biol. Med.* **2**, 337–350.

Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 2650–2654.

Pech, M. & Zachau, H. G. (1984). *Nucl. Acids Res.* **12**, 9229–9236.

Pech, M., Jaenichen, H.-R., Pohlenz, H.-D., Neumaizer, P. S., Klobeck, H.-G. & Zachau, H. G. (1984). *J. Mol. Biol.* **176**, 189–204.

Rechavi, G., Bientz, B., Ram, D., Ben-Neriah, Y., Cohen, J. B., Zakut, R. & Givol, D. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 4405–4409.

Rechavi, G., Ram, D., Glazer, L., Zakut, R. & Givol, D. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 855 859.

Reynaud, C.-A., Anquez, V., Dahan, A. & Weil, J.-C. (1985). *Cell*, **40**, 283–291.

Sakano, H., Maki, R., Kurosawa, Y., Roader, W. & Tonegawa, S. (1980). *Nature (London)*, **286**, 676–683.

Sakoyama, Y., Hong, K.-J., Byun, S. M., Hisajima, H., Ueda, S., Yaoita, Y., Hayashida, H., Miyata, T. & Honjo, T. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, in the press.

Sanger, F., Coulson, B. G., Smith, A. J. H. & Roa, B. A. (1980). *J. Mol. Biol.* **143**, 161–178.

Schiff, C., Milili, M. & Fougereau, M. (1985). *EMBO J.* **4**, 1225–1230.

Southern, E. M. (1975). *J. Mol. Biol.* **98**, 503–517.

Takahashi, N., Nakai, S. & Honjo, T. (1980). *Nucl. Acids Res.* **24**, 5983–5991.

Takahashi, N., Noma, T. & Honjo, T. (1984). *Proc. Nat. Acád. Sci., U.S.A.* **81**, 5194–5198.

Tonegawa, S. (1983). *Nature (London)*, **302**, 575–581.

Yancopoulos, G. D., Desiderio, S. V., Paskind, M., Kearney, J. F., Baltimore, D. & Alt, F. W. (1984). *Nature (London)*, **311**, 727–733.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). *Gene*, **33**, 103–119.

Yaoita, Y. & Honjo, T. (1980). *Biomed. Res.* **1**, 164–175.

Wood, C. & Tonegawa, S. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 3030–3034.

Wu, T. T. & Kabat, E. A. (1970). *J. Exp. Med.* **132**, 211–250.

Edited by P. Chambon