The EMBO Journal vol.4 no.5 pp.1225-1230, 1985

# Functional and pseudogenes are similarly organized and may equally contribute to the extensive antibody diversity of the IgV$_H$II family

Claudine Schiff, Michèle Milili and Michel Fougereau

Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, Case 906, 13288 Marseille Cédex 9, France

Communicated by M.Fougereau

Eleven germ-line immunoglobulin V_H genes have been isolated from a BALB/c genomic library, using a cDNA probe specific for the GAT/NP^a variable region. Restriction fragments of all genes were sequenced: two over 800 bp, covering signals of the 5'- and 3'-non-coding regions, three encompassing the complete coding region and part of the 5', the remaining sequences covering most of the V coding region. All sequences pertained to the V$_H$II family, and were compared with the other 13 homologous genes already published. Characteristic features defining the family are clearly visible all along the sequences analyzed, including the 5'-non-coding region, the leader fragment and the intron organization. About half of the compared genes have pseudogene characteristics, defined either by a stop codon in the coding region or the lack of an initiator codon in the leader segment. Analysis of the replacement mutations, as compared with silent ones, indicate that they are highly clustered in complementarity determining regions, for both the functional and the pseudogenes, suggesting that all genes have been submitted to similar selective pressure, and that the pseudogene repertoire may be actively used, by recombination and/or conversion process. Signals that regulate transcription are highly conserved through the family barriers. The V$_H$II group is the largest Ig V genes family, with extreme sequence divergences reaching 22% nucleotide differences. As no two genes were found identical out of the 24 members compared, and as two genes were found to differ by as little as three nucleotides, it seems that the previous estimate of 60 members might be much too low.

Key words: antibody diversity/immunoglobulin V$_H$II family/ pseudogenes

Introduction

Variable regions of immunoglobulin chains are encoded by gene segments which are separated in the germ-line, and which must rearrange in B cells before expression: V_L and J_L for the light chains (Brack *et al.*, 1978; Seidman *et al.*, 1979; Sakano *et al.*, 1979), V_H, D and J_H for the heavy chains (Early *et al.*, 1980; Sakano *et al.*, 1981). Diversity is further amplified by junctional ambiguities (Seidman *et al.*, 1979; Sakano *et al.*, 1979; Alt and Baltimore, 1982), somatic mutations (Bothwell *et al.*, 1981; Crews *et al.*, 1981; Griffiths *et al.*, 1984) and gene conversion (Seidman *et al.*, 1978; Schrier *et al.*, 1981; Clarke *et al.*, 1982b; Krawinkel *et al.*, 1983; Ollo and Rougeon, 1983). Since both the V_H and V_L regions participate in the antibody combining site (Amzel *et al.*, 1974), and because of random or semi-random pairing of H and L chains (de Préval and Fougereau, 1976), the overall number of discrete antibody molecules that may be generated is enormous.

That part of the diversity which is strictly germ-line encoded, and which allows a constant control to selective pressure, ultimately relies on the number of V_H and V_L genes. An approach to this problem has been attempted by estimating the number of DNA genomic restriction fragments that would be detected by specific V probes in Southern blots (Matthyssens and Rabbitts, 1980; Bentley and Rabbitts, 1981; Cory *et al.*, 1981; Brodeur and Riblet, 1984). Used in combination with a genetic analysis of Ig families, this allowed an estimate of ~100 genes for the mouse germ-line V_H system (Brodeur and Riblet, 1984). This number seems in agreement with sequence arguments (Dildrop, 1984), but must take into account the relatively large proportion of pseudogenes in immunoglobulin families (Givol *et al.*, 1981), although the latter may contribute significantly to diversity, through mechanisms such as gene conversion (Clarke *et al.*, 1982b; Krawinkel *et al.*, 1983).

In this paper, we report the isolation of 11 V_H germ-line genes that belong to the V$_H$II family (Kabat *et al.*, 1983) or group 1 as proposed by Dildrop (1984). Nucleotide sequences have been determined on most of the coding region of the genes, and extended to 5'- and 3'-non-coding sections for several of them. This brings to a total of 24 the number of different genes that have been sequenced in the BALB/c V$_H$II family, and permits a deeper insight to the organization of the germ-line gene diversity of immunoglobulins.

Results and Discussion

Isolation of V_H-containing clones

Twenty seven clones, out of 300 000 from the genomic library prepared in Charon 4A, gave a positive signal with the V_H (H IV 92, Schiff *et al.*, 1983) cDNA-specific probe. Twenty-four were identified as different V_H on the basis of their restriction maps. As one clone contained two V_H genes, the total number of isolated genes amounted to 25. Eleven clones that gave a strong positive signal were further analyzed (Table I). Overall sequence strategy appears in Figure 1. Two genes (H24 and H26-1) were

Table I. Characteristics of restriction fragments containing V_H genes (in kb)

	*Eco*RI	*Eco*RI + *Hind*III
H24	5.2	2.3
H26-1	3.8	1.2
H13-3	1.4	1.4
H30	3.2	1.3
H16	15	10
H26-6	7.5	6
H13-1	9.5	4
H18	6.7	5.7
H17	2.9	0.5
H8	8.2	4.2
H9	8.5	4

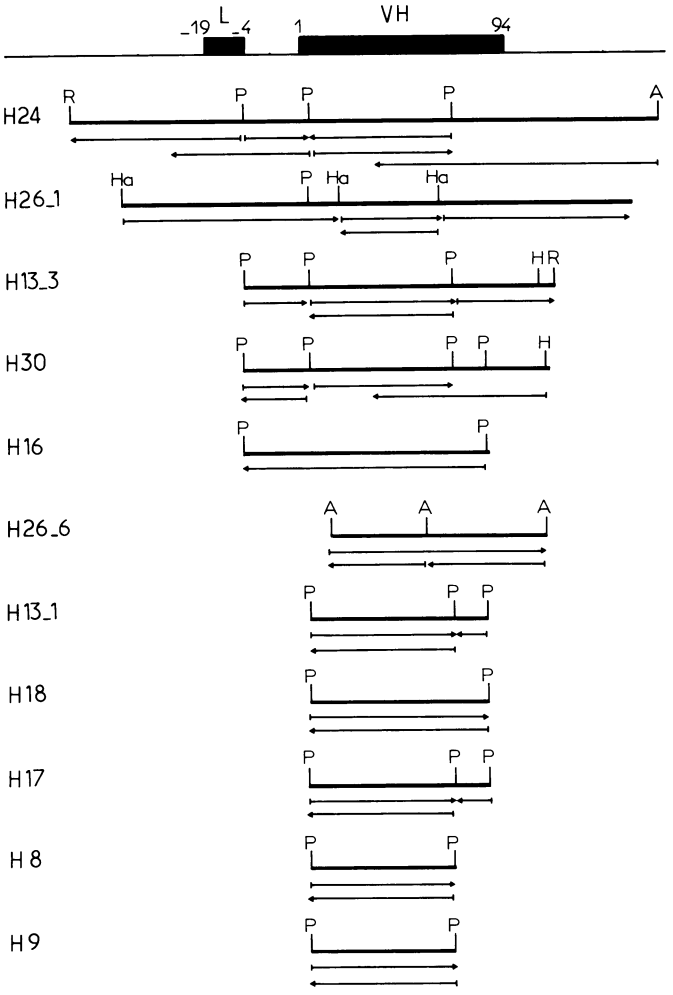

Fig. 1. Strategy used for sequencing the 11 V_H genes. Coding sections (L-leader- and V_H) are delineated on top of figure. Restriction sites: A (*Alu*I), H (*Hind*III), Ha (*Hae*III), P (*Pst*I) and R (*Eco*RI). Arrows indicate stretches that have been sequenced by the dideoxy method, in M13-derived vector.

sequenced over 800 bp, including large sections of the 5'- and 3'-non-coding regions. Sequence of three genes started from the end of the leader segment and ended in the 3'-non-coding region (H13-3 and H30) or close to the end of the coding region (H16). The remaining six sequences covered most of the coding region.

Organization of the diversity in the V_HII family

Identification of the family. Nucleotide sequences of the 11 coding regions are given in Figure 2. All sequences pertain to the V_HII family (Kabat *et al.*, 1983). The general organization of this section of the genes is highly conserved and features characteristic of this family (Rechavi *et al.*, 1983) are clearly visible all along, i.e., an intron of 83 bp which interrupts the section coding for the signal peptide (Milstein *et al.*, 1972), a coding region precisely ending with codon 98 (94 in Kabat numbering), immediately followed by the heptanucleotide recombination signal (Max *et al.*, 1979; Early *et al.*, 1980) in the 3'-non-coding region. Furthermore, comparisons of the coding regions (Figures 2 and 3) indicate that most divergent sequences differ by 25% in amino acids and 17% in nucleotides, values which are in agreement with the size and extent of divergences within this unusually large family (Brodeur and Riblet, 1984).

Estimation of the gene number in the V_HII family. Our gene sequences were translated into amino acids and compared with related BALB/c V_HII sequences as presented in Figure 3. As no sequences were found identical, this brings the total number of discrete genes of the V_HII family that have been sequenced so far to 24, thus representing to date the largest collection of germ-line Ig V genes belonging to the same family.

Current estimates of the number of V genes mostly rely on Southern blotting, using probes which may cross-hybridize only within the limits of one family (Matthyssens and Rabbitts, 1980; Cory *et al.*, 1981; Bentley and Rabbitts, 1981; Ollo *et al.*, 1983). Most recent estimates for the murine V_H repertoire average 100, with ~60 genes for the V_HII family (Brodeur and Riblet, 1984). Our data suggest that the latter value may be a much too low estimate, for the following reasons. (i) The 24 discrete V_H genes that we have isolated are far from covering the entire family. For instance, genes that would present maximal homology with the probe that was used, and which should be representative of the GAT/NP^a V_H prototype (where GAT is a (Glu^{60} Ala^{30} Tyr^{10}) random copolymer and NP^a is the BALB/c idiotype of anti-(4-hydroxy-3 nitrophenyl) acetyl antibodies), were not isolated so far; furthermore, the most diverging sequences of the family, that may be predicted on the basis of protein or cDNA sequences, as recently computed by Dildrop (1984) were still not identified amongst the germ-line genes analyzed. (ii) Out of 24 genes that have been sequenced so far (see Figure 3) no two were found identical. As clones have been randomly selected, this already points to a rather large size of this family, although various origins of the BALB/c mice that have been analyzed in separate laboratories may account for some additional polymorphism. (iii) Two genes (H16 and pCh 108A, Figure 3) differ from each other by no more than three nucleotides (three amino acids) clustered in three adjacent codons, starting at position 81 which leaves space for a large array of discrete close structures within a family for which the extreme members have a homology which may be below 78% in nucleotides (Brodeur and Riblet, 1984).

Diversity of the coding regions is germ-line organized. From amino acid comparisons given in Figure 3, a plot of variability, as defined by Kabat and Wu (1971), was derived as indicated in Figure 4. This representation clearly stresses that the organization of diversity in the germ-line absolutely parallels that reported at the expression level, with a clear-cut distinction between framework and complementarity determining regions, or CDR, in agreement with previous reports (Bothwell *et al.*, 1981; Cohen *et al.*, 1982; Blankenstein *et al.*, 1984).

Gene sequences may be grouped within discrete subfamilies, for instance H8, H9 and H13-1; pCh 111 and H24; H16, pCh 108A, 108B, H26-1 and H26-6 (Figure 3). Interestingly, whenever two genes were found linked within the same insert (pCh 108A and 108B; H26-1 and H26-6) they belonged to the same sub-family, an observation which may be suggestive of a spatial organization in the gradient of divergence. It is also striking that structural features that define a given sub-family are scattered all along the sequence, including both the framework and the CDR regions. It has long been suggested that discrete gene segments might rearrange to encode a given V_H region (Kabat *et al.*, 1978), an argument also put forward for the possible existence of gene conversion. As the actual number of different germ-line genes may be rather large and gradient organized it may be difficult to exclude that accumulated mutations may mimic conversion. For instance, Rechavi *et al.* (1983) reported that an identical 13-bp segment was identified in CDR2 codons 53–56 of two V_H genes, one belonging to the V_HII and the other to the V_HIII family (TAGTGGTGGTACC) and identical, in addition, to a D segment. It may be observed in Figure 2 that within

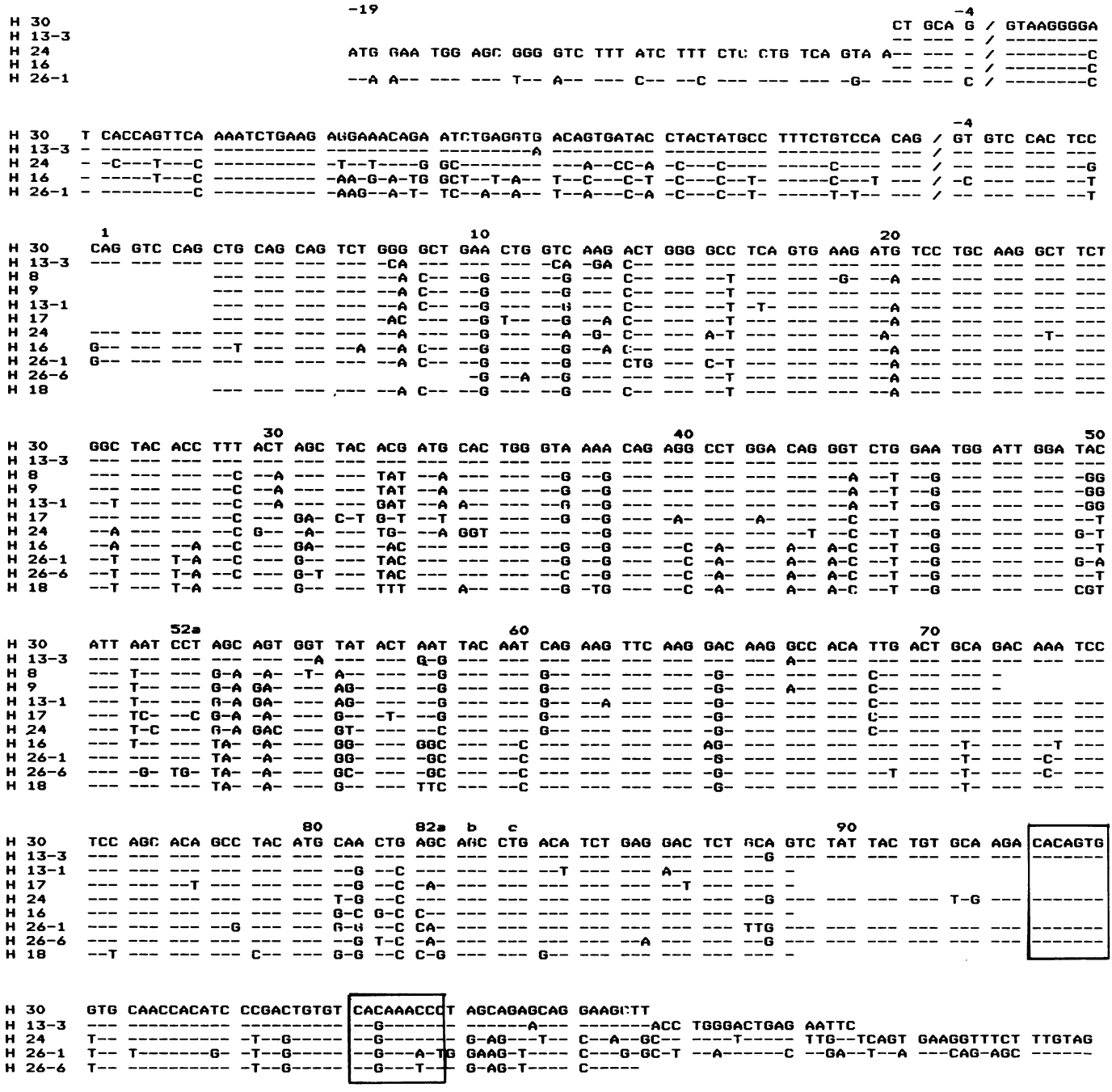

Fig. 2. Nucleotide sequences containing the V_H coding region of 11 genes of the V_HII family. Numbering is that of Kabat *et al.* (1983). 3′ rearrangement signals are boxed.

this region one finds extensive diversity and that the various nucleotides identified in the discrete genes make it likely that a distinct germ-line gene possessing the above sequence might well be isolated.

Strong homologies observed within a given family, both for nucleotide sequence and general length organization of coding and non-coding regions, would, however, certainly favor exchanges between discrete members. This may take place both at the somatic or germ-line level. An indication of stretches that have been exchanged in evolution, either upon recombination or conversion, may be gained by looking at some sequences presented in Figure 3, such as genes H26-1 and pCh 108B or H16 and pCh 108A.

Pseudogenes account for half of the V_HII gene family and must contribute to the expression of diversity. Givol *et al.* (1981) reported that a substantial number of germ-line sequences had crippling mutations. The proportion of pseudogenes was estimated at 40% by Rechavi *et al.* (1983). Among the 24 gene sequences of the V_HII family at least nine could be identified as pseudogenes, six because of a mutation leading to a stop codon within the coding region (NP^beq-5, NP^beq-3, NP^beq-31, V_H 104A, pCh 111, H24) (where NP^b is the C57 BL/6 idiotype of anti-(4 hydroxy-3 nitrophenyl) acetyl antibodies), three because of a mutation affecting the initiation codon of the leader segment (NP^beq-28, pCh 108B and H26-1). Although there is an ambiguity as to the nature of the codon at position 39 (Leu or Stop)

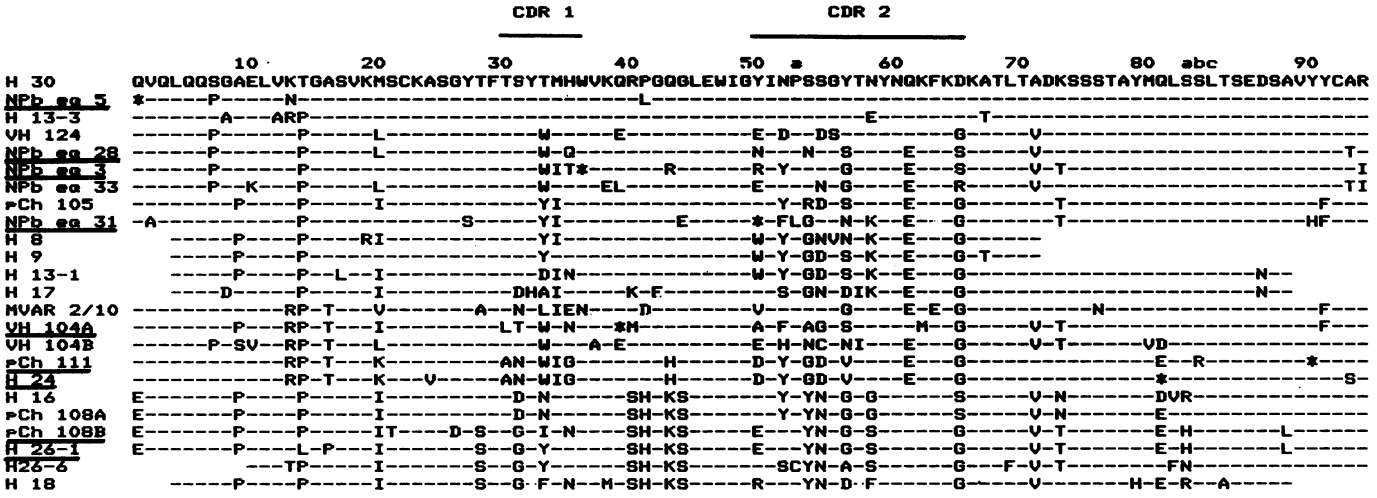

Fig. 3. Derived amino acid sequences of the coding region of 24 BALB/c V_H genes pertaining to the V_HII family. Sequences of the NP^{b}eq series are from Loh *et al.* (1983); V_H124 and V_H104A and B from Cohen and Givol (1983); pCh series from Cohen *et al.* (1982); VAR 2/10 from Blankenstein *et al.* (1984). The 11 sequences of the H series are from this publication (see Figure 2). Pseudogenes, resulting either from the occurrence of a stop codon in the coding region (indicated by an asterisk in the sequence) or from the absence of an initiator codon in the leader segment, are underlined.

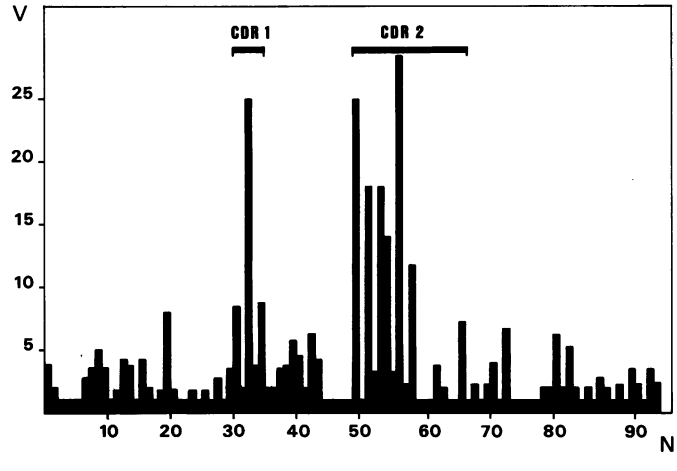

Fig. 4. Variability plot (Kabat and Wu, 1971) derived from the compilation of V_HII amino acid sequences corresponding to 24 germ-line genes as presented in Figure 3. V: variability index; N: residue number.

of NP^{b}eq-33 (Loh *et al.*, 1983), we have considered it a functional one. Leaving out the seven sequences which were not complete, this brings to 56% (nine out of 16) the proportion of pseudogenes in this family.

As already pointed out (Givol *et al.*, 1981; Blankenstein *et al.*, 1984) replacement mutations (R) are much more frequent in CDR, as appreciated from the R/S (replacement/silent mutation) ratio. This speaks in favor of a positive selective pressure which allows amplification of the germ-line encoded antigen recognition repertoire, whereas structural constraints counter select replacement mutations in framework residues. To examine whether functional and pseudogenes were submitted to similar selective pressure, we analyzed separately the distribution of replacement *versus* silent differences in both groups of sequences. To do this we adapted the Kabat and Wu (1971) diversity plot principle to nucleotide sequences. Comparison of the functional and pseudogene patterns are given in Figure 5a and b, respectively. Although the general distribution of mutation seems similar, a more objective analysis may be given by comparing the R/S values com-

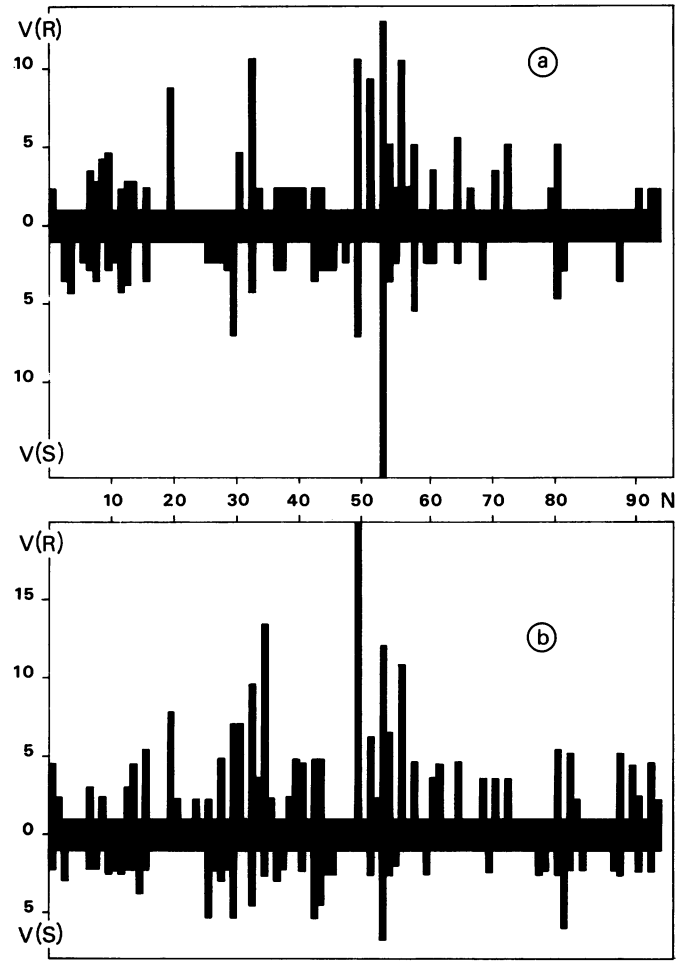

Fig. 5. Variability plot, adapted to nucleotide sequences of: (**a**) seven functional genes (V_H-124, NP^{b}eq-33, pCh 105, V_H 104B, pCh 108A, H30, H13-3) and (**b**) nine pseudogenes (NP^{b}eq-5, NP^{b}eq-28, NP^{b}eq-3, NP^{b}eq-31, V_H 104A, pCh 111, H24, pCh 108B, H26-1). Replacement mutations (R) are plotted in the upper part of each pattern, silent mutations (S) appear below. V and N as in Figure 4.

puted for each section, including all framework and CDRs (Table II).

The overall (R/S)f(R/S)ps ratio was identical for CDR1 and CDR2, and had a value (0.5) that points to the increased proportion of replacement mutations in CDRs of pseudogenes, as compared with functional ones. As these mutations are not randomly distributed, it seems reasonable to propose that the accumulation of replacement mutations in CDRs has been selected for in evolution, and that pseudogene sequences must be expressed. This imposes the necessity of frequent exchanges between pseudo and functional genes, either by recombination or conversion. These exchanges should be facilitated by the conservation in length and the marked homology of genes within a given family.

Table II. Computation of nucleotide variability index in the various regions of the coding section of V$_H$II genes

Subregions	Codons[a]	Functional genes			Pseudogenes			(R/S)f
		R	S	(R/S)$_f$	R	S	(R/S)$_{ps}$	(R/S)ps
FR1	1–29	55.35	56.6	0.97	60.57	54.6	1.12	0.86
CDR1	30–35	20.4	15.2	1.34	41.5	15.47	2.69	0.5
FR2	36–49	22.1	23	0.96	31.4	29.54	1.06	0.9
CDR2	50–65	74.35	46.95	1.58	85.35	28.46	3	0.5
FR3	66–94	49.2	42.4	1.15	63.69	49.46	1.28	0.9

[a]Kabat numbering. For each position, an index of variability was calculated, as adapted from Kabat and Wu (1971) for nucleotides (see Materials and methods). Figures indicate the cumulative values of the variability index within the corresponding stretch.
R: replacement mutation; S: silent mutations; f: functional genes and ps: pseudogenes.

Structural features of the 5′-non-coding regions

The nucleotide sequence of the 5′ regions of H24 and H26-1 genes is presented in Figure 6 and compared with other genes pertaining to the same (V$_H$II) or discrete (V$_H$I and V$_H$III) families. By looking at the various sequences, two types of subregions may clearly be identified, one highly family specific, and the other highly conserved.

Family specific residues appear clustered in large segments: (i) an intron which interrupts the gene section coding for the signal peptide (Figure 2) has, as already pointed out by Rechavi *et al.* (1983), a length which is family characteristic, with an overall homology of ~80%; (ii) the leader segments (Figure 6), and (iii) the non-coding region which starts 5′ upstream of the leader also expresses a marked homology within each family, both when length and sequence are compared.

The family-specific stretches are only interrupted by two short regions which are highly conserved (boxed in Figure 6) through the family barriers, and which are likely to represent important regulatory signals: one, ~75 nucleotides 5′ upstream from the initiation codon is compatible with the consensus sequence of the Hogness or TATA box (Breathnach and Chambon, 1981). A second of 10 nucleotides is found ending 18 nucleotides upstream of the TATA box (Parslow *et al.*, 1984). This 'cd' region, reported to be strongly homologous with a critical stretch of the SV40 enhancer by Falkner and Zachau (1984), may exert an important role in the control of the gene expression. This signal might be at the crossroads of several regulatory functions: direct action on the TATA box (Benoist *et al.*, 1980), interaction with a complementary sequence (dc) of the light chains (Falkner and Zachau, 1984), interaction with enhancer segments in the 3′-non-

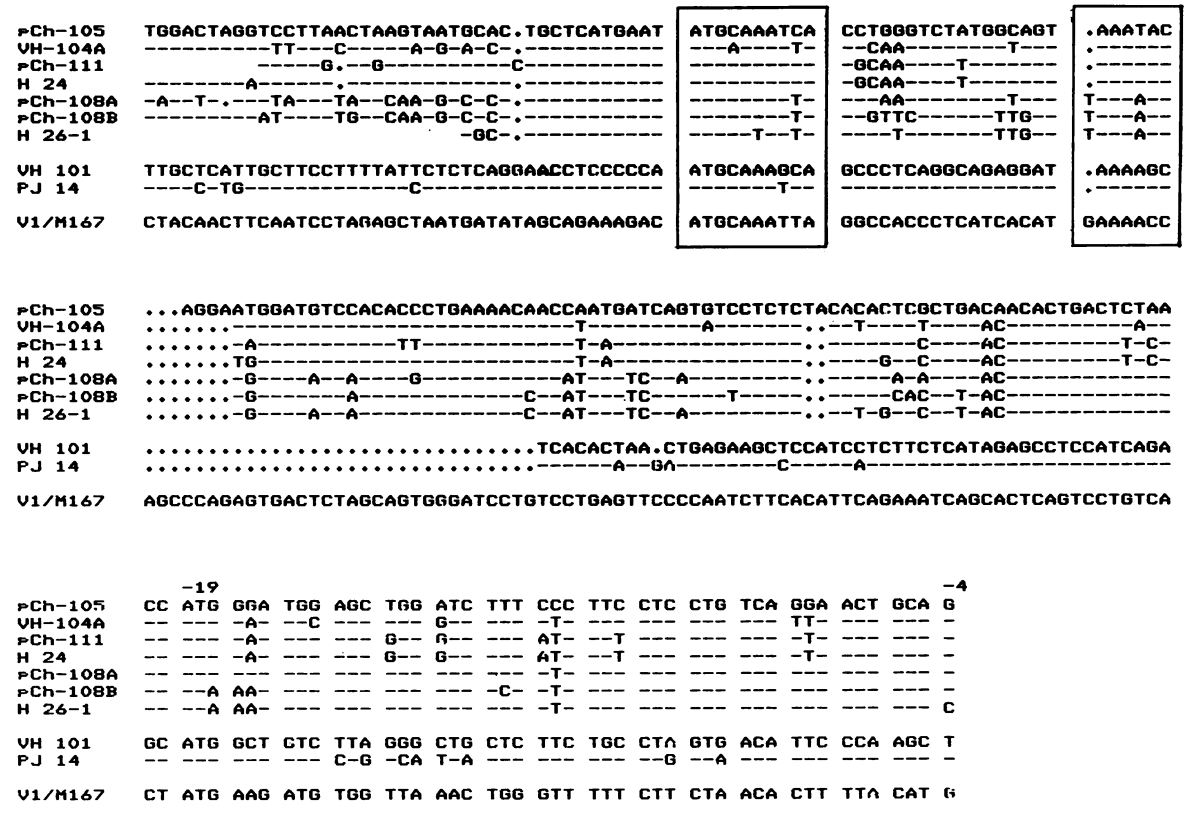

Fig. 6. Nucleotide sequences of 5′-non-coding regions and leader exon of genes pertaining to three discrete families: V$_H$II (seven first sequences, references given in Figure 3), V$_H$I (V$_H$ 101 from Kataoka *et al.*, 1982; PJ 14 from Sakano *et al.*, 1980) and V$_H$III (V1/M167 from Clarke *et al.*, 1982a). Homologies are indicated with dashes, by reference to the first sequence of each series. Deletions were introduced to ensure similar lengths. Highly conserved signal sequences are boxed.

coding region (Queen and Baltimore, 1983; Banergi *et al.*, 1983; Gillies *et al.*, 1983) or formation of a stem and loop structure by interacting with a complementary structure in 3′ of the V_H gene, allowing amplification and duplication (Rechavi *et al.*, 1983).

Besides these two critical signals, one should identify, some 23 bp downstream of the TATA box, the cap site linked to mRNA initiation of transcription (Breathnach and Chambon, 1981). This site has been reported to be various distances 5′ upstream of the initiation codon of the leader sequence. The various lengths of the region comprised between the TATA box and the initiation codon may be directly linked to the fluctuation of the location of the cap site, that thus should occur as family specific.

The overall organization of V_H genes thus obey some general and characteristic features. Besides the two main regulatory signals (cd and TATA box) all sections of the gene, including the 5′-non-coding region, the leader segment, the first intron and the main coding region are strictly family characteristic, either by sequence homologies, length of the various stretches, or both. These features strongly suggest that V_H families have evolved separately and that recombinational or conversion events may only take place within a given family.

Materials and methods

Isolation of recombinant clones

The Charon 4A *Eco*RI library containing BALB/c embryo DNA was kindly provided by Ollo *et al.* (1981). This library was screened with a nick-translated ^{32}P-labeled V_H (H IV 92, 230-bp *Pst*I fragment encoding amino acids 4 – 81 in the V_H GAT segment, Schiff *et al.*, 1983), according to the method of Benton and Davis (1977). Final washings were made at 65°C in 1 x SET (0.15 M NaCl, 30 mM Tris pH 8, 2 mM EDTA) 0.2% SDS. Isolated phage clones were mapped using single and double restriction enzyme digestions.

Subcloning into M13 vector

Double-stranded (RF) mp8 vector (Messing and Vieira, 1982) was digested with *Pst*I, *Eco*RI, *Eco*RI and *Hind*III or *Sma*I. The vector DNA was ligated with T4 ligase to recombinant DNA digested with the same enzyme, except for *Sma*I, for which the recombinant DNA was digested either by *Hae*III or *Alu*I. Colonies of transformed bacteria (JM 101) were screened for the presence of V_H sequences with the H IV 92 probe. The 5′ sequences were detected with a *Pvu*II-*Eco*RI H24 fragment starting 350 bp upstream of codon 3 and the 3′ sequences with 150-bp *Pst*I-*Eco*RI H13-3 fragment starting at codon 72.

DNA sequencing

Nucleotide sequences were determined by the dideoxy chain terminator procedure of Sanger *et al.* (1977).

Nucleotide variability index

Variability was expressed for each nucleotide position of the coding region as:

$$V = \frac{\text{Number of different nucleotides at a given position}}{\text{Frequency of the most common nucleotide at that position}}$$

as proposed by Kabat and Wu (1971) for amino acids. Results were computed for all replacement positions (leading to a change in amino acid) and silent substitutions and were plotted separately for each codon.

Acknowledgements

We express our gratitude to Drs Ollo and Rougeon for their generous gift of the genomic library, to Dr Gilbert Mazza and Françoi-Xavier Theodule for helping with computer facilities, and Dr Lee D.Leserman for critical reading of the manuscript. This work was supported by INSERM and CNRS.

References

Alt,F.W. and Baltimore,D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4118-4122.
Amzel,L.M., Poljak,R., Saul,F., Varga,J. and Richards,F. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1427-1431.
Banergi,J., Olson,L. and Schaffner,W. (1983) *Cell*, **33**, 729-740.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127-142.
Bentley,D.L. and Rabbitts,T.H. (1981) *Cell*, **24**, 613-623.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Blankenstein,T., Zoebelein,G. and Krawinkel,U. (1984) *Nucleic Acids Res.*, **12**, 6887-6900.
Bothwell,A.L.M., Paskind,M., Reth,M., Imanishi-Kari,T., Rajewsky,K. and Baltimore,D. (1981) *Cell*, **24**, 625-637.
Brack,C., Hirama,M., Lenhard-Shuller,R. and Tonegawa,S. (1978) *Cell*, **15**, 1-14.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Brodeur,P.H. and Riblet,R. (1984) *Eur. J. Immunol.*, **14**, 922-930.
Clarke,C., Berenson,J., Goverman,J., Boyer,P.D., Crews,S., Siu,G. and Calame, K. (1982a) *Nucleic Acids Res.*, **10**, 7731-7749.
Clarke,S.H., Claflin,J.L. and Rudikoff,S. (1982b) *Proc. Natl. Acad. Sci. USA*, **79**, 3280-3284.
Cohen,J.B., Effron,K., Rechavi,G., Ben-Neriah,Y., Zakut,R. and Givol,D. (1982) *Nucleic Acids Res.*, **10**, 3353-3370.
Cohen,J.B. and Givol,D. (1983) *EMBO J.*, **2**, 1795-1800.
Cory,S., Kemp,D.J., Bernard,O., Gough,N., Gerondakis,S., Tyler,B., Webb,E. and Adams,J.M. (1981) in Janeway,C., Sercarz,E.E. and Wigzell,H. (eds.), *Immunoglobulin Idiotypes*, Academic Press, pp. 49-58.
Crews,S., Griffin,J., Huang,H., Calame,K. and Hood,L. (1981) *Cell*, **25**, 59-55.
de Préval,C. and Fougereau,M. (1976) *J. Mol. Biol.*, **102**, 657-678.
Dildrop,R. (1984) *Immunol. Today*, **5**, 85-86.
Early,P., Huang,H., Davis,M., Calame,K. and Hood,L. (1980) *Cell*, **19**, 981-992.
Falkner,F.G. and Zachau,H.G. (1984) *Nature*, **310**, 71-74.
Gillies,S.D., Morrison,S.L., Oi,V.T. and Tonegawa,S. (1983) *Cell*, **33**, 717-728.
Givol,D., Zakut,R., Effron,K., Rechavi,G., Ram,D. and Cohen,J.B. (1981) *Nature*, **292**, 426-430.
Griffiths,G.M., Berek,C., Kaartinen,M. and Milstein,C. (1983) *Nature*, **312**, 271-275.
Kabat,E.A. and Wu,T.T. (1971) *Ann. N.Y. Acad. Sci.*, **190**, 382-393.
Kabat,E.A., Wu,T.T. and Bilofsky,H. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 2429-2433.
Kabat,E.A., Wu,T.T., Bilofsky,H., Reid-Miller,M. and Perry,H. (1983) *Sequences of Proteins of Immunological Interest*, NIH publications.
Kataoka,T., Nakaido,T., Miyata,T., Moriwaki,K. and Honjo,T. (1982) *J. Biol. Chem.*, **257**, 277-285.
Krawinkel,W., Zoebelein,G., Brüggemann,M., Radbruch,A. and Rajewsky,K. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4997-5001.
Loh,D.Y., Bothwell,A.L.M., White-Scharf,M.E., Imanishi-Kari,T. and Baltimore,D. (1983) *Cell*, **33**, 85-93.
Matthyssens,G. and Rabbitts,T.H. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6561-6565.
Max,E.E., Seidman,J.G. and Leder,P. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3450-3454.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 296-276.
Milstein,C., Brownlee,G.G., Harrison,T.M. and Matthews,M.B. (1972) *Nature New Biol.*, **239**, 117-120.
Ollo,R. and Rougeon,F. (1983) *Cell*, **32**, 515-523.
Ollo,R., Auffray,C., Sikorav,J.L. and Rougeon,F. (1981) *Nucleic Acids Res.*, **9**, 4099-4108.
Ollo,R., Sikorav,J.L. and Rougeon,F. (1983) *Nucleic Acids Res.*, **11**, 7887-7897.
Parslow,T.G., Blair,D.L., Murphy,W.J. and Granner,D.K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2650-2654.
Queen,C. and Baltimore,D. (1983) *Cell*, **33**, 741-748.
Rechavi,G., Ram,D., Glazer,L., Zakut,R. and Givol,D. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 855-859.
Sakano,H., Hüppi,K., Heinrich,G. and Tonegawa,S. (1979) *Nature*, **280**, 288-294.
Sakano,H., Maki,R., Kurosawa,Y., Roeder,W. and Tonegawa,S. (1980) *Nature*, **286**, 676-683.
Sakano,H., Kurosawa,Y., Weigert,M. and Tonegawa,S. (1981) *Nature*, **290**, 562-565.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Schiff,C., Milili,M. and Fougereau,M. (1983) *Nucleic Acids Res.*, **11**, 4007-4017.
Schrier,P.H., Bothwell,A.L.M., Mueller-Hill,B. and Baltimore,D. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4495-4499.
Seidman,J.G., Leder,A., Nau,M., Norman,B. and Leder,P. (1978) *Science (Wash.)*, **202**, 11-17.
Seidman,J.G., Max,E.E. and Leder,P. (1979) *Nature*, **280**, 370-375.

Received on 5 March 1985