

Dispersed localization of D segments in the human immunoglobulin heavy-chain locus

Fumihiko Matsuda, Kwang Ho Lee, Sumiko Nakai¹, Takayuki Sato, Mieko Kodaira, Shu Qin Zong, Hiroshi Ohno², Shiro Fukuhara² and Tasuku Honjo

Department of Medical Chemistry and ²Department of Medicine, Kyoto University Faculty of Medicine, Sakyo-ku, Kyoto 606 and ¹Department of Genetics, Osaka University School of Medicine, Osaka 560, Japan

Communicated by S. Numa

We have studied the organization of the human immunoglobulin heavy-chain genes by pulse field gel electrophoresis as well as by isolation of cosmid clones. The total length of the heavy-chain variable region locus was estimated to be ~3000 kb. We found that D segments including a recently isolated D₅ segment were dispersed among V_H segments. We identified a pseudo V segment 18 kb 3' to the D₅ segment in isolated cosmid clones. A 300 kb fragment produced by *Mlu*I digestion contained V_H, D, J_H segments and the distance between V_H and D was estimated to be ~240 kb. Overlapping cosmid clones containing the human D₁, D₂, D₃, D₄, J_H, C_μ and C_δ genes were isolated. Restriction maps of these regions indicated that the distance between D and J_H is about 22 kb. A partial restriction map of the V_H locus was constructed using the pulse field gel electrophoresis technique and deletion of V_H segments in B cells.

Key words: cosmid clones/deletion map/nucleotide sequence/pulse field gel electrophoresis

Introduction

The immunoglobulin heavy chain consists of the variable (V_H) region which recognizes antigens and the constant (C_H) region which mediates physiological functions. The complete V_H region gene is comprised of three germline segments, namely V_H, diversity (D) and joining (J_H) segments (reviewed by Honjo, 1983; Tonegawa, 1983; Honjo and Habu, 1985) which are located on chromosome 14 (14q32) in the human (Kirsch *et al.*, 1982) and chromosome 12 in the mouse (D'eustachio *et al.*, 1980). The number and organization of these segments provide the germline basis of the immunoglobulin repertoire and random recombination between separate sets of the three segments generates combinatorial diversity. Numbers of the V_H and D segments in the human genome are unknown whereas there are six J_H segments and 11 C_H genes. The six human J_H segments are clustered within a small region (3 kb) about 6 kb upstream to the C_μ gene (Ravetch *et al.*, 1981). Four human D segments were shown to be clustered (Siebenlist *et al.*, 1981). Although their location was not determined, it was assumed that the D segments were located between the V_H segments and the J_H segments from the physical

linkage study in mouse (Wood and Tonegawa, 1983). To evaluate precisely the germline repertoire of the human immunoglobulin, it is essential to construct the physical map of the immunoglobulin heavy-chain locus.

We have initiated isolation of the entire human V_H locus using a cosmid vector to elucidate the germline V_H repertoire. We have isolated 18 independent cosmid clones containing 55 V_H segments from a human genomic library (Kodaira *et al.*, 1986). Using these 18 clones we have shown that V_H segments belonging to three different families (V_{H-I}, V_{H-II} and V_{H-III}) are interspersed among themselves. Subsequently, we have cloned and characterized a novel family (V_{H-IV}) of human V_H genes (Lee *et al.*, 1987). We have also isolated and sequenced the fifth D segment (D₅) which is not physically linked to the D₁–D₄ cluster (Zong *et al.*, 1988).

To elucidate the general organization of the human V_H locus we employed the pulse field gel electrophoresis (PFG) technique, which allowed us to estimate that the size of the human V_H locus is ~3000 kb. Using this method we showed that D segments including the D₅ segment were dispersed among V_H clusters in agreement with our previous proposal which was based on evolutionary considerations (Honjo *et al.*, 1981). This finding was further confirmed by cloning and characterization of the 120 kb DNA fragments containing D₁–D₄, J_H, C_μ and C_δ genes. We have constructed a partial restriction map of the human V_H locus by a combination of PFG analysis and deletion of V_H segments in B cell lines.

Results and discussion

General organization of the human V_H locus

DNA from FLEB14-14, which is an Epstein–Barr virus (EBV)-transformed fetal liver cell line and has no rearrangement of the immunoglobulin genes (Katamine *et al.*, 1984; Otsu *et al.*, 1987), was digested with the restriction enzyme *Mlu*I, electrophoresed in PFG and then subjected to Southern blot analysis using the V_H probe of the V_{H-III} family (Kodaira *et al.*, 1986). This probe detected 300, 350, 700, 800 and >1000 kb DNA fragments (Figure 1A, lane 1). The same five bands were detected using V_{266BL} (V_{H-I} family), V_{CE-1} (V_{H-II} family) and V₇₁₋₂ (V_{H-IV} family) as probes (data not shown), indicating that the total length of the V_H locus is ~3000 kb.

When the same filter was hybridized with the D₅ probe which cross-hybridized the D₁, D₂, D₃ and D₄ segments, three out of the five *Mlu*I bands were detected, namely the 300, 700 and >1000 kb bands (Figure 1A, lane 2). By contrast, the J_H probe detected only the 300 kb *Mlu*I band (Figure 1A, lane 3). These results indicate that the 300 kb *Mlu*I fragment contains V_H, D and J_H segments and that D segments are dispersed among the V_H clusters.

The D₅ probe detected eight distinct bands ranging from ~6 kb up to 240 kb in *Sfi*I-digested FLEB14-14 DNA (Fig-

ure 1B). However, *Bam*HI digestion of FLEB14-14 DNA produced only five bands (18, 7.4, 6.6, 1.8 and 1.6 kb) hybridizing with the D₅ probe (Figure 1C) in agreement with the previous reports (Siebenlist et al., 1981; Zong et al., 1988). The results suggest that large chromosomal regions containing multiple D segments were duplicated and that more unknown germline D segments were present in the human genome.

Physical mapping of the D₄-J_H-C_δ region

Among 18 independent V_H or J_H containing cosmid clones from human genomic libraries (Kodaira et al., 1985), two clones, U2-2 and U7, were shown to hybridize with both D and J_H probes. Comparison of the restriction maps of the two clones showed that they were almost identical to each other and encompassed the region from ~5 kb upstream to the D₂ segment to 2 kb downstream to the J_H cluster (Figure 2). These clones thus overlapped with previously isolated clones (D25, D31 and D48) containing the D₁, D₂, D₃ and D₄ segments (Zong et al., 1988).

Another clone (64), which hybridized only with J_H probe, encompassed the region from 5 kb upstream to the J_H cluster to 20 kb downstream to the C_δ gene. Clone 64 had a 10 kb overlap with the clones U2-2 and U7. These six clones together cover the region (~12 kb) containing the D₁-D₄, J_H, C_μ and C_δ genes (Figure 2). The distance between the most 3' D segment (D₄) and the J_H cluster was found to be ~22 kb. Clone 64 has only two *Mlu*I sites, one immediately 3' to the J_H cluster and the other between the C_μ and C_δ genes. Physical mapping studies on the isolated cosmid clones are consistent with the above PFG results that all of the four D segments (D₁-D₄) and the J_H cluster are located in the 300 kb fragment produced by *Mlu*I digestion.

The distance between the 3' V_H and D segment is less than ~240 kb

A cosmid clone (35) was isolated by the V_H-I probe. The 1.1 kb *Pst*I-*Pvu*II fragment hybridizing with the V_H-I probe was isolated and sequenced (sequencing strategy not shown). The coding region sequence was identified by comparison with the V₇₁₋₅ sequence of the V_H-I family (Kodaira et al., 1986) as shown in Figure 3A. Their nucleotide and amino acid sequences were 87 and 77%

homologous, respectively. The V₃₅ gene was found to be potentially functional. The amino acid sequence of the V₃₅ segment was also homologous (78.6%) to that of V_{266BL} (Nishida et al., 1982; Kenten et al., 1982).

The 1.3 kb *Eco*RI fragment (p35 probe) isolated from clone 35 hybridized to a unique fragment of human placenta DNA on Southern blot hybridization (data not shown). The p35 probe hybridized with the 300 kb *Mlu*I fragment fractionated by PFG (Figure 1A, lane 4). Because the four D segments and the entire J_H cluster were present in the 300 kb *Mlu*I fragment, we concluded that the distance between the V₃₅ segment and the D₄ segment was less than 240 kb.

Evidence for 5'-D-V_H-3' organization

We confirmed that the D₅ segment was dispersed in V_H segments by physical mapping of clone D26, in which the D₅ segment had been found originally (Zong et al., 1988). Clone D26 had a 30 kb overlapping region with another clone 3 isolated by the V_HBV probe (Figure 4). These two clones had a common restricted fragment hybridized with

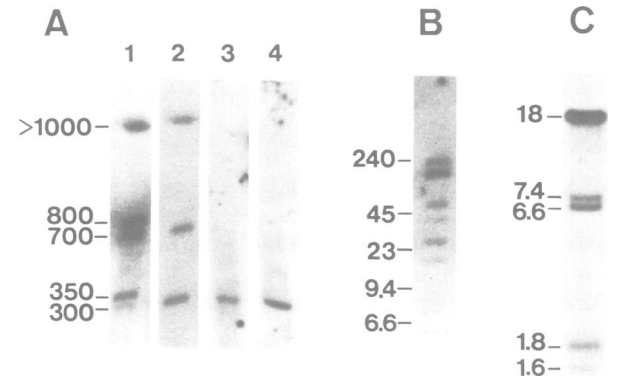


Fig. 1. PFG electrophoresis analysis of human V_H locus. Southern blot filters of *Mlu*I (A), *Sfi*I (B), *Bam*HI (C) digested DNA of FLEB14-14 were hybridized with probes indicated. Chromosomes of *Saccharomyces cerevisiae* were used as DNA size markers. Numbers in kb. Each lane contains ~1 μg (A,B) and ~2 μg (C) DNA respectively. Probes used for each lane are: (A) lane 1, V_HBV (V_H-III); lane 2, D₅; lane 3, J_H; lane 4, p35; (B), (C) D₅.

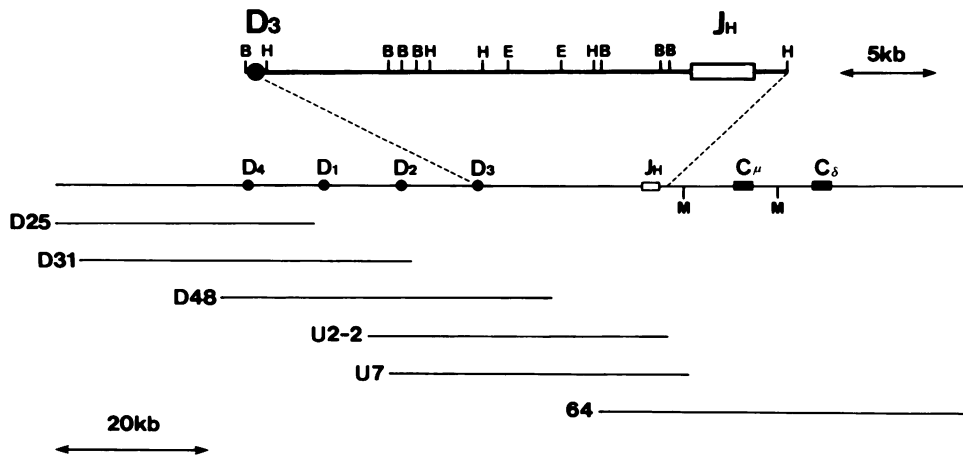


Fig. 2. Restriction map of the region containing human D-J_H-C_μ-C_δ genes. The top line represents newly mapped 22 kb region. Second line represents ~120 kb region containing D₁-D₄ (●), J_H cluster (□), C_μ and C_δ genes (■). The transcriptional orientation of all the segments is from left to right. Nucleases are abbreviated as follows: E, *Eco*RI; B, *Bam*HI; H, *Hind*III; M, *Mlu*I.

the V_HBV probe. Since restriction mapping showed that the V_H-hybridizing fragment was located ~20 kb 3' to the D₅ gene, we determined the nucleotide sequence of the V_H-hybridizing fragment (sequencing strategy not shown). The coding region sequences were identified by comparison with that of the V₇₁₋₁ gene of the V_H-III family (Kodaira *et al.*, 1986) as shown in Figure 3B. The coding region of V₃ and V₇₁₋₁ were found to be 78% homologous with each other.

There were six termination codons in frame with the initiator codon ATG in the V₃ sequence. Furthermore, the V₃ gene had eight insertion mutations and three deletion mutations in the coding region. V_H pseudogenes seem to be classified into two distinct groups; those with a few mutations (conserved) and those with drastic changes (Kodaira *et al.*, 1986). The V₃ gene seems to be a member of the latter group.

We then tested if the D₅ segment is located in the

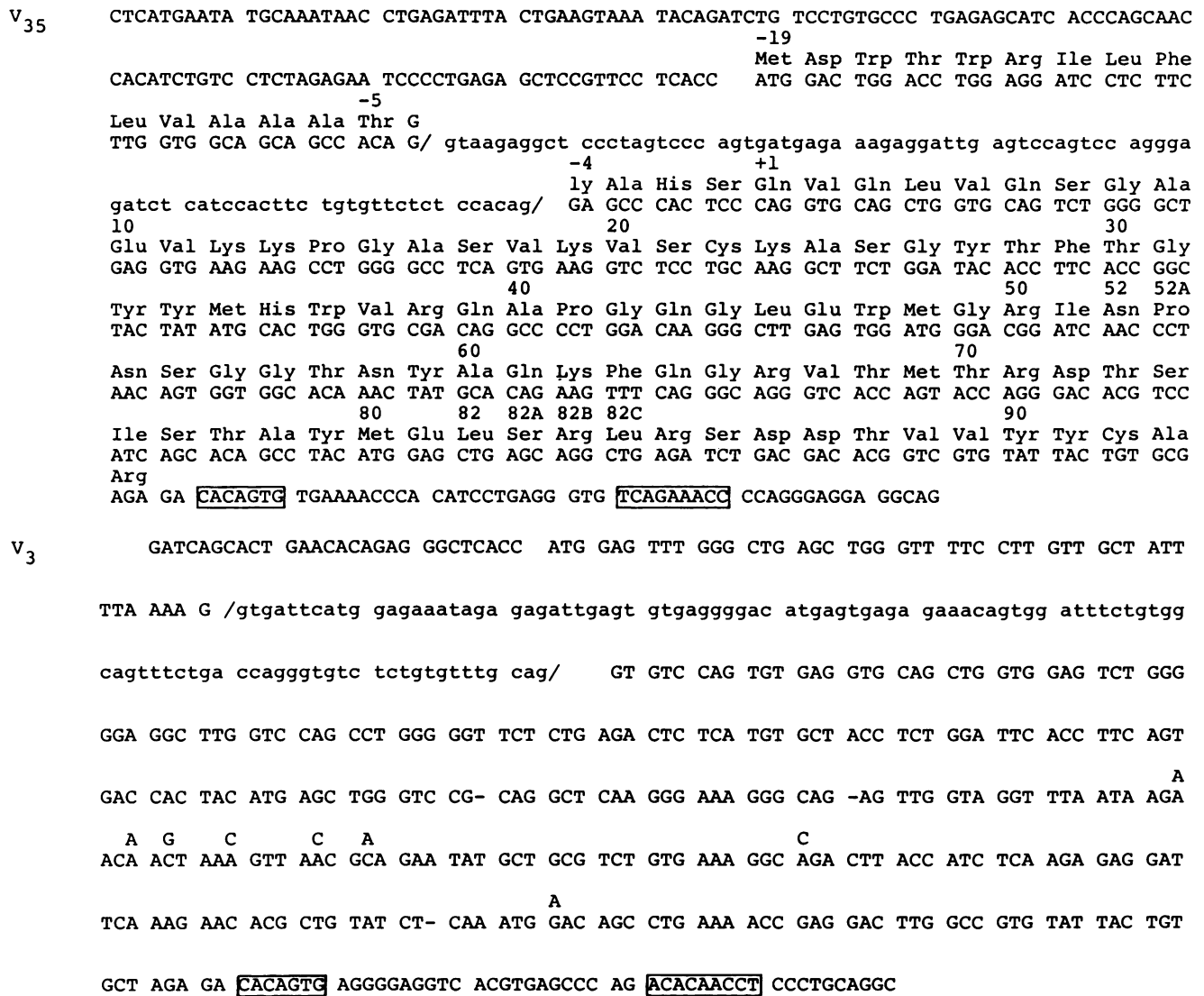


Fig. 3. Nucleotide sequence of V₃₅ (A) and V₃ (B) genes. Sequences that align with the protein coding sequence are given in triplets with numbering according to Kabat *et al.* (1987) and introns are shown in lower case letters. The amino acid sequence of V₃₅ deduced from the nucleotide sequence is also shown above the nucleotide sequence. Inserted bases are placed above the next base and deletions are shown by dashes. The signals for V_H-D joining are boxed. Exon and intron junctions are shown by slashes.

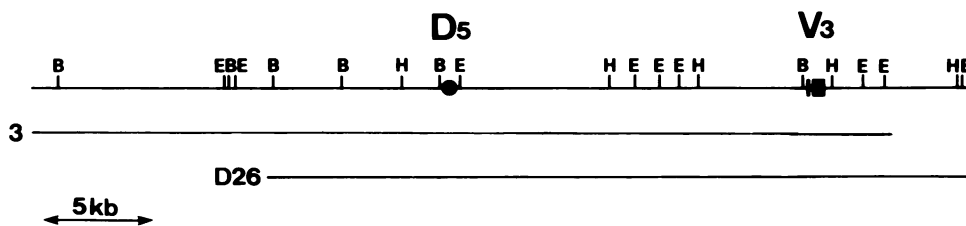


Fig. 4. Restriction map of the human D₅ locus. The transcriptional orientation of D₅ (●) and V₃ (■) genes is from left to right. Nucleases are abbreviated as follows: E, *EcoRI*; B, *BamHI*; H, *HindIII*.

immunoglobulin heavy-chain locus (band 14q32) using the human – mouse cell hybrid Rag/Go4 line which carries a single human chromosome with 14-X translocation (X;14) (p22;q32). The hybrid contained the immunoglobulin heavy-chain locus (band 14q32) which was translocated to the short arm of chromosome X (Purrello *et al.*, 1987). High mol. wt DNAs of Rag/Go4, FLEB14-14 (germline control) and Rag (mouse cell) were digested with *Bam*HI. Southern blot hybridization using D₅ probe showed the identical set of five bands, namely 18, 7.4, 6.6, 1.8 and 1.6 kb in both FLEB14-14 and Rag/Go4 DNAs. However, none of them was detected in Rag DNA (Figure 5). The D₅ probe did not hybridize with mouse – human hybrid cells which had human X chromosomes but not human chromosome 14 (data not shown). The results indicate that the D₅ segment is located in the heavy-chain locus of chromosome 14.

What mechanism brought the D segment upstream to the V_H segment? The heavy-chain genes of horned shark, a primitive elasmobranch, was shown to have a unique organization; a 10 kb unit consisting of each copy of the V_H, D, J_H and C_H segments, which might be an original form of the immunoglobulin gene, was tandemly repeated in the genome of this organism (Hinds and Litman, 1986). The human V_H locus must have evolved through repeated duplication of V_H, D and J_H segments. Assuming that a unit of duplication contained the V_H, D and J_H segments, the D and J_H segments could easily have been interspersed between V_H segments.

In fact, many fragments in V_H-containing cosmid clones were weakly hybridized with either the D or the J_H probe (unpublished observations). The D segments, which are flanked by nonamer and heptamer recombination signal sequences at both sides, might have been capable of D – J_H rearrangement even if located very far from the J_H cluster. These D segments could be conserved by functional constraint. Since the sequences of the five human germline D segments thus far isolated are not found in any of the known V – D – J sequences, it is reasonable to assume that there may be more unknown functional D fragments in the human V_H locus. The J_H segments, which had been dispersed among V_H clusters, were unable to participate in the formation of functional heavy-chain genes because of the large distance from the C_μ gene, and thus diverged rapidly without functional constraint. The fragment detected by the J_H probe might be diverged pseudo-J_H genes or their relicts.

Order of the MluI fragments of the human V_H locus

We have tested whether EBV-transformed human B cell lines might have deleted some *Mlu*I fragments containing V_H segments by V_H – D – J_H recombinations on both alleles. DNAs from these cells were digested with *Mlu*I, electrophoresed in PFG and subjected to Southern blot analysis using

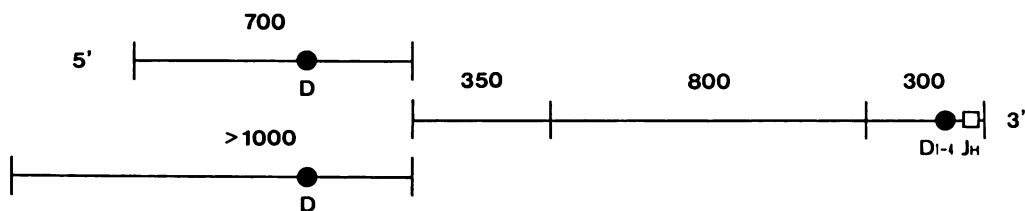


Fig. 7. Organization of human V_H locus. Horizontal lines separated by vertical lines show the *Mlu*I fragments which contain V_H segments. The number above each line represents the size of the fragment in kb. The symbols (●) and (□) indicate the D and J_H segments respectively.

V_H-III probe as shown in Figure 6. DNA from FLEB14-14 cells which gives the germline V_H profile was used as control. The profiles of some cell lines were very complex and, therefore, did not allow us to draw any conclusions. Three cell lines, namely CESS, LCL-D and EBV5, showed deletions of some *Mlu*I fragments.

The deletion of the 300 and 800 kb fragments was observed in all of the three cell lines, suggesting that these fragments are located at the most 3' part of the V_H locus. Since the J_H cluster is located in the 300 kb fragment as described above, the 800 kb fragment is likely to be located immediately upstream to the 300 kb fragment. The 350 kb fragment was deleted in two of them (LCL-D and EBV5)

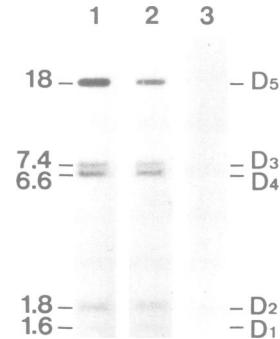


Fig. 5. Southern hybridization of the D₅ probe to DNA from a mouse – human hybrid cell line. Southern blot filters of *Bam*HI-digested DNAs of FLEB14-14, Rag/Go4 and Rag were hybridized with the D₅ probe. Each lane contains ~2 μg DNA. Origins of DNAs are: lane 1, FLEB14-14; lane 2, Rag/Go4; lane 3, Rag. Numbers are in kb.

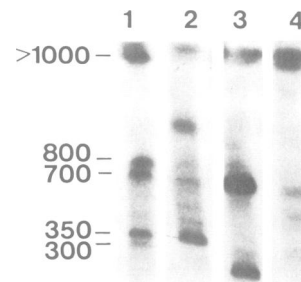


Fig. 6. Deletion of V_H segments in human B cell lines. DNAs from the cell lines were digested with *Mlu*I, electrophoresed in PFG. Filters were hybridized with the V_HIII (V_H-III family) probe. DNA size markers are indicated in kb. Each lane contains ~1 μg of DNA. DNAs used for each lane are: lane 1, FLEB14-14; lane 2, CESS; lane 3, LCL-D; lane 4, EBV5.

although it remained in the germline intensity in CESS cell line. None of the three cells showed complete loss of the >1000 and 700 kb fragments. As the >1000 and 700 kb fragments were retained in LCL-D and EBV5 cells which deleted the 800, 350 and 300 kb fragments, the >1000 and 700 kb fragments might be located 5' to those deleted fragments. DNA fragments larger than 1000 kb were staggered after PFG and caused some non-specific binding. It was, therefore, very difficult to detect the change of the intensity of this fragment. Assuming that the rearrangements in human V_H locus are mediated by DNA excision mechanism in these cells, we propose a tentative order of the *Mlu*I fragments of the V_H locus as follows: 5'-(>1000, 700)-250-800-300-3' as shown in Figure 7.

Materials and methods

FLEB14-14, EBV5, CESS and LCL-D are EBV-transformed human cell lines. FLEB14-14, which is derived from a human fetal liver cell, has several B lineage markers but no rearrangements in the immunoglobulin gene (Katamine *et al.*, 1984; Otsu *et al.*, 1987). The other three lines are derived from human peripheral blood lymphocytes and have both alleles of the JH segment rearranged (Takahashi *et al.*, 1984; our unpublished data). Genomic DNAs used for PFG were prepared as described (Poustka *et al.*, 1987). Restriction endonuclease digests of DNA were electrophoresed using Pulsaphor™ system (LKB) according to the manufacturer's directions. Gels were transferred to Genescreenplus™ nylon membrane (NEN) according to the method of Southern (1975) after treatment with 0.25 M HCl for 15 min. The filters were hybridized to ³²P-labeled nick translated probe as described (Honjo *et al.*, 1979). Filters were washed three times (30 min each) at 50°C in 0.1 M NaCl, 10 mM sodium citrate, 0.1% sodium dodecyl-sulfate.

Plasmid DNA of cosmid clones was isolated by the alkaline lysis method as described (Maniatis *et al.*, 1982). Isolated restriction fragments were cloned into pUC18 and pUC19. The chain termination method (Sanger *et al.*, 1980; Hattori and Sakaki, 1986) was used for sequencing pUC clones. Origins of DNA probes are: V_{266BL} (Nishida *et al.*, 1982), V_{CE-1} (Takahashi *et al.*, 1984), V_{HBV} and V₇₁₋₂ (Kodaira *et al.*, 1986), D₅ (Zong *et al.*, 1988) and JH (Otsu *et al.*, 1987).

Acknowledgements

We are grateful to Dr M.Siniscalco (Memorial Sloan-Kettering Cancer Center) for Rag and Rag/Go4 lines; Drs O.Niwa and M.Yanagida (Kyoto University) for advice about PFG; Ms J.Kimura and Ms K.Ikeno for their excellent technical assistance; and Ms K.Hirano for her assistance in the preparation of the manuscript. This work was supported by grants from the Ministry of Education, Science and Culture of Japan and from the Science and Technology Agency of Japan.

References

- D'eustachio,P., Pravtceva,D., Marcu,K. and Ruddle,F.H. (1980) *J. Exp. Med.*, **151**, 1545-1550.
- Hattori,M. and Sakaki,Y. (1986) *Anal. Biochem.*, **152**, 232-238.
- Hinds,K.R. and Litman,G.W. (1986) *Nature*, **320**, 546-549.
- Honjo,T. (1983) *Annu. Rev. Immunol.*, **1**, 499-528.
- Honjo,T. and Habu,S. (1985) *Annu. Rev. Biochem.*, **54**, 803-830.
- Honjo,T., Obata,M., Yamawaki-Kataoka,Y., Kataoka,T., Kawakami,T., Takahashi,N. and Mano,Y. (1979) *Cell*, **18**, 559-568.
- Honjo,T., Nakai,S., Nishida,Y., Kataoka,T., Yamawaki-Kataoka,Y., Takahashi,N., Obata,M., Shimizu,A., Yaoita,Y., Nikaido,T. and Ishida,N. (1981) *Immunol. Rev.*, **59**, 33-68.
- Kabat,E.A., Wu,T.T., Reid-Miller,M., Perry,H.M. and Gottesman,K.S. (1987) *Sequences of Proteins of Immunological Interest*. NIH Publications, Washington, DC.
- Katamine,S., Otsu,M., Tada,K., Tsuchiya,S., Sato,T., Ishida,N., Honjo,T. and Ono,Y. (1984) *Nature*, **309**, 369-371.
- Kenten,J.H., Molgaard,H.V., Houghton,M., Derbyshire,R.B., Viney,J., Bell,L.O. and Gould,H.J. (1982) *Proc. Natl. Acad. Sci. USA.*, **79**, 6661-6665.
- Kirsch,I.R., Morton,C.C., Nakahara,K. and Leder,P. (1982) *Science*, **216**, 301-303.
- Kodaira,M., Kinashi,T., Umemura,I., Matsuda,F., Noma,T., Ono,Y. and Honjo,T. (1986) *J. Mol. Biol.*, **190**, 529-541.
- Lee,K.H., Matsuda,F., Kinashi,T., Kodaira,M. and Honjo,T. (1987) *J. Mol. Biol.*, **195**, 761-768.
- Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Nishida,Y., Miki,T., Hisajima,H. and Honjo,T. (1982) *Proc. Natl. Acad. Sci. USA.*, **79**, 3833-3837.
- Otsu,M., Katamine,S., Uno,M., Yamaki,M., Ono,Y., Klein,G., Sasaki,M.S., Yaoita,Y. and Honjo,T. (1987) *Mol. Cell. Biol.*, **7**, 708-717.
- Poustka,A., Pohl,T.M., Barlow,D.P., Frischauf,A. and Lehrach,H. (1987) *Nature*, **325**, 353-355.
- Purrello,M., Alhadeff,B., Whittington,E., Buckton,K.E., Daniel,A., Arnaud,P., Rocchi,M., Archidiacono,N., Filippi,G. and Siniscalco,M. (1987) *Cytogenet. Cell. Genet.*, **44**, 32-40.
- Ravetch,J.V., Siebenlist,U., Korsmeyer,S., Waldmann,T. and Leder,P. (1981) *Cell*, **27**, 583-591.
- Sanger,F., Coulson,B.G., Smith,A.J.H. and Roa,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.
- Siebenlist,U., Ravetch,J.V., Korsmeyer,S., Waldmann,T. and Leder,P. (1981) *Nature*, **294**, 631-635.
- Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
- Takahashi,N., Noma,T. and Honjo,T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5194-5198.
- Tonegawa,S. (1983) *Nature*, **392**, 575-581.
- Wood,C. and Tonegawa,S. (1983) *Proc. Natl. Acad. Sci. USA.*, **80**, 3030-3034.
- Zong,S.Q., Nakai,S., Matsuda,F., Lee,K.H. and Honjo,T. (1988) *Immunol. Lett.*, in press.

Received on January 4, 1988