Rearranged immunoglobulin heavy chain variable region (V_H) pseudogene that deletes the second complementaritydetermining region

(molecular cloning/DNA sequence determination/human V_H gene)

NAOKI ȚAKAHASHI[†], TAKAHUMI NOMA, AND TASUKU HONJO

Department of Genetics, Osaka University Medical School, Osaka 530, Japan

Communicated by Elvin A. Kabat, March 30, 1984

ABSTRACT We have cloned two rearranged heavy chain variable region (V_H) genes from the IgG-producing human cell line CESS. The V_H gene, which is linked to the μ chain constant region (C_{μ}) gene, has two deletions at residues 45-62 and 82A-90, the former of which corresponds closely to the second complementarity-determining region (CDR2). These results could indicate that translocation of CDR2 occurred and could give support to the argument that reassortment of the V minigenes is involved in the generation of hypervariability during evolution. However, the rearranged pseudogene could have also arisen by fortuitous deletion. The other V_H gene of CESS is an expressed form and is probably linked to the C_{γ} gene. The diversity region (D) segments used in these rearranged V genes are <38% homologous to known human germline D segments, indicating the presence of more unknown germline Dsegments.

The variable (V) regions of the immunoglobulin have been shown to contain three hypervariable segments that make contact with various antigen determinants (for review, see ref. 1). These segments are called the complementarity-determining region (CDR); the rest of the V region constitutes the framework segments (FR1, FR2, FR3, and FR4). Somatic changes of immunoglobulin gene sequences contribute enormously to amplification of the V-region diversity (for reviews, see refs. 2 and 3). The somatic mechanisms involve (i) randomly paired recombination of variable (V), diversity (D), and joining (J) segments, (ii) variability at a ligation site between the recombined segments, and (iii) base replacement (somatic mutation). Although hypervariability of the third CDR is thus explained by the reassortment of D and Jsegments, the genetic basis of hypervariability of the first and second CDRs are less understood.

Evolutionary processes that increase the number and diversity of the V, D, and J segments also play important roles to augment the V region diversity. Ohta (4) calculated that the amino acid substitution rate of CDR is 3 times faster than that of the framework region. The divergence rate of CDR is as high as that of fibrinopeptide, the most rapidly diverging protein so far known, indicating that CDR is under the weakest selection pressure at the protein level. The homology in the framework region may be maintained by stronger selection pressure or by segment transfer between different loci either through gene conversion or through double unequal crossing-over (5–7). The V, D, and J segments appear to diverge by the balance of drift and recombinational homogenization.

Wu and Kabat (8) found that CDR2 of a human heavy chain variable region (V_H) segment contains 14 nucleotides identical to a human D segment. Sequence homology be-

tween a human D segment and CDR2 of a human V_H segment was also reported (9). Bernstein *et al.* (10) found a considerable homology between CDR2 of a rabbit V_H and the human V_{H26} sequences. The homology region is precisely the part of the human CDR2 in which Wu and Kabat (8) found a 14-nucleotide stretch of homology to the human D2 minigene. These results are consistent with the proposal that a minigene, like the D segment, may be involved in the generation of the CDR diversity through mechanisms of either gene conversion, reassortment, or insertion of the minigene (8, 11–13). This proposal argues that recombination, be it somatic or germline, contributes to amplify the diversity of CDR1 and CDR2 as well as of CDR3.

To study organization and diversity of the human V_H gene family, we have cloned and characterized rearranged V_H genes from an Epstein-Barr virus-transformed human B-cell line (CESS) producing IgG (14). We have found an interesting rearranged V_H pseudogene that has large deletions closely corresponding to CDR2. One attractive interpretation of the results is that the CDR2 was translocated to another V_H segment or that such a pseudogene could be an acceptor of a minigene, thus increasing the V_H segment repertoire, in agreement with the minigene theory (8, 11–13).

MATERIALS AND METHODS

Materials. CESS cell line was grown as described (14) and provided by T. Kishimoto (Osaka University). High molecular weight DNA of CESS cells was prepared as described (15).

Methods. Southern blot transfer and hybridization were carried out as described (16, 17). EcoRI and $EcoRI^*$ fragments containing human heavy chain joining region (J_H) segments were isolated by agarose gel electrophoresis and ligated with λ gtWES vector (18) with T4 ligase. The hybrid DNA was packaged *in vitro* (19), and the recombinant phage were screened using the human J_H fragment (20) as probe, according to the method of Benton and Davis (21). Cloning experiments were carried out in accordance with the Japanese guidelines for recombinant DNA experiments. DNA sequence determination was done according to the method of Maxam and Gilbert (22).

RESULTS AND DISCUSSION

Rearrangement of J_H **Segments in CESS Cell Line.** To identify rearranged V_H genes of CESS cells, Southern blot filters of restricted DNA of CESS cells were hybridized with human J_H and μ chain constant region (C_{μ}) probes (20) as

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

Abbreviations: V, D, and J, variable, diversity, and joining regions of immunoglobulin; V_H and J_H , V and J region of heavy chain; C_{μ} and C_{γ} , constant region of μ chain and γ chain of immunoglobulin; CDR, complementarity-determining region; kb, kilobase(s).

Present address: Division of Biology, California Institute of Technology, Pasadena, CA.

shown in Fig. 1A. It is clear that both alleles of the J_H segment of CESS are rearranged. EcoRI digestion of CESS DNA produced 11- and 23-kilobase (kb) fragments hybridizing to the J_H probe, whereas a 17-kb EcoRI fragment hybridizing to the J_H is found in human placenta DNA. By contrast, a single germline C_{μ} gene fragment is present in CESS (Fig. 1B). The intensity of the C_{μ} hybridizing band in CESS DNA is decreased to approximately one-half of that in human placenta DNA, suggesting that one allele of the C_{μ} genes is deleted from CESS DNA. Since Xba I digestion yielded a C_{μ} fragment identical to the germline form (data not shown), gross rearrangement did not take place in the region encompassing both the μ -chain switch region (S_{μ} region) and the C_{μ} gene. When the J_{H} -hybridized filter was washed off and probed with the 5' flanking region of the C_{μ} gene, only the 11-kb EcoRI fragment hybridized; the 23-kb EcoRI fragment did not. Instead, the 23-kb EcoRI fragment hybridized with a C_{γ} probe (data not shown). The results indicate that the 23- and 11-kb EcoRI fragments are linked to the C_{γ} and C_{μ} genes, respectively. This conclusion is supported by the nucleotide sequence determination described below. Namely, the 23-kb EcoRI fragment contains an active V_H gene, whereas the 11-kb EcoRI fragment contains a pseudogene.

Cloning the Two Rearranged J_H Segments. The 23- and 11kb EcoRI fragments hybridizing to the J_H probe were partially purified by agarose gel electrophoresis. The 11-kb fraction was ligated with λ gtWES phage arms and packaged *in vitro*. As the 23-kb fragment is too large to be incorporated in λ phage vectors, we digested the purified 23-kb fragment partially with $EcoRI^*$ under conditions that allowed one cut per 23-kb fragment, and we ligated the partial digests with λ gtWES phage arms. Phage plaques were screened with the J_H probe (Fig. 1).

The clone containing partial $EcoRI^*$ digests of the 23-kb fragment was designated as H·Ig·CE-1. The 5.4-kb EcoRI/HindIII fragment that hybridized to the J_H probe was subcloned into pBR322 (pCE-1). The clone containing the 11-kb EcoRI fragment was designated H·Ig·CE-114. The 4.5-kb EcoRI/HindIII fragment that hybridized to the J_H probe was further subcloned into pBR322 (pCE-114). The restriction



FIG. 1. Rearrangement of J_H segments in CESS DNA. Human placenta and CESS DNAs (2 μ g each) were digested with EcoRI (E), Xba I (X), or Sac I (S). Southern blots of the digests were hybridized with the J_H or C_{μ} probe. p, Placenta DNA; c, CESS DNA. (A) Autoradiogram using the J_H probe. (B) Autoradiogram using the C_{μ} probe. Restriction map around human J_H segments is shown at the bottom. Probes are indicated by horizontal arrows. H, HindIII.

endonuclease cleavage maps of pCE-1 and pCE-114 were constructed as shown in Fig. 2. Comparison of the restriction maps of pCE-1 and pCE-114 with that of the germline J_H segment indicates that the clones contain the rearranged J_H segments.

Nucleotide Sequence Determination of Rearranged and Germline V_H Segments. The nucleotide sequence of the V region of H·Ig·CE-1 was determined according to the strategy shown (Fig. 2). The V_{CE-1} gene contains a 14-residue-long D segment and the J_{H3} segments, both of which are in the frame with the V_H segment as shown in Fig. 3. The nucleotide sequence of the D segment does not bear significant resemblance to the known germline D sequences (23). The amino acid sequence of the V_{CE-1} gene is 82% homologous to that of Cor protein (24), which belongs to subgroup II. The results support the conclusion that the V_{CE-1} gene is the expressed V gene in CESS and is linked to the C_{γ} gene.

The nucleotide sequence of the V region of H·Ig·CE-114 was also determined according to the strategy shown (Fig. 2). The V_{CE-114} sequence is 70% homologous to Eu protein (subgroup I) in framework segments 1 region. The V_{CE-114} gene contains a 12-residue-long D segment and the J_{H4} segments (Fig. 4). However, there are three stop codons in the frame with the initiator codon AUG, two in the V_H segment



FIG. 2. Restriction maps of the cloned V_H segments. Restriction maps of pCE-1, pCE-114, and pV₂₀₁. Horizontal rectangles indicate the subclones of the V_H segments. Wider open, widest open, and wider closed rectangles show V_H , D, and J_H segments, respectively. Smaller segments containing structural sequences are enlarged below each clone. The directions and ranges of nucleotide-sequence determination are shown by horizontal arrows. Nucleotide sequence determination was carried out by labeling the 5' ends of each fragment. The fragment A is a probe used for cloning the germline clone H·IgV-201. Restriction sites are abbreviated as follows; E, EcoRI; H, HindIII; B, BamHI; Bg, Bg/II; Hh, Hha I; P, Pst I; Hf, HinfI; F, Fnu4HI; T, Taq I. bp, Base pairs.

:	20	40	60	80	100
GAAGATCCTGTCTGTGCGT	CAAGACTTGACCGAGAGC	CATGGTGCTGAAATGAG	ATAGATTTCCTGATGG	AGAGCACACĠTGGACTTCC	ACACCTGAGGG 100
CTCACTGCTCCTCACCACA	GATGCACTCCCCTACTGA	STCCTGAGACCTGAGT	CACCCCATAGAGTAGG	GCTCAGATGAGGGGGATGCA	AATCTCCACCA 200
GCTCCACCCTCCCCTGGGT	ICAAAAXGAGGACAGGGC(CTCGCTCAGTGAATCCT	GCTCCCACCATGGACA	TACTITIGTTCCACGCTCCT(GCTACTGACTG 300
TCCCGTCCTGTGAGTGCTG V P S	IGGTCAGGTAGTACTTCA	SAAGCAAAAAATCTATI	CTCTCCTTTGTGGGGCT	TCATCTTCTTATGTCTTCT	CCACAGGGGTC 400
TTATCCCAGGTCAACTTAA	GGGAGTCTGGTCCTGCGC R E S G P A I	IGGTGAAAGCCACACAT	ACCCTCACACTGACCT	GCACCTICTCTGGGTTGTC/ C T F S G L S	AGTCAACACTC 500 V N T
GTGGAATGTCTGTGAGCTG R G M S V S W	GATCCGTCAGCCCCCAGG	GAAGGCCCTGGAGTGGC K A L E W	CTTGCACGCATTGATTG	GGATGATGATAAGTACTAC D D D K Y Y	GGTACATCTCT 600 G T S L
GGAGACTAGGCTCACCATC E T R L T I	ICCAAGGACACCTCTAAA S K D T S K	ACCAGGTGGTCCTTA/ N Q V V L	AGTGACCAACATGGAC	CCTGCGGACACAGCCACGT	ATTACTGTGCG 700 Y Y C A
CGGATGCAGGTTACTATGG	TTCGGGAAGTTATGATAA V R E V M I	CGTCTAATGCTTTTGAT	ATCTGGGGGCCAAGGGA I W G Q G	CATGGTCACCGTCTCTTCA T M V T V S S	GGTAAGATGAC 800
	D		J3		

FIG. 3. Nucleotide sequence of the expressed V_H gene in CESS. The nucleotide sequence of pCE-1 together with the amino acid sequence (one-letter code) deduced from it is shown. Vertical lines represent the borders of the V, D, and J segments.

and one in the D segment. Furthermore, when compared with known V_H region sequences, the V_H segment has large deletions, encoding only 67 residues. The D segment sequence is different from that of V_{CE-1} and known germline D sequences (23). It would be interesting to know whether the D_{CE-114} sequence is a pseudogene in the germline.

To locate the precise positions of deletions, we cloned a germline V_H segment similar to the V_{CE-II4} gene from a phage library containing human DNA. The clone H·IgV·201 was isolated by screening the library with a *Hha* I fragment (probe A in Fig. 2) of V_{CE-II4} as probe. The 3.7-kb *Hind*III fragment hybridizing with the V_{CE-II4} probe was subcloned into pBR322 (pV₂₀₁), and the nucleotide sequence of the germline V_{201} segment was determined by the strategy shown (Fig. 2). The sequence of the V_{201} segment was aligned with that of the V_{CE-II4} gene (Fig. 4). Two sequences are 87% homologous to each other, but obviously the V_{201} segment is not the germline segment of the V_{CE-II4} gene, because the restriction sites in the 5' flanking regions of the two DNAs differ considerably. Nonetheless, comparison of the two sequences allowed us to locate the precise positions

of deletions in the V_{CE-114} gene. The 5' deletion is 57 bases long, located at positions 323–379; the 3' deletion is 34 bases long, located at positions 440–473. We have read through both deletions on two strands of pCE-114.

 V_{CE-1} , $V_{CE-1/4}$, and V_{201} are interrupted at codon -4 by introns of 86, 84, and 85 base pairs, respectively. Rechavi *et al.* (9) reported that human V_H segments of subgroups I and III had introns of 84 and 103 base pairs, respectively, at codon -4. Since $V_{CE-1/4}$ and V_{201} belong to subgroup I, our results are consistent with their report, except for a slightly longer intron. A human subgroup II V_H segment (V_{CE-1}) is more homologous to the subgroup I V_H segment than the subgroup III V_H segment in the length of an intron interrupting the leader sequence.

Comparison of Amino Acid Sequences. Three amino acid sequences were aligned according to Kabat *et al.* (1) as shown in Fig. 5. V_{CE-114} and V_{201} sequences contain termination codons at an identical place, residue 23. The V_{201} segment has two more termination codons at residues 38 and 50, indicating that the V_{201} segment is a pseudogene. The V_{CE-114} gene has two more termination codons at residues 93 and



FIG. 4. Nucleotide sequences of germline and rearranged pseudogenes of CESS. Nucleotide sequences of pCE-114 and pV201 are aligned to maximize homology. Amino acid sequences (one letter code) deduced from the nucleotide sequences are also shown above or below the nucleotide sequences. Dashes indicate deletions. Vertical lines indicate junctions of V, D, and J segments. Conserved nonamer and heptamer sequences (2, 3) are underlined. Asterisks indicate termination codons.



FIG. 5. Comparison of amino acid sequences. Amino acid sequences of V_{CE-I} , V_{CE-II4} , and V_{20I} were aligned according to Kabat *et al.* (1). Eu protein sequence (27) was included for comparison. Dashes and asterisks indicate deletion and termination codons, respectively. Frame-shift mutations due to deletions are underlined. Framework segments (FR1-FR4) and CDR are shown by vertical lines. Residues corresponding to D and J segments are shown by horizontal arrows.

100A. The stop codon at residue 93 is due to frame-shift mutation caused by the 3' deletion. Although it is not known whether all the termination codons in the V_{CE-114} gene are encoded by the germline segments, coincidence of the termination codon at residue 23 in the V_{201} segment and the V_{CE-114} gene suggests that this termination codon is encoded in the germline segment of the V_{CE-114} gene. In addition, two large deletions suggest that the germline V segment of the V_{CE-114} gene is a pseudo V segment, although this has not been proved. If it is the case, then the $V_{H}-D-J_{H}$ rearrangement system does not seem to recognize a major portion of the V_{H} segment.

Deletion Includes CDR2. It is of particular interest that the 5' deletion of the V_{CE-114} gene includes residues 45-62 (Fig. 5), corresponding to almost the whole CDR2 region (residues 50-65) assigned by Kabat et al. (1). Wu and Kabat (8) suggested that the 14 bases in the CDR2 of a human V_H gene with homology to a known human D segment might reflect amplified diversity by gene conversion between minigenes like the D segment. The $V_{CE-1/4}$ gene that lacks CDR2 raises a possibility that the CDR2 region of a V segment is transposed to another V or minigene segment. However, there are many other interpretations, including fortuitous deletion. Krawinkel et al. (26) showed that recombination between two V_H segments resulted in hypermutation in CDR2. Extensive recombination between V segments including pseudogenes, which may reach 40% of the germline V_H segments (9), would certainly contribute to increased V segment diversity. We have recently found a germline minigene encoding only residues 57-94 in Xenopus laevis DNA (unpublished data). Further extensive analyses of germline V segments will reveal whether such deletion is common to the V segment.

Human D Segments. Siebenlist *et al.* (23) isolated four germline D segments that are well conserved. Another D segment was identified within the J_H cluster (27). Neither of the present D sequences (*CE-114* and *CE-1*) bears significant homology (<38%) with those reported above, indicating that the human D family has many other members. Since the D

sequences of V_{CE-114} and V_{CE-1} do not resemble with each other, they may represent separate subgroups of the human D family. We have recently found another different D sequence in a μ chain of an Epstein-Barr virus-transformed B cell (unpublished data).

We are grateful to Miss K. Kawabata for preparation of the manuscript. We thank Misses N. Kisogawa, Y. Utsunomiya, and K. Harato for their technical assistance. This work was supported by a grant-in-aid for special distinguished research of the Ministry of Education, Science and Culture of Japan.

- 1. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. (1983) Sequences of Proteins of Immunological Interests (National Institutes of Health, Washington, DC).
- 2. Honjo, T. (1983) Ann. Rev. Immunol. 1, 499-528.
- 3. Tonegawa, S. (1983) Nature (London) 302, 575-581.
- 4. Ohta, T. (1980) Evolution and Variation of Multigene Families (Springer, Berlin).
- Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) Science 202, 11-17.
- Cohen, J. B., Effron, K., Rechavi, G., Ben-Neriah, Y., Zakut, R. & Givol, D. (1982) Nucleic Acids Res. 10, 3353-3370.
- Baltimore, D. & Loh, D. Y. (1983) in Progress in Immunology V, eds. Yamamura, Y. & Tada, T. (Academic, New York), pp. 115-121.
 Wu, T. T. & Kabat, E. A. (1982) Proc. Natl. Acad. Sci USA
- Wu, T. T. & Kabat, E. A. (1982) Proc. Natl. Acad. Sci USA 79, 5031-5032.
- Rechavi, G., Ram, D., Glazer, L., Zakut, R. & Givol, D. (1983) Proc. Natl. Acad. Sci. USA 80, 855–859.
- Bernstein, K. E., Reddy, E. P., Alexander, C. B. & Mage, R. G. (1982) Nature (London) 300, 74-76.
- 11. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1978) Proc. Natl. Acad. Sci. USA 75, 2429-2433.
- Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) J. Exp. Med. 149, 1299-1313.
- Kabat, E. A., Wu, T. T. & Bilofsky, H. (1980) J. Exp. Med. 152, 72-84.
- Muraguchi, A., Kishimoto, T., Miki, Y., Kuritani, T., Kaieda, T., Yoshizaki, K. & Yamamura, Y. (1981) J. Immunol. 127, 412-416.

- 15. Yaoita, Y. & Honjo, T. (1980) Biomed. Res. 1, 164-175.
- Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
- Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
 Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) Cell 18, 559-568.
- 18. Leder, P., Tiemeier, D. & Enquist, L. (1977) Science 196, 175-177.
- 19. Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W. & Smithies, O. (1978) Science 202, 1279-1284.
- 20. Takahashi, N., Nakai, S. & Honjo, T. (1980) Nucleic Acids Res. 24, 5983-5991.
- 21. Benton, W. D. & Davis, R. W. (1977) Science 196, 180-182.

- 22. Maxam, A. M. & Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
- Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldmann, T. 23. & Leder, P. (1981) Nature (London) 294, 631-635.
- 24. Press, E. M. & Hogg, N. M. (1970) Biochem. J. 117, 641-660.
- 25. Cunningham, B. A., Rutishauser, V., Gall, W. E., Gottlieb, P. D., Waxdal, M. J. & Edelman, G. M. (1970) Biochemistry 9, 3161–3170.
- 26. Krawinkel, V., Zoebelein, G., Bruggemann, M., Radbruch, A. & Rajewsky, K. (1983) Proc. Natl. Acad. Sci. USA 80, 4997-5001.
- 27. Ravetch, J. V., Siebenlist, V., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) Cell 27, 583-591.