Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 2995–2999, May 1983
Genetics

Structure of the human immunoglobulin $C_\varepsilon 2$ gene, a truncated pseudogene: Implications for its evolutionary origin

(C_ε–C_α linkage/nucleotide sequence determination/common tetranucleotide/class-switch recombination)

HIROSHI HISAJIMA, YASUYOSHI NISHIDA, SUMIKO NAKAI, NAOKI TAKAHASHI*, SHINTARO UEDA†, AND TASUKU HONJO‡

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

Communicated by L. L. Cavalli-Sforza, January 28, 1983

ABSTRACT **Cloning of the overlapping DNA fragments together with Southern hybridization experiments showed the organization of the human C_ε and C_α gene cluster as 5′-C_ε2-14 kilobases-C_α1----C_ε1-13 kilobases-C_α2-3′. Comparison of the nucleotide sequences of the C_ε1 and C_ε2 genes revealed that four deletions have taken place in the C_ε2 gene and its flanking regions. The three deleted regions in the 5′ side of the C_ε2 gene are partially filled with shorter inserted sequences. One of them has removed the C_H1 and C_H2 exons and a portion of the epsilon switch (S_ε) region. The S_ε region and the C_H4 exon still retain the functional structures, whereas the C_H3 exon has been inactivated by deleting its 5′ intervening sequence necessary for splicing. The tetranucleotide T-G-G-G (or T-G-G-C), which is usually found in close proximity of the class-switch recombination sites in mouse myelomas, is located 5′ to the three deletion sites. The results imply that the mechanism responsible for the heavy chain class-switch recombination might be relevant to the evolutionary mechanism of creation of the truncated C_ε2 gene. The other deletion in the 3′ flanking region of the C_ε2 gene may be due to slipped mispairing of the short direct repeat (C-C-C-C-C) at both ends.**

The immunoglobulin heavy chain (H chain) gene family consists of the variable (V_H)-, diversity (D)-, joining (J_H)-, and constant (C_H)-region gene clusters. Analysis of the mouse and human C_H genes has revealed that the C_H gene cluster has evolved by dynamic rearrangements, including duplication, insertion, deletion, and intervening sequence (IVS)-mediated domain transfer as well as by point mutations (1–5).

The C_H genes are divided into five major classes, C_μ, C_δ, C_γ, C_ε, and C_α, and the C_γ genes are further divided into four subclasses in mouse. Cloning of a number of overlapping fragments encompassing the entire C_H gene cluster revealed the order of the mouse C_H genes as 5′-C_μ-C_δ-C_γ3-C_γ1-C_γ2b-C_γ2a-C_ε-C_α-3′ (6–10). The human C_H gene cluster contains one copy each of the C_μ and C_δ genes, at least four C_γ genes, one C_γ pseudogene, and two C_α genes (5, 11–14). Recent studies on the human C_ε gene identified at least three C_ε-related genes, one of which (C_ε1) is active, whereas the remaining genes are pseudogenes (15–17). One pseudogene (C_ε2) has been truncated by deletion events (16, 18) and the other (C_ε3) has lost its IVS (19, 20). It has been suggested that the C_ε and C_α genes are linked to each other in the human and that both genes were duplicated simultaneously (16, 18). The C_ε3 gene is translocated to chromosome 9. However, it was not known which of the two C_α genes (C_α1 and C_α2) was linked to the active C_ε1 gene. The evolutionary mechanism to create the truncated C_ε2 gene also remained an intriguing question.

To address these questions we have studied the detailed organization and structure of the C_ε and C_α genes and have shown the linkage of the C_ε and C_α genes as 5′-C_ε2-C_α1----C_ε1-C_α2-3′. We have determined the nucleotide sequences of the C_ε1 and C_ε2 genes and their flanking regions, each encompassing more than 5.8 kilobases (kb). Comparison of the two sequences allowed us to locate the exact deletion sites in the C_ε2 gene and its flanking regions. The nucleotide sequences immediately 5′ to the deletion sites are usually T-G-G-G (or T-G-G-C), which is found in close proximity of the class-switch recombination sites in mouse myelomas (21); the evolutionary implication of this is discussed.

MATERIALS AND METHODS

Isolation of Recombinant Phages and Plasmids. Bacteriophage Charon 4A (22) and plasmid pBR322 (23) were used as EK2 vectors and were propagated in *Escherichia coli* DP50supF (22) and χ1776 (24), respectively, under P3 conditions, as described by the National Institutes of Health. Screening of a phage library was done as described (25). A Charon 4A library containing partial *Hae* III/*Alu* I digests of human fetal liver DNA (26) and a mouse α-cDNA clone (pABα-1) were the generous gifts of T. Maniatis (Harvard University) and A. Bothwell and D. Baltimore (Massachusetts Institute of Technology), respectively.

Other Procedures. Southern blot hybridization of restriction endonuclease-digested DNAs was performed as described (27). DNA fragments used as probes were labeled with [α-^{32}P]-dCTP (specific activity, 2,000–3,000 Ci/mmol; 1 Ci = 3.7 × 10^{10} Bq; Radiochemical Centre, Amersham, England) by nick-translation (28). Hybridization was carried out in 1 M NaCl at 65°C as described (29). Filters were washed with 15 mM NaCl/1.5 mM sodium citrate/0.1% $NaDodSO_4$ for 90 min at 65°C with two changes of the buffer. Nucleotide sequence determination was performed according to the method of Maxam and Gilbert (30). Heteroduplexes were formed by using the formamide technique as described (31).

RESULTS

Linkage of the C_ε and C_α Genes. The human C_α gene fragment was cloned from a Charon 4A library containing the partial *Hae* III/*Alu* I digests of the human fetal liver DNA (26) by using the mouse α-cDNA clone (pABα-1) as probe. A positive

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, J, and C, variable, diversity, joining, and constant regions, respectively, of immunoglobulin chains; H chain, heavy chain; IVS, intervening sequence; S region, switch region; kb, kilobase(s).

* Present address: Division of Biology, California Institute of Technology, Pasadena, CA 91125.
† Present address: Dept. of Anthropology, Faculty of Science, The University of Tokyo, Tokyo, Japan.
‡ To whom reprint requests should be addressed.

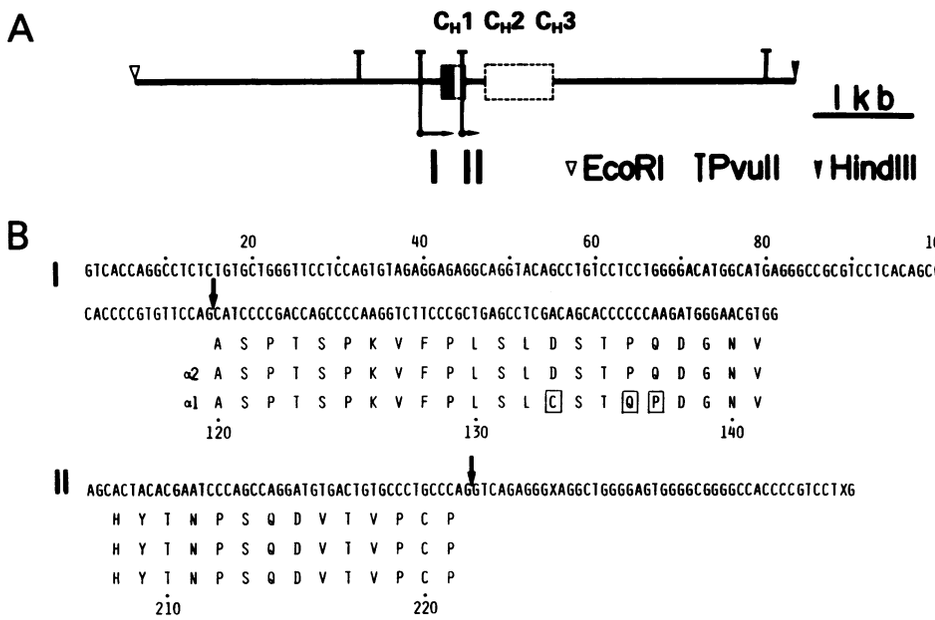

FIG. 1. Identification of the human $C_\alpha 2$ gene clone. (*A*) The 6.5-kb *Eco*RI/*Hin*dIII fragment of Ch4A·H·Igα-25 (fragment b in Fig. 2) was purified and partial nucleotide sequences were determined. The directions and ranges of sequences read are indicated by horizontal arrows I and II. The dashed box represents the tentative location of the other exons. (*B*) The nucleotide sequences of the regions I and II are shown. The amino acid sequences predicted from the nucleotide sequences were compared to those of the α2 (32) and α1 (33) chains (amino acid residues 120–141 and 208–221). The amino acid residues that differed among these three sequences are boxed. The possible splicing sites are indicated by the vertical arrows.

clone, Ch4A·H·Igα-25, was propagated and identified as the human $C_\alpha 2$ gene clone by partial nucleotide sequence determination. The nucleotide sequences predicted the amino acid sequences of the C_H1 domain (amino acid residues 120–141 and 208–221), which were identical to the corresponding regions of the known α2 chain (32) but different from those of the α1 chain (33) at three positions (Fig. 1).

Southern blot hybridization of the *Bam*HI-digested human placenta DNA produced two major bands (18 and 17 kb) hybridizing to both the 3′-flanking region probe of the $C_\varepsilon 1$ gene (fragment a in Fig. 2) and the $C_\alpha 2$ gene probe (fragment b in Fig. 2). The results suggest the linkage between the C_ε and C_α genes in the human genome and the presence of the two segments containing both C_ε and C_α genes, in agreement with previous reports (16, 18).

To prove the physical linkage of the C_ε and C_α genes, the 18- and 17-kb *Bam*HI fragments hybridizing to the C_α probe were cloned into pBR322. Detailed mapping of restriction endonuclease cleavage sites of pH·Igα-11 (18-kb fragment) and pH·Igα-21 (17-kb fragment) enabled us to conclude that pH·Igα-21 overlaps the previously isolated $C_\varepsilon 1$ gene clones (15) and the $C_\varepsilon 2$ gene clone (Igα-25), as shown in Fig. 2. The results demonstrate unequivocally the organization: 5′---$C_\varepsilon 1$-13 kb-$C_\alpha 2$---3′. Thus, the pH·Igα-11 clone was identified as the $C_\alpha 1$ clone.

The 6.5-kb *Bam*HI fragment that contains the $C_\varepsilon 2$ gene (15) was cloned into pBR322 by using the $C_\varepsilon 1$ gene as probe and was designated as pH·Igε-21. Comparative studies between the $C_\varepsilon 1$ and $C_\varepsilon 2$ genes by heteroduplex formation, restriction mapping, and Southern blot hybridization experiments revealed that at least two deletion events have taken place in the $C_\varepsilon 2$ gene (18). The 3′ deletion (2.9 kb) has removed the C_H1 and C_H2 exons from the $C_\varepsilon 2$ gene. There is another large deletion (3.7 kb) in the 5′ flanking region. The restriction map of the clone pH·Igε-21 was indistinguishable from that of the C_ε pseudogene ($\psi\varepsilon 1$) cloned previously (16).

Although both the $C_\varepsilon 2$ (pH·Igε-21) and the $C_\alpha 1$ (pH·Igα-11) gene clones are the *Bam*HI fragments and have no overlap, the direct association of Igε-21 and Igα-11 was shown by Southern blot hybridization by using the DNA fragments located at both sides of the *Bam*HI site 3′ to the $C_\varepsilon 1$ gene (fragments c and d in Fig. 2) as probes. Upon digestion of the human placenta DNA with *Xba* I, both probes hybridized to the two bands of 14 and 7 kb, the former being consistent with the restriction map of the $C_\varepsilon 1$ gene. The 7-kb fragment is expected to be derived from the $C_\varepsilon 2$ gene, assuming that Igε-21 and Igα-11 are linked directly at the *Bam*HI site. Similarly, *Sac* I digestion produced, in addition to the $C_\varepsilon 1$ gene fragment (2.9 kb), the 3.6-kb fragment hybridizing to both probes, which again coincides with the $C_\varepsilon 2$ gene fragment expected from the above assumption. The results demonstrate the direct linkage of the Igε-21 and Igα-11 clones, although the presence of a few hundred-base fragment between the two clones is not ruled out. We conclude that the $C_\varepsilon 2$ gene is located about 14 kb 5′ to the $C_\alpha 1$ gene. This conclusion also is supported by strong homology of restriction maps between the $C_\varepsilon 1$–$C_\alpha 2$ and $C_\varepsilon 2$–$C_\alpha 1$ linkage groups (Fig. 2). Because the $C_\varepsilon 2$ gene was deleted in the $C_\varepsilon 1$-expressing myeloma, the C_ε gene order was supposed to be 5′-$C_\varepsilon 2$-$C_\varepsilon 1$-3′ (5, 17). Taken together, we propose the organization: 5′-$C_\varepsilon 2$-14 kb-$C_\alpha 1$---$C_\varepsilon 1$-13 kb-$C_\alpha 2$-3′.

Nucleotide Sequences at the Deletion Sites of the $C_\varepsilon 2$ Gene. To understand the evolutionary mechanism of creation of the truncated pseudogene, the nucleotide sequences around the deletion borders in the $C_\varepsilon 2$ gene and those of the corresponding regions of the $C_\varepsilon 1$ gene were determined according to the

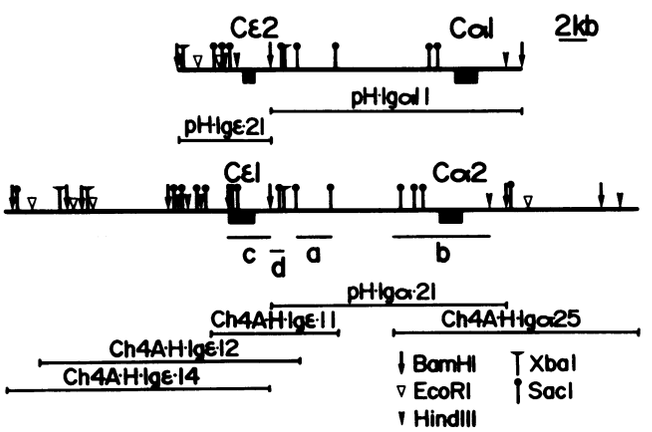

FIG. 2. Linkage of the C_ε and C_α genes. The restriction maps of the segments linking the C_ε and C_α genes were constructed from those of the overlapping clones. Cloning of Ch4A·H·Igε-11, Ch4A·H·Igε-12, and Ch4A·H·Igε-14 has been described (15). Closed boxes represent the coding regions. Fragments used as probes are indicated by horizontal bars.

strategy shown in Fig. 3. Comparison of the nucleotide sequences around the 3′ deletion region identified the breakpoints of the homology segments at nucleotide positions 5,993 and 8,049, revealing a 21-base insertion (positions 5,994–6,014) between the two borders (Fig. 4). The complete nucleotide sequence of the $C_\varepsilon 1$ gene corresponding to the 3′ deletion block of the $C_\varepsilon 2$ gene was determined (ref. 20; unpublished data) but this 21-base sequence was not found in the $C_\varepsilon 1$ gene. Instead, the 20-base sequence directly 5′ to the deletion point is highly homologous to the 21-base insertion sequence (85.7% homology), suggesting that the 21-base sequence might have been inserted by duplication of the sequence 5′ to the deletion site.

Nucleotide sequence analysis around the junction of the 5′ deletion region (positions 566–4,477) revealed once again the presence of a 13-base insertion sequence (positions 566–578) in the $C_\varepsilon 2$ gene. The origin of this sequence is not yet known because the nucleotide sequence of the 3.5-kb region of the $C_\varepsilon 1$ gene, which is deleted in the $C_\varepsilon 2$ gene, has not been determined. The nucleotide sequence determination revealed another DNA rearrangement of the $C_\varepsilon 2$ gene, in which a 149-base sequence (positions 146–299) of the $C_\varepsilon 2$ gene is replaced by a 43-base sequence in the $C_\varepsilon 1$ gene. The latter sequence showed considerable homology (67%) to a portion of the 149-base sequence (portions 176–224).

It is striking that the tetranucleotide T-G-G-G (or T-G-G-C) was always found immediately 5′ to each of the three deletion and insertion sites. The tetranucleotide also was found 5′ to the 20-base sequence that was supposed to have been duplicated at the 5′ border of the 3′ deletion region.

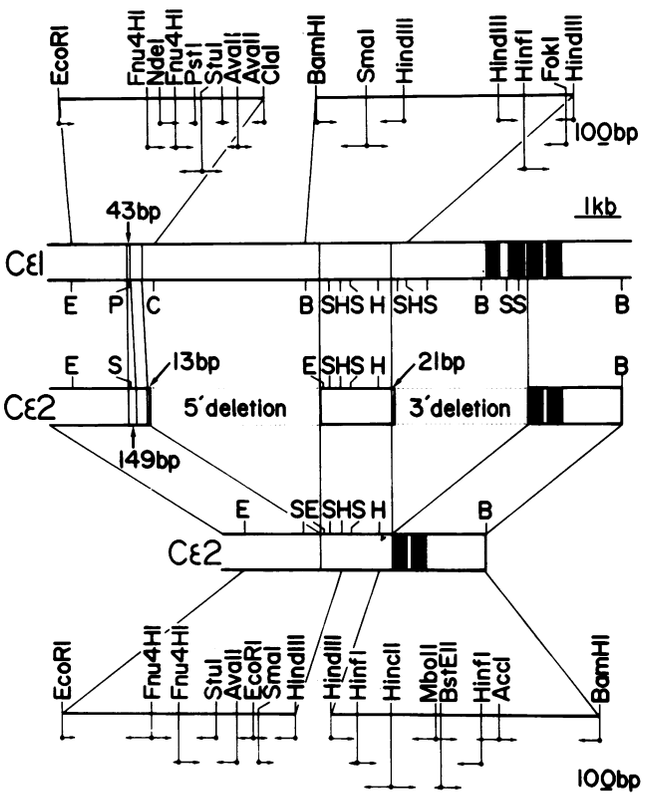

FIG. 3. Comparison of the restriction maps of the human $C_\varepsilon 1$ and $C_\varepsilon 2$ genes and strategies for nucleotide sequence analysis. The deleted regions in the $C_\varepsilon 2$ gene are represented by dotted rectangles. Shaded rectangles represent exons. The locations of the low-homology segments of 43 bases in the $C_\varepsilon 1$ gene and of 149 bases in the $C_\varepsilon 2$ gene are indicated by arrows. Strategies for the nucleotide sequence analysis of the $C_\varepsilon 1$ and $C_\varepsilon 2$ genes are represented at the top and bottom rows, respectively. The directions and ranges of the sequences read are shown by horizontal arrows. Restriction sites are abbreviated as follows: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sac* I; P, *Pst* I; and C, *Cla* I. bp, Base pairs.

To test whether or not other small deletions are found in the $C_\varepsilon 2$ gene, the 2.3-kb *Hin*dIII/*Bam*HI fragment of Igε-21 containing the pseudo-coding region was isolated and its nucleotide sequence was determined according to the strategy shown in Fig. 3. We found another small 24-base deletion (positions 9,270–9,293) in the 3′ flanking region of the $C_\varepsilon 2$ gene (Fig. 4). In this case we could not find the tetranucleotide sequence T-G-G-G at the deletion borders. Instead, direct repeats of C-C-C-C-C were found at the borders, suggesting that this deletion might have involved a different mechanism from those at the 5′ side of the $C_\varepsilon 2$ gene. The sequence 5′ to the C_H3 exon of the $C_\varepsilon 2$ gene contains a tandem array of the pentanucleotides T-G-A-G-C, T-G-G-G-G, and their related sequences, which are the common short sequences found in the mouse *S* (switch) regions (21, 34–37). Thus, in the truncated $C_\varepsilon 2$ gene, the C_H3 exon has been associated directly with the remaining S_ε region.

Because the 3′ deletion has removed the first nucleotide of the C_H3 exon together with the 3′ splicing signal of the intron between the C_H2 and C_H3 exons, the C_H3 exon is not able to be spliced. On the other hand, the C_H4 exon and the 3′ untranslated sequence remain intact. The sequence downstream from the 5′ end of the C_H3 exon is highly homologous (97.6% homology) to that of the $C_\varepsilon 1$ gene. As compared with the $C_\varepsilon 1$ gene, six nucleotide substitutions have taken place within the C_H4 exon of the $C_\varepsilon 2$ gene, four of which—nucleotide positions 8,503, 8,529, 8,655, and 8,737—have replaced the amino acid residues Pro to Leu, Ala to Thr, Val to Ile, and Ser to Ile, respectively. Thus, the $C_\varepsilon 2$ gene still retains the potentially functional S_ε region and the C_H4 exon. The present sequence disagrees at four bases with partial sequences of the pseudo-coding region determined by another group (16).

DISCUSSION

The organization of the mouse C_H genes was shown as 5′-C_μ-C_δ-C_γ-C_ε-C_α-3′ (6–10). Because several human linkage groups, C_μ–C_δ(13), $C_\gamma 2$–$C_\gamma 4$(5), $C_\varepsilon 2$–$C_\alpha 1$, and $C_\varepsilon 1$–$C_\alpha 2$, have been shown, the general order of the human C_H genes might be similar to the mouse C_H gene order.

The tetranucleotide T-G-G-G (or T-G-G-C) found immediately 5′ to each of the deletion sites in the 5′ flanking region of the $C_\varepsilon 2$ gene (Fig. 4) is reminiscent of the short common sequences (T-G-G-G-G and T-G-A-G-C) shared by *S* regions, which may be responsible for the class-switch recombination (21, 34–37). Furthermore, the tetranucleotides T-G-G-G and T-G-A-G are usually found in close proximity to the recombination sites in mouse myelomas (21). Based on these observations we proposed previously that T-G-G-G or T-G-A-G might be a part of the recognition sequence for the class-switch recombination. The occurrence of the nucleotide sequence T-G-G-G (or T-G-G-C) at the borders of the deletion sites implies that the mechanism responsible for the class-switch recombination might have been involved in the creation of the $C_\varepsilon 2$ gene or that different enzymes, but recognizing the same tetranucleotide T-G-G-G, might have been responsible for the recombination to create the $C_\varepsilon 2$ gene.

The class-switch recombination is a somatic event, whereas those rearrangements found in the $C_\varepsilon 2$ gene must have occurred in the germ-line cell in order to be fixed in the human genome. Although the somatic class-switch recombination appears unlikely to take place in the germ cell, we can envisage that the enzyme(s) responsible for the class-switch recombination might have been activated somehow in the early embryonic stage. In fact, an early embryonic cell would be more

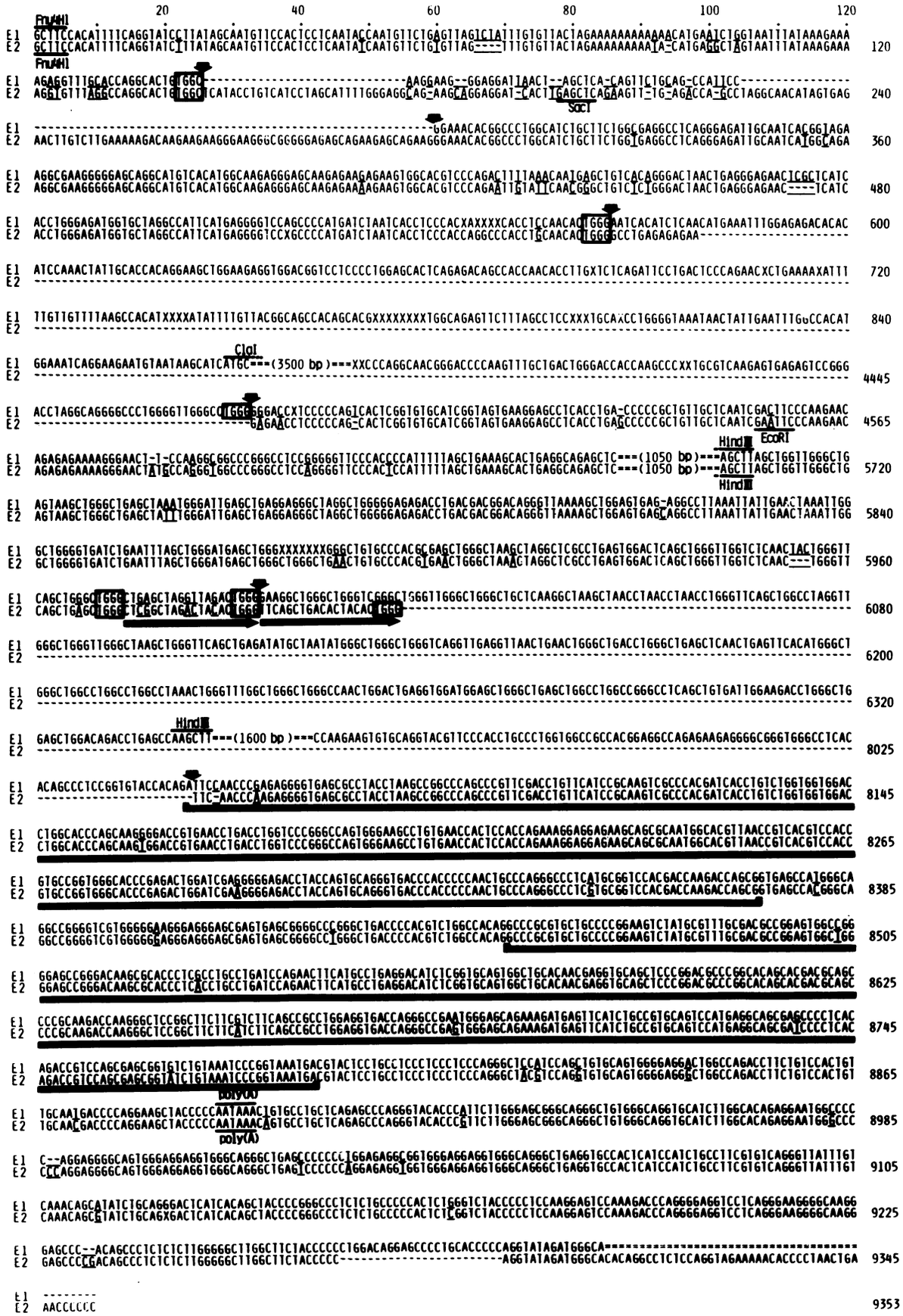

FIG. 4. Comparison of the nucleotide sequences of the human $C_\epsilon 1$ and $C_\epsilon 2$ genes. E1 and E2 represent the nucleotide sequences of the $C_\epsilon 1$ and $C_\epsilon 2$ genes, respectively. The sequence of the $C_\epsilon 1$ coding region has been reported (20). Deleted nucleotides are indicated by hyphens. The borders of the homology segments are indicated by vertical arrows. The direct repeat sequences at the 3′ deletion site are indicated by the horizontal arrows. The tetranucleotide sequences commonly found at the deletion borders are boxed. The sequences coding for the C_H3 and C_H4 exons are underlined.

efficient than a single germ cell for the inheritance of DNA rearrangements because some of early embryonic cells could be precursors to the germ cell.

Small deletions in the noncoding and coding regions of the β-like globin genes generated during evolution were suggested to be due to slipped mispairing of the short direct repeat sequences during DNA replication (38). The small deletion observed at the 3′ flanking region of the $C_\varepsilon 2$ gene might have been generated by a similar mechanism to that of the β-like globin genes, involving the short direct repeat sequence (C-C-C-C-C).

It is most puzzling why the truncated $C_\varepsilon 2$ pseudogene predominates in the human population (15, 18). Is this change neutral or advantageous? If it is neutral, it means that the amount of IgE produced by a single C_ε gene is enough and that twice the amount of IgE produced by two C_ε genes is not harmful. If the $C_\varepsilon 2$ gene is advantageous, it means that the presence of two C_ε genes might have been problematic for humans. The subsequent inactivation of one of the C_ε genes by deletion might have some advantages over two active C_ε genes.

The $C_\varepsilon 2$ gene is a truncated pseudogene unable to code for the ε chain. However, it still retains the potentially functional S_ε region and C_H4 exon. Theoretically, the $C_\varepsilon 2$ gene might be activated by the class-switch recombination to express a short polypeptide containing only the V_H and C_H4 domains. Although such a molecule has not been detected in human sera, it might be interesting to test whether it can compete with IgE for the binding to the IgE receptor on the mast cell surface.

We are grateful to Dr. T. Maniatis (Harvard University) for a human DNA library and to Drs. A. Bothwell and D. Baltimore (Massachusetts Institute of Technology) for a mouse α-cDNA clone. We thank F. Oguni for her excellent assistance in preparing the manuscript. This investigation was supported in part by grants from the Ministry of Education, Science and Culture and from the Ministry of Health and Welfare of Japan.

1. Honjo, T., Nakai, S., Nishida, Y., Kataoka, T., Yamawaki-Kataoka, Y., Takahashi, N., Obata, M., Shimizu, A., Yaoita, ı., Nikaido, T. & Ishida, N. (1981) *Immunol. Rev.* **59**, 33–67.
2. Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2143–2147.
3. Yamawaki-Kataoka, Y., Miyata, T. & Honjo, T. (1981) *Nucleic Acids Res.* **9**, 1365–1381.
4. Shimizu, A., Hamaguchi, Y., Yaoita, Y., Moriwaki, K., Kondo, K. & Honjo, T. (1982) *Nature (London)* **298**, 82–84.
5. Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S. & Honjo, T. (1982) *Cell* **29**, 671–679.
6. Liu, C.-P., Tucker, P. W., Mushinski, G. F. & Blattner, F. R. (1980) *Science* **209**, 1348–1353.
7. Shimizu, A., Takahashi, T., Yamawaki-Kataoka, Y., Nishida, Y., Kataoka, T. & Honjo, T. (1981) *Nature (London)* **289**, 149–153.
8. Nishida, Y., Kataoka, T., Ishida, N., Nakai, S., Kishimoto, T., Böttcher, I. & Honjo, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1581–1585.
9. Roeder, W., Maki, R., Traunecker, A. & Tonegawa, S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 474–478.
10. Shimizu, A., Takahashi, N., Yaoita, Y. & Honjo, T. (1982) *Cell* **28**, 499–506.
11. Ravetch, J. V., Kirsch, I. R. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6734–6738.
12. Takahashi, N., Nakai, S. & Honjo, T. (1980) *Nucleic Acids Res.* **8**, 5983–5991.
13. Rabbitts, T. H., Forster, A. & Milstein, C. P. (1981) *Nucleic Acids Res.* **9**, 4509–4524.
14. Krawinkel, U. & Rabbitts, T. H. (1982) *EMBO J.* **1**, 403–407.
15. Nishida, Y., Miki, T., Hisajima, H. & Honjo, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3833–3837.
16. Max, E. E., Battey, J., Ney, R., Kirsch, I. R. & Leder, P. (1982) *Cell* **29**, 691–699.
17. Flanagan, J. G. & Rabbitts, T. H. (1982) *EMBO J.* **1**, 655–660.
18. Nishida, Y., Ueda, S., Takahashi, N., Nakai, S., Hisajima, H. & Honjo, T. (1982) in *Proceedings of the Takeda Science Foundation Symposium*, ed. Yamamura, Y. (Academic, New York), in press.
19. Battey, J., Max, E. E., McBride, W. O., Swan, D. & Leder, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5956–5960.
20. Ueda, S., Nakai, S., Nishida, Y., Hisajima, H. & Honjo, T. (1982) *EMBO J.* **1**, 1539–1544.
21. Nikaido, T., Yamawaki-Kataoka, Y. & Honjo, T. (1982) *J. Biol. Chem.* **257**, 7322–7329.
22. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Scheldon, E. L. & Smithies, O. (1977) *Science* **196**, 161–169.
23. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynecker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2**, 95–113.
24. Curtiss, R., III, Inoue, M., Pereira, D., Hsu, J. C., Alexander, L. & Rock, L. (1977) in *Molecular Cloning of Recombinant DNA*, eds. Scott, W. A. & Werner, R. (Academic, New York), p. 99.
25. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
26. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C. & Quon, D. (1978) *Cell* **15**, 687–701.
27. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
28. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
29. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
30. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
31. Davis, R. W., Simon, M. & Davidson, N. (1971) *Methods Enzymol.* **21**, 413–428.
32. Torano, A. & Putnam, F. W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 966–969.
33. Low, T. L. K., Liu, Y. S. V. & Putnam, F. W. (1976) *Science* **191**, 350–392.
34. Kataoka, T., Kawakami, T., Takahashi, N. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 919–923.
35. Kataoka, T., Miyata, T. & Honjo, T. (1981) *Cell* **23**, 357–368.
36. Obata, M., Kataoka, T., Nakai, S., Yamagishi, H., Takahashi, N., Yamawaki-Kataoka, Y., Nikaido, T., Shimizu, A. & Honjo, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2437–2441.
37. Nikaido, T., Nakai, S. & Honjo, T. (1981) *Nature (London)* **292**, 845–848.
38. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slighton, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudhood, N. J. (1980) *Cell* **21**, 653–668.