The EMBO Journal Vol.1 No.12 pp.1539–1544, 1982

Long terminal repeat-like elements flank a human immunoglobulin epsilon pseudogene that lacks introns

Shintaro Ueda[1], Sumiko Nakai, Yasuyoshi Nishida, Hiroshi Hisajima, and Tasuku Honjo*

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

Communicated by K.Rajewsky
Received on 30 September 1982

There are at least three immunoglobulin epsilon genes ($C_{\epsilon1}$, $C_{\epsilon2}$, and $C_{\epsilon3}$) in the human genome. The nucleotide sequences of the expressed epsilon gene ($C_{\epsilon1}$) and one ($C_{\epsilon3}$) of the two epsilon pseudogenes were compared. The results show that the $C_{\epsilon3}$ gene lacks the three intervening sequences entirely and has a 31-base A-rich sequence 16 bases 3′ to the putative poly(A) addition signal, indicating that the $C_{\epsilon3}$ gene is a processed gene. The $C_{\epsilon3}$ gene sequence is homologous to the five separate DNA segments of the $C_{\epsilon1}$ gene; namely, a segment in the 5′-flanking region (100 bases) and four exons, which are interrupted by a spacer region or intervening sequences. Long terminal repeat (LTR)-like sequences which contain TATAAA and AATAAA sequences as well as terminal inverted repeats are present in both 5′- and 3′-flanking regions. The 5′ and 3′ LTR-like sequences do not, however, constitute a direct repeat, unlike transposable elements of eukaryotes and retroviruses. The 3′ LTR-like sequence is repetitive in the human genome, but is not homologous to the Alu family DNA. Models for the evolutionary origin of the processed gene flanked by the LTR-like sequences are discussed. The $C_{\epsilon3}$ gene has a new open frame which codes potentially for an unknown protein of 292 amino acid residues.

Key words: heteroduplex/nucleotide sequence/processed gene/LTR-like sequence/new open frame

Introduction

Immunoglobulin heavy chain constant region (C) genes are arranged in the mouse genome in the order 5′-C_{μ}-C_{δ}-$C_{\gamma3}$-$C_{\gamma1}$-$C_{\gamma2b}$-$C_{\gamma2a}$-C_{ϵ}-C_{α}-3′ (Liu *et al.*, 1980; Nishida *et al.*, 1981; Roeder *et al.*, 1981; Shimizu *et al.*, 1981, 1982; Takahashi *et al.*, 1981). Structural comparison of these genes has shown that they evolved through gene duplications and interverning sequence (IVS)-mediated domain transfer events (Miyata *et al.*, 1980; Yamawaki-Kataoka *et al.*, 1981, 1982). There is no extensively conserved pseudogene of the heavy chain genes in the mouse genome. In contrast, the human heavy chain gene family contains several pseudogenes of the gamma and epsilon genes (Max *et al.*, 1982; Nishida *et al.*, 1982; Takahashi *et al.*, 1982). There are two epsilon pseudogenes ($C_{\epsilon2}$ and $C_{\epsilon3}$) in addition to the expressed C_{ϵ} gene ($C_{\epsilon1}$) (Flanagan and Rabbitts, 1982; Max *et al.*, 1982; Nishida *et al.*, 1982) indicating that the C_{ϵ} locus was a target of frequent genetic rearrangements.

Several pseudogenes have been reported since the first was described in the 5S RNA gene of *Xenopus laevis* (Jacq *et al.*, 1977). Recently several so-called processed pseudogenes that lack the entire IVS have been found in the gene families of the mouse α-globin (Nishioka *et al.*, 1980; Vanin *et al.*, 1980), the human immunoglobulin lambda chain (Hollis *et al.*, 1982), and the human β-tubulin (Wilde *et al.*, 1982a, 1982b). The mouse α-globin processed gene is flanked by long terminal repeats (LTRs) of retrovirus-like intracisternal A particles on both sides, although their orientation is opposite to each other (Lueders *et al.*, 1982). The human processed genes described above have poly(A)-like tails ~20 bases 3′ to the putative poly(A) addition signal and are flanked by direct repeats of several bases on both sides (Hollis *et al.*, 1982; Wilde *et al.*, 1982a, 1982b). Such direct repeats, which were also found in human small nuclear RNA pseudogenes (Arsdell *et al.*, 1981), might have been formed by repair of staggered chromosomal breaks following insertion of pseudogenes into chromosomes. Several models have been proposed to account for the generation of such pseudogenes, including gene conversion by mRNA-DNA heteroduplex formation (Nishioka *et al.*, 1980), the insertion of cDNA into a staggered chromosomal break following the reverse transcription of mRNA (Hollis *et al.*, 1982), and insertion *via* a retroviral RNA intermediate with LTR (Lueders *et al.*, 1982).

Here, we report the nucleotide sequences of the human $C_{\epsilon1}$ and $C_{\epsilon3}$ genes and their flanking regions. Comparison of the $C_{\epsilon1}$ and $C_{\epsilon3}$ sequences indicates that the $C_{\epsilon3}$ gene lacks all of the three intervening sequences and is flanked by a 31-base A-rich sequence 16 bases 3′ to the poly(A) addition signal, indicating that it is a processed pseudogene. In addition, the $C_{\epsilon3}$ pseudogene is flanked by LTR-like elements with a direct repeat on both ends. Such a structure may suggest that the LTR-like sequence might have been involved for the generation of the $C_{\epsilon3}$ pseudogene.

Results

Comparison of the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes by heteroduplex analyses and nucleotide sequence determination

The human $C_{\epsilon1}$ and $C_{\epsilon3}$ genes (Ch4A·H·Igε-11 and WES·H.Igε-31, respectively) were isolated previously from DNA of an IgE-producing myeloma cell line 266B1 (Nishida *et al.*, 1982). The germline $C_{\epsilon1}$ gene (Ch4A·H·Igε-12) was subsequently isolated from a library containing partial *Alu*I-*Hae*III digests of human fetal liver DNA. We have examined the homology of the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes by heteroduplex analyses as shown in Figure 1. Heteroduplexes formed between the $C_{\epsilon3}$ and germline $C_{\epsilon1}$ cloned DNAs showed five homologous segments, four of which seem to correspond to the exons of the $C_{\epsilon1}$ gene from their locations and lengths. The fifth homologous segment (F in Figure 1) of ~100 bases is located 0.7 kb upstream of the CH1 exon of the $C_{\epsilon1}$ gene. The five segments seem to be separated in the $C_{\epsilon1}$ gene but contiguous in the $C_{\epsilon3}$ gene. These observations suggest that the $C_{\epsilon3}$ gene was created by the fusion of five separate segments, probably a segment in the 5′-flanking region and four exons of the prototype epsilon gene i.e., the ancestor of the $C_{\epsilon1}$ gene.

To compare the nucleotide sequences of the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes we have determined the nucleotide sequences of the

[1] Present address: Department of Anthropology, Faculty of Science, The University of Tokyo, Tokyo, Japan.

*To whom reprint requests should be sent.

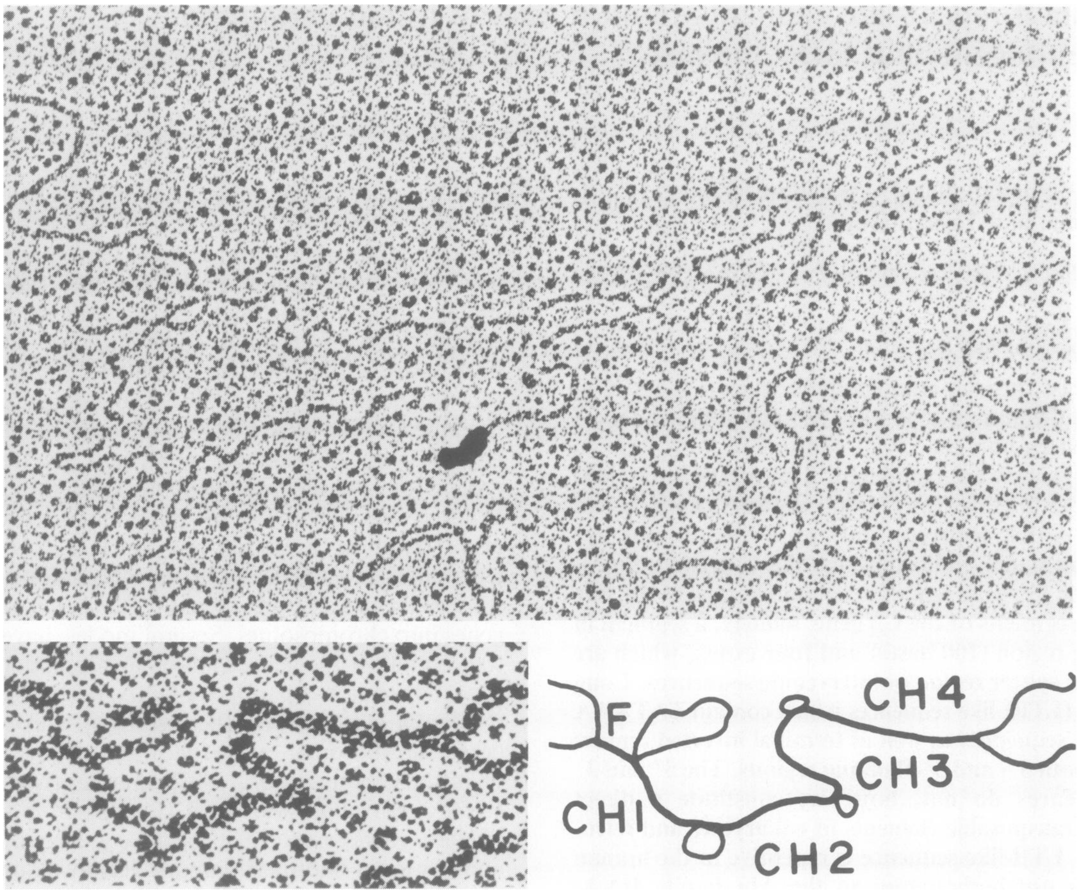

Fig. 1. Heteroduplex analyses between the $C_{\epsilon 3}$ and the germline $C_{\epsilon 1}$ genes. An electron-micrograph and diagram are shown. 31 molecules were measured. Lengths (kb) of duplexed regions are as follows; F, 0.11 ± 0.02; CH1, 0.31 ± 0.04; CH2, 0.34 ± 0.05; CH3, 0.34 ± 0.02; and CH4, 0.47 ± 0.05. The single-stranded loop between F and CH1 is 0.66 ± 0.1 kb.

structural and the flanking regions of the two genes according to the strategies shown in Figure 2. The nucleotide sequences of the $C_{\epsilon 1}$ and $C_{\epsilon 3}$ genes and the amino acid sequence predicted from the nucleotide sequence of the $C_{\epsilon 1}$ gene are shown in Figure 3. The nucleotide sequence of the $C_{\epsilon 1}$ gene was derived from the rearranged $C_{\epsilon 1}$ gene except that the germline gene sequence was used for the 5′-flanking region where the rearrangement took place in the myeloma (Nishida *et al.*, 1982). The human $C_{\epsilon 1}$ gene consists of four exons each encoding a domain and three intervening sequences like the mouse C_{ϵ} gene (Ishida *et al.*, 1982). The 5′ end of the CH1 domain of the $C_{\epsilon 1}$ gene was tentatively determined according to a consensus sequence for the 3′ splicing site (PyNPyPyPyNCAG) and because all the splicing sites fall between the first and the second bases of a codon in the immunoglobulin genes (Sharp, 1981) although the amino acid sequence of this region is not known (Bennich and von Bohr-Lindstrom, 1974).

Comparison of the nucleotide sequences of the $C_{\epsilon 1}$ and $C_{\epsilon 3}$ genes revealed the complete absence of the three IVSs from the $C_{\epsilon 3}$ gene and the presence of a 31-base A-rich sequence 16 bases downstream of the poly(A) addition signal (AATAAA) of the $C_{\epsilon 3}$ gene. A similar finding was made in another laboratory (P.Leder, personal communication). The sequence of the $C_{\epsilon 3}$ gene is >80% homologous to that of the $C_{\epsilon 1}$ gene in the coding region without counting the 26-base deletion in the $C_{\epsilon 3}$ gene. The 5′-flanking region of the $C_{\epsilon 3}$ gene does not have any sequences homologous to the V, D, or J segment. The nucleotide sequence of the 100-base region 755 bases upstream of the CH1 exon of the $C_{\epsilon 1}$ gene is homologous (85%) to that of the region immediately 5′ to the pseudo-CH1 domain of the $C_{\epsilon 3}$ gene. The sequence (GTGGG) immediately 3′ to this homology region of the $C_{\epsilon 1}$ gene is similar to the consensus sequence for the 5′ splicing site (GTPuAG). Outside these homology regions the two sequences are different from each other. These structural features of the $C_{\epsilon 3}$ gene are reminiscent of processed gene pseudogenes previously described in several gene families (Nishioka *et al.*, 1980; Vanin *et al.*, 1980; Hollis *et al.*, 1982; Wilde *et al.*, 1982a, 1982b).

The amino acid sequence predicted from the nucleotide sequence of the $C_{\epsilon 1}$ gene is consistent with that determined from the protein analysis except for 17 residues out of 427 residues; eight substitutions, eight insertions, and one deletion (Bennich and von Bohr-Lindstrom, 1974). Our $C_{\epsilon 1}$ sequence contains much longer flanking regions than those published recently (Flanagan and Rabbitts, 1982; Max *et al.*, 1982). Our sequence differs from that of the germline $C_{\epsilon 1}$ gene (Max *et al.*, 1982) at 13 positions which are located in flanking regions and IVS or at the silent positions except that G is replaced by T at position 1549, resulting in the amino acid change from Trp to Leu. At this position our sequence is consistent with the published amino acid sequence of the myeloma protein. Our sequence agrees with that of the $C_{\epsilon 1}$ clone derived from the same myeloma (Flanagan and Rabbitts, 1982) except for two positions in the flanking regions.

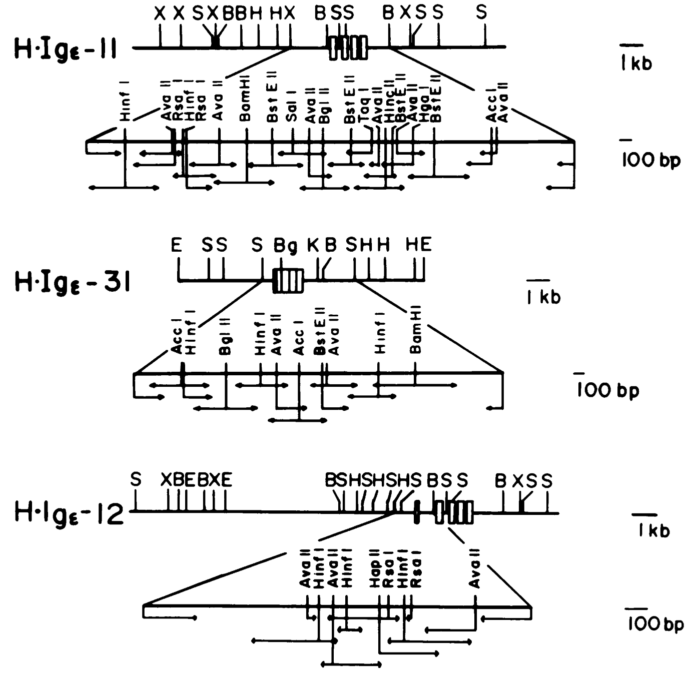

Fig. 2. Restriction cleavage site maps and sequencing strategies for the human $C_{\epsilon1}$ and $C_{\epsilon3}$ genes. The 3.0-kb *Bam*HI fragment of H·Igε-11 containing the $C_{\epsilon1}$ gene was subcloned in plasmid pBR322 and the partial nucleotide sequences of the C_{H2} and C_{H3} exons were determined previously (Nishida *et al.*, 1982). We subcloned the 4.0-kb *Bam*HI fragment of H·Igε-11 (rearranged type) and the 1.9-kb *Hind*III-*Bam*HI fragment of H·Igε-12 (germline type) containing the 5′-flanking regions of the $C_{\epsilon1}$ gene and the two *Eco*RI-*Bam*HI fragments (4.5 and 6.1 kb) of H·Igε-31. Open squares represent the C_H exons and 5′-flanking region homologous between the two genes. The directions and ranges of sequence determination are shown by horizontal arrows. Only restriction endonuclease cleavage sites used for sequencing are shown.

LTR-like sequences flank the $C_{\epsilon3}$ gene

A remarkable feature of the $C_{\epsilon3}$ gene is the presence of LTR-like elements in the flanking regions (Figure 3). The retroviral LTR is repeated at the ends of the viral genome and includes sequences that may function as promoters and terminators of transcription. The repeated sequence on either side is surrounded by inverted repeat sequences. A pair of sequences sandwiches the processed gene, which bears a close resemblance to retroviral LTR except that it does not constitute a direct repeat. The LTR-like sequences consist of a pair of 245- and 164-base sequences located 5′ and 3′ to the pseudo-coding region, respectively. Each unit contains both a TATA box and AATAAA sequences as well as terminal inverted repeats. The paired unit sequences have the same terminal inverted repeat sequences $[(T)_6$-$(A)_6]$. One unit is located immediately 5′ to the DNA segment homologous between the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes and the other is located 184 bases 3′ to the 31-base A-rich sequence. Moreover, the 5′ and 3′ units of LTR-like sequences are flanked immediately 5′ and 3′ to them, respectively, by a short direct repeat sequence (AGCT). Unlike retrovirus LTR, however, the internal sequences of the paired units are not homologous. The structures of the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes are shown schematically in Figure 4.

The distribution of the 5′ and 3′ LTR-like sequences in the human genome

The distribution of the 5′ and 3′ LTR-like sequences in the human genome was examined by Southern blot hybridization of restriction fragments of human placental DNA. The *Sac*I-*Hinf*I fragment (positions −1131 to −697 in Figure 3) and the *Hinf*I-*Bam*HI fragment (positions 2142−2468 in Figure 3) of H·Igε-31 DNA ($C_{\epsilon3}$) were used as the 5′ and 3′ LTR-like sequence probes, respectively. The *Acc*I fragment (positions −712 to 1451 in Figure 3) was used as the $C_{\epsilon3}$ structural probe which cross-hybridized weakly with the $C_{\epsilon1}$ gene fragment. Under stringent hybridization conditions, the 5′ LTR-like probe did not hybridize to any other fragments of the *Eco*RI and *Bam*HI digests than those hybridized with the $C_{\epsilon3}$ structural probe as shown in Figure 5. However, the 5′ LTR probe hybridized to the other *Hind*III fragment (6.7 kb) that did not hybridize with the $C_{\epsilon3}$ structural probe. The results indicate that there is at least one other sequence closely related to the 5′ LTR-like sequence in the human genome. Under mild conditions the 5′ LTR-like probe hybridized to several *Eco*RI fragments (21, 11, 8.6, and 7.2 kb), indicating that there are several sequences distantly related to the 5′ LTR-like sequence (data not shown). One the other hand, the 3′ LTR-like sequences in placental DNA appear blurred, indicating that the 3′ LTR-like sequence is highly repetitive in the human genome.

We have further examined whether another copy of the 5′ or 3′ LTR-like sequence is present in the vicinity of the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes using cloned DNA segments of these genes. The rearranged and germline $C_{\epsilon1}$ cloned DNAs were digested with *Bam*HI and the $C_{\epsilon3}$ cloned DNA with *Bgl*II and *Sac*I. Under stringent hybridization conditions, the 3′ LTR-like probe cross-hybridized faintly with a few bands of the $C_{\epsilon3}$ clone whereas the 5′ LTR-like probe did not hybridize with any of the cloned DNAs (data not shown). Under mild conditions, however, the 5′ LTR-like probe hybridized with a 4-kb *Bam*HI fragment immediately upstream of the coding region in the germline $C_{\epsilon1}$ gene, probably due to a small overlap with the probe, but with no fragments of rearranged $C_{\epsilon1}$ gene. The results indicate that there are no additional copies of the 5′ and 3′ LTR-like sequences around the $C_{\epsilon3}$ gene although it is flanked by several sequences distantly related to the 3′ LTR-like sequence. Furthermore, we were unable to find any sequences related to the 5′ and 3′ LTR-like sequences anywhere in the region between the $C_{\epsilon1}$ and $C_{\alpha2}$ genes, which are ~10 kb apart (Max *et al.*, 1982; Y.Nishida and T.Honjo, unpublished data).

We have also tested whether the 3′ LTR-like sequence is homologous to the Alu family sequence, which is also highly repetitive in the human genome. No restriction fragments of the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes hybridized with the Alu probe under the stringent conditions. Under mild conditions, a fragment hybridized with the Alu family probe in each phage cloned DNA. However, these fragments are different from those containing the 5′ and 3′ LTR-like sequences, indicating that the 3′ LTR-like sequence is a repetitive sequence distinct from the Alu family.

Discussion

Models for the evolutionary origin of the $C_{\epsilon3}$ gene

Remarkable structural features of the human C_ϵ pseudogene ($C_{\epsilon3}$) are summarized as follows. (1) The $C_{\epsilon3}$ gene is a processed gene. (2) It has a 31-base A-rich sequence 16 bases 3′ to the putative poly(A) addition signal. This sequence might be a descendant of either a poly(A) of mRNA, or an A-rich sequence dispersed in the genome. (3) The processed C_ϵ gene ($C_{\epsilon3}$) is not flanked by direct repeats on either side.

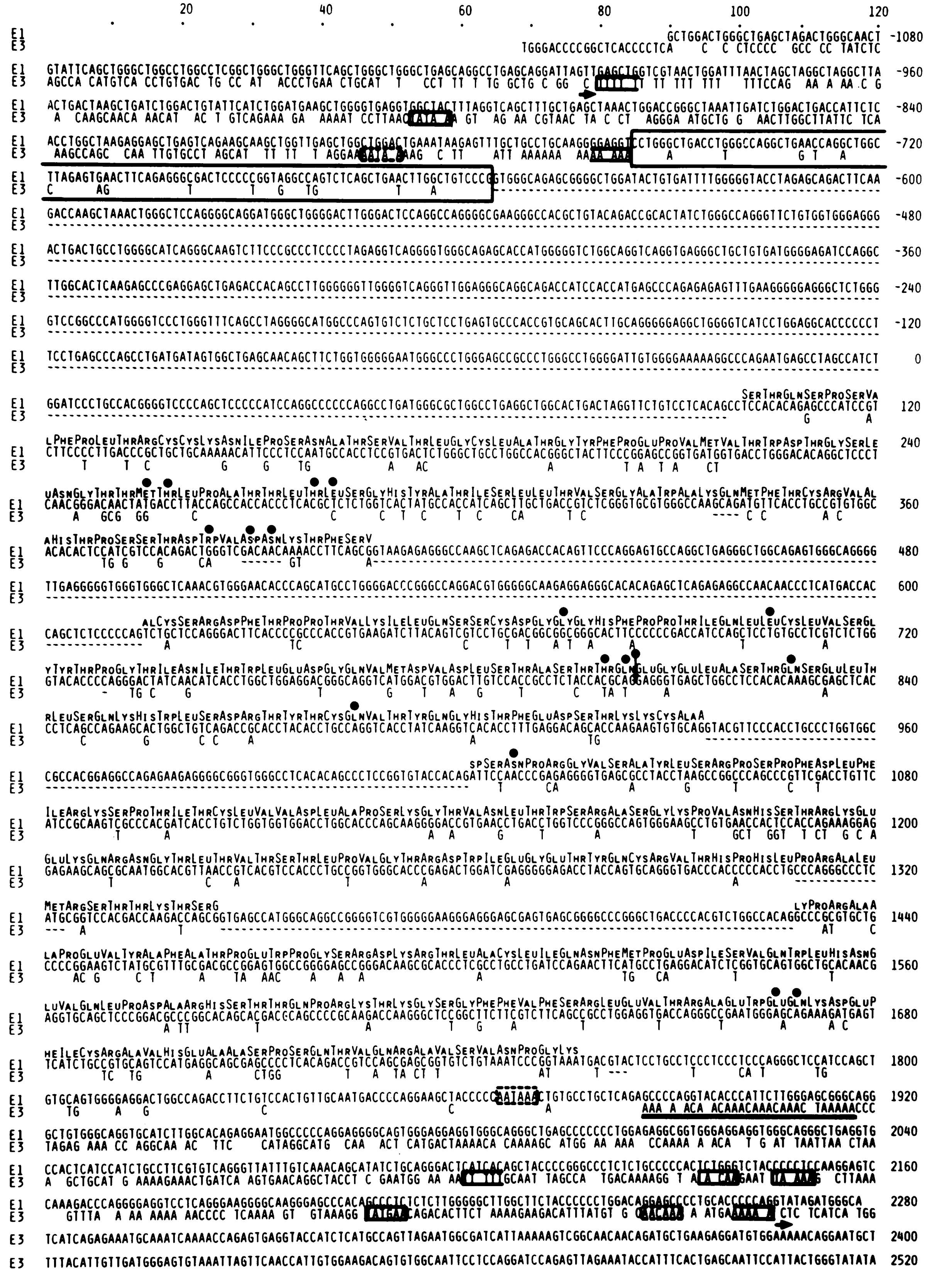

Fig. 3. Comparison of the nucleotide sequences of the human $C_{\epsilon1}$ and $C_{\epsilon3}$ genes. The nucleotide sequences upstream of the *Bam*HI site (position 2) of the $C_{\epsilon1}$ gene are determined with H·Igϵ-12 (germline type) except for 23 nucleotides (position −238 to −216) and all the others with H·Igϵ-11 (rearranged type). The two sequences were aligned with limited numbers of gaps to maximize homology. Only the nucleotides different from those of the $C_{\epsilon1}$ gene are shown in the $C_{\epsilon3}$ gene sequence. The amino acid sequence of the $C_{\epsilon1}$ gene deduced from the nucleotide sequence is shown and closed circles indicate the amino acids which do not match with the published amino acid sequences (Bennich and von Bahr-Lindstrom, 1974). The vertical line indicates the absence of an amino acid residue. The 5′-flanking DNA segment homologous between the $C_{\epsilon1}$ and $C_{\epsilon3}$ genes is boxed. Units of LTR-like sequences flanking the $C_{\epsilon3}$ gene are indicated by parentheses. TATAAA and AATAAA are boxed by solid and dotted lines, respectively. A-rich sequences are underlined. The direct repeats (AGCT) are indicated by horizontal arrows.

However, it is possible to consider that a part of the poly(A) stretch at positions −756 to −771 forms a direct repeat with a part of the A-rich sequence at positions 1887−1917. (4) LTR-like sequences which contain TATAAA and AATAAA sequences, as well as terminal inverted repeats, are present in the 5′- and 3′-flanking regions. However, this pair of LTR-like sequences does not constitute a direct repeat and we were unable to find an identical pair of either 5′ or 3′ LTR-like sequences in the H·Igε-31 DNA.

A number of speculative models for the generation of the $C_{\epsilon3}$ gene are conceivable. One, similar to that described previously (Hollis *et al.*, 1982), envisages that the $C_{\epsilon3}$ gene was transcribed aberrantly from the 5′-flanking region. The spliced mRNA may have been reverse-transcribed into DNA and integrated back to the genome. However, there are two problems. First, the presence of LTR-like sequences are not explained easily. They might perhaps fortuitously have been translocated right next to the integration site afterwards. Secondly, one would expect a promoter-like sequence 20−30 bases upstream of the transcription initiation site. Assuming that the integrated DNA contains all the transcript, we looked for a TATA box 20−30 bases upstream of the 5′ homology region but found none, only a sequence TGAAATAAGA 24 bases upstream of the homology region.

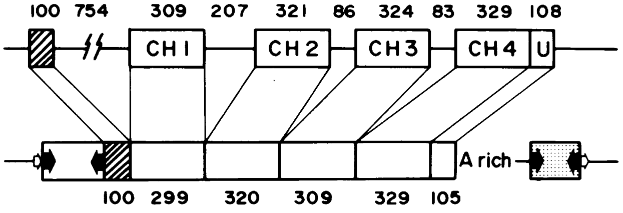

Fig. 4. Schematic representation of the structures of the human $C_{\epsilon1}$ and $C_{\epsilon3}$ genes. The 5′-flanking region homologous between the two genes and the CH exons are shown by hatched and open boxes, respectively. Numbers indicate length of each region in base pairs. The 5′ and 3′ LTR-like sequences are shown by open and dotted boxes, respectively, with closed arrows which indicate inverted repeats. Short direct repeats flanking both ends are shown by open arrows. U, 3′ untranslated sequence.

A second model (Figure 6) puts more weight on the presence of LTR-like sequences. The first event may have involved the translocation and integration of the unprocessed C_ϵ gene next to the LTR sequence, the ancestor of the current 5′ LTR-like sequence. The C_ϵ gene may then have been transcribed by the promoter in LTR, processed, and reverse-transcribed. The homologous recombination may have taken place between different LTRs, resulting in the formation of LTR-like sequences. The A-rich sequence may have been present close to the 3′ LTR unit. Since the 5′ and 3′ units of the insertion element need not necessarily be identical in order to accomplish integration (Machida *et al.*, 1982), the LTR-like sequences could have reintegrated the processed C_ϵ into the chromosome. The presence of the short direct repeat at both ends of the paired LTR-like sequences may be the hallmark of the integration. A major problem with this model, however, is that we do not find the parental gene flanked by an LTR. We have to assume that it was lost or rearranged to remove LTR. Clearly at present we cannot explain satisfactorily the origin of the $C_{\epsilon3}$ gene.

The $C_{\epsilon3}$ gene could be a gene of different function

The $C_{\epsilon3}$ gene is undoubtedly an epsilon pseudogene. The nucleotide sequence, however, shows that the $C_{\epsilon3}$ gene can encode a completely different protein of 292 amino acid

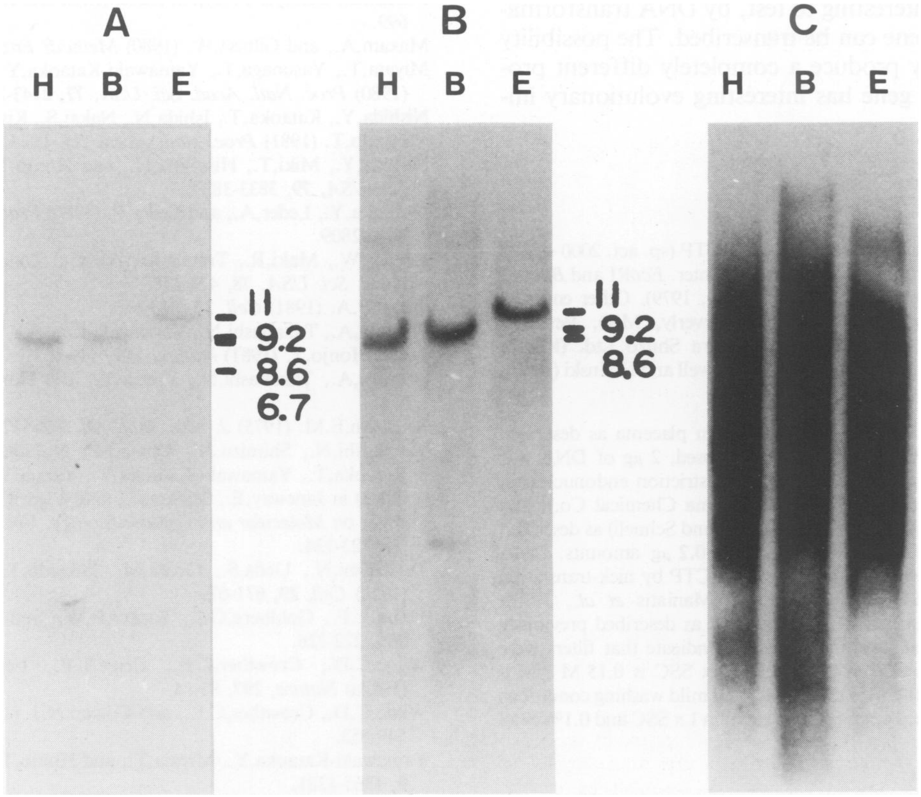

Fig. 5. Distribution of the 5′ and 3′ LTR-like sequences in the human genome. Total placental DNA was digested with *Hind*III **(lane H)**, *Bam*HI **(lane B)**, or *Eco*RI **(lane E)** and electrophoresed on a 0.5% agarose gel. The Southern blot filters were hybridized with the 5′ LTR-like DNA **(A)**, the pseudo-structure DNA **(B)**, or 3′ LTR-like DNA **(C)**. Stringent washing conditions were employed. Numbers indicate sizes of hybridized fragments in kb.

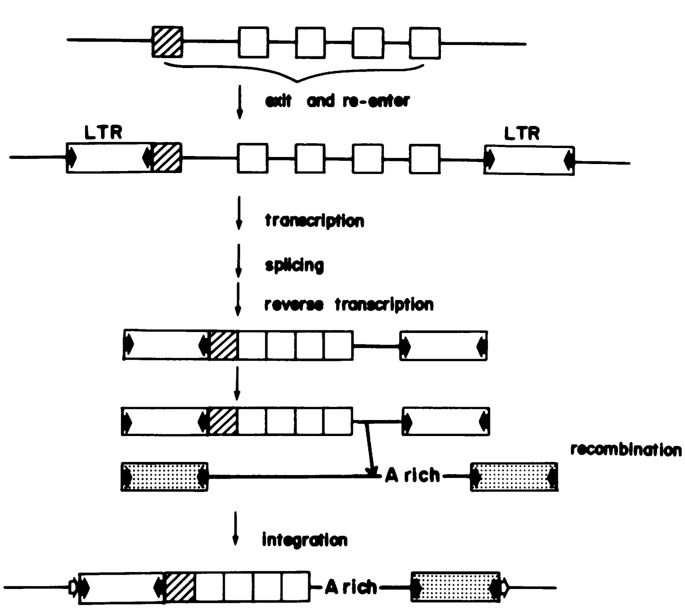

Fig. 6. A model for the generation of the C$_{\epsilon3}$ gene. The structure of the putative ancestral C$_{\epsilon}$ gene is represented in the top line. The hatched and open boxes indicate the homology segment in the 5′-flanking region and the exons, respectively. The LTR-like sequences which are responsible for the second integration step are shown by open boxes with closed arrows (inverted repeats). Short direct repeats flanking both ends are shown by open arrows.

residues, initiating at position 669 and terminating at position 1731 (Figure 3). A TATA box, a CATT box (position 110−118), and an rRNA binding site (Shine-Dalgarno sequence) (position 164) occur upstream of the open frame. Since the relative location of the TATA and CATT boxes are reversed, it would be interesting to test, by DNA transformation, whether the C$_{\epsilon3}$ gene can be transcribed. The possibility that a pseudogene may produce a completely different protein from the parental gene has interesting evolutionary implications.

Materials and methods

Materials

[γ-^{32}P]ATP (sp. act.5000 Ci/mmol) and [α-^{32}P]dCTP (sp. act. 2000−3000 Ci/mmol) were obtained from the Radiochemical Center. *Eco*RI and *Bam*HI were prepared as described previously (Honjo *et al.*, 1979). Other enzymes were purchased from New England Biolabs (Beverly, MA), Bethesda Research Laboratories (Bethesda, MD), and Takara Shuzo Ltd. (Kyoto, Japan). Plasmid DNA was prepared according to Clewell and Helinski (1969).

Southern blot hybridization

High mol. wt. DNA was obtained from a human placenta as described (Yaoita and Honjo, 1980). When total DNA was used, 2 μg of DNA was digested with the appropriate amount of various restriction endonucleases, electrophoresed on 0.5% agarose gels (Type I, Sigma Chemical Co.), and transferred to nitrocellulose filters (BA85, Schleicher and Schuell) as described by Southern (1975). Cloned DNAs were used in 0.2 μg amounts. DNA fragments used as probes were labeled with [α-^{32}P]dCTP by nick-translation to a specific activity of 200−1000 c.p.m./pg (Maniatis *et al.*, 1975). Hybridization was carried out in 1 M NaCl at 65°C as described previously (Honjo *et al.*, 1979). Stringent washing conditions indicate that filters were washed four times (30 min each) in 0.1 x SSC (1 x SSC is 0.15 M NaCl-0.015 M sodium citrate) and 0.1% SDS at 65°C, and mild washing conditions indicate that filters were washed twice (15 min each) in 1 x SSC and 0.1% SDS at 50°C.

Other methods and materials

Nucleotide sequence determination was performed according to the method of Maxam and Gilbert (1980). Heteroduplexes were formed using the formamide technique as described by Davis *et al.* (1971). Electron micrographs were taken with a JEM 100CX electron microscope (Nihondenshi Co.) at x 10 000 magnification and images were enlarged 10-fold. DNA lengths were measured with a digigramer (Mutoh Model G) using pBR322 and fd DNA as size markers. The Alu family probe was derived from the human ACTH gene clone obtained from S.Numa (Kyoto University).

Acknowledgements

This investigation was supported in part by grants from the Mitsubishi Science Foundation, from Torey the Science Foundation, and from the Ministry of Education, Science and Culture of Japan. We are grateful to Dr.S.Numa, Kyoto University, for his generous gift of an Alu family DNA, to Y.Nakayama and S.Nishida for their excellent technical assistance, and to F.Oguni for preparing this manuscript.

References

Arsdell,S.W.V., Denison,R.A., Bernstein,L.B., Weinex,A.M., Manser,T., and Gesteland,R.F. (1981) *Cell,* **26**, 11-17.

Bennich,H., and von Bahr-Lindström,H. (1974) in Brent,L., and Holborow, J. (eds.), *Progress in Immunology II,* Vol. **I**, North-Holland Publishing Co., pp. 49-58.

Clewell,D., and Helinski,D. (1969) *Proc. Natl. Acad. Sci. USA,* **62**, 1159-1166.

Davis,R.W., Simon,M., and Davidson,N. (1971) *Methods Enzymol.,* **21D**, 413-428.

Flanagan,J.G., and Rabbitts,T.H. (1982) *EMBO J.,* **1**, 655-660.

Hollis,G.F., Hieter,P.A., McBride,O.W., Swan,D., and Leder,P. (1982) *Nature,* **296**, 321-325.

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

Ishida,N., Ueda,S., Hayashida,H., Miyata,T., and Honjo,T. (1982) *EMBO J.,* **1**, 1117-1123.

Jacq,C., Miller,J.R., and Brownlee,G.G. (1977) *Cell,* **12**, 109-120.

Liu,C.P., Tucker,P.W., Mushinski,J.F., and Blattner,F.R. (1980) *Science (Wash.),* **209**, 1348-1353.

Lueders,K., Leder,A., Leder,P., and Kuff,E. (1982) *Nature,* **295**, 426-428.

Machida,Y., Machida,C., Ohtsubo,H., and Ohtsubo,E. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 277-281.

Maniatis,T., Jeffrey,A., and Kleid,D.G. (1975) *Proc. Natl. Acad. Sci. USA,* **72**, 1184-1188.

Max,E.E., Battey,J., Ney,R., Kirsch,I.R., and Leder,P. (1982) *Cell,* **29**, 691-699.

Maxam,A., and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.

Miyata,T., Yasunaga,T., Yamawaki-Kataoka,Y., Obata,M., and Honjo,T. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 2143-2147.

Nishida,Y., Kataoka,T., Ishida,N., Nakai,S., Kishimoto,T., Böttcher,I., and Honjo,T. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 1581-1585.

Nishida,Y., Miki,T., Hisajima,H., and Honjo,T. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 3833-3837.

Nishioka,Y., Leder,A., and Leder,P. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 2806-2809.

Roeder,W., Maki,R., Traunecker,A., and Tonegawa,S. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 474-478.

Sharp,P.A. (1981) *Cell,* **23**, 643-646.

Shimizu,A., Takahashi,N., Yamawaki-Kataoka,Y., Nishida,Y., Kataoka,T., and Honjo,T. (1981) *Nature,* **289**, 149-153.

Shimizu,A., Takahashi,N., Yaoita,Y., and Honjo,T. (1982) *Cell,* **28**, 499-506.

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

Takahashi,N., Shimizu,A., Obata,M., Nishida,Y., Nakai,S., Nikaido,T., Kataoka,T., Yamawaki-Kataoka,Y., Yaoita,Y., Ishida,N., and Honjo,T. (1981) in Janeway,E., Sercarz,E., and Wigzell,H. (eds.), *ICN-UCLA Symposia on Molecular and Cellular Biology, Vol.* **20C**, Academic Press, NY, pp. 123-134.

Takahashi,N., Ueda,S., Obata,M., Nikaido,T., Nakai,S., and Honjo,T. (1982) *Cell,* **29**, 671-679.

Vanin,E.F., Goldberg,G.I., Tucker,P.W., and Smithies,O. (1980) *Nature,* **286**, 222-226.

Wilde,C.D., Crowther,C.E., Cripe,T.P., Lee,M.G., and Cowan,N.J. (1982a) *Nature,* **297**, 83-84.

Wilde,C.D., Crowther,C.E., and Cowan,N.J. (1982b) *Science (Wash.),* **217**, 549-552.

Yamawaki-Kataoka,Y., Miyata,T., and Honjo,T. (1981) *Nucleic Acids Res.,* **9**, 1365-1381.

Yamawaki-Kataoka,Y., Nakai,S., Miyata,T., and Honjo,T. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 2623-2627.

Yaoita,Y., and Honjo,T. (1980) *Biomed. Res.,* **1**, 164-175.