

Similar penetration of the virus into differentiated cells induced by RA was also observed. These results indicate that the block to HCMV replication in the EC cells occurs after penetration of the virus.

Since HCMV is able to penetrate NT2/B9 EC cells, we tried to determine whether the virus could be activated by inducing differentiation after infection. Whereas less than 2 percent of the EC cells were positive for viral antigen 7 days after infection, 50 percent of the cells in cultures to which RA was added 2 hours before infection were positive 7 days later (Fig. 2). The same level of viral expression was found in cultures to which RA was added simultaneously with virus or 2 hours after infection. However, if RA was added 5 days after infection, no more cells expressing viral antigen than in untreated EC cultures were found 7 days later, even though the RA still induced cellular differentiation (8). Thus either the undifferentiated cells gradually eliminate the viral genome or the viral genome persists in a modified form that cannot be activated by later cellular differentiation.

The replication of HCMV in differentiated but not undifferentiated NT2/B9 cells is, to our knowledge, the first observation of a change in susceptibility to a virus upon differentiation in human teratocarcinomas. On the other hand, it has been known for some time that murine EC cells are resistant to several DNA (9) and RNA viruses (10), whereas their differentiated derivatives are not (9, 10). For example, murine cytomegalovirus (MCMV) penetrates PCC4 murine EC cells without subsequent transcription of viral messenger RNA, although cells differentiating in PCC4 cultures under the influence of dimethylacetamide do permit MCMV replication (11). The mechanism by which expression of the viral genome is restricted by murine EC cells may involve blocks to transcription, perhaps attributable to viral enhancer sequences (12) or host EC cells methylating newly acquired DNA (13), as well as blocks to viral messenger RNA processing (14) and to translation (15). Nevertheless, the biology of murine and human EC cells differs in many respects (5, 16), so it is not correct to extrapolate findings from one to the other.

Since HCMV penetrates undifferentiated human EC cells but replicates only in their differentiated derivatives, this system provides an opportunity to identify the cellular factors that enable HCMV to replicate. It should also now be possible to study the effect of HCMV

infection on the cellular events that occur during differentiation of cells resembling those of the early human embryo. This information might, in turn, bear on the mechanism of latent infection and reactivation by the virus and ultimately on its teratogenic effects.

EVA GÖNCZÖL

PETER W. ANDREWS

STANLEY A. PLOTKIN

Wistar Institute,

Philadelphia, Pennsylvania 19104

References and Notes

1. T. H. Weller, *N. Engl. J. Med.* **285**, 267 (1971).
2. S. Stagno, R. F. Pass, M. E. Dworsky, C. A. Alford, *Semin. Perinatol.* **7**, 31 (1983).
3. L. E. Davis *et al.*, *Pediatrics* **48**, 200 (1971); H. Stern and S. M. Tucker, *Br. Med. J.* **2**, 268 (1973); S. Stagno *et al.*, *N. Engl. J. Med.* **306**, 945 (1982).
4. G. R. Martin, *Science* **209**, 768 (1980); S. Strickland, *Cell* **24**, 277 (1981).
5. P. W. Andrews *et al.*, *Lab. Invest.* **50**, 147 (1984).
6. P. W. Andrews, *Dev. Biol.*, in press. RA also induces extensive somatic differentiation of murine EC cells [S. Strickland and Mahdavi, *Cell* **15**, 393 (1978); E. M. V. Jones-Villeneuve, M. W. McBurney, K. A. Rogers, V. I. Kalnins, *J. Cell Biol.* **94**, 253 (1982)].
7. K. I. Mathaei, P. W. Andrews, D. L. Bronson, *Exp. Cell Res.* **143**, 471 (1983); in preparation.
8. Cells developed the characteristic morphology and antigen phenotype of differentiated cells. Also, if the cells were superinfected with HCMV at a multiplicity of infection of 5 to 10, 70 to 80 percent of them expressed viral antigens 7 days later.
9. D. E. Schwartzendruber and J. M. Lehman, *J. Cell. Physiol.* **85**, 179 (1975); D. E. Schwartzendruber, T. D. Friedrich, J. M. Lehman, *ibid.* **93**, 25 (1977); M. Boccardo and F. Kelly, *Virology* **90**, 147 (1978); S. Segal and G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5611 (1979); F. K. Fujimura, P. E. Silbert, W. Eckhart, E. Linney, *J. Virol.* **39**, 306 (1981).
10. J. Peries, E. Alves-Cardoso, M. Canivet, M. C. Debous-Guillemain, J. Lasneret, *J. Natl. Cancer Inst.* **59**, 463 (1977); N. M. Teich, R. A. Weiss, G. R. Martin, D. R. Lowy, *Cell* **12**, 973 (1977); J. W. Gautsch, *Nature (London)* **285**, 110 (1980).
11. F. Dutko and M. B. A. Oldstone, *J. Exp. Med.* **159**, 1636 (1981).
12. M. Katinka, M. Yaniv, M. Vasseur, D. Blangy, *Cell* **20**, 393 (1980); F. K. Fujimura, P. L. Deininger, T. Friedman, E. Linney, *ibid.* **23**, 809 (1981); M. Katinka, M. Vasseur, N. Montreau, M. Yaniv, D. Blangy, *Nature (London)* **290**, 720 (1981).
13. C. L. Stewart, H. Stuhlman, D. Jahner, R. Jaenisch, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4098 (1982); D. Jahner *et al.*, *Nature (London)* **298**, 623 (1982).
14. S. Segal, A. J. Levine, G. Khoury, *Nature (London)* **280**, 335 (1979).
15. F. Fujimura, P. E. Silbert, W. Eckhart, E. Linney, *J. Virol.* **39**, 306 (1981).
16. P. W. Andrews, *Int. J. Cancer* **30**, 567 (1982); P. W. Andrews, P. N. Goodfellow, L. Shevinsky, D. L. Bronson, B. B. Knowles, *ibid.* **29**, 523 (1982); C. Cotte, D. Raghaven, R. A. J. McIlhinney, P. Monaghan, *In Vitro* **18**, 739 (1982); P. W. Andrews, P. N. Goodfellow, I. Damjanov, *Cancer Surv.* **2**, 41 (1983); R. A. J. McIlhinney and S. Patel, *Cancer Res.* **43**, 1282 (1983).
17. T. Furukawa, A. Fioretti, S. A. Plotkin, *J. Virol.* **11**, 991 (1973).
18. L. Vaczi and E. Gönczöl, *Acta Virol.* **17**, 189 (1973).
19. B. B. Wentworth and L. French, *Proc. Soc. Exp. Biol. Med.* **135**, 253 (1970).
20. We thank B. B. Knowles for valuable discussions and D. DiGiovanni, M. Rudenstein, J. Ianacone, G. Banting, and M. Dignazio for excellent technical assistance. Supported by a grant from the Hassel Foundation and by NIH grants AI-14927 and CA-29894.

21 December 1983; accepted 13 February 1984

Single-Copy Inverted Repeats Associated with Regional Genetic Duplications in γ Fibrinogen and Immunoglobulin Genes

Abstract. We have found that a portion (150 base pairs) of the seventh exon of the human γ fibrinogen gene is duplicated in the preceding intron. This duplicated sequence, termed a "pseudoexon," is flanked on each side by a single-copy inverted repeat sequence consisting of 102 base pairs. Frequencies of point substitutions indicate that both the pseudoexon and the inverted repeat sequence arose approximately 10 to 20 million years ago. The generality of this type of duplication is suggested by the occurrence of a similar duplication in the mouse immunoglobulin μ - δ region. As in the fibrinogen pseudoexon, the portion of the immunoglobulin μ - δ region containing the duplication and the inverted repeat was reported to be single-copy in the mouse genome. Since both of the first two single-copy inverted repeats to be sequenced are associated with regional duplications, it is likely that many of the single-copy inverted repeat sequences, which make up 1 to 2 percent of the genome, are also associated with regional duplications.

Long inverted repeat (IR) sequences, which consist of self-complementary sequences nearby on the same strand, are reported to represent approximately 5 percent of the human and other eukaryotic genomes (1). The significance of IR's in eukaryotes is uncertain, although several examples of repetitive genetic elements with terminal IR's, which function as transposable elements, have been identified in *Drosophila* (2), sea urchin (3), and *Dictyostelium* (4). The IR from the FB transposable elements of *Drosophila* and the TU1 of sea urchin have a

partial subunit structure consisting of short direct repeats. Inverted repeat sequences consist of both repetitive and single-copy IR sequences in the same ratio, as is found in total cell DNA (1). The function and significance of single-copy IR's is unknown even though they represent 1 to 2 percent of most eukaryotic genomes (1).

While analyzing the human γ fibrinogen gene, we found that a portion of the seventh exon was duplicated in an adjacent intron. This duplicated sequence or "pseudoexon" is flanked on both sides

by a single-copy IR sequence. The isolation of the gene containing the pseudoexon will be described elsewhere (5). Briefly, the complementary DNA (cDNA) (6) for γ fibrinogen was used to isolate three unique overlapping phage clones from two human genomic phage libraries (7). From these phage clones, a 32-kilobase-pair (kbp) restriction map of this genomic locus was determined, and the entire region was subcloned into plasmid pBR322 in convenient lengths for further analysis. In all three phage clones, a 0.8-kbp Hind III fragment corresponding to the sixth intron of this seven-intron gene hybridized with labeled γ fibrinogen cDNA; this was also seen by genomic Southern analysis (8) with DNA from 13 individuals (data not shown). Only one genomic locus was

identified with the γ fibrinogen cDNA.

We sequenced 3.4 kbp of the 3' portion of this gene; a map of this region is shown in Fig. 1. The solid blocks above the line represent the last three exons of the major form (γ A) of γ fibrinogen and correspond to the nucleotide sequence of γ fibrinogen cDNA (6). A minor form of the γ chain (γ B) is produced by alternate splicing, as predicted from the protein sequence (9) and from analysis of the rat gene (10); the last two exons for γ B are shown in Fig. 1 by solid blocks below the line. As shown diagrammatically in Fig. 1, a portion of the seventh exon has been duplicated in the preceding intron. This duplicated region or pseudoexon is designated in Fig. 1 by the lightly shaded block and the IR by opposing arrows. The source of the pseudoexon is indicat-

ed in Fig. 1 by the points X and Y; the pseudoexon starts 30 bp after the 5' splice site of the true exon and extends 20 bp into the unique γ B exon sequence. As shown in Fig. 2, there is strong homology between the pseudoexon and much of the seventh exon, and there is strong dyad symmetry between the two sides of the IR. Homology was weak at the junctions of the pseudoexon and the two sides of the IR.

The lower part of Fig. 2 shows that a subunit structure is discernible within the arms of the IR. Such a subunit structure has been recognized among a number of IR's in transposable elements such as FB (2) and TU1 (3). The subunit structure of the FB transposable element resembles our IR in that the sequence xGTTTGCxxx (G, guanine; T, thymine;

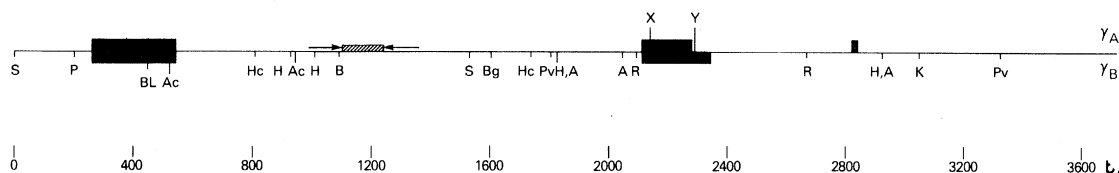


Fig. 1. The 3' terminus of the human γ fibrinogen gene. Solid boxes are exons with γ A above and γ B below the line. The pseudoexon is shown by the lightly shaded box with flanking IR indicated by arrows. X and Y indicate the limits of homology of the exon with pseudoexon. Restriction enzyme sites are Bam HI (B), Hind III (H), Sma I (S), Acc I (Ac), Pst I (P), Kpn I (K), Bal I (Bl), Alu I (A), Bgl II (Bg), Pvu II (Pv), Rsa I (R), and Hinc II (Hc); Alu I and Rsa I sites are shown only for nucleotides 1826 to 1927. This entire region was sequenced by the dideoxy chain termination method after subcloning into M13 phage (18).

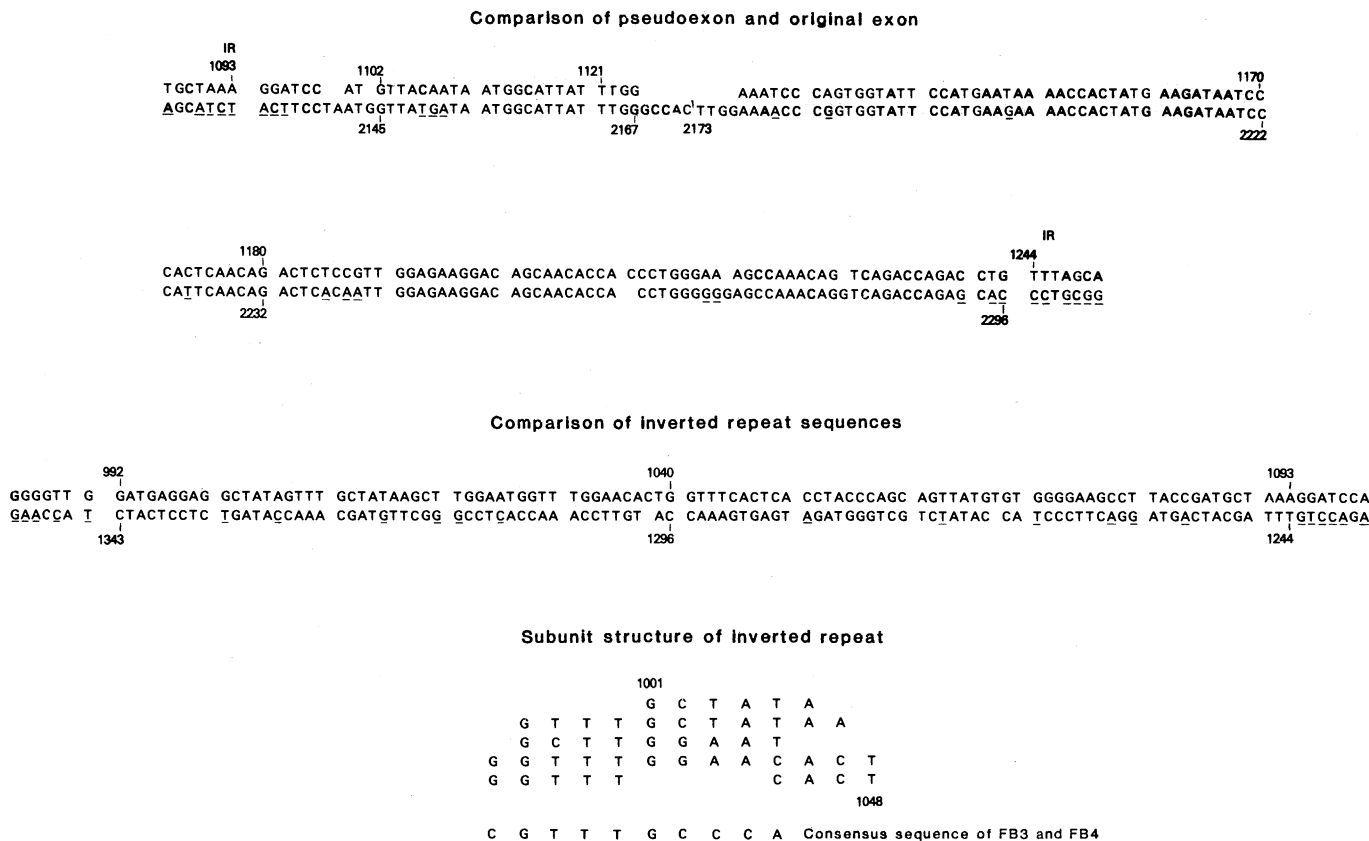


Fig. 2. Comparison of the pseudoexon, exon, and IR sequences. The exon and pseudoexon sequences are compared with the duplicated region or pseudoexon on top. The two sides of the IR are arranged to show base pairing. Nucleotides are numbered as in Fig. 1. Mismatching is indicated by underlining. Flanking sequences, 7 bp on each side, are also shown. Sequences have been aligned to show maximal homology of the subunit sequence in the lower part of the figure.

C, cytosine) is part of both IR's. Although this similarity may have arisen because this is simply an easily amplified sequence, it does raise the possibility that such transposable elements may have initially been generated by the same regional duplication mechanism.

Since this duplicated region and the flanking IR have a general structure analogous to insertion sequences and transposable elements in prokaryotes (11), we attempted to estimate the frequency of the element in human genomic DNA. The pseudoexon and IR sequences appeared to be single-copy, since under even moderately stringent conditions, this region hybridized to only a single restriction fragment in genomic DNA (Fig. 3). The IR and duplicated region hybridized less strongly to a second Hind III fragment of 1100 bp (Fig. 3) corresponding to the original exon, the origin of this pseudoexon. The same result was seen with Eco RI or Hind III digests of genomic DNA from six other individuals. Identical results were also observed when the hybridization and rinsing temperatures were lowered to 60°C (data not shown). Comparison of the IR sequence with all nucleotide sequences in the Los Alamos data bank (12) revealed no convincing similarities.

The generality of our findings is dependent on whether this is an isolated event or whether it represents an example of a class of genetic duplications associated with single-copy IR's. Recently, a second single-copy IR was found in the μ - δ region of mouse immunoglobulin (Ig) (13). From our data, we would predict that the loop (sequence flanked by the IR) is homologous to a sequence nearby. As shown in Fig. 4, this is, in fact, the case—approximately one third of the loop is homologous to a region 1 kbp away. Between the μ m and C δ 1 exons of the Ig δ gene (13), there is a 94-bp region flanked by a 143-bp IR. This 94-bp region apparently arose by a process similar to the duplication in the fibrinogen gene since there is a highly homologous region only 1 kbp on the 5' side of the IR. The similarities with the duplication in the γ fibrinogen gene are striking (Tables 1 and 2). Both duplications are oriented in the same direction relative to their probable origins. The distance between the regions of homology are almost identical in the two examples, although the IR-loop sequence is 5' to the apparent origin in the γ fibrinogen gene while 3' in Ig μ - δ . Near the junction of the IR and loop in Ig μ - δ , dyad symmetry was weak at this end of the IR (nucleotides 3658 to 3661). Although some very weak imperfect direct repeats

Table 1. The total length of the IR, loops, and portions of the loops duplicated. The distance between duplicated sequences was measured from the center of the duplication (in the loop) to the center of the neighboring homologous sequence.

Characteristic	Measure (bp)	
	γ Fibrinogen	Immunoglobulin μ - δ
Length of IR	102	143
Length of total loop	150	294
Length of total duplication	150	94
Distance between duplicated sequences	1047	818

could be found in the Ig μ - δ , no subunit structure as convincing as that found in the γ fibrinogen IR was identified.

Examination of the mismatches, which are found in these duplicated sequences, can be used to demonstrate the types of spontaneous changes occurring naturally and to estimate their times of divergence. In the area of strong homology (nucleotides 1102 to 1239), differences between the pseudoexon and the

exon consisted of five point substitutions, two double substitutions (TG to CA and AA to CG, where A represents adenine), a double substitution with a single deletion (GGG to AA), a single addition (C), a single deletion (G), and a 9-bp deletion. As shown in Fig. 2, the last 4 bp of the 9-bp deletion were identical to the 4-bp sequence preceding the deletion. Differences between the two sides of the γ fibrinogen IR sequence consisted of ten single substitutions, one double substitution (TT to CC), and two single insertion-deletion mutations. These data are summarized in Table 2 along with the Ig μ - δ results. The sequences at the IR-loop junctions were poorly matched in both cases and were not included in this part of the table. Whether this increased mismatch at the junctions was coincidental or was generated during the duplication will require more examples to resolve. If it is assumed that the mismatches listed in Table 2 have occurred since the duplications, then the two sides of the γ fibrinogen IR duplicated approximately 10 to 20 million years ago; this estimate is based on a frequency of total point substitutions of 11.8 percent (14). The frequency of total point substitutions in the pseudoexon (nucleotides 1102 to 1239) compared to the original exon is 8.0 percent; this should probably be increased when compared to the frequency in the IR, inasmuch as the original exon has been relatively invariant since the primate-rodent divergence (10). The pseudoexon clearly diverged from the exon much more recently than the primate-rat divergence, since its sequence is more homologous to the human exon than to the rat exon (10).

The duplication of the Ig μ - δ was also probably 10 to 20 million years ago, judging from the substitution frequency. In duplicated sequences such as the α -tubulin pseudogene (15), various complex mutations in addition to simple transitions occurred after the duplications; this also holds for the γ fibrinogen and Ig μ - δ genes. Implied in our analysis is the assumption that these sequences, except for the γ fibrinogen exon, have been relatively free to diverge. Richards *et al.* (13) suggested that the Ig μ - δ IR may contribute to the secondary structure of the primary RNA transcript and thus possibly act as a constraint in certain types of genetic drift such as deletion-insertions and, to a lesser extent, transversions (16). Although the transversion frequency was not reduced, the occurrence of deletion-insertions was somewhat less in the Ig μ - δ IR than in the γ fibrinogen IR.

The duplications found in the γ fibrin-

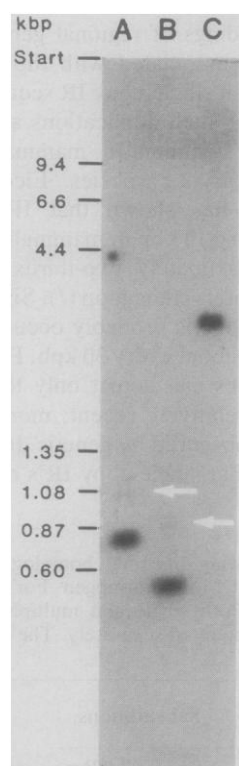


Fig. 3. Genomic Southern hybridization analysis of the pseudoexon and the IR region. Normal genomic DNA was digested with Hind III (A), Alu I (B), or Bam HI (C). This was hybridized (5) with a labeled fragment that consisted of the 0.5-kbp fragment comprising nucleotides 1017 to 1531. Following the hybridization, the final and most stringent rinse was in 1.5 $\times$ standard saline citrate for 45 minutes at 65°C.

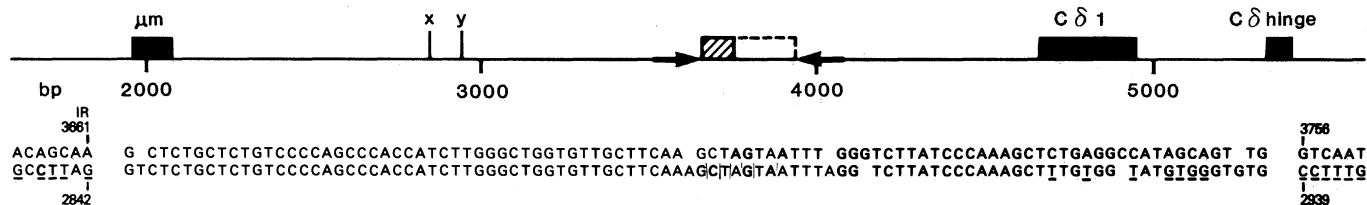


Fig. 4. Sequence analysis of the murine immunoglobulin μ - δ . Sequence determination and nucleotide numbering is from Richards *et al.* (13). The IR have been shortened by 19 bp in comparison with that shown by Richards *et al.* In the upper part of the figure, the region of the gene containing the IR and duplicated sequence are shown schematically. The sequence designated by the hatched box shares homology with the sequence designated between x and y. These two sequences are compared in the lower part of the figure.

ogen and Ig μ - δ genes differ in several respects from other examples of gene duplication in eukaryotes. (i) Most insertion duplications in eukaryotes and all of those in vertebrate DNA are flanked only by direct repeats of varying length (1-4, 17), whereas the pseudoexon and IR had no discernible flanking direct repeats. (ii) Unlike well-defined transposable elements with flanking IR's (1-4, 11), the pseudoexon and flanking IR and the Ig μ - δ loop IR are single-copy. (iii) Unlike transposition events or other duplications such as the "processed genes" (15), the duplications associated with single-copy IR's are close to their source. The above observations indicate that the pseudoexon probably did not arise by a typical transposition-like event. A gene conversion event involving unequal crossover can also probably be excluded since the generation of IR's would not be expected, and no evidence of homology can be found between the source of the duplicated region and sequences flanking the duplicated regions. By the frequency of mispairing, the γ fibrinogen IR appears to have arisen at approximately the same time as the pseudoexon but has no obvious relation to the original exon. A reasonable explanation of our findings is that by some "perturbation" of DNA replication the IR was generated, probably at the target site, along with a portion of the neighboring exon either from the same strand or a sister duplex. The same mechanism can

be proposed for the Ig μ - δ loop IR if one assumes that a major deletion of approximately 200 bp occurred in the source of the duplication, either at the time of duplication or subsequent to it. The subunit structure seen in a portion of the γ fibrinogen IR may indicate that it was produced by amplification of a smaller sequence. If the first side of the γ fibrinogen and Ig μ - δ IR was produced in this manner, then it was either a very imperfect amplification or preceded the generation of the second side of the IR by a very long interval. Further elucidation of the mechanism involved in producing single-copy IR-associated duplications could probably be obtained by examining the same loci in other primate and rodent species whose divergence from the human or mouse species preceded these duplications.

Our findings of regional genomic duplications associated with the first two mammalian single-copy IR sequences indicate that such duplications are probably very common in mammalian and other eukaryotic species. Electron microscopy has shown that IR's occur about every 10 kbp in mammalian DNA, and approximately two-thirds of these have an intervening loop (1). Single-copy IR's with loops probably occur less frequently—about every 30 kbp. Hybridization studies can detect only those IR's that are relatively recent; more ancient IR's are obscured by genetic drift, and it is likely that single-copy IR's may make

up even more than 2 percent of the genome. Therefore, it is possible that genetic duplications associated with single-copy IR's, like the two described herein, constitute a major type of regional duplication in eukaryotic evolution.

ALBERT J. FORNACE, JR.

Laboratory of Pathology,
Hematopathology Section, and
Radiobiology Section,
National Institutes of Health,
Bethesda, Maryland 20205

DAVID E. CUMMINGS

CLAUDETTE M. COMEAU

JEFFERY A. KANT

GERALD R. CRABTREE

Laboratory of Pathology,
Hematopathology Section,
National Institutes of Health

References and Notes

1. B. Lewin, *Gene Expression, Eukaryotic Chromosomes* (Wiley, New York, ed. 2, 1980), vol. 2; T. R. Cech, A. Rosenfeld, J. E. Hearst, *Mol. Biol.* **81**, 299 (1973); P. L. Deninger and C. W. Schmid, *J. Mol. Biol.* **106**, 773 (1976); W. R. Jelinek, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2679 (1978); H. J. Lin and C. L. H. Lee, *Anal. Biochem.* **96**, 144 (1979).
2. S. S. Potter, M. Truett, M. Philips, A. Maher, *Cell* **20**, 639 (1980); M. A. Truett, R. S. Jones, S. S. Potter, *ibid.* **24**, 753 (1981); S. S. Potter, *Nature (London)* **297**, 201 (1982).
3. D. Liebermann, B. Hoffman-Liebermann, J. Weinthal, G. Childs, R. Maxson, A. Mauron, S. N. Cohen, L. Kedes, *Nature (London)* **306**, 342 (1983).
4. E. Rosen, A. Silvertsen, F. A. Firtel, *Cell* **35**, 243 (1983).
5. A. J. Fornace, Jr., *et al.*, in preparation.
6. J. A. Kant, S. T. Lord, G. R. Crabtree, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3953 (1983).
7. W. D. Benton and R. W. Davis, *Science* **196**, 180 (1977).
8. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
9. C. Wolfenstein-Todel and M. W. Mosesson, *Biochemistry* **20**, 6146 (1981).
10. G. R. Crabtree and J. A. Kant, *Cell* **31**, 159 (1982).
11. N. Kleckner, *Annu. Rev. Genet.* **15**, 341 (1981).
12. W. J. Wilbur and D. J. Lipman, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 726 (1983).
13. J. E. Richards *et al.*, *Nature (London)* **306**, 483 (1983).
14. F. Perler *et al.*, *Cell* **20**, 555 (1980).
15. I. Lemischka and P. A. Sharp, *Nature (London)* **300**, 330 (1982).
16. D. M. J. Lilley, *Nucleic Acids Res.* **9**, 1271 (1981).
17. D. J. Finnegan, G. M. Rubin, M. W. Young, D. S. Hogness, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1053 (1978); G. R. Cameron, E. Y. Loh, R. W. Davis, *Cell* **16**, 739 (1979); Y. Nishioka, A. Leder, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2806 (1980).
18. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977); J. Messing, R. Crea, P. H. Seeburg, *Nucleic Acids Res.* **9**, 309 (1981).

22 November 1983; accepted 29 February 1984

Table 2. The mutation frequencies of the regions, showing maximal homology or dyad symmetry. Location refers to the nucleotide position in Fig. 1 for γ fibrinogen. For Ig μ - δ , the nucleotide position is taken from Richards *et al.* (13). Both single and multiple deletion-insertions are scored as single events. Each substitution is counted separately. The percentage of substitutions that are transversions is indicated.

Region	Location	Deletion- insertions		Substitutions		Trans- versions (per- cent)
		Num- ber	Per- cent	Num- ber	Per- cent	
<i>γ Fibrinogen</i>						
Pseudoexon	1102–1239	4	3	11	8	36
IR	992–1093	2	2	12	12	8
<i>Immunoglobulin μ-δ</i>						
Loop	3662–3755	6	6	7	7	55
IR	3543–3640	0		11	11	45