Initiation Sites for Replication of Mammalian Cell DNA

Igor Tamm, Barbara R. Jasny & Joel E. Cohen*

Replication of DNA in mammalian cells in S phase is initiated at about 50,000–100,000 sites (Hand & Tamm 1974, Jasny *et al.* 1978, Cohen *et al.* 1978). In general, replication proceeds bidirectionally from the origins (Huberman & Riggs 1968, Edenberg & Huberman 1975). There appear to be many more potential initiation sites than those activated during S phase. Evidence for this comes from work in developing organisms (Callan 1972, 1974) and with 5-fluoro-2′-deoxyuridine (Taylor & Hozier 1976, Taylor 1977). We have recently carried out a detailed examination of DNA chain initiation by light microscopic DNA fiber autoradiography in muntjac, mouse (L–929), and bovine (MDBK) cells in culture (Jasny *et al.* 1978, Cohen *et al.* 1978). Exponentially growing cells were labeled with high specific activity ^{3}H-thymidine for 10 minutes and then with low specific activity ^{3}H-thymidine for 3 hours. The modal inter-initiation site distance is 5–15 μm (15,000–45,000 base pairs) in all three cell types, but the individual measurable distances between the activated sites vary from 3.4 to 432 μm. Statistical tests of randomness, applied to the distribution of inter-initiation distances, show a nonrandom frequency distribution of the distances between activated initiation sites for all three cell types.

MEAN DISTANCE BETWEEN ACTIVATED INITIATION SITES

We have used three methods to obtain estimates of, or lower bounds on, the mean inter-initiation distances on the unbroken DNA fiber. The ranges of estimates are 8–23 μm for muntjac cells, 22–45 μm for MDBK cells, and 14–63

The Rockefeller University, New York, New York 10021 U.S.A.

* Dept. of Microbiology, State University of New York, Stony Brook, N.Y. 11794 U.S.A.

SPECIFIC EUKARYOTIC GENES, Alfred Benzon Symposium 13, Munksgaard 1979

Table I

Inter-initiation site distance and number of activated sites in mammalian DNA

Species	Cells	Mean inter-initiation distance, μm	Length of DNA, m*	Estimated number of activated sites per diploid genome	Reference
Muntjac	Muntjac	16	1.2	75,000	Cohen *et al.* 1978
Cattle	MDBK	34	1.8	53,000	Cohen *et al.* 1978
Mouse	L-929	39	1.8	46,000	Cohen *et al.* 1978
Mouse	L-929	45	1.8	40,000	Hand and Tamm 1974
Cattle	MDBK	17	1.8	106,000	Hand and Tamm 1974
Syrian hamster	BHK	30†	2.0	67,000	Hand and Tamm 1974
African green monkey	CV-1	42			Hand and Tamm 1974
Human	HEK	23	2.0	87,000	Hand and Tamm 1974

* Per diploid set of chromosomes; mouse, cattle, man from Altman & Katz (1976); muntjac from Green & Bahr (1975); Syrian hamster from S.Ohno, personal communication; a value for African green monkey was not available.

† In Chinese hamster (B14FAF28) the range of values is 5–180 μm and the range of distribution at one half peak height is 15–55 μm (Huberman & Riggs 1968); DNA length=2.0 m (S. Ohno, personal communication).

μm for L cells (Cohen *et al.* 1978). The center values of these estimates are recorded in the summary of available data in Table I. It can be seen that the estimates of the mean inter-initiation distance and of the number of activated sites in mammalian DNA (diploid) do not differ by more than a factor of 2–3 when derived by a variety of techniques. Most values for the number of activated sites fall within the 50,000–100,000 range, which corresponds to a range of 20–40 μm (60,000–120,000 base pairs) for most of the mean inter-initiation distances.

INVERTED REPEAT SEQUENCES

The physical nature of replication origins in mammalian cell DNA has not been established. It is reasonable to suppose that there are specific recognition sites in DNA for proteins that function in the initiation of new DNA chains and that such recognition sites share common nucleotide sequences. Furthermore, since DNA replication proceeds bidirectionally from most origins, the initiation sites may be expected to be structures with twofold rotational symmetry.

Regions of inverted complementary sequences, which are widely distributed in DNA (Davidson *et al.* 1973, Wilson & Thomas 1974, Schmid & Deininger 1975, Cech & Hearst 1975) may represent potential initiation sites for DNA replication (Bollum 1975, Edenberg & Huberman 1975). Inverted complementary sequences with twofold rotational symmetry are of the type

... A B C C′ B′ A′ ...
... A′ B′ C′ C B A ...

where A and A′, B and B′ and C and C′ represent complementary sequences. The structure ... A B C C′ B′ A′ ... in a single strand of DNA is commonly referred to as a pair of inverted repeat sequences and we shall follow this practice. In double-stranded DNA, a pair of inverted repeat sequences in one strand has a complement in the other strand. Numerical estimates of pairs of inverted repeats in nuclear DNA are commonly given in terms of one strand. Such estimates thus refer to regions in the double-stranded DNA containing inverted repeats. Two kinds of inverted repeat DNA structures have been recognized: 1) those which contain no or only a few intervening nucleotides between members of a pair of inverted repeats (Wilson & Thomas (1974) likened the inverted repeats without spacers to palindromes); 2) those in which the members of a pair of inverted repeats are separated by a spacer sequence of variable length.

The mean length of a member of a pair of repeat sequences is 300 bases in the human genome (Deininger & Schmid 1976). Such sequences show considerable homology, but not complete sequence identity (Robertson *et al.* 1977, Jelinek 1977, Jelinek *et al.* 1978, Jelinek 1978). These important findings are summarized elsewhere in this volume (Jelinek 1979).

MEAN DISTANCE BETWEEN PAIRS OF INVERTED REPEAT SEQUENCES

A key question is whether there are enough of the homologous inverted repeats in DNA to account for initiation sites for replication. The average center-to-center distance between pairs of inverted repeat sequences may be estimated for several mammalian species (Table II). Wherever comparison is possible, there are enough inverted repeats to account for initiation sites. Indeed, if initiation takes place in inverted repeat regions, there appears to be redundancy of such sites which would protect against failure in the critical process of genome reduplication (see data in human and hamster cells). The most extensive evidence is available for human cells (Wilson & Thomas 1974, Deininger & Schmid 1975, Jelinek *et al.* 1978) and it appears that there are 400,000–900,000

Table II

Center-to-center distance between pairs of inverted repeats and number of inverted repeat regions in mammalian DNA

Species	Cells	Mean center-to-center distance, μm	Estimated number per diploid genome: Inverted repeats Total	Inverted repeats without spacers	Reference
Human	Placental	2.1–4.5	440,000–950,000	160,000–340,000	Deininger & Schmid 1976
Human	HeLa	5*	unknown	400,000	Wilson & Thomas 1974
Human	HeLa (hnRNA)		400,000–800,000		Jelinek *et al.* 1978
Chinese hamster	CHO		600,000		Jelinek 1978
Mouse†			~200,000	~80,000	Cech & Hearst 1975, Wilson & Thomas 1974, Ryskov *et al.* 1973

*Adjacent pairs of inverted repeats lacking spacers.

† In mouse cell (SVT2) DNA, Cech & Hearst (1975) reported pairs of inverted repeats to be separated by a mean distance of 25 μm, which is equivalent to 72,000 structures in the diploid genome. About 40% (29,000) lack spacers. Wilson & Thomas (1974) reported that the mean center-tocenter distance between clusters of inverted repeats without spacers in mouse cell (L) DNA is 62.1 μm, which gives an estimate of 30,000 for the total number of clusters. This estimate of the number of clusters is essentially the same as that of Cech & Hearst (1975) for individual regions (29,000). The reasons for this discrepancy are not clear. An analysis of the hnRNA of mouse Ehrlich's ascites cells suggested (Ryskov *et al.* 1973) that the total number of pairs of inverted repeats in the diploid mouse cell genome may be 600,000. This estimate is probably too high, because the length of the repeated sequences may be greater (Cech & Hearst 1975) than the value (100 nucleotides) used in making this calculation (Ryskov *et al.* 1973). As an intermediate figure we suggest that there may be about 200,000 pairs of inverted repeats in mouse DNA.

inverted repeat regions per genome, of which 36% lack spacers (Deininger & Schmid 1975).

The data from *Triturus* and *Xenopus laevis* are also compatible with the possibility that inverted repeat sequences are potential origins for DNA replication. The distances between activated initiation sites in DNA of somatic cells of mature *Triturus* are in the range of 100–350 μm or greater (Callan 1972, 1974), while in the embryos in the neurula stage most of the initiation sites are about 40 μm apart. Inverted repeats without spacers are distributed in *Triturus* DNA at an average center-to-center distance of 8.7 μm (Wilson & Thomas 1974). In *Xenopus* the inter-initiation distances are in the range of 18–128 μm (Callan 1972) while the mean distance between pairs of inverted repeats is about 10 μm (Perlman *et al.* 1976).

COMMENTS

Clearly, the inverted repeat sequences possess the fundamental properties expected of replication initiation sites in eukaryotic DNA. Homologous inverted repeats appear to be interspersed between stretches of unique nonrepetitious sequences in eukaryotic DNA. Inverted repeats without spacers would appear to be the preferred structures for the initiation of bidirectional replication. Unidirectional replication from origins has been reported to occur with low frequency (Edenberg & Huberman 1975). Whether this is entirely physiological is an open question. It should be noted that inverted repeat sequences have been demonstrated in segments of viral DNA containing initiation sites for DNA replication, as for example in simian virus 40 (SV40) (Subramanian *et al.* 1977, Hsu & Jelinek 1977).

The hypothesis that multifocal initiation of mammalian DNA replication occurs in homologous inverted repeat regions in DNA can be tested experimentally. This hypothesis does not exclude a possible role for inverted repeat sequences in the regulation of transcription (Bollum 1975) or in posttranscriptional processing of RNA (Jelinek *et al.* 1978).

ACKNOWLEDGMENTS

We thank Dr. Warren Jelinek for permission to quote unpublished results. This work was supported by Research Grant CA-18608 and by Program Project Grant CA-18213 awarded by the National Cancer Institute, and by National Science Foundation Grant BMS (DEB) 74–13276. B. R. J. was a Predoctoral Trainee under the Institutional National Research Service Award CA-09256 from the National Cancer Institute.

REFERENCES

Altman, P.L. & Katz, D.D. (1976) *Biological Handbooks I, Cell Biology,* Fed. Amer. Soc. Exp. Biol.. 454 pages. Bethesda, Maryland.

Bollum, F.J. (1975) Mammalian DNA polymerases. In: *Progress in Nucleic Acid Research and Molecular Biology,* ed. Cohn, W.E., pp. 109–144. Academic Press, New York.

Callan, H.G. (1972) Replication of DNA in the chromosomes of eukaryotes. *Proc. R. Soc. Lond. B 181,* 19–41.

Callan, H.B. (1974) DNA replication in the chromosomes of eukaryotes. *Cold Spring Harbor Symp. Quant. Biol. 38,* 195–203.

Cech, T.R. & Hearst, J.E. (1975) An electron microscopic study of mouse foldback DNA. *Cell 5,* 429–446.

Cohen, J.E., Jasny, B.R. & Tamm, I. (1978) The spatial distribution of initiation sites for mammalian DNA replication: A statistical analysis. *J. Molec. Biol.*, in press.

Davidson, E.H., Hough, B.R., Amenson, C.S. & Britten, R.J. (1973) General interspersion of repetitive with nonrepetitive sequence elements in the DNA of *Xenopus*. *J. Molec. Biol. 77,* 1–23.

Deininger, P.L. & Schmid, C.W. (1976) An electron microscope study of the DNA sequence organization of the human genome. *J. Molec. Biol. 106,* 773–790.

Edenberg, H.J. & Huberman, J.A. (1975) Eukaryotic chromosome replication. *Ann. Rev. Genet. 9,* 245–284.

Green, R. & Bahr, G.F. (1975) Comparison of G-, Q-, and EM-banding patterns exhibited by the chromosome complement of the Indian Muntjac, *Muntiacus muntjak,* with reference to the nuclear DNA content and chromatin ultrastructure. *Chromosoma (Berl.) 50,* 53–67.

Hand, R. & Tamm, I. (1974) Initiation of DNA replication in mammalian cells and its inhibition by reovirus infection. *J. Molec. Biol.* 82, 175–183.

Hsu, M.-T. & Jelinek, W.R. (1977) Mapping of inverted repeated DNA sequences within the genome of simian virus 40. *Proc. Natl. Acad. Sci. U.S.A. 74,* 1631–1634.

Huberman, J.A. & Riggs, A.D. (1968) On the mechanism of DNA replication in mammalian chromosomes. *J. Molec. Biol. 32,* 327–341.

Jasny, B.R., Cohen, J.E. & Tamm, I. (1978) The organization of DNA replication in a mammalian cell line. In: *DNA Synthesis: Present and Future,* ed. Kohiyama, M. & Molineaux, I., pp. 175–188. Plenum Press, New York.

Jelinek, W.R. (1977) Specific nucleotide sequences in HeLa cell inverted repeated DNA: Enrichment of sequences found in double-stranded regions of heterogeneous nuclear RNA. *J. Molec. Biol. 115,* 591–601.

Jelinek, W.R. (1978) Inverted, repeated DNA from CHO cells studied with cloned DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* 75, 2679–2683.

Jelinek, W.R. (1979) Inverted repeats in the genomes of mammalian cells. In: *Alfred Benzon Symposium XIII, Specific Eukaryotic Genes,* ed. Engberg, J., Klenow, H. & Leick, V. Munksgaard, Cohagen.

Jelinek, W., Evans, R., Wilson, M., Salditt-Georgieff, M. & Darnell, J.E. (1978) Oligonucleotides in hnRNA: similarity of inverted repeats and RNA from repetitious DNA sites. *Biochemistry* 17, 2776–2783.

Perlman, S., Phillips, C. & Bishop, J.O. (1976) A study of foldback DNA. *Cell 8,* 33–42.

Robertson, H.D., Dickson, E. & Jelinek, W. (1977) Determination of nucleotide sequences from double-stranded regions of HeLa cell nuclear RNA. *J. Molec. Biol. 115,* 571–589.

Ryskov, A.P., Saunders, G.F., Farashyan, V.R. & Georgiev, G.P. (1973) Double-helical regions in nuclear precursor of mRNA (pre-mRNA). *Biochim. Biophys. Acta 312,* 152–164.

Schmid, C.W. & Deininger. P.L. (1975) Sequence organization of the human genome. *Cell 6,* 345–358.

Subramanian, K.N., Dhar, R. & Weissman, S.M. (1977) Nucleotide sequence of a fragment of SV40 DNA that contains the origin of DNA replication and specifies the 5′ ends of "early" and "late" viral RNA. *J. Biol. Chem. 252,* 355–367.

Taylor, J.H. (1977) Increase in DNA replication sites in cells held at the beginning of S phase. *Chromosoma (Berl.) 62,* 291–300.

Taylor, J.H. & Hozier, J.C. (1976) Evidence for a four micron replication unit in CHO cells. *Chromosoma (Berl.) 57,* 341–350.

Wilson, D.A. & Thomas, C.A.Jr. (1974) Palindromes in chromosomes. *J. Molec. Biol. 84,* 115–144.