TIG — October 1987, Vol. 3, no. 10

Fragile sites in human chromosomes as regions of late-replicating DNA

Charles Laird, Elizabeth Jaffe, Gary Karpen, Mary Lamb and Rebecca Nelson

We review data indicating that fragile sites in chromosomes of humans, Drosophila *and* Microtus *represent regions where DNA is late-replicating in the cell cycle. We suggest that rare fragile sites in human chromosomes represent* cis*-acting alterations to DNA that confer or accentuate late replication at that site. The possible connection between late replication and the fragile X syndrome is also discussed.*

Fragile sites in human chromosomes are characterized cytologically as specific regions that exhibit constrictions, gaps or breaks when cells are cultured and karyotyped (see Ref. 1 for review). In human pedigrees, fragile sites are inherited co-dominantly: normal and gapped homologues (Fig. 1a, b) are observed in the same cell. 'Rare' fragile sites appear to be caused by infrequent alleles (less than 5%) that are segregating in human populations; 'common' fragile sites are observed at higher frequencies in populations[1]. A disproportionate share of fragile sites are reported to be in or near chromosome bands that are also breakpoints for rearrangements found in tumor cells[1-3], although this correlation may be limited to tumors of lymphohematopoietic origin[4]. High frequencies of recombination between sister chromatids at fragile sites, and between homologues at one fragile site, have also been reported[5-8]. Over half of the breakpoints that have occurred during chromosome evolution in primates are reported to be at or near fragile sites[9]. These observations indicate that fragile sites are recombinogenic. One of the rare fragile sites on the X chromosome, Xq27 (the 'fragile X'), is associated with the most frequent form of inherited mental retardation[10,11]. Thus, there are several questions raised by these observations.

(1) What is the genetic basis of fragile sites?

(2) What DNA or chromosome structure results in the observed chromosome gaps and fragility at fragile sites?

(3) What causes chromosomes to rearrange and recombine at fragile sites?

(4) What is the cause of the mental retardation syndrome associated with the X-linked fragile site, Xq27?

We find considerable support for earlier suggestions that fragile sites in human chromosomes represent regions where DNA is late or delayed in its replication[12-17]. Our data on *Drosophila* chromosomes[18], in which 'intercalary heterochromatin' provides a useful model for human fragile sites, support these previous suggestions. We will illustrate one way in which late replication at fragile site Xq27 relates to a mechanism of the fragile X syndrome proposed elsewhere[19].

'Fragile sites' of *Drosophila* chromosomes occur at regions delayed or late in replication

Our literature review of human fragile sites was stimulated by our recent analysis of three sites of intercalary heterochromatin in *Drosophila melanogaster*[18]. The term intercalary heterochromatin was first used to describe a number of sites in euchromatin that exhibit some of the properties of centromeric heterochromatin[20]. The properties of intercalary heterochromatin in *Drosophila* include the occasional presence of ectopic fibers, which connect heterologous chromosome regions in polytene nuclei; higher than normal rates of breakage of polytene chromosomes; and, in meiotic cells, increased frequencies of chromosome translocations with breakpoints at such sites[21,22]. Chromosomal regions containing intercalary heterochromatin often replicate during the latter part of the DNA synthesis (S) phase of the cell cycle[23]. Weak sites in polytene chromosomes also can be induced by an inhibitor of DNA replication, fluorodeoxyuridine[24]. The similarity of these properties to those of human fragile sites suggested to us that *Drosophila* chromosomes may provide a useful model for fragile sites in mammalian chromosomes.

We have analysed the replication level and cytogenetic properties of DNA at three sites of intercalary heterochromatin: 89E, 39DE and 11A (Ref. 18). DNA sequences that have cytogenetic properties of intercalary heterochromatin are under-represented in DNA of polytene chromosomes, but they apparently share no other property with centromeric heterochromatin. On the basis of these data, we have proposed that some sequences of intercalary heterochromatin, such as parts of the

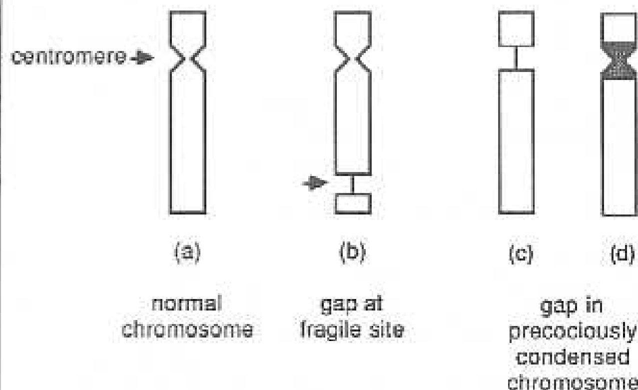

Fig. 1. *Schematic diagram of a normal chromosome (a), a chromosome with a fragile site (b), and a* Microtus *chromosome that has a gap (c) at a region of late DNA replication (d). For simplicity, only one of the two sister chromatids is shown. The comparison in (c) and (d) is after Sperling and Rao[26]. The shaded area in (d) represents a chromosome region with silver grains resulting from incorporation of tritiated thymidine late in S phase. This region corresponds to a gap in the condensed chromosome at or near the centromere (c), indicating that chromatin condensation, when induced close to the time of DNA synthesis, leads to chromosome gaps at regions of the most recent DNA synthesis.*

© 1987, Elsevier Publications, Cambridge 0168 – 9525/87/$02.00

histone and *Ultrabithorax* genes at regions 39DE and 89E, respectively, represent normal euchromatic DNA that is delayed in replication in some cell types; this delayed replication is sufficient to generate the properties of intercalary heterochromatin. A corollary is that any euchromatic sequence, in cells where it is late-replicating, could be classified as intercalary heterochromatin.

Sites of intercalary heterochromatin in *Drosophila* can be cell-type specific. Region 89E, which usually exhibits a constriction in salivary gland cells, is reported not to have a constriction in polytene cells of fat body[25]. This change is correlated with the finding that the *Ultrabithorax* sequences we examined are being replicated at the normal time in fat body rather than delayed as in salivary gland chromosomes[18], thus supporting our conclusion that delayed replication is sufficient to establish a site of intercalary heterochromatin in *Drosophila* chromosomes. These observations in *Drosophila* provide a possible explanation for the observed differences in the frequency with which human fragile sites can be induced in different cell types[1]: such differences may reflect variations in the timing of replication of a specific sequence, as will be discussed below.

Gaps in *Microtus* chromosomes occur at regions where DNA is late-replicating

Fragile sites in human cells are usually detected as gaps that occur at specific sites in condensed chromosomes (Fig. 1a, b). Similar gaps can be induced by condensing chromatin prematurely, using cell fusion techniques. For example, *Microtus* (vole) cells in different phases of the cell cycle were fused after replicating DNA was labeled with tritiated thymidine[26]. Condensed mitotic chromosomes showed gaps that coincided with the regions where DNA synthesis was most active in the period just before condensation (Fig. 1c, d). It was concluded that most DNA replication had been completed in the chromosomes that showed gaps, because two sister chromatids were observed in condensed, non-gapped regions of the chromosomes. Thus, the cells from which the gapped chromosomes originated were in late DNA synthesis (S) phase or in the subsequent G-2 phase of the cell cycle. Moreover, the gaps appeared to result from late replicative synthesis of DNA rather than from repair synthesis because gaps were often at centromeres, which are late-replicating regions in normal cells (figure 8 of Ref. 26). Thus, the major cytogenetic feature of fragile sites – chromosome gaps – can be experimentally induced in normal mammalian chromosomes by shortening the time between completion of normal DNA synthesis and chromosome condensation.

Previous discussions of human fragile sites have considered alterations in both replicative and repair DNA synthesis as potential explanations for fragile sites[12–17]. The results just described for *Drosophila* and *Microtus* chromosomes indicate that delayed or late replicative synthesis of DNA, rather than late repair synthesis, is responsible for these examples of fragile sites; we will therefore use 'late replication' in the following discussion to refer to replicative rather than to repair DNA synthesis.

Detection of fragile sites in cultured human cells

Nature of agents used to induce or suppress human fragile sites

Two methods of increasing the observed frequency of fragile sites in cultured cells involve procedures that shorten G-2 phase, thus decreasing the time between completion of DNA replication and chromosome condensation. Addition of caffeine after folate deprivation[27], or fusion of human and hamster cells, results in a shortened G-2 phase relative to that in normal human cells[28]. Thus, fragile sites, like chromosome gaps that arise from premature chromatin condensation, can be induced by decreasing the time between completion of normal DNA synthesis and chromosome condensation.

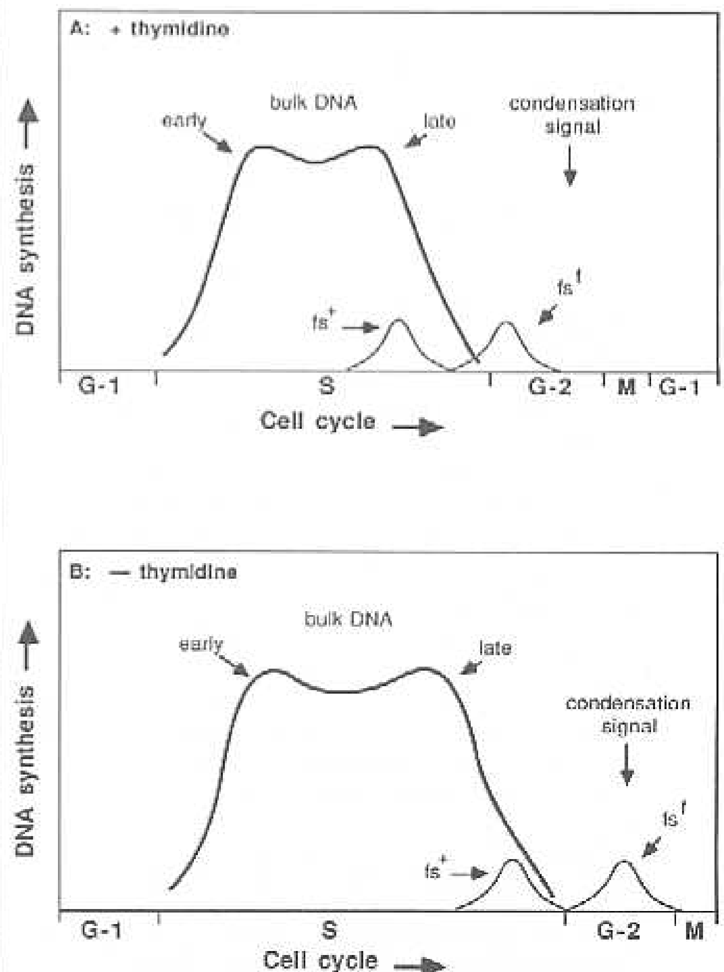

Fig. 2. *Proposed timing of DNA synthesis of fragile site DNA during the cell cycle. The early and late components of 'bulk' DNA synthesis are indicated, together with the replication of DNA at a fragile site. The cell cycle is divided into the two 'gap' stages (G-1 and G-2), the 'DNA synthesis' phase (S) and 'mitosis' phase (M). A normal allele of a fragile site (fs^+), is proposed to replicate late in S phase. A fragile site allele, fs^f, (either a common or a rare allele) replicates even later in the cell cycle, perhaps extending into what is conventionally called G-2. As long as the concentration of thymidine is normal, few or no fragile sites are expressed (A). However, under conditions of thymidine deprivation, for example, replication of the fragile site allele is shifted further into G-2 phase (B). It is proposed that DNA will remain uncondensed if it has not replicated by the time that chromosome condensation is signaled (arrow) in G-2. This failure to condense will lead to a chromosome gap at the late-replicating region.*

A number of other agents influence the frequency of fragile sites in cultured human cells[1]. These agents are used to classify fragile sites as folate-sensitive (i.e. folate-suppressible), distamycin-inducible, bromodeoxyuridine-inducible and aphidicolin-inducible[1]. A particular site is usually induced or suppressed by only a subset of these agents. Folate-suppressible sites may be induced by thymidine deprivation or by inhibition of thymidine synthetase with, for example, fluorodeoxyuridine[29]. Most of the agents that induce fragile sites are known to inhibit DNA synthesis. Some agents affect the concentration of a substrate, such as thymidine triphosphate, or of the enzyme DNA polymerase alpha (aphidicolin; Ref. 13). Others bind directly to DNA (distamycin; Ref. 30). Thus, the conditions used in culture to increase the frequency of detection of fragile sites are ones that can retard or block the normal replication pattern of DNA; conditions that suppress fragile sites are ones that would alleviate a block or delay in replication.

We propose that these conditions of replication inhibition that induce fragile sites have that effect because, like caffeine and cell fusion, they shorten the effective length of G-2 for those sequences at fragile sites; when G-2 is too short, chromosome condensation is incomplete at the late-replicating region, leading to a chromosome gap. This proposed relationship between the cell cycle and the timing of replication for normal (fs^+) and fragile site (fs^f) alleles is illustrated for conditions of suppression (Fig. 2a) and induction (Fig. 2b).

The timing of suppressibility of fragile sites is late in S or early in G-2

The expression of fragile sites in cultured cells can be suppressed by agents that reduce constraints on DNA synthesis, for example by maintaining normal levels of thymidine or folate[31]. The timing of the suppressibility, i.e. that part of the cell cycle during which the expression of the fragile site may be suppressed, is late S or early G-2 phase[32]. This result has been obtained for at least three fragile sites that are folate-suppressible (Xq27, 10q23, 2q13), when either folate or thymidine is the suppressing agent. These data strongly support the conclusion that these three fragile sites are late-replicating.

Expression of common fragile sites is also increased when components of DNA replication are limiting

Most of the information in the above paragraphs refers to rare fragile sites. Less is known about the common fragile sites, a term conventionally used to refer to fragile sites found in more than a few per cent of humans[1]. Some common fragile sites are classifiable as folate-suppressible[29]; many appear to be induced by aphidicolin, an inhibitor of DNA polymerase alpha[13]. Thus, some common fragile sites, like rare fragile sites, are expressed more frequently when components of DNA replication are limiting. As will be discussed below, common fragile sites are generally observed not to be deleterious. Thus the late replication proposed for these regions is either constitutive in cells used for cytogenetic analysis and not deleterious during the reproductive span (as is the case for *Drosophila* sites of intercalary heterochromatin, see above), or it is only induced by the conditions of cell culture rather than occurring *in vivo*. A detailed analysis of the timing during the cell cycle of inducibility and suppressibility would be useful in distinguishing between these two possibilities.

How would late DNA replication lead to the observed properties of fragile sites?

Chromosome gaps

Gaps apparently arise at fragile sites because chromatin is incompletely condensed[33]. This incomplete condensation would occur if there is a limited period of time during G-2 phase in which chromatin is 'marked' for normal condensation (represented by the 'condensation signal' arrow in Fig. 2). DNA that replicates unusually late may miss the marking signal proposed here to be necessary for normal condensation, thus creating a visible chromosome gap (Fig. 3a).

Alternatively, incomplete condensation may result from steric inhibition. The DNA polymerase complex, if still present on late-replicating DNA, could block condensation and cause a chromosome gap (Fig. 3b). An example of such steric inhibition of chromatin condensation is the 'secondary constriction' that occurs apparently as a consequence of the nucleolus remaining associated with chromatin until condensation is almost complete[34].

Chromosome breakage

Fragile X chromosomes occasionally appear, in cell culture, to have lost the chromosome tip distal to the fragile site at Xq27 (Ref. 1). Such terminal deletions may arise from chromosome breakage, which is expected to occur preferentially at gaps when force is applied. (At gaps, there is greater force per unit area on individual chromatin strands, compared with that on the condensed region of a chromatid.) Some of this force may be generated by the process of condensation if ends of an extended, interphase chromosome remain anchored to the nuclear matrix or envelope. Force on late-replicating regions would also be present during anaphase if sister chromatids were not yet fully formed (Fig. 3c). Late-replicating DNA is flanked by replication forks, which are preferentially susceptible to breakage *in vitro*[35], and presumably *in vivo* under conditions of fragile site expression. This fragility is probably caused by the presence of single-stranded regions and a reduced number of double strands at replication junctions (Fig. 3d,e).

Recombinogenicity of fragile sites

Fragile sites are recombinogenic, as evidenced by higher-than-expected frequencies of mitotic sister chromatid exchange, meiotic recombination (for Xq27) and chromosome rearrangements in tumor cells and throughout evolution[2-9]. Relating human fragile sites to intercalary heterochromatin of *Drosophila* suggests a mechanism for this property of recombinogenicity. In *Drosophila* the phenomenon of non-homologous associations observed in polytene chromosomes of salivary glands is termed 'ectopic pairing'[20]. Preferred sites of ectopic pairing are observed; some regions of ectopic pairing in polytene cells have an increased likelihood of being sites of chromosome rearrangements that occur in meiotic cells[16,17].

Ashburner[36] has suggested that the structural basis

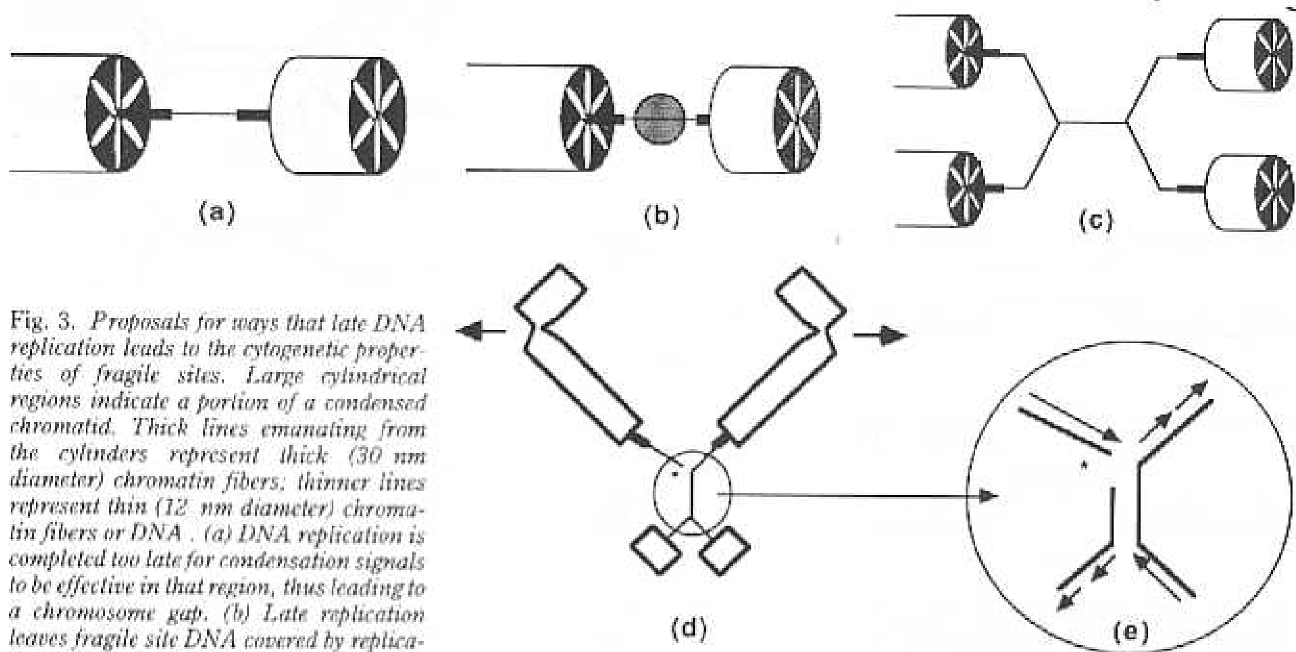

Fig. 3. *Proposals for ways that late DNA replication leads to the cytogenetic properties of fragile sites. Large cylindrical regions indicate a portion of a condensed chromatid. Thick lines emanating from the cylinders represent thick (30 nm diameter) chromatin fibers; thinner lines represent thin (12 nm diameter) chromatin fibers or DNA. (a) DNA replication is completed too late for condensation signals to be effective in that region, thus leading to a chromosome gap. (b) Late replication leaves fragile site DNA covered by replication proteins, which may interfere with chromatin condensation. Such interference occurs at nucleolar organizing regions, leading to a 'secondary constriction', or gap, at sites where a nucleolus was active prior to metaphase*[34]. *(c) Replication is not completed before condensation (sister chromatids remain attached). (d) The attached chromatids in (c) are broken during anaphase, leading to chromatid breaks and 'triradial' structures*[1]. *(e) Structure of broken DNA at site where sister chromatids remain attached. Replication forks are weak points in DNA, probably because of single-stranded regions*[35].

for ectopic pairing is the dispacement of single strands by branch migration at replication forks, followed by pairing of such single strands from different chromosome regions (Fig. 4). The basis of the pairing is not yet known, although limited sequence homology of strands emanating from different chromosome regions (Fig. 4b), or the binding of two single strands by a protein that binds DNA without sequence specificity (Fig. 4c; Ref. 37), are two possibilities. Late-replicating DNA would be flanked, until late in S or G-2 phase, by replication forks involved in the synthesis of earlier-replicated DNA. We suggest that these replication forks at human fragile sites increase the availability of single-stranded regions, leading to increased recombination and chromosomal rearrangement. Thus, exchange events are expected to be increased at replication forks that flank late-replicating DNA, more than in the late-replicating DNA itself.

What types of alteration in DNA could lead to its late replication?

The data described in the previous sections support the conclusion that DNA sequences at fragile sites are, or can be induced to be, late-replicating during the S or G-2 phases of the cell cycle. We now explore possible causes of this late replication.

Two conclusions can be made about the types of alteration that lead to fragile sites: (1) there must be several different kinds of alteration to DNA that can lead to late replication of rare fragile sites, because a particular fragile site may be induced or suppressed by some but not all of the agents described above[1], and (2) the alterations in DNA that lead to a fragile site are *cis*-acting rather than *trans*-acting, because most individuals with a rare fragile site do not carry or express other rare fragile sites[1]. Direct support for this latter conclusion comes from an experiment in which the distal end of a fragile X chromosome was translocated to a hamster chromosome[38]. The fragile X site Xq27 was expressed in these hybrid cells in the absence of any other particular human chromosome. Thus, the mutation is either at or closely linked to fragile site Xq27.

There are several kinds of *cis*-acting alteration to DNA that are expected to result in late replication. (As justified in the section on precocious chromosome condensation, we are limiting our discussion to alterations affecting replicative rather than repair DNA synthesis.) Mutations may affect the efficiency with which an origin or terminus of replication is used, as well as the time during the cell cycle at which it is active[39–48].

In addition to *cis*-acting mutations that directly affect DNA sequences involved in the timing of replication, we expect other genetic alterations to be important. Greatly expanding the distance between two replication origins by unequal recombination[49] or by insertion of DNA could result in delayed replication for the region. Some methylation events, for example those involved with X chromosome inactivation[50], are expected to lead to or maintain late replication. Spontaneous methylation of DNA may occur and be propagated by a 'maintenance methylase', an enzyme that can methylate the non-methylated strand of half-methylated DNA[50–54].

This latter suggestion – that methylation at a fragile site may be involved in late replication – is especially plausible for fragile site Xq27. The cytological expression of this fragile site in hybrid cells is inhibited when 5-azacytidine is added two cell divisions prior to harvesting cells for cytological examination[55]. This result indicates that methylation of DNA may play a role in initiating and/or maintaining the defect present

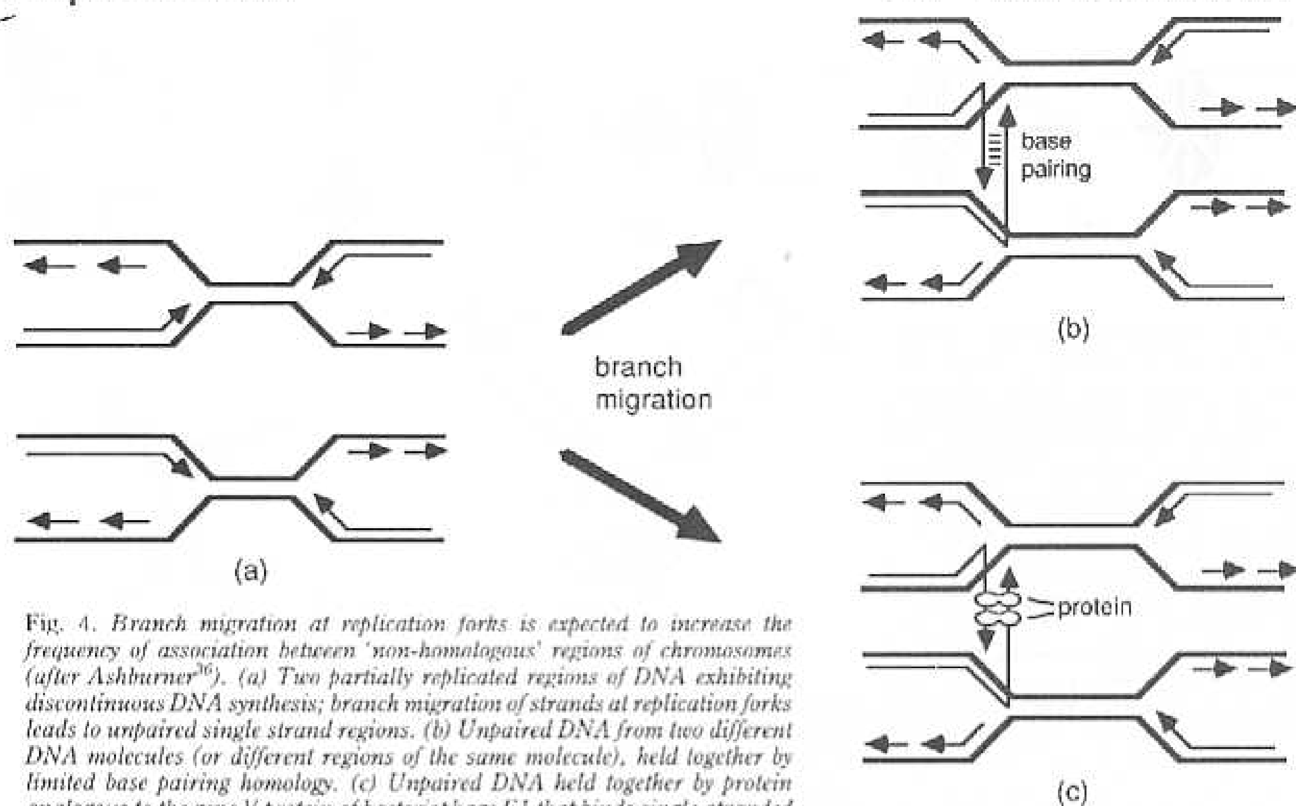

Fig. 4. Branch migration at replication forks is expected to increase the frequency of association between 'non-homologous' regions of chromosomes (after Ashburner[36]). (a) Two partially replicated regions of DNA exhibiting discontinuous DNA synthesis; branch migration of strands at replication forks leads to unpaired single strand regions. (b) Unpaired DNA from two different DNA molecules (or different regions of the same molecule), held together by limited base pairing homology. (c) Unpaired DNA held together by protein analogous to the gene V protein of bacteriophage F1 that binds single-stranded DNA without sequence specificity[37]. Little or no sequence homology is observed for ectopic pairing sites in Drosophila*[18,70], suggesting either that regions of sequence homology are very short or that some other component such as protein is mediating the association of heterologous chromosome regions.*

at Xq27. This is an intriguing observation because methylation appears to be involved in maintaining late replication and transcriptional inactivation of the inactive X chromosome, which is a component of normal dosage compensation in female mammals[50,56]. Reduced methylation caused by 5-azacytidine is correlated with transcriptional reactivation and, on occasion, a return to early replication[57,58]. Another strong parallel with the late-replicating X chromosome is that the 5-azacytidine suppression of the fragile X site does not require the continued presence of 5-azacytidine. Cells grown for only one day in 5-azacytidine gave rise to cells that no longer express the fragile site even after 17 days growth in the absence of 5-azacytidine (Table II of Ref. 59: total of five expressed fragile sites expected in the 143 chromosomes examined in six cell lines; the absence of any expressed fragile X site in 5-azacytidine treated cells is significant at a level of 0.02, according to a one-tailed χ^2 test).

These results are consistent with the suggestion that the fragile site Xq27 is late-replicating due to hypermethylation, and that a stable return to earlier replication can occur after treatment with hypomethylating agents. (Recently it has been concluded that 5-azacytidine suppresses fragile X expression only under conditions in which methylation of DNA is not affected[60]. However, these authors examined methylation of DNA at several autosomal loci and not at fragile X site Xq27. The altered pattern of DNA replication proposed here for fragile site Xq27 may well affect the conditions of 5-azacytidine treatment that are necessary for demethylation.)

Phenotypic effects of late replication

The regions in *Drosophila* polytene chromosomes that behave as fragile sites represent normal, wild-type alleles at various loci. These alleles exhibit no unusual phenotypes other than the cytogenetic effects described here and elsewhere[18]. Similarly, at least three autosomal fragile sites in human chromosomes are without major phenotypic effect even when homozygous[61–63]. Other fragile sites of human autosomes, in heterozygous individuals, also are not correlated with major phenotypic changes beyond the cytogenetic and recombinogenic effects mentioned above. This result would be expected for rare recessive or neutral alleles at autosomal loci.

For rare fragile sites on the human X chromosome, however, even recessive phenotypes should be detectable because of the hemizygosity of the X chromosome in males, and the functional hemizygosity of the X chromosome in females, brought about by X chromosome inactivation[64]. As mentioned above, a major phenotypic consequence has been described for the X chromosome fragile site Xq27, which represents a rare allele present at a frequency greater than 5×10^{-4}. This fragile X site has been well characterized because of its associated syndrome of mental retardation[1]. The very unusual inheritance and expression of the fragile X mutation (it must first be passed through a female to be expressed, and there is incomplete penetrance and variable expressivity in heterozygous females and hemizygous males – Refs 65, 66), have led to its characterization as 'unique'[17] and 'enigmatic'[67].

Recently, we have proposed a formal explanation

for the fragile X mutation at Xq27: the mutation behaves as if it interferes locally with the reactivation of an inactivated fragile X chromosome that normally occurs prior to oogenesis[19]. This mutation has little or no detectable effect until the chromosome is inactivated in pre-oogonial cells as part of the normal mechanism of dosage compensation in female mammals. (This explains why the fragile X chromosome must be passed through a female for expression of the mutant phenotype: only females inactivate an X chromosome for dosage compensation.) In subsequent preparation for oogenesis, reactivation is attempted but is only partially effective because of a block at Xq27. This block to reactivation leads to an 'imprinted' fragile X chromosome that is deleterious in progeny because the inactivated region of the X chromosome exhibits greatly reduced transcriptional activity relative to the active X. Reduced transcriptional activity at this region is the proposed basis of the phenotypes of the fragile X syndrome[19]. Thus, any late replication that blocks X chromosome reactivation could have major phenotypic effects in subsequent generations. We propose here that the basis of the block to reactivation is the late replication of DNA at fragile site Xq27.

How can an alteration that leads to late replication block chromosome reactivation?

Little is known about the normal reactivation process for X chromosomes. Speculations have focused on removal of methyl groups or of methylated bases from DNA by enzymes or by replication[51–53,68,69]. Reactivation of X chromosome genes with base analogues that lead to hypomethylation of DNA is consistent with these speculations[50,56]. As mentioned earlier, 5-azacytidine will suppress the appearance of fragile site Xq27, suggesting that methylation of cytosine in DNA is necessary for cytological expression of the site[38,55]. Since folate also will suppress the appearance of a fragile site at Xq27, it appears that replication at Xq27 can be delayed by interfering with DNA replication at the level of both nucleoside pools and methylated cytosine. It has been proposed that in normal chromosomes, Xq27 is a common fragile site, based on its inducibility at low frequency in hybrid cells[49]. Taken together, these data suggest that Xq27 is late-replicating in many apparently normal chromosomes, and that an alteration can retard its replication even further.

Our conclusions concerning the nature of fragile sites as regions of late-replicating DNA, and the proposed block to complete reactivation of the X chromosome by fragile site Xq27, are supported by indirect experimental evidence including human pedigree data[19]. How might the processes of DNA replication and chromosome reactivation be related? We propose the following model to illustrate how late DNA replication could block chromosome reactivation. In oogonial cells, at the last cell division prior to reactivation, the maintenance methylase becomes inactive, resulting in half-methylated DNA (Fig. 5). A demethylase that removes methyl groups from 5-methylcytosine[68], or a glycosylase that replaces 5-methylcytosine by cytosine[69], acts late in S phase or in G-2 to remove the remaining methyl groups, but only from half-methylated DNA. Fragile site DNA has not

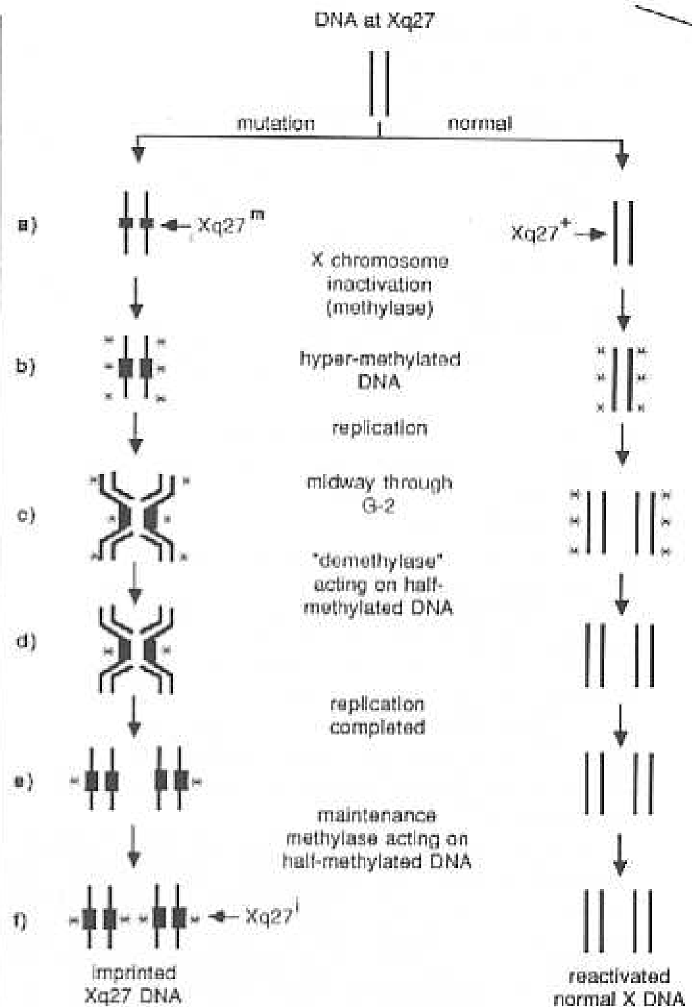

Fig. 5. Proposed mechanism by which late replication blocks the process of X chromosome reactivation prior to oogenesis. (a) A mutation (called 'pre-mutation' in Refs 65, 66) changes normal DNA at Xq27 (Xq27^{+}) to a mutated allele Xq27^{m}, which replicates later in the cell cycle. (b) As part of the normal process of X chromosome inactivation in female cells[64], the inactive X chromosome becomes methylated. (c) In preparation for the reactivation that occurs prior to oogenesis[71,72], one cell division occurs without maintenance methylase activity. (d) A 'demethylase'[68], or a DNA glycosylase that replaces 5-methylcytosine by cytosine[69], removes methyl groups from half-methylated DNA. (e) DNA synthesis is finally complete at Xq27, but it is too late for reactivation of that region because half-methylated DNA was not available during the time that 'demethylase' was present. (f) In a subsequent cell cycle, maintenance methylase activity is again present, restoring full methylation to the DNA at fragile site Xq27. At this time the fragile X chromosome is considered to be 'imprinted'[19]. The imprinted allele, Xq27^{i}, is even later-replicating, and it is now more easily detectable as a chromosome gap under conditions of fragile site induction. The imprinted allele Xq27^{i} is then stably inherited through both males and females[19]. For this illustration, the mutated fragile X chromosome is indicated as imprinted upon completion of the cycle of inactivation/reactivation. It is not known, however, whether or not the entire cycle must be completed before imprinting is observable cytologically.

completed replication and thus does not present half-methylated DNA at the time of enzyme activity. After replication is completed, the residual methyl groups on the late-replicating DNA provide signals for maintenance methylase activity in subsequent cell divisions. Thus, the DNA around fragile site Xq27 remains marked, or imprinted, by the reactivation step prior to oogenesis.

Concluding remarks

We now summarize our proposed answers to the questions raised in our introduction.

(1) What is the genetic basis of fragile sites? Rare fragile sites in human chromosomes represent *cis*-acting alterations to DNA that lead to late replication; common fragile sites represent DNA sequences that are induced to be, or are constitutively, late-replicating.

(2) What DNA or chromosome structure results in the observed chromosome gaps and fragility at fragile sites? Incomplete chromatin condensation caused by late replication, and incomplete replication, account for gaps and fragility.

(3) What causes chromosomes to rearrange and recombine at fragile sites? Recombinogenicity is a consequence of the replication junctions that flank late-replicating DNA.

(4) What is the cause of the mental retardation syndrome associated with the X-linked fragile site, Xq27? An alteration to or mutation in DNA leads to late replication at fragile site Xq27; this mutation locally blocks the reactivation, prior to oogenesis, of a previously inactive fragile X chromosome. This block to reactivation results in an imprinted fragile X chromosome; transcriptional inactivity of genes at Xq27 in an imprinted fragile X is the cause of mental retardation[19].

Direct tests of the conclusions and proposals presented here and elsewhere[19] are needed. An analysis of the timing of replication of specific DNA sequences at fragile sites, under conditions that induce or suppress their appearance, would be especially informative. The interference of late replication with chromosome condensation[26] needs further analysis, as does the possible recombinogenicity of DNA at replication junctions.

For the fragile X site Xq27, the analyses of the timing of DNA replication and the pattern of transcription may be more complex than for most other fragile sites. The ideas presented here and elsewhere[19] are consistent with there being at least four 'alleles', or 'chromosome states', at Xq27. (1) A 'normal' allele would not lead to a chromosome gap at Xq27 under conditions of fragile X induction. (2) A 'common' fragile site allele at Xq27 would result in a chromosome gap in a small fraction of cells (about 2% under conditions described in Ref. 59). (3) A 'mutated' allele, such as proposed for transmitting fragile X males and some heterozygous fragile X females[19], would be inducible at intermediate levels of about 10% (Ref. 59). (4) An 'imprinted' allele, found in affected males and some heterozygous females[19], would be inducible at high levels (about 50%, Ref. 59).

We predict that each allele represents DNA with a characteristic time of replication during the cell cycle; replication of DNA is successively later in the cell cycle for the four alleles, in the order described above. Only the imprinted allele is expected to result in significantly decreased levels of transcription of genes in this region. Experiments directed at testing replication and transcription patterns of DNA at fragile site Xq27 should therefore be carried out with fragile X chromosomes representing each of the four alleles or chromosome states.

We suggest in this article that late replication at Xq27 is the basis of a local block to reactivation of an inactivated fragile X chromosome[19]. One model is proposed here for the basis of this relationship (Fig. 5); this and other models should be evaluated with experiments directed at elucidating the molecular mechanisms of chromosome inactivation and reactivation. The discovery of methods to reverse chromosome inactivation, and to restore normal replication and transcription properties to sequences at fragile sites, could have significant clinical implications.

Acknowledgements

H. Hoehn, M. Goldman, J. Shepherd, M. Champe, V. Schweickart and J. Schaefer provided helpful comments. Support was provided by the National Institutes of Health (General Medical Sciences) and the National Science Foundation (Eukaryotic Genetics).

References

1 Sutherland, G. and Hecht, F. (1985) *Fragile Sites on Human Chromosomes (Oxford Monographs on Medical Genetics No. 13)*, Oxford University Press
2 Yunis, J. (1983) *Science* 221, 227–240
3 LeBeau, M. and Rowley, J. (1984) *Nature* 308, 607–608
4 Muleris, M., Dutrillaux, A-M., Lombard, M. and Dutrillaux, B. (1987) *Cancer Genet. Cytogenet.* 25, 7–13
5 Szabo, P. *et al.* (1984) *Proc. Natl Acad. Sci. USA* 81, 7855–7859
6 Davies, K. *et al.* (1985) *Hum. Genet.* 70, 249–255
7 Glover, T. and Stein, C. (1986) *Am. J. Hum. Genet.* 39, A114
8 Wenger, S., Hennessey, J. and Steele, M. (1987) *Am. J. Med. Genet.* 26, 909–914
9 Miro, R., Clemente, I., Fuster, C. and Egozcue, J. (1987) *Hum. Genet.* 75, 345–349
10 Lubs, H. (1969) *Am. J. Hum. Genet.* 21, 231–244
11 Opitz, J., Reynolds, J. and Spano, L., eds (1986) *Am. J. Med. Genet.* 23, 1–10
12 Taylor, W. and Hagerman, P. (1983) in *The Fragile X Syndrome* (Hagerman, R. and McBogg, P., eds), pp. 115–152, Spectra Publishing Co.
13 Glover, T., Berger, C., Coyle, J. and Echo, B. (1984) *Hum. Genet.* 67, 136–142
14 Gollin, S., Holmquist, G. and Ledbetter, D. (1985) *Am. J. Hum. Genet.* 37, 208–214
15 Sutherland, G., Baker, E. and Fratini, A. (1985) *Am. J. Med. Genet.* 22, 433–443
16 Sutherland, G. and Baker, E. (1986) *Am. J. Med. Genet.* 23, 409–417
17 Nussbaum, R. and Ledbetter, D. (1986) *Annu. Rev. Genet.* 20, 109–145
18 Lamb, M. and Laird, C. (1987) *Chromosoma* 95, 227–235
19 Laird, C. *Genetics* (in press)
20 Kaufmann, B. (1939) *Proc. Natl Acad. Sci. USA* 25, 571–577
21 Zhimulev, I., Semeshin, V., Kulichkov, V. and Belyaeva, E. (1982) *Chromosoma* 87, 197–228
22 Mukhina, L., Kulitchkov, V. and Zhimulev, I. (1981) *Drosoph. Inf. Serv.* 56, 98–103
23 Arcos-Teran, L. (1972) *Chromosoma* 37, 233–296
24 Hagle, E. (1972) *Chromosoma* 33, 297–301
25 Richards, G. (1980) *Chromosoma* 79, 241–250
26 Sperling, K. and Rao, P. (1974) *Chromosoma* 45, 121–131
27 Yunis, J. and Soreng, A. (1984) *Science* 226, 1199–1204
28 Ledbetter, D., Airhart, S. and Nussbaum, R. (1986) *Am. J. Med. Genet.* 23, 445–455
29 Glover, T. (1981) *Am. J. Hum. Genet.* 33, 234–242
30 Schmid, M., Klett, C. and Niederhofer, A. (1980) *Cytogenet. Cell Genet.* 28, 87–94
31 Sutherland, G. (1977) *Science* 197, 265–266
32 Sutherland, G. (1979) *Am. J. Hum. Genet.* 31, 125–135
33 Harrison, C., Jack, E. T., Allen, T. and Harris, R. (1983) *J. Med. Genet.* 20, 280–285
34 Goessens, G. (1984) *Int. Rev. Cytol.* 87, 107–158
35 Pettijohn, D. and Hanawalt, P. (1964) *J. Mol. Biol.* 8, 170–174
36 Ashburner, M. (1980) in *Insect Cytogenetics* (Blackman, R., Hewitt, G. and Ashburner, M., eds), pp. 65–84, Symposia of the Royal Entomological Society of London No. 10

37 Model, P., Russel, M. and Boeke, J. (1981) in *Bacteriophage Assembly* (Dubow, M. S., ed.), pp. 389–400, Alan R. Liss
38 Nussbaum, R., Airhart, S. and Ledbetter, D. (1986) *Am. J. Med. Genet.* 23, 457–466
39 Nagley, P. (1985) *Trends Genet.* 1, 290–291
40 Nagley, P. (1986) *Trends Genet.* 2, 221–222
41 Baran, N., Neer, A. and Manor, H. (1983) *Proc. Natl Acad. Sci. USA* 80, 105–109
42 Shortle, D. and Nathans, D. (1979) *J. Mol. Biol.* 131, 801–817
43 Brouch, J. *et al.* (1983) *Cold Spring Harbor Symp. Quant. Biol.* 47, 1165–1173
44 Orr-Weaver, T. and Spradling, A. (1986) *Mol. Cell. Biol.* 6, 4624–4633
45 Callan, H. (1973) *Cold. Spring Harbor Symp. Quant. Biol.* 38, 195–198
46 Blumenthal, A., Kriegstein, H. and Hogness, D. (1973) *Cold Spring Harbor Symp. Quant. Biol.* 38, 205–223
47 Hammond, M. and Laird, C. (1985) *Chromosoma* 91, 267–278
48 Holmquist, G. (1987) *Am. J. Hum. Genet.* 40, 151–173
49 Ledbetter, D., Ledbetter, S. and Nussbaum, R. (1986) *Nature* 324, 161–163
50 Mohandas, T., Sparkes, R. and Shapiro, L. (1981) *Science* 211, 393–395
51 Riggs, A. (1975) *Cytogenet. Cell Genet.* 14, 9–25
52 Holliday, R. and Pugh, J. (1975) *Science* 187, 226–232
53 Sager, R. and Kitchin, R. (1975) *Science* 189, 426–433
54 Jones, P. and Taylor, S. (1981) *Nucleic Acids Res.* 9, 2933–2947 2947
55 Mixon, C. and Dev, V. (1983) *Am. J. Hum. Genet.* 35, 1270–1275
56 Venolia, L. *et al.* (1982) *Proc. Natl Acad. Sci. USA* 79, 2352–2354
57 Schmidt, M., Wolf, S. and Migeon, B. (1985) *Exp. Cell Res.* 158, 301–310
58 Shafer, D. and Priest, J. (1984) *Am. J. Hum. Genet.* 36, 534–545
59 Ledbetter, D., Airhart, S. and Nussbaum, R. (1986) *Am. J. Med. Genet.* 23, 429–444
60 Glover, T. *et al.* (1986) *Am. J. Hum. Genet.* 38, 309–318
61 Berg, J. *et al.* (1969) *Lancet* i, 531
62 Sutherland, G. (1981) *Am. J. Hum. Genet* 33, 946–949
63 Izakovic, V. (1984) *Hum. Genet.* 68, 340–341
64 Lyon, M. (1961) *Nature* 190, 372–373
65 Pembrey, M., Winter, R. and Davies, K. (1985) *Am. J. Med. Genet.* 21, 709–717
66 Sherman, S. *et al.* (1985) *Hum. Genet.* 69, 289–299
67 Sutherland, G. (1985) *Trends Genet.* 1, 108–112
68 Gjerset, R. and Martin, D. (1982) *J. Biol. Chem.* 257, 8581–8583
69 Razin, A. *et al.* (1986) *Proc. Natl Acad. Sci. USA* 83, 2827–2831
70 Pardue, M. L., Kedes, L., Weinberg, E. and Birnstiel, M. (1977) *Chromosoma* 63, 135–151
71 Gartler, S. *et al.* (1972) *Cell. Differ.* 1, 215–223
72 Gartler, S., Rivest, M. and Cole, R. (1980) *Cytogenet. Cell Genet.* 28, 203–207

C. Laird, M. Lamb and R. Nelson are at the Dept of Zoology, and E. Jaffe and G. Karpen are at the Dept of Genetics, University of Washington, NJ-15, Seattle, WA 98195, USA.

Chloroplast gene organization in plants

Kazuhiko Umesono and Haruo Ozeki

Complete sequences are now available for the chloroplast genomes of two green plants. The information that can be gleaned from these sequences should help us to understand not only how chloroplasts function within present-day plants, but may also yield insights into the evolutionary relationships of photosynthetic organisms and into the gene movement that has occurred among the various genetic compartments of eukaryotic cells.

Chloroplasts, the organelles within the cells of green plants that carry out the reactions of photosynthesis, contain their own unique DNA molecules (the chloroplast genome). The genetic system of the chloroplast has many features in common with prokaryotic organisms and distinct from the (eukaryotic) nuclear–cytoplasmic system. These features have been invoked in support of the hypothesis that green-plant cells evolved from a symbiosis between a eukaryotic host cell and a photosynthetic prokaryote. Various mutations resulting in photosynthesis defects have been shown to be inherited maternally through the cytoplasm, indicating that the chloroplast genes play an essential role in the normal growth of whole plants. However, the chloroplast is not completely autonomous: many nuclear genes are also known to be required for the maintenance of the functional organelle.

In order to understand the genetic control of chloroplasts in plant cells, it is necessary first of all to learn how many and what kinds of genes are accommodated in the chloroplast genome, and to determine the basic mechanism for their expression. Within the last year, we have moved much closer towards these goals, with the complete sequencing of the chloroplast genome from two green plants: the bryophyte *Marchantia polymorpha* (a liverwort)[1] and the tracheophyte *Nicotiana tabacum* (tobacco)[2].

Structure of plant chloroplast genomes

A typical plant chloroplast genome consists of a unique circular dsDNA molecule, 120–200 kbp in length, that in many cases contains a pair of inverted repeat (IR) sequences separating large single-copy (LSC) and small single-copy (SSC) regions (reviewed in Refs 3 and 4). Each chloroplast contains tens to hundreds of copies of the genome and there may be as many as 20–50 chloroplasts per cell; thus each cell is highly polyploid with respect to its chloroplast gene complement.

A comparison of the two sequenced chloroplast genomes shows that, although the tobacco chloroplast genome (155 844 bp) is roughly 25% larger than that of the liverwort (121 024 bp), the corresponding LSC and SSC regions are almost the same size (81–87 kbp and 18–20 kbp respectively), and the difference in total size is due mainly to the length of the IR region. Although these two plant species are evolutionarily very distant from one another, their deduced gene organization is remarkably similar, despite an inversion of about 30 kbp in the LSC region (see centrepage Fig.). This implies that the genomes of chloroplasts in all green plants may have arisen from a unique ancestor and that the basis of the present nuclear–chloroplast relationship was established before the divergence of bryophytes and tracheophytes.

The numbers of structural genes encoded by the chloroplast genome are estimated to be 136 in liverwort (see Fig.) and about 150 in tobacco, with 9 and

© 1987, Elsevier Publications, Cambridge 0168 – 9525/87/$02.00