

Inversions with Deletions and Duplications

Alasdair J. E. Gordon* and Jennifer A. Halliday†

*Section de Biologie, Institut Curie, 75231 Paris, France, and †Groupe de Mutagenèse, Institut Monod, 75251 Paris, France

Manuscript received August 4, 1994

Accepted for publication February 7, 1995

ABSTRACT

Complex mutational events, including *de novo* inversion with deletion and duplication of sequence, have been observed but are difficult to model. We propose that nascent leading-strand misalignment upon the lagging-strand template during DNA replication can result in the inversion of sequence. The positioning of this misalignment and of the realignment of the leading strand back onto the leading-strand template will determine if the inversion is accompanied by deletion and duplication of sequence. We suggest that such strand misalignment-realignment events may occur at the replication fork during concurrent DNA replication.

A number of mutational events involving inversion of DNA sequence have recently been characterized in model eukaryotic systems and in humans suffering genetic disease. Such events include inversion leaving flanking sequence intact (LAKICH *et al.* 1993; Class A), inversion in conjunction with deletion of sequence at one end (LI and BRAY 1993; Class B) or at both ends (SOMMER and KETTERLING 1993; Class C), and inversion of sequence, some of which is also present in the normal orientation (*i.e.*, partial duplication), with deletion of sequence at one end (ROTHSTEIN *et al.* 1987; Class D); see Figure 1. In all cases, inverted-repeat sequences that range from a few to hundreds of bases are present in the wild-type configuration and are associated with the rearranged end points. Although a variety of mechanistically different models have been proposed to account for each inversion class, none appears sufficiently robust to explain all such events. Intrachromosomal homologous recombination (LAKICH *et al.* 1993) followed by an unequal crossover (LI and BRAY 1993) may account for Classes A and B, respectively, but cannot account for Classes C and D. Moreover, patient 9 of LAKICH *et al.* (1993), suffering severe hemophilia A due to a disruption of the factor VIII gene, may represent an inversion in conjunction with a partial duplication of sequence (Figure 1), an event difficult to model based on intrachromosomal homologous recombination. The double-loop model (SOMMER and KETTERLING 1993), which involves a looping single-stranded DNA replication intermediate that is resolved by staggered endonucleolytic cleavage and subsequent gap repair, can only account for Class C events, because the replication utilizing the cut ends that produces the inversion would always result in

deletion of the flanking sequence. A model proposed for Class D events, involving double-strand-break repair intermediates resulting in multiple Holliday structures that are subsequently resolved (ROTHSTEIN *et al.* 1987), in effect inverted gene conversion, may also explain the other classes, but this is difficult to assess due to the complexity of the required structures. Here we describe a strand misalignment-realignment model that incorporates possibilities intrinsic to concurrent DNA replication (WAGA and STILLMAN 1994) and that can readily account for all the above classes of inversion events.

We propose that during concurrent DNA replication, the leading strand misaligns with the lagging-strand template, facilitated by the complementarity provided by an inverted-repeat series (Figure 2). Misaligned replication proceeds, thereby inverting the sequence replicated with respect to its original orientation. The leading strand ultimately realigns onto the leading-strand template, again facilitated by the complementarity provided by a second inverted-repeat series. Realigned replication continues, producing a region of profound noncomplementarity (the nascent leading strand differing from the leading-strand template in sequence and size) bounded by regions of complete complementarity, *i.e.*, double-stranded sequence. A subsequent round of repair or chromosomal replication across the altered strand fixes the mutational event. All the inversion classes can be modeled as the rearrangement of blocks of sequence (inverted, deleted or duplicated) defined by inverted-repeat series whose order of replication is determined by leading-strand misalignments and realignments.

This model incorporates the single-stranded nature of the lagging-strand template immediately preceding the replisome complex and the processive nature of the replisome itself. During concurrent replication, the two polymerase functions may alternatively

Corresponding author: Alasdair J. E. Gordon, Centre de Recherche, Hopital Maisonneuve-Rosemont, 5415 boul. de l'Assomption, Montréal, Québec, Canada HIT 2M4.

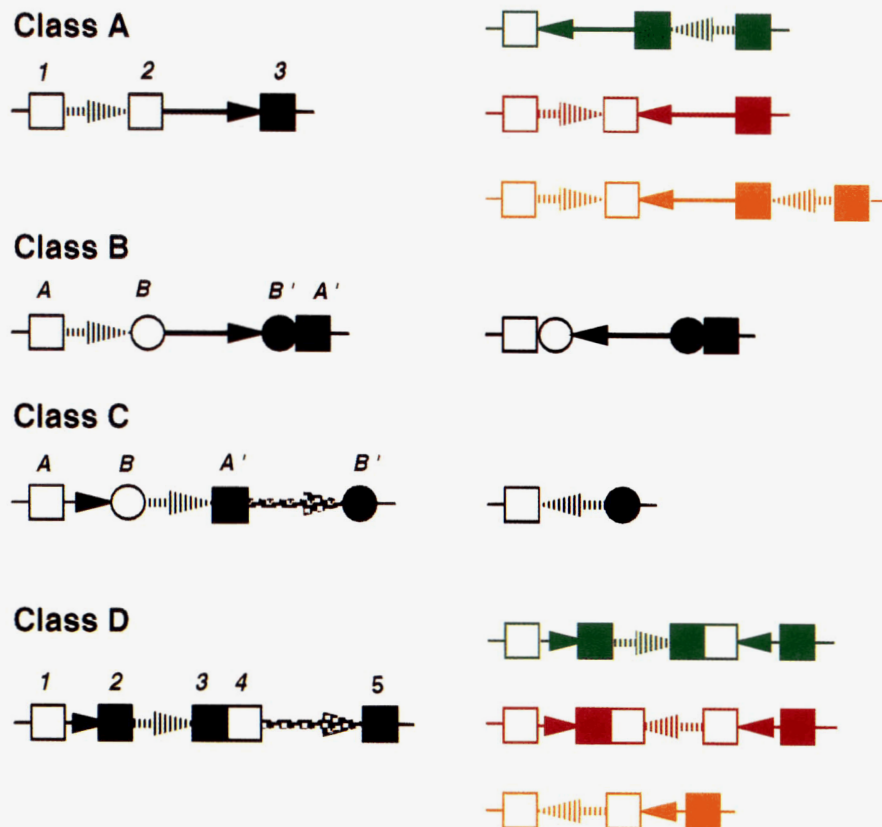


FIGURE 1.—Inverted repeat series associated with sequence inversion. Numbering indicates identical or similar sequence (with stretches of identity), letters represent repeat pairs, and filled and open symbols represent opposite orientations of the repeat (direct *vs.* inverted orientation). Arrows indicate orientation of the sequence blocks defined by repeats that undergo inversion, deletion or duplication. The left and right halves depict the wild-type and altered configuration(s), respectively. Class A: A 500-kb inversion disrupting the gene encoding human factor VIII (LAKICH *et al.* 1993) is bounded by three copies of gene A (1.8 kb). Both copy 1 and copy 2 have independently been involved in inversion with copy 3 (green and red configurations, respectively). Interestingly, patient 9 of LAKICH *et al.* (1993) may represent an inversion between 1 and 3, which includes a partial duplication of the region between 1 and 2 because this region is also present in the original orientation (yellow structure, the initial number and location of gene A copies being required to address this possibility). Class B: A 1-kb deletion (A to B) associated with a 15-kb inversion (B to B') in the human *GP1IIa* gene (LI and BRAY 1993). Repeats A and B are 4 and 5 bp, respectively (see Figure 2D); repeat A, repeat B and repeats B' and A' are contained within separate *Alu* sequences. Class C: A 92-bp inversion (B to A') associated with flanking 5.4-kb deletions (A to B and A' to B') in the gene encoding human factor IXHB209 (SOMMER and KETTERLING 1993). Both repeats are 6 bp long. Class D: In *S. cerevisiae* a genetic system was designed in a region containing five δ sequences (δ sequences being 330 bp long); selection was for loss of loci within the region bounded by δ sequences 4 and 5, with retention of the region between δ sequences 2 and 3 (ROTHSTEIN *et al.* 1987). In pairwise comparisons, the five δ sequences range from 69.4 to 97.3% identity and therefore a variety of repeat series can be arbitrarily defined depending on how much nonidentity is included; here we will use only identity in our repeat series classifications. Therefore, the boxes represent both δ sequences and, for our purposes, repeat series (of complete identity) imbedded within the δ sequences. Five inversion classes were seen. The green configuration depicts both classes I and II (in the nomenclature of ROTHSTEIN *et al.*), the red configuration depicts both classes III and IV, and the yellow configuration depicts class VII. Classes V and VI were simple deletions and are therefore not discussed here. The difference between classes I and II, and between classes III and IV, is only that different repeat series within the same δ sequences comprise the observed mutant junctions. For example, in the red configuration, a 1.9-kb *SUP4o* deletion (between repeats 4 and 5) was associated with an inversion of 3.5 kb (between repeats 1 and 2), because the region between repeats 1 and 2 is also present in its original orientation, this region has been duplicated. Repeat-sequence lengths for the class III event would be 4 bp (within repeats 1 and 5) and 239 bp (within repeats 2 and 4), while for the class IV event the repeat-sequence lengths would be 26 bp (within repeats 1 and 5) and 38 bp (within repeats 2 and 4), based on sequence identity.

retard and accelerate each other's progression due to the nature and integrity of the sequence being replicated, *e.g.*, one polymerase stalling after encountering DNA damage and then resuming synthesis after the damage is repaired. Such flux may permit misalignment (a stalled lagging strand polymerase exposing the lagging-strand template) and precipi-

tate realignment (resumption of lagging strand synthesis displacing the leading strand from the lagging strand template).

Concurrent DNA synthesis has been considered in relation to complex mutational events found in *Escherichia coli dnaE173*, where half of a duplicated sequence was found in an inverted orientation (MO *et*

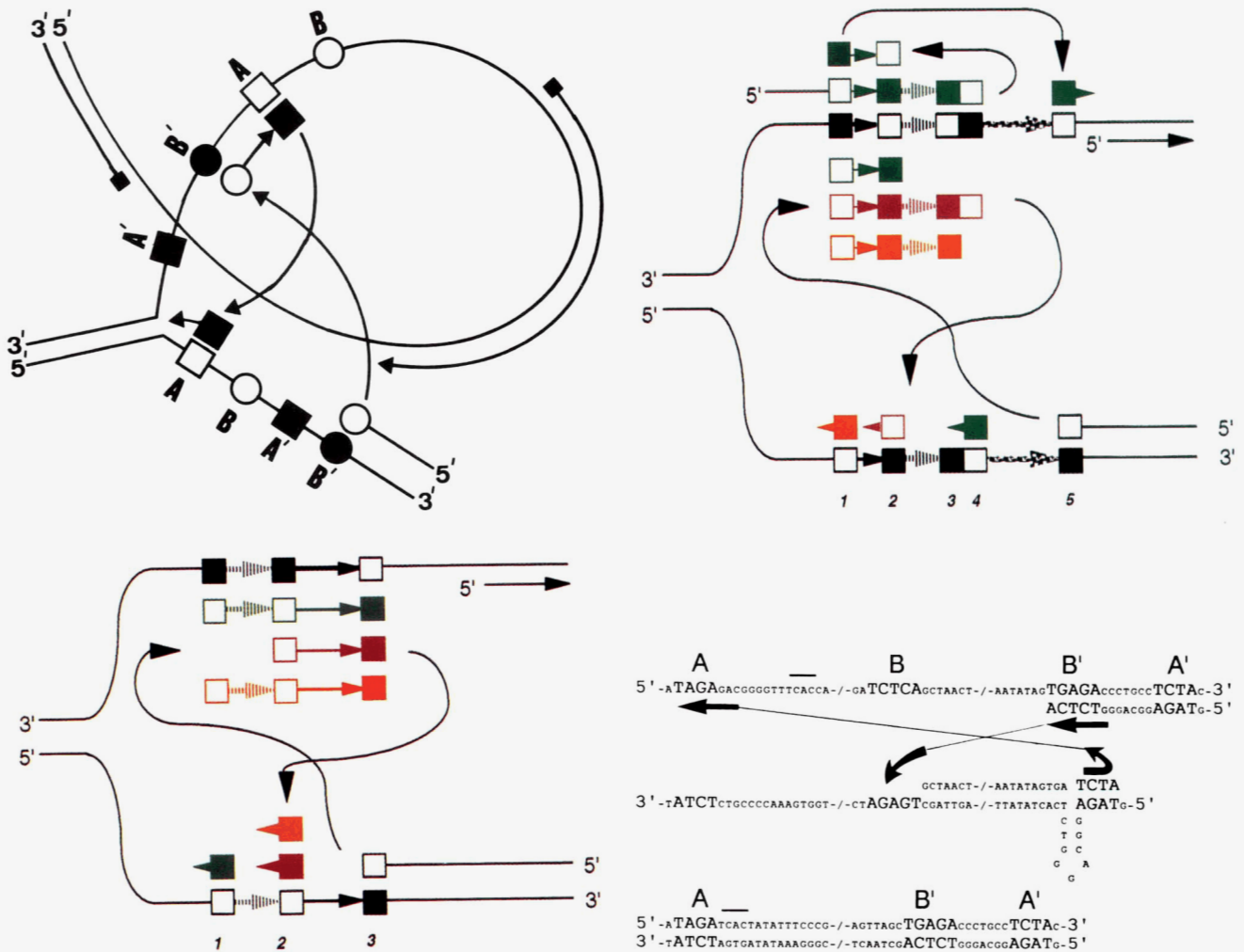


FIGURE 2.—A strand misalignment-realignment model for inversion events facilitated by concurrent DNA replication. (A) Leading-strand misalignment from B' on the leading-strand template to B' on the lagging-strand template followed by chain growth and ultimately realignment from A on the lagging-strand template to A on the leading-strand template and continued synthesis would generate a Class C event (inversion of the region bounded by B and A' and deletion of A to B , and of A' to B' , reading from the leading-strand template). To model the other inversion classes we will unwind the looped DNA replication fork (B-D). (B) The Class D events can be modeled as leading-strand replication proceeding along the leading-strand template from the bottom right until repeat 5, misaligning upon the lagging-strand template via repeat 1, continuing misaligned replication until repeat 2 (green), repeat 4 (red) or repeat 3 (yellow), and ultimately realigning onto the leading-strand template via repeat 4 (green), repeat 2 (red) or repeat 1 (yellow). Such misalignments and realignments would produce all five inversion classes of ROTHSTEIN *et al.* (1987), each exhibiting inversion, deletion and duplication features. The green mutant configuration (classes I and II) can additionally be modeled as an intrastrand event as depicted above the lagging-strand template: here, lagging-strand replication proceeding along the lagging strand template from the top left until repeat 4, folding back upon itself and misaligning upon repeat 2, continuing misaligned replication until repeat 1, and ultimately realigning upon the lagging-strand template via repeat 5 would generate the replacement of one region (4 to 5) by an inverted duplication of another noncontiguous region (1 to 2). The Rad52 dependence of such events in yeast is compatible with a replication-based mechanism. Indeed, the strand switching as invoked in our model here to occur during concurrent DNA replication must also occur in the model of ROTHSTEIN *et al.* (1987) by one of the 3' ends generated from the gap formed after a double-strand break which leads to inverted gene conversion. (C) The Class A events can be similarly modelled: leading-strand replication proceeding along the leading-strand template from the bottom right until repeat 3, misaligning upon the lagging strand template via repeat 1 (green and yellow) or repeat 2 (red), continuing misaligned replication until repeat 3 (green, red and yellow), and ultimately realigning onto the leading-strand template via repeat 1 (green) or repeat 2 (red and yellow) would produce the mutant structures in Figure 1. (D) Junction sequences involved in disruption of the human *GP111a* gene (LI and BRAY 1993). After replicating B' (5'-TGAGA-3'; repeats in large letters), the nascent leading strand (taking the top strand to be the leading-strand template) misaligns onto the lagging-strand template at B' (5'-TGAGA-3'), continues growing until it replicates A (5'-TAGA-3') [invoking a small hairpin so that A' (5'-TCTA-3') forms without incorporating the preceding nine bases into the nascent strand; such secondary structures may preferentially form in the lagging strand during replication (TRINH and SINDEN 1991)], and then realigns onto the leading strand template at A (5'-TAGA-3'). This can explain the mutation (shown below) which includes two deletions (A to B including the loss of 9 bp and 1 kb) separated by four bases of original sequence (overlined), followed by a 15-kb inversion (B to B'). The original proposal (LI and BRAY 1993) involves two independent steps and cannot easily account for the observed 5' junction.

al. 1991). However, such events can also be modeled by an intrastrand misalignment mechanism (RIPLEY 1991), in which the growing strand folds back upon itself, resumes synthesis using itself as a template, and then snaps back onto the correct template and continues. Although a subset of class D events (ROTHSTEIN *et al.* 1987) involves replacement of one region by an inverted duplication of another, noncontiguous region and can be modeled through intrastrand as well as interstrand misalignments (Figure 2B, green events), none of the other events described here can. By using itself as the template, such fold-backs cannot generate simple inversion or inversion associated with a one-sided or two-sided deletion of sequence or to inversion that includes only partial duplication of sequence.

Support was provided by a Human Frontiers Science Program Long-Term Fellowship (A.J.E.G.) and a French CNRS post-rouge Fellowship (J.A.H.). We thank the reviewers for their helpful comments.

LITERATURE CITED

- LAKICH, D., H. H. KAZAZIAN, S. E. ANTONARAKIS and J. GITSCHIER, 1993 Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat. Genet.* **5**: 236–241.
- LI, L., and P. F. BRAY, 1993 Homologous recombination among three intragene Alu sequences causes an inversion-deletion resulting in the hereditary bleeding disorder Glanzmann thrombasthenia. *Am. J. Hum. Genet.* **53**: 140–149.
- MO, J.-Y., H. MAKI and M. SEKIGUCHI, 1991 Mutational specificity of the *dnaE173* mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. *J. Mol. Biol.* **222**: 925–936.
- RIPLEY, L. S., 1991 Frameshift mutation: determinants of specificity. *Annu. Rev. Genet.* **24**: 189–213.
- ROTHSTEIN, R., C. HELMS and N. ROSENBERG, 1987 Concerted deletions and inversions are caused by mitotic recombination between delta sequences in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **7**: 1198–1207.
- SOMMER, S. S., and R. P. KETTERLING, 1993 A postulated mechanism for deletions with inversions. *Am. J. Hum. Genet.* **52**: 1016–1018.
- TRINH, T. Q., and R. R. SINDEN, 1991 Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352**: 544–547.
- WAGA, S., and B. STILLMAN, 1994 Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. *Nature* **369**: 207–212.

Communicating editor: P. L. FOSTER