Chromosome replication in eukaryotes has been an issue of fascination for more than a century. The remarkable lengths of the DNA molecules are part of their fascination. Each chromatid of the metaphase chromosome contains a single, very long molecule of DNA, more than a centimeter in length in humans. This long DNA molecule is confined within a tight space in the interphase nucleus, and yet, replication successfully avoids any states of permanent entanglement. Observation and experiment within the dense chromatin material are difficult, and thus many issues regarding the topological separation of these large DNA molecules in the cell nucleus remain unresolved. In this paper we present a general line of topological reasoning relating to the mechanism of chromosome reproduction, informed by recent experimental results, but going beyond what has been strictly proven. This reasoning leads to the proposal of a mechanism that greatly reduces the probabilities of DNA entanglement during replication.

The basic assumptions and their consequences

We assume two solid principles: (1) the double-stranded constitution of DNA, with template-directed reproduction; and (2) the existence of multiple bidirectional replication initiation centers located at irregular intervals along the double helix, each being the initiation point of a pair of diverging replication forks, operating under the coordinated control of the cell cycle. More speculatively, we also adopt a third hypothesis originally proposed by Sundin and Varshavsky7 that (3) each pair of replication forks that initiate at the same origin remains physically attached throughout the reproductive process (the connected-fork model).

The consequences of these three simple assumptions are so striking that we feel they should be presented even in the absence of conclusive experimental evidence that assumption (3) is correct. Only with all three of these assumptions operating can one construct a reproductive topology that assures the continuous integrity of each parental DNA template strand throughout the norma

The growing fork

Assumption (1) implies that DNA replication always involves a Y-shaped topology called a growing fork, in which the two original strands are split apart and a new complementary strand is constructed for each, using each original strand as a template. The structural and mechanistic details of the fork may well differ in various phyla. However, one feature that appears to be consistent is that known DNA polymerases move towards the 5’ end of the template strand, which means that synthesis goes smoothly and continuously on one parent strand (the leading strand) and must move in a discontinuous, back-and-forth manner on the other strand (the lagging strand).

The polymerase complex at the growing fork is structurally complex. The lagging strand synthesis involves an RNA primase, at least one DNA polymerase, an RNAse H that recycles the RNA primer material, and a DNA ligase which knits together the short (2000 bp) sections of discontinuously synthesized DNA. The polymerase complexes are assisted in propagating along the DNA duplex by a helicase that unwinds the DNA duplex, and by ‘sliding clamps’3–5, which hold each DNA polymerase close to the DNA as it moves along the polymer.

The issues of DNA topology within the fork have received valuable consideration by others6. These include important unresolved issues such as the degree of torsional tension created in the DNA as replication proceeds, and the distribution of that torsional tension during the replication process. This paper, however, focuses on entanglement, not torsional tension. In this context, it suffices for us to accept that the fork exists and somehow functions efficiently. Nothing in the known details of fork structure is inconsistent with the topological ideas which we set forth below.

The origin recognition complex

Assumptions (1) and (2) imply the existence of the replication fork. Assumption (2) also introduces a firmly established fact: the existence of multiple replication origins within the eukaryotic chromosome. At each origin,
two replication forks (oppositely oriented, as shown in Fig. 1a) are generated at a definite site of replication initiation, defined by the binding of an origin recognition complex (ORC). The ORC concept, appropriately elaborated\[12\], seems to hold the key to the once-only initiation of replication during each division cycle. The basic idea is that replication involves far more than a simple polymerase attachment to DNA. It requires the painstaking assembly of a complex ORC at the initiation site, with different parts appearing at different times in the cell cycle, culminating in a one-time triggering of S (synthesis) phase.

We accept the fork and the ORC concepts as sketched above, and pursue the consequences for topological entanglement of the two nascent duplex daughter strands, and their ultimate effect on mitosis. Figure 1c is a graphical representation of replication in which the fork pairs move away from each other in physical space as well as on the DNA contour.

How does entanglement occur and how might it be prevented?

In Fig. 1b, one fork has rotated with respect to the other by one full turn. This might happen from one of two causes. First, the parental DNA might remain stationary. If so, a simple fork model requires the forks themselves to make one turn for approximately every 10 bp processed. If one uses a more complex fork model that allows parental DNA rotation about its helical axis, assisted by topoisomerase activities, then independent fork pairs might still rotate diffusively with respect to each other, creating entanglements. Once entanglement has occurred, independent of mechanism, Fig. 1c shows the consequences of just one relative rotation when mitosis time arrives. The two daughter strands are topologically entwined and cannot separate without a pass-through.

In a classic paper, Sundin and Varshavsky\[9\] make the following statement: "If, during replication, the two growing forks were to rotate with respect to each other, entanglement of daughter chromatin strands could occur. To avoid this difficulty, we suggest that both replication forks are tied together and uniquely oriented within a binary complex (the connected-fork model)." In addition, Wessel, Schweizer and Stahl\[6\] have reported experimental evidence for such an attachment in SV40 bi-directional replication. No theoretical work can prove the correctness of such an hypothesis, but it can be used to follow out its consequences. In this work we show that if one adopts the connected-fork hypothesis of Sundin and Varshavsky, one not only finds a simple, unentangled separation of chromatin strands at mitosis, but one also finds that linear chromosome condensation is a natural consequence of replication.

The connected-fork model and other mechanisms of entanglement avoidance

We are well aware that much experimental work suggests mechanisms other than the connected-fork model; these have been largely in the category of knot-untangling mechanisms that would undo the consequences of independently swirling fork pairs (Class II topoisomerase activities\[13\]). However, we remain distinctly uneasy with the idea that there must be one duplex pass-through for every ten base pairs replicated. The integrity of the information in DNA is best guaranteed by a mechanism that does not normally break the DNA duplex once it is formed. Every exception increases the danger of mutation, and while mutation is a normal and necessary part of life, it must still be rare. Replication is above all very fast and very accurate, and it is important to show the existence of a simple and transparent mechanism that could maintain the integrity of the DNA parent strands in their normal operation.

A second source of our unease with the concept of untangling mechanisms being utilized exclusively is the following consideration: in a complex knot extending over a volume too large for structure recognition, the random passage of one duplex through another is just as likely to increase the knotting as to decrease it\[14\]. Disentanglement requires a global knowledge of the topology of the knot, but enzyme systems must sense and act within a very local volume, limited by the short...
range of intermolecular forces. A solution to this Gordian disentanglement problem has been suggested, in which the disentanglement enzyme, a topoisomerase II/IV, attaches to two separate binding sites on the duplex DNA and migrates along the contour, collecting catenations into a locally recognizable structure. Such a mechanism is most easily applied to circular DNAs. It might be applied between diverging growing forks, but again, this would require the physical attachment of the forks, one to the other, at least at or near the end of the fork migration. An alternative is that this activity is used between converged, nearly touching growing forks, providing the context in which the Class II topoisomerase acts in a local domain of DNA contour. There is experimental evidence for the importance of topo II activity at this point.

Francis Crick has written that 'Evolution is not a clean designer… It is opportunistic. If a new device works, in however odd a manner, evolution will promote it, [and] the final design may be a rather messy accumulation of interacting gadgets.' The connected-fork model and the occasional pass-through model are both simple gadgets, and it is likely that evolution has made use of them both. We emphasize that our work is intended to reintroduce an alternative rather than to imply the rejection of other models. We demonstrate here the topological consequences of the connected-fork hypothesis.

Chromosome replication by a connected-fork mechanism

The essential features of the connected-fork model are shown in Fig. 2. The protein precursors of the connected fork are bound to DNA at the origins of replication, forming an ORC. But the ORC is incomplete in some respects and is therefore incapable of forming a fork pair. As S (synthesis) phase begins, a signal (symbolically depicted here by the doubling of the protein content at the origin – not implied that this is literally true) 'licenses' the ORC to unwind the DNA duplex locally, creating two small single-stranded regions in each strand of the DNA. This enables replication to initiate on each loop, generating back-to-back DNA replication forks, as in Fig. 2b. Replication now proceeds by drawing double-strand DNA into the connected replication fork complex from both ends, and by extruding loops of nascent double-strand from opposite sides of the complex, as in Fig. 2c.

Figure 3 is a representation of the replication of a very small chromosome with simultaneous ORC initiation. It begins in an extended state with ORCs in place, but incomplete (Fig. 3a). In Fig. 3b, the ORCs have been completed and the initiation signal has been received, and the connected-fork pairs are created, inevitably back-to-back. As replication proceeds in Figs 3c and 3d, the nascent loops grow and the parental duplex sections between the converging replication fork complexes shrink until finally all that is left is an extended row of connected-fork replication complexes (Fig. 3e), each with two long unknotted loops of nascent duplex DNA attached to their sides.

The length contraction seen in this figure could contribute to chromosome condensation prior to mitosis. From experiment, it is known that other factors contribute to chromosome condensation. We propose that the linear condensation suggested by the connected-fork model is one of several mechanisms contributing to chromosome condensation. This mechanism also suggests that chromosome banding patterns may reflect the linear frequency of origins of replication in that region of the chromosome. In the final frame of Fig. 3, the replication complexes split up again, with half the protein content assigned to each double strand (again this is symbolic and has not been experimentally established). At this point mitosis can begin, inherently free of any entanglement. A detail of the fully collided configuration in Fig. 3e needs discussion. A small length of DNA in direct contact with the collided fork complexes has not completed replication when collision first occurs, and it must be completed by a special replication mechanism inherent to these collided structures. This 'end game' in DNA replication has been studied in some detail using SV40 as a model, and, if left unperturbed, is efficient at avoiding entanglement (2). This process is reported to require a Class II topoisomerase activity in these locally confined regions.

Abstracting chromosome replication

We have generated a computer model of the connected-fork chromosome replication process discussed in.
above, using Mathematica. The output is represented in
Figs 1, 2, and 5. The details of this process will be pre-
presented elsewhere (J.E. Hearst, L. Kauffman, W.M. McClain
and Y. Shi, unpublished). A full movie version of chromo-
some reproduction has been converted to JAVA script and is
berkeley.edu/~jehgrp/symbiologic/DNA_Replication.html

The abstract chromosome model is susceptible to
numerous experimental evaluations. For example deter-
mination of the segregation pattern of the protein sub-
units in the replication complexes or alternatively in the
origin recognition complexes following cell division
would establish which subunits remained permanently
in location on the DNA and which subunits bind and
release during the cell cycle. The model itself suggests
that the results of a DNA location-specific transfer
experiment establishing how the permanently bound
protein subunits segregate following cell division would
reveal molecular details about the connected fork mecha-
nism. The subunit segregation pattern used in our fig-
ures is arbitrary. The pattern must be determined by
experiment. It would also be of interest to compare the
frequencies of ORCs on the DNA associated with dense
and clear bands in the condensed chromosomes.

Acknowledgements

This study was supported in part (J.E.H.) by National
Institutes of Health Grant ROI GM 41911 and by the
Office of Energy Research, Office of Health and
Institutes of Health Grant AI22636.

References

Acad. Sci. U. S. A. 71, 135–139


Biol. Chem. 258, 4087–4094

4. Stukenberg, P.T., Studwill-Vaughan, P.S. and


5, 981–992


(Kaufman, L., ed.) pp. 1–80, World Scientific

16. Rybenkov, V.V., Ullsperger, C. Vologodskii, A. and

Charles Scribner’s Sons

J.E. Hearst is in the Department of Chemistry, University
of California, Berkeley, CA 94720–1460, USA, and the
Division of Structural Biology, Lawrence Berkeley
National Laboratory, USA.

L. Kauffman is in the Department of Mathematics,
University of Illinois, Chicago, IL, USA.

W.M. McClain is in the Department of Chemistry, Wayne
State University, Detroit, MI, USA.

NEW PUBLIC DATABASE OF EXPRESSED HUMAN GENOME

The National Center for Genome Resources (NCGR) and the South African National Bioinformatics Institute (SANBI) recently announced
the launch of the Sequence Tag Alignment and Consensus Knowledgebase, or STACK, a public database of gene sequences expressed

Only a small fraction of the more than 50 000 human genes have been completely sequenced, and the majority of existing gene
data are available primarily as gene fragments. STACK provides an independent method for processing the gene fragments, detecting
errors and creating joined sets of consensus sequences for each gene sequence.

STACK also features expressed gene sequences organized according to tissue type and provides a comprehensive representation of
each gene with alignments of its expressed fragments. STACK can potentially be used to distinguish between the alternative products
of a given gene.

Contacts

Marjan P. Skupski
marjan@ncgr.org

Winston Hide
wminhide@sanbi.ac.za

Contributed by Janine Sieja Hagerman

Contributed by Janine Sieja Hagerman

1800 Old Pecos Trail, Suite A, Santa Fe, NM 87505, USA.

247