

Comparing DNA Damage-Processing Pathways by Computer Analysis of Chromosome Painting Data

DAN LEVY,¹ MARIEL VAZQUEZ,¹ MICHAEL CORNFORTH,³ BRADFORD LOUCAS,³
RAINER K. SACHS,¹ and JAVIER ARSUAGA^{1,2}

ABSTRACT

Chromosome aberrations are large-scale illegitimate rearrangements of the genome. They are indicative of DNA damage and informative about damage processing pathways. Despite extensive investigations over many years, the mechanisms underlying aberration formation remain controversial. New experimental assays such as multiplex fluorescent *in situ* hybridization (mFISH) allow combinatorial “painting” of chromosomes and are promising for elucidating aberration formation mechanisms. Recently observed mFISH aberration patterns are so complex that computer and graph-theoretical methods are needed for their full analysis. An important part of the analysis is decomposing a chromosome rearrangement process into “cycles.” A cycle of order n , characterized formally by the cyclic graph with $2n$ vertices, indicates that n chromatin breaks take part in a single irreducible reaction. We here describe algorithms for computing cycle structures from experimentally observed or computer-simulated mFISH aberration patterns. We show that analyzing cycles quantitatively can distinguish between different aberration formation mechanisms. In particular, we show that homology-based mechanisms do not generate the large number of complex aberrations, involving higher-order cycles, observed in irradiated human lymphocytes.

Key words: karyotype, chromosome aberration, repair/misrepair pathway, cyclic graphs, radiation damage.

1. INTRODUCTION

CHROMOSOME ABERRATIONS ARE ILLEGITIMATE REARRANGEMENTS of the genome involving large (>1 Mb) DNA segments and occurring during the early part of the cell cycle. Such large-scale structural changes are frequently associated with genetic diseases (Lee *et al.*, 2000), chromosomal instability (Limoli *et al.*, 2000), or cancer (Mitelman *et al.*, 2002). Chromosome aberrations in a cell alter its *karyotype*, which is characterized by the number of chromosomes and the large-scale structure of each chromosome (Fig. 1, bottom).

¹Mathematics Department, University of California at Berkeley, Berkeley, CA 94720.

²Molecular and Cell Biology Department, University of California at Berkeley, Berkeley, CA 94720.

³Department of Radiation Oncology, The University of Texas Medical Branch, Galveston, TX 77555.

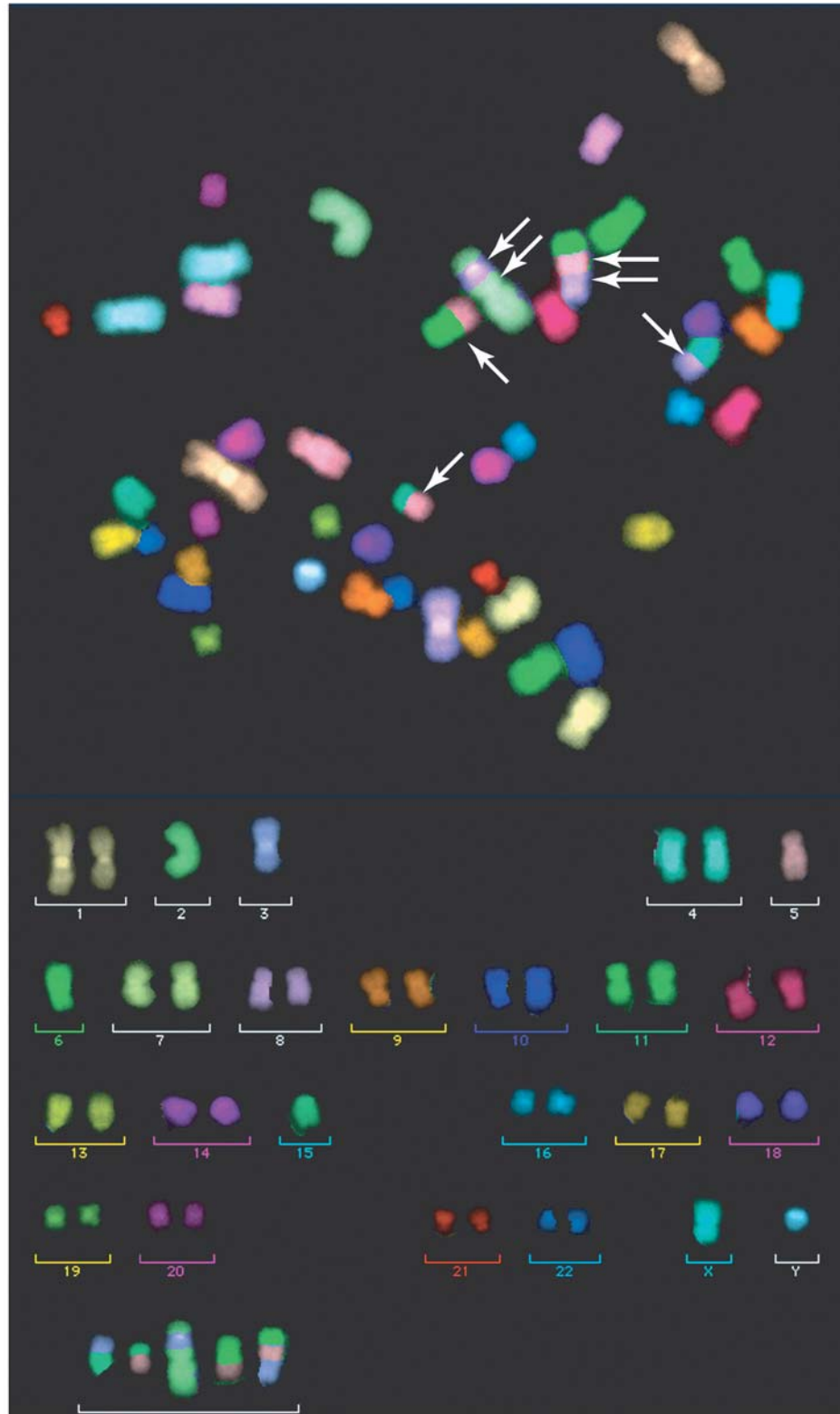


FIG. 1. mFISH patterns for damaged genomes. **Top.** When DNA DSBs are erroneously processed during the G0/G1 phase of the cell cycle, chromosome aberrations are introduced. Many of these aberrations (white arrows) can be detected at the next subsequent metaphase by mFISH. Sister chromatids, attached near the center but visually resolvable in some cases here, are “painted” the same color, which is also used for the homologous chromosomes. **Bottom.** The karyotype of the top figure.

Many karyotype changes can be detected by multiplex fluorescence *in situ* hybridization (mFISH, Speicher *et al.*, 1996) or spectral karyotyping (SKY, Schröck *et al.*, 1996). SKY and mFISH are assays where heterologous chromosomes are “painted” different colors. These multicolor protocols extend FISH (Bauchinger *et al.*, 1993) and complement other classical karyotyping methods (reviewed by Wang [2002] and Savage [1999]). Both mFISH and SKY have been highly informative in the classification of complex chromosome aberrations (e.g., Schröck *et al.*, 2000, Anderson *et al.*, 2003; Savage, 2002); in particular, mFISH has been successfully used in recent analyses of irradiated human cells (Anderson *et al.*, 2002; Durante *et al.*, 2002; Greulich *et al.*, 2000; Loucas *et al.*, 2001; Hlatky *et al.*, 2002), in clinical studies concerning different types of leukemias, lymphomas, and some congenital disorders (reviewed by Lee *et al.* [2001]) and in studies of interphase nuclear organization (Cornforth *et al.*, 2002).

Aberrations result from misrepair of DNA *double strand breaks* (DSBs), where both DNA sugar-phosphate backbones of a double helix are broken at nearby sites. The subsequent DNA repair/misrepair pathways, whose products are chromosome aberrations (Fig. 1, top), are so complicated that biologically based computer models of chromosome aberration production (Sachs *et al.*, 2000a; Edwards, 2002; Holley *et al.*, 2002; Ballarini *et al.*, 2002) and mathematical descriptions of aberration patterns detected by mFISH are called for. In these studies, the concept of a *cycle*, related to concepts familiar from comparative genomics (Bafna *et al.* [1996], for a review, see Pevzner [2000] or Waterman [1995]), plays a key role. An *n*-cycle describes a situation in which *n* different DSBs interact in one single misrejoining reaction that cannot be decomposed into simpler reactions involving disjoint subsets of DSBs.

Here, we present the mathematical foundations of cycles, their relationship with aberration multigraphs (Sachs *et al.*, 2002), computer algorithms to analyze cycle structures, and the application of cycles to the study of DNA damage processing in human lymphocyte cells. First, we give an introduction to the concepts and terminology used in analyzing chromosome aberrations and describe through examples concepts such as obligate cycle structure. Second, we present a graph-theoretical formulation of cycles and discuss their properties. Third, we describe algorithms to compute the obligate cycle structure of an observed mFISH aberration pattern. Next, we apply these algorithms to simulated mFISH aberration patterns, computed with a previously developed program based on biophysical models of aberration production (Sachs *et al.*, 2000a; Chen *et al.*, 1998). We then compare theoretical cycle distributions with experimentally observed ones (Loucas *et al.*, 2001) and draw conclusions about DNA damage processing pathways in human lymphocytes. For comparison, we reanalyze the same experiments without considering cycle structures.

From this study, we conclude that cycles can distinguish between different mechanisms of chromosome aberration production during the G0/G1 phase of the cell cycle and that a breakage-and-reunion pathway, presumably based on a nonhomologous end-joining mechanism (reviewed by Ferguson *et al.* [2001]), dominates homology-based mechanisms of aberration formation in the experiments considered.

2. BACKGROUND ON CYTOGENETICS AND CHROMOSOME ABERRATIONS

2.1. Chromosome structure during the cell cycle

Inside the nucleus of a typical somatic human cell shortly after a cell division (i.e., during the G0/G1 phase of the cell cycle), DNA is organized in 22 pairs of homologous chromosomes plus two sex chromosomes (XX for female and XY for male). Each of the 46 chromosomes is composed of a DNA double helix and proteins. At each end of the chromosome is a “cap,” the *telomere*, which usually prevents the chromosome end from interacting with other chromosome ends or with broken chromosomes. Somewhere in the chromosome, often near the middle, is the *centromere*, a region used by the cell for the segregation of chromosomes at the time of cell division. The chromosomes appear as long multiply-folded filaments that are mainly localized in regions of the nucleus called *chromosome territories* (reviewed by Cremer *et al.* [2001] and by Holley *et al.* [2002]).

During the next phase (i.e., after G0/G1), called the *S-phase*, chromosomes replicate. The *G2 phase* follows, where pairs of identical chromosomes present after replication remain attached, as *sister chromatids*. Finally, during *mitosis*, cells divide into two daughter cells, each carrying a whole set of chromosomes. At *metaphase*, a part of mitosis, chromosomes condense and can be seen with the two sister chromatids attached to each other near their centromeres by proteins, so that there are four *arms*, each capped at its end by a telomere (Fig. 1).

Here, we will study DNA damage and damage processing that occur during the G0/G1 phase of the cell cycle, with the damage identified by the observed mFISH pattern at the first subsequent metaphase (Fig. 1). An *observed mFISH pattern for karyotype changes* refers to properties, such as color junctions, extra or missing centromeres, etc., that occur in rearranged chromosomes as a consequence of DNA damage and damage processing (Fig. 1, top and Fig. 1, last row); the rest of the karyotype is the same as in the undamaged cell (Fig. 1, bottom).

In general, an observed mFISH pattern for karyotype changes may contain several different, essentially independent chromosome aberrations, involving disjoint sets of chromosomes. In Sections 2–4, we henceforth assume, for brevity, that there is just one chromosome aberration, i.e., that all the chromosomes involved in changes are connected, directly or indirectly, by misjoinings. In this case, we use the term *observed mFISH aberration pattern* to describe the karyotype alterations seen at metaphase.

2.2. DSB misrejoining and rearranged chromosomes

A number of external factors (e.g., radiation, certain chemicals) as well as endogenous factors (e.g., free radicals from metabolic reactions, incomplete recombination events) can induce DSBs. Each DSB has two free ends (for example, the DSB ends numbered 1 and 2 in Fig. 2A belong to the same DSB). The number and distribution of DSBs along chromosomes determine the *initial configuration*. Rejoining of the ends is important for the integrity of the genome. If a given DSB end rejoins with its original partner, we say that the DSB has *restituted* (even if there are local alterations in the DNA on scales $\ll 1$ Mb). However, in a situation where several DSBs (in one or more chromosomes) occur almost simultaneously and close to each other in space, a DSB end can *misrejoin* with another DSB end, different from its original partner (Fig. 2). The collection of all misrejoining reactions is the *exchange process* whose products are *rearranged chromosomes* (Figs. 2B and 2C). The set of all rearranged chromosomes defines the *final configuration*. If all DSB ends interact with other DSB ends, such that no end is left without a partner after the process terminates, the *restitution/exchange process* is called *complete*. A rearranged chromosome produced by a complete restitution/exchange process either contains a telomere at each end or forms a ring, since no DSB end is left without a partner. Rearranged chromosomes obtained as the product of complete restitution/exchange processes and observed with mFISH are the focus of our study.

2.3. Cycle structures of exchange processes: Examples

A complete exchange process consists of one or more irreducible complete exchange processes and has a corresponding *cycle structure* that is determined by its irreducible processes (Sachs *et al.*, 1999; Cornforth, 2001), as we will now illustrate with some examples (Figs. 2–4). The complete exchange process $A \rightarrow B$ in Fig. 2 includes a smaller complete exchange process where 1 joins with 6 and 2 with 5 (the other possible complete exchange process for these two DSBs, 1 with 5 and 2 with 6, is not shown). This smaller complete exchange process is *irreducible* in the sense that it cannot be further decomposed into still smaller complete exchange processes. If the situation for Fig. 2A \rightarrow 2B is redrawn as in Fig. 3A, ignoring the chromosomes and focussing attention on the exchange process itself, cyclic patterns are uncovered. In Fig. 3A, end 1 is shown as the vertex labeled 1. It has an edge, labeled “i” (for “initial”), going to its original partner 2, and has an edge going to its final partner 6, labeled “f” (for “final”). Filling in the other entries, we get a square, which we call a *2-cycle* or *cycle of order 2* because there are 2 DSBs involved (giving 4 DSB ends). Such a 2-cycle is denoted by c_2 in the biology literature. The exchange process $2A \rightarrow 2B$ involves another 2-cycle, also shown in Fig. 3A. Analogously, a cycle of order 3 (denoted by c_3 in the biology literature) results when 3 DSBs undergo a complete exchange process without any restitutions. When 4 DSBs take part in a complete exchange process, the process may be irreducible, involving all 4 DSBs interacting in a “musical chairs” type arrangement (e.g., $A \rightarrow C$ in Fig. 2) forming a 4-cycle c_4 (Fig. 3B, where we have omitted the numbers 1–8 and the labels “i” or “f”). Or the cycle structure may involve two different irreducible complete exchange processes and have *cycle structure* c_2+c_2 as shown in the complete exchange process $A \rightarrow B$ in Fig. 2 and in Fig. 3A.

2.4. Obligate cycle structures: An example

In general, given an observed mFISH aberration pattern, one may be able to find a number of different complete exchange processes that could have generated the pattern. An example is given in Fig. 4. The

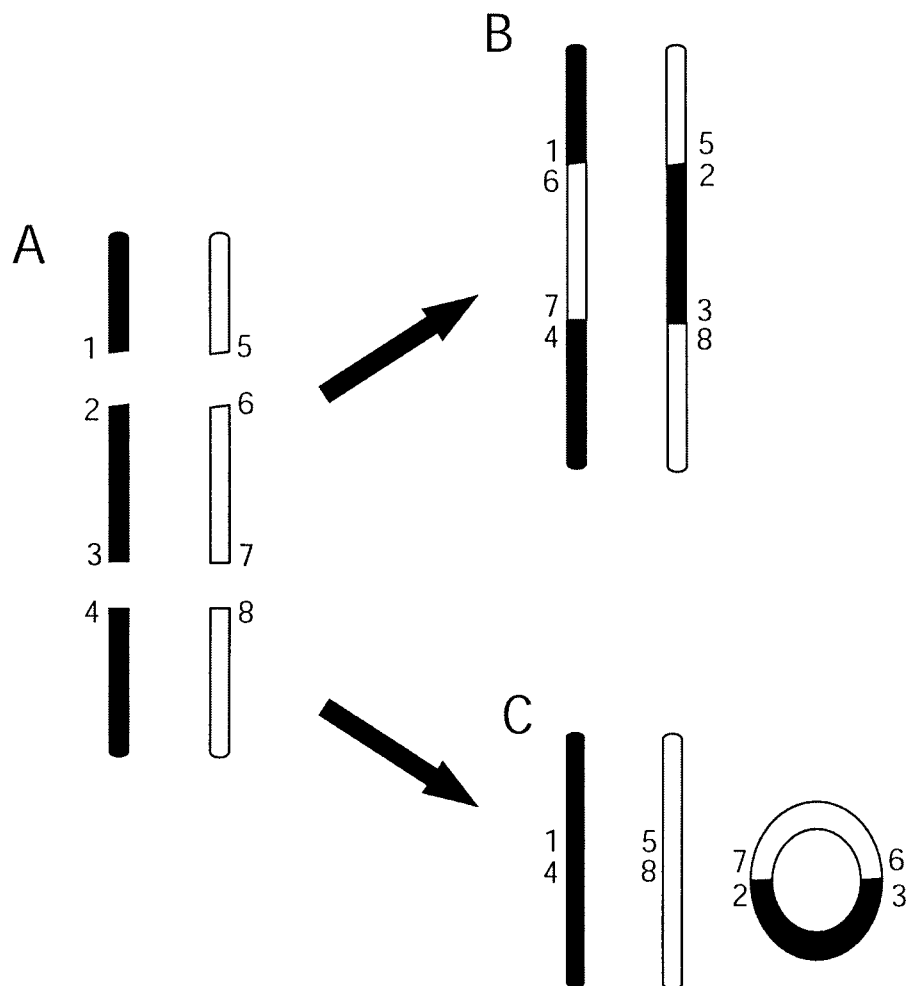


FIG. 2. Examples of misjoinings. **Panel A** shows two different chromosomes, one solid black, the other white. Each chromosome has two initial DSBs represented by gaps. Free ends are labeled with numbers. The exchange processes (i.e., the set of misjoining reactions) from A to B is decomposable into two 2-cycles (Fig. 3A), and the exchange process for A to C forms one 4-cycle (Fig. 3B). **Panels B and C** show the final configurations resulting from the exchange processes.

observed mFISH aberration pattern shown in Fig. 4A has two possible exchange processes that can generate it (Figs. 4B and 4D). The ambiguity implies that more than one cycle structure can be associated to this observed mFISH aberration pattern, in one case $c2+c3$ (Fig. 3C) and in the other $c5$ (Fig. 3D). The reader should check that in fact Fig. 4 gives rise to the cycles shown in Figs. 3C and 3D. In this situation, $c5$ may seem like an unnecessarily complicated way of interpreting the observed mFISH aberration pattern, so we emphasize the exchange process whose cycle structure is $c2+c3$ (Cornforth, 2001) and call $c2+c3$ the *obligate cycle structure* of the observed mFISH aberration pattern.

3. MATHEMATICAL CHARACTERIZATION OF CYCLES

We now show that the pattern suggested by Fig. 3A–D is general. We consider any complete exchange process, generalizing Figs. 2 and 3A–D. Specifically, suppose there are some DSBs. Suppose each DSB end is misrejoined with an end different from its original partner (the more general case, where some free ends may rejoin with their original partners, corresponding to restitution, is discussed below). We show how to define the cycle structure of the complete exchange process. The reader familiar with graph theory

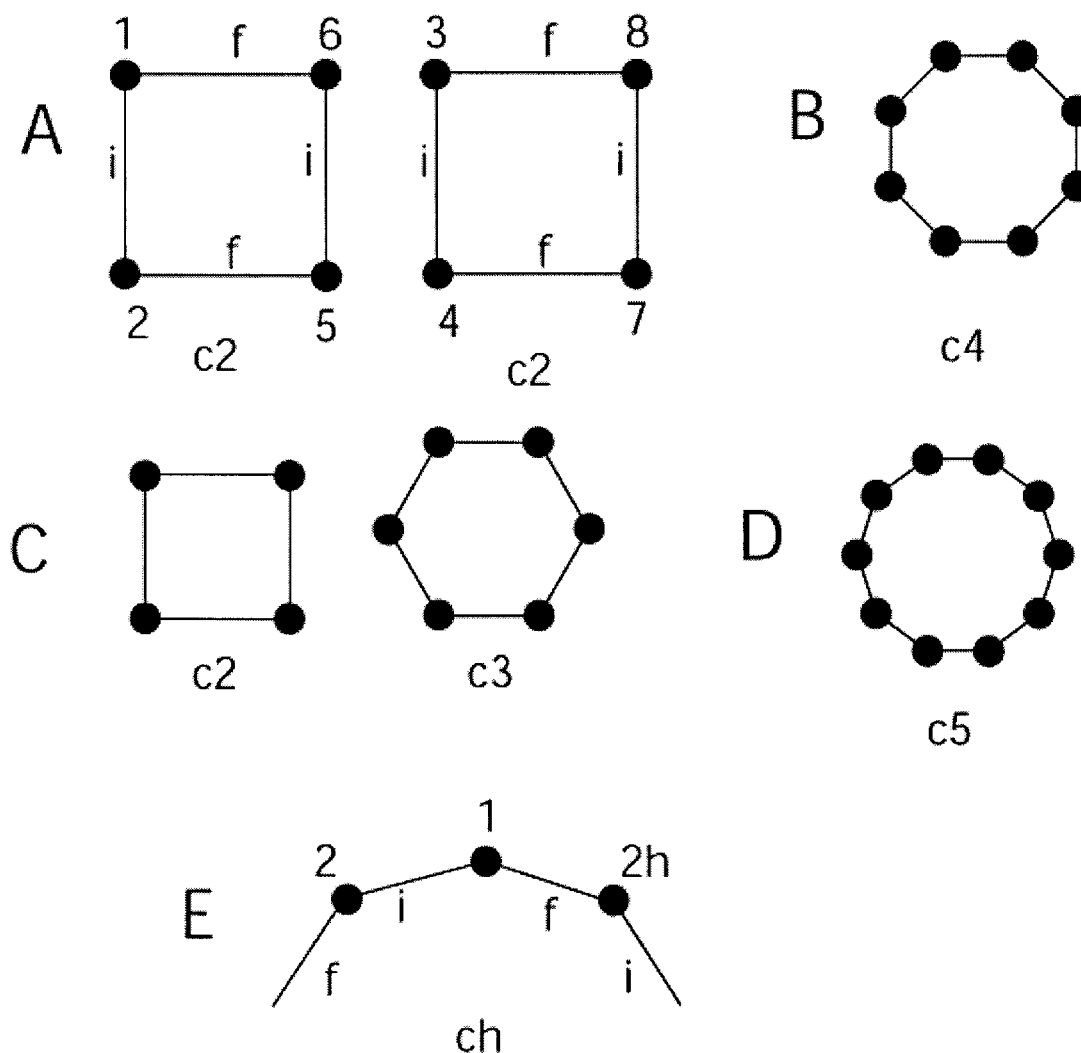


FIG. 3. **Figure 3A** shows the cyclic graphs associated to the interaction of DSB ends shown in Fig. 2 (A \rightarrow B) whose cycle decomposition is c_2+c_2 . **Figure 3B** shows the corresponding cyclic graph for a c_4 and represents the misrejoining reaction shown in figure 2(A \rightarrow C). **Figures 3C and D** represent the two possible cyclic graphs associated to the case when 5 DSBs have occurred without restitution (Fig. 4). **Figure 3E** shows the generalized picture of the previous examples (Fig. 3A–D).

will note that our next paragraph is just an informal proof of the well-known fact that for a 2-regular graph that is properly edge colorable with two colors, each connected component consists of a cyclic graph with an even number of vertices.

Consider any complete exchange process F. Each DSB end has one initial partner (from the same DSB) and has one final partner (because of completeness), with the final partner different from the initial partner (because we are considering only exchange processes, rather than restitution/exchange processes). Generalizing Fig. 3A, form a graph H by representing the free ends as vertices, using edges labeled “i” for initial partners and using edges labeled “f” for final partners (Fig. 3E). Graph H is called the graph of the complete exchange process. We now show that H consists of cycles. Starting with any vertex (say, 1 in Fig. 3E) choose either the “i” edge or the “f” edge—without essential loss of generality we may suppose it is the “i” edge. This edge takes us to a different vertex (which we label 2), and we can then leave vertex 2 via its other edge, “f” in this particular case. Iterating, the process can terminate only when we reach a vertex which already has one and only one edge used, and thus that terminating vertex must be vertex 1. Thus, we get a cyclic graph (Fig. 3E). When we arrive back at 1, we must come in along “f” edge, so the

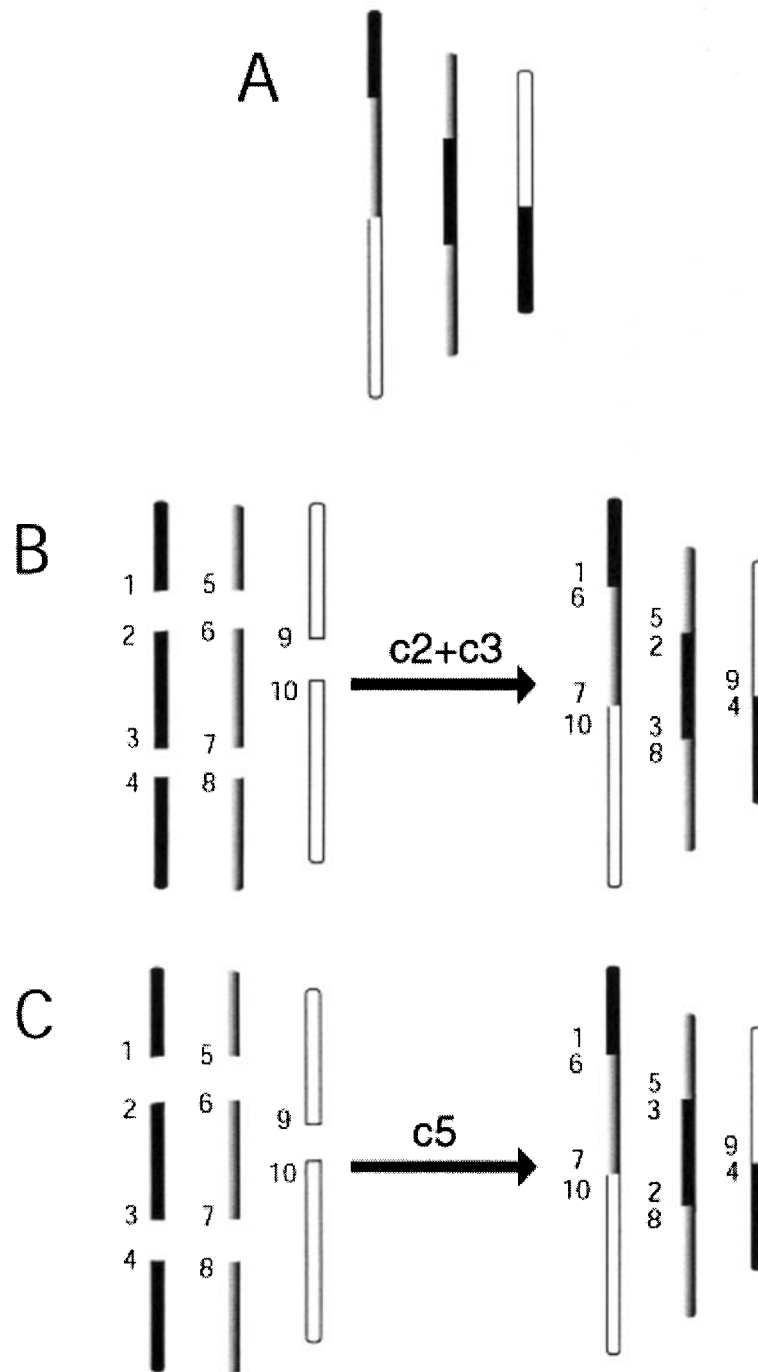


FIG. 4. The obligate cycle decomposition of an mFISH pattern for karyotype changes. Three chromosomes of different colors, with 2, 2, and 1 DSBs, respectively, misrejoin in two different ways (B,C) to produce the same mFISH pattern for karyotype changes (A). The obligate cycle decomposition in this example is $c2+c3$, the cycle decomposition of the misrejoining process shown in B (compare Fig. 3C).

cyclic graph must have an even number of vertices, say, $2h$. For consistency with the biology literature, we designate the cyclic graph as ch . In graph terminology, we may regard ch as the standard cyclic graph C_{2h} . Now, pick a vertex not yet used, if there is one. Proceeding as before, we get another cyclic graph with an even number of vertices. Iterate till all the vertices are used. The result is a decomposition of the complete exchange process into cyclic graphs, each of which has an even number of vertices.

Remarks. The following are asides:

1. We could have allowed restitutions, i.e., considered the more general case of a complete restitution/exchange process. Then, the above discussion would have gone through without any essential change, although we would have been dealing with multigraphs (Hartsfield *et al.*, 1994) rather than graphs, since the possibility that two vertices are connected by two different edges (both “i” and “f”) would have been allowed.
2. It can be shown that given any 2-regular graph H that is properly edge colorable by two colors, there exists a complete exchange process of which H is the graph.
3. Intuitively speaking, the DSBs participating in one cycle are expected to be close in space and time in order for the corresponding irreducible reaction to occur.
4. The concepts of cycle and obligate cycle structure are similar to those introduced earlier in the context of comparative genomics (Bafna *et al.* [1996], reviewed by Pevzner [2000] or Waterman [1995]). In our study, we treat cycles as descriptors of an actual biophysical process: the exchange process. In contrast with previous work in comparative genomics, the interest when analyzing observed mFISH aberration patterns is mainly on the complex cycles ($n > 2$), as will be seen in Section 5.
5. In radiation cytogenetics, the cyclic multigraphs of Fig. 3 are regarded as submultigraphs of an aberration multigraph; the latter arises naturally when adding vertices and edges that correspond to telomeres and chromosome segments, respectively (Sachs *et al.*, 2002).

Let H be the graph of a complete exchange process; generalizing our examples gives the following definition.

Definition 3.1. *The cycle structure of a complete exchange process is the set of connected components of H , $(ch_1, ch_2, \dots, ch_m)$, ordered such that $h_1 \leq h_2 \leq \dots \leq h_m$. To denote the cycle structure we write $ch_m + \dots + ch_1$.*

As we have shown by an example, given an observed mFISH aberration pattern, there may exist more than one complete exchange process that can generate the observed pattern and correspondingly more than one cycle structure associated to the aberration pattern (e.g., Fig. 4). This ambiguity led to the concept of obligate cycle structure (Cornforth, 2001), which we now discuss formally, starting with a definition.

Definition 3.2. *Let F_1 and F_2 be two complete exchange processes that produce the same observed mFISH aberration pattern. Let $(ch_1, ch_2, \dots, ch_m)$ be the cycle structure of F_1 and $(cg_1, cg_2, \dots, cg_k)$ be the cycle structure of F_2 . We will say that $F_1 < F_2$ if: either $h_1 + h_2 + \dots + h_m < g_1 + g_2 + \dots + g_k$, or $h_1 + h_2 + \dots + h_m = g_1 + g_2 + \dots + g_k$ and there exists an index r such that $g_r > h_r$ and $g_i = h_i$ for all $i < r$.*

The set of complete exchange processes associated to an observed mFISH aberration pattern, with the order relation $<$ defined in definition 3.2, has at least one minimum, and all minima must have the same cycle structure, so the following definition makes sense.

Definition 3.3. *The obligate cycle structure $(ch_1, ch_2, \dots, ch_m)$ of an observed mFISH aberration pattern is the cycle structure of any minimum of the complete exchanges that could have generated the mFISH pattern.*

Intuitively speaking, the obligate cycle structure involves the lowest-order cycles allowed for a given mFISH pattern and is therefore in some sense the least complex cycle structure consistent with observation.

4. AN ALGORITHM FOR COMPUTING OBLIGATE CYCLE STRUCTURES

We developed an algorithm for computing obligate cycle structures, to be described next.

4.1. Calculating the minimum number of manifest misrejoinings

From experimental mFISH data, one can determine only the minimum number of chromosomes involved in an aberration and the minimum number of misrejoinings per pair of homologous chromosomes since there can always be cryptic complications (Simpson *et al.*, 1995; Sachs *et al.*, 1999; Cornforth, 2001). To determine these minimum values from the observed mFISH aberration pattern and get information about exchange processes consistent with the pattern, we analyzed the set of rearranged chromosomes as follows.

We assumed an idealized mFISH pattern, of the kind generated by computer simulations and also in many (though not all) cells from a biological experiment, where all color junctions are visible and the exchange process is complete. We first assigned a minimum number of misrejoinings that must have occurred for the aberration to arise. Misrejoinings formed by the joining of two DSB ends from two nonhomologous chromosomes appear as mFISH color junctions (Fig. 1). Some other misrejoinings can be inferred rather than observed directly. For example, in a single color ring, one can infer, assuming minimal number of DSBs, that the initial configuration had a chromosome with two DSBs (three segments). Two of these segments had a telomere and a DSB end, and one segment had two DSB ends that misrejoined to form a ring (compare Fig. 2C). Those misrejoinings that can be either directly detected or inferred from the observed mFISH aberration pattern are called *manifest misrejoinings*. Manifest misrejoinings occur at color junctions and between different centromeres on a one-color stretch of a rearranged chromosome, with one extra manifest misrejoining within any one-color rearranged chromosome.

Chromosome regions between two manifest misrejoining ends or between one manifest misrejoining end and a telomere are called *manifest chromosome segments*. Each manifest chromosome segment can be classified by its color and by one of the four following labels: (TC) is assigned to a segment if it has one telomere and one centromere, (TNC) if it has one telomere and no centromere, (NTC) if it has no telomere but has one centromere, and (NTNC) if it has no telomere and no centromere. The collection of manifest chromosome segments of the same color arranged according to the above categories is called a *segment decomposition* for that color. Since we have assumed that the aberration is complete, the segment decomposition for any color must have exactly two telomeres for each centromere.

Since mFISH cannot distinguish between homologous chromosomes, it is not possible to know from mFISH patterns alone whether two segments of the same color came from the same chromosome or not unless both segments contain a centromere (in practice, experimentalists sometimes use gross discrepancies in length as an additional indicator, but we are not counting length measurements as part of the mFISH pattern in the present analysis because their use is highly variable among different laboratories). Therefore homologous chromosomes were taken into account if and only if two centromeres (or equivalently four telomeres) of the same color were detected in the observed mFISH aberration pattern. In this situation, we considered all possible distributions of breaks across the two homologues consistent with the chromosome segment data.

4.2. Calculating cycle structures

Given an observed mFISH aberration pattern, and having calculated the segment decomposition for each color, we generated all possible complete exchange processes that reshuffle the set of chromosome segments into the set of rearranged chromosomes observed in the mFISH pattern. Since manifest misrejoinings and chromosome segments are given by the observed mFISH aberration pattern, the computation involves finding all consistent sets of initial partnerships for DSB ends, where initial partners are those ends that were originally the two ends of a single DSB (compare Fig. 2). Representing initial partnership by an *initial line* (Fig. 5), an *initial line configuration* is said to be valid if the set of initial lines, in union with the manifest chromosome segments, give rise to essentially the original, unperturbed karyotype, i.e., if each connected component has a single color, two telomere ends, and exactly one centromere. Accordingly, initial lines may connect only ends on chromosome segments of the same mFISH color.

To determine valid initial lines, we proceed one color at a time, taking a segment decomposition for the color and stepwise producing a set of pairs composed of an initial line and a *reduced segment decomposition*.

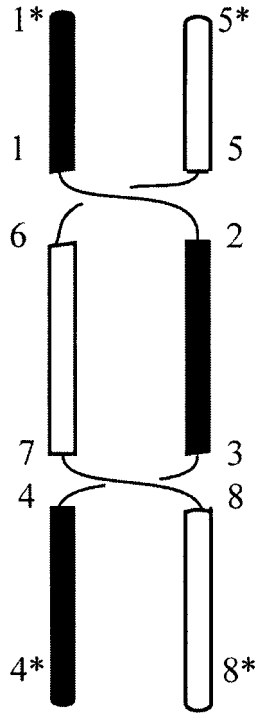


FIG. 5. The rearranged chromosomes of Fig. 2Bii with gaps. The thin lines represent the initial edges. Telomeric ends have been marked with asterisks to distinguish them from DSB free ends. Searching for the cycle structure consists of finding all possible initial edges (thin lines) that in union with the chromosome edges give rise to the original, unperturbed karyotype.

The reduced segment decomposition associated to each initial line (v, w) results from annealing the two chromosome segments, (v, v^*) and (w, w^*) , into the single segment (v^*, w^*) labeled appropriately. Should (v^*, w^*) comprise a complete chromosome (i.e., v^*, w^* are telomere ends), we omit this rearranged chromosome from the reduced segment decomposition. Proceeding recursively, one generates a tree of initial interactions terminating in empty segment decompositions, with each path from the root to a leaf specifying a unique, valid initial line configuration.

For instance, the final aberration pattern depicted in Fig. 5 (2B), where centromeres are assumed to be in the middle of the chromosome, would have a segment decomposition for chromosome 1 (Fig. 5, left) of {TNC: $(1^*, 1)$, $(4^*, 4)$; NTC: $(2, 3)$ }. There are two valid initial lines incident on vertex 2, namely $(2, 1)$ and $(2, 4)$. The line $(2, 1)$ has an associated reduced segment decomposition {TC: $(1^*, 3)$; TNC: $(4^*, 4)$ }. For this segment decomposition, vertex 3 has only one valid initial line, namely, $(3, 4)$ which results in an empty segment decomposition. Hence, $\{(2, 1), (3, 4)\}$ is a valid initial line configuration for chromosome 1. The reduced segment decomposition for the line $(2, 4)$ is {TC: $(4^*, 3)$; TNC: $(1^*, 1)$ } which has as its only available initial line $(3, 1)$. So $\{(2, 4), (3, 1)\}$ is another valid initial line configuration for chromosome 1. Similarly, chromosome 2 (Fig. 5, right) would have as valid configurations $\{(6, 5), (7, 8)\}$ and $\{(6, 8), (5, 7)\}$. We note that $(1, 4)$ and $(2, 3)$ in chromosome 1 as well as $(5, 8)$ and $(6, 7)$ in chromosome 2 are not valid initial lines since they would result in an initial configuration inconsistent with the original unperturbed karyotype.

To obtain all valid initial line configurations, we take all combinations of valid initial line configurations for each chromosome. Hence, in this example, there are four valid initial line configurations. To compute the cycle structures, we proceed as described in Sections 2 and 3. For instance, for the initial line set pictured, $\{(2, 1), (3, 4), (6, 5), (7, 8)\}$, we obtain two cycles: $\{(1, 2, 5, 6)\}$ and $\{(3, 4, 7, 8)\}$, or $c2+c2$. Another valid initial line configuration, $\{(2, 4), (3, 1), (6, 5), (7, 8)\}$ gives a single $c4$ cycle: $\{(1, 3, 8, 7, 4, 2, 5, 6)\}$.

For a given segment decomposition, there are several ways of choosing a vertex and establishing all of its possible valid initial lines. The method we will describe aims at minimizing the number of tests to determine a valid initial line. For a fixed color, we first consider lines labeled NTNC. Let (v, v^*) be such a line. Vertex v may form an initial line with any other free end w besides v^* (with v^* it would form a ring, which is not allowed). This results in the initial line (v, w) and a reduced segment decomposition omitting (v, v^*) and (w, w^*) but adding the segment (v^*, w^*) labeled appropriately. Once all the segments labeled with NTNC have been used, we proceed similarly with NTC segments. If (v, v^*) is such a segment, then vertex v may form an initial line (v, w) with the free end of any TNC segment (w, w^*) . The associated reduced segment decomposition omits (v, v^*) and (w, w^*) and adds the segment (v^*, w^*) to the category TC. Finally, if there are no NTNC or NTC segments left, we examine TC labeled segments and look for possible initial lines formed with the free end of any TNC segment. By induction it is possible to show this algorithm gives all valid configurations.

This process is repeated for each color. Each combination of valid initial line configurations of each color results in a unique overall initial line configuration. Since the misrejoinings are already given by the aberration pattern, we may compute the cycle structure of the exchange process for each valid initial line configuration by alternating from the list of initial lines and final misrejoinings and counting the number of edges in each connected component, as in Fig. 3. Searching over the set of all possible initial configurations, we determine the obligate cycle structure. In cases where computing all configurations is too computationally expensive, we randomly sampled the space of initial configurations and took the minimal cycle structure for an approximation of the obligate cycle structure.

Following the algorithm above, one can analytically describe the number of valid initial line configurations for a given observed mFISH aberration pattern as follows (Equation [1]):

$$\prod_{l \in Ab} \left[\left(\prod_{i=1}^{NTNC(l)} (FE(l) - 2i) \right) \left(\prod_{j=1}^{NTC(l)} (TNC(l) - j + 1) \right) \right. \\ \left. \left(\prod_{k=1}^{TC(l)+NTC(l)} (TNC(l) - NTC(l) - k + 1) \right) \right].$$

Here, the first product is over all colors l occurring in the observed aberration pattern Ab . $NTNC(l)$, $TNC(l)$, and $TC(l)$ are the number of NTNC, TNC, and TC segments of color l . $FE(l)$ is the total number of DSB ends with color l . Note that each of the three interior products corresponds to a step in the reconstruction algorithm described above.

We remark, as an aside, that the entire discussion above can easily be generalized to any whole-chromosome painting protocol. That is, the condition that at most two chromosomes have the same color is not an essential restriction.

5. COMPARISONS TO EXPERIMENT

5.1. Some radiation biology

Ionizing radiations, for example X-rays or high energy alpha particles, are commonly used as probes to study DNA repair and misrepair mechanisms. When ionizing radiation tracks cross a cell nucleus, it releases enough energy to disrupt the atomic structure of the DNA and induce DSBs. The dose is measured in Gray (Gy), with $1 \text{ Gy} = 1 \text{ Joule/kilogram}$. Sparsely ionizing radiations, such as gamma-rays, create DSBs randomly and independently throughout the genome. Such radiations induce ~ 40 DSBs per Gy (Löbrich *et al.*, 1994). At doses < 5 Gy, more than 90% of the DSBs restitute, so that less than 10% undergo misrejoining (Radivoyevitch *et al.*, 1998). Our comparisons with experiment use chromosome aberrations detected by mFISH in human lymphocytes after they have been irradiated with 1, 2, or 4 Gy of gamma rays (Loucas *et al.*, 2001).

There are several biophysical models for the pathways of misrejoining and chromosome aberration production. Here, we will consider two: the breakage-and-reunion model and the recombinational misrepair

model (both reviewed by Savage [1998]). Breakage-and-reunion requires more than one radiation-induced DSB for misrejoining to occur. If the ends of unrestituted DSBs are in spatial proximity at any time, they can misrejoin through an enzymatic process, presumably nonhomologous end-joining (reviewed by Ferguson [2001]) and produce a chromosome aberration (reviewed by Savage [1998]).

Recombinational misrepair, on the other hand, is based on homology of the DNA sequence that contains a DSB with other sequences in the genome. In this model, when a single DSB is produced by radiation, regions that share some sequence homology with the region that contains the DSB are searched throughout the genome and used as templates for repair by means of homologous recombination (reviewed by Modesti *et al.* [2001]). It is believed that when errors occur under this homology-based mechanism chromosome aberrations can result (reviewed by Savage [1998]).

5.2. Aberration formation simulation

Both biophysical models have been previously implemented computationally (e.g., Sachs *et al.*, 2000a; Edwards, 2002; Holley *et al.*, 2002; Ottolenghi *et al.*, 2001). Here, we use CAS (Chromosome Aberration Simulator, Sachs *et al.*, 2000a; Chen *et al.*, 1998). CAS is a Monte Carlo based computer program, which simulates aberration production by ionizing radiation. It has two adjustable parameters: the number of interaction sites S in each cell (S accounts for “proximity” effects, which favor misrejoining among spatially nearby DSBs) and the number of reactive DSBs per genome per Gy (δ), i.e., of DSBs which are not systematically restituted prior to the start of another process where restitution competes with misrejoining. In our study, $S = 10$ (Sachs *et al.*, 2000b), and δ was adjusted using experimental data from Loucas and Cornforth (2001) as explained below. We extended CAS to give mFISH patterns for karyotype changes, using an extra module to provide the relevant information on colors, centromeres, etc. in each simulated cell. CAS output was processed by the Java cycle program described in Section 4 above, so that the obligate cycle structure of each CAS-simulated observed mFISH aberration pattern could be computed. We compared theoretical results for both aberration formation pathways with experimental data.

5.3. Cycle results

After determining obligate cycle structures for all simulated cells, cycles were divided into two groups: 2-cycles (c2) and cycles of order $n > 2$ (cn, $n > 2$). In both cases, cm ($m = 2$ or $m > 2$) represents m DSBs whose ends have interacted in a “musical-chairs” exchange process (compare Figs. 2 and 4). The average number of 2-cycles per cell for 1, 2, and 4 Gy observed experimentally by Loucas and Cornforth (2001) was used to estimate the parameter δ in CAS, assuming as usual that the number of reactive DSBs per cell is linear in dose. We found $\delta = 2.8$ breaks/Gy for the breakage-and-reunion model and $\delta = 0.5$ for the recombinational misrepair model by minimizing $\chi^2 = \Sigma(obs-theor)^2/obs$ where the sum goes over the three doses (1, 2, and 4 Gy), obs is the average number of 2-cycles per cell observed by Loucas and Cornforth (2001), and $theor$ is the average number of 2-cycles per cell obtained from CAS. The χ^2 values at the minima were 0.028 for the breakage-and-reunion model and 0.23 for the recombinational misrepair model. Both values of δ are consistent with values found in previous investigations (Sachs *et al.*, 2000b). The cycle frequencies obtained for both biophysical models were then compared with each other and with experimental data. Results are shown in Fig. 6.

Both models follow the trend of the data for 2-cycles with minor discrepancies. The recombinational misrepair model overpredicted the number of 2-cycles for low doses (1 and 2 Gy) and underpredicted it for the highest dose (4 Gy). The breakage-and-reunion model predictions, on the other hand, fell within the error bars for 1 and 2 Gy and were somewhat too small for 4 Gy.

We found the frequency of n -cycles/cell, $n > 2$, to be more revealing (Fig. 6B). In this case, a major difference was found between the predictions of the two biophysical models. When we compared both models against the experimental data from Loucas and Cornforth (2001), breakage-and-reunion provided a good approximation. On the other hand, the recombinational misrepair model severely underpredicted the average number of n -cycles, $n > 2$, for all doses.

5.4. Study using the number of colors per cell

We also compared our computer predictions with experimental mFISH results expressed in terms of a somewhat different experimental endpoint, originally introduced for a different dataset (Greulich *et al.*,

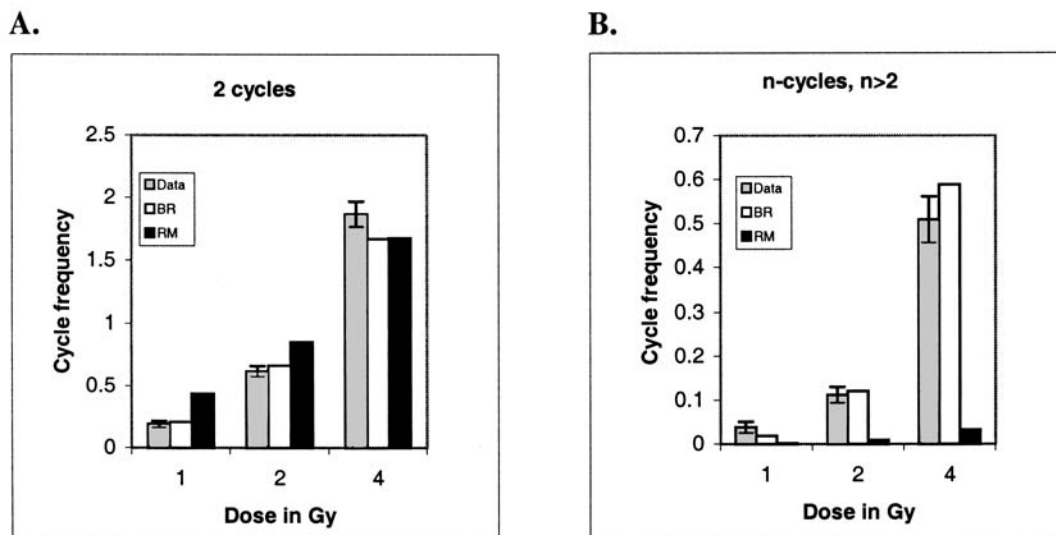


FIG. 6. 2-cycles and n -cycles, $n > 2$. The average number of 2-cycles (**Fig. 5A**) and n -cycles, $n > 2$ (**Fig. 5B**) per exposed cell versus dose, for the two aberration models (BR = breakage-and-reunion, RM = recombinational misrepair) and for the experimental data. Data error bars represent one standard deviation assuming Poisson statistics. Simulations used samples of 20,000 cells.

2000) but also applicable to the present data. This endpoint is the number of different colors involved in chromosome aberrations in each cell. Though less mechanistic than information on cycles, this experimental endpoint is more robust, e.g., as regards chromatin segments too small to be detected experimentally.

We quantified the number of colors per cell for both experimental and simulated data. We determined δ by minimizing χ^2 for the frequency of cells with no chromosome aberrations for different doses. As before, the sum was taken over all three doses, the *obs* value was computed from the experimental data, and the *theor* value was obtained from CAS. We found $\delta = 2.92$ breaks/Gy for the breakage-and-reunion model, reasonably consistent with the value found in the cycle analysis, and $\delta = 0.82$ breaks/Gy for recombinational misrepair, substantially different from the value for cycle analysis. Results for 4 Gy are shown in Fig. 7. In this case, predictions of both biophysical models followed the trend of the experimental data with certain discrepancies. Both models underpredict the number of cells involving 3 colors and overpredict for 4 colors. The breakage-and-reunion model also overpredicts 7 colors and recombinational misrepair underpredicts for 9 and 10 colors. Similar results were obtained for 1 and 2 Gy (results not shown). Overall, the color analysis, in contrast to the cycle analysis, does not provide a sharp distinction between the breakage-and-reunion pathway and the one-hit pathway.

6. DISCUSSION

Chromosome aberrations, indicators of DNA damage, can be readily detected by means of multicolor chromosome painting assays such as mFISH. In particular, the discovery of more complex chromosome aberrations than were detected by other methods has revealed these multicolor assays as an important step towards the understanding of mechanisms for aberration production. Experimental results are intricate, and accurate descriptors of the data, such as cycle structures, that can capture underlying biological information are in demand. The cycle structure of a complete exchange process, and likewise the obligate cycle structure of an observed mFISH aberration pattern, describe, intuitively speaking, the complexity of the reactions.

Here, we developed a graph-theoretical and computer simulation framework for analyzing cycles, and applied cycles to the study of chromosome aberration formation in irradiated human lymphocytes. We showed that cycles can distinguish between the breakage-and-reunion and the recombinational misrepair

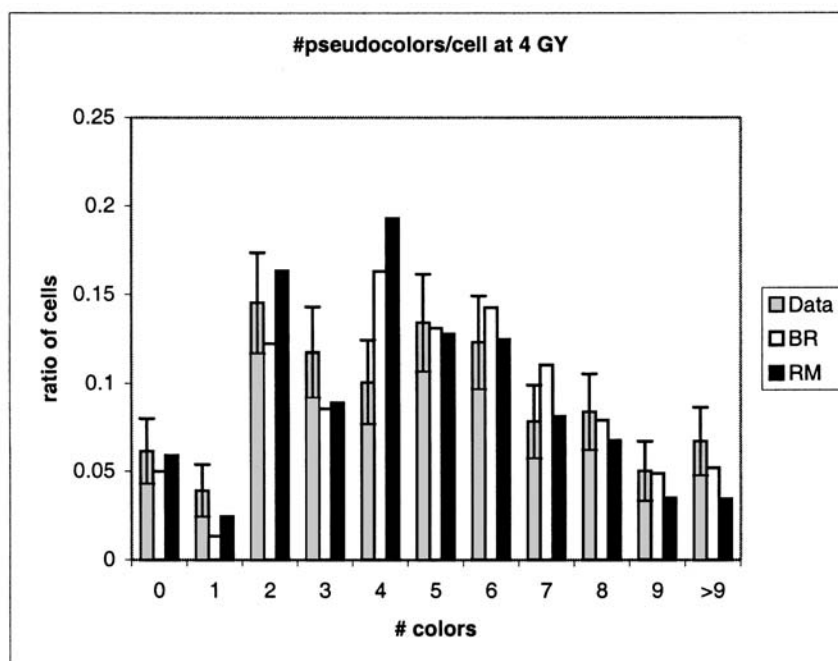


FIG. 7. Number of colors for experimental data and models. The number of colors involved in aberrations in each cell. BR and RM are as in Fig. 5. A few cells had more than 10 different colors involved and are here grouped with those that contained 10. The ratio of cells is given by the ratio of cells with a given number of colors over the total number of cells. The error bars shown were computed assuming Poisson distributions.

pathways (Fig. 6B), being more informative than an alternate endpoint given by the number of colors per cell (Fig. 7). We found that homology-based pathways (modeled by recombinational misrepair) do not generate as many higher-order cycles (cn , $n > 2$) as observed experimentally (Fig. 5b), showing, in agreement with previous studies (Richardson *et al.*, 2000; Sachs *et al.*, 2000b), that such mechanisms cannot be the main underlying mechanism for aberration production during the G0/G1 phase of the cell cycle. On the other hand, molecular biology studies have shown, consistent with our present radiobiology results, that the main repair mechanism during the G0/G1 phase of the cell cycle is nonhomologous end-joining (Takata *et al.*, 1998) and that nonhomologous end-joining can generate chromosome aberrations (Richardson *et al.*, 2000; Rothkamm *et al.*, 2001).

Whether breakage-and-reunion is the unique pathway acting during the G0/G1 phase of the cell cycle is an open question. In a more detailed calculation (not shown), we found that for 4 Gy, the breakage-and-reunion pathway also shows some discrepancies with the experimental data. The number of 3-cycles was overpredicted and the number of cycles of order greater than 5 was underpredicted. This observation suggests that a small percentage of the breaks could be repaired by homology-based mechanisms or by some other, as yet uncharacterized mechanism (Vazquez *et al.*, 2002). Using radiation as a probe and analyzing the resulting aberrations with graph theory or computer algorithms, hold promise of additional insights into DNA repair/misrepair mechanisms.

ACKNOWLEDGMENTS

Research supported by National Institute of Environmental Health Sciences, NIH, EHS Center Grant P30 ES01896 (JA), NSF grant DMS 9971169 (DL), NIH-GM 68423 (MV), NIH grant CA76260 (MC and BL), and the Low Dose Radiation Research Program, Biological and Environmental Research (BER), U.S. Department of Energy, grants ER-62684 (MC and BL) and DE-FG03-00-ER62909 (RKS). We are grateful to L. Hlatky, J. R. Savage, and S. Hannenahalli for many detailed discussions. In memory of N. Rosen.

REFERENCES

- Anderson, R.M., Marsden, S.J., Paice, S.J., Bristow, A.E., Kadhim, M.A., Griffin, C.S., and Goodhead, D.T. 2003. Transmissible and nontransmissible complex chromosome aberrations characterized by three-color and mFISH define a biomarker of exposure to high-LET alpha particles. *Radiat. Res.* 159, 40–48.
- Anderson, R.M., Stevens, D.L., Goodhead D.T. 2002. M-FISH analysis shows that complex chromosome aberrations induced by alpha-particle tracks are cumulative products of localized rearrangements. *Proc. Natl. Acad. Sci. USA* 99 (19), 12167–12172.
- Bafna, V., and Pevzner, P.A. 1996. Genome rearrangements and sorting by reversals. *SIAM J. Comput.* 25, 272–289.
- Ballarini F., Biaggi, M., and Ottolenghi, A. 2002. Nuclear architecture and radiation induced chromosome aberrations: Models and simulations. *Radiation Protection Dosimetry* 99, 175–182.
- Bauchinger, M., Schmid, E., Zitselsberger, H., Braselmann, H., and Nahrstedt, U. 1993. Radiation-induced chromosome aberrations analyzed by two-color fluorescence in situ hybridization with composite whole chromosome-specific DNA probes and a pancentromeric DNA probe. *Int. J. Radiat. Biol.* 64 (2), 179–184.
- Chen, A.M., Hahnfeldt, P., and Sachs, R.K. 1998. Chromosome aberration simulator (CAS) user's manual, source code, and executables for UNIX or Windows95/98 available on request: sachs@math.berkeley.edu.
- Cornforth, M.N. 2001. Analyzing radiation induced chromosome rearrangements by combinatorial painting. *Radiat. Res.* 155 (5), 643–659.
- Cornforth, M.N., Greulich-Bode, K.M., Loucas, B.D., Arsuaaga, J., Vazquez, M., Sachs, R.K., Bruckner, M., Molls, M., Hahnfeldt, P., Hlatky, L., and Brenner, D.J. 2002. Chromosomes are predominantly located randomly with respect to each other in interphase human cells. *J. Cell. Biol.* 159 (2), 237–244.
- Cremer, T., and Cremer, C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Rev. Genet.* 2 (4), 292–301.
- Durante M., George, K., Wu, H., and Cucinotta, F.A. 2002. Karyotypes of human lymphocytes exposed to high-energy iron ions. *Radiat. Res.* 158, 581–590.
- Edwards, A.A. 2002. Modeling radiation-induced chromosome aberrations. *Int. J. Radiat. Biol.* (78) 551–558.
- Ferguson, D.O., and Alt, F.W. 2001. DNA double strand break repair and chromosomal translocation: Lessons from animal models. *Oncogene* 20 (40), 5572–5579.
- Greulich, K.M., Kreja, L., Heinze, B., Rhein, A.P., Weier, H.U.G., Brückner, M., Fuchs, P., and Molls, M. 2000. Rapid detection of radiation-induced chromosomal aberrations in lymphocytes and hematopoietic progenitor cells by mFISH. *Mutat. Res.* 452 (1), 73–81.
- Hartsfield, N., and Ringel, G. 1994. *Pearls in Graph Theory*, Academic Press, Harcourt Brace, San Diego, CA.
- Hlatky, L., Sachs, R., Vazquez, M., Cornforth, M. 2002. Radiation-induced chromosome aberrations: Insights gained from biophysical modeling. *Bioessays* 24, 714–723.
- Holley, W.R., Mian, L.S., Park, S.J., Rydberg, B., and Chatterjee, A. 2002. A model for interphase chromosomes and evaluation of radiation induced aberrations. *Radiat. Res.* 158 (5), 568–580.
- Lee, C., Lemyre, E., Miron, P.M., Morton, C.C. 2001. Multicolor fluorescence in-situ hybridization in clinical cytogenetic diagnostics. *Curr. Opin. Pediatr.* 13 (6), 550–555.
- Limoli, C.L., Ponnaiya, B., Corcoran, J.J., Giedzinski, E., Kaplan, M.I., Hartmann, A., Morgan, W.F. 2000. Genomic instability induced by high and low LET ionizing radiation. *Adv. Space Res.* 25(10), 2107–2117.
- Löbrich, M., Ikpe, S., and Kiefer, J. 1994. Measurement of DNA double-strand breaks in mammalian cells by pulsed-field gel electrophoresis: A new approach using rarely cutting restriction enzymes. *Radiat. Res.* 65, 623–630.
- Loucas, B.D., and Cornforth, M.N. 2001. Complex chromosome exchanges induced by gamma rays in human lymphocytes: An mFISH study. *Radiat. Res.* 155 (5), 660–671.
- Mitelman, F., Johansson, B., and Mertens, F., eds. 2002. *Mitelman Database of Chromosome Aberrations in Cancer*. (www.cgap.nci.nih.gov/Chromosomes/Mitelman).
- Modesti, M., and Kanaar, R. 2001. Homologous recombination: From model organisms to human disease. *Genome Biol.* 2 (5) Reviews 1014, 1–5.
- Ottolenghi, A., Ballarini, F., and Biaggi, M. 2001. Modelling chromosomal aberration induction by ionising radiation: The influence of interphase chromosome architecture. *Adv. Space Res.* 27, 369–382.
- Pevzner, P.A. 2000. *Computational Molecular Biology: An Algorithmic Approach*, MIT Press, Cambridge, MA.
- Radvoyevitch, T., Hoel, D.G., Chen, A.M., Sachs, R.K. 1998. Misrejoining of double strand breaks after X radiation: Relating moderate to high doses by a Markov model. *Radiat. Res.* 149 (1), 59–67.
- Richardson, C., and Jasin, M. 2000. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 405 (6787), 697–700.
- Rothkamm, K., Kühne, M., Jeggo, P. A., Löbrich, M. 2001. Radiation-induced genomic rearrangements formed by non-homologous end-joining of DNA double-strands breaks. *Cancer Res.* 61 (10), 3886–3893.
- Sachs, R.K., Arsuaaga, J., Vazquez, M., Hlatky, L., and Hahnfeld, P. 2002. Using graph theory to describe and model chromosome aberrations. *Radiat. Res.* 158, 556–567.

- Sachs, R.K., Chen, A.M., Simpson, P.J., Hlatky, L.R., Hahnfeldt, P., and Savage, J.R.K. 1999. Clustering of radiation-produced breaks along chromosomes: Modeling the effects on chromosome aberrations. *Int. J. Radiat. Biol.* 75 (6), 657–672.
- Sachs, R.K., Levy, D., Chen, A.M., Simpson, P.J., Cornforth, M.N., Ingeman, E.A., Hahnfeld, P., and Hlatky, L.R. 2000a. Random breakage-and-reunion chromosome aberration formation model; an interaction-distance version based on chromatin geometry. *Int. J. Radiat. Biol.* 76 (12), 1579–1588.
- Sachs, R.K., Rogoff, A., Chen, A.M., Simpson, P.J., Savage, J.R.K., Hahnfeldt, P., Hlatky, L.R. 2000b. Underprediction of visibly complex chromosome aberrations by a recombinational-repair ('one-hit') model. *Int. J. Radiat. Biol.* 76 (2), 129–148.
- Savage, J.R.K. 1998. A brief survey of aberration origin theories. *Mutat. Res.* 404 (1–2), 139–147.
- Savage, J.R.K. 1999. An introduction to chromosome aberrations. www.infobiogen.fr/services/chromcancer/Deep/chromaber.html
- Savage, J.R.K. 2000. Reflections and meditations upon complex chromosomal exchanges. *Mutat. Res.* 512 (2–3), 93–109.
- Schröck, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M.A., Ning, Y., Ledbetter, L.D., Bar-Am, I., Soenksen, D., Garini, Y., Ried, T. 1996. Multicolor spectral karyotyping of human chromosomes. *Science* 273 (5274), 494–497.
- Schröck, E., and Padilla-Nash, H. 2000. Spectral karyotyping and multicolor fluorescence in situ hybridization reveal new tumor-specific chromosomal aberrations. *Seminars in Hematology* 37 (4), 334–347.
- Simpson, P.J., and Savage, J.R. 1995. Detecting 'hidden' exchange events within X-ray-induced aberrations using multicolor chromosome paints. *Chromosome Res.* 3 (1), 69–72.
- Speicher, M.R., Ballard, S. G., and Ward, D.C. 1996. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet.* 12 (4), 368–375.
- Takata, M., Sasaki, M., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., Takeda, S. 1998. Homologous recombination and non-homologous end-joining pathways DNA double-strand break repair have overlapping roles in the maintenance of chromosome integrity in vertebrate cells. *EMBO J.* 17 (18), 5497–5508.
- Vazquez, M., Greulich-Bode, K.M., Arsuaga, J., Cornforth, M.N., Bruckner, M., Sachs, R.K., Hlatky, L., Molls, M., Hahnfeldt, P. 2002. Computer analysis of mFISH chromosome aberration data uncovers an excess of very complicated metaphases. *Int. J. Radiat. Biol.* 78 (12), 1103–1115.
- Wang, N. 2002. Methodologies in cancer cytogenetics and molecular cytogenetics. *Am. J. Med. Genet.* 115, 118–124.
- Waterman M. 1995. *Introduction to Computational Biology: Maps, Sequences and Genomes*, Chapman and Hall, London.

Address correspondence to:

Javier Arsuaga

Mathematics Department

University of California at Berkeley

Berkeley, CA 94720

E-mail: jarsuaga@math.berkeley.edu