

Artem L. Ponomarev · David Brenner
Lynn R. Hlatky · Rainer K. Sachs

A polymer, random walk model for the size-distribution of large DNA fragments after high linear energy transfer radiation

Received: 19 July 1999 / Accepted in revised form: 10 December 1999

Abstract DNA double-strand breaks (DSBs) produced by densely ionizing radiation are not located randomly in the genome: recent data indicate DSB clustering along chromosomes. Stochastic DSB clustering at large scales, from >100 Mbp down to <0.01 Mbp, is modeled using computer simulations and analytic equations. A random-walk, coarse-grained polymer model for chromatin is combined with a simple track structure model in Monte Carlo software called DNAbreak and is applied to data on alpha-particle irradiation of V-79 cells. The chromatin model neglects molecular details but systematically incorporates an increase in average spatial separation between two DNA loci as the number of base-pairs between the loci increases. Fragment-size distributions obtained using DNAbreak match data on large fragments about as well as distributions previously obtained with a less mechanistic approach. Dose-response relations, linear at small doses of high linear energy transfer (LET) radiation, are obtained. They are found to be non-linear when the dose becomes so large that there is a significant probability of overlapping or close juxtaposition, along one chromosome, for different DSB clusters from different tracks. The non-linearity is more evident for large fragments than for small. The DNAbreak results furnish an example of the RLC (randomly located clusters) ana-

lytic formalism, which generalizes the broken-stick fragment-size distribution of the random-breakage model that is often applied to low-LET data.

Introduction

Double-strand break clustering along chromosomes

High linear energy transfer (LET) radiation consists of tracks (i.e., events) which are statistically independent of each other [1]. Recent high-LET pulsed field gel electrophoresis (PFGE) experiments [2, 3, 4, 5, 6] measured distributions of DNA fragment sizes, where “size” is used, here and throughout, to mean DNA content. PFGE fragment-size data are informative about the clustering of double-strand breaks (DSBs) along DNA, dependent on the chromosome geometry during the G_0/G_1 phase of the cell cycle and on LET or other aspects of the radiation track structure [7, 8, 9, 10, 11, 12, 13].

To elucidate the implications of the observed fragment-size distributions for DNA DSB locations along chromosomes, numerical and/or analytical modeling is required [2, 3, 4, 5, 12, 14]. In many cases, modeling has used Monte Carlo computer simulations based on the detailed track structure and the detailed geometry of chromatin [8, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. For high LET, such models have usually emphasized one-track action. They have been applied mainly to data on comparatively small sizes, analyzing ‘locally multiply damaged sites’ on the 10 base-pair (bp) scale of the underlying DNA double helix [22, 23, 24, 25] or ‘regionally multiply damaged sites’ on chromatin scales of 10 bp to several kbp ($= 10^3$ bp), corresponding to nucleosomes and the 30 nm chromatin fiber [12, 13]. For low-LET radiation and soft x-rays, Friedland and co-workers have considered much larger scales [17, 18].

The recent high-LET PFGE data on mammalian cells also include results on large fragment sizes. There are ‘globally’ multiply damaged sites. Given that a particular locus on a chromosome has a DSB, the probability that

A.L. Ponomarev (✉)
NASA ISC, Mail Code SN, Houston, TX 77058, USA
e-mail: artemp@bem.tme.edu
Tel.: +1-286-461-6191, Fax: +1-281-461-6194

R.K. Sachs
Center for Pure and Applied Mathematics, 969 Evans Hall, #1083,
University of California, Berkeley, CA 94720, USA
e-mail: alp17@math.berkeley.edu
Tel.: +1-510-2346723, Fax: +1-510-4123529
URL: <http://math.berkeley.edu/~sachs>

D. Brenner
Center for Radiological Research, Columbia University,
New York, NY 10032, USA

L.R. Hlatky
Joint Center for Radiation Therapy, Harvard Medical School,
Boston, MA 02115, USA

problem will be presented, compared to PFGE data on fragment sizes for alpha-particle irradiation of V-79 cells [4], and used as a specific example of the general RLC formalism.

We shall replace one-track DNA fragment-size distributions chosen for mathematical convenience with mechanistically calculated distributions, using the simplest available geometric model of large-scale chromosome structure, namely a random walk. The geometric model is what polymer physicists call ‘coarse-grained’ [30]: it neglects all of the molecular details of the chromatin. The model does take into account the continuous nature of the DNA thread, i.e. the fact that even at large scales loci separated by fewer base-pairs tend to be closer in space at least on average, as well as the multiple doubling back and folding which interphase chromosomes show. The random walk model and some extensions to include chromatin loops have been applied to aberrations and mutations (see overview in [31]).

There is in fact evidence of considerable randomness in chromatin structure at scales from 0.1 Mbp to more than 100 Mbp [26, 32, 33, 34]. However, there is also evidence for more regular structures such as loops (e.g. [34, 35, 36, 37, 38, 39]). Thus, the random walk model is an intentionally oversimplified idealization, which captures some of the main features of large-scale interphase chromosome geometry in a well-known, systematic, consistent, comparatively elementary construct. Very much more detailed chromatin models have been used (e.g., [12, 17, 22]). However, apart from a recent model of Friedland et al. [18], these neglect spatial correlations among chromatin regions separated by more than 0.1 Mbp, so they are not directly applicable to DSB clustering on larger scales, spanning more than three orders of magnitude, up to >100 Mbp.

Quite sophisticated track codes are now available (survey in [40]), but we suggest that an appropriate DSB model is one which has about the same level of detail for radiation track structure as for chromatin structure. The coarse-grained polymer chromosome geometry model (implemented as a discrete random walk on a cubic lattice [41]) will be combined with a very simple model of an ionizing radiation track. Essentially, we shall assume that each intersection of track and chromatin has a probability p of producing a DSB, where p is an adjustable parameter which summarizes the effects of track structure and radiochemistry in a single number. Bridging the gap between this coarse-grained approach, useful for large DNA fragment sizes, and the more detailed chromatin/track-structure calculations applicable to smaller sizes will be briefly discussed as a generalization of our track-structure model within the framework of a coarse-grained approach.

The models for chromatin and track will be used to determine one-track DNA cluster patterns by Monte Carlo simulations, checked with analytic formulae. Multi-track Monte Carlo computer simulations, cross-checked by combining the one-track simulations with the RLC equations, will be compared to PFGE data. Estimates for

the number of DSBs per Gy and for multitrack cluster patterns will be obtained. The domain of validity of the RLC formalism’s approximation that clusters due to different tracks are statistically independent will be investigated in this special example of the formalism.

Methods

We describe the computer programs used. These have been combined into a software package called DNA-break [42].

Programming random walks

Random walk models for chromatin, motivated in the Introduction section, can be continuous or discrete [43]. For computer use, we here adopt a discrete version, as follows. Consider ‘monomers’ (in the terminology of polymer physics) equally spaced along a chromosome. The monomers are regarded as DNA segments, numbered from 1 to N and then used to specify locations along the chromosome, from one telomere to the other. When CPU resources are sufficient, the number of monomers per chromosome can be increased to a maximum, determined by a condition that the monomer size is about the persistence length of chromatin, which is at least 5 kbp and perhaps more [32, 33, 39, 44]. The scales at which chromatin behaves more nearly like an elastic rod [45] are presumably at the lower margin of the experimental data (≈ 10 kbp) and need not be resolved in this calculation. For monomers which are sufficiently big, the chromatin stiffness does not have a major effect on scales larger than one monomer. We used $N=52,500$ such monomers per chromosome, so that for a typical V-79 chromosome, of size $S=245$ Mbp [11], each monomer has size ≈ 4.7 kbp. This scale determines the lower limit of resolution of the calculation; at such sizes ‘discrete-size’ effects, i.e., artifacts due to the discrete approximation, become significant.

Chromosome geometry is approximated by using a random walk rule and a random number generator to place the monomers at the points (X,Y,Z) of a cubic lattice. X , Y , and Z are integers, and a length scale L for the distance between nearest neighbors is implied. For instance, for the data considered below, we will estimate L to be about 50 nm. The random walk rule is that if one monomer is at (X,Y,Z) , the next monomer has a location given by one step of the random walk, with probability $1/6$ for each of the 6 nearest neighbors, i.e., a $1/6$ chance of being at $(X+1,Y,Z)$ $1/6$ for $(X-1,Y,Z)$, etc.

Radiation tracks

For alpha-particles at LET ≈ 100 keV/ μm , a radiation track is modeled as follows. It is assumed that the track follows a lattice line in the Z -direction and has probabili-

ty p of placing one DSB on any monomer that has a matching set of (X, Y) coordinates. For example, one track might have a chance $p=0.5$ of making one DSB on any monomer located at any lattice site of the form $(20, 30, Z)$, where Z is an arbitrary integer. The probability of one track making more than one DSB on one monomer is taken as zero; this amounts to saying that two or more one-track DSBs spaced at less than the monomer scale of several kilobase-pairs are not resolved, being counted as a single DSB in our calculation for larger-scale fragments.

It is in effect assumed that no DSBs are made at a distance more than $L/2$ from the track center. For ≈ 100 keV/ μm alpha-particles, whose tracks have a core radius <1 nm and a penumbra maximum radius of ≈ 71 nm [8, 12], this is a reasonable approximation for a coarse-grained approach. The single adjustable parameter p replaces the complicated details of the track structure and radiochemistry such as random walk parameters like L and N replace complicated molecular details of DNA and chromatin.

Multitrack action and periodic boundary condition boxes

DNA fragment-size distributions for multitrack action are obtained by Monte Carlo computations, in the following steps: simulation of a chromosome as a random walk; simulation of irradiation; recording the sizes of the resulting fragments; and then repeating simulation many times (e.g., at least 10,000) with different random walks for Monte Carlo averaging.

The simulation of irradiation assumes that tracks strike lattice locations in the (X, Y) plane at random. For example, the number of tracks hitting an $A \times A$ square [e.g., with corners $(1,1)$, $(A,1)$, (A,A) , and $(1,A)$] is assumed to be Poisson distributed; the average of the Poisson distribution is taken to be proportional to A^2 and to radiation dose; and if the number of tracks that hit the square in simulating a particular irradiated chromosome is m (chosen probabilistically in accordance with the Poisson distribution), then each of the m tracks is equally likely to hit any of the A^2 sites, independently of what the other $m-1$ tracks do. Figure 1 shows a schematic example.

In the multitrack simulations, PBC boxes of 100×100 in the (X, Y) plane were used for computational convenience. That is, the number and location of tracks in one PBC box were determined as described above, with $A=100$, and then extrapolated periodically over the entire (X, Y) plane. Use of PBC boxes for radiation is an extension of a standard trick [30]. In the present calculation, it means some artificial correlations in number and location for tracks located far apart; however, sample calculations with larger values of A , up to 1000, showed that with $A > 100$ the results do not depend on A significantly, i.e., the artificial correlations with periodicity 100 lattice units have a negligible effect.

Once simulated irradiation of a chromosome has been finished, the number of fragments of a given size j is

counted, where $j=0, 1, \dots, N$, $N=52,500$. Here $j=0$ corresponds to the situations, possible though quite rare at the doses of interest, in which two or more tracks each make one DSB on the same monomer, and size 0 is interpreted as any size smaller than the monomer size, ρ , of several kilobase-pairs. For example, suppose there is one DSB on monomer 20, two DSBs on monomer 26, and one DSB on monomer 36. Then for the fragments between adjacent DSBs we count the respective sizes 6, 0, 10. Size N corresponds to an intact chromosome.

Repeating simulated irradiation for many chromosomes and averaging give the cumulative size probability B_j , i.e., the probability that a fragment has size $\leq j$. Here $B_j \leq 1$ and $B_N = 1$. The average number of DSBs per chromosome is also computed.

Telomere effects

Most of the DNA fragments determined in the multitrack simulations have DSBs at both ends. However, a few fragments have a telomere at one end, and there is even a theoretical possibility (with negligible probability at the doses considered here) of an intact chromosome, with telomeres at both ends. Thus, the multitrack simulations systematically incorporate 'telomere effects' [14, 27], i.e., effects due to the finite size of a chromosome. In most of our calculations, telomere effects are small; for example, we consider below data for 100 Gy of alpha-particles acting on V-79 cells and estimate that in this case less than 1% of the fragments has a telomere at one end, rather than DSBs at both ends. Throughout we assume for simplicity that all chromosomes have the same size S , equal to the average V-79 size, $S=245$ Mbp. Because the telomere effects are small, using other reasonable chromosome sizes would not affect the overall conclusions.

One-track distributions and RLC formalism

The DNAbreak multitrack simulations are sufficient for direct comparisons with experiments. However, the simulations are also useful as examples of RLC formalism, and that formalism can be used to check the DNAbreak code. Descriptions of one-track DSB patterns are needed for comparing the multitrack simulations with RLC formalism. As discussed in the Appendix, two quantities are needed: the one-track fragment-size distribution neglecting telomere effects and the average number of DSBs per one-track DSB cluster. Taking advantage of the facts that a random walk has no memory and that a two-dimensional random walk is recurrent (returns to its starting place sooner or later with probability 1), Monte Carlo simulations for one-track fragment-size distributions were carried out as follows.

A three-dimensional random walk is started with monomer no. 1 located at lattice site $(1, 1, 1)$ and runs until it either returns in two dimensions, i.e., to a lattice

site of the form $(1, 1, Z)$, or reaches a cutoff monomer number $J \geq N$; we here took $J=N=52,500$. If the walk returns in two dimensions at monomer $k < J$, there is probability p of having a second DSB, in which case the fragment size is recorded as $j=k-1$ and the run is broken off. Otherwise, with probability $1-p$, the walk is continued, starting with monomer k at $(1, 1, Z)$. Iterating, one eventually either reaches a second DSB at monomer number $k' \leq J$ and records the fragment size $k'-1$, or one reaches monomer number J and the run is terminated. Repeating the whole procedure, e.g., 10,000 times, one obtains the frequencies g_j for fragments of length, $j, j=1, 2, \dots, J-1$.

Regarding a cluster as terminating whenever a segment of size $\geq J$ occurs, a renormalization factor C , the one-track fragment-size distribution f_j , and the cumulative fragment-size distribution F_j are taken to be:

$$C = \sum_{j=1}^{J-1} g_j; \quad f_j = g_j / C \quad (j=1, 2, \dots, J-1);$$

$$F_j = \sum_{k=1}^j f_k \quad (j=1, 2, \dots, J-1). \quad (1)$$

In this construction, the average cluster multiplicity M can be deduced from C , as follows. Consider the leftmost DSB in a cluster. We note that $(1-C)$ is the probability of having a fragment ending at $j \geq J$. To have such a fragment is the same as having exactly one DSB in the cluster. The probability of having exactly two DSBs is $(1-C)C$. The probability of having exactly three DSBs is $(1-C)C^2$, and so on. Each of these probabilities should be multiplied by the cluster multiplicity. Therefore, the average cluster multiplicity is

$$M = 1 - C + 2(1-C)C + 3(1-C)C^2 + \dots = 1/(1-C). \quad (2)$$

A check on the entire construction is to regard the one-track cumulative distribution F_j as a discrete version of $F(s)$ in RLC formalism (Appendix). Then the results should be independent of J provided $J \geq N$. It can be proved that this independence in fact holds (calculation not given).

One-track distributions with telomere effects included

The one-track fragment-size distribution F_j obtained by the above methods describes only fragments which have DSBs at both ends, not a telomere at one end (or both), and is the distribution needed in the RLC formalism (Appendix). For comparison, and to have some examples available for intuitive estimates, one-track clusters with telomere effects included were also generated, as follows. In a random walk of 52,500 monomers, a monomer was chosen at random and a track was put through that monomer and therefore through all other monomers with the same (X, Y) coordinates. The resulting DSB distribution on the chromosome was recorded. Figure 1 shows some examples.

Cross-checks

With any computational software, it is important to have benchmarks [30], i.e., independent ways of checking that the program is working properly, without bugs, and that the underlying algorithms are consistent and reasonable. For DNAbreak, the following cross-checks were used.

For a one-dimensional random walk, explicit forms for the return probability (essentially g_j in the case $p=1$) are known [41], with $g_j \sim j^{-3/2}$ for large j . These forms were checked successfully against g_j calculated by a one-dimensional version of the one-track Monte Carlo program described above. For two or three dimensions, precise checks are possible, at least for small j , using analytic information on the generating function for g_j [43, 46], and these checks were successfully carried out.

In the three-dimensional case, fragment sizes are determined by the two-dimensional projection of a three-dimensional random walk onto the XY plane. Asymptotically (at large j), fragments correspond to two-dimensional random walks with average size $(2/3)j$. The asymptotic behavior of g_j for $p=1$ can thus be deduced from results on two-dimensional random walks [43] as $g_j \sim 1/(j \ln^2 j)$. This equation was also confirmed by a numerical experiment. The best fitting asymptotic line in a plot of $\ln(g_j \ln^2 j)$ against $\ln j$ for a very extensive set of numerical data with $p=1$ was found to have the slope -0.97 ± 0.01 , close to the theoretical value of -1 . A similar result appears to hold for $p < 1$.

Additional cross-checks were obtained by comparing the multitrack simulations with RLC formalism in several ways, as described in the next section.

Results

We discuss the results, first parameter choices in multitrack simulations and then comparisons of multitrack simulations for alpha-particles with observations of irradiated V-79 cells [4]. The way in which the simulations exemplify RLC formalism is described next, and then some benchmarks are considered. Finally, some predicted dose-response curves are discussed and compared to the data.

Parameters

As described in the Methods section, the parameters used for the multitrack simulations are the following: the number N of monomers per chromosome; the nearest neighbor monomer spacing ρ ; the distance L between lattice sites; the length A of one side of a PBC box; the average number β of tracks per lattice site; and the probability p a track makes a DSB when it intersects the chromosome, with $0 \leq p \leq 1$. Here ρ is measured, e.g., in kbp; L has units of length, e.g. nm; the other parameters are dimensionless. There are two relations among these six parameters, and it was also found by numerical ex-

periments that certain changes in the parameters leave the results essentially unchanged, apart from discrete-size effects which apply to very small sizes [42]. Overall, two relations and two invariance transformations for six parameters mean that there are essentially only two independent parameters.

The two relations are:

$$\text{LET} \cdot \beta / L^2 = D \cdot \text{density}, \quad \rho = S/N \quad (3)$$

Here LET, dose D , and density are all measured in some consistent set of units, and S is again the size of a chromosome.

The two invariance transformations of the parameters and their interpretations are the following. First, increasing A above 100 and leaving all the other parameters invariant has no perceptible effect. The interpretation is that the use of a PBC box is practically equivalent to assuming an indefinitely large plane being struck by the radiation, as one needs for the validation of the PBC box approach.

Second, suppose N is increased, ρ is decreased with ρN constant, A is increased with A^2/N constant, β is decreased with βN constant, L is decreased with $L^2 N$ constant, and p is held constant. Then there is no change in the simulation results except for very small sizes. The interpretation is that choosing a finer mesh for the chromatin random walk, and a correspondingly finer mesh for the lattice, does not change the results except at the scale of about one mesh unit.

Comparison of multitrack simulations to experiments

This leaves us with two adjustable parameters, which we take to be the intensity of radiation β at 100 Gy (the average number of tracks striking one lattice site, here much less than 1) and p , the probability of creating a DSB when a monomer is hit by a track. By varying β and p in the multitrack simulation, the best least-square two-parameter fit was found for the empirical data obtained for V-79 cells irradiated by 100 Gy of alpha-particles. The fit exhibits acceptable agreement with the data (Fig. 2A). The parameters for the best fit, $p=0.75$, $\beta=0.0143$, were found in the simulations to give about 0.026 DSBs per Mbp per Gy. As discussed above, in the case of a one-track locally or regionally multiply damaged site on a scale less than 4.7 kbp, only one DSB is counted as contributing to 0.026 DSBs/(Mbp Gy). If we count adjacent DSBs (made by the same or different tracks) which are separated by fragments of size two monomers or less (i.e., ≈ 9 kbp or less) as a single DSB, to take into account the lower limits of resolution of the experiment and of the calculation, the DSB number decreases to about 0.019 DSBs/(Mbp Gy).

Inserting the value of β into Eq. (3) gives the distance L between nearest neighbor lattice sites as $L \approx 50$ nm. Thus, in the present calculation, the value of L is determined from a fit to the data. L in turn can be used to characterize the chromosome geometry. For large j , the average end-to-end distance for a segment of j mono-

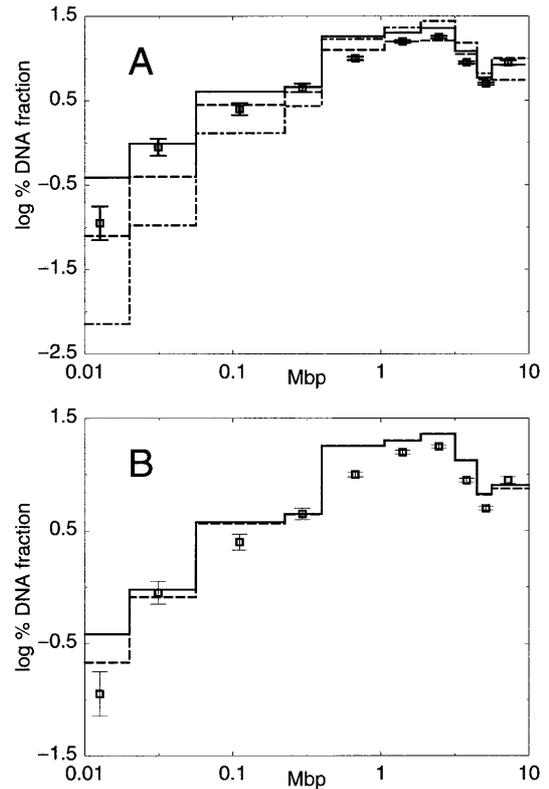


Fig. 2 Comparison of the empirical data to models. In **A**, the size-fraction, in percent, of DNA is shown plotted against DNA fragment-size bins (in Mbp) on a log-log plot. The boxes are experimental data [4] for V-79 cells, irradiated with 100 Gy of He-4 ions. The best fit obtained from multitrack DNAbreak simulations, with adjustable parameters $\beta=0.0143$ tracks per lattice site and DSB probability $p=0.75$, is shown as the *solid line*. The fit is about as good as the two-parameter fit [27] previously obtained using RLC formalism (*dashed line*), based on a Weibull fragment-size distribution function (chosen *ad hoc*). The *dot-dashed line* is the best (one-parameter) fit for the random-breakage model (Appendix), which disregards complex cluster structure, with $\lambda = 0.0056$ DSB per Mbp. In **B**, a cross-check, described in the text. *Solid curve* repeats the multitrack simulation from **A**. *Dashed curve* was obtained using one-track simulations and the RLC formalism. Near agreement is seen. The discrepancies at small sizes are due to discrete-size artifacts (see text). Those at larger sizes result from the fact that only a finite number of runs are made during the simulations

mers is $j^{1/2}L$ [47]. For a 1-Mbp DNA segment, the corresponding end-to-end distance is about 0.7 μm . For a whole chromosome, one obtains about 12 μm .

Also shown in Fig. 2A for comparison is a two-parameter fit previously given [27], based on RLC formalism (Appendix) with a one-track fragment-size distribution chosen as Weibull for mathematical convenience. It shows fewer small fragments (< 200 kbp), more intermediate-size ones (between 200 kbp and 6 Mbp), and fewer large ones (> 6 Mbp); intuitively speaking, it is thus less clustered. The two fits are comparable in quality. The third curve in Fig. 2A is a one-parameter fit using the standard random-breakage model. Symptomatic of the fact that DSBs are clustered, not random, this third fit is much less accurate than the other two.

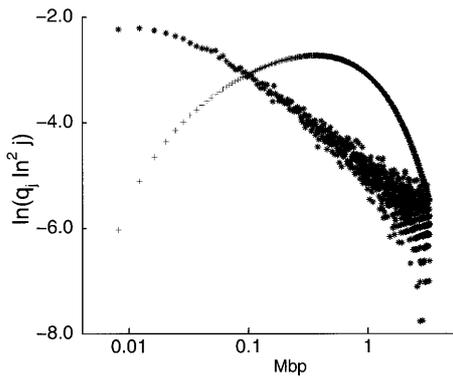


Fig. 3 Fragment sizes. The one-track fragment-size distribution g_j with $p=0.75$ for a 3-dimensional random walk, found by DNA-break simulations described in the text, is shown (*stars*). The choice of vertical axis was suggested by the asymptotic behavior for large j discussed under ‘cross-checks’ in the Methods section. The standard error of each computer-generated point is of the order of the fluctuations seen on the figure. For comparison, a curve for the standard random-breakage model, which assumes each track makes at most one DSB, is also shown (*pluses*). The parameter of the random-breakage model was adjusted to give the same average number of DSBs. It is seen that the random-breakage model predicts fewer small and more intermediate-size fragments

Multitrack simulations compared to the RLC formalism

The DNAbreak software was cross-checked using the RLC formalism. Using the parameter $p=0.75$, determined from the comparison of the multitrack simulations to experiment, the distribution g_j of Eq. (1) was calculated as described in the Methods section, with the results shown in Fig. 3. The fragment-size distribution F_j and the one-track cluster multiplicity M were then calculated from Eq. (1). F_j was interpreted as the discrete approximation to the one-track cumulative fragment-size distribution function $F(s)$ of the RLC formalism (Appendix). The RLC parameter λ (Appendix) was taken as $1/M$ times the average number of simulated DSBs per unit size, determined by simulations for the above parameter values $p=0.75$ and $\beta=0.0143$.

Given $F(s)$, M , and λ , the RLC formalism determines, via Eq. (4), corresponding DNA fractions for the given size bins. Figure 2B shows the comparison of these fractions to the fractions obtained from the multitrack simulations. As required for consistency, there is agreement within the accuracy of the calculations.

Proportionality of β and dose

Conceptually, β , the average number of tracks per lattice site in the (X, Y) plane, should be directly proportional to dose. This expectation was checked as follows. Let P be the probability that a given stretch of chromatin contains no DSBs. A proportionality between $\ln P$ and dose is found experimentally [2], predicted by the random-breakage model [29], and predicted by the RLC formalism [27]. The DNAbreak multitrack simulations showed

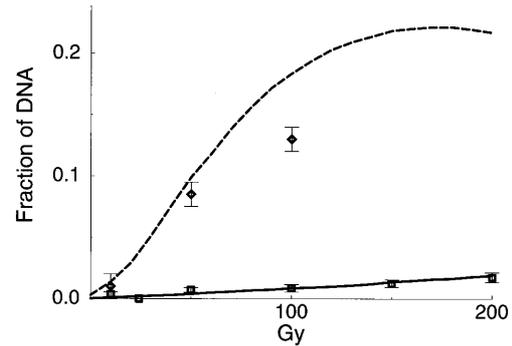


Fig. 4 Predicted dose-response curves for two different fragment-size bins. The percentage of DNA in each size bin is plotted vs dose. *Squares* show sizes in the bin 1.0–1.8 Mbp. It is seen that for relatively small doses, the response is linear. There is superlinear behavior (i.e., upward curvature) at somewhat larger doses, due to several tracks cooperating in making fragments of the appropriate sizes for this bin. For still larger doses, the predicted response levels off and ultimately decreases, as multiple tracks cooperate to cut the chromatin into sizes smaller than those in this bin. *Circles* show the relation for small sizes in a bin 20–60 kbp. Also shown are experimental points for the two different size bins, from the V-79 data [4]

that $\ln P$ is proportional to β , i.e., that choosing β proportional to dose is consistent.

Dose-response relations

With the parameter p fixed as described in the caption to Fig. 2, the multitrack simulations determine the dose-response relations for fragment-size distributions. The relations are obtained by taking β proportional to dose. Examples of dose-response relations are shown in Fig. 4. For sufficiently low doses, the relations are linear, but they become nonlinear when the dose is high enough for significant overlapping or close juxtaposition on a chromosome of different one-track DSB clusters.

Discussion

Review

In the present paper, we used Monte Carlo simulations for a geometric model of chromatin on large scales and a corresponding track-structure model, to analyze DNA fragment-size distributions after high-LET radiation. The approach was coarse-grained: chromatin was represented by a random walk, in which all molecular details are hidden behind ρ , the number of base-pairs per monomer. Similarly, details of the track structure and radiochemistry were hidden behind an adjustable parameter p , the probability that a DSB will be created if a track hits the chromatin. Such an approach has a number of weaknesses, some of which are discussed below, but it can cope with two high-LET phenomena which have not been adequately analyzed up to now: high-LET DSB clustering at large chromatin scales, up to the full length of a chro-

mosome (e.g. Fig. 1); and dose-dependent, nonlinear, multitrack DNA fragment-size distributions (e.g., Fig. 4).

Using the DNAbreak software, multitrack DSB patterns and the associated fragment-size distributions were determined, compared to experiments, and compared to other models (Fig. 2). Dose-response relations were determined (Fig. 4). The Monte Carlo simulations were cross-checked in a number of ways (e.g., Fig. 2B), thereby validating the DNAbreak software to a large extent, and also supplying a specific example of RLC formalism. As discussed in the Appendix, it was found that in this example, one of the main approximations of RLC formalism, statistical independence of different clusters, holds to high accuracy but not exactly.

Some numerical results

Agreement with experiments was adequate, but for alpha-particles the estimated total number of DSBs per Mbp per Gy is larger than in other models (e.g., [4, 14]), corresponding to an overestimate of the number of comparatively small fragments (Fig. 2A). Presumably, the reason is that a random walk doubles back very frequently at small scales; in other words, the data suggest that the persistence length of the chromatin fiber may be considerably larger than 5 kbp.

The fit to the data gave the average end-to-end distance of a 1 Mbp stretch in a V-79 chromosome as about 0.7 μm . This estimate is roughly in line with direct, non-radiobiological determinations of chromatin scale (e.g., [26]). It is substantially larger than the diameter estimate of 0.4 μm for a region containing several Mbp, recently made for V-79 cells with different radiations and a different model [5]. The corresponding average end-to-end distance for a whole V-79 chromosome is unexpectedly large, about 12 μm . We can suggest a plausible reason for this large value. In our calculation, the chromatin scale is determined mainly by data near 1 Mbp or less. However, at still larger scales, the spreading out of chromatin appears to be decreased by very large-scale loops or other structure [26, 48]. This feature would have comparatively little influence on the data, due to the upper cutoff on DNA that escapes from the plug. Thus, our estimate of total chromosome size, based on extrapolations from the 1-Mbp scale, is apparently an overestimate.

Limitations

The present approach has a number of problems and limitations. One limitation, intrinsic to the entire coarse-grained approach, is similar to that of corresponding approaches to polymer physics: Small-scale phenomena, here phenomena on a chromatin scale substantially less than ~ 10 kbp and/or a spatial scale substantially less than ≈ 100 nm, are not adequately treated. It would be possible to make the mesh finer, as pointed out at the start of the Results section, and this would not noticeably strain CPU resources. There would be no real gain, however,

because then we would be approximating chromatin as a random walk on a scale where it behaves much more like a stiff rod instead. In addition, a main feature of the present analysis, the systematic interrelation of multitrack action with one-track action, becomes irrelevant at these small scales, where, for high LET, one-track action dominates at even the largest doses used in practice. Thus, when considering comparatively small sizes, one of the much more detailed approaches (e.g., [8, 16, 17, 22]), which consider geometry of tracks and chromatin on nanometer and subnanometer scales, is superior to the present approach. The situation closely parallels the situation in polymer physics, where one must interrelate a coarse-grained with a molecular approach (e.g., [49]).

Some other limitations are less basic and could perhaps be removed by extensions of the formalism without increasing the number of adjustable parameters. Chromatin at the largest scales should perhaps be represented by a self-avoiding walk, corresponding to chromatin-chromatin interactions [47, 48], and/or with loops of somewhat less than 100 kbp [16], and/or loops on much larger scales of >1 Mbp [26, 32, 39], and/or as a constrained polymer [34]. All these extensions are technically feasible using standard polymer methods [30], provided enough detailed biological information becomes available.

The limitation, discussed in the Methods section, that a track creates no DSBs beyond a distance $L/2$ from the track center is not very onerous in the case considered here, where more than 90% of the alpha-particle energy is deposited within such distances [35]. In the framework of a coarse-grained approach, this limitation could be removed, without increasing the number of adjustable parameters, by allowing for a profile of probabilities surrounding the track center, based on the known energy profile of a track penumbra [8].

Conclusions

In addition to locally and regionally multiply damaged DNA sites, globally multiply damaged sites also occur at high LET. They can be analyzed by assuming polymer models of chromatin. This leads to a coarse-grained approach, not useful for small DNA fragment sizes but capable of dealing at least approximately with size scales spanning more than four orders of magnitude. The analysis gives the patterns for one-track DSB clusters and a systematic way to see what happens when different DSB clusters, from different tracks, overlap or come close to overlapping. It supplies a specific, fairly realistic example of a very general, RLC, formalism. It also helps relate the high-dose experimental data to the one-track action of primary interest in such applications as biodosimetry, or risk estimation for carcinogenesis.

Acknowledgments We are grateful to D. Aldous, B. Holley, A. Kellerer, H. Paretzke, K. Prise, B. Rydberg and B. O'Shaughnessy for discussions. Research supported by NIH grant GM 57245 (RKS and ALP) by NSF grant DBI-9904842 (LRH), and is supported by NIH grants RR-11623, CA-49062, CA-77285, CA-24232, ES-07361, and DOE grant FG02-98ER62682 (DIB).

Appendix: RLC formalism

Cluster intensity

The RLC formalism [27] considers situations in which one track can make a cluster of DSBs on one chromosome. For multitrack irradiation, consider the average number, λ , of such one-track clusters per unit size of the genome. λ , called the cluster intensity, is directly proportional to dose. The RLC formalism uses low-dose, one-track, one-cluster properties, corresponding to a very small cluster intensity λ , to derive multitrack, multicluster effects occurring at higher values of λ .

The basic pattern of the argument is somewhat similar to the microdosimetry argument [1], which derives a multitrack, dose-dependent specific energy distribution $f(z,D)$ from a (dose-independent) one-track specific energy distribution $f_1(z)$. However, the detailed arguments are quite different [27]. Mathematically, the RLC formalism uses the stochastic point processes called stationary Poisson cluster processes, which have been studied extensively over the years [50, 51].

Properties of individual clusters

For the moment, consider only DNA fragments from those chromosomes which are hit by just one track, so by assumption there is one and only one cluster, containing one or more DSBs, on the chromosome. For a small enough λ , virtually all damaged chromosomes would obey this criterion. Temporarily ignore telomere effects, i.e., imagine a one-track DSB cluster near the middle of a long chromosome, and consider only DNA fragments with DSBs at both ends, rather than a telomere at one or both ends. Telomere effects are put into the formalism later. Thus, for the time being, any DNA fragment considered has a DSB at each end, both coming from the same cluster made by one radiation track.

There will be a certain average cluster multiplicity $M \geq 1$. M is one more than the average number of DNA fragments per one-track cluster. Moreover, suppose one chooses a DNA fragment at random, where the choice is from among all the one-cluster fragments, as described above, i.e., the chances of using a particular cluster is proportional to the number of fragments in that cluster.

There will be a certain probability $F(s)$ that the fragment has a size less than or equal to s . Here F is a cumulative probability function, with $F(0)=0$ and $F(s)=1$ for a sufficiently large s . Given F and the size of a chromosome, the actual one-track cluster distribution, with telomere effects included, is readily obtained [27]. We shall assume for simplicity that all chromosomes have the same length, which we designate by S . The case of chromosomes having various lengths is similar and involves no new ideas.

RLC equations

The RLC equations determine multitrack action in terms of the cluster intensity λ , one-track cluster multiplicity M , one-track cluster size distribution $F(s)$, and chromosome size S . The result needed for the main text is the equation giving Φ , the DNA fraction (by size, i.e., by DNA content) for those DNA fragments that have sizes in the range S_1 to S_2 , where $S_1 < S_2 \leq S$. Φ is given by [27]:

$$\Phi = -[[s\lambda[1-(s/S)][M-(M-1)F(s)] + 1]\exp[-\lambda[s+(M-1)H(s)]]]^2 \quad (4)$$

Here $\int_0^s [1-F(s')]ds'$ and, for any function $k(s)$, $[k(s)]_1^2$ denotes $k(S_2)-k(S_1)$. Equation (4) involves the dose only via λ ; the value of λ for 1 Gy, i.e., the number of one-track DSB clusters per unit genomic size per unit dose, is usually taken as an adjustable parameter. The basic one-track quantities F and M are in principle given by some track-structure/chromatin-geometry model, such as the DNAbreak one-track simulations.

The function $b(s)$, defined by taking $b(S_1)$ as the limit of $\Phi/(S_2-S_1)$ when $S_2 \leftarrow S_1$, is the multitrack, dose-dependent, frag-

ment-size distribution, i.e., the relation of $b(s)$ to dF/ds is analogous to the relation [1] of $f(z,D)$ to $f_1(z)$ in standard microdosimetry.

Setting $M=1$ in RLC equations such as Eq.(4), i.e., assuming that one track makes at most one DSB on one chromosome, gives the broken-stick equations, depending only on λ and S , of the familiar random-breakage model, often applied to low LET [28, 29]. In particular, Eq.(4) becomes

$$\Phi = -[[s\lambda[1-(s/S)]+1]\exp[-\lambda s]]_1^2 \quad (5)$$

Equations (4) and (5) are the results needed for the present paper.

Cluster correlation effects

RLC formalism assumes that different one-track DSB clusters on the same chromosome, due to different tracks, are statistically independent [27]. DNAbreak allows one to estimate the actual magnitude of correlations between different clusters, as follows. In the multitrack simulations described in the Methods section, all hits produce DSBs on one particular chromatin configuration, and it is this commonality which produces some correlations between different clusters. We found that these correlations are present, but are small, by comparing to a simulation where chromatin 'moves' between track arrivals so that every successive hit occurs on re-randomized chromatin. In a multitrack simulation, one introduces the tracks one at a time and, before each successive track is simulated, simulates a new random walk configuration for the chromosome (carrying along whatever DSBs the chromosome already has). Then different one-track DSB clusters are completely independent of each other, since tracks and rearranged chromatin are both independent.

Specifically, let us introduce a quantity that measures the relative deviation of the cumulative fragment-size distribution B_j for the moving and fixed chromatin:

$$\Delta = \frac{B_j^{moving} - B_j^{fixed}}{B_j^{fixed}} \quad (6)$$

It was found that Δ is systematically slightly less than zero for small values of j and slightly greater than zero for intermediate values of j , but for the parameters used in Fig. 2A, its magnitude is substantially less than 1% at all j , i.e., much less than typical uncertainties in the data. Thus, intercluster dependencies, in contrast to intracluster dependencies, are negligible in the special example of RLC formalism given by DNAbreak.

References

1. Kellerer AM (1985) Fundamentals of microdosimetry. In: Kase K, Bjarnagard B, Attix F (eds) The dosimetry of ionizing radiation, Vol I. Academic Press, Orlando, pp 77-162
2. Löbrich M, Cooper PK, Rydberg B (1996) Non-random distribution of DNA double-strand breaks induced by particle irradiation. Int J Radiat Biol 70: 493-503
3. Kraxenberger A (1996) Verbesserungen der Methoden zur Pulsfeld-Gelelektrophorese und Anwendung der Technik auf die Analyse von DNA-Schäden durch dicht ionisierende Strahlung. Thesis, Technical University of Munich, Germany, pp 74-91
4. Newman HC, Prise KM, Folkard M, Michael BD (1997) DNA double-strand break distributions in X-ray and alpha-particle irradiated V79 cells: evidence for non-random breakage. Int J Radiat Biol 71: 347-363
5. Kraxenberger F, Weber KJ, Friedl AA, Eckardt-Schupp F, Flentje M, Quicken P, Kellerer AM (1998) DNA double-strand breaks in mammalian cells exposed to gamma-rays and very heavy ions. Fragment-size distributions determined by pulsed-field gel electrophoresis. Radiat Environ Biophys 37: 107-115
6. Höglund E, Blomquist E, Carlsson J, Stenerlöw B (2000) DNA damage induced by radiation of different linear energy transfer. Initial fragmentation. Radiat Res (in press)

7. Brenner DJ (1990) Track structure, lesion development, and cell survival. *Radiat Res* 124: S29–S37
8. Chatterjee A, Holley WR (1991) Energy deposition mechanisms and biochemical aspects of DNA strand breaks by ionizing radiation. *Int J Quantum Chem* 391: 709–727
9. Rydberg B, Löbrich M, Cooper PK (1994) DNA double strand breaks induced by high-energy neon and iron ions in human fibroblasts. I. Pulsed field gel electrophoresis method. *Radiat Res* 139: 133–141
10. Cedervall B, Wong R, Albright N, Dynlacht J, Lambin P, Dewey WC (1995) Methods for the quantification of DNA double-strand breaks determined from the distribution of DNA fragment sizes measured by pulsed-field gel electrophoresis. *Radiat Res* 143: 8–16
11. Friedl AA, Kraxenberger A, Eckardt-Schupp F (1995) An electrophoretic approach to the assessment of the spatial distribution of DNA double-strand breaks in mammalian cells. *Electrophoresis* 16: 1865–1874
12. Holley WR, Chatterjee A (1996) Clusters of DNA induced by ionizing radiation: formation of short DNA fragments. I. Theoretical modeling. *Radiat Res* 145: 188–199
13. Rydberg B (1996) Clusters of DNA damage induced by ionizing radiation: formation of short DNA fragments. II. Experimental detection. *Radiat Res* 145: 200–209
14. Sachs RK, Brenner DJ, Hahnfeldt PJ, Hlatky LR (1998) A formalism for analyzing large-scale clusters of radiation-induced breaks along chromosomes. *Int J Radiat Biol* 74: 185–206
15. Moiseenko VV, Edwards AA, Nikjoo H, Prestwich WV (1997) Modeling of chromosome exchanges in human lymphocytes exposed to radiations of different quality. In: Goodhead DT, O'Neill P, Menzel HG (eds) *Microdosimetry: an interdisciplinary approach*. Royal Society of Chemistry, Cambridge, UK, pp 152–155
16. Friedland W, Jacob P, Paretzke HG (1997) Biophysical simulation of DNA fragments from low-LET radiation. *Radiat Res* 148: 486–487
17. Friedland W, Jacob P, Paretzke HG, Stork T (1998) Monte Carlo simulation of the production of short DNA fragments by low-linear energy transfer radiation using higher-order DNA models. *Radiat Res* 150: 170–182
18. Friedland W, Jacob P, Paretzke HG, Merzagora M, Ottolenghi A (1999) Simulation of DNA fragment distributions after irradiation with photons. *Radiat Environ Biophys* 38: 39–47
19. Andreev SG, Khvostunov IK, Spitkovsky DM, Chepel V (1997) Clustering of DNA breaks in chromatin fiber: dependence on radiation quality. In: Goodhead DT, O'Neill P, Menzel HG (eds) *Microdosimetry: an interdisciplinary approach*. Royal Society of Chemistry, Cambridge, UK, pp 133–136
20. Brahme A, Rydberg B, Blomquist P (1997) Dual spatially correlated nucleosomal double strand breaks in cell inactivation. In: Goodhead DT, O'Neill P, Menzel HG (eds) *Microdosimetry: an interdisciplinary approach*. Royal Society of Chemistry, Cambridge, UK, pp 125–128
21. Prise KM (1997) Ionizing radiation induced clustered damage to DNA – a review of the experimental evidence. In: Goodhead DT, O'Neill P, Menzel HG (eds) *Microdosimetry: an interdisciplinary approach*. Royal Society of Chemistry, Cambridge, UK, pp 111–116
22. Ottolenghi A, Merzagora M, Paretzke HG (1997) DNA complex lesions induced by protons and alpha-particles: track structure characteristics determining linear energy transfer and particle type dependence. *Radiat Environ Biophys* 36: 97–103
23. Brenner DJ, Ward JF (1992) Constraints on energy deposition and target size of multiply damaged sites associated with DNA double-strand breaks. *Int J Radiat Biol* 61: 737–748
24. Michalik V (1993) Energy deposition clusters in nanometer regions of charged-particle tracks. *Radiat Res* 134: 265–270
25. Moiseenko VV, Edwards AA, Nikjoo N (1996) Modeling the kinetics of chromosome exchange formation in human cells exposed to ionizing radiation. *Radiat Environ Biophys* 35: 31–35
26. Sachs RK, Yokota H, Engh G van den, Trask B, Hearst JE (1995) A random-walk/giant loop model for interphase chromosomes. *Proc Natl Acad Sci USA* 92: 2710–2714
27. Sachs RK, Hahnfeldt PJ, Ponomarev AL, Hlatky LR (1999) Genomic locations of radiation-produced DNA double strand breaks: a stochastic cluster process formalism. *Math Biosci* 159: 165–187
28. Kraxenberger A, Friedl AA, Kellerer AM (1994) Computer simulation of pulsed field gel runs allows the quantitation of radiation-induced double-strand breaks in yeast. *Electrophoresis* 15: 128–136
29. Radivoyevitch T, Cedervall B (1996) Mathematical analysis of DNA fragment distribution models used with pulsed-field gel electrophoresis. *Electrophoresis* 17: 1087–1093
30. Binder K (1995) Introduction: general aspects of computer simulation techniques and their applications in polymer physics. In: Binder K (ed) *Monte Carlo and molecular dynamics simulations in polymer science*. Oxford University Press, New York, p 8
31. Wu H, Sachs RK, Yang TC (1998) Radiation-induced total-deletion mutations in the human hprt gene: a biophysical model based on random walk interphase chromatin geometry. *Int J Radiat Biol* 73: 149–156
32. Hahnfeldt P, Hearst JE, Brenner DJ, Sachs RK, Hlatky LR (1993) Polymer models for interphase chromosomes. *Proc Natl Acad Sci USA* 90: 7854–7858
33. Ostashevsky JY, Lange CS (1994) The 30 nm chromatin fiber as a flexible polymer. *J Biomol Struct Dynam* 11: 813–820
34. Dernburg AF, Broman KW, Fung JC, Marshall WF, Philips J, Agard DA, Sedat JW (1996) Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* 85: 745–759
35. Chatterjee A, Schaefer HJ (1976) Microdosimetric structure of heavy ion tracks in tissue. *Radiat Environ Biophys* 13: 215–227
36. Cremer T, Kurz A, Zirbel R, Dietzel S, Rinke B, Schrock E, Speicher MR, Mathieu U, Jauch A, Emmerich P, Schertan H, Ried T, Cremer C, Lichter P (1993) Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harbour Symp Q Biol* 58: 777–792
37. Strouboulis J, Wolffe AP (1996) Functional compartmentalization of the nucleus. *J Cell Sci* 109: 1991–2000
38. Marshall WF, Fung JC, Sedat JW (1997) Deconstructing the nucleus: global architecture from local interactions. *Curr Opin Genet Dev* 7: 259–263
39. Ostashevsky J (1998) A polymer model for the structural organization of chromatin loops and minibands in interphase chromosomes. *Mol Biol Cell* 9: 3031–3040
40. Nikjoo H, Uehara S, Wilson WE, Hoshi M, Goodhead DT (1998) Track structure in radiation biology: theory and applications. *Int J Radiat Biol* 73: 355–364
41. Chandrasekhar S (1943) Stochastic problems in physics and astronomy. *Rev Mod Phys* 15: 1–20
42. Ponomarev AL, Sachs RK (1999) Polymer chromosome models and Monte Carlo simulations of radiation breaking DNA. *Bioinformatics* (in press). (Code also available at URL <http://math.berkeley.edu/~sachs/ponomarev/artem.index.html>)
43. Weiss GH (1994) Aspects and applications of the random walk. North Holland, Amsterdam
44. Gennes P-G de (1979) *Scaling concepts in polymer physics*. Cornell University Press, Ithaca, p 31
45. Shi Y, He S, Hearst JE (1996) Statistical mechanics of the extensible and shearable elastic rod and of DNA. *J Chem Phys* 105: 714–731
46. Bhattacharya RN, Waymire EC (1990) *Stochastic processes with applications*. Wiley, New York
47. Doi M (1996) *Introduction to polymer physics*. Clarendon Press, Oxford, p 10
48. Munkel C, Eils R, Dietzel S, Zink D, Mehring C, Wedemann G, Cremer T, Langowski J (1999) Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol* 285: 1053–1065
49. Tries V, Paul W, Baschnagel J, Binder K (1997) Modeling polyethylene with the bond fluctuation model. *J Chem Phys* 106: 738–748
50. Daley DJ, Vere-Jones D (1988) *An introduction to the theory of point processes*. Springer, New York Berlin Heidelberg
51. Koenig D, Schmidt V (1992) *Zufällige Punktprozesse*. Teubner BG, Stuttgart