

A formalism for analysing large-scale clustering of radiation-induced breaks along chromosomes

RAINER K.SACHS†*, DAVID J.BRENNER‡, PHILIP J.HAHNFELDT§ and LYNN R. HLATKY§§

(Received 17 October 1997; accepted 10 February 1998)

Purpose: To model intrachromosomal clustering of DSB (DNA double strand breaks) induced by ionizing radiation. That DSB are located non-randomly along chromosomes after high LET irradiation, with clustering even at extremely large scales, has been confirmed by recent pulsed field gel electrophoresis data for size distributions of DNA fragments. We therefore extend the standard random-breakage model for DNA fragment-size distributions to a more general ‘clustered-breakage’ formalism, which can take correlations of DSB locations along a chromosome into account.

Methods: The new formalism is based mainly on a one-track probability distribution, describing the DNA fragment-size pattern due to a single primary high-energy particle, a pattern determined by track structure and chromatin geometry. Multi-track fragment-size distributions are derived mathematically from the one-track distribution, so that dose–response relations are obtained.

Results: The clustered-breakage formalism is applicable to any chromosomal geometry and any radiation track structure. It facilitates extrapolations of high-dose data to the much lower doses of interest for most applications. When applied to recently published data for irradiation of mammalian cells with ions of $\text{LET} \approx 100 \text{ keV } \mu\text{m}^{-1}$ it indicates a pattern of Mbp-scale DSB clusters, each containing a number of DSB and corresponding to a very large-scale, multiply-damaged chromatin site. Although DSB are bunched, DSB clusters are scattered almost at random throughout the genome. Estimates of DSB yield are markedly increased by resolving such clusters into individual DSB. The dose–response relation for fragments of a given size becomes non-linear when clusters from different tracks interlace or adjoin, as can occur for high doses and large sizes.

Conclusions: DSB clustering along chromosomes, which influences important radiobiological endpoints, is described quantitatively by the clustered-breakage formalism.

1. Introduction

Double strand breaks (DSB) are an important form of ionizing radiation damage (Ward 1994).¹ DSB cut

a chromosome into smaller fragments. Recent pulsed field gel electrophoresis (PFGE) data on DNA fragment-size distributions² are informative about DSB clustering, dependent on chromatin geometry in mammalian cells during cell cycle interphase and on LET or other aspects of radiation track structure (Brenner 1990, Chatterjee and Holley 1991, Rydberg *et al.* 1994, Cedervall *et al.* 1995, Friedl *et al.* 1995, Holley and Chatterjee 1996, Löbrich *et al.* 1996, Kraxenberger 1996, Rydberg 1996, Newman *et al.* 1997, Prise 1997). It has long been clear that high LET ionizations are spatially clustered (Lea 1955), and that such clustering strongly influences the fate of irradiated cells (Goodhead 1985). The PFGE data are now starting to indicate how spatial clustering of ionizations is expressed in terms of DSB clustering along chromosomes.

DSB locations along chromosomes fully determine the DNA fragment-size distribution; but to elucidate the implications of the observed fragment-size distributions for the DSB locations requires modelling (Holley and Chatterjee 1996, Löbrich *et al.* 1996, Kraxenberger 1996, Newman *et al.* 1997). In some cases, modelling has used detailed information on chromatin geometry (Chatterjee and Holley 1991, Holley and Chatterjee 1996, Rydberg 1996, Moiseenko *et al.* 1997, Friedland *et al.* 1997, Andreev *et al.* 1997, Prise 1997, Brahme *et al.* 1997, Ottolenghi *et al.* 1997). Such models have been applied mainly to data on comparatively small sizes, analysing locally multiply damaged sites on the 10bp scale of the underlying double helix (Brenner and Ward 1992, Michalik 1993, Moiseenko *et al.* 1996), or ‘regionally multiply damaged sites’ on scales of 10bp to several kbp, corresponding to nucleosomes and the

* Author for correspondence.

†Department of Mathematics, University of California, Berkeley, CA 94720, USA.

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

§Joint Center for Radiation Therapy, Harvard Medical School, Boston, MA 02115, USA.

¹Abbreviations: DSB = double-strand break; PFGE = pulsed field gel electrophoresis; LET = linear energy transfer; RBE = relative biological effectiveness; kbp = 10^3 base pairs; Mbp = 10^6 base pairs.

²Here and throughout we use ‘size’ to mean DNA content, measured in base pairs, or kbp, or Mbp. In this context, ‘size’ is synonymous with ‘molecular weight’ (Newman *et al.* 1997) and with ‘molecular length’ (Kraxenberger *et al.* 1994). Size is an indicator of the chromatin contour length along a chromosome, between two DSB, or between a DSB and a telomere, or between two telomeres, or between two restriction enzyme cutting sites, etc.

30nm fibre (Holley and Chatterjee 1996, Rydberg 1996). However, the recent PFGE data include results on much larger sizes, up to more than 5 Mbp, i.e. more than 3% of the size of an average-sized chromosome. Such sizes are so large the detailed chromatin geometry is not well characterized, the ability to view chromatin on a large scale during interphase being rather recent (Cremer *et al.* 1993). There is evidence of considerable randomness in chromatin structure at scales from 0.1 Mbp to more than 100 Mbp (Sachs *et al.* 1995, Dernburg *et al.* 1996), and also evidence for systematic structures (e.g. Cremer *et al.* 1997, Marshall *et al.* 1997), but comparatively little is known about these large scales. Thus analyses based on detailed chromatin geometry are problematical for the larger size fragments observed in the PFGE experiments.

Other analyses have used the standard random-breakage model, which assumes DSB located randomly within the genome and does not require an explicit model of interphase chromatin geometry. The random-breakage equations are based on the 'broken stick' and exponential fragment-size distributions (reviews: Kraxenberger *et al.* 1994, Radivoyevitch and Cedervall 1996). However, the random-breakage model provides a poor fit for some high LET PFGE data (Löbrich *et al.* 1996, Kraxenberger 1996, Newman *et al.* 1997), so some recent analyses have generalized it. These generalizations treat the basic parameter of the random-breakage model, namely the average number of DSB per unit size and dose, as a variable. Newman *et al.* (1997) use a different value for each measured fragment-size bin, a method which involves a large number of adjustable parameters. Kraxenberger (1996), analysing his data for heavy ions irradiating V79 cells, gives models in which the basic parameter of the random-breakage model is regarded as a stochastic quantity, with a probability distribution directly determined by a track-structure calculation. This approach has the advantage that it minimizes the number of adjustable parameters, but it does not give analytic dose-response relations, and it may not be applicable to all chromatin geometries.

The purpose of the present paper is to derive and illustrate a systematic extension of the random-breakage model, which we call the 'clustered-breakage' formalism. The new formalism results from dropping one assumption of the random-breakage model, namely the assumption that locations of different DSB in the genome are independent of each other. The formalism is based primarily on a probability distribution for DNA fragment sizes under one-track action, involving the location along chromosomes of

DSB made by one event.³ The clustered-breakage approach can be applied to any geometric configuration of the chromosomes, and to any type of radiation. The chromatin geometry and the radiation track structure in principle determine the one-track fragment-size distribution. The formalism allows size correlations among the fragments made by a single track, but, because events are statistically independent, the approximation that DSB made by different tracks are independent will be made, neglecting any correlations among one-track clustering patterns for different tracks. Then predictions for multi-track effects can be derived mathematically in terms of the one-track distribution, determining dose-dependent DSB clustering patterns and giving dose-response relations for DNA fragment-size measurements.

The clustered-breakage formalism will be illustrated by applying it to two published data sets on DNA fragment-size distributions after irradiation of mammalian cells at high doses, using nitrogen ions or α particles, each having an LET of approximately $100 \text{ keV } \mu\text{m}^{-1}$ (Löbrich *et al.* 1996, Newman *et al.* 1997). The data suggest DSB yields of more than $0.01 \text{ DSB Mbp}^{-1} \text{ Gy}^{-1}$, corresponding to approximately 80 DSB per Gy per human genome, in contrast to markedly lower estimates obtained when DSB clusters smaller than a Mbp are not resolved into individual DSB. Even for X-rays some increase in estimates of DSB yield is obtained in some experiments when the small fragments are taken into account (Löbrich *et al.* 1996, Newman *et al.* 1997). However, the high LET RBE, which appears to be less than one when analysing only larger fragments, is considerably more than one when the smaller fragments are taken into account. That is, high LET radiations make more small fragments relative to large fragments than do X-rays.

As discussed above, applying the clustered-breakage approach to the data involves specifying the distribution of DNA fragment sizes for one-track action. Since the formalism is applicable for any one-track distribution, we here, to illustrate the approach concisely, use as the one-track distribution a two-parameter form which was chosen for mathematical convenience, rather than being derived by analysing chromatin geometry and radiation track structure. For each of the two data sets, the two model parameters are determined by the fragment-size distribution at one dose. This results in predicted

³An event is one primary radiation track, with different events statistically independent. For the data considered here, an event constitutes all the energy depositions due either to one nitrogen ion (Löbrich *et al.* 1996) or to one primary α particle (Newman *et al.* 1997). We shall here use 'one-track' as synonymous with 'single-event'.

dose–response relations, which are compared to data at other doses; estimates are also obtained for the average size of a one-track DSB cluster on a chromosome and for the average number of DSB such a cluster contains. The formalism allows a mechanistically based extrapolation of the high-dose PFGE results down to the much smaller doses of interest in most applications.

Applying the model to the data shows that most DSB are situated in extremely large-scale DSB clusters, corresponding to multiply damaged sites on the Mbp scale. This behaviour would agree with the suggestion of Newman *et al.* (1997) that α irradiation may tend to produce an all-or-nothing outcome, with hits on a chromosome rare but each hit tending to produce a severe cluster. The particular one-track fragment-size distribution used to illustrate the clustered-breakage formalism is phenomenological, so the detailed numerical results obtained on clustering are less well grounded than would be the case for a distribution based on radiation track structure and chromatin geometry. However, we will argue that the basic conclusions, on the frequency of extremely large-scale clusters and their structure, are robust.

The main results of the paper are presented in §3.1, which summarizes the clustered-breakage formalism. The other parts of the paper motivate, derive, elaborate, generalize, illustrate or apply the formalism, and discuss its implications.

2. Background

2.1. Spatial localization and DSB clustering along chromosomes

Non-random DSB clustering along a chromosome results from the interplay between two kinds of localization. First, ionizations are spatially localized; at high LET they are predominantly near the line representing the centre of a track rather than being spread randomly over a whole cell nucleus, and even X-rays produce some spatial clustering of ionizations (Michalik 1993, Moiseenko *et al.* 1996). Second, a chromosome consists of a continuous thread of DNA, so that loci which are nearby along the chromosome tend to be nearby in space. Figure 1 indicates the resulting DSB clustering along a chromosome schematically, in a 2-dimensional representation. In figure 1 (A) a high LET radiation track is shown as a straight line; the interphase chromosome is shown as a wormlike random coil (Hahnfeldt *et al.* 1993; Ostashevsky and Lange 1994); and for illustrative purposes a DSB is assumed to occur wherever track and chromatin intersect. In figure 1 (B) the chromosome is shown as it would appear if straightened out

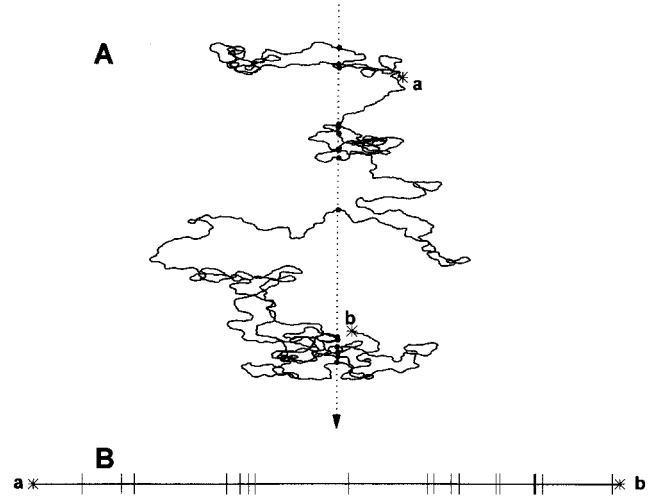


Figure 1. The figure schematically illustrates spatial DSB clustering, DSB clustering along a chromosome, and how the two kinds of clustering are related. In 1A, the chromosome is represented schematically by a computer generated wormlike random coil in two dimensions, and a track is represented schematically as a straight line. In 1B the chromosome is shown schematically straightened out along the x axis, which represents size intervals along the chromatin contour; a much smaller scale than in figure 1A is needed to fit the entire chromosome into panel 1B. The marks show genomic locations for the intersections in 1A of the lines representing chromatin and track. It is seen that these marks cluster non-randomly along the chromosome.

along the x axis. It can be seen that the DSB are clustered non-randomly along the chromosome—many occur in tight groups, with different groups separated by comparatively large stretches. In reality the situation is more complicated than the figure portrays, for several reasons: track and chromatin both have 3-dimensional structure rather than just being lines in two dimensions; even at large scales the chromatin is somewhat more ordered than a random coil (Dernburg *et al.* 1996; Marshall *et al.* 1997); and the production of DSB depends on a variety of chemical factors so that DSB production is not solely governed by geometric overlap of track and chromatin. But the basic phenomenon, of DSB clustering along the chromosome (figure 1 (B)), should still hold.

2.2. Data

Two data sets which indicate that such clustering does occur will be used to illustrate the clustered-breakage formalism. One data set is for nitrogen irradiation of human fibroblasts, at an LET of $\approx 97 \text{ keV } \mu\text{m}^{-1}$ (Löbrich *et al.* 1996). The main data concern the fraction of DNA in fragments of various

sizes up to 1.12 Mbp (table 1). Various doses, between 31.5 and 600 Gy, were used in the experiments, with larger doses used for obtaining data on the smaller sizes. For the time being we shall, as in table 1, focus attention on a dose of 189 Gy, a value for which direct measurements of 8 fragment-size bins are available and interpolation or extrapolation from nearby dose points is possible for the remaining two fragment-size bins.

Using a hybridization assay, Löbrich *et al.* also obtained data on a specific 3.2 Mbp chromatin segment defined by two consecutive restriction enzyme cutting sites. These additional data determine the probability that the given segment contains no DSB, an indicator of how many DSB per Mbp are present on this larger size scale. Figure 2, adapted from Löbrich *et al.*, shows the data. Confirmatory data (not shown here) were obtained by measurements of the fraction of activity released from the plug, an estimator of the total DNA content of fragments which have a size smaller than about 9 Mbp.

Some order of magnitude estimates supply context for analyses of the nitrogen ion data. The human G1 genome contains 46 chromosomes totalling about

Table 1. DNA fragment-size distribution.

L_2	L_1	Mean	%	Ave#
1120	780	950	13.5	19.8
780	365	573	23.2	56.7
365	225	295	7.9	37.4
225	145	185	3.4	25.7
145	97	121	1.95	22.6
97	48	73	1.78	34.1
48	23	36	0.52	20.2
23	9	16	0.57	49.9
9	2	5.5	0.089	22.6
2	0.1	0.95	0.013	19.2
				308.2

The first column gives the maximum size, in kbp, for a DNA fragment in a given size zone. The second gives the minimum size. The third gives the mean $(L_1 + L_2)/2$ for that zone. The fourth gives the observed percentage of DNA for fragments within the size zone at a dose of 189 Gy (Löbrich *et al.* 1996); for the two smallest size zones, where 189 Gy was not one of the doses used, values are obtained from values at other doses using the straight line approximations of Löbrich *et al.* The fifth column gives the average number of fragments for this size zone in a chromosome 140 Mb long, estimated directly by formula (Löbrich *et al.* 1996) $\text{Ave\#} = 140\,000 \times (\% / 100) / (\text{mean size})$. It is seen that on average the formula gives about 308 fragments with sizes between 0.1 kbp and 1.12 Mb. Assuming there are exactly 308 fragments and assuming none of these fragments have a telomere as an end, the number of DSBs implied is at least 309. 309 would apply if all the fragments were in one giant cluster of multiplicity 308 (see text). For other, more realistic, patterns of clustering, more DSBs are involved, e.g. ≈ 330 according to the clustered-breakage model of the text.

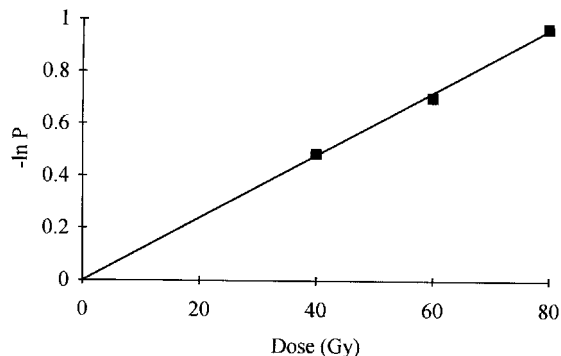


Figure 2. The dose-response relation for the probability P that a 3.2 Mbp segment of chromatin contains no DSB. Data from Löbrich *et al.* (1996). Both random-breakage and clustered-breakage models predict a straight line with zero intercept for $\ln P$, and the data are consistent with the prediction.

6400 Mbp. Thus the average size of a human chromosome is $S_C \approx 140$ Mbp. After irradiation with LET $100 \text{ keV } \mu\text{m}^{-1}$, the average number n of tracks which traverse (i.e. geometrically intersect) a human fibroblast cell nucleus is, very roughly, $n = 10D$, where D is dose in Gy; this is the relation that would hold for a cell which is a right cylinder having its axis parallel to the beam and having a cross-sectional area of $\approx 160 \mu\text{m}^2$. We shall here assume the average is exactly $n = 10D$, but values of the same order of magnitude (e.g. $n = 5D$ or $n = 15D$) could be used without any essential change, since in the data analysis n will appear together with an adjustable parameter (namely λ , introduced in equation 14 below) which determines the probability that a track hits (i.e. makes at least one DSB on) a chromosome. It will here be assumed that the track number is Poisson-distributed from cell to cell, as would occur if all the cell nuclei presented the same area to the beam; for the doses considered, where $n \gg 1$, any other distribution whose variance has order of magnitude n (or less) would give quite similar results.

The other PFGE data set to be considered here is for α -particle irradiation of V79 cells, with a LET of $\approx 110 \text{ keV } \mu\text{m}^{-1}$ (Newman *et al.* 1997). Doses from 10–200 Gy were used and the percentage of DNA in 9 size bins was measured, the smallest sizes being in a bin from 8.2 kbp to 20 kbp. The percentage of DNA in fragments larger than 5.7 Mbp was also measured. The data will be analysed similarly. In the analysis, the nominal value $n = 10D$ will be retained for the average number of tracks per nucleus, but the exact proportionality constant between D and n again has no essential bearing on the conclusions.

2.3. Telomere effects

Fragment size patterns such as those in table 1 are most readily related to DSB patterns if most fragments have DSB at both ends. Some fragments will have a telomere at one end instead, and at lower doses there are even fragments with telomeres at both ends, i.e. intact chromosomes. However, at the doses of main interest here, fragments with a telomere at one or both ends constitute a small minority, i.e. 'telomere effects' are not very significant. Telomere effects result from the finite size of a chromosome and correspond to edge effects; they are important for large fragments and small doses. In most arguments of the present paper they are analysed only to insure that they are indeed negligible. To compute telomere effects requires either extra information about the DSB production process, in addition to the information furnished by the one-track fragment size distribution, or an extra approximation (see Appendix A, Examples, and Appendix B).

2.4. Probability distributions

The size pattern for the DNA fragments formed by DSB can be described mathematically by a probability density, $f(x)$. For very small dx , $f(x)dx$ is the probability that a DNA fragment has a size in the range from x to $x+dx$. It is often convenient to work instead with an integral, $F^*(x)$, of $f(x)$:

$$F^*(x) = \int_x^\infty f(x') dx' \rightarrow f(x) = -\frac{dF^*(x)}{dx} \quad (1)$$

$F^*(x)$ is uniquely determined by $f(x)$ and conversely, so either can be taken as the basic function of interest. The PFGE experiments can be regarded as measuring $F^*(x)$, which specifies the probability that a fragment has size greater than x ; $F^*(x)=1$ when $x=0$ and F^* equals or approaches 0 for sufficiently large x . $F^*(x)$ is sometimes called a 'complementary distribution'. It complements the cumulative distribution $F(x)=1-F^*$, which specifies the probability that a given fragment has size x or less. Cumulative distributions are used more often than complementary distributions, but in the present context it is simpler to work with F^* .

2.5. The random-breakage model

Before presenting the more general clustered-breakage formalism, we first summarize some results about the often-used random-breakage model, which describes DSB that are located, independently of each other, with equal probability throughout the

genome⁴ (reviewed by Radivoyevitch and Cedervall 1996). Systematic mathematical treatments of the random-breakage model often start by considering DSB on a hypothetical segment of chromatin which is very much longer than an actual chromosome—indeinitely long in principle (Appendix A)—and is straightened out along an x axis after the DSB have been formed (compare figure 1 (B)). Here x represents size in units of base pairs. A chromosome (or a restriction enzyme fragment) is then considered as a segment of the appropriate size on the x axis, located at random within the larger segment. Thus suppose DSB are located at random along the entire non-negative x axis, with an average of κ DSB per Mbp, where κ is linearly proportional to dose. Then the probability $P(x)$ that a given segment of chromatin with size x is free from DSB is given by Poisson statistics as (Wolff 1988)

$$P(x) = \exp(-\kappa x). \quad (2)$$

It is seen in figure 2 that for $x=3.2$ Mbp, $\ln P(x)$ is, to good approximation, proportional to dose, consistent with equation (2) since κ is proportional to dose; it will later be shown that the general, clustered-breakage, formalism also predicts this linearity of $\ln P$ in dose, but with a more general interpretation for the slope. For nitrogen ion irradiation, the restriction enzyme data of Löbrich *et al.* give a numerical value of $(\kappa/D) \approx 3.8 \times 10^{-3}$ DSB per Mbp per Gy, corresponding to an average of about 100 DSB in a 140 Mbp chromosome at 189 Gy and to about 24 DSB per human genome per Gy during the G1 phase of the cell cycle. This value, as emphasized by Löbrich *et al.*, is substantially smaller than the value indicated by the data on smaller fragment sizes (table 1).

For a chromosome having length S_C prior to irradiation, the fragment-size density obeys $f(x)=0$ for $x>S_C$. Random location of DSB implies that $f(x)$ for $x \leq S_C$ is given by the 'broken stick' form (Kendall and Moran 1963, Cook and Mortimer 1991, Kraxenberger *et al.* 1994, Radivoyevitch and Cedervall 1996, Newman *et al.* 1997):

$$f(x) = \frac{\delta(x - S_C) + 2\kappa + \kappa^2(S_C - x)}{\kappa S_C + 1} \exp(-\kappa x), \quad (3)$$

$$(x \leq S_C).$$

⁴Such uniform probability does not hold at the nucleosome scale, where some sites are clearly more vulnerable to attack by free radicals than other sites, but it is more reasonable when averaging over sizes of the order of 1 kbp or more, as is appropriate for the present data. Applying the random-breakage model to the measured size distributions also assumes relatively homogeneous migration behaviour in PFGE gels for DNA fragments of a given size, as is reasonable for irradiation during G1, though not during S (Dewey and Albright 1997).

Here δ is the Dirac delta function. The normalizing factor $\kappa S_C + 1$ in the denominator represents the average number of fragments per chromosome, being one more than the average number κS_C of DSB per chromosome.

For fragments much shorter than a chromosome and for doses so high that a chromosome is on average broken many times by DSB, i.e. if $x \ll S_C$ and $1/\kappa \ll S_C$, randomness implies that the fragment-size density and the complementary distribution are both exponential (Radivoyevitch and Cedervall 1996):

$$\begin{aligned} f(x) &= \kappa \exp(-\kappa x), \\ F^*(x) &= \int_x^\infty f(x') dx' = \exp(-\kappa x), \end{aligned} \quad (4)$$

$(x \ll S_C, 1/\kappa \ll S_C).$

Equation (4) neglects telomere effects (§2.3), i.e. effects due to finite chromosome size. Specifically, taking the limit $S_C \rightarrow \infty$ in equation (3) gives equation (4).

Theoretical values for the fraction, Φ , of DNA in a size range from S_1 to S_2 are obtained by integration (Blöcher 1990):

$$\Phi = \frac{\int_{S_1}^{S_2} x f(x) dx}{\int_0^\infty x f(x) dx}. \quad (5)$$

In particular, for the random-breakage model with negligible telomere effects equation (4) holds and performing the integrals gives the fraction of DNA as (Blöcher 1990, Newman *et al.* 1997):

$$\left. \begin{aligned} \Phi &= -[\kappa x + 1] \exp(-\kappa x) \Big|_{S_1}^{S_2}, \\ \text{where } \int g(x) \Big|_{S_1}^{S_2} &= g(S_2) - g(S_1). \end{aligned} \right\} \quad (6)$$

Equation (6), for the parameter value $\kappa = 0.71 \text{ Mbp}^{-1}$ obtained by using equation (2) and the data in figure 2 as described above, is shown as the lower curve in figure 3. It is seen that this curve is a very poor fit, predicting too few fragments, especially too few small fragments. This discrepancy shows that, as emphasized by Löbrich *et al.*, their data do not fit a random-breakage model. Other data (Kraxenberger 1996; Newman *et al.* 1997) likewise show major deviations from the random-breakage model.

3. The clustered-breakage formalism

Since the random-breakage model is not adequate, we developed a formalism which is applicable more generally. The following two assumptions of the random breakage model are retained: first, that the radiation sensitivity averaged over kbp scales is

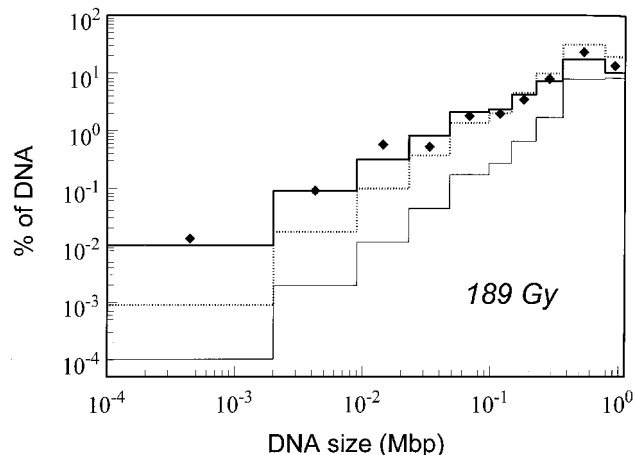


Figure 3. Data is for percentage of DNA in various DNA fragment-size bins after nitrogen ion irradiation with a dose of 189 Gy (table 1 and Löbrich *et al.* 1996). Data points are plotted in the centre of each size bin. Three calculated histograms are also shown. The lowest histogram (light solid line) was generated using the random-breakage equation (6), with the parameter $\kappa = 0.71 \text{ DSB/Mbp}$, equal to the value determined by the slope of the line in figure 2. The dotted histogram was generated using the random-breakage equation (6), with its parameter κ allowed to vary to provide the best fit in figure 3, resulting in $\kappa = 1.07 \text{ DSB Mbp}^{-1}$. The top histogram (heavy solid line) is the best fit for the two-parameter clustered-breakage model of §3.3, with the data of figure 2 used as a constraint. The corresponding parameter values are $s_w = 0.555 \text{ Mbp}$ and $\lambda = 3.01 \times 10^{-4} \text{ DSB clusters per track per Mbp}$.

approximately uniform throughout the genome, so that any one DSB, considered by itself, is equally likely to occur anywhere in the genome; and second, that the DSB pattern for any one radiation track is independent of the DSB made by other tracks. However, the new formalism takes into account the fact that a single track may make a number of different DSB with correlated locations on a chromosome, which would contradict the random-breakage model. This more general approach will be called the clustered-breakage formalism.

3.1. The basic formalism

The basic clustered-breakage formalism is now summarized, with mathematical proofs for the equations relegated to Appendix A. The formalism is based primarily on a one-track complementary distribution, $F_i^*(x)$, giving the probabilities of sizes for DNA fragments made by two DSB from the same track. F_i^* depends on chromosome geometry and on particle track parameters such as the LET. One can obtain from $F_i^*(x)$, which has no dose-dependence because it describes one-track action, the other quant-

ities of main interest. These other quantities include the dose-dependent, multi-track complementary fragment-size distribution $F^*(x)$ and theoretical estimates for experimental results such as those in table 1 or figure 2.

Generalizations are obtained for the equations of the random-breakage model by again considering the non-negative x axis populated by DSB, in some probabilistic pattern (Appendix A). Consider DSB made by one track. The one-track fragment-size complementary distribution $F_i^*(x)$ is the probability that, starting from one DSB at a location x_0 on the axis, there is no DSB within the interval $(x_0, x_0 + x]$. $F_i^*(x)$ depends only on the fragment size x , not on x_0 (Appendix A). Denote by N_1 the average number of DSB per Mbp produced by a track.⁵ As discussed in Appendix A, N_1 is given by an integral involving F_i , as follows:

$$(1/N_1) = \int_0^\infty x f_1(x) dx = \int_0^\infty F_i^*(x) dx. \quad (7)$$

The intuitive interpretation of equation (7) is that, neglecting the telomere effects discussed in §2.3, the average fragment size is the inverse of the average number of DSB per Mbp. Modifications required when telomere effects are significant are discussed in the Appendices.

When DSB are located non-randomly along a chromosome, it is important to distinguish between two quantities: the first is the number of base pairs from one DSB to the next DSB (whose distribution is described by F_i^*); the second is the number of base pairs from a randomly chosen chromatin location to the next DSB. When there is clustering, the latter can be substantially larger on average than the former, contrary to the offhand expectation that it should be half as large. This feature is a variant of the famous ‘waiting time paradox’ (Wolff 1988); it occurs because the presence of one DSB signals the likelihood of being in a cluster, where DSB are closely spaced, whilst a randomly chosen chromatin point can, on the contrary, often be outside a DSB cluster. It is shown in Appendix A that, starting from a random point on the chromatin, the cumulative distribution function, here denoted by $E_1(x)$, for the size of the DNA segment between the *random* point

and the next DSB is

$$E_1(x) = N_1 \int_0^x F_i^*(x') dx'. \quad (8)$$

The cumulative distribution E_1 is the probability that a segment of size x starting from an arbitrary point contains at least one DSB, i.e. $E_1 = 1 - P_1$, where P_1 is the probability of no DSB in the segment. For example, for a very small size x , one may approximate $F_i^* \approx 1$ in equation (8); integrating then gives $E_1 \approx N_1 x$, i.e. for a very small size x the probability that a randomly located segment of size x contains at least one DSB is simply the size times the average number N_1 of DSB per unit size.

Equations (7) and (8) refer to one-track action. They can be used to derive multi-track effects, as follows. The average number of DSB per unit size of chromatin is linearly proportional to average track number n (Appendix A), i.e.

$$\text{number of DSB per Mbp} = nN_1. \quad (9)$$

Thus, in the clustered-breakage formalism, nN_1 is the generalization of the random-breakage model parameter κ . It is shown in Appendix A that the probability $P(x)$ for a specific chromatin segment of size x to be free of DSB, equal to $1 - E(x)$ where E is the probability for the segment to contain at least one DSB, is given by

$$P(x) = \exp[-nE_1(x)]. \quad (10)$$

generalizing equation (2) of the random-breakage model. The intuitive picture is that the chromatin segment can contain a DSB if at least one DSB cluster centre hits the segment or scores a near miss, with $\exp[-nE_1(x)]$ systematically accounting for both hits and near misses as well as the internal DSB pattern of a cluster.

To generalize equation (4) one must consider a DSB and calculate the multi-track probability for the size of a corresponding DNA fragment, taking into account that the next DSB may be either from the same track or from some other track. Denoting the multi-track complementary distribution by $F^*(x)$ (§2.4), one finds (Appendix A)

$$F^*(x) = F_i^*(x) \exp[-nE_1(x)]. \quad (11)$$

Since F_i^* and E_1 are one-track, dose-independent functions the dose dependence in equations [restriction] and [F] is carried entirely by the linear dependence of n on dose, e.g. $n = 10D$.

To generalize equation (6) of the random-breakage model, one can calculate, neglecting telomere effects, the fraction of DNA in fragments having a certain size range (table 1 and figure 3). Using the notation

⁵ N_1 is averaged over the entire genome, including chromosomes not hit by the track, and is also averaged over many cells. Thus $N_1 = (\nu D)/(nL_G)$, where L_G is the total size of a G1 genome and ν is the frequently used quantity (average number of DSB per cell per Gy).

of equation (6), the fraction of DNA is (Appendix A):

$$\Phi = \left[-\exp[-nE_1(x)] / [nN_1 x F_1^*(x) + 1] \right]_{x=s_1}^{x=s_2} \quad (12)$$

Overall, given the one-track quantity $F_1^*(x)$, properties of the DSB yield and of the DNA fragment-size pattern, including their dose dependence, are determined. The clustered-breakage formalism has four basic dose-dependent, multi-track equations, equations (9)–(12); the last two assume telomere effects are negligible (Appendix A). If telomere effects are not negligible, additional information about the process is needed or an extra approximation must be used (Appendix B).

It is also proved in Appendix A that the random-breakage equations (4) and (6) are a limiting special case of a clustered-breakage model in which the track number n becomes very large and the number N_1 of DSB per track per unit size becomes very small, with nN_1 fixed as κ . In other words each cell nucleus is traversed by a very large number of tracks, with almost all tracks either missing the genome entirely or making just one DSB.

3.2. Extremely large-scale clusters of DSB

In detailed discussions it is convenient to quantify clustering in terms of a maximum size S_0 , such as the maximum size $S_0 = 1.12$ Mbp in table 1. A group of DSB with no two consecutive DSB separated by more than S_0 will be termed a ‘cluster’. The ‘multiplicity’ M of the cluster is the number of DSB it contains. For example, a ‘singlet’ (multiplicity $M = 1$) is merely an isolated DSB; and a ‘doublet’ (multiplicity $M = 2$) is formed when two DSB are closer to each other than S_0 but are further than S_0 from all other DSB on the same chromosome. Suppose the telomere effects discussed in §2.3 are negligible, i.e. that we can ignore the possibility that the cluster is flanked by a chromosome segment which is smaller than S_0 , the segment being terminated by a telomere rather than a DSB. Then a cluster of multiplicity M always generates $M - 1$ DNA fragments having sizes less than S_0 .

3.3. A two-parametric clustered-breakage model

In principle the single-track, fragment-size distribution $F_1^*(x)$ which determines the other quantities of the clustered-breakage formalism is itself determined by LET or other features of the radiation track structure and by chromatin geometry; at present, however, too little is known about large-scale chromatin geometry to calculate an analytic form

for F_1^* . Monte Carlo estimates of F_1^* , using track codes and assumptions about large-scale chromatin structure could be given but would require considerable development. Instead, we now illustrate the clustered-breakage formalism by choosing, phenomenologically and *ad hoc*, a 2-parameter form of $F_1^*(x)$.

For x less than the maximum measured size S_0 , e.g. $S_0 = 1.12$ Mbp in the case of the nitrogen ion data (table 1), F_1^* is chosen as a Weibull complementary distribution of shape parameter 0.5, i.e. (Thompson 1988)

$$F_1^*(x) = \exp[-(x/s_W)^{0.5}], \quad \text{if } x \leq S_0. \quad (13)$$

Here s_W is a positive adjustable parameter, the size scale of the Weibull distribution. The Weibull form in equation (13) was suggested by inspection of table 1 and by mathematical convenience, not by mechanistic considerations of track structure and chromatin geometry. Compared to random-breakage, for which the size density $f = \kappa \exp(-\kappa x)$ is approximately constant at small x , the Weibull form (13) describes enhanced frequency of small fragments, since the density $f_1 = -dF_1^*/dx$ is proportional to $x^{-1/2}$ at small x , and figure 3 shows such an enhancement is needed to match the data.

For large x the data are consistent with random breakage (Löbrich *et al.* 1996, Newman *et al.* 1997). The simplest way to incorporate this fact approximately into F_1^* is to assume $F_1^* = B \exp(-\lambda x)$ at large x , where λ is the other adjustable parameter of the model, while B is determined by the condition that F_1^* is continuous at S_0 .⁶ Specifically, this assumption gives

$$F_1^* = \exp[-(S_0/s_W)^{0.5}] \exp[-\lambda(x - S_0)], \quad (x \geq S_0). \quad (14)$$

Equations (13), (14) and (7) now give the number of DSB per Mbp per track as

$$\left. \begin{aligned} N_1 &= \frac{\lambda \exp(h)}{A + 1}, \quad \text{where } h = (S_0/s_W)^{0.5} \\ \text{and } A &= 2\lambda s_W [\exp(h) - 1 - h]. \end{aligned} \right\} \quad (15)$$

The other basic one-track function of the clustered-breakage formalism, determining the probabilities for

⁶Equations (13) and (14) involve a discontinuity in slope at $x = S_0$. Versions of F_1^* based on a more mechanistic model or suitably using extra adjustable parameters would not have such a discontinuity.

An intuitive interpretation of λ is given below, in the discussion of equation (17). It turns out that the most natural way to specify λ is to specify the probability that a given track misses any one average-sized chromosome.

a randomly chosen chromatin segment of size x to be free of DSB, is now obtained from equation (8) as

$$\left. \begin{aligned} E_1 &= 2N_1 s_W / [1 - (1+y) \exp(-y)] & \text{if } x \leq s_0; \\ E_1 &= 1 - \frac{\exp[-\lambda(x - S_0)]}{1+A} & \text{if } x \geq S_0. \end{aligned} \right\} \quad (16)$$

Here $y = (x/S_0)^{0.5}$.

F_i^* is used not only to describe clustering of DSB on a chromosome, but also to determine the probability that one particular radiation track hits (i.e. makes at least one DSB on) any one particular chromosome (compare Appendix A). Specifically, equation (16) gives the probability that a particular track produces no DSB on a particular chromosome of size S_C as

$$\frac{\exp(\lambda S_0)}{1+A} \exp(-\lambda S_C). \quad (17)$$

In practice, $\lambda S_0 \ll 1$ and $A \ll 1$. Neglecting λS_0 and A in equation (17) gives the approximate, but simple, expression $\exp(-\lambda S_C)$. This simple expression corresponds to the following intuitive interpretations: cluster centre locations, unlike DSB locations, are (approximately) independent of each other; moreover, the adjustable parameter λ is (approximately) the average number of DSB clusters per unit length of chromatin for single-track action.

For the data of Newman *et al.* equations (13) and (14) were also used for F_i^* , but with $S_0 = 5.7$ Mbp, the largest fragment size measured in the α -particle experiment.

3.4. Renewal approximations

In the clustered-breakage formalism the distribution F_i^* determines DNA fragment sizes, but not the order of the fragments; for example it does not specify if small fragments tend to group together (with several large fragments between different groups of small fragments), or if, instead, small and large fragments tend to intersperse. For most purposes using F_i^* is adequate, but in a few calculations, involving detailed analyses of clustering patterns, extra information must be given, not just F_i .

One simple approximation which always supplies the missing information involves neglecting correlations between the sizes of adjacent DNA fragments. Specifically, one can assume that whenever a DSB occurs the entire probabilistic situation 'renews' itself (Appendix B). Any F_i^* can be supplemented by using renewal approximation; analysing the renewal approximation then gives all relevant details about

clustering (Appendix B). It is argued in Appendix B that in the present context renewal approximation determines a lower limit for clustering, i.e. predicts smaller clusters of lower multiplicity than does any other reasonable clustered-breakage model having the same F_i^* .

4. Comparison to data

We first describe results obtained for human cells irradiated with nitrogen ions (Löbrich *et al.* 1996), then similar results for V79 cells irradiated with α particles (Newman *et al.* 1997).

4.1. The nitrogen ion data

The clustered-breakage model of §3.3 has two adjustable parameters, s_W and λ . For the data on human cells, these parameters were determined as follows. The model predicts a straight line dose-response relation with zero intercept in figure 2, since n in equation (10) is linearly proportional to dose and $E_1(x)$ is a single-track quantity, independent of dose; using the slope of the straight line shown in figure 2 shows that at $D = 189$ Gy, $nE_1(3.2) \approx 2.27$ DSB, where E_1 is determined by equations (7), (8), (13) and (14) involving the two parameters and n is assigned the value $10D$. The two parameters were obtained by fitting the data in figure 3 to equations (13) and (14) using a constrained least squares algorithm, the constraint being $nE_1(3.2) = 2.27$. The values obtained in this way were $s_W = 0.555$ Mbp and $\lambda = 3.01 \times 10^{-4}$ DSB clusters per track per Mbp.

It is seen in figure 3 that the clustered-breakage model with these parameters (heavy solid curve) is a dramatic improvement over the random-breakage model (light solid curve). In part the improvement is merely due to the fact that the clustered-breakage model uses two adjustable parameters (s_W and λ) to fit the data of figures 2 and 3, whereas the random-breakage model has only κ available, which is fixed by the data of figure 2. However, one can put the two models on a par as far as number of adjustable parameters is concerned, by allowing other values of κ in figure 3, i.e. by dropping the constraint implied by figure 2 for the random-breakage model while retaining the constraint for the clustered-breakage model. Even in this case, no fit comparable to that of the clustered-breakage model can be obtained from the random-breakage model. The figure shows a random-breakage curve for that value of κ which gives the best least-squares fit to the data of figure 3, ignoring the data of figure 2, $\kappa = 1.07$ DSB per Mbp. The fit is still quite poor. This discrepancy means that the *relative fractions* of DNA in the various size

bins of table 1 are inconsistent with randomly located DSB, even apart from the overall average number of DSB per Mbp judged from figure 2.

With the two parameters fixed, the model makes various predictions. First, dose-response curves for the percentage of DNA in the various size bins of table 1 can be obtained, neglecting telomere effects, by allowing n in equations (11) and (12) to scale linearly with dose, $n = 10D$ where D is in Gy. These curves are shown in figure 4, compared to the data. The fit in some of the panels is mediocre, but it should be recalled that all the data, for figure 2 and

all the panels in figure 4, is being fitted by adjusting two parameters at 189 Gy and leaving the parameters fixed thereafter.

The main features of dose-response curves in the clustered-breakage formalism neglecting telomere effects can be seen from figure 4. For low doses almost all fragments in one size bin are due to single-track action and the dose-response is therefore linear (not quadratic as it would be in the random-breakage model with telomere effects neglected; this linear/quadratic difference between high/low LET parallels the corresponding differences in exchange-type chromosome aberration dose-response curves). For higher doses there can be a super-linear behaviour, due to different tracks co-operating to make fragments of the given size. For still higher doses, the yield increases more slowly and ultimately starts to decrease, as interlacing of DSB clusters from many tracks starts to cut DNA segments into sizes smaller than the sizes in the given bin.

The total DSB yield implied by the parameters $s_w = 0.555$ Mbp and $\lambda = 3.01 \times 10^{-4}$ Mbp $^{-1}$ in the one-track function F_i^* is given by equation (9) as approximately 0.012 DSB Mbp $^{-1}$ Gy $^{-1}$, in agreement with the estimate made more directly from the same data by Löbrich *et al.* (1996), who found in addition that 0.012 DSB Mbp $^{-1}$ Gy $^{-1}$ corresponds to an RBE greater than 1 compared to X-rays in a companion experiment. At 189 Gy, 0.012 DSB Mbp $^{-1}$ Gy $^{-1}$ would give a total of about 330 DSB on a chromosome 140 Mbp long; this number is somewhat larger than the number of fragments smaller than $S_0 = 1.12$ Mbp (table 1), just as one would expect. Some of the extra DSB are singlets (§3.2) and some come from the fact that in each cluster the number of DSB is one larger than the number of fragments.

Here the contribution corresponding to fragments smaller than 0.1 kbp, predicted by the model when the model is extrapolated down to zero sizes, is negligible. In principle these small sizes must be discussed separately, for the following two reasons. First, there may be cryptic DSB multiplets in locally multiply damaged sites, corresponding to fragments too small to be resolved in the present experiments. Second, the particular two-parameter form of F_i^* used here is not expected to be a reasonable approximation at such small sizes, where systematic geometric properties of chromatin dominate the fragment size pattern (Holley and Chatterjee 1996, Rydberg 1996), though the general clustered-breakage formalism should still apply.

For single-track action the above estimates correspond to about 8 DSB per genome. Moreover, the probability that a track misses a particular average-

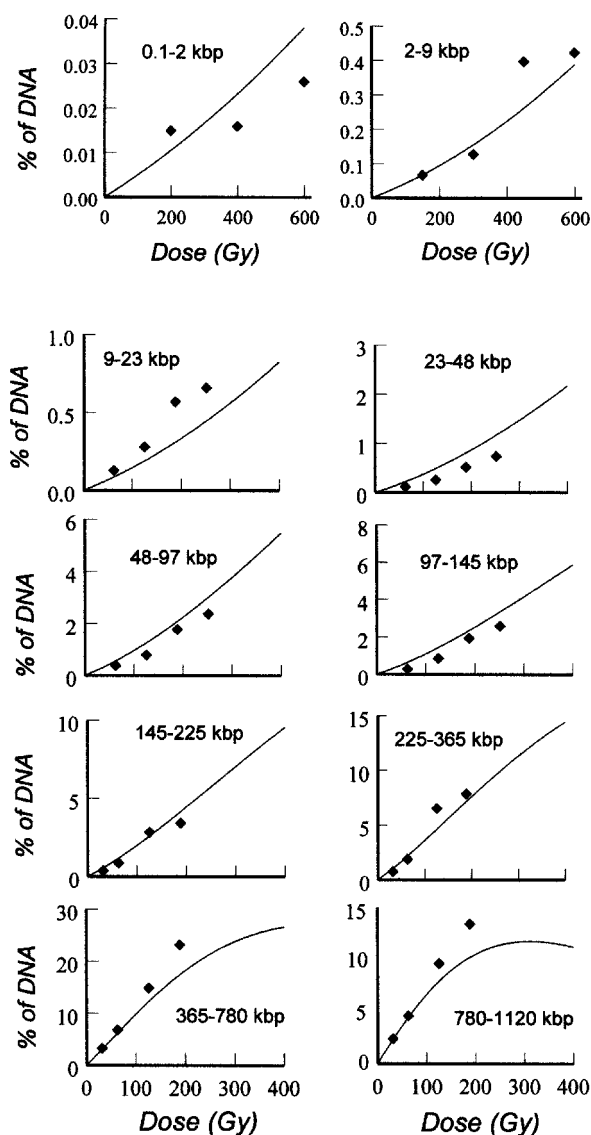


Figure 4. Data points and dose-response curves for the 9 DNA fragment-size bins used in the nitrogen ion experiments of Löbrich *et al.* (1996). Theoretical curves are for the clustered-breakage model of equations (13) and (14), with the two parameters determined by hybridization data (figure 2) and by data at 189 Gy (figure 3).

sized chromosome is given by equation (17) and the parameter values as about 0.96. Thus the model predicts that typically one track hits only about two of the 46 chromosomes. Extrapolating, this would mean that at low doses, e.g. 0.2 Gy, most chromosomes are missed altogether by all tracks, but a hit chromosome usually has one or two Mbp-scale clusters with about 4 DSB each.

4.2. The α -particle data

The data of Newman *et al.* (1997) for V79 cells was analysed similarly, one difference being that there were no extra data corresponding to figure 2 for a restriction enzyme fragment.

Figure 5 shows the data for 100 Gy. The clustered breakage model, equation (13) with $S_0 = 5.7$ Mbp and equation 14, was compared to the data (solid curve). The two model parameters were obtained by a least-squares fit in figure 5; in contrast to the similar procedure for figure 3, the fit was unconstrained, and the availability of error bars on the data points allowed a weighted fit to be performed. The resulting parameter values were $s_w = 1.285$ Mbp, $\lambda = 1.3 \times 10^{-4}$ (DSB clusters per Mbp per track). These parameter values correspond via equation (9) to approximately 0.011 DSB Mbp $^{-1}$ Gy $^{-1}$, remarkably similar to the value given above for nitrogen ion irradiation of human fibroblasts.

Also shown in figure 5 is the best-fitting random-

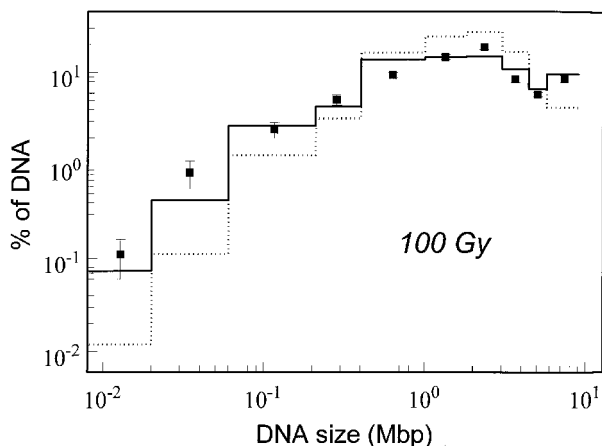


Figure 5. Data points are for percentage of DNA in various DNA fragment-size bins following a 100 Gy dose of α particles (Newman *et al.* 1997). Data is plotted as in figure 3, except that data for all sizes larger than 5.7 Mbp was plotted using the bin 5.7–10 Mbp for convenience. The dotted histogram is the best fitting random-breakage model, with parameter $\kappa = 0.85$ DSB/Mbp. The solid histogram is the best fit for the clustered-breakage model of equations (13) and (14), corresponding to the parameter values $s_w = 1.285$ Mbp, $\lambda = 1.3 \times 10^{-4}$ DSB clusters per track per Mbp.

breakage model, with parameter $\kappa = 0.85$ DSB Mbp $^{-1}$. As in figure 3, the random-breakage model drastically underestimates the number of small fragments, whereas the clustered-breakage model approximates the data more closely. Figure 6 shows dose-response curves predicted by the clustered-breakage model for the various size bins, compared to the data. As in figure 4 the theoretical curves reproduce the main features of the data. The most obvious discrepancy is that the model underestimates the number of fragments measured for the 20–60 kbp bin. This discrepancy could be an indication of some chromatin feature, such as loops, at this size range (Newman *et al.* 1997). However, in figure 4 the most nearly corresponding underestimate occurs for the 9–23 kbp bin whereas the next larger bin is overestimated.

Figure 7 presents the theoretical predictions of the clustered-breakage model for the α particle data as a 3-dimensional graph. By tracing lines it can be seen that at a fixed dose, the fragment-size distribution $xf(x)$ typically has a maximum at a size which decreases somewhat as dose increases.

4.3. Clustering patterns

We now describe some details that assume minimal clustering, i.e. use renewal approximation (Appendix B). The definitions given for clusters in §3.2 and the above parameter values will also be assumed. Then, for the nitrogen ion data of Löbrich *et al.* (1996), the following statistics on cluster size and cluster multiplicity for one-track clusters follow from the equations of Appendix B, the average multiplicity is $M_1 \approx 4.1$ DSB, with standard deviation ≈ 3.6 and considerable skewing toward high-multiplicity clusters; the average cluster size, including zero-size singlets, is ≈ 0.79 Mbp; the standard deviation is $\sigma \approx 1.04$ Mbp, σ being larger than the average because there is marked skewing toward clusters of large size. On average a track which traverses the nucleus hits two different chromosomes, making one cluster on each, and misses all the other chromosomes. It is argued in Appendix B that modifying the renewal approximation in any plausible way would lead to predictions of more pronounced clustering, i.e. predict still larger one-track clusters having still higher multiplicity, coupled with an even smaller probability that a track hits a chromosome.

At higher doses, overlap of the clusters from different tracks becomes increasingly important, so the clusters become larger, with higher multiplicity, and with a greater proportion of smaller fragments. The average cluster multiplicity M increases exponentially with dose as clusters interlace and coalesce,

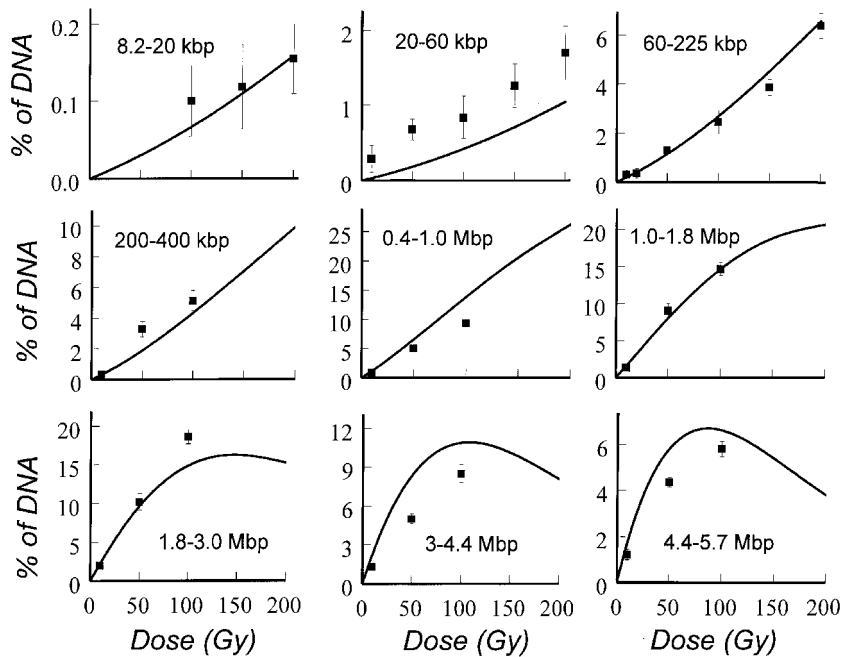


Figure 6. Theoretical dose-response curves compared to data for the 9 DNA fragment-size bins measured in the α particle experiment of Newman *et al.* (1997). The theoretical curves are for the clustered-breakage model of equations (13) and (14), with the two adjustable parameters established by fitting to the data for 100 Gy (see figure 5).

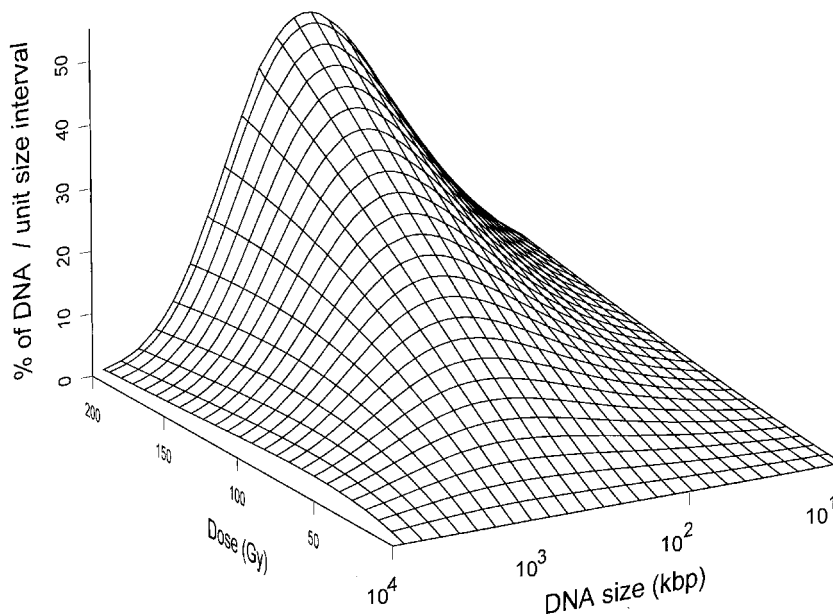


Figure 7. Calculated percentage of DNA per unit bin width as a function of dose and of DNA fragment size for the α particle experiments of Newman *et al.* (1997). The height at a given dose D and given DNA fragment size x is $xf(x)/\int_0^\infty xf(x)dx$, where $f(x)$ is the probability density for DNA fragment sizes. The calculation uses the clustered-breakage model of equations (13) and (14), with the two adjustable parameters established by fitting to the data for 100 Gy (see figure 5).

$M = M_1 \exp(\alpha n)$, where $n = 10D$ and $\alpha = E_1(S_0)$. Figure 8, obtained by Monte Carlo calculations as described in Appendix B, shows a sample predicted DSB pattern for 30 Gy irradiation with nitrogen ions. On the representative chromosome shown, there are

11 clusters, some extending several Mbp and having rather high multiplicity. The clusters are separated from each other and from the telomeres by stretches averaging roughly 10 Mbp. Such a DSB pattern is of course stochastic, varying considerably from chromo-

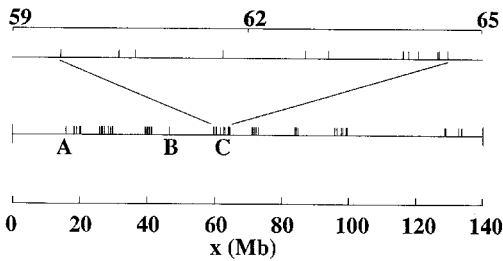


Figure 8. A 140 Mbp chromosome, with a sample pattern, for a dose of 30 Gy, of 72 DSB in 11 clusters. Telomeres are shown at each end. The sample pattern was obtained by one Monte Carlo run (Appendix B), using renewal approximation to the clustered-breakage model of §3.3, with the two parameters chosen as described in figure 3. The largest DSB cluster for this particular case, C, is shown magnified in an insert. It has multiplicity 12 and stretches over almost 5 Mbp. Note also the singlets at A and B. Such singlets are 'extra' DSB, contributing nothing to the fragments tabulated in table 1.

some to chromosome and cell to cell. For example, in a sample of 100 Monte Carlo runs for 30 Gy, one run included a cluster of multiplicity 26, formed by superposition of clusters from three different tracks.

For the α particle data of Newman *et al.* (1997) a cluster can be defined by DSB spaced no more than 5.7 Mbp (rather than 1.12 Mbp) apart. Assuming minimal clustering, the clustered-breakage model predicts an average of about 8.3 DSB per cluster, with the average size of a cluster about 7.5 Mbp. According to the model, one α particle which traverses the nucleus on average hits a G1 V79 genome only about once, i.e. typically makes just one DSB cluster on one chromosome, missing all the other chromosomes.

5. Discussion

5.1. Summary

Radiation induced DSB were analysed by generalizing the standard random-breakage model, allowing for the possibility that DSB along a chromosome can be tightly bunched in groups, with different groups widely scattered. The resulting clustered-breakage formalism gives predictions for dose-response relations and for multi-track effects, starting from a dose-independent, single-track, fragment-size distribution F_i^* (§3.1). Single-track DSB fragment-sizes, characterized by F_i^* , depend on chromatin geometry, radiation track structure, and radiation chemistry; however, once F_i^* is specified, and assuming independence of DSB patterns from different tracks, analysing multi-track effects becomes a mathematical exercise (equations (7)–(12)).

Analysing published data on human fibroblasts

irradiated with nitrogen ions, dose-response curves for DNA percentage in 9 different fragment-size ranges (table 1) were obtained using two adjustable parameters in an illustrative clustered-breakage model, and were compared to the data (figures 2–4). It was found that the clustered-breakage model describes the data much better than does the random-breakage model. A different data set, For V79 cells subjected to α -particle radiation, was analysed similarly (figures 5–7) and gave results consistent with those for the first data set. In both data sets, the yield of DSB was estimated and found to be quite large, more than $0.01 \text{ DSB Mbp}^{-1} \text{ Gy}^{-1}$. This result is consistent with the estimates in the original papers of an RBE considerably greater than 1 compared to X-rays.

Details of DSB clustering were investigated using analytic and Monte Carlo techniques (§4.3). For single-track action, even assuming minimal clustering, estimates for the nitrogen ion data of Löbrich *et al.* (1996) gave the average cluster size as almost 1 Mbp, with a few much larger clusters, where a cluster was defined by the criterion that neighbouring DSB are closer than 1.12 Mbp. On this picture, the average DSB multiplicity of a cluster is more than 4, and most DSB are part of a cluster, rather than being singlets. Assuming an average of 10 tracks per Gy ($n=10D$) a given track has a probability of ≈ 0.04 of hitting any given chromosome. Clusters, but not dSBs, are scattered almost at random throughout the genome. The basic clustering pattern at doses of 30 Gy or less is single-track clusters of DSB, separated by large DSB-free stretches (figure 8). There is some interlacing of clusters from different tracks, which increases at higher doses; the clustered-breakage formalism systematically keeps track of such interlacing. For the α -particle data, clustering was also found to be very pronounced.

5.2. Robustness of the results

It is useful to distinguish between features inferred directly from the clustered-breakage formalism, compared to features which invoke the specific 2-parametric form introduced in §3.3, the latter features being less robust than the former. The overall pattern, of extremely large-scale clusters and a total DSB yield greater than for low LET when fragments of small sizes are included, is suggested directly by high LET PFGE data (Newman *et al.* 1997, Löbrich *et al.* 1996). The pattern is implied (Appendices A and B) by any reasonable clustered-breakage model applied to the data. On the other hand, particular numbers (e.g. that typically each nitrogen ion makes one cluster of DSB, rather than

two, on each chromosome it hits) are in part consequences of the particular proportionality factor 10 tracks per Gy used in $n=10D$ and/or of the special form for F_i^* given in §3.3. This special form was chosen *ad hoc* and has a discontinuity in the slope at $x=S_0$. It should be regarded mainly as a feasibility demonstration: it is feasible to produce a simple, explicit clustered-breakage model which is consistent with the data.

Improved fits with the data could, trivially, be obtained by increasing the number of adjustable parameters in F_i^* . A more mechanistic approach would be to use Monte Carlo simulations for single-track action, with not only the track but also the chromatin geometry modelled stochastically.

5.3. Influence of clustering on other endpoints

The prevalence at high LET of extremely large-scale DSB clusters has significant implications for various biological endpoints that are believed to have DSB, or at least a subset of DSB that participate in misrejoining (Radivoyevitch *et al.* 1998), as precursors. Because of proximity effects (Sachs *et al.* 1997a, Chen *et al.* 1997), clustering might enhance the fraction of DSB that undergo pairwise misrejoinings as opposed to restitution. Moreover, for a given number of pairwise DSB misrejoinings, the balance between cell killing and mutation is presumably different for a pattern where most misrejoinings occur within high-multiplicity, Mbp-scale clusters of DSB, compared to a pattern of randomly located misrejoinings. In the latter case one might expect more killing and expect fewer viable mutations, because asymmetric interchromosomal exchanges (i.e. dicentrics) are usually lethal clonogenically, while a sufficiently small asymmetric intrachromosomal exchange (i.e. a small deletion which forms a ring) may often leave the cell clonogenically viable (Durante *et al.* 1992, Cornforth and Bedford 1993).

DSB clustering along chromosomes must also markedly affect other size distributions—of mutations at a defined locus (assuming many of these arise from misrejoinings), and of ‘intrachanges’, i.e. intrachromosomal exchange-type aberrations such as rings or inversions (Wu *et al.* 1997, 1998). Using chromosome aberration data, and theoretical modelling of pairwise DSB misrejoining, we have previously argued the following four points (Sachs *et al.* 1997b): that intrachanges which are small by cytogenetic standards (i.e. 10 Mbp or less) are far more frequent compared to large intrachanges than randomness would predict; that the bias for extra small intrachanges is probably greater at high LET than at low LET; that these small intrachanges may well be

significant for carcinogenesis; and that they might serve, in retrospective biodosimetry, to help distinguish between low LET and high LET past exposure. The PFGE data and the present analysis somewhat strengthen these arguments by demonstrating LET dependent enhancement of the frequency of such small sizes even before DSB misrejoining; proximity effects, for the misrejoinings which produce intrachanges, would be expected to compound this enhancement.

5.4. Conclusions

PFGE data clearly show that after high LET irradiation DSB are clustered non-randomly along chromosomes (Löbrich *et al.* 1996, Kraxenberger 1996, Newman *et al.* 1997). Experiments and theory both suggest there may be some, albeit less pronounced, deviations from randomness even for X-rays (Löbrich *et al.* 1996, Newman *et al.* 1997, Michalik 1993, Moiseenko *et al.* 1996). To work out the implications of non-randomness for DSB yield estimates and for dose-response relations of fragment-size distributions requires modelling. The clustered-breakage formalism introduced here is a generalization of the standard random-breakage model, applicable even when a single primary radiation track can produce multiple DSB with *interdependent* genomic locations. The new formalism can facilitate systematic analysis of DNA fragment-size distributions for any radiation quality and for any chromosomal geometry. Applied to two data sets for mammalian cells, the formalism shows that a single track of $LET \approx 100 \text{ keV } \mu\text{m}^{-1}$ typically produces Mbp-scale DSB clusters of average multiplicity substantially larger than two. Such clustering means that if all the DSB in a Mbp-scale cluster are counted as a single DSB, which is the case for many current assays, the high LET DSB yield and RBE are considerably underestimated.

For $LET \approx 100 \text{ keV } \mu\text{m}^{-1}$ almost all clustering effects at doses of a few Gy or less are due to one-track action, with interlacing of different one-track clusters negligible at such doses. This means that the main use for the clustered-breakage formalism, apart from applications to the high-dose PFGE experiments themselves, is to try to extricate a one-track distribution from the multi-track effects observed in a high-dose PFGE experiment. Once the one-track distribution is known, it can be applied to much lower doses and to other endpoints. Extrapolations to lower doses using the formalism indicates that large-scale DSB clustering significantly affects major radiobiological endpoints that are thought to have DSB as

precursors—chromosome aberrations, cell killing, mutations, and carcinogenesis.

Acknowledgements

We are grateful to K. Prise and B. Rydberg for kindly supplying details on previously published data and for useful discussions. We thank A. Kellerer for pointing out to us the importance of length correlations among nearby fragments, and for several significant improvements in the analysis. We also thank D. Aldous, W. Holley, V. Moiseenko and T. Radivoyevitch for useful discussions and corrections. Research supported by NSF grants DMS-9532055 (PH) and BIR-963-0735 (LRH), and by NIH grants GM-57245 (RKS), RR-11623, and ES07361 (DJB).

Appendix A: mathematical details

Preview

This Appendix formally defines, discusses, and illustrates the clustered breakage model; the formulae of § 3.1 will be derived.

DSB locations on a chromosome can be specified as points on an x axis, where x represents size measured in Mbp (figure 8). Such a picture corresponds to mathematical models called random point processes (Wolff 1988, Thompson 1988). The points represent probabilistic DSB locations along the chromatin contour after uncorrelated tracks with stochastic energy depositions hit (or miss) a chromosome that itself can have many different geometric configurations. Neglecting the telomere effects defined in § 2.3, DNA fragment sizes are determined by the intervals from one DSB to the next (figure 8). The formalism can take correlations among the sizes of nearby fragments into account. The situation where many tracks traverse a cell nucleus is modelled by superimposing random one-track DSB patterns.

For a cell nucleus traversed by a specified number of tracks the random point process representing DSB locations will here be taken as *stationary*, which very markedly simplifies the analysis (Thompson 1988). Stationarity involves analysing DSB along the entire non-negative axis $0 \leq x < \infty$, with the stochastic pattern of fragment sizes invariant under translations to the right (Wolff 1988, Thompson 1988); a chromosome can then be regarded as a typical segment, of the appropriate size, along the axis (figure 9). However, considering a particular chromosome, of a particular size, is needed only if the telomere effects defined in § 2.3 must be estimated (Appendix B). For the time being, in Appendix A, we neglect telomere



Figure 9. DSB for one-track action modelled as a stationary point process. One segment of the region $x \geq 0$, is shown, with 19 DSB scattered stochastically along this particular segment. $j=3$ labels a single DSB; $j=4-9$ label 6 DSB in a tight cluster; $j=10-12$ a cluster of 3; etc. S_j is the size between the j th DSB and the $(j+1)$ th. The origin $x=0$ is off the picture to the left; the stationary point process approach considers the picture as extending indefinitely to the right. Neglecting telomere effects, the DNA fragment-size pattern can be calculated just from the DSB pattern along the x axis: one can compute the average number of DSB per genome, the average multiplicity of a tight cluster, the probability a track misses a chromosome altogether, etc. However, to calculate telomere effects one must imagine a chromosome put into the picture. AB shows one possible location for a chromosome, with one telomere at A, the other telomere at B, and 3 DSB. CD and EF are other possible locations for the same chromosome, with no DSB on the chromosome in the case CD. This construction, where one puts in the DSB first and either omits telomeres entirely or puts them in later, was originally suggested by Radivoyevitch (1997). Any one location for the chromosome is acceptable as long as we allow the DSB pattern to fluctuate appropriately, i.e. consider the full ensemble of DSB patterns. Equivalently, and somewhat more vividly, one can consider a single DSB pattern such as the one shown and regard all chromosome locations as equally probable. Then the technique is to take the left telomere of a chromosome as a Poisson arrival process, appealing to the stationarity of the point process, to ergodicity and to theorems on PASTA—‘Poisson Arrivals See Time Averages’ (Wolff 1988). The picture shown here is indicative of one-track action; for a dose of, say, 100 Gy, a corresponding picture would hold, with the number of DSB about 1000 times as great (compare figure 8). For a dose of 100 Gy, telomere effects are negligible for all fragment sizes of interest, and the probability for fragment sizes on the x -axis that are comparable to or larger than a chromosome can be neglected. On the other hand, in the one-track pattern shown here, telomere effects are negligible only for the smaller fragment sizes—sizes so small the dominant contribution comes from within the tight clusters—and there is a significant probability of sizes along the axis much larger than any chromosome (such as S_1). The implications of these qualitative differences when extrapolating from high to low doses are discussed at the end of Appendix A.

effects, focusing on the statistical properties of chromatin fragments that have DSB on both ends, and analysing the entire region $0 \leq x < \infty$.

The formalism is defined in the next section, and the derivations of the basic formulae are then given. Specific examples of stationary point processes are then given, and then we derive the random-breakage distribution as a limit of a clustered-breakage distribu-

tion. Extrapolations to low doses are discussed in the final subsection.

DSB locations and stationary point processes

For one-track action, consider DSB located at various values of $x \geq 0$ and labelled from left to right with j (where $j = 1, 2, \dots$). It will be assumed that the DSB locations \mathbf{X}_j form a random point process which is 'orderly', i.e. (Wolff 1988) each \mathbf{X}_j is a non-negative random variable and $\mathbf{X}_{j+1} \geq \mathbf{X}_j$, where inequality holds a.s. Here the increment $\mathbf{S}_j = \mathbf{X}_{j+1} - \mathbf{X}_j$ represents the size (measured in Mbp) of the chromatin fragment between the DSB labelled with j and $j+1$ (figure 9).

We further assume the point process is stationary (Wolff 1988, Thompson 1988). Stationarity involves translational invariance for the DNA fragment size pattern. For example, with $x_0 \geq 0$ stationarity implies that the conditional probability no DSB occurs within the interval $(x_0, x_0 + x]$, given that there is a DSB at x_0 , is independent of x_0 ; this probability is $F_1^*(x)$. The stationarity assumption does *not* imply that \mathbf{S}_j and $\mathbf{S}_{j'}$ are independent for $j \neq j'$. For example if the size \mathbf{S}_7 of the seventh fragment is below average, this signals that the 7th and 8th DSB, bounding the fragment, may well be near the radiation track core, and then there should be a bias for \mathbf{S}_8 to be below average as well. Thus for one-track action one may expect any two consecutive fragment sizes, such as \mathbf{S}_7 and \mathbf{S}_8 , to be positively correlated. The clustered-breakage formalism allows such correlations and also allows more complicated interdependencies.

Deriving the clustered-breakage formalism

The assumptions will now be used to prove the basic equations of the clustered-breakage formalism, equations (7)–(12).

Assuming single-track action, and neglecting the telomere effects defined in §2.3, the fraction of fragments in a small size range dx is $f_1(x) dx$, where $f_1 = -dF_1^*/dx$. Consequently the number-average fragment size is the integral:

$$\int_0^\infty x f_1(x) dx = \int_0^\infty F_1^* dx. \quad (18)$$

Here integration by parts was used. Equation (18) implies (Thompson 1988) the first basic equation of the clustered-breakage formalism, equation (7), giving the average one-track number of DSB per unit size, N_1 , as the inverse of the integral in equation (18). Mathematically speaking, N_1 is what is usually called the 'intensity' or 'arrival rate' of the stationary point process, generalizing κ in the expo-

nential distribution $\exp(-\kappa x)$, equation (4), for random breakage.

Our second basic equation involves the size from a random chromatin point to the next DSB. The distribution can be found by conditioning on the size of the segment \mathbf{S}_j in which the random point is located and computing the probability $E_1(x)$ that the next DSB is at incremental size x or less. The probability of the point being located in a segment having size between x' and $x' + dx'$ is

$$N_1 x' f_1(x') dx'. \quad (19)$$

Given that the random point is in a segment of size x' , then either $x' \leq x$ and the next DSB is certainly within x ; or $x' > x$ and then the probability for the next DSB to be within x is x/x' . Thus, conditioning on x' gives:

$$\begin{aligned} E_1(x) &= N_1 \left[\int_0^x x' f_1(x') dx' + \int_x^\infty \frac{x}{x'} x' f_1(x') dx' \right] \\ &= N_1 \int_0^x F_1^*(x') dx', \end{aligned} \quad (20)$$

where the relation $f_1 = -dF_1^*/dx$ and integration by parts were used in the calculation. Equation (20) is the second main equation of the clustered-breakage formalism (§3.1). It is a special case of a general equation called Palm's formula (Thompson 1988).

To derive the other, multi-track, equations of the clustered-breakage formalism, we assume that if a cell nucleus is traversed by j tracks, each track independently gives rise to a stationary process, placing DSB along the non-negative x axis. This assumption of independence involves an approximation. For example, suppose one chromosome in one cell at one instant is in a particularly vulnerable conformation. Then there is an enhanced probability for a track striking the cell nucleus to make extra DSB in the chromosome. Such fluctuations are, at least in principle, appropriately included in the one-track pattern described by F_1 . But now suppose a second track independently traverses the cell nucleus. It will likewise tend to make extra DSB on the same chromosome. Thus in general there are some correlations among the DSB patterns made by different tracks, even though tracks are independent, because different tracks act on the same chromosomes; these particular correlations are neglected in what follows.

By assumption, the probability that exactly j tracks traverse the nucleus is

$$\exp(-n) n^j / j!. \quad (21)$$

The intensity of j independent stationary point processes superimposed is just j times the intensity, N_1 , of one process, so averaging over j using equation (21)

gives nN_1 as the average number of DSB per unit length, i.e. equation (9) holds.

For the next result, suppose, a specified number of tracks, j , traverse a cell nucleus and suppose an arbitrary point x_0 on the x axis is chosen. By the definition of $E_1(x)$, the probability $P_1(x)$ that a particular track will not make any DSB within the interval from x_0 to $x_0 + x$ is $P_1(x) = 1 - E_1(x)$ (independent of x_0), so the probability that there are no DSB in this stretch is $[1 - E_1(x)]^j$; here $j=0$ is allowed. Conditioning on j , the probability $P(x)$ of no DSB in a segment of size x and the probability $E(x)$ for at least one DSB are:

$$\begin{aligned} P(x) &= \exp(-n) \sum_{j=0}^{\infty} [1 - E_1(x)]^j n^j / j! \\ &= \exp(-n) \exp[n - nE_1(x)] \\ &= \exp[-nE_1(x)]; \quad E = 1 - P. \end{aligned} \quad (22)$$

$P = \exp(-nE_1)$ is another key equation of clustered-breakage theory (§3.1). As consistency checks, note the following: (1) for x very large $\exp[-nE_1(x)]$ approaches, as it should, the probability $\exp(-n)$ that no tracks traverse a cell nucleus; (2) for x very small, $E_1 \approx N_1 x \ll 1$ so $P \approx 1 - nN_1 x$, the correct limiting form when different tracks act independently; (3) for n very large, P is different from zero only when x is so small that $E_1 \ll 1$, in which case P has approximately the random-breakage form $P = \exp[-nN_1 x]$ (i.e. when patterns from sufficiently many independent tracks interlace the result is simply random breakage with very many small pieces, even if the pattern made by any one track is highly clustered); and (4) for n very small, $P \approx 1 - nE_1$, which is the correct limiting result for a situation where at most one track traverses a cell nucleus, the probability of one traversal being $n \ll 1$.

For the next key equation, equation (11), consider a situation where exactly $j \geq 1$ tracks traverse a cell nucleus and consider a DSB from one track located at x_0 ; the probability that a following segment, of length x , contains no DSB from the same track is $F_1^*(x)$, and the probability that it contains no DSB from any of the $j-1$ other tracks is $[1 - E_1(x)]^{j-1}$, so one has

$$F_j^*(x) = F_1^*(x) [1 - E_1(x)]^{j-1}, \quad (j \geq 1). \quad (23)$$

Here F_j^* is the complementary distribution for the size of a fragment between two DSB when j tracks traverse a cell nucleus and telomere effects are neglected. The probability that a given DSB came from a cell nucleus traversed by exactly j tracks is weighted by j , thus proportional to $j \exp(-n) n^j / j!$

or, with the proper normalization, is

$$(j/n) \exp(-n) n^j / j!. \quad (24)$$

Combining equation (24) with equation (23) for F_j^* , the distribution $F^*(x)$ for fragment sizes is given by

$$\begin{aligned} F^*(x) &= \sum_{j=1}^{\infty} \frac{j}{n} \exp(-n) n^j F_1^*(x) [1 - E_1(x)]^{j-1} \\ &= F_1^*(x) \exp[-nE_1(x)]. \end{aligned} \quad (25)$$

This is another main result of the clustered-breakage formalism, equation (11).

To obtain the last main result of §3.1, note that the fraction of DNA in size range $S_1 \leq x \leq S_2$ for those cells traversed at least once is proportional to

$$\begin{aligned} \int_{S_1}^{S_2} x f(x) dx &= - \int_{S_1}^{S_2} x (dF^*/dx) dx \\ &= [-x F^*]_{S_1}^{S_2} + \int_{S_1}^{S_2} F^* dx \\ &= \{ -\exp(-nE_1) [x F_1^* + (1/nN_1)] \}_{S_1}^{S_2}. \end{aligned} \quad (26)$$

Here $(dE_1/dx) = N_1 F_1^*$ was used to integrate F^* explicitly. Taking $S_2 = \infty$ and $S_1 = 0$ in equation (26) gives as the average fragment length among cells traversed at least once

$$\frac{[1 - \exp(-n)]}{nN_1}. \quad (27)$$

As a consistency check, note that for $n \ll 1$ equation (27) reduces to the one-track result with telomere effects neglected, namely $1/N_1$, as it should. From equation (27) and the fact that the probability for no tracks striking a nucleus is $\exp(-n)$, equation (26) gives the last main result, equation (12).

There are useful parallelisms between the one-track and multiple-track equations, summarized in the following identities: $f_1 = -dF_1^*/dx$, $f = -dF^*/dx$; $F_1^* = (1/N_1) dE_1/dx$, $F^* = (1/nN_1) dE/dx$; $P_1 = 1 - E_1$, $P = 1 - E$.

Examples

We give some more explicit examples of one-track stationary point processes which could model DSB patterns. Each such point process determines a corresponding distribution F_1^* , but the same F_1^* can correspond to various one-track stationary point processes, differing in the way lengths for nearby fragments are correlated. Thus F_1^* carries only part of the information that can be obtained from a fully specified one-track stationary point process. As described in the section above, the multi-track pro-

cesses are determined once the one-track process is given.

- (1) **Renewal Processes.** Suppose a complementary distribution F_i^* is given. Let the first DSB location, \mathbf{X}_1 , be chosen according to the associated distribution E_1 , i.e. $P = E_1(x)$ where P denotes probability and E_1 is defined by equation (20). Let each fragment size \mathbf{S}_j be chosen according to F_i^* , i.e. $P = F_i^*(x)$. Locate the remaining DSB appropriately, i.e. $\mathbf{X}_{j+1} = \mathbf{X}_j + \mathbf{S}_j$, $j = 1, 2, \dots$. Then $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots$ forms a stationary point process, the *renewal* process with complementary distribution F_i^* (Wolff 1988). The distinctive feature here is that there are no correlations between the sizes of successive fragments; each size is chosen at random from the same distribution, without regard to previous choices.
- (2) **Poisson Processes.** In the special case $F_i^* = \exp(-\kappa x)$ for some constant $\kappa > 0$, the renewal process is a Poisson process (Wolff 1988). In this case one-track DSB are scattered at random throughout the genome.
- (3) **Processes with Markov-Related Fragments.** Suppose we have a set of fragment lengths $S_\alpha > 0$, ordered according to increasing size. For simplicity take the index α to be from a finite set, $\alpha = 1, 2, \dots, M$ (generalizations are straightforward). Let $T_{\alpha\beta}$ be the transition probabilities of an irreducible, time-reversible Markov chain with equilibrium t^α , i.e. (Wolff 1988):

$$\left. \begin{aligned} \sum_{\beta=1}^M T_{\alpha\beta} &= 1, & \sum_{\beta=1}^M T_{\alpha\beta} t_\beta &= t^\alpha, \\ \sum_{\alpha=1}^M t^\alpha &= 1, & T_{\alpha\beta} t^\alpha &= T_{\beta\alpha} t_\beta, \end{aligned} \right\} \quad (28)$$

where $T_{\alpha\beta}$ is non-negative and t^α is positive. Let the first DSB location \mathbf{X}_1 be chosen as follows. The probability that the origin $x=0$ is in a segment of length S_α is taken as $s^\alpha = S_\alpha t^\alpha / \sum S_\beta t_\beta$. The density for the location of the origin in such a stretch is taken as uniform. These conditions determine \mathbf{X}_1 . To determine the remaining \mathbf{X}_i run the Markov chain with initial state s^α , and choose fragment sizes according to the length of the state α_m obtained after m steps, $m = 1, 2, 3, \dots$. This procedure determines a stationary point process. The distinctive feature is that correlations are allowed between successive fragments (a pair of fragments, determined by 3 DSB), the probabilities for the lengths of a particular fragment being uniquely determined by the length of its predecessor to the left. One can form the corresponding F_i^* , i.e.

$F_i^*(x) = \sum_{\alpha=1}^{M'} t^\alpha$, where M' is the largest value of α for which $S_\alpha > x$.

- (4) **Cluster Processes.** Consider any stationary point process \mathbf{X}_i , for example a Poisson process (2) above. With each \mathbf{X}_i associate a cluster of DSB according to some probabilistic laws for the multiplicity of the cluster and the location of its points, with \mathbf{X}_i on average at the centre of its cluster, but the cluster properties otherwise independent of i . This gives a stationary point process, usually called a *cluster process*. The intuitive idea is that the \mathbf{X}_i are rather far apart, with a tight cluster around each. This example is a tempting one, since the experiments do suggest tight clusters whose centres follow approximately the random-breakage model.

Specializing the clustered-breakage formalism to obtain the random-breakage model

To prove that the random-breakage model is a limiting special case of the clustered-breakage formalism, consider the renewal model with $F_i^* = \exp(-\lambda x)$, which implies $N_1 = \lambda$ by equation (7) and implies $E_1 = 1 - F_i^*$ by equation (8). Define $\kappa = n\lambda$ and consider the limit $n \rightarrow \infty$, $\lambda \rightarrow 0$, with $n\lambda$ fixed as $n\lambda = \kappa$. Taking these limits in the three basic multi-track equations of the clustered-breakage formalism, equations (10)–(12), gives directly three basic equations of the random-breakage model, equations (2), (4) and (6), since $nE_1(x) \rightarrow \kappa x$ in this limit.

High and low doses

We make some remarks on our extrapolation of high dose results to one-track estimates.

Formally, the one-track fragment-size complementary distribution F_i^* describes some segments which are larger than any actual chromosome, such as the segment in figure 9 from DSB 9 to DSB 10 or especially the segment from 12 to 13; that is, in general the DSB probability density $f_1(x) = -dF_i^*/dx$ will be non-zero even for x as large as the size between DSB 12 and DSB 13. An example is given by equation (14). Such large segments are never expressed in experiments, the maximum size being the size of the largest chromosome, i.e. neglecting telomere effects is certainly invalid when considering such large sizes.

For high doses, the average number of DSB per chromosome is much larger than the number (namely 2) of telomeres per chromosome, telomere effects are negligible, and questions about large fragments do not arise; but for one-track action the opposite is true, with the average number of DSB per chromo-

some (including undamaged chromosomes) considerably less than 2. Then the presence of large sizes in a pattern such as that shown in figure 9 insures that the large-scale statistical properties, such as the probability of a track missing a chromosome or the average number of DSB per chromosome, are the appropriate ones. Such information, together with the information given by F_i^* on the frequency of small sizes, suffices for many applications when extrapolating from high to low doses. However, if telomere effects must be considered explicitly, then modelling involves using a specific stationary point process, which not only determines a distribution F_i^* but also gives additional information (see Examples section). The next Appendix illustrates this procedure.

Appendix B: DSB clustering and telomere effects in renewal approximation

Renewal approximation

To calculate telomere effects or details on DSB clustering requires more information than just a complementary function F_i^* . We now analyse renewal models, which can supply the extra information given any F_i^* ; in general an extra approximation or assumption is then involved, neglecting certain (but not all) correlations among DSB locations.

In the arguments of Appendix A, any interdependence of nearby fragment sizes \mathbf{S}_i was allowed. Simplifications occur if one has a Markov chain, where \mathbf{S}_i depends only on \mathbf{S}_{i-1} , or if one specializes still more to a renewal picture, where the \mathbf{S}_i are all mutually independent (Examples section). Given any F_i^* , even one generated by a point process with correlations between nearby \mathbf{S}_i , there is a corresponding renewal process which has the same F_i^* (Examples section). In this renewal process, correlations between DSB locations (i.e. clusters) are still described but additional correlations between nearby sizes are neglected. The renewal process can be used to approximate the given process. Intuitively speaking, the approximation is that every time one track makes one DSB on one chromosome, the statistical situation 'renews' itself, 'forgetting' about other nearby fragment sizes \mathbf{S}_i . If one visualizes a single-track cluster of multiplicity M as being formed by selecting $M-1$ chromatin segments from a distribution and laying the segments end to end, the renewal selection rule is that previous selections do not influence the current selection. In figure 1 (A), if the persistence length (Cantor and Schimmel 1980) of the random coil is much less than the average size of a fragment, renewal approximation would hold for this 2-dimensional, schematic picture.

In the present context, renewal approximation presumably underestimates clustering, for the following reason: for one-track action, there should be a positive correlation between successive fragment sizes, with small sizes grouped together close to the track centre and large sizes grouped together for locations far from the track. Neglecting this positive correlation, as renewal approximation requires, means allowing more alternation of large sizes and small sizes, i.e. means underestimating the degree of clustering. The opposite behaviour, of negative correlations and renewal approximation overestimating the clustering, could occur only in conjunction with systematic, non-random chromatin structure. At scales smaller than those emphasized in this paper such behaviour may well occur (Holley and Chatterjee 1996, Rydberg 1996). For example, systematic, rigid chromatin loops might lead to a pattern of small doublets (clusters of multiplicity 2) with different doublets separated by large sizes, and this pattern could involve negative correlations between successive fragment sizes. Current pictures of large-scale chromatin structure involve considerable randomness (Sachs *et al.* 1995, Dernburg *et al.* 1996), suggesting that, at the size scales considered in this paper, negative correlations are not present and renewal approximation therefore underestimates clustering, rather than overestimating it.

Clustering parameters

Renewal approximation can be used to specify clustering properties numerically. The method will now be illustrated with the data of Löbrich *et al.*, using the particular definitions of clusters given in §3.2 and neglecting telomere effects. The average multiplicity of a one-track cluster and the variance of the multiplicity can be computed as follows. The probability a DNA fragment has size less than $S_0 = 1.12$ Mbp is $p = 1 - F_1^*(S_0) \approx 0.758$, where the numerical value uses the parameter values obtained above for the clustered-breakage model. With a chromosome straightened out along the x -axis, consider a cluster and consider the DSB at the left end of the cluster. The probability that the cluster has multiplicity one is $q = 1 - p$, i.e. is the probability the first DNA fragment is larger than S_0 . The probability for this first fragment to be, instead, smaller than S_0 is p , and the probability for the second fragment to be larger than S_0 is q , so that there is a probability pq for a cluster of multiplicity two; here we have used renewal approximation by assuming that the probabilities for the first fragment and the second fragment are independent of each other. Continuing this argument, the probability the cluster has multipli-

city j is $p^{j-1}q$. Consequently the average multiplicity, which we shall denote by M_1 , is

$$M_1 = \sum_{j=1}^{\infty} j p^{j-1} q = q/(1-p)^2 = 1/q. \quad (29)$$

Similarly the variance V_1 in multiplicity is

$$V_1 = \sum_{j=1}^{\infty} j^2 p^{j-1} q - M_1^2 = p/q^2. \quad (30)$$

These equations were used in the main text when discussing average cluster multiplicity and its variance. To get corresponding results for cluster size, we first note that the number of fragments in a cluster is $M-1$, where M is a random positive integer (Wolff 1988), with mean M_1 and variance V_1 specified above. Consequently, the cluster size, regarded as a random variable S has the form

$$S = S_1 + S_2 + \dots + S_{M-1}, \quad (31)$$

where the various S_i are independent, identically distributed random variables, representing fragment sizes, as discussed in Appendix A. The average cluster size S , which is the mean of S , and the variance V_s of S are consequently given by (Wolff 1988)

$$S = M_1 X, \quad V_s = V_1 X^2 + V_x (M_1 - 1), \quad (32)$$

where X and V_x are the mean and variance of S_1 , conditioned on S_1 being less than S_0 . With $f_1(x) = -dF_1^*/dx$ and $h = (S_0/s_w)^{0.5}$ as before, the explicit expressions for X and V_x are given by

$$X = \frac{\int_0^{s_0} x f_1(x) dx}{\int_0^{s_0} f_1(x) dx} = (1/p) \int_0^{s_0} x f_1(x) dx \\ = (s_w/p) [2 - e^{-h} (h^2 + 2h + 2)], \quad (33)$$

$$p(V_x + X^2) = \int_0^{s_0} x^2 f_1(x) dx \\ = s_w^2 [24 - e^{-h} (h^4 + 4h^3 + 12h^2 + 24h + 24)]. \quad (34)$$

Monte Carlo methods for calculating telomere effects

Renewal approximations can also be used to generate sample clustering patterns, with telomere effects included, as follows. The first step is to pick a size for a chromosome, say S_C . Next the computer chooses a number j of tracks which traverse the cell nucleus in question, using a Poisson distribution of average n and a random number generator. If $j=0$ (which can occur at low doses though it is highly improbable at 100 Gy), the nucleus is not traversed so the chromosome is not hit, and there is just one fragment, having the chromosome size S_C .

If $j>0$ the next step is to consider one track and consider either of the two chromosome telomeres.

The probability that the track makes a DSB within size x of the telomere is governed by the distribution $E_1(x)$ (Appendix A), and Monte Carlo simulation gives the relevant value of x . In most simulations the track will miss altogether (i.e. the simulation will give $x > S_C$). But if the first track does hit, the distribution for its next DSB, if any, and (by the renewal assumption) for all subsequent DSB, if any, made by this one track on this chromosome is given by F_1^* . F_1^* is then used to compute additional fragment sizes until finally one DSB falls outside the other telomere, i.e. at total size $> S_C$. The locations of the DSB for this track are then tabulated. One next iterates on the track number. When all j tracks have been considered, the DSB locations are ordered, giving the simulated DSB clustering pattern on the chromosome.

Iterating for many chromosomes gives statistics for all desired estimates, with all telomere effects properly included. In addition, each chromosome simulated in this way is a representative sample, e.g. figure 8. Looking at a number of chromosomes gives clues on general patterns of DSB clustering.

References

- Andreev, S. G., Khvostunov, I. K., Spitkovsky, D. M., and CHEPEL, V., 1997, Clustering of DNA breaks in chromatin fibre: dependence on radiation quality. In *Microdosimetry: An Interdisciplinary Approach*, edited by D. T. Goodhead, P. O'Neill and H. G. Menzel (Cambridge, UK: Royal Society of Chemistry), pp. 133–136.
- BLÖCHER, D., 1990, In Chef electrophoresis a linear induction of dsb corresponds to a nonlinear fraction of extracted DNA with dose. *International Journal of Radiation Biology*, **57**, 7–12.
- BRAHME, A., RYDBERG, B., and BLOMQUIST, P., 1997, Dual spatially correlated nucleosomal double strand breaks in cell inactivation. In *Microdosimetry: An Interdisciplinary Approach*, edited by D. T. Goodhead, P. O'Neill and H. G. Menzel (Cambridge, UK: Royal Society of Chemistry), pp. 125–128.
- BRENNER, D. J., 1990, Track structure, lesion development, and cell survival. *Radiation Research*, **124**, S29–37.
- BRENNER, D. J., and WARD, J. F., 1992, Constraints on energy deposition and target size of multiply damaged sites associated with DNA double-strand breaks. *International Journal of Radiation Biology*, **61**, 737–48.
- CANTOR, C. R., and SCHIMMEL, P. R., 1980, *The Behavior of Biological Macromolecules* (San Francisco: W.H. Freeman).
- CEDERVALL, B., WONG, R., ALBRIGHT, N., DYNLACHT, J., LAMBIN, P., and DEWEY, W. C., 1995, Methods for the quantification of DNA double-strand breaks determined from the distribution of DNA fragment sizes measured by pulsed-field gel electrophoresis. *Radiation Research*, **143**, 8–16.
- CHATTERJEE, A., and HOLLEY, W. R., 1991, Energy deposition mechanisms and biochemical aspects of DNA strand breaks by ionizing radiation. *International Journal of Quantum Chemistry*, **391**, 709–727.

- CHEN, A. M., LUCAS, J. N., SIMPSON, P. J., GRIFFIN, C. S., SAVAGE, J. R. K., BRENNER, D. J., HLATKY, L. R., and SACHS, R. K., 1997, Computer simulation of fish data on chromosome aberrations produced by X-rays or alpha particles. *Radiation Research*, **148**, S93–S101.
- COOK, V. E., and MORTIMER, R. K., 1991, A quantitative model of DNA fragments generated by ionizing radiation, and possible experimental applications. *Radiation Research*, **125**, 102–106.
- CORNFORTH, M. N., and BEDFORD, J., 1993, Ionizing radiation damage and its early development in chromosomes. In *Advances in Radiation Biology*, vol. 17, *DNA and chromatin damage caused by radiation*, edited by J. T. Lett and W. K. Sinclair (San Diego: Academic press), pp. 423–496.
- CREMER, T., KURZ, A., ZIRBEL, R., DIETZEL, S., RINKE, B., SCHROCK, E., SPEICHER, M. R., MATHIEU, U., JAUCH, A., EMMERICH, P., SCHERTAN, H., RIED, T., CREMER, C., and LICHTER, P., 1993, Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harbor Symposia on Quantitative Biology*, **58**, 777–792.
- CREMER, C., MUENKEL, C., DIETZEL, S., EILS, R., GRANZOW, M., BORNFLETH, H., JAUCH, A., ZINK, D., LANGOWSKI, T., and CREMER, T., 1997, Nuclear architecture and its role in radiation-induced aberrations. In *Microdosimetry: An Interdisciplinary Approach*, edited by D. T. Goodhead, P. O'Neill and H. G. Menzel (Cambridge, UK: Royal Society of Chemistry), pp. 143–151.
- DEWEY, W. C., and ALBRIGHT, N., 1997, Developing a model of DNA replication to be used for Monte Carlo calculations that predict the sizes and shapes of molecules resulting from DNA double-strand breaks induced by X irradiation during DNA synthesis. *Radiation Research*, **148**, 421–434.
- DERNBURG, A. F., BROMAN, K. W., FUNG, J. C., MARSHALL, W. F., PHILIPS, J., AGARD, D. A., and SEDAT, J. W., 1996, Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell*, **85**, 745–59.
- DURANTE, M., GROSSI, G. F., NAPOLITANO, M., PUGLIESE, M., and GIALANELLA, G., 1992, Chromosome damage induced by high-Let alpha-particles in plateau-phase C3H 10T1/2 cells. *International Journal of Radiation Biology*, **62**, 571–80.
- FRIEDL, A. A., KRAXENBERGER, A., and ECKARDT-SCHUPP, F., 1995, An electrophoretic approach to the assessment of the spatial distribution of DNA double-strand breaks in mammalian cells. *Electrophoresis*, **16**, 1865–74.
- FRIEDLAND, W., JACOB, P., PARETZKE, H. G., PERZL, M., and STORK, T., 1997, Simulation of strand breaks and short DNA fragments in the biophysical model Partrac. In *Microdosimetry: An Interdisciplinary Approach*, edited by D. T. Goodhead, P. O'Neill and H. G. Menzel (Cambridge, UK: Royal Society of Chemistry), pp. 43–46.
- GOODHEAD, D. T., 1985, Relationship of microdosimetric techniques to applications in biological systems. In *The Dosimetry of Ionizing Radiation*, vol. II, edited by K. Kase, B. Bjarngard and F. Attix (Orlando, FL: Academic Press), pp. 1–89.
- HAHNFELDT, P., HEARST, J. E., BRENNER, D. J., SACHS, R. K., and HLATKY, L. R. 1993, Polymer models for interphase chromosomes. *Proceedings National Academy Sciences US*, **90**, 7854–7858.
- HOLLEY, W. R., and CHATTERJEE, A., 1996, Clusters of DNA induced by ionizing radiation: formation of short DNA fragments. I. Theoretical modeling. *Radiation Research*, **145**, 188–99.
- KENDALL, M. G., and MORAN, P. A. P., 1963, *Geometrical Probability* (London: C. Griffin).
- KOENIG, D., and SCHMIDT, V., 1992, *Zufällige Punktprozesse* (Stuttgart Germany: B. G. Teubner).
- KRAXENBERGER, A., 1996, Verbesserungen der Methoden zur Pulsfeld-Gelelektrophorese und Anwendung der Technik auf die Analyse von DNA-Schaden durch dicht ionisierende Strahlung. Thesis, Technische Universität München, Munich, Germany, pp. 74–91.
- KRAXENBERGER, A., FRIEDL, A. A., and KELLERER, A. M., 1994, Computer simulation of pulsed field gel runs allows the quantitation of radiation-induced double-strand breaks in yeast. *Electrophoresis*, **15**, 128–36.
- LEA, D. E., 1955, *Actions of Radiations on Living Cells*, second edition (Cambridge: Cambridge University Press).
- LOBRICH, M., COOPER, P. K., and RYDBERG, B., 1996, Non-random distribution of DNA double-strand breaks induced by particle irradiation. *International Journal of Radiation Biology*, **70**, 493–503.
- MARSHALL, W. F., FUNG, J. C., and SEDAT, J. W., 1997, DECONSTRUCTING THE NUCLEUS: GLOBAL ARCHITECTURE FROM LOCAL INTERACTIONS. *Current Opinion in Genetics and Development*, **7**, 259–63.
- MICHALIK, V., 1993, Energy deposition clusters in nanometer regions of charged-particle tracks. *Radiation Research*, **134**, 265–70.
- MOISEENKO, V. V., EDWARDS, A. A., and NIKJOO, N., 1996, Modelling the kinetics of chromosome exchange formation in human cells exposed to ionising radiation. *Radiation and Environmental Biophysics*, **35**, 31–5.
- MOISEENKO, V. V., EDWARDS, A. A., NIKJOO, H., and PRESTWICH, W. V. 1997, Modeling of chromosome exchanges in human lymphocytes exposed to radiations of different quality. In *Microdosimetry: An Interdisciplinary Approach*, edited by D. T. Goodhead, P. O'Neill and H. G. Menzel (Cambridge, UK: Royal Society of Chemistry), pp. 152–155.
- NEWMAN, H. C., PRISE, K. M., FOLKARD, M., and MICHAEL, B. D., 1997, DNA double-strand break distributions in X-ray and alpha-particle irradiated V79 cells: evidence for non-random breakage. *International Journal of Radiation Biology*, **71**, 347–63.
- OSTASHEVSKY, J. Y., and LANGE, C. S., 1994, The 30 nm chromatin fiber as a flexible polymer. *Journal of Biomolecular Structure and Dynamics*, **11**, 813–20.
- OTTOLENGHI, A., MERZAGORA, M., and PARETZKE, H. G., 1997, DNA complex lesions induced by protons and alpha-particles: track structure characteristics determining linear energy transfer and particle type dependence. *Radiation and Environmental Biophysics*, **36**, 97–103.
- PRISE, K. M., 1997, Ionising radiation induced clustered damage to DNA—A review of the experimental evidence. In *Microdosimetry: An Interdisciplinary Approach*, edited by D. T. Goodhead, P. O'Neill and H. G. Menzel (Cambridge, UK: Royal Society of Chemistry), pp. 111–116.
- RADIVOEVITCH, T., 1997, Mathematical models of misrejoining DNA double-strand breaks and DNA fragment size distributions. PhD thesis, Medical University of South Carolina, USA.
- RADIVOEVITCH, T., HOEL, D. G., HAHNFELDT, P. J., RYDBERG, B., CHEN, A. M., and SACHS, R. K., 1998, Double-strand break misrejoining after X-irradiation: relating moderate to very high doses by a Markov model. *Radiation Research*, **149**, 52–58.
- RADIVOEVITCH, T., and CEDERVALL, B., 1996, Mathematical analysis of DNA fragment distribution models used with

- pulsed-field gel electrophoresis for DNA double-strand break calculations. *Electrophoresis*, **17**, 1087–93.
- RYDBERG, B., 1996, Clusters of DNA damage induced by ionizing radiation: formation of short DNA fragments. II. Experimental detection. *Radiation Research*, **145**, 200–209.
- RYDBERG, B., LÖBRICH, M., and COOPER, P. K., 1994, DNA double strand breaks induced by high-energy neon and iron ions in human fibroblasts. I. Pulsed field gel electrophoresis method. *Radiation Research*, **139**, 133–141.
- SACHS, R. K., VAN DEN ENGH, G., TRASK, B., YOKOTA, H., and HEARST, J. E., 1995, A random-walk/giant-loop model for interphase chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 2710–2714.
- SACHS, R. K., CHEN, A. M., and BRENNER, D. J., 1997 a, Review: proximity effects in the production of chromosome aberrations by ionizing radiation. *International Journal of Radiation Biology*, **71**, 1–19.
- SACHS, R. K., BRENNER, D. J., CHEN, A. M., HAHNFELDT, P., and HLATKY, L. R., 1997 b, Intra-arm and inter-arm chromosome intrachanges: tools for probing chromatin geometry and dynamics. *Radiation Research*, **148**, 330–339.
- THOMPSON, W. A., 1988, *Point Process Models with Applications to Safety and Reliability* (New York: Chapman and Hall).
- WARD, J. F., 1994, The complexity of DNA damage: relevance to biological consequences. *International Journal of Radiation Biology*, **66**, 427–32.
- WOLFF, R. W., 1988, *Stochastic Modeling and the Theory of Queues* (Englewood Cliffs, NJ: Prentice Hall).
- WU, H., DURANTE, M., SACHS, R. K., and YANG, T. C., 1997, Centric rings, acentric rings and excess acentric fragments based on a random-walk interphase chromosome model. *International Journal of Radiation Biology*, **71**, 487–96.
- WU, H., SACHS, R. K., and YANG, T., 1998, Radiation-induced total-deletion mutations in the human hprt gene: A biophysical model based on random walk interphase chromatin geometry. *International Journal of Radiation Biology*, **73**, 149–156.