Genomic Locations of Radiation-Produced DNA Double Strand Breaks: A Stochastic Cluster Process Formalism

Rainer K. Sachs[‡], Artem L. Ponomarev[‡], Philip Hahnfeldt[†], and Lynn R. Hlatky[†]

[‡]Department of Mathematics, University of California, Berkeley, CA 94720;

[†]JCRT, DFCI, Harvard Medical School, Boston, MA 02115.

Abstract

Ionizing radiation produces DSBs (DNA double strand breaks) in chromosomes. For densely ionizing radiation, the DSBs are not spaced randomly along a chromosome: recent data for size distributions of DNA fragments indicate break clustering on kbp-Mbp scales. Different DSB clusters on a chromosome are typically made by different, statistically independent, stochastically structured radiation tracks, and the average number of tracks involved can be small. We therefore model DSB positions along a chromosome as a stationary Poisson cluster process, i.e. a stochastic process consisting of secondary point processes whose locations are determined by a primary point process that is Poisson. Each secondary process represents a break cluster, typically consisting of 1-10 DSBs in a comparatively localized stochastic pattern determined by chromatin geometry and radiation track structure. Using this model, which we call the "RLC" (randomly-located-clusters) formalism, theorems are derived for how the DNA fragment-size distribution depends on radiation dose. RLC dose-response relations are expressed in terms of a dose-independent, one-cluster, fragment-size distribution function. The RLC formalism generalizes previous models, fits current data adequately, and allows mechanistically-based extrapolations from high-dose experiments to the much lower doses of interest for most applications.

Running title: Clustering of DNA Breaks Along Chromatin

Corresponding author: Rainer K. Sachs, Dept. Math., UCB, Berkeley, CA 94720.

Phone 510-642-4384; Fax 510-642-6726;

Email sachs@math.berkeley.edu; Internet

http://math.berkeley.edu/ sachs/

Key words: DNA double strand breaks; non-random chromosome breakage; Poisson cluster point process; ionizing radiation; LET.

1 Introduction

1.1 Background

We will analyze mammalian cells exposed to ionizing radiation during the G1 phase of the cell cycle. The radiation causes DSBs (DNA double strand breaks) [1].[1] DSBs can be lethal to a cell, or cause it to undergo malignant transformation.

The radiation consists of stochastically structured tracks which are independent of each other [2]. For example, some recent data concern irradiation with a beam of high-energy nitrogen nuclei [3], and then one track consists of all the energy depositions (e.g. ionizations) produced in the biological matter by one nitrogen nucleus from the beam. In this case, the linear energy transfer (LET), characterizing average energy deposition per unit track length [2, 4], is about 100 keV/μm, i.e. one is dealing with densely ionizing ("high LET") radiation. Gamma rays, by way of contrast,

[1] Abbreviations: DSB = DNA double-strand break; kbp = 10^3 base pairs; Mbp = 10^6 base pairs; LET = linear energy transfer; Gy = Gray = radiation dose in Joules per kilogram; RLC formalism = randomly-located-clusters formalism.

are sparsely ionizing, with an LET of less than 1 keV/μm.

For radiation with an LET as high as 100 kev/μm, even a single radiation track passing through the nucleus of a cell has a significant probability of causing serious damage. The track makes many ionizations, and a few of the ionizations may produce DSBs. Unrepaired DSBs cut chromosomes into smaller fragments. Recent high-LET pulsed field gel electrophoresis experiments [3, 5, 6] measure the corresponding distribution of DNA fragment sizes (here and throughout we use *size* to mean DNA content, measured, e.g., in base pairs).[2]

DNA fragment-size data are informative about DSB clustering along the DNA, dependent on chromosome geometry during cell cycle interphase and on LET or other aspects of radiation track structure [8 – 14]. It has long been known that the stochastically-placed ionizations due to one high-LET radiation track are spatially localized [15], being predominantly near the line representing the center of the track rather than being spread randomly over a whole cell nucleus. This localization strongly influences the fate of irradiated cells [4]. Spatial localization of ionizations produces DSB clustering along chromosomes, because an undamaged chromosome is based on a continuous, though multiply folded and winding, thread of DNA. Loci which are separated by fewer base pairs along the DNA tend on average to be closer in space. With this geometry, a bias for ionizations close to each other in space produces a bias for DSBs close to each other along the DNA [16].

DSB locations along chromosomes fully determine the DNA fragment-size distribution; but to elucidate the implications of the observed fragment-size distributions for DSB locations along chromosomes requires some modeling [3, 5, 6, 13, 16]. For example a pattern of interspersed small and large DNA fragments can produce the same size distribution as a different DSB pattern, in which most small (respectively large) DNA fragments are adjacent to other small (respectively large) fragments. Modeling of DNA fragment-size-distributions has typically used computer simulations

[2]In this context, "size" is synonymous with "molecular weight" [6] and with "molecular length" [7]. Size is an indicator of contour length along the DNA: between two DSBs; or between a DSB and a telomere (chromosome end); or between two telomeres; or between two restriction enzyme cutting sites; etc.

based on radiation track structure and on the detailed geometry of chromatin, i.e. of the DNA and its associated proteins [9, 13i, 14 17– 21]. Such models have been applied mainly to data on comparatively small sizes, analyzing "locally multiply damaged sites" on the 10 bp scale of the underlying DNA double helix [21 – 24] or "regionally multiply damaged sites" on chromatin scales of 10 bp to several kbp, corresponding to nucleosomes and the 30 nm chromatin fiber [13, 14, 18 – 20].

However, the recent high-LET fragment-size data on mammalian cells include results on much larger fragment sizes, up to several Mbp. Such sizes are so large the detailed chromatin geometry is not well characterized, the ability to view chromatin on a large scale during cell-cycle interphase being rather recent [25]. There is evidence of considerable randomness in chromatin structure at scales from 0.1 Mbp to more than 100 Mbp [26 – 29] and also evidence for systematic structures (e.g. [19, 25, 29 – 31]), 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 experiments.

Even without detailed information on chromatin geometry, stochastic models can give information on the implications of measured DNA fragment-size distributions [5]. A recent discussion of the high-LET fragment-size data introduced a "clustered-breakage" formalism [16], generalizing the well-known "random-breakage" formalism [7, 32]. Both formalisms represent DSB locations along a chromosome by a stochastic process involving points on the x-axis (with x regarded as DNA content). That is, one uses a stochastic point process [33, 34].

Applying the clustered-breakage formalism to some of the high-LET data showed that most DSBs are situated in clusters, corresponding to multiply damaged chromatin sites on the kbp-Mbp scale [6, 16]. It was also concluded that different *clusters* are located more or less randomly in the genome (though different DSBs within one cluster are highly correlated) and that one track typically makes only one or two clusters per genome, having a virtually negligible probability of making more than one distinct cluster on any one chromosome [16].

1.2 The RLC Formalism: Motivations and Preview

These clustering properties suggest describing the DSB pattern by a standard kind of stochastic point process — a Poisson cluster process [33, 34]. DSB cluster "locations" (e.g. centers) on a chromosome are then represented as an ordinary Poisson process, called the *primary* process of the Poisson cluster process, whilst any one cluster is represented as a localized point process, called a *secondary* process of the Poisson cluster process. The intuitive picture is that primary points are located at random on the DNA, and each primary point specifies the location of a comparatively tight cluster of DSBs, with different clusters due to different radiation tracks.

Because radiation tracks are independent, the standard assumption that the different clusters (i.e. the different secondary processes) are probabilistically independent will be made. If irradiation is sufficiently prolonged that Brownian motion significantly rearranges a chromosome during the time between the arrivals of one track and the next track, this assumption is appropriate. For very rapid irradiation, however, dependencies could in principle arise because different tracks are hitting the same chromosome. For example, if two different tracks each make a DSB cluster on a particular chromatin stretch which is more condensed than average, both clusters will tend to have an above-average number of DSBs; such local correlations among DSB clusters produced by different tracks are neglected in the formalism. Preliminary calculations (to be reported on elsewhere) suggest that, even in the case of no chromatin motion, the magnitude of the correlations between different clusters is negligibly small in the situations of interest.

Our analysis will use point processes which are *stationary* (i.e. translation invariant [33, 34]), because, when averaging over a sufficiently large size scale, chromosomes appear to be more or less uniformly sensitive to DSB induction [20, 35]. Stationarity implies that the secondary processes (i.e. the clusters of the Poisson cluster process) are stochastically equivalent, apart from being located (e.g. centered) at the different points specified by the Poisson primary process.

A stationary point process involves points along the entire infinite x-axis. In our case the number of points in any one cluster is finite a.s. (almost surely), but we must envisage an infinite number of clusters spaced out randomly along the x-axis. Correspondingly, we must sometimes be careful with effects due to the finite size of a chromosome [7, 16, 32]. Such effects are called "telomere effects", since the ends of a chromosome consist of telomeres. In the present context, telomere effects are somewhat analogous to the boundary effects of solid state physics.

Our stationary Poisson cluster process approach to describing DSB locations along a chromosome will be called the *RLC* (randomly-located-clusters) formalism. In the special case where each secondary "cluster" always consists just of a single DSB, the RLC formalism is the familiar random-breakage formalism, based on an ordinary Poisson process and applicable to low-LET radiation. This special case of the random-breakage formalism leads to what is called the "broken-stick" distribution for DNA fragment sizes, which reduces to an exponential distribution when telomere effects are negligible [7, 32].

We shall show that the "clustered-breakage" formalism introduced earlier for analyzing high-LET radiation [16] is also related to the RLC formalism, being a limiting case. However, the new RLC formalism is conceptually clearer. For example in the earlier, clustered-breakage, formalism one must provisionally specify the average number of tracks that strike the cell nucleus per unit radiation dose, for which only fairly rough estimates may be available. One then shows that to very good approximation this parameter always appears multiplying a second parameter, the number of DSB clusters per track, forming a single relevant combination, the number of clusters per unit dose, which is easier to determine from the data than is either factor separately. In the RLC formalism, on the other hand, one never needs to introduce the number of tracks per cell nucleus in the first place.

Another advantage of the RLC formalism is that the mathematical framework, of stationary Poisson cluster processes, is very well known, with useful theorems available in the literature [33, 34]. For example, we will show that telomere effects, which arise because a few DNA segments terminate in telomeres instead of DSBs, can readily be analyzed by using standard results on the sum of two stochastic point processes. In the clustered-breakage formalism, however, telomere effects have only been calculated in special cases, using extra assumptions and numerical methods [16]. In addition, various generalizations of the RLC formalism are well known; for example if one wanted to take into account variable sensitivity to breakage along a chromosome, it would be possible to use non-stationary cluster processes.

However, these advantages of the RLC formalism over the earlier clustered-breakage formalism are more conceptual than numerical. We will show by an example that for the data sets of main interest, the two formalisms give virtually the same predictions numerically, provided a simple method of associating models is used. The predictions fit the data much better than do the results of the random-breakage formalism.

The main purpose of the RLC formalism is relating low-dose, one-track effects to large-dose, multi-track effects. The RLC equations determine dose-dependent DSB clustering patterns, and predict dose-response relations for measured DNA fragment-size distributions. The formalism thereby allows mechanistically based extrapolations of the high-LET fragment-size data, obtained using large doses, to the much smaller doses relevant for most related radiobiological applications. Such applications include estimating the risk of carcinogenesis, estimating cell killing in radiotherapy, and analyzing chromosome aberrations as indicators of past radiation exposure.

2 Methods

2.1 Fragment Sizes

Suppose a cumulative distribution function $F(x)$ is given (with $0 < x < \infty$), perhaps interpreted as the probability an arbitrarily selected DNA fragment has size x or less. One relevant example is when F is Weibull, specifically $F = 1 - \exp(-\sqrt{x/S_W})$, where $S_W > 0$ is an adjustable parameter [16]. In what follows, unless explicitly indicated otherwise, we shall assume that F is smooth, that

$F \to 0$ for $x \to 0$, and that $1 - F$ is of rapid decay for x large. The average size S of a fragment is given by

$$S = \int_0^\infty x dF(x) = \int_0^\infty [1 - F(x)]dx. \tag{1}$$

Here we have adopted the usual notation $dF(x)$, e.g. $dF = (dF/dx)dx$ for smooth F. Integration by parts was used to obtain the second equality.

2.2 Point Processes

In the present context, an outcome of a point process is an ordered set of points on the x-axis, $... \leq X_{-1} \leq X_0 \leq X_1 \leq X_2 \leq ...$, where there are only finitely many points in any bounded interval and the inequalities are strict almost surely. For a stationary point process we may here assume there are infinitely many points on both sides of $x = 0$ and can make the convention that X_0 is the largest negative value [34].

The *intensity* of a stationary point process is the average number of points per unit size; in the cases of interest to us, the intensity of a stationary process modeling DSBs will be directly proportional to the radiation dose. The average fragment size S for a stationary point process, i.e. the average DNA content of an interval $X_{j+1} - X_j$ is the inverse of the intensity [33, 34].

2.3 Remaining Size

Suppose we have a stationary point process in which, given that we are located at any one point of the process, the probabilities for the location of the next point are governed by $F(x)$, i.e. the probability distribution for $X_{j+1} - X_j$ is $F(x)$ for all j, the intervals being independent and identically distributed. Then the point process is called a stationary *renewal* process [34]. A standard result on stationary renewal processes will here be useful, though we shall not actually use any stationary renewal processes other than Poisson processes. For a stationary renewal process, there is a distribution $E(x)$, called the "remaining size" distribution, which can be interpreted in either of two ways: (1), for $x > 0$ and for a randomly selected location x_0 the x-axis, $E(x)$ is the probability that there is a point of the point process in the interval $(x_0, x_0 + x]$; or (2), 1-$E(x)$ is the probability a randomly selected interval of size x contains no points. The equation for $E(x)$ is the following [34]:

$$E(x) = (1/S)\int_0^x [1 - F(x')]dx'; \quad 1 - F = S(dE/dx). \tag{2}$$

An intuitive derivation of this equation starts by noting that for infinitesimal dy the probability a randomly chosen point lies within a DNA fragment with size in the range $(y, y + dy)$ is proportional to $ydF(y)$, where the proportionality constant must be $(1/S)$, the average number of fragments per unit size. Conditional on the randomly chosen point lying in a fragment of size y, the probability that there is no DSB within size x to the right is $(y - x)/y$ for $y > x$ and zero for $y \leq x$. Thus $E(x) = 1 - (1/S)\int_x^\infty (y - x)dF(y) = (1/S)\int_0^x [1 - F(y)]dy$, i.e. Eq. 2 holds. The average remaining size, i.e. $\int_0^\infty xdE$, can readily be larger than S, contrary to naive expectations that it should be $S/2$, a famous phenomenon sometimes called the sampling paradox.

2.4 Secondary Point Processes Representing Individual Clusters

As discussed above, the point processes of main interest in the present paper are stationary Poisson cluster processes, which involve a secondary point process modeling a local cluster of DSBs. We take the secondary process to have at least one DSB, and to have a finite average for the number of DSBs, the *average multiplicity* M. $M \geq 1$, with $M = 1$ iff a secondary cluster almost surely contains only one DSB.

The secondary point process representing one cluster is determined in principle by radiation track structure and the geometry of chromosomes, but we will here work mainly with less complicated secondary processes, chosen for mathematical convenience. A simple mathematical example of a secondary point process is the process that always contains exactly two points, on opposite sides of the location x_0 specified by the primary Poisson process and at equal distances from x_0, with the size of the intervening DNA fragment distributed according to some distribution function $F(x)$.

A somewhat more realistic example is given by a renewal-type construction, where a cluster whose primary point is the origin $x = 0$ is determined as follows.[3] There is a DSB at the origin $x = 0$, and there is a probability $p_1 = p$, with $0 < p \leq 1$, that the cluster contains only this one DSB. There is a probability $p_2 = pq$, where $q = 1 - p$, that the cluster has exactly 2 DSBs, and in that case the second DSB is located to the right of the origin, being within the interval $(0, x]$ with probability $F(x)$. There is a probability $p_3 = pq^2$ for exactly three DSBs in the cluster, and in that case, given that the second DSB was at x, the third DSB is within the interval $(x, x+y]$ with probability $F(y)$. And so on for 4, 5, ... DSBs. The average cluster multiplicity M and average cluster size are thus:

$$M = \sum_{j=1}^{\infty} j p_j = p \sum_{j=1}^{\infty} j q^{j-1} = \frac{1}{p}; \quad \text{Average Cluster Size } = (M-1)\int_0^{\infty}(1-F)dx. \tag{3}$$

This example is special in various ways. In general the probability p_j of a cluster having multiplicity j need not be given in terms of a single parameter p; any discrete probability distribution would be acceptable. Moreover, in the example there are no correlations between the sizes of successive DNA fragments in one cluster, nor any correlation between the fragment sizes and cluster multiplicity; each fragment size is chosen at random from the same distribution $F(x)$, without regard to the number or size of previous choices. "Markov" clusters, where there are correlations between the sizes of two successive fragments, are perhaps even more realistic, since, for example, the occurrence

[3] In this case the secondary process is not centered at the location specified by the primary Poisson process; rather, the primary location specifies the leftmost point of the local cluster. However, this off-center property (which could be avoided) makes no essential difference to the arguments.

of a fragment smaller than average may signal being near the center of a radiation track, which in turn could bias for the next fragment to be smaller than average. The RLC formalism, Theorems 1-5 below, applies to secondary clusters with any kind of interdependencies, but current data is insufficiently detailed to fix even roughly the extra adjustable parameters one must introduce if using a Markov or other secondary cluster with interdependent fragment sizes.

Given any secondary cluster process for which $M > 1$, one can define a fragment-size distribution function $F(x)$ as the probability an arbitrarily selected fragment bounded by DSB free ends from the cluster has size x or less. Clusters of higher multiplicity are to be weighted more heavily, i.e.

$$F(x) = [1/(M-1)] \sum_{j=2}^{\infty} (j-1) p_j F(x|j), \quad M - 1 = \sum_{j=2}^{\infty} (j-1) p_j. \tag{4}$$

Here p_j is the probability a given cluster has multiplicity j, and $F(x|j)$ is the fragment-size probability distribution conditional on there being exactly $j \geq 2$ DSBs in the cluster. Thus $F(x)$ describes the overall frequency for a secondary point process to make fragments of size x or less, while the secondary process itself may specify more detailed information — e.g. information on which sized fragments occur in clusters of a particular multiplicity, on possible correlations in the sizes of successive fragments, etc. Assuming x is much smaller than an unbroken chromosome, and neglecting telomere effects, $F(x)$ is in principle measured by measuring histograms of DNA fragment sizes, using a radiation dose so low that the chance of two diferent clusters overlapping, or adjoining within size x, is negligible.

2.5 Generating Functionals

Generating functionals play an important role in the theory of point processes. Consider a function $f(x)$ from the reals to [0,1], and suppose f differs from 1 only within a bounded region of the x-axis. The generating functional $G[f]$ of a point process assigns to any such f the expected value (i.e. the ensemble average) of the product

$$...f(X_{-1})f(X_0)f(X_1)f(X_2).... \tag{5}$$

For example, any generating functional must assign to the function $f \equiv 1$ the value 1. A point process is uniquely characterized by its generating functional [33,34].

The following examples supply context and some results useful later.

1. If a point process consists merely of a point at the origin $x = 0$ with probability one, then its generating functional is $G[f] = f(0)$.
2. As above, consider a pair of points, with the origin $x = 0$ halfway between them and the interval governed by a cumulative probability function $F(x)$; then

$$G[f] = \int_0^\infty f(-x/2)f(x/2)dF(x). \tag{6}$$

3. Suppose, for a stationary Poisson cluster process, the secondary process located at the origin $x = 0$ has generating functional $G^{[0]}[f]$. Then when the process is located at the point $x = a$ instead of the origin the generating functional is $G^{[a]}[f] = G^{[0]}[g]$ where $g(x) = f(x + a)$ for all x. For example if $G^{[0]}$ has the form specified in Eq. 6 then

$$G^{[a]}[f] = \int_0^\infty f(a - x/2)f(a + x/2)dF(x). \tag{7}$$

4. Consider the renewal-type secondary process defined in Subsection 2.4, located at the origin. Then the generating functional is

$$G^{[0]}[f] = f(0)\left[p + pq\int_0^\infty f(x)dF(x) + pq^2\int_0^\infty f(x)dF(x)\int_0^\infty f(x + y)dF(y) + ...\right]. \tag{8}$$

5. A stationary Poisson process whose intensity is λ has generating functional [33, 34]

$$G[f] \quad = \quad \exp\left[\lambda \int_{-\infty}^{\infty} [f(x) - 1] dx\right]. \tag{9}$$

6. Consider a stationary Poisson cluster process whose secondary process located at the origin has generating functional $G^{[0]}[f]$. Then the generating functional for the Poisson cluster process is [33, 34]

$$G[f] \quad = \quad \exp\left[\lambda \int_{-\infty}^{\infty} \left(G^{[x]}[f] - 1\right) dx\right], \tag{10}$$

   Here λ is the intensity of the primary Poisson process.

7. A stationary process places DSBs all along the x axis, but of course a chromosome has a given, finite size. In the present paper we shall for brevity consider only chromosomes of one fixed size, S_C, although no essentially new ideas are needed to extend the results to more realistic cases, e.g. if there are 46 different chromosomes having 23 different sizes. Because telomeres are like DSBs in that they terminate a DNA fragment, a convenient way to take telomere effects, and thus finite chromosome size, into account is to introduce a stationary point process of telomeres, with the left telomere located at random and the right telomere located S_C to the right. Specifically, we shall, for later convenience, define as the *telomere process* the stationary point process with generating functional

$$T[f] \quad = \quad (1/S_C) \int_0^{S_C} \quad dx \quad [...f(x - S_C)f(x)f(x + S_C)f(x + 2S_C)...]. \tag{11}$$

2.6 Sums

The *sum* of two point processes has as its outcomes the point-sets which are the unions of the point-sets for the outcomes of the two processes, and the probabilities are adjusted appropriately.

A useful theorem is that the generating functional of the sum is the product of the individual generating functionals [33, 34]. For example, when considering DNA fragment sizes, a telomere and a DSB play similar roles [36], since either can terminate a fragment. Correspondingly, we shall be interested in the sum of a Poisson cluster process as in example 6 of Subsection 2.5 and a telomere process as in example 7. The corresponding generating functional is given by the product of G in Eq. 10 and T in Eq. 11.

2.7 Palm Processes

Intuitively speaking, the Palm process corresponding to a point process specifies the pattern obtained if a randomly chosen point of the original process is translated to the origin [34]. Palm processes will be used when we want to analyze fragment sizes, since a randomly chosen fragment starts with a randomly chosen DSB or telomere and ends with the next DSB or telomere. Rather than reviewing the formal definition of a Palm process [34], we give some examples.

1. For the process in example 2 of Subsection 2.5, the generating functional of the corresponding Palm process is

$$G[f] = (1/2)f(0)\int_0^\infty dF(x)[f(x) + f(-x)]. \tag{12}$$

2. Consider a secondary process, representing one local cluster, with multiplicity $M = 1/p$ and with generating functional given by Eq. 8. Then the generating functional $P[f]$ of the corresponding Palm process is

$$P[f] = (1/M)f(0)\Big[p + pq\int_0^\infty dF(x)[f(x) + f(-x)]$$
$$+ pq^2\int_0^\infty dF(x)f(x)\int_0^\infty dF(y)f(x+y) + pq^2\int_0^\infty dF(x)f(-x)\int_0^\infty dF(y)f(y)$$
$$+ pq^2\int_0^\infty dF(x)f(-x)\int_0^\infty dF(y)f(-x-y) + pq^3...\Big]. \quad (13)$$

3. Consider a stationary Poisson cluster process, and suppose the generating functional of the Palm process corresponding to the secondary process is $P[f]$ (compare the previous example). Denote the generating functional for the stationary Poisson cluster process (example 6 of Subsection 2.5) by $G[f]$. Then a rather long proof [34] shows that the Palm process for the stationary Poisson cluster process is just the process whose generating functional is the product

$$P[f]G[f] \quad (14)$$

Such a product of generating functionals corresponds to a sum of two point processes, and the intuitive interpretation here is rather elegant, as follows. To form the Palm process we must consider a DSB being translated to the origin. That DSB must belong to one of the secondary processes (i.e. to one local cluster). The over-all process is the sum of that particular secondary process and all the other secondary processes. However, by the independence properties of a Poisson process, the sum of all the remaining secondary process itself forms a stationary Poisson cluster process, which has generating functional G because it has exactly the same stochastic structure as the original Poisson cluster process; for example, the average number of primary points in a given stretch of size x is λx, and the average number of additional primary points in such a stretch, conditional on having found and removed one primary point in the stretch, is also λx (not $\lambda x - 1$ as naive intuition would have it). We thus have a sum of one secondary process and the original stationary Poisson cluster process itself, so the generating functional is the product.

3 Results

3.1 The RLC (Randomly-Located-Clusters) Formalism

We now present the basic RLC formalism, in the form of 5 theorems on stationary Poisson cluster processes. We start with a lemma that applies when examining a specific stretch of DNA and measuring the probability that the stretch contains no DSBs. This kind of measurement is sometimes used as a supplement to measuring fragment-size distributions [3].

Lemma. For a point process, the probability that there is no point in the interval $(a, a+b]$ is the generating functional $G[f]$ for the following special choice of f:

$$f = 1, \quad x \leq a; \qquad f = 0, \quad a < x \leq b; \qquad f = 1, \quad x > b. \tag{15}$$

For proof, note that with this choice of f the product in Eq. 5, whose expected value is the generating function, has value 1 whenever no point lies in the interval $(a, a+b]$ and has value 0 otherwise. Thus the probability there is no point in the interval is $G[f]$. ■

Combining the Lemma with our formula above for the generating functional of a stationary Poisson cluster process gives our first main result, as follows.

Theorem 1. Suppose we have a stationary Poisson cluster process, with λ the intensity of the primary Poisson process. Suppose the secondary process has average multiplicity $M > 1$, and has a fragment-size distribution given by $F(x)$. Let S be the average fragment size defined by Eq. 1 and let $E(x)$ be the remaining size distribution defined in Eq. 2. Denote by $\widehat{A}(x)$ the probability that a given DNA stretch of size x has at least one DSB, so that $1 - \widehat{A}$ is the probability for none.[4] Then

[4]In what follows, hats as in $\widehat{A}$ will be used to connote that telomere effects are (temporarily) being neglected.

$$1-\widehat{A}(x)=\exp\bigl[-\lambda[x+(M-1)SE(x)]\bigr]. \tag{16}$$

The proof consists of applying the Lemma to the generating functional, Eq. 10, for a stationary Poisson cluster process, as follows. Since the process is stationary, one may for simplicity take the given stretch to be $(0,x]$. Then we must evaluate the integral $\int[G^{[y]}[f]-1]dy$ for f given by 15 with $a=0$, $b=x$. By example 3 of Subsection 2.5, $G^{[y]}[f]=G^{[0]}[g]$, where $g(u)$ is zero for $y<u\leq y+x$ and is 1 otherwise. Now $G^{[0]}[g]$ consists of an ensemble average of terms like $g(X_1)g(X_2)...g(X_{m-1})g(X_m)$. Given y each such term is either zero or one, according as one (or more) of the X_j lies in the interval $(y,y+x]$ or not. The total size of the intervals where the product is zero is thus

$$x+X_m-X_1-\sum_{j=1}^{m-1}z_j,\quad \text{where}\quad z_j=\begin{cases}X_{j+1}-X_j-x, & \text{if } x<X_{j+1}-X_j\ ;\\ 0, & \text{otherwise.}\end{cases} \tag{17}$$

Averaging, using Eq. 10 and using the definition of $E(x)$ now gives the theorem. ■

We next give a theorem on fragment size-distributions, applicable if telomere effects are negligible.

Theorem 2. For a stationary Poisson cluster process, the probability $\widehat{B}(x)$ that an arbitrarily chosen fragment has size less than or equal to x is given by

$$1-\widehat{B}(x)=(1/M)\Bigl[1+(M-1)[1-F(x)]\Bigr]\exp\Bigl[-\lambda[x+(M-1)SE(x)]\Bigr]=(1/\lambda M)(d\widehat{A}/dx). \tag{18}$$

The proof consists of applying the Lemma to the product generating functional PG of the Palm process (Eq. 14), using Theorem 1 to evaluate the factor G and using Eq. 4 to evaluate the factor P. ■

Both equalities in Eq. 18 have useful intuitive interpretations. The interpretation of the first equality is the following. Suppose one DSB is chosen at random; it belongs to some one secondary cluster. The probability the next DNA stretch, of size x, is free of DSBs is the product of the probability that there is no DSB from from the same secondary cluster and the probability there is no DSB from any other secondary cluster. The second factor of the product is just the distribution of Eq. 16 since the chosen DSB is merely a random point as far as the stationary Poisson cluster process formed from all the other secondary clusters is concerned. The first factor of the product is $1 - F(x)$ in an average fraction of $(M - 1)/M$ cases, namely the cases where the selected DSB is not the rightmost DSB of its own secondary cluster. In an average fraction of $1/M$ cases, the factor is unity since the DSB involved is the rightmost DSB of its own cluster. Combining these possibilities gives the first equality in Eq. 18.

The intuitive interpretation of the second equality is that the relation between $\widehat{A}(x)$ and $\widehat{B}(x)$ is, conceptually and computationally, wholly similar to the standard relation, discussed in connection with Eq. 2, between $E(x)$ and $F(x)$. Note, in particular, that the average fragment size for a stationary Poisson cluster process is $1/\lambda M$.

Example. Suppose each secondary process consists of a single point at the location specified by the primary process. Then the Poisson cluster process is merely an ordinary Poisson process and $M = 1$. In this case $\widehat{A} = 1 - \exp(-\lambda x) = \widehat{B}$ by Eqs. 16 and 18. This is essentially the standard result that for an ordinary stationary Poisson process, fragment sizes and remaining sizes are both exponentially distributed, with the same parameter [34].

Theorems 1 and 2 ignore telomere effects. We now give corresponding results when telomere effects, for a chromosome of size S_C, are included. Essentially, telomeres will be regarded as just extra DSBs [36], regularly spaced along the x-axis. We will thus use the process with generating functional $G[f] = \widehat{G}[f]T[f]$, where $\widehat{G}$ is the generating functional of a stationary Poisson cluster process, Eq. 10, and T is the generating functional of the telomere process, Eq. 11. The functions corresponding to $\widehat{A}$ and $\widehat{B}$ in Theorems 1 and 2 will be denoted by A and B respectively.

Theorem 3. Denote by $I(x)$ the function defined by $I = 1$ for $x \leq S_C$ and $I = 0$ otherwise; denote $1/S_C$ by τ. Then

$$1 - A(x) = [1 - \widehat{A}(x)][1 - \tau x]I(x) \qquad \text{(where} \quad \tau = 1/S_C\text{)}. \tag{19}$$

Proof. Apply the Lemma to the generating functional $G = \widehat{G}T$. ■ The intuitive interpretation of Eq. 19 is that, starting from a randomly chosen point, a subsequent stretch of size x is free from "breaks" (i.e. contains neither DSBs nor telomeres) iff it is free from both DSBs and telomeres. The probability the stretch is free from DSBs is $[1-\widehat{A}(x)]$ by the interpretation of $\widehat{A}$, while the probability it is free of telomeres is $[1 - (x/S_C)] = 1 - \tau x$ for x smaller than the size of a chromosome and is zero otherwise. This reasoning gives Eq. 19. The fragment size distribution including telomere effects is now obtained much as in Theorem 2, as follows.

Theorem 4.

$$1 - B(x) = \frac{dA/dx}{\lambda M + \tau} = \frac{I(x)}{\lambda M + \tau}[\lambda M[1 - \widehat{B}(x)](1 - \tau x) + \tau[1 - \widehat{A}(x)]] \tag{20}$$

Proof. The average number of "breaks" (i.e. DSBs plus telomeres) per unit size is $\lambda M + \tau$. The first equality thus follows by the same reasoning used above to interpret the second equality in Eq. 18. The second equality then follows from differentiating Eq. 19, using Eq. 18 and simplifying. ■

Example. Consider again the case where the Poisson cluster process is merely an ordinary Poisson process (i.e. $M = 1$). Then Eq. 20 becomes the standard "broken stick" distribution of the familiar random-breakage model, i.e. [7, 32]

$$1 - B(x) = \frac{I(x)\exp(-\lambda x)}{\lambda + \tau}\big[\lambda(1 - \tau x) + \tau\big]. \tag{21}$$

Theorem 5. Suppose we measure the fraction, by size, of all DNA fragments that have sizes in the range S_1 to S_2. Then the fraction Φ is

$$\Phi = \Big[A - x(dA/dx)\Big]_{x=S_1}^{x=S_2} = -\Big[I(x)[1-\widehat{A}(x)]\big[x\lambda(1-\tau x)[M(1-F(x))+F(x)]+1\big]\Big]_{x=S_1}^{x=S_2}. \quad (22)$$

Proof. The fraction in an infinitesimal interval is proportional to $x(dB/dx)dx$, and the proportionality constant is $\lambda M + \tau$, the inverse of the average size; integrating $\int xB'$ by parts, using $(\lambda M + \tau)[1 - B] = dA/dx$, and then simplifying gives the theorem. ■

Theorems 1-5 define the RLC (i.e. randomly-located-clusters) formalism. The formalism relates one-cluster quantities (M, F, S, E) to measured size distributions, using the number λ of clusters per unit size.

In practice, the main measurements are those which determine $F(x)$ for small x, while quantities such as S (whose computation requires knowing F for all x, including large values), M, and λ are inferred more indirectly. Notice in Theorem 1 that the exponent can be rewritten in the form

$$-\lambda M\big[x + [1-(1/M)]\int_0^x dF(x')\big]. \quad (23)$$

Here S does not appear explicitly; moreover if M is large and λM is held fixed, there is only a weak dependence on M. Since the only dependence on one-cluster quantities in Theorems 2-5 is via $widehat A$, the same comment applies to all the subsequent equations. The effect of these properties is that it is easier to estimate, from experiments, $F(x)$ for small x and λM than it is to estimate M, λ, $F(x)$ for large x, or S.

3.2 Comparing Formalisms

We have shown in examples above that the random-breakage model is that special case of the RLC formalism where each "cluster" is merely an isolated DSB, so that the average multiplicity has its minimum possible value $M = 1$. This random-breakage model is often used to analyze fragment-size data for low-LET radiations (reviews: [7, 32]). We now show that an opposite limit, involving large M in the RLC formalism, gives what is called the clustered-breakage formalism, used previously to analyze fragment-size data for high-LET radiations. Our treatment will be condensed; interpretations and motivations, which were given in the earlier paper on the clustered-breakage formalism [16], will for the most part be omitted.

In the clustered-breakage formalism, the one-track fragment length distribution function F above is modified, to include a few extra, very long fragments. As will be illustrated concretely in the next Subsection, $F(x)$ is then used not only to describe clustering of DSBs, but also to determine the probability that one particular radiation track makes at least one DSB on any one particular chromosome. Formally we shall write $F(x) \to F_1(x)$, where the subscript 1 was used to specify one-track quantities in the clustered-breakage formalism [16]. Corresponding notations were used for average size and for the remaining size distribution, defined in Eqs. 1 and 2, and N_1 was used to denote the inverse of the average size, i.e.

$$S \to S_1 = \int_0^\infty (1 - F_1)dx = 1/N_1, \qquad E(x) \to E_1(x) = N_1 \int_0^x (1 - F_1)dx. \tag{24}$$

In the clustered-breakage formalism, the breaks due to a single track were modelled as a stationary point process. This means that we must consider a limit where the multiplicity M used in the previous Subsection becomes infinite. Then one must also take the limit $\lambda \to 0$, with λM finite, specifically $\lambda M \to n/S_1$, where n was interpreted as the average number of tracks which strike a cell nucleus [16]. Finally, telomere effects were neglected in the main equations of the clustered-breakage formalism. Thus the limit $S_C \to \infty$ (i.e. $\tau \to 0$) is also needed.

In Theorems 1-5 we therefore insert $F \to F_1$, $E \to E_1$, $S \to S_1 = 1/N_1$, $M \to \infty$, $\lambda M \to nN_1$, and $\tau \to 0$. For ease of comparison with the notation of a previous paper [16] we also use the notations $1 - A \to E_{cb}$, $B \to F_{cb}$, and $\Phi \to \Phi_{cb}$. A calculation using the RLC formalism, Theorems 1-5, then gives the main equations of the clustered-breakage formalism [16], namely:

$$E_{cb}(x) = \exp\left[-nE_1(x)\right]; \tag{25}$$

$$1 - F_{cb}(x) = [1 - F_1(x)] \exp\left[-nE_1(x)\right] = -(1/nN_1)(dE_{cb}/dx); \tag{26}$$

$$\Phi_{cb} = \left[\exp\left[-nE_1(x)\right] \left[nN_1 x[(1 - F_1(x)] + 1\right]\right]_{x=S_1}^{x=S_2}. \tag{27}$$

We have thus shown that an appropriate limiting process gives the clustered-breakage formalism from the RLC formalism. The next Subsection considers the relation between the two formalisms when they are applied to a specific data set.

3.3 Comparison to Data

Using the data of Löbrich et al. [3] for human cells irradiated with nitrogen ions we now give an example of how the RLC formalism (Theorems 1-5) compares to the data, to the clustered-breakage formalism (Eqs. 24-27), and to the more familiar random-breakage formalism (Eq. 21).

In an earlier paper [16], the clustered-breakage formalism was compared to the nitrogen ion data, neglecting telomere effects and using the following form for F_1. For $x \le S_0$, where $S_0 = 1.12$ Mbp was a maximum measured size for fragments, the form of F_1 was given by

$$1 - F_1(x) = \exp[-(x/S_W)^{0.5}] \qquad (x \le S_0). \tag{28}$$

Here S_W was a positive adjustable parameter, the size scale of the Weibull distribution in Eq. 28. The Weibull form was chosen by inspection of the data and for mathematical convenience, not by mechanistic considerations of track structure and chromatin geometry. Intuitively, the distribution 28 corresponds to the distribution for a cluster, where a cluster is defined by the property that no DNA segment from within a cluster has size greater than S_0.

For large x the following form, motivated by inspection of the data and a requirement that F_1 be continuous at S_0, was used.

$$1 - F_1 = \exp[-(S_0/S_W)^{0.5}] \exp[-\lambda_{cb}(x - S_0)], \qquad (x \ge S_0). \tag{29}$$

Here λ_{cb} was the second and last adjustable parameter of the model. An intuitive interpretation of λ_{cb} was given as follows. It was proved that the probability that a particular track produces no DSBs on a given DNA segment of size S_S is

$$\frac{\exp(\lambda_{cb} S_0)}{1+\alpha} \exp(-\lambda S_S), \quad \text{where} \quad h = (S_0/S_W)^{0.5}, \quad \alpha = 2\lambda_{cb} S_W[\exp(h) - 1 - h]. \tag{30}$$

It was shown that to good approximation this probability becomes $\exp(-\lambda_{cb} S_S)$, i.e. λ_{cb} is approximately the intensity for clusters (not for DSBs) produced by one track.

The number, n, of tracks per nucleus was assigned the nominal value $n = 10D$, where D is radiation dose in Gray (Gy). The two adjustable parameters, S_W and λ_{cb} were then determined, from fragment-size data at 189 Gy and from the measured probability that no DSBs occur within one specially labeled stretch of the genome. The values obtained were S_W =0.555 Mbp and $\lambda_{cb} = 3.01 \times$

10^{-4} per Mbp. The clustered-breakage model with these parameters was a significant improvement over the best fitting random-breakage model (Table 1) and gave acceptable fits to data at other doses [16].

One can get a RLC model "corresponding" to this clustered-breakage model as follows. First, take

$$F(x) = \begin{cases} F_1(x)/F_1(S_0) & \text{if } 0 \leq x \leq S_0, \\ 1 & \text{if } x \geq S_0, \end{cases} \tag{31}$$

i.e. regard F_1, suitably rescaled, as applying to a secondary cluster. Next, assume a cluster terminates iff F_1 dictates that a large size is to be chosen, i.e. take $M - 1 = 1/F_1(S_0)$. And finally, take $\lambda = n\lambda_{cb} = 10D\lambda_{cb}$, i.e. the intensity λ of the Poisson primary point process is directly proportional to radiation dose D, corresponding to the average number, $10D$, of tracks per cell nucleus times an intensity, λ_{cb}, for a single track.

With F given, S and $E(x)$ for $x \leq S_0$ are now determined by Eqs. 1 and 2 as

$$S = S_W \frac{2 - (2 + 2h + h^2)\exp(-h)}{1 - \exp(-h)}, \quad E = \frac{2 - 2(1+y)\exp(-y) - y^2\exp(-h)}{2 - (2 + 2h + h^2)\exp(-h)} \quad (x \leq S_0). \tag{32}$$

Here h was defined in Eq. 30 and $y = (x/S_0)^{0.5}$; $E = 1$ for $x \geq S_0$. Given F, E, S and M, the multi-track fragment-size distribution is determined by Theorems 1-5.

Table 1 compares the experimental fragment-size distribution at 189 Gy and the predictions of the clustered-breakage model or random-breakage model [16] with the results obtained from the RLC model, the latter being specified as discussed above, i.e. by Eqs. 28 and 31 with $S_0 = 1.12$ Mbp, $S_W = 0.555$ Mbp, $M = 1 + [1/F_1(S_0)]$, and $\lambda = 10\lambda_{cb}D = 0.569$ per Mbp. It is seen that the random-breakage model gives a very poor fit to the data, while the other two models virtually coincide. In particular, telomere effects, included in the RLC formalism but not in the clustered-breakage formalism, are seen to be negligible for such small fragments at such a high dose. Results at other doses, and for other data sets such that as that of Newman et al. [6], are similar to those shown in Table 1.

Of course one could obtain a better fit to the data in Table 1 by adjusting the parameters (S_W, M, and λ) of the RLC model. However, in view of the experimental uncertainties the fit shown, obtained by direct transcription of the clustered-breakage parameters rather than curve fitting, is already acceptable. The main point is that, while the two models differ conceptually, they are often almost the same numerically when applied to the actual data of interest.

3.4 Predicted Dose-Response Curves

.

Eq. 22 gives the fraction of DNA predicted by the RLC model for a given size range. For small doses and small sizes, the predicted dose-response relation is essentially linear (Figure 1A), and the measured dose-response does appear to be linear [3, 6]. The interpretation is that fragments of the given size are being made by individual radiation tracks, i.e. typically both ends of a fragment are DSBs from the same secondary process (same one-track cluster) when the dose is so low that overlapping of different clusters is negligible. This one-track type of action results in a linear dose-response since the average number of tracks is directly proportional to dose. A linear response involving pairs of DSBs (one DSB at each end of a small chromatin fragment) parallels well-known linear high-LET dose-response relations for cellular log-survival or chromosome aberrations, both also interpreted as one-track, two-DSB effects [2, 37]

For a somewhat larger dose, an upward curvature is predicted (Fig. 1A). The interpretation is that two or more tracks are cooperating to produce fragments of the given size, as different clusters overlap or adjoin closely. At still higher doses, a downward curvature and ultimately a negative slope are predicted (Figure 1B). The interpretation is that multi-track cluster overlaps are cutting fragments of the given size into smaller sizes. The experimental results are consistent with, though not strongly indicative of, these predictions on higher doses [3, 6, 16]. At very high doses the prediction is that the size-distribution is the same as for the random-breakage formalism at very high intensity – if the two DSBs at the ends of each fragment are almost always due to different tracks, the corresponding point process is essentially an ordinary Poisson process as far as independence for two consecutive points is concerned. However, no empirical evidence is available for such very high doses.

3.5 Moments

One overall parameter of interest is the variance in fragment sizes. A convenient equation for calculating the variance or various moments, obtained through using Theorem 4 and integrating by parts twice, is the following.

$$\int_0^\infty \exp(-kx)dB(x) = 1 - \frac{k}{\lambda M + \tau} + \frac{k^2}{\lambda M + \tau}\int_0^\infty \exp(-kx)[1 - A(x)]dx. \tag{33}$$

The coefficient of $-k$ is the mean $1/(\lambda M)$ and using the coefficient of k^2 gives the variance as

$$\frac{1}{\lambda M + \tau}\left[2\int_0^\infty [1 - A(x)]dx - \frac{1}{\lambda M + \tau}\right]. \tag{34}$$

Some examples assuming telomere effects are negligible (i.e. setting $\tau = 0$) are the following. If $M = 1$ the variance is that for an ordinary Poisson process, $1/\lambda^2$. For clustering, i.e. for $M > 1$, and with λM held fixed, the variance is larger. Larger variance follows from Eq. 34, since the function $E(x)$ in Theorem 1 obeys, in view of Eqs. 1 and 2, $SE(x) \leq x$ for all x, with inequality holding for some values (e.g. $x > S$). In particular, if $F(x)$ is a step function, i.e. if $F = 0$ for $x < S$

and $F = 1$ otherwise, then the variance is $1/(\lambda M)^2[1 + (M-1)\exp(-\lambda MS)]$; here the second term in the square brackets represents the extra variance due to clustering.

4 Discussion

4.1 Review

We showed that well known results, on stochastic point processes that are stationary Poisson cluster processes, can be applied to data on DNA fragment-size distributions for mammalian cells exposed to densely ionizing radiation. The results for DSB locations on chromosomes, Theorems 1-5, were called the RLC (i.e. randomly-located-clusters) formalism. The RLC formalism identifies those aspects of the DSB pattern which influence observed fragment-size distributions; these aspects are the average multiplicity of a cluster (i.e. a secondary process), the one-cluster fragment-size distribution, and the intensity, proportional to the radiation dose, of the primary Poisson process which determines the locations of the clusters.

Explicit relations for the dose response of DNA fragment-size distributions were obtained (Theorem 4, Theorem 5, and Fig. 1). Most of the data is at high doses, but if the secondary process (i.e. the one-cluster DSB pattern) can be disentangled from the data, extrapolation to arbitrarily low doses is feasible, since at sufficiently low doses there will be one-track action almost exclusively. That is, when the dose drops low enough, the DSB clustering pattern for those cells that are damaged is always given by the same stochastic process; smaller and smaller doses then manifest themselves merely in fewer and fewer damaged cells, not in changes of the DSB clustering pattern within one damaged cell [2].

The familiar random-breakage model is that special case of the RLC formalism for which each cluster always has multiplicity one (Subsection 3.1). The RLC formalism matches recent data sets on DNA fragment-size distributions after high LET irradiation much more closely than does the random-breakage model (compare Table 1), i.e. the data strongly favor randomly-located clusters of DSBs compared to randomly-located DSBs. A clustered-breakage formalism previously introduced is a limiting case of the RLC formalism (Subsection 3.2, and, with a direct transcription of model parameters, gives virtually the same numerical results for the currently available data sets of interest (e.g. Table 1). However, the RLC formalism has the advantages of much greater mathematical familiarity (Subsection3.1), of somewhat greater conceptual clarity, and of slightly greater generality.

4.2 The RLC Formalism's Domain of Validity

The RLC formalism, Theorems 1-5, is rather general, being applicable to any radiation track structure, and to any chromatin geometry at any scale from about 1 kbp to the full length of a mammalian chromosome, 5 orders of magnitude larger. A lower size limit of about 1 kbp is indicated because the RLC formalism assumes uniform radiosensitivity. Chromatin may have non-uniform radiosensitivity at smaller scales, with cold spots in those DNA locations well shielded, e.g. by histone proteins, from indirect effects, i.e. from attack by chemical radicals produced nearby [13]. Typically, in any case, only one-track action is relevant for fragments a few kbp or less in size, and then the dose-response is linear (compare Fig. 1); for doses small enough that the dose-response is linear, the RLC formalism, even if applicable, is not really needed, since simple linear extrapolations then serve equally well.

4.3 Summary

The RLC formalism is a systematic way to see what happens when different DSB clusters, from different tracks, overlap or come close to overlapping. Because the formalism leads to explicit equations for the dose-response of DNA fragment-size distributions, it relates the high-dose experimental data to the very much lower doses of interest in such applications as risk estimation for carcinogenesis.

Acknowledgements.

We are grateful to D. Aldous, D. Brenner, and A. Kellerer for discussions. Research supported by NIH grant GM 57245-01 (RKS and ALP) and by NSF grants DMS 9532055 (PJH) and BIR 963-0735 (LRH).

References

1. J.F. Ward, The complexity of DNA damage: relevance to biological consequences, *Int. J. Radiat. Biol.* 66, 427-32 (1994).

2. A.M. Kellerer, Fundamentals of Microdosimetry. In: *The Dosimetry of Ionizing Radiation, Vol. I,* K. Kase, B. Bjarngard and F. Attix, eds., Academic Press, Orlando, 1985, pp. 77-162.

3. M. Löbrich, P.K. Cooper, and B. Rydberg, Non-random distribution of DNA double-strand breaks induced by particle irradiation, *Int. J. Radiat. Biol.* 70:493–503 (1996).

4. D.T. Goodhead, Relationship of Microdosimetric Techniques to Applications in Biological Systems. In: *The Dosimetry of Ionizing Radiation, Vol. II,* K. Kase, B. Bjarngard and F. Attix, eds., Academic Press, Orlando, 1987, pp. 1-89.

5. A. Kraxenberger, Verbesserungen der Methoden zur Pulsfeld-Gelektrophorese and Anwendung der Technik auf die Analyse von DNA-Schaeden durch dicht ionisierende Strahlung, *Thesis*, Technische Universitaet Muenchen, Munich Germany, 74-91 (1996).

6. H.C. Newman, K.M. Prise, M. Folkard and B.D. Michael, DNA double-strand break distributions in X-ray and alpha-particle irradiated V79 cells: evidence for non-random breakage, *Int. J. Radiat. Biol.* 71, 347-63 (1997).

7. A. Kraxenberger, A.A. Friedl and A.M. Kellerer, Computer simulation of pulsed field gel runs allows the quantitation of radiation-induced double-strand breaks in yeast, *Electrophoresis* 15, 128-36 (1994).

8. D.J. Brenner, Track structure, lesion development, and cell survival, *Radiat. Res.* 124, S29-37 (1990).

9. A. Chatterjee and W.R. Holley, Energy deposition mechanisms and biochemical aspects of DNA strand breaks by ionizing radiation, *Int. J. Quantum Chem.* 391, 709-727 (1991).

10. B. Rydberg, M. Löbrich and P.K. Cooper, DNA double strand breaks induced by high-energy neon and iron ions in human fibroblasts. I. Pulsed field gel electrophoresis method, *Radiat. Res.* 139, 133-141 (1994).

11. B. Cedervall, R. Wong, N. Albright, J. Dynlacht, P. Lambin and W.C. Dewey, Methods for the quantification of DNA double-strand breaks determined from the distribution of DNA fragment sizes measured by pulsed-field gel electrophoresis, *Radiat. Res.* 143, 8-16 (1995).

12. A.A. Friedl, A. Kraxenberger and F. Eckardt-Schupp, An electrophoretic approach to the assessment of the spatial distribution of DNA double-strand breaks in mammalian cells, *Electrophoresis* 16, 1865-74 (1995).

13. W.R. Holley and A. Chatterjee, Clusters of DNA induced by ionizing radiation: formation of short DNA fragments. I. Theoretical modeling, *Radiat. Res.* 145, 188-99 (1996).

14. B. Rydberg, Clusters of DNA damage induced by ionizing radiation: formation of short DNA fragments. II. Experimental detection, *Radiat. Res.* 145, 200-9 (1996).

15. D.E. Lea, *Actions of radiations on living cells*, University Press, Cambridge, 1946.

16. R.K. Sachs, D.J. Brenner, P.J. Hahnfeldt, and L.R. Hlatky, A formalism for analyzing large-scale clusters of radiation-induced breaks along chromosomes, *Int. J. Radiat. Biol.* 73, 000-000 (1998).

17. V.V. Moiseenko, A.A. Edwards, H. Nikjoo and W.V. Prestwich, Modeling of Chromosome Exchanges in Human Lymphocytes Exposed to Radiations of Different Quality. In: *Microdosimetry: An Interdisciplinary Approach,* D.T. Goodhead, P. O'Neill and H.G. Menzel, eds., Royal Society of Chemistry, Cambridge, UK, 1997, pp. 152-155.

18. W. Friedland, P. Jacob, H.G. Paretzke, M. Perzl and T. Stork, Simulation of Strand Breaks and Short DNA Fragments in the Biophysical Model PARTRAC. In: *Microdosimetry: An*

Interdisciplinary Approach, D.T. Goodhead, P. O'Neill and H.G. Menzel, eds., Royal Society of Chemistry, Cambridge, UK, 1997, pp. 43-46.

19. S.G. Andreev, I.K. Khvostunov, D.M. Spitkovsky and V. Chepel, Clustering of DNA Breaks in Chromatin Fibre: Dependence on Radiation Quality. In: *Microdosimetry: An Interdisciplinary Approach,* D.T. Goodhead, P. O'Neill and H.G. Menzel, eds., Royal Society of Chemistry, Cambridge, UK, 1997, pp. 133-136.

20. A. Brahme, B. Rydberg and P. Blomquist, Dual Spatially Correlated Nucleosomal Double Strand Breaks in Cell Inactivation. In: *Microdosimetry: An Interdisciplinary Approach,* D.T. Goodhead, P. O'Neill and H.G. Menzel, eds., Royal Society of Chemistry, Cambridge, UK, 1997, pp. 125-128.

21. A. Ottolenghi, M. Merzagora and H.G. Paretzke, DNA complex lesions induced by protons and alpha-particles: track structure characteristics determining linear energy transfer and particle type dependence, *Radiat. and Environ. Biophys.* 36, 97-103 (1997).

22. D.J. Brenner and J.F. Ward, Constraints on energy deposition and target size of multiply damaged sites associated with DNA double-strand breaks, *Int. J. Radiat. Biol.* 61, 737-48 (1992).

23. V. Michalik, Energy deposition clusters in nanometer regions of charged-particle tracks, *Radiat. Res.* 134, 265-70 (1993).

24. V.V. Moiseenko, A.A. Edwards and N. Nikjoo, Modelling the kinetics of chromosome exchange formation in human cells exposed to ionising radiation, *Radiat. and Environ. Biophys.* 35, 31-5 (1996).

25. T. Cremer, A. Kurz, R. Zirbel, S. Dietzel, B. Rinke, E. Schrock, M.R. Speicher, U. Mathieu, A. Jauch, P. Emmerich, H. Schertan, T. Ried, C. Cremer and P. Lichter, Role of chromosome territories in the functional compartmentalization of the cell nucleus, *Cold Spring Harbor Symposia on Quantitative Biol.* 58, 777-792 (1993).

26. P. Hahnfeldt, J.E. Hearst, D.J. Brenner, R.K. Sachs and L.R. Hlatky, Polymer models for interphase chromosomes, *Proc. Nat. Acad. Sci. U.S.* 90, 7854-7858 (1993).

27. J.Y. Ostashevsky and C.S. Lange, The 30 nm chromatin fiber as a flexible polymer, *J. Biomolec. Struct. and Dynam.* 11, 813-20 (1994).

28. R.K. Sachs, H. Yokota, G. van-den-Engh, B. Trask and J.E. Hearst, A random-walk/giant loop model for interphase chromosomes, *Proc. Nat. Acad. Sci. U.S.* 92, 2710-2714 (1995).

29. A.F. Dernburg, K.W. Broman, J.C. Fung, W.F. Marshall, J. Philips, D.A. Agard and J.W. Sedat, Perturbation of nuclear architecture by long-distance chromosome interactions, *Cell* 85, 745-59 (1996).

30. J. Strouboulis and A.P. Wolffe, Functional compartmentalization of the nucleus, *Journal of Cell Science* 109, 1991-2000 (1996).

31. W.F. Marshall, J.C. Fung and J.W. Sedat, Deconstructing the nucleus: global architecture from local interactions, *Current Opinion in Genetics and Develop.* 7, 259-63 (1997).

32. T. Radivoyevitch and B. Cedervall, Mathematical analysis of DNA fragment distribution models used with pulsed-field gel electrophoresis, *Electrophoresis* 17, 1087-1093 (1996).

33. D.J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, Springer, New York, 1988.

34. D. Koenig and V. Schmidt, *Zufaellige punktprozesse*, B.G. Teubner, Stuttgart, Germany, 1992.

35. R.K. Sachs, A.M. Chen and D.J. Brenner, Review: proximity effects in the production of chromosome aberrations by ionizing radiation, *Int. J. Radiat. Biol.* 71, 1-19 (1997).

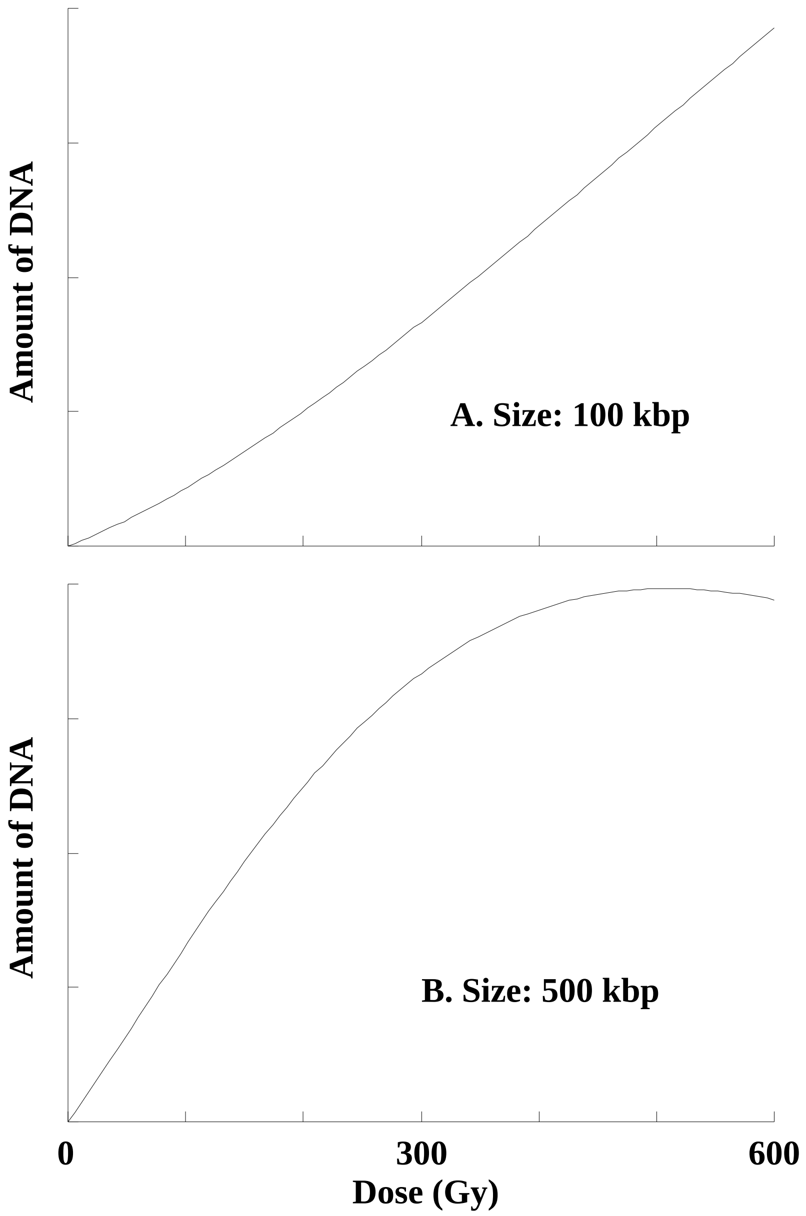
36. T. Radivoyevitch, Mathematical models of misrejoining DNA double-strand breaks and DNA fragment size distributions, *PhD Thesis*, Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, SC 29425 (1997).

37. R.K. Sachs, P. Hahnfeldt, and D.J. Brenner, Review: The link between low-LET dose-response relations and the underlying kinetics of damage production/repair/misrepair, *Int. J. Radiat. Biol.* 72, 351-374 (1997).

Figure 1. Predicted dose-response curves. The panels show amount of DNA, as predicted by the RLC (randomly-located-clusters) model with the parameters given in the text, in a narrow bin centered either on 100 kbp (A) or on 500 kbp (B). 600 Gy was the maximum radiation dose used in the experiments [3]. The scales for the vertical axes are arbitrary and are not the same in panels A and B. In panel A, the upward curvature (see text) is readily visible but a clear-cut downward curvature occurs only at doses larger than those shown; in panel B the downward curvature at large doses is obvious but the upward curvature at smaller doses (though present) is hard to see.

Table 1. DNA fractions. The first column gives the minimum size, in kbp, of the sizes in a bin used when measuring DNA fractions after nitrogen ion irradiation. The second column indicates the maximum size for that bin. The third column indicates the measured fraction, by size (i.e. by mass), of DNA which was found in that bin after irradiation with dose D=189 Gy. Specifically 6 + Log_{10}(fraction) is tabulated; for example the entry 2.95 means the bin contains $10^{2.95-6} \approx 0.00089$ of the total DNA mass. The fourth column gives the predictions of the randomly-located-clusters (RLC) model for that bin. The fifth and sixth columns give the predictions of the clustered-breakage (CB) and random-breakage (RB) models respectively. To enable a direct comparison, the clustered-breakage and random-breakage models were both taken as having 1 degree of freedom when fitting the data shown; the parameters of the RLC model were directly transcribed from the parameters of the clustered-breakage model, rather than being fit to the data.

Min	Max	Data	RLC	CB	RB
0.1	2	2.11	1.99	1.99	0.23
2	9	2.95	2.94	2.95	1.61
9	23	3.76	3.49	3.49	2.37
23	48	3.71	3.90	3.91	2.96
48	97	4.25	4.32	4.32	3.54
97	145	4.30	4.37	4.37	3.73
145	225	4.53	4.62	4.62	4.11
225	365	4.90	4.85	4.86	4.50
365	780	5.37	5.24	5.24	5.13
780	1120	5.13	5.01	5.01	5.09



Amount of DNA
A. Size: 100 kbp
Amount of DNA
B. Size: 500 kbp
0
300
600
Dose (Gy)