Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 245 (2007) 161–168
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Modelling the probability distribution of the number of DNA double-strand breaks due to sporadic alkylation of nucleotide bases

K. Walters*

Division of Genomic Medicine, School of Medicine and Biomedical Sciences, University of Sheffield, Sheffield S10 2RX, UK

Received 27 June 2006; received in revised form 26 September 2006; accepted 26 September 2006
Available online 1 October 2006

Abstract

Metabolites and certain chemical agents (for example methyl methanesulfonate) can induce nucleotide bases on chromosomal strands to become alkylated. These alkylated sites have the potential to become single-strand chromosomal breaks, a form of DNA damage, if they are exposed to a sufficient temperature in vitro. It has been proposed that a single-strand break (SSB) sufficiently close to another SSB on the opposite chromosomal strand will form a double-strand break (DSB). DNA repair mechanisms are less able to repair DSBs compared to SSBs. Because of the complex three-dimensional structure of DNA, some chromosomal regions are more susceptible to alkylation than others. A question of interest is therefore whether these alkylated bases are randomly distributed or tend to be clustered. Pulsed-field gel electrophoresis allows the number of DNA fragments (and hence the number of DSBs) to be observed directly. The randomness of alkylation events can therefore be tested using the standard statistical hypothesis-testing framework. Under the null hypothesis, that the SSBs are randomly distributed on each of the strands, we can calculate the probability of observing a number of DSBs at least as large as that observed and hence the associated p-value. Previously, the probability distribution of the number of DSBs has been determined by Monte Carlo simulations; when considering the whole genome this can be very time consuming. In this paper, we theoretically derive an approximation to the distribution enabling appropriate probabilities to be calculated quickly. Based on previous findings we assume that the number of breaks on each strand is small compared to the number of nucleotide bases. We show that our method can give the correct probability distribution when alkylation events are relatively rare, discuss how rare these events have to be and suggest potential extensions to the model when a greater proportion of bases are alkylated.

Keywords: Probability model; Alkylation; DNA damage

1. Introduction

Mathematical modelling of DNA double-strand breaks (DSBs) is a well established area of research (Holley and Chatterjee, 1996; Kraxenberger et al., 1998; Löbrich et al., 1996; Newman et al., 1997; Ponomarev et al., 2000; Sachs et al., 1998, 1999). Most of the literature, however, focuses on modelling the DNA fragment size distribution under various models of DSB formation occurring as a result of ionizing radiation (Ponomarev et al., 2000; Radivoyevitch and Cedervall, 1996; Sachs et al., 1998). Ionizing radiation is known to cause DSBs directly (Ward, 1994) but the process of DSB formation from alkylation events appears to be less well understood. There is, however, an important postulated distinction between the breaks caused by alkylation events and those occurring as a result of ionizing radiation damage: alkylation of bases occurs on individual strands and if an alkylation event resolves itself as a break, it will form a single-strand break (SSB), not a DSB directly. DSBs are more difficult for DNA maintenance machinery to repair than SSBs. DSBs can be repaired by homologous recombination (van Gent et al., 2001), but critically the level of homologous recombination impacts upon genome stability (Richardson et al., 2004) which has been implicated in tumour progression (Liu et al., 2004). The mechanism by which SSBs become DSBs is not clear but

*Mathematical Modelling and Genetic Epidemiology Group, Division of Genomic Medicine, EU 28, School of Medicine & Biomedical Sciences, Beech Hill Road, SHEFFIELD, S10 2RX, United Kingdom.
Tel.: 44 114 271 3046; fax: 44 114 271 1711.
E-mail address: k.walters@sheffield.ac.uk.

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.09.028

one study has suggested that it is the proximity of SSBs on opposite chromosomal strands that cause a DSB to occur (Lundin et al., 2005).

The mathematical models of DSB formation occurring as a result of ionizing radiation fall into two categories: low and high linear energy transfer models. In low linear energy transfer models, each ionizing event is assumed to induce at most one DSB. The DSBs are modelled by a Poisson process and this naturally leads to the well-known "broken sticks" approach, first formulated in the species abundance literature (MacArthur, 1957) and subsequently applied to problems in other research fields (Bhattacharya et al., 2003; Clark, 1999; Walters and Cannings, 2005). In high linear energy transfer models, exposure to ionization radiation is assumed to create multiple ionization events that are physically close to one another (Sachs et al., 1999). This spatial proximity causes DSBs resulting from ionizing radiation to be clustered along chromosomes. As a result of this clustering, the pattern of DSBs observed in high linear energy transfer models has been described by a Poisson cluster process (Daley and Vere-Jones, 1988) where the cluster locations follow a homogeneous Poisson process (i.e. are randomly positioned along the chromosome) and within the clusters the DSBs are modelled by some localized point process (Sachs et al., 1999).

The models in both the low and high linear energy transfer fields assume a (single or clustered) Poisson process along the chromosome where the DNA fragment lengths are determined by the distances between DSB. In our model of DSB formation occurring as a result of alkylation events, we assume a Poisson process along each chromosomal strand so that alkylation events are randomly distributed on each strand. For simplicity, we also assume that every alkylated base becomes a SSB although this assumption is subject to some debate (Lundin et al., 2005). Since we assume that all alkylation events become SSBs, we can refer to the alkylizing process and the process of SSB formation interchangeably. If breaks on opposite strands are sufficiently close then a DSB will occur. If DSBs could result from "close" SSBs occurring on the same strand, then we could model the DSBs using a single Poisson process (as with ionizing radiation damage) with rate twice that of the rate along a single-strand. The models developed in the ionizing radiation literature cannot be correctly applied to our problem because DSBs result only from sufficiently close SSBs on *opposite* strands.

In the models developed in the ionizing radiation literature, the interest usually focuses on the fragment size distribution (Ponomarev et al., 2000; Radivoyevitch and Cedervall, 1996; Sachs et al., 1998). The observed fragment size distribution is then compared to that expected under various models of DSB formation to determine which model best fits the data. In this paper, we assume that the only available data are the number of fragments rather their lengths. We do this because determining the DNA fragment sizes using PFGE is difficult to do with much accuracy, as in fact is estimating the number of DNA fragments. The number of fragments is considered to be the more reliable quantity of the two using PFGE, especially when the fragment sizes are small as in the case of Lundin et al. 2005. Alkylation events often result from contact with chemical agents. Because of the complex folding and winding of DNA during various phases of the cell cycle it follows that certain regions of chromosomal DNA are more prone to alkylation. It is therefore reasonable to consider that alkylated bases are clustered. The motivation for this paper is to derive the distribution of the number of DSBs which will allow statistical testing of the null hypothesis that alkylation events are randomly distributed on each chromosomal strand. Our simulations indicate that the distribution of DSBs is skewed for much of the parameter space (see Figs. 3a–d) and therefore it is not sufficient to calculate the lower moments of the distribution and appeal to standard Normal distribution theory.

Monte Carlo simulations have been extensively used in the ionizing radiation literature in relation to formation of DSBs (Friedland et al., 1998; Friedland et al., 1999; Ottolenghi et al., 1997; Ponomarev and Sachs, 1999; Ponomarev et al., 2001) and have recently been used in the context of alkylation events (Lundin et al., 2005). Lundin et al. (2005) simulate the alkylation process to estimate the distribution of the number of DSBs, given the number or alkylation events, under the null hypothesis of a poisson process for SSB formation. They do this to determine if there is statistical evidence that alkylations induced by two separate alkylating agents (methly methanesulfonate (MMS) and *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG)) are randomly distributed. They simulate 2,700,000 and 830,000 alkylation events for MMS and MMG, respectively. Because of the time required to simulate such a large number of events (considered further in the discussion), the authors only sample 1000 replicates in the Monte Carlo simulations. This modest number of replicates could potentially yield a relatively large standard error for the probabilities. Here we seek an analytical approximation to the distribution of the number of DSBs to circumvent these problems. The most important issue is the accuracy of the derived approximation; we must ensure that it is more accurate to use our analytically derived approximated values than to use simulated probabilities with their associated uncertainties.

In deriving our approximation, we make a simplifying assumption: the maximum separation of SSBs on opposite strands that leads to a DSB is small in relation to the expected distance between SSBs in the combined Poisson process. This assumption is consistent with the values Lundin et al. (2005) used in their study. They assumed 0.044% and 0.014% of bases would become alkylated when treated with MMS and MNNG, respectively; equivalent to expected distances between SSBs of 1111 and 3614 for MMS and MNNG, respectively. They used 14 base pairs as the distance that single-strand breaks on opposite strands would have to be within to form a DSB. These quantities certainly satisfy the assumption used in