On the consistency of Monte Carlo track structure DNA damage simulations

Piotr Pater,[a] Jan Seuntjens, and Issam El Naqa
McGill University, Montreal, Quebec H3G 1A4, Canada

Mario A. Bernal
Instituto de Fisica Gleb Wataghin, Universidade Estadual de Campinas, Campinas 13083-859, Brazil

(Received 8 April 2014; revised 13 October 2014; accepted for publication 31 October 2014; published 18 November 2014)

Purpose: Monte Carlo track structures (MCTS) simulations have been recognized as useful tools for radiobiological modeling. However, the authors noticed several issues regarding the consistency of reported data. Therefore, in this work, they analyze the impact of various user defined parameters on simulated direct DNA damage yields. In addition, they draw attention to discrepancies in published literature in DNA strand break (SB) yields and selected methodologies.
Methods: The MCTS code Geant4-DNA was used to compare radial dose profiles in a nanometer-scale region of interest (ROI) for photon sources of varying sizes and energies. Then, electron tracks of 0.28 keV–220 keV were superimposed on a geometric DNA model composed of 2.7×10^6 nucleosomes, and SBs were simulated according to four definitions based on energy deposits or energy transfers in DNA strand targets compared to a threshold energy E_{TH}. The SB frequencies and complexities in nucleosomes as a function of incident electron energies were obtained. SBs were classified into higher order clusters such as single and double strand breaks (SSBs and DSBs) based on inter-SB distances and on the number of affected strands.
Results: Comparisons of different nonuniform dose distributions lacking charged particle equilibrium may lead to erroneous conclusions regarding the effect of energy on relative biological effectiveness. The energy transfer-based SB definitions give similar SB yields as the one based on energy deposit when $E_{TH}\approx10.79$ eV, but deviate significantly for higher E_{TH} values. Between 30 and 40 nucleosomes/Gy show at least one SB in the ROI. The number of nucleosomes that present a complex damage pattern of more than 2 SBs and the degree of complexity of the damage in these nucleosomes diminish as the incident electron energy increases. DNA damage classification into SSB and DSB is highly dependent on the definitions of these higher order structures and their implementations. The authors' show that, for the four studied models, different yields are expected by up to 54% for SSBs and by up to 32% for DSBs, as a function of the incident electrons energy and of the models being compared.
Conclusions: MCTS simulations allow to compare direct DNA damage types and complexities induced by ionizing radiation. However, simulation results depend to a large degree on user-defined parameters, definitions, and algorithms such as: DNA model, dose distribution, SB definition, and the DNA damage clustering algorithm. These interdependencies should be well controlled during the simulations and explicitly reported when comparing results to experiments or calculations.
 [http://dx.doi.org/10.1118/1.4901555]

Key words: Monte Carlo track structure, strand break yields, DNA, genome structure

1. INTRODUCTION

Ionizing radiation causes single and double strand breaks (SSBs and DSBs) in DNA either through direct interactions with the DNA itself, such as ionizations of the phosphodiester groups (PDGs), or through indirect interactions with chemical species produced by water radiolysis, notably hydroxide radicals (OH•). A damaged DNA molecule can potentially lead to lethal consequences for the cell, to mutations, or it can also be flawlessly repaired.

Radiation therapy makes use of the cell-killing ability of radiation to treat cancer patients. The study of the effects induced by ionizing radiation on DNA, through experiments or simulations, is of great interest in the medical physics community where various ionizing radiation types and energies are used to treat cancer patients with advanced techniques. In previous work,[1] we argued that Monte Carlo track structure (MCTS) simulations, validated by adequate experimental data, and implemented in a bottom-up framework linking microdosimetric quantities such as physical interactions with DNA to biological end-points such as cell survival, could be possible[2–4] and would allow for estimating tumor control probability (TCP) and normal tissue complication probability (NTCP), from first principles.

MCTS codes were developed (see review in Ref. 5) to track particles to subionization energies and DNA damage yields can be obtained by superimposing these electron tracks on to a geometric DNA model,[6,7] which could be atomistic[8,9] or by

postprocessing these tracks with a probabilistic model.[10–12] Some codes also simulate indirect effects that can amount to an important proportion of the damage.[13,14] For instance, Holley and Chatterjee[15] proposed a general theoretical model to simulate direct and indirect DNA damage, without relying on event-by-event tracking of particles. Common MCTS codes such as Partrac,[16] PITS,[17] Kurbuc,[5] Geant4-DNA,[18] and Penelope[19] rely on somewhat different empirical, semiempirical, and/or experimental models of interaction cross-sections that influence the simulated DNA damage yields. For instance, Li *et al.*[20] compared DNA damage simulated using six different electron inelastic cross-sections in liquid water and concluded that significant differences in SSBs and DSBs are expected. It is now well documented that electrons of kinetic energies around 100 eV are the major contributors to direct radiation damage of DNA in a cell.[21] Although inelastic cross-sections for DNA molecules were measured and theoretically modeled,[22–27] many MCTS codes still rely on liquid or gaseous water cross-sections. The measurement and theoretical modeling of cross-sections in liquid water and/or DNA is challenging, as high-energy theories gradually fail as energy decreases.[28,29] Zhang and Tan[30] proposed a calculation method to incorporate knowledge of electron cross-sections in base-pair molecules, and show for instance, that guanine–cytosine (GC) base pairs experience more frequent damage than adenine–thymine (AT) pairs.

It has been shown that a step-by-step tracking of low energy electrons (LEEs) may violate the Heisenberg uncertainty principle, which questions the experimental significance of such simulations.[31] Nevertheless, recent work by Liljequist and Nikjoo provide a different perspective on this paradox through the concept of circumstantial validity.[32] They estimate that electrons of 100 eV in liquid water have a relative error in position or in momentum of 1%–2%, under certain conditions. In the view of these results, and previous experimental validations of MCTS simulations,[33,34] we presume they can be applied to simulate relative DNA damage as a function of irradiation setup and parameters.

We acknowledge that the choice of the MCTS code, with associated cross-sections, impacts the simulated direct strand break (SB) yields, however user-controlled parameters also have an important role. For instance, users create region of interests (ROI), define types and dimensions of sources, implement different classifications and clustering of DNA damage, and set limits on the tracking cutoff energy of electrons (E_{cutoff}) and the threshold energy for the creation of an SB (E_{TH}). In this work, we investigate the impacts of these parameters on simulated direct DNA damage yields. Specifically, we analyze the impact of the following parameters: dose distribution dependence, SB definition, choice of E_{TH}, nucleosome damage patterns, and SB classification in clusters. We also compare simulated SSB, DSB, and complex SSB (SSB+) after irradiation with electron sources of 0.28 keV, 1.5 keV, 5 keV, 10 keV, and 220 keV. This work explores the effects of user-defined parameters and their impacts on simulated DNA damage yields for given scenarios. We mainly focus on examples and parameters governing the direct effect of electron irradiation, however conclusions are also valid for mixed direct–indirect simulations and simulations with other particles types.

2. MATERIALS AND METHODS

2.A. Track structure simulations

MCTS simulations were carried out using the Geant4 (4.9.5) simulation toolkit[35] with the Geant4-DNA processes.[18] The simulation phantom was composed of three enclosed volumes filled with liquid water: the water slab, the low energy process region (LEPR), and the ROI. Simulations are conducted in a liquid water medium, as DNA material represents a small fraction of the ROI, and cross-sections for DNA targets are not so detailed as those used in Geant4-DNA for liquid water. In recent work by de Vera *et al.*,[36] it was found that despite they accounted for differences of chemical composition and cross-sections, the specific energy deposited by ions and secondary electrons in the cytoplasm and nucleus was practically equal. Nevertheless, the liquid water density was scaled to 1.06 g cm^{-3} to approximate the density of a cell nucleus.[37] The ROI and LEPR are centered in the water slab, which is a cube with sides of 1 cm. The LEPR is a virtual volume surrounding the ROI in which the following Geant4-DNA models and processes, from the G4EMLOW6.23 data file, were active: Champion elastic,[38] Born ionization and excitation (see details in Ref. 18), Melton attachment, and Sanche excitation. Geant4 distinguishes itself from other codes notably because all created particles are tracked to zero range and there is no tracking cutoff. Therefore, an additional process (G4eCapture) was used to stop electrons with kinetic energy lower than a cutoff energy (E_{cutoff}) and deposit locally their energy. A value of $E_{\text{cutoff}} = 10.79$ eV, equal to the lowest ionization potential of liquid water, was used in this work. Therefore, subionization electrons were not tracked, and their energy deposited locally. These low energy electrons have a residual range that could allow them to reach DNA strand targets and potentially lead to additional, nonsimulated, SBs through resonant effects.[39,40] Outside the LEPR, electrons are tracked using multiple scattering, Livermore ionization, and bremsstrahlung models to preserve accuracy, while reducing the calculation time. Photons are followed using the Livermore models for photoelectric effect, Compton, and Rayleigh scatterings. Fluorescence and Auger electron de-excitations are active and particles are produced if their energy is higher than 14 eV. All interactions depositing energy in the ROI are saved to a binary file, which is further analyzed with MATLAB (R2012a, The MathWorks, Natick, MA).

2.B. Impact of dose distribution

The objective of this section is to demonstrate that the choice of source size and type changes the dose distribution with consequences on simulated DNA damage. A cylindrical ROI of 15 nm radius and half-height of 525 μm, is irradiated by photon sources of 1.5 keV or 1.25 MeV. We used a similar irradiation setup as presented by Bernal and Liendo to emulate

their results.[6] Ten simulations of dose distributions binned radially into 100 equal-volume cylindrical shells in the ROI were obtained, each for a total ROI dose of 100 Gy, and averaged. The mean dose per bin is reported as a function of the bin's maximum radius. For 1.5 keV, photons were generated isotropically and uniformly throughout a cylindrical volume of size (a) equal to the ROI or (b) exceeding the ROI by 0.6 μm isotropically, with the LEPR size equal to the source size. For 1.25 MeV, photons were generated in a beam centered on the ROI central axis. The beam was either (c) a pencil beam with the LEPR size equal to the ROI, or (d) a 1 μm-radius circular parallel beam with the LEPR size exceeding that of the ROI by 1 μm isotropically. In irradiation setup (b), charged particle equilibrium (CPE) is achieved in the ROI, whereas it was not achieved for all other cases.

2.C. DNA models and strand break definition

This section discusses different DNA models in the literature by comparing the size of the DNA strand target and contrasts four definitions of the simulated SB. DNA models consist of spatial arrangements on several organizational levels of volumes corresponding to base pairs (bps) and PDGs. In Bernal and Liendo,[6] the PDGs were modeled as prisms with circular sector base with a volume of 0.24 nm^3. In Bernal *et al.*,[41] the volume was reduced to 0.13 nm^3 to avoid overlapping, while preserving the overall shape. Friedland *et al.*[8,42] model the PDG as the union of spheres centered at the positions of all constituent atoms (1 Phosphor, 5 Oxygen, 7 Hydrogen, and 5 Carbon) with radii corresponding to the atoms van der Waals radii (respectively, 0.19, 0.14, 0.12, 0.17 nm). We calculated the volume of the union of these spheres and obtained 0.13 nm^3. In addition, 60% of interactions occurring within the water shell (union of spheres of 0.35 nm radius centered on each atoms center) also contributed to the creation of SBs. We calculated that the total effective volume of the direct DNA strand target in their model is close to 0.36 nm^3. Charlton *et al.*[43] and Nikjoo *et al.*[44] simulated DNA strand targets as half cylindrical shells of volume 1.73 nm^3 inside small cylinders. In their probabilistic model, Francis *et al.*[10] used an adjustable empirical parameter to sample interactions located in DNA strand targets. On the other hand, in the model presented by Semenenko and Stewart,[11,12] DNA strand targets volumes are not simulated, and only genomic distances are distributed.

In addition to differences in size and geometry of the DNA strand target and the whole geometric DNA model, the SB definition varies also from one study to another. One definition (Sum E_{dep}) requires that the sum of all energy deposits in a DNA strand target exceeds a threshold energy E_{TH} in order to create an SB. To define the SB, various authors used E_{TH} of 10–2000,[7] 10.79,[37,45] and 17.5 eV.[43] A ramp probability function (linearly increasing from 0 to 1 between 5 keV and 37.5 eV[46] or 40 eV[8]) was also proposed. Francis *et al.*[10] used this approach to look at maximal energy deposits in DNA strand targets. Alternative definitions include (Max E_{dep}) requiring that the maximal energy deposit in a DNA strand target exceeds E_{TH}, (Sum E_{trans}) requiring that the sum of all energy transfers exceeds E_{TH} and (Max E_{trans}) requiring that the maximal energy transfer exceeds E_{TH}.[6,41]

The DNA model we used is described in details by Bernal *et al.*[6,41] It consists of 2.7×10^6 independent DNA nucleosomes, arranged in groups of 6 around a 30 nm diameter circle and stacked 500 times to form the 30 nm chromatin fiber. The ROI is filled with 900 copies of the 30 nm chromatin fiber, totaling 524.6×10^6 bps, arranged around a cylindrical shell of 2625 nm half-height, with inner and outer radii of 4984.76 and 5015.24 nm, respectively. Each nucleosome consists of a double stranded two-turn structure of 198 bps. Each PDG volume is equal to 0.1344 nm^3 and each nucleosome turn has a diameter of 10.5 nm and a thickness of 2.37 nm. The ROI was irradiated with monoenergetic electrons generated isotropically and uniformly throughout a cylindrical volume with radial and half-height dimensions exceeding those of the ROI by 0.4 μm for 0.28 keV, 0.6 μm for 1.5 keV, 1.0 μm for 5 keV, and 2.5 μm for both 10 keV and 220 keV electrons sources. The LEPR volume was equal to the source volume for each energy. The source size was made large enough to achieve CPE inside the ROI except for the 220 keV source. For this case, a source exceeding the ROI by 0.5 mm would be necessary to achieve CPE. Such a source is larger by an order of 10^5 than the one we used and would be computationally too expensive to simulate. The full set of inelastic interactions inside the ROI is processed by an in-house algorithm to rapidly determine which interactions happened inside the DNA strand targets. Our algorithm also identifies SBs according to the four strand break definitions (Sum E_{dep}, Max E_{dep}, Sum E_{trans}, and Max E_{trans}) and was used to compare the total number of SBs (TSB) for all simulations. In six consecutive and similar steps, our algorithm finds all interactions occurring within six subunits: (1) a 30 nm chromatin fiber, (2) a group of 6 nucleosomes, (3) a nucleosome, (4) a nucleosome turn, (5) a base pair, and (6) a DNA strand target. In each step, a state is associated with every inelastic interaction at once. For instance, in step (1), interactions will be associated with state "0" if they are outside of all 30 nm chromatin fibers or to a state between "1" and "900" corresponding to the position of the 30 nm chromatin fiber that was hit. Once a state is obtained, we use symmetry considerations to move all interactions within a representative structure for the given subunit. This is repeated for each subunits and the set of six states for each interaction determines if and which DNA strand target will be hit.

2.D. Nucleosome damage patterns

In this part, we estimated the frequency and the complexity of damage in nucleosomes. Each nucleosome is a circular double-stranded DNA fragment of 198 bps. These nucleosome damage patterns give an estimate of the complexity of the damage created and are independent of SB clustering algorithms. We used the SBs determined with Sum E_{dep} and $E_{\mathrm{TH}} = 10.79$ eV. Our algorithm determines how many nucleosomes per unit dose show at least one, exactly one, two, three, four, or more than five SBs as a function of the incident electron energy.

2.E. Differences in strand break clustering algorithms

We implemented three algorithms from the literature to cluster SBs into complex DNA damage types in addition to one other algorithm proposed by us. Typically, SB distributions are clustered using inter-SB distances and the knowledge of the strand that was hit. Charlton and Humm[47] definitions (SSB, SSB+, 2 SSB, DSB, DSB+, DSB++) of clustered damage are commonly used, notably the DSB is defined as a pair of SBs on opposite strands located within 10 bps. These definitions are adapted and implemented differently depending on the authors and the DNA models adopted. Figure 1 compares the expected clustered DNA damage as obtained by three different algorithms from the literature (Bernal,[6,41] Friedland,[37] and Charlton and Nikjoo[7,43]) and our own proposed clustering algorithm. Bernal's algorithm looks for pairs of SBs in a unidirectional manner, and categorizes each SB in more than one complex type of damage. This results in scoring 2 DSBs if 3 alternating SBs are found within 10 bps. The algorithm also scores a complex SSB (SSB+) when a pair of SBs within 10 bps on the same strand is found. Also, all SBs not related to DSBs are counted as SSBs, including those only leading to SSB+. On the other hand, Friedland's algorithm starts by scoring all DSBs with a preference given to the closest pairs of SBs. Then, all remaining SBs are classified as SSBs, except when two directly adjacent SBs on the same strand are found : this case is counted as one SSB. Using Charlton's and Nikjoo's definitions, the example damage pattern in Fig. 1 is simply classified as a DSB++, a type of damage where both strands are hit twice or more, creating a short unbound DNA fragment that can potentially lead to a deletion after incomplete repair. To compare their classification created for linear DNA fragments, we defined the $\mathrm{SSB}_{\mathrm{NIKJOO}} = (\mathrm{SSB}) + 2*(2\ \mathrm{SSB}) + 2*(\mathrm{SSB+}) + (\mathrm{DSB+})$, to count the total number of SSBs, i.e., 1 for SSB, 2 for 2SSB, 2 for SSB+, and 1 for DSB+, which is composed of 1 SSB and 1 DSB. We also defined the $\mathrm{DSB}_{\mathrm{NIKJOO}} = (\mathrm{DSB}) + (\mathrm{DSB+}) + 2*(\mathrm{DSB++})$, to count the total number of DSBs, i.e., 1 for DSB, 1 for DSB+, and 2 for DSB++, which could potentially be composed of more than 2 DSBs.

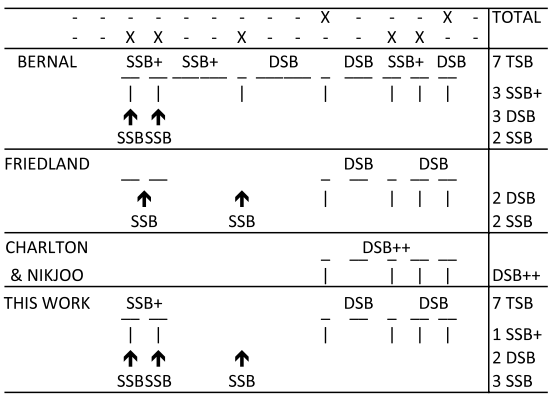

FIG. 1. Comparison of various SSB, DSB, complex single strand breaks (SSB+), and TSB clustering conditions, including the ones from Bernal and Liendo (Ref. 6), Friedland *et al.* (Ref. 37), Charlton and Nikjoo (Refs. 7, 43, and 47), and our own implementation. The top part represents two DNA strands with x representing a strand break and - representing an unaffected DNA strand target. The DSB++ can be understood as a double DSB where each strand is hit at least two times within 10 bps.

In addition, a modified algorithm is proposed that finds the same DSB yields as the Friedland's algorithm, and the same SSB yields as Bernal's one. Our approach, also shown in Fig. 1, starts by looking for the closest pairs of opposite strand SBs and categorizes each as a DSB. When no more DSBs are present, the algorithm looks for pairs of SBs on the same strand and categorizes them as SSB+. These pairs have to be within 10 bps to be counted in these complex damage types. In addition, all SBs unrelated to DSBs are counted as an SSB and the sum of all individual SBs gives the TSB. Using the Sum E_{dep} SB definition (see Sec. 2.C), we compare the percent difference of SSB and DSB yields between all these classification rules for our set of data. In addition, we report the TSB, SSB, SSB+, and DSB yields according to our own clustering conditions as average values over ten batches. The total dose in each simulation was 1000 Gy, except for 10 keV and 220 keV where only 100 Gy was delivered to reduce simulation time.

3. RESULTS

3.A. Impact of dose distribution

The mean dose in the ROI cylindrical shells irradiated with photon sources described in Sec. 2.B as a function of the outer radius of the shell is given in Fig. 2. In all cases, the total dose to the ROI was 100 Gy, but the radial dose distributions differ. For 1.5 keV photons, a source that produces CPE in the ROI (●) yields a constant dose of 100 Gy in all shells, within the statistical uncertainty. In contrast, when CPE is not achieved

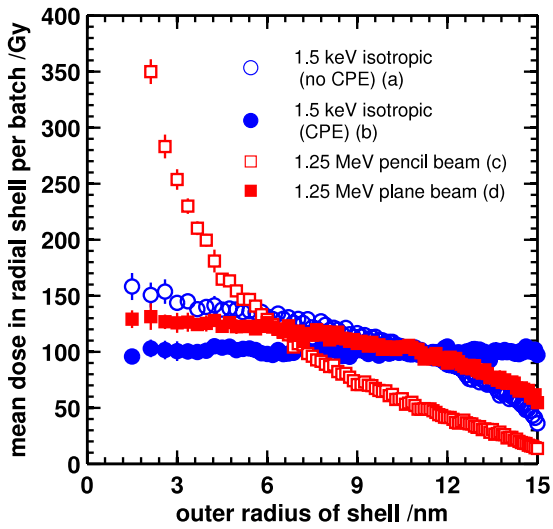

FIG. 2. Radial dose distribution in the 15 nm-radius cylindrical ROI after the irradiation with isotropic sources of 1.5 keV photons (○, ●) or plane beams sources of 1.25 MeV photons (□, ■) with the following dimensions: (a) ○ source size = LEPR size = ROI size, (b) ● source size = LEPR size = ROI size + 0.6 μm in all directions, (c) □ pencil beam with LEPR size = ROI size, and (d) ■ plane beam (radius = 1 μm) with LEPR size = ROI size + 1 μm in all directions. CPE in the ROI is attained only for ●, the source (b). All shells have equal volume. Error bars correspond to two times the standard deviation of the mean.

(○), a nonuniform radial dose distribution is obtained, with a peak dose of 160 Gy in the center and a low dose of 38 Gy around the edge of the ROI. For 1.25 MeV photons, none of the presented simulations yields an uniform dose in the ROI. The plane beam with 1 μm radius (■) produces a dose of 130 Gy near the center and of 60 Gy near the ROI boundary. A pencil beam of 1.25 MeV photons (□) produces a dose distribution that is heavily peaked around the center of the ROI. At this energy, a 1 cm-diameter beam should be simulated in order to achieve CPE within the ROI.

3.B. Impact of the strand break definition

The TSB is dependent on the definition of the SB and on the value of E_{TH}. Figure 3 compares the simulated TSB as a function of E_{TH} for the Sum E_{dep}, Max E_{dep}, and Max E_{trans} SB definitions. The Sum E_{trans} definition is not shown but follows a similar trend as Max E_{trans}. For all cases, the TSB decreases monotonically as E_{TH} increases. The Max E_{dep} curves follow step-like functions related to the different ionization and excitation potentials of liquid water.[20] The Sum E_{dep} curves decrease continuously and no significant dependence on the incident electron energy is observed. For both sets of curves based on energy deposited virtually, no SBs are created for $E_{TH} \geq 100$ eV. Both set of curves based on energy transfer exhibit an energy dependence, i.e., for a fixed E_{TH} value, more SBs are produced for higher incident energies. An interesting point is that if $E_{TH} \leq 10.79$ eV energy transferred and energy deposited are now equal for all interactions. Finally, definitions based on maximum and total energies differ only when more than one interaction occurs inside a given DNA strand target.

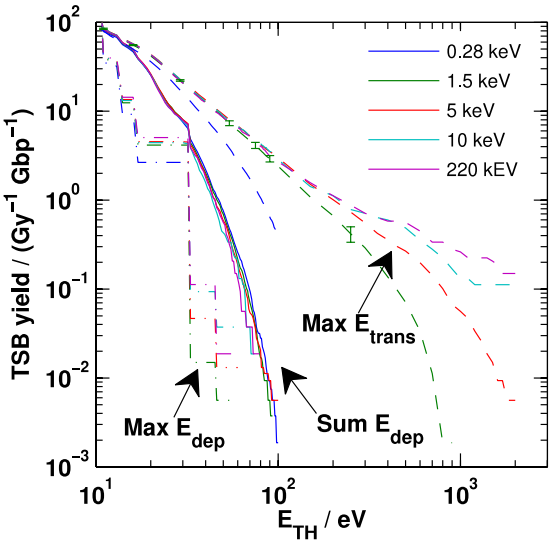

FIG. 3. TSB yields as a function of the thresholding energy E_{TH} for simulations with 0.28, 1.5, 5, 10, and 220 keV electron sources. Dashed–dotted curves represent the SB definition Max E_{dep}, solid curves represent Sum E_{dep}, and dashed curves represent Max E_{trans} (see Sec. 2.C). Curves based on Sum E_{trans} were omitted for clarity but mainly follows similar trends as Max E_{trans}. Representative error bars are shown on one curve (Max E_{trans} at 1.5 keV) and equal to one standard deviation of the mean over ten simulations.

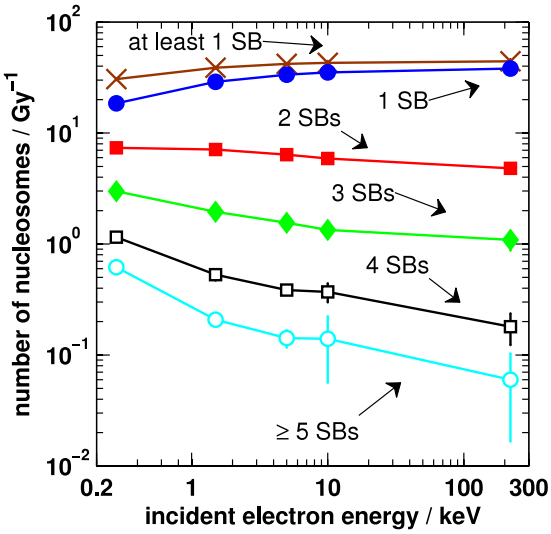

FIG. 4. Number of nucleosomes (per Gy) presenting at least 1 (×), exactly 1 (●), 2 (■), 3 (◆), 4 (□), and 5 or more (○) SBs as a function of the incident electron energy. Error bars correspond to two times the standard deviation of the mean for ten batches.

3.C. Nucleosome damage patterns

Figure 4 shows that only 30.6 nucleosomes at 0.28 keV and 44.2 nucleosomes at 220 keV presented at least one SB per Gy (×) out of 2.7×10^6 nucleosomes in total. Also, the number of nucleosomes hit exactly once (●) increases with incident electron energy, whereas the number of nucleosomes presenting complex damage patterns [2 (■), 3 (◆), 4 (□), and 5+ (○) SBs] decreases with increasing incident electron energy (i.e., decreasing linear energy transfer). Finally, the fraction of highly complex damage (>5 SBs) over the sum of all complex damage diminishes as the incident energy increases. This last observation entails that both the frequency and the complexity of the damage are higher for incident electrons of lower energy.

3.D. Differences in strand break clustering algorithms

Figure 5 presents the percent differences of SSB and DSB yields obtained when using different SB clustering algorithms from the literature compared to our proposed algorithm, using the same initial nucleosome damage patterns, based on the Sum E_{dep} SB definition with $E_{TH} = 10.79$ eV. As expected, our method finds the same SSB yields as in Bernal's and the same DSB yields as in Friedland's. However, our classification detects between 6% and 16% more SSBs than Friedland's, and between 8% and 12% less DSBs than Bernal's as a function of the incident electron energy. Charlton and Nikjoo's algorithm classifies damage into five different categories, and for purposes of comparison with this work, we defined SSB_{NIKJOO} and DSB_{NIKJOO} in Sec. 2.E. Using these definitions, our classification method finds between 12% and 24% more DSBs than using DSB_{NIKJOO} and between 18% and 48% less SSBs than using SSB_{NIKJOO} as a function of energy.

Table I presents TSB, SSB, SSB+, and DSB yields obtained with our proposed SB clustering algorithm and Fig. 6

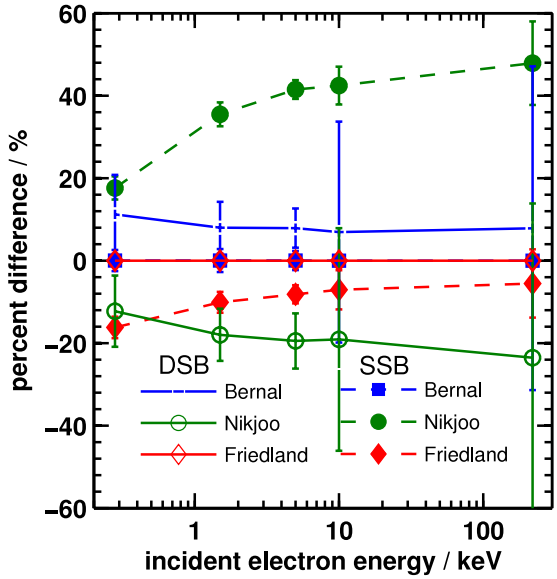

FIG. 5. Percent difference of SSB and DSB yields obtained with strand break clustering algorithms from Bernal (Refs. 6 and 41), Friedland *et al.* (Ref. 37), and Charlton and Nikjoo (Refs. 7 and 43) as compared to our own algorithm. It is defined as % diff = 100% ∗ (value − this work)/this work. Error bars represent one standard deviation of the sample.

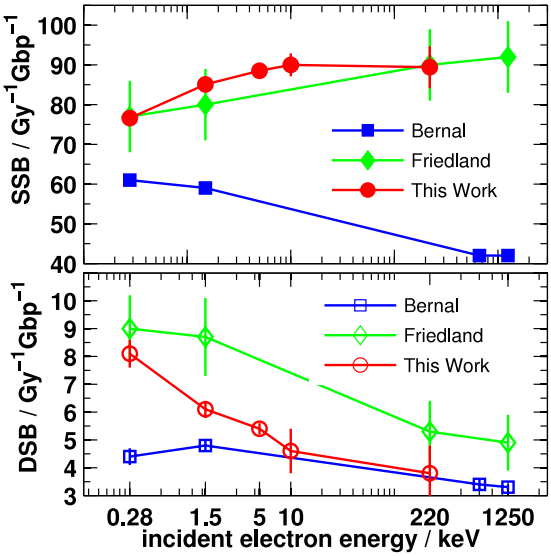

FIG. 6. Simulated direct SSB (●, ◆, ■) and DSB (○, ◇, □) yields comparisons between this work (●, ○), Bernal and Liendo (Ref. 48) (■, □), and Friedland *et al.* (Ref. 37) (◆, ◇). Error bars for "This Work" equal one standard deviation of the sample, whereas literature values are shown with one standard deviation (unspecified).

compares these values (●, ○) to SB yields published by Bernal and Liendo[48] (■, □) and Friedland *et al.*[37] (◆, ◇).

4. DISCUSSION

Table II presents a summary of the investigated parameters and their impact on simulated direct DNA damage from irradiation with electrons sources. Calculations should be seen valid from the relative instead of the absolute point of view. The conclusions presented are valid for irradiations with other particles such as ions or photons, however the absolute numerical differences may differ. Sections 4.A–4.D discuss these issues in depth.

4.A. Impact of dose distribution

In order to obtain full CPE in the ROI, the dimensions of a uniform isotropic source must be larger than the ROI by the maximal secondary electron range, and electrons need to be tracked with track structure processes (i.e., down to subionization or subexcitation energies) throughout the source volume. At incident energies of the order of the kilo-electron-volt, this is computationally achievable, but as the maximum range of electrons increases, this becomes challenging. For instance, looking at the 1.5 keV simulations of Fig. 2, when CPE is achieved, doses in all radial bins are equal, and thus there is an equal interaction density in all bins. When nonuniform interaction densities are superimposed on nonuniform DNA densities in the ROI, this can lead to erroneous conclusions. In the work of Bernal and Liendo,[6] no DNA strand targets were present in the central 4 nm-radius of the ROI. They used a pencil beam for the 1.25 MeV radiation to simulate SBs, which created a strongly peaked radial dose distribution around the center, where dose was deposited, but no SBs created and thus underestimated SSB yields (see Fig. 6). In addition, they compared results by calculating the relative biological effectiveness (RBE) as a function of energy, overlooking the nonuniform interaction densities that contributed to these differences. If a beam is used instead of a volumetric source, the depth-attenuation in the microscopic ROI can be neglected for higher electron energies, but is significant for low energy. This effect was observed by Friedland[37] when homogeneous irradiation yielded an increase of about 10% of the SSB and DSB yields when compared to irradiations using a beam. They used a source of the same size for all simulations, not accounting for lateral CPE which produced slightly different dose distributions in the ROI at each energy. The spectrum of secondary

TABLE I. Direct DNA strand break yields in Gy^{-1} Gbp^{-1}, including TSB, SSB, DSB, and SSB+ obtained using our classification presented in Sec. 2.E for incident electron energies of 0.28, 1.5, 5, 10, and 220 keV.

	0.28 keV	1.5 keV	5 keV	10 keV	220 keV
TSB	93.8 ± 1.9	97.7 ± 1.4	99.7 ± 1.4	99.5 ± 3.7	97.4 ± 6.0
SSB	76.6 ± 1.4	85.1 ± 1.7	88.5 ± 1.4	90.0 ± 2.9	89.4 ± 5.3
DSB	8.1 ± 0.5	6.1 ± 0.3	5.4 ± 0.2	4.6 ± 0.8	3.8 ± 1.0
SSB+	9.7 ± 0.4	7.5 ± 0.3	6.4 ± 0.3	5.9 ± 0.9	5.0 ± 0.6

TABLE II. Impacts of parameter choices in Monte Carlo track structure simulations on DNA strand break yields.

Parameter choices	Impact on DNA strand break yields
Dose distribution in the ROI	Achieving CPE in the ROI guarantees an uniform dose distribution. This can be achieved by using a source with dimensions that exceeds the ROI by at least the range of the highest energy electrons. Similar, but nonuniform dose distributions can also be compared relatively. However, attempting to compare results from different dose distributions in the ROI can lead to erroneous conclusions on the causes of DNA damage differences (see Sec. 3.A).
Cutoff energy E_{cutoff}	This parameter is limited by the MCTS cross-sections. It should be as low as possible to avoid underestimation of the total number of ionizations and thus of SBs. Higher values of E_{cutoff} can potentially lead to lower incidences of complex damage patterns. In addition, electrons with kinetic energy below E_{cutoff}, may also contribute to DNA damage through resonant effects that may or may not be simulated (see Sec. 2.A).
ROI dimensions, genome size and dose	Larger ROI requires longer simulations, and larger sources to achieve CPE, but has no influence on yields. Increasing genome size for fixed ROI dimensions or increasing the dose will decrease the statistical uncertainty on the yields.
DNA strand target size and positions	The DNA strand target volume affects the number of simulated SBs as multiple interactions in the same DNA target are more probable for larger target sizes (Ref. 41). Relative distances between neighboring DNA strand targets can affect clustered yields such as DSBs.
Strand Break definition and threshold energy E_{TH}	SBs defined based on the sum of all energy deposits in a DNA strand target (Sum E_{dep}) or on the maximal energy transfer (Max E_{trans}) are both used in the literature and produce decreasing TSB yields as a function of E_{TH} (see Fig. 3). These definitions are equivalent for E_{TH} = 10.79 eV. For higher values of E_{TH}, energy transfer-based definitions overestimates TSB compared to energy deposit based definitions. In addition, TSB yields defined with energy transfer definitions vary as a function of the incident electron energy.
DNA damage classification	Implementations of DSB and SSB classifications of SBs differ among authors. This can result in differences of up to 54% for SSB yields and of up to 32% for DSB yields, as a function of the incident electrons energy and of the models compared (see Fig. 5). Comparison of DNA damage yields between different authors, or with experimental values should account for this effect.

electrons incident on the ROI is highly dependent on the irradiation setup and blinded comparison based on nominal beam energy can lead to misrepresentation of DNA damage.

4.B. Impact of the strand break definition

The SB definition as well as the choice of E_{TH} impacts the TSB yield as shown in Fig. 3. At all values of $E_{TH} > 10.79$ eV, energy transferred based definitions yield more SBs than energy deposited definitions. This is in accordance with the definition of these energies. The actual value of the E_{TH} parameter is still unknown and it can be seen as a fitting parameter. The user can chose an E_{TH} value that predicts a given experimental or simulated yield of direct strand breaks, as was previously done in other work (see references in Sec. 2.C). One could argue that energy transfer-based definitions are a closer approximation of the amount of energy available to create an SB in a DNA molecule. Most simulations are based on interactions in liquid water, and energy depositions are calculated based on the ionization potentials of liquid water, not DNA molecule constituents. However, as E_{TH} increases, energy transferred definitions exhibit a dependence on the initial spectrum. In other words, such definitions do not conserve the invariance of the TSBs as a function of the incident energy and is not an expected behavior.[41] It is worth to note that for E_{TH} values close to 10.79 eV, TSBs are less dependent on the definition.

4.C. Nucleosome damage patterns

We chose to use the Sum E_{dep} with $E_{TH} = 10.79$ eV definition of an SB as this corresponds to the lowest ionization potential of liquid water, and thus all simulated ionizations in PDGs produce an SB. We did not attempt to reproduce an experimental data set and thus our results can be seen as relatively valid with respect to each other. Although DNA SBs from direct effects of radiation have a low probability of occurring, they are still likely to induce biological effects. As seen in Fig. 4, at all energies, less than 8 nucleosomes/Gy exhibit more than two SBs, becoming even less probable as the incident electron energy is increased. In addition, the complexity of the damage is increased with lower incident electron energies. These results predict that complex DNA damage is more frequent and more complex as the incident electron energy decreases. Nucleosome damage patterns can also be seen as strand break frequencies in DNA fragments of around 200 bps. We believe that nucleosome damage patterns, which depend only on the SB definition, on the value of E_{TH} and on the geometrical DNA model could be good candidates for consistency checks between different MCTS codes. They do not suffer from the added complexity of classification of SBs into SSBs and DSBs, yet they predict similar trends.

4.D. Differences in strand break clustering algorithms

Figure 4 gives only the total number of SBs in a nucleosome, but provides no information on the SBs distribution in the nucleosome. For example, it is well known that two isolated SBs are easier to repair than two clustered SBs on opposite strands.[49] In addition, experimental results are often presented as SSB and DSB yields, defined as either a change in the conformation of the molecule or as the creation of fragments of the DNA molecule. Authors mostly agree on the definition of the DSB, regarded as a pair of SBs on both strands within 10 bps, but implementations

of this definition differ and can lead to discrepancies as shown by our results (see Fig. 5). These differences become more apparent for lower energy electrons, where the damage complexity increases. From Fig. 5, differences in DNA damage clustering of as much as 32% difference in DSB yields and of up to 54% for SSB yields are expected as a function of the clustering implementations and electron energy. In publish data, these relative differences in the clustering algorithms can be masked by differences in MCTS codes, cross-sections, or other parameter choices. In addition, when simulation data are validated against experiments, these differences could be not significant. Nevertheless, we believe that standardizing the reporting of the clustering methods and possibly the methods themselves, with adequate experimental validations, would allow easier interpretation, intercomparisons, and fewer floating-parameters in simulations that may increase the chance of over-fitting the experimental data. Our simulation data shown in Table I confirm that the yields of complex damage (i.e., DSB and SSB+) decreases with increasing energy. This is concordant with our nucleosome damage results that showed a decrease in multiple hits in the same nucleosome and a decrease of the complexity of the damage as the incident electron energy is increased. The TSB yields are invariant with energy except for 0.28 keV, where multiple ionizations in the same DNA strand target become more probable, reducing the TSB yield. This invariance has already been studied in detail for ion beams, using the site-hit probability.[41] For incident electrons, our results show that the SSB yields increase with energy. This is an indirect effect of the TSB invariance with incident energy and the DSB yield decrease with incident energy. Finally, for all energies, SSB+ are slightly more probable than DSBs. This last result is interesting and will be studied further in a future study. We believe it is linked to the fact that the mean distance between two targets on the same strands that can create a SSB+ is shorter than the one between two targets on opposite strands in the DNA model we used. In other words, a second SB (within 10 bps) is more probable to occur in a DNA strand target in the same strand than in the opposite-strand. This result contradicts models that use equal probability to hit either strand.[10–12] Our simulations show an overall good agreement (Fig. 6) with Friedland's results. A slight overestimation of SSBs in our work compared to Friedland could be explained by the fact that Friedland's DNA damage scoring algorithm counts as one SSB two adjacent SBs, whereas they are counted as 2 SSBs (and an SSB+) in our work. Our work underestimates DSB yields compared to those reported by Friedland and we hypothesize that this is the consequence of different sizes of the DNA strand targets and of nonuniform dose distributions used in their work. Bernal and Liendo's results overall underestimate SSB and DSB yields and show a decreasing SSB yield with energy, in contradiction to the expected behavior. Their results are a direct effect of the use of a pencil beam at high energies and of a small volumetric source for low energies, that creates inhomogeneous dose distributions in the ROI, with possibly high dose regions created outside of DNA strand target locations. Finally, we remind the reader that this study did not compare differences in MCTS codes, cross-sections or models, nor did it compare simulations to experimental results, but it was rather focused on end-user controlled parameters and their impacts on end results.

5. CONCLUSION

The main focus of this study was to compare choices of various user-defined MCTS simulation parameters and understand their impact on calculated direct DNA damage yields. Our findings suggest that significant differences arise from subtle modifications of definitions, algorithms, or parameter values, which may or may not impact the validation by experimental data. More specifically, we showed that achieving a uniform dose is ideal for multiple energy studies, and it can be obtained by accounting for the range of the secondary electrons created when defining the incident particle source size, or achieving similar dose distributions for all energies. We compared the impact of the SB definition on TSB yields. We also showed that differences of up to 54% for SSB yields and of up to 32% for DSB yields, can result from slight variations in SB clustering algorithms implemented in the literature. This paper shows quantitatively why we need a forum for "standardization" of MCTS simulation parameters.

ACKNOWLEDGMENTS

The authors would like to acknowledge support from the Fonds de recherche du Québec - Santé, support from the CREATE Medical Physics Research Training Network grant of NSERC (Grant No. 432290), and support from CIHR-MOP-114910. M. Bernal thanks the FAPESP foundation in Brazil for financing his research activities through the 2011/51594-2 project.

[a]Electronic mail: piotr.pater@mail.mcgill.ca

[1]I. El Naqa, P. Pater, and J. Seuntjens, "Monte Carlo role in radiobiological modelling of radiotherapy outcomes," Phys. Med. Biol. **57**(11), R75–R97 (2012).
[2]R. B. Hawkins, "A Microdosimetric-kinetic model for the effect of non-Poisson distribution of lethal lesions on the variation of RBE with LET," Radiat. Res. **160**(1), 61–69 (2003).
[3]R. B. Hawkins and T. Inaniwa, "A microdosimetric-kinetic model for cell killing by protracted continuous irradiation including dependence on LET I: Repair in cultured mammalian cells," Radiat. Res. **180**(6), 584–594 (2013).
[4]T. Friedrich, M. Durante, and M. Scholz, "Modeling cell survival after photon irradiation based on double-strand break clustering in megabase pair chromatin loops," Radiat. Res. **178**(5), 385–394 (2012).
[5]H. Nikjoo, S. Uehara, D. Emfietzoglou, and F. A. Cucinotta, "Track-structure codes in radiation research," Radiat. Meas. **41**(9–10), 1052–1074 (2006).
[6]M. A. Bernal and J. A. Liendo, "An investigation on the capabilities of the PENELOPE MC code in nanodosimetry," Med. Phys. **36**(2), 620–625 (2009).
[7]H. Nikjoo, D. T. Goodhead, D. E. Charlton, and H. G. Paretzke, "Energy deposition in small cylindrical targets by ultrasoft x-rays," Phys. Med. Biol. **34**(6), 691–705 (1989).
[8]W. Friedland, P. Bernhardt, P. Jacob, and H. G. Paretzke, "Simulation of DNA damage after proton and low LET irradiation," Radiat. Prot. Dosim. **99**(1–4), 99–102 (2002).

[9]M. A. Bernal, D. Sikansi, F. Cavalcante, S. Incerti, C. Champion, V. Ivanchenko, and Z. Francis, "An atomistic geometrical model of the B-DNA configuration for DNA–radiation interaction simulations," Comput. Phys. Commun. **184**(12), 2840–2847 (2013).
[10]Z. Francis, C. Villagrasa, and I. Clairand, "Simulation of DNA damage clustering after proton irradiation using an adapted DBSCAN algorithm," Comput. Methods Programs Biomed. **101**(3), 265–270 (2011).
[11]V. A. Semenenko and R. D. Stewart, "A fast Monte Carlo algorithm to simulate the spectrum of DNA damages formed by ionizing radiation," Radiat. Res. **161**(4), 451–457 (2004).
[12]V. A. Semenenko and R. D. Stewart, "Fast Monte Carlo simulation of DNA damage formed by electrons and light ions," Phys. Med. Biol. **51**(7), 1693–1706 (2006).
[13]H. Nikjoo, P. O'Neill, M. Terrissol, and D. T. Goodhead, "Quantitative modelling of DNA damage using Monte Carlo track structure method," Radiat. Environ. Biophys. **38**(1), 31–38 (1999).
[14]J. F. Ward, "Biochemistry of DNA lesions," Radiat. Res. **104**(2s), S103–S111 (1985).
[15]W. R. Holley and A. Chatterjee, "Clusters of DNA damage induced by ionizing radiation: Formation of short DNA fragments. I. Theoretical modeling," Radiat. Res. **145**(2), 188–199 (1996).
[16]W. Friedland, M. Dingfelder, P. Kundrát, and P. Jacob, "Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC," Mutat. Res. **711**(1–2), 28–40 (2011).
[17]W. E. Wilson and H. Nikjoo, "A Monte Carlo code for positive ion track simulation," Radiat. Environ. Biophys. **38**(2), 97–104 (1999).
[18]S. Incerti, A. Ivanchenko, M. Karamitros, A. Mantero, P. Moretto, H. N. Tran, B. Mascialino, C. Champion, V. N. Ivanchenko, M. A. Bernal, Z. Francis, C. Villagrasa, G. Baldacchino, P. Guèye, R. Capra, P. Nieminen, and C. Zacharatou, "Comparison of GEANT4 very low energy cross section models with experimental data in water," Med. Phys. **37**(9), 4692 (2010).
[19]J. M. Fernández-Varea, G. González-Muñoz, M. E. Galassi, K. Wiklund, B. K. Lind, A. Ahnesjö, and N. Tilly, "Limitations (and merits) of PENELOPE as a track-structure code," Int. J. Radiat. Biol. **88**(1–2), 66–70 (2012).
[20]J.-L. Li, C.-Y. Li, R. Qiu, C.-C. Yan, W.-Z. Xie, Z. Zeng, and C.-J. Tung, "Comparison of direct DNA strand breaks induced by low energy electrons with different inelastic cross sections," Nucl. Instrum. Methods Phys. Res., Sect. B **311**(C), 27–36 (2013).
[21]H. Nikjoo and P. Girard, "A model of the cell nucleus for DNA damage calculations," Int. J. Radiat. Biol. **88**(1–2), 87–97 (2012).
[22]A. Kumar and M. D. Sevilla, "Radiation effects on DNA: Theoretical investigations of electron, hole and excitation pathways to DNA damage," in *Radiation Induced Molecular Phenomena in Nucleic Acids*, edited by M. K. Shukla and J. Leszczynski (Springer, Dordrecht, Netherlands, 2008), pp. 577–617.
[23]C. Champion, H. Lekadir, M. E. Galassi, O. Fojón, R. D. Rivarola, and J. Hanssen, "Theoretical predictions for ionization cross sections of DNA nucleobases impacted by light ions," Phys. Med. Biol. **55**(20), 6053–6067 (2010).
[24]I. Abril, R. Garcia-Molina, C. D. Denton, I. Kyriakou, and D. Emfietzoglou, "Energy loss of hydrogen- and helium-ion beams in DNA: Calculations based on a realistic energy-loss function of the target," Radiat. Res. **175**(2), 247–255 (2011).
[25]I. Abril, C. D. Denton, P. de Vera, I. Kyriakou, D. Emfietzoglou, and R. Garcia-Molina, "Effect of the Bethe surface description on the electronic excitations induced by energetic proton beams in liquid water and DNA," Nucl. Instrum. Methods Phys. Res., Sect. B **268**(11–12), 1763–1767 (2010).
[26]I. Abril, R. Garcia-Molina, P. de Vera, I. Kyriakou, and D. Emfietzoglou, "Inelastic Collisions of Energetic Protons in Biological Media," in *Advances in Quantum Chemistry* (Elsevier, Inc., Oxford, UK, 2013), pp. 129–164.
[27]Z. Tan, Y. Xia, X. Liu, M. Zhao, Y. Ji, F. Li, and B. Huang, "Cross sections of electron inelastic interactions in DNA," Radiat. Environ. Biophys. **43**(3), 173–182 (2004).
[28]D. Liljequist, T. Liamsuwan, and H. Nikjoo, "Elastic scattering cross section models used for Monte Carlo simulation of electron tracks in media of biological and medical interest," Int. J. Radiat. Biol. **88**(1–2), 29–37 (2012).
[29]D. Emfietzoglou, I. Kyriakou, I. Abril, R. Garcia-Molina, and H. Nikjoo, "Inelastic scattering of low-energy electrons in liquid water computed from optical-data models of the Bethe surface," Int. J. Radiat. Biol. **88**(1–2), 22–28 (2012).
[30]L. Zhang and Z. Tan, "A new calculation on spectrum of direct DNA damage induced by low-energy electrons," Radiat. Environ. Biophys. **49**(1), 15–26 (2009).
[31]R. M. Thomson and I. Kawrakow, "On the Monte Carlo simulation of electron transport in the sub-1 keV energy range," Med. Phys. **38**(8), 4531 (2011).
[32]D. Liljequist and H. Nikjoo, "On the validity of trajectory methods for calculating the transport of very low energy (<1 keV) electrons in liquids and amorphous media," Radiat. Phys. Chem. **99**(C), 45–52 (2014).
[33]A. Akkerman, T. Boutboul, A. Breskin, R. Chechik, and A. Gibrekhterman, "Low-energy electron transport in alkali halides," J. Appl. Phys. **76**(8), 4656 (1994).
[34]J. Kuhr and H.-J. Fitting, "Monte Carlo simulation of electron emission from solids," J. Electron Spectrosc. Relat. Phenom. **105**(2), 257–273 (1999).
[35]S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, and G. Barrand, "GEANT4—A simulation toolkit," Nucl. Instrum. Methods Phys. Res., Sect. A **506**(3), 250–303 (2003).
[36]P. de Vera, E. Surdutovich, I. Abril, R. Garcia-Molina, and A. V. Solov'yov, "Analytical model of ionization and energy deposition by proton beams in subcellular compartments," Eur. Phys. J. D **68**(4), 96 (2014).
[37]W. Friedland, P. Jacob, H. G. Paretzke, M. Merzagora, and A. Ottolenghi, "Simulation of DNA fragment distributions after irradiation with photons," Radiat. Environ. Biophys. **38**(1), 39–47 (1999).
[38]C. Champion, S. Incerti, H. Aouchiche, and D. Oubaziz, "A free-parameter theoretical model for describing the electron elastic scattering in water in the Geant4 toolkit," Radiat. Phys. Chem. **78**(9), 745–750 (2009).
[39]R. Panajotovic, F. Martin, P. Cloutier, D. Hunting, and L. Sanche, "Effective cross sections for production of single-strand breaks in plasmid DNA by 0.1 to 4.7 eV electrons," Radiat. Res. **165**, 452–459 (2006).
[40]B. Boudaïffa, P. Cloutier, D. Hunting, M. A. Huels, and L. Sanche, "Resonant formation of DNA strand breaks by low-energy (3 to 20 eV) electrons," Science **287**(5458), 1658–1660 (2000).
[41]M. A. Bernal, C. E. deAlmeida, C. Sampaio, S. Incerti, C. Champion, and P. Nieminen, "The invariance of the total direct DNA strand break yield," Med. Phys. **38**(7), 4147–4153 (2011).
[42]W. Friedland, P. Jacob, H. G. Paretzke, and T. Stork, "Monte Carlo simulation of the production of short DNA fragments by low-linear energy transfer radiation using higher-order DNA models," Radiat. Res. **150**(2), 170 (1998).
[43]D. E. Charlton, H. Nikjoo, and J. L. Humm, "Calculation of initial yields of single- and double-strand breaks in cell nuclei from electrons, protons and alpha particles," Int. J. Radiat. Biol. **56**, 1–19 (1989).
[44]H. Nikjoo, D. E. Charlton, and D. T. Goodhead, "Monte Carlo track structure studies of energy deposition and calculation of initial DSB and RBE," Adv. Space Res. **14**(10), 161–180 (1994).
[45]S. Incerti, C. Champion, H. N. Tran, M. Karamitros, M. Bernal, Z. Francis, V. Ivanchenko, A. Mantero, and Members of the Geant4-DNA collaboration, "Energy deposition in small-scale targets of liquid water using the very low energy electromagnetic physics processes of the Geant4 toolkit," Nucl. Instrum. Methods Phys. Res., Sect. B **306**(C), 158–164 (2013).
[46]W. Friedland, H. G. Paretzke, F. Ballarini, A. Ottolenghi, G. Kreth, and C. Cremer, "First steps towards systems radiation biology studies concerned with DNA and chromosome structure within living cells," Radiat. Environ. Biophys. **47**(1), 49–61 (2008).
[47]D. E. Charlton and J. L. Humm, "A method of calculating initial DNA strand breakage following the decay of incorporated 125I," Int. J. Radiat. Biol. **53**(3), 353–365 (1988).
[48]M. A. Bernal and J. A. Liendo, "Erratum: An investigation on the capabilities of the PENELOPE MC code in nanodosimetry [Med. Phys. 36, 620–625 (2009)]," Med. Phys. **37**(8), 4520–4520 (2010).
[49]A. G. Georgakilas, P. O'Neill, and R. D. Stewart, "Induction and repair of clustered DNA lesions: What do we know so far?," Radiat. Res. **180**(1), 100–109 (2013).