“dsbandrepair”- an updated Geant4-DNA simulation tool for evaluating the radiation-induced DNA damage and its repair

Le Tuan Anh[1], Tran Ngoc Hoang[2], Yann Thibaut[1], Konstantinos Chatzipapas[3], Dousatsu Sakata[4], Sébastien Incerti[2], Carmen Villagrasa[1], Yann Perrot[1]

[1] Institut de Radioprotection et de Sûreté Nucléaire (IRSN), BP 17, 92262 Fontenay-aux-Roses, France

[2] CNRS/IN2P3, CENBG, UMR 5797, Bordeaux University, 33170 Gradignan, France

[3] University of Brest, INSERM, LaTIM, UMR1101, 29200 Brest, France

[4] Division of Health Science, Osaka University, Osaka, Japan

Abstract

Purpose: Interdisciplinary scientific communities have shown large interest to achieve a mechanistic description of radiation-induced biological damage, aiming to predict biological results produced by different radiation quality exposures. Monte Carlo track-structure simulations are suitable and reliable for the study of early DNA damage induction used as input for assessing DNA damage. This study presents the most recent improvements of a Geant4-DNA simulation tool named “dsbandrepair”.

Methods: “dsbandrepair” is a Monte Carlo simulation tool based on a previous code (FullSim) that estimates the induction of early DNA single-strand breaks (SSBs) and double-strand breaks (DSBs). It uses DNA geometries generated by the DNAFabric computational tool for simulating the induction of early single-strand breaks (SSBs) and double-strand breaks (DSBs). Moreover, the new tool includes some published radiobiological models for survival fraction and un-rejoined DSB. Its application for a human fibroblast cell and human umbilical vein endothelial cell containing both heterochromatin and euchromatin was conducted. In addition, this new version offers the possibility of using the new IRT-syn method for computing the chemical stage.

Results: The direct and indirect strand breaks, SSBs, DSBs, and damage complexity obtained in this work are equivalent to those obtained with the previously published simulation tool when using the same configuration in the physical and chemical stages. Simulation results on survival fraction and un-rejoined DSB are in reasonable agreement with experimental data.

Conclusions: “dsbandrepair” is a tool for simulating DNA damage and repair, benchmarked against experimental data. It has been released as an advanced example in Geant4.11.2.

1. Introduction

Studying radiobiological effects at the subcellular scale continues to receive increasing attention from the interdisciplinary scientific community for applications in various fields (medical, space, radiation protection, effects of low absorbed doses). A better understanding of biological processes induced by ionizing radiation will help in establishing improved approaches to minimize the biological consequences of the absorbed dose not only in radiotherapy but also in other medical applications, spatial exploration, or other professional contexts.

As mentioned in several recent studies, radiation-induced biological effectiveness is believed to depend on patterns of microscopic and nanoscopic energy deposition [1], that's why Monte Carlo simulations, which allow transporting the charged particles in a way that is adapted to different scales, have been considered a useful and reliable approach in this frame [2].

Over the years, dedicated Monte Carlo simulation codes have been developed for radiobiological research at the subcellular scale such as PARTRAC [3], KURBUC [4], RITRACKS [5], IDDRRA [6], and Geant4-DNA [7, 8, 9, 10]. These codes utilize the so-called track-structure approach to simulate the detailed interaction topology which can be coupled to target geometries with molecular description. These track-structure codes describe not only the physical interactions leading to energy deposition and spatial distribution of interactions (physical stage) but also the physicochemical and chemical subsequent processes. Indeed, during the physical stage, the transport of radiation and the local energy depositions are simulated to form ionized and excited water molecules. These hot water fragments then dissociate and produce chemical species during physico-chemistry processes. Then, these chemical species diffuse and react with each other and eventually the DNA components in the chemical stage. A combination with realistic geometric DNA models allows for determining the amount and complexity of DNA damage. More details on the review mainly of the physical stage of these codes can be found in [11]. The first three mentioned codes are mainly distributed within specific scientific communities, while Geant4-DNA offers free access to all users.

Among these different codes, Geant4-DNA is developed and maintained by an international collaboration to extend the functionalities of the general-purpose Monte Carlo code Geant4, initially developed for High Energy Physics. This project was initially launched in 2001 by the European Space Agency to simulate the biological damage induced by cosmic radiation at the cellular and subcellular levels but the relevance and the stability of the code have rapidly led to biomedical applications such as the simulation of direct and indirect effects on geometries including the whole genome of cell nuclei, yeast, or bacteria. As a consequence of these effects, single and double-strand breaks (SSBs and DSBs) can be obtained, especially DSBs which are more difficult to repair, and their complexity is one of the causes of malignancies [12, 13].

The first version of this application was initially developed by S. Meylan et al. [14] in 2017 (FullSim). This code includes a DNA geometrical model obtained from the output of the DNAFabric software [15] that is used to build the complete genome content

in a nuclear geometry including the continuous chromatin fiber down to the base pair (bp) precision level of human cells in the G0/G1 phase. It also allows to include different heterochromatin and euchromatin distributions [16] and has been recently refined to account for the isochore theory [17]. The upgraded version of this simulation tool, presented in this article and now called "dsbandrepair" has recently been released in Geant4 and proposes an alternative approach to the one proposed in the "moleculardna" example [18, 19, 20] which uses fractal geometry to model the cell nucleus structure of the DNA.

This paper introduces a detailed description of the "dsbandrepair" simulation tool, including the use of MPI parallelism [21] and synchronous independent reaction time (IRT-syn) [22] technique to speed up computing, and a new code structure for simpler maintenance and exploitation. Furthermore, to calculate the survival fraction and un-rejoined DSBs, we additionally incorporate repair models such as the Two Lesion Kinetic model (TLK) [23] and the Local Effect Model IV (LEM-IV) [24]. The "dsbandrepair" simulation results in this work used Geant4.11.1 version. To benchmark the code, comparisons with previous simulated results and existing experimental data on SB yield, survival fraction, and un-rejoined DSB were conducted and are presented here.

2. Materials and methods

2.1. "Fullsim" - The previous simulation tool

The preceding version ("Fullsim") was published in 2017 [14] and released in Geant4-DNA as an extended example with a simplified geometry of only one DNA chromatin voxel of 40 nm side and containing 18 histones. The main reason for this was that, to run the complete cell nucleus geometry, modifications in the Geant4 code kernel classes were needed preventing the example from being delivered as a user application. This problem has now been resolved, and the new 'dsbandrepair' simulation tool can be released, considering the complete eukaryote genome.

Fullsim consisted of seven subprograms that needed to be executed in a specific order to compute the yields of DSBs from both the direct and indirect effects:

- "Phys_geo": This subprogram is dedicated to simulating the physical stage.
- "Extract_chem_input_data": This subprogram is responsible for preparing the input data for the "Chem_geo" subprogram from the output of "Phys_geo".
- "Chem_geo": The physico-chemical and chemical stages are simulated by this application.
- "extractSSB", "DBScan", and "Post_analysis": These three subprograms (a clustering algorithm and several analysis routines) are used to calculate and extract the DSB yield from the outputs of "Phys_geo" and "Chem_geo".
- "statistics": This subprogram controls the statistics of the simulation. If the uncertainty on the DSB yield is higher than a desired value, additional simulations will be invoked.

This modular design enables the users to separately launch each module. Strand breaks (SBs) are registered if the cumulative energy imparted into a volume that combines phosphate and 2-deoxyribose (and their corresponding hydration shells) of a nucleotide pair for each event in the physical stage is at least 17.5 eV in a direct effect. This energy threshold is a commonly accepted value [25] and has enabled us to better reproduce the experimental DSB induction data [14]. In addition, a probability of 40% is used to determine whether a reaction between hydroxyl radicals (OH*) and sugar generates a SB during the chemical stage. This is because the structure of the DNA chain allows hydroxyl radicals to reach two of the five reaction sites of the sugar [26]. The authors validated the code by comparing the simulated DSB yield to the experimental data for proton-induced damages at different energies. The simulation was conducted for a fibroblast cell bombarded by monoenergetic proton beams. An end-time of 2.5 ns was set for the chemistry stage. This approach allows to simulate the scavenging role of non-histone proteins associated with chromatin [27]. Generally, good agreement between these simulation results and the experimental data was observed. For more details, please refer to [14].

Subsequently, the simulation tool was validated in terms of DSB yields with experimental data induced by irradiating human umbilical vein endothelial cells (HUVEC) with 40 kVp, 220 kVp, and 4 MV X-rays [28]. Based on the work in [16], the end-time of 2.5 ns was kept for the simulation of the endothelial cell nucleus filled with heterochromatin and euchromatin.

In all the above-mentioned simulations, the evolution of an initially heterogeneously distributed reaction-diffusion system in the chemical stage was handled by the so-called step-by-step (SBS) approach [29]. The SBS method employs time steps to model the diffusion of the chemical species. The time steps can be chosen statically or dynamically. The main drawback of this method is its long computation time in comparison with the IRT-syn method recently implemented [17]. It is worth noting that the IRT-syn method was not available at the time the previous simulation tool was developed.

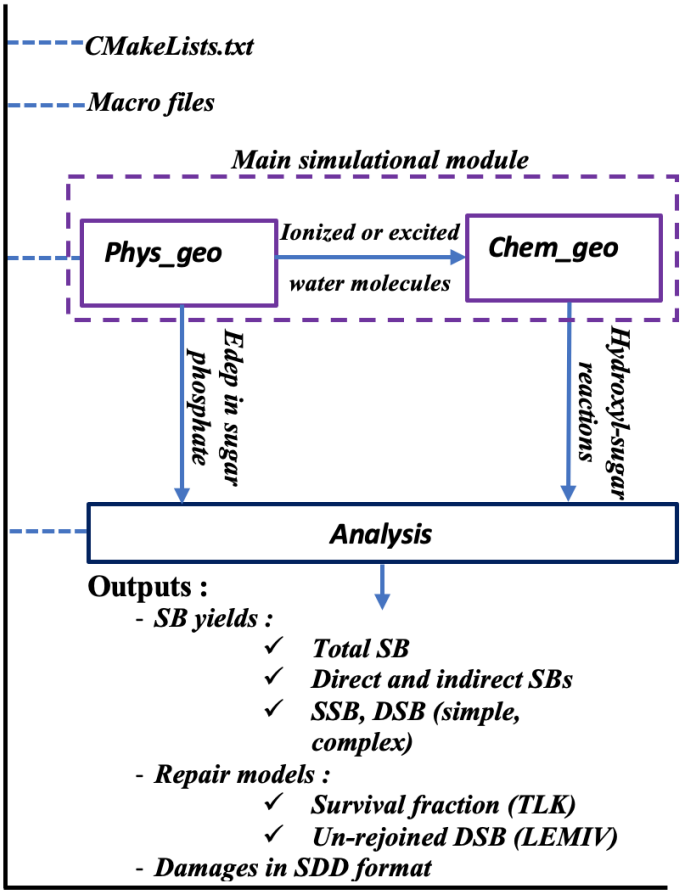


Figure 1: Structure of the "dsbandrepair" simulation tool.

2.2. The new simulation tool - "dsbandrepair"

As shown in Figure 1, the simulation tool named "dsbandrepair" contains three subprograms instead of seven as in the previous version. Some subprograms were merged to ease the use of the code. The first module (Phys_geo) is responsible for the simulation of the physical stage, while the physicochemical and chemical stages are simulated with the second module (Chem_geo). The last module (Analysis) takes the results of Phys_geo and Chem_geo to calculate the damage yields (direct SB, indirect SB, SSB, DSB yield and complexity) and run repair models based on these inputs.

In this work, we also utilize Message Passing Interface (MPI) parallelism to reduce the computational time. MPI is a standardized library for parallel computing on distributed computing systems [21]. Indeed, thanks to the work of K. Murakami and A. Dotti [30], an MPI interface was introduced into Geant4, and it's now used in this work.

2.2.1. Geometries

Following the work of S. Meylan et al., a bottom-up approach was chosen to model DNA structure, i.e. more complex DNA compaction geometries were constructed based on six different spherical volumes representing fundamental DNA constituents which are filled with water for simulation purposes. These volumes are phosphate, deoxyribose, adenine, guanine, thymine, and cytosine (see [14] for details). Namely, these constituents were first used to form nucleotide pairs. Then, B-DNA double helixes built from a stack of nucleotide pairs were used to create nucleosomes which then were helically linked

together to form pieces of chromatin fiber. These pieces of fiber later were placed and oriented to build five distinct voxel types: “straight”, “right”, “left”, “up”, and “down” voxels. More detailed descriptions for these voxel types can be found elsewhere [15]. Finally, the cell nucleus was built and filled with the five abovementioned voxel types for both heterochromatin and euchromatin compactions. “dsbandrepair” supports all types of DNA geometries (yeast, bacteria, cell nuclei) constructed with the DNAFabric software we have developed, an example of which to generate simple geometries based on one voxel is publicly available (https://bitbucket.org/sylMeylan/opendnafabric/src/master/). “dsbandrepair” includes at the moment 3 different cell nucleus geometries representing, in terms of dimensions, the human cell nuclei of fibroblast, lymphocyte, and endothelium.

The complete cell nucleus geometry usually contains several million voxels. Therefore, introducing the modelled cell geometry as the simulation target requires a huge amount of memory. To keep the consumed memory at an acceptable level (below 10 Gb), the Geant4 parameterization method was considered to parameterize 10 different physical volumes (5 for heterochromatin and 5 for euchromatin) corresponding to 5 voxel types. However, parameterization of physical volume is not fully supported in Geant4. That means it hinders launching the simulation of the physical stage in multi-threaded mode. In this work, to overcome this problem, MPI parallelism was considered by including the MPI interface of Geant4 into “dsbandrepair” to enable it to run the physical stage in parallel, especially across multi-nodes on High-Performance Computing resources.

2.2.2. Simulation of the chemistry stage

In the chemical phase, radicals can react with each other or with DNA molecules. To describe these interactions, Geant4-DNA used the so-called diffusion-controlled reactions [29]. Currently, Geant4-DNA offers three methods to simulate diffusion-control reactions in the chemical stage, namely, step-by-step (SBS) [29], independent reaction time (IRT) [31], and synchronous IRT (IRT-syn) [22] methods. Note that the IRT method does not provide the exact positions of the radicals in time, which are required by “dsbandrepair” for later analysis. Therefore, our simulation tool includes only the SBS and IRT-syn options for simulating the chemical stage. IRT-syn uses the IRT method to determine the time step to the next chemical reaction. However, after a reaction is finished, the remaining molecules and the reactive products will all be synchronized for the subsequent diffusion in the next time step. This synchronization allows IRT-syn to provide the spatiotemporal information of the reactive species after each time step.

G4EmDNAChemistry_option2 is a chemical constructor developed specifically for the SBS chemistry approach in liquid water. Besides, chemistry parameters for simulating reactions with DNA components were also included in this constructor in our previous work. The G4EmDNAChemistry_option2 [14] consists of 30 chemical reactions 13 of which involve DNA components (phosphate, deoxyribose, and base pairs) and 7 concern histones.

Reaction	**Reaction rate [$m^3 mol^{-1} s^{-1}$]**

$OH^{\cdot}$ + Deoxyribose	1.8×10^{6}
$OH^{\cdot}$ + Adenine	6.1×10^{6}
$OH^{\cdot}$ + Guanine	9.2×10^{6}
$OH^{\cdot}$ + Thymine	6.4×10^{6}
$OH^{\cdot}$ + Cytosine	6.1×10^{6}
e_{aq} + Deoxyribose	0.01×10^{6}
e_{aq} + Adenine	9.0×10^{6}
e_{aq} + Guanine	14.0×10^{6}
e_{aq} + Thymine	18.0×10^{6}
e_{aq} + Cytosine	13.0×10^{6}
$H^{\cdot}$ + Deoxyribose	0.029×10^{6}
$H^{\cdot}$ + Adenine	0.10×10^{6}
$H^{\cdot}$ + Thymine	0.57×10^{6}
$H^{\cdot}$ + Cytosine	0.092×10^{6}
$OH^{\cdot}$ + Histone	--
OH^{-} + Histone	--
e_{aq} + Histone	--
H2 + Histone	--
$H3O^{+}$ + Histone	--
H^{-} + Histone	--
H2O2 + Histone	--

Table 1: Chemical reactions for DNA components and histones taken from G4EmDNAChemistry_option2. The last reactions involving histones simulate the histone as a scavenger, i.e. histones absorb water radicals that encounter its surface (no reaction rates are attached to these reactions).

G4EmDNAChemistry_option3 [19], on the other hand, was implemented initially for the IRT method. Since Geant4.10.7, IRT-syn has been available and then used by this constructor. Geant4-DNA users can easily switch between these two methods. G4EmDNAChemistry_option3 includes 72 chemical reactions for 15 species from water radiolysis. However, it does not contain reactions for DNA components that are needed by "dsbandrepair". Therefore, a new constructor derived from G4EmDNAChemistry_option3 was created in "Chem_geo". Twenty-one reactions for DNA components and histones taken from G4EmDNAChemistry_option2 were added to the inherited constructor. Details of these reactions are listed in Table 1.

Henceforth, "dsbandrepair" users can use either the SBS approach by choosing G4EmDNAChemistry_option2, or the IRT-syn approach by invoking the extended version of G4EmDNAChemistry_option3 by declaring it in a macro file.

2.2.3. Analysis and outputs

Phys_geo and Chem_geo record all necessary information (see *Figure 1*) in several ntuples which are written in ".root" files, and can be later accessed by the ROOT-CERN library [32].

After the end of the physical and chemical stages, the simulation results will be analyzed by the module "Analysis" to obtain the SB yields.

First, the module identifies direct SBs via the output of the physical stage. Direct damage is usually calculated from the accumulation of energy deposited in backbone volume. Two approaches are widely accepted in the literature to translate the deposited energy into direct SB: a single threshold method and a linear probability method. In this current implementation, as in our previous work, we use a threshold of 17.5 eV for the deposited energy in the backbone volume (including its hydration shell volume) to create a direct SB.

Next, the module "Analysis" extracts indirect SBs from the output of the chemical stage. As in previous work, it is assumed that only chemical reactions between OH^* and a sugar group (2-deoxyribose and phosphate) can produce indirect SB. A probability of 40 % was still applied to such a reaction to determine whether it induced an indirect SB.

Finally, a clustering algorithm is applied to both direct and indirect resulting SBs to address the number of DSBs and their components. A new cluster is registered if the location of its first damage is at least 10 bp far away from the location of the last damage of the previous cluster. We define a DSB as a cluster containing at least one SB on each opposite strand. If a DSB contains extra SBs, it is classified as a complex DSB, the number of total SBs being the complexity; otherwise, it is categorized as a simple DSB. Any cluster that contains SBs on only one strand is defined as an SSB. More details on the scoring damage approach can be found in [14]. It is important to note that the method of calculating DSBs may vary depending on the repair model under consideration, particularly in relation to the distance between damages. It is the responsibility of the user to ensure that the default parameters proposed in "dsbandrepair" are suitable and, if necessary, to make the necessary adjustments.

The uncertainty of simulated SB yield is estimated through the following equation:

$$\sigma = \sqrt{\frac{\left|\sum_{i=0}^{m} x_i^2 - \left(\sum_{i=0}^{m} x_i\right)^2\right|}{m}} \qquad (1)$$

where m denotes the total number of events at the origin of the type of SB that is being analyzed (SSB, DSB, direct SB, or indirect SB, etc.). Meanwhile, x_i represents the number of the analyzed SB type in the ith event.

The results of SB yields (SSB, DSB, direct and indirect SB), then, are written in a text file. In addition, we also included in the new "dsbandrepair", as in the "dnadamage1" example, a function that allows exporting output in a simple SDD format, which is a standardized format proposed in 2019 by J. Schuemann et al. [33], to serve as the foundation for interdisciplinary studies of DNA damage induction and subsequent kinetics of DNA repair processes in particular, making it easier to compare results from different simulation codes.

Optionally, the new "dsbandrepair" offers two radiobiological models that allow users to calculate survival fraction (TLK) [23] and/or un-rejoined DSB (LEM-IV) [24]. Users can invoke these models via a macro file in the module "Analysis". A brief introduction of the two models is presented in the following subsection.

2.2.4. Radiobiological models

2.2.4.1. Survival fraction with TLK model

Cell death is the most noticeable issue in radiation-induced DNA damage. Especially, unrepaired DSBs are widely considered as the main cause leading to cell killing. The Two-Lesion Kinetic (TLK) model proposed by R. D. Stewart [23] was developed to provide a method capable of linking DNA damage, and more specifically DSBs (formation, repair, misrepair), with cell death. It should be noted that the TLK model has already been coupled to DNA damage calculations performed with Geant4-DNA [34] [35] [36]. In this study, we included this model in the module "Analysis" of "dsbandrepair. The DNA fragments associated with DSBs can interact with each other in pairs and form lethal or non-lethal chromosomal aberrations. In the TLK model, DNA repair is performed by using first-order non-linear differential equations to predict the enzymatic kinetics of DNA repair. Even if no specific hypothesis is made about the biochemical steps involved in the repair of DSBs, this model suggests that DSB repair depends on the severity of the lesion.TLK model incorporates both simple DSBs and complex DSBs in its calculation to deduce the survival fraction (SF) of the cell and consists of three simplified differential equations, as follows:

$$\frac{dL_1(t)}{dt} = -\lambda_1 L_1(t) - \eta L_1(t)\big(L_1(t) + L_2(t)\big) \quad (2)$$

$$\frac{dL_2(t)}{dt} = -\lambda_2 L_2(t) - \eta L_2(t)\big(L_1(t) + L_2(t)\big) \quad (3)$$

$$\frac{dL_f(t)}{dt} = \beta_1\lambda_1 L_1(t) + \beta_2\lambda_2 L_2(t) + 0.25\eta[L_1(t) + L_2(t)]^2 \quad (4)$$

Here, L1(t) and L2(t) are respectively the expected number of simple DSBs and complex DSBs per irradiated cell with absorbed dose D at a given time after the irradiation. $L_f(t)$ is the time-dependent accumulation of lethal DNA damage in a cell and λ_1 and λ_2 are respectively simple DSB and complex repair probability. Meanwhile, β_1 and β_2 are respectively simple DSB and complex misrepair probability. The last parameter, η represents a binary misrepair probability.

With parameters λ_1 and λ_2 representing repair kinetics and β_1 and β_2 representing probabilities of poor repair, these parameters are not expected to change between different cell lines involving the same repair mechanisms. On the other hand, the parameter η representing the probability of interaction between two DSBs should be influenced by the average proximity of the DSB, and therefore by the concentration of DNA in the nuclear volume. The first part of this study focused on fibroblast cell nucleus, and the parameters of the TLK model were taken from those proposed by Stewart [23] who studied this cell line. If another cell line is envisaged, which is also the case in this study with human umbilical vein endothelial cell (HUVEC), the η parameter can therefore be adjusted according to the nuclear volume used in the simulation. To fit this TLK parameter for this cell line, the model parameters were therefore kept constant except for the η parameter. Survival curves were calculated for values of η varying over the range allowed by the model (from 0.0002 h^{-1} to 0.01 h^{-1}) and according to a constant-step absorbed dose sampling, every 0.25 Gy, over a dose range from 0 to 6 Gy. In this

study, the value of the parameter η was chosen to minimize the deviation from the experimental survival curve for 220 kV X-Ray irradiation from the literature [37].

In "dsbandrepair", we implemented the fourth-order Runge-Kutta method to solve this system of ordinary differential equations (ODEs). After solving ODEs, the SF(t) was easily obtained by using the following equation: $SF(t) = e^{-L_f(t)}$.

Comparison of simulated and measured data enabled us to set the distance between two SBs at 15 bp to obtain the best agreement. This distance is within the generally accepted range of between 10 and 25 bp [38].

2.2.4.2. Local effect model – LEM

The local effect model was initially developed in the 1990s at GSI in Darmstadt in the context of heavy ion radiotherapy. The LEM allows the prediction of RBE values for a variety of ions and linear energy transfer (LET) values. Due to its predictive power, the LEM is used nowadays as an input for heavy ion treatment plans in various hadrontherapy centers. Successive developments have led to different improved versions of the LEM. Among them, the most recent version (2010) is the LEM-IV [24, 39]. In this work, we included LEM-IV in "dsbandrepair" using Equations (5-7) hereafter.

The main assumption of LEM is that a similar distribution of DNA damage should lead to equivalent biological damage regardless of the quality of the radiation that causes it. The LEM-IV is therefore based on the spatial distribution of DSBs within the cell nucleus. For this purpose, this model considers the number of DSBs present in 2 Mbp chromatin loops called "Giant loops" or domains representing sensitive DNA structures. When a single DSB is present in a loop, it is then considered an isolated DSB that will be repaired with high fidelity. In the case where two or more DSBs are in a loop, these DSBs are then considered as a cluster of DSBs.

Secondly, the model is also based on the assumption that the fast and slow components of rejoining observed in the experiments can be related to the processing of isolated DSBs and clustered DSBs, respectively. The fraction of unrepaired DSBs at a time t, U(t), is calculated by a two-phase exponential decay with components F_{fast} and F_{slow}, characterized by half-lives τ_{fast} and τ_{slow}:

$$U(t) = F_{fast} e^{-\frac{ln(2)}{\tau_{fast}}t} + \left(F_{slow} - F_{unrej}\right) e^{-\frac{ln(2)}{\tau_{slow}}t} + F_{unrej} \qquad (5)$$

with:

$$F_{fast} = \frac{n_i}{n_i + n_c \lambda_c} \qquad (6)$$

and:

$$F_{slow} = \frac{n_c \lambda_c}{n_i + n_c \lambda_c} \qquad (7)$$

where n_i is the number of domains containing an isolated DSB, n_c is the number of domains containing a clustered DSB, and λ_c the mean number of DSBs in a cluster of DSB. F_{unrej} is the fraction of DSBs that are not repaired even for late times (> 24 h).

2.2.5. *Configurations of the simulation for validating "dsdandrepair"*

2.2.5.1. *Cell nuclei geometrical models*

Unlike in previous work [14], here eleven DNAFabric output files were used for providing the geometrical description of DNA content. The first file contains the description of the cell nucleus as well as other information on the position and type of each voxel placed in it. The rest of the files can be grouped in five pairs each corresponding to a voxel type ("straight", "right", "left", "up", and "down") with one file for euchromatin and one file for heterochromatin. These files contain information on the dimension (all are 40 nm side) and content of the voxels, as well as the placement and dimension of DNA constituents within the voxels. Each DNA molecule has a spherical shape covered by a spherical hydration shell. Additionally, a cut algorithm is applied to the spherical shapes to ensure that they do not overlap. More details are listed in Table 2 and Table 3.

Two types of human cell nuclei were considered in this work: a human fibroblast nucleus to enable comparison with existing experimental data in terms of DSB induction as well as cell fate after ion irradiation and a human umbilical vein endothelial cell (HUVEC) nucleus to test the simulation tool for low LET irradiations. Both nuclei are filled with the complete DNA genome including ~6.2 Gbp in the G0/G1 phase of the cell cycle [40] and contain both euchromatin (34%) and heterochromatin (66%). The fibroblast nucleus has an ellipsoidal shape described by the equation $\frac{x^2}{(9.85\ \mu m)^2} + \frac{y^2}{(7.1\ \mu m)^2} + \frac{z^2}{(2.5\ \mu m)^2} = 1$ as illustrated in Figure 2, and this geometry results in a base pair density of ~8.5 Mbp/µm^3. The HUVEC nucleus is an elliptical cylinder with an ellipsoid section described by the equation $\frac{x^2}{(9.5\ \mu m)^2} + \frac{y^2}{(5.1\ \mu m)^2} = 1$ and a height of 2 µm as shown in Figure 2. The base pair density for this geometry is ~20.4 Mbp/µm^3.

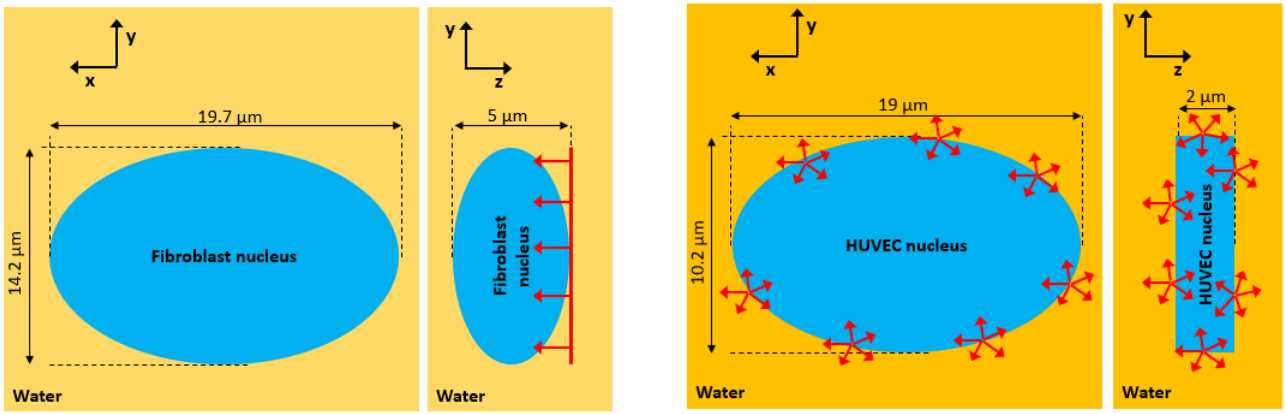


Figure 2: Schematic illustration of the simulated setup used in this work: a human fibroblast cell nucleus irradiated by parallel ion beams (in red) and a HUVEC nucleus isotropically irradiated by secondary electrons (in red) from X Rays.

	Voxels									
	Heterochromatin					Euchromatin				
Characteristics	**Straight**	**Up**	**Down**	**Right**	**Left**	**Straight**	**Up2**	**Down2**	**Right2**	**Left2**

						2				
Number of histones/nucleosomes	18	12	12	12	12	10	8	8	8	8
Number of nucleotide pairs	3594	2408	2415	2404	2402	2011	1637	1660	1646	1646

Table 2: Characteristics of the voxels for heterochromatin and euchromatin.

DNA molecules and histone	**Radius (nm)**	**Hydration shell radius (nm)**
2-deoxyribose	0.29	0.493
Phosphate	0.27	0.459
Adenine	0.3	0.51
Thymine	0.3	0.51
Guanine	0.3	0.51
Cytosine	0.3	0.51
Histone	2.4	0

Table 3: Characteristics of DNA constituents and histone.

2.2.5.2. Configurations of the simulations

For the fibroblast cell nucleus, the primary particles are mono-energetic protons or alphas. The energies of the particles were set to cover a wide LET range with values from 1.24 to 132 keV/µm as listed in Table 4. The primary source is generated in an ellipsoidal surface ($\frac{x^2}{(9.85\ \mu m)^2} + \frac{y^2}{(7.1\ \mu m)^2} = 1$) placed above the cell nucleus, having a direction parallel to the Z-axis as shown in Figure 2.

For the HUVEC cell nucleus, photon irradiations (kilovoltage and megavoltage X-Rays) were considered. This is why the spectrum of simulated secondary electrons, whose implementation details can be found in a previous study [28], was generated isotropically on the surface of the nucleus as shown in Figure 2. Geant4-DNA models do not allow the transport of electrons above 1 MeV. This is why the kinetic energy of electrons above 1 MeV has been forced to 1 MeV. To justify this hypothesis, it is important to remember that over the energy range 1 MeV - 3 MeV (maximum energy of the secondary electrons recorded in the simulation) the variation in stopping power in liquid water is less than 4%, going from 0.1862 keV/µm at 1 MeV to 0.1889 keV/µm at 3 MeV [41].

In this paper, the new "dsbandrepair" was tested with the latest version of Geant4 (11.1). Two configurations (in terms of chemical stage options) for the calculation of DNA damage induction in a human fibroblast cell were considered:

- Aconfig: For the physical and chemical stages, respectively, this setup used G4EmDNAPhysics_option2 and G4EmDNAChemistry_option2. The SBS method was automatically used by G4EmDNAChemistry_option2 for describing the evolution of the reaction-diffusion system.

- Bconfig: Similar to Aconfig, this setting used G4EmDNAPhysics_option2 for the physical stage. However, for the chemical stage, the extended version of G4EmDNAChemistry_option3 mentioned in Subsection 2.2.2 was applied to control the chemical reactions. Instead of SBS, IRT-syn was utilized. The use of this configuration aims to see how "dsbandrepair" adapts to the IRT-syn approach.

The results obtained with these two configurations were then benchmarked by comparison with the former simulation tool FullSim. As mentioned in Subsection 2.1, FullSim can only be compiled and executed with a modified version of Geant4 (10.6). Similar to Aconfig, FullSim uses G4EmDNAPhysics_option2 for the physical stage and G4EmDNAChemistry_option2 for the chemical stage.

A value of 5 ns was set to the chemical final time in all cases. This time limit in the chemical step was chosen because it allows for a more accurate representation of the average path of the hydroxyl radical in the biological environment [17]. The numbers of primary events were the same for all configurations with equal LET. Moreover, they were chosen to reach a compromise among DSB statistical uncertainty (< 10%), simulation time, and computing resources. The detailed values are listed in Table 4 in the Results section.

All the runs were executed on a High-Performance Computing system. Each node in this cluster consists of 16 CPUs and has a memory of 31 GB. Except for the results shown in Figure 14, the rest was computed with 4 or 5 nodes. The physical stage was set to use 5 CPUs per node, while all CPUs available in each node were used for the chemical stage (80 CPUs in total).

3. Results

3.1. DNA Strand breaks for fibroblast

Figure 3 shows the strand breaks (SBs) yield for incident monoenergetic protons and alphas as a function of LET and normalized per Gy and Gb in the cell nucleus. An agreement was observed between Aconfig and the result obtained with the previous simulation tool, both making use of the SBS method for the chemical stage. For protons, the total number of SBs from Aconfig and the "Fullsim" tool increases and reaches a peak at around 8 keV/µm, and then decreases with LET. Bconfig, on the other hand, decreases from the first investigated LET of 1.24 keV/µm (50 MeV) to 81.61 keV/µm (0.1 MeV). Bconfig produced more SBs than Aconfig and the previous tool for protons with LET below around 8 keV/µm. However, for LET above 8 keV/µm, Bconfig generates fewer SBs compared to the other results. In the case of alphas, the number of SBs always decreases with increasing LET values. An agreement is found between Aconfig and the previous tool. Meanwhile, there is a visible discrepancy between Bconfig and the other results. Figure 3 also shows that the SBs yield induced by protons are less numerous than the ones induced by alphas at the same LET value, as expected.

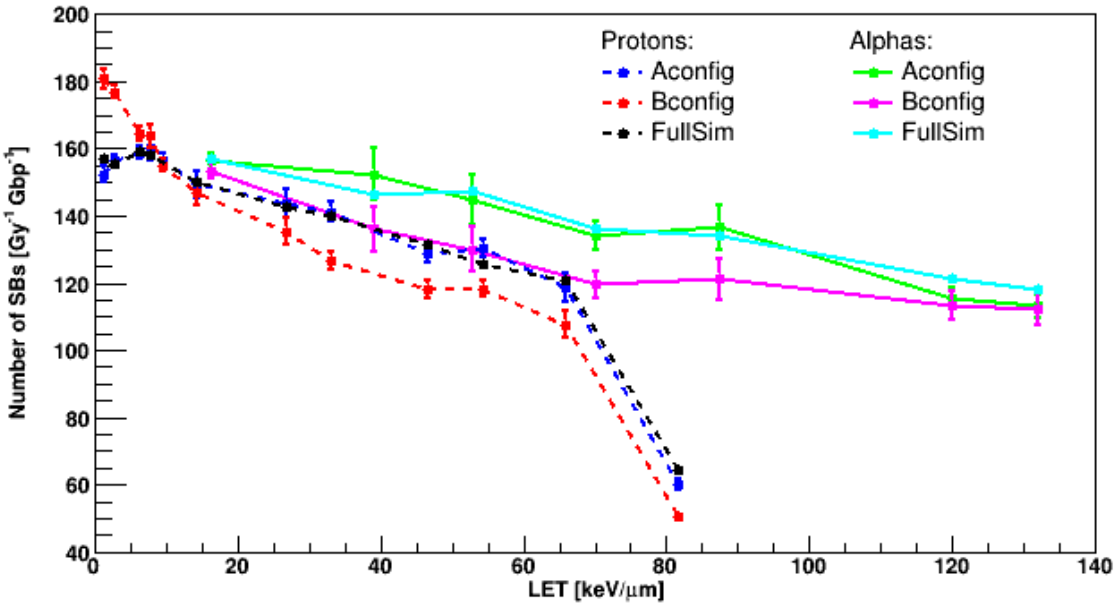
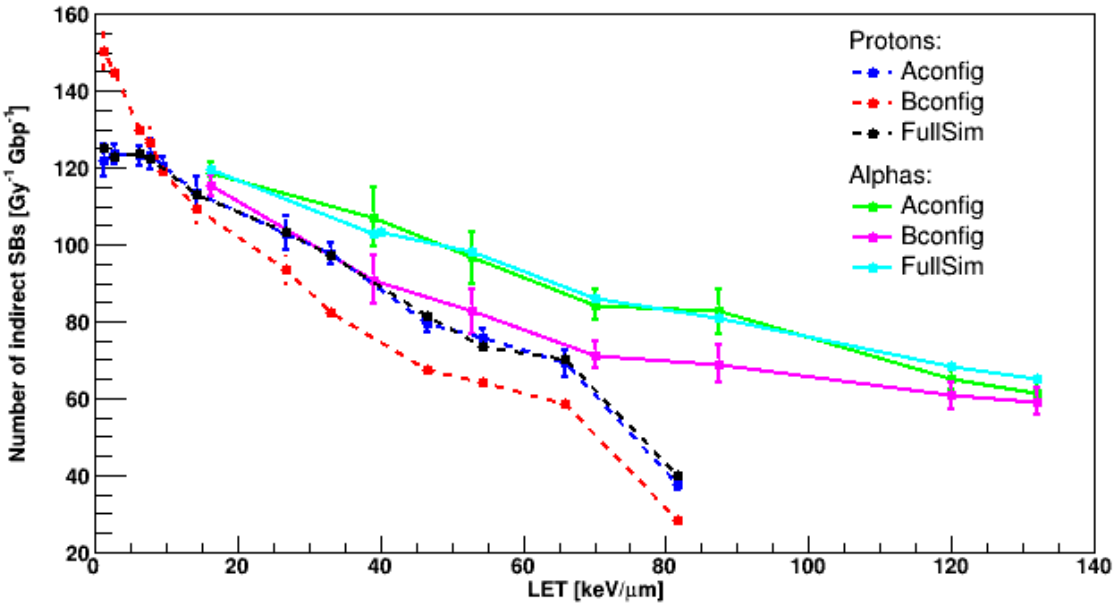
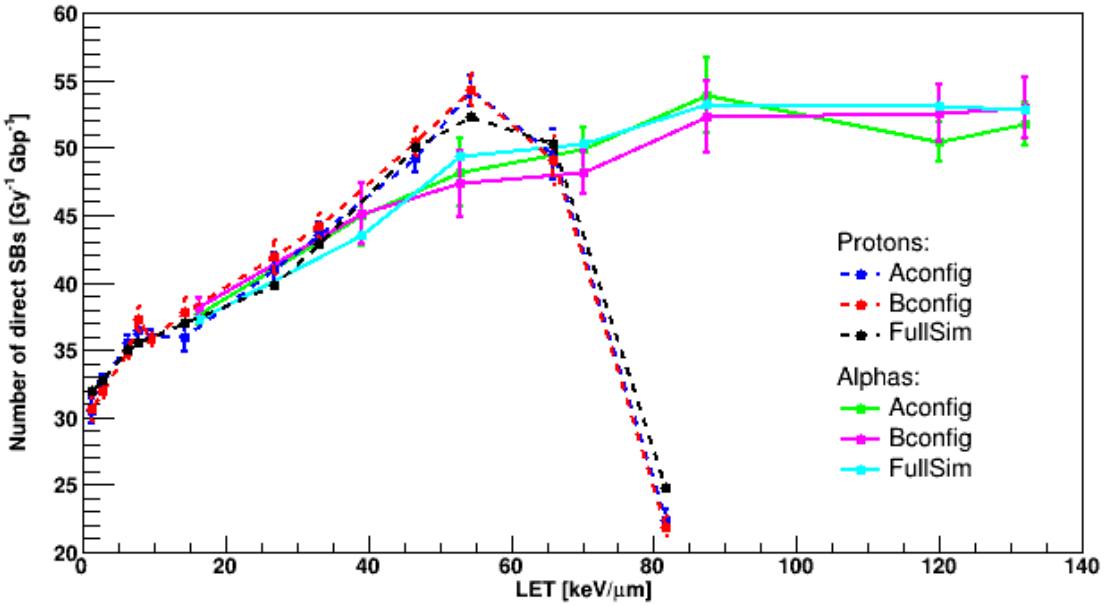


Figure 3: Strand breaks yield as a function of LET for alphas (solid lines) and protons (dotted lines) with different LET values. The results obtained by the previous simulation tool were used to benchmark the simulated yields generated by "dsbandrepair" using Aconfig and Bconfig, respectively. Note that the connected lines in this figure (as well as in Figures 4-8) are just guides for the eye.

Figure 4 presents the indirect and direct SB yields per Gy and Gbp in the cell nucleus for protons and alphas as a function of LET. The indirect SB damages present a similar tendency to the total number of SBs in Figure 3. On the contrary, there is a good agreement of direct SBs among Aconfig, Bconfig, and the previous "Fullsim" for both protons and alphas. This is easy to understand as all these simulations use the same physics list. Direct damages induced by protons dramatically increase from 1.24 keV/µm to ~54.41 keV/µm but drop sharply with LET above ~54.41 keV/µm. Meanwhile, direct SBs of alpha saturated above ~90 keV/µm. From 20 keV/µm to 30 keV/µm, protons generated a number of direct SBs similar to alphas; however, this trend deviated for LET from 30 keV/µm to 70 keV/µm where protons produced more direct damages compared to alphas with the same LET.



a) Indirect strand breaks

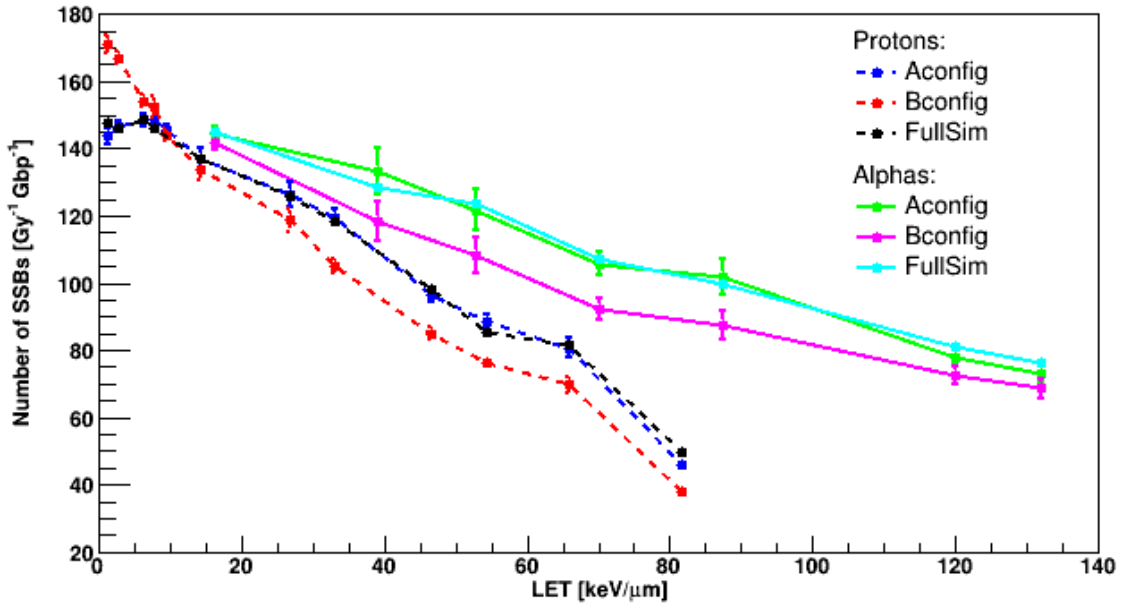


b) Direct strand breaks

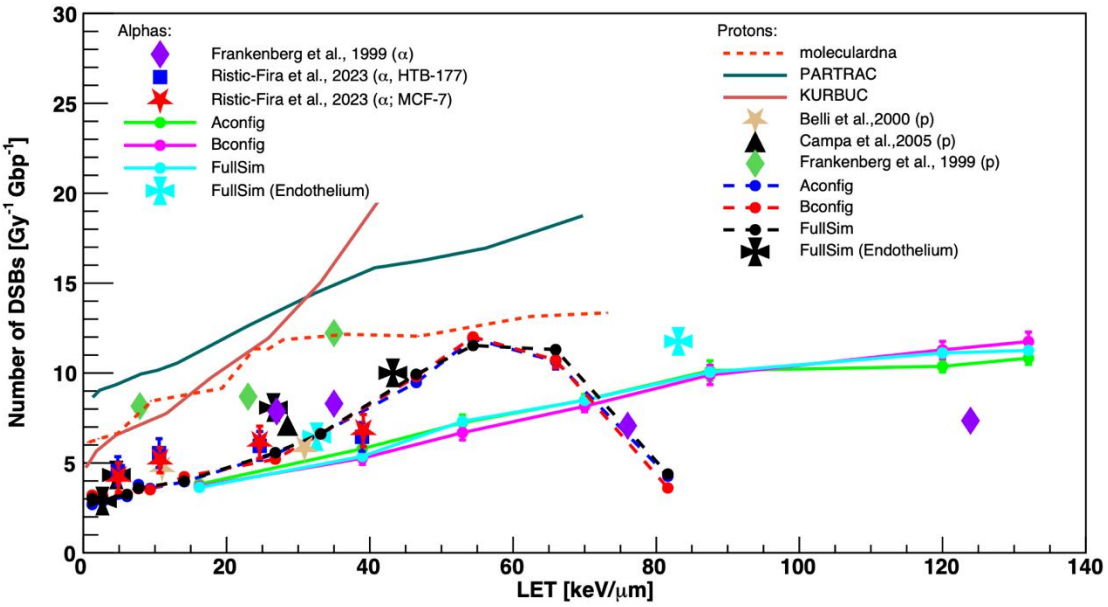
Figure 4: The indirect strand breaks (a), and direct strand breaks (b) as a function of the primary ion LET.

Like in the case of indirect SBs, single SBs (SSBs) shown in Figure 5a) had a tendency that follows the total SB curves in Figure 3. Meanwhile, the double-strand breaks (DSBs Gy^{-1}• Gbp^{-1}) increased against LET for alphas as displayed in Figure 5b); and the simulated results obtained by Aconfig were reasonably the same as the results of Bconfig and the previous tool for both protons and alphas. The DSB yields for protons reach a peak (at ~54.41 keV/µm) before decreasing drastically. Compared to "moleculardna" [20], the DSB yields in this work were lower. However, it has to be noted that "moleculardna" modelled a fibroblast cell having a base pair density of 0.012

bp/nm^3 which is higher than our present work (~0.0085 bp/nm^3). Therefore, Figure 5b also presents the data simulated by FullSim for the endothelium cell nucleus which had the same chromatin mixture as the fibroblast in this work but had a smaller volume. It can be seen that the number of DSBs Gy^{-1}• Gbp^{-1} was higher than our current results for both protons and alphas due to the higher density of DNA in the nuclear volume which has a strong impact on the DSB yields. Compared to KURBUC [42] and PARTRAC [43], our results were lower. The DSB yields induced by protons in this work were close to those experimentally obtained by Belli et al. [44] for Chinese hamster fibroblast cells (V79) and Campa et al. [45] for normal human fibroblasts (AG1522); however, other data obtained by Frankenberg et al. [46] for human skin fibroblasts were higher than the results of this work for all the configurations. The DSB yields caused by alpha in this study were comparable to those that Ristic-Fira et al. [47, 48] obtained experimentally for the HTB-177 (large cell lung carcinoma) and the MCF-7 (breast adenocarcinoma) human cells. The simulated DSB yields in this work also show that for LET values between 20 - 70 keV/µm, protons out-generated DSBs than alphas at the same LET.



a) Single-strand breaks



b) Double-strand breaks

Figure 5: The single-strand break (SSB) (a) and double-strand break (DSB) (b) yields as a function of LET.

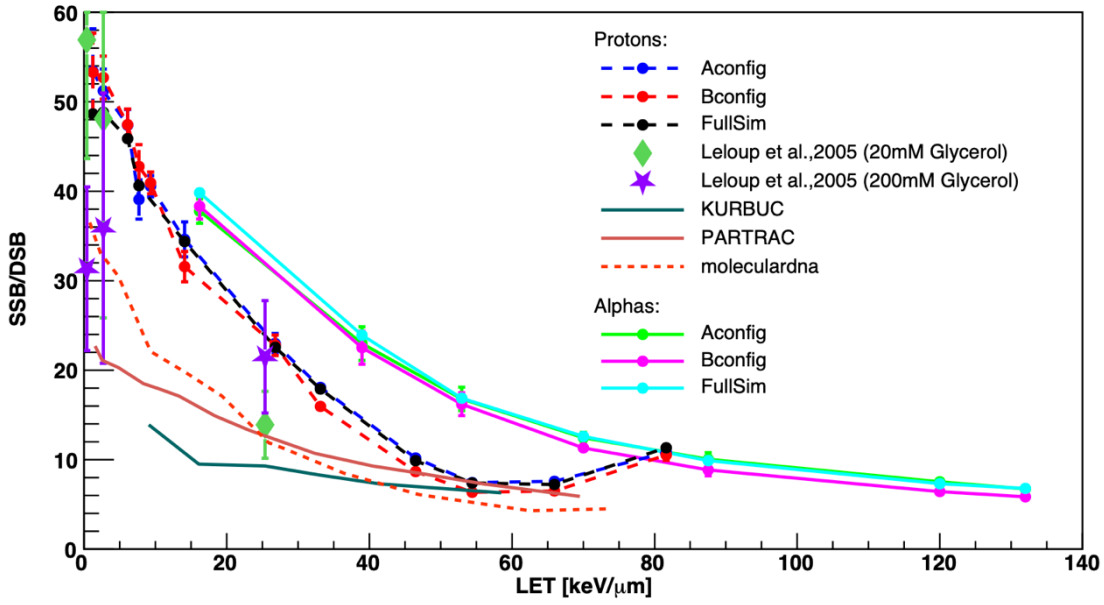
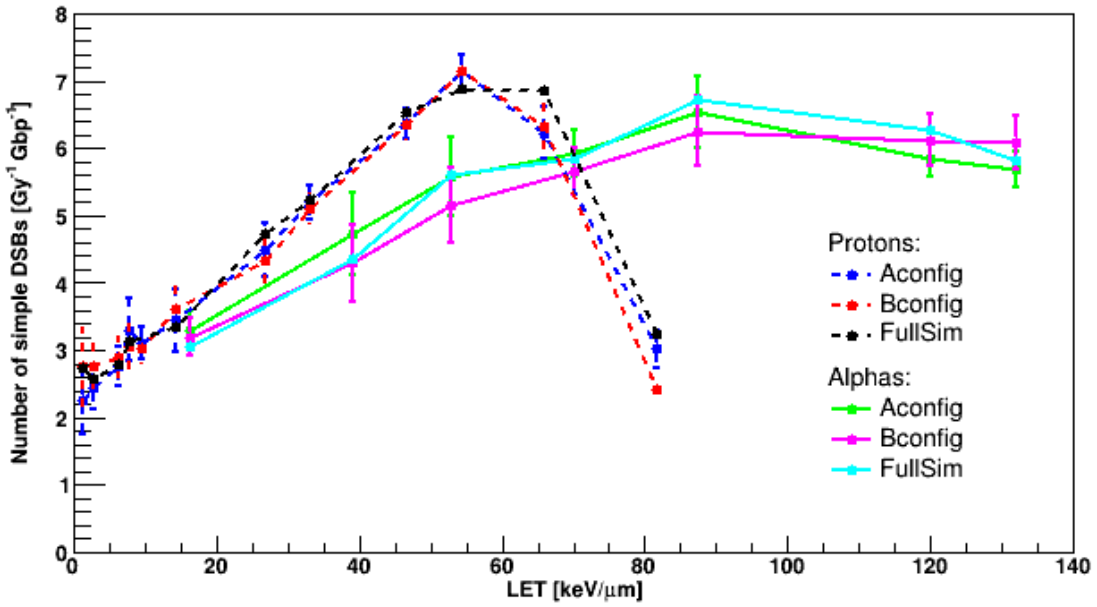
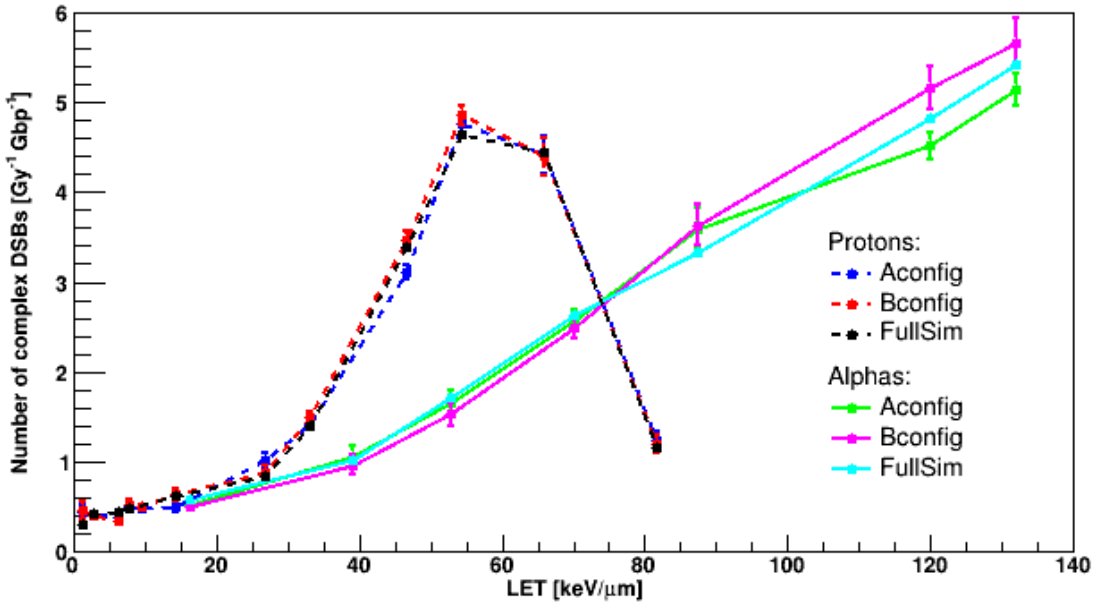


Figure 6: SSB/DSB ratio as a function of LET.

Despite their opposite tendencies with LET, the ratio of SSB over DSB decreases exponentially for all configurations. These simulated ratios for proton were decently close to the experimental data of Leloup et al. [49], even though the experiments were conducted for plasmids instead of eucaryote cells. However, the simulated ratios in this work were higher in comparison with the results simulated by KURBUC [42], PARTRAC [43], and "moleculardna" [20] for the same kind of cell nucleus.



a) Simple double-strand breaks



b) Complex double-strand breaks

Figure 7: The simple DSB (a) and complex DSB (b) yields as a function of LET.

Figure 7 displays two types of DSB resulting from the "Analysis" module, namely, simple and complex DSB $Gy^{-1}Gbp^{-1}$ which are mainly used as input for repair models. A reasonable agreement was observed among Aconfig, Bconfig, and the previous simulation tool for both protons and alphas, as well as for both DSB types. The simple and complex DSBs of protons in this work noticeably increase with LET and reach a peak at ~54.41 keV/µm before going down sharply. A similar increasing tendency was found for complex DSBs of alphas. In the meantime, simple DSBs induced by alphas increase rapidly for LET values below 80 keV/µm, then saturate for LET ranging from 80 keV/µm onwards. The contribution of simple and complex components to the total DSBs varies with the LET values. This dependence is presented in Figure 8. For alphas,

the contribution of simple component decreases steadily with LET values, whilst an increase is observed for complex component. In the case of protons, the same tendency can be seen for both simple and complex components for LET below ~65 keV/µm; however, the tendency is reversed for the last studied point above ~65 keV/µm which corresponds to a kinetic energy of 0.2 MeV.

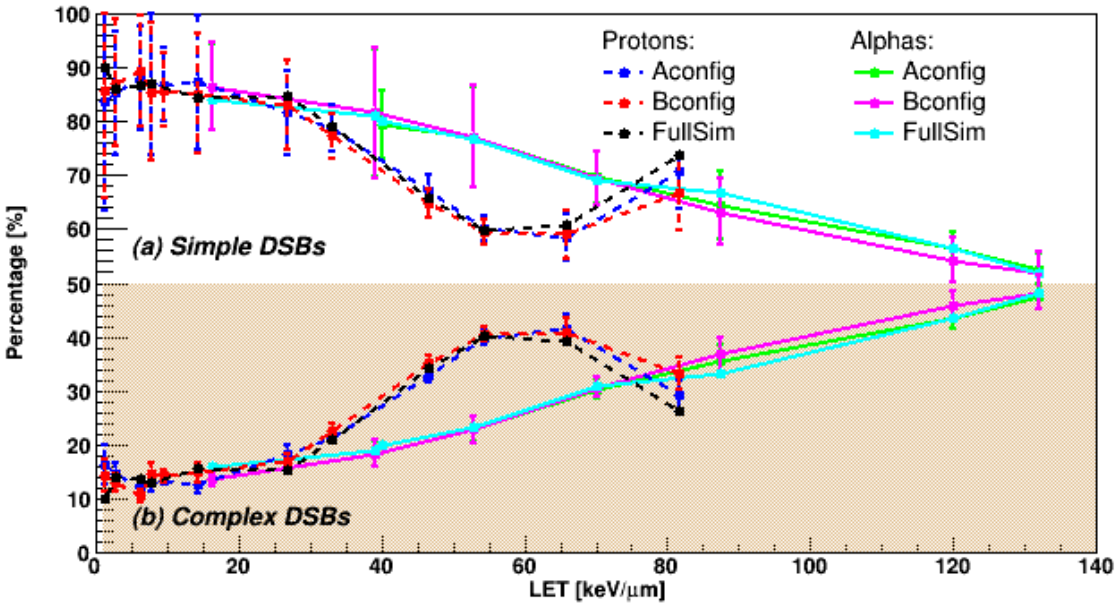


Figure 8: The proportion of simple DSB (a) and complex DSB (b – plotted in the filled area) as a function of LET.

In nature, radiation-induced damages are random processes. This natural variation introduces an uncertainty in the measurement of the interested quantities. In our implementation, the simulated SB yields also obey the random distributions. Figure 9 demonstrates the distribution of DSBs Gy^{-1}• Gbp^{-1} obtained for 10,000 protons of 50 MeV when we repeated the simulation 80 times. The DSB values can be well-described with a Gaussian distribution.

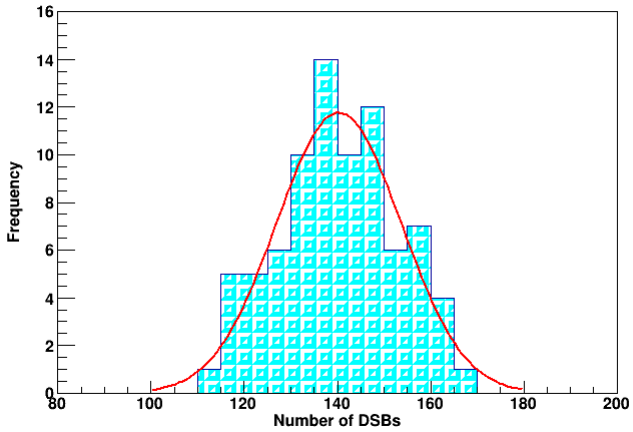
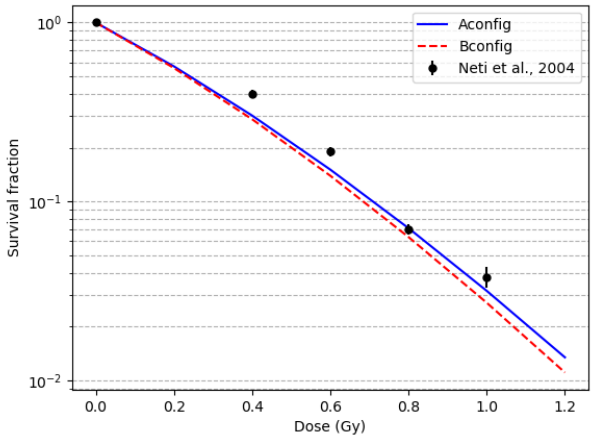


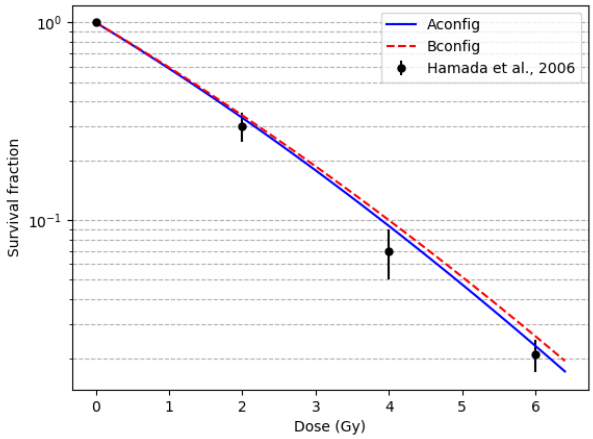
Figure 9: DSB values were obtained when the simulation for 50-MeV proton was repeated 80 times. These simulations were executed with Aconfig for 10,000 primary protons. The red line is the fitting curve using the Gaussian distribution function.

3.2. Repair Models results for fibroblast

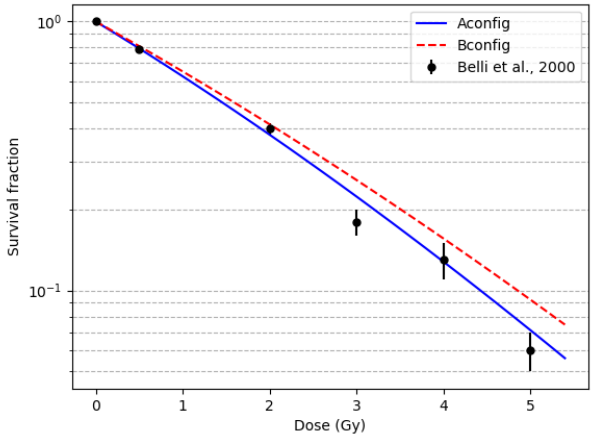
Next, we used the TLK model presented in Subsect. (2.2.3.1) to calculate the survival fraction curve for fibroblast. TLK takes the simulated complex and simple DSB values as input to solve Equations. (3-5). In Figure 10, the survival fraction curves are plotted as a function of the simulated absorbed dose in the cell nucleus. The two upper plots were obtained by simulating the irradiation with alphas at 2.9 MeV (a) and 35 MeV (b), while the lower left plot (c) was done for 4.9 MeV protons. The TLK model which keeps the parameters proposed for fibroblast by Stewart [29] was applied to the experimental data taken from [50, 51, 44] for the human fibroblast cell. It can be seen that there is nearly no difference between the different configurations as well as with FullSim in all cases. A good agreement can be observed between the "dsbandrepair" results and the published data for all beam qualities.



a) Alpha, 132 keV/µm

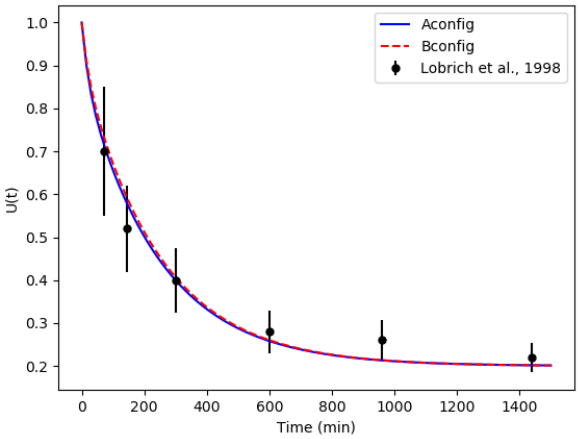


b) Alpha, 16.2 keV/µm

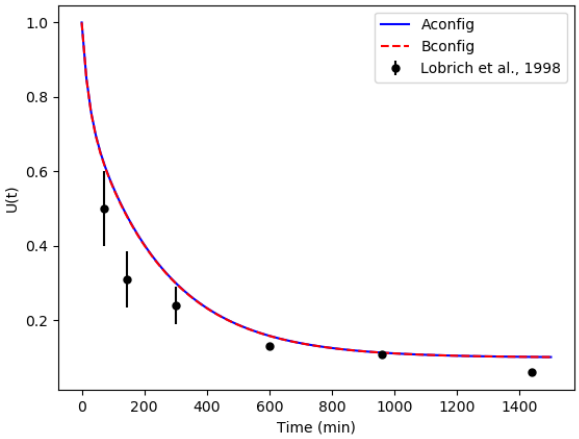


c) Proton, 7.7 keV/µm

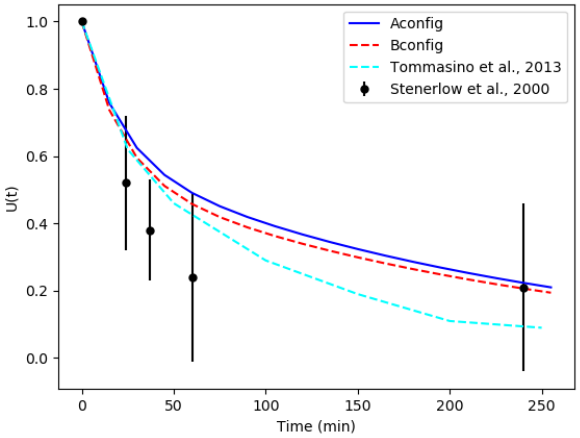
Figure 10: Survival fraction curves for fibroblast computed with TLK model included in the "Analysis" module of "dsbandrepair" compared to published data [50, 51, 44].



a) Alpha, 120 keV/µm



b) Alpha, 70 keV/µm



c) Alpha, 40 keV/µm

Figure 11: Un-rejoined DSB computed with LEM-IV model included in the "Analysis" module of "dsbandrepair" compared to published data [52, 53].

In Figure 11, three plots respectively display the fraction of unrepaired U(t) DSBs for alphas at 120 keV/µm (a), 70 keV/µm (b), and 40 keV/µm (c). Funrej was set to 0.2 for 120 keV/µm alpha, and to 0.1 for 70 keV/µm alpha, whereas Funrej = 0 was used for the last case as proposed by Lobrich [52]. Similar to TLK, an agreement can be observed among Aconfig, Bconfig, and FullSim. For the comparison with experimental measurements, data obtained by [52, 53] for human fibroblasts were utilized. A good agreement was found between the simulation and data obtained by M. Lobrich et al. [52]. Meanwhile, our result is higher than the results calculated by F. Tommasino et al. [54] and the experimental data [53] for 40 keV/µm alpha.

3.3. Results for HUVEC cells

By varying η over the range allowed by the model (from 0.0002 h-1 to 0.01 h-1) and keeping constant other parameters that are not cell specific, we obtained 0.0011 h^{-1} as the optimum value for η to reproduce the experimental survival curve when considering 220 keV irradiation.

The results in terms of cell survival are shown in Figure 12. While it should be noted that the experimental data at 220 kV, for which the parameter η of the TLK model has

been adjusted, is well reproduced, the assumptions made for irradiation at 4 MV make it possible to reproduce the experimental data.

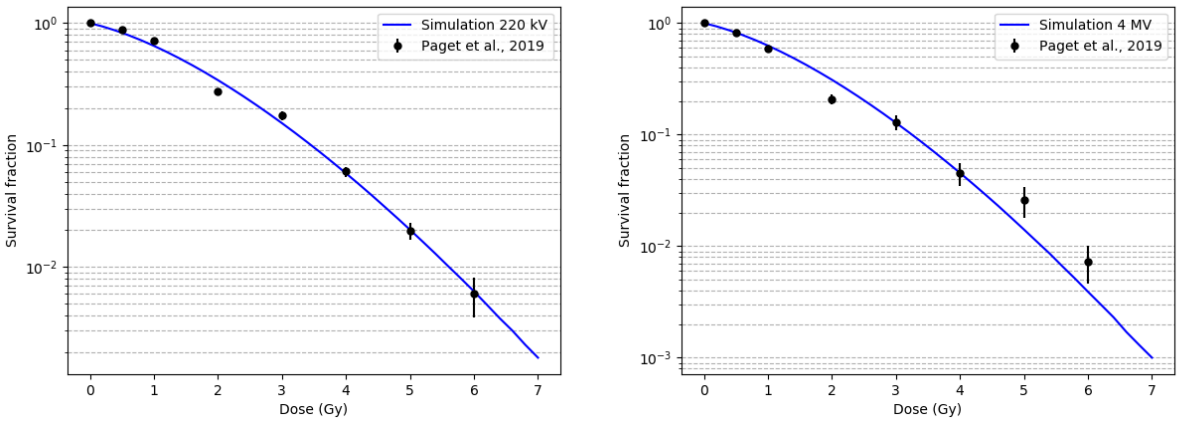


Figure 12: Survival fraction curves for HUVEC computed with TLK model included in the "Analysis" module of "dsbandrepair" compared to published data for X-Rays irradiations [37].

3.4. Simulation time

E (MeV)	LET (keV/µm)	Number of particles	Aconfig			Bconfig		
			T_p (h)	T_c (h)	T (h)	T_p (h)	T_c (h)	T (h)
Alphas								
2.9*	132	700	3.14	28.98	32.11	3.36	90.99	94.36
3.3*	120	800	3.05	27.67	30.73	3.41	83.51	86.92
5.0	87.54	500	2.45	20.15	22.60	2.49	33.24	35.73
6.4*	70	1200	4.21	35.85	40.06	3.70	43.00	46.70
10	52.94	500	1.86	13.38	15.24	1.96	12.15	14.11
12.8*	40	1500	2.98	29.07	32.06	3.82	23.90	27.71
15	38.96	500	1.40	12.22	13.62	1.54	7.10	8.64
35*	15.3	3000	1.58	26.33	27.91	1.64	12.04	13.68
50*	12	5000	3.04	42.55	45.59	3.63	14.17	17.80
Protons								
0.1	81.61	3000	1.20	9.47	10.67	1.18	10.0	11.18
0.2	65.94	1000	1.91	17.27	19.18	1.76	33.43	35.19
0.3	54.41	3000	6.61	77.89	84.50	6.69	154.68	161.37
0.4	46.48	3000	5.58	78.70	84.28	6.27	115.80	122.07
0.7	33.14	3000	4.20	57.60	61.80	4.12	48.90	53.02
1.0*	26.8	1500	1.67	20.49	22.16	1.66	13.09	14.74
2.3*	14.1	2000	1.94	17.42	19.36	1.93	6.70	8.63
4.0	9.33	10000	5.62	78.08	83.70	5.60	24.02	29.62
4.9*	7.7	3000	1.47	17.56	19.03	1.47	5.61	7.07
7.0	6.11	10000	3.56	46.10	49.66	4.38	15.88	20.26
20*	2.7	15000	1.98	31.24	33.22	1.98	7.95	9.93
50	1.24	10000	1.00	11.01	12.01	1.19	3.90	1.00

* : 5 nodes

Table 4. The energies of protons and alphas along with the corresponding LET in water, the number of primary particles and the simulation time. The total simulation time T(h) is defined as the sum of execution time in the physical stage T_p(h) and chemical stage T_c(h). Computation time for the analysis part is not considered here but is relatively the same for the two configurations.

Compared to the previous simulation tool, the current code can run across multi-nodes and be able to use the IRT-syn method to describe the evolution of an initially heterogeneously distributed reaction-diffusion system. Table 4 lists the information on the number of protons and alphas used in our studies as well as the corresponding simulation times. All the considered simulations lasted from a few hours to several days. Aconfig and Bconfig have statistically similar computational time in the physical stage (T_p) since they use the same physics list (also see Figure 14). In contrast, this resemblance is broken in the chemical stage (T_c) as Aconfig and Bconfig use distinct time techniques, i.e., SBS and IRT-syn respectively.

It's interesting to note that in our investigation, IRT-syn does not always have the benefit of providing less computing time compared to SBS. The time ratios of the SBS to IRT-syn approach are depicted on the vertical axis of Figure 13. Here, only T_c was considered for calculating the ratios. If these ratios exceed 1 (the purple dashed line), it signifies that IRT-syn provides a faster simulation time than SBS. We found that IRT-syn completed the simulations more quickly than SBS for LET values below ~50 keV/µm for alpha and below ~33 keV/µm for proton. On the other hand, SBS offered a quicker calculating time for higher LET values of both types of particles.

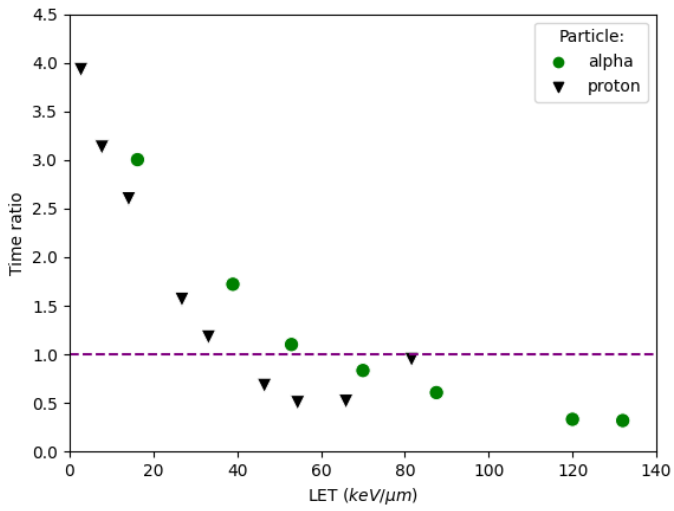


Figure 13: Comparison between SBS (Aconfig) and IRT-syn (Bconfig) in terms of simulation time. The vertical axis is the time ratio of the SBS to IRT-syn method, while the horizontal one expresses the LET value.

Adding more nodes to the simulation is another method for speeding up the computation. Figure 14 illustrates the dependence of computing time on the exploited nodes for 3000 protons at 4.9 MeV. As the number of nodes increased, both T_p and T_c, as well as their sum $T = T_p + T_c$, decreased significantly between 1 node and 8 nodes. It is worth mentioning that simulation time will eventually reach saturation, at which adding

more nodes will be useless. The shape of the dependence might slightly change for each LET value, but the tendency will always hold.

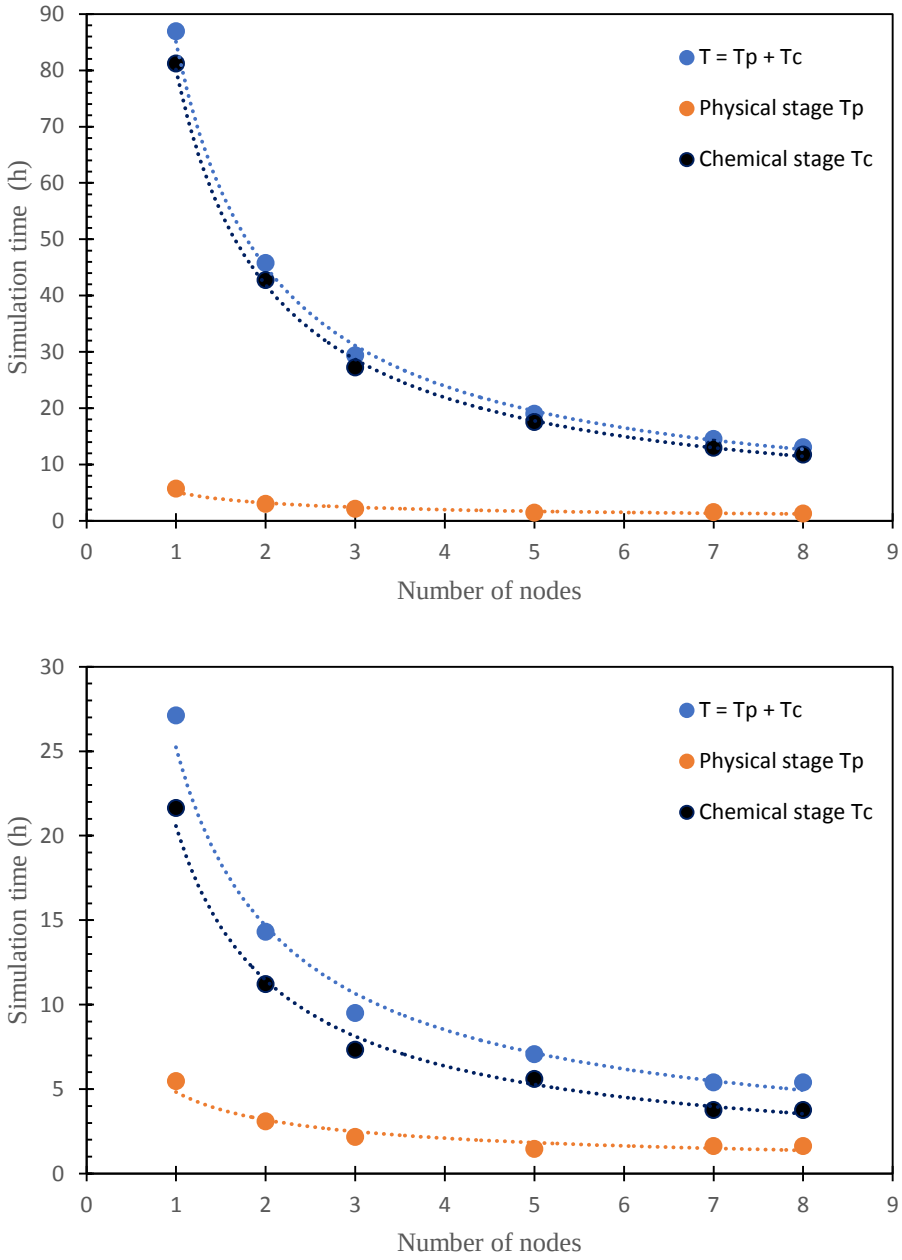


Figure 14: The simulation times, which were determined with Aconfig (upper) and Bconfig (lower) for 3000 protons of 4.9 MeV, as a function of required nodes. The dashed lines are only there for eye-guiding.

4. Discussion

When running with Aconfig, "dsbandrepair" produced yields of total, direct, indirect, single, and double-strand breaks which offer a good agreement with results obtained by the previous " Fullsim" simulation tool (Figure 3-Figure 7) for both protons and alphas. This can be explained by the fact that Aconfig uses the same physics list (G4EmDNAPhysics_option2), chemistry list (G4EmDNAChemistry_option2), and SBS

method as in the previous simulation tool. Therefore, the differences, if exist, can come from the use of different Geant4 versions, namely versions 11.1 and 10.6.

As indirect SBs are generated during the chemical stage, the difference between the indirect SB yields obtained with Bconfig and the other configurations (Figure 3) comes from the different chemistry lists as well as the use of IRT-syn. Indeed, not only the approach is different but also the reaction list that is taken into account. As expected, because the same physics list is used in both configurations, the direct SBs are always similar.

Interestingly, the simulated DSB yields of the three configurations are similar to each other, as illustrated in Figure 5. In addition, they all increase with LET like the direct SB yields. This might imply that most DSBs have at least one SB of an opposite strand that relates to the direct SBs. Further analysis of our simulated results confirms this statement as more than 80 % of DSBs consist of both direct and indirect damages in high LET regions.

Our simulated DSB and direct SB yields of protons reach a peak at ~54.4 keV/µm corresponding to 0.3 MeV (this value can vary when changing the dimensions of the cell). This tendency is also observed from the result obtained with "moleculardna", though the locations of maximum values are slightly different. Generally, the same tendency can be observed for alpha and other ions if the incident energy is low enough. Various factors may contribute to the shape of DSB and direct SB yields Gy^{-1}• Gbp^{-1} depicted in figures 3-7. First, when the energy of primary protons is low enough, the protons stop inside the cell nucleus within a short path-length. Namely, they do not have enough kinematic energy to reach further DNA domains. The path-length distribution of protons and their secondary electrons in a cell nucleus depends on the geometry of the cell nucleus and the incidence of primary particles (parallel or isotropic). This affects the locations of maximum value in DSB and direct curves. Second, the normalization method also affects the shape of DSB and direct SB yields. The normalization is done by dividing the count of DSBs (or other analyzed SB type) by the deposited energy and the total number of base pairs placed in the cell nucleus. When the energy of primary protons is too low, the protons deposit all their energy in a subvolume much smaller than the whole cell nucleus. In this case, the normalization leads to an underestimation. It should be noted that a different normalisation method could lead to a different trend. Normalisation by track length of the primary particle results in a continuous increase in the DSB yield of at least up to 200 keV/µm [34].

Comparison with experimental data is still challenging in terms of DNA damage. As shown in Figure 5 and Figure 6, a large deviation was observed among the experimental datasets. This deviation may come from lots of factors like the difference in cell type and cycle, the uncertainties in the experimental methods to determine damage, or the difference among experimental conditions. Lacking consistency among existing experimental data makes it onerous to deduce a reasonable conclusion for the validity of a simulation code. Nevertheless, in this work, the agreement between Aconfig and the

previous tool is satisfactory to confirm the plausibility of the improvements made for the new dsbandrepair simulation tool.

Additionally, regarding DSB (simple and complex), the decent agreement among Aconfig, Bconfig, and the previous tool implies that the changes implemented for new dsbandrepair make no major difference to the quantification of observables. That means the new version of dsbansrepair gains efficiency while keeping the same validity.

Generally, the contribution of the simple component to DSB changes reversely with the contribution of the complex component. This holds for both proton and alpha (see Figure 7). It's worth mentioning that when the incident energy of the proton is getting low, the number and kinetic energy of secondary electrons created in the cell nucleus will decrease. Consequently, it leads to a decrease in the probability of forming complex DSB. This can explain why the contribution of complex DSB decreases with LET above ~65 keV/µm as shown in Figure 8 for protons.

Another factor that also influences DSB yields is chromatin compaction. Tang et al. [16] used FullSim to study the influence of chromatin compaction on DSBs. Their results show that a cell nucleus filled with only tightly packed chromatin (heterochromatin) produces fewer DSBs than a cell nucleus filled with both heterochromatin (48 %) and euchromatin (52 %). Note that euchromatin is a lightly packed form of chromatin. The same observation was experimentally obtained by H. Takata et al [55] for γ-ray irradiation. In their measurements, the decondensed chromatin produced more DNA damage than condensed chromatin. According to the authors, the compacted chromatin has fewer water molecules per chromatin, thereby generating fewer reactive radicals.

Speaking of simulation time, the chemical stage requires a longer running time than the physical stage even when it uses IRT-syn. Moreover, IRT-syn does not always offer the best performance compared to SBS, especially at high LET values (see Figure 13 and Table 4). The reason for this is the non-optimization of the synchronization algorithm in the IRT-syn model at high LET, where a large number of reactions occur simultaneously within the same time step. However, efforts are underway to investigate and improve IRT-syn at high LET. However, in comparison with the previous simulation tool which usually takes weeks to finish, this work could significantly reduce the calculation time.

For the results on survival fraction for fibroblast cells, the three configurations give quite similar results with the TLK model. This is due to the agreement among those configurations for the number of simple DSBs and complex DSBs, as discussed above. The survival curves obtained are in good agreement with the experimental data for protons and alpha for all beam qualities. The complementary study for photon irradiation shows that our simulation tool is also suitable for low LET irradiation

In our implementation, the LEM-IV model takes the number of isolated DSBs and clustered DSBs in 2 Mbp chromatin loops as input. Although this model is simpler than TLK, it could provide acceptable output of un-rejoined DSBs, as presented in Figure 11 where the results of all the configurations were in good agreement with each other and with experimental data. It should be kept in mind that the agreement with published data is not relevant for all the irradiation setups and configurations. Therefore, users should

investigate the radiobiological data to better understand the irradiation configuration and the corresponding dosimetry before using LEM-IV (or TLK), as well as radiobiological models in general. This holds not only for simulation tool validation but also for other fields that use radiobiology data.

5. Conclusion

Monte Carlo simulations will continue to be a useful and reliable approach for radiobiological research. In this work, we introduced the improvement of a full simulation tool based on the Geant4-DNA toolkit that is compatible with DNAFabric geometries that facilitate the construction of all types of DNA geometries (yeast, bacteria, cell nuclei), as well as allow including both euchromatin and heterochromatin. Not only providing the calculation of many types of DNA damages (SSB, DSB, direct and indirect SB) and their complexity, this new "dsbandrepair" tool also includes the possibility of using the new IRT-syn approach for the chemical stage simulation as well as the extension to calculate latter radio-induced effects by providing published radiobiological models for repair process (TLK, LEM-IV).

The simulation tool introduced in this work has been provided publicly to the user community in the Geant4 11.2.0 release.

Acknowledgments

This work was partially funded by the European Spatial Agency (ESA) as part of the BioRad III project (grant DAR n°4000132935/21/NL/CRS).

References

[1] R. M. Harrison, E. Ainsbury, J. Alves, J. Bottollier-Depois, B. Breustedt, M. C. I. Clairand et al. EURADOS Strategic Research Agenda 2020: Vision For The Dosimetry Of Ionising Radiation. Radiat. Prot .Dosimetry 2021; *194: 42-56*

[2] K. P. Chatzipapas, P. Papadimitroulas, D. Emfietzoglou et al. Ionizing Radiation and Complex DNA Damage: Quantifying the Radiobiological Damage Using Monte Carlo Simulations. Cancers 2021; *12: 799*

[3] W. Friedland, M. Dingfelder, P. Kundrat and P. Jacob. Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC. Mutat. Res. 2011; *711: 28-40*

[4] H. Nikjoo, D. Emfietzoglou, T. Liamsuwan, R. Taleei, D. Liljequist and S. Uehara. Radiation track, DNA damage and response-a review. Rep. Prog. Phys. 2016; *79: 116601*

[5] I. Plante, A. Ponomarev, T. Slaba and M. Hada. RITCARD: Radiation-Induced Tracks, Chromosome Aberrations, Repair and Damage. Radiat. Res. 2019; *192: 282-298*

[6] K. P. Chatzipapas, P. Papadimitroulas, G. Loudos et al. IDDRRA: A novel platform, based on Geant4-DNA to quantify DNA damage by ionizing radiation. Med. Phys. 2021; , *48: 2624-2636*

[7] S. Incerti, G. Baldacchino, M. Bernal, R. Capra, C. Champion, Z. Francis et al. The Geant4-DNA project. Int. J. Model. Simul. Sci. Comput. 2010; *1: 157-158*

[8] M. Karamitros, S. Incerti and C. Champion. THE GEANT4-DNA PROJECT. Radiother. Oncol. 2012; *102: S191-S192*

[9] S. Incerti, I. Kyriakou, M. Bernal, M. Bordage, Z. Francis, S. Guatelli, V. Ivanchenko, M. Karamitros, N. Lampe, S. Lee et al. Geant4-DNA example applications for track structure simulations in liquid water: A report from the Geant4-DNA Project. Med. Phys. 2018; *45: e722-e739*

[10] M. Bernal, M. Bordage, J. Brown, M. Davidkova, E. Delage, Z. El Bitar, S. Enger, Z. Francis, S. Guatelli, V. Ivanchenko et al. Track structure modeling in liquid water: A review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit. Phys. Med. 2015; *31: 861-874*

[11] I. Kyriakou, D. Sakata, H. N. Tran et al. Review of the geant4-DNA simulation toolkit for radiobiological applications at the cellular and DNA level. Cancers (Basel) 2022; *14: 35*

[12] P. L. Olive. The Role of DNA Single- and Double-Strand Breaks in Cell Killing by Ionizing Radiation. Radiat. Res. 1998; *150:S42-S51*

[13] D. Goodhead. Initial Events in the Cellular Effects of Ionizing Radiations: Clustered Damage in DNA. Int. J. Radiat. Biol. 1994; *65: 7-17*

[14] S. Meylan, S. Incerti, M. Karamitros, N. Tang, M. Bueno, I. Clairand and C. Villagrasa. Simulation of early DNA damage after the irradiation of a fibroblast cell nucleus using Geant4-DNA. Sci. Rep. 2017; *7: 11923*

[15] S. Meylan, U. Vimont, S. Incerti, I. Clairand and C. Villagrasa. Geant4-DNA simulations using complex DNA geometries generated by the DnaFabric tool. Computer Physics Communications 2016; *204: 159-169*

[16] N. Tang, M. Bueno, S. Meylan, S. Incerti, H. Tran, A. Vaurijoux, G. Gruel and C. Villagrasa. Influence of chromatin compaction on simulated early radiation-induced DNA damage using Geant4-DNA. Med. Phys. 2019; *46: 1501–1511*

[17] Y. Thibaut, N. Tang, H. N. Tran, A. Vaurelie, C. Villagrasa and Y. Perrot. Nanodosimetric Calculations of Radiation-Induced DNA Damage in a New Nucleus Geometrical Model Based on the Isochore Theory. Int. J. Mol. Sci. 2022; *23: 3770*

[18] N. Lampe. The Long Term Impact of Ionising Radiation on Living Systems. Ph.D. Thesis, Université Clermont Auvergne, Clermont-Ferrand, France, 2017.

[19] W. Shin, D. Sakata, N. Lampe, O. Belov, N. H. Tran et al. A Geant4-DNA Evaluation of Radiation-Induced DNA Damage on a Human Fibroblast. Cancer 2021; *13: 4940*

[20] K. Chatzipapas, N. H. Tran, M. Dordevic, S. Zivkovic et al. Simulation of DNA damage using Geant4-DNA: an overview of the "molecularDNA" example application. Prec. Radiat. Oncol. 2023; *7: 4-14*

[21] T. M. Forum. MPI: a message passing interface. Proceedings of the 1993 ACM/IEEE conference on Supercomputing, *Portland, Oregon, USA, 1993.*

[22] H. Tran, J. Ramos-Mendez, W-G. Shin, Y. Perrot et al. Assessment of DNA damage with an adapted

independent reaction time approach implemented in Geant4-DNA for the simulation of diffusion-controlled reactions between radio-induced reactive species and a chromatin fiber. Med. Phys. 2021; *48: 890-901*

[23] R. D. Stewart. Two-Lesion Kinetic Model of Double-Strand Break Rejoining and Cell Killing. Radiat. Res. 2001; *156: 365-378*

[24] T. Elsasser, T. F. W. K. Weyrather, M. Durante et al. Quantification of the relative biological effectiveness for ion beam radiotherapy: direct experimental comparison of proton and carbon ion beams and a novel approach for treatment planning. Int. J. Radiat. Oncol. Biol. Phys. 2010; *78: 1177-1183*

[25] H. Nikjoo, N. P. O, D. Goodhead and M. Terrissol. Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events. Int. J. Radiat. Biol. 1997; *71: 467-483*

[26] M. Dizdaroglu and P. Jaruga. Mechanisms of free radical-induced damage to DNA,» Free Rad. Res. 2012; *46: 382-419*

[27] M. C. Elia. and M. O. Bradley. Influence of chromatin structure on the induction of DNA double strand breaks by ionizing radiation. Cancer Res. 1992; *52: 1580-1586*

[28] N. Tang, M. Bueno, S. Meylan, Y. Perrot, H. N. Tran et al. Assessment of Radio-Induced Damage in Endothelial Cells Irradiated with 40 kVp, 220 kVp, and 4 MV X-rays by Means of Micro and Nanodosimetric Calculations. Int. J. Mol. Sci. 2019; *2: 6204*

[29] M. Karamitros, S. Luan, M. A. Bernal et al. Diffusion-controlled reactions modeling in Geant4-DNA. J. Comput. Phys. 2014; *274: 841-882*

[30] A. Dotti, M. Asai, G. B. I. Hrivnacova and K. Murakami. Extending Geant4 Parallelism with external libraries (MPI, TBB) and its use on HPC resources. IEEE Nucl. Scie. Sym. Med. Ima. Conf., *2015.*

[31] J. Ramos-Mendez, W-G.. Shin, M. Karamitros et al. Independent reaction times method in Geant4-DNA: Implementation and performance. Med. Phys. 2020; *47: 5919-5930*

[32] R. Brun and F. Rademarkers. ROOT — An object oriented data analysis framework,» Nucl. Instrum. Methods. Phys. Res. A 1997; *389: 81-86*

[33] J. Schuemann, A. McNamara, J. Warmenhoven et. al. A new standard DNA damage (SDD) data format. Radiat. Res. 2019; *191: 76-92*

[34] N. Tang. Ph.D. Thesis, Université de Bordeaux; Fance, 2019. Available: https://www.theses.fr/2019BORD0160.

[35] D. Sakata, M. Suzuki, R. Hirayama, Y. Abe et al. Performance Evaluation for Repair of HSGc-C5 Carcinoma Cell Using Geant4-DNA. Cancers 2021; *13: 6046*

[36] D. Sakata, R. Hirayama, W. G. Shin, M. Belli et al. Prediction of DNA rejoining kinetics and cell survival after proton irradiation for V79 cells using Geant4-DNA. Phys. Med. 2023; *106: 102508*

[37] V. Paget, M. B. Kacem, M. D. Santos et al. Multiparametric radiobiological assays show that variation of X-ray energy strongly impacts relative biological effectiveness: comparison between 220 kV and 4 MV. Scientific Reports 2019; *9: 14329*

[38] W. Friedland, P. Jacob and P. Kundrat. Mechanistic simulation of radiation damage to DNA and its repair: on the track towards systems radiation biology modelling. Radiat. Prot. Dosimetry 2011; *143: 542-548*

[39] T. Friedrich, U. Scholz, T. Elsässer, M. Durante and M. Scholz. Calculation of the biological effects of ion beams based on the microscopic spatial damage distribution pattern. Int. J. Radiat. Biol. 2012; *88: 103-107*

[40] S. Nurk, S. Koren, A. Rhie et al. The complete sequence of a human genome. Science 2022; *376: 44-53*

[41] M. Berger, J. Coursey, M.A. Zucker and J. Chang. ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions (version 1.2.3). Available: http://physics.nist.gov/Star.

[42] Y. Matsuya, T. Kai, A. Parisi, Y. Yoshii and T. Sato. Application of a simple DNA damage model developed for electrons to proton irradiation. Phys. Med. Biol. 2022; *67: 215017*

[43] W. Friedland, P. Jacob, P. Bernhardt, H. G. Paretzke and M. Dingfelder. Simulation of DNA damage after proton irradiation. Radiat. Res. 2003; *159: 401-410*

[44] M. Belli, D. Bettega, P. Calzolari, F. Cera, R. Cherubini, M. D. Vecchia, M. Durante, S. Favaretto, G. Gialanella and G. Gross. Inactivation of human normal and tumour cells irradiated with low energy protons. Int. J. Radiat. Biol. 2000; *76: 831-839*

[45] A. Campa, R. Ballarini, M. Belli et al. DNA DSB induced in human cells by charged particles and gamma rays: Experimental results and theoretical approaches. Int. J. Radiat. Biol. 2005; *81: 841-854*

[46] D. Frankenberg, H. Brede, U. J. Schrewe et al. Induction of DNA double-strand breaks by 1H and 4He ions in primary human skin fibroblasts in the LET range of 8 to 124 keV/um. Radiat. Res. 1999; *151: 540-549*

[47] A. Ristic-Fira, O. Keta, Petkovic V., Dordevic M. et al. Experimental validation of helium ions as a function of linear energy transfer in radioresistant human malignant cells. Int. J. Radiat. Biol. 2023; *Submitted for publication, 2023*

[48] K. Chatzipapas, M. Dordevic, S. Zivkovic, N.H. Tran et al. Geant4-DNA simulation of human cancer cells irradiation with helium ion. Phys. Med. 2023; *112: 102613*

[49] C. Leloup, G. Garty, G. Assaf, A. Cristovão, A. Breskin, R. Chechik et al. Evaluation of lesion clustering in irradiated plasmid DNA. Int. J. Radiat. Biol. 2005; *81: 41-54*

[50] P. S. V. Neti, S. M. d. Toledo, V. Perumal, E. I. Azzam and R. W. Howell A Multi-port Low-Fluence Alpha-Particle Irradiator: Fabrication, Testing and Benchmark Radiobiological Studies. Radiat. Res. 2004; *161: 732-738*

[51] N. Hamada, T. Funayama, S. Wada, T. Sakashita, M. N. T. Kakizaki and Y. Kobayashi. LET-Dependent Survival of Irradiated Normal Human Fibroblasts and Their Descendents. Radiat. Res. 2006; *166: 24-30*

[52] M. Löbrich, P. K. Cooper and B. Rydberg. Joining of correct and incorrect DNA ends at double-strand breaks produced by high-linear energy transfer radiation in human fibroblasts. Radiat. Res. 1998; *150: 619-626*

[53] B. Stenerlow, E. Hoglund, J. Carlsson and E. Blomquist. Rejoining of DNA fragments produced by radiations of different linear energy transfer. Int. J. Radiat. Biol. 2000; *76: 549-557*

[54] F. Tommasino, T. Friedrich, U. Schol and G. Taucher-Scholz. A DNA Double-Strand Break Kinetic Rejoining Model. Radiat. Res. 2013; *180: 524-538*

[55] H. Takata, M. T. Hanafusa et al. Chromatin Compaction Protects Genomic DNA from Radiation Damage. PLoS ONE 2013; *8: e75622*